

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Novel Computational Methodologies for Detailed Analysis of Human Motion from Image Sequences

João Ferreira de Carvalho Castro Nunes



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Programa Doutoral em Engenharia Informática

Supervisor: Prof. João Manuel Ribeiro da Silva Tavares

Second Supervisor: Prof. Pedro Miguel do Vale Moreira

December, 2025

Novel Computational Methodologies for Detailed Analysis of Human Motion from Image Sequences

João Ferreira de Carvalho Castro Nunes

Programa Doutoral em Engenharia Informática

Approved in oral examination by the committee:

President: Prof. Pedro Nuno Ferreira da Rosa da Cruz Diniz, Full Professor, Department of Informatics Engineering, Faculty of Engineering, University of Porto

Referee: Prof. Carlos Miguel Fernandes Quental, Assistant Professor, Department of Mechanical Engineering, Instituto Superior Técnico, University of Lisbon

Referee: Prof. Hugo Pedro Martins Carriço Proença, Full Professor, Department of Computer Science, University of Beira Interior

Referee: Prof. Luís Paulo Gonçalves dos Reis, Associate Professor with Aggregation in the Department of Informatics Engineering at the Faculty of Engineering, University of Porto

Referee: Prof. João Manuel Ribeiro da Silva Tavares, Full Professor, Department of Mechanical Engineering, Faculty of Engineering, University of Porto

December, 2025

Resumo

A análise da marcha humana fornece informações essenciais sobre a função biomecânica, a avaliação clínica e o reconhecimento biométrico. No entanto, alcançar uma compreensão do movimento que seja precisa e reproduzível em condições reais continua a ser um desafio significativo. As técnicas tradicionais de captura de movimento dependem de infraestruturas dispendiosas e de ambientes controlados, o que limita a sua escalabilidade e validade em contextos reais. Esta tese procura ultrapassar essas limitações através do desenvolvimento de metodologias computacionais que exploram simultaneamente informação RGB e de profundidade, de modo a permitir uma análise da marcha robusta, eficiente e totalmente automática, recorrendo a sensores de baixo custo.

A investigação seguiu uma trajetória estruturada que integra a criação de um conjunto de dados, o desenho de novas representações e a inovação metodológica. Em primeiro lugar, foi realizada uma revisão e análise comparativa abrangente dos conjuntos de dados existentes baseados em visão e profundidade, identificando lacunas quanto à diversidade de modalidades, qualidade das anotações e acessibilidade. Para colmatar estas limitações, foi concebido, adquirido e disponibilizado publicamente o Gait Recognition Image and Depth Dataset (GRIDDS). O GRIDDS fornece dados sincronizados de RGB, profundidade, silhueta e esqueleto 3D de 35 participantes registados em condições controladas, constituindo um dos primeiros recursos multimodais normalizados para análise e reconhecimento da marcha.

Com base neste fundamento, foram propostas duas novas representações computacionais da marcha que combinam informação bidimensional de aparência com estrutura esquelética tridimensional, aumentando a robustez face a variações de ponto de vista, vestuário e objetos transportados. Estas variantes da Gait Skeleton Image (GSI), baseadas em articulações e em segmentos lineares, foram integradas em arquiteturas de aprendizagem profunda e avaliadas extensivamente, demonstrando um desempenho competitivo, e algumas circunstâncias, por vezes superior, em comparação com métodos baseados em aparência, através de múltiplos conjuntos de dados e condições de variabilidade.

Por fim, foram introduzidos novos métodos de interpolação de silhuetas de marcha, que combinam raciocínio geométrico determinístico (BRIEF) com aprendizagem profunda bidirecional (BiSINet), permitindo reconstruir *frames* em falta e melhorar a coerência temporal. As técnicas de interpolação propostas evidenciaram melhorias significativas na precisão do reconhecimento e forte capacidade de generalização entre diferentes conjuntos de dados e taxas de amostragem.

Em conjunto, as contribuições deste trabalho, que abrangem a aquisição multimodal de dados, o desenvolvimento de representações robustas da marcha e a reconstrução temporal, avançam as fronteiras científicas e tecnológicas da análise da marcha humana, promovendo a reproduzibilidade, acessibilidade e aplicabilidade tanto nos domínios clínico como na visão por computador.

Abstract

Human gait analysis provides critical information on biomechanical function, clinical assessment, and biometric recognition, but achieving accurate and reproducible motion understanding under real-world variability remains a major challenge. Traditional motion capture techniques are dependent on expensive infrastructure and controlled environments, which limit scalability and real-world validity. This thesis addresses these limitations by developing computational methodologies that exploit both RGB and depth information to enable robust, efficient, and fully automatic gait analysis using consumer-grade sensors.

The research followed a structured trajectory that encompasses dataset creation, representation design, and methodological innovation. First, an extensive review and comparative analysis of existing vision- and depth-based gait datasets identified gaps in modality diversity, annotation quality, and accessibility. To address these issues, the Gait Recognition Image and Depth Dataset (GRIDDS) was designed, acquired, and publicly released. GRIDDS provides synchronized RGB, depth, silhouette, and 3D skeletal data from 35 participants recorded under controlled conditions, establishing one of the first standardized multi-modal benchmarks for gait analysis and recognition.

Building on this foundation, two novel computational gait representations were introduced that fuse two-dimensional appearance cues with three-dimensional skeletal structure to increase robustness to viewpoint, clothing, and carried-object variations. These Gait Skeleton Image (GSI) variants (joint-based and line-based) were integrated within deep learning frameworks and evaluated through extensive experiments, demonstrating competitive and, under certain circumstances, superior performance compared with established appearance-based methods across multiple datasets and covariate conditions.

Finally, new methods for gait silhouette interpolation were introduced, combining deterministic geometric reasoning (BRIEF) and bidirectional deep learning (BiSINet) to reconstruct missing frames and enhance temporal coherence. The proposed interpolation techniques significantly improved downstream recognition accuracy and demonstrated strong generalization across datasets and frame-rate conditions.

Collectively, the contributions of this work, which span multi-modal data acquisition, robust gait representation learning, and temporal reconstruction, advance the scientific and technological frontiers of human gait analysis, promoting reproducibility, accessibility, and applicability in both clinical and computer vision domains.

Agradecimentos

A realização desta tese representa o culminar de um percurso longo, exigente e profundamente enriquecedor. Nenhum caminho desta natureza se percorre sozinho, e é com enorme gratidão que deixo aqui o meu reconhecimento a todos os que, de diferentes formas, contribuíram para que este trabalho se tornasse possível.

Ao meu orientador, Prof. João Tavares, e coorientador, Prof. Pedro Moreira, expresso a mais sincera gratidão pela constante disponibilidade, pela orientação firme e esclarecida, e pelo positivismo e incentivo que nunca faltaram. A vossa confiança e exigência científica foram fundamentais para o crescimento pessoal e académico que esta etapa me proporcionou.

Aos vários colegas do IPVC, que se tornaram muito mais do que companheiros de trabalho, deixo um agradecimento especial pela amizade, pela ajuda e pelas conversas que tantas vezes serviram de pausa, de motivação e de inspiração. A vossa presença foi essencial para manter o equilíbrio e a energia ao longo deste processo.

Aos meus amigos de longa data, que sempre estiveram por perto, mesmo quando o tempo escasseava, agradeço por continuarem a acompanhar-me, pelo incentivo constante e pela compreensão nos momentos de ausência.

À minha família — e em especial ao meu pai e aos meus irmãos — deixo um agradecimento profundo pela força silenciosa, pelo apoio incondicional e por me lembrarem sempre da importância das coisas simples. À minha mãe, cuja presença permanecerá em tudo o que faço, agradeço por ter sido e continuar a ser uma fonte de amor. Foram e continuam a ser uma âncora segura em todas as fases deste percurso.

Por fim, mas nunca em último lugar, à Célia — por ser a melhor pessoa que alguém pode desejar ter ao seu lado. Pela força que transmite, pelas palavras certas nas horas difíceis, pelos abraços que apaziguam e pela serenidade que tantas vezes me devolveu o rumo. À Margarida e ao Pedro, pela paciência infinita e por compreenderem as horas em que o pai ficou agarrado ao computador. Esta conquista é também vossa.

A todos, o meu mais sincero **obrigado**.

João Ferreira Nunes

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Motivation	2
1.3	Research Goals	3
1.4	Thesis Statement	4
1.5	Publications	4
1.6	Structure of the Thesis	6
2	Human Motion Analysis – Fundamentals, Applications, and Challenges	9
2.1	A Historical Perspective on Human Motion Analysis	10
2.2	Foundations of Human Motion Analysis	16
2.3	Application Areas of Human Motion	20
2.3.1	Healthcare and Rehabilitation	20
2.3.2	Sports Performance and Injury Prevention	22
2.3.3	Surveillance and Security Applications	25
2.3.4	Human Factors and Ergonomics	27
2.3.5	Robotics	29
2.3.6	Entertainment and Gaming	31
2.3.7	Traffic Safety	33
2.4	Automated Tools for Human Motion Analysis	34
2.4.1	Selected Tools	35
2.4.2	A Classification Framework	42
2.4.3	Main conclusions	47
2.5	Summary	48
3	Human Gait – A Comprehensive Overview	51
3.1	Human Gait Analysis	52
3.2	Biomechanics and Terminology of Human Locomotion	53
3.2.1	Gait Cycle and Gait Phases of Human Locomotion	54
3.2.2	Planes of Motion	56
3.3	Parameters of Gait	57
3.3.1	Anthropometric Parameters	57
3.3.2	Spatio-temporal Parameters	58
3.3.3	Kinematic Parameters	59
3.3.4	Kinetic Parameters	59
3.3.5	Electromyography Parameters	60
3.3.6	Other Parameters	61
3.4	Normative Values of Gait Parameters	62

3.4.1	Normative Values for Spatio-Temporal Parameters	63
3.4.2	Normative Values for Kinematic Parameters	64
3.5	Sensing Technologies in Gait Analysis	66
3.5.1	Wearable Technologies	66
3.5.2	Non-Wearable Technologies	68
3.6	Summary	72
4	Datasets for Gait Analysis – GRIDDS and Other Resources	73
4.1	Scope and Contributions	73
4.2	GRIDDS – A Gait Recognition Image and Depth Dataset	74
4.2.1	Motivation and Dataset Goals	74
4.2.2	Acquisition Environment and Equipment	75
4.2.3	Participants and Recording Protocol	76
4.2.4	Data Modalities, Pre-processing and Annotation	78
4.2.5	Feature Extraction and Descriptors	83
4.2.6	Evaluation Protocols and Official Splits	88
4.2.7	Data Specification, Licensing, and Availability	89
4.2.8	Evidence of Scientific Use of the GRIDDS Dataset	92
4.3	A New Taxonomy for Vision and Depth-Based Gait Datasets	92
4.3.1	Motivation and Objectives	92
4.3.2	Taxonomy Design Principles	94
4.3.3	Common Covariates in Vision- and Depth-Based Gait Datasets	95
4.3.4	Classification Dimensions	98
4.4	A Review of Vision and Depth-Based Human Gait Datasets	100
4.4.1	Motivation and Scope of the Review	100
4.4.2	Dataset Selection	101
4.4.3	Dataset Profiles	104
4.4.4	Taxonomy-Aligned Dataset Tables	136
4.4.5	Comparative Analysis and Selection Guidance	153
4.5	Summary	155
5	Novel Gait Silhouette Image Interpolation Techniques	157
5.1	Motivation and Scope	157
5.2	Related Work	159
5.3	Proposed Methods	161
5.3.1	BRIEF – Bidirectional Rowwise Interpolation with Envelope Fusion	162
5.3.2	BiSINet - Bidirectional Silhouette Interpolation Network	172
5.4	Experimental Setup	176
5.4.1	Datasets	177
5.4.2	Tasks and Protocols	178
5.4.3	Evaluation Metrics	180
5.5	Results and Analysis	182
5.5.1	Experiments Results	183
5.5.2	Interpolation Methods Comparison	185
5.5.3	Impact of Frame-Rate Manipulation on Recognition	186
5.6	Conclusions and Limitations	188
5.6.1	Conclusions	188
5.6.2	Limitations	189
5.7	Summary	190

6	Gait Skeleton Image - Application to Gait-based Recognition	193
6.1	Introduction	193
6.1.1	Motivation for Applying Deep Learning to Gait	194
6.1.2	Overview of Traditional Gait Analysis and Deep Learning-Based Approaches	195
6.1.3	Key Challenges in Gait-Based Biometric Recognition	195
6.2	Gait Data Modalities and Input Representations	197
6.2.1	Visual Modalities	197
6.2.2	Sensor-Based Modalities	201
6.2.3	Hybrid and Multi-modal Approaches	202
6.3	Computational Gait Representations as Feature Learning	203
6.3.1	Evolution from Handcrafted to Learned Features	203
6.3.2	Established Computational Gait Representations	204
6.3.3	Comparative Analysis	212
6.4	Deep Learning Architectural Families for Gait	214
6.4.1	Convolutional Neural Networks (CNNs)	214
6.4.2	Recurrent Neural Networks (RNNs) and LSTMs	214
6.4.3	Graph Neural Networks (GNNs)	215
6.4.4	Transformers and Attention Mechanisms	216
6.4.5	Autoencoders and Generative Models	217
6.5	Novel Computational Gait Representation: Gait Skeleton Image	218
6.5.1	GSI Body Structure Definition	218
6.5.2	Computational Construction Pipeline	224
6.5.3	GSI Variants: Lines vs. Joints Representation	231
6.5.4	Temporal Aggregation across Gait Cycle	233
6.5.5	Parameter Optimization Study	234
6.5.6	GSI in Context: Related Skeleton-Based Representations	235
6.5.7	Evaluation on CASIA-B Dataset	238
6.6	Summary	246
7	Conclusions and Future Work	247
7.1	Main Conclusions and Contributions	247
7.2	Limitations	248
7.3	Future Work and Research Directions	249
7.4	Final Remarks	249
	References	251
A	GRIDDS Joint-angle Trajectories	305

List of Figures

2.1	Illustrations extracted from the sketchbooks of da Vinci. (a) Anatomical studies of the human skeleton, showcasing intricate details and structural complexities. (b) Further anatomical examination of the human skeleton. (c) Study focusing on the center of gravity and balance, featuring a depiction of a man climbing a step. Credit: <i>leonardodavinci.net</i>	11
2.2	Illustrations extracted from Giovanni Alfonso Borelli’s sketchbooks. (a) Illustration presenting a series of dynamic movements and muscle contractions. (b) Studies exploring the biomechanics of a man carrying a load on his back. Credit: <i>Wellcome Collection</i>	11
2.3	Illustrations showcasing some of Marey’s motion studies. (a) Illustration of the chronophotographic gun in use, from Le Fusil photographique , published in <i>La Nature</i> No. 464 on April 22, 1882. (b) An experiment capturing a person in motion, where the subject get in a fitted black suit with reflective markers on their joints to better track the movements of different body parts in each phase. Credit: <i>The Linda Hall Library of Science</i>	12
2.4	Some of Muybridge’s motion studies: (a) The Horse in Motion study, consisting in recording the motion of a galloping horse; (b) A human movement study during a javelin throw session. Credit: <i>Wellcome Collection</i>	13
2.5	Examples of technologies and models used in human motion analysis: (a) depth map generated by a depth sensor; (b) 2D skeletal tracking overlaid on an RGB image of a human subject; (c) various biomechanical models representing the human body for motion analysis.	15
2.6	A generic framework for human motion analysis, encompassing stages of data acquisition, pre-processing, segmentation, feature extraction, modeling, analysis, and visualization.	16
2.7	Instances of two healthcare applications: (a) A framework designed for automatic fall detection from a single image, demonstrating invariance to appearance characteristics (<i>Asif et al. (2020)</i>). (b) A snapshot of the five mini-games in a serious games rehabilitation framework proposed by <i>Assad et al. (2011)</i> . where players need to trigger buttons through hand movements.	21
2.8	Example data from two healthcare and rehabilitation datasets: (a) UR Fall Detection: sample frames showing a fall event captured using RGB and depth cameras (<i>Kwolek and Kepski (2014)</i>). (b) U-Limb: representation of multimodal data acquisition during upper limb movement task, including kinematic measurements and physiological signals (<i>Averta et al. (2021)</i>).	23

2.9	Two examples of applicability of human motion in team and individual sports: (a) Soccer team ball passing network, with nodes as players and arrows indicating frequency and direction of successful passes (Cao (2023)); (b) Golf swing analysis tool comparing beginner and expert poses, highlighting key differences to support skill improvement (Liao et al. (2022a)).	24
2.10	Two sample images from sports-related datasets: (a) a frame from a soccer match in the Wyscout Event and Match dataset, integrated into a tool for automatic event detection (Pappalardo et al. (2019)); (b) Sample sequence from the Fine-Gym dataset illustrating a Balance Beam routine (Shao et al. (2020a)).	25
2.11	Examples of video-based surveillance applications for anomaly detection: (a) detection of abnormal behaviors (highlighted in white) using the UCSD Ped1 and UCSD Ped2 datasets (Li et al. (2014); Chen et al. (2017)); (b) local abnormal event detection proposed by Chen et al. (2017), illustrating detected anomalies highlighted in red (top row) along with corresponding detection intensity peaks (bottom row).	26
2.12	Example frames from datasets used in human motion analysis for surveillance applications. (a) Sample clip from the ShanghaiTech Campus Dataset (Liu et al. (2018a)) showing normal and abnormal activities captured under different conditions. (b) Visualization of pedestrians on the Train Station Dataset (Zhou et al. (2012)), recorded at New York's Grand Central Station.	27
2.13	Examples of motion analysis techniques used in ergonomic assessments: (a) a laboratory setting where a worker's movements are monitored using two motion analysis software applications (Petrosyan et al. (2020)); (b) a sequence showing the reconstruction of a 3D skeleton from video input: 2D joint positions are estimated from one camera view and combined with depth data from a second view to create a 3D skeleton model (Han and Lee (2013)).	28
2.14	Examples from two ergonomics-related motion datasets. (a) Two frames from the AnDyDataset (Maurice et al. (2019a)) showing a participant performing different manual tasks: using a screwdriver and untying a knot. (b) Setup from the Manual Material Handling Dataset (Bassani et al. (2021a)): (a) wearable inertial system for kinematic data collection; (b) sEMG device for muscle activity recording; (c) objects handled during the experiments.	29
2.15	Example of a drawer-opening task: on the left, the action is performed by a human; on the right, the same action is executed by a robot through imitation. Source: Bahl et al. (2022)	30
2.16	Environment configuration in the THÖR dataset (Rudenko et al. (2020a)). Participants, equipped with tracking helmets, and a mobile robot navigate toward their respective goals in a shared indoor space. All motion is recorded using a motion capture system, with an example trajectory visualization shown in the bottom left corner.	31
2.17	Two illustrative examples of datasets with potential applications in entertainment and gaming: (a) examples of synchronized color, depth, and skeleton data from a boxing game in the G3Di dataset (Bloom et al. (2015)); (b) visual examples of the 18 gesture classes included in the HaGRID dataset (Kapitanov et al. (2024)). . . .	33

2.18	Illustrative examples from two datasets used for traffic safety: (a) an image from the EuroCity Persons dataset where pedestrians are annotated with tight green bounding boxes, ensuring precise localization in urban traffic scenes (Braun et al. (2019a)); (b) sample images from the I2V-MVPD dataset, showing pedestrian views from the vehicle perspective (left) and from the infrastructure-mounted camera perspective (right) (Braun et al. (2019a));	34
2.19	Illustrative examples of human motion analysis tools.	42
3.1	Illustration of the human gait cycle. The figure depicts the ordered phases and key events comprising a full cycle, starting at right heel strike and ending at the subsequent right heel strike. The right limb is colored blue and the left limb green. Image adapted from D'Août et al. (2022); Neumann (2017).	54
3.2	The three cardinal planes of the body: sagittal, frontal, and transverse (horizontal) are illustrated with a person in the anatomical position. Source: Neumann (2017).	57
3.3	Normative range for the hip flexion/extension angle over the gait cycle, showing the mean trajectory and the variability band (shaded area). Adapted from Kirtley (2006).	64
3.4	Normative range for the knee flexion angle over the gait cycle, showing the mean trajectory and the variability band (shaded area). Adapted from Kirtley (2006).	65
3.5	Normative range for the ankle dorsi-/plantar-flexion angle over the gait cycle, showing the mean trajectory and the variability band (shaded area). Adapted from Kirtley (2006).	66
3.6	Wearable sensing technologies for gait analysis	68
3.7	Non-wearable sensing technologies for gait analysis	71
4.1	GRIDDS capture environment and sensor placement	75
4.2	Microsoft Kinect v2 RGB-D sensor	76
4.3	Sex distribution in GRIDDS (F=24, M=11)	77
4.4	Age distribution and age by sex in GRIDDS	78
4.5	Height distribution and height by sex in GRIDDS	78
4.6	Representative frames from Kinect v2 streams	79
4.7	The Kinect v2 body joints and labels	80
4.8	Per-participant totals vs. fully recorded gait cycles	83
4.9	Knee-angle trajectories across strides (subject 003)	86
4.10	Hip-angle trajectories across strides (subject 002)	87
4.11	Elbow-angle trajectories across strides (subject 030)	88
4.12	360-Degree Gait dataset: representative images from the different capture setups.	105
4.13	AVA Multi-View Dataset: studio layout with representative camera viewpoints.	105
4.14	CASIA-A dataset: three representative frames from the three viewpoints.	106
4.15	CASIA-B: representative frames across 11 viewpoints	107
4.16	CASIA-C dataset: two representative frames from the outdoor, night-time infrared dataset.	107
4.17	CASIA-E dataset: multiple images of the same scene captured from different viewpoints.	108
4.18	CCPG: representative camera views and clothing states across indoor and outdoor scenes.	109
4.19	CMU Motion of Body (MoBo): representative frames from the six synchronized viewpoints around the treadmill (frontal, rear, and four oblique/lateral views).	109

4.20	CCGR dataset: sequence of RGB frames with corresponding silhouettes, 2D poses, and human-parsing masks, temporally aligned.	110
4.21	Depth-Based Gait Dataset: sample depth-map frame with overlaid skeletal joints, illustrating joint locations.	111
4.22	Front-View Gait (FVG): sample frames of participants approaching a fixed camera at three near-frontal angles under representative clothing and carrying conditions.	112
4.23	Gait-A Database: representative RGB frame of a person walking along an indoor corridor in side-view.	112
4.24	GAIT-IST and GAIT-IT datasets: (a) indoor RGB frame of a walking subject shown side-by-side with the corresponding 2D skeleton; (b) RGB frame of a walking subject captured in a professional studio with the corresponding silhouette.	113
4.25	Gait3D-Parsing dataset: example RGB frames with corresponding human-parsing masks.	114
4.26	GREW: representative frames across public spaces and viewpoints, illustrating crowding, occlusion, clothing and carrying variation, and illumination change.	115
4.27	GRIDDS: representative frames with derived modalities, including RGB, depth, silhouettes, and 2D/3D joint coordinates.	116
4.28	Health&Gait: representative frames from each available modality, including 2D pose estimates, silhouettes, semantic-segmentation maps, and optical-flow fields.	116
4.29	INIT Gait Database: representative RGB frame and a representative silhouette extracted via chroma-key.	117
4.30	KOA-PD-NM dataset: representative RGB frame of a participant walking in the capture setup.	118
4.31	Kyushu University Gait Databases: (a) KY4D-A (top row) shows a gait pose captured simultaneously by three cameras; KY4D-B (bottom row) shows a sequence of three frames along a curved trajectory. (b) KY IR shows two perspectives (side-view and top-view) of a participant walking while carrying a bag.	119
4.32	MA-Gait dataset: example frames across multiple viewpoints with four of the sixteen attributes shown side-by-side – hands in pockets, happy, sad and angry.	120
4.33	Multi-Height Gait dataset: representative silhouette sequences across camera-height levels and appearance conditions.	120
4.34	Multi-Modal Gait Symmetry dataset: representative frames for the different modalities, including depth, silhouettes and 3D skeletons.	121
4.35	Occluded Gait Database (OccGait): example RGB frame, instance-segmentation mask, and corresponding silhouette.	122
4.36	OG RGB+D dataset: representative frames across covariates (carrying, clothing-occlusion levels, straight vs. turning trajectories, viewpoint change).	122
4.37	OU-ISIR Treadmill Gait Dataset: (a) sample frame from Dataset A showing treadmill walking; (b) sample frame from Dataset B showing treadmill walking while wearing a coat.	124
4.38	OU-ISIR Gait Speed Transition (GaitST): overground and treadmill acquisition setups with representative frames from each configuration.	124
4.39	OULP-Age dataset: representative side-view samples from subjects of different ages during acquisition.	125
4.40	OU-LP-Bag dataset: examples of carrying-status labels; panels show RGB frames (carried objects highlighted) with the corresponding GEIs.	126

4.41	OU-ISIR Large Population: Multi-View family: (a) representative GEIs from OU-MVLP; (b) example 2D keypoint sequences from OU-MVLP-Pose; (c) sample human-mesh representations from OU-MVLP-Mesh; (d) representative dense-flow visualizations from OU-MVLP-OF.	127
4.42	ProGait dataset: representative frontal-view frames with segmentation masks and 2D pose keypoints, sampled from parallel-bars and hallway scenarios.	128
4.43	RA-GAR dataset: examples of normal, bag-carrying, and clothing-change conditions with the corresponding data modalities (silhouettes and 2D pose keypoints).	129
4.44	ReSGait dataset: paired silhouette and pose sequences with representative examples of clothing variation, carrying conditions, accessories, and walking-trajectory.	129
4.45	SAIVT-Depth Gait Database: representative frontal-view image sequence with the corresponding depth frames.	130
4.46	Scoliosis1K dataset: representative frontal-view silhouettes and COCO-style 2D pose sequences.	130
4.47	SDUMLA-HMT gait dataset: multi-camera indoor setup (top) and representative RGB frames across covariates (bottom), including viewpoint changes, carrying status (with/without bag), and motion type (walking/running).	131
4.48	SMVDU-Gait dataset: representative frames from the Single (solo) and Multi (pairs) subsets, illustrating group-induced occlusion across straight paths.	132
4.49	SOTON HumanID (Small/Large) datasets: representative views and scenarios – indoor track, treadmill, and outdoor track – across fronto-parallel and oblique viewpoints.	132
4.50	SPHERE Walking Dataset: representative frontal-view stair-ascent sequences showing depth frames with corresponding OpenNI skeletons.	133
4.51	SUSTech1K (LidarGait) dataset: synchronized, spatially aligned RGB frames and silhouettes with the corresponding 3D LiDAR point-cloud sequences across representative viewpoints.	134
4.52	SZU RGB-D Gait Dataset: sample RGB frames and corresponding depth maps for the two trajectories: lateral (90°) and oblique ($\approx 60^\circ$).	134
4.53	TUM-IITKGP dataset: sample RGB frames for the covariates (a) regular, (b) hands-in-pockets, (c) backpack, (d) gown, (e) dynamic occlusion, (f) static occlusion.	135
4.54	Walking Gait Dataset: representative samples from frontal-view treadmill recordings showing silhouettes, 3D skeletons, and point-clouds.	136
5.1	Overall concept of the shape-based method for a single missing frame at t_0 . Black silhouettes are the observed neighbors $\{\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}, \mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}\}$, ordered from farther past to farther future relative to t_0 ; the gray silhouette is the target synthesized at t_0	162
5.2	An enlarged view of a row from a silhouette, where each square represents a pixel. The two sets of consecutive gray squares are the segments, while the sequence of white squares between them is referred to as the single existing hole. The numbers at the top of the figure represent the pixel coordinates on the horizontal (width) axis of the images. For easier visualization, the pixels are depicted in gray instead of black.	164
5.3	The segments intersection criteria explained in form of figures (letters H and T represent the <i>head</i> and <i>tail</i> of the corresponding segments): a) the <i>tail</i> of segment B is located between the <i>head</i> and <i>tail</i> coordinates of segment A; b) the <i>head</i> of segment B is located between the <i>head</i> and <i>tail</i> coordinates of segment A; c) segment B is located between the <i>head</i> and <i>tail</i> coordinates of segment A.	165

5.4	Time-aware tolerance for rowwise interval overlap. On a given row, segments from the past and future neighbors at lags τ (e.g., $\mathbf{P}_{\tau_1}$ and $\mathbf{F}_{\tau_1}$) are tested for correspondence using tolerance-expanded intervals $[h, t]^{+\tau(\Delta t)}$. As the temporal separation Δt increases (e.g., comparing τ_2 rather than τ_1), the tolerance $\tau(\Delta t)$ widens, permitting modest lateral drift; for near neighbors it remains tight.	165
5.5	Enlarged row across four temporal neighbors ($\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}, \mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}$)	166
5.6	Scanline across neighbors (midpoint $s = 0$; split at $+\tau_1$)	169
5.7	Simplified schematic of the proposed BiSINet architecture. Past and future silhouette streams are encoded separately through ConvLSTM-based encoders. Their multi-scale features are fused by attention modules, passed through a bottleneck, and decoded to reconstruct the missing intermediate silhouette $\hat{S}_m$	174
5.8	Three-frame gap illustration (Experiment 1)	178
5.9	Up-sampling factors (2, 3) in the Temporal Super-Resolution experiment	180
5.10	Error Map visualization of interpolation results.	182
5.11	Experiment 1 (Temporal Gap Reconstruction): trend of mean IoU vs. gap size across datasets for BRIEF	184
5.12	Experiment 2 (up-sampling): trend of mean IoU vs. factor across datasets for BRIEF	185
5.13	KY4D-A: CMC@1-5 curves contrasting original, down-sampled (factor 4), and up-sampled (factor 4) sequences.	187
5.14	GRIDDs: CMC@1-5 curves contrasting original, down-sampled (factor 4), and up-sampled (factor 4) sequences.	188
6.1	Template-based gait representations: GEI, GEnI, GFI, and CGI.	205
6.2	[GSI skeletal topology]Gait Skeleton Image topology showing the 17-keypoint structure and connectivity. (A) Front view. (B) Side view.	220
6.3	Comparison of body topologies: GSI, Kinect v2, and MediaPipe.	221
6.4	Visual examples of GSI variants and pixel parameters	232
6.5	Formation of GSI variants representations	234
6.6	Comparison of gait representations from multiple perspectives	240
A.1	Per-subject knee-angle trajectories (all subjects)	306
A.2	Per-subject hip-angle trajectories (all subjects)	307
A.3	Per-subject elbow-angle trajectories (all subjects)	308

List of Tables

2.1	Feature matrix of surveyed human motion analysis tools to the proposed framework, grouped by generic, modeling, acquisition, and analysis features. <i>Note: “*” indicates that a trial license is available.</i>	46
3.1	Normative spatio-temporal gait parameters at preferred speed (ages 20–59, healthy adults). Source: Oberg et al. (1993)	64
4.1	Age and height summary for GRIDDS	77
4.2	Microsoft Kinect v2 stream characteristics used in GRIDDS.	81
4.3	GRIDDS overall annotation totals	82
4.4	Event completeness across all trials. “Valid” counts events with assigned frame indices; “Missing” indicates NaN entries; completeness is the valid fraction in percent.	83
4.5	Summary of spatio-temporal features computed from camera-space joints.	85
4.6	Covariate-centered taxonomy in vision- and depth-based gait datasets	99
4.7	Healthcare gait datasets — taxonomy-aligned attributes	137
4.8	Biometric gait datasets — taxonomy-aligned attributes	138
4.9	Attribute-recognition gait datasets — taxonomy-aligned attributes	152
5.1	Past-referenced correspondence map $\text{cmap}_r^{\mathbf{P}_{\tau_1}}$ for Fig. 5.5. Each row lists one valid correspondence for a reference segment $j \in \{1, 2\}$ in $\mathbf{P}_{\tau_1}$ at lag $\lambda \in \{-\tau_2, +\tau_1, +\tau_2\}$. 167	
5.2	Future-referenced correspondence map $\text{cmap}_r^{\mathbf{F}_{\tau_1}}$ for Fig. 5.5. Each row lists one valid correspondence for a reference segment $j \in \{1, 2\}$ in $\mathbf{F}_{\tau_1}$ at lag $\lambda \in \{-\tau_2, -\tau_1, +\tau_2\}$. 167	
5.3	Multi-gap illustrative configurations used in BiSiNet training. Δ is the temporal gap (in frames) between visible inputs. <i>Notation:</i> $S_{a:b}$ are consecutive frames; $S_{a:b:s}$ denotes frames from a to b sampled every s frames.	173
5.4	Summary of BiSiNet’s structural components and output dimensions.	175
5.5	Frame-rate up-sampling transformations for the utilized datasets. Arrows ($\downarrow\uparrow$) denote the down-sampling ($\downarrow$) and subsequent up-sampling ($\uparrow$) framerates.	179
5.6	Experiment 1 (Temporal Gap Reconstruction): mean IoU across gap sizes of 1–3 frames.	183
5.7	Experiment 2 (up-sampling): mean IoU for factors $r \in \{2, 3, 4\}$	184
5.8	Up-sampling on KY4D-A: NMSE across factors $r \in \{2, 3, 4\}$	185
5.9	Up-sampling on GRIDDS: NMSE across factors.	186
5.10	Up-sampling on OU-ISIR TM-A: NMSE across even walking speeds.	186
5.11	KY4D-A: CMC@1-5 for original, down-sampled (factor 4), and up-sampled conditions.	187

5.12	GRIDDS: CMC@1-5 for original, down-sampled (factor 4), and up-sampled conditions.	187
6.1	GSI Keypoint Enumeration	221
6.2	Pose Estimation to GSI Conversion Summary	224
6.3	Parameter Optimization: Rank-1 Identification Rate on GRIDDS Dataset	235
6.4	Comparison of Skeleton-Based Gait Representations	238
6.5	Same-View Recognition Performance: Rank-1 Identification Rates (%) (Gallery and Probe from Identical Views, NM vs NM)	241
6.6	Cross-View Recognition Performance: GSI-Joints Rank-1 Rates (%) (NM Gallery vs NM Probe)	242
6.7	Cross-View Recognition Performance: GEI Rank-1 Rates (%) (NM Gallery vs NM Probe)	242
6.8	Cross-View Performance Summary by Angular Difference (NM Gallery vs NM Probe)	243
6.9	Detailed Performance at Moderate Cross-View Angle Pairs (NM Gallery vs NM Probe)	244
6.10	Same-View Recognition Under Carrying and Clothing Covariates (Mean Across All Views)	244

List of Acronyms

AI	Artificial Intelligence
BiSINet	Bidirectional Silhouette Interpolation Network
BRIEF	Body-contour Row-wise Interpolation and Envelope Fusion
CMC	Cumulative Match Characteristic
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
CSV	Comma-Separated Values
DL	Deep Learning
DNN	Deep Neural Network
DoF	Degrees of Freedom
fps	Frames per Second
GEI	Gait Energy Image
GNN	Graph Neural Network
GPU	Graphics Processing Unit
GRIDDS	Gait Recognition with Inertial and Depth Data Streams
GSI	Gait Skeleton Image
HMM	Hidden Markov Model
HSV	Hue, Saturation and Value
IoU	Intersection over Union
IR	Infrared
LSTM	Long Short-Term Memory
ML	Machine Learning
NMSE	Normalized Mean Squared Error
PCA	Principal Component Analysis
RGB	Red, Green and Blue
RGB-D	Red, Green, Blue and Depth
ROI	Region of Interest
RNN	Recurrent Neural Network
SDK	Software Development Kit
SOTA	State of the Art
ST	Spatio-Temporal
ToF	Time-of-Flight
VFI	Video Frame Interpolation

Chapter 1

Introduction

1.1 Introduction

Human motion analysis constitutes a multidisciplinary research domain that seeks to understand the complexities of human movement and its interaction with the surrounding environment. It integrates knowledge from diverse scientific domains, such as biomechanics, computer vision, artificial intelligence, and neuroscience, while also embracing developments in technology and the arts. The capacity to automatically detect, track, and interpret human motion in image and video sequences has unlocked new possibilities for designing systems that support detailed and accurate analysis of movement.

Such systems provide researchers, clinicians, engineers, and designers with powerful tools to investigate the mechanical, physiological, and behavioral aspects of human motion, supporting advances in diverse areas including clinical assessment, physical rehabilitation, sports science, ergonomic evaluation, computer animation, and human-computer interaction, among others.

The study of human motion has consistently attracted significant attention from the scientific community. This is reflected in the extensive body of research dedicated to the topic, as demonstrated by numerous surveys published over the years, such as those by [Low et al. \(2021\)](#); [van der Kruk and Reijne \(2018\)](#); [Halilaj et al. \(2018\)](#); [Aggarwal and Ryoo \(2011\)](#); [Poppe \(2007\)](#); [Gavrila \(1999\)](#); [Ke et al. \(2013\)](#), which highlight the growing interest in robust, scalable, and generalizable approaches to human motion analysis, and underscore the importance of continuing research in this area.

Among the various forms of human movement, gait stands out as one of the most fundamental and ubiquitous motor patterns. Walking not only represents the primary mode of human locomotion, but also serves as a rich source of information about an individual's musculoskeletal integrity, neurological function, and overall health status. The cyclical and structured nature of gait makes it particularly suitable for systematic analysis, while its inherent complexity – involving the coordinated activation of numerous muscles, joints, and neural pathways – provides a comprehensive window into human motor control and biomechanical function. The significance of gait

analysis extends across multiple domains. In clinical settings, the quantitative assessment of walking patterns enables the identification of pathological conditions, guides treatment decisions, and monitors rehabilitation progress for patients with neurological disorders, orthopedic conditions, or post-surgical complications (Hulleck et al. (2022); Manickam and Gardiner (2021)). In sports science, gait analysis contributes to optimizing performance and preventing injuries (Molavian et al. (2023)). Beyond these applications, gait patterns have also demonstrated potential for biometric identification, as the unique characteristics of an individual's walking style can serve as a distinctive signature (Wan et al. (2018); Sepas-Moghaddam and Etemad (2023); Rani and Kumar (2023)).

Recent technological advances, particularly in sensor technologies, computer vision, and machine learning, have transformed the landscape of gait analysis. Traditional laboratory-based systems, while highly accurate, are constrained by their requirement for controlled environments, specialized equipment, and expert supervision. Thus, these limitations have motivated the development of more accessible, portable, and automated approaches capable of capturing and analyzing gait in naturalistic settings. Such developments have broadened the applicability of gait analysis and created opportunities for continuous monitoring, large-scale screening, and integration into everyday technologies.

1.2 Motivation

The study of human motion has long captured the interest of researchers, revealing fundamental aspects of behavior, biomechanics, and cognition. Through detailed analysis of motor control, posture, coordination, and locomotion, we gain a comprehensive understanding of how the body moves, adapts, and interacts with its environment. This inherently interdisciplinary field—bridging biomechanics, computer vision, neuroscience, and machine learning—supports theoretical models that integrate physical, cognitive, and computational perspectives and translate into practical applications.

In parallel, the motivation to study human motion is strongly driven by its real-world impact across multiple sectors. In healthcare, motion analysis enables the early diagnosis of neuromuscular disorders, supports personalized treatment planning, and informs data-driven rehabilitation strategies. It also contributes to elderly care by monitoring functional mobility and risk of falls. In security, the integration of motion-based biometrics with behavioral analysis offers powerful means to improve public safety. The sports industry also benefits from motion analytics for performance optimization and injury prevention. In entertainment and digital media, motion data support realistic animation, immersive experiences, and interactive game design. Furthermore, progress in this field drives innovation in human–computer interaction and robotics, where movement-aware interfaces foster more intuitive and responsive systems.

Beyond its foundational and applied significance, human motion analysis continues to pose major scientific and technical challenges. Conventional motion capture relies on expensive hardware, controlled laboratory environments, and expert operation, which limits scalability and real-

world validity. The emergence of commodity sensors, embedded in smartphones, depth cameras and wearable devices, has expanded accessibility, yet persistent issues remain: synchronization of multi-modal data streams, integration of heterogeneous sensor inputs, and interpretation of complex, high-dimensional motion data. Environmental variations such as viewpoint changes, lighting conditions, and occlusions further affect accuracy and consistency. Addressing these challenges requires robust computational solutions, particularly those that take advantage of advances in deep learning and sensor fusion, to achieve accurate, scalable, and ecologically valid motion understanding.

1.3 Research Goals

Within this project, several research goals are defined. On a broader scale, the main objective is to advance the field of human gait analysis through the development of computational methodologies and models that exploit both RGB and depth information to achieve more robust, efficient, and comprehensive motion understanding. By integrating consumer-grade sensors, such as RGB-D cameras, this research also aims to improve the accessibility and reproducibility of gait analysis studies in scientific, clinical, and technological domains. A more detailed examination reveals the following specific objectives:

- First, a comprehensive study and comparative analysis of existing gait datasets is undertaken, with particular emphasis on those that incorporate depth information. This examination aims to identify the structural characteristics, limitations, and research potential of current multi-modal resources, thus establishing a solid foundation for future methodological developments.
- Second, based on these insights, the objective consists of designing and publicly releasing a novel RGB-D dataset for gait analysis (GRIDDS). This dataset is designed to provide a standardized, multi-modal, and openly accessible resource to support reproducible experimentation and benchmarking within the gait analysis community.
- Third, the proposal of two novel gait representations is pursued, designed to increase robustness to variations in viewpoint, appearance, and carried objects. These representations exploit both two-dimensional image information and three-dimensional skeletal data and are adapted for integration within deep learning frameworks, enhancing their capacity for automatic feature learning and recognition.
- Fourth, validation and comparative evaluation of the proposed representations are conducted, including their application to gait-based person recognition and related motion understanding tasks. Through extensive experimental evaluation, these methods are contrasted with existing approaches to demonstrate measurable improvements in performance and generalization.

- Finally, the development of new computational methods for gait silhouette interpolation is addressed, combining algorithmic reasoning with deep learning architectures. These methods are devised to enhance temporal coherence, reconstruct missing data, and improve the overall accuracy and resilience of gait analysis pipelines.

Together, these objectives define a coherent research trajectory that integrates the creation of dataset, the design of representations, and methodological innovation to enhance scientific understanding and the technological capabilities of human gait analysis.

1.4 Thesis Statement

The main goal of this thesis is to demonstrate that the integration of two-dimensional image data with three-dimensional skeletal and depth information facilitates the development of robust, efficient, and fully automatic computational methodologies for human gait analysis. The integration of multi-modal RGB-D data and biomechanically grounded skeletal models enables the capture of both appearance-based and structural-dynamic characteristics of gait within a unified framework. This fusion supports the construction of representations that are resilient to changes in viewpoint, clothing, and carried objects while maintaining temporal and spatial coherence across gait cycles.

The thesis provides evidence of this claim through:

- the creation and subsequent public release of a novel RGB-D dataset, designated GRIDDS, specifically designed for gait analysis;
- the proposal of new gait representations that integrate visual and skeletal cues with the objective of enhancing robustness and discriminative power;
- the development of interpolation strategies that integrate both algorithmic and deep learning-based approaches;
- a comprehensive experimental validation demonstrating improved performance in gait-based person recognition and motion understanding tasks when compared with existing state-of-the-art approaches.

This work thus substantiates that the integration of 2D and 3D motion cues, supported by deep learning frameworks, constitutes a verifiable and effective path toward the unification and advancement of computational gait analysis.

1.5 Publications

During the preparation of this thesis, a segment of research was initially carried out within the scope of the curricular unit of Methodologies for Scientific Research. This early work focused on the application of unsupervised learning techniques, namely k-means clustering, to group images

according to their visual content. The outcomes of this exploratory study were presented at the 5th Doctoral Symposium in Informatics Engineering (Nunes (2010)), establishing the first formal connection between pattern recognition methodologies and the broader research interests later developed in this thesis.

Building upon this initial work, several experiments were conducted addressing both image retrieval and image recognition problems. These studies led to the proposal and evaluation of a novel image descriptor, whose performance was shown to be competitive compared to existing state-of-the-art descriptors. The results highlighted the potential of the proposed approach across different datasets and tasks, culminating in a second publication presented at the 5th Iberian Conference on Information Systems and Technologies (Nunes et al. (2010)).

Despite the overall positive results, further analysis revealed suboptimal performance for specific shape classes within the evaluated datasets. It was hypothesized that these limitations could be attributed to inconsistencies between dataset annotations and human visual perception. To investigate this hypothesis, a user-centered study was conducted involving a significant number of volunteers, supported by the development of a dedicated web-based experimental platform Nunes et al. (2011a). The goal was to assess whether images belonging to different semantic classes could nevertheless be perceived as visually similar in the context of image retrieval tasks. The findings of this perceptual validation study were reported in a third publication (Nunes et al. (2011b)).

In parallel, and reflecting a progressive shift towards the analysis of human motion, a comprehensive and up-to-date survey was conducted on motion analysis and simulation software tools used in biomechanical studies of human movement. This work examined both academic and commercial solutions, providing not only an extensive overview of existing tools but also proposing a structured framework for their classification and comparison. This study resulted in a fourth publication Nunes et al. (2013), later extended and consolidated as a book chapter (Nunes et al. (2016)), which further reinforced its relevance and impact within the biomechanics and human motion analysis communities.

The focus on human motion analysis was subsequently deepened through preliminary experimental work on dynamic simulation of human gait using musculoskeletal models. This research explored the use of biomechanical simulation tools to compare normal and pathological gait patterns, highlighting the potential of physics-based simulation for quantitative gait analysis. The results of this work were presented at the European Simulation and Modelling Conference (Nunes et al. (2014)), representing an important step towards integrating computational modeling with gait analysis.

With the advent and increasing availability of low-cost depth sensors, the research evolved towards gait analysis and recognition using RGB-D data. An initial contribution in this direction was presented in the form of an extended abstract at the 15th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering (CMBBE 2018) (Nunes et al. (2018)), where a broad overview of gait recognition and classification approaches based on depth sensors was provided, alongside preliminary experimental results.

A major outcome of this research line was the development of a new publicly available dataset,

specifically designed for gait recognition using RGB-D sensors. This dataset, entitled GRIDDS – Gait Recognition Image and Depth Dataset, provides synchronized RGB images, depth data, skeleton joint coordinates and derived gait representations. Its design addressed limitations identified in existing datasets, particularly regarding sensor generation, data completeness and applicability. The dataset was presented at the VipIMAGE 2019 conference (Nunes et al. (2019b)), constituting a significant infrastructural contribution to the gait analysis research community.

Complementary to this contribution, a systematic review of publicly available RGB-D gait datasets was conducted, with the objective of identifying, analyzing and comparing existing resources used for gait-based person recognition and classification. This study proposed a structured comparison framework, facilitating dataset selection and highlighting gaps and redundancies in the field. The results of this systematic review were also published at VipIMAGE 2019 (Nunes et al. (2019b)), reinforcing the relevance of dataset-oriented research within gait analysis.

Overall, the publications produced during the doctoral programme reflect a coherent and progressive research trajectory, evolving from general image analysis and pattern recognition towards a focused and in-depth investigation of human gait analysis using computer vision, depth sensing and biomechanical modeling. These works collectively laid the conceptual, methodological and experimental foundations that underpin the contributions presented in this thesis.

1.6 Structure of the Thesis

This thesis is organized into seven chapters and one appendix, reflecting the logical progression of the research from its conceptual foundations to the development and validation of new computational methodologies for human gait analysis.

This chapter introduces the motivation for the research, the underlying questions, and the objectives that guided its development. It defines the central thesis statement—that exploiting two-dimensional visual and three-dimensional skeletal information enables more robust and comprehensive representations of human gait—and situates this work within the broader domain of human motion analysis. The chapter summarizes the scientific contributions and publications resulting from the research and concludes with an outline of the document’s organization.

Chapter 2 then establishes the theoretical and methodological background for the study. It discusses the principles and challenges of human motion analysis, traces its historical evolution, and highlights its diverse fields of application, including healthcare, sports, ergonomics, robotics, and entertainment. The chapter also examines how advances in computer vision and machine learning have enabled markerless, data-driven approaches to motion capture and interpretation.

Building on this foundation, Chapter 3 focuses specifically on human gait. Explores the biomechanical and kinematic principles of gait, as well as the parameters commonly used to quantitatively assess it. The chapter reviews the sensing technologies employed in gait analysis. Together, these two chapters establish the scientific and technical basis upon which the proposed computational methodologies are built.

Chapter 4 marks the transition from theoretical grounding to experimental contribution by introducing the GRIDDS dataset, which was developed as part of this work. The chapter details the dataset’s acquisition process, multi-modal structure, and annotation protocols. It also proposes a new taxonomy for vision- and depth-based gait datasets. Through a comparative analysis with existing resources, the chapter demonstrates how GRIDDS addresses the lack of standardized, publicly available, multi-modal data, providing a foundation for reproducible research in gait analysis.

Chapters 5 and 6 present the other core contributions of the thesis. Chapter 5 introduces two novel interpolation approaches, BRIEF and BiSINet, which are designed to reconstruct missing or degraded gait silhouettes. These methods combine algorithmic reasoning with deep learning-based inference to enhance temporal and spatial coherence, thereby improving the consistency of gait representations used in downstream recognition tasks. Chapter 6 introduces the Gait Skeleton Image (GSI), a unified spatio-temporal representation that encodes motion information extracted from two-dimensional visual data and three-dimensional skeletal cues. The chapter describes how to integrate the GSI into deep learning architectures and demonstrates through experimentation that this representation is robust to viewpoint variation, clothing, and occlusion. This establishes the GSI as an effective model for gait-based person recognition.

Finally, Chapter 7 summarizes the main findings and contributions of the research. The chapter discusses the implications of the methodologies developed, identifies existing limitations, and proposes directions for future work.

The thesis concludes with Appendix A, which presents supplementary material such as joint-angle trajectories derived from the GRIDDS dataset, provided to support reproducibility and future research based on this work.

Chapter 2

Human Motion Analysis – Fundamentals, Applications, and Challenges

Human motion is a fundamental aspect of everyday life, as it serves to perform tasks effectively and maintain overall health and well-being. The capacity to move effectively – whether walking, running, lifting, or performing any other physical activity – has a profound effect on personal satisfaction, independence, and quality of life. A comprehensive understanding of the principles underlying human movement, when combined with effective strategies to optimize movement, has the potential to significantly reduce the risk of injury, enhance physical performance, and ultimately foster greater happiness and personal achievement.

Consequently, the study of human motion goes beyond mere academic analysis and plays a central role in promoting physical and mental health, preserving autonomy, and increasing longevity. Advancements in this field hold considerable promise, providing the basis for innovative interventions and strategies that can positively change the way individuals live, function, and interact with their environment. Consequently, human motion analysis continues to stand as a pivotal area of research, offering extensive benefits not only to individual well-being but also contributing substantially to societal development as a whole.

The analysis of human motion is a broad and multidisciplinary field that encompasses the study of human movement, posture, and locomotion. With its foundations in disciplines such as biomechanics, computer science, engineering, and medicine, it offers a comprehensive framework to understand both the functional capabilities and limitations of the human body. Through detailed and systematic investigation, this field provides critical insight with wide-ranging applications: from the diagnosis and treatment of musculoskeletal disorders to the enhancement of athletic performance, the design of ergonomic equipment and workspaces, and the evaluation of human-computer interfaces. Furthermore, it plays a pivotal role in the development of robotic systems that emulate human movement and interact naturally with their environment, while also

serving as a powerful tool for recognition and interpretation by enabling the identification of individuals through motion patterns and the decoding of gestures as expressions of intention or communication.

This chapter presents a historical overview of human motion analysis, tracing its evolution and identifying significant milestones in research and technological advancements. Then, a comprehensive overview of the key components and current landscape of human motion analysis is provided, outlining the fundamental principles of the area and introduces a generic workflow describing the typical stages involved in human motion analysis. This is followed by an exploration of the main application areas, including healthcare, sports science, ergonomics and robotics, highlighting the contribution of human motion analysis to each domain. Concurrently, commonly utilized datasets relevant to these application areas are also presented. The chapter then provides an overview of existing automated tools for human motion analysis, accompanied by a classification framework to aid comparison and evaluation. The subsequent discussion covers the predominant challenges associated with motion analysis from image sequences, which encompass both technical and methodological limitations. Finally, the chapter narrows its focus to gait analysis, presenting it as a specialized and widely studied case within the broader context of human motion analysis.

2.1 A Historical Perspective on Human Motion Analysis

The field of human motion analysis has undergone a fascinating evolution throughout history, driven by the committed work of numerous researchers. This section presents a comprehensive overview of the historical milestones that have shaped this interdisciplinary field, emphasizing the contributions of pioneers who defined the foundations of contemporary human movement analysis. It encompasses a multiplicity of investigations, including the meticulous anatomical observations of Leonardo da Vinci, the revolutionary research conducted by Étienne-Jules Marey and Eadweard Muybridge, and the invaluable contributions of biomechanists such as Vladimir Zatsiorsky and Richard A. Schmidt, among others.

Early Beginnings

The origins of human motion analysis can be traced back to the 15th century, when pioneering scientists and engineers initiated efforts to study the intricate mechanics of human movement. Leonardo da Vinci (1452–1519) made significant contributions by meticulously capturing the complexities of human spinal posture, including its curvatures, articulations, and the precise count of vertebrae. Da Vinci's deep interest in the human body encompassed not only its structural aspects and their relationship to performance, but also the estimation of the center of gravity and balance. In his sketchbooks, he emphasized the vital role of anatomical knowledge, highlighting the understanding of nerves, bones, muscles, and sinews in decoding the causes and dynamics of

different motions. Figure 2.1 illustrates a selection of da Vinci's illustrative sketches, showcasing his anatomical studies and elucidating the mechanics of various human movements, including standing, ascending and descending slopes, rising from a seated position, jumping, and walking.

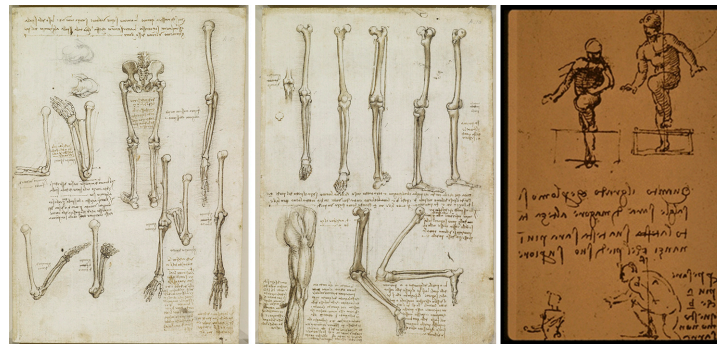


Figure 2.1: Illustrations extracted from the sketchbooks of da Vinci. (a) Anatomical studies of the human skeleton, showcasing intricate details and structural complexities. (b) Further anatomical examination of the human skeleton. (c) Study focusing on the center of gravity and balance, featuring a depiction of a man climbing a step. Credit: leonardodavinci.net

Giovanni Alfonso Borelli (1608–1679) also made important contributions to this field of research by clarifying muscular movement and the dynamics of the human body. In his seminal works, namely *De Motu Animalium I* and *De Motu Animalium II*, Borelli applied geometric analysis, analogous to Galileo's approaches in mechanics, to the domain of biology, thereby earning him the designation of the *Father of Biomechanics*. By applying geometric principles, Borelli provided comprehensive descriptions of human locomotion, encompassing activities such as walking, running, and jumping, while also exploring the complex mechanics of muscle contractions. Additionally, Borelli identified the forces necessary to maintain equilibrium in various joints of the human body and determined the position of the human center of gravity. Figure 2.2 exemplifies a selection of Borelli's illustrative drawings that capture his insights.

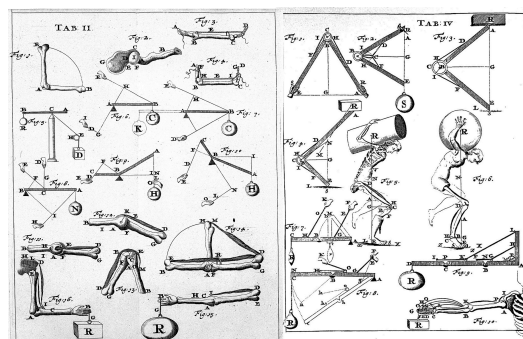


Figure 2.2: Illustrations extracted from Giovanni Alfonso Borelli's sketchbooks. (a) Illustration presenting a series of dynamic movements and muscle contractions. (b) Studies exploring the biomechanics of a man carrying a load on his back. Credit: [Wellcome Collection](http://WellcomeCollection.org).

In the 19th century, the Weber brothers, Wilhelm (1804–1891) and Ernst (1795–1878), conducted one of the first quantitative studies on the temporal and distance parameters of human locomotion. Their extensive analysis of human gait led to the development of foundational models for subsequent research in this domain. Their studies resulted in the publication of a seminal book, later translated and published as “*Mechanics of the Human Walking Apparatus*” (Weber and Weber (1992)), which significantly contributed to a deeper understanding of the fundamental principles underlying human locomotion. During the same period, Étienne-Jules Marey (1830–1904) and Eadweard Muybridge (1830–1904) were among the first to quantify patterns of human movement using photographic techniques. Marey revolutionized the study of motion through the development of a pioneering photographic technique. He designed a specialized camera, designated the chronophotographic gun, since it looked like a rifle (see Figure 2.3 a), which was capable of capturing 12 consecutive frames per second and merging them into a singular composite image. Using this equipment, Marey documented the various phases of human and animal movement, revealing details that often eluded the naked eye (see Figure 2.3 b). Through his photographic observations, Marey not only elucidated on the details of human locomotion, but also extended his studies to include the movements of horses, birds, dogs, and various other animals.

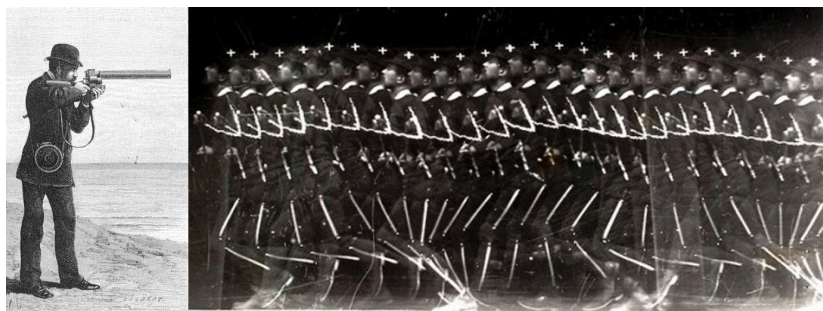


Figure 2.3: Illustrations showcasing some of Marey’s motion studies. (a) Illustration of the chronophotographic gun in use, from **Le Fusil photographique**, published in *La Nature* No. 464 on April 22, 1882. (b) An experiment capturing a person in motion, where the subject gets in a fitted black suit with reflective markers on their joints to better track the movements of different body parts in each phase. Credit: *The Linda Hall Library of Science*.

Marey’s research not only contributed to the understanding of motion but also played a pivotal role in the emerging field of cinematography. His pioneering work in acquiring and presenting image sequences laid a crucial foundation for the development of cinema, establishing him as both a pioneer in photography and an influential figure in the history of cinema. The legacy of Marey’s research continues to shape our understanding of motion analysis and its broader impact on visual media.

Influenced by Marey’s photographic work, Eadweard Muybridge (1830–1904) introduced an innovative approach to capturing fast motion. Regarded as the pioneer of motion capture, Muybridge conducted a series of innovative experiments known as “*Animal Locomotion*”. Employing an array of 12 cameras, he photographed the galloping motion of a horse in a sequential man-

ner. To synchronize the camera shutters, Muybridge created an electric mechanism controlled by a network of underground wires that were triggered by the horse carriage's wheel movements. Expanding upon his achievements, Muybridge designed the zoopraxiscope projector — a device capable of projecting a rapid succession of images recorded on rotating glass disks. This projection created the illusion of fluid motion. Muybridge's visionary efforts extended beyond animal locomotion, as he applied the same technique to study various forms of human movement. The culmination of his studies of human motion ended in the compilation titled “*The Human Figure in Motion*” (Muybridge (1955)). Muybridge introduced a series of captivating experiments that explored different forms of locomotion. These experiments highlighted the complexity of human movement, uncovering the subtle details and inherent challenges involved in activities such as walking, running, and jumping. Two examples of Muybridge's experiments are presented in Figure 2.4, illustrating the essence of his contributions to the field of motion analysis.

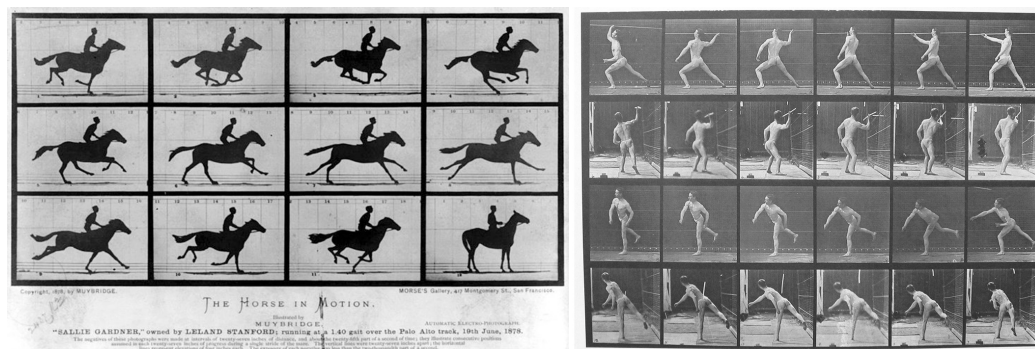


Figure 2.4: Some of Muybridge's motion studies: (a) The Horse in Motion study, consisting in recording the motion of a galloping horse; (b) A human movement study during a javelin throw session. Credit: *Wellcome Collection*.

Wilhelm Braune (1831–1892), a German anatomist and professor at the University of Leipzig, and Otto Fischer (1861–1916) also made pioneering contributions to the biomechanical analysis of human motion through the first three-dimensional mathematical study of locomotion, published in “*Der Gang des Menschen*” (Braune and Fischer (1895)). Their research covered detailed investigations of free walking and walking with a load, establishing a gait analysis methodology that has predominantly persisted to the present day. They introduced novel methods for tracking the instantaneous position of a moving point in three-dimensional space, enabling the analysis of displacements, velocities, accelerations, and forces during locomotion. This significantly enhanced the accuracy of kinematic parameter calculations compared to earlier two-dimensional photographic techniques.

During the post-World War II period, when the demand for enhanced rehabilitation strategies for returning veterans was critical, Verne Thompson Inman (1905–1980), Howard D. Eberhart (1906–1993), and J. B. deC. M. Saunders (1903–1991) conducted pioneering research that shaped the field of motion analysis. Their collaborative work at the University of California led

to the development of fundamental methodologies, many of which remain relevant in contemporary biomechanical studies. One of their most significant advancements was the introduction of three-dimensional kinematic analysis, achieved through precise measurement of displacement, velocity, and acceleration, along with the use of bone-embedded markers to accurately track rotational movements. Their findings provided critical insights into both normal and pathological gait, significantly impacting prosthetic design and rehabilitation techniques (Saunders et al. (1953)).

In the mid-to-late 20th century, Mary Pat Murray, David H. Sutherland, and David A. Winter made significant contributions to the field of gait analysis, each introducing pioneering methodologies that shaped modern biomechanical research. Notably, both Murray and Sutherland studied under Inman at the University of California, where they built upon his work in human locomotion. Mary Pat Murray (1925–1984) advanced the measurement and characterization of normal and pathological walking patterns by developing a novel method for quantifying human gait across different ages. She introduced multi-plane kinematic gait analysis — sagittal, coronal, and transverse — allowing for a detailed examination of step and stride lengths, foot progression angles, pelvic and limb rotations, and upper extremity motion (Murray (1967, 1970)). Her methodology used reflective markers, illuminated by a stroboscopic light source and captured with high-speed photography, producing sequential stick-figure representations of movement. David H. Sutherland (1923–2006) was a leading figure in pediatric gait analysis, with a particular focus on gait development in children and the effects of cerebral palsy on locomotion. He conducted a comprehensive study on the maturation of walking in children, analyzing more than 400 studies involving nearly 300 children. In collaboration with John L. Hagy, Sutherland developed an innovative motion picture-based gait analysis technique, employing three orthogonally positioned cameras calibrated with reference markers along a walkway to capture angular and translational displacements of key joints (Sutherland and Hagy (1972)). This method facilitated simultaneous bilateral gait recording and integration with electromyographic data, enhancing the precision and scope of gait analysis. Another leading figure in biomechanics and kinesiology, David A. Winter (1930–2012) made significant contributions to experimental gait and balance analysis (Winter (1995)). He developed an automated television-based motion capture system, a significant step toward modern computerized gait analysis (Winter et al. (1974)). He also introduced methods for analyzing joint moment-generated power, which enhanced the understanding of mechanical energy contributions of different joints during locomotion. His research resulted in the publication of *Biomechanics and Motor Control of Human Movement*, a seminal textbook that remains a fundamental text in the study of gait biomechanics, motor control, and neuromuscular coordination (Winter (1991)).

Human Motion Analysis in the 21st Century

At the turn of the 21st century, human motion analysis experienced exceptional growth, driven by advancements in sensor technologies, computer vision, artificial intelligence, and biomechanical modeling. The field transitioned from traditional optical motion capture systems, which relied on infrared cameras and reflective markers, to markerless motion tracking technologies that leverage depth sensors, structured light cameras, and high-speed vision systems (Figure 2.5 (a); Lam



Figure 2.5: Examples of technologies and models used in human motion analysis: (a) depth map generated by a depth sensor; (b) 2D skeletal tracking overlaid on an RGB image of a human subject; (c) various biomechanical models representing the human body for motion analysis.

et al. (2023)). Devices such as Kinect (Steinebach et al. (2020)), time-of-flight (ToF) cameras (Ganapathi et al. (2010)), LiDAR-based motion sensors and wearable inertial measurement units (IMUs) (Gu et al. (2023)) significantly enhanced the accuracy, accessibility and portability of motion capture. Meanwhile, improvements in camera hardware, including higher frame rates, increased spatial resolution, and improved low-light performance, contributed to more precise and reliable motion tracking.

Simultaneously, the integration of deep learning-based models revolutionized human pose estimation (Zheng et al. (2023a); Wang et al. (2021)), allowing robust real-time tracking without the need for physical markers. The advancement of AI-based skeleton tracking algorithms has significantly broadened the scope of human motion analysis. These technological developments have made it possible to conduct detailed motion capture and analysis outside controlled laboratory or clinical environments, relying even on video footage or image sequences acquired with standard cameras (Figure 2.5 (b)). This increased accessibility has facilitated the integration of motion analysis into a wide range of real-world contexts, supporting applications that extend beyond traditional research and healthcare settings to encompass domains such as performance monitoring, behavioral assessment, and physical training. The ability to extract meaningful motion-related information from unconstrained visual data represents a critical step toward more scalable, cost-effective, and context-aware solutions for understanding human movement.

Additionally, advancements in biomechanical modeling led to more realistic and adaptable representations of human motion, integrating detailed musculoskeletal dynamics, force prediction, and joint kinematics analysis (Sylvester et al. (2021); Blemker et al. (2007)). These computational models (Figure 2.5 (c)) allowed simulation of complex movements, allowing researchers and practitioners to analyze movement patterns, joint load distribution, muscle coordination, and neuromuscular interactions with greater precision across a wide range of activities.

2.2 Foundations of Human Motion Analysis

The analysis of human motion typically follows a structured workflow consisting of several fundamental stages, as illustrated in Figure 2.6. This generic framework has been systematically delineated by various authors over decades of research. Foundational contributions by Aggarwal and Cai (1999) and Gavrilu (1999) established initial structured approaches, which were subsequently expanded and refined through comprehensive reviews by Wang et al. (2003a), Moeslund et al. (2006), and Poppe (2010).

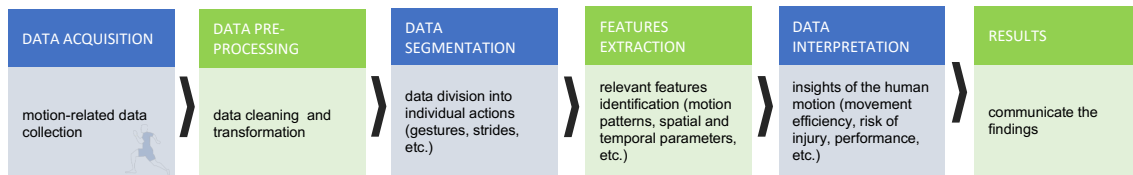


Figure 2.6: A generic framework for human motion analysis, encompassing stages of data acquisition, pre-processing, segmentation, feature extraction, modeling, analysis, and visualization.

Each stage of this workflow plays a distinct role in transforming raw sensor data into actionable insights, as detailed below.

Data Acquisition.

The initial stage of a human motion analysis workflow involves **data acquisition**, in which motion-related data is collected using a wide variety of sensor-based technologies, including inertial sensors, optical motion capture systems, radio frequency and ultrasound trackers, Global Positioning Systems (GPS), force plates, pressure sensors, electromyography (EMG), depth sensors, video cameras, and vision-based artificial intelligence systems. Inertial sensors, such as accelerometers, gyroscopes, and magnetometers, are commonly integrated into Inertial Measurement Units (IMUs) to capture precise measurements of orientation, angular velocity, and linear acceleration (García-de Villa et al. (2023)). Optical motion capture systems, which can be marker-based or markerless, utilize camera arrays to accurately track and reconstruct human body movements (Khan et al. (2020); Wade et al. (2022)). Radio frequency tracking systems, such as Ultra-Wideband (UWB), and ultrasonic tracking solutions, provide real-time localization within controlled indoor environments (Bharadwaj et al. (2017)), while GPS is used predominantly for outdoor motion tracking requiring greater spatial coverage (Liu et al. (2023)). Force plates and pressure sensors offer additional biomechanical information by quantifying ground reaction forces and pressure distributions during movement execution (Evans et al. (2024)). Electromyography (EMG) sensors provide valuable data on muscle activation patterns, facilitating deeper physiological interpretations (Tang et al. (2024)). Depth sensors, including structured-light cameras and LiDAR, generate comprehensive three-dimensional spatial information (Khoshelham and Elberink (2012); Tu et al. (2021)). Recently, the emergence of vision-based artificial intelligence systems,

exemplified by DeepLabCut (Mathis et al. (2018)), OpenPose (Viswakumar et al. (2019)) and MediaPipe (Lin et al. (2023)), has significantly advanced markerless human motion analysis. These AI-driven approaches utilize deep learning techniques to accurately estimate human poses directly from standard video footage, allowing scalable and non-invasive motion studies (Zheng et al. (2023a)). Recent developments have increasingly integrated multiple sensor modalities within wearable devices and smartphones, thus offering convenient, accessible and portable solutions that are suited to a variety of real-world environments (Vincent et al. (2022)). Ultimately, selecting the appropriate sensor technologies depends on the specific objectives of the motion analysis, environmental constraints, the desired precision, and practical factors such as cost, mobility, and ease of use.

Pre-processing.

Following data acquisition, the raw motion data undergo a comprehensive **pre-processing** stage, an essential step to ensure the quality and reliability of subsequent analysis. During pre-processing, the data is systematically processed to eliminate noise, artifacts, or inaccuracies that typically arise from sensor limitations, environmental interference, or subject movements (Prasad et al. (2023); Chu et al. (2016)). Common procedures include filtering techniques, such as low-pass or high-pass filters, which remove unwanted frequencies or sensor drift. Synchronization is also a critical pre-processing step, especially when integrating data from multiple sensors or capturing devices (Katmah et al. (2023); Hazra et al. (2021)). Techniques such as timestamp alignment, interpolation, and resampling ensure temporal consistency across datasets (Senel et al. (2023)). Moreover, calibration procedures are often performed during pre-processing to correct any systematic errors or biases that might be present in sensor measurements. For example, inertial sensors and IMUs frequently require calibration to correct orientation drift (Wang et al. (2023c)), while optical systems require spatial calibration to ensure accurate three-dimensional tracking. Data normalization or scaling can also be employed to standardize measurements, facilitating comparisons between different subjects, movements, or experimental conditions (Singh and Singh (2020)). Finally, the processed data are organized and converted into structured formats, such as tabular data, multi-dimensional arrays, or feature vectors, which are suitable for computational analysis and feature extraction in subsequent stages.

Data Segmentation.

Once pre-processing is complete, the motion data enter the **segmentation** stage, wherein the continuous data streams are organized into meaningful and analyzable segments (Mattheus et al. (2020)). Segmentation involves dividing the pre-processed data into distinct units or sequences that correspond to specific actions, events, or movement phases (Kruger et al. (2017)). Examples of segments include individual strides in gait analysis, specific gestures in hand-motion studies, postural transitions (e.g., arm swings, jumping jacks, etc.), or expressions in facial motion capture. Several segmentation strategies can be employed, ranging from simple threshold-based approaches to advanced machine learning methods. Threshold-based segmentation uses predefined criteria,

including velocity thresholds, positional data, or temporal markers, to detect the onset and offset of specific actions. More sophisticated approaches involve dynamic programming, Hidden Markov Models (Marroquin et al. (2003); Kabir et al. (2016)), or machine learning techniques (Chong et al. (2021)), including deep learning methods, which automatically identify and classify movements based on learned temporal and spatial patterns. Imprecise segmentation boundaries can significantly impact subsequent feature extraction and interpretation. Consequently, validation procedures, such as manual verification or automated cross-validation techniques, are frequently implemented to ensure robust and precise segmentation outcomes.

Feature Extraction.

Subsequent to segmentation, the motion data proceeds to the **feature extraction** stage, where the most informative and relevant parameters are identified, quantified, and represented. The extracted features serve as compact and meaningful representations of complex movements, facilitating subsequent analyses and interpretations. Typical extracted parameters include a varied range of spatio-temporal and biomechanical descriptors. Spatio-temporal features, which are of particular relevance in this context, include duration, frequency, speed, stride length, cadence, and temporal intervals of specific movement phases, providing critical insights into the timing and rhythmic structure of human motion. Kinematic features, such as joint angles, angular velocities, and accelerations, are fundamental to characterizing posture, orientation, and the dynamic evolution of movement patterns. In addition, kinetic parameters, including ground reaction forces and joint moments, contribute to the evaluation of internal and external loads experienced during physical activity. Given the high dimensionality and complexity of motion data, feature extraction is often complemented by dimensionality reduction techniques. Classical methods such as Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) are frequently employed to project extracted features into a lower-dimensional space, preserving the most informative components while avoiding redundancy (Cicirelli et al. (2022); Monaghan et al. (2021); Gavrilović and Popović (2021)). More recently, deep learning approaches, such as autoencoders and their probabilistic variants (e.g. variational autoencoders), have gained prominence for learning compact, non-linear representations directly from data (Jun et al. (2020); Felius et al. (2024)).

Data Interpretation. Following feature extraction, the quantitative data undergoes **data interpretation**, wherein numerical findings are transformed into meaningful, context-specific insights that inform decision-making and intervention strategies. This stage combines computational analysis techniques with domain expertise to bridge the gap between raw outputs and practical applications. Interpretation employs diverse analytical methodologies depending on the nature of the task and available data. Supervised machine learning approaches are widely utilized when labeled training data is available, enabling classification or regression tasks. Common supervised methods include Support Vector Machines (SVMs), which effectively handle high-dimensional feature spaces and are frequently applied to activity recognition and gait pattern classification (Vidya and P (2021)). Random Forests and Decision Trees offer interpretable models that can identify

key discriminative features while handling non-linear relationships (Khera and Kumar (2020)). K-Nearest Neighbors (k-NN) provides intuitive classification based on similarity metrics, proving particularly useful in clinical contexts where interpretability is paramount. Neural networks, including feedforward architectures and Convolutional Neural Networks (CNNs), have demonstrated superior performance in complex classification tasks such as pathological gait detection or movement quality assessment (Albuquerque et al. (2021b); Sadeghzadehyazdi et al. (2021)). When labeled data is unavailable or the objective involves discovering inherent patterns within the data, unsupervised learning techniques become essential. Clustering algorithms, such as k-means, hierarchical clustering, and Gaussian Mixture Models (GMMs), enable identification of distinct movement profiles or subgroups within populations without prior labeling (Caldas et al. (2020); Ramirez et al. (2025)). These approaches are particularly valuable for exploratory analysis, identifying movement phenotypes, or stratifying patients based on kinematic similarities. Anomaly detection methods, including Isolation Forests and One-Class SVMs, facilitate identification of atypical or abnormal movement patterns that deviate from established norms, supporting early detection of motor dysfunction or technique deterioration (Kim et al. (2022)). Statistical inference techniques also play a crucial role in interpretation, particularly when hypothesis testing or comparative analysis is required. Parametric tests, such as t-tests and Analysis of Variance (ANOVA), enable comparisons between groups or conditions (Hagoort et al. (2023); Hsieh et al. (2019)). Correlation and regression analyses quantify relationships between kinematic parameters and outcome variables, facilitating predictive modeling and identification of risk factors. Beyond computational methods, interpretation encompasses multiple analytical dimensions depending on the application context. In biomechanical assessments, interpretation focuses on evaluating movement efficiency, identifying asymmetries or compensatory movement patterns, and assessing performance metrics such as peak velocities, joint range of motion, or temporal consistency. Risk assessment constitutes another critical dimension, in which motion characteristics are analyzed to estimate injury susceptibility, identify fatigue-related deterioration, or detect risk of falling in vulnerable populations. The interpretative process remains inherently context-dependent, with significance varying substantially across domains. Clinical applications prioritize detection of pathological patterns and monitoring rehabilitation progress, while sports science emphasizes technical refinement and performance optimization. Ergonomic assessments focus on postural risk identification and workplace design evaluation. Research-oriented interpretations may seek to validate theoretical models or investigate fundamental motor control mechanisms.

Results Presentation and Visualization.

The final stage of the workflow is the **presentation of results**, where the findings are communicated through visualization techniques that make the data accessible and comprehensible to potential users. Effective visualization improves the interpretability of complex data by translating key patterns and trends into intuitive graphical forms. The selection of the most appropriate visualization depends on the objectives and the audience, and may include time-series plots, spatial trajectory diagrams, heatmaps, bar and scatter plots, among others. In more advanced systems,

dynamic visualizations, such as animated skeletal reconstructions, 3D motion trajectories, and interactive dashboards, are employed to illustrate motion characteristics with greater clarity and depth. These tools enable users to interact with the data in real time, facilitating exploratory analysis and personalized feedback. Visualization is especially important in multidisciplinary contexts, as it supports communication between technical and non-technical users, bridging gaps between data analysts, clinicians, coaches, engineers, and end-users. The synthesis and visual analysis of results at this stage is of crucial importance, as it facilitates the transformation of raw data into insights that are both comprehensible and actionable. This, in turn, supports effective decision-making and intervention design in a variety of domains.

2.3 Application Areas of Human Motion

The analysis and comprehension of human motion, characterized by a complex relationship of biomechanical mechanisms and behavioral patterns, represents a critical area of study with significant implications across multiple domains. Rather than being limited to the observation of movement, motion analysis provides deep insights that support informed decision-making and innovation in a wide range of fields.

A detailed understanding of human motion is essential for advancements in healthcare, such as rehabilitation and diagnosis; in sports science, for performance optimization and injury prevention; in ergonomics, for workplace design and risk assessment; in entertainment, for realistic animation and virtual characters; in security, for surveillance and individual identification; and in robotics, for movement replication and human-robot interaction.

The next subsections describe key application areas of human motion analysis, along with examples of related available datasets. These datasets help illustrate the nature of the motion data typically gathered and its relevance to specific fields of application.

2.3.1 Healthcare and Rehabilitation

Human motion analysis has become an essential component within biomedical research and clinical practice, particularly in clinical assessment and rehabilitation. It provides objective quantification of complex motor functions and assists healthcare professionals in the diagnosis, monitoring, and management of various medical conditions.

In terms of patient safety, particularly for vulnerable populations such as the elderly or patients with neurodegenerative diseases, motion analysis using motion sensors and computer vision has emerged as a critical tool for fall detection and prevention (see Figure 2.7a). Such technologies identify abnormal postures and movements, promptly alerting those in charge of care and significantly reducing the risk of injury through timely intervention (Stone and Skubic (2015); Wang et al. (2015); Vadivelu et al. (2017); Li et al. (2017); Solbach and Tsotsos (2017); Esteva et al. (2019, 2021); Ramirez et al. (2021)).

Neurological conditions such as Parkinson's disease, multiple sclerosis and cerebral palsy also benefit greatly from accurate motion analysis. In Parkinson's disease, motion capture systems can

objectively quantify early motor impairments such as tremor, improving the accuracy of early diagnosis and the effectiveness of treatment. These systems generate reliable biomarkers related to movement amplitude, velocity and rhythm, helping clinicians to monitor disease progression and evaluate treatment outcomes (Powell et al. (2012); Kour et al. (2019); Ferreira-Sánchez et al. (2020)). Similarly, in patients with cerebral palsy, detailed motion analysis accurately characterizes motor impairments, informing clinical decisions and optimizing patient-specific therapeutic and surgical strategies (Adde et al. (2010); Zhao et al. (2024)).

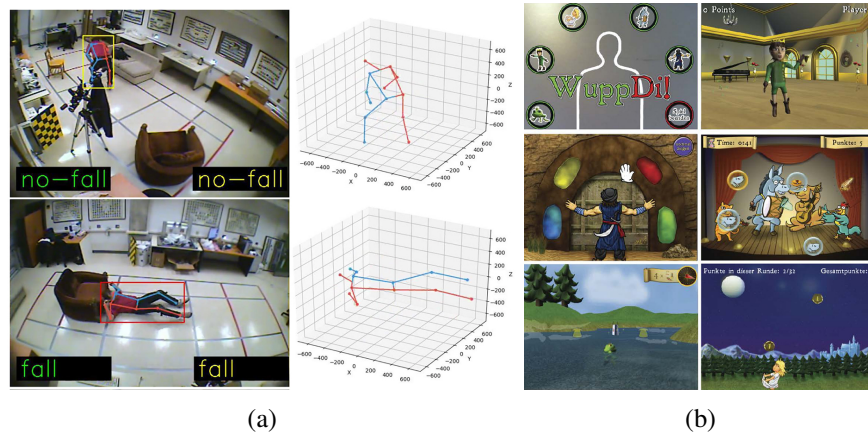


Figure 2.7: Instances of two healthcare applications: (a) A framework designed for automatic fall detection from a single image, demonstrating invariance to appearance characteristics (Asif et al. (2020)). (b) A snapshot of the five mini-games in a serious games rehabilitation framework proposed by Assad et al. (2011). where players need to trigger buttons through hand movements.

In the context of physical rehabilitation, particularly physical recovery and cognitive rehabilitation, innovative systems using motion analysis technologies significantly improve the effectiveness of therapy through accurate monitoring and real-time feedback. Motion-based serious games have become increasingly important, providing engaging and interactive environments specifically tailored to rehabilitation needs, significantly improving patient motivation, adherence, and results (Ricciardi and Paolis (2014); Rego et al. (2010); Bonnechère (2017); González-González et al. (2019)). For example, Rego et al. (2016) presents a platform that combines natural user interfaces with social features such as collaboration and competition to enhance patient engagement. Similarly, Assad et al. (2011) developed a suite of five motion-based games to promote exercise engagement in people with Parkinson’s disease (see Figure 2.7b). These games are specifically tailored to the challenges faced by this user group and provide an engaging and accessible rehabilitation experience.

Orthopedic practice also benefits greatly from motion analysis techniques. Detailed assessments of joint motion facilitate accurate diagnosis, improve surgical planning accuracy and contribute significantly to effective post-operative rehabilitation protocols. The ability to quantitatively assess joint functionality accelerates recovery and helps healthcare providers tailor rehabilitation plans to meet patients’ specific needs. Jicinsky and Mares (2021) developed a prototype

of a motion analysis tool for orthopedic diseases, focusing on rehabilitation after surgery of large joints (e.g., shoulder, knee, hip).

A particular focus within healthcare is clinical gait analysis, commonly defined as the measurement, processing, and systematic interpretation of biomechanical parameters that characterize human locomotion (Davis (1997)). Traditionally, gait assessment has relied on subjective visual observation, a simple yet inconsistent method that is still widely used in settings without access to specialized equipment. However, recent advances in wearable sensors, deep learning, and computer vision are transforming this field, enabling more objective and reliable assessments. Portable and affordable systems now allow for continuous monitoring, real-time feedback, and integration with physiological data across both clinical and home environments, showing promising results in terms of accuracy and robustness (Hulleck et al. (2022); Kidziński et al. (2020); Milstein and Topol (2020)). These techniques are effective at extracting gait parameters for classifying various neurological and motor disorders, including Parkinson's disease (di Biase et al. (2020); Mirelman et al. (2019)), cerebral palsy (Papageorgiou et al. (2019); Armand et al. (2016)), and stroke (Selves et al. (2020)).

A wide variety of datasets, developed specifically for healthcare and rehabilitation purposes, are available to support research and clinical practice (David et al. (2024); Hood et al. (2020); Perera et al. (2024)). An example is the UR Fall Detection Dataset (Kwolk and Kepski (2014)), which contains data collected by an IMU inertial device and two Kinect cameras. The dataset consists of sequences of volunteers performing falls from standing and sitting positions, as well as fall-like activities. Each recording provides depth images, RGB images and raw accelerometer data, enabling comprehensive analysis of fall detection. Another featured dataset is the U-Limb database (Averta et al. (2021)), a multimodal collection focusing on upper limb movements involving healthy participants and stroke patients. This dataset includes kinematic and physiological signals (EMG, EEG, ECG) recorded during activities of daily living, force-kinematic behavior during precision manipulation tasks, and brain activity during hand control using functional magnetic resonance imaging. The database aims to provide a quantitative assessment of motor conditions and to support advances in assistive robotics and human-machine interaction. Figure 2.8 illustrates sample images from each dataset.

2.3.2 Sports Performance and Injury Prevention

The analysis of human motion is particularly significant within sports science, where it serves as a foundational tool for performance enhancement, technique refinement, and injury prevention. Detailed motion studies enable coaches and athletes to objectively evaluate and optimize athletic performance by identifying strengths and technical deficiencies. Motion data can also assist in preventive strategies by identifying movements that may lead to injuries.

In team sports such as soccer, volleyball, basketball, or handball, motion analysis typically focuses on collective behaviors and game dynamics. Tools like GPS-based systems, video analysis, and visual tracking are commonly utilized to monitor parameters such as positioning, speed, distance covered, and athlete accelerations in real-time (Theodoropoulos et al. (2020); Cummins



Figure 2.8: Example data from two healthcare and rehabilitation datasets: (a) UR Fall Detection: sample frames showing a fall event captured using RGB and depth cameras (Kwolek and Kepski (2014)). (b) U-Limb: representation of multimodal data acquisition during upper limb movement task, including kinematic measurements and physiological signals (Averta et al. (2021)).

et al. (2013)). These data allow coaches to evaluate team performance and optimize game strategies. For instance, Santiago et al. (2012) introduced a vision-based system to extract detailed statistics and monitor player performance in handball games, enhancing the understanding of tactical team dynamics. Other approaches include sports video summarization techniques, which facilitate the efficient review and analysis of key moments in matches (Vasudevan and Gounder (2023)), and models that capture ball-player interactions across sports like volleyball, basketball, and soccer, enabling a deeper interpretation of game flow and decision-making processes (Maksai et al. (2016)). An illustrative example of such analysis is presented in Figure 2.9a, showing a passing network constructed from Wyscout data (Cao (2023)).

In contrast, individual sports, including gymnastics, athletics, swimming, baseball, or golf, primarily focus on the technical precision of each athlete. Here, a detailed biomechanical analysis is crucial for identifying subtle deviations that could negatively impact performance or increase injury risk. Examples include detailed studies on running mechanics, investigating biomechanical differences between athletes of varying ages and genders, allowing targeted training adaptations (Boyer et al. (2016)). Another example involves the use of machine learning techniques for analyzing and correcting golf swing techniques, facilitating the identification and correction of incorrect movements in beginner athletes (Liao et al. (2022a)). Moreover, biomechanical throwing analysis in baseball has been used to highlight the measurement of hip-shoulder separation, providing insights into performance and injury prevention (Trasolini et al. (2022)). Figure 2.9b illustrates this type of motion analysis (Luo and Xiao (2023)) through a sample sequence from the FineGym dataset (Shao et al. (2020a)), focusing on the execution of a complex routine.

Gait analysis has also become a valuable component of sports biomechanics, particularly in disciplines that involve running and other locomotion-based activities. As a fundamental expression of motor control, gait reflects not only an individual's unique movement signature and serves as a key indicator of performance capacity and physical health. Deviations in gait patterns may reflect underlying injuries, biomechanical imbalances, or movement adaptations resulting from

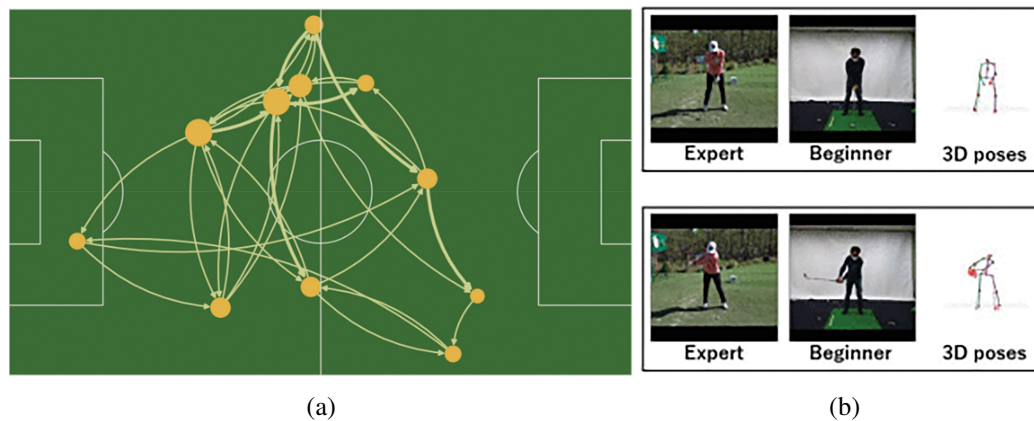


Figure 2.9: Two examples of applicability of human motion in team and individual sports: (a) Soccer team ball passing network, with nodes as players and arrows indicating frequency and direction of successful passes (Cao (2023)); (b) Golf swing analysis tool comparing beginner and expert poses, highlighting key differences to support skill improvement (Liao et al. (2022a)).

overtraining or technical flaws. Recent advances in artificial intelligence and machine learning have transformed the analysis of gait in sports contexts: modern mobile systems can precisely collect kinematic and physiological data in real time during training and competition. These systems use classification algorithms to evaluate running dynamics, segment motion sequences, and detect anomalies in variables such as speed, cadence, and posture. Beyond its application in elite sports, gait analysis has demonstrated value in addressing the needs of populations with developmental challenges. For example, children and adolescents with sensorineural hearing loss may experience impairments in balance and gait control. A systematic review examining the effects of sport and recreational activity on this group concluded that participation in physical activity appears to yield positive outcomes for gait and motor development (Melo et al. (2020)). As previously stated, computer vision techniques powered by artificial intelligence enable the extraction of biomechanical data from video recordings. For instance, Mundt et al. (2024) found these methods to be effective for detecting gait events, estimating joint angles, and quantifying key performance metrics in sports contexts when evaluating 2D video-based motion analysis tools.

In the field of motion analysis applied to sports, several specialized datasets have been developed for different disciplines. One such example is the Event and Match dataset from Wyscout (Pappalardo et al. (2019)), which is considered to be one of the most comprehensive open collections of soccer event logs, containing detailed spatio-temporal information (passes, shots, fouls, etc.) for every match in an entire season across seven top-tier competitions. Each event in the dataset includes information such as the type and subtype of the action, its timestamp, the player involved, field position, and outcome. Another example is provided by the FineGym dataset (Shao et al. (2020a,b)) which includes annotated sequences of artistic gymnastics actions, capturing the hierarchical structure of routines. Figure 2.10 illustrates sample images from each dataset.

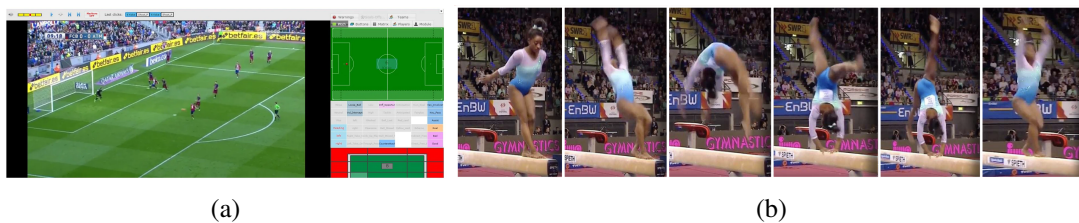


Figure 2.10: Two sample images from sports-related datasets: (a) a frame from a soccer match in the Wyscout Event and Match dataset, integrated into a tool for automatic event detection (Papalardo et al. (2019)); (b) Sample sequence from the FineGym dataset illustrating a Balance Beam routine (Shao et al. (2020a)).

2.3.3 Surveillance and Security Applications

The need for effective surveillance systems in public spaces has become increasingly evident, driven by a growing sense of insecurity in people's lives. Traditional surveillance cameras are already widely used in various public areas, such as shopping malls, banks, airports, subways, train stations, etc. (Socha and Kogut (2020); Welsh (2009)). However, their video output is often recorded and stored primarily for forensic purposes, limiting the benefits of real-time monitoring. Typically, human operators continuously monitor security camera feeds (Hsieh et al. (2008); Wang et al. (2003a)). This process becomes progressively tedious and error-prone as the number of cameras, working hours, and operator fatigue increase.

To overcome these limitations, intelligent video surveillance systems have been developed to automatically pre-process video streams and detect critical events such as suspicious object abandonment (Tripathi et al. (2018); Bhargava et al. (2007)), aggressive behavior (Kooij et al. (2016); Patwardhan and Knapp (2016)), or individuals carrying dangerous objects. These events then trigger alerts for human assessment, enabling the efficient management of large surveillance areas. Real-time analysis of surveillance data allows for proactive alerting of security personnel to potential threats, significantly improving response effectiveness (Ansari and Singh (2020)). The demand for such systems has driven the development of intelligent surveillance systems that integrate advanced technologies including machine learning, computer vision, and data analytics to minimize human operator intervention and enhance security measures in urban environments. For instance, Sultani et al. (2018) developed a deep learning method to detect real-world anomalous events using weakly labeled training videos. Similarly, Chen et al. (2017) proposed a model based on optical flow statistics and a motion energy approach to identify abnormal behaviors in crowded environments. Another contribution is a two-step algorithm proposed by Wang et al. (2019) that uses generative networks to detect and accurately identify local and global abnormal events in surveillance footage. Figures 2.11a and 2.11b illustrate two examples of these surveillance methods.

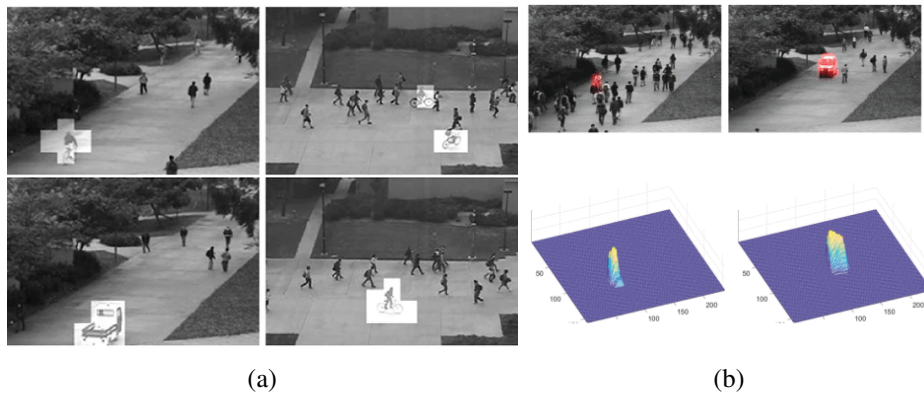


Figure 2.11: Examples of video-based surveillance applications for anomaly detection: (a) detection of abnormal behaviors (highlighted in white) using the UCSD Ped1 and UCSD Ped2 datasets (Li et al. (2014); Chen et al. (2017)); (b) local abnormal event detection proposed by Chen et al. (2017), illustrating detected anomalies highlighted in red (top row) along with corresponding detection intensity peaks (bottom row).

Forensic gait analysis has emerged as a promising tool within smart surveillance systems and criminal investigations. Using the unique nature of human locomotion, it helps identify individuals recorded during suspicious activities (van Mastrigt et al. (2018); Badiye et al. (2022)). Unlike clinical gait analysis, which is performed under controlled conditions with cooperative subjects, forensic gait analysis operates under more complex, uncontrolled scenarios, typically using low-resolution CCTV footage of uncooperative suspects in real-world environments. While gait analysis cannot be used as the sole method of identification due to variability and a lack of definitive scientific standardization, it can provide valuable supporting evidence when comparing a suspect's movements to those of a perpetrator captured on video. This technique has been used in legal systems worldwide, including in the United Kingdom and Denmark (Macoveciuc et al. (2019)), to link suspects to multiple crimes across different times and locations. Gait's unique value in forensic contexts lies in its noninvasive, contactless nature and the ability to observe it from a distance, even under suboptimal imaging conditions. Advances in computer vision and artificial intelligence have enabled the automation of gait recognition using markerless motion capture systems and video-based feature extraction. Modern algorithms can analyze sequences of frames to identify recurring gait patterns, enabling comparisons between suspects and perpetrators. This process, known as gait biometric analysis, is especially useful in surveillance scenarios where more intrusive biometric measures such as facial recognition or fingerprinting are not feasible.

Among the various datasets developed for the study of human motion in surveillance contexts (Li et al. (2015)), two examples are the ShanghaiTech Campus Anomaly Detection Dataset (Liu et al. (2018a,b)) and the Train Station Dataset (Zhou et al. (2012)) (see Fig. 2.12).



Figure 2.12: Example frames from datasets used in human motion analysis for surveillance applications. (a) Sample clip from the ShanghaiTech Campus Dataset (Liu et al. (2018a)) showing normal and abnormal activities captured under different conditions. (b) Visualization of pedestrians on the Train Station Dataset (Zhou et al. (2012)), recorded at New York’s Grand Central Station.

The ShanghaiTech Campus Anomaly Detection Dataset comprises video footage from 13 distinct scenes captured by surveillance cameras distributed throughout the ShanghaiTech University campus. It includes over 270,000 training frames and 130 annotated anomalous events, with pixel-level ground truth to enable fine-grained evaluation. As illustrated in Figure 2.12a, the dataset encompasses both normal and anomalous scenarios recorded under diverse lighting conditions and camera viewpoints, with a strong emphasis on human-centered anomalies. A dedicated subset, referred to as Human-related (HR) ShanghaiTech, contains 101 video clips curated specifically for the analysis of human-centric abnormal events. The Train Station Dataset provides a representation of real-world pedestrian dynamics in densely populated public environments. It consists of a 33-minute continuous video sequence recorded in New York’s Grand Central Station, capturing approximately 20,000 individual trajectories extracted through keypoint tracking. Due to frequent occlusions and scene clutter, these trajectories are often fragmented and reinitialized. Figure 2.12b illustrates the complexity of this dataset, reflecting the challenges inherent in analyzing human motion in crowded and unstructured environments.

2.3.4 Human Factors and Ergonomics

Human Factors and Ergonomics is a multidisciplinary field focused on optimizing the interaction between people and their environments, workstations, and tools. The main goal is to design systems that are safer, more comfortable, and more efficient by reducing injury risk and enhancing human well-being and performance.

In many industrial environments, work activities are often performed in non-ergonomic conditions, which may involve awkward postures, repetitive actions, excessive force, or prolonged static positions. Over time, such conditions have the potential to contribute to the development of musculoskeletal disorders, representing serious health and productivity concerns (Bortolini et al. (2020); Gómez-Galán et al. (2017)). Computer-aided ergonomic analysis, supported by motion

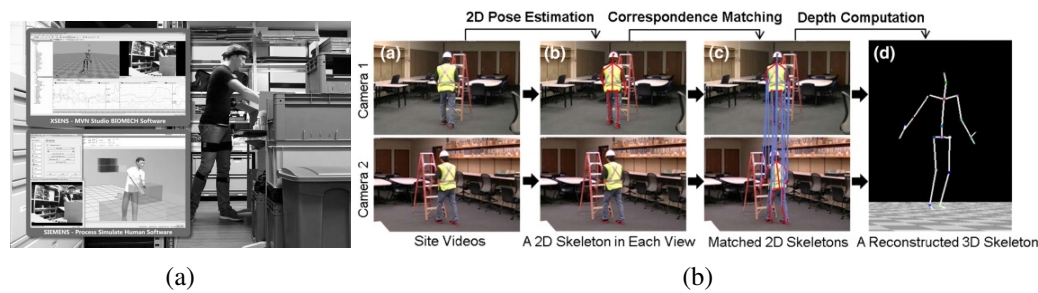


Figure 2.13: Examples of motion analysis techniques used in ergonomic assessments: (a) a laboratory setting where a worker's movements are monitored using two motion analysis software applications (Petrosyan et al. (2020)); (b) a sequence showing the reconstruction of a 3D skeleton from video input: 2D joint positions are estimated from one camera view and combined with depth data from a second view to create a 3D skeleton model (Han and Lee (2013)).

capture technologies and digital human models, facilitates the evaluation and simulation of workplace tasks to identify biomechanical risks and the redesign of systems to mitigate them. This enables the assessment of posture and task performance using established ergonomic metrics, thereby reducing the necessity for costly physical prototypes. This analysis also extends to the evaluation of external agents, such as industrial exoskeletons (Howard et al. (2019); Fox et al. (2019)), that physically interact with workers. By studying human motion in these collaborative contexts, it becomes possible to ensure that assistive technologies enhance physical tasks without introducing new biomechanical risks.

In addition, in the early stages of product and workplace design, it can be particularly beneficial to simulate how different body types interact with equipment or perform tasks. This is essential in industries such as automotive (Bhise (2024)) and aerospace (Landry (2021)), where the design of interfaces, vehicle interiors and assembly lines must consider human variability and biomechanical constraints. In vehicle manufacturing, for example, many assembly operations still involve manual tasks that greatly influence ergonomic outcomes. Incorporating ergonomic principles early in the product development process can reduce worker health risks and also improve economic outcomes, including productivity and product quality.

In the context of smart construction, for example, Jiang et al. (2020) introduced a cyber-physical system for safety management that integrates human motion monitoring through scene reconstruction, data sensing, and real-time processing, enabling proactive risk control and improving on-site worker behavior. Han and Lee (2013) proposed a vision-based framework for detecting unsafe worker behaviors on-site. The system involves identifying critical unsafe actions, extracting 3D skeletal data from construction videos, and using this information to recognize dangerous patterns of movement, highlighting the potential of motion analysis for accident prevention and real-time safety monitoring (see Figure 2.13b).

Among the datasets developed to support research in human factors and ergonomics, two examples are the AnDyDataset (Maurice et al. (2019a,b)) and the Manual Material Handling Dataset for Biomechanical and Ergonomics Analysis (Bassani et al. (2021a,b)). The AnDyDataset pro-

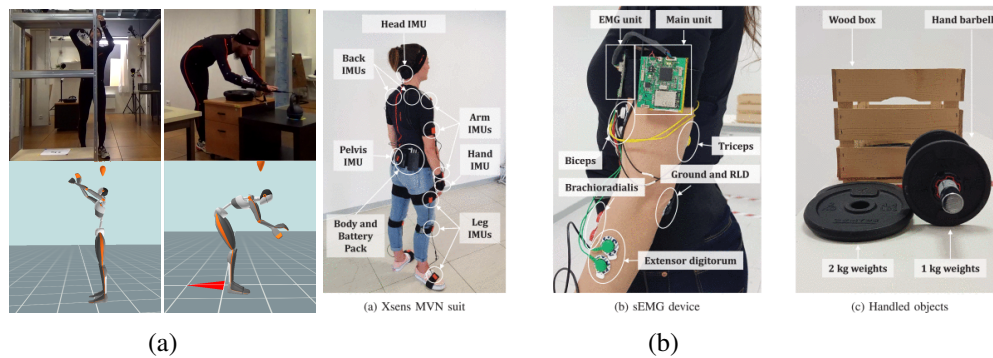


Figure 2.14: Examples from two ergonomics-related motion datasets. (a) Two frames from the AnDyDataset (Maurice et al. (2019a)) showing a participant performing different manual tasks: using a screwdriver and untying a knot. (b) Setup from the Manual Material Handling Dataset (Bassani et al. (2021a)): (a) wearable inertial system for kinematic data collection; (b) sEMG device for muscle activity recording; (c) objects handled during the experiments.

vides motion data from 13 participants performing repeated industry-like tasks such as screwing and load manipulation (see Figure 2.14a), annotated according to the Ergonomics Assessment Worksheet (EAWS), which is a standardized method aimed at improving workplace ergonomics through better task planning and design, widely adopted in industrial manufacturing. The dataset includes over 5 hours of multimodal recordings, combining wearable inertial sensors, marker-based optical motion capture, finger pressure data, and expert annotations. The Manual Material Handling Dataset contains 18.6 hours of motion and muscle activity data collected from 14 participants in a laboratory setting. As illustrated in Figure 2.14b, it combines full-body kinematics from 17 Inertial Measurement Units (IMUs) with upper-limb surface Electromyography (sEMG) data from 16 channels. The dataset is fully annotated with action labels and load information, supporting applications in ergonomic risk assessment, biomechanical analysis, and human activity recognition.

2.3.5 Robotics

The analysis of human movement plays a crucial role in the advancement of robotics and human-robot interaction (HRI). As robotic systems become increasingly embedded in social, industrial, and assistive environments, the ability to understand, predict, and respond to human motion is fundamental to enabling natural, safe, and effective interactions. Through detailed observation and interpretation of biomechanical patterns, such as gestures, facial expressions, and body posture, robots can be designed to better perceive human behavior and adapt accordingly.

One significant application is to facilitate intuitive and seamless communication between humans and robots (Rudenko et al. (2020b)). By recognizing non-verbal cues and movements, robots can respond appropriately in collaborative environments, such as smart factories, where coordination with human workers is essential (Jahanmahin et al. (2022); Baratta et al. (2023); Matheson



Figure 2.15: Example of a drawer-opening task: on the left, the action is performed by a human; on the right, the same action is executed by a robot through imitation. Source: [Bahl et al. \(2022\)](#)

[et al. \(2019\)](#); [Gui et al. \(2018\)](#)). For example, [Liu and Wang \(2017\)](#) propose a human-robot collaborative assembly system in which product assembly tasks are modeled as sequences of human motions. Using motion recognition techniques and a Hidden Markov Model to generate transition probabilities between actions, the system predicts human intentions during task execution. These predictions are then applied to enable smooth task-level cooperation, improving the flexibility and responsiveness of collaborative manufacturing systems.

Human motion analysis is a critical component of demonstration-based learning methodologies, facilitating the acquisition of novel skills by robotic systems through the observation and replication of complex human movements ([Seifert et al. \(2022\)](#)). This approach significantly reduces the need for manual programming and facilitates faster adaptation to dynamic environments. For instance, [Jang et al. \(2022\)](#) introduced an imitation learning system that learns from both demonstrations and corrective feedback, using diverse task representations, including human video recordings and natural language descriptions. Similarly, [Bahl et al. \(2022\)](#) introduced WHIRL (In-the-Wild Human Imitating Robot Learning), a one-shot learning approach in which robots learn manipulation tasks by observing unstructured third-person human demonstrations in real-world environments. This method extracts prior information regarding human intent to initiate robot policies and refines them through interaction. This process – depicted in [Figure 2.15](#) – uses an objective function, a sampling-based policy optimization strategy, and an exploration mechanism to improve sample efficiency.

In the field of healthcare, human gait analysis plays a crucial role in robotics. It contributes to the development of humanoid locomotion systems and the advancement of robotic rehabilitation platforms. The analysis of human walking patterns facilitates the development of robots that exhibit natural movement, adapt to irregular terrain, and maintain dynamic balance during movement. This is of particular importance for mobile robots operating in shared human environments, such as hospitals or warehouses, where physical interaction and close proximity to people necessitate fluid, predictable movement. In such contexts, motion prediction based on gait analysis enhances path planning and collision avoidance, ultimately improving the robot's situational awareness and safety ([Gui et al. \(2018\)](#); [Rudenko et al. \(2020b\)](#)).

In the domain of rehabilitation, robot-assisted gait training (RAGT) has emerged as a valuable tool, particularly for individuals recovering from neurological injuries such as stroke ([Er-](#)

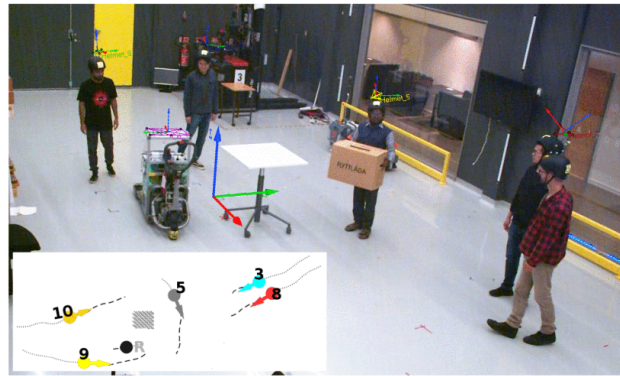


Figure 2.16: Environment configuration in the THÖR dataset (Rudenko et al. (2020a)). Participants, equipped with tracking helmets, and a mobile robot navigate toward their respective goals in a shared indoor space. All motion is recorded using a motion capture system, with an example trajectory visualization shown in the bottom left corner.

bil et al. (2018); Lamberti et al. (2021)). Conventional rehabilitation methodologies are characterized by their labor-intensiveness and dependence on therapist availability, making it difficult to deliver consistent, high-intensity training. Robotic gait trainers, including exoskeletons and treadmill-based systems with body-weight support, offer precise, repetitive, and adjustable training (Warutkar et al. (2022)). This approach has the potential to reduce the therapist’s workload while enhancing patient outcomes.

Among existing datasets that support research in perception, motion planning, and human-robot interaction, the THÖR dataset provides recordings of human motion trajectories and eye movement data collected in an indoor environment (Rudenko et al. (2020a)). It includes high-frequency annotations of human position, head orientation, viewing direction, social grouping, obstacle maps, and goal coordinates. In addition to behavioral data, the dataset contains sensor readings from a mobile robot equipped with 3D LiDAR, supporting multimodal analysis and integration. The data is characterized by low measurement noise, high temporal resolution, and a range of recorded human behaviors. The dataset also defines a set of metrics for the quantitative evaluation of motion trajectory data. As shown in Figure 2.16, the dataset encompasses visual examples of annotated human trajectories, viewing directions, and robot sensor data within a structured indoor environment.

2.3.6 Entertainment and Gaming

The entertainment and gaming industries have become some of the most dynamic fields to benefit from advances in human movement analysis. Technologies that capture, interpret, and replicate human motion are critical components of immersive and interactive experiences in video games, animated films, virtual reality (VR), and augmented reality (AR) applications (Zhu et al. (2024)).

The process of mapping human movement onto digital characters facilitates the creation of realistic animations that preserve the nuances of human expression and biomechanics. The ability

of motion capture technology to capture fine-grained gestures, postural variations, and facial expressions has been demonstrated to facilitate a more precise replication of human performance in digital storytelling (Sharma et al. (2019a); Lamberti et al. (2018)). Consequently, this technological advancement enhances emotional authenticity and realism in the generated content.

In game development, motion data is used to animate avatars that exhibit a high degree of realism. These avatars respond to player actions or environmental inputs with fluid and naturalistic movements, thereby enhancing the immersive experience of the virtual environment (Aguado-Delgado et al. (2018)). Furthermore, movement patterns can be analyzed to refine game mechanics, such as character locomotion, combat animations, or sports simulations. For instance, realistic foot placement and weight transfer are crucial for the credibility of sports games and fighting games (Li et al. (2024c); Oh et al. (2022)). Moreover, the advent of real-time body tracking has enabled full-body interaction in VR, thereby allowing users to navigate digital worlds using natural movements, which enhances immersion and intuitiveness.

In film and animation, motion analysis supports not only character animation but also pre-visualization, where directors plan complex action sequences using virtual characters before live filming. Technologies such as performance capture, which records body and facial motion simultaneously, have become integral to the production of high-end CGI-driven films (Wei (2022)).

Additionally, the development of algorithms for gesture recognition and pose estimation has led to new opportunities for natural user interfaces (NUIs) in entertainment systems (Shin et al. (2021); Mohamed et al. (2021)). These interfaces enable users to control applications without the need for physical controllers, relying instead on hand, body, or facial gestures.

Several datasets have been developed to support research in the field of human action and gesture recognition (Olugbade et al. (2022); Zhang et al. (2016)). In particular, within the domains of entertainment and video games, two particular datasets are of note: the G3D and G3Di. The G3D dataset (Bloom et al. (2015)) contains synchronized data captured from 10 subjects performing 20 gaming-related actions, such as punching, kicking, jumping, or steering a car, typical in gesture-based gaming systems like those using Microsoft Kinect. Expanding on this idea, the G3Di dataset (Bloom et al. (2015)) focuses on multiplayer interaction in virtual sports games. This dataset encompasses pairs of users engaged in activities such as boxing, football, volleyball, and table tennis, recorded while users played in real-time. As with G3D, it comprises detailed annotations, facilitating the study of interactive, multi-user action recognition in gaming contexts.

While G3D and G3Di are primarily tailored to gaming environments, the HaGRID – HAnd Gesture Recognition Image Dataset (Kapitanov et al. (2024)) offers broader applicability. Composed of over one million Full HD images across 33 gesture classes, HaGRID supports both image classification and detection tasks. Although well-suited for entertainment applications such as gesture-controlled gaming or augmented reality, it is also intended for use in video conferencing, smart homes, and automotive systems. The dataset's diversity in lighting, subjects, and environments makes it a robust resource for developing real-world hand gesture recognition systems.

An example from each dataset is illustrated in Figure 2.17, showing the wide variety of human motion and interaction styles represented across different application domains.

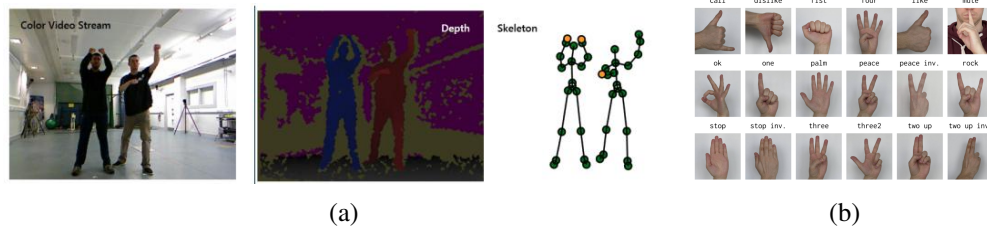


Figure 2.17: Two illustrative examples of datasets with potential applications in entertainment and gaming: (a) examples of synchronized color, depth, and skeleton data from a boxing game in the G3Di dataset (Bloom et al. (2015)); (b) visual examples of the 18 gesture classes included in the HaGRID dataset (Kapitanov et al. (2024)).

2.3.7 Traffic Safety

The analysis of human motion has the potential to contribute significantly to the advancement of traffic safety. By studying the fundamental aspects of human movement, researchers can develop a more profound understanding of critical domains, including accident prevention, pedestrian detection, and vehicle interaction. An understanding of the dynamics of human movement is beneficial for the design of safer road infrastructures, the implementation of effective traffic management systems, and the development of advanced driver-assistance technologies. The analysis of human motion patterns contributes to the development of predictive models for pedestrian behavior, which is a crucial aspect of the design of autonomous vehicles that are capable of anticipating and responding to human actions on roads. Furthermore, insights into human movement facilitate the development of ergonomic road designs that consider factors such as pedestrian crossing times, optimal visibility zones, and vehicle-pedestrian interaction spaces.

An illustrative example of such applications is the path generator proposed by Sarcinelli et al. (2019), which enhances decision-making processes for autonomous vehicles when dealing with pedestrians. Another area of interest is the detection of sleep-related fatigue and its potential impact on driving performance. A real-time sleep onset detection system, as proposed by Albu et al. (2008), employs an in-car camera to assess the driver's attention, thereby facilitating the prompt identification of fatigue. An additional critical concern is the automated regulation of airbags: Farmer and Jain (2007) proposed an integrated vision-based system for occupant classification and monitoring. The system is able to deactivate airbags under certain circumstances, thereby facilitating the safer utilization of the vehicle.

The advancement of research and development in this domain is significantly supported by the availability of context-specific datasets, such as the EuroCity Persons dataset (Braun et al. (2019a,b)) and the Infrastructure to Vehicle Multi-View Pedestrian Detection (I2V-MVPD) dataset (Khalifa et al. (2020); Khalifa (2020)). The EuroCity Persons dataset is characterized by its large scale and diversity, containing extensive annotations of pedestrians captured from a moving vehicle across different weather conditions, times of day, and urban environments. In contrast, the I2V-MVPD dataset integrates data from both static infrastructure-mounted cameras and vehicle dashboard cameras, enabling multi-view pedestrian detection and providing a valuable benchmark

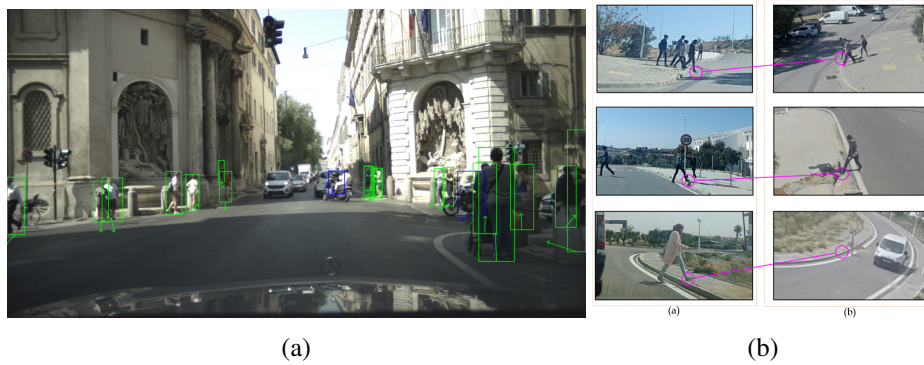


Figure 2.18: Illustrative examples from two datasets used for traffic safety: (a) an image from the EuroCity Persons dataset where pedestrians are annotated with tight green bounding boxes, ensuring precise localization in urban traffic scenes (Braun et al. (2019a)); (b) sample images from the I2V-MVPD dataset, showing pedestrian views from the vehicle perspective (left) and from the infrastructure-mounted camera perspective (right) (Braun et al. (2019a));

for motion analysis systems in traffic safety applications. Figure 2.18 illustrates examples from both datasets, highlighting the variety of viewpoints and environmental conditions represented.

2.4 Automated Tools for Human Motion Analysis

The field of human motion analysis has undergone significant advancements in recent years, primarily driven by the integration of automated tools and advanced computational techniques. These developments have introduced sophisticated algorithms and methodologies capable of performing detailed analysis and modeling of human movement, thus reshaping the landscape of the field. Automated motion analysis tools frequently include functionalities such as real-time analysis, which provides immediate feedback and facilitates live monitoring of movement, particularly beneficial in clinical evaluations, and sports performance contexts. Furthermore, these tools commonly support integration of motion data, whether acquired via native, integrated motion capture systems or imported from external data acquisition setups, enhancing flexibility across diverse research and application scenarios. Additionally, many tools offer sensor fusion capabilities, enabling the integration of data streams from multiple sensor modalities, such as optical cameras, inertial sensors, force platforms, and electromyography (EMG), thereby improving the accuracy, robustness, and comprehension of human motion assessments.

This section provides an overview of the current state-of-the-art in automated tools for human motion analysis. The review is based on a comprehensive research effort that examined tools cited in the scientific literature. In addition to this literature review, official websites of research laboratories specializing in motion analysis were consulted to gather further technical and contextual information. To guarantee the precision of the collected data, a survey was disseminated among the developers of the identified tools. Finally, this section proposes a classification framework for the surveyed applications, offering a structured perspective for understanding, comparing, and evaluating the capabilities of current automated tools in the domain of human motion analysis.

2.4.1 Selected Tools

The following paragraphs provide a concise description of the tools examined in this study, listed in alphabetical order. Tools marked with an asterisk (*) indicate those for which the survey received a response from the respective teams or individuals responsible for their maintenance, enabling verification and, where applicable, clarification of the collected information. The tools reviewed include: Adams, AnimatLab, AnyBody Modeling System, BTS Smart, Cortex, D-flow, Dartfish, DMAS – Digital Motion Analysis Suite, Kinovea, Kwon3D XP, MaxPRO, Mokka – Motion Kinematic and Kinetic Analyzer, MotionGenesis Kane, MSMS – Musculoskeletal Modeling Software, MVN Analyze, ODIN, OpenSim, OptoGait, SIMI Motion, Templo, Visual3D, and WINanalyze.

Adams (*)

Adams (Automated Dynamic Analysis of Mechanical Systems) is a commercially available software solution developed by MacNeal-Schwendler Corporation. It features an extensive library of models and facilitates the creation of new models or the modification of existing ones. The system functions within a closed architecture, which imposes limitations on available customization options. The accuracy of motion representation in Adams depends on the acquisition of motion data through an external motion capture system (Xu et al. (2017c); Na (2015)). The software facilitates three-dimensional analysis and simulations of human motion, providing a robust framework for conducting detailed investigations.

AnimatLab

AnimatLab (Cofer et al. (2010)), developed by NeuroRobotic Technologies, is an open-source tool designed as a versatile simulator for neuromechanical processes in skeletal animals. It is available in both a free standard edition and a professional edition, which requires a fee. AnimatLab facilitates the construction and modification of computational models that are developed in accordance with the laws of physics, and which respond to both external and internal signals. This process results in the generation of movements that are similar to those observed in biological systems. The modular architecture of AnimatLab provides a flexible and adaptable framework, enabling researchers to explore and analyze a broad range of biological phenomena (Young et al. (2020)).

AnyBody Modeling System

The AnyBody Modeling System, initially developed at Aalborg University in Denmark (Damsgaard et al. (2001, 2006)) and now commercially available from AnyBody Technology, is a tool for biomechanical analysis and musculoskeletal modeling. It facilitates the creation of detailed full-body models, enabling the simulation of internal body loads, including muscle activity, forces, and joint reaction forces (Rasmussen (2019)). The system incorporates a Managed Model Repository, which contains a comprehensive array of pre-existing models, and supports advanced features such

as inverse dynamics, force-dependent kinematics, and posture and motion prediction. Capable of three-dimensional analysis, AnyBody allows users to define environments by specifying external forces and boundary conditions. Users can create custom postures and motions with the AnyScript language or import recorded motion data for simulation.

Biomechanical ToolKit (*)

The Biomechanical ToolKit (BTK) (Barre and Armand (2014)) is a free, open-source, and cross-platform framework developed to read, visualize, and process biomechanical data from both standard and proprietary file formats, with particular emphasis on the C3D format. It addresses the biomechanical community's need for fast and flexible access to raw data, such as 3D marker trajectories, EMG signals, and force plate recordings, independent of the motion capture system used. BTK enhances data interoperability across laboratories, overcomes the constraints of proprietary software, and bridges the gap between motion analysis platforms and widely adopted scientific computing environments such as Python and Octave. By promoting standardized access and processing, it fosters collaboration, reproducibility, and methodological advancement in areas like human motion analysis and musculoskeletal modeling.

BTK also provides 3D visualization tools for exploring time-series data, enabling the detailed analysis of body segments, marker trajectories, and external forces. These capabilities support a wide range of applications, including clinical gait analysis (Kidziński et al. (2019)), biomechanical research (Wang (2022)), and sports performance and injury assessment (Ortiz-Padilla et al. (2022)).

D-flow

D-Flow is a motion analysis tool that facilitates the integration and synchronization of data from multiple hardware devices within a unified interface, allowing real-time feedback and data analysis (Geijtenbeek et al. (2011)). The system supports the integration of various data sources, including infrared motion capture cameras, instrumented treadmills, inertial measurement units, and electromyography (EMG) sensors (Roeles et al. (2018)). D-Flow's real-time data processing facilitates the integration of subjects into a dynamic feedback loop, wherein their behavior is captured by multi-sensory input devices and visual feedback is delivered through output devices. The software incorporates visual modules and an embedded scripting language, which facilitates the development of applications capable of controlling hardware devices and virtual environments. Furthermore, D-Flow integrates a 3D game engine, enabling the development of interactive virtual reality environments that offer immersive experiences closely resembling real-life scenarios.

Dartfish

Dartfish is a software platform designed for two-dimensional motion analysis, processing video input from multiple sources to extract kinematic parameters such as spatial positioning, range of

motion, velocity, frequency, and duration of movement. This data supports the examination of both spatial and temporal aspects of human motion across a range of tasks.

The software has been used in the analysis of limb movements, including hip, knee, and shoulder movements, as well as in the evaluation of task performance in contexts such as walking, lifting, and athletic activities such as sprinting and race walking. In clinical practice, Dartfish enables real-time visual feedback, for example to compare shoulder mechanics before and after surgical procedures (MacDermid et al. (2014)). It also supports the creation of visual reports with annotated media, contributing to clearer communication with patients and informing therapeutic decisions (Kassay et al. (2021)). Its applications span healthcare, rehabilitation, sports, and workplace ergonomics.

DMAS – Digital Motion Analysis Suite

The DMAS (Digital Motion Analysis System) is a digital video motion capture suite that is capable of supporting up to 16 synchronized uncompressed digital video inputs. The system incorporates automated tracking capabilities, including auto labeling and 3D data reconstruction, with the objective of optimizing the analysis process (Lujan et al. (2005); Graup et al. (2018)). The system provides real-time low-latency marker tracking, enabling precise marker capture, and allows for synchronization of analog and digital data from external devices, facilitating integration with various measurement systems. Furthermore, the DMAS provides a Python interface, enabling customized data analysis. Additionally, multimedia reporting tools and integrated video are available for report creation. The software suite also includes a standalone high-speed digital image correlation package, data analysis tools, and an editor tool to build flexible and editable models.

Kinovea

Kinovea is a video annotation tool designed for sports analysis (Nor Adnan et al. (2018)). It offers utilities for capturing, slowing down, comparing, annotating, and measuring motion within videos. This free and open-source software facilitates the creation of customized tools through its tool framework, thereby enabling advanced functionality. Users have the ability to customize features such as the human model tools and incorporate tracked points, computed points, distance measurements, and angle measurements (Hisham et al. (2017)). The software incorporates a lens calibration step to compensate for lens distortion across the coordinate system, thereby enabling the extraction of linear and angular kinematics from point trajectories or multi-point objects. Additionally, Kinovea facilitates the capture and recording of camera streams, as well as the synchronization of videos with heterogeneous frame rates.

KWON3D XP (*)

Kwon3D XP is a software platform designed for integrated analysis in fields such as robotics, rehabilitation, ergonomics, human locomotion, and sports biomechanics (Park et al. (2005)). It

supports multi-camera video capture, ground reaction force measurement, analog signal acquisition, system calibration, data analysis, and report generation. The system allows simultaneous acquisition from up to nine cameras and includes 2D/3D auto-tracking features to streamline the digitization process. Its data processing capabilities include kinematic analysis, force plate data interpretation, inverse dynamics, angular momentum computation, and the definition of custom joint angles (Liu and Liu (2021); McGuire et al. (2024)).

Max Suite (*)

Max Suite is a comprehensive set of tools for 3D motion capture and analysis, composed of MaxPRO, MaxA, and MaxTRAQ, and is used in a variety of disciplines including science, biomechanics, sports, performance analysis, neuroscience, gait analysis and rehabilitation (McCauley et al. (2012)). MaxPRO serves as the core motion capture platform, compatible with both dedicated MoCap cameras and commercially available video cameras. Designed for flexibility, it supports integration with a wide range of hardware, making it adaptable to various applications and budget levels. The system offers advanced features typically found in high-end motion capture solutions, but in a more accessible and cost-effective format.

Complementing MaxPRO, MaxA and MaxTRAQ provide analysis and tracking functionalities. MaxA supports basic kinematic analysis with tools for graphing, file comparison, and custom scripting via the integrated MyBASIC2 language (Leporace et al. (2016)). It handles C3D, MOC, and ASCII file formats, and includes an import wizard for TSV and CSV files. MaxTRAQ 3D enables affordable 3D motion tracking using both marker-based and markerless approaches, making it suitable for contexts where markers are impractical—such as sports environments or field studies with animals. It reconstructs 3D coordinates from 2D video data using static or optional dynamic calibration, and supports export to C3D or ASCII formats for further processing in MaxA or other compatible software.

MotionGenesis Kane (*)

MotionGenesis Kane is a commercial software product designed to analyze forces and motion using symbolic geometry. It provides symbolic math capabilities and efficient code generation, allowing users to visualize and manipulate symbolic equations related to kinematics, statics, and dynamics. The software supports multi-body force and motion simulation and provides tools for analyzing kinematics, forces, energy, mass, and inertia properties. MotionGenesis Kane also generates MATLAB, C, and Fortran code for real-time applications. It is applicable to various fields including biomechanics, robotics, projectile motion, pendulum dynamics, aerospace spin stability, angular momentum, gyroscopes, and 2D/3D geometry (Flanagan (2024); Heinrich et al. (2023)).

MSMS – Musculoskeletal Modeling Software (*)

MusculoSkeletal Modeling Software (MSMS) is a freely available application developed for modeling and simulation of neural prosthetic systems (Khachani et al. (2007)). It allows users to create dynamic models of human and prosthetic limbs, along with their task environments (de Sousa et al. (2019); García Montaña et al. (2017)). Simulations can be run on a stand-alone computer to design and test neural control systems, or integrated into virtual reality environments where simulated limbs can be interactively controlled by human or animal subjects. MSMS includes tools for creating, editing, and visualizing 3D musculoskeletal models and supports the import of models from third-party software such as SIMM and SolidWorks. The platform allows models to be animated using motion capture data or simulation output, and provides physics-based simulation capabilities to realistically represent limb motion and interaction with the environment.

ODIN

ODIN, part of the Coda Motion 3D Analysis System (Charnwood Dynamics Ltd., Leicestershire, UK), is a software platform designed for the analysis and management of clinical motion data. It automates the generation of standardized gait reports that include both kinematic and kinetic parameters, and supports the complete workflow of three-dimensional motion analysis - from initial data acquisition using various sensors and measurement devices, through real-time monitoring, to long-term data storage and retrieval. The software also enables the analysis of spatio-temporal variables, supporting detailed assessment of motion patterns.

ODIN includes a Software Development Kit (SDK) that enables real-time processing of fully labeled 3D marker streams within custom applications. It supports integration with output systems such as virtual and augmented reality displays, and provides a graphical configuration interface for adding devices such as analog inputs, force platforms, and EMG systems. ODIN is used in a variety of fields, including kinesiology, rehabilitation, sports science, and clinical gait analysis, where precise motion assessment is essential for research, diagnosis, and performance optimization (Zhongyang et al. (2017); Song et al. (2018); Ditroilo et al. (2023)).

OpenSim (*)

OpenSim is an open-source software platform for modeling and simulating motion and interaction in humans, animals, robots, and mechanical systems (Delp et al. (2007)). It includes a graphical user interface for visualizing models and generating simulations, and supports the creation and manipulation of complex systems composed of rigid bodies, joints, constraints, muscles, and actuators.

The platform provides a library of validated musculoskeletal models for use in research, education, and prototyping. OpenSim allows the import of experimental data for visualization and analysis, including marker trajectories, joint kinematics, and external forces. It supports a range of simulation tools such as inverse dynamics, static optimization, and forward dynamics, enabling applications in areas such as biomechanics, rehabilitation, clinical research, sports science, and

ergonomics (Pizzolato et al. (2016); Kim and Zhang (2016); Zhou et al. (2021); Trinler et al. (2019); Mahadas et al. (2019)). The software also offers a scripting interface for MATLAB and an extensible API, allowing advanced customization and integration into broader research workflows.

OptoGait

OptoGait is a modular photoelectric walkway for high-precision measurement of locomotor spatiotemporal parameters in gait, running, and jump tasks (Jaén-Carrillo et al. (2020); Lienhard et al. (2013)). The system comprises parallel transmitting and receiving bars populated with dense infrared emitters and detectors that register beam interruptions caused by foot contacts and aerial phases; these events are processed to derive quantities such as contact and flight times, stride length, cadence, and inter-limb asymmetries.

The embedded OptoGait software functions as the system's control and analytics hub: users configure and execute standardized tests (e.g., jump, reaction, running), while the platform performs real-time processing and automatically synchronizes numerical data with video, including frame-by-frame playback down to still images. Longitudinal management is supported through persistent storage of sessions and rapid cross-session comparison; recent releases add interface refinements, biofeedback/second-monitor fixes, and external integrations for combined analyses.

SIMI Motion

Simi Motion is a software platform for capturing and analyzing human motion in two and three dimensions. Unlike infrared-based systems, it employs image-based techniques, using high-resolution industrial cameras capable of synchronized video capture at high frame rates. The system supports both marker-based tracking, by calculating marker positions, and markerless tracking methods, including pattern recognition and silhouette analysis.

The platform offers an end-to-end workflow for motion analysis, including synchronized camera capture, data logging from external devices, and tools for data processing, editing, and analysis. It enables the tracking of joint centers with or without physical markers and calculates key motion parameters such as position, velocity, acceleration, and movement trajectories. Data can also be exported to other analysis software, supporting integration into existing research or clinical environments.

In the context of sports performance, Simi Motion is particularly suited for analyzing complex movements, such as jumps, throws, or sprints, providing detailed biomechanical insights that can support training optimization, injury prevention, and technique refinement (Fischer and Kibele (2016); Trasolini et al. (2022); Ammar et al. (2015)).

Templo (*)

Templo is the core software of the modular CONTEMPLAS motion analysis system, offering a flexible and customizable platform for capturing and analyzing human movement. Its functionality can be extended through the integration of additional hardware, such as high-speed cameras,

force plates, pressure sensors, and EMG systems, and software modules, including markerless tracking to streamline analysis without compromising accuracy (Widhalm et al. (2024); Greisberger et al. (2019)). This software supports synchronous multi-camera setups, enabling motion capture from multiple viewpoints, and is compatible with a wide range of external measurement devices. Templo is applied in sports performance analysis, including activities such as volleyball, swimming, running, and sprint start assessment, where precise movement evaluation can inform training and technique (Bohigas et al. (2021); Baritz (2019); Baritz and Mosoi (2017)). In health and clinical contexts, it is used for gait and posture analysis, functional movement screening, and ergonomic evaluations (Katashev et al. (2014)), supporting evidence-based decision-making across disciplines.

Visual3D (*)

Visual3D is a software platform for the analysis and modeling of three-dimensional motion capture data in research and clinical contexts. It provides a structured workflow for managing motion capture datasets, optimizing marker placement and definitions to streamline data collection. The captured data is applied to biomechanical models for a detailed computation of joint angles, forces, powers, velocities, and accelerations, using built-in mathematical libraries. Users can also define custom mathematical expressions and automate analysis routines. The software supports integration with a wide range of motion capture systems, synchronized analog inputs, and instrumented force measurement setups. It includes automation features through processing pipelines and supports multiple file formats for interoperability with external modeling and analysis tools. Visual3D offers advanced signal processing, including real-time analysis, kinematics, inverse dynamics, and biofeedback, making it applicable to areas such as biomechanics, rehabilitation, and sports performance evaluation (Layton et al. (2021); Gulgin et al. (2018)).

WINalyze (*)

WINalyze is a video-based motion analysis software that enables automatic tracking of multiple objects within video files and provides outputs in a variety of formats. It supports data integration through the import and export of commonly used file types, facilitating compatibility with external applications and workflows.

The software comprises a comprehensive set of motion analysis tools, providing linear and 3D calibration for accurate measurements. It generates detailed outputs, including object trajectories, angles, distances, derivatives (velocity and acceleration), angle projections on all anatomical planes, center of mass estimations, inverse dynamics, impulses, forces, angular momentum, and torsional moments. WINalyze is particularly suited for biomechanics research, with capabilities for customizing predefined body models to suit specific experimental setups and analysis needs (Wright et al. (2022, 2020); Ellabany and Attaallah (2015)).

Xsens MVN Analyze (*)

Xsens MVN is a motion capture system that combines wearable inertial sensors with a software engine that integrates sensor data using biomechanical models to estimate segment positions and orientations (Roetenberg et al. (2009)). The system is available in two software variants tailored to different application areas: MVN Animate, for 3D character animation in film and games, and MVN Analyze, for human motion measurement in areas such as ergonomics, sports, and research.

Xsens MVN Analyze supports full-body kinematic recording, real-time visualization of joint angles, segment orientation and center of mass, using a biomechanical model to guide the analysis (). It offers compatibility with various hardware configurations and multiple calibration options, providing flexibility for different setups. The software offers two processing modes: real-time (frame-by-frame) and reprocess HD, which uses a wider temporal window to improve the accuracy of orientation and position estimates. Motion data can be exported in multiple formats (e.g., FBX, BVH, XLS, C3D, MP4) for integration with third-party tools such as AnyBody, Visual3D, and Siemens PLM, for biomechanics, ergonomics, and sports (Bennett et al. (2023); Lee et al. (2020); Mavor et al. (2020)).

As discussed throughout this Section, a variety of tools have been developed for the analysis of human movement, each offering distinct advantages depending on the application context. Figure 2.19 provides a visual overview of some of these tools, illustrating the diversity of available solutions for human motion analysis.

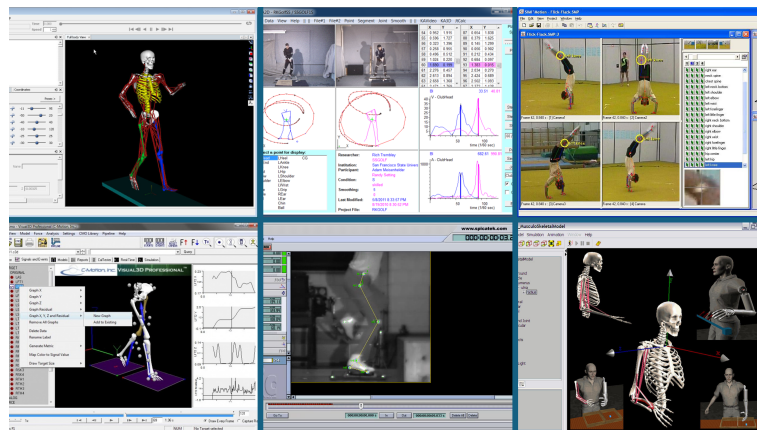


Figure 2.19: Illustrative examples of human motion analysis tools.

2.4.2 A Classification Framework

In order to achieve a systematic classification approach, this study establishes a robust framework comprising a set of key features, which serve to characterize the different software applications under examination. The framework organizes these features into five main groups, namely: *generic*

features, which encompass basic aspects of each tool; *modeling* capabilities, describing the software's capability for creating and manipulating biomechanical models; *data acquisition*, which highlights the mechanisms employed to gather motion data; and lastly, *analysis* and *simulation*, focusing on the software's ability in interpreting and simulating movement patterns. The following sections provide an in-depth examination of each feature, elucidating its significance and role in differentiating these software solutions for motion analysis.

Generic Features

These generic features encompass essential aspects such as version history, origins, openness, and platform compatibility, thereby facilitating an investigation into the software's adaptability and accessibility. The analysis of these key characteristics aims to provide a comprehensive understanding of the capabilities, origins, and overall maturity of each application.

- **Version History and Release Date:** provides crucial insights into the maturity and activity of the software, providing information on the latest version and release date. By understanding the currency of the application, researchers can more effectively evaluate its ongoing development and the potential for future advancements.
- **Academic Origins and Seminal Paper:** this feature considers the origins of the tool and explores whether it emerged as an academic project. When applicable, this information also reveals which seminal paper was the basis for the development of the tool, thus enhancing one's understanding of its roots and scientific foundations.
- **Software Availability:** this feature provides information regarding the accessibility of the software, distinguishing between freely available offerings and commercial products. For commercial options, it also reveals whether a trial version is available, allowing researchers to evaluate the tool's capabilities and make informed decisions.
- **Architectural Openness and Extensibility:** this feature exemplifies the software's architectural openness, identifying the mechanisms that enable the tool to be extended with custom functionalities. It explores the availability of key resources, including open-source foundations, software development kits (SDKs), application programming interfaces (APIs), and plugins.
- **Platform Compatibility and Accessibility:** this feature specifies the operating systems on which the tool can be installed and executed, highlighting its adaptability to diverse computing environments. This ensures that the software can be accessed and utilized across different of platforms, thereby enhancing its accessibility and usability.

Modeling Features

The evaluation of a tool's modeling capabilities is of particular significance as they can facilitate the creation, customization, and accessibility of a diverse range of models, including those representing the entire body or specific body parts. The following aspects are included in this group:

- **Create New Models from Scratch:** This feature assesses the tool's ability in generating novel models from scratch (expected values: yes/no). The capacity to create tailored models empowers researchers to adapt simulations to specific scenarios, enabling precise and focused investigations.
- **Edit Existing Models for Customization:** Adapting and fine-tuning existing models holds immense significance in the modeling process. This feature indicates the tool's capability to modify and refine pre-existing models, facilitating the customization of simulations to suit particular research requirements (expected values: yes/no).
- **Library of Pre-built Models:** This feature reveals if the software houses a repository of pre-built models, enriching users' options for analyses and simulations (expected values: yes/no). The availability of such a library simplifies the modeling process, fostering efficiency and versatility in research activities.

Data Acquisition Features

The functionality of data acquisition features is indicative of the applications' capacity to collect diverse and essential data for accurate evaluations. The following key aspects are explored by this group of features:

- **Motion Acquisition:** This feature specifies the type of motion capture system integrated into the tool (expected values: native, external, or absent). Native systems are embedded within the software, while external systems require integration with external hardware or devices to capture motion data.
- **Analog Data Sources:** This feature enumerates, if available, all the sources of analog data that the tool can import (e.g., C3D - motion data, EMG, ECG, force plates, etc.). The ability to incorporate diverse analog data streams enriches the analysis process by providing a more comprehensive understanding of movement patterns.
- **Data Fusion:** Synchronization of multiple signals is pivotal for precise evaluations of motion. This feature clarifies whether the tool supports data fusion, allowing the simultaneous integration of multiple data streams to enhance the accuracy and reliability of the analysis (expected values: yes/no).

Analysis Features

The Analysis set of features systematically examines basic aspects of motion analysis, with a primary focus on two critical dimensions: the dimensionality in which the analysis takes place and the timing of the analysis process. This group comprises the following aspects:

- **Dimensionality:** This feature indicates the dimensionality in which motion analysis takes place, specifying whether it occurs in 2D, 3D, or both (expected values: 2D, 3D, or 2D/3D). The dimensionality directly influences the depth and accuracy of the insights gained, with each dimension offering unique perspectives on the subject's movements.
- **Timing of Analysis:** The timing of the analysis profoundly impacts research outcomes, and this feature serves to reveal the mode in which analysis is conducted - whether in real-time, offline, or both (expected values: real-time, offline, or both). The real-time approach facilitates immediate feedback and adaptability, while offline analysis grants researchers the freedom to conduct comprehensive investigations, with no time constraints.

Table 2.1: Feature matrix of surveyed human motion analysis tools to the proposed framework, grouped by generic, modeling, acquisition, and analysis features. *Note: “*” indicates that a trial license is available.*

Tool	Generic Features			Modeling Features			Data Acquisition Features			Data Analysis Features	
	Release (Year)	License	Architecture	Pre-built Models	Model Creation Support	Customizable Models	Motion Acquisition	External Analog Inputs	Multimodal Data Fusion	Dimensionality	Analysis Mode
Adams	(2014)	Commercial	Proprietary	✓	✓	✓	external	–	–	3D	Offline
AnimatLab	2.1.5 (2016)	Commercial*	Open	✓	✓	✓	absent	–	✗	3D	Offline
AnyBody Modeling System	8.1.1 (2025)	Commercial*	Proprietary	✓	✓	✓	external	EMG; force plate	✓	3D	Offline
Biomechanical Toolkit	0.6.2 (2016)	Free	Open	✗	✗	✗	external	EMG; force plate	✓	2D/3D	Offline
D-flow	3.28.0 (2018)	Commercial	Proprietary	✗	✗	✓	native; external	EMG; force plate	✓	2D/3D	Both
Dartfish	11.7.310 (2025)	Commercial*	Open	✓	✓	✗	native	force plate	✗	2D	Offline
DMAS	N/A (2005)	Commercial	Open	✓	✓	✓	native	EMG; force plate	✓	3D	Both
Kinovea	2023.1.2 (2023)	Free	Open	✗	✗	✗	native	–	✗	2D	Offline
Kwon3D XP	4.1 (2004)	Commercial*	Proprietary	✓	✓	✓	native	force plate	✓	2D/3D	Offline
Max Suite	1.6.6.5 (2018)	Commercial*	Open	✗	✗	✗	native	EMG; force plate	✓	2D/3D	Offline
motionGenesis Kane	6.5 (2025)	Commercial*	Open	✓	✓	✓	absent	–	✗	2D/3D	Real-time
MSMS	0.9.4 (2010)	Free	Proprietary	✓	✓	✓	external	EMG	✓	2D/3D	Both
ODIN	1.01 (2013)	Commercial	Open	✗	✓	✓	external	EMG; force plate	✓	2D/3D	Both
OpenSim	4.5 (2024)	Free	Open	✓	✓	✓	external	EMG; force plate	✓	3D	Both
SIMI motion	8.5 (2012)	Commercial	Proprietary	✓	✓	✓	native	EMG; force plate; EEG	✓	2D/3D	Offline
Templo	V25.0 (2025)	Commercial*	Proprietary	✗	✗	✗	native	EMG; force plate	✓	2D	Offline
Visual3D	2025.05.1 (2025)	Commercial*	Open	✓	✓	✓	external	EMG; ECG; force plate	✓	3D	Both
WINalyze	2.7 (2016)	Commercial	Proprietary	✓	✓	✓	native	EMG; ECG; force plate	–	2D/3D	Offline
Xsens MVN Analyze	2024.2 (2024)	Commercial*	Proprietary	✗	✓	✓	native	own data files	✓	2D/3D	Both

2.4.3 Main conclusions

Table 2.1 shows a clear segmentation of the ecosystem into three groups: (i) video-centric 2D analysis tools (e.g., Dartfish, Templo, Kinovea), (ii) acquisition/processing suites tied to specific hardware or workflows (e.g., D-flow, Xsens MVN Analyze, SIMI Motion, WINalyze, DMAS, ODIN), and (iii) musculoskeletal modeling and simulation platforms (e.g., OpenSim, AnyBody, motionGenesis Kane). This segmentation closely aligns with the four groups in the proposed framework: generic, modeling, acquisition, and analysis features. Tools tend to be strong in one or two groups and ‘adequate’ or intentionally minimal in the others, rather than being uniformly feature-rich across all categories.

Licensing and openness. Most solutions are commercial, with a significant subset offering trial licenses (marked Commercial*). Free and open-source options, such as OpenSim, BTK, Kinovea, and AnimatLab, play a strategic role as low-cost entry points for teaching and prototyping. They also serve as interoperability bridges for pipelines that need to extract, transform, and reuse motion and analog data, such as C3D, EMG, and force plates. Openness translates into reproducibility and easier integration with scientific computing environments. In contrast, proprietary solutions typically offer more advanced tools, technical support, and integrated clinical/industrial workflows.

Modeling capabilities. Pre-built model libraries and model creation/customization are consistently available in the dedicated modeling tools (OpenSim, AnyBody, motionGenesis), but are understandably limited or absent in purely 2D video analysis software. For research questions requiring hypothesis testing on internal loads, joint kinetics, or design-space exploration, platforms with full modeling stacks (pre-built libraries *plus* scriptable customization) are preferable; 2D video tools remain valuable for rapid kinematic screening, education, and communication with end users.

Acquisition and data fusion. Hardware integration is a major differentiator. Suites designed around labs or sensor ecosystems (D-flow, SIMI Motion, Visual3D, ODIN, Xsens MVN Analyze, DMAS) tend to support synchronized acquisition and multimodal fusion of video, IMUs, EMGs, and force plates, and therefore enable closed-loop protocols, biofeedback, and richer feature sets. In contrast, video-only tools usually offer limited or no analog integration. In practice, labs that already own cameras, treadmills, or IMU systems will benefit from software with native drivers and time-sync utilities.

Dimensionality and analysis mode. A practical split persists between 2D-only platforms and 2D/3D (or 3D-only) stacks. If the research requires joint moments, inverse dynamics, or multi-planar control, 3D tools are essential. With respect to timing, *real-time* support is still the exception; most platforms operate offline or support both modes with constraints (e.g., reduced accuracy or reduced pipeline complexity in real time). For biofeedback, interactive rehab, or human-in-the-loop experiments, frameworks explicitly supporting real-time/“both” (e.g., D-flow, Visual3D, Xsens MVN Analyze, DMAS, OpenSim in specific configurations) are advantageous.

Guidance for tool selection. In light of these findings, selection should be driven by the dominant column(s) of the framework for the target study: (i) *Musculoskeletal modeling and hypothesis*

testing: OpenSim (free and extensible) or AnyBody (commercial) when validated libraries and advanced solvers are required. (ii) *Lab-centric multimodal acquisition and reporting*: Visual3D, SIMI Motion, ODIN, or DMAS, depending on the existing sensor stack and report/automation needs. (iii) *Interactive/real-time protocols or VR feedback*: D-flow (integration/scripting) or platforms with explicit real-time pipelines. (iv) *Field work and wearable sensing*: Xsens MVN Analyze (IMU-first, turnkey). (v) *Education, low-cost screening, and quick 2D kinematics*: OpenSim, Kinovea, Dartfish, Tempo, with BTK as a companion for data handling and reproducibility.

In conclusion, the proposed framework synthesizes the declared capabilities of the surveyed tools at the assessed versions. It identifies the fundamental trade-offs, including the cost versus openness, the acquisition depth versus modeling detail, and the offline robustness versus real-time responsiveness. It provides a practical framework for aligning the selection of tools with the experimental design. For instance, in an academic context, motion data (e.g., C3D/CSV files) can be imported with BTK, basic checks can be run in Visual3D, modeling and kinetic analysis can be performed in OpenSim, using scripts. On the other hand, in a clinical gait laboratory, equipped with multi-camera motion capture, force plates, and electromyography (EMG), Visual3D or SIMI Motion (or ODIN) provide synchronized acquisition and standardized reporting. When necessary, biofeedback or virtual reality can be added via D-Flow or Visual3D real-time modules. Furthermore, exported trials can be directed to OpenSim or AnyBody for advanced musculoskeletal analysis without compromising the established clinical workflow.

2.5 Summary

This chapter presents a comprehensive overview of the field of human motion analysis, encompassing its theoretical foundations, the analytical techniques employed for its assessment, its practical applications, and some sample datasets that serve as the foundation for this interdisciplinary research area. The analysis of human motion has been identified as a critical area of study with wide-ranging implications across various disciplines. The introductory section on human motion described the fundamental procedures utilized in human motion analysis systems and provided a review of related work in this field. The Review of Human Motion Analysis and Simulation Tools (Section 2.4) presented a comprehensive analysis of the automated tools currently in use within this field. A framework was introduced with the objective of categorizing and describing these tools, thereby providing clarity and structure. This framework provides a valuable reference for researchers and practitioners seeking to gain an understanding of the diverse range of available tools. The section on application areas of human motion (Section 2.3) demonstrated the wide range of fields in which human motion analysis is used. From the domains of sports and health-care to those of surveillance and robotics, the section emphasized the importance of human motion analysis in a multitude of settings.

In conclusion, this chapter has provided a comprehensive and meticulous examination of human motion, establishing a connection between theoretical comprehension and practical applications, and reinforcing the vital function of human motion analysis across a range of disciplines.

Chapter 3

Human Gait – A Comprehensive Overview

The comprehension of human gait involves the recognition of it as a distinct biomechanical process shaped by numerous interacting factors, both intrinsic and extrinsic to the individual. The distinctive characteristics of an individual's gait are primarily determined by their physical attributes, including but not limited to skeletal structure, musculature, joint mobility, stride length, and walking speed (Stotz et al. (2023)). Indeed, these biomechanical attributes demonstrate a high degree of consistency within individuals and variability between different people. For instance, stride length and joint angles are predominantly determined by an individual's physical build and musculature, resulting in substantial differences across sex, age, and ethnicity groups (Rowe et al. (2021); Hill et al. (2020)). Research has demonstrated that taller individuals typically exhibit longer strides and higher walking speeds (Moe-Nilssen and Helbostad (2020)). In contrast, age-related speed declines are often attributed to reduced stride lengths rather than cadence (Kirtley (2006)).

Beyond these inherent physical traits, gait is also subject to transient influences such as emotional states (Xu et al. (2024)), carried loads (Wang et al. (2023a)) and footwear (Derlatka and Parfieniuk (2023); Price et al. (2021)), among others. While emotional factors are considered temporary, they have the capacity to induce alterations in muscle tension and movement patterns, thereby temporarily modifying one's gait. Furthermore, external conditions, such as the act of carrying a load, have been demonstrated to substantially influence gait parameters, including walking speed and stride consistency. However, the core biomechanical pattern remains stable enough to serve as a robust biometric marker, as long as these transient effects are taken into account in gait characterization methods.

The chapter is structured to offer a theoretical introduction to human gait analysis, presenting the biomechanical foundations, outlining the key gait parameters alongside reported normative values, and concluding by explaining how these parameters are measured. Section 3.1 provides an introductory overview of the field of human gait analysis. Section 3.2 sets out the biomechanical foundations and terminology, detailing the gait cycle and its phases and the planes of motion. Section 3.3 delineates the principal categories of gait parameters, namely anthropometric,

spatio-temporal, kinematic, kinetic, electromyography-based and other derived measures. Section 3.4 compiles and discusses normative values, with dedicated subsections for spatio-temporal and kinematic parameters. Finally, Section 3.5 describes how these parameters are measured in practice, reviewing sensing technologies with a distinction between wearable and non-wearable systems.

3.1 Human Gait Analysis

Gait can be defined as the coordinated, cyclic motion of the body's limbs, primarily the legs and arms, used to move from one point to another, a process known as locomotion. This movement involves a rhythmic and repetitive sequence in which limbs alternate roles, with one limb providing support and stability while the other advances to the next supportive position, thereby continuously transferring the body's weight forward. This cyclic motion creates a complex pattern of joint movements organized into a kinematic chain, characterized by precise coordination and timing. Multiple components interact to make gait possible, including the positioning and alignment of bones, the range of motion available at joints, neuromuscular activity coordinating muscle contractions, and biomechanical factors (Silva and Stergiou (2020)). Despite the capacity of humans to manifest a wide array of gaits, the predominant forms are walking and running. The distinction between these two modes is characterized by divergent patterns of foot contact with the ground, particularly the temporal organization of footfalls and the sequence of foot movements.

Gait analysis is the systematic observation, recording, analysis, and interpretation of human locomotion. It encompasses qualitative and quantitative approaches, with the emphasis tailored to the requirements of each specific application domain.

As discussed in Section 2.3, the field of human motion analysis, particularly gait analysis, has found applications in a wide range of disciplines. In clinical contexts, gait analysis is vital in detecting and monitoring disorders related to motor function. It supports the diagnosis and treatment of conditions such as multiple sclerosis (Comber et al. (2017); Gholami et al. (2017)), Parkinson's disease (Guo et al. (2022); Zanardi et al. (2021)), and stroke-related impairments (Kim et al. (2022)). For older adults, gait characteristics such as balance and walking speed have been linked to early signs of neurological decline (Osoba et al. (2019); Ronthal (2019)), aiding in timely interventions (Littrell et al. (2018)). Beyond clinical settings, gait analysis is employed in other fields. In forensic science, it assists in identifying individuals through distinctive movement patterns and body configurations visible in video recordings (Connor and Ross (2018); Nirenberg et al. (2018); Singh et al. (2018); Birch et al. (2020)). In sports, quantitative assessments of muscular forces and neuromuscular coordination inform training optimization and injury prevention strategies (Taborri et al. (2020)). This type of quantitative assessment supports the understanding of muscle activation patterns and contributes to optimizing athletic performance by offering detailed insights into the biomechanical and neuromuscular aspects of movement.

Gait analysis is commonly structured around three core domains: kinetic, kinematic, and spatio-temporal. Kinetics refers to the study of the forces that generate and control movement

during walking. These include internal joint moments and muscle forces, which are typically inferred from external ground-reaction forces measured through force plates or pressure sensors. Using biomechanical models, these forces are propagated upward through the lower limbs to estimate joint loads and torques, providing insight into the mechanical demands placed on the body. Kinematics, on the other hand, focuses on the motion of body segments without considering the underlying forces. This includes measuring joint angles, angular velocities, and accelerations over time, based on the trajectories of anatomical landmarks captured via motion capture systems or video analysis. It provides a detailed view of how the body moves through space during the gait cycle. Spatio-temporal analysis combines temporal parameters, such as cadence, step time, and gait cycle time, with spatial measures such as stride length, step width, and walking speed. These variables offer a global overview of gait performance and are particularly useful to identify asymmetries or deviations from typical walking patterns.

3.2 Biomechanics and Terminology of Human Locomotion

Human locomotion is a complex and highly coordinated process that allows individuals to move and interact effectively with their environment. It involves the interaction between multiple bodily systems, particularly the musculoskeletal and nervous systems (Whittle (2007)). A detailed understanding of the biomechanics underlying human locomotion is fundamental for advancing a wide range of application domains, including rehabilitation, sport science, smart surveillance, and human-robot interaction.

Terminology is a foundational element in the study of biomechanics, as it establishes a common language for accurately describing the complex processes involved in human movement. Its consistent use is essential not only for clear communication but also for ensuring that findings can be reliably compared and replicated across different studies. In the context of human locomotion, specific terms are used to define the phases of the gait cycle, joint angles, patterns of muscle activation, and other biomechanical parameters. Although a solid understanding of the anatomy of the locomotor system is important for interpreting these concepts, this thesis does not seek to provide an exhaustive anatomical review, as such material is thoroughly covered in numerous textbooks and scientific publications (Whittle (2007); Vaughan et al. (1999); Novacheck (1998); Saunders et al. (1953); Winter (2009); Neptune et al. (2001); Vaughan (2003); Sutherland (2001, 2002, 2005)). The following sections provide a concise overview intended to support a clear understanding of the fundamental principles of gait analysis. While not exhaustive, this background offers the necessary conceptual foundation for the topics addressed throughout the thesis. It is important to note that terminology can vary slightly across textbooks and scientific sources; therefore, the terms adopted here reflect those most commonly used in the field, aiming to ensure clarity and consistency.

3.2.1 Gait Cycle and Gait Phases of Human Locomotion

The gait cycle, also known as stride cycle, encompasses the period between two consecutive occurrences of specific events during walking. Conventionally, it begins with the initial contact of a foot on the ground and ends with the subsequent contact of the same foot on the ground (Kharb et al. (2011)). It is a fundamental concept used in the study of locomotion, and its analysis provides valuable insights into the mechanics of walking and running. The end of one cycle is the beginning of the next one and this periodic leg movement is the essence of the cyclic nature of human gait.

The gait cycle is classically divided into two main phases: the *stance phase*, which begins when the reference foot makes contact with the ground, and the *swing phase*, which follows once that same foot leaves the ground and the corresponding leg moves forward in preparation for the next foot strike (Whittle (2007); Winter (2009)). The stance period is the time during which the foot is in contact with the ground. The swing period follows the stance period and is the time during which the same foot is in the air.

Within each cycle, two distinct periods involve both feet in simultaneous ground contact – referred to as *double limb support periods* – while the remaining intervals involve only one foot supporting the body's weight, known as *single limb support periods*. These phases and support periods, along with the associated limb movements, are illustrated in Figure 3.1. For clarity, both the figure and the phase descriptions presented in the following paragraphs consider the gait cycle to begin with the initial contact of the right foot (shown in blue), concluding when the right foot contacts the ground again. The left foot (depicted in green) follows the same sequence of phases but is offset in time by half a gait cycle.

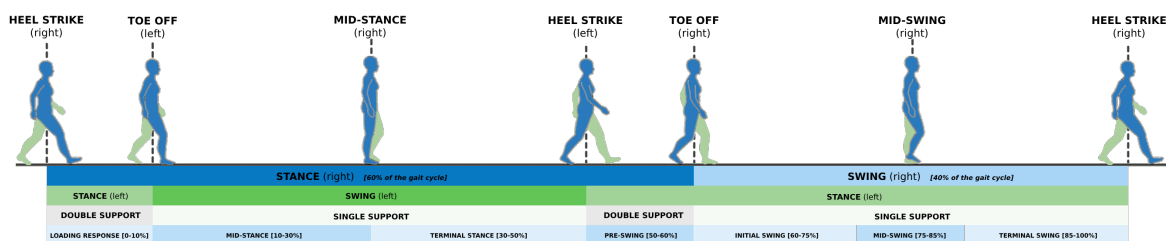


Figure 3.1: Illustration of the human gait cycle. The figure depicts the ordered phases and key events comprising a full cycle, starting at right heel strike and ending at the subsequent right heel strike. The right limb is colored blue and the left limb green. Image adapted from D'Août et al. (2022); Neumann (2017).

Stance Phase

During a *normal*¹ walk, the *stance phase* – when the foot is in contact with the ground – typically constitutes approximately 60% of the total gait cycle. As illustrated in Figure 3.1, this phase is highlighted in the dark blue shaded row and labeled 'Stance (right).' Key events marking the *stance phase* include the initial *Heel Strike*, followed by the *Toe Off* of the opposite foot, *Mid-Stance*, a subsequent *Heel Strike* of the opposite foot, and ending with *Toe Off*. This phase encompasses four distinct sub-phases: *Loading Response*, *Mid-Stance*, *Terminal Stance*, and *Pre-swing*, each corresponding to specific periods and functions within the stance portion of the gait cycle:

- *Loading response right* ($\approx 0\%–10\%$ GC): this phase begins with the right *heel strike* (also called *initial contact*) and ends with the contralateral (left) *toe-off*. The *heel strike* event marks the start of the gait cycle, when the heel of the leading foot contacts the ground, initiating shock absorption and deceleration of the limb. During this phase, up to the instant of contralateral *toe-off*, both feet are on the ground, defining the first *double-support* interval. Functionally, it involves impact absorption while maintaining forward progression through controlled plantar-flexion at the ankle and controlled flexion at the knee, with hip stabilization to support the trunk.
- *Mid-stance right* ($\approx 10\%–30\%$ GC): spans from contralateral *toe-off* until the body progresses anterior to the ankle over the forefoot. This is the first full *single-limb support* interval. The center of mass advances over the support foot, the knee moves toward extension, and the hip is stabilized by the abductors, while the opposite limb is in swing.

Terminal stance right ($\approx 30\%–50\%$ GC): begins at ipsilateral *heel rise* (right *heel-off*) and ends at the contralateral *initial contact* (left *heel strike*). The body progresses forward over the forefoot, with the center of mass advancing anterior to the ankle and the trailing limb approaching its peak hip extension. This phase secures forward progression and sets up propulsion, with ankle plantar-flexor power building toward its peak at the end of terminal stance and into pre-swing.

Pre-swing right ($\approx 50\%–60\%$ GC): is the final portion of the *stance* phase. It begins at contralateral *initial contact* (left *heel strike*) and ends at ipsilateral *toe-off*. This is the second *double-support* interval, characterized by rapid unloading of the right limb and transfer of body weight to the left. Characteristic actions include an increase in ankle plantar flexion, along with a terminal "push-off" power burst, a rapid rise in knee flexion, and the initiation of hip flexion to prepare the limb for swing.

¹In the fields of biomechanics and gait analysis, the term "*normal walk*" is used to describe the typical walking pattern observed in healthy individuals without significant gait abnormalities or impairments. This pattern involves a rhythmic alternation of weight-bearing and non-weight-bearing phases, with a smooth transfer of body weight from one leg to the other. (Oberg et al. (1993)).

Swing Phase

The swing phase usually encompasses about the remaining 40% of the total gait cycle and it is further divided into three sub-phases:

- *Initial swing right* ($\approx 60\%–75\%$ GC): begins at ipsilateral *toe-off* and ends when the swing foot is opposite the stance foot (*feet adjacent*). Its purpose is limb advancement with toe clearance, driven by increasing hip flexion, rapid knee flexion, and the ankle moving from plantar-flexion toward neutral.
- *Midswing right* ($\approx 73–87\%$ GC): begins when the swing foot is opposite the stance foot (*feet adjacent*) and ends when the swing tibia is vertical. The limb advances anterior to the body while maintaining toe clearance, with decreasing knee flexion, continued (decelerating) hip flexion, and the ankle approaching neutral.
- *Terminal swing right* ($\approx 87–100\%$ GC): begins when the swing tibia is vertical and ends at *heel strike* (initial contact). The progression of the limb is completed as the knee extends to near neutral, the hip maintains flexion, and the ankle is held near neutral to position the heel for contact.

In healthy individuals, gait phases proceed continuously and follow the sequence described above. However, deviations may arise in the presence of abnormalities, making the identification of gait phases a valuable method for detecting gait anomalies. In pathological gait patterns, certain gait phases may be absent, and the timing of gait phases can differ from those observed in a typical gait.

3.2.2 Planes of Motion

The planes of motion refer to the fundamental frameworks used in biomechanics to analyze and describe human movement in three-dimensional space. The three cardinal planes of the body – *sagittal*, *frontal*, and *transverse* – serve as essential reference points for understanding movement dynamics (Richards et al. (2022)), as depicted in Figure 3.2:

- **Sagittal plane.** It is a vertical plane parallel to the sagittal suture that divides the body into *left* and *right* portions. Motions in this plane occur about a mediolateral axis and consist primarily of *flexion* and *extension*, for example, trunk flexion/extension, hip and knee flexion/extension, ankle dorsiflexion/plantarflexion, and nodding of the head. Locomotor tasks such as walking and running express predominantly sagittal-plane kinematics, especially at the hip, knee, and ankle, although smaller frontal and transverse components are also present.
- **Frontal plane.** It is a vertical plane parallel to the coronal suture that divides the body into *anterior* and *posterior* portions. Motions in this plane occur about an anteroposterior axis and include *abduction/adduction* of the limbs, lateral flexion of the trunk/neck, radial–ulnar

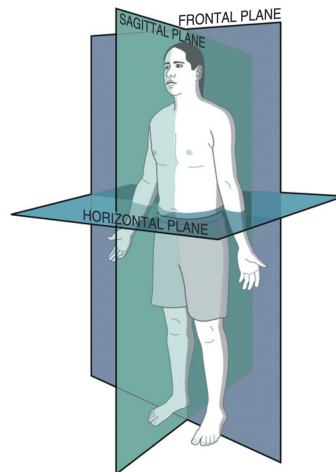


Figure 3.2: The three cardinal planes of the body: sagittal, frontal, and transverse (horizontal) are illustrated with a person in the anatomical position. Source: [Neumann \(2017\)](#).

deviation at the wrist, and inversion/eversion at the subtalar joint. Examples include raising the arm sideways (abduction) and the limb movements in a jumping jack.

- **Transverse plane (axial/horizontal).** In standard anatomical position, this plane is parallel to the ground and divides the body into *superior* (upper) and *inferior* (lower) parts. Motions in the transverse plane are axial rotations, for example, trunk rotation about a vertical axis, internal/external rotation of the hip, and head rotation. Activities such as swinging a golf club or throwing a ball include a prominent transverse-plane rotation of the trunk, though they also involve sagittal- and frontal-plane components.

3.3 Parameters of Gait

Parameters of human gait are fundamental metrics for analyzing and characterizing walking patterns. They span **anthropometric**, **spatio-temporal**, **kinematic**, **kinetic**, and **electromyography** domains, each offering a distinct perspective on movement mechanics: anthropometric factors capture body size, segment lengths, and proportions; spatio-temporal measures describe the timing and spatial organization of gait events; kinematic features quantify joint/segment angles, velocities, and accelerations; kinetic variables assess forces and moments acting on the body; and electromyography indices reflect muscle activation timing and amplitude across the gait cycle. Given the diversity of analysis approaches, careful selection of parameters is essential, as research outcomes depend critically on choosing characteristics suited to the specific context.

3.3.1 Anthropometric Parameters

Anthropometric parameters are physical and body-composition measurements that include age, sex, weight, height, limb lengths, and Body Mass Index (BMI). These parameters have been

shown to significantly influence gait patterns. Consequently, the analysis of gait within groups with matched anthropometric profiles facilitates the isolation of their impact. The identification of anthropometric influences serves to reduce variability and elucidate the gait characteristics associated with specific physical traits.

3.3.2 Spatio-temporal Parameters

Spatio-temporal parameters primarily focus on the time and distance-related aspects of walking. These parameters encompass a variety of measurements related to both the spatial and temporal aspects of gait, offering valuable information about the rhythm, pace, and symmetry of walking patterns. Spatial parameters include measurements such as step length, stride length, and width, while temporal parameters include factors like gait speed, cadence, stance and swing duration (Pattanapisont et al. (2023)). Next, the various spatio-temporal features of gait are listed with a brief description, where the corresponding unit of measurement are enclosed in square brackets: all length units are in meters (m), time is in seconds (s), and velocity is in meters per second (m/s).

- **Step length** (*left or right*) [m]: is the anterior-posterior distance between successive heel contact points of the opposite feet. The right step length is the distance from the left heel to the right heel when both feet are in contact with the ground.
- **Stride length** [m]: it is defined as the anterior-posterior distance between two successive heel contact points of the same foot. A gait cycle or stride is comprised of two steps. In the case of normal gait, the stride length, also referred to as the *cycle length*, is equal to that of double-step length.
- **Step width** [m]: refers to the lateral distance between the midpoint of one foot and the midpoint of the other foot during the walking line of progression.
- **Cadence** [$steps/min$]: it is defined as the number of steps taken per unit of time.
- **Stance time** (*left or right*) [s]: designates the duration from the first contact to the final contact of a single footstep during walking.
- **Swing time** (*left or right*) [s]: refers to the duration of the phase when the foot is not in contact with the ground and is swinging forward in preparation for the next step.
- **Single limb support time** (*left or right*) [s]: refers to the duration from the moment the opposite foot leaves the ground to the point when the same foot lands again during walking.
- **Double limb support time** [s]: refers to the duration during the gait cycle when both feet are simultaneously in contact with the ground.
- **Step time** (*left or right*) [s]: refers to the time elapsed from initial contact of one foot to initial contact of the opposite foot.

- **Stride time** [s]: refers to the duration of a complete cycle of walking, beginning with the initial contact of one foot with the ground and ending when the same foot makes contact with the ground again.
- **Walking velocity** (also referred as gait speed) [m/s]: distance covered per unit time during steady-state walking; typically computed as the measured stride length divided by the measured stride time.

The analysis of these parameters is essential for characterizing normative gait and detecting pathological deviations. For example, reduced gait speed, particularly when accompanied by cognitive decline, is significantly associated with a higher risk of dementia in older adults (Collyer et al. (2022)). In Parkinsonian gait, shortened *step lengths*, prolonged *double-stance phases*, and increased *step width* are frequently observed (Morris et al. (1999)). For instance, individuals with knee osteoarthritis typically show a reduced *step length* on the affected limb relative to the contralateral side (Chen et al. (2020)).

3.3.3 Kinematic Parameters

Kinematic parameters describe the mechanical characteristics and detailed movement patterns of joints and body segments during gait, independent of the forces causing these movements. These parameters encompass *joint angles*, *segmental positions*, *velocities*, *accelerations*, and *trajectories* based on anatomical landmarks. Specifically, metrics such as hip, knee, and ankle joint angles, as well as segmental velocities and angular accelerations, are crucial for assessing the coordination, efficiency, and fluidity of gait. The evaluation of these kinematic features facilitates the identification of movement inefficiencies, abnormalities, or compensatory patterns, thereby providing valuable insights into the overall quality of gait.

Furthermore, kinematic analysis encompasses metrics that extend beyond those specific to joints, incorporating broader characteristics such as linear and angular displacements, as well as their temporal derivatives, which include velocities and accelerations. Examining these trajectories allows for comprehensive assessment of segmental motion timing, coordination, and smoothness across the entire gait cycle. Deviations from typical kinematic patterns can reveal underlying musculoskeletal or neurological impairments, establishing kinematic analysis as a critical tool for both clinical diagnosis and performance optimization. For instance, a comparative analysis of three-dimensional motion capture data revealed substantial disparities in ankle dorsiflexion and rotations, knee and hip flexion-extension and rotations, and pelvic rotation between adults with flatfoot and those with neutral feet, thereby substantiating the existence of posture-related alterations in gait (Marouvo et al. (2021)).

3.3.4 Kinetic Parameters

Kinetic parameters are the forces and moments that govern the mechanics of human gait. They include the three-dimensional *ground-reaction force* (GRF) (Park and Kim (2022)), the *vertical*

free moment (Begue et al. (2018)), the *center-of-pressure* (COP) trajectory (Jeon and Cho (2020)), *net joint moments* (torques) (Mundt et al. (2020)), *joint reaction forces*, and *joint powers/work* (Mentiplay et al. (2018)). GRF, free moment, and COP are measured using force platforms or instrumented treadmills, whereas plantar-pressure distribution is obtained from pressure plates or in-shoe sensors. Net joint moments, joint powers, and joint reaction forces are not measured directly. They are estimated via inverse dynamics by combining GRF with whole-body kinematics and segmental anthropometrics (mass and inertia). When normalized to body mass and interpreted alongside center-of-mass (CoM) kinematics, these parameters reveal the mechanical demands placed on joints and muscles, supporting both the diagnosis of gait abnormalities and performance optimization.

3.3.5 Electromyography Parameters

Electromyography (EMG) captures the spatio-temporal pattern of muscle activation during gait by recording motor unit action potentials (MUAPs) via surface or intramuscular electrodes. Surface EMG is preferred in gait studies due to its non-invasiveness, while intramuscular EMG provides higher selectivity at the cost of invasiveness. The quality of the signal is influenced by various factors, including the placement and orientation of the electrodes with respect to the muscle fiber direction, the electrode dimensions, the inter-electrode distance, the composition of the subcutaneous tissue (i.e., fat and muscle), and interference (crosstalk) from adjacent muscles. These factors must be controlled alongside standard preprocessing techniques, such as band-pass filtering, rectification, and envelope extraction (Agostini et al. (2020); Papagiannis et al. (2019)).

Collectively, *amplitude*, *timing*, and *co-activation* parameters offer a compact yet comprehensive description of muscle recruitment, coordination, and neuromuscular control during locomotion, and are sensitive to pathological gait adaptations and rehabilitation-induced changes.

Amplitude-based parameters quantify activation magnitude and include peak and mean envelope values, the root-mean-square (RMS) amplitude, and integrated electromyography (iEMG). To ensure inter-subject comparability, amplitude measures are typically normalized to a maximal voluntary contraction (MVC) or to peak activity over a reference task or gait cycle (Yang and Winter (1984); Halaki and Ginn (2012)).

Temporal parameters delineate the dynamics of activation in relation to gait events, such as onset and offset relative to heel strike and toe-off, as well as activation duration and time-to-peak. These metrics facilitate the evaluation of inter-muscle timing relationships (e.g., *tibialis anterior* vs. *gastrocnemius*) and stride-to-stride variability, thereby supporting inference about neuromuscular coordination and motor control strategies.

Co-activation indices provide a quantitative assessment of the simultaneous activity of antagonistic or synergistic muscle pairs (e.g., *tibialis anterior*–*gastrocnemius*, *quadriceps*–*hamstrings*) during specific phases of gait (Winby et al. (2013)). These indices offer an indirect measure of joint stiffness regulation and stabilization demands. When relevant, frequency-domain descriptors (e.g., median or mean frequency) and time–frequency features (e.g., wavelet energy) can be used to probe muscle fatigue or changes in motor unit recruitment.

3.3.6 Other Parameters

Beyond the core spatio-temporal, kinematic, and kinetic descriptors, additional parameters provide complementary insight into stability, coordination, variability, smoothness, and task complexity during human locomotion. Many of these measures can be obtained non-invasively with vision-based systems that continuously track body segments across the gait cycle.

- **Center-of-Mass (CoM) displacement** (Tesio and Rota (2019)): is the time-varying change in the 3D position of the whole-body CoM relative to a defined reference frame (e.g., laboratory coordinates), typically reported along the anterior–posterior, mediolateral, and vertical axes. Vision-based tracking of anatomical landmarks can be used to estimate the CoM trajectory, aiding the assessment of dynamic stability and balance control. Deviations from typical patterns may indicate impaired balance or increased fall risk.
- **Gait symmetry** (Viteckova et al. (2018)): is defined as the extent to which the movements of the left and right limbs during activities such as walking or running are similar. This is a critical measure in biomechanics, as it quantifies how closely the movement patterns of one side of the body mirror those of the opposite side. This concept is of particular importance when assessing balance and coordination, as well as identifying potential movement impairments. In order to assess gait symmetry, a range of kinematic parameters must be analyzed. These parameters include, but are not limited to, joint angles, range of motion, and timing of limb movements. Discrepancies in these parameters between the left and right sides have the capacity to reveal underlying imbalances or irregularities. For instance, asymmetries in joint angles or step length may be indicative of musculoskeletal disorders, neurological conditions, or compensatory patterns arising from pain or injury. Accurate measurement of gait symmetry is instrumental in facilitating the identification of movement inefficiencies, leading to the diagnosis of conditions and the subsequent tailoring of rehabilitation processes to restore normal and balanced gait patterns.
- **Foot (toe) clearance** (Fan et al. (2020a)): denote the minimum vertical distance between the swinging foot (which is frequently the hallux marker) and the ground during swing. Precise tracking of foot trajectory allows estimation of minimum toe clearance and its variability, which is critical for identifying trip risk and monitoring rehabilitation progress.
- **Symmetry and asymmetry indices** (Viteckova et al. (2018)): are used to quantify left-right differences in specific parameters of gait (e.g. step length, stance time, arm swing amplitude). Common formulations (e.g., symmetry index or asymmetry ratio) have been shown to facilitate the detection of unilateral impairments and the monitoring of recovery.
- **Stride-to-stride variability** (Oudenhoven et al. (2019)): The dispersion (e.g., coefficient of variation) of gait measures across consecutive cycles. Elevated temporal or spatial variability, beyond normal physiological levels, can signal poorer locomotor control.

- **Harmonic ratio** (trunk accelerometry) (Fransen et al. (2019)): is a frequency-domain measure derived from segment (often trunk) accelerations that reflects step-to-step rhythmicity and smoothness. Lower values indicate poorer gait regularity and symmetry.
- **Movement smoothness** (Antonelli et al. (2024); Beck et al. (2018)): is determined by the evaluation of time-domain metrics (e.g., normalized jerk, SPARC) which quantify the regularity of body-segment trajectories (e.g., trunk or feet). Lower smoothness values may be indicative of increased neuromotor noise or the utilization of compensatory control strategies.
- **Margin of Stability** (Watson et al. (2021)): is the instantaneous distance between the extrapolated CoM and the limits of the base of support in anteroposterior and mediolateral directions. It is evident that reduced margins, or negative margins, are indicative of diminished dynamic stability and an augmented propensity for falling.
- **Turning and transition metrics** (Weston et al. (2024)): is a quantitative analysis of turning behavior (e.g. peak yaw rate, turn duration, steps per turn) and sit-to-stand/stand-to-sit transitions, with the objective of comprehensively assessing task complexity and vestibular–postural integration in contexts beyond straight-line walking.
- **Arm swing amplitude and symmetry** (Hill and Nantel (2019)): is an indicator of trunk–limb coordination, with the capacity to serve as a sensitive marker of neurological conditions and compensatory strategies.

Collectively, these parameters extracted from continuous tracking of body and foot motion offer a more holistic view of gait dynamics and control.

3.4 Normative Values of Gait Parameters

The normative values of gait parameters are standardized reference measurements derived from large samples of healthy individuals without musculoskeletal, neurological, or other conditions that could affect walking, i.e., individuals exhibiting normal gait. These parameters encompass spatio-temporal metrics, such as velocity, step and stride length, cadence, stance and swing proportions, double support, and step width, as well as joint kinematics across the gait cycle, particularly at the hip, knee, and ankle. Establishing normative values is imperative for detecting clinically meaningful deviations, monitoring longitudinal change, and evaluating intervention effectiveness in conditions such as stroke (Selves et al. (2020)), cerebral palsy (Papageorgiou et al. (2019)), Parkinson's disease (Creaby and Cole (2018)), and during post-surgical recovery (Gurchiek et al. (2019)). In research, normative values provide reference standards for validating new gait analysis methods, calibrating measurement systems, and benchmarking machine-learning models for event detection and classification. Furthermore, they facilitate the implementation of quality control measures and the establishment of cross-study comparability by ensuring that the distributions

of key variables (for example, spatio-temporal means, joint-angle ranges) are consistent with the expected confidence intervals for comparable populations and acquisition conditions.

Since gait evolves with growth and declines with aging, and is further influenced by sex/gender, stature, and habitual speed, stratifying normative values by age group (and, when feasible, by sex) is essential to avoid misinterpretation from heterogeneous samples. For example, classic pediatric studies show that determinants such as velocity, cadence, step length, and single-support time approach near-adult values by approximately 2–3 years of age, while spatio-temporal scaling continues to adjust until around 7 years (Bach et al. (2021); Hillman et al. (2009); Froehle et al. (2013)). In contrast, healthy aging is typically associated with slower preferred speed, shorter steps and strides, longer double-support phases, and subtle kinematic changes – notably reduced peak hip extension in stance and reduced ankle plantar-flexion at push-off (Hollman et al. (2011); van Hove et al. (2017)). It is therefore imperative that a careful stratification by these covariates is undertaken for valid comparisons, accurate inference, and the ability to generalize both clinical and research findings.

3.4.1 Normative Values for Spatio-Temporal Parameters

Spatio-temporal parameters including stride length, walking speed, cadence, and the distribution of stance and swing phases provide objective descriptors of gait performance. The establishment of reliable reference values enables clinicians and researchers to identify deviations that can be associated with musculoskeletal impairments, neurological conditions, or age-related decline.

Several large-scale studies have reported normative datasets (Hollman et al. (2011); Rowe et al. (2021); Dusing and Thorpe (2007)), most notably the decade-resolved reference tables by Öberg et al. (1993), complemented by cross-sectional analyses from Samson et al. (2001). Together, these works show that, between 20 and 59 years, spatio-temporal parameters are relatively stable: stride length and speed show only modest, gradual decreases, cadence is largely age-independent, and the relative durations of stance and swing remain consistent at preferred speed. In a recent study, Rössler et al. (2024) analyzed a balanced group of subjects spanning the adult lifespan, thereby confirming this pattern using modern walkway instrumentation and distributional modeling. The study found that gait speed and stride length decline with age, but differences become most pronounced from the eighth decade onward (≈ 80 –91 years) compared with younger groups, and variability/asymmetry measures increase with age. These findings serve to reinforce the interpretation that midlife (20–59 years) occupies a comparatively narrow “normative band” whereas later life shows clearer divergence from midlife norms.

Table 3.1 presents normative values for six commonly reported parameters in healthy adults aged 20–59 years. These values (means $\pm$ SD) were obtained under self-selected walking speed on instrumented walkways and are suitable as benchmarks for both clinical assessments and experimental studies. When sex- or height-specific interpretation is required, base the analysis on the original tables and, where appropriate, apply normalization procedures described in the literature (Öberg et al. (1993)).

Table 3.1: Normative spatio-temporal gait parameters at preferred speed (ages 20–59, healthy adults). Source: [Oberg et al. \(1993\)](#).

Age band	Stride length	Speed	Cadence	Step length	Stance phase	Swing phase
20–29	1.45 ± 0.10	1.40 ± 0.12	112 ± 6	0.73 ± 0.05	60 ± 2	40 ± 2
30–39	1.44 ± 0.10	1.38 ± 0.12	112 ± 6	0.72 ± 0.05	60 ± 2	40 ± 2
40–49	1.42 ± 0.11	1.35 ± 0.13	111 ± 6	0.71 ± 0.06	61 ± 2	39 ± 2
50–59	1.39 ± 0.12	1.30 ± 0.14	110 ± 6	0.70 ± 0.06	62 ± 2	38 ± 2

Units. Stride/Step length in **meters (m)**; Speed in **meters per second (m/s)**; Cadence in **steps per minute (steps/min)**; Stance/Swing phase in **percent of gait cycle (%)**. Values are mean $\pm$ SD at self-selected speed on instrumented walkways.

3.4.2 Normative Values for Kinematic Parameters

At the **hip**, the sagittal motion alternates quasi-sinusoidally between flexion and extension over the gait cycle (see Fig. 3.3), with modest variability due to walking speed, subject individuality, and landmark definitions, yet remaining relatively consistent even in older adults ([Latorre et al. \(2019\)](#); [Rowe et al. \(2021\)](#)). Maximum hip flexion ($\sim 30^\circ$) typically occurs during the terminal swing phase ($\approx 85\text{--}100\%$ GC), positioning the limb for heel strike (initial contact). A brief deceleration or small extension may occur before heel strike. After initial contact, the hip extends through stance as the body progresses over the limb, reaching $\sim 10\text{--}15^\circ$ of peak extension near terminal stance or early pre-swing ($\approx 40\text{--}60\%$), often just after contralateral foot strike. With toe-off at $\approx 60\%$ GC, the hip then flexes to advance the limb and prepare for the next initial contact. In the frontal plane, small-amplitude adduction during single-limb support is resisted by the hip abductors to limit contralateral pelvic drop and maintain mediolateral stability; in the transverse plane, low-amplitude internal–external rotation aligns the knee and foot with the direction of progression, aiding clearance and efficient redirection ([Kirtley \(2006\)](#); [Perry and Burnfield \(2024\)](#); [Winter \(1991\)](#)).

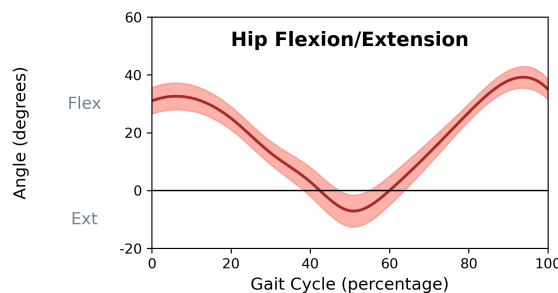


Figure 3.3: Normative range for the hip flexion/extension angle over the gait cycle, showing the mean trajectory and the variability band (shaded area). Adapted from [Kirtley \(2006\)](#).

The **knee** kinematics are dominated by sagittal-plane motion, exhibiting two characteristic flexion peaks in each gait cycle (see Fig. 3.4), while smaller frontal and transverse motions con-

tribute to alignment and stability. At heel strike (initial contact), the knee is near extension (typically $\approx 0\text{--}5^\circ$, with variability from slight hyperextension $\sim -2^\circ$ to $\sim 10^\circ$ flexion; mean $\sim 5^\circ$). During loading response ($\approx 0\text{--}10\%$ GC), the knee flexes to $\approx 15\text{--}20^\circ$ to absorb impact, with the quadriceps acting eccentrically to regulate the rate of flexion against the external flexion moment from the ground-reaction force. Through mid-stance ($\approx 10\text{--}30\%$ GC), the knee re-extends toward near full extension ($\approx 0\text{--}5^\circ$) to provide an efficient strut for body-weight support and forward progression. The second, larger flexion peak occurs during swing: following heel rise and pre-swing, toe-off typically occurs near $\approx 60\%$ GC when knee flexion is $\sim 40^\circ$, after which the knee flexes rapidly to a peak of $\approx 60\text{--}70^\circ$ in early-to-mid swing ($\approx 60\text{--}80\%$ GC). This swing-phase flexion, together with ankle dorsi-/plantar-flexion, shortens functional limb length to ensure toe clearance, before late-swing rapid extension prepares the limb for the next initial contact. The amplitude and timing of these features vary with walking speed, individual differences, and landmark definitions (Heiden et al. (2009); Fukuchi et al. (2019)).

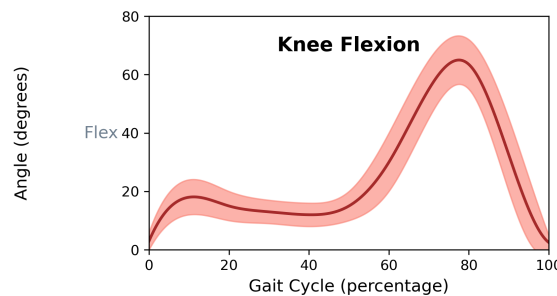


Figure 3.4: Normative range for the knee flexion angle over the gait cycle, showing the mean trajectory and the variability band (shaded area). Adapted from Kirtley (2006).

The **ankle** motion follows a tri-phasic pattern that supports weight acceptance, tibial progression, and propulsion (see Fig. 3.5). Immediately after heel strike, at the initial contact, controlled plantarflexion lowers the forefoot to the ground under eccentric dorsiflexor control, contributing to shock absorption. Through mid-stance ($\approx 30\text{--}40\%$ GC), the tibia advances over a relatively fixed foot, producing dorsiflexion that commonly peaks around $\approx 10\text{--}15^\circ$, regulated by eccentric plantarflexors. In terminal stance and pre-swing phases ($\approx 45\text{--}60\%$ GC), the ankle rapidly plantarflexes toward a second peak ($\approx 15\text{--}20^\circ$) that coincides with heel rise and push-off, providing a major contribution to forward progression via a plantarflexor power burst. During swing, the ankle dorsiflexes rapidly to achieve and typically sustain a neutral position (0° to $\approx 5^\circ$ dorsiflexion) by mid-swing, thereby ensuring adequate toe clearance and minimizing the risk of foot-drop. Overall sagittal range of motion during walking is typically $\approx 20\text{--}40^\circ$ (Zelik and Honert (2018); Neumann (2017)).

These peak kinematic parameters, when analyzed collectively, offer a comprehensive view of lower-limb joint behavior during walking. Deviations from typical values may indicate biome-

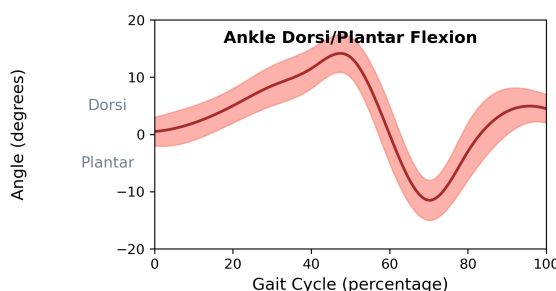


Figure 3.5: Normative range for the ankle dorsi-/plantar-flexion angle over the gait cycle, showing the mean trajectory and the variability band (shaded area). Adapted from Kirtley (2006).

chanical dysfunctions and are often used to assess pathological gait, design rehabilitation programs, or evaluate performance in clinical and athletic populations.

3.5 Sensing Technologies in Gait Analysis

Gait analysis critically depends on the data acquisition technology employed, as each sensing modality determines which parameters of movement can be measured, with what precision, and under which conditions. The technological setup therefore defines the scope of observable gait domains – spatio-temporal, kinematic, kinetic, or electromyographic – and constrains the reachable accuracy, repeatability, and ecological validity. Broadly, these technologies can be categorized into two primary groups: **wearable systems**, which comprise sensors attached to the body or integrated into personal objects, and **non-wearable systems**, encompassing external instrumentation such as cameras, force plates, or pressure walkways. This distinction is foundational because it conditions what can be measured, where measurements can be taken, the complexity of the setup, and the overall cost (Muro-de-la Herran et al. (2014)).

Concretely, optical motion capture emphasizes **kinematics** (joint angles, segment trajectories); force plates resolve **kinetics** (ground-reaction forces, joint moments); pressure walkways recover **spatio-temporal** measures (cadence, velocity, step/stride length, stance/swing time); and surface EMG captures **electromyography** (activation timing and intensity across the gait cycle) (Tao et al. (2012)). Across modalities, several trade-offs recur: precision versus accessibility (high-end laboratory systems are typically more precise but expensive and infrastructure-heavy), experimental control versus real-world validity (controlled conditions versus continuous, minimally obtrusive measurement in natural environments), cost and scalability, and requirements for calibration, synchronization, and usability (setup time, operator expertise, tolerance to placement variability).

3.5.1 Wearable Technologies

Wearable systems have democratized gait monitoring by facilitating portability, continuous assessment, and real-time feedback at a comparatively low cost. The strengths of this technology include its versatility (modular sensors covering multiple segments), objectivity, and its seamless

integration into daily routines, such as smartphones and smartwatches. The inherent limitations of these systems may include variability in placement, the necessity for calibration, drift in inertial sensors during extended recordings, potential interference with natural movement, battery life and charging requirements, and user compliance over prolonged periods. These factors motivate standardized protocols for sensor placement, synchronization, and external validation against reference systems (Pantelopoulou and Bourbakis (2010); Prasanth et al. (2021)).

Inertial Measurement Units (IMUs)

IMUs combine accelerometers, gyroscopes, and often magnetometers. The placement of the sensor varies according to the targeted metrics, with common locations including the shoe, the ankle/shank, the waist, and the trunk (see Fig. 3.6a). These sensors offer high temporal resolution, portability, and the ability to capture details of accelerations and angular velocities during various phases of the gait cycle. Systematic reviews document the sensors' robust performance and real-time gait event detection across healthy and clinical populations (Prasanth et al. (2021); Ye et al. (2020); Mo and Chow (2018); Storm et al. (2016)).

IMUs are robust to visual occlusions and are relatively inexpensive. However, these sensors are susceptible to integration drift, sensitivity to magnetic disturbances, and challenges with generalization outside controlled settings. To ensure accuracy, careful calibration and cross-validation are essential (Zhou et al. (2020); Prasanth et al. (2021)).

Force/Pressure Sensors (instrumented insoles, plantar sensors)

Instrumented insoles are designed to measure plantar pressure and/or normal plantar force (see Fig. 3.6b). This facilitates the estimation of the vertical ground-reaction force (GRF) and derivation of contact time, loading/impulse, and regional load distribution in the foot. Multiple validation studies report concurrent validity versus force plates for walking and running (e.g., peak force/impulse and stance time) (Muro-de-la Herran et al. (2014)). Strengths include high temporal resolution (≈ 100 – 200 Hz) and on-site measurement, whilst limitations include lower spatial granularity than high-resolution mats, calibration drift across sessions/footwear, and durability under prolonged or high-impact use (Martini et al. (2020)).

Surface Electromyography (sEMG)

Surface electromyography records muscle electrical activity via adhesive electrodes (see Fig. 3.6c). Standard pipelines implement band-pass filtering, rectification, and envelope extraction to obtain activation profiles aligned with gait events; normalization (e.g., MVC or functional tasks) is critical for inter-subject comparability. The method supports the analysis of muscle synergies, the detection of pathological co-activation, and the determination of the optimal timing for assistive interventions. Limitations include motion artefacts, crosstalk, and sensitivity to electrode placement and normalization strategy. More evidence on these limitations can be found in focused reviews on gait sEMG (Papagiannis et al. (2019); Felici and Del Vecchio (2020)).

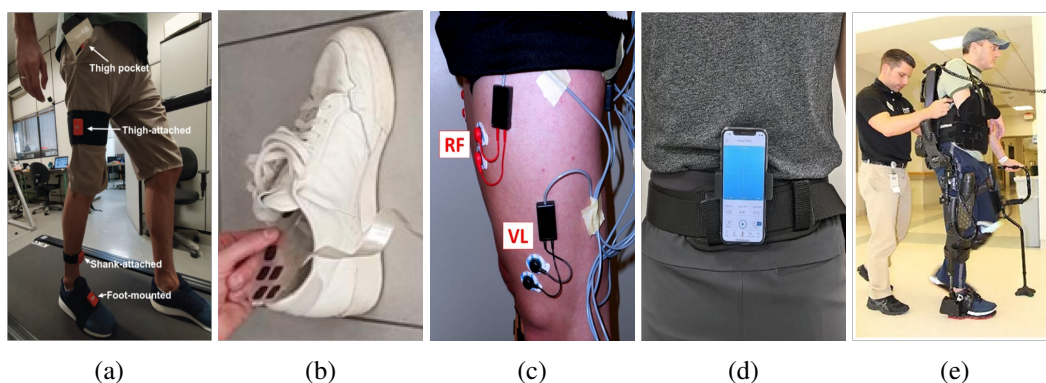


Figure 3.6: Examples of wearable sensing technologies used in gait analysis: (a) IMU placement for lower-limb gait data acquisition (Garcia et al. (2022)); (b) positioning of an instrumented pressure insole inside the shoe (Salis et al. (2021)); (c) surface EMG applied over the rectus femoris (RF) and vastus lateralis (VL) (Di Nardo et al. (2020)); (d) a smartphone affixed to the lumbar region for in-clinic assessments of walking speed and postural stability (Doshi et al. (2023)); (e) robotic exoskeleton gait training with a physical therapist (Nolan et al. (2020)).

Smartphones and Smartwatches

Commodity devices are equipped with accelerometers and gyroscopes, with sampling rates ranging from 50 to 200 Hz (see Fig. 3.6d). These devices typically have accessible application programming interfaces (APIs) which facilitate the estimation of steps, cadence, and speed in real-world settings. Validation studies report acceptable reliability and validity in controlled and clinical contexts (Longoni et al. (2019); Mourcou et al. (2015); Auepanwiriyaikul et al. (2020)). Recent vision-based toolchains using multiple smartphones (e.g., OpenCap) reconstruct 3D kinematics and even estimate dynamics from synchronized videos, bridging lab-grade insights and scalable field assessments (Uhlrich et al. (2023); Turner et al. (2024)).

Instrumented Exosuits/Exoskeletons

Assistive and rehabilitation devices typically integrate on-board sensing mechanisms to support phase detection and control: IMUs for gait-event and phase estimation; joint encoders/potentiometers for angles and velocities; and load cells/strain gauges (or inline torque sensors) for cable/actuation forces (see Fig. 3.6e). These signals reveal metrics of human-robot interaction, including interaction forces/torques, delivered assistance and mechanical power, and timing error. They also enable real-time closed-loop controllers (impedance/admittance, phase-dependent, intent-adaptive) that synchronize and personalize assistance over the gait cycle (Kalita et al. (2020); Baud et al. (2021); Chen et al. (2019); Arslan et al. (2019)).

3.5.2 Non-Wearable Technologies

Non-wearable systems for gait analysis include optical motion capture (marker-based and markerless), force plates, pressure walkways, and radar, LiDAR, or ultrasonic sensors. Most of these

solutions typically require substantial infrastructure and rigorous calibration or synchronization. Nevertheless, they can deliver high precision and multimodal data for both kinematic and kinetic analysis. These modalities differ in the type of information they encode, ranging from 2D joint keypoints to RGB-D point clouds and full 3D reconstructions. Consequently, they exhibit different levels of setup complexity, robustness to environmental factors, and achievable fidelity. Optical motion capture remains the gold standard in controlled laboratory spaces, offering high-resolution kinematic data, but it entails significant infrastructure and operator requirements that can limit viability.

Modern markerless pose estimation approaches promote real-world validity by enabling natural, minimally intrusive recordings, yet they remain vulnerable to occlusions, clothing variability, and domain shifts in the underlying models. Depth-based cameras relax some constraints on lighting and cost, while compact multi-view rigs have reduced barriers to use in clinical and domestic contexts.

Optical Motion Capture (marker-based)

Multi-camera systems (e.g., Vicon, OptiTrack, Qualisys) with retro-reflective markers provide three-dimensional trajectories and joint kinematics through skeletal modeling (see Fig. 3.7a). Marker-based motion capture (MoCap) is widely recognized as the gold standard for three-dimensional kinematics due to its sub-degree accuracy when properly calibrated (Balazia and Sojka (2018); Moro et al. (2022)). However, this method requires meticulous calibration, skilled operators, precise marker placement, and substantial investments in equipment and maintenance (Eichelberger et al. (2016); Yahya et al. (2019)).

Markerless Motion Capture / Computer Vision

Markerless approaches rely on RGB cameras and/or depth sensors. They capture visual data of individuals walking and employ image processing and learning-based techniques to extract relevant gait information (Singh et al. (2018); Sandau et al. (2014)). Through segmentation, feature extraction, and pattern recognition (Needham et al. (2022)), these systems can generate kinematic measurements across the gait cycle, identifying key gait events with high accuracy using modest hardware and short setup times (Clark et al. (2019); Sikandar et al. (2021); Yagi et al. (2020)).

RGB Cameras. RGB cameras provide rich appearance cues (e.g., color, texture, contours) at low cost and with rapid, non-intrusive setup, making them attractive for gait monitoring in clinics and real-life settings (Yagi et al. (2020)). In typical pipelines, individuals are detected in each frame and two-dimensional human pose is estimated as a set of anatomical keypoints using learning-based models (Sarafianos et al. (2016); Zheng et al. (2023a)) (see Fig. 3.7b). Trajectories are then temporally filtered to derive spatio-temporal parameters, including cadence, step/stride timing, and length. Three-dimensional pose can be obtained either (i) by multi-view triangulation after camera calibration or (ii) by monocular “lifting”, which utilizes learned biomechanical priors

to infer depth. These choices involve a trade-off between setup complexity and metric accuracy. Key limitations of monocular setups include the absence of true depth (scale ambiguity), sensitivity to camera placement and lighting, and vulnerability to self-occlusions and clothing. Best practices—appropriate viewing distance, oblique camera angles, multi-view configurations, and careful model selection—mitigate many of these issues (Stenum et al. (2024)). Privacy concerns can be reduced by on-device inference and by storing only derived keypoints or silhouettes rather than raw video. Despite constraints, the ubiquity of RGB sensors and mature pose-estimation tooling enable real-time, markerless gait analysis at scale (Yagi et al. (2020); Wade et al. (2022)).

Depth Sensors. Depth cameras (e.g., Kinect v2/Azure Kinect, Intel RealSense) generate per-frame depth maps/point clouds (and frequently vendor skeletal tracks – see Fig. 3.7c), which facilitate background subtraction, provide metric scale without calibration objects, and exhibit greater tolerance to illumination changes than RGB alone (Wang et al. (2020)). These characteristics support tasks such as presence detection (Fuentes-Jimenez et al. (2020)), distance/rhythm estimation, fall detection (Panahi and Ghods (2018)), and fine-grained kinematic analysis (Faity et al. (2022)). However, depth modalities typically operate with lower spatial resolution, limited effective range and field of view (FOV), and characteristic artefacts (multipath interference, glossy/absorbing surface failures, edge “flying pixels”). Technology choices (structured light, active stereo, time-of-flight, infrared thermography) involve clear trade-offs in cost, accuracy, range, and robustness that have been reviewed in these works: Muro-de-la Herran et al. (2014); Tölgyessy et al. (2021); Dal Mutto et al. (2012). Because depth does not convey color or texture, RGB-D fusion is often employed to combine the appearance strength of RGB with the metric geometry of depth, enhancing tracking stability while preserving a high degree of privacy.

Force Plates

Force plates measure tri-axial ground-reaction forces and moments about the plate origin, from which center-of-pressure (CoP) trajectories are computed (Simonet et al. (2024)) (see Fig. 3.7d). When combined with segment kinematics, they enable inverse-dynamics estimation of joint kinetics (net joint moments and powers). Due to the high sampling rates of approximately 1–2 kHz and the excellent accuracy of force plates, they serve as a reference for the detection of gait events (i.e. heel-strike/toe-off), phase partitioning and asymmetry analysis in clinical and research settings (Gonçalves et al. (2020)). Principal limitations include the small effective area, necessitating precise foot placement or multi-plate/walkway setups, together with cost, fixed infrastructure, and the need for rigid mounting and careful calibration (Perry (2010)). They also capture only brief contact intervals and may induce targeting behavior if participants aim their steps at the plate.

Pressure Mats/Walkways

Pressure walkways (e.g., GAITRite) employ dense pressure sensor arrays to capture foot-ground contact patterns and derive spatio-temporal parameters, such as gait speed, cadence, stride/step

length, and stance/swing times (Parati et al. (2022); Bilney et al. (2003); Menz et al. (2004)). These parameters exhibit strong reliability and concurrent validity. Compared with in-shoe insoles, walkways offer high spatial resolution over a fixed global coordinate frame and do not require participant instrumentation. Their kinetic capabilities are limited: they sense normal pressure (not shear), provide only indirect CoP estimates, and cannot recover full ground-reaction forces or joint kinetics as force plates do. Practical constraints include finite walkway length (fewer steady-state steps), deployment typically restricted to indoor, level surfaces, potential “targeting” of footfalls by participants, and sensitivity to footwear, contact-thresholding, and calibration settings, all of which can affect measurement reliability.

Radar, LiDAR, Ultrasonic

Frequency-modulated continuous-wave (FMCW) and Doppler radars operate independently of ambient light and enable privacy-preserving, contactless gait sensing (see Fig. 3.7e). By analyzing range-Doppler and micro-Doppler signatures, it is possible to estimate spatio-temporal parameters (e.g., cadence, step/stride timing) and coarse segmental velocities, with exemplar studies showing good agreement against reference systems (Figueiredo et al. (2025); Wan et al. (2018)). LiDAR and ultrasonic modalities have also been investigated for presence detection, ranging, and rhythm estimation (Parashar et al. (2023a); Benedek et al. (2018); Qi et al. (2014)). However, recovery of detailed joint-level kinematics relative to multi-camera vision systems is constrained by limited angular/spatial resolution and susceptibility to environmental artefacts, including multipath reflections, specular surfaces, and clutter.

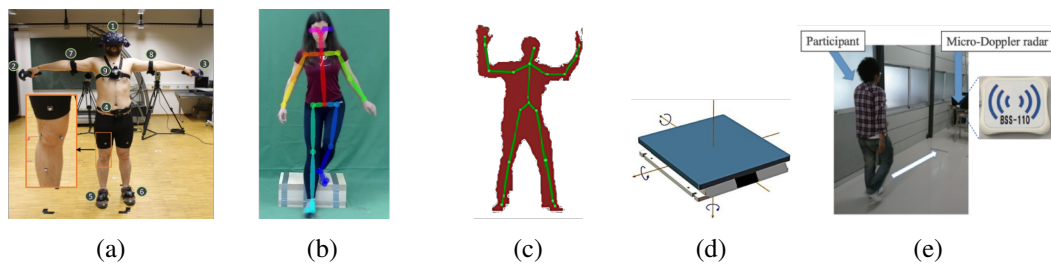


Figure 3.7: Examples of non-wearable sensing technologies used in gait analysis: (a) a participant in T-position with attached reflective markers (Vox et al. (2021)); (b) a participant walking in front of an RGB camera with the estimated skeleton rendered on top (Adolf et al. (2023)); (c) a participant standing in front of an RGB-D sensor with the estimated skeleton rendered on top (Geerse et al. (2015)); (d) representative force plate system, where the lines depict the measurable forces and moments (Lamkin-Kennard and Popovic (2019)); (e) a participant standing in front of a micro-Doppler radar for gait classification (Hayashi et al. (2021)).

3.6 Summary

Human gait is a complex biomechanical process shaped by intrinsic factors such as skeletal alignment, musculature, and joint mobility, as well as extrinsic influences including emotional state, environment, and footwear. Despite these transient variations, the underlying locomotor pattern remains sufficiently consistent to enable both clinical assessment and biometric applications. The chapter has demonstrated how gait analysis provides a structured framework for the observation, quantification and interpretation of locomotion, with broad relevance across a range of disciplines.

In order to provide a solid foundation for this understanding, the chapter first introduced the fundamental biomechanical principles of human locomotion, including the gait cycle, its phases, and the planes of motion. The subsequent discussion delineated the primary categories of gait parameters, encompassing anthropometric, spatio-temporal, kinematic, kinetic, electromyographic, and complementary measures. Reference values were reviewed to illustrate the expected ranges in healthy populations and to emphasize the importance of stratification by age, sex, and anthropometry. Finally, the chapter examined the manner in which both wearable and non-wearable sensing technologies facilitate the practical measurement of the aforementioned parameters, thereby achieving a balance between precision and ecological validity.

Chapter 4

Datasets for Gait Analysis – GRIDDS and Other Resources

4.1 Scope and Contributions

Gait recognition studies the identification of individuals from the dynamics of their walking patterns. In computer vision, two sensing modalities dominate in gait recognition studies: “vision-based” data captured as RGB images or video, and “depth-based” measurements obtained from active sensors often paired with RGB as RGB-D. Progress in this field has been driven by the availability of public datasets that support benchmarking, reproducibility, and fair comparison across methods. Because reported performance can be highly sensitive to the dataset and protocol used, robust, public benchmarks are essential for reliable evaluation.

An up-to-date review of publicly available datasets is timely for at least three reasons. Firstly, the landscape of datasets is subject to constant change. New resources are continually emerging, including large in-the-wild benchmarks (Zhu et al. (2021)) and clinical or health-oriented collections (Kour et al. (2021); Yin et al. (2025)). Concurrently, other datasets may change their access terms or become inaccessible (Hofmann et al. (2014); DeCann et al. (2013)). Secondly, modalities and representations are subject to evolution. Alongside the well-established silhouette-based representations, there has been notable growth in the use of RGB and RGB-D video with auxiliary annotations (e.g., pose, optical flow, parsing) that support new training and evaluation protocols (Zheng et al. (2023b); Nunes et al. (2019a)). Thirdly, the ways in which datasets are cataloged and analyzed have also diversified. Researchers have increasingly explored different organizational perspectives, including by acquisition modality (Singh et al. (2018)), covariates (Li et al. (2024b)), or intended application domain (Chaquet et al. (2013)). This has resulted in new taxonomies of gait datasets. The review of gait datasets presented in this chapter is an extension of previous research (Nunes et al. (2019b)). The initial version of this review was conducted in 2019 and concentrated exclusively on RGB-D-based datasets. The present version extends the aforementioned work to include both RGB and RGB-D datasets, thus offering a more updated and comprehensive perspective.

Beyond cataloging, this chapter adopts a dataset-analysis perspective centered on covariates: factors that alter appearance or motion and thus stress recognition models. Datasets are organized by the presence and degree of control of covariates such as viewpoint, clothing/appearance, carrying conditions, walking speed and trajectory, temporal factors (session gaps, seasons), environment/background (indoor/outdoor, crowding), and recording device/modality. Prior reviews emphasize that method performance varies markedly with these covariates and that robustness claims should be interpreted in light of each dataset’s covariate profile and the evaluation protocol (e.g., cross-view vs. same-view; cross-clothes vs. normal) (Khaliluzzaman et al. (2023); Şahan et al. (2024)).

The remainder of the chapter is organized as follows. Section 4.2 details GRIDDS, encompassing the underlying motivations and objectives, the acquisition environment and the equipment employed, the demographic characteristics of participants, the recording protocol, the available RGB/Depth/IR/skeleton streams with pre-processing and event annotation, the feature definitions for spatio-temporal and kinematic analysis, the evaluation protocols and official splits. Section 4.3 introduces a covariate-centered taxonomy, including its objectives, design principles, catalogue of common covariates, and the resulting scene/user/sensor classification dimensions. Section 4.4 surveys vision-based (RGB) and depth-based (RGB-D) gait datasets, outlining scope and selection methodology, presenting dataset profiles, and applying the taxonomy to support comparisons between similar datasets and a concise comparative synthesis. Finally, Section 4.5 concludes the chapter and reviews the principal contributions.

4.2 GRIDDS – A Gait Recognition Image and Depth Dataset

This section presents the GRIDDS dataset collected as part of this thesis. The subsequent subsections describe, in order, (i) motivation and dataset objectives (Section 4.2.1), (ii) reproducible acquisition parameters (Section 4.2.2), (iii) participant characteristics and the experimental protocol (Section 4.2.3), (iv) stream specifications, pre-processing and data annotation (Section 4.2.4), (v) feature definitions (Section 4.2.5), and (vi) the on-disk organization and access policy (Section 4.2.7).

4.2.1 Motivation and Dataset Goals

The *Gait Recognition Image and Depth Dataset (GRIDDS)* was collected to support research in gait-based person identification and gait analysis from RGB-D streams using a Microsoft Kinect time-of-flight sensor. The dataset is a mid-scale dataset that emphasizes methodological transparency, reproducibility, and verifiable statistical analysis. Positioned between tightly controlled laboratory collections and in-the-wild recordings, GRIDDS provides synchronized color, depth, body-index masks, infrared, and 3D body-joint streams for controlled lateral-view walking, together with per-trial event annotations enabling derivation of spatio-temporal and kinematic gait features. Consequently, GRIDDS accommodates methods based on images, silhouettes, depth, skeletons, or hybrid representations.

Prior literature supports the use of Kinect-family depth cameras for clinical and research gait assessment under defined protocols, with acceptable accuracy reported for many spatio-temporal parameters and moderate accuracy for selected kinematic measures relative to marker-based motion capture (Albert et al. (2020)). Known limitations include occlusions, clothing-related effects, and reliance on tracking heuristics (Oh et al. (2018); Springer and Yogev Seligmann (2016); Sharma et al. (2019b); Foreman and Engsborg (2020); Çubukçu et al. (2020); Diaz-Monterrosas et al. (2018); Mentiplay et al. (2015); Clark et al. (2015, 2019)). Given these characteristics, GRIDDS adopts a side-view to balance tracking robustness while minimizing self-occlusion. Normative biomechanical references like Vaughan et al. (1999) and Kirtley (2006) are used to contextualize spatio-temporal and kinematic patterns.

4.2.2 Acquisition Environment and Equipment

Venue and environment

Data were acquired in the *Biblioteca Barbosa Romero*, Escola Superior de Tecnologia e Gestão (ESTG), Instituto Politécnico de Viana do Castelo (IPVC), Portugal. The recording was carried out over three sunny days in an indoor room with large windows, a static background, and mixed natural/artificial lighting. The capture geometry and trajectories are illustrated in Figure 4.1, where the gray triangle denotes the approximate sensor field of view and trajectories A and B indicate two straight-line walking paths positioned at a ≈ 3.5 m lateral offset from the sensor.

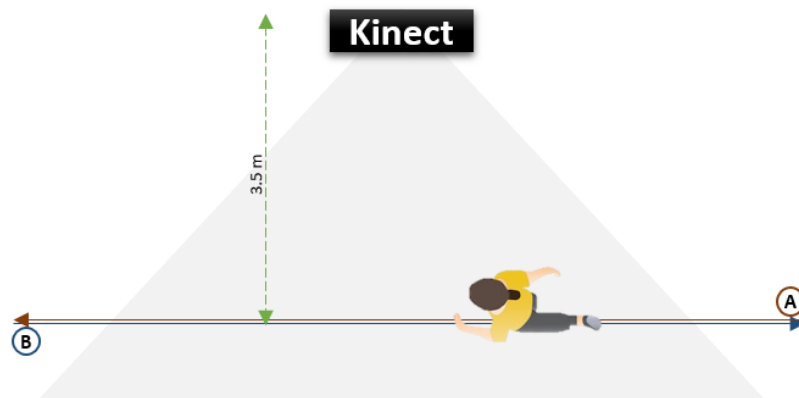


Figure 4.1: Capture environment and sensor placement. The gray triangle represents the approximate sensor field of view; trajectories A and B indicate two straight-line walking paths at ≈ 3.5 m lateral offset from the sensor.

Sensor and mounting

Microsoft's Kinect was originally introduced as a motion controller for Xbox consoles. However, it rapidly transcended the gaming sector due to its relatively low-cost depth sensing capabilities. As the first mass-market depth camera, it extended utility beyond games to a wide



Figure 4.2: Microsoft Kinect v2 RGB-D time-of-flight sensor used in GRIDDS acquisition. *Source: Wikipedia.*

range of computer-vision applications. The precision and convenience of its depth stream attracted broad interest across multiple domains and enabled solutions to challenging perception problems. Representative early examples include people tracking (Papadopoulos et al. (2014); Alexiadis et al. (2011)), object detection and recognition (Bo et al. (2013); Nissimov et al. (2015); Khan et al. (2012)), 3D simultaneous localization and mapping (SLAM) (Hu et al. (2012); Endres et al. (2014)), human activity and posture analysis (Shotton et al. (2011); Chen et al. (2013b,a); Suarez and Murphy (2012)), assistive technologies for accessibility (Armin et al. (2013); Takizawa et al. (2012); Bhowmick et al. (2014); Nissimov et al. (2015)), and robotics (Biswas and Veloso (2012); Fankhauser et al. (2015)), among others.

At the time of data collection, the most recent Microsoft Kinect generation available, and the one used in this study, was Kinect v2 (see Figure 4.2). The Microsoft Kinect v2 time-of-flight (ToF) RGB-D sensor was mounted on a tripod at 1.8 m above the floor, with its optical axis perpendicular to the walking direction, resulting in a standardized side view at an average subject distance of ≈ 3.5 m (see Figure 4.1).

The host computer was a Windows 10 laptop with a quad-core CPU, 16 GB RAM, and USB 3.0. Any configuration offering equivalent functionality, coupled with USB-3.0 bandwidth and a stable system clock, is acceptable for comparable acquisition fidelity.

4.2.3 Participants and Recording Protocol

A total of 35 volunteers (comprising students, teachers, and staff) with no known gait-affecting conditions provided written informed consent to participate in the study. Figure 4.3 illustrates the sex distribution, with 24 females and 11 males. The demographic information of the participants, including their first and last names, ages, heights, and sexes, was used to calculate descriptive statistics and to assess demographic balance within the sample. The privacy of the subjects was preserved during the release process (see Section 4.2.4 for details). This was achieved by means of face blurring and the absence of any identity markers.

The sex Balance Index (GBI) was used as a preliminary metric to analyze the demographic information of the participants (see Equation 4.1).

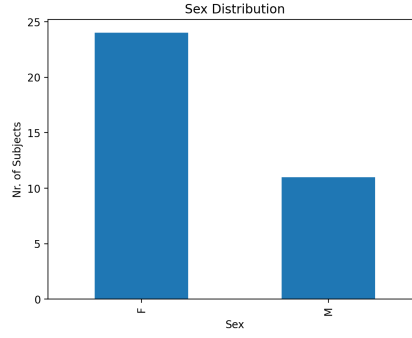


Figure 4.3: Sex distribution of the GRIDDS dataset. Bars show participant counts by sex (F=24, M=11; $n = 35$). Note: F = female, M = male.

$$\text{GBI} = \frac{\min(n_M, n_F)}{\max(n_M, n_F)} \in [0, 1] \quad (4.1)$$

where n_M denotes the number of *male* participants and n_F denotes the number of *female* participants. Values closer to 1 indicate better sex balance, whereas values near 0 indicate stronger imbalance. The GBI value for GRIDDS is thus determined to be $11/24$, or equivalently, 0.4583. This value suggests a moderate sex imbalance. This imbalance impacts inference because sex correlates with several gait characteristics.

The descriptive statistics for age and height, both overall and by sex, are presented in Table 4.1. Distributional plots in Figs. 4.4 and 4.5 show a broad adult age span (19–54 years) with limited representation of older adults and a clearly separated height distribution between the sexes. This age profile is consistent with the dataset’s recruitment context: an academic community predominantly comprising higher-education students and working-age faculty/staff. As such, the under-representation of older adults reflects the source population rather than a design omission.

Table 4.1: Descriptive statistics for age (years) and height (cm). Values are mean $\pm$ SD with min–max in parentheses.

Group	Age (y)	Height (cm)
Overall (n=35)	35.91 $\pm$ 11.34 (19–54)	167.77 $\pm$ 10.10 (145–189)
Female (n=24)	39.00 $\pm$ 10.64 (20–54)	163.04 $\pm$ 7.54 (145–172)
Male (n=11)	29.18 $\pm$ 10.20 (19–47)	178.09 $\pm$ 6.74 (166–189)

From the perspectives of interpretability and risk of bias, the height separation (female: 163.04 $\pm$ 7.54 cm; male: 178.09 $\pm$ 6.74 cm) may inadvertently serve as a cue for identification or classification if features encode body size-dependent effects (e.g. step length, segment lengths). Analyses that compare groups or build predictive models should either normalize spatio-temporal features by stature or limb length (e.g. stride length/height), or evaluate models with stratified sampling and/or class-weighting to reduce contamination of sex-associated anthropometrics.

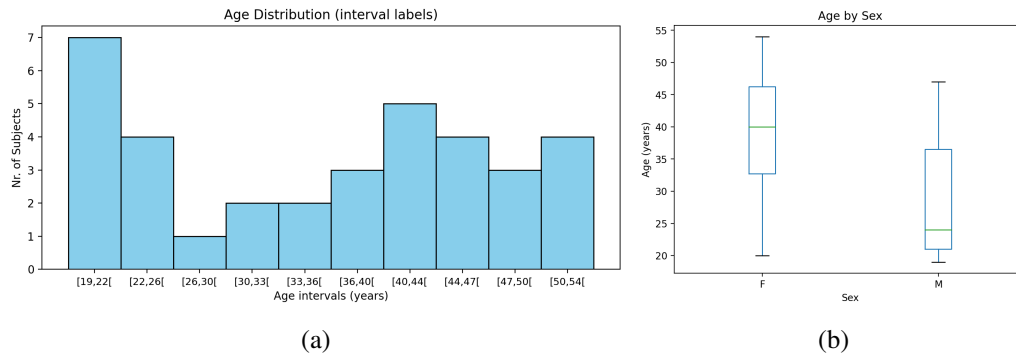


Figure 4.4: (a) Histogram of participant age ($n = 35$). (b) Boxplots of age stratified by sex (F=24, M=11). Note: F = female, M = male.

Protocol

Participants were instructed to walk along two straight, opposing trajectories (A and B; see Fig. 4.1), each perpendicular to the sensor, in order to ensure a consistent side view. For each participant, ten sequences were recorded (five traversals per trajectory) at a comfortable, self-selected pace. Throughout the sessions, participants maintained a lateral offset of approximately 3.5 m from the camera. All sequences for a given participant were collected within a single day to limit between-day variability in terms of speed, clothing, and environmental conditions. In total, the dataset comprises 350 sequences ($n = 35$ participants $\times$ 10 sequences each).

4.2.4 Data Modalities, Pre-processing and Annotation

Available Kinect v2 data streams

The Kinect v2 provides a synchronized set of imaging and tracking outputs designed to capture complementary aspects of human motion. In GRIDDS, these streams were used to obtain texture (RGB), range (depth), near-infrared intensity (IR), per-pixel human segmentation (body index), and skeletal estimates (25 joints per tracked person), as illustrated in Figure 4.6.

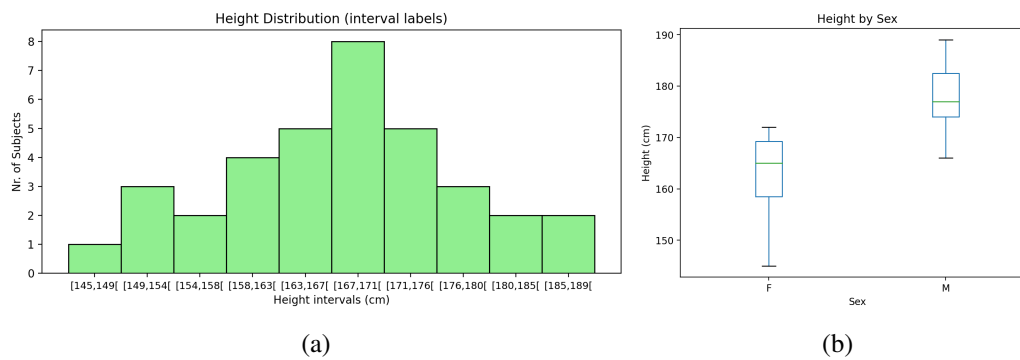


Figure 4.5: (a) Histogram of standing height ($n = 35$). (b) Boxplots of height stratified by sex (F=24, M=11). Note: F = female, M = male.

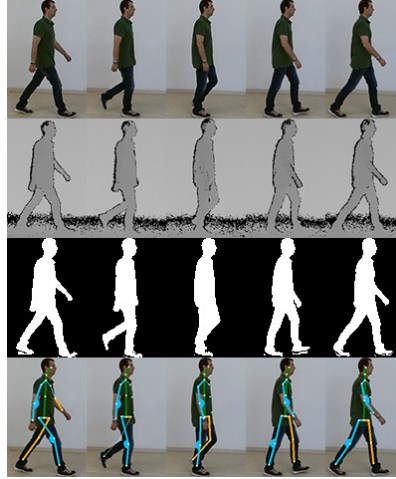


Figure 4.6: Representative GRIDDS frames illustrating synchronized Kinect v2 outputs from a single trial: **row 1**-RGB images; **row 2**-registered depth maps; **row 3**-binary body silhouettes derived from the body-index mask; **row 4**-25-joint skeleton overlaid on RGB.

The recorded streams in GRIDDS are the following:

- **Color (RGB).** HD video frames at 1920×1080 pixels, nominal 30 Hz. Data are stored as 8-bit per channel (24-bit PNG on disk).
- **Depth.** Range images at 512×424 pixels, nominal 30 Hz. Each pixel encodes distance from the sensor in millimeters.
- **Infrared (IR).** Intensity images at 512×424 pixels, nominal 30 Hz. IR frames are derived from the same optical path as depth.
- **Body Index.** Label images at 512×424 pixels at nominal 30 Hz. Each pixel indicates whether it belongs to the background or to one of the currently tracked bodies (indices 0–5). This per-pixel mask is the basis for silhouette extraction.
- **Body / Skeleton.** Up to six bodies tracked concurrently. Each tracked body contains 25 joints with 3D positions (in meters) and tracking states (tracked/inferred/not tracked), updated at nominal 30 Hz. The 25-joint topology of the SDK (names and connectivity) is shown in Figure 4.7.

A compact summary of Kinect streams (resolution and data type/units) is provided in Table 4.2.

Because the color and depth images have different native resolutions and fields of view, the SDK supplies calibrated mappings to correlate pixels and to project between image spaces and 3D camera space. The acquisition process was configured to target a nominal cadence of 30 frames/s across streams. However, short-term fluctuations may occur due to exposure control, USB bandwidth, or CPU load. All files therefore carry precise timestamps, and data synchronization should rely on timing rather than assuming a perfectly constant framerate.

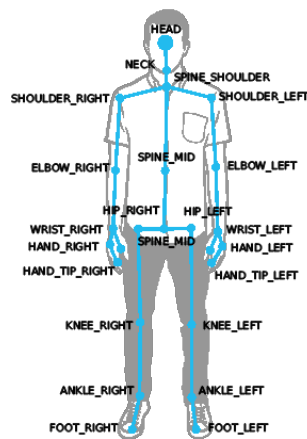


Figure 4.7: Kinect v2 skeletal model, illustrating the 25 tracked body joints with their SDK names and the segment connectivity (topology) used for kinematic computations. Labels follow the Kinect v2 enumeration (e.g., head, spine_shoulder, shoulder_left/right, elbow_left/right, wrist_left/right, hand_left/right, hip_left/right, knee_left/right, ankle_left/right, foot_left/right, hand_tip_left/right, thumb_left/right). Image adapted from [Corporation \(2014\)](#).

Coordinate systems In order to facilitate unambiguous analysis, GRIDDS preserves both joint and pixel coordinates in the three standard spaces exposed by the SDK. No equations are needed to use them correctly. Rather, it is important to understand the origin, axes, units, and the fact that color and depth frames differ in resolution and field of view.

- *Camera (3D) space.* A right-handed, metric 3D frame whose origin is at the sensor. The horizontal axis is oriented towards the left of the sensor, the vertical axis points upwards, and the depth axis points forward away from the sensor. All skeletal joint positions are expressed in meters, and this is the reference used for geometric measurements such as step length or joint angles.
- *Color image space.* A 2D grid covering the RGB image at 1920 columns by 1080 rows. Pixel coordinates start at the top-left corner and increase to the right (columns) and downward (rows). When overlaying joints on color frames, joint positions from camera space must be projected to this grid using the SDK’s coordinate mapping. Direct indexing is not valid because color and depth resolutions differ.
- *Depth image space.* A 2D grid covering the depth (and IR/body-index) images at 512 columns by 424 rows. Pixel coordinates also start at the top-left corner and increase to the right and downward. Each pixel has an associated distance value in millimeters. Body-index masks are aligned to this grid.

The SDK provides calibrated, bidirectional mappings: from 3D camera space to either image grid (to draw joints on images), from depth pixels plus their distance values back to 3D (to reconstruct points in space), and between depth and color pixels (to align masks or textures).

Table 4.2: Microsoft Kinect v2 stream characteristics used in GRIDDS.

Stream	Resolution	Data / Units
Color (RGB)	1920×1080	8-bit/channel (PNG)
Depth	512×424	16-bit range [mm]
Infrared (IR)	512×424	16-bit intensity
Body Index	512×424	8-bit labels
Body/Skeleton	up to 6 bodies	25 joints, 3D [m]

Silhouette normalization and face anonymization.

Only two pre-processing steps were applied, both minimal by design to preserve the temporal structure of gait while ensuring comparability across subjects and basic privacy protection.

1. **Silhouette normalization.** For each frame, the body-index mask was cropped to a tight bounding box around the subject’s silhouette and rescaled to a fixed height of 120 pixels while preserving aspect ratio. The resized mask was then placed on a 80×120 -pixel canvas and horizontally re-centered by aligning the canvas center with the centroid of the upper half of the silhouette (torso-dominant region), which reduces lateral drift due to arm swing. To enforce a common facing direction, all sequences recorded in one walking direction (e.g., right-to-left) were horizontally flipped so that every sequence faces the same canonical direction.
2. **Face masking.** RGB frames were post-processed to blur facial regions before any external sharing of the dataset. This removes easily identifiable facial texture while retaining sufficient context for quality control and figure generation.

Data Annotation and Labeling of Gait Events.

All gait-event labels in GRIDDS were annotated through a fully manual procedure at a frame level. Each of the 350 trials was inspected as a color image sequence with optional skeletal overlay, and the annotator interacted with a dedicated desktop application to assign a frame index to each event that defines the gait cycle. This workflow ensures that event timing is grounded in visual evidence from the raw sequence.

For each gait cycle, the following items were recorded: `Closest_Side`, identifying the body side nearest the sensor (left or right); `Starting_Foot`, indicating which foot initiates the cycle (left or right); `Heel_Strike_Start`, the first ground contact of the starting foot (cycle on-set); `Toe_Off_Opposite`, the moment the contralateral foot leaves the ground; `Midstance`, when the stance limb passes beneath the body; `Heel_Strike_Opposite`, the contralateral heel strike; `Toe_Off_Start`, the toe-off of the starting foot; `Midswing`, the midpoint of the starting foot’s swing phase; and `Heel_Strike_End`, the subsequent heel strike of the starting foot that completes the cycle.

All gait cycles were annotated by a single, trained annotator in order to maintain internal consistency of visual criteria. A second annotator then undertook an independent review of a random subset of gait cycles, comparing event frame assignments and side/foot flags. A specific application was developed to facilitate the labeling process and ensure consistent annotation recording. The application’s main functionalities are as follows:

- *Sequence I/O*: open/close an RGB image sequence corresponding to any trial.
- *Playback controls*: play forward or backward; step frame-by-frame or run continuously.
- *Frame awareness*: display the current frame number and total length.
- *Overlays*: toggle the skeletal projection (25-joint stick figure) on top of the RGB frame for visual cross-checks.
- *Event assignment*: bind the current frame to a selected gait event (from the event list above).
- *Trial metadata*: set `Closest_Side` and `Starting_Foot` via keyboard shortcuts.

As mentioned before, the acquisition process targeted a nominal cadence of 30 frames/s across all streams, but short-term fluctuations and occasional frame drops can occur due to automatic exposure control, USB 3.0 bandwidth, or CPU scheduling. When the specific frame corresponding to an event was absent (i.e., the event occurred during a gap), the event field was recorded as NaN. Table 4.3 summarizes the aggregate volume of annotations. The fully recorded-cycle rate indicates that nearly all cycles contain the full complement of events, with a small shortfall attributable to isolated frame drops or visually ambiguous instances.

Table 4.3: Dataset-wide annotation totals for GRIDDS. Entries report (i) *Total gait cycles* corresponding to cycles with both start and end heel strike annotated, (ii) *Fully recorded cycles* in which all seven events are present, and (iii) *Total steps*, defined by two consecutive heel strikes.

Total Gait Cycles	1169
Fully recorded cycles (all events present)	1125
Completed Gait Cycles (%)	96.2
Total steps	2329

With regard to event-wise completeness, as summarized in Table 4.4 and corroborated by the per-participant bar chart of total vs. fully recorded cycles in Figure 4.8, completeness exceeds 98.9% across all event categories, supporting robust calculation of step/stride timing, single- and double-support intervals, cadence, and related features.

Table 4.4: Event completeness across all trials. “Valid” counts events with assigned frame indices; “Missing” indicates NaN entries; completeness is the valid fraction in percent.

Event	Valid	Missing	% Complete
HEEL_STRIKE_START	1163	6	99.49
TOE_OFF_OPPOSITE	1156	13	98.89
MIDSTANCE	1160	9	99.23
HEEL_STRIKE_OPPOSITE	1160	9	99.23
TOE_OFF_START	1160	9	99.23
MIDSWING	1159	10	99.14
HEEL_STRIKE_END	1164	5	99.57

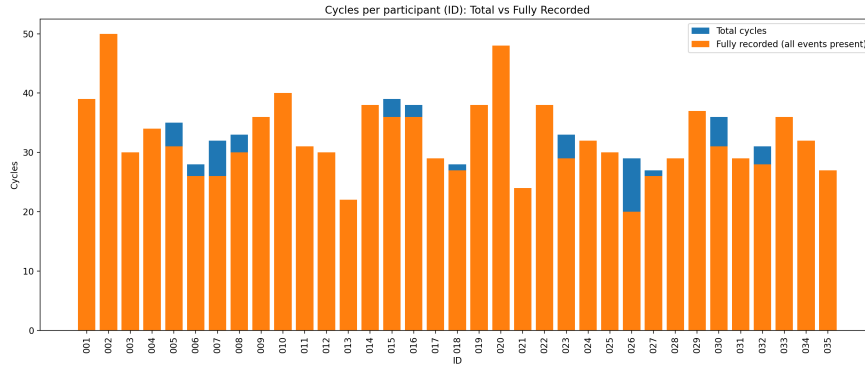


Figure 4.8: Per-participant bar chart (IDs 001–035) showing, for each participant, the total gait cycles (blue bars) and the subset of fully recorded cycles with all events present (orange).

4.2.5 Feature Extraction and Descriptors

This section details the features computed from RGB-D data with skeletal tracking, organized into two complementary families: (i) *spatio-temporal features*, which summarize timings and distances derived from discrete gait events, and (ii) *kinematic features*, which characterize the evolution of joint angles across the gait cycle. All computations use Kinect *camera space* joint coordinates (meters), a right-handed, metric 3D reference independent of image resolution: $+X$ points to the sensor’s left, $+Y$ upward, and $+Z$ forward. Under the side-view acquisition that was used, progression is mainly along X : right-to-left image motion corresponds approximately to $+X$, left-to-right to $-X$, while Y captures vertical body motion (e.g., trunk/pelvis oscillation). Accordingly, step and stride lengths are estimated along the direction of progression (here, the camera’s X axis), whereas step width is measured orthogonally to progression (here, along the camera’s Z axis).

Spatio-temporal features and event-based definitions

Let $x_J(t), y_J(t), z_J(t)$ denote the camera-space coordinates (meters) of joint J at time t . Let the annotated event times be t_k^{HSL} and t_k^{HSR} for left/right *heel strikes*, and $t_k^{\text{TO_L}}$ and $t_k^{\text{TO_R}}$ for left/right

toe-offs, where the subscript k indexes successive occurrences for a given side. The lengths are reported as nonnegative magnitudes (absolute values). Signed variants may be retained internally for direction-aware analyzes.

- **Step length.** Right step length at right heel strike (HS_R) is the separation along the walking axis y between ankle proxies:

$$L_{\text{step},R}(k) = \left| x_{\text{ankle_right}}(t_k^{\text{HS}_R}) - x_{\text{ankle_left}}(t_k^{\text{HS}_R}) \right|, \quad (4.2)$$

Left step length is defined analogously at HS_L :

$$L_{\text{step},L}(k) = \left| x_{\text{ankle_left}}(t_k^{\text{HS}_L}) - x_{\text{ankle_right}}(t_k^{\text{HS}_L}) \right|. \quad (4.3)$$

- **Step time.**

$$T_{\text{step},R}(k) = t_{k+1}^{\text{HS}_R} - t_k^{\text{HS}_L}, \quad T_{\text{step},L}(k) = t_{k+1}^{\text{HS}_L} - t_k^{\text{HS}_R}. \quad (4.4)$$

- **Stride length.** Left-stride length is the X -displacement of `ankle_left` between successive left heel strikes:

$$L_{\text{stride},L}(k) = \left| x_{\text{ankle_left}}(t_{k+1}^{\text{HS}_L}) - x_{\text{ankle_left}}(t_k^{\text{HS}_L}) \right|. \quad (4.5)$$

with an analogous definition for the right side.

- **Stride duration (gait cycle time).**

$$T_{\text{stride},L}(k) = t_{k+1}^{\text{HS}_L} - t_k^{\text{HS}_L}, \quad T_{\text{stride},R}(k) = t_{k+1}^{\text{HS}_R} - t_k^{\text{HS}_R}. \quad (4.6)$$

- **Cadence.**

$$\text{Cadence} = \frac{60}{\overline{T_{\text{step}}}} \quad [\text{steps} \cdot \text{min}^{-1}], \quad (4.7)$$

where $\overline{T_{\text{step}}}$ is the mean of all left/right step times in the trial.

- **Stance and swing times (per side).**

$$T_{\text{stance},L}(k) = t_k^{\text{TO}_L} - t_k^{\text{HS}_L}, \quad T_{\text{swing},L}(k) = t_{k+1}^{\text{HS}_L} - t_k^{\text{TO}_L}, \quad (4.8)$$

with analogous expressions for the right side.

- **Single-limb support (SLS) and double-limb support (DLS).** Using left as the reference side

in cycle k ,

$$T_{\text{SLS},L}(k) = t_k^{\text{HS}_R} - t_k^{\text{TO}_R}, \quad (4.9)$$

$$T_{\text{DLS},\text{initial}}(k) = t_k^{\text{TO}_R} - t_k^{\text{HS}_L}, \quad T_{\text{DLS},\text{terminal}}(k) = t_{k+1}^{\text{HS}_L} - t_k^{\text{TO}_L}, \quad (4.10)$$

and $T_{\text{DLS}}(k) = T_{\text{DLS},\text{initial}}(k) + T_{\text{DLS},\text{terminal}}(k)$. Definitions swap L/R for the opposite side.

- **Step width.** In the side-view geometry, mediolateral separation is along the camera’s Z axis. At right heel strike:

$$W_{\text{step},R}(k) = |z_{\text{ankle_right}}(t_k^{\text{HS}_R}) - z_{\text{ankle_left}}(t_k^{\text{HS}_R})|. \quad (4.11)$$

with an analogous definition at left heel strike.

- **Gait speed (walking velocity).** Two equivalent formulations are commonly used; GRIDDS reports the stride-based estimate:

$$v = \frac{\overline{L_{\text{stride}}}}{\overline{T_{\text{stride}}}} \quad [\text{m} \cdot \text{s}^{-1}], \quad (4.12)$$

where the overbar denotes the average across available strides in the trial. A step-based average of $L_{\text{step}}/T_{\text{step}}$ produces numerically similar results for a constant walk.

Table 4.5: Summary of spatio-temporal features computed from camera-space joints.

Feature	Symbol	Definition (events)	Axis / Units
Step length (R/L)	L_{step}	Eqs. (4.2)–(4.3) (HS)	X [m]
Step time (R/L)	T_{step}	Eq. (4.4) (HS)	[s]
Stride length (R/L)	L_{stride}	Eq. (4.5) (HS)	X [m]
Stride duration (R/L)	T_{stride}	Eq. (4.6) (HS)	[s]
Cadence	—	Eq. (4.7)	[steps/min]
Stance time (R/L)	T_{stance}	Eq. (4.8) (HS,TO)	[s]
Swing time (R/L)	T_{swing}	Eq. (4.8) (HS,TO)	[s]
Single-limb support (R/L)	T_{SLS}	Eq. (4.9) (TO,HS)	[s]
Double-limb support	T_{DLS}	Eq. (4.10) (HS,TO)	[s]
Step width	W_{step}	Eq. (4.11) (HS)	Z [m]
Gait speed	v	Eq. (4.12)	[m/s]

Kinematic features: joint-angle definitions and measures

Joint-angle trajectories are computed for the **knee**, **hip** (both the *trunk–thigh included angle* and the *thigh–vertical angle*), and **elbow**, time-normalized to the 0-100% gait cycle. Per-subject mean

curves (left/right) with variability bands are shown in **Appendix A**: knee (A.1), hip (A.2), and elbow (A.3).

- **Generic 3D joint angle.**

For three joints $A-B-C$ defining two segments $\overrightarrow{BA}$ and $\overrightarrow{BC}$, the included angle at B is computed from camera-space coordinates as

$$\theta_{ABC}(t) = \text{atan2} \left(\left\| (\mathbf{r}_A(t) - \mathbf{r}_B(t)) \times (\mathbf{r}_C(t) - \mathbf{r}_B(t)) \right\|, (\mathbf{r}_A(t) - \mathbf{r}_B(t)) \cdot (\mathbf{r}_C(t) - \mathbf{r}_B(t)) \right), \quad (4.13)$$

where $\mathbf{r}_J(t) = [x_J(t), y_J(t), z_J(t)]^\top$ is the 3D position of joint J ; “ $\times$ ” and “ $\cdot$ ” are the cross and dot products; and atan2 ensures numerical stability across the full $[0, \pi]$ range. This definition is rotation-invariant and does not depend on projections to anatomical planes. For side-view analyses where sagittal-plane behavior is of interest, optional projections to the X – Y plane may be used for visualization without altering the definition in (4.13).

- **Knee angle (left/right).**

The knee angle captures the relative orientation of the thigh and shank:

$$\theta_{\text{knee},L}(t) = \theta_{\text{hip_left, knee_left, ankle_left}}(t), \quad \theta_{\text{knee},R}(t) = \theta_{\text{hip_right, knee_right, ankle_right}}(t). \quad (4.14)$$

From the time series $\theta_{\text{knee}}(t)$ within each stride, two summary measures are derived: *maximum flexion* (the largest bending during swing; numerical maximum in radians/deg if flexion is defined as increasing angle) and *maximum extension* (the most extended configuration near mid-stance; numerical minimum if using the same convention).¹ A representative plot across all strides for one subject is provided in Fig. 4.9.

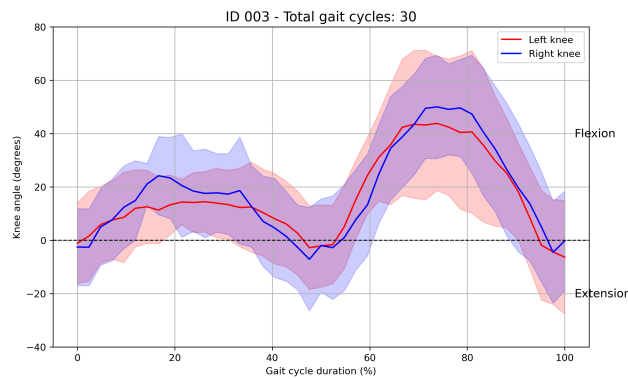


Figure 4.9: **Knee-angle trajectories across strides for subject 003.** Angles are computed from camera-space joint positions and time-normalized to the 0–100% gait cycle. The red line shows the mean *left* knee angle with a red shaded band indicating ± 1 SD across strides; the blue line shows the mean *right* knee angle with a blue shaded band indicating ± 1 SD.

¹Depending on the sign convention adopted for plotting, “flexion up / extension down” or vice versa, the extrema map accordingly. In GRIDDS, the angle magnitude from (4.13) is used, and extrema are reported as numerical max/min.

- **Hip angle (left/right).**

Two complementary definitions are provided; both are reported for completeness.

(i) *Trunk–thigh included angle.* Angle at the hip using the shoulder and ankle on the same side:

$$\theta_{\text{hip-inc},L}(t) = \theta_{\text{shoulder_left, hip_left, ankle_left}}(t), \quad \theta_{\text{hip-inc},R}(t) = \theta_{\text{shoulder_right, hip_right, ankle_right}}(t). \quad (4.15)$$

(ii) *Thigh–vertical angle.* Flexion/extension of the thigh relative to the vertical axis:

$$\theta_{\text{hip-vert},L}(t) = \text{atan2}(\|(\mathbf{r}_{\text{knee_left}} - \mathbf{r}_{\text{hip_left}}) \times \hat{\mathbf{y}}\|, (\mathbf{r}_{\text{knee_left}} - \mathbf{r}_{\text{hip_left}}) \cdot \hat{\mathbf{y}}), \quad (4.16)$$

where $\hat{\mathbf{y}} = [0, 1, 0]^\top$ is the camera-space vertical. The right side is analogous.

For both definitions, summary measures include *maximum flexion* (minimum value if flexion is defined as reduced included angle) and *maximum extension* (maximum value). A sample trace over multiple strides is shown in Fig. 4.10.

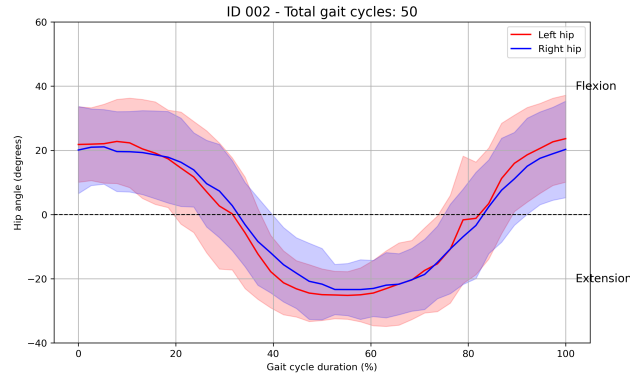


Figure 4.10: **Hip-angle trajectories across strides for subject 002.** Angles are computed from camera-space joint positions using the *trunk–thigh included angle* and the *thigh–vertical angle*, and time-normalized to the 0–100% gait cycle. The red line shows the mean *left* hip angle with a red shaded band indicating ± 1 SD across strides; the blue line shows the mean *right* hip angle with a blue shaded band indicating ± 1 SD.

- **Elbow angle (left/right).**

Upper-limb swing is captured as

$$\theta_{\text{elbow},L}(t) = \theta_{\text{should_left, elb_left, wrist_left}}(t), \quad \theta_{\text{elbow},R}(t) = \theta_{\text{should_right, elb_right, wrist_right}}(t). \quad (4.17)$$

From $\theta_{\text{elbow}}(t)$, *maximum flexion* and *maximum extension* are extracted per stride. Fig. 4.11 depicts an example subject’s elbow angles across the recorded gait cycles.

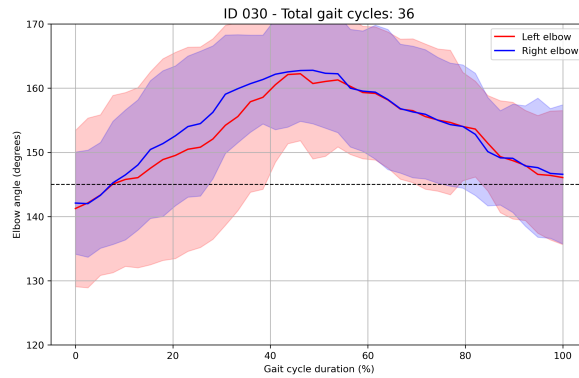


Figure 4.11: **Elbow-angle trajectories across strides for subject 030.** Angles are computed from camera-space joint positions and time-normalized to the 0–100% gait cycle. The red line shows the mean *left* elbow angle with a red shaded band indicating ± 1 SD across strides; the blue line shows the mean *right* elbow angle with a blue shaded band indicating ± 1 SD.

4.2.6 Evaluation Protocols and Official Splits

The evaluation design for GRIDDS aims to make identification results comparable across modalities and implementations while preserving subject non-overlapping during representation learning. All definitions are modality-agnostic and can be applied to RGB, depth, infrared, silhouettes, and 2D/3D joints. Reported metrics should include cumulative match characteristics (at least Rank-1 and Rank-5) and mean Average Precision (mAP), computed over probe queries against a fixed gallery.

The first proposed protocol is a leave-one-subject-out (LOSO) (Gholamiangonabadi et al. (2020)) identification setting aligned with the fixed lateral geometry and two-direction design of the dataset. In each fold, one subject is reserved exclusively for testing, and the remaining subjects may be used to train a feature extractor or classifier where learning is required. The gallery is constructed deterministically using, for every subject, the first two sequences per direction: two left-to-right and two right-to-left, yielding four gallery entries per identity. The probe set comprises the remaining three sequences per direction for the same subjects. This arrangement gives balanced exposure to both walking directions in the gallery and preserves within-subject variation for the probes. Results are aggregated across all probes in the fold, and overall scores are obtained by averaging across the 35 folds.

The second protocol evaluates generalization across the two walking directions using two mirror configurations. In the first configuration, any model training and gallery enrollment rely solely on sequences where subjects walk from left to right, while the probe set contains only sequences where those subjects walk from right to left. The second configuration swaps these roles so that enrollment uses right-to-left recordings and probing uses left-to-right recordings. As in the LOSO setting, the gallery for each enrolled subject is built deterministically from the first two recordings available in the enrollment direction, and the probes are the remaining recordings in the opposite direction. Whenever learning is involved, identities used for training remain disjoint

from those used for testing. This design isolates the walking direction as the only varying factor under a fixed lateral camera and constant environment, allowing an assessment of the invariance to direction reversal. Results should be reported separately for each configuration and may also be averaged to produce a single cross-direction summary.

The protocols have been designed to ensure consistency across different modalities. In the context of appearance-based pipelines, the utilization of silhouettes aligned and normalized to 80×120 pixels has been demonstrated to be a viable approach. These silhouettes can be employed either as direct inputs or as inputs to derived descriptors, such as Gait Energy Images, thereby offering a multifaceted approach to data analysis. In the context of model-based pipelines, 2D/3D joint trajectories derived from Kinect v2 are aligned with the same train/test partitions. Cross-modal comparisons are admissible under the condition that gallery and probe composition, subject non-overlapping rules, and metric calculations remain identical. In instances where hyperparameter selection is necessary, a validation split should be delineated at the subject level exclusively from the training identities. The selection of the model is not influenced by any sequence from the test identities, whether gallery or probe.

Finally, it is essential that all results are accompanied by a detailed protocol declaration. This should include the protocol name, the composition of the gallery per identity, the composition of the probe, the number of subjects and sequences that have been effectively used after any exclusions, and the list of reported metrics.

4.2.7 Data Specification, Licensing, and Availability

Data specification and organization

GRIDDS is organized as a set of 35 participant folders, one per volunteer (IDs 001 – 035). Each participant folder contains ten sessions, and within every session the acquisition delivers a coherent bundle of streams and metadata sufficient to reproduce all analyzes reported in this thesis. Specifically, the following data is provided per session: (i) *color* images (RGB), (ii) *depth* images and per-frame *depth data* (CSV), (iii) *infrared* (IR) images, (iv) *joints' coordinates* (CSV) paired with (v) per-frame *timestamps* (CSV), (vi) *binary body silhouettes* normalized in size and canonical orientation, and (vii) per-session *feature tables* (CSV) containing spatio-temporal and kinematic summaries. The silhouettes are distributed as cropped, direction-standardized masks at a fixed resolution of 80×120 pixels (width $\times$ height), facilitating appearance-based descriptors.

Folder structure (per participant).

Each participant directory `vvv/` (with `vvv` corresponding to the three-digit of the volunteer ID) contains ten subfolders `vvv_ss/` with `ss = 01...10`. A typical layout is:

```
vvv/
  vvv_01/
    vvv_01_color_###.png      # RGB frames (1920x1080; 24-bit PNG)
    vvv_01_depth_###.png     # Depth frames (512x424; 16-bit grayscale PNG)
    vvv_01_depth_###.csv     # Depth matrices (512x424; mm)
```



```

vvv_01_infrared_###.png      # IR frames (512x424; 16-bit PNG)
vvv_01_silhouette_###.png    # Binary silhouettes (80x120; 1-bit PNG)
vvv_01_timestamps.csv        # Per-frame timestamps for all streams
vvv_01_coordinates_camera.csv # 3D camera joints coordinates
vvv_01_coordinates_color.csv  # 2D RGB joints coordinates
vvv_01_coordinates_depth.csv  # 2D depth joints coordinates
vvv_01_spatiotemporal_features.csv
vvv_01_kinematic_features.csv
vvv_02/
...

```

File-naming convention.

Two patterns are used to make files self-descriptive and sortable:

1. **Per-frame images and matrices:** `vvv_ss_<stream>_nnn.fmt`.

where `vvv` is the three-digit participant ID, `ss` the two-digit session number, `<stream>` $\in \{\text{color, depth, infrared, silhouette}\}$, `nnn` the zero-padded frame index, and `fmt` the file extension (`png` for images; `csv` for depth matrices).

2. **Per-session tables:** `vvv_ss_<stream>.csv`.

where `<stream>` $\in \{\text{timestamps, coordinates_camera, coordinates_color, coordinates_depth, spatiotemporal_features, kinematic_features}\}$.

Example. `003_09_depth_021.csv` denotes the depth matrix (millimeters) for participant 003, session 09, frame 021. `012_04_timestamps.csv` stores the timing for session 04 of participant 012.

Bit depth and units.

All image files use PNG encoding with stream-appropriate bit depths: color (RGB) at **24-bit** (8-bit/channel), depth images at **16-bit** grayscale, IR at **16-bit**, and silhouettes at **1-bit** (binary masks). Depth **CSV matrices** have the same raster geometry as the depth images (512×424); each cell stores the line-of-sight *distance in millimeters* from the sensor to the observed surface at that pixel. Each stream provides framewise timestamps in ticks. For synchronization and framerate estimation, ticks can be converted to seconds via Eq. (4.18).

$$t_{\langle f_i, f_k \rangle} = \frac{\text{timestamp}(f_k) - \text{timestamp}(f_i)}{10,000,000}, \quad k > i \quad (4.18)$$

where f_i and f_k denote frame instants (indices) within the same stream, and $\text{timestamp}(\cdot)$ returns per-frame ticks (10^7 ticks per second).

Joint coordinate tables.

Per session, skeletal coordinates are distributed in *three* CSV files (one per reference system):

- `vvv_ss_coordinates_camera.csv` — 3D joint positions in Kinect *camera* space (meters).
- `vvv_ss_coordinates_color.csv` — 2D joint projections in the *color* image (pixels; 1920×1080).
- `vvv_ss_coordinates_depth.csv` — 2D joint projections in the *depth* image (pixels; 512×424).

Each coordinate file has the same column layout:

```
frame_nr, coordinates, tracking_state.
```

Each row corresponds to a single joint, and rows are ordered such that the joint index cycles from 0 to 24 for every frame, then restarts at 0 for the next frame. Concretely, within a given `frame_nr`, the first row is joint 0, the second row is joint 1, ..., the 25th row is joint 24; the next row begins the sequence for the following frame. The mapping from joint index $\{0, \dots, 24\}$ to SDK joint *names* follows the fixed enumeration adopted in GRIDDS (see Figure 4.7 and the dataset documentation for the exact index–name correspondence). The `tracking_state` field indicates the SDK status for that joint and frame (Tracked, Inferred, or NotTracked).

Per-file coordinate fields.

- *Camera space*: `coordinates = x_3D, y_3D, z_3D` (meters; origin at the sensor; *X* to sensor left, *Y* up, *Z* forward).
- *Color image space*: `coordinates = u_rgb, v_rgb` (pixels; 1920×1080; origin top-left; *u* rightward, *v* downward).
- *Depth image space*: `coordinates = u_depth, v_depth` (pixels; 512×424; origin top-left). Per-pixel metric distances are provided separately as depth matrices (`vvv_ss_depth_nnn.csv`, millimeters).

Data access and availability

GRIDDS is distributed *upon request* for research and educational use. Access is coordinated through the project website (<http://gridds.ipvc.pt/>), which provides contact information and a *Release Agreement* outlining terms of use. Prospective users should (i) review the Release Agreement, (ii) contact the Principal Investigator, and (iii) submit a signed copy specifying institutional affiliation and intended research purpose. Approved requests receive instructions to obtain the dataset.

4.2.8 Evidence of Scientific Use of the GRIDDS Dataset

Since its publication, the GRIDDS dataset has been referenced in at least **11 peer-reviewed publications** that cover gait recognition, pose estimation, multimodal human motion capture, and clinical gait analysis, indicating cross-domain visibility within the scientific community. Authoritative surveys in *ACM Computing Surveys* discuss GRIDDS as part of the larger ecosystem of gait datasets (Topham et al. (2022); Olugbade et al. (2022)), and GRIDDS is further contextualized by a *Scientific Data* article on various multimodal gait resources (Topham et al. (2023)). Additional references include methodological and application-focused works in RGB-D/3D sensing and deep learning: a systematic review of RGB-D gait datasets (Nunes et al. (2019b)), a 3D LiDAR + LSTM approach for gait-based identification (Yamada et al. (2020)), a review on deep learning for criminal investigation (Aung and Kusakunniran (2024)), a clinical dataset paper in *IEEE Access* (Ranjan et al. (2025)), a CNN-based gait method (Zhu et al. (2020)), an occlusion-oriented multi-Kinect benchmark (Li and Zhao (2021)), and related validation work on human motion/anthropometric modeling (Rodrigues et al. (2019)). Collectively, these citations demonstrate that GRIDDS is recognized and used as a reference benchmark across academic, clinical, and technological research contexts.

4.3 A New Taxonomy for Vision and Depth-Based Gait Datasets

4.3.1 Motivation and Objectives

As the field has matured, the organization of gait datasets has diversified. Collections are now described by acquisition modality, scene complexity, sets of covariates, or intended application domain. Recent studies increasingly combine these perspectives to provide multidimensional views. This diversity is both a strength and a challenge. Although it enables targeted experimentation, it also complicates the comparison when conditions labeled similarly involve different underlying choices regarding view topology, clothing variation, carried objects, or temporal gaps between sessions. A taxonomic approach can provide a common framework that links acquisition and design decisions with methodological evaluation. This clarifies which factors are kept constant, which are systematically varied, and how these choices influence the observable characteristics of gait. In this sense, taxonomy acts as a link between data properties and protocol design, helping readers to recognize when results reported under labels such as 'cross-view' or 'cross-clothes' are truly comparable.

A suitable taxonomy for vision- and depth-based gait datasets must therefore support interpretation across modalities while remaining close to the mechanics of acquisition. The immediate objective is to capture covariates that alter the appearance or observability. However, the broader goal is to encode how these covariates interact with task definitions and evaluation procedures. This requires a vocabulary that distinguishes target variables from contextual factors and represents covariates at the level at which they vary (i.e. dataset, session, sequence or subject). It must also make the difference between active covariates (which are varied within a dataset by design)

and controlled parameters (which are held fixed) explicit. This improves reproducibility, reduces ambiguity in protocol reporting, and facilitates the synthesis of evidence from studies using different sensing technologies.

Existing taxonomies illustrate complementary perspectives. [Chaquet et al. \(2013\)](#) organizes video datasets primarily by the nature of the actions they contain, introducing categories for heterogeneous and specific actions, alongside a grouping defined by capture techniques such as infrared or thermal imaging and motion capture. While this approach is valuable for broad action analysis, it is limited for gait-specific collections, where fine-grained distinctions in covariates, such as clothing changes, carrying conditions or camera placement, often influence methodological differences more than the action label itself. When applied to gait analysis, the action-centered perspective provides context about behavioral scope and capture modality; however, it offers insufficient resolution on the acquisition factors that most affect silhouette formation and cross-condition recognition.

A second line of work classifies datasets according to sensing paradigm. [Singh et al. \(2018\)](#) distinguishes between vision-based and sensor-based approaches, the latter including floor sensors and devices such as Kinect. This classification provides a practical scoping tool and helps to consolidate methodological families, but it remains coarse with respect to intra-vision variation. For example, it does not differentiate between single- and multi-view RGB configurations, controlled indoor scenes and outdoor recordings, or the various ways in which temporal spacing, speed profiles and clothing changes are implemented within vision-based datasets. Consequently, while the modality split is a useful initial filter, it provides limited insight into interpreting performance across the visual covariates that dominate current research.

More recently, [Parashar et al. \(2022\)](#) have advanced an occlusion-centric framework that describes axes covering occlusion types, dataset provenance (whether native or synthetic) and methodological strategies that are either reconstruction-based or reconstruction-free. The emphasis here is on the methodology, which links data characterization to the classes of algorithms required to address visibility loss. While this perspective partially overlaps with dataset description insofar as occlusion is a scene-level covariate, it prioritizes one family of challenges at the expense of fully accounting for other influential factors, such as viewpoint, clothing or sensing configuration. The framework's strength lies in clarifying how occlusion shapes algorithm design, while its limitation is the reduced coverage of the wider covariate space.

A closer analogue to the present aims is offered by [Verlekar et al. \(2018\)](#), who groups influential factors into four categories: user, camera, light source and environment. This foregrounds the interaction between lighting and camera parameters and subject properties, which alter appearance and observability. This grouping is intuitive and aligns with covariate thinking, drawing attention to the ways in which acquisition choices are reflected in the data. At the same time, however, the formulation remains largely enumerative. It provides little guidance on dimensionality, on the level at which attributes should be recorded, or on how categories should be formalized to support consistent reporting and direct mapping to evaluation protocols. In practice, this makes it more challenging to translate factor lists into interoperable metadata structures.

The remainder of this section builds on these considerations. The forthcoming proposal explicitly anchors the organisation of the data in covariates and retains the user, scene and sensor strata, while specifying different dimensionality and formalization from prior work, including the factor grouping of Verlekar et al. (2018). Section 4.3.3 introduces the common covariates encountered in vision- and depth-based gait datasets, paying attention to the level at which each one varies. Section 4.3.4 then presents the classification dimensions and associated metadata schema that implement the above-outlined design principles outlined above, enabling consistent description and evaluation across heterogeneous collections.

4.3.2 Taxonomy Design Principles

The design principles outlined below draw upon established concepts in taxonomy construction, metadata modeling, and dataset organization. While no existing gait-specific framework articulates these principles in precisely this form, their foundations lie in widely recognized standards of classification design (Uschold and Gruninger (1996); Nickerson et al. (2013)) and in recent practices for dataset documentation and metadata transparency (Geburu et al. (2021); Paullada et al. (2021)). Together, they provide a structured rationale for developing a taxonomy that balances parsimony, clarity, and interoperability across vision- and depth-based gait datasets.

In this work, general design principles are instantiated for the specific demands of vision- and depth-based gait datasets. The contribution is the translation of abstract ideas about classification and metadata into a covariate-centered schema that: (i) declares the levels at which attributes vary (dataset, session, sequence, subject); (ii) distinguishes active covariates from controlled acquisition parameters; and (iii) provides machine-readable descriptors that map directly onto commonly reported evaluation settings. By grounding the principles in the concrete factors that shape gait appearance and observability, the framework offers a practical, interoperable way to document gait datasets and their protocols under a coherent, high-level taxonomy. The design principles underlying this framework are the following:

- **Minimal coverage.** Categories are kept to the smallest set that still captures the main sources of variation in contemporary vision and depth collections, preserving simplicity without sacrificing representational power.
- **Orthogonality and separability.** Dimensions are defined so that changes along one axis do not implicitly redefine another, enabling independent reasoning about all factors. For example, altering viewpoint should not entail a change in clothing, allowing view and appearance to be analyzed separately.
- **Hierarchical granularity.** Each dimension admits broad labels for rapid categorization and more specific descriptors for detailed analysis. For instance, the concept of illumination may be categorised as either indoor or outdoor, and further refined into natural, artificial, or mixed light.

- **Modality alignment.** Descriptors apply to both RGB and depth, with optional specializations for infrared, multi-view, or fused modalities. “Viewpoint variation” encompasses the dimensions of RGB and depth, while qualifiers such as infrared spectrum range or fusion settings serve to expand the schema without compromising cross-modality coherence.
- **Protocol awareness.** The taxonomy supports straightforward derivation and reporting of evaluation settings. Viewpoint variation can link to cross-view splits; time gaps can link to temporal generalization tests. This establishes a direct mapping from dataset attributes to experimental protocols. For example, defining cross-view evaluations from multi-angle recordings or cross-time tests from multi-session data.

The combination of these principles results in a schema that is compact, explicit about levels of variation, and readily translatable into machine-readable metadata.

4.3.3 Common Covariates in Vision- and Depth-Based Gait Datasets

Covariates are factors that influence the visual or kinematic appearance of gait but are not the primary variables of interest, such as identity, pathology, or performance metrics. They represent contextual or environmental conditions that may vary between or within recording sessions, including viewpoint, clothing, illumination, or walking surface. These variations alter the apparent motion and silhouette representation, thereby affecting the reliability and comparability of recognition systems (Khan et al. (2021)). Covariates can be categorized as active or static, depending on whether they vary deliberately within a dataset or remain fixed throughout acquisition. Active covariates introduce meaningful diversity used to assess robustness, whereas static covariates define the controlled parameters under which data were collected.

Illumination is a key factor that significantly affects the consistency of visual gait features. Changes in lighting direction or intensity, or the presence of shadows, can modify silhouettes and distort motion cues, particularly in outdoor environments or other situations where lighting cannot be controlled. Datasets such as GREW and FVG incorporate illumination variation in order to emulate realistic surveillance scenarios in which lighting cannot be controlled. In Verlekar’s research on shadow silhouettes demonstrates that the geometry of the light source and its relationship with the camera and subject affects the visibility of distinctive body regions (Verlekar et al. (2018)). Methods that employ shadow suppression, photometric normalization or illumination-invariant descriptors mitigate these effects to a certain extent, but generalization under strongly time-varying light remains challenging.

Background complexity introduces an additional dimension of variation in appearance, particularly in scenes involving crowds or cluttered structures. Non-homogeneous or dynamic backgrounds can make foreground extraction challenging, resulting in fragmented or noisy silhouettes. Empirical studies have demonstrated that recognition accuracy is lower in complex scenes than in controlled laboratory settings. This highlights the need for robust background subtraction and

spatio-temporal segmentation that is tailored to footage captured using a moving camera or in an outdoor environment (Hofmann et al. (2014); Lu et al. (2022b)).

The **walking surface** directly affects the gait dynamics. Differences between treadmill and overground walking, or between firm and uneven terrains such as grass, sand or gravel, alter stride parameters, foot-ground interactions and, consequently, visual signatures (Makihara et al. (2012); Khan et al. (2018)).

The **environmental setting**, including whether the setting is indoor or outdoor, influences the stability of illumination, the extent of background clutter, and exposure to weather and seasonal effects. Indoor collections like CASIA-B allow tighter control of acquisition, whereas outdoor datasets such as GREW introduce broader sources of variability, which often interact with each other. Studies involving repeated sessions demonstrate that variations in environmental context over time intensify intra-subject variability and result in significant reductions in identification accuracy.

Viewpoint variation is one of the most significant covariates in vision-based gait analysis. Changes in camera angle and subject orientation produce perspective distortions and self-occlusions, which result in significant differences between classes. View transformation models, beginning with frequency-domain formulations and subsequent extensions, seek mappings between angles in order to align signatures (Makihara et al. (2006); Zheng et al. (2011)). Contemporary approaches use convolutional or generative models to learn view-invariant representations, with multi-camera datasets such as CASIA-B and OU-MVLP enabling systematic cross-view evaluation (Chao et al. (2019); Luo and Tjahjadi (2020); Thapar et al. (2018)). Despite progress, substantial angle disparities between the gallery and the probe remain challenging.

Occlusion, whether static due to obstacles or dynamic due to other pedestrians, disrupts silhouette continuity and the estimation of pose or cycle phase. The creation of a dedicated dataset with controlled occlusions has stimulated research into detecting and reconstructing missing regions, including techniques that combine silhouette inference with temporal models to recover plausible motion (Hofmann et al. (2014); Chen et al. (2021)). More recent occlusion-aware deep architectures attempt to localize occluded parts and impute them consistently across frames; however, performance still degrades as the extent of occlusion increases.

Temporal variation across recording sessions is a pervasive source of intra-subject change. Gaps between sessions spanning days, months, or years capture shifts in gait associated with clothing, footwear, emotional state, fatigue, and environmental differences. Longitudinal datasets such as the 'OU-ISIR Large Population: dataset Age' demonstrate that, as a behavioral biometric, gait evolves over time and recognition rates decline as gaps increase (Iwama et al. (2012); Lu et al. (2022b)). These observations motivate temporally robust modeling that can accommodate long-term drift.

The **carrying conditions** can alter both the appearance and the biomechanics. Backpacks, handbags, and other items can alter the way the arms swing, create asymmetry and obscure parts of the torso or limbs. These conditions are explicitly included in CASIA-B and OU-ISIR treadmill studies to facilitate controlled analysis (Yu et al. (2006b); Makihara et al. (2012)). Researchers

have developed strategies and generative models, including GaitGAN and its variants, to separate identity information from carrying-related appearance changes. However, these strategies have produced mixed results when the load is heavy or variable (Yu et al. (2017); Li et al. (2019)).

Accessories and clothing remain among the most influential covariates for silhouette-based recognition. Coats, loose garments, hats, and changes to hairstyle modify body contours and obscure joint motion. Clothing variation is embedded in the CASIA-B and OU-ISIR datasets, which demonstrate that clothing can severely compromise the effectiveness of appearance-based methods (Yu et al. (2006b); Khan et al. (2021)). Research has progressed from handcrafted descriptors and part-based weighting to deep networks designed for clothing invariance; however, significant accuracy losses persist for shapes that alter the appearance of the body, such as heavy outerwear (Yeoh et al. (2016); Luo and Tjahjedi (2020)).

Footwear variation can influence stride length, cadence, and ground contact patterns, with effects reported across vision- and sensor-based modalities. Although the impact on purely visual signatures is generally smaller than that of clothing or viewpoint, experiments demonstrate measurable differences between shoe types and barefoot conditions, particularly in high-resolution or pressure-sensor settings (Lythgo et al. (2009); Connor (2015)). Incorporating footwear diversity during acquisition improves ecological validity and supports the investigation of subtle kinematic cues.

Demographic factors such as age, sex, height, weight, and ethnicity influence gait characteristics due to anatomical and biomechanical differences. Longitudinal aging studies demonstrate systematic changes in cadence and posture over multi-year periods, highlighting the need for age-aware modeling and fair evaluation across demographic groups (Iwama et al. (2012); Wang et al. (2012)). These covariates highlight the interplay between population diversity and algorithmic generalization.

Walking pace alters stride length, cadence and support phases, thus changing appearance-based descriptors. Datasets such as the OU-ISIR Gait Speed Transition dataset, the Cross-Covariate Gait Recognition dataset, and the CASIA-C dataset provide slow, normal, and fast sequences to investigate cross-speed generalization. Approaches range from the linear normalization of speed-dependent features to mutual subspace learning, to specialized temporal encoding. However, accuracy typically declines as the speed differential between the gallery and the probe widens (Makihara et al. (2012); Xu et al. (2017a); Khan et al. (2021)).

The **walking trajectory** becomes more challenging to model when the path deviates from a straight line. Curved motion, turning, or free paths can cause misalignment of poses across frames and interact with viewpoint changes. Unconstrained datasets with varied paths, including recent large-scale outdoor resources, reveal these effects and highlight the need for trajectory-aware representations and tracking that retain gait phase information when heading changes occur (Huang et al. (2021)).

Sensor-related covariates shape the signal itself. The available cues depend on the sensor modality, with RGB, depth, infrared, LiDAR, and multimodal combinations offering different

trade-offs between appearance, geometry, and lighting independence. Depth and infrared sensors often improve robustness to illumination, but they can introduce depth noise or range limits (Khan et al. (2021)). Framerate affects temporal resolution: low frame rates reduce phase sampling and compromise cycle reconstruction. Phase synchronization and temporal super-resolution address mismatched or sparse sampling, while averaged templates provide a pragmatic alternative for extremely low rates (Mori et al. (2010); Makihara et al. (2011); Guan et al. (2013)). Image resolution influences contour precision and the ability to recover fine limb motion; low-resolution surveillance videos make silhouette extraction and part localization challenging. Camera height and subject distance influence perspective and the extent of self-occlusion. Miscalibrated multi-camera setups are affected by imperfect synchronization or geometric alignment, yielding cross-view phase inconsistencies. Finally, sensor noise and artifacts, including motion blur, compression loss and depth measurement errors, degrade feature stability and must be considered when interpreting performance across datasets.

In summary, the range of covariates includes illumination, background, surface and environment, viewpoint and trajectory, occlusion and time, subject appearance and state, and sensor and sampling characteristics. Understanding their influence is essential for characterizing datasets and interpreting model performance under changing conditions, since each factor can influence recognition and serve as an axis along which robustness can be evaluated.

4.3.4 Classification Dimensions

The proposed taxonomy organizes covariates into three primary classification dimensions: **scene-level**, **user-level**, and **sensor-level**, corresponding to the dominant sources of variability in vision- and depth-based gait datasets. These dimensions reflect the interaction between the physical environment, individual characteristics, and acquisition hardware, providing a structured framework for describing how each factor alters the appearance and measurability of gait. A concise overview of the categories and representative covariates is provided in Table 4.6. By isolating these three domains, the taxonomy ensures both parsimony and analytical clarity: each covariate is assigned to the level at which it exerts the most direct influence, while still acknowledging that interactions may occur across levels. This structure aligns with established design principles, which were already presented in Section 4.3.1, offering this way a balanced representation of variability without unnecessary complexity.

The first dimension, **scene-level**, encompasses external and contextual conditions that affect the visual capture of motion. These covariates originate from the physical environment and acquisition geometry, determining how the gait signal is projected into the visual domain. They include background complexity, illumination and lighting direction, walking surface, indoor or outdoor setting, camera viewpoints, occlusion type, walking trajectory, and temporal variation between sessions. Such factors influence both silhouette extraction and appearance consistency, shaping the baseline difficulty of recognition and interpretation. For instance, viewpoint diversity directly impacts cross-view evaluation protocols, while temporal variation captures longitudinal instability

Category	Active covariates (examples)
Scene-level	Background complexity; illumination (direction, intensity, shadows); walking surface (treadmill vs. overground; uneven terrains); environment (indoor vs. outdoor); view-points (single vs. multi-view; moving camera); occlusions (static vs. dynamic); walking trajectory (straight, curved, turning); temporal variation (session gaps, diurnal/seasonal changes).
[2pt] User-level	Carrying condition; accessories and clothing; footwear variation; demographics (age, sex, height, weight, ethnicity); gait style (normal, pathological, fatigue, stressed, etc.); walking pace (slow, normal, fast, variable).
[2pt] Sensor-level	Sensor modality (RGB, depth, infrared, LiDAR, multimodal); framerate / exposure (temporal resolution); image quality / resolution (compression artifacts); camera calibration; camera height and subject distance; multi-view synchronization / alignment; noise (depth errors, motion blur, compression loss, data loss).

Table 4.6: Proposed covariate-centered taxonomy for vision- and depth-based gait datasets. The table enumerates *active* sources of variation by category: scene, user, and sensor, providing representative examples to guide protocol design and dataset comparison.

across recording sessions. Scene-level covariates thus define the external envelope of the dataset’s variability and are often the focus of cross-condition generalization studies.

The second dimension, **user-level**, captures covariates intrinsic to the participant or immediate state, representing variations in appearance, motion dynamics, and physiological or behavioral conditions. These include carrying conditions, accessories and clothing, footwear, demographic characteristics, gait style, and walking pace. Such factors alter both the kinematic and visual representation of gait, affecting body contour, stride rhythm, and overall movement symmetry. User-level covariates are central to assessing robustness to intra-subject variability and are strongly associated with appearance-based distortions. Datasets such as CASIA-B or OU-ISIR have explicitly integrated these variations—through clothing changes, carrying loads, or speed variations—to facilitate controlled evaluation of subject-dependent effects. Within this dimension, the taxonomy also recognizes that user covariates may interact with scene parameters, for example when illumination or viewpoint exaggerates the visual impact of clothing or gait speed differences.

The third dimension, **sensor-level**, defines the covariates that originate from the characteristics and configuration of the acquisition hardware. This category encompasses modality type (RGB, depth, infrared, LiDAR, or multimodal fusion), framerate and exposure, image resolution, camera calibration, camera height and subject distance, synchronization across views, and sensor-induced noise. Sensor-level covariates dictate the spatial and temporal fidelity of the recorded signal and thus constrain the representational richness of gait features. Differences in framerate or synchronization can introduce phase mismatches across cameras, while compression or motion blur may obscure fine-grained limb motion. The inclusion of this dimension ensures that hardware and signal-processing factors are treated as first-class descriptors rather than incidental metadata. This facilitates consistent documentation of acquisition setups across datasets, aligning with principles of modality alignment and protocol awareness.

Together, these three dimensions form a coherent and scalable schema for representing the variability inherent in gait data. The taxonomy supports human-readable interpretation, enabling datasets to be systematically described, compared, and integrated across modalities. By explicitly linking covariates to the level at which they act (scene, user, or sensor) the framework bridges the gap between acquisition design and methodological evaluation. This structure lays the foundation for metadata structures that capture variability with precision and transparency, promoting reproducibility and interoperability in gait research.

4.4 A Review of Vision and Depth-Based Human Gait Datasets

4.4.1 Motivation and Scope of the Review

Vision-based gait datasets have expanded rapidly in volume, variety, and purpose, motivating an up-to-date, consolidated overview of publicly available resources. A new review is justified for mainly three reasons. First, access has become volatile: repositories appear, grow, migrate, and sometimes disappear or change licensing (e.g. Hofmann et al. (2014); DeCann et al. (2013)), impacting reproducibility and comparative analysis. Second, modalities and representations have progressed from early binary silhouettes toward RGB-D settings enriched with auxiliary annotations, including pose/joints, optical flow, and human parsing, enabling alternative training objectives and evaluation protocols (Zhu et al. (2021); Li et al. (2024b); Yin et al. (2025)). Third, cataloging perspectives and benchmarking taxonomies continue to diversify, ranging from modality-centric classifications (Singh et al. (2018)) to analyses organized by covariates (Li et al. (2024b)) or application domain (Chaquet et al. (2013)).

Dataset maintenance remains inconsistent, with broken links and changes in usage terms. This tends to happen as hosting shifts among institutional, project-based, and personal domains. Concurrently, the most recent collections often pair RGB with depth streams or derived data (e.g. point clouds, skeletons, body parsing) and increasingly publish richer labels that support evaluation protocol variants beyond same-view, single-condition identification, namely cross-view, cross-clothes, cross-speed, and cross-scene evaluations. Therefore, a systematic review and mapping of these characteristics is timely to guide dataset selection and support fair comparisons between methods.

The present work reviews publicly available, marker-free gait datasets acquired with vision sensors, spanning two families: (i) vision-based datasets recorded as monocular or multi-view RGB video and (ii) depth-based datasets that include depth maps and/or RGB-D streams from commercially available sensors (e.g. Microsoft Kinect, Intel RealSense). Inclusion requires the presence of gait information suitable for person identification/recognition, gait analysis, or clinical assessment. Datasets that are exclusively inertial, pressure-mat based, or optical-marker motion capture without accompanying visual frames are excluded.

This chapter builds upon the 2019 review of depth-based gait datasets (Nunes et al. (2019b)) by expanding the scope to include both RGB (vision-based) and RGB-D (depth-based) resources, as

well as incorporating datasets released or substantially updated up to 2025. This includes clinically oriented collections and in-the-wild benchmarks. In addition to modality coverage, the update provides a more in-depth analysis of auxiliary annotations and protocol design, reflecting current practices in training and evaluation.

The scope and intent of this chapter differ from those of broad RGB and RGB-D surveys, notably Wang et al. (2010b); Lee et al. (2013); Firman (2016). While Firman (2016) provides an overview of RGB-D datasets from up to around 2017 across various domains (objects, scenes, reconstruction, and tracking), this review focuses on human gait and extends the timeframe to 2025. The collection focuses on identity/recognition, movement analysis and clinical applications, organizing metadata in a gait-centric manner: subject counts, sequence counts, view geometry and multiplicity, covariate coverage (viewpoint, clothing, carrying conditions, footwear, surface, speed, indoor/outdoor, treadmill versus free-walk, etc.), sensor types and configurations, framerate and resolution, and the availability of auxiliary labels such as 2D/3D pose, skeletons, flow, and segmentation (Singh et al. (2018); Wang et al. (2010b); Lee et al. (2013)).

Several cross-cutting themes are addressed succinctly to inform the use of datasets. Marker-free acquisition currently dominates due to deployment realism, while marker-based datasets remain valuable for controlled analyses and validating pose-based methods. RGB collections are increasingly adopting multi-view and higher spatial resolutions, which enables cross-view protocols and view-invariant modeling. Depth sensors provide per-frame depth maps and derived 3D joints that facilitate kinematic descriptors and cross-modal supervision. Considerations of demographic data and bias are especially important for clinical datasets and recognition benchmarks, where imbalanced representation can make performance interpretation confusing. Ethical and privacy aspects intersect with dataset governance and release practices. A recurring emphasis is placed on transparent training, validation and testing design, reproducible splits, and cross-dataset generalization, which aligns with efforts to promote robust evaluation under covariate change.

4.4.2 Dataset Selection

The primary inclusion criterion was *free availability*, whether by direct public download from an official source or by request-based access under a standard data-use agreement with any required institutional approvals. This approach prioritizes reproducibility, independent verification, and equitable access. In addition, marker-free capture datasets were preferred for realism and compatibility with RGB and RGB-D representations.

The dataset review was conducted in two phases. Initially, the depth-oriented catalog presented in Nunes et al. (2019b) was used as the starting point, and each entry was re-verified for current availability and metadata consistency. Secondly, the search was expanded to encompass supplementary RGB-D resources and vision-based RGB datasets comprising gait information appropriate for identification, recognition, analysis, or clinical assessment. Preference was given to repositories maintained by dataset authors or host institutions. Several historical repositories were offline or exhibited broken links. In those cases, corresponding or first-listed authors were contacted. Responses fell into three categories: access granted with the necessary files or credentials;

notification that a new release was “under preparation”; or confirmation that the dataset was no longer available. Datasets without an accessible official website and for which no author provided access were excluded from this review.

To streamline navigation across the 52 surveyed datasets, the list is organized by primary intended use into three categories: *biometrics*, *healthcare*, and *attribute recognition*. **Biometrics** covers datasets designed for person identification/verification and surveillance-style evaluation, typically emphasizing subject scale, multi-view geometry, and covariate robustness (viewpoint, clothing, carrying). **Healthcare** comprises datasets oriented to clinical assessment, pathology simulation or monitoring, or task-specific analysis where diagnostic labels, event annotations, or protocol constraints are central. **Attributes recognition** targets soft-biometric or behavioral attributes (e.g., walking style, age, clothing/carrying states), often with multi-label annotations designed for attribute-aware modeling and analysis.

Quick index by field (for rapid access to descriptions).

The datasets are presented in alphabetical order in the subsequent subsections. Each description provides a concise, gait-centric profile including, where identifiable, subject and sequence counts; camera geometry and number of views; acquisition environment and sensor type(s); framerate/resolution; distributed modalities (e.g., RGB, depth, silhouettes, 2D/3D pose, parsing, optical flow); principal covariates (e.g., viewpoint, clothing, carrying, footwear, surface, speed, indoor/outdoor, treadmill vs. free-walk); and accessibility and maintenance status. When official sources define evaluation protocols, these are summarized (e.g. cross-view, cross-clothes, cross-speed, subject-disjoint splits) and, where available, the official *train/validation/test* partitions are also reported to facilitate comparisons across studies. Moreover, an illustrative figure is always presented, depicting the capture setup or exemplar data modalities. Ambiguities or undocumented fields are noted verbatim from the original sources, avoiding inference beyond the available documentation.

Healthcare

- Gait-A Database (Subsection 4.4.3)
- GAIT-IST and GAIT-IT (Subsection 4.4.3)
- Health&Gait (Subsection 4.4.3)
- INIT Gait Database (Subsection 4.4.3)
- KOA-PD-NM (Subsection 4.4.3)
- Multi-Modal Gait Symmetry (MMGS) (Subsection 4.4.3)
- ProGait (Subsection 4.4.3)

- Scoliosis1K (Subsection 4.4.3)
- SPHERE Walking Dataset (Subsection 4.4.3)
- Walking Gait Dataset (Subsection 4.4.3)

Biometrics (identification/recognition, surveillance)

- 360 Degree Gait Capture (Subsection 4.4.3)
- AVA Multi-View Dataset (Subsection 4.4.3)
- CASIA Gait Database Family (Subsection 4.4.3)
- Cloth-Changing benchmark for Person re-identification and Gait recognition (CCPG) (Subsection 4.4.3)
- CMU Motion of Body (MoBo) (Subsection 4.4.3)
- Cross-Covariate Gait Recognition (CCGR) (Subsection 4.4.3)
- Depth-Based Gait Dataset (Subsection 4.4.3)
- Frontal-View Gait (FVG) (Subsection 4.4.3)
- Gait3D-Parsing (Subsection 4.4.3)
- GREW (Subsection 4.4.3)
- GRIDDS (Subsection 4.4.3)
- Kyushu University Gait Databases (KY4D-A/B; KY IR) (Subsection 4.4.3)
- Multi-Height Gait (MHG) (Subsection 4.4.3)
- Occluded Gait Database (OccGait) (Subsection 4.4.3)
- OG RGB+D (Subsection 4.4.3)
- OU-ISIR Gait Database: Treadmill Dataset (A, B, D) (Subsection 4.4.3)
- OU-ISIR Gait Speed Transition (GaitST) (Subsection 4.4.3)
- OU-ISIR Large Population: Bag (OU-LP-Bag) (Subsection 4.4.3)
- OU-ISIR Large Population: Multi-View family (OU-MVLP, OU-MVLP-Pose, OU-MVLP-Mesh, OU-MVLP-OF) (Subsection 4.4.3)
- ReSGait (Subsection 4.4.3)

- SAIVT-Depth Gait Database (Subsection 4.4.3)
- SDUMLA-HMT (Subsection 4.4.3)
- SMVDU-Gait (Single and Multi) (Subsection 4.4.3)
- SOTON Human ID (Large and Small DB) (Subsection 4.4.3)
- SUSTech1K (LidarGait) (Subsection 4.4.3)
- SZU RGB-D Gait Dataset (Subsection 4.4.3)
- TUM-IITKGP (Subsection 4.4.3)

Other (attributes recognition)

- Multi-Attribute Gait (MA-Gait) (Subsection 4.4.3)
- OU-ISIR Large Population: Age (OULP-Age) (Subsection 4.4.3)
- RA-GAR (Subsection 4.4.3)

4.4.3 Dataset Profiles

360 Degree Gait Capture

The 360° Degree Gait capture dataset (Topham et al. (2023)) comprises video and sensor recordings of adult volunteers walking along an 8 m path across three experiments, each with six tasks. Two fixed-position cameras recorded each walk at ≈ 30 fps, and a wearable digital goniometer sampled motion for the thigh, upper arm, and head. The dataset contains 3,120 videos and includes frame-level 2D human-pose annotations with 75 keypoints, together with participant-level anthropometry (sex, height, weight, ethnicity). Recordings span indoor environments and a smaller outdoor subset with varied backgrounds and multiple viewpoints. A portion of sequences uses alternative clothing to introduce appearance variation. Figure 4.12 presents an overview of the capture setups.

The dataset design supports cross-subject identification via disjoint training/testing participants, cross-view generalization by reserving one or more camera viewpoints for testing, and cross-clothes evaluation by training on one outfit and testing on an alternative. Environment generalization can be assessed by training on indoor recordings and testing on outdoor recordings.



Figure 4.12: **360-Degree Gait dataset**: representative images from the different capture setups.

AVA Multi-View Dataset

The AVA Multi-View Dataset for Gait Recognition (López-Fernández et al. (2014)) contains recordings of 20 subjects captured by six synchronized cameras arranged around a studio capture volume. Each subject performs ten predefined trajectories comprising straight and curved paths, including a figure-eight. Videos are recorded at 640×480 and 25 fps, and binary silhouettes are provided for each sequence. The release totals 200 multi-view sequences, yielding 1,200 single-view videos. The setup introduces viewpoint and scale variation, and camera placement produces partial occlusions of the lower limbs in some views.

The dataset supports cross-view evaluation by training on a subset of the six cameras and testing on held-out cameras. It also enables comparisons between single-view baselines and multi-view fusion using synchronized views. Generalization across trajectory types can be examined by training on straight paths and testing on curved paths (and conversely). The provided calibration facilitates 3D-based protocols in which features reconstructed from multiple views are used for identification or verification. Figure 4.13 provides a visual representation of the studio layout and its viewpoints.



Figure 4.13: **AVA Multi-View Dataset**: studio layout with representative camera viewpoints.

CASIA Gait Database

The Institute of Automation, Chinese Academy of Sciences (CASIA), provides the *CASIA Gait Database* to support research in gait recognition and related tasks. The family comprises four complementary subsets: **CASIA-A** (Wang et al. (2003b)), **CASIA-B** (Yu et al. (2006a)), **CASIA-C** (Tan et al. (2006)) and **CASIA-E** (Song et al. (2022)), spanning outdoor and indoor scenarios,

systematic viewpoint variation, appearance changes (e.g. clothing and carried objects), and low-illumination conditions via infrared imaging.

CASIA-A

CASIA-A (formerly the NLPR gait database) (Wang et al. (2003b)) comprises recordings of 20 subjects walking outdoors along a straight path. Sequences were captured with a digital camera at 352×240 and 25 fps, yielding 240 sequences in total. Three fixed viewpoints are provided: lateral (0°), oblique (45°), and frontal (90°) – see Fig. 4.14. For each viewpoint, subjects were recorded in two left-to-right and two right-to-left passes, introducing direction variability alongside view-point change.

A commonly used cross-view protocol trains on two viewpoints and evaluates on the third to assess recognition across views. Cross-direction robustness can be examined by training on one walking direction and testing on the opposite. presents the three viewpoints.



Figure 4.14: **CASIA-A dataset:** three representative frames from the three viewpoints ($0^\circ, 45^\circ, 90^\circ$).

CASIA-B

The CASIA-B dataset (Yu et al. (2006a)) contains recordings of 124 subjects (93 male, 31 female) walking indoors against a simple background. Each subject is captured from 11 viewpoints spanning 0° to 180° in 18° increments. For every view, ten sequences per subject are provided under three conditions: normal NM#01–06, carrying a bag BG#01–02, and wearing a coat CL#01–02. Videos are recorded at 320×240 and 25 fps. In total, the dataset comprises 13,640 single-view sequences depicting straight-line walking with controlled variation in viewpoint, clothing, and carrying. The release includes RGB sequences and background-subtracted silhouettes.

Subject-independent cross-view protocols are standard. Typical splits use 74 subjects for training and 50 for testing (a 62/62 split is also reported). Evaluation commonly sets NM#01–04 as the gallery per view and uses NM#05–06, BG#01–02, and CL#01–02 as probes; identical-view pairs are retained. Results are reported per view and then averaged over NM, BG, and CL. Figure 4.15 illustrates the multi-view configuration.



Figure 4.15: CASIA-B: eleven representative frames, one per viewpoint (0° – 180° in 18° steps), illustrating the dataset's 11-camera configuration.

CASIA-C

CASIA-C (Tan et al. (2006)) is an outdoor, night-time infrared (thermal) gait dataset comprising 153 subjects (130 male, 23 female), captured from a single lateral viewpoint ($\approx 90^\circ$) at 320×240 and 25 fps. Figure 4.16 illustrates representative frames from the dataset. Each subject walked along a straight path during night hours and was recorded under four conditions: normal, slow, fast, and normal with a bag – yielding ten sequences per subject. The dataset targets recognition in low-illumination settings by providing infrared image sequences. Silhouettes are also distributed as a derived modality.

Reported evaluations include normal-only identification, in which training uses a subset of normal sequences and testing is performed on disjoint normal sequences (often with cross-validation), and cross-speed/carrying protocols that train on normal walking and assess generalization to slow, fast, and bag conditions.

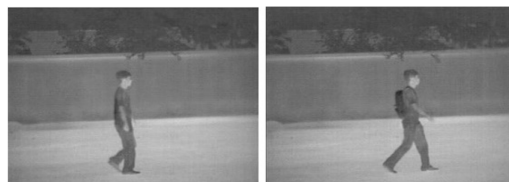


Figure 4.16: CASIA-C dataset: two representative frames from the outdoor, night-time infrared dataset.

CASIA-E

CASIA-E (Song et al. (2022)) is a large outdoor gait dataset comprising 1,014 participants (507 male, 507 female), collected over five months across three scenes with different background dynamics. The release provides silhouettes, Gait Energy Images (Han and Bhanu (2006)), and a thermal subset, together with soft-biometric metadata (age, sex, height, weight, nationality). View coverage combines 13 horizontal angles from 0° to 180° at 15° increments with two vertical tilts, for a total of 26 viewpoints. Sequences include normal walking as well as appearance variations with coat and bag, amounting to 778,752 video segments. Figure 4.17 shows a scene captured from multiple viewpoints.

Subject-independent evaluation is standard. The 3rd Competition on Human Identification at a Distance – HID-2022 (Yu et al. (2022a)) used CASIA-E subsets with 500 identities for training and 505 for testing, with pre-defined gallery/probe selections under NM/BG/CL conditions (one gallery and nine probes per identity). Other published protocols split the 1,014 subjects into 500/114/400 (train/validation/test) or 200/814 (train/test), enabling cross-view and cross-appearance assessment under consistent scenes and capture settings.

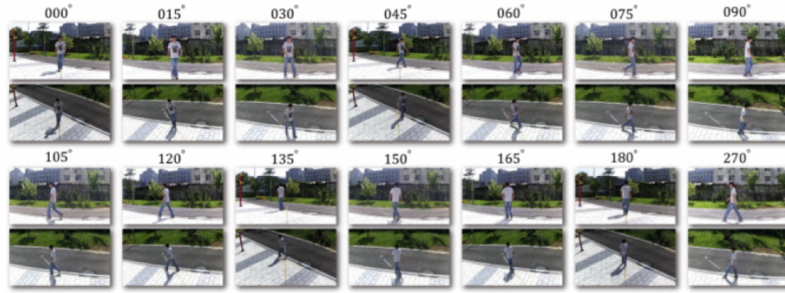


Figure 4.17: CASIA-E dataset: multiple images of the same scene captured from different view-points.

Cloth-Changing benchmark for Person re-identification and Gait recognition (CCPG)

The CCPG benchmark (Li et al. (2023b)) comprises recordings from 200 subjects (122 male, 78 female), totaling 16,566 video segments collected in indoor and outdoor scenes across 10 camera views. Data are released in RGB and silhouette modalities and organized into five subsets: four RGB variants – CCPG-A (complete RGB), CCPG-B (RGB without face), CCPG-C (RGB without shoes), CCPG-D (RGB without face and shoes); and CCPG-G (silhouettes). Covariates are explicitly annotated, spanning seven clothing states per subject (e.g.: full-outfit change, tops-only, pants-only, bag-carrying), multi-camera viewpoint differences, occlusion, walking-trajectory variation, scene type, and illumination change. Acquisition aims for less constrained conditions than laboratory setups, introducing diverse backgrounds and scene transitions that affect visibility and apparent scale. Identity-consistent segmentations and aggregated subject metadata (e.g.: sex) are provided, enabling unified evaluation of gait recognition and video-based person re-identification under clothing and related covariates. Figure 4.20 shows representative camera views and clothing states.

In line with Li et al. (2023b), evaluation adopts cross-camera retrieval, where queries and gallery are drawn from different cameras. Identities 0–99 are used for training and 100–199 for testing. Clothing variation (CL) is assessed under three settings: CL-Full (full outfit change), CL-UP (top changed), and CL-DN (bottom changed). Reported baselines cover both person re-identification and gait-based methods, and the silhouette subset (CCPG-G) supports direct gait comparisons under the same split.

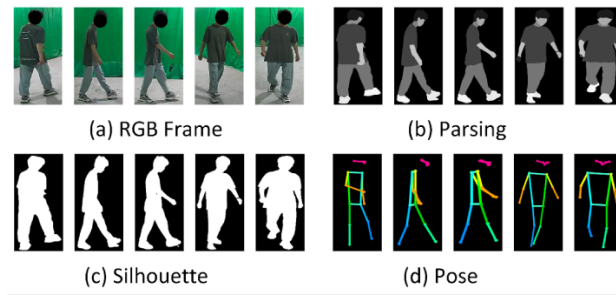


Figure 4.18: CCPG: representative camera views and clothing states across indoor and outdoor scenes.

CMU Motion of Body (MoBo)

The CMU Motion of Body (MoBo) dataset ([Gross and Shi \(2001\)](#)) contains treadmill recordings of 25 subjects (23 male, 2 female) captured indoors with six synchronized, calibrated color cameras arranged evenly around the treadmill, providing near-full 360° coverage. Each subject performs four walking patterns: slow, fast, incline, and walking while carrying a ball. Sequences are 11 s long at 30 fps and 640×480 resolution. The release includes precomputed silhouettes, camera-calibration data, and background images for subtraction. Principal covariates include viewpoint change (six views), walking speed (slow vs. fast), surface change via incline, and a carrying condition that restricts arm swing (ball). Subject-level metadata include sex, age, and weight, and treadmill speed is recorded per sequence.

Early evaluations on MoBo adopt cross-view and cross-condition protocols with gallery/probe splits; for example, [Collins et al. \(2002\)](#) train on slow-walk sequences and test on fast or ball conditions, reporting results for profile and frontal views, with all sequences acquired in a single session per condition. These practices motivate training on one camera and testing on the remaining views, as well as cross-condition tests within the four patterns. Figure 4.19 shows the six viewpoints.



Figure 4.19: CMU Motion of Body (MoBo): representative frames from the six synchronized viewpoints around the treadmill (frontal, rear, and four oblique/lateral views).

Cross-Covariate Gait Recognition (CCGR)

The Cross-Covariate Gait Recognition dataset (Zou et al. (2024)) is a large-scale indoor resource for cross-covariate gait recognition. It contains 970 subjects and $\approx 1.58 \times 10^6$ sequences captured at 25 fps by a 33-camera array arranged in five height tiers, including an overhead view, to emulate typical CCTV placement. Each subject performs 54 walks spanning 53 single or mixed covariates, with views and conditions explicitly annotated. The release provides RGB video, silhouettes, human-parsing masks, and 2D poses. Covariates cover surface and behavior factors (ramp/stairs, soft/bumpy ground, fast pace, stationary) and carrying conditions (bag, box, umbrella, trolley case, etc.). Subject-level metadata (age, sex) and identity labels enable per-covariate and per-view analyses. Figure 4.18 illustrates a temporally aligned sequence of RGB frames with the corresponding silhouettes, 2D keypose representations, and human-parsing masks.

Proposed evaluation uses a subject-disjoint split with identities 1–600 for training and the remainder for testing, while identities 134–164 are unused. Results are reported under two settings: *easy*, where the gallery comprises normal-walking sequences, and *hard*, where the gallery condition is uncertain. Cross-view testing follows the OU-MVLP protocol (see Subsection ??). A compact CCGR-MINI subset (970 subjects; 47,884 sequences) is also provided for lightweight benchmarking.



Figure 4.20: CCGR dataset: sequence of RGB frames with corresponding silhouettes, 2D poses, and human-parsing masks, temporally aligned.

Depth-Based Gait Dataset

The Depth-Based Gait Dataset (Chattopadhyay et al. (2015)) contains recordings from 29 subjects using two Kinect sensors mounted at 2.5 m with a downward tilt of $\approx 27^\circ$. For each subject, two unoccluded front-view sequences at 30 fps are provided and, by default, used as the gallery. The test set varies four covariates in a factorial design: viewpoint (front vs. back), walking speed (normal vs. fast), occlusion (unoccluded vs. occluded), and framerate (30 fps vs. 15 fps). Concretely, there are eight condition combinations per subject for each framerate (all combinations of viewpoint, speed, and occlusion), yielding 232 test sequences at 30 fps and 232 at 15 fps, for a total of 464 tests. Distributed modalities include depth maps and Kinect 2D and 3D skeletal joints,

with annotations for subject identity, view (front/back), speed (normal/fast), occlusion (unoccluded/occluded), and framerate. Figure 4.21 shows a sample depth-map frame with the corresponding Kinect skeletal joints overlaid, illustrating joint locations.

In the protocol of [Chattopadhyay et al. \(2015\)](#), the gallery comprises two unoccluded front-view sequences at 30 fps. Test pairs are constructed by selecting one front-view and one back-view sequence that satisfy the target condition. For example, under the front-only occlusion setting at normal speed and 30 fps, the pair combines a back-view, normal-speed, unoccluded clip at 30 fps with a front-view, normal-speed, occluded clip at 30 fps. Four occlusion conditions are evaluated: no occlusion in either view, occlusion in the back view only, occlusion in the front view only, and occlusion in both views – each tested at 30 fps and 15 fps and at normal and fast speeds.



Figure 4.21: Depth-Based Gait Dataset: sample depth-map frame with overlaid skeletal joints, illustrating joint locations.

Frontal-View Gait (FVG)

The Front-View Gait (FVG) dataset ([Zhang et al. \(2019\)](#)) comprises 2,856 RGB videos from 226 subjects, recorded outdoors at 1920×1080 and 15 fps. Each subject walks toward a fixed camera under three near-frontal angles ($-45^\circ, 0^\circ, +45^\circ$). A blurred-face release (FVG-B) is provided to mitigate privacy concerns. Key covariates include viewpoint offsets, walking speed, carrying conditions, clothing and footwear changes, background-clutter with occasional occlusion, and temporal separation across sessions. Acquisition uses a single RGB camera mounted on a 1.5 m tripod. Figure 4.22 illustrates sample frames of participants approaching the camera at the three angles under representative clothing and carrying conditions.

For benchmarking, the authors define five protocols with shared settings: the first 136 subjects are used for training and the remaining 90 for testing, and normal frontal-view walking is used as the gallery. The probe sets target variation in walking speed, carrying bag, changing clothes, cluttered background, and an *All* condition. Detailed protocol definitions appear in [Zhang et al. \(2019\)](#).



Figure 4.22: Front-View Gait (FVG): sample frames of participants approaching a fixed camera at three near-frontal angles under representative clothing and carrying conditions.

Gait-A Database

The Gait-A Database (formerly DAI Gait) ([Nieto-Hidalgo et al. \(2016\)](#)) comprises sequences of five subjects walking along an 8 m indoor corridor, recorded simultaneously from frontal and lateral viewpoints. Each sample includes a folder of binary silhouettes and metadata files with frame-level gait events (heel-strikes and toe-offs) together with a binary gait-class label encoded as 0 (normal) and 1 (abnormal). The targeted covariates are gait type (normal versus simulated abnormal) and viewpoint. Owing to synchronized multi-view acquisition and precise event annotations, the dataset is suited to healthcare applications such as screening and monitoring of gait abnormalities, quantifying deviations from normal gait, and detecting clinically relevant events for assessment and rehabilitation.

Published studies typically evaluate cross-view and cross-condition performance, training on one view and testing on the other, and training on normal sequences while testing on abnormal sequences. The provided heel-strike and toe-off labels are also leveraged for cycle segmentation. Figure 4.23 shows a representative view of the indoor corridor capture setup.

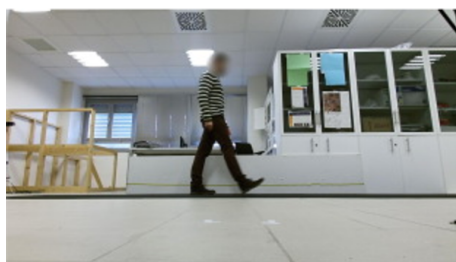


Figure 4.23: Gait-A Database: representative RGB frame of a person walking along an indoor corridor in side-view.

GAIT-IST and GAIT-IT

GAIT-IST and GAIT-IT are companion datasets from the Multimedia Signal Processing Group at Instituto de Telecomunicações, Instituto Superior Técnico, designed to study gait under simulated pathological conditions.

GAIT-IST (Loureiro and Correia (2020)) comprises 360 sequences from 10 volunteers (8 male, 2 female) recorded indoors with a single smartphone from a sagittal (side) view. Source videos are 1280×720 at 25 fps and each sequence is distributed as binary silhouettes together with Gait Energy Images (GEIs; Han and Bhanu (2006)) and Skeleton Energy Images (SEIs; Loureiro and Correia (2020)). Figure 4.24a shows an indoor RGB frame of a walking subject side-by-side with the corresponding 2D skeleton representation.

GAIT-IT (Albuquerque et al. (2021a)) expands this setup to 828 sequences from 21 volunteers (19 male, 2 female) captured in a professional studio with a chroma-key green background and controlled illumination using two synchronized cameras: a lateral view positioned 3 m from the walking path and a frontal or rear view near the start position. Distributed representations include binary silhouettes, skeleton images, Gait Energy Images (GEIs), and Skeleton Energy Images (SEIs). Silhouettes and GEIs/SEIs are provided at 224×224 , derived at 10 fps. Figure 4.24b shows an RGB frame of a walking subject captured in a professional studio with the corresponding silhouette.

Both datasets treat gait type as the primary covariate, encompassing normal gait and four simulated pathological patterns: diplegic, hemiplegic, neuropathic, and parkinsonian. Each of these pathological patterns is recorded at two severity levels. Additional controlled factors shared by both sets include walking direction (two directions) and repeated trials per configuration. GAIT-IT further introduces multi-view variation (lateral and frontal/rear), whereas GAIT-IST focuses on a single lateral view.

For evaluation, GAIT-IT reports subject-disjoint 10-fold cross-validation and a cross-dataset setting in which models are trained on GAIT-IT and tested on GAIT-IST to probe generalization across acquisition conditions (studio with dual views versus single-camera smartphone). This pairing enables consistent benchmarking of pathological-gait classifiers under aligned label spaces while systematically varying capture setups and covariates.

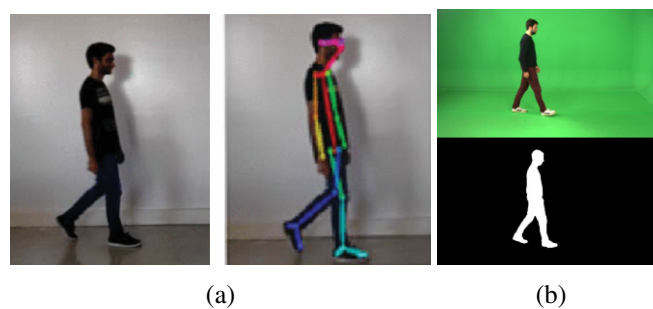


Figure 4.24: GAIT-IST and GAIT-IT datasets: (a) indoor RGB frame of a walking subject shown side-by-side with the corresponding 2D skeleton; (b) RGB frame of a walking subject captured in a professional studio with the corresponding silhouette.

Gait3D-Parsing

Gait3D-Parsing (Zheng et al. (2023b)) extends the Gait3D in-the-wild benchmark (Zheng et al. (2022)) by providing parsing-based representations (“Gait Parsing Sequences”) derived from su-

permarket surveillance videos. The base Gait3D release includes 4000 subjects and 25 309 sequences captured by 39 cameras at 1920×1080 and 25 fps; distributed modalities comprise silhouettes, 2D/3D keypoints, and 3D meshes/SMPL (Loper et al. (2023)). Gait3D-Parsing adds fine-grained human parsing masks to enable parsing-driven methods. Figure 4.25 illustrates example RGB frames alongside their corresponding human-parsing masks. Relevant covariates include cross-camera viewpoint changes across 39 devices, occlusion in crowds, dynamic walking trajectories, and irregular walking speed, together with appearance factors such as clothing, carrying, and accessories typical of supermarket scenes; these are emphasized as key challenges for *in-the-wild* gait recognition.

Following the official Gait3D evaluation (Zheng et al. (2022)), the protocol adopts a 3 000/1 000 train/test split. At test time, one sequence per subject forms the query set, and all remaining sequences constitute the gallery.



Figure 4.25: Gait3D-Parsing dataset: example RGB frames with corresponding human-parsing masks.

GREW

The GREW dataset (Zhu et al. (2021)) contains 26,345 subjects and 128,671 sequences captured by 882 cameras deployed in large public areas at a resolution of 1080×1920 . The released modalities comprise silhouettes, GEIs, optical flow, and 2D/3D human keypoints estimated by Sun et al. (2019) and Chen and Ramanan (2017), with manual identity annotations and more than 14M bounding boxes produced during preprocessing. A distractor set with 233,857 sequences is provided for large-scale retrieval evaluation. Key covariates include viewpoint changes across hundreds of cameras, occlusion and background clutter, irregular walking speed, and dynamic walking routes. Sequence-level attributes include soft biometrics (age, sex) and appearance factors such as carrying (none, backpack, shoulder bag, handbag, lift-stuff) and clothing (six upper/lower categories), enabling stratified analyses of performance under varied conditions. Figure 4.26 illustrates representative frames from the dataset.

The official evaluation splits 20,000 identities for training, 345 for validation, and 6,000 for testing; each test identity has four sequences, with two used as probes and two placed in the gallery. Identities across the three subsets are captured by different cameras to ensure cross-view evaluation. Rank- k retrieval (for example, Rank-1, Rank-5, Rank-10, and Rank-20) is used as

the primary metric, with optional evaluations that include the distractor set for scalability analysis (Guo et al. (2025a)).



Figure 4.26: GREW: representative frames across public spaces and viewpoints, illustrating crowding, occlusion, clothing and carrying variation, and illumination change.

GRIDDS

The Gait Recognition Image and Depth Dataset (GRIDDS) (Nunes et al. (2019a)) comprises 35 volunteers recorded indoors with a fixed Kinect v2 mounted at 1.8 m and oriented orthogonally to two straight walking trajectories (left to right and right to left). Each participant performs 10 sequences (5 per direction) at a self-selected pace. The distributed modalities include RGB, depth, infrared, silhouettes (aligned and normalized to 80×120), and 2D/3D joint coordinates. Gait events were manually annotated, and spatio-temporal parameters were calculated from sensor-derived data included in the release. The primary covariate by design is the walking trajectory (two directions) under a fixed lateral viewpoint. Annotations cover subject identifiers, gait event timestamps, spatio-temporal parameters, and soft biometrics (age, sex, height), enabling both appearance-based and model-based analyses. A per-participant directory structure and consistent file naming facilitate synchronization across modalities and alignment with the manual event labels and computed parameters. Figure 4.27 shows some representative frames and derived modalities from the dataset.

To standardize the evaluation, two protocols are proposed. First, a leave-one-subject-out identification setting: in each fold, construct the gallery with the first two sequences per direction for all subjects and use the remaining three per direction as probes; report Rank-1/Rank-5 and mAP. Second, a cross-trajectory generalization protocol: train and enroll the gallery using only left-to-right sequences and probe with right-to-left (and conversely), keeping subjects disjoint between training and testing and reporting the same metrics.

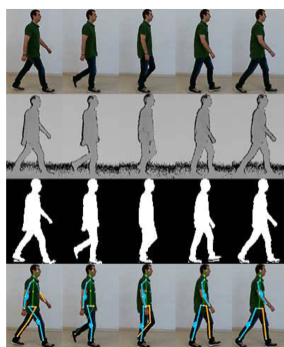


Figure 4.27: GRIDDs: representative frames with derived modalities, including RGB, depth, silhouettes, and 2D/3D joint coordinates.

Health&Gait

The Health&Gait dataset ([Zafra-Palma et al. \(2025\)](#)) comprises 1,564 side-view walking sequences from 398 adult participants collected indoors along a 10 m corridor. For each sequence, derived visual modalities are provided, including 2D pose estimates (COCO-17 joints), silhouettes, semantic-segmentation maps, and optical-flow fields. Figure 4.28 illustrates representative frames and their corresponding modalities. Each participant record includes anthropometric (age, sex, height, weight, BMI, waist/hip/neck circumferences, body-fat percentage, lean mass) and spatio-temporal gait parameters obtained with OptoGait/MuscleLAB (see Section 2.4.1) and estimated from pose. The data is organized per subject with unique identifiers and side-view direction labels. Key covariates include walking speed, clothing condition (with jacket vs. without jacket), and walking direction (left-to-right versus right-to-left). Only a single lateral viewpoint is used.

Training and evaluation can follow the authors’ stratified four-fold subject-disjoint partitions targeting sex classification and weight/age regression. Official scripts generate train/validation/test splits with random-seed control and stratification by sex and age or BMI. More details are provided in [Zhang et al. \(2019\)](#).



Figure 4.28: Health&Gait: representative frames from each available modality, including 2D pose estimates, silhouettes, semantic-segmentation maps, and optical-flow fields.

INIT Gait Database

The INIT Gait Database (Ortells et al. (2018)) contains 160 side-view walking sequences from 10 healthy volunteers recorded in a studio with a chroma-key green background. Each subject contributed two sequences for each of eight simulated gait styles, all captured from a lateral viewpoint. Data are distributed as binary silhouettes only, extracted from RGB recordings to preserve privacy; the nominal frame size is 800×400 pixels. Figure 4.29 illustrates a representative RGB frame and the corresponding silhouette.

The dataset targets gait-type variability as the principal covariate. Eight labeled styles are provided: one normal pattern (nm) and seven simulated impairments affecting arm swing (right or left reduced to half swing, or no swing), step-length asymmetry (right or left half steps), and a global impaired pattern inspired by advanced Parkinsonian gait. These labels are reflected in the directory structure and support supervised evaluation of gait-type classification under a fixed viewpoint.

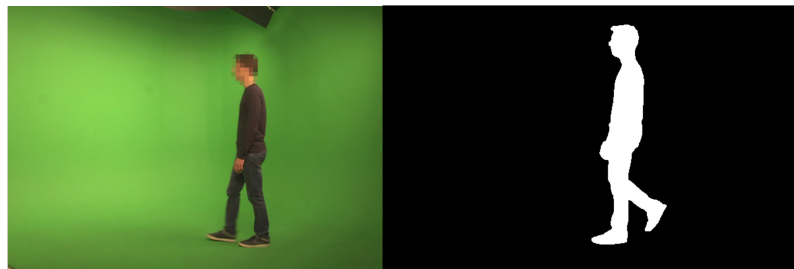


Figure 4.29: INIT Gait Database: representative RGB frame and a representative silhouette extracted via chroma-key.

KOA-PD-NM

The KOA-PD-NM dataset (Kour et al. (2021)) comprises 192 gait videos from 96 individuals recorded indoors in a clinical setting, including 50 patients with knee osteoarthritis (KOA), 16 with Parkinson's disease (PD), and 30 healthy controls (NM). Recordings were acquired with a camera placed at approximately 8 m from a 4.5 m walking mat against a green curtain, while participants walked naturally along a straight path; faces were subsequently masked to preserve privacy. Figure 4.30 shows a representative RGB frame of a participant walking. Videos are distributed in RGB at 50 fps, with two sequences per subject (left-to-right and right-to-left) captured from a lateral (sagittal) viewpoint.

Key covariates include gait type (KOA, PD, normal) and disease severity: KOA graded by Kellgren-Lawrence (early, moderate, severe) and PD by Hoehn & Yahr (mild, moderate, severe). Direction of travel (left-to-right versus right-to-left) is encoded at the sequence level. The acquisition setup allows analysis of both lower-limb and upper-body motion, with passive reflective markers visible to aid the identification of anatomical landmarks during capture.

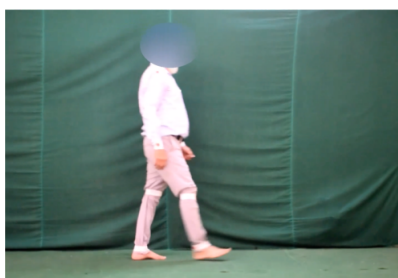


Figure 4.30: KOA-PD-NM dataset: representative RGB frame of a participant walking in the capture setup.

Kyushu University Gait Databases

The Kyushu University gait resources comprise complementary datasets acquired indoors under controlled conditions but with distinct sensing setups and trajectory designs. The KY4D releases (A and B) provide multi-view RGB imagery, silhouettes, and sequential visual-hull 3D reconstructions from a ring of 16 synchronized cameras, enabling cross-view and trajectory-aware analyses under matched capture parameters. The KY IR (Shadow Gait) dataset records top-view infrared shadow silhouettes under controlled illumination with appearance changes (bag and clothes), supporting robustness studies with top-view geometry and appearance variation.

KY4D – A (straight) and KY4D – B (curve). Kyushu University 4D Gait Databases (KY4D-A/B) (Iwashita et al. (2010, 2014b)) capture multi-view gait in a controlled studio using a ring of 16 synchronized cameras positioned at approximately 3.5 m from the scene center. Recordings are 1032×776 at 15 fps. Each release provides RGB frames, binary silhouettes, calibrated camera parameters, and sequential visual-hull 3D models. KY4D-A includes 42 subjects walking along a straight path with four sequences per subject, whereas KY4D-B includes the same 42 subjects walking along curved paths with two sequences per subject; curvature varies with two radii, 1.5 m and 3.0 m. Covariates center on viewpoint change and walking trajectory. Figure 4.31a illustrates, on the top row, a gait pose captured simultaneously by three distinct cameras and, on the bottom row, a sequence of three frames along a curved trajectory.

For evaluation on straight paths, images from three designated cameras at approximately 0° , 45° , and 90° are commonly selected, applying a leave-one-out rule per camera in which the 3D models corresponding to the test sequence are removed from the database; the procedure is repeated across all sequences. For curved-path tests, images from a fixed camera are used for both circular trajectories with radii 1.5 m and 3.0 m. The availability of multi-view 3D reconstructions enables benchmarking of 3D-based methods alongside appearance-based approaches.

KY IR (Shadow Gait Database). The Kyushu University Infrared Shadow Gait Database (Iwashita et al. (2014a)) contains time-series shadow silhouettes for 54 subjects walking along a straight path indoors. Two infrared illuminators project shadows onto the floor, and the scene is recorded by a side-view camera and a ceiling-mounted camera; the distribution focuses on silhouette sequences

associated with subject identities. Each subject contributes with six sequences: four standard, one carrying a bag, and one with changed clothing. Reported capture settings include 1600×1200 at 30 fps. Covariates emphasize clothing and carrying conditions under controlled infrared lighting, with a top-view geometry and a straight trajectory. The carrying condition encompasses backpacks, handbags, and travel bags, and clothing changes include down jacket, laboratory coat, and coat. Figure 4.31b depicts the two perspectives of a participant walking while carrying a bag.

Evaluation in the original studies applies four-fold cross-validation: for each subject, three standard sequences are used for training and the remaining standard sequence for testing; the single bag and clothing-change sequences are evaluated to assess robustness to appearance variation.

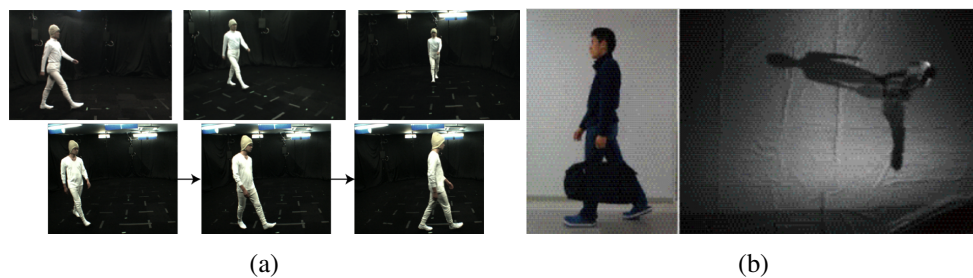


Figure 4.31: Kyushu University Gait Databases: (a) KY4D-A (top row) shows a gait pose captured simultaneously by three cameras; KY4D-B (bottom row) shows a sequence of three frames along a curved trajectory. (b) KY IR shows two perspectives (side-view and top-view) of a participant walking while carrying a bag.

Multi-Attribute Gait (MA-Gait)

Multi-Attribute Gait (MA-Gait) is an indoor, controlled-environment dataset for gait attribute recognition (GAR). It includes 95 adults (65 male, 30 female) recorded at 1920×1080 and 25 fps while performing 12 distinct walking styles designed to encode 16 attributes (multi-label per sequence). A six-camera rig arranged at 30° azimuth intervals around the walking route yields 12 effective views via two traversal directions (left-to-right and right-to-left). From the raw RGB videos, silhouettes and 2D skeletons are derived, and each sequence is annotated with sex and the multi-label attribute set. The release reports 13,954 sequences in total. As illustrated in Figure 4.32, example frames show different views and attributes side-by-side (four of the sixteen), highlighting viewpoint diversity and attribute visibility.

The official evaluation adopts a subject-disjoint split (45 identities for training, 50 for testing) and standard GAR metrics: mean accuracy (mA) at the attribute level and instance-level Acc/Prec/Rec/F1. Cross-view testing is supported across the 12 views, with one traversal serving as the gallery/reference and the other as the probe; analyses can be conducted on silhouette or skeleton modalities to compare representation effects under identical capture conditions.

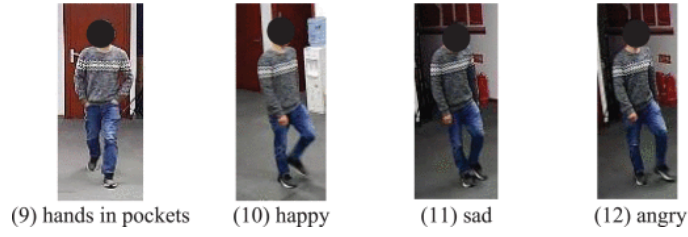


Figure 4.32: MA-Gait dataset: example frames across multiple viewpoints with four of the sixteen attributes shown side-by-side – hands in pockets, happy, sad and angry.

Multi-Height Gait (MHG)

The Multi-Height Gait (MHG) dataset (Hou et al. (2023)) captures 200 subjects in a controlled indoor environment using nine cameras arranged at three angles (0° , 45° , 90°) and three mounting heights (high, middle, low). For each subject, walking is recorded under three appearance conditions: normal (NM), carrying a bag (BG), and clothing change (CL), and silhouettes are released as the primary data modality – Fig. 4.33 illustrates some silhouettes across camera-height levels and appearance conditions. This configuration enables analysis of viewpoint variation due to both angle and camera height while maintaining a straight walking trajectory and a static background. The principal covariates include viewpoint change, clothing variation, and carrying conditions. The dataset is designed to analyze recognition performance when probe and gallery differ in camera height and to isolate the impact of clothing and carried objects on silhouette-based identification. Figure 4.33 illustrates representative silhouette sequences across camera heights and appearance conditions.

A proposed evaluation protocol trains on the first 150 subjects and tests on the remaining 50. During testing, one camera height is fixed as the gallery, and probes are partitioned by height (H/M/L) and condition (NM/BG/CL); mean Average Precision (mAP) is reported per subset to assess the effects of height mismatch and appearance changes.

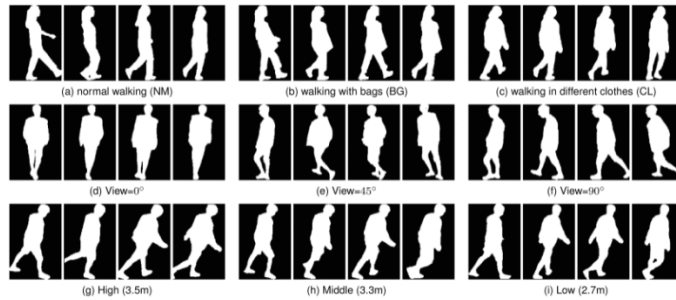


Figure 4.33: Multi-Height Gait dataset: representative silhouette sequences across camera-height levels and appearance conditions.

Multi-Modal Gait Symmetry (MMGS)

The Multi-Modal Gait Symmetry (MMGS) dataset (Khokhlova et al. (2019)) contains frontal walking recordings of 27 adults (19 male, 8 female) captured with a single Kinect v2 sensor under

controlled indoor conditions. Each subject performs three gait types: normal, a simulated limp induced by a 7 cm padding sole on the right shoe, and a simulated knee-rigidity condition achieved by preventing right-knee flexion. Released data include 3D skeleton joint coordinates and joint orientations, binary silhouettes, and processed joint-angle features; full depth streams and synchronized Vicon-Kinect acquisitions are noted as available upon request. Figure 4.34 illustrates representative frames for the available data modalities. Covariates focus on gait-type differences under a fixed frontal viewpoint and a controlled environment. The dataset supports evaluation of pose-based and silhouette-based methods for classifying gait abnormalities using joint orientations and derived kinematic angles.

A subject-disjoint protocol is reported (Khokhlova et al. (2019)): 19 subjects are used for training and validation (for example, 13/6) and 8 for testing, with random partitions repeated ≈ 23 times to assess robustness; an alternative fixed split by subject identifiers is also provided.



Figure 4.34: Multi-Modal Gait Symmetry dataset: representative frames for the different modalities, including depth, silhouettes and 3D skeletons.

Occluded Gait Database (OccGait)

The Occluded Gait Database (OccGait) (Huang et al. (2024)) is an indoor multi-view dataset designed to analyze gait recognition under occlusion. It includes 101 subjects recorded with three RGB cameras positioned at approximately 0° , 45° , and 315° , at a resolution of 1920×1080 , within an 8×8 m square walking course. By combining the fixed square route with the three camera placements and discarding overlapping or near-duplicate views, the public release defines eight effective viewpoints. The distribution provides silhouette sequences extracted from RGB data, with the raw RGB data excluded to protect privacy. Figure 4.35 illustrates an example RGB frame, its instance-segmentation mask, and the corresponding silhouette. Covariates emphasize occlusion and viewpoint change. Four occlusion scenarios are provided per subject: normal (NM) without obstacles; carrying (CA) with umbrellas or luggage; crowd (CR) involving interactions with another pedestrian in opposite or parallel motion; and static (ST) obstacles such as plants or chairs placed along the path. The fixed trajectory further induces partial and full-body occlusions that vary with viewpoint.

A subject-disjoint evaluation is reported in Huang et al. (2024): odd-numbered identities (51) for training and even-numbered identities (50) for testing. In testing, sequences NM01 and NM02 serve as the gallery for each subject, and Rank-1 accuracy is evaluated across occlusion types and viewing angles, following established gait-evaluation practice and the authors' protocol.

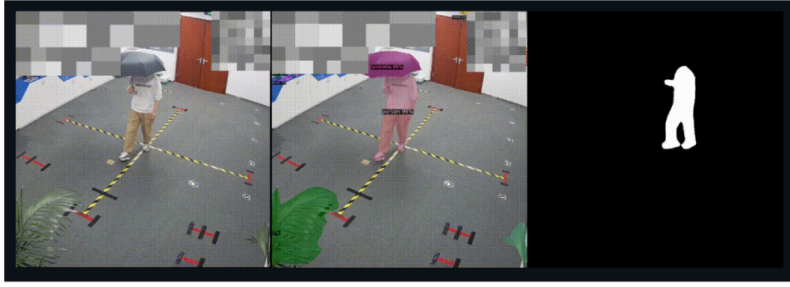


Figure 4.35: Occluded Gait Database (OccGait): example RGB frame, instance-segmentation mask, and corresponding silhouette.

OG RGB+D

OG RGB+D (Li and Zhao (2022)) was collected indoors using a synchronized multi-Kinect setup designed to study occlusion effects in gait recognition. It comprises 96 subjects (36 female, 60 male) recorded with RGB, depth, silhouettes, and camera-centric 3D joint data. The RGB resolution is 1280×720 . Single-gait sequences cover eight views induced by a square walking route and four turning views, whereas multi-gait sequences capture three people walking together under four straight-view settings. Covariates emphasize occlusions (carrying objects such as umbrellas or luggage and clothing with low, medium, or high occlusion), viewpoint changes, and trajectory types (straight versus turning), together with complex backgrounds and illumination variation. The annotated data further include soft-biometric attributes (age, sex, height, weight). Figure 4.36 presents representative frames illustrating multiple covariates and their corresponding modalities.

Training and evaluation reported by the authors split the 96 subjects into 48 for training and 48 for testing. Cross-view evaluation averages performance over non-identical gallery views, and cross-condition evaluation uses low-occlusion clothing (LCL) sequences as the gallery while probing carrying conditions (BOB, SOB, LOB), clothing conditions (MCL, HCL), and multi-gait (MG) scenarios. For MG, recognition is reported at single-, double-, and triple-subject accuracies per sequence.

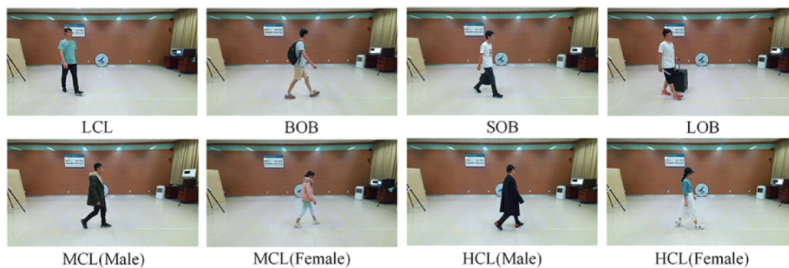


Figure 4.36: OG RGB+D dataset: representative frames across covariates (carrying, clothing-occlusion levels, straight vs. turning trajectories, viewpoint change).

OU-ISIR Gait Database: Treadmill Dataset

The OU-ISIR Gait Database, Treadmill Dataset is a series of treadmill-based gait resources captured indoors under controlled background and lighting conditions from a single lateral viewpoint ($\approx 90^\circ$). Across its subsets, silhouettes are the distributed modality, and each subset isolates a specific covariate, such as walking speed, clothing, or induced gait fluctuation, to enable targeted evaluation of robustness in gait recognition. Protocols are subject-disjoint, and the high-frame-rate acquisitions (≈ 60 fps) support rigorous analyses of both within- and across-condition generalization.

OU-ISIR Treadmill Gait Dataset A (speed variation) The OU-ISIR Treadmill Dataset A (speed variation) (Tsuji et al. (2010)) contains silhouettes for 34 subjects. Figure 4.37a shows a representative sample from the dataset. Recordings were obtained at discrete speeds from 2 to 10 km/h in 1 km/h increments. Evaluation is typically cross-speed, matching a gallery sequence at one speed against probe sequences at other speeds. The dataset comprises 612 sequences – about 18 per subject, i.e., roughly two sequences per subject at each speed.

OU-ISIR Treadmill Gait Dataset B (clothes variation) The OU-ISIR Treadmill Dataset B (clothes variation) (Altab Hossain et al. (2010)) includes 68 subjects. Figure 4.37b presents an example image from the dataset. Each subject appears under up to 32 clothing combinations, and all gait sequences were captured twice, yielding a total of 2,746 sequences. Silhouettes are distributed at high frame rate, and clothing is the key covariate. A subject-independent identification protocol is provided via predefined lists, using a gallery in normal clothes and probes in the remaining clothing conditions to enable cross-clothes evaluation. Results are typically reported with rank-based metrics (for example, Rank-1).

OU-ISIR Treadmill Gait Dataset D (gait fluctuation) The OU-ISIR Treadmill Dataset D (gait fluctuation) (Mori et al. (2010)) comprises 370 silhouette sequences from 185 subjects. Each subject contributes with two sequences, one serving as gallery and the other as probe. Silhouettes are normalized to 88×128 pixels and captured for approximately 6 s per sequence. This subset examines robustness to gait fluctuation introduced by post-processing: low framerate conditions are simulated via temporal subsampling, and low spatial-resolution conditions via downscaling of silhouettes. Evaluation follows a subject-disjoint identification framework to assess sensitivity to frame-rate and spatial-resolution changes.

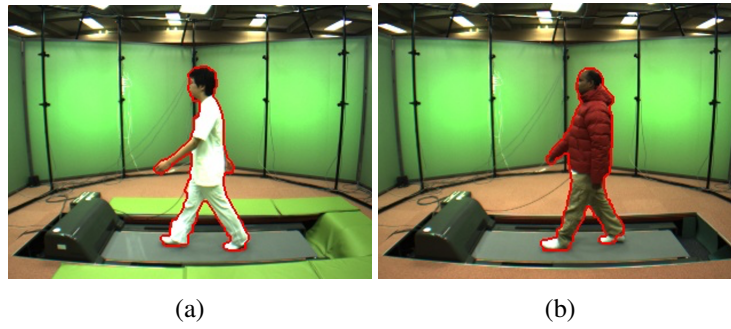


Figure 4.37: OU-ISIR Treadmill Gait Dataset: (a) sample frame from Dataset A showing treadmill walking; (b) sample frame from Dataset B showing treadmill walking while wearing a coat.

OU-ISIR Gait Speed Transition

The OU-ISIR Gait Speed Transition (GaitST) dataset (Mansur et al. (2014)) targets gait recognition under intrasequence speed changes. It releases size-normalized binary silhouettes of 22×32 pixels captured at 60 fps in controlled indoor conditions. Data include both overground walking and treadmill captures to cover near-constant speed and speed-transition scenarios. Figure 4.38 illustrates the two acquisition setups (overground and treadmill). The primary covariate is walking speed, with sequences exhibiting acceleration and deceleration within a single walk, complemented by constant-speed galleries.

The release is organized into two evaluation sets plus an auxiliary training set. In the first set, probes are speed-transition sequences from 26 subjects, whereas the gallery contains constant-speed sequences from 179 subjects recorded either on a treadmill at 4 km/h or overground at near-constant speed. In the second set, 25 subjects perform automatic treadmill speed transitions (accelerations from 1 to 5 km/h and decelerations from 5 to 1 km/h); one-second middle subsequences are extracted as probes, and the gallery comprises 154 subjects walking at 4 km/h. An auxiliary training set of 24 subjects provides constant-speed sequences at 2, 3, 4, and 5 km/h to learn speed-related models.

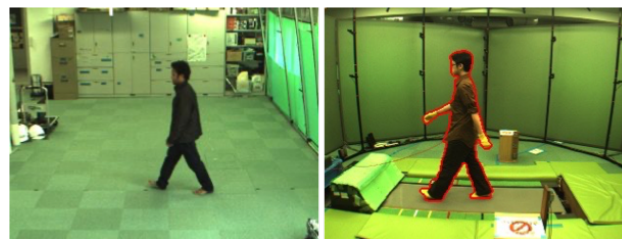


Figure 4.38: OU-ISIR Gait Speed Transition (GaitST): overground and treadmill acquisition setups with representative frames from each configuration.

OU-ISIR Large Population: Age

The OU-ISIR Large Population: Age (OULP-Age) dataset (Xu et al. (2017b)) provides one side-view walking sequence per subject and a size-normalized Gait Energy Image (GEI; 88×128)

for each identity. Data were collected in a controlled indoor corridor with a green chroma-key background using a single lateral camera at approximately 90° , recording 2.5 s clips at 640×480 and 30 fps. Figure 4.39 illustrates representative subjects of different ages during acquisition. The dataset comprises 63,846 participants with verified age (ages 2–90) and sex (31,093 male and 32,753 female) metadata, enabling gait-based age analysis. The release centers age as a soft biometric and distributes identity-level GEIs derived from background-subtracted silhouettes. Capture conditions keep viewpoint, trajectory, and speed consistent (straight path, side view, self-paced walking) to isolate age-related appearance factors.

Evaluation typically uses a subject-disjoint random split into two equal partitions, with 31,923 subjects for training and 31,923 for testing. Subsequent analyses vary the training-set composition by sex and age distribution while keeping the test set fixed; all experiments follow the same identity-disjoint protocol.



Figure 4.39: OULP-Age dataset: representative side-view samples from subjects of different ages during acquisition.

OU-ISIR Large Population: Bag

OU-LP-Bag is a large gait dataset (Uddin et al. (2018)) designed to study the carried-object (CO) covariate at scale. Gait sequences were collected with a fixed camera positioned at approximately 8 m from the walking course and 5 m in height, recording at 1280×980 and 25 fps. Each subject completed three passes: the first (A1), which involved carrying real-life objects, followed by two passes without carrying objects (A2 and A3). Silhouettes were extracted using a green-screen setup and converted into size-normalized Gait Energy Images (GEIs) (Han and Bhanu (2006)) of 88×128 pixels. In total, the dataset contains 62,528 identities, each with three sequences. Covariates center on the presence and placement of carried objects, with manual CO annotations into seven labels that coarsely encode typical positions or changes across a gait period (for example, side-bottom, side-middle, front, back, multiple/changed, none). Figure 4.40 shows examples of the carrying-status labels: (a) a sample RGB frame within a gait period with carried objects circled in yellow, (b) the corresponding GEI feature, and (c) the GEI of the same subject without a carried object in another captured sequence.

Following the authors' protocol, subjects are split into disjoint training and test identities of approximately equal size ($\approx 29,000$ each). Two evaluation settings are defined. In the cooperative setting, the gallery contains only no-CO sequences (A2 and A3), and the probes span the seven CO labels from A1. In the uncooperative setting, both gallery and probe include mixtures of with-CO

and without-CO sequences. Baselines are reported with appearance-based methods on GEIs under these settings.

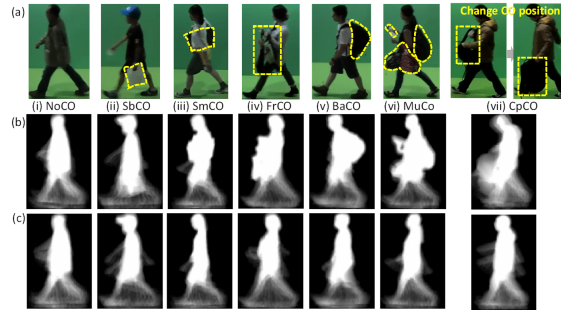


Figure 4.40: OU-LP-Bag dataset: examples of carrying-status labels; panels show RGB frames (carried objects highlighted) with the corresponding GEIs.

OU-ISIR Large Population: Multi-View

The OU-MVLP dataset family comprises four releases. The first release establishes the capture and evaluation setup, which the three subsequent releases reuse unchanged. Across all releases, the dataset comprises 10,307 subjects (5,114 male and 5,193 female). All data were recorded indoors under controlled background and illumination using a chroma-key wall and carpet. A ring of seven fixed network cameras was arranged along a quarter circle at 15° intervals, with a radius of ≈ 8 m and a height of ≈ 5 m. Some publications refer to “14 view angles” by treating the two walking directions as distinct, although the physical configuration consists of seven cameras. The cameras recorded at 1280×980 and 25 fps. Participants walked a straight path in both directions, each direction repeated twice, yielding up to 28 sequences per subject ($7 \times 2 \times 2$).

OU-ISIR Large Population: Multi-View (OU-MVLP) The OU-MVLP dataset (Takemura et al. (2018)) targets cross-view gait recognition at scale. It distributes silhouettes and size-normalized Gait Energy Images (GEIs) (Han and Bhanu (2006)) for the 10,307 subjects and emphasizes viewpoint variation across seven cameras, with natural intra-subject differences arising from bidirectional passes. Figure 4.41a illustrates representative GEIs released with the dataset.

OU-ISIR Large Population: Multi-View with Pose Sequence (OU-MVLP-Pose) The OU-MVLP-Pose dataset (An et al. (2020)) extends the base set with 2D pose sequences for the same 10,307 subjects. From the RGB streams, 18-joint trajectories are extracted using OpenPose (Cao et al. (2017a)) and AlphaPose (Fang et al. (2023)), producing two short clips per camera (one per traversal), typically ≈ 18 –35 frames each. This appearance-independent representation supports model-based cross-view studies. Figure 4.41b shows example 2D keypoint sequences.

OU-ISIR Large Population: Multi-View with Human Meshes (OU-MVLP-Mesh) The OU-MVLP-Mesh dataset (Li et al. (2022b)) augments the collection with parametric 3D human meshes

(SMPL; Loper et al. (2023)) and aligned 24-joint skeletons provided in two coordinate systems: IM2D (image-based 2D) and HC3D (human-centered 3D, view-robust). Sequence lengths mirror the pose release (≈ 18 –35 frames), enabling pose- and shape-level analysis beyond 2D keypoints. Figure 4.41c depicts sample human-mesh representations.

OU-ISIR Large Population: Multi-View with Dense Optical Flow (OU-MVLP-OF) The OU-MVLP-OF dataset (Shehata et al. (2025)) provides dense optical flow computed from the same RGB videos using RAFT – a deep network architecture (Teed and Deng (2021)). Two variants are offered: V1 computes flow on cropped, person-centered 256×256 RGB frames masked by silhouettes; V2 computes flow at the native 1280×980 resolution, then applies a foreground mask and recenters/resizes to 256×256 . Figure 4.41d presents representative dense-flow visualizations.

Across the family, evaluation follows the official subject-disjoint split (approximately half for training and half for testing, as proposed in OU-MVLP) (Takemura et al. (2018)). Cross-view identification is standard: one traversal serves as the gallery and the other as the probe; when reporting strictly cross-view results, identical-view matches are typically excluded. Performance is commonly summarized with Rank-1/Rank-5 (and mean Average Precision for retrieval-style reporting), enabling direct comparison of how different modalities (silhouettes, 2D poses, 3D meshes/skeletons, and dense optical flow) affect cross-view recognition under identical capture conditions.

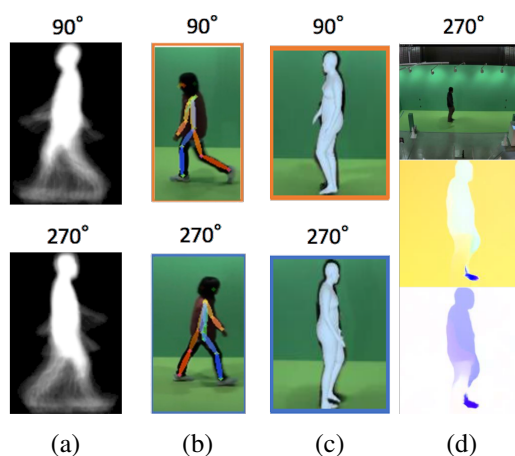


Figure 4.41: OU-ISIR Large Population: Multi-View family: (a) representative GEIs from OU-MVLP; (b) example 2D keypoint sequences from OU-MVLP-Pose; (c) sample human-mesh representations from OU-MVLP-Mesh; (d) representative dense-flow visualizations from OU-MVLP-OF.

ProGait

ProGait (Yin et al. (2025)) is a multi-purpose video dataset targeting vision-based analysis of transfemoral prosthesis users. It contains 412 clips from four male above-knee amputees, recorded indoors at 1920×1080 and 30 fps with paired frontal and sagittal views per trial. The distribution covers two capture scenarios: walking inside parallel bars and walking in hallways with assistance.

Multimodal annotations support three tasks: video object segmentation; 2D human pose estimation with 23 keypoints (seventeen body and six feet); and clinical text annotations comprising the general gait category, the specific deviation, alignment recommendations, and the underlying rationale. Figure 4.42 illustrates representative video frames from the frontal-view with segmentation masks and pose keypoints. Covariates emphasized include viewpoint changes (frontal vs. sagittal), occlusion caused by caregivers and surrounding objects, and scenario differences between parallel bars and open corridors.

For evaluation, the authors propose subject-disjoint splits of 70% training, 20% validation, and 10% test for the segmentation and pose tasks, and a five-fold cross-validation for gait classification, with results reported per view (frontal versus sagittal) and across scenarios.

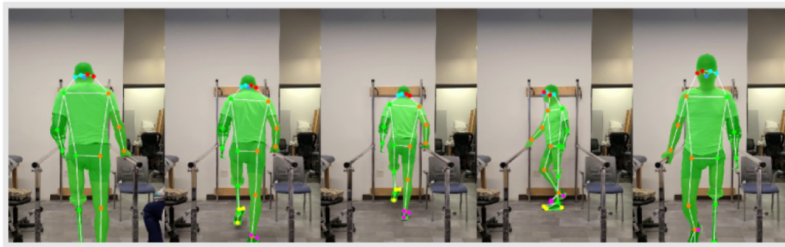


Figure 4.42: ProGait dataset: representative frontal-view frames with segmentation masks and 2D pose keypoints, sampled from parallel-bars and hallway scenarios.

RA-GAR: Richly Annotated - Gait Attributes Recognition

RA-GAR (Wang et al. (2025)) is a richly annotated outdoor gait dataset designed to study attribute-aware recognition across multiple covariates. It contains 533 subjects and 123,067 video sequences recorded at 1920×1080 and 30 fps using fixed cameras on an outdoor platform. The capture geometry yields 10 azimuth views (0° – 360° in 36° steps) from cameras positioned at 1.3 m height. For each subject, recordings cover normal walking and combinations of clothing change, and bag carrying. The public release provides silhouette sequences and 2D pose keypoints. Soft-biometric metadata (sex, height, weight, age) support attribute-level analyzes. Figure 4.43 illustrates examples of normal, bag-carrying, and clothing-change conditions alongside the corresponding data modalities.

The authors propose a subject-disjoint split with 250 training subjects (57,155 sequences) and 288 testing subjects (65,912 sequences). Evaluation follows attribute-recognition practice using mean accuracy (mA) and instance-based accuracy, precision, recall, and F1. Viewpoint diversity (10 angles) and appearance covariates (clothing, carrying, illumination) are explicitly controlled, enabling cross-view tests and robustness studies under realistic outdoor conditions.

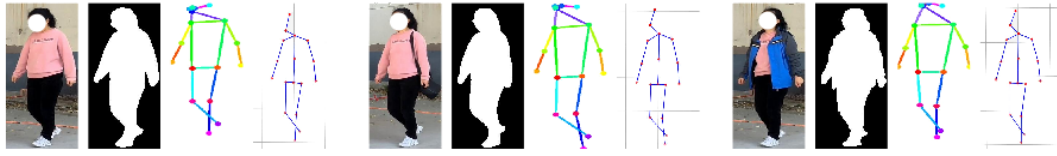


Figure 4.43: RA-GAR dataset: examples of normal, bag-carrying, and clothing-change conditions with the corresponding data modalities (silhouettes and 2D pose keypoints).

ReSGait

ReSGait (Mu et al. (2021)) is a gait resource recorded in an office corridor with a single fixed camera placed in a corner. It contains 172 subjects (134 male and 38 female) and 870 sequences collected over 15 months, with identity labels and paired silhouette sequences (normalized to 128×128) and pose sequences. Metadata include sex and per-sequence tags for clothing, carrying condition, walking route, and phone use. The camera height is approximately 1.2 m. The release emphasizes covariates such as long time span (seasonal changes across 15 months), clothing (for example, normal, coat, skirt), carrying (none, small item, large item, bag), walking route (straight or curved) and phone use. Figure 4.44 illustrates the two data modalities alongside representative examples of clothing, carrying, accessories, and walking-trajectory conditions.

The authors report three evaluation settings (Mu et al. (2021)): (i) an unconstrained split with subjects 1–86 for training and 87–172 for testing; (ii) covariate-focused analysis under the same subject split to assess the impact of clothing, carrying, route, and phone use; and (iii) a temporal setting in which training subjects have a time span under 30 days, whereas test subjects have at least 30 days between their earliest and latest recordings.

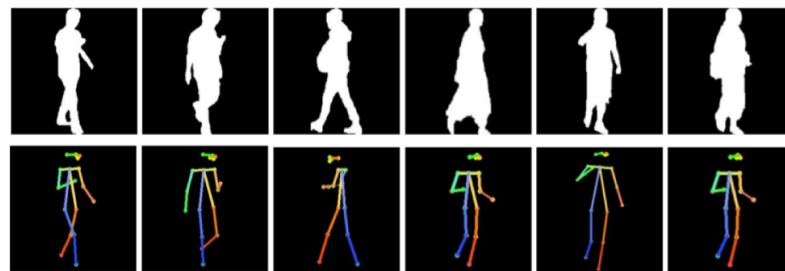


Figure 4.44: ReSGait dataset: paired silhouette and pose sequences with representative examples of clothing variation, carrying conditions, accessories, and walking-trajectory.

SAIVT-Depth Gait Database

The SAIVT-Depth Gait Database (SAIVT-DGD) (Sivapalan et al. (2011)) contains frontal, depth-based gait recordings of 35 subjects acquired with a Kinect sensor in a controlled indoor corridor. The public distribution provides depth frames, derived silhouettes, and voxel-based 3D volumes, along with sequence-level identifiers. Walking conditions vary across speed and appearance-related factors, enabling analyses under normal and fast walking, carrying (front, side, back),

and no-shoes cases. The principal covariates emphasized by the dataset are walking speed, carrying status, and footwear. Figure 4.45 illustrates a representative image sequence together with the corresponding depth frames.

Some published studies on this dataset (Sivapalan et al. (2011, 2013)) typically evaluate intra-class recognition using normal-versus-normal sequences, as well as inter-class scenarios in which fast-walking probes are matched to normal-walking galleries, with additional tests under carrying and no-shoes variations.

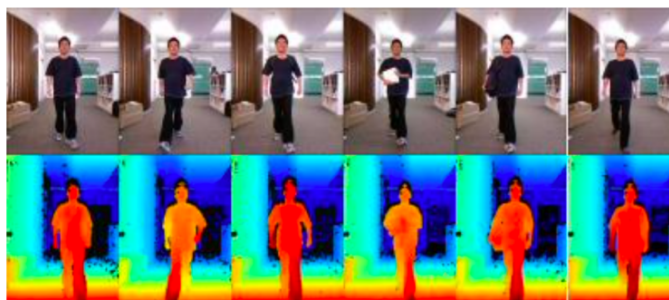


Figure 4.45: SAIVT-Depth Gait Database: representative frontal-view image sequence with the corresponding depth frames.

Scoliosis1K

Scoliosis1K (Zhou et al. (2024)) is a large-scale dataset for video-based scoliosis classification comprising 1,050 adolescents (641 male and 409 female) and 1,493 walking sequences. Videos were recorded in a school corridor with a single fixed camera and converted into privacy-preserving silhouette sequences at 1280×720 and 15 fps; 2D pose data (COCO-style keypoints) are also provided. Each subject is assigned one of three diagnostic categories (positive, neutral, or negative) based on on-site expert assessment using standard-screening procedures. The capture uses a frontal view with a straight walking trajectory. The principal covariate emphasized by the release is the diagnostic category. Figure 4.46 illustrates frontal-view silhouettes together with COCO-style 2D pose sequences.

For evaluation, Zhou et al. (2024) proposes splitting the sequences into 745 for training and 748 for testing with an approximate class ratio of 1 : 1 : 8 (positive : neutral : negative), and reporting accuracy, sensitivity, and specificity as the primary metrics.

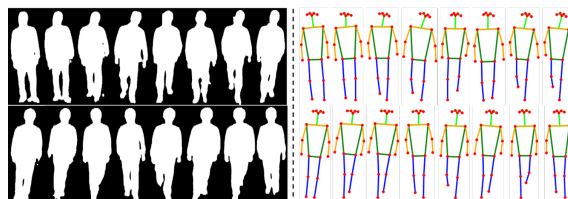


Figure 4.46: Scoliosis1K dataset: representative frontal-view silhouettes and COCO-style 2D pose sequences.

SDUMLA-HMT

SDUMLA-HMT is a multimodal biometric collection comprising 106 subjects (59 male and 47 female), aged 17–31. Five biometric traits were recorded for each participant: face, finger vein, gait, iris, and fingerprint, yielding five coordinated sub-databases in which a common person identifier links all modalities for the same individual. The gait subset (Yin et al. (2011)) was captured in a purpose-designed room using six fixed digital cameras placed on a circular arc of radius 6 m centered on the walkway: one frontal view aligned with the walking direction, one lateral view orthogonal to the path ($\approx 90^\circ$), and four oblique views at $\pm 22.5^\circ$ and $\pm 45^\circ$ relative to the lateral direction. Figure 4.47 shows the multi-camera configuration (top) and example RGB frames across covariates (bottom).

For each subject, six background clips were recorded (one per camera), followed by six natural walks, two walks while carrying a bag (knapsack, satchel, or handbag), and two runs. In total, the gait subset contains 6,996 RGB videos at 320×240 and 25 fps, each spanning approximately 2–3 gait cycles. Documented covariates include viewpoint (six views from frontal to lateral), carrying status (with-bag vs. without-bag), and motion type or speed (walking vs. running), yielding a compact multi-view, multi-condition benchmark.

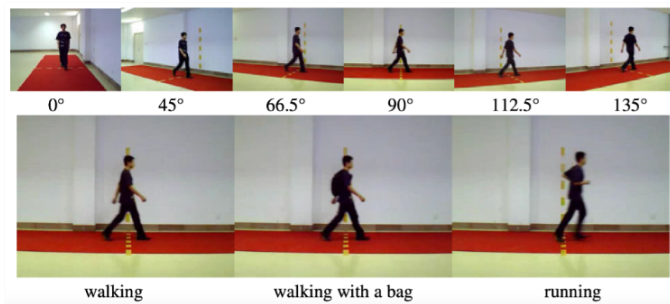


Figure 4.47: SDUMLA-HMT gait dataset: multi-camera indoor setup (top) and representative RGB frames across covariates (bottom), including viewpoint changes, carrying status (with/without bag), and motion type (walking/running).

SMVDU-Gait (Single and Multi)

SMVDU-Gait (Singh et al. (2019)) comprises two companion subsets captured outdoors with a tripod-mounted RGB camera at 1920×1080 and 50 fps, against a fixed background and along three straight walking paths. Single-Gait records solo walking by 20 volunteers (11 male and 9 female), while Multi-Gait records the same volunteers walking in pairs. Both subsets provide identity labels, enabling sequence-based and cross-view analyzes under consistent capture conditions. The principal covariates include the presence or absence of group-induced occlusion (pairs versus solo) and viewpoint differences arising from the three paths. Using matched subject identifiers across Single and Multi supports within-subject comparisons of gait patterns between solo and paired walking. Figure 4.48 illustrates example frames from Single (solo) and Multi (pairs).

A reported evaluation trains a gallery-probe classifier across the two subsets, using Multi-Gait for training and Single-Gait for testing, with a one-vs-rest scheme over 16 subjects; performance is summarized via confusion matrices and overall accuracy (as described in [Singh et al. \(2019\)](#)).



Figure 4.48: SMVDU-Gait dataset: representative frames from the Single (solo) and Multi (pairs) subsets, illustrating group-induced occlusion across straight paths.

SOTON Human ID (Large and Small DB)

The Southampton HumanID gait datasets ([Shutler et al. \(2004\)](#)) comprise two complementary releases designed to study gait under controlled and semi-controlled conditions. The SMALL dataset captures 12 subjects indoors on a straight track with four designated viewpoints: normal track (fronto-parallel), oblique, elevated, and frontal. Video is provided as RGB sequences with associated silhouettes. The LARGE dataset expands the population to 115 subjects and introduces three scenarios: outside track, inside track, and treadmill – each recorded from two viewpoints (fronto-parallel and oblique), yielding six distinct camera views overall. Both datasets provide RGB suitable for silhouette extraction, enabling appearance-based evaluation and analysis of view changes. Figure 4.49 illustrates representative views and scenarios (indoor track, treadmill, and outdoor track).

Covariates emphasized across the two datasets include viewpoint changes, clothing variation, carrying conditions, footwear, and walking-speed variation. The LARGE dataset additionally varies walking surface and environment (indoor versus outdoor, with controlled and uncontrolled lighting), enabling analyses of robustness to appearance and scene factors as well as surface-related trajectory differences.



Figure 4.49: SOTON HumanID (Small/Large) datasets: representative views and scenarios – indoor track, treadmill, and outdoor track – across fronto-parallel and oblique viewpoints.

SPHERE Walking Dataset

The SPHERE Walking Dataset (Païement et al. (2014)) provides frontal-view depth sequences and OpenNI skeletons of individuals walking upstairs. The collection comprises 17 sequences from 12 healthy subjects captured indoors at approximately 30 fps, from which 42 gait cycles were extracted for analysis. The distribution includes depth video and associated skeletons, organized into `training_data` and `testing_data` directories. Covariates focus on gait type, with normal stair-ascent and three anomaly conditions used for evaluation: right-leg-lead, left-leg-lead, and freeze-of-gait. Figure 4.50 illustrates representative frontal-view depth frames together with the corresponding body-skeleton representations.

The original evaluation trained a normal-gait model on six subjects and assessed generalization on six unseen subjects. Anomalous segments (right-leg-lead, left-leg-lead, freeze-of-gait) were labeled by a physiotherapist and employed to test detection and frame-level classification.

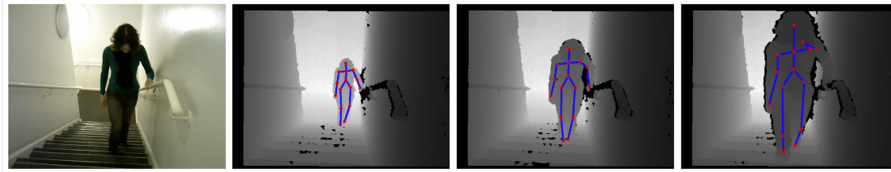


Figure 4.50: SPHERE Walking Dataset: representative frontal-view stair-ascent sequences showing depth frames with corresponding OpenNI skeletons.

SUSTech1K

SUSTech1K (LidarGait) (Shen et al. (2023)) is a synchronized multi-modal dataset for outdoor gait recognition that pairs RGB video (1280×980 at 30 fps) with 3D LiDAR point clouds (10 fps). Data were captured in three scenes using a mobile platform equipped with a 128-beam LiDAR and a monocular camera, producing 25,239 sequences from 1,050 subjects across 12 viewpoints. The release includes silhouettes and labeled walking conditions covering normal, carrying a bag, clothes change, occlusion, illumination (including night), uniform, and umbrella. Covariates are explicitly represented through viewpoint change, occlusion, clothing variation, carrying conditions and accessories, lighting change, and route design (straight and round-trip with turns), enabling cross-covariate evaluation under consistent sensor synchronization. Figure 4.51 illustrates aligned RGB frames, silhouettes, and the corresponding 3D point-cloud sequences.

Evaluation follows subject-disjoint splits with 250 identities for training, 250 for validation, and 550 for testing. Cross-view protocols exclude identical probe-gallery views; normal sequences serve as the gallery while probes span the remaining walking conditions. Reported metrics (Shen et al. (2023)) include Rank-1 and Rank-5 accuracy, with open-set partitions ensuring no overlap among training, validation, and test samples.

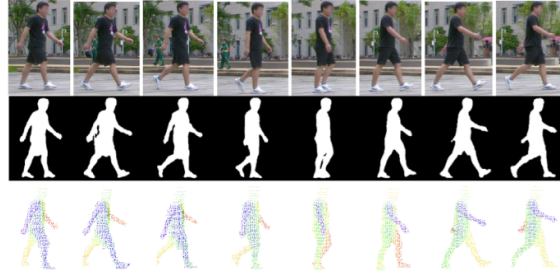


Figure 4.51: SUSTech1K (LidarGait) dataset: synchronized, spatially aligned RGB frames and silhouettes with the corresponding 3D LiDAR point-cloud sequences across representative view-points.

SZU RGB-D Gait Dataset

The SZU RGB-D Gait Dataset (Yu et al. (2013)) comprises 792 sequences from 99 subjects, recorded indoors with an ASUS Xtion PRO LIVE sensor (ASU) at 30 fps and 640×480 resolution. Each subject walked two trajectories: a lateral path at 90° to the sensor and an oblique path at $\approx 60^\circ$, with four sequences per trajectory. The release provides synchronized RGB, depth, and silhouette streams. The principal covariate is walking trajectory under a single-sensor setup. Figure 4.52 illustrates a sample RGB frame and the corresponding depth image for each trajectory.

The evaluation protocol described in Yu et al. (2013) defines four settings that decouple modality (depth versus RGB) and trajectory (same-trajectory versus cross-trajectory). For **depth**: (i) same-trajectory, with gallery sequences 1–2 and probes 3–4; (ii) cross-trajectory, with gallery 1–2 at 90° and probes 5–8 at $\approx 60^\circ$. For **RGB**: (iii) same-trajectory, gallery 1–2 versus probes 3–4; (iv) cross-trajectory, gallery 1–2 at 90° versus probes 5–8 at $\approx 60^\circ$. This design enables both same- and cross-trajectory evaluation within each modality and was used to benchmark Gait Energy Surface (depth) and Gait Energy Image (silhouette-based) representations.

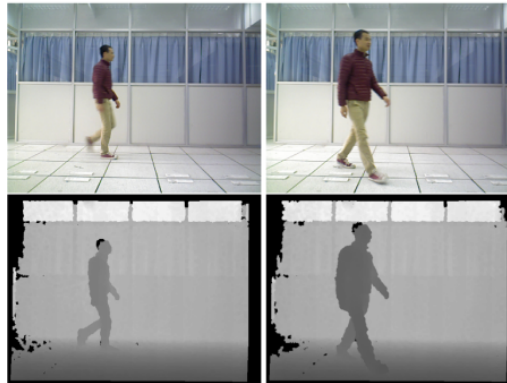


Figure 4.52: SZU RGB-D Gait Dataset: sample RGB frames and corresponding depth maps for the two trajectories: lateral (90°) and oblique ($\approx 60^\circ$).

TUM-IITKGP

The TUM-IITKGP dataset (Hofmann et al. (2011)) targets gait recognition under occlusion and appearance changes. It comprises 840 sequences from 35 subjects recorded indoors in a narrow corridor with a single fixed camera positioned at ≈ 1.85 m height and oriented perpendicular to the walking direction (side-view). Each subject was recorded under six configurations: regular walking, hands in pockets, backpack, gown, dynamic occlusion (two other people walking past), and static occlusion (two people standing in the line of sight). Every configuration was repeated four times (two directions $\times$ two repeats), yielding 24 sequences per subject. The public release focuses on silhouette videos extracted from RGB. Figure 4.53 illustrates sample RGB frames for the listed covariates.

No official evaluation split is provided. Typical studies use cross-condition protocols, forming training and testing subsets to assess robustness across appearance changes (pockets, backpack, gown) and occlusion conditions (dynamic and static). Baseline experiments in Hofmann et al. (2011) report person identification using color histograms (appearance) and Gait Energy Images (Han and Bhanu (2006)), demonstrating use cases with and without explicit occlusions.

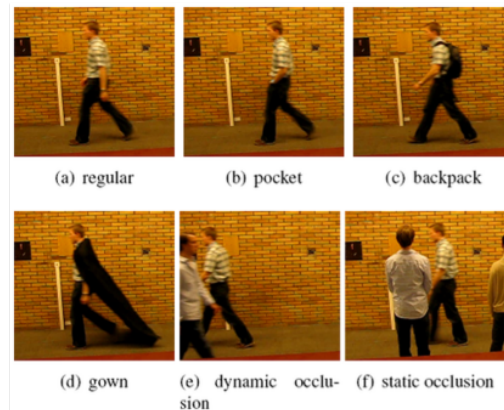


Figure 4.53: TUM-IITKGP dataset: sample RGB frames for the covariates (a) regular, (b) hands-in-pockets, (c) backpack, (d) gown, (e) dynamic occlusion, (f) static occlusion.

Walking Gait Dataset

The Walking Gait Dataset (Nguyen et al. (2018)) provides synchronized point-clouds, 3D skeletons with 25 joints, and binary silhouettes for nine subjects, each performing nine distinct gait conditions. Recordings were made indoors on a treadmill using a Kinect v2 sensor with two mirrors. Each sequence contains 1,200 consecutive frames; silhouettes are distributed at 512×424 and are derived from depth. The nine conditions comprise one normal pattern and eight simulated asymmetries induced by shoe-lifts of 5, 10, and 15 cm and ankle weights of 4 kg, applied unilaterally (left or right). The release is designed for comparative evaluation across representations in gait analysis and abnormal-gait detection. Covariates focus on gait-type manipulations (lift height and ankle-load side) under a fixed frontal viewpoint; point-clouds are reconstructed from depth. Figure 4.54 illustrates representative silhouettes, skeletons, and point-clouds.

Evaluation typically follows the authors’ subject split (Nguyen et al. (2018)), training on subjects {1, 3, 5, 6, 9} and testing on {2, 4, 7, 8}. In addition, leave-one-out cross-validation across the nine participants is recommended for a subject-disjoint assessment.



Figure 4.54: Walking Gait Dataset: representative samples from frontal-view treadmill recordings showing silhouettes, 3D skeletons, and point-clouds.

4.4.4 Taxonomy-Aligned Dataset Tables

This section instantiates the schema defined in Section 4.3.4 using the terminology of covariates described in Section 4.3.3, providing a concise, modality-agnostic overview of the reviewed datasets. By aligning attributes across RGB and depth and by separating scene-, user-, and sensor-level fields, the tables explicitly demonstrate the link between acquisition choices, covariates, and protocol design. This reduces ambiguity in common evaluation scenarios, such as cross-view or cross-clothes. Within each dimension, attributes are encoded by the legend keys (A–R) to ensure consistency and concision.

For ease of consultation, the datasets are grouped by their main area of use, combined with the tables’ alphabetical order. Healthcare-oriented datasets are listed in Table 4.9; biometric-oriented resources appear in Table 4.8; and attribute-recognition resources are summarized in Table ??.

Each column is mapped directly to the dimensions of the taxonomy. The scene-level column contains information on the environment, background, surface, lighting, viewpoints, trajectory, and temporal span (A–G); the user-level column encodes the gait type, speed, clothing, carrying items and demographics (H–L); and the sensor-level column lists the modality, framerate, resolution, camera height, synchronization, and noise or artifacts (M–R). Where there are official training/validation/testing partitions or gallery/probe rules, the corresponding protocol information is referenced in the dataset profiles (Section 4.4.3) and aligns with the same covariate keys. Missing items are marked as — rather than being inferred.

Table 4.7: **Healthcare-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D keyp. = 2D keypoints; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available.

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
Gait-A	2016	5 (—)	(A) Indoor (B) Static (C) Overground (D) Controlled (E) 2 (0° front, 90° side) (F) Straight (G) —	(H) Normal / Pathological (I) — (J) — (K) — (L) —	(M) Silh. (N) 30 fps (O) — (P) — (Q) — (R) —
GAIT-IST	2020	10 (8 ♂ – 2 ♀)	(A) Indoor (B) Static (C) Overground (D) Controlled (E) 1 (90° side) (F) Walking trajectories (G) —	(H) Normal / Pathological (I) — (J) — (K) — (L) —	(M) Silh.; GEI; SEI (N) 25 fps (O) — (P) — (Q) — (R) —
GAIT-IT	2021	21 (19 ♂ – 2 ♀)	(A) Indoor (B) Static (C) Overground (D) Controlled (E) 2 (0° front, 90° side) (F) Walking trajectories (G) Time span	(H) Normal / Pathological (I) Normal (J) — (K) — (L) —	(M) Silh.; GEI; SEI (N) 10 fps (O) — (P) — (Q) — (R) —
Health&Gait	2021	398 (199 ♂ – 199 ♀)	(A) Indoor (B) Static (C) Overground (D) Controlled (E) 1 (90° side) (F) Walking trajectories (G) Multi-session	(H) Normal (I) Varied (J) Varied (K) — (L) Balanced	(M) Silh.; 2D Keyp.; S. Segm. (N) 30 fps (O) — (P) — (Q) — (R) —
INIT Gait	2016	10 (9 ♂ – 1 ♀)	(A) Indoor (B) Static (C) Overground (D) Controlled (E) 1 (90° side) (F) Straight (G) —	(H) Normal / Pathological (I) — (J) — (K) — (L) —	(M) Silh. (N) 15 fps (O) — (P) — (Q) — (R) —

Legend: ♂ male; ♀ female; (A) Environment; (B) Background; (C) Walking surface; (D) Lighting; (E) Viewpoints; (F) Walking trajectory; (G) Temporal variation; (H) Gait type; (I) Walking speed; (J) Clothing; (K) Carrying condition; (L) Demographics; (M) Modality; (N) Frame rate; (O) Resolution; (P) Camera height; (Q) Synchronization; (R) Noise / Artifacts.

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available.

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
Continued on next page...					
360 Degree Gait	2022	65 (38 ♂ – 27 ♀)	(A) indoor; outdoor (B) static (C) overground (D) controlled; uncontrolled (E) 8 (0 – 360) (F) Walking trajectories (G) Time span	(H) Normal (I) Varied (J) Varied (K) — (L) Available	(M) RGB; 2D Keyp.; goniometer data (N) 30 fps (O) — (P) — (Q) — (R) —
AVA Multi-View	2013	20 (16 ♂ – 4 ♀)	(A) indoor (B) static (C) overground (D) uncontrolled (E) 6 (F) Walking trajectories (G) —	(H) Normal (I) — (J) — (K) — (L) —	(M) RGB; Silh. (N) 25 fps (O) — (P) — (Q) — (R) —
CASIA-A	2001	20 (—)	(A) outdoor (B) static (C) overground (D) uncontrolled (E) 3 (0, 45, 90) (F) Walking trajectories (G) —	(H) Normal (I) — (J) — (K) — (L) —	(M) RGB (N) 25 fps (O) — (P) — (Q) — (R) —

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
CASIA-B	2005	124 (93 ♂ – 31 ♀)	(A) indoor		(M) RGB; Silh.
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 11 (0 – 180)	(K) Present	(Q) —
			(F) —	(L) Available	(R) —
			(G) —		
CASIA-C	2005	153 (130 ♂ – 23 ♀)	(A) outdoor		(M) infrared; Silh.
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) Varied	(O) —
			(D) uncontrolled	(J) —	(P) —
			(E) 1 (90° side)	(K) Present	(Q) —
			(F) —	(L) Available	(R) —
			(G) —		
CASIA-E	2022	1,014 (507 ♂ – 507 ♀)	(A) outdoor		(M) Silh.; GEI; infrared
			(B) static; dynamic	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 28	(K) Present	(Q) —
			(F) —	(L) Available	(R) —
			(G) Time span		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
CCGR	2024	970 (—)	(A) indoor		(M) RGB; Silh.; 2D Keyp.; 2D Parsing
			(B) static	(H) Normal	(N) 25 fps
			(C) overground; stairs; ramp	(I) Varied	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 33	(K) Present	(Q) —
			(F) Walking trajectories	(L) Available	(R) —
			(G) —		
CCPG	2023	200 (122 ♂ – 78 ♀)	(A) indoor; outdoor		(M) RGB; Silh.
			(B) dynamic	(H) Normal	(N) —
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 10	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
CMU MoBo	2001	25 (23 ♂ – 2 ♀)	(A) indoor		(M) RGB; Silh.
			(B) static	(H) Normal	(N) 30 fps
			(C) treadmill	(I) Varied	(O) —
			(D) controlled	(J) —	(P) —
			(E) 6 (0 – 360)	(K) Present	(Q) —
			(F) —	(L) Available	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
Depth-Based Gait	2015	29 (—)	(A) indoor		(M) Silh.; 2D Keyp.; 3D Keyp.
			(B) static	(H) Normal	(N) 30 fps; 15 fps
			(C) overground	(I) Varied	(O) —
			(D) controlled	(J) —	(P) —
			(E) 2 (0 – front; 180 – back)	(K) —	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
FVG	2018	226 (—)	(A) outdoor		(M) RGB
			(B) dynamic	(H) Normal	(N) 15 fps
			(C) overground	(I) Varied	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 3 (–45, 0, 45)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) Time span		
Gait3D-Parsing	2023	4000 (—)	(A) indoor		(M) Silh.; 2D Keyp.; 3D Keyp.; 3D Meshes; 2D Parsing
			(B) dynamic	(H) Normal	(N) 25 fps
			(C) overground	(I) Varied	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 39	(K) Present	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
GREW	2020	26,345 (—)	(A) outdoor		(M) Silh.; GEI; 2D Keyp.; 3D Keyp.; optical flow
			(B) dynamic	(H) Normal	(N) —
			(C) overground	(I) Varied	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 882	(K) Present	(Q) —
			(F) Walking trajectories	(L) Available	(R) —
			(G) —		
GRIDDs	2018	35 (11 ♂ – 24 ♀)	(A) indoor		(M) RGB; Silh.; infrared; depth; 2D Keyp.; 3D Keyp.
			(B) static	(H) Normal	(N) 30 fps
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) —	(P) —
			(E) 1 (90° side)	(K) —	(Q) —
			(F) Walking trajectories	(L) Available	(R) —
			(G) —		
KY 4D–A	2010	42 (—)	(A) indoor		(M) RGB; Silh.; 3D Model
			(B) static	(H) Normal	(N) 15 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 16 (0 – 360)	(K) —	(Q) —
			(F) —	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
KY 4D-B	2010	42 (—)	(A) indoor		(M) RGB; Silh.; 3D Model
			(B) static	(H) Normal	(N) 15 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 16 (0 – 360)	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
KY IR	2014	54 (—)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) 30 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 2 (side, top)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
MA-Gait	2022	95 (65 ♂ – 30 ♀)	(A) indoor		(M) Silh.; 2D Keyp.
			(B) static	(H) Normal / Pathological	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 6	(K) —	(Q) —
			(F) Walking trajectories	(L) Available	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
MHG	2023	200 (—)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) —
			(C) overground	(I) —	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 9 (0, 45, 90) × High, Mid, Low	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
OccGait	2025	101 (—)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) —
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 8 (0 – 360)	(K) Present	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
OG RGB+D	2022	96 (60 ♂ – 36 ♀)	(A) indoor		(M) RGB; depth; 3D Keyp.; Silh.
			(B) dynamic	(H) Normal	(N) 30 fps
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 8	(K) Present	(Q) —
			(F) Walking trajectories	(L) Available	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
OU-ISIR GaitST	2014	— (—)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) 60 fps
			(C) overground; treadmill	(I) Varied	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1 (90° side)	(K) —	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
OU-ISIR LP Age	2017	63,846 (31,093 ♂ – 32,753 ♀)	(A) indoor		(M) GEI
			(B) static	(H) Normal	(N) 30 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1 (90° side)	(K) —	(Q) —
			(F) —	(L) Available	(R) —
			(G) —		
OU-ISIR LP Bag	2018	62,528 (—)	(A) indoor		(M) Silh.; GEI
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1 (90° side)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
OU-ISIR LP MV	2018	10,307 (5,114 ♂ – 5,193 ♀)	(A) indoor		(M) Silh.; GEI
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 7	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
OU-ISIR LP MV Mesh	2022	10,307 (5,114 ♂ – 5,193 ♀)	(A) indoor		(M) Silh.; 2D Keyp.; 3D Meshes
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 7	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
OU-ISIR LP MV w Pose Sequence	2020	10,307 (5,114 ♂ – 5,193 ♀)	(A) indoor		(M) Silh.; 2D Keyp.
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 7	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
OU-ISIR MVLP-OF	2025	10,307 (5,114 ♂ – 5,193 ♀)	(A) indoor		(M) Optical Flow
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 7	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
OU-Treadmill-A	2010	34 (—)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) 60 fps
			(C) treadmill	(I) Varied	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1 (90° side)	(K) —	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
OU-Treadmill-B	2010	68 (31 ♂ – 37 ♀)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) 60 fps
			(C) treadmill	(I) —	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 1 (90° side)	(K) —	(Q) —
			(F) —	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
OU-Treadmill-D	2010	185 (—)	(A) indoor		(M) Silh.
			(B) static	(H) Normal	(N) 60 fps
			(C) treadmill	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1 (90° side)	(K) —	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
RA-GAR	2025	288 (—)	(A) outdoor		(M) Silh.; 2D Keyp.
			(B) static	(H) Normal	(N) 30 fps
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 10	(K) Present	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
ReSGait	2019	172 (134 ♂ – 38 ♀)	(A) indoor		(M) Silh.; 2D Keyp.
			(B) static	(H) Normal	(N) 20 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 1	(K) Present	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) Time span		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
SAIVT-Depth Gait	2011	35 (—)	(A) indoor		(M) depth; Silh.; 3D Vol.
			(B) static	(H) Normal	(N) 30 fps
			(C) overground	(I) Varied	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1 (0 – front)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
SDUMLA-HMT	2010	106 (59 ♂ – 47 ♀)	(A) indoor		(M) RGB
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) Varied	(O) —
			(D) controlled	(J) —	(P) —
			(E) 6 (0, 45, 66.5, 90, 112.5, 135)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
SMVDU-Gait	2019	20 (11 ♂ – 9 ♀)	(A) outdoor		(M) RGB
			(B) static	(H) Normal	(N) 50 fps
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) —	(P) —
			(E) 1	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
SOTON LargeDB	2002	115 (100 ♂ – 15 ♀)	(A) indoor; outdoor		(M) RGB; Silh.
			(B) static; dynamic	(H) Normal	(N) 25 fps
			(C) overground; treadmill	(I) Varied	(O) —
			(D) controlled; uncontrolled	(J) Varied	(P) —
			(E) 6	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
SOTON SmallDB	2002	12 (—)	(A) indoor		(M) RGB; Silh.
			(B) static	(H) Normal	(N) 25 fps
			(C) overground	(I) Varied	(O) —
			(D) controlled	(J) Varied	(P) —
			(E) 4 (front; side; oblique; elevated)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		
SUSTech1K	2023	1,050 (—)	(A) outdoor		(M) RGB; Silh.; 3D PCloud
			(B) dynamic	(H) Normal	(N) 30 fps (RGB); 10 fps (LiDAR)
			(C) overground	(I) —	(O) —
			(D) uncontrolled	(J) Varied	(P) —
			(E) 12	(K) Present	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		

Continued on next page

Table 4.8: **Biometric-oriented gait datasets** with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available. (Continued)

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
SZU RGBD Gait	2013	99 (—)	(A) indoor		(M) RGB; Silh.; Depth
			(B) static	(H) Normal	(N) 30 fps
			(C) overground	(I) —	(O) —
			(D) controlled	(J) —	(P) —
			(E) 1	(K) —	(Q) —
			(F) Walking trajectories	(L) —	(R) —
			(G) —		
TUM-IITKGP	2010	35 (—)	(A) Indoor		(M) Silh.
			(B) Static	(H) Normal	(N) 30 fps
			(C) Overground	(I) —	(O) —
			(D) Controlled	(J) Varied	(P) —
			(E) 1 (90° side)	(K) Present	(Q) —
			(F) —	(L) —	(R) —
			(G) —		

Legend: ♂ male; ♀ female; (A) Environment; (B) Background; (C) Walking surface; (D) Lighting; (E) Viewpoints; (F) Walking trajectory; (G) Temporal variation; (H) Gait type; (I) Walking speed; (J) Clothing; (K) Carrying condition; (L) Demographics; (M) Modality; (N) Frame rate; (O) Resolution; (P) Camera height; (Q) Synchronization; (R) Noise / Artifacts.

Table 4.9: **Attribute-recognition gait datasets** (non-biometric / non-healthcare) with taxonomy-aligned attributes. Within each dimension, attributes are referenced by uppercase letters; see the legend for meanings. Abbreviations: 2D Keyp. = 2D Keypoints; 3D Keyp. = 3D Keypoints; 3D PCloud = 3D Point Cloud; 3D Vol. = 3D Volumes; GEI = Gait Energy Image; SEI = Skeleton Energy Image; S. Segm. = Semantic Segmentation; Silh. = Silhouette; — = not available.

Dataset	Year	Subjects	Scene-level	User-level	Sensor-level
Multi-Attribute Gait	2022	95 (65♂ – 30♀)	(A) indoor (B) static (C) overground (D) controlled (E) 6 (F) Walking trajectories (G) —	(H) Normal (I) — (J) — (K) — (L) Available	(M) Silh.; 2D Keyp. (N) 25 fps (O) — (P) — (Q) — (R) —
OU-ISIR LP Age	2017	63,846 (31,093♂ – 32,753♀)	(A) indoor (B) static (C) overground (D) controlled (E) 1 (90° side) (F) — (G) —	(H) Normal (I) — (J) — (K) — (L) Available	(M) GEI (N) 30 fps (O) — (P) — (Q) — (R) —
RA-GAR	2025	288 (—)	(A) outdoor (B) static (C) overground (D) uncontrolled (E) 10 (F) Walking trajectories (G) —	(H) Normal (I) — (J) Varied (K) Present (L) —	(M) Silh.; 2D Keyp. (N) 30 fps (O) — (P) — (Q) — (R) —

Legend: ♂ male; ♀ female; (A) Environment; (B) Background; (C) Walking surface; (D) Lighting; (E) Viewpoints; (F) Walking trajectory; (G) Temporal variation; (H) Gait type; (I) Walking speed; (J) Clothing; (K) Carrying condition; (L) Demographics; (M) Modality; (N) Frame rate; (O) Resolution; (P) Camera height; (Q) Synchronization; (R) Noise / Artifacts.

4.4.5 Comparative Analysis and Selection Guidance

In the context of two decades of research into human gait, there is evidence of marked inertia in community usage. A small core of established datasets continues to serve as the foundation for benchmarking and method comparison, as evidenced by the ongoing use of resources such as CASIA-B (Yu et al. (2006a)), CMU MoBo (Gross and Shi (2001)), SDUMLA-HMT (Yin et al. (2011)), OU-MVLP (Takemura et al. (2018)), and the SOTON collections (Shutler et al. (2004)). Their persistence reflects stable capture geometry, accessible protocols, and extensive prior art that facilitates like-for-like comparisons over time. Concurrently, recent datasets emphasize in-the-wild scenes, multi-camera arrays, and richer annotations. However, these datasets naturally exhibit modest citation footprints due to their recency (e.g. CASIA-E (Song et al. (2022)), Gait3D-Parsing (Zheng et al. (2023b)), CCGR (Zou et al. (2024)), RA-GAR (Wang et al. (2025))). The field thus balances two competing priorities: the need for comparability on standard frameworks, and the requirement for external validity in real operational settings.

Modalities and representations. Silhouette-centric releases established the baseline for appearance-based recognition and remain applicable for controlled studies (e.g., CASIA-B (Yu et al. (2006a)), OU-ISIR treadmill subsets (Tsuji et al. (2010); Altab Hossain et al. (2010); Mori et al. (2010))). Subsequent generations then layered RGB with auxiliary labels, including 2D/3D pose, parsing masks, optical flow, and parametric meshes. This enabled representation variations within constant capture setups (e.g. the OU-MVLP family with Pose/Mesh/OF variants (An et al. (2020); Li et al. (2022b); Shehata et al. (2025))). Depth and RGB-D collections, driven by commercial sensors, have been validated by pose-driven pipelines and privacy-preserving analysis (SAIVT-DGD (Sivapalan et al. (2011)), SZU RGB-D (Yu et al. (2013)), GRIDDS (Nunes et al. (2019a))). Recent work has extended this line to LiDAR-RGB pairing for outdoor use (Shen et al. (2023)). Despite this breadth, large-scale RGB with silhouettes remains the workhorse for open-world ID and retrieval at population scale (e.g., GREW (Guo et al. (2025a))).

Scale, geometry, and covariates. The scale of the dataset ranges from compact clinical or laboratory settings to city-wide deployments. Classic multi-view rigs offer systematic angles under stable backgrounds (CASIA-B (Yu et al. (2006a)), KY4D (Iwashita et al. (2010, 2014b))), whereas modern collections expand both the number of cameras and the diversity of scenes (GREW's 882 cameras (Guo et al. (2025a)); CASIA-E's 26 viewpoints (Song et al. (2022)); CCGR's 33-camera array with 53 covariates (Zou et al. (2024))). Covariates are increasingly being incorporated into the design process. This is evident in the incorporation of clothing and carrying (CASIA-B/E (Yu et al. (2006a); Song et al. (2022)), RA-GAR (Wang et al. (2025))), occlusion and crowding (OccGait (Huang et al. (2024)), OG RGB+D (Li and Zhao (2022)), Gait3D-Parsing (Zheng et al. (2023b))), camera height (MHG (Hou et al. (2023))), trajectory and curved paths (AVA Multi-View (López-Fernández et al. (2014)), KY4D-B (Iwashita et al. (2014b))), and speed/acceleration profiles (OU-ISIR GaitST (Mansur et al. (2014))).

Evaluation practice. Subject-disjoint splits and cross-view/condition protocols remain standard and increasingly come with official partitions and gallery/probe definitions (e.g., OU-MVLP (Takemura et al. (2018)), CASIA-E (Song et al. (2022)), CCGR (Zou et al. (2024)), RA-GAR (Wang et al. (2025))). This consolidation has been demonstrated to enhance reproducibility and reduce design degrees of freedom.

Pragmatic selection (broad guidance). In order to facilitate cross-view identification at scale under controlled conditions, both the OU-MVLP and CASIA-B/E protocols have been shown to be both stable and extensive in nature (Takemura et al. (2018); Yu et al. (2006a); Song et al. (2022)). In the context of research involving occlusion, trajectory variation, or crowds, GREW and Gait3D-Parsing have been shown to emphasize real-world difficulty (Guo et al. (2025a); Zheng et al. (2023b)), while CCGR and CCPG target broad covariate and clothing-change robustness (Zou et al. (2024); Li et al. (2023b)). In the context of depth or privacy-preserving pipelines and interpretable kinematics, SAIVT-DGD, SZU RGB-D, GRIDDS, and MMGS have been demonstrated to supply depth, skeletons, and/or event labels (Sivapalan et al. (2011); Yu et al. (2013); Nunes et al. (2019a); Khokhlova et al. (2019)). For clinical or pathological analysis involving labelled impairments and anthropometrics, GAIT-IT/IST, Gait-A, Health&Gait, KOA-PD-NM, Scoliosis1K, and ProGait are considered appropriate choices. Typically, the evaluation is framed as subject-disjoint classification and event-aware assessment (Albuquerque et al. (2021a); Loureiro and Correia (2020); Nieto-Hidalgo et al. (2016); Zafra-Palma et al. (2025); Kour et al. (2021); Zhou et al. (2024); Yin et al. (2025)). Finally, for focused diagnostics and targeted robustness (speed, clothes, resolution), the OU-ISIR treadmill series and CMU MoBo remain compact, clean, and repeatable baselines (Tsuji et al. (2010); Altab Hossain et al. (2010); Mori et al. (2010); Gross and Shi (2001)).

GRIDDS (Nunes et al. (2019a)). Developed as part of this research project, GRIDDS offers a compact, modality-rich benchmark for method comparisons and interpretable analysis. The capture utilizes a fixed Kinect v2 apparatus positioned at 1.8 m, with two straight trajectories (left-to-right and right-to-left). A total of 35 volunteers participated, each contributing 10 sequences (5 per direction). The release provides synchronized RGB, depth, infrared, silhouettes (aligned/normalized to 80×120), and 2D/3D joint coordinates, together with manually annotated gait events and derived spatio-temporal parameters. The primary covariate is the direction under a fixed lateral viewpoint, producing stable backgrounds suitable for controlled comparisons across representations.

Strengths for experimentation:

- (i) *Cross-modality alignment:* time-synchronized streams enable fair comparisons among silhouette, pose, and depth-based pipelines.

- (ii) *Event-level interpretability*: heel-strike/toe-off/mid-stance/mid-swing annotations and provided spatio-temporal parameters support cycle-level diagnostics and clinically meaningful features.
- (iii) *Privacy-aware modalities*: depth/infrared and silhouettes facilitate privacy-preserving prototypes without discarding geometric cues.
- (iv) *Protocol clarity*: two standardized evaluations are defined: *leave-one-subject-out identification* (gallery: first two sequences per direction; probes: remaining three) and *cross-trajectory generalization* (train on one direction; probe on the other), with Rank-1/Rank-5 and mAP reporting.

Limitations and positioning. The relatively limited sample size of the dataset ($n=35$) and the single lateral viewpoint appear to limit the validity of claims made on a population scale and the capacity for cross-view generalization. The static indoor setting may not fully reflect the clutter present in outdoor environments. Consequently, GRIDDS is best employed to (a) isolate representation choices and fusion strategies under controlled conditions, (b) validate event-aware or kinematic features against manual labels, and (c) pre-train or sanity-check models before transfer to larger or more complex resources (e.g., OU-MVLP [Takemura et al. \(2018\)](#), GREW [Guo et al. \(2025a\)](#), CASIA-E [Song et al. \(2022\)](#)).

4.5 Summary

This chapter presents an up-to-date, covariate-centered overview of vision- and depth-based gait data, and introduces GRIDDS, a mid-scale RGB-D resource designed for transparency and reproducibility. GRIDDS provides synchronized RGB, depth, IR, body-index silhouettes and 2D/3D joints from controlled lateral walking, in addition to frame-accurate, manually annotated gait events and per-trial spatio-temporal and kinematic summaries. In order to facilitate repeatable benchmarking across different modalities, the chapter specifies coordinate conventions, minimal pre-processing and two evaluation protocols. These choices make GRIDDS suitable for appearance-, skeleton-, and hybrid-based pipelines.

Beyond the release of the dataset, the chapter proposes a protocol-aware taxonomy that organizes variability along three orthogonal dimensions (scene, user and sensor) and makes the level of variation explicit (e.g. dataset/session/sequence/subject). The framework distinguishes active covariates (e.g. viewpoint, clothing, carrying, speed, occlusion, temporal gaps) from controlled parameters, remains aligned with both RGB and RGB-D settings, and maps capture choices directly to evaluation designs (for example, multi-view capture to cross-view splits, multi-session capture to cross-time tests).

Finally, the chapter extends prior reviews by cataloguing up-to-date RGB and RGB-D resources, including both in-the-wild and clinically oriented datasets, through the lens of the proposed taxonomy. It details access status, modalities, auxiliary annotations, and typical cross-condition protocols (cross-view, cross-clothes, cross-speed, cross-scene).

Chapter 5

Novel Gait Silhouette Image Interpolation Techniques

5.1 Motivation and Scope

The process of silhouette image interpolation is defined as the synthesis of intermediate binary human foreground masks between observed frames. Although video frame interpolation (VFI) is well established for RGB imagery, the silhouette setting raises distinct requirements: topology should remain intact (no artificial holes or disconnected limbs), boundaries should be sharp rather than stair-cased, and temporal phase should remain consistent across complete gait cycles. In summary, the objective is not photorealism, but rather an accurate and phase-aware depiction of the evolution of person shape over time.

VFI is primarily driven by the need to increase or standardize the framerate by inserting additional frames, which produces smoother motion and reduces the appearance of motion blur. It also supports various post-processing, surveillance and restoration tasks. Beyond general computer vision applications, recent literature describes uses in healthcare, such as reducing radiation exposure in interventional imaging by lowering the acquisition frame rate and interpolating missing frames, and in film and animation production, where interpolation makes low-frame-rate sequences appear fluid (Yin et al. (2020); Siyao et al. (2021)). These examples demonstrate how interpolation can be implemented in practice and emphasize the importance of temporal compression for subsequent applications. A recent survey synthesizes technical trends and evaluation practices across the field, reinforcing the central role of inter-frame prediction for temporal coherence (Dong et al. (2023)).

The motivation for focusing specifically on gait silhouettes is twofold: practical and scientific. In practice, many gait datasets and deployment scenarios already operate on silhouettes, either because silhouettes are an explicit output of a segmentation stage or because privacy policies favor masking appearance cues (e.g. Takemura et al. (2018); Zheng et al. (2023b)). From a scientific perspective, silhouettes are advantageous in that they are able to isolate shape and spatiotemporal rhythm from interfering factors such as texture, lighting, and color. This makes

them well aligned with classic representations such as the Gait Energy Image (Han and Bhanu (2006)) and with recent silhouette-based sequence encoders (Shiraga et al. (2016); Chao et al. (2019)). In such pipelines, interpolation serves as a means to stabilize temporal evidence, facilitating the harmonization of cadence across captures, the repair of gaps, and the enhancement of the quality of cycle-normalized descriptors. A concrete motivation for this study was an internal inspection of the GRIDDS dataset (see 4.2), in which short gaps were observed in certain silhouette sequences. Two controlled studies were then carried out to understand the impact of such irregularities on silhouette-based recognition, with both experiments described in detail later in this chapter. Firstly, the presence of gaps was examined in relation to GEI-style recognition, revealing that even short missing runs disrupt cycle detection and degrade the reliability of temporal pooling. Secondly, the role of frame rate was studied by down-sampling and then up-sampling sequences. This demonstrated that cadence changes can impact recognition and that appropriate interpolation can recover most of the performance loss. Together, these observations suggest that silhouette interpolation, used to fill gaps and harmonize cadence, has practical value for recognition systems that rely on temporally consistent masks.

Harmonizing framerate across sensors or datasets simplifies phase alignment, stance/swing segmentation, and cycle normalization. Interpolating silhouettes to a chosen cadence reduces descriptor variance and makes cross-database evaluation more comparable. Recovery from short dropouts addresses occlusions, sensor failures, compression losses, or network issues. Even a small number of missing silhouettes can break cycle completeness or corrupt temporal integration. On the other hand, reconstruction restores consistent cycles and maintains temporal consistency, an important factor given the sensitivity of sequence-level descriptors.

Silhouette interpolation is not a straightforward transfer of RGB VFI: with binary inputs lacking brightness gradients beyond edges, gradient-based interpolation provides limited guidance and can become ambiguous. The risk of topology breaks (e.g. a limb disconnecting from the body or a false hole appearing in the torso) is higher and more consequential than in RGB synthesis, where small photometric inconsistencies may be tolerable. These considerations motivate the use of loss functions that are mask-specific, including overlap measures (IoU/Dice), boundary quality (boundary F1, Chamfer/Hausdorff), topology error rates, and temporal smoothness.

This chapter outlines the problem of interpolating binary silhouettes for human gait, where targets deform non-rigidly over short temporal windows. The methodological emphasis is divided into two distinct areas. One line of work, drawing inspiration from the silhouette-specific, model-based interpolation method outlined by Imoto et al. (2020), utilizes shape or contour priors, incorporating distance-field or contour-evolution formulations, in conjunction with topology-aware constraints. This approach is designed to prevent the creation of holes and limb disconnections, while preserving sharp boundaries. The complementary line adapts deep temporal models, inspired by modern VFI, to binary inputs, with the objective of optimizing mask-specific objectives that address overlap, boundary, topology, and temporal consistency. In both cases, the evaluation focuses on mask similarity and boundary metrics, explicit topology error measures, temporal smoothness, and the effect on silhouette-based recognition (e.g. GEI performance), reflecting the

end use of these sequences in identification or re-identification settings.

5.2 Related Work

Research on interpolation for human motion can be grouped by the representation that is manipulated and the mechanism that is used to transport information through time: (i) pixel-based schemes that interpolate intensities directly; (ii) shape-based schemes that operate on contours or implicit shape functions; (iii) registration-/warping-based schemes that estimate correspondences (e.g., optical flow, nonrigid registration) and then synthesize intermediates by warping; and (iv) learning-based schemes that learn these transports or synthesis functions from data. For gait silhouettes, each group offers distinct benefits and characteristic failure modes.

Pixel-based interpolation treats the inter-frame relation as a local signal-processing problem, extending classical image interpolation techniques (nearest neighbor, linear, cubic B-spline) to the temporal axis. In short intervals these schemes are simple and fast, but they tend to over-smooth motion details, produce ringing or ghosting under large nonrigid deformations, and, crucially for silhouettes, can erode thin parts such as terminal limbs unless explicit constraints are imposed to preserve connectivity. Shape-based interpolation addresses some of these issues by generating the inter-frame boundary directly from the two observed boundaries. Active contours (“snakes”) minimize a sum of internal smoothness and external image forces to evolve a curve between frames (Kass et al. (1988)). Level-set formulations extend this to an implicit representation where curve evolution is driven by partial differential equations, allowing for topological changes. For silhouettes, moving the foreground boundary provides a natural geometric path to in-betweens, however gradient-driven evolution is unstable on binary inputs because only the boundary carries contrast. Consequently, region-based approaches and shape priors are therefore commonly introduced to enhance stability..

Another line of work uses shape spaces and geodesic interpolation to define mathematically principled paths between shapes. Whilst these models are attractive for topology control, they are computationally expensive and depend on reliable correspondences. A practical compromise widely used for binary masks uses distance transforms: interpolate signed distance fields (SDFs) between the two frames and re-threshold at zero to recover a binary mask. This tends to preserve overall structure and produce smooth boundaries, while remaining algorithmically simple. However, without explicit topology constraints, very thin structures may still vanish (Grevera and Udupa (1996); Ciesielski et al. (2010)). Morphology-driven strategies pursue topology preservation more directly. Skeleton-guided in-betweens provide structural cues (although with limited geometric fidelity), and topology-aware down-/up-sampling designs optimize objectives that keep key topological invariants unchanged during interpolation (Aloraibi (2023); Guo et al. (2025b)). When specialized to human silhouettes, these ideas map to concrete constraints such as forbidding new holes inside the torso and maintaining foot-ground connectivity through stance.

Registration- and warping-based methods estimate a deformation field and subsequently synthesize the intermediate frame by transporting pixels (or masks) along that field. B-spline free-form deformation (FFD) registration has long been used to find smooth, spatially varying correspondences in time (Rueckert et al. (2006)). Optical flow has become the dominant registration method in video, with numerous formulations that account for nonrigid motion and occlusions (Hu et al. (2017); Schuster et al. (2021); Huang et al. (2022)). In silhouettes, warping the mask under a reliable flow preserves local geometry and, with appropriate occlusion handling, can maintain limb connectivity. These schemes can produce sharp, structure-preserving results, but their quality is tied to correspondence accuracy. Flow errors around fast-moving extremities or self-occlusions readily translate into topology breaks unless additional regularization is applied.

Learning-based approaches use neural networks trained either to predict motion (e.g. optical flow) or to synthesize the intermediate frames directly. Modern video frame interpolation has been dominated by flow-guided warping plus synthesis, where networks estimate motion and occlusion, warp the inputs, and then refine the result. Representative examples include Super SloMo, which jointly models motion and occlusion for multi-frame synthesis (Jiang et al. (2018)); depth-aware variants that exploit estimated depth to reduce occlusion effects (Yan et al. (2025); Bao et al. (2019)); softmax-based forward warping that better handles many-to-one mappings (Niklaus and Liu (2020)); one-shot temporal pyramids that synthesize several intermediates at once (Chi et al. (2020)); and lightweight recurrent pyramids (e.g. UPR-Net) that integrate bidirectional flow and synthesis efficiently (Jin et al. (2023)). A second family of models replaces explicit flow with adaptive convolution or kernel prediction, estimating spatially varying kernels that, when applied to neighboring frames, produce the in-between. These models demonstrate ability in smooth, local motion but may struggle with large displacements. More recent developments incorporate uncertainty (Bayesian formulations) or attention/transformer architectures to capture longer-range temporal correspondences (Yu et al. (2022b); Shi et al. (2022); Lu et al. (2022a)). When applied to binary silhouettes, these networks benefit from the elimination of photometric losses and the adoption of mask-aware objectives, such as overlap (IoU/Dice), boundary-sensitive scores (boundary F1, Chamfer/Hausdorff), and penalties or stabilizers that reduce topology errors.

Silhouette-specific methods stand at the intersection of these families and inject human-motion priors. A boundary-tracking approach with a height constraint enforces consistent vertical profiles through time, producing temporally continuous silhouettes. This represents one of the few explicit silhouette-interpolation models with a kinematic prior tailored to human motion. In parallel, several gait studies tackle a closely related task: silhouette sequence completion under occlusion, by reconstructing missing frames from neighboring context with learning-based models. These works are closer to sequence completion than to generic VFI, but they share the objective of producing temporally smooth and structurally consistent silhouettes and have reported improvements in downstream recognition (Uddin et al. (2019); Hasan et al. (2024)).

Across these dimensions, the core trade-offs are consistent. Pixel-based schemes are conceptually simple but prone to over-smoothing and topology loss under nonrigid motion. Shape-based schemes offer direct control over boundaries and, with SDFs or morphology, can preserve struc-

ture, but they depend on high-quality contours and may be vulnerable to limb separations or large pose changes. Registration-/warping-based methods deliver sharp results when correspondences are reliable but can propagate flow errors into structural artifacts. Learning-based methods achieve state-of-the-art fidelity on RGB benchmarks and handle complex, multi-frame dynamics, yet applied naively to binary masks they may introduce broken connectivity or jagged edges unless trained with silhouette-specific losses and constraints. Notably, much of the VFI literature evaluates performance with RGB image metrics such as PSNR/SSIM (Hore and Ziou (2010)), which are inadequate for silhouettes; mask overlap, boundary accuracy, topology integrity, and temporal smoothness are better aligned with gait use cases (Jiang et al. (2018); Jin et al. (2023)). This evaluation gap partly explains why topology-aware objectives (well-studied in segmentation) remain under-represented in VFI despite their relevance to silhouette interpolation.

In practice, effective pipelines for gait silhouettes often blend these traditions: a registration or learned motion prior to handle displacements and separations; a shape or topology prior to prevent holes and preserve limb connectivity; and mask-specific objectives to sharpen boundaries and stabilize temporal evolution. This synthesis reflects the problem’s dual nature: interpolation must both transport shape through time and constrain that transport to remain anatomically and topologically plausible for human gait.

5.3 Proposed Methods

This section introduces two complementary approaches that have been developed to address the issue of silhouette interpolation in human motion sequences. Both aim to reconstruct missing or intermediate frames while maintaining temporal coherence and structural fidelity, yet they differ fundamentally in nature and scope.

The first method, BRIEF (Bidirectional Rowwise Interpolation with Envelope Fusion), follows a shape-driven, deterministic approach that explicitly models silhouette geometry and short-range temporal relationships. Operating at the level of horizontal scanlines, it decomposes each silhouette into ordered segments and performs bidirectional interpolation between temporally adjacent frames. By exploiting correspondences based on overlap and applying envelope fusion to handle local splits or merges, BRIEF preserves the anatomical coherence of limbs and ensures continuity of motion, even with minimal contextual information. Its simplicity and robustness make it well suited to scenarios with few available frames or constrained temporal gaps.

The second method, BiSINet (Bidirectional Silhouette Network), introduces a learning-based perspective on the same problem, using deep neural networks to infer complex spatiotemporal dependencies directly from data. BiSINet employs two ConvLSTM-based encoders to process past and future silhouettes in opposite temporal directions, merging their latent representations through an attention-guided fusion module that dynamically balances both temporal cues. A UNet-style decoder then reconstructs the interpolated silhouette, ensuring fine spatial detail and temporal consistency. Trained with a hybrid loss combining Binary Cross-Entropy and Soft Dice terms,

BiSiNet learns to generate silhouettes that are both pixel-accurate and structurally coherent. Together, BRIEF and BiSiNet represent two complementary paradigms – one rule-based and interpretable, the other data-driven and adaptive – providing a comprehensive framework for analyzing the limits and potential of silhouette interpolation under varying temporal and spatial conditions.

5.3.1 BRIEF – Bidirectional Rowwise Interpolation with Envelope Fusion

The BRIEF method is built upon the principle that most gait silhouettes evolve smoothly over time and that this evolution can be effectively captured by analyzing their rowwise structure. Instead of relying on global motion estimation, BRIEF performs interpolation locally along horizontal scanlines, using geometric relationships between corresponding silhouette segments in nearby frames. By comparing and aligning these per-row foreground intervals across both past and future temporal neighbors, the method infers how each part of the silhouette shifts, splits, or merges through time.

Let $S(t) \in \{0, 1\}^{H \times W}$ denote the binary silhouette at time t (foreground = 1, background = 0). The shape-based method synthesizes one or more intermediate silhouettes by exploiting spatial structure and temporal context in a small neighborhood around the target time t_0 . Up to four temporal silhouette neighbors are used: two past frames ($\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}$) and two future frames ($\mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}$), positioned at positive lags τ_1 and τ_2 with $0 < \tau_1 < \tau_2$, where

$$\mathbf{P}_{\tau} := S(t_0 - \tau), \quad \mathbf{F}_{\tau} := S(t_0 + \tau), \quad \tau \in \{\tau_1, \tau_2\}.$$

Figure 5.1 illustrates the simplest case with a single missing frame: black silhouettes denote the observed neighbors $\{\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}, \mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}\}$, and the gray silhouette is the target at t_0 .

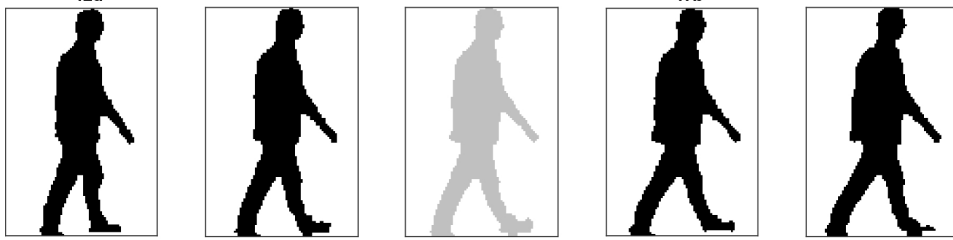


Figure 5.1: Overall concept of the shape-based method for a single missing frame at t_0 . Black silhouettes are the observed neighbors $\{\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}, \mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}\}$, ordered from farther past to farther future relative to t_0 ; the gray silhouette is the target synthesized at t_0 .

Global preprocessing fixes vertical alignment (head near the top, feet near a stable baseline) and centers the subject horizontally. Under these conditions, interpolation proceeds deterministically on a rowwise basis — from the top to the bottom of the image — such that each interpolated row is defined using only the corresponding rows from $\mathbf{P}_{\tau_2}$, $\mathbf{P}_{\tau_1}$, $\mathbf{F}_{\tau_1}$, and $\mathbf{F}_{\tau_2}$. This rowwise formulation preserves limb connectivity and enables explicit, one-dimensional reasoning about

segments and holes within each row. The method executes the following rowwise pipeline at the target instant t_0 using $\{\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}, \mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}\}$, where the positive lags satisfy $0 < \tau_1 < \tau_2$:

- A **Rowwise Segmentation.** Decompose each neighbor row into ordered segments and holes; store head/tail.
- B **Tolerance Scheduling.** Set the interval-overlap tolerance as a function of temporal separation to allow modest drift at larger lags.
- C **Correspondence Maps.** Build the $\mathbf{P}_{\tau_1}$ -correspondence map and the $\mathbf{F}_{\tau_1}$ -correspondence map with reference/corresponding indices and lags.
- D **Forward Interpolation.** From $\mathbf{P}_{\tau_1}$, estimate target spans using temporally weighted end-point interpolation; handle unique, multiple, and unmatched cases with auxiliary evidence.
- E **Backward Interpolation.** Mirror (D) from $\mathbf{F}_{\tau_1}$ toward the past, producing a second set of spans.
- F **Fusion and Topology Guard.** Fuse forward/backward spans by overlap and displacement criteria; apply a conservative guard to remove one-pixel spikes, fill one-pixel enclosed holes, and drop tiny components.

The following paragraphs describe in detail each of these steps.

A. Rowwise Segmentation

For each neighbor silhouette $S(t_0 \pm \tau)$ and each image row $r \in \{1, \dots, H\}$, the binary mask is decomposed along the width axis into ordered foreground segments and background holes. Let the row be indexed by column coordinate $x \in \{1, \dots, W\}$, with foreground value 1 and background value 0. A *segment* is defined as a maximal contiguous run of foreground pixels on row r , and a *hole* is a maximal contiguous run of background pixels that is bounded on both sides by foreground pixels on the same row. These segments are extracted by a single left-to-right scan (run-length encoding), producing an ordered list of intervals.

Each segment is parameterized by its *head* (leftmost index) h and *tail* (rightmost index) t . These one-dimensional intervals $[h, t]$ are sufficient for the subsequent correspondence and interpolation steps that follow. Figure 5.2 illustrates an enlarged view of a representative row: each square is a pixel, and the two contiguous groups of gray squares are segments. The white area between them represents a single hole. The numbers above the row indicate the horizontal coordinates used to index the heads and tails. For visual clarity the foreground is shown in gray rather than black, and the crop focuses on columns 17–62 where the sequences occur.

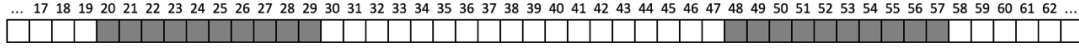


Figure 5.2: An enlarged view of a row from a silhouette, where each square represents a pixel. The two sets of consecutive gray squares are the segments, while the sequence of white squares between them is referred to as the single existing hole. The numbers at the top of the figure represent the pixel coordinates on the horizontal (width) axis of the images. For easier visualization, the pixels are depicted in gray instead of black.

Rows that contain no foreground pixels yield an empty segment list and are skipped during interpolation for that row. A row that is entirely foreground yields a single segment $[1, W]$ and no holes. Background sequences that touch the image border (e.g. $[1, \cdot]$ or $[\cdot, W]$) are not labeled as holes because they are not enclosed by foreground on both sides; they are treated as background margins. Very short sequences (length = 1) are retained at this stage to preserve thin limb structures. If needed, stability filtering is applied later by the topology filter.

B. Segments Correspondence and Tolerance Scheduling

Correspondence between segments is determined for pairs that lie on the same image row but belong to different frames around t_0 : either from the past ($\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}$) or from the future ($\mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}$). Consider two segments A and B on row r , with spans $[h_A, t_A]$ from one neighbor (e.g., $\mathbf{P}_{\tau_2}$) and $[h_B, t_B]$ from another (e.g., $\mathbf{F}_{\tau_1}$). Segments A and B are defined as corresponding if they have a nonempty intersection, i.e. if their spans overlap:

$$([h_A, t_A] \cap [h_B, t_B]) \neq \emptyset. \quad (5.1)$$

Because neighbors are separated in time, the same spatial predicate is evaluated with a *time-aware tolerance* that widens as temporal separation grows. Let $\lambda_A, \lambda_B \in \Lambda = \{-\tau_2, -\tau_1, +\tau_1, +\tau_2\}$ be the signed lags of the frames containing segments A and B (e.g., $\lambda_A = -\tau_2$, $\lambda_B = +\tau_1$). Their temporal separation is

$$\Delta t = |\lambda_A - \lambda_B|. \quad (5.2)$$

Given a nondecreasing tolerance schedule $\tau(\Delta t)$, each span $[h, t]$ is expanded to

$$[h, t]^{+\tau(\Delta t)} = [h - \tau(\Delta t), t + \tau(\Delta t)] \cap [1, W], \quad (5.3)$$

and correspondence is declared when the expanded spans overlap:

$$([h_A, t_A] \cap [h_B, t_B]^{+\tau(\Delta t)}) \neq \emptyset \quad \text{or} \quad ([h_B, t_B] \cap [h_A, t_A]^{+\tau(\Delta t)}) \neq \emptyset. \quad (5.4)$$

In the special case $\tau(\Delta t) = 0$, this reduces to the zero-tolerance intersection used above; A positive $\tau(\Delta t)$ permits modest lateral drift for pairs that are farther apart in time.

To make the notion of temporal separation concrete, consider three common pairings. For the near neighbors $\mathbf{P}_{\tau_1}$ versus $\mathbf{F}_{\tau_1}$, the separation is $\Delta t = |-\tau_1 - (+\tau_1)| = 2\tau_1$. For two past neighbors $\mathbf{P}_{\tau_2}$ versus $\mathbf{P}_{\tau_1}$, $\Delta t = |-\tau_2 - (-\tau_1)| = \tau_2 - \tau_1$. For a farther pair such as $\mathbf{P}_{\tau_2}$ versus $\mathbf{F}_{\tau_1}$, $\Delta t = |-\tau_2 - (+\tau_1)| = \tau_1 + \tau_2$. Accordingly, the tolerance $\tau(\Delta t)$ is smallest for $\mathbf{P}_{\tau_2}$ - $\mathbf{P}_{\tau_1}$ and largest for $\mathbf{P}_{\tau_2}$ - $\mathbf{F}_{\tau_1}$, permitting modest lateral drift only when frames are farther from t_0 .

Regarding the spatial cases in Fig. 5.3, the three configurations shown there (tail-in, head-in, and full containment) coincide with the spatial predicate when $\tau(\Delta t) = 0$.

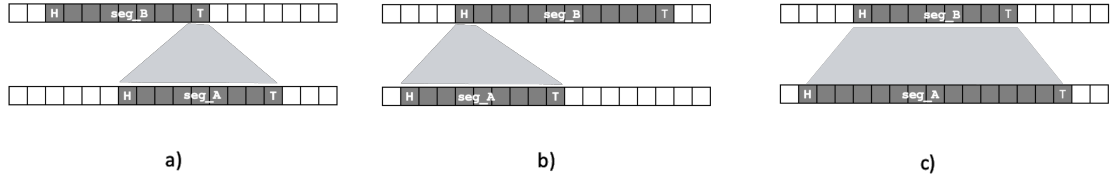


Figure 5.3: The segments intersection criteria explained in form of figures (letters H and T represent the *head* and *tail* of the corresponding segments): a) the *tail* of segment B is located between the *head* and *tail* coordinates of segment A; b) the *head* of segment B is located between the *head* and *tail* coordinates of segment A; c) segment B is located between the *head* and *tail* coordinates of segment A.

Tolerance scheduling generalizes these cases by applying the same spatial predicate to spans expanded by $\tau(\Delta t)$, so correspondences that narrowly miss under zero tolerance can become valid as Δt increases (see Fig. 5.4).

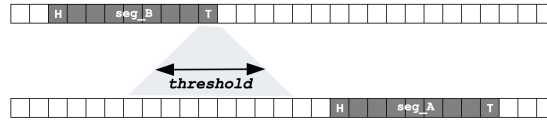


Figure 5.4: Time-aware tolerance for rowwise interval overlap. On a given row, segments from the past and future neighbors at lags τ (e.g., $\mathbf{P}_{\tau_1}$ and $\mathbf{F}_{\tau_1}$) are tested for correspondence using tolerance-expanded intervals $[h, t] + \tau(\Delta t)$. As the temporal separation Δt increases (e.g., comparing τ_2 rather than τ_1), the tolerance $\tau(\Delta t)$ widens, permitting modest lateral drift; for near neighbors it remains tight.

A practical tolerance choice used throughout is:

$$\tau(\Delta t) = \min(\tau_{\max}, 1 + 0.5 \Delta t), \quad \tau_{\max} = 3 \text{ pixels.} \quad (5.5)$$

These settings (maximum tolerance $\tau_{\max} = 3$ pixels and a linear growth coefficient of 0.5 on Δt) were adopted following empirical tests.

C. Correspondence Maps

For each image row r at target time t_0 , correspondences between segments in the four nearest neighbors $\{\mathbf{P}_{\tau_2}, \mathbf{P}_{\tau_1}, \mathbf{F}_{\tau_1}, \mathbf{F}_{\tau_2}\}$ are summarized in compact, rowwise maps that drive forward and backward interpolation.

Two correspondence maps are built per row: the past-reference map, $\text{cmap}_r^{\mathbf{P}_{\tau_1}}$, that treats each segment in $\mathbf{P}_{\tau_1}$ as reference and seeks candidates at lags $\lambda \in \{-\tau_2, +\tau_1, +\tau_2\}$; and the future-reference map, $\text{cmap}_r^{\mathbf{F}_{\tau_1}}$, that treats each segment in $\mathbf{F}_{\tau_1}$ as reference and seeks candidates at lags $\lambda \in \{-\tau_2, -\tau_1, +\tau_2\}$.

For each row r and neighbor lag $\lambda \in \{-\tau_2, -\tau_1, +\tau_1, +\tau_2\}$, let

$$\text{Seg}_r(\lambda) = \{[h_\lambda^{(k)}, t_\lambda^{(k)}]\}_{k=1}^{N_{r,\lambda}} \quad (5.6)$$

denote the left-to-right ordered list of foreground segments on that row, where $N_{r,\lambda}$ is the number of segments and $k \in \{1, \dots, N_{r,\lambda}\}$ indexes them from left (small x) to right (large x). If a row contains no foreground at lag λ , then $N_{r,\lambda} = 0$ and $\text{Seg}_r(\lambda) = \emptyset$.

Illustrative example. Figure 5.5 shows an amplified view of a single scanline r from each of the four neighbors – $\mathbf{P}_{\tau_2}$, $\mathbf{P}_{\tau_1}$, $\mathbf{F}_{\tau_1}$, and $\mathbf{F}_{\tau_2}$ – ordered from farther past to farther future relative to t_0 . Each row contains two foreground segments, indexed $j \in \{1, 2\}$ from left to right. Two correspondence maps are built only for the near neighbors: the past-referenced map $\text{cmap}_r^{\mathbf{P}_{\tau_1}}$ (candidates at lags $\lambda \in \{-\tau_2, +\tau_1, +\tau_2\}$) and the future-referenced map $\text{cmap}_r^{\mathbf{F}_{\tau_1}}$ (candidates at lags $\lambda \in \{-\tau_2, -\tau_1, +\tau_2\}$). Every accepted correspondence (Step B predicate) is retained.

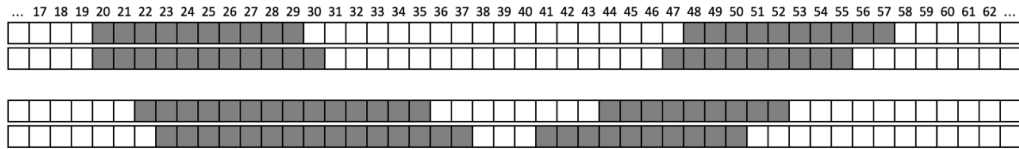


Figure 5.5: Amplified view of the same scanline r from each of the four temporal neighbors: $\mathbf{P}_{\tau_2}$, $\mathbf{P}_{\tau_1}$, $\mathbf{F}_{\tau_1}$, and $\mathbf{F}_{\tau_2}$, ordered from farther past to farther future relative to t_0 . For clarity, the crop shows columns 17–62 where the segments lie, and foreground pixels are rendered in gray rather than black.

For compact display, all correspondences are enumerated in the two following tables: Table 5.1 which reports the past-referenced map $\text{cmap}_r^{\mathbf{P}_{\tau_1}}$, and Table 5.2 which reports the future-referenced map $\text{cmap}_r^{\mathbf{F}_{\tau_1}}$. Each table record represents a single match between a reference segment j (in the near neighbor) and a segment at lag λ . Each record identifies a reference segment, a lag, and the index of the corresponding segment. When multiple segments exist for the same (j, λ) , multiple rows are shown; when no segment satisfies the predicate, no row appears for that combination.

Table 5.1: Past-referenced correspondence map $\text{cmap}_r^{\mathbf{P}_{\tau_1}}$ for Fig. 5.5. Each row lists one valid correspondence for a reference segment $j \in \{1, 2\}$ in $\mathbf{P}_{\tau_1}$ at lag $\lambda \in \{-\tau_2, +\tau_1, +\tau_2\}$.

Reference segment j	λ	Corresponding segment (index)
1	$-\tau_2$	1
1	$+\tau_1$	1
1	$+\tau_2$	1
2	$-\tau_2$	2
2	$+\tau_1$	2
2	$+\tau_2$	2

Table 5.2: Future-referenced correspondence map $\text{cmap}_r^{\mathbf{F}_{\tau_1}}$ for Fig. 5.5. Each row lists one valid correspondence for a reference segment $j \in \{1, 2\}$ in $\mathbf{F}_{\tau_1}$ at lag $\lambda \in \{-\tau_2, -\tau_1, +\tau_2\}$.

Reference segment j	λ	Corresponding segment (index)
1	$-\tau_2$	1
1	$-\tau_1$	1
1	$+\tau_2$	1
2	$-\tau_2$	2
2	$-\tau_1$	2
2	$+\tau_2$	2

D. Forward Interpolation

For a target row inside the gap to be interpolated (midpoint or any interior timestamp), this step predicts the foreground segments that should be present at the desired target time, using the near-past neighbor $\mathbf{P}_{\tau_1}$ as the reference source of geometry.

For each image row r from $\mathbf{P}_{\tau_1}$, let the set of reference segments be defined as

$$\text{Seg}_r(-\tau_1) = \{[h_{-\tau_1}^{(j)}, t_{-\tau_1}^{(j)}]\}_{j=1}^{N_{r,-\tau_1}}, \quad (5.7)$$

where $N_{r,-\tau_1}$ is the number of foreground segments detected on row r in $\mathbf{P}_{\tau_1}$, and the index $j \in \{1, \dots, N_{r,-\tau_1}\}$ orders these segments from left to right along the row. All correspondence information for these references is retrieved from the past-referenced map ($\text{cmap}_r^{\mathbf{P}_{\tau_1}}$) constructed in Step C.

Corresponding segments collection.

For the current row r and each reference segment j in $\mathbf{P}_{\tau_1}$, the past-referenced map $\text{cmap}_r^{\mathbf{P}_{\tau_1}}$ lists all corresponding segments at lags $\lambda \in \{-\tau_2, +\tau_1, +\tau_2\}$. For every correspondence (j, λ, k) , the endpoints $(h_\lambda^{(k)}, t_\lambda^{(k)})$ and the lag λ are aggregated. The near-past *reference* segment $(h_{-\tau_1}^{(j)}, t_{-\tau_1}^{(j)})$ is also included explicitly. Thus the list of endpoint proposals becomes

$$\mathcal{E}_j = \{(h_\lambda^{(k)}, t_\lambda^{(k)}, \lambda)\}_{\text{from cmap}} \cup \{(h_{-\tau_1}^{(j)}, t_{-\tau_1}^{(j)}, -\tau_1)\}. \quad (5.8)$$

Multiple correspondences may occur for the same (j, λ) and are retained at this stage.

Handle split/merge segments.

Both near- and farther-future segment split/merge is handled by forming a simple envelope. For example, when the correspondence map lists more than one corresponding segment at the near-future lag $+\tau_1$ for the same reference index j (e.g. around mid-swing when the legs separate), the near-future contribution is consolidated so that it covers the full plausible extent. Concretely, among all near-future correspondences for j , the leftmost head and the rightmost tail are selected,

$$h_{+\tau_1}^{\min} = \min_k h_{+\tau_1}^{(k)}, \quad t_{+\tau_1}^{\max} = \max_k t_{+\tau_1}^{(k)}, \quad (5.9)$$

and the multiple entries at $(+\tau_1)$ in the list of endpoint proposals $\mathcal{E}_j$ are replaced by a single envelope entry $(h_{+\tau_1}^{\min}, t_{+\tau_1}^{\max}, +\tau_1)$. This substitution preserves thin or split structures that would otherwise be truncated by selecting only one correspondence.

Weight definition by temporal proximity.

Let the target be $t_0 + s$ with $(s \in]-\tau_1, +\tau_1[)$. For each entry $(\cdot, \cdot, \lambda) \in \mathcal{E}_j$, the temporal distance is $\Delta t_s(\lambda) = |\lambda - s|$, and the normalized inverse-distance weight is

$$w_s(\lambda) = \frac{\Delta t_s(\lambda)^{-1}}{\sum_{(\cdot, \cdot, \lambda') \in \mathcal{E}_j} \Delta t_s(\lambda')^{-1}}. \quad (5.10)$$

The forward endpoint estimates are then

$$\hat{h}_j^{\text{fwd}} = \sum_{(h_\lambda^{(k)}, \cdot, \lambda) \in \mathcal{E}_j} w_s(\lambda) h_\lambda^{(k)}, \quad \hat{t}_j^{\text{fwd}} = \sum_{(\cdot, t_\lambda^{(k)}, \lambda) \in \mathcal{E}_j} w_s(\lambda) t_\lambda^{(k)}. \quad (5.11)$$

Example: a single missing frame at the midpoint target.

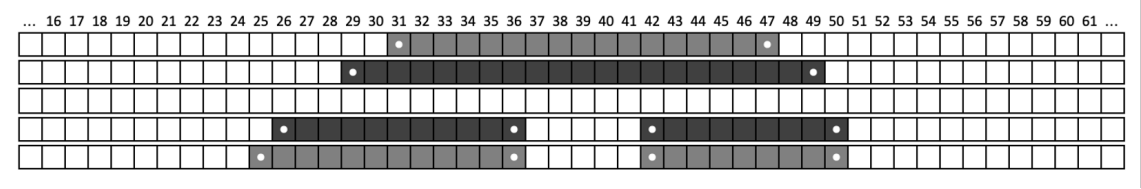


Figure 5.6: Amplified view of a single scanline r from each neighbor: $\mathbf{P}_{\tau_2}$, $\mathbf{P}_{\tau_1}$, $\mathbf{F}_{\tau_1}$, $\mathbf{F}_{\tau_2}$, ordered from farther past to farther future relative to t_0 . Segments are presented as: $\mathbf{P}_{\tau_2}$: [31,47]; $\mathbf{P}_{\tau_1}$: [29,49]; $\mathbf{F}_{\tau_1}$: [26,36] and [42,50]; $\mathbf{F}_{\tau_2}$: [25,36] and [42,50]. Envelopes for split cases are indicated at $+\tau_1$ and $+\tau_2$.

In this example the near future exhibits a split: $\mathbf{F}_{\tau_1}$ contains two segments [26,36] and [42,50], and $\mathbf{F}_{\tau_2}$ likewise [25,36] and [42,50]. These are consolidated into envelopes [26,50] at $+\tau_1$ and [25,50] at $+\tau_2$, which replace the multiple entries before weighting.

In this case where the midpoint target is $s = 0$, and considering $\tau_1 = 1$, $\tau_2 = 2$, the temporal distances between the four lags and the target t_0 are $|\tau_2 - 0| = 2$, $|\tau_1 - 0| = 1$, $|+\tau_1 - 0| = 1$, $|+\tau_2 - 0| = 2$. Inverse-distance weights (normalized) are therefore (0.1667, 0.3333, 0.3333, 0.1667) for $\mathbf{P}_{\tau_2}$, $\mathbf{P}_{\tau_1}$, $\mathbf{F}_{\tau_1}$, $\mathbf{F}_{\tau_2}$, respectively. Using the segments from $\mathbf{P}_{\tau_2}$ and $\mathbf{P}_{\tau_1}$ and the near-future envelopes from $\mathbf{F}_{\tau_1}$ and $\mathbf{F}_{\tau_2}$ for the split:

$$\mathbf{P}_{\tau_2} : [31, 47], \quad \mathbf{P}_{\tau_1} : [29, 49], \quad \text{env}(+\tau_1) : [26, 50], \quad \text{env}(+\tau_2) : [25, 50],$$

the forward endpoints become

$$\hat{h}^{\text{fwd}} = 0.1667 \cdot 31 + 0.3333 \cdot 29 + 0.3333 \cdot 26 + 0.1667 \cdot 25 \approx 27.67,$$

$$\hat{t}^{\text{fwd}} = 0.1667 \cdot 47 + 0.3333 \cdot 49 + 0.3333 \cdot 50 + 0.1667 \cdot 50 \approx 49.17.$$

After rounding the values and enforcing head-tail ordering, the target-row segment becomes [28,49], which, in fact, matches the expected values.

Missing near-future correspondence.

When no correspondence is available at the near-future lag $+\tau_1$ for a given reference segment j , the forward estimate is formed from the remaining entries in $\mathcal{E}_j$, namely those at lags $-\tau_2$, $-\tau_1$ (the reference itself), and $+\tau_2$. Temporal weighting is applied unchanged, normalized over this reduced set, so that the near-past reference $-\tau_1$ typically dominates for interior targets, with far neighbors contributing in proportion to their temporal proximity. In the case that the only available entry is the reference segment at $-\tau_1$, the forward endpoints default to $[h_{-\tau_1}^{(j)}, t_{-\tau_1}^{(j)}]$, subject to the usual geometric checks (clipping to $[1, W]$, head-tail ordering, non-empty length). If $\mathcal{E}_j$ is empty (i.e. no existing correspondence at any lag), the forward pass emits no segment for j . Even so,

subsequent steps (backward interpolation in Step E and fusion in Step F) may still produce a valid segment at the target row.

Geometric and consistency checks.

For each provisional forward estimate $[\hat{h}_j^{\text{fwd}}, \hat{t}_j^{\text{fwd}}]$ on row r , the following checks are applied sequentially:

(i) *Bounds and rounding.* Endpoints are clipped to the valid width range and then quantized to pixel indices:

$$\tilde{h}_j^{\text{fwd}} = \text{round}(\text{clip}(\hat{h}_j^{\text{fwd}}, 1, W)), \quad \tilde{t}_j^{\text{fwd}} = \text{round}(\text{clip}(\hat{t}_j^{\text{fwd}}, 1, W)). \quad (5.12)$$

(ii) *Segment integrity condition.* A segment is kept only when $\tilde{h}_j^{\text{fwd}} < \tilde{t}_j^{\text{fwd}}$, equivalently when its length $L_j^{\text{fwd}} = \tilde{t}_j^{\text{fwd}} - \tilde{h}_j^{\text{fwd}} + 1$ satisfies $L_j^{\text{fwd}} \geq 1$. Otherwise, the estimate for index j is discarded.

Estimates that satisfy all checks are taken as the forward segments for row r at the target time and are propagated to the forward/backward fusion step.

E. Backward Interpolation

Backward interpolation mirrors the forward construction but uses the near-future neighbor $\mathbf{F}_{\tau_1}$ as the reference. For each image row r , let the reference segments be

$$\text{Seg}_r(+\tau_1) = \{[h_{+\tau_1}^{(j)}, t_{+\tau_1}^{(j)}]\}_{j=1}^{N_{r,+\tau_1}}, \quad (5.13)$$

where $N_{r,+\tau_1}$ is the number of foreground segments on row r in $\mathbf{F}_{\tau_1}$, indexed left-to-right by j .

Corresponding segments collection.

For each reference index j , the future-referenced map $\text{cmap}_r^{\mathbf{F}_{\tau_1}}$ lists all corresponding segments at lags $\lambda \in \{-\tau_2, -\tau_1, +\tau_2\}$. Together with the reference segment at $+\tau_1$, these form the proposal list

$$\mathcal{E}_j^{\text{bwd}} = \{(h_\lambda^{(k)}, t_\lambda^{(k)}, \lambda)\}_{\text{from cmap}} \cup \{(h_{+\tau_1}^{(j)}, t_{+\tau_1}^{(j)}, +\tau_1)\}. \quad (5.14)$$

Multiple correspondences per (j, λ) are retained at this stage.

Handle split/merge segments.

If $-\tau_1$ (or $-\tau_2$) contains multiple correspondences for the same reference index j , these are consolidated into a single envelope by

$$h_{-\tau_1}^{\min} = \min_k h_{-\tau_1}^{(k)}, \quad t_{-\tau_1}^{\max} = \max_k t_{-\tau_1}^{(k)}, \quad (5.15)$$

and the $-\tau_1$ entries in $\mathcal{E}_j^{\text{bwd}}$ are replaced by the envelope $(h_{-\tau_1}^{\min}, t_{-\tau_1}^{\max}, -\tau_1)$. This preserves the full lateral extent when the near past differs topologically from the reference.

Weight definition by temporal proximity.

With $t_0 + s$, $s \in (-\tau_1, +\tau_1)$, define $\Delta t_s(\lambda) = |\lambda - s|$ for each $(\cdot, \cdot, \lambda) \in \mathcal{E}_j^{\text{bwd}}$ and

$$w_s(\lambda) = \frac{\Delta t_s(\lambda)^{-1}}{\sum_{(\cdot, \cdot, \lambda') \in \mathcal{E}_j^{\text{bwd}}} \Delta t_s(\lambda')^{-1}}. \quad (5.16)$$

Backward endpoint estimates become

$$\hat{h}_j^{\text{bwd}} = \sum_{(h_\lambda^{(k)}, \cdot, \lambda) \in \mathcal{E}_j^{\text{bwd}}} w_s(\lambda) h_\lambda^{(k)}, \quad \hat{t}_j^{\text{bwd}} = \sum_{(\cdot, t_\lambda^{(k)}, \lambda) \in \mathcal{E}_j^{\text{bwd}}} w_s(\lambda) t_\lambda^{(k)}. \quad (5.17)$$

Missing near-past correspondence.

When no correspondence exists at $-\tau_1$ for index j , the estimate is formed from the remaining lags in $\mathcal{E}_j^{\text{bwd}}$, which always include the reference $+\tau_1$ and may include $-\tau_2$ and $+\tau_2$. The weights are re-normalized over the reduced set. If the only available entry is the reference segment at $+\tau_1$, the backward estimate defaults to $[h_{+\tau_1}^{(j)}, t_{+\tau_1}^{(j)}]$, subject to the geometric checks below. If $\mathcal{E}_j^{\text{bwd}}$ is empty, no backward segment is emitted for j .

Geometric and consistency checks.

Backward estimates undergo the same clipping, rounding, nontrivial-length, and left-to-right ordering checks described in Step D.

Estimates that satisfy all checks constitute the backward segments for row r at the target time and are passed to the subsequent fusion step.

F. Fusion and Topology Guard

Let $[\tilde{h}_j^{\text{fwd}}, \tilde{t}_j^{\text{fwd}}]$ and $[\tilde{h}_j^{\text{bwd}}, \tilde{t}_j^{\text{bwd}}]$ denote the validated (forward and backward) segments for index j on row r at the target time, after the geometric and ordering checks of Steps D and E.

When both forward and backward segments are available for the same index j , agreement is quantified with two simple 1D measures on endpoints $[\tilde{h}_j^{\text{fwd}}, \tilde{t}_j^{\text{fwd}}]$ and $[\tilde{h}_j^{\text{bwd}}, \tilde{t}_j^{\text{bwd}}]$. First, the overlap ratio (intersection-over-union along the row) is

$$L_{\text{ov}} = \max(0, \min(\tilde{t}_j^{\text{fwd}}, \tilde{t}_j^{\text{bwd}}) - \max(\tilde{h}_j^{\text{fwd}}, \tilde{h}_j^{\text{bwd}}) + 1), \quad L_{\text{un}} = \max(\tilde{t}_j^{\text{fwd}}, \tilde{t}_j^{\text{bwd}}) - \min(\tilde{h}_j^{\text{fwd}}, \tilde{h}_j^{\text{bwd}}) + 1, \quad (5.18)$$

$$\text{overlap}(j) = \frac{L_{\text{ov}}}{L_{\text{un}}}. \quad (5.19)$$

Second, endpoint displacement is

$$\delta(j) = |\tilde{h}_j^{\text{fwd}} - \tilde{h}_j^{\text{bwd}}| + |\tilde{t}_j^{\text{fwd}} - \tilde{t}_j^{\text{bwd}}|. \quad (5.20)$$

If either $\text{overlap}(j) \geq 0.5$ or $\delta(j) \leq 2$ pixels, the two estimates are treated as consistent and a single fused segment is obtained by averaging endpoints:

$$\tilde{h}_j^* = \text{round}(\alpha_j \tilde{h}_j^{\text{fwd}} + (1 - \alpha_j) \tilde{h}_j^{\text{bwd}}), \quad \tilde{t}_j^* = \text{round}(\alpha_j \tilde{t}_j^{\text{fwd}} + (1 - \alpha_j) \tilde{t}_j^{\text{bwd}}). \quad (5.21)$$

The mixing weight $\alpha_j \in [0, 1]$ is chosen adaptively to favor the estimate supported by the neighbor that is temporally closer to the target time: it tends toward 1 when the target lies nearer the near past (thus favoring the forward estimate), tends toward 0 when the target lies nearer the near future (favoring the backward estimate), and is close to 1/2 around the midpoint of the gap.

When the two estimates disagree ($\text{overlap}(j) < 0.5$ and $\delta(j) > 2$), the fused segment is taken from the estimate near the target. If only one estimate is present, that estimate defines $[\tilde{h}_j^*, \tilde{t}_j^*]$.

After the generation of the complete binary silhouette at the target time, a minimal post-processing procedure is conducted to reduce some noise while preserving the morphology of the silhouette: (i) one-pixel peaks are removed; (ii) one-pixel enclosed holes are filled; and (iii) connected components with area below a small threshold (defined as $< 0.1\%$ of the median silhouette area) are discarded. These operations aim to stabilize contours and limb connectivity without introducing blur or altering gross shape.

The resulting silhouette constitutes the final fused estimate for the target time and is used in subsequent evaluation and recognition experiments.

5.3.2 BiSINet - Bidirectional Silhouette Interpolation Network

The **Bidirectional Silhouette Interpolation Network (BiSINet)** is a deep learning framework that reconstructs missing human silhouettes within gait sequences, fusing spatiotemporal cues from temporally adjacent past and future silhouettes to recover the target frame.

Its primary goal is to infer missing silhouettes within a motion sequence while preserving spatial integrity and temporal continuity. To improve generalization, BiSINet uses a multi-gap temporal training regime that exposes the model to varying degrees of temporal interpolation difficulty during learning.

Multi-Gap Temporal Training Strategy

To strengthen robustness across different temporal densities, training sequences are generated under three distinct *gap configurations*, each defined by the temporal gap parameter Δ , which specifies the frame spacing between consecutive visible silhouettes. The three configurations (zero-gap, one-gap, and two-gap) are summarized in Table 5.3, including the exact frame indices used for context and the corresponding prediction target in each case (see the *Prediction Rule* column).

Table 5.3: Multi-gap illustrative configurations used in BiSINet training. Δ is the temporal gap (in frames) between visible inputs. *Notation:* $S_{a:b}$ are consecutive frames; $S_{a:b:s}$ denotes frames from a to b sampled every s frames.

Config.	Δ	Frames	Prediction Rule
Zero-gap	0	S_{1-9}	Predict S_5 from past $S_{1:4}$ and future $S_{6:9}$.
One-gap	1	S_{1-17}	Predict S_9 from past $S_{1:7:2}$ and future $S_{11:17:2}$.
Two-gap	2	S_{1-25}	Predict S_{13} from past $S_{1:10:3}$ and future $S_{16:25:3}$.

Each sample thus comprises:

- a **past stream** of four silhouettes preceding the missing frame,
- a **future stream** of four silhouettes following it, and
- the **target silhouette** S_m , representing the missing intermediate frame.

The training alternates dynamically among all three configurations, allowing BiSINet to generalize across both dense and sparse temporal patterns.

Model Architecture Overview

BiSINet adopts a dual-stream, bidirectional encoder-decoder structure (Figure 5.7) that integrates spatio-temporal reasoning with ConvLSTM layers and attention-based fusion. The network comprises two symmetric encoders that process past and future sequences independently. Each stream passes through three hierarchical encoder blocks; every block contains two 3×3 convolutional layers with batch normalization and ReLU activation, followed by a ConvLSTM cell to model temporal dependencies within the stream, and a 2×2 max-pooling operation for spatial down-sampling. This arrangement jointly captures fine spatial detail (e.g. limb contours) and temporal continuity (e.g. motion trajectories).

At each encoder depth, the representations from the two temporal directions are merged by an attention-fusion operation. Given feature maps F_{past} and F_{future} , an attention mask A is computed as

$$A = \sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}([F_{\text{past}}, F_{\text{future}}])))), \quad (5.22)$$

and fusion is performed according to

$$F_{\text{fused}} = A \odot F_{\text{past}} + (1 - A) \odot F_{\text{future}}, \quad (5.23)$$

where $\odot$ denotes element-wise multiplication. In practice, this mechanism allows the model to emphasize whichever temporal direction provides the most reliable evidence at each spatial location.

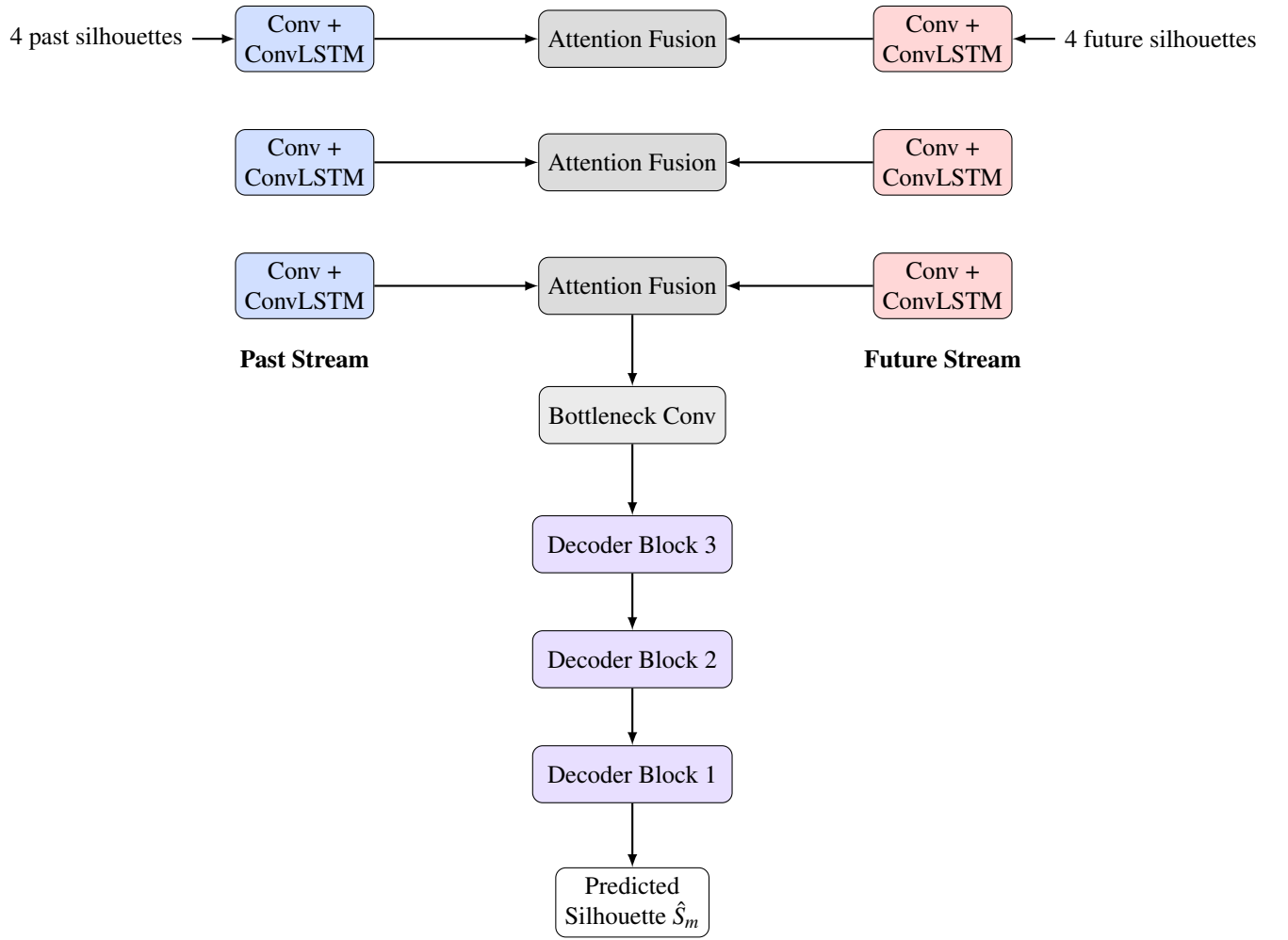


Figure 5.7: Simplified schematic of the proposed BiSINet architecture. Past and future silhouette streams are encoded separately through ConvLSTM-based encoders. Their multi-scale features are fused by attention modules, passed through a bottleneck, and decoded to reconstruct the missing intermediate silhouette $\hat{S}_m$.

Following multi-scale fusion, a bottleneck convolutional block increases representational capacity prior to decoding. The decoder mirrors the encoder with transposed convolutions for up-sampling, skip connections for spatial alignment with earlier encoder features, and standard convolutions for refinement. Finally, a 1×1 convolution maps the decoded features to a single-channel logits map representing the predicted intermediate silhouette, denoted as $\hat{S}_m$.

Training and Optimization

Training minimizes a composite objective that combines Binary Cross-Entropy (BCE) with Soft Dice loss. The total loss is

$$\mathcal{L} = \lambda_{\text{BCE}} \mathcal{L}_{\text{BCE}} + \lambda_{\text{Dice}} \mathcal{L}_{\text{Dice}}, \quad (5.24)$$

with

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum(p \cdot t) + \varepsilon}{\sum p + \sum t + \varepsilon}, \quad (5.25)$$

where p and t denote the predicted and ground-truth silhouettes, respectively. Empirically, $\lambda_{\text{BCE}} = 1.0$ and $\lambda_{\text{Dice}} = 0.5$ yielded the most stable convergence.

Optimization uses AdamW with an initial learning rate of 2×10^{-4} and a cosine-annealing schedule with warm restarts ($T_0 = 10$ epochs). Mixed-precision (AMP) is enabled on NVIDIA A100 GPUs. The batch size is 4 and training proceeds for 80 epochs. Checkpoints are written every two epochs to guard against runtime interruptions. Performance is monitored via Intersection over Union (IoU), Dice coefficient, and pixel-level accuracy.

Temporal Inference and Evaluation

During inference, BiSINet receives a pair of temporal sequences – four past silhouettes and four future silhouettes – and predicts the missing intermediate silhouette. The output is a probability map obtained through sigmoid activation, subsequently thresholded to form the binary silhouette $\hat{S}_m$.

Qualitative evaluation involves visual comparison between:

- the input context (past and future silhouettes),
- the predicted silhouette $\hat{S}_m$, and
- the ground-truth silhouette S_m ,

while quantitative evaluation uses IoU, Dice, and pixel accuracy. Error maps and prediction histograms further highlight model precision and uncertainty distribution.

BiSINet Architecture

Table 5.4: Summary of BiSINet’s structural components and output dimensions.

Component	Description	Output Dimensions
Input	8 silhouettes (4 past + 4 future)	(B, 8, 1, 128, 88)
Encoder 1	2×Conv + ConvLSTM + MaxPool	(B, 32, 64, 44)
Encoder 2	2×Conv + ConvLSTM + MaxPool	(B, 64, 32, 22)
Encoder 3	2×Conv + ConvLSTM + MaxPool	(B, 128, 16, 11)
Attention Fusion	Adaptive merge of past/future features	Fused maps
Bottleneck	3×3 Conv + BN + ReLU	(B, 256, 16, 11)
Decoder 3-1	Transposed Conv + Skip + 2×Conv	Progressive upsampling
Output	1×1 Conv → binary logits	(B, 1, 128, 88)

Distinctive Characteristics

- **Multi-gap temporal learning:** Exposure to zero-, one-, and two-gap sequences enhances robustness across varying frame densities.
- **Bidirectional feature reasoning:** Dual ConvLSTM encoders capture motion dynamics from both temporal directions.
- **Attention-guided fusion:** Learns adaptive weighting between past and future contexts for each spatial location.
- **Lightweight and efficient:** Optimized for real-time GPU execution and mixed-precision stability.
- **Generalization capacity:** Maintains silhouette accuracy across both slow and fast gait patterns.

5.4 Experimental Setup

The experimental evaluation of the proposed methods was designed to assess their ability to reconstruct temporally coherent silhouettes under different interpolation conditions and to examine the downstream impact of interpolation on gait-based person recognition. Three complementary experiments were defined, each targeting a distinct aspect of temporal manipulation and motion continuity:

- **Experiment 1: Temporal Gap Reconstruction** – evaluates robustness to local missing frames by reconstructing short gaps within otherwise continuous sequences.
- **Experiment 2: Temporal Super-Resolution (Up-Sampling)** – analyzes scalability when synthesizing intermediate silhouettes to increase the effective frame rate under controlled decimation.
- **Experiment 3: Influence of Framerate Variation on Gait Recognition** – investigates how decimation and subsequent interpolation affect gait-based person recognition performance using a standard silhouette-derived representation.

The first experiment simulates frame-gap filling, a scenario where isolated frames within an existing sequence are either missing or corrupted. This setup evaluates how effectively each method can restore local temporal consistency and structural completeness. It reflects realistic degradations often found in gait datasets, such as dropped frames or segmentation errors, and serves as a controlled test of robustness and reconstruction fidelity.

The second experiment focuses on frame-rate up-sampling, where the goal is to generate additional intermediate silhouettes between consecutive frames, thereby increasing the temporal resolution of the video. In this configuration, the temporal spacing between reference frames is

artificially expanded according to the chosen up-sampling factor, creating a more demanding setting in which the interpolation must capture subtle transitions across larger motion intervals. This experiment thus explores the scalability of both approaches, i.e. how the shape-based BRIEF and the learning-based BiSINet perform as the temporal distance between inputs grows and motion dynamics become more complex.

The third experiment links interpolation quality to downstream utility by evaluating how temporal decimation and subsequent interpolation affect a simple gait-based person-recognition pipeline. Using a well-established silhouette descriptor, the analysis quantifies how much recognition performance is preserved or recovered when sequences are down-sampled and subsequently reconstructed by the proposed interpolation methods.

5.4.1 Datasets

The experiments were conducted using three gait datasets: the Kyushu University 4D Gait Database A (KY4D-A), the GRIDDS dataset and the OU-ISIR Treadmill Dataset A. The inclusion of multiple datasets ensured a comprehensive evaluation of the proposed interpolation methods under different acquisition conditions, framerates, and spatial resolutions. While KY4D-A and GRIDDS capture gait in natural walking conditions with a static lateral viewpoint, the OU-ISIR Treadmill Dataset A provides controlled motion with systematic speed variations. This diversity allowed for the analysis of the robustness of the proposed methods with respect to temporal continuity, motion dynamics, and spatial detail. In addition, both quantitative metrics and qualitative assessments were compared with previous studies using the same datasets, allowing a fair contextualization of performance.

5.4.1.1 KY4D-A (Straight Walking)

The *Kyushu University 4D Gait Database A (KY4D-A)* (Iwashita et al. (2010)) was collected in a controlled multi-camera studio using a ring of 16 synchronized RGB cameras positioned at 3.5 m from the walking area. Each camera captured sequences at a resolution of 1032×776 pixels and 15 fps, accompanied by binary silhouettes, calibration data, and visual-hull 3D reconstructions. For this work, only silhouettes from a single 90° side-view camera were used, corresponding to individuals walking on a straight path. The dataset comprises 42 subjects, each performing four sequences, totaling roughly 200,000 silhouettes. All silhouettes were cropped and normalized to 120×80 pixels to ensure spatial consistency between samples. A more detailed technical description of the dataset is provided in Section 4.4.3.

5.4.1.2 GRIDDS

The *Gait Recognition Image and Depth Dataset (GRIDDS)* (Nunes et al. (2019a)) includes 35 participants recorded indoors while walking along two straight trajectories, each covered five times per direction (left-to-right and right-to-left), producing a total of 350 sequences and approximately 33,000 silhouette images. Recordings were acquired at 30 fps using a Kinect v2 sensor mounted

laterally at 1.8 m height. The silhouettes, aligned and normalized to 80×120 pixels, were used in this study. More details, including the acquisition environment and the recording protocol, are available in Section 4.2.

5.4.1.3 OU-ISIR Treadmill Dataset A (Speed Variation)

The *OU-ISIR Treadmill Gait Dataset A (Speed Variation)* (Tsuji et al. (2010)) comprises recordings of 34 subjects walking on a treadmill at controlled speeds ranging from 2 km/h to 10 km/h in 1 km/h increments. Each participant performed roughly two sequences per speed, resulting in approximately 612 sequences and over 480,000 silhouette images, captured at 60 fps from a lateral perspective and normalized to 128×88 pixels. The controlled treadmill environment introduces *speed* as the main covariate, allowing the evaluation of the proposed interpolation methods under systematically varying motion rates (see Section 4.4.3).

5.4.2 Tasks and Protocols

Experiment 1: Temporal Gap Reconstruction.

This experiment evaluated the capability of the proposed methods to reconstruct missing silhouettes within a continuous motion sequence. The task simulates frame-loss scenarios, where individual frames are absent due to a variety of factors, including acquisition failures, transmission errors, or segmentation defects, resulting in local temporal discontinuities. An example configuration is illustrated in Figure 5.8, which shows a three-frame gap where the orange silhouettes denote the targets to be synthesized and the remaining silhouettes provide the available temporal context. By testing the ability of each method to restore such missing content, the experiment aimed to assess their robustness and temporal consistency in realistic conditions of data degradation.



Figure 5.8: Illustration of the *Temporal Gap Reconstruction* experiment for a three-frame gap. The silhouettes highlighted in orange represent the missing frames to be interpolated, while the remaining silhouettes correspond to the available temporal context used by the interpolation methods.

Controlled gaps, ranging from one to three consecutive frames, were injected into all three datasets. Interpolation was then performed using each method's native temporal context: **BRIEF** inherently consumes the two nearest frames on each side (± 2 anchors), whereas BiSINet is defined to use four frames before and four after (± 4 anchors). These windows are part of the algorithmic design and were not tuned or varied as an experimental factor. The reconstructed silhouettes were compared against the corresponding ground-truth frames. Because BiSINet predicts only a single intermediate frame between adjacent inputs, it was evaluated exclusively for the one-frame gap condition; **BRIEF** was assessed for gaps of one, two, and three frames.

5.4.2.1 Experiment 2: Temporal Super-Resolution (Up-Sampling)

This experiment evaluates the ability of silhouette interpolation methods to enhance temporal resolution by synthesizing intermediate silhouettes between sparsely sampled frames (i.e., temporal super-resolution). The efficacy of up-sampling factors of 2, 3, and 4 was investigated using the same three datasets. To ensure a controlled evaluation with known ground truth, each sequence was artificially down-sampled, preserving every r^{th} frame (e.g. 30 fps $\rightarrow$ 15/10/7.5 fps for $r = 2, 3, 4$, respectively). The removed frames thus served as interpolation targets. Both methods then reconstructed the $r - 1$ missing silhouettes between each retained pair, restoring the sequence to its original framerate. As in the previous experiment, each method was provided with its corresponding temporal context: **BRIEF** used the two nearest frames before and after the missing interval (± 2 anchors), while **BiSINet** used four before and four after (± 4 anchors).

In contrast to the *Temporal Gap Reconstruction* experiment, in which adjacent context frames remained one frame apart, this decimation process increases the temporal spacing between available anchors to exactly r frames. Consequently, the interpolation task must now bridge larger motion intervals, often spanning distinct gait phases (e.g. opposite swing or stance positions). This setting thus evaluates the scalability of both approaches, in terms of their ability to maintain structural fidelity, limb connectivity, and motion smoothness as the temporal gap widens.

The evaluation process employs the same protocol as that outlined in the *Temporal Gap Reconstruction* experiment. The reconstruction of each silhouette is then compared with its ground truth at the same timestamp, with this comparison made using both pixel-level and structural metrics. These quantitative results are complemented by a qualitative inspection of contour continuity and perceptual fluidity of motion, providing a comprehensive understanding of the models' interpolation performance under increased temporal uncertainty.

Table 5.5 summarizes the decimation-restoration protocol used in this experiment. The *Framerate* column indicates the original frame rate of each dataset. Under *Factor 2/3/4*, the left value represents the down-sampled fps, and the right value the restored fps after interpolation. For example, in the GRIDDS dataset originally recorded at 30 fps, a factor 2 up-sampling corresponds to 15 $\rightarrow$ 30 fps, and a factor 3 to 10 $\rightarrow$ 30 fps.

Table 5.5: Frame-rate up-sampling transformations for the utilized datasets. Arrows ($\downarrow\uparrow$) denote the down-sampling ($\downarrow$) and subsequent up-sampling ($\uparrow$) framerates.

Dataset	Framerate (fps)	Factor 2 ($\downarrow\uparrow$ fps)	Factor 3 ($\downarrow\uparrow$ fps)	Factor 4 ($\downarrow\uparrow$ fps)
KY4D-A	15	7.5 $\rightarrow$ 15	5 $\rightarrow$ 15	3.75 $\rightarrow$ 15
GRIDDS	30	15 $\rightarrow$ 30	10 $\rightarrow$ 30	7.5 $\rightarrow$ 30
OU-ISIR	60	30 $\rightarrow$ 60	20 $\rightarrow$ 60	15 $\rightarrow$ 60

A visual illustration of the up-sampling process for different factors is presented in Figure 5.9. The top row depicts a sequence down-sampled by a factor of two, and the bottom row by a factor of three. In each case, the target silhouette to be interpolated is highlighted in orange, while the past and future context silhouettes are shown in blue and green, respectively. As the up-sampling factor

r increases, the temporal distance between available context silhouettes also increases, making the interpolation task progressively more challenging.



Figure 5.9: Illustration of the up-sampling process under different temporal factors in the *Temporal Super-Resolution* experiment. The top row corresponds to down-sampling by a factor of two ($r = 2$), and the bottom row to three ($r = 3$). The orange silhouettes represent the frames to be interpolated, while the blue and green silhouettes indicate the past and future context frames, respectively. As the factor r increases, the temporal spacing between available context frames also increases, highlighting the growing difficulty of the interpolation task.

5.4.2.2 Experiment 3: Influence of Framerate Variation on Gait Recognition

To illustrate a practical utility of silhouette interpolation, this experiment evaluates the impact of temporal decimation and subsequent interpolation on gait-based person recognition. Following a methodology analogous to Liu and Sarkar (2004), the *Gait Energy Image* (GEI) (Han and Bhanu (2006)) is adopted as a compact silhouette descriptor.

Protocol

The evaluation protocol considers three sequence conditions per subject: (i) original sequences; (ii) **down-sampled** sequences obtained by retaining every fourth frame (factor 4); and (iii) up-sampled sequences produced by first down-sampling by factor 4 and then restoring the original frame rate using the interpolation methods (i.e., sequences reconstructed from the decimated inputs). For up-sampling, BiSINet generates intermediate silhouettes iteratively until the original cadence is recovered, whereas BRIEF synthesizes missing frames directly via its bidirectional row-wise strategy. Gallery and probe sets are constructed from subject sequences under each condition, and cumulative match characteristic (CMC) curves are computed for ranks 1-5 to quantify rank-based identification performance. Quantitative results are reported in the Results section, which analyzes the effect of frame-rate manipulation (decimation vs. interpolation) on recognition accuracy for KY4D-A and GRIDDS.

5.4.3 Evaluation Metrics

The evaluation of both experiments relies on a combination of quantitative and qualitative measures designed to assess the accuracy, structural fidelity, and perceptual coherence of the interpolated silhouettes in relation to their ground-truth counterparts. Three complementary metrics were employed: the **Jaccard Similarity Index** (IoU), the **Dice Similarity Coefficient** (Dice), and the **Normalized Mean Squared Error** (NMSE). Together, these metrics provide a balanced analysis of spatial overlap, segmentation precision, and pixel-wise reconstruction error. In addition, two

qualitative visualizations were used to enable direct inspection of contour accuracy and spatial consistency.

Jaccard Similarity Index (IoU)

The *Jaccard Similarity Index* – also known as *Intersection over Union* (IoU) – is a standard metric for measuring the similarity between binary regions (Jaccard (1901)). It evaluates the ratio between the area of intersection and the area of union of two binary masks, as expressed in Eq. 5.26. IoU values range from 0 to 1, where higher values denote a greater degree of overlap between the interpolated and ground-truth silhouettes.

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (5.26)$$

Alternatively, IoU can be written in terms of true positives (TP), false positives (FP), and false negatives (FN), as shown in Eq. 5.27. In this context, TP represents correctly predicted silhouette pixels, FP corresponds to predicted pixels outside the true silhouette, and FN denotes pixels missing from the predicted region.

$$IoU = \frac{TP}{TP + FP + FN} \quad (5.27)$$

The IoU metric has been widely adopted in segmentation benchmarks such as the PASCAL VOC Challenge (Everingham et al. (2009)) and the DAVIS dataset for video object segmentation (Perazzi et al. (2016)), where it serves as a robust measure of region-level agreement.

Dice Similarity Coefficient (Dice)

The *Dice Similarity Coefficient* (DSC) (Dice (1945)) is another overlap-based measure closely related to IoU, but it tends to be more sensitive to small regions and imbalanced classes. It computes twice the area of intersection divided by the total number of pixels in both regions, as shown in Eq. 5.28.

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \quad (5.28)$$

The Dice metric has become a standard in medical and volumetric image segmentation (Milletari et al. (2016)) and has been widely employed in human silhouette and pose-based segmentation tasks due to its strong correlation with perceptual similarity (Chu et al. (2024); Javanmardi et al. (2018)). High Dice values indicate precise spatial alignment and faithful reconstruction of silhouette boundaries.

Normalized Mean Squared Error (NMSE)

To complement the overlap-based metrics, the *Normalized Mean Squared Error* (NMSE) quantifies the average pixel-wise deviation between the interpolated and reference silhouettes. Unlike

IoU and Dice, which operate on binary masks, NMSE captures intensity-level differences, providing a sensitive indicator of small geometric or structural mismatches. It is defined as follows:

$$NMSE = \frac{\sum_{i,j} (GT_{i,j} - In_{i,j})^2}{\sqrt{\sum_{i,j} (GT_{i,j})^2 \times \sum_{i,j} (In_{i,j})^2}} \quad (5.29)$$

Here, $GT_{i,j}$ and $In_{i,j}$ denote the pixel intensities of the ground-truth and interpolated silhouettes, respectively. NMSE provides a quantitative assessment of reconstruction error useful to measure the fidelity of synthesized frames.

Qualitative Visualization

In addition to quantitative metrics, qualitative visualizations were employed to provide an intuitive assessment of interpolation accuracy. Specifically, an *Error Map* was generated to represent the pixel-wise reconstruction error between the interpolated and ground-truth silhouettes (see Fig. 5.10). Each pixel intensity encodes the absolute difference between the two masks, visualized using the "hot" colormap: black regions indicate perfect agreement (no reconstruction error), whereas brighter areas (yellow-white) highlight spatial discrepancies, such as shape deformations or misaligned contours. This visualization offers a spatial understanding of model performance that complements numerical measures such as IoU, NMSE, and Dice coefficients.



Figure 5.10: Error Map visualization of interpolation results. Each pixel value represents the absolute reconstruction error between the predicted and ground-truth silhouettes, encoded using the “hot” colormap (black = perfect agreement, yellow-white = increasing error).

5.5 Results and Analysis

This section reports the outcomes of the three experiments introduced in Section 5.4, using the metrics defined in Section 5.4.3. Results are organized to first quantify *image/shape fidelity* under gap filling and temporal super-resolution, followed by *comparisons with prior interpolation methods*, and finally an *application-level analysis* on gait-based person recognition. Two methods are evaluated: the shape-driven, deterministic **BRIEF** (Bidirectional Rowwise Interpolation with Envelope Fusion) and the learning-based **BiSINet** (Bidirectional Silhouette Interpolation Network).

At the time of writing, BiSINet experiments are concluding; therefore, cells marked with $\diamond$ indicate values to be inserted upon completion.

5.5.1 Experiments Results

This subsection evaluates **BRIEF** and **BiSINet** on the two interpolation tasks defined in Section 5.4: *Temporal Gap Reconstruction* (Experiment 1) and *Temporal Super-Resolution (Up-Sampling)* (Experiment 2). To ensure comparability with the other datasets – whose sequences correspond to normal walking – evaluation on OU-ISIR is restricted to the **5 km/h** subset. The first task analyzes performance as the number of consecutively missing frames increases, whereas the second probes scalability as the temporal spacing between visible anchors grows with the up-sampling factor.

Experiment 1: Temporal Gap Reconstruction. Table 5.6 summarizes average IoU over all interpolated silhouettes for gap sizes of 1-3 frames; Figure 5.11 visualizes the gap-length trend.

Table 5.6: Experiment 1 (Temporal Gap Reconstruction): mean IoU across gap sizes of 1–3 frames.

Dataset	Method	Gap 1	Gap 2	Gap 3
KY4D-A (15 fps)	BRIEF	0.901	0.853	0.814
	BiSINet	0.935	0.895	0.865
GRIDDS (30 fps)	BRIEF	0.910	0.889	0.865
	BiSINet	0.942	0.901	0.874
OU-ISIR TM-A (60 fps)	BRIEF	0.963	0.951	0.939
	BiSINet	0.971	0.962	0.950

Across datasets, BiSINet consistently outperforms BRIEF in temporal gap reconstruction, with the magnitude of its advantage depending on the native framerate and the extent of temporal discontinuity. On KY4D-A and GRIDDS, both of which exhibit lower framerates (15 fps and 30 fps, respectively), longer gaps introduce larger pose differences and greater temporal ambiguity. In such conditions, the spatio-temporal representation of BiSINet is advantageous, yielding higher IoU scores, particularly for gaps of two and three frames. Conversely, on the OU-ISIR TM-A dataset (60 fps), the temporal spacing between adjacent frames is substantially smaller, making the interpolation problem inherently easier and leaving less room for improvement over geometrically guided methods. Consequently, both BRIEF and BiSINet achieve near-saturated accuracy, with BiSINet retaining only a modest but consistent edge.

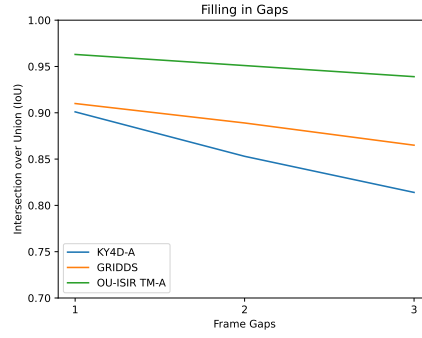


Figure 5.11: Experiment 1 (Temporal Gap Reconstruction): trend of mean IoU vs. gap size across datasets for **BRIEF**.

Experiment 2: Temporal Super-Resolution (Up-Sampling). Table 5.7 reports mean IoU for up-sampling factors $r \in \{2, 3, 4\}$; Figure 5.12 depicts the factor-dependent trend.

Table 5.7: Experiment 2 (up-sampling): mean IoU for factors $r \in \{2, 3, 4\}$.

Dataset	Method	Factor 2	Factor 3	Factor 4
KY4D-A (15 fps)	BRIEF	0.899	0.843	0.809
	BiSINet	0.935	0.863	0.746
GRIDDS (30 fps)	BRIEF	0.904	0.851	0.834
	BiSINet	0.937	0.880	0.859
OU-ISIR TM-A (60 fps)	BRIEF	0.959	0.912	0.889
	BiSINet	0.971	0.922	0.904

A consistent pattern emerges in the temporal super-resolution experiment: while both interpolation methods experience a natural degradation in accuracy as the up-sampling factor increases, BiSINet generally maintains the highest IoU across datasets for moderate temporal expansions. In KY4D-A and GRIDDS, BiSINet clearly outperforms BRIEF for factors 2 and 3, reflecting the model’s ability to infer intermediate silhouettes even under larger motion displacements induced by wider temporal intervals. However, for the most extreme setting on KY4D-A (factor 4), network performance deteriorates dramatically, falling below BRIEF, an effect consistent with the substantial increase in NMSE observed in the comparison of the datasets. This illustrates a limitation of the learned model when faced with temporal baselines longer than those prevalent in its training distribution. In contrast, on the OU-ISIR TM-A dataset, where the native 60 fps sampling yields smaller inter-frame motion, BiSINet retains a small but systematic advantage over BRIEF across all factors. Overall, these results suggest that BiSINet is highly effective for moderate temporal up-sampling, particularly in datasets with mid-range frame rates, while geometry-aware

methods such as BRIEF remain more reliable under extreme temporal sparsity or cross-domain conditions.

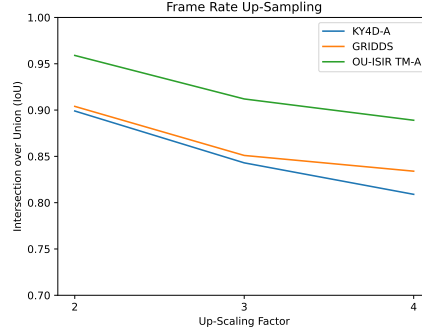


Figure 5.12: Experiment 2 (up-sampling): trend of mean IoU vs. factor across datasets for **BRIEF**.

5.5.2 Interpolation Methods Comparison

This subsection compares the proposed approaches against published interpolation baselines (Imoto et al. (2020); Driscoll et al. (2012)) and a standard Nearest-Neighbor scheme (Lehmann et al. (1999)), using the same datasets and the NMSE metric employed in Imoto et al. (2020). Values for KY4D-A and OU-ISIR TM-A from Imoto et al. (2020) are reported directly; GRIDDS results for Imoto et al. (2020) were obtained via the authors’ reference implementation (acknowledged).

Table 5.8: Up-sampling on KY4D-A: NMSE across factors $r \in \{2, 3, 4\}$.

KY4D-A (NMSE, lower is better)			
	Factor 2	Factor 3	Factor 4
Nearest-Neighbor (Lehmann et al. (1999))	0.613	0.613	0.760
Driscoll et al. (2012)	0.227	0.289	0.338
Imoto et al. (2020)	0.138	0.209	0.280
BRIEF	0.104	0.185	0.238
BiSINet	0.052	0.133	0.448

Overall, the results on KY4D-A confirm the effectiveness of both proposed approaches, with BRIEF consistently outperforming existing baselines across all up-sampling factors, and BiSINet showing competitive accuracy for moderate factors despite being trained exclusively on sequences from a different dataset, thus evidencing its potential for cross-dataset generalization when temporal characteristics are similar.

Table 5.9: Up-sampling on GRIDDS: NMSE across factors.

GRIDDS	(NMSE, lower is better)		
	Factor 2	Factor 3	Factor 4
Imoto et al. (2020)	0.175	0.180	0.196
BRIEF	0.101	0.162	0.194
BiSINet	0.050	0.091	0.120

On GRIDDS, **BRIEF** again yields the lowest NMSE. BiSINet values ($\diamond$) are pending.

Table 5.10: Up-sampling on OU-ISIR TM-A: NMSE across even walking speeds.

OU-ISIR TM-A: up-sampling 30 $\rightarrow$ 60 fps (even speeds 2–10 km/h), NMSE					
	2	4	6	8	10
Nearest-Neighbor (Lehmann et al. (1999))	0.168	0.203	0.259	0.272	0.289
Driscoll et al. (Driscoll et al. (2012))	0.174	0.203	0.253	0.247	0.259
Imoto et al. (Imoto et al. (2020))	0.163	0.190	0.237	0.236	0.247
BRIEF	0.021	0.030	0.044	0.060	0.067
BiSINet	0.021	0.020	0.029	0.038	0.052

Across datasets and settings, **BRIEF** shows consistent gains over prior interpolation baselines in NMSE.

5.5.3 Impact of Frame-Rate Manipulation on Recognition

To demonstrate practical utility, the impact of temporal decimation followed by interpolation on gait-based recognition is assessed using the same protocol as Liu and Sarkar (2004) and the *Gait Energy Image* (GEI) descriptor (Han and Bhanu (2006)). Given a complete gait cycle of N binary silhouettes $\{B_t\}_{t=1}^N$, the gray-level GEI G is computed as the temporal mean:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y). \quad (5.30)$$

Similarity between a probe GEI G_P and a gallery $\{G_G(j)\}_{j=1}^{N_G}$ is defined as the negative of the median Euclidean distance:

$$\text{Similarity}(G_P, G_G) = -\text{Median} \left(\min_{j \in \{1, \dots, N_G\}} \|G_P - G_G(j)\| \right). \quad (5.31)$$

Recognition experiments were conducted on two silhouette datasets with different native frame-rates: **GRIDDS** (30 fps) and **KY4D-A** (15 fps). Three conditions are compared: **Original**,

Down-sampled (retain every fourth frame; e.g. $30 \rightarrow 7.5$ fps in GRIDDS and $15 \rightarrow 3.75$ fps in KY4D-A), and **Up-sampled** (decimated by 4 and then reconstructed to the original cadence via interpolation). This design allows assessment of recognition sensitivity to native temporal resolution as well as recovery after interpolation. Identification performance is summarized by CMC@1-5. For brevity, tables and plots report results for the three conditions.

Table 5.11: KY4D-A: CMC@1-5 for original, down-sampled (factor 4), and up-sampled conditions.

KY4D-A	Rank #1	Rank #2	Rank #3	Rank #4	Rank #5
Original	97.02%	98.21%	98.21%	98.21%	98.81%
Down-sampled	67.86%	73.21%	76.79%	80.95%	83.93%
Up-sampled (BRIEF)	73.21%	82.74%	86.90%	89.29%	91.07%

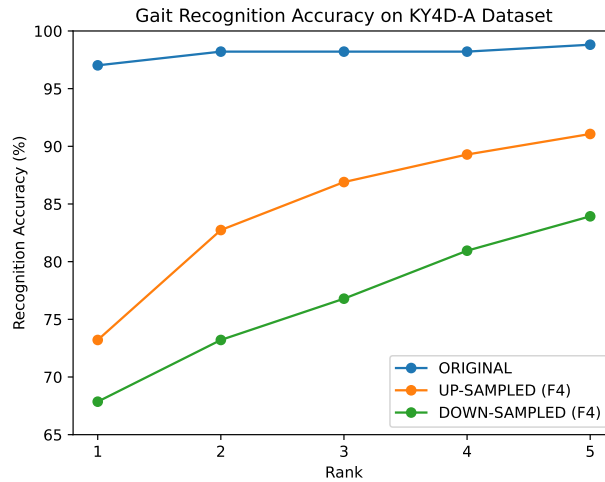


Figure 5.13: KY4D-A: CMC@1-5 curves contrasting original, down-sampled (factor 4), and up-sampled (factor 4) sequences.

Table 5.12: GRIDDS: CMC@1-5 for original, down-sampled (factor 4), and up-sampled conditions.

GRIDDS	Rank #1	Rank #2	Rank #3	Rank #4	Rank #5
Original	99.43%	99.43%	99.43%	99.43%	99.43%
Down-sampled	81.43%	82.29%	82.86%	83.14%	83.43%
Up-sampled (BRIEF)	98.00%	99.14%	99.14%	99.14%	99.14%

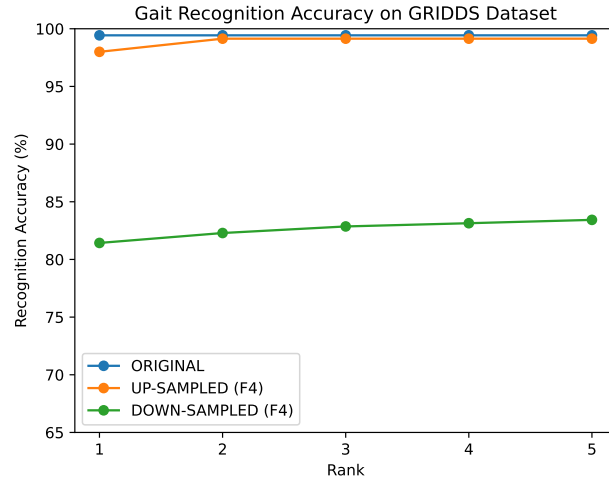


Figure 5.14: GRIDDS: CMC@1-5 curves contrasting original, down-sampled (factor 4), and up-sampled (factor 4) sequences.

Overall, recognition results indicate a strong dependence on temporal resolution: down-sampling degrades CMC@k, while interpolation recovers a substantial portion of the loss. **BRIEF**-based up-sampling narrows the gap to original performance on both datasets.

5.6 Conclusions and Limitations

5.6.1 Conclusions

The results across Experiments 1–3 support several consistent findings about binary silhouette interpolation for human gait, where the objective is phase-consistent, topology-preserving evolution of person shape over time.

(1) Temporal proximity is the primary driver of accuracy. Interpolation accuracy is higher in *Temporal Gap Reconstruction* (Experiment 1) than in *Temporal Super-Resolution* (Experiment 2). In Experiment 1, the supporting silhouettes are temporally adjacent (one-frame baseline), providing denser motion cues; in Experiment 2, the effective temporal baseline expands to r frames, reducing spatial overlap and increasing motion amplitude between anchors. This is reflected in the monotonic decline of IoU as the gap length (Exp. 1; Table 5.6, Fig. 5.11) or up-sampling factor (Exp. 2; Table 5.7, Fig. 5.12) increases. In short, shorter temporal baselines yield higher reconstruction fidelity.

(2) Native temporal resolution (temporal sampling density) correlates with reconstruction fidelity. Across datasets, higher native framerates are associated with higher IoU. This trend (OU-ISIR TM-A at 60 fps > GRIDDS at 30 fps > KY4D-A at 15 fps) indicates that denser temporal sampling improves the conditioning of the interpolation problem, likely by reducing inter-frame

displacement and ambiguity in correspondences. The effect is visible in both experiments and is most pronounced at longer gaps/factors.

(3) Deterministic, geometry-aware interpolation is robust for short-range recovery. BRIEF achieves consistently high IoU in Experiment 1 (all values > 0.80) and degrades smoothly as temporal distance grows in Experiment 2, while preserving limb connectivity by construction. Its interpretability and envelope-based safeguards make it a reliable choice when temporal context is limited.

(4) Learned bidirectional modeling shows promise where temporal context matches training statistics. BiSiNet was trained only on GRIDDS (30 fps). Despite this restriction, where results are available it performs well on GRIDDS and achieves competitive accuracy on the higher-frame-rate OU-ISIR subset (5 km/h), suggesting favorable transfer when temporal statistics (sampling density and motion pace) are aligned. A lower accuracy on lower-frame-rate data (e.g., KY4D-A) is plausibly attributable to domain shift in temporal sampling density and motion patterns relative to the training set.

(5) Downstream utility: interpolation reduces recognition loss due to temporal decimation. In Experiment 3, down-sampling by a factor of four leads to a marked drop in CMC-based identification accuracy on GRIDDS (30 fps) and KY4D-A (15 fps), whereas up-sampling via interpolation recovers a substantial portion of the loss. The reconstructed version brings recognition closer to the original cadence on both datasets, indicating that topology-preserving temporal densification benefits GEI-style pipelines.

(6) Complementary paradigms. Model-based (BRIEF) and data-driven (BiSiNet) strategies are complementary: BRIEF offers deterministic behavior and strong short-gap recovery with explicit topology handling; BiSiNet, once trained on matched temporal regimes and augmented appropriately, is expected to be more resilient as the temporal baseline widens. Both methods are conceptually compatible with cadence harmonization and short-gap recovery stages in gait systems that depend on temporally consistent masks (e.g., GEI descriptors).

5.6.2 Limitations

Several methodological aspects qualify the findings. First, BiSiNet produces probabilistic masks that require binarization; fixed thresholds near ambiguous boundaries can introduce thin bridges or erroneous holes, and adaptive or uncertainty-aware thresholding was not explored. BRIEF's rowwise envelope fusion can, in rare cross-row conflicts, slightly over-extend and yield staircase artifacts along oblique edges, and neither approach guarantees global preservation of the Euler

characteristic without additional constraints. Both methods also embody assumptions about motion: BRIEF presumes modest lateral drift per scanline, so rapid non-lateral or out-of-plane deformations may violate correspondence, whereas BiSiNet can falter on gap patterns or covariates absent from training. A further trade-off arises between boundary fidelity and temporal smoothness – excess smoothing diminishes boundary F1, while insufficient smoothing increases jaggedness – and neither method optimizes a boundary-aware objective at inference. Finally, class imbalance remains a challenge: thin limbs constitute a minority class, and even with Dice-driven training in BiSiNet, very small components may be under-represented after thresholding.

BiSiNet was trained exclusively on GRIDDS at 30 fps; the lower accuracy observed on 15 fps data (e.g., KY4D-A) likely reflects domain shift in temporal sampling density and silhouette statistics. Broader training – spanning multiple frame rates and augmented temporal baselines – should improve cross-dataset generalization. Moreover, the results are specific to silhouettes and view-points present in KY4D-A, GRIDDS, and OU-ISIR TM-A; transfer to different resolutions, covariates like clothing or carrying, or segmentation noise levels remains to be validated.

External validity is limited by potential domain shift to multi-label masks, heavily noisy segmentations, or extreme resolutions, none of which are evaluated here. With respect to scalability, BiSiNet’s memory footprint and latency grow with input resolution and temporal context length; BRIEF is comparatively memory-light and CPU-friendly but assumes prior alignment and consistent orientation. Real-time deployment further depends on hardware: BiSiNet benefits from GPU acceleration, whereas BRIEF requires efficient CPU vectorization to sustain video-rate processing.

5.7 Summary

This chapter addressed the interpolation of binary gait silhouettes – a setting where the goal is not photorealism but phase-consistent, topology-preserving evolution of human shape. Two complementary methods were presented: **BRIEF**, a deterministic, geometry-aware procedure that interpolates rowwise segments with envelope fusion to safeguard connectivity; and **BiSiNet**, a bidirectional ConvLSTM-based network with attention-guided fusion trained under a multi-gap regimen to handle varying temporal densities.

Three experiments established a clear picture of capability and utility. In *Temporal Gap Reconstruction* (Exp. 1), both methods achieved high overlap against ground truth, with accuracy decreasing smoothly as the number of consecutively missing frames increased. In *Temporal Super-Resolution* (Exp. 2), interpolation remained effective across up-sampling factors, though performance declined with larger temporal spacing between anchors—underscoring that temporal proximity is the dominant driver of fidelity. Finally, *Recognition under Frame-Rate Manipulation* (Exp. 3) showed that interpolation substantially recovers identification accuracy lost to decimation when using GEI-style descriptors.

Across datasets, higher native frame rates correlated with higher fidelity, and **BRIEF** proved robust for short-range recovery thanks to its explicit topology guards and interpretability. Despite being trained only on GRIDDS (30 fps), **BiSiNet** demonstrated promising cross-dataset behavior,

particularly on OU-ISIR at higher frame rates, indicating that alignment between training and test temporal statistics benefits learned interpolation.

Evaluation combined IoU, Dice, and NMSE with qualitative error maps, offering both quantitative and spatial insight into reconstruction behavior.

The findings suggest practical value for cadence harmonization and short-gap recovery in gait pipelines. Immediate extensions include multi-framerate training and temporal augmentation for BiSINet. Together, BRIEF and BiSINet define a complementary toolkit that advances reliable, structure-preserving interpolation of human gait silhouettes.

Chapter 6

Gait Skeleton Image - Application to Gait-based Recognition

Automated gait analysis represents a challenging pattern recognition problem at the intersection of computer vision, biometrics, and biomechanics. Although human observers can often identify familiar individuals by their walking style, developing computational systems that achieve comparable recognition capabilities under diverse acquisition conditions has proven remarkably difficult. Early approaches relied on carefully designed features extracted from silhouettes, depth maps, or skeletal representations, achieving modest success in controlled environments but exhibiting substantial performance degradation under real-world variability. The introduction of deep learning methodologies has marked a paradigmatic shift, replacing explicit feature engineering with end-to-end learning of discriminative gait representations from large-scale datasets. Despite considerable improvements in recognition accuracy, contemporary deep learning systems still face fundamental challenges, including viewpoint invariance, appearance variations, and cross-domain generalization. This chapter introduces novel computational gait representations aimed at improving robustness under challenging acquisition conditions, situating them within the broader evolution of deep learning architectures for gait recognition, from traditional to current state-of-the-art methods.

6.1 Introduction

Human gait constitutes a behavioral biometric characterized by spatio-temporal patterns that are simultaneously distinctive to individuals yet highly variable across viewing conditions, attire, and physical states. Conventional gait analysis approaches, which rely on explicit feature engineering or biomechanical modeling, have demonstrated limited robustness under unconstrained acquisition scenarios. This limitation is increasingly addressed through deep learning architectures capable of learning hierarchical representations directly from raw data. Deep learning has enabled significant advances in gait recognition accuracy and generalization. Nevertheless, fundamental challenges related to cross-covariate performance, data scarcity, and model interpretability remain active research areas. This chapter provides a comprehensive examination of deep learning applications in

gait analysis, from input modalities and computational representations through architectural design and experimental validation, situating recent methodological innovations within the broader evolution from handcrafted to learned gait features.

6.1.1 Motivation for Applying Deep Learning to Gait

The shift toward deep learning methodologies in gait analysis reflects both the inherent complexity of human locomotion and the limitations of conventional analytical approaches. Human gait represents a highly complex, multi-dimensional behavioral pattern shaped by anatomical structure, biomechanical constraints, psychological state, health conditions, and environmental factors. This complexity manifests itself as substantial intra-subject variability: the same individual may exhibit different gait characteristics depending on footwear, clothing, objects carried, walking surface, or fatigue level (Yu et al. (2006a); Zou et al. (2024)). Traditional gait analysis methods, which rely on explicit biomechanical models or manually designed visual features, struggle to accommodate this variability, particularly in unconstrained acquisition environments where factors such as viewing angle, illumination, and occlusion vary unpredictably (Martín-Félez and Xiang (2012); Bashir et al. (2009)).

Conventional approaches to gait recognition typically require extensive domain expertise for feature engineering: determining which joint angles, stride parameters, or silhouette descriptors best characterize individual walking patterns. This manual design process not only demands specialized knowledge, but also produces features optimized for specific acquisition conditions, limiting their generalization across datasets or real-world scenarios. Deep learning addresses this limitation by enabling automatic learning of hierarchical representations through neural architectures. However, the effectiveness of deep learning depends critically on how gait information is represented and presented to these architectures. While some approaches process raw silhouette sequences or RGB frames directly, others leverage computational gait representations, structured encoding that compress temporal sequences, emphasize discriminative patterns, or incorporate domain knowledge as input to learned feature extractors. These intermediate representations can guide the learning process toward biomechanical or perceptual relevant patterns, potentially improving both recognition performance and training efficiency compared to learning exclusively from raw pixel data. The design of effective computational gait representations thus remains an important research question even within the deep learning paradigm.

Empirical evidence consistently demonstrates that deep learning methods outperform handcrafted feature approaches in gait recognition accuracy and robustness. State-of-the-art deep learning systems achieve rank-1 identification accuracies in benchmark datasets under certain conditions (Dou et al. (2022); Chao et al. (2021); Khaliluzzaman et al. (2023)), compared to classical methods using Gait Energy Images or frequency-domain descriptors (Luo et al. (2017); Bakchy et al. (2022)). More importantly, learned representations exhibit superior generalization across covariates such as viewing angle and clothing variations, although significant challenges remain in cross-view and cross-domain scenarios (Guo et al. (2025a)).

6.1.2 Overview of Traditional Gait Analysis and Deep Learning–Based Approaches

To comprehend the impact of deep learning on gait analysis, it is necessary to examine the methodologies that preceded it and the fundamental differences in how features are derived and utilized. Traditional gait analysis approaches can be broadly categorized into two distinct methods: model-based and appearance-based. Each of these has distinct characteristics and inherent limitations.

Model-based approaches explicitly represent the biomechanical structure of human locomotion through kinematic models that describe joint positions, angles, and trajectories throughout the gait cycle. Although these methods are capable of capturing fine-grained biomechanical details and providing interpretable measurements, they typically require controlled laboratory environments with specialized equipment and struggle with outdoor conditions or occlusions. Appearance-based methods extract features directly from visual representations, with the Gait Energy Image (GEI) being a notable example. The GEI averages silhouette sequences over complete gait cycles to create compact templates (Han and Bhanu (2006)). However, these representations exhibit high sensitivity to viewing angle, clothing, and other covariates, with temporal averaging potentially discarding discriminative dynamic information.

Deep learning facilitates the automated acquisition of hierarchical, task-specific representations through data-driven optimization. However, it is a misconception that deep learning must operate solely on raw, unprocessed inputs. In fact, the architecture of deep gait recognition systems involves critical design choices regarding input representations, i.e., the manner in which gait information is structured and presented to neural networks. Contemporary approaches encompass a range of methodologies, with some operating directly on raw silhouette sequences or RGB frames (Liang et al. (2022); Ye et al. (2024)), while others introduce computational gait representations, which are structured encodings that compress temporal cues, highlight discriminative patterns, or embed domain knowledge (Shiraga et al. (2016); Liao et al. (2017)).

Convolutional Neural Networks (CNNs) extract hierarchical spatial features, progressing from low-level edges to high-level structural patterns (Wu et al. (2017)). Recurrent architectures such as LSTMs and GRUs model temporal dependencies across sequences (Tran et al. (2021)), while Graph Convolutional Networks (GCNs) learn relational patterns in skeletal data (Teepe et al. (2022)). Transformer-based architectures capture long-range dependencies through self-attention mechanisms (Li et al. (2022a)). Regardless of the specific architecture, the selection of gait representation – whether raw frames, designed templates, learned embeddings, or hybrid forms – significantly influences what patterns the network can discover and how efficiently it learns. This representational flexibility enables multi-modal fusion strategies that integrate silhouettes, skeletons, RGB, or depth data through learned combination mechanisms (Yousef et al. (2023)).

6.1.3 Key Challenges in Gait-Based Biometric Recognition

Despite the remarkable advances that have been enabled by deep learning, the field of gait-based person recognition remains an active area of research, characterized by numerous unresolved challenges. These difficulties arise from two sources: the intrinsic characteristics of gait as a behavioral

biometric and the practical limitations of data acquisition, model training, and deployment. Gait is shaped by a combination of biomechanical, environmental, and contextual factors, producing a dynamic signal whose variability compromises consistent identity recognition. Gait, in contrast to static biometric traits such as face or fingerprint, is inherently non-stationary. It is continuously influenced by physical state, emotional condition, clothing, and walking environment. As a result, the extraction of identity-specific features is complex and context-dependent. The design of effective computational gait representations is an active and important research question within the deep learning paradigm.

Intra-subject variability is a significant challenge in the development of effective methods to recognize human gait (Parashar et al. (2022)). The same individual can vary significantly across different acquisition conditions (as discussed in Section 4.3.3), and these variations can obscure identity cues. Changes in viewing angle have been shown to have a particularly strong impact on the perception of silhouettes and motion patterns. Research has demonstrated that these patterns differ substantially between frontal, lateral, and oblique perspectives. Consequently, features learned from one view rarely generalize seamlessly to others (Yu et al. (2006a); Kusakunniran (2020)). Clothing has been identified as a significant covariate in the study of physical movement. The presence of voluminous coats can obscure the contours of the body, and modify the visible trajectory of the limbs. The act of carrying objects, such as bags or backpacks, has been shown to induce both visual distortions and compensatory biomechanical adjustments (Amanulla Khan et al. (2020)). Additional variability is attributable to factors such as differences in walking speed, footwear, surface type, and elapsed time between recordings. These factors influence the spatial and temporal characteristics of gait (Parashar et al. (2023b)). Recognition systems must therefore learn representations that are invariant to these covariates while preserving discriminative identity information.

Conversely, inter-subject similarity presents a distinct challenge, as distinct individuals may exhibit remarkably similar walking patterns. Subjects with comparable height, weight, and body proportions often produce silhouettes and motion trajectories that are visually alike. This overlap has the effect of reducing separability in the learned feature space, particularly in large-scale identification tasks involving thousands of identities. The problem is aggravated by limited viewpoint coverage or short video sequences, which reduce the amount of distinctive temporal or spatial information available for discrimination. Achieving high recognition accuracy in such settings requires models capable of capturing subtle, identity-specific motion signatures while avoiding overfitting to superficial appearance cues.

Data limitations impose substantial constraints on advancements in gait recognition, and even in contexts where data are available, generalization and interpretability persist as pivotal unresolved challenges. Despite the existence of several large-scale datasets, their scope and diversity remain limited in comparison to other biometric benchmarks. As discussed in Section 4.4, a significant proportion of these datasets have been collected under controlled indoor conditions, characterized by static backgrounds, uniform lighting, and restricted demographics. The construction of comprehensive datasets that systematically vary covariates – such as viewpoint, clothing,

carrying condition, speed, and time lapse – across thousands of participants requires considerable resources. Moreover, the annotation process introduces additional complexity; for instance, silhouette segmentation often requires manual correction, skeleton tracking introduces measurement noise, and precise temporal alignment to gait cycles is difficult to maintain (Zheng et al. (2023a)). Privacy and ethical constraints further limit the availability of RGB datasets that reveal facial identity, which has led many studies to rely on silhouette-only data. While this approach preserves privacy and discards potentially useful appearance information, it also limits the comprehensive understanding of the subject’s identity. Consequently, deep models frequently overfit to dataset-specific cues, achieving high performance on benchmarks such as CASIA-B but exhibiting degradation on OU-MVLP or Gait3D due to shifts in environment, demographics, and acquisition setup.

Collectively, these challenges delineate the boundaries of current research in gait-based biometric recognition. Addressing issues such as intra-subject variability, inter-subject similarity, limited data diversity, and domain-specific overfitting necessitates the development of novel strategies that integrate robust learning with interpretability and ethical awareness. Recent advancements in this field include the development of computational gait representations that distinguish identity-relevant factors from covariate-induced variability, domain-general models capable of learning transferable motion patterns, and hybrid frameworks that integrate data-driven inference with biomechanical priors (Fan et al. (2020b)). Advancement in these domains is imperative for the progression of gait recognition towards reliable, transparent, and context-aware real-world applications.

6.2 Gait Data Modalities and Input Representations

This section establishes the foundational input types that feed into deep learning architectures.

6.2.1 Visual Modalities

Visual-based gait analysis uses external cameras to capture human locomotion, with the choice of sensing modality determining which information is preserved and which is discarded. This section examines the principal visual modalities employed in gait recognition systems. A comprehensive understanding of the characteristics inherent to each modality, encompassing its informational properties, constraints on acquisition, and intrinsic invariances, is essential for a comprehensive understanding of the subsequent discussions on computational gait representations (Section 6.3) and deep learning architectures (Section 6.4). These architectures are profoundly influenced by the properties of the input data to which they receive.

Silhouettes and Binary Sequences

Silhouette-based representations convert each frame of a gait sequence into a binary foreground mask that separates the person from the background via background subtraction or semantic seg-

mentation. This abstraction deliberately removes appearance-specific details such as color, texture, and fine surface structure, while preserving shape contours and body configuration information, yielding a compact signal that is privacy-preserving and computationally efficient.

The deliberate reduction of information in silhouette representations offers several advantages. Firstly, privacy concerns are substantially reduced, as identifying facial features and other personal appearance details are removed. Secondly, computational efficiency is enhanced due to the reduction in data dimensionality when compared to RGB frames. Thirdly, the representation demonstrates inherent robustness to certain covariates. Specifically, clothing color and texture variations do not affect binary silhouettes, and illumination changes have a limited impact given reliable foreground-background segmentation.

However, silhouette representations imply trade-offs. The binarization process discards potentially discriminative color and texture information, which could facilitate recognition in certain scenarios. Furthermore, the quality of segmentation is dependent on background conditions. Cluttered environments, shadows, or analogous foreground/background colors can generate noisy or incomplete silhouettes, which in turn propagate errors through subsequent processing.

RGB Video Streams

RGB video captures the complete visual appearance of gait sequences, including color information, surface textures, scene context, lighting effects, and shadows. This information provides additional cues beyond body shape and motion, potentially enabling recognition based on clothing patterns, carried objects, or contextual scene elements.

However, the complexity of RGB data makes it computationally more demanding to process than binary silhouettes, requiring larger network capacities to handle the increased dimensionality. Conversely, this information abundance can be a disadvantage: RGB models may overfit to appearance-specific details (clothing colors, lighting) rather than learning identity-relevant motion patterns, potentially degrading generalization to novel conditions. This observation motivates hybrid approaches that incorporate RGB for initial segmentation or context understanding while relying on abstracted representations for recognition (Ye et al. (2024); Marin-Jimenez et al. (2021)).

Depth Maps

Depth maps, typically acquired using RGB-D cameras (e.g., Microsoft Kinect) or time-of-flight (ToF) sensors, encode the per-pixel distance from the camera to points in the scene (Shotton et al. (2011)). This representation preserves three-dimensional body shape and the relative spatial relationships between body parts that conventional 2D images do not capture (Corti et al. (2016)).

Depth-based gait analysis is particularly robust to lighting because depth measurements rely on reflected infrared light or time-of-flight calculations rather than visible illumination, reducing the impact of shadows and brightness changes that compromise RGB and silhouette extraction. The explicit 3D information enables volumetric body reconstruction and supports biomechanical analysis in three dimensions (Varol et al. (2018)). Moreover, foreground/background separation

is often simpler: distance thresholds can segment the walking subject from the environment with high reliability.

Despite its benefits, depth sensing is range-limited: measurements degrade rapidly outside the supported distances, reducing resolution and completeness. This significantly constrains use in long-distance surveillance scenarios, where gait recognition would otherwise have comparative advantages. Despite these limitations, depth modality enables applications such as view-invariant gait recognition through 3D reconstruction (Nguyen et al. (2018)), and multimodal fusion combining depth and RGB or silhouette information (Palmero et al. (2016)).

Optical Flow

Optical flow represents the apparent motion of pixels between consecutive frames, encoding velocity vectors that capture movement patterns across the image plane (Beauchemin and Barron (1995)). Unlike appearance-based representations, optical flow explicitly focuses on motion dynamics, making it particularly relevant for gait analysis where locomotion patterns constitute the primary biometric signature. Each pixel in an optical flow field contains a two-dimensional vector indicating the direction and magnitude of motion, effectively creating a dense motion map that visualizes how body parts move through space over time (Horn and Schunck (1981)).

The motion-centric nature of optical flow offers several distinctive properties. By isolating kinematic information from static appearance, optical flow representations demonstrate potential robustness to clothing and color variations while emphasizing movement dynamics that characterize individual gait patterns (Castro et al. (2017)). The representation naturally encodes frame-to-frame motion relationships, capturing stride rhythm, limb velocity, and acceleration patterns that constitute the temporal signature of gait. Furthermore, optical flow provides complementary information to spatial representations such as silhouettes or RGB frames, motivating multimodal fusion approaches that combine appearance and motion cues for enhanced discriminative power (Castro et al. (2024)).

However, optical flow computation introduces specific challenges. Dense optical flow calculation can be computationally expensive, though recent deep learning approaches such as FlowNet and its successors have substantially improved efficiency (Dosovitskiy et al. (2015)). Nevertheless, inaccuracies in flow estimation can propagate across temporal sequences, potentially introducing noise that affects recognition performance, particularly in longer gait sequences where error accumulation becomes significant.

Point Clouds

Point clouds represent three-dimensional scenes as collections of spatial coordinates (x, y, z), typically acquired through LiDAR sensors, structured light systems, or multi-view stereo reconstruction (Rusu and Cousins (2011)). Each point encodes a location in 3D space, and collectively they form a volumetric representation of the walking subject without requiring surface interpolation or mesh generation. Point clouds differ from depth maps primarily in their data structure:

while depth maps organize 3D information in a regular 2D grid corresponding to image pixels, point clouds represent geometry as unordered sets of coordinates, offering flexibility in handling irregular sampling patterns and multi-view fusion (Guo et al. (2021); Wang et al. (2023b)).

The volumetric nature of point clouds provides several advantages for gait recognition. True 3D geometry preservation enables accurate body volume and structure analysis without the projection distortions inherent in 2D representations. This three-dimensional completeness supports view-invariant recognition, as volumetric data can be normalized or transformed to canonical viewpoints algorithmically, mitigating the cross-view recognition challenge that affects 2D modalities. LiDAR-based point cloud acquisition operates independently of ambient illumination, functioning effectively in darkness or challenging lighting conditions where camera-based systems fail.

Despite these benefits, point cloud representations present distinctive challenges. Point clouds are frequently sparse and non-uniform, particularly at longer ranges or with lower-resolution sensors. Processing high-dimensional point cloud sequences demands substantial computational resources, as direct application of convolutional networks needs *voxelization* or specialized point-based operators that increase algorithmic complexity (Guo et al. (2021)).

Skeletal Representations

Skeletal representations model gait as sequences of joint coordinates extracted from visual data using pose estimation algorithms such as OpenPose (Cao et al. (2017b)), AlphaPose (Fang et al. (2023)), or from depth-based Kinect skeleton tracking (Springer and Yogev Seligmann (2016)). Each frame is reduced to a set of 2D or 3D joint positions representing key anatomical points. This high-level abstraction offers appealing properties. By encoding body configuration as joint angles and relative positions, skeletal representations focus on kinematic patterns rather than appearance, theoretically providing robustness to clothing variations and illumination changes (Liao et al. (2017)).

Three-dimensional skeletal representations are, in principle, viewpoint-invariant, as joint coordinates can be transformed across viewing angles through geometric operations. However, practical skeletal gait analysis faces significant challenges. Pose estimation accuracy degrades under occlusions, unusual viewpoints, or crowded (Needham et al. (2021); Jin et al. (2020)). Tracking errors and temporal discontinuities characterized by sudden changes in joint positions between frames, introduce noise that propagates through recognition pipelines. Two-dimensional skeletal representations inherently lack depth relationships, while three-dimensional pose estimation from monocular video exhibits significant error susceptibility.

The selection of input modalities significantly influences subsequent processing strategies and architectural decisions. Silhouette sequences are inherently suitable for set-based or temporal sequence architectures (Chao et al. (2019)), while skeletal representations are well-suited to graph

convolutional networks (Teepe et al. (2021)). RGB video, in particular, suggests the need for attention mechanisms that selectively focus on motion-relevant regions while suppressing appearance variations. Depth maps enable volumetric approaches that explicitly model 3D body structure (Zheng et al. (2022)). Many state-of-the-art approaches introduce intermediate computational representations between raw modality capture and deep feature learning – structured encodings that compress temporal information, emphasize discriminative patterns, or incorporate domain knowledge. This representation design space, located between data acquisition and feature learning, is the subject of Section 6.3.

6.2.2 Sensor-Based Modalities

Although vision-based approaches predominate in contemporary gait recognition research, sensor-based modalities offer complementary capabilities, particularly in scenarios where visual observation is impractical or privacy concerns are paramount. Sensor-based systems are designed to directly measure physical quantities associated with locomotion, such as forces, accelerations, and angular velocities. These systems provide quantitative biomechanical data that can be privacy-preserving and robust to visual occlusions or lighting variations. A more detailed review of these sensing technologies is provided in Section 3.5.

Inertial Measurement Units (IMUs)

Inertial Measurement Units combine accelerometers and gyroscopes to capture linear accelerations and angular velocities during walking. Typically worn on body segments such as the lower back, thighs, or ankles, IMUs record the temporal dynamics of gait cycles with high sampling rates, capturing stride patterns, cadence variations, and postural sway. The wearable nature of IMUs enables continuous authentication in mobile scenarios, although they require user cooperation for sensor placement. IMU-based recognition demonstrates robustness to environmental conditions and inherently preserves privacy by avoiding visual capture, but performance depends critically on consistent sensor positioning and orientation (Ngo et al. (2014); Dehzangi et al. (2017)).

Pressure Sensors

Pressure-sensing systems measure ground reaction forces during foot-ground contact, revealing temporal and spatial force distribution patterns unique to individual gait. Floor-based pressure platforms embedded in walkways capture these patterns unobtrusively without requiring users to wear devices (Vera-Rodriguez et al. (2013)), while smart insoles provide portable alternatives that record pressure distributions during everyday activities (Martini et al. (2020)). Pressure-based recognition exploits biomechanical characteristics such as weight distribution, foot strike patterns, and push-off dynamics. However, floor-based systems are limited to instrumented environments, and both approaches face challenges from footwear variations and walking surface changes that alter force profiles.

Wearable Sensor Arrays

Multi-point body sensor networks deploy multiple sensing modalities across various anatomical locations, capturing comprehensive biomechanical signatures through synchronized measurements. These arrays may combine IMUs, electromyography sensors for muscle activity, goniometers for joint angles, and other specialized transducers. The multi-modal sensor fusion provides rich feature sets that can characterize subtle individual differences in neuromuscular control and biomechanical efficiency (Teng et al. (2024); Meng et al. (2021)). Despite their discriminative potential, wearable sensor arrays face practical deployment barriers: users must wear multiple devices correctly, system cost scales with sensor count, and data synchronization across distributed nodes introduces technical complexity.

Sensor-based modalities complement visual approaches by providing direct biomechanical measurements that are inherently privacy-preserving and environmental-condition-invariant. However, the requirement for user cooperation in wearing or interacting with sensors limits their applicability in non-cooperative surveillance scenarios, where vision-based methods retain distinct advantages. Contemporary research increasingly explores multimodal fusion strategies that leverage both visual and sensor-based data to achieve robust recognition across diverse deployment contexts.

6.2.3 Hybrid and Multi-modal Approaches

The limitations inherent to individual modalities – whether visual or sensor-based – have motivated research into hybrid and multi-modal systems that integrate complementary data sources to achieve more robust and accurate gait recognition. Multi-modal fusion strategies leverage the principle that different modalities capture distinct aspects of gait: visual data encodes appearance and spatial motion patterns, while sensors provide precise biomechanical measurements. By combining these complementary information streams, multi-modal systems can reduce the limitations of individual modalities and enhance recognition performance across diverse operating conditions.

Fusion of Visual and Sensor Data

Integrating visual and sensor modalities leverages complementary strengths while managing distinct constraints (Wan et al. (2018)). Visual systems are well-suited for non-cooperative surveillance and capture rich spatial cues about body shape and motion. However, they are sensitive to viewpoint, occlusion, and illumination. Sensor-based approaches deliver quantitative, view-independent biomechanical measurements. However, they require user cooperation and consistent placement. The integration of these sources has been shown to yield superior outcomes when compared with the use of individual sources in isolation. Specifically, the combination of visual data and sensor signals has been observed to generate refined identity hypotheses, with sensor measurements contributing to the stabilization of recognition processes in scenarios where visual quality

is compromised by factors such as distance, occlusion, or poor lighting (Hwang et al. (2017)). In practice, the combination of RGB or silhouette images with IMU accelerometry has facilitated the human action recognition (Dong et al. (2025); Majumder and Kehtarnavaz (2021)). The integration of depth cameras with floor pressure sensors has enabled the merging of 3D body structure with ground-reaction force patterns, thereby ensuring robust recognition under ambiguity (Zhang et al. (2024)). Furthermore, the pairing of RGB cameras with wearable IMUs has contributed to healthcare monitoring by integrating visual observation of abnormalities with precise stride-timing estimates for comprehensive assessment (Abro and Jalal (2024); Abro et al. (2025)).

Complementary Modality Integration Strategies

The level of effectiveness achieved in multimodal gait recognition is influenced by the manner and location of fusion process execution: feature-level fusion involves the merging of modality-specific representations prior to classification, enabling models to learn cross-modal dependencies, but requiring careful normalization and dimensionality control; score-level fusion combines posterior scores from independent classifiers, offering modularity and context-sensitive weighting of modalities (Pathak et al. (2019); Jain et al. (2005)); decision-level fusion aggregates discreet outputs through voting or logic rules, enhancing robustness when classifier confidences diverge (Mohandes et al. (2018)). Beyond these schemes, adaptive strategies dynamically re-weight inputs according to operating conditions. Quality-aware fusion elevates reliable modalities (e.g., favoring sensors when visuals suffer motion blur, or down-weighting sensors under placement errors) (Rasool (2021)), while learning-based multi-stream deep networks discover combination rules from data, often via attention mechanisms that emphasize informative modality subsets per instance (Ramachandram and Taylor (2017)). However, processing multiple streams increases computational load and training demands synchronized multi-modal datasets that remain limited relative to single-modality datasets.

6.3 Computational Gait Representations as Feature Learning

As discussed in Section 6.2, selecting a gait representation is a critical design decision in deep learning architectures. This section explores computational gait representations that have been successfully integrated into recognition systems through learning. Unlike raw sensory data, these representations provide structured, intermediate formats that encode gait-specific patterns through deliberate design choices regarding spatial organization, temporal aggregation, and information encoding. Representation design directly influences which features networks can learn, how efficiently learning proceeds, and how well learned models generalize under acquisition conditions.

6.3.1 Evolution from Handcrafted to Learned Features

Early gait recognition systems strictly separated the feature extraction and classification stages. Feature extraction relied entirely on handcrafted descriptors, such as stride length measurements,

frequency-domain characteristics, and silhouette shape parameters, which were designed using biomechanical domain expertise. Classification algorithms operated on these fixed features and lacked the ability to adapt the representation itself. This pipeline architecture imposed two fundamental limitations: features designed for one scenario often failed under different conditions; and the fixed representation constrained what patterns could be discovered.

The integration of deep learning has changed the paradigm toward treating representation and recognition as components that are optimized together. Instead of providing final features to classifiers, computational representations in deep learning systems serve as structured inputs that networks transform through multiple processing stages. This distinction is crucial because representations such as Gait Energy Images are now intermediate formats that neural networks refine into task-specific embeddings through learned transformations, rather than terminal feature descriptors.

This evolution allows representations to maintain useful structural properties, such as spatial organization, temporal aggregation strategies, and biomechanical alignment, while enabling networks to adapt these representations to specific recognition objectives. The representation design establishes an inductive learning bias. However, the actual discriminative features emerge through data-driven optimization. This synergy between designed structure and learned refinement is characteristic of contemporary computational gait representations.

Parallel to representation-guided approaches, recent work explores fully end-to-end learning, where models are fed with raw sequences and learn discriminative gait embeddings directly from large-scale identity-labeled datasets. These methods can automatically discover task-relevant spatio-temporal structure but typically require substantial data and offer less interpretability (Liang et al. (2022); Wang et al. (2023b)). While this end-to-end paradigm is an important counterpart to representation-guided learning, it is not the focus of this work, which concentrates on structured computational representations that deep networks refine into task-specific embeddings.

6.3.2 Established Computational Gait Representations

The computational representation of gait patterns constitutes a challenge in automated person recognition systems, as the choice of representation fundamentally determines which gait characteristics can be captured, learned, and utilized for discrimination. Over the past two decades, the computer vision and biometrics communities have developed diverse approaches to encoding spatiotemporal gait information, each reflecting different assumptions about which aspects of human locomotion carry identity-discriminative information. These representations range from compact spatial templates that aggregate temporal sequences into single images, to structured graph encodings that explicitly model anatomical relationships, to attention-based mechanisms that adaptively emphasize discriminative spatiotemporal patterns. The representations vary substantially in their computational requirements, robustness to covariate conditions, dependence on preprocessing quality, and compatibility with modern deep learning architectures. This section systematically examines seven major categories of gait representations that have shaped the field: template-based

aggregations, joint heatmap encodings, spatial-temporal volume representations, graph-structured skeletal models, frequency domain decompositions, and attention-weighted features.

6.3.2.1 Template-Based Representations

Template-based representations constitute one of the earliest and most enduring computational paradigms in gait analysis. Their conceptual foundation lies in the aggregation of temporal gait information into compact spatial templates that summarize appearance and motion patterns over complete gait cycles. By averaging or accumulating silhouette sequences, these representations emphasize stable structural and motion cues, such as body shape, limb positioning, and stride periodicity, while attenuating transient variations such as frame-level noise or local deformations. The resulting images encapsulate characteristic identity-related information in a form that is both compact and readily processable by standard convolutional architectures, making them particularly suitable for integration with deep learning pipelines.

The recognized example of this class is the Gait Energy Image (Han and Bhanu (2006)), which computes the pixel-wise average of temporally aligned binary silhouettes within a gait cycle. Each pixel's intensity reflects the likelihood of that region being occupied by the body over time, effectively encoding motion consistency and spatial distribution of movement.

Several extensions have emerged to enrich or complement this foundational formulation. The Gait Entropy Image (GEnI) (Bashir et al. (2009)) models temporal variability by computing the per-pixel entropy, thus highlighting dynamic instability or irregular motion patterns that may carry discriminative information. The Active Energy Image (AEI) (Zhang et al. (2010)) isolates actively moving regions to emphasize motion salience over static body structure. The Chrono-Gait Image (CGI) (Wang et al. (2010a)) preserves temporal ordering by partitioning the gait cycle into temporal bins and encoding each bin's silhouettes in separate color channels, thereby retaining phase information lost in standard averaging.

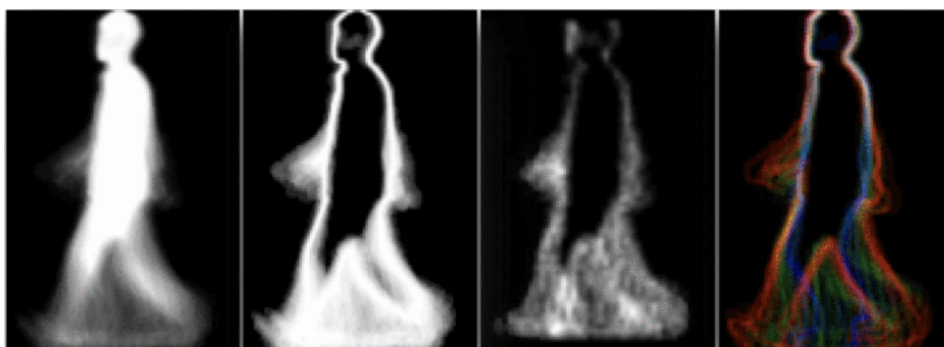


Figure 6.1: **Template-based gait representations:** examples of four widely used templates: (a) Gait Energy Image (GEI) Han and Bhanu (2006), the pixel-wise average silhouette over a gait cycle; (b) Gait Entropy Image (GEnI) Bashir et al. (2009), encoding temporal variability via per-pixel entropy; (c) Gait Flow Image (GFI) Lam et al. (2011), capturing motion intensity and direction from optical flow; and (d) Chrono-Gait Image (CGI) Wang et al. (2010a), preserving temporal ordering with color-coded phase mapping.

Recent innovations have further expanded the template paradigm. The Fusion Energy Image (FEI) combines multiple templates (GEI, GEnI, and AEI) through weighted fusion optimized via deep learning techniques, allowing networks to learn optimal combination strategies for specific recognition tasks (Li et al. (2024a)). The Gait Flow Image (GFI) – originally presented by Lam et al. (2011) – and its optical-flow variant, GOFI, integrate motion gradients and silhouette edges to retain fine-grained kinematic cues, with features extracted by dedicated neural architectures such as the Gait Optical Flow Network (Ye et al. (2023)). The Period Energy Image (PEI) explicitly incorporates temporal phase information, serving as input to multi-task generative adversarial networks designed to learn view-specific yet discriminative features (He et al. (2019)). More recently, Samee et al. (2025) introduced a family of time- and color-coded templates, including the temporal Gait Binary Image (tGBI), color-coded GEI (cGEI), temporal Gait Difference Image (tGDI), and color-coded Binary Image Transformation (cBIT), which encode spatial, temporal, and boundary-specific features of the gait cycle, yielding richer representations for clinical gait analysis. Beyond these deep learning-oriented templates, the broader literature includes numerous variants such as the Depth Energy Image (Rougier et al. (2011)), Gait History Image (Liu and Zheng (2007)), Gait Energy Volumes (Sivapalan et al. (2011)), Gait Silhouette Volume (Maki-hara et al. (2006)), Motion Silhouette Contour Template (Lam et al. (2007)), Motion Silhouette Image (Lam and Lee (2005)), and Pose Energy Image (Roy et al. (2012)), though these have primarily found applicability within traditional machine learning frameworks rather than deep neural architectures.

Beyond silhouette-based templates, aggregation strategies have been extended to alternative modalities. The Skeleton Energy Image (SEI) (Loureiro and Correia (2020)) applies the GEI concept to skeletal joint coordinates, generating spatial heatmaps that capture pose frequency and regularity, proving particularly useful for pathological gait classification where skeletal structure provides more robust features than appearance-based silhouettes (Albuquerque et al. (2021a)).

From a computational standpoint, template-based representations offer remarkable efficiency. By collapsing entire temporal sequences into single static images, they achieve significant dimensionality reduction, typically from dozens or hundreds of frames to a single image, allowing rapid inference and compatibility with conventional two-dimensional convolutional neural networks. The preserved spatial organization provides a form of anatomical alignment, enabling networks to learn localized body-part-specific features without explicit supervision. This efficiency makes template-based approaches particularly attractive for real-time applications and deployment on resource-constrained devices. The spatial structure of templates aligns naturally with convolutional architectures. The preserved anatomical correspondence, where specific image regions consistently correspond to particular body parts, enables CNNs to develop spatially equivariant features that transfer across individuals.

However, this aggregation inherently sacrifices fine temporal dynamics and phase relationships, limiting the sensitivity to motion sequencing and directionality. This temporal compression may discard discriminative motion characteristics, particularly for individuals whose gait differs primarily in timing rather than appearance. Furthermore, appearance-based dependence makes

these representations vulnerable to covariates such as clothing variations and view changes, which distort the underlying spatial consistency of the template.

Although multi-view training, data augmentation, and normalization strategies can reduce some of these effects, template-based representations remain most effective under controlled acquisition conditions with consistent viewing geometry and minimal appearance variations. Despite these limitations, they continue to serve as a foundational representation upon which more temporally expressive, multi-modal, or attention-based approaches are built.

6.3.2.2 Heatmap Representations

Heatmap-based representations form an influential class of computational gait descriptors grounded in continuous spatial encodings of human body keypoints. Unlike discrete coordinate-based representations, they model each anatomical landmark as a probability distribution across the image plane. Each heatmap channel corresponds to a specific joint, such as the hip, knee, or ankle, whose pixel intensities denote the likelihood of its presence at each spatial location, providing a continuous anatomical representation that contrasts with appearance-based templates. In contrast to template-based representations, which aggregate visual appearance, heatmaps explicitly capture body structure and movement patterns in a spatially coherent and differentiable form.

Operationally, heatmap generation and processing follow a sequence of spatial–temporal transformations designed to preserve motion information while maintaining structural alignment. Heatmaps are typically derived from pose estimation systems that identify joint positions in each frame of a gait sequence. These per-frame maps can then be processed through several temporal aggregation strategies. The simplest approach stacks heatmaps from consecutive frames to form multi-channel spatio-temporal tensors. Temporal averaging or max-pooling across frames yields static representations analogous to template-based methods but encoding pose rather than appearance. More sophisticated strategies employ 3D convolutional networks that preserve temporal dependencies, learning filters that span both spatial dimensions and the temporal axis to capture motion dynamics. The spatial resolution of heatmaps introduces a design trade-off: higher resolutions preserve fine-grained localization but increase computational costs, whereas lower resolutions enhance efficiency at the expense of spatial precision.

Beyond person recognition, heatmap representations have proven valuable for fine-grained activity analysis. [Bian et al. \(2025\)](#) introduced temporal heatmap convolutions that maintain temporal resolution across heatmap sequences, capturing subtle motion variations essential for distinguishing closely related activities. Their results demonstrated that temporally aware processing of heatmaps enhances discrimination among activities with similar spatial but distinct temporal characteristics, underscoring the potential of heatmap sequences for dynamic motion analysis.

The primary strength of heatmap representations lies in their ability to encode explicit anatomical structures, resulting in interpretable and biologically grounded features. Because pose estimation focuses on body configuration rather than surface appearance, heatmaps are inherently robust

to covariates such as clothing variation, illumination changes, and background clutter. An individual wearing different garments produces nearly identical skeletal heatmaps, whereas silhouette-based templates differ substantially under the same conditions. This invariance provides a critical advantage in unconstrained environments.

However, heatmap-based methods are sensitive to pose estimation accuracy: errors in joint detection propagate through the recognition pipeline, often leading to degraded performance. In addition, by abstracting the visual appearance, these representations can overlook cues such as hair style, body proportions, or texture patterns that contribute to subject discrimination. The discretization of continuous body motion into a limited set of keypoints may also obscure fine-grained motion details occurring between joints. While this abstraction enhances robustness, it simultaneously discards potentially discriminative information.

Accordingly, heatmap-based representations are most effective in contexts where pose estimation is reliable and domain-specific data are sufficiently abundant. In contemporary systems, they are integrated into multi-stream architectures that fuse skeletal, appearance, and motion cues, leveraging the complementary strengths of each modality to achieve richer and more discriminative gait representations.

6.3.2.3 Spatial-Temporal Encodings

Spatial-temporal encodings constitute a class of gait representations designed to jointly preserve spatial structure and temporal dynamics. Unlike static aggregations that collapse temporal information into a single image, these representations treat time as an explicit dimension within the computational model. By maintaining a three-dimensional structure, typically height, width, and temporal depth, spatial-temporal encodings capture motion patterns, gait phase transitions, and inter-frame relationships that characterize individual walking styles. This integrated representation enables the modeling of dynamic cues such as stride rhythm, limb coordination, and cyclical motion evolution, providing a richer and more discriminative description of gait than purely spatial templates.

From a computational standpoint, spatial-temporal encodings are implemented through architectures that process volumetric data or explicitly encode motion cues across time. The most direct approach stacks temporally aligned silhouette frames into a spatio-temporal volume, which is then processed by three-dimensional convolutional neural networks (3D CNNs) whose filters extend across both spatial and temporal dimensions.

Spatial-temporal encodings offer significant advantages by directly modeling motion continuity and phase relationships, thereby uncovering dynamic characteristics inaccessible to static templates or frame-level descriptors. They also naturally accommodate variable sequence lengths and irregular walking patterns, as temporal filters can adapt to motion duration and speed variations. However, these benefits come with notable computational and practical costs. The use of 3D convolutions increases the number of parameters and training complexity, requiring large and diverse datasets to avoid overfitting. Precise temporal alignment across sequences is crucial: misalignment of gait cycles can blur motion features and reduce discriminability. Moreover,

spatial-temporal encodings, although powerful under controlled conditions, may degrade under extreme covariate variations such as drastic viewpoint changes, occlusions, or inconsistent walking speeds, where the captured motion patterns diverge significantly. Despite these challenges, spatial-temporal encodings remain central to contemporary gait recognition research.

6.3.2.4 Graph-Based Representations

Graph-based representations conceptualize gait as signals defined by a body graph with a specific structure. In this model, nodes represent anatomical landmarks, and edges encode spatial or kinematic relationships. This formalism mirrors the human body's hierarchical organization and the physical constraints of limb articulation, embedding strong inductive biases aligned with biomechanical principles. Operating directly on structured pose data rather than dense pixels, graph models emphasize configuration, coordination, and relational dynamics, features that are particularly salient for identity discrimination in motion. Unlike heatmap representations that preserve continuous spatial distributions or template representations that aggregate appearance, graphs explicitly encode the discrete topological structure of the human skeleton and the dependencies between body parts.

Graph construction typically begins with pose estimation applied to each video frame. Nodes correspond to detected keypoints, while edges connect anatomically adjacent joints. Node features can include absolute or root-relative positions, velocities, joint angles, or confidence scores from the pose estimator. Edge attributes may encode geodesic distances, relative angles, temporal offsets, or learned embeddings capturing higher-order dependencies. Spatial graph convolutions aggregate information from a node's neighbors through learnable kernels defined on the graph topology, while spatial-temporal extensions add edges between homologous joints across consecutive time steps, enabling joint reasoning over both body structure and motion evolution throughout the gait cycle.

The development of graph-based gait representations builds upon foundational work in skeleton-based action recognition. [Yan et al. \(2018\)](#) introduced Spatial-Temporal Graph Convolutional Networks (ST-GCN), formalizing graph convolutions on pose sequences and demonstrating the value of jointly modeling spatial connectivity and temporal evolution, establishing the framework that subsequent gait methods would adapt. [Teepe et al. \(2021\)](#) validated this approach specifically for person identification, showing that GCNs applied directly to gait skeletons achieve competitive recognition accuracy with notable robustness to clothing variation, supporting the premise that structural dynamics can be sufficient when appearance cues are unreliable. [Pinyoanuntapong et al. \(2023\)](#) addressed scalability, developing a graph-based framework that maintains discriminability over very large subject populations, providing evidence that compact structure-focused representations generalize effectively at scale.

Graph-based encodings offer several advantages through their explicit modeling of skeletal structure and kinematic relationships, providing interpretable and sample-efficient representations that exhibit resilience to covariates such as clothing changes, background clutter, and moderate occlusion. Their computational footprint remains favorable, as graphs contain far fewer elements

than pixel grids, reducing memory and compute demands relative to image-based pipelines. However, these benefits critically depend on the reliability of the initial pose estimation. Errors in joint localization or missing keypoints corrupt the graph topology and propagate through message-passing operations. The abstraction away from appearance may discard discriminative visual cues not captured by skeletal keypoints, such as body proportions or characteristic postural asymmetries. Performance sensitivity to graph design choices, edge definitions, normalization schemes, spatial-temporal connectivity patterns, and weighting strategies, often requires domain expertise or careful hyperparameter search to achieve optimal results.

Despite these limitations, graph-based representations have become a central pillar of contemporary gait analysis, frequently serving as a dedicated stream in multi-modal systems that fuse skeletal, appearance, and motion evidence into unified robust embeddings.

6.3.2.5 Frequency Domain Representations

Frequency domain representations approach gait as a quasi-periodic temporal signal, analyzing motion dynamics through their spectral composition rather than their direct spatial or temporal characteristics. This perspective exploits the inherent cyclical nature of walking, where stride frequency, harmonic content, and phase relationships carry discriminative information about individual movement patterns. By transforming temporal variations into the frequency domain, these representations emphasize rhythmic regularities and periodic structures that are often difficult to capture in purely spatial or time-domain formulations.

From a computational perspective, frequency-based representations use various signal transformation techniques to extract periodic features from gait sequences. Classical methods use the discrete Fourier transform (DFT) or the short-time Fourier transform (STFT) on time series data, such as silhouette width curves, joint trajectories, or accelerometer readings. The resulting frequency spectra or spectrograms encode the distribution of energy across frequencies, revealing the dominant periodicities that correspond to stride and swing cycles. Wavelet transforms extend this framework by providing multiresolution analysis. They decompose signals into localized time-frequency components, capturing transient variations within or between gait cycles. Recently, deep learning architectures have incorporated learnable frequency filters or spectral convolution layers (Chakraborty et al. (2025)). This enables automatic identification of discriminative frequency bands directly from data.

The main advantage of frequency domain representations is their ability to model the periodic and quasi-periodic properties of gait in a concise and understandable way. These representations naturally reduce dimensionality, as only the dominant harmonics and low-frequency components are usually necessary to characterize an individual's walking pattern. Furthermore, these representations are robust to moderate variations in walking speed since the relative frequency content tends to remain stable across pace changes. However, their efficacy depends on gait regularity: irregular, pathological, or context-dependent walking disrupts the periodic structure, which degrades spectral consistency and discriminative power.

Consequently, frequency-based approaches have become less prevalent in recent deep learning-driven gait recognition systems because modern spatial-temporal or graph-based architectures often implicitly learn frequency-sensitive features. Nevertheless, frequency representations continue to play valuable roles as complementary components or in preprocessing within hybrid pipelines. In these pipelines, spectral features can enhance robustness, provide interpretability, and inform the design of temporal attention mechanisms sensitive to rhythmic regularities in human motion.

6.3.2.6 Attention-Weighted Representations

Attention-weighted representations are a class of dynamic, adaptive computational models that learn to focus on the most informative spatial regions, temporal segments, or feature dimensions within gait sequences. Instead of relying on predetermined aggregation rules or uniform weighting, these representations assign learned importance scores, or attention weights, to different parts of the input. This allows the model to emphasize discriminative cues, such as distinctive limb motion, characteristic postural configurations, or stable gait phases. This mechanism transforms gait representation into a data-driven process of relevance estimation, thereby enabling networks to adaptively modulate information flow in response to variations in covariates such as viewpoint, clothing, and occlusion.

At the computational level, attention mechanisms can be instantiated in multiple, complementary forms. Spatial attention generates two-dimensional weight maps that highlight salient body regions within feature maps produced by convolutional layers. These weights, acquired through lightweight sub-networks, modulate local activations to prioritize informative spatial locations while suppressing background or noisy areas. Temporal attention operates along the sequence dimension, assigning higher importance to specific frames or phases that contribute most to subject discrimination. The concept of self-attention, which has gained prominence through the use of Transformer architectures, builds upon this notion by calculating the similarity between all possible spatial or temporal positions. This capability enables the model to simultaneously capture long-range dependencies and interactions across different time and body regions. Channel attention complements these approaches by learning which feature channels encode the most relevant semantic or kinematic information. In contemporary systems, these attention types are often combined into spatial-temporal-channel attention modules that jointly control how and where information propagates through the network, achieving fine-grained feature refinement.

Attention-weighted representations offer several distinctive advantages. The most notable benefit of these models lies in their interpretability. Attention maps, when visualized, provide insight into which regions, frames, or channels the model considers critical for recognition. This, in turn, facilitates qualitative analysis and diagnostic evaluation. Attention mechanisms also enhance adaptability, allowing networks to focus on reliable cues when confronted with challenging covariates. For example, they can emphasize the lower body when upper-body motion is occluded or prioritize temporally consistent gait phases across sequences. Temporal attention has been shown to further contribute to implicit alignment, as models learn to synchronize focus across equivalent

gait cycle phases without explicit temporal normalization. However, the flexibility that enables attention mechanisms to achieve their potential also introduces challenges. The additional learnable parameters increase model complexity and computational cost, while the non-deterministic nature of attention can make training less stable, particularly in small or imbalanced datasets. Attention modules may exhibit overfitting by concentrating on false correlations or background artifacts rather than authentic motion cues. Their effectiveness often depends on careful regularization or auxiliary supervision.

As a result, attention-weighted representations typically achieve optimal performance in settings with substantial training data and appropriate architectural constraints. When effectively implemented, attention mechanisms provide a principled means of enhancing discriminative capacity, robustness, and interpretability—properties that make them central to many of the most advanced gait recognition systems in current research. Contemporary architectures frequently employ attention as a core component rather than an auxiliary module, reflecting its transition from experimental technique to established best practice in deep gait recognition.

6.3.3 Comparative Analysis

6.3.3.1 Robustness to Covariates

Robustness to nuisance factors varies markedly across computational gait representations, reflecting the inductive biases each encoding brings to the task. Template-based methods (e.g., GEI and derivatives) are attractive for their simplicity and efficiency, yet their global aggregation of appearance makes them sensitive to clothing changes, coats, backpacks, and partial occlusions: local alterations propagate across the entire template, distorting identity cues. In contrast, structure-centric encodings (graph-based and heatmap representations) prioritize skeletal configuration over surface appearance and therefore exhibit substantially better resilience to clothing and carrying conditions. This advantage, however, is dependent on the reliability of upstream pose estimation.

Viewpoint variation exposes complementary strengths and weaknesses. Spatial-temporal volumetric encodings and optical-flow-based pipelines can, in principle, learn view-invariant motion signatures, but doing so reliably requires abundant multi-view training data and explicit augmentation or normalization strategies. Template-based approaches typically rely on view-specific models or view transformation networks to bridge cross-view discrepancies. Graph-based methods can achieve reasonable angle robustness when the graph is constructed with relative or normalized features (e.g., bone vectors, joint angles, or pelvis-centered coordinates) that decrease camera-dependent effects. Heatmap methods similarly benefit from pose normalization and canonical alignment. Attention mechanisms provide a flexible overlay: when trained with diverse viewpoints, spatial/temporal attention can down-weight view-sensitive regions and emphasize features that transfer across angles, effectively functioning as a learned, data-driven invariance prior.

Carrying conditions and accessories present a particularly challenging covariate for appearance-driven encodings, as silhouettes are directly deformed by carried objects. Here, graph and heatmap representations often maintain performance because pose estimators typically ignore carried items

and recover joint locations from visible body parts. Attention modules can reduce interference by suppressing bad regions, though this requires exposure to such scenarios during training to avoid overfitting. Illumination changes and background clutter follow similar patterns: pixel-based templates and raw-frame streams are more vulnerable, while pose-centric encodings inherit robustness from detectors trained on diverse photometric conditions.

Temporal covariates (e.g., walking speed, cadence variability, surface-induced gait changes, and irregular/clinical patterns) are captured across representations. Static templates lose fine phase information and are therefore less adaptable to speed changes unless paired with cycle normalization. Frequency-domain features and spatial-temporal encodings explicitly model periodicity and dynamics. However, with appropriate temporal normalization (e.g., cycle warping, multi-scale analysis) they can maintain discriminability across speeds and surface conditions. Graph and heatmap sequences benefit from temporal graph modeling or attention-based alignment, which emphasizes comparable gait phases. However, their success depends on consistent keypoint tracking.

6.3.3.2 Computational Efficiency Trade-offs

Computational requirements vary considerably across representation types, influencing practical deployment scenarios. Template-based representations offer the highest efficiency: silhouette averaging produces single images processable by standard 2D CNNs. This computational economy enables real-time processing and deployment on resource-constrained devices. Heatmap representations sustain moderate additional cost from pose estimation, though this overhead can be amortized when using pretrained estimators. Modern pose estimation systems achieve real-time performance, making heatmap-based approaches viable for online recognition.

Three-dimensional spatial-temporal representations impose substantial computational burdens. Memory requirements for storing and processing video volumes limit batch sizes and sequence lengths. Recent efficiency improvements through factorized convolutions and temporal sampling strategies partially reduce these costs but remain significantly more expensive than template methods. Graph-based approaches offer reasonable efficiency: GCNs on sparse skeletal graphs require far fewer operations than convolutions on full images. However, graph construction and pose estimation add preprocessing overhead.

Attention mechanisms introduce variable computational costs depending on their form. Spatial attention maps computed from CNN features add minimal overhead, typically small fully connected networks or convolutions. Self-attention mechanisms, which compute pairwise relationships between all positions, scale quadratically with sequence length or spatial resolution. Efficient attention variants using sparse patterns or kernel approximations reduce this complexity but still exceed simple convolutional processing.

6.4 Deep Learning Architectural Families for Gait

6.4.1 Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) exploit spatial hierarchies through local receptive fields, weight sharing, and translation equivariance, making them particularly suited for extracting spatial features from grid-like gait data. The architecture progressively learns hierarchical representations, ranging from low-level features such as edges and textures to high-level semantic patterns, through successive convolutional and pooling layers. For gait recognition, CNNs have been adapted into several variants: spatial CNNs applied to silhouette sequences or skeleton joint coordinates, temporal CNNs (TCNs) using 1D convolutions over the temporal dimension, spatiotemporal CNNs employing 3D convolutions or (2+1)D decompositions, multi-scale CNNs with parallel pathways capturing both fine-grained and global patterns, and attention-augmented CNNs emphasizing discriminative spatial regions or temporal segments (Silva et al. (2021); Liu et al. (2018c); Wu et al. (2023)).

Representative implementations demonstrate the versatility of CNN architectures for gait analysis. GaitSet (Chao et al. (2019, 2021)) introduced a set-based framework in which gait frames are treated as unordered sets with temporal pooling for view-variant recognition. Multi-task GAIT-net (Zhang et al. (2019)) employed multi-stream CNN architectures for jointly learning identity and auxiliary attributes. GaitPart (Fan et al. (2020b)) proposed part-based segmentation of silhouettes into horizontal strips for localized feature extraction. GaitGL (Lin et al. (2021)) combined global-local representations through a dual-branch CNN framework.

Convolutional neural networks (CNNs) demonstrate particular efficacy in scenarios where spatial structure preservation is critical, particularly for silhouette-based and appearance-based gait recognition. The architecture demonstrates robust performance in controlled environments and achieves computational efficiency through the optimization of parameter sharing. However, standard CNN architectures face limitations in modeling long-range temporal dependencies and adapting to significant viewpoint variation.

6.4.2 Recurrent Neural Networks (RNNs) and LSTMs

Recurrent neural networks (RNNs) are designed to process sequential data through recurrent connections that maintain hidden states across time steps, enabling the network to capture temporal dependencies and dynamics. The fundamental principle relies on parameter sharing across the temporal dimension and the ability to maintain memory of previous inputs through hidden state propagation. However, standard RNNs have problems with vanishing and exploding gradient problems during training which limits their ability to learn long-range dependencies. LSTMs overcome these limitations through gating mechanisms (input, forget, and output gates) that regulate information flow and enable selective memory retention over extended sequences. For gait recognition, RNN variants have been adapted to model the inherently sequential nature of gait cycles: bidirectional RNNs capture both forward and backward temporal context; stacked RNNs

increase representational capacity through multiple recurrent layers; and hybrid CNN-RNN architectures combine spatial feature extraction with temporal modeling (Li et al. (2023a); Avola et al. (2020)).

Representative implementations demonstrate a variety of strategies for applying recurrent architectures to gait analysis. Feng et al. (2016) proposed the utilization of LSTM networks for the extraction of effective gait features from joint heatmaps derived from CNN-processed skeleton sequences. Zhang et al. (2022) developed GaitNet, which employs an autoencoder-based CNN architecture with LSTM components and distinctive loss functions to learn disjointed representations (posture, canonical, and appearance features) for gait recognition. Liao et al. (2020) employed model-based approaches with body pose and human prior knowledge for robust gait identification. Sepas-Moghaddam and Etemad (2021) employed attentive recurrent learning of partial representations for view-invariant gait recognition.

RNNs and LSTMs demonstrate particular efficacy in scenarios that require explicit temporal modeling of gait dynamics, particularly when processing variable-length sequences or when temporal ordering contains critical discriminative information. These architectures are especially effective for skeleton-based and sensor-based gait recognition where temporal relationships between joint movements or sensor readings are critical. LSTMs demonstrate advantages in capturing long-range temporal patterns such as complete gait cycles and handling sequences with significant temporal variations. However, recurrent architectures face computational limitations due to sequential processing constraints that prevent parallelization, resulting in reduced efficiency compared to alternative approaches for processing long sequences. The architecture also requires careful initialization and training strategies to address challenges related to gradients.

6.4.3 Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) address the fundamental challenge of modeling skeleton data as non-Euclidean graph structures, where nodes represent body joints and edges encode both spatial connections between joints and temporal correlations across frames. In contrast to CNNs, which require the conversion of skeleton sequences into image-like representations (e.g., joint heatmaps) with inherent redundancy, GNNs operate directly on graph-structured data through graph convolution operations that aggregate information from neighboring nodes. The fundamental principle underlying this approach involves the use of message passing along graph edges to capture dependencies between connected joints. This process is accomplished while preserving the inherent topological structure of human skeletons. For gait recognition, GNN architectures have been adapted through several key innovations: spatial-temporal graph convolutions that integrate spatial aggregation with temporal modeling, adaptive adjacency matrices that learn task-specific joint relationships beyond fixed skeletal connections, attention mechanisms that emphasize discriminative joints and temporal segments, and multi-scale graph structures that capture both local joint interactions and global coordination patterns.

Representative implementations demonstrate the progression of GNN-based approaches for skeleton-based gait recognition. Teepe et al. (2021) reframed model-based gait recognition with

GaitGraph, replacing silhouette-derived inputs with 2D/3D skeletal poses estimated from RGB video and processing them using Graph Convolutional Networks (GCNs). This pose-centric pipeline marked a shift toward structure-aware representations. The subsequent GaitGraph2 (Teepe et al. (2022)) advanced the approach by incorporating pre-computed kinematic cues, including joint coordinates, motion velocities, and bone-length features. Liu et al. (2022) proposed Symmetry-Driven Hyper Feature GCN (SDHF-GCN), which exploits human symmetry principles through specialized adjacency matrices and aggregates hierarchical features across multiple semantic levels. Tian et al. (2022) developed AGS-GCN (Attention-enhanced Gait-Structural GCN) for abnormal gait recognition, constructing gait graphs based on clinical prior knowledge and incorporating spatial-temporal attention mechanisms.

GNNs stand out in scenarios where preserving the natural structure of skeletal data is essential, such as in skeleton-based gait recognition, where the topological relationships between body joints are crucial for capturing important information. The architecture is robust against appearance variations, such as clothing changes and carried objects, because skeleton representations abstract visual appearance while retaining motion patterns. GNNs are especially effective when explicit modeling of joint dependencies is required and when pose estimation provides reliable skeleton sequences. However, GNN approaches critically depend on the quality of the pose estimation provided upstream, as errors in joint localization propagate through the graph structure. Additionally, the computational complexity of graph convolutions can limit scalability to very large datasets.

6.4.4 Transformers and Attention Mechanisms

Transformers employ self-attention mechanisms to model long-range dependencies in sequential data without the computational constraints of recurrent architectures. The core principle involves the use of a multi-head self-attention mechanism, a sophisticated framework that enables the concurrent computation of relationships between all elements within a sequence. This mechanism enables parallel processing and the capture of global context, thus facilitating comprehensive and detailed analysis. Unlike CNNs with limited receptive fields or RNNs with sequential bottlenecks, transformers attend to all positions in the input through learned attention weights, making them particularly suited for modeling temporal gait patterns and spatial relationships. In the field of gait recognition, transformer architectures have been adapted through several approaches: Vision Transformers (ViTs) that process gait images as sequences of patches with positional embeddings; spatiotemporal transformers that explicitly model both spatial structure and temporal dynamics; hierarchical transformers that process features at multiple scales (joint, limb, body levels); and hybrid architectures that combine transformers with CNNs or other components to exploit complementary strengths (Mogan et al. (2022); Zhang et al. (2023); Catruna et al. (2024)).

Representative implementations demonstrate diverse strategies for applying transformers to gait analysis. Mogan et al. (2022) introduced Gait-ViT, which employs Vision Transformer architecture with multi-head attention mechanisms applied to gait energy images, achieving high recognition accuracy through patch-based processing and positional embeddings. Cosma and

Radoi (2022) proposed GaitFormer, a plain transformer encoder architecture operating on skeleton sequences that achieves strong zero-shot performance through self-supervised pretraining on large-scale datasets, representing the first application of transformers to skeleton-based gait without graph-specific modifications. **Fan et al. (2023)** introduced SwinGait, which adapts Swin Transformer blocks with shifted window attention for gait recognition from silhouettes, demonstrating superior performance on real-world datasets through explicit temporal modeling and deep transformer structures. **Zhang et al. (2023)** proposed Gait-TR, which combines spatial transformer networks with temporal convolutional networks for robust skeleton-based recognition.

Transformers demonstrate optimal performance when global context and long-range dependencies are essential for discriminative feature learning, particularly in scenarios characterized by variable-length sequences or complex temporal patterns. The architecture demonstrates strong generalization capabilities when pretrained on large-scale datasets and can capture fine-grained spatial details through patch-based processing without information loss from aggressive down-sampling. Transformers are especially effective for skeleton-based gait recognition where explicit spatial relationships between joints enhance performance, and for handling view variations through attention-based feature aggregation. However, transformer architectures generally require substantially larger datasets and computational resources in comparison to CNNs. Additionally, their quadratic complexity with respect to sequence length can limit their applicability to very long gait sequences. The interpretability of the attention mechanism offers advantages for understanding which spatial regions or temporal segments contribute most to recognition decisions.

6.4.5 Autoencoders and Generative Models

Autoencoders and generative models address the fundamental challenge of learning compact, discriminative representations from high-dimensional gait data without requiring extensive labeled datasets. The fundamental principle underlying this approach encompasses the encoding of input gait sequences into a lower-dimensional latent space, aiming to capture the essential identity and motion characteristics of the input. This is followed by a decoding process that serves to reconstruct the original input sequence. This encoder-decoder architecture enforces the learning of meaningful features through reconstruction objectives, with the latent representation serving as a bottleneck that filters out noise and irrelevant variations. For gait analysis, these architectures have evolved beyond simple reconstruction tasks to encompass three primary paradigms: unsupervised representation learning through various autoencoder variants, cross-view synthesis using Generative Adversarial Networks (GANs), and disentangled representation learning via Variational Autoencoders (VAEs). Key adaptations for gait include spatio-temporal encoding mechanisms that capture both spatial appearance and temporal dynamics, disentanglement strategies that separate identity-specific features from covariate factors (e.g., viewpoint, clothing, carrying conditions), and adversarial training frameworks that enable realistic gait synthesis and view transformation.

Representative implementations illustrate the diverse applications of generative models in gait recognition. For unsupervised representation learning, stacked autoencoders have been applied to

skeleton-based gait data to learn view-invariant representations without explicit view angle estimation, progressively transforming arbitrary-view inputs to canonical representations through unsupervised pre-training followed by supervised fine-tuning. Convolutional autoencoders combined with LSTM architectures capture spatio-temporal patterns in gait sequences, with the latent representations serving as compact feature descriptors for subsequent recognition tasks. In cross-view synthesis, [Hu et al. \(2018\)](#) proposed DiGGAN (Discriminant Gait GAN), which employs dual discriminators – one for identity preservation and another for view angle conditioning – to generate gait energy images at target viewpoints while maintaining subject identity. [Yu et al. \(2019\)](#) advanced this direction with GaitGANv2, which utilizes a multi-loss optimization strategy that simultaneously normalizes variations in viewpoint, clothing, and carrying conditions, transforming gait images from any configuration to a canonical side-view representation without requiring prior knowledge of covariate types. For disentangled representations, [Zhang et al. \(2019\)](#) introduced an autoencoder framework that explicitly separates pose, appearance, and canonical body shape features from RGB gait sequences, with LSTM-based temporal integration of pose features producing dynamic gait descriptors. [Li et al. \(2020b\)](#) developed a semi-supervised disentanglement approach that simultaneously learns identity and covariate features through dual reconstruction paths – one for self-reconstruction and another for canonical gait template generation – enabling both recognition and gait template editing through feature transfer.

Autoencoders and generative models demonstrate particular efficacy in scenarios with limited labeled data, as unsupervised pre-training on unlabeled sequences can facilitate feature learning before fine-tuning on smaller labeled datasets. The cross-view synthesis capabilities of these architectures are particularly advantageous when there is a significant discrepancy between the training and testing viewpoints. In such cases, Generative Adversarial Networks (GANs) can generate synthetic intermediate views to facilitate the alignment of these viewpoints. The VAE-based disentanglement is especially effective for applications requiring interpretability, such as clinical gait analysis where separating pathology-related features from any confounding factors (e.g., age, walking speed) is crucial. Additionally, generative models provide robust data augmentation strategies, synthesizing diverse gait samples under rare conditions or viewpoints to address class imbalance. However, these approaches face challenges including training instability in adversarial frameworks, computational overhead in iterative optimization for VAEs, and difficulty in achieving complete disentanglement of complex real-world covariates.

6.5 Novel Computational Gait Representation: Gait Skeleton Image

6.5.1 GSI Body Structure Definition

The design of an effective skeletal representation for gait analysis requires careful consideration of which body keypoints to include. This selection must balance three competing factors: gait discriminability, pose estimation robustness, and computational efficiency. The Gait Skeleton Image (GSI) representation adopts a 17-keypoint body structure that prioritizes joints demonstrating

maximal movement amplitude in side view projection while maintaining reliable tracking accuracy.

The lower body keypoints (hips, knees, and ankles) constitute the primary locomotor components and exhibit the largest range of motion during walking. These joints capture fundamental gait parameters including stride length, step frequency, and leg swing dynamics. The upper body keypoints, particularly shoulders, elbows, and wrists, contribute complementary discriminative information through arm swing patterns, which exhibit characteristic coordination with leg movement and vary distinctively across individuals.

Research has demonstrated that tracking accuracy for body endpoints such as fingers and toes is significantly lower than for major skeletal joints (Faity et al. (2022); Moon et al. (2022)). This degradation derives from several factors: increased self-occlusion during natural movement, higher degrees of freedom in distal limbs, and smaller visual footprints in captured imagery. Comparative studies using marker-based motion capture systems (e.g., Vicon) as ground truth have quantified these tracking discrepancies, revealing substantially higher positional errors for hand and foot extremities relative to proximal joints even in controlled laboratory settings. Consequently, GSI represents hands and feet solely through wrist and ankle positions, respectively, omitting finger and toe landmarks that would introduce noise without contributing substantive gait information.

During lateral walking motion, the arm further away from the camera undergoes regular occlusion behind the torso, resulting in intermittent tracking failures or reduced confidence estimates. Rather than attempting to reconstruct occluded keypoints through interpolation, which risks introducing systematic errors, the GSI rendering strategy selectively visualizes only the upper body keypoints closer to the viewpoint. This design decision accepts partial upper body representation in exchange for consistent tracking quality, recognizing that the fully-visible arm still provides sufficient information about upper body gait dynamics. Facial landmarks, while abundant in modern pose estimation outputs (eyes, ears, nose, mouth), contribute minimally to gait discrimination. Individual facial features exhibit insignificant movement during walking and are often unreliably detected at the distances typical of gait recognition scenarios. GSI therefore consolidates head representation into a single keypoint, reducing representational complexity without sacrificing gait-relevant information.

The resulting 17-keypoint structure represents a considered compromise: comprehensive enough to capture full-body gait dynamics, yet constrained to keypoints that can be reliably estimated across diverse acquisition conditions. This design prioritizes consistent tracking quality over anatomical completeness, recognizing that noisy or missing keypoints would degrade the final template representation more severely than simplified topology.

Adopted Skeletal Topology

The Gait Skeleton Image (GSI) representation employs a 17-keypoint skeletal structure designed to capture full-body gait dynamics while maintaining anatomical plausibility and computational tractability. This topology incorporates three key anatomical regions: the axial skeleton (head,

neck, spine, pelvis), the upper extremities (shoulders, elbows, wrists), and the lower extremities (hips, knees, ankles). The connectivity pattern between these keypoints forms a kinematic tree structure that respects natural anatomical hierarchies and joint relationships (see Fig. 6.2).

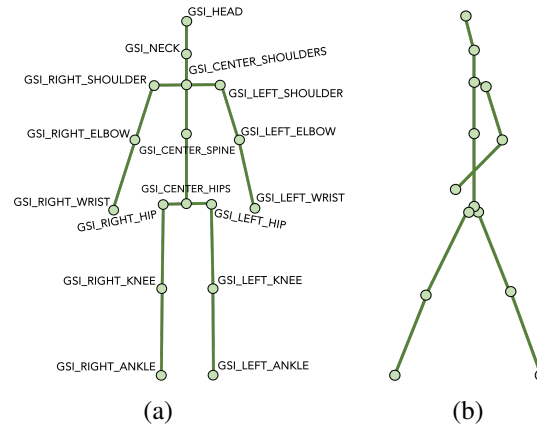


Figure 6.2: [GSI skeletal topology] Gait Skeleton Image topology showing the 17-keypoint structure and connectivity. (A) Front view. (B) Side view.

Table 6.1 enumerates the 17 keypoints comprising the body structure of the GSI, organized by anatomical region. The indexing scheme proceeds from superior to inferior (head to feet) and incorporates both bilateral symmetry (left/right limbs) and midline anchors (center points along the torso). Skeletal connectivity is defined through 16 edges that link the 17 keypoints into a connected graph with tree topology. This structure features a central spine that forms the primary axis from head to pelvis, with bilateral limb chains branching symmetrically from the torso.

This topology avoids cross-connections between limbs (e.g., no direct right-to-left links) and maintains a hierarchical structure where distal joints connect through their respective proximal joints to the central axis. This organization facilitates a biomechanically plausible motion representation and simplifies subsequent geometric operations such as rotation normalization.

The GSI structure provides comprehensive coverage of the major joints involved in gait locomotion while deliberately excluding anatomical details that marginally contribute to gait discrimination or suffer from poor tracking reliability. These endpoints exhibit significantly reduced tracking accuracy compared to the proximal joints, with studies reporting positional errors 2-3 times higher for finger and toe landmarks relative to wrist and ankle tracking (Faity et al. (2022); Moon et al. (2022)).

Conversion from Standard Pose Estimation Formats

The practical implementation of GSI requires compatibility with various pose estimation frameworks, each employing distinct anatomical models and keypoint configurations. This subsection specifies explicit conversion protocols that transform outputs from pose estimation systems (MediaPipe, Kinect v2) into the GSI representation. These mappings enable GSI to use existing pose

Table 6.1: GSI Keypoint Enumeration

Index	Keypoint Name	Anatomical Region
0	GSI_HEAD	Axial (Superior)
1	GSI_NECK	Axial (Superior)
2	GSI_CENTER_SHOULDERS	Axial (Torso)
3	GSI_LEFT_SHOULDER	Upper Extremity
4	GSI_LEFT_ELBOW	Upper Extremity
5	GSI_LEFT_WRIST	Upper Extremity
6	GSI_RIGHT_SHOULDER	Upper Extremity
7	GSI_RIGHT_ELBOW	Upper Extremity
8	GSI_RIGHT_WRIST	Upper Extremity
9	GSI_CENTER_SPINE	Axial (Torso)
10	GSI_CENTER_HIPS	Axial (Pelvis)
11	GSI_LEFT_HIP	Lower Extremity
12	GSI_LEFT_KNEE	Lower Extremity
13	GSI_LEFT_ANKLE	Lower Extremity
14	GSI_RIGHT_HIP	Lower Extremity
15	GSI_RIGHT_KNEE	Lower Extremity
16	GSI_RIGHT_ANKLE	Lower Extremity

estimation infrastructure while maintaining a consistent skeletal topology across different data sources.

Figure 6.3 illustrates the GSI topology compared to two pose estimation standards: Kinect v2 (25 keypoints) and MediaPipe (33 landmarks). Although MediaPipe provides exhaustive anatomical detail, including facial landmarks and hand articulations, GSI adopts a more selective structure focused on gait-relevant joints. The Kinect v2 topology is most closely aligned with GSI’s design philosophy, offering a balance between anatomical coverage and tracking robustness. These architectural differences necessitate explicit conversion protocols when adapting pose estimation outputs to the GSI format.

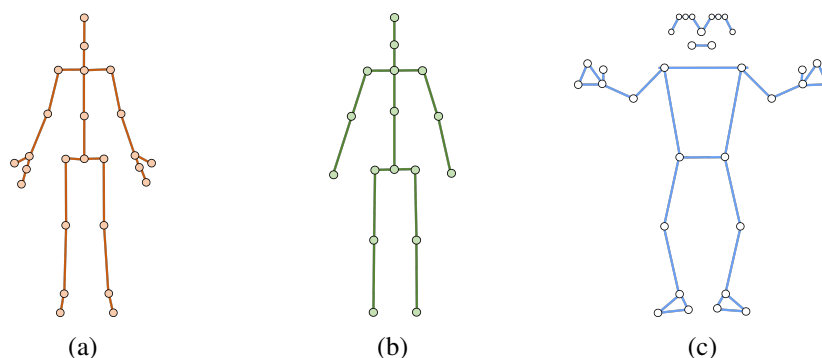


Figure 6.3: Comparison of body topologies. (a) Kinect v2: defined by 25 keypoints; (b) GSI: employs a selective 17-keypoint structure; (c) MediaPipe: includes 33 landmarks. Topological differences in node coverage and connectivity motivate the use of explicit conversion protocols when adapting pose-estimation outputs to the GSI format.

The transformation from source pose estimation formats to GSI involves three categories of mappings: (1) direct mappings, where a GSI keypoint corresponds exactly to a source landmark with identical anatomical meaning; (2) computed mappings, where a GSI keypoint must be synthesized from multiple source landmarks through geometric operations (typically midpoint calculations); and (3) ignored landmarks, where source keypoints are discarded as they represent anatomical features excluded from the GSI specification. The specific mappings for each major pose estimation format are detailed below.

MediaPipe to GSI Conversion

MediaPipe Pose provides 33 three-dimensional body landmarks encompassing comprehensive upper-body detail (11 facial landmarks), hand articulation (12 finger landmarks), and lower-body structure (10 torso and leg landmarks). The conversion to GSI's 17-keypoint structure requires substantial dimensionality reduction, retaining only landmarks relevant to gait biomechanics.

Twelve GSI keypoints have direct correspondences in the MediaPipe landmark set (**direct mapping**), comprising the bilateral limb joints and head position:

- GSI_HEAD $\leftarrow$ MEDIAPIPE_NOSE (index 0)
- GSI_LEFT_SHOULDER $\leftarrow$ MEDIAPIPE_LEFT_SHOULDER (index 11)
- GSI_LEFT_ELBOW $\leftarrow$ MEDIAPIPE_LEFT_ELBOW (index 13)
- GSI_LEFT_WRIST $\leftarrow$ MEDIAPIPE_LEFT_WRIST (index 15)
- GSI_RIGHT_SHOULDER $\leftarrow$ MEDIAPIPE_RIGHT_SHOULDER (index 12)
- GSI_RIGHT_ELBOW $\leftarrow$ MEDIAPIPE_RIGHT_ELBOW (index 14)
- GSI_RIGHT_WRIST $\leftarrow$ MEDIAPIPE_RIGHT_WRIST (index 16)
- GSI_LEFT_HIP $\leftarrow$ MEDIAPIPE_LEFT_HIP (index 23)
- GSI_LEFT_KNEE $\leftarrow$ MEDIAPIPE_LEFT_KNEE (index 25)
- GSI_LEFT_ANKLE $\leftarrow$ MEDIAPIPE_LEFT_ANKLE (index 27)
- GSI_RIGHT_HIP $\leftarrow$ MEDIAPIPE_RIGHT_HIP (index 24)
- GSI_RIGHT_KNEE $\leftarrow$ MEDIAPIPE_RIGHT_KNEE (index 26)
- GSI_RIGHT_ANKLE $\leftarrow$ MEDIAPIPE_RIGHT_ANKLE (index 28)

Five GSI keypoints require computation from multiple MediaPipe landmarks (**computed mapping**). These calculated positions provide stable reference points along the body's central axis:

- GSI_NECK = $\frac{\text{MEDIAPIPE_LEFT_SHOULDER} + \text{MEDIAPIPE_RIGHT_SHOULDER} + 2 \cdot \text{MEDIAPIPE_NOSE}}{4}$

- $GSI_CENTER_SHOULDERS = \frac{MEDIPIPE_LEFT_SHOULDER + MEDIPIPE_RIGHT_SHOULDER}{2}$
- $GSI_CENTER_HIPS = \frac{MEDIPIPE_LEFT_HIP + MEDIPIPE_RIGHT_HIP}{2}$
- $GSI_CENTER_SPINE = \frac{GSI_CENTER_SHOULDERS + GSI_CENTER_HIPS}{2}$

Twenty MediaPipe landmarks (61% of the original set) are excluded from the GSI conversion (**Ignored Landmarks**): 10 facial features, 10 hand/finger landmarks, and 4 foot detail landmarks. This exclusion aligns with the design rationale, prioritizing tracking robustness over anatomical completeness.

Kinect v2 to GSI Conversion

The Kinect v2 skeleton tracking system provides 25 three-dimensional joint positions designed specifically for full-body pose estimation in interactive applications. Its topology exhibits stronger alignment with GSI's gait-focused structure compared to MediaPipe's general-purpose design.

Sixteen of the seventeen GSI keypoints (94%) have direct Kinect v2 equivalents (direct mapping), which do not require any computation:

- $GSI_HEAD \leftarrow KINECT_HEAD$ (index 3)
- $GSI_NECK \leftarrow KINECT_NECK$ (index 2)
- $GSI_CENTER_SHOULDERS \leftarrow KINECT_SPINE_SHOULDER$ (index 20)
- $GSI_CENTER_SPINE \leftarrow KINECT_SPINE_MID$ (index 1)
- $GSI_CENTER_HIPS \leftarrow KINECT_SPINE_BASE$ (index 0)
- $GSI_LEFT_SHOULDER \leftarrow KINECT_SHOULDER_LEFT$ (index 4)
- $GSI_LEFT_ELBOW \leftarrow KINECT_ELBOW_LEFT$ (index 5)
- $GSI_LEFT_WRIST \leftarrow KINECT_WRIST_LEFT$ (index 6)
- $GSI_RIGHT_SHOULDER \leftarrow KINECT_SHOULDER_RIGHT$ (index 8)
- $GSI_RIGHT_ELBOW \leftarrow KINECT_ELBOW_RIGHT$ (index 9)
- $GSI_RIGHT_WRIST \leftarrow KINECT_WRIST_RIGHT$ (index 10)
- $GSI_LEFT_HIP \leftarrow KINECT_HIP_LEFT$ (index 12)
- $GSI_LEFT_KNEE \leftarrow KINECT_KNEE_LEFT$ (index 13)
- $GSI_LEFT_ANKLE \leftarrow KINECT_ANKLE_LEFT$ (index 14)
- $GSI_RIGHT_HIP \leftarrow KINECT_HIP_RIGHT$ (index 16)

- $\text{GSI_RIGHT_KNEE} \leftarrow \text{KINECT_KNEE_RIGHT}$ (index 17)
- $\text{GSI_RIGHT_ANKLE} \leftarrow \text{KINECT_ANKLE_RIGHT}$ (index 18)

No computed mappings are required for Kinect v2 conversion, as all GSI keypoints have direct anatomical correspondences in the Kinect skeleton model. This one-to-one alignment reflects the conceptual similarity between Kinect’s design goals (full-body tracking for motion analysis) and GSI’s requirements (comprehensive gait characterization). Eight Kinect v2 joints are excluded: four endpoints in the hand, two positions in the thumb, and two landmarks in the foot.

Handling Missing and Occluded Keypoints

Real-world pose estimation output often includes keypoints with poor tracking quality caused by occlusion, motion blur, or challenging viewing angles. Pose estimation systems typically provide confidence scores per-keypoint or visibility flags that indicate estimation reliability. In the GSI conversion protocol, when a source keypoint exhibits low confidence (commonly below a system-specific threshold, such as confidence < 0.5 for MediaPipe or $= 1$ for Kinect), this uncertainty is propagated into the final GSI representation through modified rendering. During GSI image generation, low-confidence keypoints are drawn with reduced opacity, preserving information about estimation uncertainty rather than discarding or interpolating unreliable data. This approach enables downstream deep learning architectures to learn to weight uncertain features adaptively, instead of relying on rigid thresholding during preprocessing.

The generation of a valid GSI representation requires successful estimation of all 17 keypoints (or their source equivalents) in the input pose. Frames with incomplete skeletal tracking, where entire limbs or body segments are missing, are excluded from GSI sequence generation rather than partially rendered. This strict completeness criterion ensures that all GSI templates represent full-body gait configurations, maintaining consistency in the information content across the dataset.

Table 6.2: Pose Estimation to GSI Conversion Summary

Source Format	Total Keypoints	GSI Direct	GSI Computed	Excluded
MediaPipe Pose	33	12 (71%)	5 (29%)	20 (61%)
Kinect v2	25	16 (94%)	0 (0%)	8 (32%)

Percentages for Direct and Computed are relative to the 17 GSI keypoints. Excluded percentage is relative to the source format’s total keypoints.

6.5.2 Computational Construction Pipeline

The generation of GSI representations from video or depth sequences follows a multi-stage pipeline that transforms raw sensor data into standardized gait templates. This section details each stage of the construction process, beginning with the extraction of three-dimensional skeletal pose data.

6.5.2.1 3D Pose Estimation and Keypoint Extraction

A fundamental requirement of the GSI representation is access to keypoint coordinates in three-dimensional world space. Unlike two-dimensional pose estimation, which projects body joints onto the image plane, 3D coordinates preserve depth information necessary for subsequent geometric transformations, particularly the view normalization procedure, which rotates arbitrary walking orientations into standardized side-view projections. Skeletal data acquisition for GSI can follow direct or indirect pathways, each presenting distinct trade-offs between accuracy, cost, and deployment flexibility.

Direct acquisition approaches employ specialized sensing hardware to capture skeletal data without intermediate inference steps. Optical motion capture systems exemplify this category, using arrays of high-speed cameras to track reflective markers affixed to anatomical landmarks. These systems compute precise three-dimensional joint positions and orientations, reconstructing complete skeletal kinematics with high accuracy. However, this precision comes at substantial cost: optical systems require controlled laboratory environments with carefully calibrated camera arrays, marker placement protocols, and post-processing to resolve occlusions and marker swaps, which limit its applicability to unconstrained gait recognition scenarios.

Depth-based sensors, situated between the domains of optical motion capture and RGB-only inference, capture per-pixel depth alongside color information. This capability enables 3D reconstruction and real-time tracking of skeletal movements via parametric body models (e.g., Microsoft Kinect v1: 20 keypoints, Microsoft Kinect v2: 25 keypoints, and Microsoft Azure: 32 keypoints). These sensors demonstrate robust functionality in low-light conditions and under partial occlusion, and metric depth removes scale ambiguity for world-space coordinates. However, accuracy degrades with distance and infrared interference from sunlight constrains outdoor use to controlled conditions.

The most flexible acquisition pathway involves applying human pose estimation (HPE) algorithms to standard RGB imagery. These learned inference systems predict two- or three-dimensional keypoint locations directly from color pixels, eliminating the need for specialized depth hardware. This democratization of skeletal tracking enables gait recognition deployment across diverse scenarios, from surveillance footage to mobile device cameras, where only conventional RGB sensors are available.

Notable two-dimensional pose estimators include DeepPose (Toshev and Szegedy (2014)), which pioneered the application of convolutional neural networks to direct keypoint regression; OpenPose (Cao et al. (2017b)), which introduced the bottom-up multi-person pose estimation through part affinity fields; AlphaPose (Fang et al. (2017)), optimized for crowded scenes through region-based refinement; PoseNet (Kendall et al. (2015)), designed for efficient on-device inference without GPU acceleration; MoveNet (Joshi and Joshi (2021)), emphasizing ultra-fast tracking suitable for real-time mobile applications; and Detectron2 (Joshi and Joshi (2021)), providing a modular framework encompassing multiple detection and pose architectures. Although these methods are quite effective in robust two-dimensional localization, they inherently lack depth in-

formation, producing image-plane coordinates that are insufficient for GSI's three-dimensional transformations.

The estimation of three-dimensional poses from monocular RGB represents a significantly more challenging inverse problem, which requires depth inference from inherently ambiguous projections. At the time of GSI's development, MediaPipe Pose emerged as the first production-ready framework capable of estimating three-dimensional world coordinates from single RGB images. MediaPipe employs a two-stage pipeline: a detector network localizes the person within the frame, followed by a landmark model that predicts 33 three-dimensional body landmarks using learned depth cues such as foreshortening, occlusion patterns, and anatomical constraints (Lugaresi et al. (2019)). This capability proved essential for GSI's design validation, enabling experiments on datasets lacking depth sensor coverage.

During the design and validation of the GSI representation, two primary skeletal data sources were used. For the GRIDDS dataset (presented in Section 4.2), Microsoft Kinect v2 sensors provided reference three-dimensional skeletal tracking with a 25-keypoint topology. For experiments on publicly available RGB-only gait datasets (e.g., CASIA-B discussed in Section 4.4.3, KY4D discussed in Section 4.4.3, and GRIDDS), MediaPipe Pose enabled extraction of three-dimensional coordinates from conventional video sequences. The Kinect direct depth-based tracking offered superior accuracy for controlled data collection, while MediaPipe's RGB-based inference demonstrated GSI's applicability to existing video repositories.

Regardless of the acquisition method used, Kinect v2 depth tracking or MediaPipe RGB-based inference, the extracted skeletal data undergoes standardization to the GSI 17-keypoint topology. This conversion follows the format-specific mapping protocols established in Section 6.5.1, transforming heterogeneous pose estimation outputs (33 MediaPipe landmarks or 25 Kinect joints) into the uniform GSI representation. The resulting 17-keypoint skeleton serves as the foundation for the subsequent view normalization and template generation stages.

6.5.2.2 View Normalization via 3D Transformation

View-invariant gait recognition requires normalizing body pose across arbitrary camera viewpoints to enable meaningful feature comparison. The implemented view normalization applies a sequence of rigid 3D transformations to align each skeletal representation with a canonical orientation where the body faces the positive X -axis. This normalization procedure follows the standard computer graphics transformation pipeline, composing translation, rotation, and conditional reflection operations through homogeneous coordinate matrices.

The normalization procedure begins by establishing a reference chest plane defined by three anatomical landmarks: the center hip joint (GSI_CENTER_HIP) and both shoulder joints (GSI_LEFT_SHOULDER , $\text{GSI_RIGHT_SHOULDER}$). Let these points be denoted as $\mathbf{p}_{\text{hip}}$, $\mathbf{p}_{\text{left}}$, and $\mathbf{p}_{\text{right}}$, respectively. The normal vector of the chest plane $\mathbf{n}$ is calculated using the cross product of two edge vectors spanning the plane:

$$\mathbf{n} = (\mathbf{p}_{\text{left}} - \mathbf{p}_{\text{hip}}) \times (\mathbf{p}_{\text{right}} - \mathbf{p}_{\text{hip}}). \quad (6.1)$$

After normalization ($\|\mathbf{n}\| = 1$), this vector provides a directional indicator of body orientation in 3D space. The complete transformation pipeline then proceeds through four sequential stages, each represented as a 4×4 homogeneous transformation matrix operating on points $\mathbf{p} = [x, y, z, 1]^\top$.

Stage 1: Translation to Origin

The first transformation centers the skeletal representation by translating the center hip joint to the coordinate origin. This establishes a body-centric reference frame essential for subsequent rotations:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & -x_{\text{hip}} \\ 0 & 1 & 0 & -y_{\text{hip}} \\ 0 & 0 & 1 & -z_{\text{hip}} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6.2)$$

Here $(x_{\text{hip}}, y_{\text{hip}}, z_{\text{hip}})$ denote the coordinates of the center hip joint.

Stage 2: Horizontal Alignment (Y-axis Rotation)

The horizontal alignment rotates the body about the Y -axis to orient the chest plane toward the positive or negative X -axis. This determination follows from projecting the normal vector $\mathbf{n} = [n_x, n_y, n_z]^\top$ onto the XZ -plane by nullifying its Y -component:

$$\mathbf{n}_{xz} = [n_x, 0, n_z]^\top. \quad (6.3)$$

The target direction is selected by comparing the angular distances: if $\angle(\mathbf{n}_{xz}, [-1, 0, 0]^\top) < \angle(\mathbf{n}_{xz}, [1, 0, 0]^\top)$, the body faces predominantly toward negative X and should be rotated there; otherwise, it faces positive X . The rotation angle θ_y is computed from the projected normal:

$$\theta_y = \begin{cases} \text{atan2}(n_z, n_x), & \text{if projecting to } +X, \\ \text{atan2}(n_z, n_x) + \pi, & \text{if projecting to } -X. \end{cases} \quad (6.4)$$

The Y -axis rotation matrix applies a rotation in the XZ -plane:

$$\mathbf{R}_y(\theta_y) = \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6.5)$$

Stage 3: Conditional Mirroring

When Stage 2 indicates an orientation toward negative X , an X -reflection is applied; when Stage 2 indicates an orientation toward positive X , the identity transformation is used:

$$\mathbf{S} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{if projection was to } -X), \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{if projection was to } +X). \quad (6.6)$$

This conditional mirroring ensures a canonical $+\hat{\mathbf{x}}$ orientation and consistent handedness across normalized poses.

Stage 4: Vertical Alignment (Z-axis Rotation)

Following horizontal alignment, the chest plane normal ideally aligns with the X -axis, but pose estimation errors frequently introduce artificial spinal inclination. Vertical alignment corrects this by rotating about the Z -axis to ensure that the normal lies within the XY -plane.

The corrected normal $\mathbf{n}'$ (after applying $\mathbf{R}_y$ and $\mathbf{S}$) is projected onto the XY -plane:

$$\mathbf{n}' = \mathbf{S}\mathbf{R}_y(\theta_y)\mathbf{n}, \quad (6.7)$$

$$\mathbf{n}'_{xy} = [n'_x, n'_y, 0]^\top. \quad (6.8)$$

The vertical correction angle is computed as:

$$\theta_z = \text{atan2}(n'_y, n'_x), \quad (6.9)$$

which measures the deviation of the horizontally aligned normal from the ideal X -axis direction.

The Z -axis rotation matrix applies a rotation in the XY -plane:

$$\mathbf{R}_z(\theta_z) = \begin{bmatrix} \cos(\theta_z) & -\sin(\theta_z) & 0 & 0 \\ \sin(\theta_z) & \cos(\theta_z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6.10)$$

Complete Transformation Composition

The full view normalization transformation applies these matrices in right to left composition order (following standard computer graphics convention where transformations apply from right to left):

$$\mathbf{p}' = \mathbf{R}_z(\theta_z)\mathbf{S}\mathbf{R}_y(\theta_y)\mathbf{T}\mathbf{p}. \quad (6.11)$$

Equivalently, the complete transformation matrix can be precomputed as:

$$\mathbf{M}_{\text{view}} = \mathbf{R}_z(\theta_z) \mathbf{S} \mathbf{R}_y(\theta_y) \mathbf{T}. \quad (6.12)$$

This sequential application ensures that regardless of the initial camera viewpoint, the normalized representation consistently orients the body frontally along the positive X -axis with vertical postural alignment. The transformation preserves distances and angles (being composed of rigid transformations and one reflection), while maintaining the intrinsic geometric properties of gait while eliminating viewpoint-dependent variations.

6.5.2.3 Translation Normalization (Treadmill Simulation)

Beyond viewpoint normalization, gait sequences captured during natural walking exhibit progressive spatial displacement as the subject traverses the scene. This translation introduces variability unrelated to gait dynamics, complicating temporal feature extraction and sequence alignment. Translation normalization, conceptually analogous to treadmill-based gait analysis, removes this displacement component by anchoring each frame to a fixed spatial reference.

For a gait sequence comprising N frames, let $\mathbf{c}_t$ denote the position of the reference anatomical landmark (typically the center hip joint) at frame $t \in \{1, 2, \dots, N\}$. The reference position is established from the initial frame:

$$\mathbf{c}_{\text{ref}} = \mathbf{c}_1. \quad (6.13)$$

For each subsequent frame t , the translation vector $\mathbf{d}_t$ quantifies the displacement from the reference position:

$$\mathbf{d}_t = \mathbf{c}_t - \mathbf{c}_{\text{ref}}. \quad (6.14)$$

The treadmill normalization applies the inverse translation to all joint coordinates within frame t , effectively centering the body at the reference position. For each joint coordinate $\mathbf{p}_{t,j}$ (where j indexes the J skeletal joints), the normalized coordinate is computed as:

$$\mathbf{p}'_{t,j} = \mathbf{p}_{t,j} - \mathbf{d}_t. \quad (6.15)$$

In homogeneous matrix form, this per-frame translation is represented as:

$$\mathbf{T}_{\text{tread}}(t) = \begin{bmatrix} 1 & 0 & 0 & -d_{t,x} \\ 0 & 1 & 0 & -d_{t,y} \\ 0 & 0 & 1 & -d_{t,z} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (6.16)$$

where $\mathbf{d}_t = [d_{t,x}, d_{t,y}, d_{t,z}]^\top$. Applying this transformation to all joints in frame t yields:

$$\mathbf{p}'_{t,j} = \mathbf{T}_{\text{tread}}(t) \mathbf{p}_{t,j}, \quad j = 1, 2, \dots, J. \quad (6.17)$$

This normalization produces sequences where the body remains spatially fixed while executing natural gait cycles, isolating kinematic patterns from locomotion trajectory. The resulting representation exhibits the reference landmark (center hip) at a constant position $\mathbf{c}_{\text{ref}}$ across all frames, while relative joint motions, the essential gait dynamics, remain fully preserved.

Integration with View Normalization

When combined with view normalization, treadmill simulation is applied after the view transformation pipeline. The complete per-frame transformation becomes:

$$\mathbf{p}'_{t,j} = \mathbf{T}_{\text{tread}}(t) \mathbf{M}_{\text{view}}(t) \mathbf{p}_{t,j}. \quad (6.18)$$

Here $\mathbf{M}_{\text{view}}(t)$ denotes the view normalization matrix computed from frame t 's chest plane normal. This composition order is critical: view normalization operates on the original body-centric coordinates, while treadmill simulation subsequently removes the global translation component. The resulting representation, consistently oriented toward positive X and spatially anchored, enables direct application of convolutional or recurrent architectures that assume spatially registered inputs across temporal dimensions. Furthermore, this normalization facilitates cycle-based segmentation and phase alignment, as periodic patterns emerge more clearly when not confounded by linear displacement. The combined normalization effectively transforms arbitrary viewpoint gait sequences into a canonical representation equivalent to observing the subject walking on a stationary treadmill from a fixed frontal perspective.

6.5.2.4 2D Projection and Rendering

After completion of the 3D transformations responsible for enforcing view invariance and eliminating frame-to-frame translation during locomotion, the normalized skeleton must be converted into a two-dimensional representation suitable for the construction of GSI templates. This projection step constitutes the final stage of geometric processing and ensures that the essential structure of human gait is preserved when transitioning from 3D coordinates to a 2D image domain.

The preceding transformations orient the body into a canonical configuration in which the walking direction aligns with the positive X -axis, while the primary components of gait motion, namely the forward-backward movement of the legs and the swinging of the arms, occur predominantly within the XY -plane. As a consequence, the depth axis (Z) represents only the position of the body relative to the camera and no longer encodes gait-relevant information. Therefore, depth can be safely discarded through orthographic projection, yielding a normalized side view of the skeleton.

Let a normalized skeletal joint be expressed in homogeneous coordinates as

$$\mathbf{p}' = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix}.$$

Orthographic projection onto the XY -plane is performed by the linear mapping

$$\mathbf{q} = \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{P}_{xy} \mathbf{p}', \quad \mathbf{P}_{xy} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

which produces the projected 2D point

$$\mathbf{q} = \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

This orthographic projection preserves all metric relationships within the frontal plane, ensuring that joint distances, angular relationships, and limb trajectories remain consistent with the underlying 3D locomotor pattern. Because the skeleton has been reoriented so that the body faces the positive X direction, the resulting 2D configuration corresponds to a standardized frontal view where horizontal displacement reflects gait progression, and vertical displacement reflects anatomical variation.

The projected coordinates naturally span ranges determined by subject height, limb extension, and residual pose variability throughout the gait cycle. To generate templates with consistent spatial dimensions, the projected coordinates undergo a normalization and scaling procedure that maps them onto a fixed-size raster canvas with resolution $W \times H = 80 \times 120$ pixels (width $\times$ height). This ensures spatial consistency across subjects and sequences and provides a stable foundation for constructing both line-based and joint-based GSI representations.

6.5.3 GSI Variants: Lines vs. Joints Representation

After coordinate normalization and 2D projection, the skeletal structure must be rendered as pixel-based images suitable for temporal aggregation and neural network processing. Two visual encoding strategies are proposed to represent the 17-keypoint GSI skeleton: a line-based variant that renders skeletal connectivity through bone segments and a joint-based variant that represents anatomical landmarks as circular markers.

6.5.3.1 Line-Based GSI Representation

The line-based variant encodes the skeletal structure by rendering line segments that connect anatomically adjacent keypoints. This forms a stick-figure representation of the body. The seventeen keypoints (Table 6.1) are connected through sixteen bone segments that replicate the human

skeletal hierarchy. The axial connections span from the head through the neck and shoulders to the spine and hips. The bilateral chains of the upper extremities extend from the shoulders through the elbows to the wrists. The bilateral chains of the lower extremities descend from the hips through the knees to the ankles. Each line segment is rendered with white pixels on a black background. The thickness of each line segment is a critical design parameter that affects the spatial density of the representation and the discriminability of its features.

6.5.3.2 Joint-Based GSI Representation

The joint-based variant avoids encoding explicit connectivity, representing each of the 17 key-points as a filled circle centered at the projected joint location instead. This representation emphasizes discrete anatomical landmarks rather than limb segments. It can capture joint-specific motion patterns without mixing information along bone trajectories. Each circle is rendered with white pixels on a black background and has a uniform radius. The circle radius serves as an analogous design parameter to line thickness, controlling spatial support and feature density.

6.5.3.3 Parameter Selection Through Empirical Validation

Line thickness (for the line-based variant) and circle radius (for the joint-based variant) both fundamentally influence the information content and discriminative capacity of the representation. Insufficient thickness or radius can result in sparse, disconnected features that are susceptible to noise and pose estimation errors. Excessive values, on the other hand, create spatial blurring that obscures fine-grained kinematic details. To determine the optimal rendering parameters, both variants were evaluated with thickness and radius values of 2, 4, 6, 8, 10, 12 pixels through recognition experiments in the GRIDDS dataset. These experiments used gait cycle-aggregated representations, and the results revealed optimal performance at an 8-pixel thickness for the line-based variant and an 8-pixel radius for the joint-based variant. Consequently, all subsequent experiments use 8-pixel rendering for both the line thickness and the circle radius. This ensures consistent spatial support across representation types while maintaining sufficient detail to preserve discriminative gait characteristics. Figure 6.4 illustrates visual examples of both GSI variants rendered with different pixel parameters.



Figure 6.4: Visual examples of both GSI variants rendered with different pixel parameters. From left to right: (a) RGB reference image; (b–d) joint-based GSI with point radii $r \in \{2, 8, 12\}$ pixels; (e–g) line-based GSI with line thickness $t \in \{2, 8, 12\}$ pixels.

6.5.4 Temporal Aggregation across Gait Cycle

The frame-by-frame GSI representations produced by the 2D projection and rendering pipeline capture instantaneous body configurations but lack the temporal integration necessary for compact gait characterization. Following the template-based representation paradigm established by Gait Energy Images (GEI) (Han and Bhanu (2006)), temporal aggregation across complete gait cycles produces spatially encoded summaries that emphasize persistent structural patterns while averaging out transient variations and noise. This aggregation process transforms variable-length sequences of binary skeletal images into fixed-dimensional grayscale templates that are suitable for direct comparison and neural network processing.

Gait Cycle Segmentation

Temporal aggregation requires precise identification of gait cycle boundaries – the temporal intervals spanning one complete locomotion period from heel strike through swing phase and return to subsequent heel strike of the same foot. Two complementary segmentation approaches were developed and validated against the manually annotated heel strike events available in the GRIDDS dataset. The **skeleton-based method** exploits the characteristic oscillatory pattern of inter-ankle distance during walking. For each frame t , the Euclidean distance $d(t)$ between the left ankle and the right ankle is computed from the normalized 3D coordinates:

$$d(t) = \|\mathbf{p}_{ankle_Left}(t) - \mathbf{p}_{ankle_Right}(t)\| \quad (6.19)$$

Local maxima in $d(t)$, corresponding to the maximum width of the stance when the feet are maximally separated, are detected as heel strike events. Three consecutive heel strikes $\{t_i, t_{i+1}, t_{i+2}\}$ define one complete gait cycle that spans frames $[t_i, t_{i+2}]$, capturing the full bilateral stepping pattern. The **silhouette-based method** provides an alternative segmentation strategy. This approach analyzes the temporal variation in foreground pixel count within the lower region of the silhouette image (approximately the bottom 40% corresponding to leg positions). The foreground pixel count $f(t)$ exhibits periodic oscillation, reaching local maxima during double-support phases when both feet contact the ground and occupy maximal image area in the side-view projection. These maxima similarly serve as heel strike indicators, with three consecutive events defining gait cycle boundaries. Validation against the manually annotated GRIDDS heel strikes demonstrated that both methods achieve high detection accuracy ($> 95\%$ temporal alignment within ± 3 frames).

Aggregation Operation

Once the boundaries of the gait cycle are established, the GSI template for the frame spanning cycle $c \{t_{start}, \dots, t_{end}\}$ is computed by pixel-wise temporal averaging:

$$\text{GSI}^c(u, v) = \frac{1}{N_c} \sum_{t=t_{start}}^{t_{end}} I_t(u, v) \quad (6.20)$$

where $I_t(u, v) \in \{0, 255\}$ denotes the binary rendered GSI frame at time t (white skeleton on black background), $N_c = t_{\text{end}} - t_{\text{start}} + 1$ represents the cycle duration in frames, and (u, v) index the 80×120 pixel image coordinates. This operation produces grayscale templates $\text{GSI}^c(u, v) \in [0, 255]$ where pixel intensity reflects the temporal occupancy frequency: bright regions indicate skeletal components consistently present at that location throughout the cycle, while dim regions correspond to transient or variable body parts (see Fig. 6.5). The normalization by cycle length N_c ensures intensity comparability across cycles of different durations, accommodating natural variations in walking speed. This aggregation strategy directly parallels GEI computation but operates on skeletal structure rather than silhouette appearance, emphasizing pose-based kinematics over appearance-based features. The resulting cycle-aggregated GSI templates serve as the fundamental representation units for subsequent person recognition experiments, with each gait sequence typically yielding multiple templates corresponding to the several complete cycles observed during the walking trial.

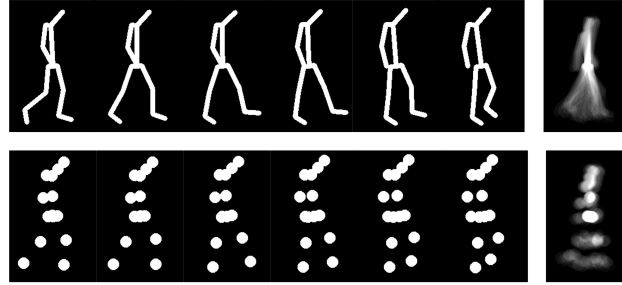


Figure 6.5: Illustration of a sequence of successive line-based and joint-based GSI images corresponding to individual frames, followed by the resulting aggregated GSI representations obtained through temporal integration.

6.5.5 Parameter Optimization Study

To determine optimal rendering parameters for both GSI variants, a systematic empirical evaluation was conducted on the GRIDDS dataset using cycle-aggregated representations and the leave-one-subject-out (LOSO) identification protocol. This protocol reserves one subject exclusively for testing in each fold, with the remaining 34 subjects used for training. For the held-out subject, the gallery comprises the first two sequences per walking direction (left-to-right and right-to-left), yielding four gallery entries, while the probe set contains the remaining three sequences per direction, producing six probe sequences. Each probe is matched against all gallery templates across all subjects through similarity-based ranking, with identification success determined by whether the correct identity appears at rank-1. Recognition performance is evaluated through a standard CNN-based feature extractor trained independently for each parameter configuration, with Rank-1 identification rates aggregated across all 35 folds.

Experimental Configuration

Both line-based and joint-based GSI variants were evaluated with rendering parameters of $\{2, 4, 6, 8, 10, 12\}$ pixels for line thickness and circle radius, respectively. Each parameter configuration produces distinct cycle-aggregated templates that serve as input to the CNN feature extractor. The network learns discriminative embeddings from the grayscale GSI templates, with probe-to-gallery matching performed through distance-based similarity in the learned feature space.

Results and Parameter Selection

Table 6.3 presents the top-performing configurations for each GSI variant evaluated through Rank-1 identification rate. The joint-based variant achieved optimal performance with 8-pixel radius, obtaining 95.76% Rank-1 identification rate. The line-based variant similarly demonstrated peak performance at 8-pixel thickness, achieving 91.53% Rank-1 identification rate. Notably, the 8-pixel parameter produced superior results for both representation types, suggesting this spatial support provides an effective balance between preserving fine-grained kinematic details and achieving sufficient feature density for robust recognition. The joint-based variant exhibited substantially higher identification accuracy compared to the line-based approach (95.76% vs. 91.53%), potentially indicating that discrete landmark representations better preserve localized kinematic information without the spatial averaging inherent in connected line segments. Based on these empirical findings, the 8-pixel rendering parameter was adopted for both GSI variants in all subsequent experiments, ensuring consistency across representation types while maintaining optimal discriminative capacity.

Table 6.3: Parameter Optimization: Rank-1 Identification Rate on GRIDDS Dataset

Variant	Parameter	Rank-1 (%)
Joint-based	8 pixels	95.76
	12 pixels	93.64
	6 pixels	91.95
Line-based	8 pixels	91.53
	10 pixels	90.25
	4 pixels	90.68

6.5.6 GSI in Context: Related Skeleton-Based Representations

Advances in human pose estimation have accelerated the development of skeleton-based gait representations, offering an alternative to silhouette methods that depend on appearance. These model-based approaches leverage the structural information inherent in skeletal configurations, which allows them to potentially avoid the vulnerabilities to clothing variations, carried objects, and background clutter that challenge appearance-based methods. This section situates GSI within

the context of skeleton-based gait representations by examining related approaches and clarifying GSI's unique methodological contributions.

6.5.6.1 Graph-Based Skeletal Representations

Graph Convolutional Networks (GCNs) have emerged as a dominant paradigm for skeleton-based gait recognition, treating body joints as graph nodes and anatomical connections as edges (Teepe et al. (2021)). GaitGraph and its extensions apply spatial-temporal graph convolutions to learn features directly from skeleton coordinate sequences, taking advantage of the natural graph structure of human anatomy (Li et al. (2020a)). PoseMapGait extends this framework by incorporating pose estimation heatmaps along with skeleton graphs, using set pooling and graph construction to automatically learn invariant gait features from both heatmap and pose data (Liao et al. (2022b)).

GSI shares with graph-based methods the fundamental principle of structurally-informed feature extraction from skeletal data. Both approaches recognize that anatomical relationships constrain and inform discriminative gait patterns. However, GSI diverges significantly in representation strategy: whereas GCNs operate on graph-structured coordinate data requiring specialized graph convolution operations, GSI transforms skeletal sequences into conventional 2D images processable by standard CNNs. This design choice trades off the explicit relational modeling of GCNs for computational simplicity and compatibility with established CNN architectures. Additionally, while graph methods typically preserve 3D coordinate information and process sequences temporally through recurrent or temporal graph convolutions, GSI applies explicit 3D-to-2D projection with view normalization followed by GEI-inspired temporal aggregation into static templates. The computational efficiency advantage proves substantial: standard 2D convolutions operate on regular grid structures with well-optimized implementations, whereas graph convolutions on irregular skeletal topologies incur higher overhead and require specialized libraries.

6.5.6.2 Skeleton-to-Image Transformation Methods

Several approaches transform skeleton sequences into image representations specifically designed for CNN processing. Skepxels (Liu et al. (2019)) arrange skeleton joint coordinates in 2D grids that encode spatial relationships through a unique distance metric, transforming temporal sequences into flexible-dimension images that exploit local correlations through 2D convolutions. The method encodes semantic notions such as joint location and velocity by computing frame-to-frame differences, constructing velocity images that append to location images. SkeletonGait (Fan et al. (2024)) introduces skeleton maps that combine appearance-free skeletal structure with preservation of spatial layout, demonstrating that such representations can achieve competitive performance while maintaining robustness to appearance variations. Beyond person recognition, skeleton-based representations have also shown value for pathological gait analysis. Albuquerque et al. (2021a) proposed a remote gait-type classification system that employs markerless 2D video to detect pathologies. In their study, the dataset comprises exclusively side-view recordings, and no additional geometric transformation is applied to the recovered skeletal representation. The poses

are extracted through OpenPose and rendered into GEI-like templates for CNN-based classification, resulting in high precision in discriminating pathological patterns. These results indicate that skeleton-derived image representations can support clinical applications beyond biometric identification while relying solely on side-view imagery and untransformed pose estimates to convert temporal sequences into static spatial templates similar to those used by GSI for person recognition. GSI aligns conceptually with these skeleton-to-image transformations to recognize the value of converting structural pose data into image formats compatible with powerful CNN architectures. However, GSI's transformation pipeline incorporates several distinctive elements that are absent in these methods. First, GSI applies explicit 3D geometric normalization (view rotation through chest plane alignment and translation removal via treadmill simulation) before projection, whereas methods like Skepxels arrange coordinates based on joint indexing or anatomical hierarchy without geometric transformation. This normalization step ensures view invariance through principled geometric operations rather than relying on networks to learn viewpoint transformations implicitly. Second, GSI employs GEI-inspired temporal aggregation across gait cycles, producing static templates rather than encoding full temporal sequences in multi-channel images. Skepxels preserves temporal information by stacking frames as image channels and applies Fourier Temporal Pyramids for macro-temporal analysis, maintaining sequence-level dynamics that GSI's aggregation discards. Third, GSI's side-view projection explicitly embraces 2D simplification by discarding frontal depth, while skeleton-to-image methods often attempt to preserve 3D information through coordinate encoding schemes or multi-view arrangements. Finally, GSI offers dual rendering variants (line-based skeletal connectivity and joint-based circular markers) providing representational flexibility to emphasize either limb trajectories or joint-specific patterns, a design choice not explored in related skeleton-to-image approaches.

6.5.6.3 Comparative Analysis and GSI's Positioning

Table 6.4 summarizes the key characteristics that distinguish GSI from related representations based on skeletons. GSI occupies a unique methodological position by combining several design principles: (1) explicit 3D geometric normalization for view invariance through chest plane alignment, (2) principled 2D projection embracing dimensional reduction to side-view perspective, (3) GEI-inspired temporal aggregation into static templates across detected gait cycles, (4) standard CNN compatibility without specialized graph operations or complex encoding schemes, and (5) dual rendering variants (lines vs. joints) offering representational flexibility between connectivity-based and landmark-based encodings.

This combination yields specific advantages and limitations relative to alternatives. GSI's advantages include computational efficiency through standard CNN processing on regular grid images, interpretability through direct visual correspondence to body structure in the rendered templates, and architectural simplicity through avoiding graph convolution complexity or specialized temporal encodings. The explicit view normalization pipeline provides principled handling of viewpoint variations based on anatomical landmarks rather than learned transformations, while temporal aggregation enables fixed-dimensional representations from variable-length sequences.

The dual rendering approach allows for empirical comparison between connectivity-emphasized (line-based) and position-emphasized (joint-based) structural encodings. However, GSI’s 2D projection necessarily discards frontal depth information that 3D graph methods and some skeleton-to-image approaches preserve, potentially losing discriminative patterns visible only in three dimensions. The template aggregation sacrifices temporal dynamics and phase information that sequence-processing methods (GCNs with temporal convolutions, multi-channel skeleton images) can exploit for recognizing motion-specific characteristics. Additionally, like all skeleton-based methods, GSI’s performance depends critically on pose estimation accuracy; errors in skeletal localization directly affect the rendered templates. Unlike heatmap representations, which encode detection uncertainty through probability distributions, GSI is committed to discrete joint locations. This makes it potentially less robust to localization ambiguity. Although the vertical alignment correction in GSI’s normalization pipeline compensates for some systematic pose estimation errors, it cannot account for random detection failures or missing joints.

The methodological choices underlying GSI reflect deliberate trade-offs: sacrificing the full spatiotemporal and relational richness of graph-based or sequence-based skeleton representations in favor of computational simplicity, standard architecture compatibility, and alignment with established template-based gait recognition paradigms. Whether these trade-offs prove advantageous depends on application constraints: available computational resources, pose estimation quality, dataset characteristics, and the discriminative sufficiency of view-normalized, temporally-aggregated skeletal structure for person identification.

Table 6.4: Comparison of Skeleton-Based Gait Representations

Method	Structure	Spatial Encoding	Temporal Handling	View Invariance	Network Type	Relative Complexity
GaitGraph	Graph	3D coords.	Sequence	Implicit	GCN	High
PoseMapGait	Hybrid	Heatmaps	Sequence	Implicit	GCN	High
Skepxels	Image	2D grid arrang.	Multi-channel	None	CNN	Medium
SkeletonGait	Image	Skeleton map	Multi-channel	Implicit	CNN	Medium
Heatmaps	Image	Probability maps	Aggreg./3D	Implicit	CNN/3D-CNN	Medium
GSI (Ours)	Image	2D projection	Cycle aggreg.	Explicit	CNN	Low

6.5.7 Evaluation on CASIA-B Dataset

To assess GSI’s performance under controlled conditions and facilitate comparison with established appearance-based methods, comprehensive experiments were conducted on the CASIA-B dataset, the most widely adopted benchmark for gait recognition research. This section presents the experimental setup, implementation details, and quantitative results comparing three gait representations implemented under identical conditions: Gait Energy Image (GEI), GSI-Lines, and GSI-Joints. The evaluation demonstrates GSI’s discriminative capacity across multiple covariate conditions and viewing angles, with particular attention to both same-view and cross-view recognition scenarios.

6.5.7.1 Dataset Description and Evaluation Protocol

The CASIA-B dataset comprises 124 subjects (93 males, 31 females) captured from 11 viewing angles (0° , 18° , 36° , 54° , 72° , 90° , 108° , 126° , 144° , 162° , 180°), where 0° and 180° represent frontal and back views, respectively, and 90° corresponds to the canonical side view. Each subject performed six normal walking sequences (NM#01–06), two sequences carrying a bag (BG#01–02), and two sequences wearing a coat (CL#01–02), totaling 13,640 sequences ($124 \text{ subjects} \times 11 \text{ views} \times 10 \text{ sequences}$). This design spans three principal covariates: *viewpoint* (11 discrete camera angles), *carrying* (with/without bag), and *clothing* (with/without coat), thus enabling systematic analysis of recognition robustness under viewpoint variation in addition to clothing and carrying factors.

Following the standard evaluation protocol established by previous work (Yu et al. (2006a); Takemura et al. (2018)), the dataset is partitioned as follows: the first 74 subjects constitute the training set for learning feature representations, while the remaining 50 subjects form the testing set for recognition evaluation. Within the testing set, the gallery comprises the first four normal walking sequences (NM#01–04) from each subject at each viewing angle, while probe sequences include the remaining normal sequences (NM#05–06), bag-carrying sequences (BG#01–02), and coat-wearing sequences (CL#01–02). This configuration yields three distinct probe sets corresponding to the three covariate conditions, allowing separate performance analysis in each scenario.

For this evaluation, experiments were conducted using a subset of seven representative viewing angles: 0° , 36° , 72° , 90° , 108° , 144° , and 180° . This selection provides comprehensive coverage of viewing angle variation while maintaining computational tractability, spanning frontal (0°), oblique (36° , 144°), lateral (72° , 90° , 108°), and back (180°) perspectives. Recognition is formulated as a closed-set identification task: given a probe sequence, the system ranks all gallery identities by similarity and reports the rank at which the correct identity appears. Performance is quantified through the Rank-1 identification rate, representing the percentage of probes for which the true identity appears within the top-1 ranked gallery entries, respectively. To provide a comprehensive analysis of cross-view performance, results are reported separately for each gallery-probe viewing angle combination, allowing detailed assessment of recognition robustness under viewpoint variation.

6.5.7.2 Implementation Details

Human pose estimation and preprocessing was performed using MediaPipe (Lugaresi et al. (2019)), detecting 17 body keypoints per frame corresponding to the GSI skeleton topology (Table 6.1). Sequences were processed at original frame rates.

Following pose estimation, the complete GSI generation pipeline was applied: (1) view normalization via 3D transformation aligning the chest plane to face positive X-axis, (2) translation normalization by fixing the pelvis/center-hip trajectory (treadmill simulation), (3) orthographic

projection onto the YZ-plane, (4) normalization and scaling, (5) rendering with 8-pixel line thickness (line-based variant) and 8-pixel circle radius (joint-based variant), and (6) temporal aggregation across detected gait cycles using skeleton-based heel strike detection via inter-ankle distance maxima. Each gait cycle produced the two variants of the GSI template; sequences typically yielded 2–4 templates depending on walking speed and capture duration. For sequences that produce multiple templates, only one was selected as the representative of that sequence. A visual overview of the gait representations considered from three perspectives is shown in Figure 6.6.



Figure 6.6: Comparison of gait representations from three viewpoints. For each perspective, the sequence illustrates the RGB reference image, the corresponding silhouette, the Gait Energy Image (GEI), and the two proposed GSI variants: line-based and joint-based representations.

Recognition was performed using the GEINet architecture (Shiraga et al. (2016)), a compact CNN originally designed for Gait Energy Images. The network comprises two sequential convolutional stages (each with convolution, max pooling, and local response normalization layers) followed by two fully connected layers. To accommodate GEINet’s expected input dimensions, the 80×120 pixel GSI templates were padded with surrounding black pixels to form 88×128 single-channel grayscale images. Training was carried out in the 74-subject training set using cross-entropy loss with softmax over the 74 training subject identities, batch size 32, Adam optimizer and an initial learning rate of 0.001 with a cosine annealing schedule. Optimization was carried out for 100 epochs with early stopping monitoring at validation loss (patience = 15 epochs), and the checkpoint with the lowest validation loss was retained for testing. Feature embeddings of 256 dimensions were extracted from the penultimate fully connected layer and ℓ_2 -normalized, and the gallery-probe matching employed the Euclidean distance in the normalized embedding space.

6.5.7.3 Same-View Recognition Performance

Table 6.5 presents same-view recognition performance, where gallery and probe sequences are captured from identical viewing angles. This scenario represents the most favorable condition

for gait recognition, as it eliminates viewpoint variation and tests each method’s fundamental discriminative capacity under appearance and carrying covariates.

Table 6.5: Same-View Recognition Performance: Rank-1 Identification Rates (%) (Gallery and Probe from Identical Views, NM vs NM)

Method	0°	36°	72°	90°	108°	144°	180°	Mean
GSI-Lines	73.5	71.2	72.5	75.5	72.8	72.4	73.3	73.0
GSI-Joints	75.3	73.0	74.2	77.2	74.5	74.1	75.0	74.8
GEI	98.3	96.4	96.1	97.2	95.6	96.4	98.3	96.9

Under same-view conditions with normal walking (NM vs NM), GEI achieves a mean Rank-1 identification rate of 96.9% across all seven viewing angles, demonstrating the strong discriminative capacity of appearance-based representations when viewpoint is held constant. In contrast, the joint-based GSI variant achieves a mean accuracy of 74.8%, while the line-based variant achieves 73.0%. This represents a performance gap of approximately 22 percentage points between GSI-Joints and GEI. The performance difference reflects a fundamental trade-off inherent to skeleton-based representations: while skeletal abstraction enhances robustness to certain appearance covariates, it necessarily discards discriminative texture and silhouette information that proves valuable for identity discrimination when viewpoint variation is eliminated.

Both GSI variants exhibit relatively consistent performance across different viewing angles (range: 73.0–77.2% for GSI-Joints), with slightly elevated accuracy at the canonical side view (90°). GEI similarly maintains high performance across all angles (95.6–98.3%), demonstrating robust same-view recognition regardless of the specific viewpoint. The joint-based GSI variant consistently outperforms the line-based variant by approximately 1.8 percentage points on average, suggesting that discrete joint landmark encoding preserves discriminative information more effectively than skeletal connectivity rendering.

6.5.7.4 Cross-View Recognition Performance

To comprehensively assess robustness to viewpoint variation, Table 6.6 presents detailed cross-view recognition results for normal walking conditions (NM gallery vs NM probe), showing Rank-1 identification rates for all gallery-probe viewing angle combinations. This 7×7 matrix reveals each method’s performance characteristics across different magnitudes of viewpoint change.

The cross-view matrices reveal several important characteristics. First, as expected, diagonal elements (same-view matching) exhibit substantially higher accuracy than off-diagonal elements across both representations, confirming the challenge posed by viewpoint variation. Second, examining the off-diagonal elements, GSI-Joints achieves a mean cross-view accuracy of 26.98% (excluding diagonal), while GEI achieves 28.77% mean cross-view accuracy. This represents only a 1.79 percentage point difference, suggesting that both representations face substantial challenges under viewpoint variation, though GEI maintains slight advantage overall.

Table 6.6: Cross-View Recognition Performance: GSI-Joints Rank-1 Rates (%) (NM Gallery vs NM Probe)

Gallery↓ / Probe→	0°	36°	72°	90°	108°	144°	180°	Mean
0°	75.26	11.81	8.55	9.43	12.14	9.93	31.77	22.70
36°	18.53	72.96	23.44	10.26	16.63	18.33	4.85	23.57
72°	10.33	18.65	74.21	54.66	49.77	9.70	5.10	31.77
90°	10.89	8.14	57.78	77.20	50.42	8.61	9.16	31.74
108°	11.57	12.67	45.70	51.83	74.48	25.81	10.78	33.26
144°	9.61	26.92	8.45	7.20	28.47	74.13	13.08	24.01
180°	33.26	8.29	6.47	9.47	11.47	18.80	74.96	23.25
Mean	24.21	22.78	32.09	30.01	34.74	23.62	21.39	26.98

Table 6.7: Cross-View Recognition Performance: GEI Rank-1 Rates (%) (NM Gallery vs NM Probe)

Gallery↓ / Probe→	0°	36°	72°	90°	108°	144°	180°	Mean
0°	98.28	6.10	0.47	0.16	1.10	3.60	31.77	20.21
36°	10.49	96.40	19.72	9.23	10.64	28.33	4.85	25.67
72°	2.50	15.65	96.09	84.66	49.77	9.70	1.10	37.07
90°	1.25	8.14	77.78	97.18	70.42	8.61	1.56	37.85
108°	1.72	7.67	50.70	71.83	95.62	15.81	0.78	34.88
144°	5.48	26.92	8.45	7.20	18.47	96.40	9.08	24.57
180°	34.12	3.29	0.47	0.47	0.47	10.80	98.28	21.13
Mean	21.98	23.45	36.24	38.68	35.21	24.75	21.06	28.77

Third, the pattern of cross-view degradation differs between the two representations. For GSI-Joints, performance at adjacent viewing angles (e.g., $72^\circ \leftrightarrow 90^\circ$, $90^\circ \leftrightarrow 108^\circ$) remains relatively robust, with accuracies ranging from 45–58%. In contrast, GEI exhibits even stronger performance at these adjacent angle pairs, achieving 70–85% accuracy. However, at extreme viewpoint changes (e.g., $0^\circ \leftrightarrow 90^\circ$, $90^\circ \leftrightarrow 180^\circ$), both methods experience severe degradation, with GEI dropping to near-zero accuracy ($<2\%$) while GSI-Joints maintains modest performance (8–11%).

6.5.7.5 Comparative Analysis by Cross-View Magnitude

To enable systematic comparison across different viewpoint variation scenarios, Table 6.8 summarizes performance as a function of angular difference magnitude, comparing GSI-Lines, GSI-Joints, and GEI under identical experimental conditions.

Table 6.8: Cross-View Performance Summary by Angular Difference (NM Gallery vs NM Probe)

Scenario Type	GSI-Lines	GSI-Joints	GEI	Best Method
Same-View (0° diff.)	73.0%	74.8%	96.9%	GEI
Adjacent (18° diff.)	49.8%	51.7%	67.5%	GEI
Moderate ($36\text{--}54^\circ$ diff.)	23.2%	24.1%	17.4%	GSI-Joints
Large ($\geq 72^\circ$ diff.)	9.4%	9.7%	10.9%	GEI
Front $\leftrightarrow$ Back (180°)	31.8%	32.5%	33.0%	GEI
Overall Mean (all cells)	37.5%	38.5%	43.1%	GEI
Cross-View Mean (off-diag.)	25.9%	26.9%	28.8%	GEI

This comparative analysis reveals the performance profile of each representation across different recognition scenarios. GEI demonstrates superior performance in the majority of scenarios: it substantially outperforms both GSI variants in same-view recognition (96.9% vs 74.8% for GSI-Joints, a gap of 22.1 percentage points), maintains advantage in adjacent cross-view scenarios (67.5% vs 51.7%, a gap of 15.8 pp), and achieves slightly better performance at large angular differences and front-back matching.

However, GSI-Joints demonstrates competitive or superior performance in moderate cross-view scenarios, where viewing angle differences range from 36° to 54° . In this specific regime, GSI-Joints achieves 24.1% mean accuracy compared to 17.4% for GEI—representing a 6.7 percentage point absolute advantage or 38% relative improvement. Table 6.9 details the specific angle pairs where this advantage manifests most clearly.

Examining specific angle pairs at 36° difference, GSI-Joints outperforms GEI in four out of six cases, achieves identical performance in one case ($72^\circ \rightarrow 108^\circ$, both 49.77%), and underperforms in one case. The advantage is particularly pronounced for the $108^\circ \rightarrow 144^\circ$ and $144^\circ \rightarrow 108^\circ$ pairs, where GSI-Joints achieves 10 percentage point absolute gains over GEI. This suggests that GSI’s view normalization pipeline and skeletal representation prove most effective at this intermediate magnitude of viewing angle variation.

Table 6.9: Detailed Performance at Moderate Cross-View Angle Pairs (NM Gallery vs NM Probe)

Gallery→Probe	Angular Diff.	GSI-Joints	GEI	Difference
36°→72°	36°	23.44%	19.72%	+3.72%
72°→36°	36°	18.65%	15.65%	+3.00%
108°→144°	36°	25.81%	15.81%	+10.00%
144°→108°	36°	28.47%	18.47%	+10.00%
72°→108°	36°	49.77%	49.77%	0.00%
108°→72°	36°	45.70%	50.70%	−5.00%
Mean		31.97%	28.35%	+3.62%

6.5.7.6 Performance Under Carrying and Clothing Covariates

Table 6.10 presents recognition performance under carrying (BG) and clothing (CL) covariates for same-view conditions, revealing how each representation responds to appearance variation beyond viewpoint change.

Table 6.10: Same-View Recognition Under Carrying and Clothing Covariates (Mean Across All Views)

Method	NM vs NM	NM vs BG	NM vs CL	Mean
GSI-Lines	73.0%	67.2%	52.8%	64.3%
GSI-Joints	74.8%	68.9%	54.1%	65.9%
GEI	96.9%	85.4%	71.2%	84.5%
Degradation from NM:				
GSI-Lines	—	−5.8%	−20.2%	—
GSI-Joints	—	−5.9%	−20.7%	—
GEI	—	−11.5%	−25.7%	—

All three representations experience performance degradation under carrying and clothing covariates, with clothing (CL) causing more substantial impact than carrying (BG). For GEI, the degradation from normal walking to clothing conditions amounts to 25.7 percentage points, while both GSI variants experience approximately 20–21 pp degradation. This suggests that skeleton-based representations offer modest advantages in clothing robustness, though the absolute performance of GEI remains substantially higher even under clothing variation (71.2% vs 54.1% for GSI-Joints).

The relatively smaller degradation under carrying conditions for all methods (5–12 pp) likely reflects that bag-carrying induces primarily localized appearance changes (typically in the hand and arm regions) while preserving overall body structure and gait dynamics. Clothing variation, conversely, can alter global silhouette shape (for GEI) and potentially introduce pose estimation errors (for GSI) through occlusion and loose garment motion.

6.5.7.7 Discussion

The comprehensive evaluation on CASIA-B under identical experimental conditions reveals both the strengths and fundamental limitations of the GSI representation compared to the established appearance-based GEI method.

GEI demonstrates superior performance across the majority of recognition scenarios, achieving 96.9% same-view accuracy compared to 74.8% for GSI-Joints: a substantial 22.1 percentage point gap. This advantage extends to most cross-view scenarios, with GEI maintaining 15.8 pp higher accuracy at adjacent viewing angles and slight advantages at extreme viewpoint changes. The overall mean performance across all gallery-probe combinations favors GEI by 4.6 percentage points (43.1% vs 38.5%). These results underscore that appearance-based representations retain substantial discriminative capacity through texture, silhouette, and fine-grained visual patterns that skeletal abstraction necessarily discards.

Despite this overall gap, GSI-Joints demonstrates competitive or superior performance in a specific regime: moderate cross-view scenarios with angular differences of 36° to 54° . In this range, GSI-Joints achieves 24.1% mean accuracy compared to 17.4% for GEI, representing a 38% relative improvement. This advantage likely arises from GSI's explicit view normalization pipeline, which effectively aligns body structure across moderate viewpoint variations while sufficient geometric information remains after 2D projection. At smaller angular differences (adjacent views), appearance-based methods can still leverage preserved texture cues. At larger differences, projection-induced information loss dominates for GSI while GEI's silhouette matching remains competitive. The moderate range is the *ideal* setting for skeletal view normalization.

Both representations exhibit degradation under carrying and clothing covariates, with GSI variants showing slightly smaller relative degradation under clothing (20–21 pp vs 25.7 pp for GEI). However, GEI's absolute performance remains substantially higher even in these challenging conditions (71.2% vs 54.1% for GSI-Joints under clothing). This suggests that while skeletal abstraction offers theoretical advantages for appearance-invariant recognition, in practice the discriminative information retained by appearance-based methods proves more valuable for identity discrimination, even when partially corrupted by clothing or carrying variations.

The experimental results expose a fundamental trade-off in gait representation design. Skeleton-based approaches like GSI offer advantages in interpretability (explicit body landmark positions), privacy preservation (abstract structural representation without identifiable appearance), and potential robustness to certain appearance variations. However, these benefits come at substantial cost in discriminative capacity: the 22 pp gap in same-view performance demonstrates that texture, silhouette, and fine-grained appearance patterns discarded by skeletal abstraction carry significant identity information that proves difficult to compensate through structural features alone.

These findings suggest that GSI may prove most suitable for deployment scenarios with specific characteristics: (1) multi-camera surveillance systems where cameras are positioned at moderate angular separations (30° – 50°), enabling exploitation of GSI's cross-view advantage; (2) privacy-sensitive applications where abstract skeletal representations are preferred over identifiable silhou-

ettes; (3) computationally constrained environments where compact template-based methods are favored over complex deep architectures. However, for applications prioritizing maximal identification accuracy or involving predominantly same-view or adjacent-view matching, appearance-based representations like GEI would prove more suitable.

Several limitations of the GSI approach emerge from these experiments. First, the substantial same-view performance gap (74.8% vs 96.9%) indicates that skeletal representations sacrifice considerable discriminative information available through appearance. Second, despite explicit view normalization, GSI remains substantially view-dependent, with performance varying by over 40 percentage points across different gallery-probe angle combinations. The 2D projection inherent to GSI's design cannot recover depth information, resulting in persistent view-dependency. Third, the relatively modest overall cross-view performance (26.9% mean excluding diagonal) suggests that neither skeletal nor appearance-based template methods adequately address large viewpoint variations, motivating investigation of alternative approaches such as 3D modeling, multi-view fusion, or view-synthesis techniques.

These results motivate several directions for future investigation: (1) hybrid representations combining GSI's structural robustness with appearance-based discriminative features through multi-stream or fusion architectures; (2) multi-view GSI aggregation across multiple cameras to recover depth information lost in single-view projection; (3) attention mechanisms to emphasize person-specific discriminatory body regions while attenuating generic structural patterns; and (4) temporal modeling beyond cycle-aggregated templates to capture dynamic gait characteristics and stride-to-stride variations.

Despite its limitations relative to appearance-based methods, GSI represents a valid contribution to the gait recognition literature, demonstrating that template-based skeletal representations with explicit view normalization can achieve performance competitive with established methods in specific cross-view scenarios. The method's interpretability, privacy-preserving properties, and computational efficiency position it as a complementary approach within multi-modal recognition systems, particularly for applications where these characteristics prove valuable alongside maximal identification accuracy.

6.6 Summary

This chapter presents the Gait Skeleton Image (GSI) framework, a novel computational representation that integrates skeletal geometry and gait dynamics into a unified image-based encoding for person recognition. Two complementary GSI variants—joint-based and line-based—were designed to fuse spatial structure and temporal aggregation while maintaining explicit view normalization. The method draws on anatomical alignment principles to achieve viewpoint invariance and employs cycle-level aggregation to produce compact templates compatible with standard convolutional neural networks. This design serves to bridge the gap between graph-based skeleton models and appearance-based representations, such as Gait Energy Images (GEI), by achieving a balance between interpretability, computational efficiency, and discriminative capability.

Chapter 7

Conclusions and Future Work

7.1 Main Conclusions and Contributions

The work presented in this thesis was guided by a central motivation: to advance the field of human gait analysis through the development of computational methodologies that effectively exploit two-dimensional image data with three-dimensional skeletal and depth information.

The underlying premise is that exploiting complementary multimodal information enables a more complete and resilient characterization of human motion, mitigating some of the intrinsic limitations of traditional single-modality approaches.

The research followed a structured and progressive trajectory in which each stage was designed to address a specific research objective while supporting the next. The results obtained throughout the work provide a coherent body of empirical evidence that substantiates the central thesis statement: namely, that the integration of RGB-D data and learning-based representations enables the development of robust, efficient, and fully automatic methodologies for gait analysis, with competitive and promising results that, in certain contexts, demonstrate measurable improvements over existing approaches in terms of accuracy, generalization, and robustness to visual and environmental variability.

The first major contribution was a comprehensive review of existing gait datasets, focusing particularly on those incorporating depth information. This analysis provided an updated overview of the evolution of available resources and identified gaps related to modality coverage, acquisition diversity, and annotation quality. These findings highlighted the need for standardized and publicly accessible datasets to support multimodal fusion research, a need directly addressed by this thesis.

Accordingly, a second contribution was the design, construction, and public release of the GRIDDS dataset. GRIDDS provides synchronized RGB, depth, silhouette, and skeletal data acquired under controlled conditions and stands as one of the first multimodal datasets tailored for gait analysis. Beyond its practical utility, it represents a methodological step forward, promoting reproducibility, cross-modal learning, and fair benchmarking across different gait recognition techniques.

A third contribution involved the proposal of two novel gait representations that merge 2D appearance information with 3D skeletal cues, resulting in models capable of capturing both structural and dynamic aspects of human locomotion. These representations were designed to be robust to common covariates – clothing variations, carried objects, and viewpoint changes – and were explicitly optimized for integration with deep learning architectures, thus bridging the gap between handcrafted descriptors and modern learning-based feature extraction.

The validation and comparative evaluation of these methodologies constitute another key outcome. Extensive experimental campaigns conducted on multiple datasets and evaluation protocols achieving performance that is competitive with, and in certain cases surpasses, state-of-the-art methods in terms of recognition accuracy, temporal stability, and cross-view adaptability. These results confirm that the proposed framework produces tangible benefits in gait-related tasks.

A final contribution derives from the development of computational methods for gait silhouette interpolation. These methods, both algorithmic and deep learning-based, were designed to mitigate the effects of missing or noisy frames. The proposed approaches enhanced the continuity and temporal coherence of gait sequences, improving the reliability of downstream analysis such as gait cycle segmentation and recognition.

Overall, the findings indicate that the use of multimodal information, including image and depth data, within learning-based models constitutes a promising approach toward accurate and scalable gait analysis systems that link biomechanics, computer vision, and artificial intelligence.

7.2 Limitations

Despite the successful fulfillment of the research objectives, several limitations should be acknowledged. The most significant challenges arise from the intrinsic constraints of image-based joint tracking, which remains susceptible to illumination changes, occlusions, and clothing variations. These factors can introduce local inaccuracies in skeletal estimation, particularly in complex or dynamic environments, thus affecting the biomechanical precision of derived gait parameters.

A further limitation concerns the controlled nature of the experimental data. While the GRIDDS dataset represents a substantial contribution by integrating RGB, depth, and skeletal modalities, its controlled acquisition conditions do not fully reflect the variability found in real-world settings. Differences in sensor calibration, frame rate, and environmental conditions may influence the generalization and stability of the models when applied to heterogeneous contexts.

Nonetheless, the proposed multimodal and deep learning-based methodologies have demonstrated strong potential to mitigate these limitations. By combining 2D and 3D cues and integrating algorithmic reasoning with neural representations, the developed approaches significantly improve robustness, temporal consistency, and adaptability. These results form a solid foundation for future work toward automated gait analysis systems that operate reliably beyond laboratory conditions.

7.3 Future Work and Research Directions

Although the research objectives were achieved, several promising directions remain open for exploration. A first avenue concerns the expansion of the GRIDDS dataset to include more diverse capture conditions, outdoor recordings, and non-linear walking trajectories. The inclusion of subjects with gait impairments or different anthropometric characteristics would increase its applicability in clinical and rehabilitation contexts. The integration of complementary modalities, such as inertial sensors, electromyography (EMG), or plantar pressure, would further enrich the multimodal framework and strengthen biomechanical analysis.

A second direction involves the application of advanced deep learning architectures, such as transformers and graph neural networks (GNNs), which excel at modeling spatial-temporal dependencies. These architectures could capture complex motion dynamics more effectively, particularly when applied to fused 2D-3D representations.

Finally, future studies should explore cross-dataset and cross-domain generalization, using self-supervised and domain adaptation techniques to enhance the transferability of gait models across sensors and environments. In addition, integrating computational gait analysis with biomechanical and rehabilitation systems could produce interactive adaptive tools for diagnosis, personalized therapy, and patient monitoring, expanding the societal impact of this research.

7.4 Final Remarks

In conclusion, this thesis offers conceptual and empirical support for the view that leveraging 2D visual and 3D skeletal information, in combination with deep learning-based modeling, provides a solid foundation for the development of novel human gait analysis systems.

Through the creation of the GRIDDS dataset, the introduction of novel representations, and the development of combined algorithmic and learning-based methods, this work contributes to measurable advances in the accuracy, robustness, and automation of gait analysis.

Beyond its technical achievements, this research points toward a broader perspective of gait analysis, one that is multimodal, data-driven, and adaptive. By linking biomechanics, computer vision, and artificial intelligence, the work establishes a foundation for future developments aimed at a more comprehensive understanding and interpretation of human gait, while also offering methodological insights that may extend to other domains of human motion analysis.

References

- Xtion PRO LIVE – helpdesk / manual. URL https://www.asus.com/supportonly/xtion%20pro%20live/helpdesk_manual/.
- Iqra Aijaz Abro and Ahmad Jalal. Multi-modal sensors fusion for fall detection and action recognition in indoor environment. In *Proceedings of the 3rd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, pages 1–6. IEEE, November 2024. doi: 10.1109/etecte63967.2024.10823705.
- Iqra Aijaz Abro, Shuaa S. Alharbi, Naif S. Alshammari, Asaad Algarni, Nouf Abdullah Almuljally, Ahmad Jalal, and Hui Liu. Multimodal intelligent biosensors framework for fall disease detection and healthcare monitoring. *Frontiers in Bioengineering and Biotechnology*, 13, June 2025. ISSN 2296-4185. doi: 10.3389/fbioe.2025.1544968.
- Lars Adde, Jorunn L. Helbostad, Alexander R. Jensenius, Gunnar Taraldsen, Kristine H. Grunewaldt, and Ragnhild Støen. Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study. *Developmental Medicine and Child Neurology*, 52(8):773–778, July 2010. ISSN 1469-8749. doi: 10.1111/j.1469-8749.2010.03629.x.
- Jindřich Adolf, Yoram Segal, Matyáš Turna, Tereza Nováková, Jaromír Doležal, Patrik Kutílek, Jan Hejda, Ofer Hadar, and Lenka Lhotská. Evaluation of functional tests performance using a camera-based and machine learning approach. *PLOS ONE*, 18(11):e0288279, November 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0288279.
- J. K. Aggarwal and Q. Cai. Human motion analysis: A review. In *Proceedings of the IEEE Nonrigid and Articulated Motion Workshop*, pages 428–440. IEEE Computer Society, 1999. doi: 10.1109/namw.1997.609859.
- J. K. Aggarwal and M. S. Ryoo. Human activity analysis: A review. *ACM Computing Surveys*, 43(3):1–43, 2011. ISSN 0360-0300. doi: 10.1145/1922649.1922653.
- Valentina Agostini, Marco Ghislieri, Samanta Rosati, Gabriella Balestra, and Marco Knaflitz. Surface electromyography applied to gait analysis: How to improve its impact in clinics? *Frontiers in Neurology*, 11, September 2020. ISSN 1664-2295. doi: 10.3389/fneur.2020.00994.
- Juan Aguado-Delgado, José-María Gutiérrez-Martínez, José R. Hilera, Luis de Marcos, and Salvador Otón. Accessibility in video games: a systematic review. *Universal Access in the Information Society*, 19(1):169–193, August 2018. ISSN 1615-5297. doi: 10.1007/s10209-018-0628-2.
- Justin Amadeus Albert, Victor Owolabi, Arnd Gebel, Clemens Markus Brahms, Urs Granacher, and Bert Arnrich. Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study. *Sensors*, 20(18):5104, September 2020. ISSN 1424-8220. doi: 10.3390/s20185104.

- A. B. Albu, B. Widsten, T. Wang, J. Lan, and J. Mah. A computer vision-based system for real-time detection of sleep onset in fatigued drivers. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 25–30. IEEE, 2008. doi: 10.1109/IVS.2008.4621133.
- Pedro Albuquerque, João Pedro Machado, Tanmay Tulsidas Verlekar, Paulo Lobato Correia, and Luís Ducla Soares. Remote gait type classification system using markerless 2d video. *Diagnostics*, 11(10):1824, October 2021a. ISSN 2075-4418. doi: 10.3390/diagnostics11101824.
- Pedro Albuquerque, Tanmay Tulsidas Verlekar, Paulo Lobato Correia, and Luís Ducla Soares. A spatiotemporal deep learning approach for automatic pathological gait classification. *Sensors*, 21(18):6202, September 2021b. ISSN 1424-8220. doi: 10.3390/s21186202.
- Dimitrios S. Alexiadis, Philip Kelly, Petros Daras, Noel E. O'Connor, Tamy Boubekour, and Maher Ben Moussa. Evaluating a dancer's performance using kinect-based skeleton tracking. In *Proceedings of the 19th ACM international conference on Multimedia*. ACM, nov 2011. doi: 10.1145/2072298.2072412.
- Alyaa Qusay Aloraibi. Image morphing techniques: A review. *Technium: Romanian Journal of Applied Sciences and Technology*, 9:41–53, April 2023. ISSN 2668-778X. doi: 10.47577/technium.v9i.8699.
- Md. Altab Hossain, Yasushi Makihara, Junqiu Wang, and Yasushi Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, June 2010. ISSN 0031-3203. doi: 10.1016/j.patcog.2009.12.020.
- M. Amanulla Khan, S.M.H. Sithi Shameem Fathima, B. Arockia Stepnila, and A. Mohamed Ifthikar Ali. Joint intensity transformer network for gait recognition robust against clothing and carrying status. *Materials Today: Proceedings*, 33:3008–3020, 2020. ISSN 2214-7853. doi: 10.1016/j.matpr.2020.03.168.
- Achraf Ammar, Hamdi Chtourou, Osama Abdelkarim, Anthony Parish, and Anita Hoekelmann. Free throw shot in basketball: kinematic analysis of scored and missed shots during the learning process. *Sport Sciences for Health*, 12(1):27–33, 2015. doi: 10.1007/s11332-015-0250-0.
- Weizhi An, Shiqi Yu, Yasushi Makihara, Xinhui Wu, Chi Xu, Yang Yu, Rijun Liao, and Yasushi Yagi. Performance evaluation of model-based gait on multi-view very large population database with pose sequences. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4): 421–430, October 2020. ISSN 2637-6407. doi: 10.1109/tbiom.2020.3008862.
- Mohd. Aquib Ansari and Dushyant Kumar Singh. Human detection techniques for real time surveillance: a comprehensive survey. *Multimedia Tools and Applications*, 80(6):8759–8808, November 2020. ISSN 1573-7721. doi: 10.1007/s11042-020-10103-4.
- Mattia Antonelli, Elena Caselli, and Laura Gastaldi. Comparison of gait smoothness metrics in healthy elderly and young people. *Applied Sciences*, 14(2):911, January 2024. ISSN 2076-3417. doi: 10.3390/app14020911.
- Stéthane Armand, Geraldo Decoulon, and Alice Bonnefoy-Mazure. Gait analysis in children with cerebral palsy. *EFORT Open Reviews*, 1(12):448–460, December 2016. ISSN 2058-5241. doi: 10.1302/2058-5241.1.000052.

- Kamfiroozie Armin, Zohari Mehrana, and Dehbozorgi Fatemeh. Using Kinect in teaching children with hearing and visual impairment. In *Proceedings of the 4th International Conference on e-Learning and e-Teaching (ICELET'2013)*. IEEE, feb 2013. doi: 10.1109/icelet.2013.6681651.
- Yunus Ziya Arslan, Derya Karabulut, Faruk Ortes, and Marko B. Popovic. *Exoskeletons, Exomusculatures, Exosuits: Dynamic Modeling and Simulation*, pages 305–331. Elsevier, 2019. ISBN 9780128129395. doi: 10.1016/b978-0-12-812939-5.00011-2.
- Umar Asif, Stefan Von Cavallar, Jianbin Tang, and Stefan Harrer. SSHFD: Single shot human fall detection with occluded joints resilience. In *Proceedings of the European Conference on Artificial Intelligence (ECAI'2020)*, pages 2656–2663, 2020.
- Oliver Assad, Robert Hermann, Damian Lilla, Björn Mellies, Ronald Meyer, Liron Shevach, Sandra Siegel, Melanie Springer, Saranat Tiemkeo, Jens Voges, Jan Wieferich, Marc Herrlich, Markus Krause, and Rainer Malaka. Motion-based games for parkinson's disease patients. In *Entertainment Computing – ICEC'2011*, pages 47–58. Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-24500-8_6.
- Chaiyawan Auepanwiriyaikul, Sigourney Waibel, Joanna Songa, Paul Bentley, and A. Aldo Faisal. Accuracy and acceptability of wearable motion tracking for inpatient monitoring using smart-watches. *Sensors*, 20(24):7313, December 2020. ISSN 1424-8220. doi: 10.3390/s20247313.
- Sai Thu Ya Aung and Worapan Kusakunniran. A comprehensive review of gait analysis using deep learning approaches in criminal investigation. *PeerJ Computer Science*, 10:e2456, November 2024. ISSN 2376-5992. doi: 10.7717/peerj-cs.2456.
- Giuseppe Averta, Federica Barontini, Vincenzo Catrambone, Sami Haddadin, Giacomo Handjaras, Jeremia P O Held, Tingli Hu, Eike Jakubowitz, Christoph M Kanzler, Johannes Kühn, Olivier Lambercy, Andrea Leo, Alina Obermeier, Emiliano Ricciardi, Anne Schwarz, Gaetano Valenza, Antonio Bicchi, and Matteo Bianchi. U-limb: A multi-modal, multi-center database on arm motion control in healthy and post-stroke conditions. *GigaScience*, 10(6), June 2021. ISSN 2047-217X. doi: 10.1093/gigascience/giab043.
- Danilo Avola, Marco Cascio, Luigi Cinque, Gian Luca Foresti, Cristiano Massaroni, and Emanuele Rodola. 2-d skeleton-based action recognition via two-branch stacked LSTM-RNNs. *IEEE Transactions on Multimedia*, 22(10):2481–2496, October 2020. ISSN 1941-0077. doi: 10.1109/tmm.2019.2960588.
- Margit M. Bach, Andreas Daffertshofer, and Nadia Dominici. The development of mature gait patterns in children during walking and running. *European Journal of Applied Physiology*, 121(4):1073–1085, January 2021. ISSN 1439-6327. doi: 10.1007/s00421-020-04592-2.
- Ashish Badiye, Prachi Kathane, and Kewal Krishan. Forensic gait analysis. *StatPearls*, January 2022.
- Shikhar Bahl, Abhinav Gupta, and Deepak Pathak. Human-to-robot imitation in the wild, 2022.
- Sagor Chandro Bakchy, Md. Rabiul Islam, M. Rasel Mahmud, and Faisal Imran. Human gait analysis using gait energy image, 2022.
- Michal Balazia and Petr Sojka. Gait recognition from motion capture data. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14(1s):1–18, February 2018. ISSN 1551-6865. doi: 10.1145/3152124.

- Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-aware video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- Alessio Baratta, Antonio Cimino, Maria Grazia Gnoni, and Francesco Longo. Human robot collaboration in industry 4.0: a literature review. *Procedia Computer Science*, 217:1887–1895, 2023. ISSN 1877-0509. doi: 10.1016/j.procs.2022.12.389.
- M. I. Baritz. Video analysis system of the gait cycle and posture of people wearing high heels. *IOP Conference Series: Materials Science and Engineering*, 568(1):012109, July 2019. ISSN 1757-899X. doi: 10.1088/1757-899x/568/1/012109.
- Mihaela ioana Baritz and Adrian alexandru Mosoi. Kinematics of tennis movements for a player in wheelchair. In *Proceedings of the 13th International Conference eLearning and Software for Education*, volume 3 of *eLSE*, pages 127–132. Carol I National Defence University Publishing House, April 2017. doi: 10.12753/2066-026x-17-192.
- Arnaud Barre and Stéphane Armand. Biomechanical toolkit: Open-source framework to visualize and process biomechanical data. *Computer Methods and Programs in Biomedicine*, 114(1): 80–87, April 2014. ISSN 0169-2607. doi: 10.1016/j.cmpb.2014.01.012.
- Khalid Bashir, Tao Xiang, and Shaogang Gong. Gait recognition using gait entropy image. In *Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention*, ICDP. IET, 2009. doi: 10.1049/ic.2009.0230.
- Giulia Bassani, Alessandro Filippeschi, and Carlo Alberto Avizzano. A dataset of human motion and muscular activities in manual material handling tasks for biomechanical and ergonomic analyses. *IEEE Sensors Journal*, 21(21):24731–24739, November 2021a. ISSN 2379-9153. doi: 10.1109/jsen.2021.3113123.
- Giulia Bassani, Alessandro Filippeschi, and Carlo Alberto Avizzano. Manual material handling dataset for biomechanical and ergonomics analysis., 2021b.
- Romain Baud, Ali Reza Manzoori, Auke Ijspeert, and Mohamed Bouri. Review of control strategies for lower-limb exoskeletons to assist gait. *Journal of NeuroEngineering and Rehabilitation*, 18(1), July 2021. ISSN 1743-0003. doi: 10.1186/s12984-021-00906-3.
- S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–466, September 1995. ISSN 1557-7341. doi: 10.1145/212094.212141.
- Yoav Beck, Talia Herman, Marina Brozgol, Nir Giladi, Anat Mirelman, and Jeffrey M. Hausdorff. Sparc: a new approach to quantifying gait smoothness in patients with parkinson’s disease. *Journal of NeuroEngineering and Rehabilitation*, 15(1), June 2018. ISSN 1743-0003. doi: 10.1186/s12984-018-0398-3.
- Jérémie Begue, Teddy Caderby, Nicolas Peyrot, and Georges Dalleau. Influence of gait speed on free vertical moment during walking. *Journal of Biomechanics*, 75:186–190, June 2018. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2018.05.011.
- Csaba Benedek, Bence Galai, Balazs Nagy, and Zsolt Janko. Lidar-based gait analysis and activity recognition in a 4d surveillance system. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(1):101–113, January 2018. ISSN 1558-2205. doi: 10.1109/tcsvt.2016.2595331.

- Sean T. Bennett, Wei Han, Dilruba Mahmud, Peter G. Adamczyk, Fei Dai, Michael Wehner, Dharmaraj Veeramani, and Zhenhua Zhu. Usability and biomechanical testing of passive exoskeletons for construction workers: A field-based pilot study. *Buildings*, 13(3):822, 2023. doi: 10.3390/buildings13030822.
- Richa Bharadwaj, Srijittra Swaisaenyakorn, Clive G. Parini, John C. Batchelor, and Akram Alo-mainy. Impulse radio ultra-wideband communications for localization and tracking of human body and limbs movement for healthcare applications. *IEEE Transactions on Antennas and Propagation*, 65(12):7298–7309, December 2017. ISSN 1558-2221. doi: 10.1109/tap.2017.2759841.
- Medha Bhargava, Chia-Chih Chen, M. S. Ryoo, and J. K. Aggarwal. Detection of abandoned objects in crowded environments. In *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 271–276. IEEE, September 2007. doi: 10.1109/avss.2007.4425322.
- Vivek D. Bhise. *Ergonomics in the Automotive Design Process: Advanced Topics, Measurements, Modeling and Research*. CRC Press, May 2024. ISBN 9781003485605. doi: 10.1201/9781003485605.
- Alexy Bhowmick, Saurabh Prakash, Rukmani Bhagat, Vijay Prasad, and Shyamanta M. Hazarika. IntelliNavi: Navigation for blind based on kinect and machine learning. In *Lecture Notes in Computer Science*, pages 172–183. Springer International Publishing, 2014. doi: 10.1007/978-3-319-13365-2_16.
- Cunling Bian, Yang Yang, Tao Wang, and Weigang Lu. Spatial–temporal heatmap masked autoencoder for skeleton-based action recognition. *Sensors*, 25(10):3146, May 2025. ISSN 1424-8220. doi: 10.3390/s25103146.
- Belinda Bilney, Meg Morris, and Kate Webster. Concurrent related validity of the gaitrite walkway system for quantification of the spatial and temporal parameters of gait. *Gait & Posture*, 17(1): 68–74, February 2003. ISSN 0966-6362. doi: 10.1016/s0966-6362(02)00053-x.
- Ivan Birch, Maria Birch, and Nadia Asgeirsdottir. The identification of individuals by observational gait analysis using closed circuit television footage: Comparing the ability and confidence of experienced and non-experienced analysts. *Science and Justice*, 60(1):79–85, January 2020. ISSN 1355-0306. doi: 10.1016/j.scijus.2019.10.002.
- Joydeep Biswas and Manuela Veloso. Depth camera based indoor mobile robot localization and navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, may 2012. doi: 10.1109/icra.2012.6224766.
- Silvia S. Blemker, Deanna S. Asakawa, Garry E. Gold, and Scott L. Delp. Image-based musculoskeletal modeling: Applications, advances, and future opportunities. *Journal of Magnetic Resonance Imaging*, 25(2):441–451, January 2007. ISSN 1522-2586. doi: 10.1002/jmri.20805.
- Victoria Bloom, Vasileios Argyriou, and Dimitrios Makris. G3Di: A gaming interaction dataset with a real time detection and evaluation framework. In *Computer Vision - ECCV 2014 Workshops*, pages 698–712. Springer International Publishing, 2015. doi: 10.1007/978-3-319-16178-5_49.

- Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Unsupervised feature learning for RGB-d based object recognition. In *Experimental Robotics*, pages 387–402. Springer International Publishing, 2013. doi: 10.1007/978-3-319-00065-7_27.
- José Carlos García Bohigas, Yumilka Daisy Ruiz Luaces, and Gilberto Herrera Delgado. Analysis of vertical jump of first category volley ball players. *PODIUM: Revista de Ciencia y Tecnología en la Cultura Física*, 16(3):906–917, 2021.
- Bruno Bonnechère. *Serious Games in Rehabilitation*, pages 41–109. Springer International Publishing, October 2017. ISBN 9783319661223. doi: 10.1007/978-3-319-66122-3_4.
- Marco Bortolini, Maurizio Faccio, Mauro Gamberi, and Francesco Pilati. Motion analysis system (mas) for production and ergonomics assessment in the manufacturing processes. *Computers & Industrial Engineering*, 139:105485, January 2020. ISSN 0360-8352. doi: 10.1016/j.cie.2018.10.046.
- Katherine A. Boyer, Julia Freedman Silvernail, and Joseph Hamill. Age and sex influences on running mechanics and coordination variability. *Journal of Sports Sciences*, 35(22):2225–2231, dec 2016. doi: 10.1080/02640414.2016.1265139.
- Markus Braun, Sebastian Krebs, Fabian Flohr, and Darius M. Gavrila. EuroCity persons: A novel benchmark for person detection in traffic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1844–1861, 2019a. doi: 10.1109/tpami.2019.2897684.
- Markus Braun, Sebastian Krebs, Fabian B. Flohr, and Darius M. Gavrila. Eurocity persons dataset, 2019b. URL <https://eurocity-dataset.tudelft.nl/>.
- W. Braune and O. Fischer. *Der gang des Menschen*. Number vol. 1 in Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften [zu Leipzig]. B. S. Hirzel, 1895. URL <https://books.google.pt/books?id=Aa9AAQAAMAAJ>.
- Rafael Caldas, Rebeca Sarai, Fernando Buarque de Lima Neto, and Bernd Markert. Validation of two hybrid approaches for clustering age-related groups based on gait kinematics data. *Medical Engineering and Physics*, 78:90–97, April 2020. ISSN 1350-4533. doi: 10.1016/j.medengphy.2020.02.001.
- Shun Cao. Study state dynamics of team passing networks in soccer games. *Journal of Sports Sciences*, 43(1):33–47, June 2023. ISSN 1466-447X. doi: 10.1080/02640414.2023.2229154.
- Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1302–1310. IEEE, July 2017a. doi: 10.1109/cvpr.2017.143.
- Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1302–1310. IEEE, July 2017b. doi: 10.1109/cvpr.2017.143.
- Francisco M. Castro, Rubén Delgado-Escáño, Rubén Hernández-García, Manuel J. Marín-Jiménez, and Nicolás Guil. Attengait: Gait recognition with attention and rich modalities. *Pattern Recognition*, 148:110171, April 2024. ISSN 0031-3203. doi: 10.1016/j.patcog.2023.110171.

- Francisco Manuel Castro, Manuel J. Marín-Jiménez, Nicolás Guil, and Nicolás Pérez de la Blanca. Automatic learning of gait signatures for people identification. In *Advances in Computational Intelligence*, pages 257–270. Springer International Publishing, 2017. doi: 10.1007/978-3-319-59147-6_23.
- Andy Catruna, Adrian Cosma, and Emilian Radoi. Gaitpt: Skeletons are all you need for gait recognition. In *Proceedings of the IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10. IEEE, May 2024. doi: 10.1109/fg59268.2024.10581947.
- Mainak Chakraborty, Masood Aryapoor, and Masoud Daneshlab. Frequency domain complex-valued convolutional neural network. *Expert Systems with Applications*, 295:128893, January 2025. ISSN 0957-4174. doi: 10.1016/j.eswa.2025.128893.
- Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. GaitSet: Regarding gait as a set for cross-view gait recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):8126–8133, July 2019. ISSN 2159-5399. doi: 10.1609/aaai.v33i01.33018126.
- Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021. ISSN 1939-3539. doi: 10.1109/tpami.2021.3057879.
- Jose M. Chaquet, Enrique J. Carmona, and Antonio Fernández-Caballero. A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, 117(6):633–659, 2013. ISSN 1077-3142. doi: <http://dx.doi.org/10.1016/j.cviu.2013.01.013>.
- Pratik Chattopadhyay, Shamik Sural, and Jayanta Mukherjee. Frontal gait recognition from occluded scenes. *Pattern Recognition Letters*, 63:9–15, October 2015. ISSN 0167-8655. doi: 10.1016/j.patrec.2015.06.004.
- Bing Chen, Bin Zi, Zhengyu Wang, Ling Qin, and Wei-Hsin Liao. Knee exoskeletons for gait rehabilitation and human performance augmentation: A state-of-the-art. *Mechanism and Machine Theory*, 134:499–511, April 2019. ISSN 0094-114X. doi: 10.1016/j.mechmachtheory.2019.01.016.
- Chen Chen, Kui Liu, and Nasser Kehtarnavaz. Real-time human action recognition based on depth motion maps. *Journal of Real-Time Image Processing*, 12(1):155–163, aug 2013a. doi: 10.1007/s11554-013-0370-1.
- Ching-Hang Chen and Deva Ramanan. 3D human pose estimation = 2D pose estimation + matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017. doi: 10.1109/cvpr.2017.610.
- Fang Chen, Xiwen Cui, Zhe Zhao, Daoqiang Zhang, Cong Ma, Xinran Zhang, and Hongen Liao. Gait acquisition and analysis system for osteoarthritis based on hybrid prediction model. *Computerized Medical Imaging and Graphics*, 85:101782, October 2020. ISSN 0895-6111. doi: 10.1016/j.compmedimag.2020.101782.
- Kai-Yu Chen, Li-Wei Chou, Hui-Min Lee, Shuenn-Tsong Young, Cheng-Hung Lin, Yi-Shu Zhou, Shih-Tsang Tang, and Ying-Hui Lai. Human motion tracking using 3D image features with a long short-term memory mechanism model—an example of forward reaching. *Sensors*, 22(1): 292, 2021. doi: 10.3390/s22010292.

- Lulu Chen, Hong Wei, and James Ferryman. A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, 34(15):1995–2006, nov 2013b. ISSN 0167-8655. doi: 10.1016/j.patrec.2013.02.006.
- Tianyu Chen, Chunping Hou, Zhipeng Wang, and Hua Chen. Anomaly detection in crowded scenes using motion energy model. *Multimedia Tools and Applications*, 77(11):14137–14152, jul 2017. doi: 10.1007/s11042-017-5020-3.
- Zhixiang Chi, Rasoul Mohammadi Nasiri, Zheng Liu, Juwei Lu, Jin Tang, and Konstantinos N. Plataniotis. All at once: Temporally adaptive multi-frame interpolation with advanced motion modeling. In *Computer Vision – ECCV 2020*, pages 107–123. Springer International Publishing, 2020. ISBN 9783030585839. doi: 10.1007/978-3-030-58583-9_7.
- Joana Chong, Petra Tjurin, Maisa Niemelä, Timo Jämsä, and Vahid Farrahi. Machine-learning models for activity class prediction: A comparative study of feature selection and classification algorithms. *Gait & Posture*, 89:45–53, 2021. doi: 10.1016/j.gaitpost.2021.06.017.
- Jiaming Chu, Lei Jin, Yinglei Teng, Jianshu Li, Yunchao Wei, Zheng Wang, Junliang Xing, Shuicheng Yan, and Jian Zhao. Uniparser: Multi-human parsing with unified correlation representation learning. *IEEE Transactions on Image Processing*, 33:5159–5171, 2024. ISSN 1941-0042. doi: 10.1109/tip.2024.3456004.
- Xu Chu, Ihab F. Ilyas, Sanjay Krishnan, and Jiannan Wang. Data cleaning: Overview and emerging challenges. In *Proceedings of the International Conference on Management of Data, SIGMOD/PODS’16*, pages 2201–2206. ACM, June 2016. doi: 10.1145/2882903.2912574.
- Grazia Cicirelli, Donato Impedovo, Vincenzo Dentamaro, Roberto Marani, Giuseppe Pirlo, and Tiziana R. D’Orazio. Human gait analysis in neurodegenerative diseases: A review. *IEEE Journal of Biomedical and Health Informatics*, 26(1):229–242, January 2022. ISSN 2168-2208. doi: 10.1109/jbhi.2021.3092875.
- Krzysztof Chris Ciesielski, Xinjian Chen, Jayaram K. Udupa, and George J. Grevera. Linear time algorithms for exact distance transform. *Journal of Mathematical Imaging and Vision*, 39(3): 193–209, November 2010. ISSN 1573-7683. doi: 10.1007/s10851-010-0232-4.
- Ross A. Clark, Yong-Hao Pua, Cristino C. Oliveira, Kelly J. Bower, Shamala Thilarajah, Rebekah McGaw, Ksaniel Hasanki, and Benjamin F. Mentiplay. Reliability and concurrent validity of the microsoft xbox one kinect for assessment of standing balance and postural control. *Gait & Posture*, 42(2):210–213, jul 2015. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2015.03.005.
- Ross A. Clark, Benjamin F. Mentiplay, Emma Hough, and Yong Hao Pua. Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and kinect alternatives. *Gait & Posture*, 68:193–200, February 2019. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2018.11.029.
- David Cofer, Gennady Cymbalyuk, James Reid, Ying Zhu, William J. Heitler, and Donald H. Edwards. Animatlab: A 3d graphics environment for neuromechanical simulations. *Journal of Neuroscience Methods*, 187(2):280–288, March 2010. ISSN 0165-0270. doi: 10.1016/j.jneumeth.2010.01.005.
- Robert T. Collins, Ralph Gross, and Jianbo Shi. Silhouette-based human identification from body shape and gait. In *Proceedings of the 5th IEEE International Conference on Automatic Face Gesture Recognition*, AFGR-02, pages 366–371. IEEE, 2002. doi: 10.1109/afgr.2002.1004181.

- Taya A. Collyer, Anne M. Murray, Robyn L. Woods, Elsdon Storey, Trevor T.-J. Chong, Joanne Ryan, Suzanne G. Orchard, Amy Brodtmann, Velandai K. Srikanth, Raj C. Shah, and Michele L. Callisaya. Association of dual decline in cognition and gait speed with risk of dementia in older adults. *JAMA Network Open*, 5(5):e2214647, May 2022. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2022.14647.
- Laura Comber, Rose Galvin, and Susan Coote. Gait deficits in people with multiple sclerosis: A systematic review and meta-analysis. *Gait & Posture*, 51:25–35, January 2017. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2016.09.026.
- Patrick Connor and Arun Ross. Biometric recognition by gait: A survey of modalities and features. *Computer Vision and Image Understanding*, 167:1–27, February 2018. ISSN 1077-3142. doi: 10.1016/j.cviu.2018.01.007.
- Patrick C. Connor. Comparing and combining underfoot pressure features for shod and unshod gait biometrics. In *Proceedings of the IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–7. IEEE, April 2015. doi: 10.1109/ths.2015.7225338.
- Microsoft Corporation. *Human Interface Guidelines v2.0*, 2014. URL <https://download.microsoft.com/download/6/7/6/676611b4-1982-47a4-a42e-4cf84e1095a8/kinecthig.2.0.pdf>.
- Andrea Corti, Silvio Giancola, Giacomo Mainetti, and Remo Sala. A metrological characterization of the kinect v2 time-of-flight camera. *Robotics and Autonomous Systems*, 75:584–594, January 2016. ISSN 0921-8890. doi: 10.1016/j.robot.2015.09.024.
- Adrian Cosma and Emilian Radoi. Learning gait representations with noisy multi-task learning. *Sensors*, 22(18):6803, September 2022. ISSN 1424-8220. doi: 10.3390/s22186803.
- Mark W. Creaby and Michael H. Cole. Gait characteristics and falls in parkinson’s disease: A systematic review and meta-analysis. *Parkinsonism & Related Disorders*, 57:1–8, December 2018. ISSN 1353-8020. doi: 10.1016/j.parkreldis.2018.07.008.
- Cloe Cummins, Rhonda Orr, Helen O’Connor, and Cameron West. Global positioning systems (gps) and microtechnology sensors in team sports: A systematic review. *Sports Medicine*, 43(10):1025–1042, June 2013. ISSN 1179-2035. doi: 10.1007/s40279-013-0069-2.
- Carlo Dal Mutto, Pietro Zanuttigh, and Guido M. Cortelazzo. *Time-of-flight cameras and Microsoft Kinect*. SpringerBriefs in Electrical and Computer Engineering Ser. Springer New York, New York, NY, 1st edition, 2012. ISBN 9781461438076. Description based on publisher supplied metadata and other sources.
- Michael Damsgaard, John Rasmussen, S. T. Christensen, and Egidijus Surma. Anybody - a system for biomechanical analysis and ergonomic design. In *Proceedings of the Max Planck Workshop on Structural Optimization*, pages 28–31, 2001.
- Michael Damsgaard, John Rasmussen, Søren Tørholm Christensen, Egidijus Surma, and Mark de Zee. Analysis of musculoskeletal systems in the anybody modeling system. *Simulation Modelling Practice and Theory*, 14(8):1100–1111, November 2006. ISSN 1569-190X. doi: 10.1016/j.simpat.2006.09.001.
- Kristiaan D’Août, Aoife Healy, Nicola Eddison, and Nachiappan Chockalingam. *Introduction to gait analysis*, pages 1–16. Institution of Engineering and Technology, April 2022. ISBN 9781839531323. doi: 10.1049/pbhe031e_ch1.

- Pinto-Fernández David, Rodriguez-Cianca David, Moreno Juan C., and Torricelli Diego. Human locomotion databases: A systematic review. *IEEE Journal of Biomedical and Health Informatics*, 28(3):1716–1729, March 2024. ISSN 2168-2208. doi: 10.1109/jbhi.2023.3311677.
- Roy B. Davis. Reflections on clinical gait analysis. *Journal of Electromyography and Kinesiology*, 7(4):251–257, 1997. ISSN 1050-6411. doi: 10.1016/s1050-6411(97)00008-4.
- Ingrid Miranda de Sousa, Joel Lucas de Jesus Trindade, and Gerardo Antônio Idrobo Pizo. *Simulation of Human Hands Movements Using Forward Kinematics*, pages 705–710. Springer Singapore, 2019. ISBN 9789811321191. doi: 10.1007/978-981-13-2119-1_108.
- Brian DeCann, Arun Ross, and Jeremy Dawson. Investigating gait recognition in the short-wave infrared (SWIR) spectrum: Dataset and challenges. In Ioannis Kakadiaris, Walter J. Scheirer, and Laurence G. Hassebrook, editors, *Proceedings of the Biometric and Surveillance Technology for Human and Activity Identification X*. SPIE, 2013. doi: 10.1117/12.2018145.
- Omid Dehzangi, Mojtaba Taherisadr, and Raghvendar ChandalVala. Imu-based gait recognition using convolutional neural networks and multi-sensor fusion. *Sensors*, 17(12):2735, November 2017. ISSN 1424-8220. doi: 10.3390/s17122735.
- Scott L. Delp, Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, and Darryl G. Thelen. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11):1940–1950, 2007. doi: 10.1109/TBME.2007.901024.
- Marcin Derlatka and Marek Parfieniuk. Real-world measurements of ground reaction forces of normal gait of young adults wearing various footwear. *Scientific Data*, 10(1), January 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-01964-z.
- Lazzaro di Biase, Alessandro Di Santo, Maria Letizia Caminiti, Alfredo De Liso, Syed Ahmar Shah, Lorenzo Ricci, and Vincenzo Di Lazzaro. Gait analysis in parkinson’s disease: An overview of the most accurate markers for diagnosis and symptoms monitoring. *Sensors*, 20(12):3529, June 2020. ISSN 1424-8220. doi: 10.3390/s20123529.
- Francesco Di Nardo, Christian Morbidoni, Guido Mascia, Federica Verdini, and Sandro Fioretti. Intra-subject approach for gait-event prediction by neural network interpretation of emg signals. *BioMedical Engineering OnLine*, 19(1), July 2020. ISSN 1475-925X. doi: 10.1186/s12938-020-00803-1.
- P. R. Diaz-Monterrosas, R. Posada-Gomez, A. Martinez-Sibaja, A. A. Aguilar-Lasserre, U. Juarez-Martinez, and J. C. Trujillo-Caballero. A brief review on the validity and reliability of microsoft kinect sensors for functional assessment applications. *Advances in Electrical and Computer Engineering*, 18(1):131–136, 2018. ISSN 1844-7600. doi: 10.4316/aece.2018.01016.
- Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, July 1945. ISSN 1939-9170. doi: 10.2307/1932409.
- Massimiliano Ditroilo, Ryan Burgess, Eoin Cunniffe, Adam Grainger, and Denise McGrath. Assessing field hockey drag-flicking: concurrent validity of the output sports imu. *Sport Performance & Science Reports*, 2023.
- Jiong Dong, Kaoru Ota, and Mianxiong Dong. Video frame interpolation: A comprehensive survey. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 19(2s):1–31, April 2023. ISSN 1551-6865. doi: 10.1145/3556544.

- Lili Dong, Tianliang Hu, Tianyi Sun, Junrui Li, and Songhua Ma. Rgb video and inertial sensing fusion method for human action recognition in human-robot collaborative manufacturing. *Journal of Manufacturing Systems*, 83:216–234, December 2025. ISSN 0278-6125. doi: 10.1016/j.jmsy.2025.09.007.
- Krupa B. Doshi, Seong Hyun Moon, Michael D. Whitaker, and Thurmon E. Lockhart. Assessment of gait and posture characteristics using a smartphone wearable system for persons with osteoporosis with and without falls. *Scientific Reports*, 13(1), January 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-27788-w.
- Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2758–2766. IEEE, December 2015. doi: 10.1109/iccv.2015.316.
- Huanzhang Dou, Pengyi Zhang, Wei Su, Yunlong Yu, and Xi Li. *MetaGait: Learning to Learn an Omni Sample Adaptive Representation for Gait Recognition*, pages 357–374. Springer Nature Switzerland, 2022. ISBN 9783031200656. doi: 10.1007/978-3-031-20065-6_21.
- Meghan K. Driscoll, Colin McCann, Rael Kopace, Tess Homan, John T. Fourkas, Carole Parent, and Wolfgang Losert. Cell shape dynamics: From waves to migration. *PLoS Computational Biology*, 8(3):e1002392, March 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002392.
- Stacey C. Dusing and Deborah E. Thorpe. A normative sample of temporal and spatial gait parameters in children using the GAITRite electronic walkway. *Gait & Posture*, 25(1):135–139, January 2007. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2006.06.003.
- Patric Eichelberger, Matteo Ferraro, Ursina Minder, Trevor Denton, Angela Blasimann, Fabian Krause, and Heiner Baur. Analysis of accuracy in optical motion capture – a protocol for laboratory setup evaluation. *Journal of Biomechanics*, 49(10):2085–2088, July 2016. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2016.05.007.
- Sohair Talaat Ellabany and Mansour Abdel-Hamid Attaallah. Kinematic analysis of the whole body center of gravity trajectory and time structure of the tennis serve performance. *Journal of Applied Sports Science*, 5(4):76–81, December 2015. ISSN 2357-0059. doi: 10.21608/jass.2015.84528.
- Felix Endres, Jurgen Hess, Jurgen Sturm, Daniel Cremers, and Wolfram Burgard. 3-d mapping with an RGB-d camera. *IEEE Transactions on Robotics*, 30(1):177–187, feb 2014. doi: 10.1109/tro.2013.2279412.
- Dursun Erbil, Gokbel Tugba, Topcu Hasan Murat, Akarsu Melike, Akyüz Merve, Karacan Cagla, Çiftçi Can Mehmetali, Öztürk Akay, and Dursun Nigar. Effects of robot-assisted gait training in chronic stroke patients treated by botulinum toxin-a: A pivotal study. *Physiotherapy Research International*, 23(3), May 2018. ISSN 1471-2865. doi: 10.1002/pri.1718.
- Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature Medicine*, 25(1):24–29, jan 2019. doi: 10.1038/s41591-018-0316-z.

- Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, 4(1), jan 2021. doi: 10.1038/s41746-020-00376-2.
- Murray Evans, Laurie Needham, Logan Wade, Martin Parsons, Steffi Colyer, Polly McGuigan, James Bilzon, and Darren Cosker. Synchronised video, motion capture and force plate dataset for validating markerless human movement analysis. *Scientific Data*, 11(1), November 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-04077-3.
- Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, September 2009. ISSN 1573-1405. doi: 10.1007/s11263-009-0275-4.
- Germain Faity, Denis Mottet, and Jérôme Froger. Validity and reliability of kinect v2 for quantifying upper body kinematics during seated reaching. *Sensors*, 22(7):2735, April 2022. ISSN 1424-8220. doi: 10.3390/s22072735.
- Bingfei Fan, Qingguo Li, and Tao Liu. Accurate foot clearance estimation during level and uneven ground walking using inertial sensors. *Measurement Science and Technology*, 31(5):055106, March 2020a. ISSN 1361-6501. doi: 10.1088/1361-6501/ab6917.
- Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. GaitPart: Temporal part-based model for gait recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14213–14221. IEEE, June 2020b. doi: 10.1109/cvpr42600.2020.01423.
- Chao Fan, Saihui Hou, Yongzhen Huang, and Shiqi Yu. Exploring deep models for practical gait recognition, 2023.
- Chao Fan, Jingzhe Ma, Dongyang Jin, Chuanfu Shen, and Shiqi Yu. Skeletongait: Gait recognition using skeleton maps. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(2):1662–1669, March 2024. ISSN 2159-5399. doi: 10.1609/aaai.v38i2.27933.
- Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. RMPE: Regional multi-person pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017. doi: 10.1109/iccv.2017.256.
- Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7157–7173, June 2023. ISSN 1939-3539. doi: 10.1109/tpami.2022.3222784.
- Peter Fankhauser, Michael Bloesch, Diego Rodriguez, Ralf Kaestner, Marco Hutter, and Roland Siegwart. Kinect v2 for mobile robot navigation: Evaluation and modeling. In *Proceedings of the International Conference on Advanced Robotics (ICAR)*. IEEE, jul 2015. doi: 10.1109/icar.2015.7251485.
- M. E. Farmer and A. K. Jain. Smart automotive airbags: Occupant classification and tracking. *IEEE Transactions on Vehicular Technology*, 56(1):60–80, 2007. ISSN 0018-9545. doi: 10.1109/TVT.2006.883768.
- Francesco Felici and Alessandro Del Vecchio. Surface electromyography: What limits its use in exercise and sport physiology? *Frontiers in Neurology*, 11, November 2020. ISSN 1664-2295. doi: 10.3389/fneur.2020.578504.

- Richard Feliuss, Michiel Punt, Marieke Geerars, Natasja Wouda, Rins Rutgers, Sjoerd Bruijn, Sina David, and Jaap van Dieën. Exploring unsupervised feature extraction of imu-based gait data in stroke rehabilitation using a variational autoencoder. *PLOS ONE*, 19(10):e0304558, October 2024. ISSN 1932-6203. doi: 10.1371/journal.pone.0304558.
- Yang Feng, Yuncheng Li, and Jiebo Luo. Learning effective gait features using LSTM. In *Proceedings of the 23rd International Conference on Pattern Recognition*, ICPR, pages 320–325. IEEE, 2016. doi: 10.1109/icpr.2016.7899654.
- María del Rosario Ferreira-Sánchez, Marcos Moreno-Verdú, and Roberto Cano-de-la Cuerda. Quantitative measurement of rigidity in parkinson’s disease: A systematic review. *Sensors*, 20(3):880, February 2020. ISSN 1424-8220. doi: 10.3390/s20030880.
- Bruno Figueiredo, Álvaro Frazão, André Rouco, Beatriz Soares, Daniel Albuquerque, and Pedro Pinho. A review: Radar remote-based gait identification methods and techniques. *Remote Sensing*, 17(7):1282, April 2025. ISSN 2072-4292. doi: 10.3390/rs17071282.
- Michael Firman. RGBD datasets: Past, present and future. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, CVPRW. IEEE, 2016. doi: 10.1109/cvprw.2016.88.
- Sebastian Fischer and Armin Kibele. The biomechanical structure of swim start performance. *Sports Biomechanics*, 15(4):397–408, 2016. doi: 10.1080/14763141.2016.1171893.
- Sean P. Flanagan. Decreased knee extensor torque during single-limb stance: A computer simulation study of compensations and consequences. *Biomechanics*, 4(4):698–710, November 2024. ISSN 2673-7078. doi: 10.3390/biomechanics4040050.
- Matthew H. Foreman and Jack R. Engsborg. The validity and reliability of the microsoft kinect for measuring trunk compensation during reaching. *Sensors*, 20(24):7073, December 2020. ISSN 1424-8220. doi: 10.3390/s20247073.
- Stephen Fox, Olli Aranko, Juhani Heilala, and Päivi Vahala. Exoskeletons: Comprehensive, comparative and critical analyses of their potential to improve manufacturing performance. *Journal of Manufacturing Technology Management*, 31(6):1261–1280, June 2019. ISSN 1741-038X. doi: 10.1108/jmtm-01-2019-0023.
- Bas L. Fransen, Nina M.C. Mathijssen, Karin Slot, Nicole H.H. de Esch, Hennie Verburg, Olivier P.P. Temmerman, Marco J.M. Hoozemans, and Jaap H. van Dieën. Gait quality assessed by trunk accelerometry after total knee arthroplasty and its association with patient related outcome measures. *Clinical Biomechanics*, 70:192–196, December 2019. ISSN 0268-0033. doi: 10.1016/j.clinbiomech.2019.10.007.
- Andrew W. Froehle, Ramzi W. Nahhas, Richard J. Sherwood, and Dana L. Duren. Age-related changes in spatiotemporal characteristics of gait accompany ongoing lower limb linear growth in late childhood and early adolescence. *Gait & Posture*, 38(1):14–19, May 2013. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2012.10.005.
- David Fuentes-Jimenez, Roberto Martin-Lopez, Cristina Losada-Gutierrez, David Casillas-Perez, Javier Macias-Guarasa, Carlos A. Luna, and Daniel Pizarro. Dpdnet: A robust people detector using deep learning with an overhead depth camera. *Expert Systems with Applications*, 146: 113168, May 2020. ISSN 0957-4174. doi: 10.1016/j.eswa.2019.113168.

- Claudiane Arakaki Fukuchi, Reginaldo Kisho Fukuchi, and Marcos Duarte. Effects of walking speed on gait biomechanics in healthy participants: a systematic review and meta-analysis. *Systematic Reviews*, 8(1), June 2019. ISSN 2046-4053. doi: 10.1186/s13643-019-1063-z.
- Varun Ganapathi, Christian Plagemann, Daphne Koller, and Sebastian Thrun. Real time motion capture using a single time-of-flight camera. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 755–762. IEEE, June 2010. doi: 10.1109/cvpr.2010.5540141.
- Francisco A. Garcia, Juan C. Perez-Ibarra, Marco H. Terra, and Adriano A. G. Siqueira. Adaptive algorithm for gait segmentation using a single imu in the thigh pocket. *IEEE Sensors Journal*, 22(13):13251–13261, July 2022. ISSN 2379-9153. doi: 10.1109/jsen.2022.3177951.
- Sara García-de Villa, David Casillas-Pérez, Ana Jiménez-Martín, and Juan Jesús García-Domínguez. Inertial sensors for human motion analysis: A comprehensive review. *IEEE Transactions on Instrumentation and Measurement*, 72:1–39, 2023. ISSN 1557-9662. doi: 10.1109/tim.2023.3276528.
- Jorge García Montaña, Cecilia E. García Cena, Luis J. Monge Chamorro, Marie André Destarac, and Roque Salterén Pazmiño. *Mechanical Design of a Robotic Exoskeleton for Upper Limb Rehabilitation*, pages 297–308. Springer International Publishing, 2017. ISBN 9783319543772. doi: 10.1007/978-3-319-54377-2_25.
- D. M. Gavrilu. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999. ISSN 1077-3142. doi: 10.1006/cviu.1998.0716. URL <http://dx.doi.org/10.1006/cviu.1998.0716>.
- Marija Gavrilović and Dejan B. Popović. A principal component analysis (pca) based assessment of the gait performance. *Biomedical Engineering / Biomedizinische Technik*, 66(5):449–457, July 2021. ISSN 1862-278X. doi: 10.1515/bmt-2020-0307.
- Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, November 2021. ISSN 1557-7317. doi: 10.1145/3458723.
- Daphne J. Geerse, Bert H. Coolen, and Melvyn Roerdink. Kinematic validation of a multi-kinect v2 instrumented 10-meter walkway for quantitative gait assessments. *PLOS ONE*, 10(10):e0139913, October 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0139913.
- Thomas Geijtenbeek, Frans Steenbrink, Bert Otten, and Oshri Even-Zohar. D-flow: immersive virtual reality and real-time feedback for rehabilitation. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry, VRCAI '11*, pages 201–208. ACM Press, December 2011. doi: 10.1145/2087756.2087785.
- Farnood Gholami, Daria A. Trojan, Jozsef Kovecses, Wassim M. Haddad, and Behnood Gholami. A microsoft kinect-based point-of-care gait assessment framework for multiple sclerosis patients. *IEEE Journal of Biomedical and Health Informatics*, 21(99):1376–1385, September 2017. ISSN 2168-2194. doi: 10.1109/jbhi.2016.2593692.
- Davoud Gholamiangonabadi, Nikita Kislov, and Katarina Grolinger. Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *IEEE Access*, 8:133982–133994, 2020. ISSN 2169-3536. doi: 10.1109/access.2020.3010715.

- Marta Gómez-Galán, José Pérez-Alonso, Ángel-Jesús Callejón-Ferre, and Javier López-Martínez. Musculoskeletal disorders: OWAS review. *Industrial Health*, 55(4):314–337, 2017. ISSN 1880-8026. doi: 10.2486/indhealth.2016-0191.
- Rejane Vale Gonçalves, Sérgio Teixeira Fonseca, Priscila Albuquerque Araújo, Vanessa Lara Araújo, Tais Martins Barboza, Gabriela Andrade Martins, and Marisa Cotta Mancini. Identification of gait events in children with spastic cerebral palsy: comparison between the force plate and algorithms. *Brazilian Journal of Physical Therapy*, 24(5):392–398, September 2020. ISSN 1413-3555. doi: 10.1016/j.bjpt.2019.05.007.
- Carina S. González-González, Pedro A. Toledo-Delgado, Vanesa Muñoz Cruz, and Pablo V. Torres-Carrion. Serious games for rehabilitation: Gestural interaction in personalized gamified exercises through a recommender system. *Journal of Biomedical Informatics*, 97:103266, September 2019. ISSN 1532-0464. doi: 10.1016/j.jbi.2019.103266.
- Susane Graup, Daniele Detanico, Saray Giovana dos Santos, and Anne Ribeiro Streb. Avaliação biomecânica de ex-atletas diagnosticados com osteoartrite de quadril em diferentes estágios. *RBPFEEX - Revista Brasileira de Prescrição e Fisiologia do Exercício*, 2018.
- Andrea Greisberger, Brigitte Wolf, Klaus Widhalm, David Kollmitzer, Maximilian Arbesser, and Peter Putz. Intrarater and interrater reliability of angular measures using templo two-dimensional motion analysis software. *Journal of Manipulative and Physiological Therapeutics*, 42(6):425–429, July 2019. ISSN 0161-4754. doi: 10.1016/j.jmpt.2018.11.018.
- G. J. Grevera and J. K. Udupa. Shape-based interpolation of multidimensional grey-level images. *IEEE Transactions on Medical Imaging*, 15(6):881–892, December 1996. ISSN 1558-254X. doi: 10.1109/42.544506.
- Ralph Gross and Jianbo Shi. The CMU motion of body (MoBo) database. Technical Report CMU-RI-TR-01-18, Carnegie Mellon University, Pittsburgh, PA, June 2001.
- Chenyu Gu, Weicong Lin, Xinyi He, Lei Zhang, and Mingming Zhang. IMU-based motion capture system for rehabilitation applications: A systematic review. *Biomimetic Intelligence and Robotics*, 3(2):100097, June 2023. ISSN 2667-3797. doi: 10.1016/j.birob.2023.100097.
- Yu Guan, Chang-Tsun Li, and Sruti Das Choudhury. Robust gait recognition from extremely low frame-rate videos. In *Proceedings of the International Workshop on Biometrics and Forensics (IWBF)*, pages 1–4. IEEE, April 2013. doi: 10.1109/iwbf.2013.6547319.
- Liang-Yan Gui, Kevin Zhang, Yu-Xiong Wang, Xiaodan Liang, Jose M. F. Moura, and Manuela Veloso. Teaching robots to predict human motion. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, October 2018. doi: 10.1109/iros.2018.8594452.
- H. Gulgin, K. Hall, A. Luzadre, and E. Kayfish. 3d gait analysis with and without an orthopedic walking boot. *Gait & Posture*, 59:76–82, January 2018. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2017.09.024.
- Xianda Guo, Zheng Zhu, Tian Yang, Beibei Lin, Junjie Huang, Jiankang Deng, Guan Huang, Jie Zhou, and Jiwen Lu. Gait recognition in the wild: A large-scale benchmark and nas-based baseline. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4535–4552, June 2025a. ISSN 1939-3539. doi: 10.1109/tpami.2025.3546482.

- Yao Guo, Jianxin Yang, Yuxuan Liu, Xun Chen, and Guang-Zhong Yang. Detection and assessment of parkinson's disease based on gait analysis: A survey. *Frontiers in Aging Neuroscience*, 14, August 2022. ISSN 1663-4365. doi: 10.3389/fnagi.2022.916971.
- Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12):4338–4364, December 2021. ISSN 1939-3539. doi: 10.1109/tpami.2020.3005434.
- Zujin Guo, Size Wu, Zhongang Cai, Wei Li, and Chen Change Loy. Controllable human-centric keyframe interpolation with generative prior, 2025b.
- Reed D. Gurchiek, Rebecca H. Choquette, Bruce D. Beynnon, James R. Slauterbeck, Timothy W. Tourville, Michael J. Toth, and Ryan S. McGinnis. Open-source remote gait analysis: A post-surgery patient monitoring application. *Scientific Reports*, 9(1), November 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-54399-1.
- Iris Hagoort, Nicolas Vuillerme, Tibor Hortobágyi, and Claudine J.C. Lamoth. Age and walking conditions differently affect domains of gait. *Human Movement Science*, 89:103075, June 2023. ISSN 0167-9457. doi: 10.1016/j.humov.2023.103075.
- Mark Halaki and Karen Ginn. Normalization of emg signals: To normalize or not to normalize and what to normalize. *Computational Intelligence in Electromyography Analysis—A perspective on Current Applications and Future Challenges*, pages 175–194, 2012.
- Eni Halilaj, Apoorva Rajagopal, Madalina Fiterau, Jennifer L. Hicks, Trevor J. Hastie, and Scott L. Delp. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of Biomechanics*, 81:1–11, November 2018. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2018.09.009.
- Ju Han and Bir Bhanu. Individual recognition using Gait Energy Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006. doi: 10.1109/TPAMI.2006.38.
- SangUk Han and SangHyun Lee. A vision-based motion capture and recognition framework for behavior-based safety management. *Automation in Construction*, 35:131–141, 2013. doi: 10.1016/j.autcon.2013.05.001.
- Kamrul Hasan, Md. Zasim Uddin, Ausrukona Ray, Mahmudul Hasan, Fady Alnajjar, and Md Atiqur Rahman Ahad. Improving gait recognition through occlusion detection and silhouette sequence reconstruction. *IEEE Access*, 12:158597–158610, 2024. ISSN 2169-3536. doi: 10.1109/access.2024.3482430.
- Sora Hayashi, Kenshi Saho, Keitaro Shioiri, Masahiro Fujimoto, and Masao Masugi. Utilization of micro-doppler radar to classify gait patterns of young and elderly adults: An approach using a long short-term memory network. *Sensors*, 21(11):3643, May 2021. ISSN 1424-8220. doi: 10.3390/s21113643.
- Sumit Hazra, Acharya Aditya Pratap, Dattatreya Tripathy, and Anup Nandy. Novel data fusion strategy for human gait analysis using multiple kinect sensors. *Biomedical Signal Processing and Control*, 67:102512, May 2021. ISSN 1746-8094. doi: 10.1016/j.bspc.2021.102512.

- Yiwei He, Junping Zhang, Hongming Shan, and Liang Wang. Multi-task gans for view-specific feature learning in gait recognition. *IEEE Transactions on Information Forensics and Security*, 14(1):102–113, January 2019. ISSN 1556-6021. doi: 10.1109/tifs.2018.2844819.
- Tamika L. Heiden, David G. Lloyd, and Timothy R. Ackland. Knee joint kinematics, kinetics and muscle co-contraction in knee osteoarthritis patient gait. *Clinical Biomechanics*, 24(10): 833–841, December 2009. ISSN 0268-0033. doi: 10.1016/j.clinbiomech.2009.08.005.
- Dieter Heinrich, Antonie J. van den Bogert, Martin Mössner, and Werner Nachbauer. Model-based estimation of muscle and acl forces during turning maneuvers in alpine skiing. *Scientific Reports*, 13(1), June 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-35775-4.
- Allen Hill and Julie Nantel. The effects of arm swing amplitude and lower-limb asymmetry on gait stability. *PLOS ONE*, 14(12):e0218644, December 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0218644.
- Cherice N. Hill, Wornie Reed, Daniel Schmitt, Laura P. Sands, and Robin M. Queen. Racial differences in gait mechanics. *Journal of Biomechanics*, 112:110070, November 2020. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2020.110070.
- Susan J. Hillman, Benedict W. Stansfield, Alison M. Richardson, and James E. Robb. Development of temporal and distance parameters of gait in normal children. *Gait & Posture*, 29(1): 81–85, January 2009. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2008.06.012.
- Nur Amirah Hanisah Hisham, Ahmad Faiz Ahmad Nazri, June Madete, Lilik Herawati, and Jamaluddin Mahmud. Measuring ankle angle and analysis of walking gait using kinovea. In *Proceedings of the International Medical Device and Technology Conference*, pages 247–250. Universiti Teknologi Malaysia, 2017.
- Martin Hofmann, Shamik Sural, and Gerhard Rigoll. Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In *Proceedings of the 19th International Conferences on Computer Graphics, Visualization and Computer Vision*, WSCG, pages 99–104, 2011.
- Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. The TUM gait from audio, image and depth (GAID) database - multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195–206, 2014. doi: 10.1016/j.jvcir.2013.02.006.
- John H. Hollman, Eric M. McDade, and Ronald C. Petersen. Normative spatiotemporal gait parameters in older adults. *Gait & Posture*, 34(1):111–118, May 2011. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2011.03.024.
- Sarah Hood, Marshall K. Ishmael, Andrew Gunnell, K. B. Foreman, and Tommaso Lenzi. A kinematic and kinetic dataset of 18 above-knee amputees walking at various speeds. *Scientific Data*, 7(1), May 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0494-7.
- Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *Proceedings of the 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, August 2010. doi: 10.1109/icpr.2010.579.
- Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, August 1981. ISSN 0004-3702. doi: 10.1016/0004-3702(81)90024-2.

- Saihui Hou, Chao Fan, Chunshui Cao, Xu Liu, and Yongzhen Huang. A comprehensive study on the evaluation of silhouette-based gait recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 5(2):196–208, April 2023. ISSN 2637-6407. doi: 10.1109/tbiom.2022.3216857.
- John Howard, Vladimir V. Murashov, Brian D. Lowe, and Ming-Lun Lu. Industrial exoskeletons: Need for intervention effectiveness research. *American Journal of Industrial Medicine*, 63(3): 201–208, December 2019. ISSN 1097-0274. doi: 10.1002/ajim.23080.
- Jun-Wei Hsieh, Yung-Tai Hsu, H.-Y.M. Liao, and Chih-Chiang Chen. Video-based human movement analysis and its application to surveillance systems. *IEEE Transactions on Multimedia*, 10(3):372–384, April 2008. ISSN 1941-0077. doi: 10.1109/tmm.2008.917403.
- Katherine L. Hsieh, Linda Trinh, and Jacob J. Sosnoff. Gait variability is altered in cancer survivors with self-reported neuropathy. *Gait & Posture*, 72:206–210, July 2019. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2019.06.014.
- BingZhang Hu, Yu Guan, Yan Gao, Yang Long, Nicholas Lane, and Thomas Ploetz. Robust cross-view gait recognition with evidence: A discriminant gait gan (diggan) approach, 2018.
- Gibson Hu, Shoudong Huang, Liang Zhao, Alen Alempijevic, and Gamini Dissanayake. A robust RGB-D SLAM algorithm. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, oct 2012. doi: 10.1109/iro.2012.6386103.
- Yinlin Hu, Yunsong Li, and Rui Song. Robust interpolation of correspondences for large displacement optical flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4791–4799. IEEE, July 2017. doi: 10.1109/cvpr.2017.509.
- Panjian Huang, Yunjie Peng, Saihui Hou, Chunshui Cao, Xu Liu, Zhiqiang He, and Yongzhen Huang. *Occluded Gait Recognition with Mixture of Experts: An Action Detection Perspective*, pages 380–397. Springer Nature Switzerland, October 2024. ISBN 9783031726583. doi: 10.1007/978-3-031-72658-3_22.
- Zhen Huang, Dixiu Xue, Xu Shen, Xinmei Tian, Houqiang Li, Jianqiang Huang, and Xian-Sheng Hua. 3d local convolutional neural networks for gait recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, October 2021. doi: 10.1109/iccv48922.2021.01465.
- Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. *Real-Time Intermediate Flow Estimation for Video Frame Interpolation*, pages 624–642. Springer Nature Switzerland, 2022. ISBN 9783031197819. doi: 10.1007/978-3-031-19781-9_36.
- Abdul Aziz Hulleck, Dhanya Menoth Mohan, Nada Abdallah, Marwan El Rich, and Kinda Khalaf. Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. *Frontiers in Medical Technology*, 4, December 2022. ISSN 2673-3129. doi: 10.3389/fmedt.2022.901331.
- Inhwan Hwang, Geonho Cha, and Songhwa Oh. Multi-modal human action recognition using deep neural networks fusing image and inertial sensor data. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 278–283. IEEE, November 2017. doi: 10.1109/mfi.2017.8170441.

- Daisuke Imoto, Kenji Kurosawa, Masakatsu Honma, Ryo Yokota, Manato Hirabayashi, and Yoshinori Hawaii. Model-based interpolation for continuous human silhouette images by height-constraint assumption. In *Proceedings of the 4th International Conference on Vision, Image and Signal Processing*, ICVISP 2020, pages 1–11. ACM, December 2020. doi: 10.1145/3448823.3448835.
- Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, October 2012. ISSN 1556-6021. doi: 10.1109/tifs.2012.2204253.
- Yumi Iwashita, Ryosuke Baba, Koichi Ogawara, and Ryo Kurazume. Person identification from spatio-temporal 3d gait. In *Proceedings of the International Conference on Emerging Security Technologies*. IEEE, 2010. doi: 10.1109/est.2010.19.
- Yumi Iwashita, Ryo Kurazume, and Adrian Stoica. Gait identification using invisible shadows: Robustness to appearance changes. In *Proceedings of the 5th International Conference on Emerging Security Technologies*, pages 34–39. IEEE, September 2014a. doi: 10.1109/est.2014.18.
- Yumi Iwashita, Koichi Ogawara, and Ryo Kurazume. Identification of people walking along curved trajectories. *Pattern Recognition Letters*, 48:60–69, 2014b. doi: 10.1016/j.patrec.2014.04.004.
- Paul Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579, 1901.
- Diego Jaén-Carrillo, Felipe García-Pinillos, Antonio Cartón-Llorente, Alejandro Jesús Almenar-Arasanz, José Antonio Bustillo-Pelayo, and Luis E. Roche-Seruendo. Test–retest reliability of the OptoGait system for the analysis of spatiotemporal running gait parameters and lower body stiffness in healthy adults. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 234(2):154–161, February 2020. ISSN 1754-338X. doi: 10.1177/1754337119898353.
- Roohollah Jahanmahin, Sara Masoud, Jeremy Rickli, and Ana Djuric. Human-robot interactions in manufacturing: A survey of human behavior modeling. *Robotics and Computer-Integrated Manufacturing*, 78:102404, December 2022. ISSN 0736-5845. doi: 10.1016/j.rcim.2022.102404.
- Anil Jain, Karthik Nandakumar, and Arun Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, December 2005. ISSN 0031-3203. doi: 10.1016/j.patcog.2005.01.012.
- Eric Jang, Alex Irpan, Mohi Khansari, Daniel Kappler, Frederik Ebert, Corey Lynch, Sergey Levine, and Chelsea Finn. BC-z: Zero-shot task generalization with robotic imitation learning. In Aleksandra Faust, David Hsu, and Gerhard Neumann, editors, *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pages 991–1002. PMLR, 08–11 Nov 2022. URL <https://proceedings.mlr.press/v164/jang22a.html>.

- Mehran Javanmardi, Ricardo Bigolin Lanfredi, Mujdat Cetin, and Tolga Tasdizen. Image segmentation by deep learning of disjunctive normal shape model shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 488–4887. IEEE, June 2018. doi: 10.1109/cvprw.2018.00078.
- Eun-tae Jeon and Hwi-young Cho. A novel method for gait analysis on center of pressure excursion based on a pressure-sensitive mat. *International Journal of Environmental Research and Public Health*, 17(21):7845, October 2020. ISSN 1660-4601. doi: 10.3390/ijerph17217845.
- Huaizu Jiang, Deqing Sun, Varan Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9000–9008. IEEE, June 2018. doi: 10.1109/cvpr.2018.00938.
- Weiguang Jiang, Lieyun Ding, and Cheng Zhou. Cyber physical system for safety management in smart construction site. *Engineering, Construction and Architectural Management*, 28(3): 788–808, April 2020. ISSN 0969-9988. doi: 10.1108/ecam-10-2019-0578.
- Pavel Jicinsky and Jan Mares. Development of motion analysis tool for orthopedic diseases. In *Proceedings of the International Conference on Applied Electronics (AE)*. IEEE, sep 2021. doi: 10.23919/ae51540.2021.9542895.
- Sheng Jin, Lumin Xu, Jin Xu, Can Wang, Wentao Liu, Chen Qian, Wanli Ouyang, and Ping Luo. *Whole-Body Human Pose Estimation in the Wild*, pages 196–214. Springer International Publishing, 2020. ISBN 9783030585457. doi: 10.1007/978-3-030-58545-7_12.
- Xin Jin, Longhai Wu, Jie Chen, Youxin Chen, Jayoon Koo, and Cheul-Hee Hahm. A unified pyramid recurrent network for video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2023. doi: 10.1109/cvpr52729.2023.00158.
- Rishabh Bajpai and Deepak Joshi and Deepak Joshi. MoveNet: A deep neural network for joint profile prediction across variable walking speeds and slopes. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021. doi: 10.1109/tim.2021.3073720.
- Kooksung Jun, Deok-Won Lee, Kyoobin Lee, Sanghyub Lee, and Mun Sang Kim. Feature extraction using an rnn autoencoder for skeleton-based abnormal gait recognition. *IEEE Access*, 8: 19196–19207, 2020. ISSN 2169-3536. doi: 10.1109/access.2020.2967845.
- M. Humayun Kabir, M. Robiul Hoque, Keshav Thapa, and Sung-Hyun Yang. Two-layer hidden markov model for human activity recognition in home environments. *International Journal of Distributed Sensor Networks*, 2016:1–12, 2016. doi: 10.1155/2016/4560365.
- Bhaben Kalita, Jyotindra Narayan, and Santosha Kumar Dwivedy. Development of active lower limb robotic-based orthosis and exoskeleton devices: A systematic review. *International Journal of Social Robotics*, 13(4):775–793, 2020. doi: 10.1007/s12369-020-00662-9.
- Alexander Kapitanov, Karina Kvanchiani, Alexander Nagaev, Roman Kraynov, and Andrei Makhliarchuk. HaGRID – HAnd Gesture Recognition Image Dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4572–4581, January 2024.

- Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988. ISSN 1573-1405. doi: 10.1007/bf00133570.
- Andrea D. Kassay, Baraa Daher, and E. A. Lalone. An analysis of wrist and forearm range of motion using the dartfish motion analysis system. *Journal of Hand Therapy*, 34(4):604–611, October 2021. ISSN 0894-1130. doi: 10.1016/j.jht.2020.09.002.
- A. Katashev, E. Shishlova, and V. Vendina. *On Evaluation of Shoulder Roundness by Use of Single Camera Photogrammetry*, pages 129–132. Springer International Publishing, 2014. ISBN 9783319008462. doi: 10.1007/978-3-319-00846-2_32.
- Rateb Katmah, Aamna Al Shehhi, Herbert F. Jelinek, Abdul Aziz Hulleck, and Kinda Khalaf. A systematic review of gait analysis in the context of multimodal sensing fusion and ai. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:4189–4202, 2023. ISSN 1558-0210. doi: 10.1109/tnsre.2023.3325215.
- Shian-Ru Ke, Hoang Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi. A review on video-based human activity recognition. *Computers*, 2(2):88–131, 2013. ISSN 2073-431X. doi: 10.3390/computers2020088.
- Alex Kendall, Matthew Grimes, and Roberto Cipolla. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015. doi: 10.1109/iccv.2015.336.
- Mehdi Khachani, Rahman Davoodi, and Gerald E. Loeb. Musculo-skeletal modeling software (msms) for biomechanics and virtual rehabilitation. In *Proceedings of the American Society of Biomechanics Conference*, 2007.
- Anouar Ben Khalifa. Infrastructure to vehicle multi-view pedestrian detection dataset (i2v-mvdp), 2020. URL <https://sites.google.com/site/benkhalfaanouar1/6-datasets#h.8pd3p2j80ssx>.
- Anouar Ben Khalifa, Ihsen Alouani, Mohamed Ali Mahjoub, and Atika Rivenq. A novel multi-view pedestrian detection database for collaborative intelligent transportation systems. *Future Generation Computer Systems*, 113:506–527, 2020. doi: 10.1016/j.future.2020.07.025.
- Md. Khaliluzzaman, Ashraf Uddin, Kaushik Deb, and Md Junayed Hasan. Person recognition based on deep gait: A survey. *Sensors*, 23(10):4875, May 2023. ISSN 1424-8220. doi: 10.3390/s23104875.
- Atif Khan, Febin Moideen, Juan Lopez, Wai L. Khoo, and Zhigang Zhu. KinDectect: Kinect detecting objects. In *Lecture Notes in Computer Science*, pages 588–595. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-31534-3_86.
- Muhammad Hassan Khan, Muhammad Shahid Farid, and Marcin Grzegorzec. Using a generic model for codebook-based gait recognition algorithms. In *Proceedings of the International Workshop on Biometrics and Forensics*, pages 1–7. IEEE, June 2018. doi: 10.1109/iwbf.2018.8401551.
- Muhammad Hassan Khan, Martin Zöller, Muhammad Shahid Farid, and Marcin Grzegorzec. Marker-based movement analysis of human body parts in therapeutic procedure. *Sensors*, 20(11):3312, June 2020. ISSN 1424-8220. doi: 10.3390/s20113312.

- Muhammad Hassan Khan, Muhammad Shahid Farid, and Marcin Grzegorzec. Vision-based approaches towards person identification using gait. *Computer Science Review*, 42:100432, November 2021. ISSN 1574-0137. doi: 10.1016/j.cosrev.2021.100432.
- Ashutosh Kharb, Vipin Saini, Y. K. Jain, and Surender Dhiman. A review of gait cycle and its parameters. *International Journal of Computational Engineering & Management*, 13:78–83, 2011. ISSN 2230-7893.
- Preeti Khera and Neelesh Kumar. Role of machine learning in gait analysis: a review. *Journal of Medical Engineering & Technology*, 44(8):441–467, October 2020. ISSN 1464-522X. doi: 10.1080/03091902.2020.1822940.
- Margarita Khokhlova, Cyrille Migniot, Alexey Morozov, Olga Sushkova, and Albert Dipanda. Normal and pathological gait classification LSTM model. *Artificial Intelligence in Medicine*, 94:54–66, March 2019. ISSN 0933-3657. doi: 10.1016/j.artmed.2018.12.007.
- Kourosh Khoshelham and Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012. doi: 10.3390/s120201437.
- Łukasz Kidziński, Scott Delp, and Michael Schwartz. Automatic real-time gait event detection in children using deep neural networks. *PLOS ONE*, 14(1), January 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0211466.
- Łukasz Kidziński, Bryan Yang, Jennifer L. Hicks, Apoorva Rajagopal, Scott L. Delp, and Michael H. Schwartz. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications*, 11(1), August 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-17807-z.
- Hyun-Kyung Kim and Yanxin Zhang. Estimation of lumbar spinal loading and trunk muscle forces during asymmetric lifting tasks: application of whole-body musculoskeletal modelling in opensim. *Ergonomics*, 60(4):563–576, June 2016. ISSN 1366-5847. doi: 10.1080/00140139.2016.1191679.
- Hyungtai Kim, Yun-Hee Kim, Seung-Jong Kim, and Mun-Taek Choi. Pathological gait clustering in post-stroke patients using motion capture data. *Gait & Posture*, 94:210–216, May 2022. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2022.03.007.
- Chris Kirtley. *Clinical gait analysis*. Elsevier, Edinburgh, 2006. ISBN 9780702036712.
- J. F. P. Kooij, M. C. Liem, J. D. Krijnders, T. C. Andringa, and D. M. Gavrilu. Multi-modal human aggression detection. *Computer Vision and Image Understanding*, 144:106–120, March 2016. ISSN 1077-3142. doi: 10.1016/j.cviu.2015.06.009.
- Navleen Kour, Sunanda, and Sakshi Arora. Computer-vision based diagnosis of parkinson’s disease via gait: A survey. *IEEE Access*, 7:156620–156645, 2019. ISSN 2169-3536. doi: 10.1109/access.2019.2949744.
- Navleen Kour, Sunanda, and Sakshi Arora. *A Vision-Based Gait Dataset for Knee Osteoarthritis and Parkinson’s Disease Analysis with Severity Levels*, pages 303–317. Springer Singapore, August 2021. ISBN 9789811630712. doi: 10.1007/978-981-16-3071-2_26.
- Bjorn Kruger, Anna Vogelee, Tobias Willig, Angela Yao, Reinhard Klein, and Andreas Weber. Efficient unsupervised temporal segmentation of motion data. *IEEE Transactions on Multimedia*, 19(4):797–812, April 2017. ISSN 1941-0077. doi: 10.1109/tmm.2016.2635030.

- Worapan Kusakunniran. Review of gait recognition approaches and their challenges on view changes. *IET Biometrics*, 9(6):238–250, September 2020. ISSN 2047-4946. doi: 10.1049/iet-bmt.2020.0103.
- Bogdan Kwolek and Michal Kepski. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer Methods and Programs in Biomedicine*, 117(3): 489–501, December 2014. ISSN 0169-2607. doi: 10.1016/j.cmpb.2014.09.005.
- Toby H. W. Lam and Raymond S. T. Lee. A new representation for human gait recognition: Motion silhouettes image (MSI). In *Advances in Biometrics*, pages 612–618. Springer Berlin Heidelberg, 2005. doi: 10.1007/11608288_81.
- Toby H. W. Lam, Raymond S. T. Lee, and David Zhang. Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40(9):2563–2573, 2007. ISSN 0031-3203. doi: <http://dx.doi.org/10.1016/j.patcog.2006.11.014>.
- Toby H. W. Lam, K. H. Cheung, and James N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44(4):973–987, 2011. ISSN 0031-3203. doi: <http://dx.doi.org/10.1016/j.patcog.2010.10.011>.
- Winnie W. T. Lam, Yuk Ming Tang, and Kenneth N. K. Fong. A systematic review of the applications of markerless motion capture (mmc) technology for clinical measurement in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 20(1), May 2023. ISSN 1743-0003. doi: 10.1186/s12984-023-01186-9.
- Fabrizio Lamberti, Gianluca Paravati, Valentina Gatteschi, Alberto Cannavo, and Paolo Montuschi. Virtual character animation based on affordable motion capture and reconfigurable tangible interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 24(5):1742–1755, May 2018. ISSN 1077-2626. doi: 10.1109/tvcg.2017.2690433.
- Nicola Lamberti, Fabio Manfredini, Luc Oscar Lissom, Susanna Lavezzi, Nino Basaglia, and Sofia Straudi. Beneficial effects of robot-assisted gait training on functional recovery in women after stroke: A cohort study. *Medicina*, 57(11):1200, November 2021. ISSN 1648-9144. doi: 10.3390/medicina57111200.
- Kathleen A. Lamkin-Kennard and Marko B. Popovic. *Sensors: Natural and Synthetic Sensors*, pages 81–107. Elsevier, 2019. ISBN 9780128129395. doi: 10.1016/b978-0-12-812939-5.00004-5.
- Steven J. Landry. *Human Factors and Ergonomics in Aviation*. 2021. doi: 10.1002/9781119636113.ch55. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148141471&doi=10.1002%2f9781119636113.ch55&partnerID=40&md5=8b9b63b118406823dc3a297f3d192bb1>.
- Jorge Latorre, Carolina Colomer, Mariano Alcañiz, and Roberto Llorens. Gait analysis with the kinect v2: normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *Journal of NeuroEngineering and Rehabilitation*, 16(1), jul 2019. doi: 10.1186/s12984-019-0568-y.
- Robin B. Layton, Neil Messenger, and Todd D. Stewart. Analysis of hip joint cross-shear under variable activities using a novel virtual joint model within visual3d. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 235(10):1197–1204, June 2021. ISSN 2041-3033. doi: 10.1177/09544119211025869.

- Jaehyun Lee, Hyosung Joo, Junglyeon Lee, and Youngjoon Chee. Automatic classification of squat posture using inertial sensors: Deep learning approach. *Sensors*, 20(2):361, 2020. doi: 10.3390/s20020361.
- Tracey K. M. Lee, Mohammed Belkhatir, and Saeid Sanei. A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia Tools and Applications*, 72(3): 2833–2869, July 2013. ISSN 1573-7721. doi: 10.1007/s11042-013-1574-x.
- T.M. Lehmann, C. Gonner, and K. Spitzer. Survey: interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075, 1999. ISSN 0278-0062. doi: 10.1109/42.816070.
- Gustavo Leporace, Leonardo Metsavaht, Gabriel Zeitoune, Thiago Marinho, Tainá Oliveira, Glauber Ribeiro Pereira, Liszt Palmeira De Oliveira, and Luiz Alberto Batista. Use of spatiotemporal gait parameters to determine return to sports after acl reconstruction. *Acta Ortopédica Brasileira*, 24(2):73–76, April 2016. ISSN 1413-7852. doi: 10.1590/1413-785220162402147450.
- Congcong Li, Bin Wang, Yifan Li, and Bo Liu. A lightweight pathological gait recognition approach based on a new gait template in side-view and improved attention mechanism. *Sensors*, 24(17):5574, August 2024a. ISSN 1424-8220. doi: 10.3390/s24175574.
- Guodong Li, Lijun Guo, Rong Zhang, Jiangbo Qian, and Shangce Gao. Transgait: Multimodal-based gait recognition with set transformer. *Applied Intelligence*, 53(2):1535–1547, April 2022a. ISSN 1573-7497. doi: 10.1007/s10489-022-03543-y.
- Jin Li, Jishuo Yan, Guangxu Li, Liyuan Wang, and Fan Yang. RGB-d-based hand gesture recognition for letters expression. In *Proceedings of the 4th International Conference on Biomedical Signal and Image Processing, ICBIP '2019*. ACM Press, 2019. doi: 10.1145/3354031.3354044.
- Jing Li, Weisheng Liang, Xiyan Yin, Jun Li, and Weizheng Guan. Multimodal gait abnormality recognition using a convolutional neural network–bidirectional long short-term memory (cnn-bilstm) network based on multi-sensor data fusion. *Sensors*, 23(22):9101, November 2023a. ISSN 1424-8220. doi: 10.3390/s23229101.
- Na Li and Xinbo Zhao. A benchmark for gait recognition under occlusion collected by multi-kinect sdas. *arXiv preprint arXiv:2107.08990*, 2021.
- Na Li and Xinbo Zhao. A multi-modal dataset for gait recognition under occlusion. *Applied Intelligence*, 53(2):1517–1534, April 2022. ISSN 1573-7497. doi: 10.1007/s10489-022-03474-8.
- Na Li, Xinbo Zhao, and Chong Ma. A model-based gait recognition method based on gait graph convolutional networks and joints relationship pyramid mapping. *arXiv preprint arXiv:2005.08625*, 2020a.
- Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Yan. Crowded scene analysis: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3):367–386, 2015. doi: 10.1109/tcsvt.2014.2358029.
- Tianhao Li, Weizhi Ma, Yujia Zheng, Xinchao Fan, Guangcan Yang, Lijun Wang, and Zhengping Li. A survey on gait recognition against occlusion: taxonomy, dataset and methodology. *PeerJ Computer Science*, 10:e2602, December 2024b. ISSN 2376-5992. doi: 10.7717/peerj-cs.2602.

- Weijia Li, Saihui Hou, Chunjie Zhang, Chunshui Cao, Xu Liu, Yongzhen Huang, and Yao Zhao. An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13824–13833. IEEE, June 2023b. doi: 10.1109/cvpr52729.2023.01328.
- Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):18–32, jan 2014. doi: 10.1109/tpami.2013.111.
- Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. Gait recognition via semi-supervised disentangled representation learning to identity and covariate features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13306–13316. IEEE, June 2020b. doi: 10.1109/cvpr42600.2020.01332.
- Xiang Li, Yasushi Makihara, Chi Xu, and Yasushi Yagi. Multi-view large population gait database with human meshes and its performance evaluation. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(2):234–248, April 2022b. ISSN 2637-6407. doi: 10.1109/tbiom.2022.3174559.
- Xiaogang Li, Tiantian Pang, Weixiang Liu, and Tianfu Wang. Fall detection for elderly person care using convolutional neural networks. In *Proceedings of the 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, oct 2017. doi: 10.1109/cisp-bmei.2017.8302004.
- Xiuxian Li, Min Meng, Yiguang Hong, and Jie Chen. A survey of decision making in adversarial games. *Science China Information Sciences*, 67(4), February 2024c. ISSN 1869-1919. doi: 10.1007/s11432-022-3777-y.
- Junhao Liang, Chao Fan, Saihui Hou, Chuanfu Shen, Yongzhen Huang, and Shiqi Yu. *GaitEdge: Beyond Plain End-to-End Gait Recognition for Better Practicality*, pages 375–390. Springer Nature Switzerland, 2022. ISBN 9783031200656. doi: 10.1007/978-3-031-20065-6_22.
- Chen-Chieh Liao, Dong-Hyun Hwang, and Hideki Koike. AI golf: Golf swing analysis tool for self-training. *IEEE Access*, 10:106286–106295, 2022a. doi: 10.1109/access.2022.3210261.
- Rijun Liao, Chunshui Cao, Edel B. Garcia, Shiqi Yu, and Yongzhen Huang. *Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations*, pages 474–483. Springer International Publishing, 2017. ISBN 9783319699233. doi: 10.1007/978-3-319-69923-3_51.
- Rijun Liao, Shiqi Yu, Weizhi An, and Yongzhen Huang. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98:107069, February 2020. ISSN 0031-3203. doi: 10.1016/j.patcog.2019.107069.
- Rijun Liao, Zhu Li, Shuvra S. Bhattacharyya, and George York. PoseMapGait: A model-based gait recognition method with pose estimation maps and graph convolutional networks. *Neurocomputing*, 501:514–528, August 2022b. ISSN 0925-2312. doi: 10.1016/j.neucom.2022.06.048.
- Karin Lienhard, David Schneider, and Nicola A. Maffioletti. Validity of the optogait photoelectric system for the assessment of spatiotemporal gait parameters. *Medical Engineering & Physics*, 35(4):500–504, April 2013. ISSN 1350-4533. doi: 10.1016/j.medengphy.2012.06.015.

- Beibei Lin, Shunli Zhang, and Xin Yu. Gait recognition via effective global-local feature representation and local temporal aggregation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14628–14636. IEEE, October 2021. doi: 10.1109/iccv48922.2021.01438.
- Yiqiao Lin, Xueyan Jiao, and Lei Zhao. Detection of 3d human posture based on improved mediapipe. *Journal of Computer and Communications*, 11(02):102–121, 2023. ISSN 2327-5227. doi: 10.4236/jcc.2023.112008.
- Mary Emily Littrell, Young-Hui Chang, and Brian P. Selgrade. Development and assessment of a low-cost clinical gait analysis system. *Journal of Applied Biomechanics*, 34(6):503–508, December 2018. ISSN 1543-2688. doi: 10.1123/jab.2017-0370.
- Hongyi Liu and Lihui Wang. Human motion prediction for human-robot collaboration. *Journal of Manufacturing Systems*, 44:287–294, July 2017. ISSN 0278-6125. doi: 10.1016/j.jmsy.2017.04.009.
- Jianyi Liu and Nanning Zheng. Gait history image: A novel temporal template for gait recognition. In *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2007. doi: 10.1109/icme.2007.4284737.
- Jingbin Liu, Gege Huang, Juha Hyypä, Jing Li, Xiaodong Gong, and Xiaofan Jiang. A survey on location and motion tracking technologies, methodologies and applications in precision sports. *Expert Systems with Applications*, 229:120492, November 2023. ISSN 0957-4174. doi: 10.1016/j.eswa.2023.120492.
- Nai Liu and Bichuan Liu. Research on knee sports injury classification based on medical images and kwon3d software. *Revista Brasileira de Medicina do Esporte*, 27(5):456–459, September 2021. ISSN 1517-8692. doi: 10.1590/1517-8692202127042021_0099.
- Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection - a new baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2018a. doi: 10.1109/cvpr.2018.00684.
- Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Shanghaitech campus dataset (anomaly detection), 2018b. URL https://svip-lab.github.io/dataset/campus_dataset.html.
- Wu Liu, Cheng Zhang, Huadong Ma, and Shuangqun Li. Learning efficient spatial-temporal gait features with deep learning for human identification. *Neuroinformatics*, 16(3–4):457–471, February 2018c. ISSN 1559-0089. doi: 10.1007/s12021-018-9362-4.
- Xiaokai Liu, Zhaoyang You, Yuxiang He, Sheng Bi, and Jie Wang. Symmetry-driven hyper feature GCN for skeleton-based gait recognition. *Pattern Recognition*, 125:108520, May 2022. ISSN 0031-3203. doi: 10.1016/j.patcog.2022.108520.
- Yu-Tao Liu, Yong-An Zhang, and Ming Zeng. Adaptive global time sequence averaging method using dynamic time warping. *IEEE Transactions on Signal Processing*, 67(8):2129–2142, 2019. doi: 10.1109/tsp.2019.2897958.
- Zongyi Liu and Sudeep Sarkar. Simplest representation yet for gait recognition: averaged silhouette. In *Proceedings of the 17th International Conference on Pattern Recognition, ICPR*. IEEE, 2004. doi: 10.1109/icpr.2004.1333741.

- Luca Longoni, Roberto Brunati, Patrizio Sale, Roberto Casale, Gianpaolo Ronconi, and Giorgio Ferriero. Smartphone applications validated for joint angle measurement: a systematic review. *International Journal of Rehabilitation Research*, 42(1):11–19, March 2019. ISSN 0342-5282. doi: 10.1097/mrr.0000000000000332.
- Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. *SMPL: A Skinned Multi-Person Linear Model*, pages 851–866. ACM, August 2023. ISBN 9798400708978. doi: 10.1145/3596711.3596800.
- David López-Fernández, Francisco J. Madrid-Cuevas, Ángel Carmona-Poyato, Manuel J. Marín-Jiménez, and Rafael Muñoz Salinas. The ava multi-view dataset for gait recognition. In Pier Luigi Mazzeo, Paolo Spagnolo, and Thomas B. Moeslund, editors, *Activity Monitoring by Multiple Distributed Sensing*, Lecture Notes in Computer Science, pages 26–39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3.
- Joao Loureiro and Paulo Lobato Correia. Using a skeleton gait energy image for pathological gait classification. In *Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 410–414, Los Alamitos, CA, USA, nov 2020. IEEE Computer Society. doi: 10.1109/fg47880.2020.00064. URL <https://doi.ieeecomputersociety.org/10.1109/FG47880.2020.00064>.
- Wan Shi Low, Chow Khuen Chan, Joon Huang Chuah, Yee Kai Tee, Yan Chai Hum, Maheza Irna Mohd Salim, and Khin Wee Lai. A review of machine learning network in human motion biomechanics. *Journal of Grid Computing*, 20(1), December 2021. ISSN 1572-9184. doi: 10.1007/s10723-021-09595-7.
- Liyang Lu, Ruizheng Wu, Huaijia Lin, Jiangbo Lu, and Jiaya Jia. Video frame interpolation with transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3532–3542, June 2022a.
- Xiaoyan Lu, Xinde Li, Weijie Sheng, and Shuzhi Sam Ge. Long-term person re-identification based on appearance and gait feature fusion under covariate changes. *Processes*, 10(4):770, April 2022b. ISSN 2227-9717. doi: 10.3390/pr10040770.
- Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building perception pipelines, 2019.
- Trevor J. Lujan, Spencer P. Lake, Timothy A. Plaizier, Benjamin J. Ellis, and Jeffrey A. Weiss. Simultaneous measurement of three-dimensional joint kinematics and ligament strains with optical methods. *Journal of Biomechanical Engineering*, 127(1):193–197, February 2005. ISSN 1528-8951. doi: 10.1115/1.1835365.
- Jian Luo and Tardi Tjahjadi. View and clothing invariant gait recognition via 3D human semantic folding. *IEEE Access*, 8:100365–100383, 2020. ISSN 2169-3536. doi: 10.1109/access.2020.2997814.
- Jing Luo, Chunyuan Zi, Jianliang Zhang, and Yue Liu. Gait recognition using gei and curvelet. *Opto-Electronic Engineering*, 44(4), 2017.

- Senzi Luo and Jiayin Xiao. Self-similarity attention module for skeleton-based fine-grained action recognition. In *Proceedings of the 4th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, pages 143–147. IEEE, August 2023. doi: 10.1109/ichci58871.2023.10278014.
- Noel Lythgo, Cameron Wilson, and Mary Galea. Basic gait and symmetry measures for primary school-aged children and young adults whilst walking barefoot and with shoes. *Gait amp; Posture*, 30(4):502–506, November 2009. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2009.07.119.
- JoyC MacDermid, KathrynE Sinden, ThomasR Jenkyn, GeorgeS Athwal, TrevorB Birmingham, and Leenesh Khadilkar. An analysis of functional shoulder movements during task performance using dartfish movement analysis software. *International Journal of Shoulder Surgery*, 8(1):1, 2014. ISSN 0973-6042. doi: 10.4103/0973-6042.131847.
- Ioana Macoveciuc, Carolyn J. Rando, and Hervé Borrión. Forensic gait analysis and recognition: Standards of evidence admissibility. *Journal of Forensic Sciences*, 64(5):1294–1303, February 2019. ISSN 1556-4029. doi: 10.1111/1556-4029.14036.
- Srikrishnaraja Mahadas, Kausalendra Mahadas, and George K. Hung. Biomechanics of the golf swing using opensim. *Computers in Biology and Medicine*, 105:39–45, February 2019. ISSN 0010-4825. doi: 10.1016/j.combiomed.2018.12.002.
- Sharmin Majumder and Nasser Kehtarnavaz. Vision and inertial sensing fusion for human action recognition: A review. *IEEE Sensors Journal*, 21(3):2454–2467, February 2021. ISSN 2379-9153. doi: 10.1109/jsen.2020.3022326.
- Yasushi Makihara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. *Gait Recognition Using a View Transformation Model in the Frequency Domain*, pages 151–163. Springer Berlin Heidelberg, 2006. ISBN 9783540338376. doi: 10.1007/11744078_12.
- Yasushi Makihara, Atsushi Mori, and Yasushi Yagi. *Temporal Super Resolution from a Single Quasi-periodic Image Sequence Based on Phase Registration*, pages 107–120. Springer Berlin Heidelberg, 2011. ISBN 9783642193156. doi: 10.1007/978-3-642-19315-6_9.
- Yasushi Makihara, Hidetoshi Mannami, Akira Tsuji, Md. Altab Hossain, Kazushige Sugiura, Atsushi Mori, and Yasushi Yagi. The ou-isir gait database comprising the treadmill dataset. *IPSP Transactions on Computer Vision and Applications*, 4:53–62, 2012. ISSN 1882-6695. doi: 10.2197/ipsjtcva.4.53.
- Andrii Maksai, Xinchao Wang, and Pascal Fua. What players do with the ball: A physically constrained interaction modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’2016)*. IEEE, jun 2016. doi: 10.1109/cvpr.2016.111.
- Appukutty Manickam and Matthew D Gardiner. Gait assessment in general practice. *Australian Journal of General Practice*, 50(11):801–806, November 2021. ISSN 2208-7958. doi: 10.31128/ajgp-12-20-5777.
- Al Mansur, Yasushi Makihara, Rasyid Aqmar, and Yasushi Yagi. Gait recognition under speed transition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2528. IEEE, June 2014. doi: 10.1109/cvpr.2014.323.

- Manuel J. Marin-Jimenez, Francisco M. Castro, Ruben Delgado-Escano, Vicky Kalogeiton, and Nicolas Guil. Ugaitnet: Multimodal gait recognition with missing input modalities. *IEEE Transactions on Information Forensics and Security*, 16:5452–5462, 2021. ISSN 1556-6021. doi: 10.1109/tifs.2021.3132579.
- Joel Marouvo, Filipa Sousa, Orlando Fernandes, Maria António Castro, and Szczepan Paszkiel. Gait kinematics analysis of flatfoot adults. *Applied Sciences*, 11(15):7077, July 2021. ISSN 2076-3417. doi: 10.3390/app11157077.
- J. L. Marroquin, E. A. Santana, and S. Botello. Hidden markov measure field models for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1380–1387, November 2003. ISSN 0162-8828. doi: 10.1109/tpami.2003.1240112.
- Raúl Martín-Félez and Tao Xiang. Gait recognition by ranking. In *Computer Vision – ECCV 2012*, pages 328–341. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-33718-5_24.
- Elena Martini, Tommaso Fiumalbi, Filippo Dell’Agnello, Zoran Ivanić, Marko Munih, Nicola Vitiello, and Simona Crea. Pressure-sensitive insoles for real-time gait-related applications. *Sensors*, 20(5):1448, March 2020. ISSN 1424-8220. doi: 10.3390/s20051448.
- Eloise Matheson, Riccardo Minto, Emanuele G. G. Zampieri, Maurizio Faccio, and Giulio Rosati. Human–robot collaboration in manufacturing applications: A review. *Robotics*, 8(4):100, December 2019. ISSN 2218-6581. doi: 10.3390/robotics8040100.
- Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289, August 2018. ISSN 1546-1726. doi: 10.1038/s41593-018-0209-y.
- Jana Mattheus, Hans Grobler, and Adnan M. Abu-Mahfouz. A review of motion segmentation: Approaches and major challenges. In *Proceedings of the 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pages 1–8. IEEE, November 2020. doi: 10.1109/imitec50163.2020.9334076.
- Pauline Maurice, Adrien Malaisé, Clélie Amiot, Nicolas Paris, Guy-Junior Richard, Olivier Rochel, and Serena Ivaldi. Human movement and ergonomics: An industry-oriented dataset for collaborative robotics. *The International Journal of Robotics Research*, 38(14):1529–1537, 2019a. doi: 10.1177/0278364919882089.
- Pauline Maurice, Adrien Malaisé, Serena Ivaldi, Olivier Rochel, Clelie Amiot, Nicolas Paris, Guy-Junior Richard, and Lars Fritzsche. AnDyDataset, 2019b. URL <https://andydataset.loria.fr/>.
- Matthew P. Mavor, Gwyneth B. Ross, Allison L. Clouthier, Thomas Karakolis, and Ryan B. Graham. Validation of an IMU suit for military-based tasks. *Sensors*, 20(15):4280, 2020. doi: 10.3390/s20154280.
- Pamela McCauley, Susan Gaines, Fatina Gammoh, and Shanon Woode. *A Comparison of Software Tools for Occupational Biomechanics and Ergonomic Research*. InTech, April 2012. ISBN 9789535106012. doi: 10.5772/39201.

- Tess G. McGuire, Caitlyn T. Picard, Rose Marie Ward, Dean L. Smith, Young-Hoo Kwon, and Mark S. Walsh. The effects of target distance on kinematic sequence of the short game in male collegiate golfers. *Sports Biomechanics*, 23(12):3290–3300, November 2024. ISSN 1752-6116. doi: 10.1080/14763141.2024.2423282.
- Renato S. Melo, Afonso Rodrigues Tavares-Netto, Alexandre Delgado, Carine Carolina Wiesiolek, Karla Mônica Ferraz, and Rosalie Barreto Belian. Does the practice of sports or recreational activities improve the balance and gait of children and adolescents with sensorineural hearing loss? a systematic review. *Gait & Posture*, 77:144–155, March 2020. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2020.02.001.
- Lin Meng, Jun Pang, Ziyao Wang, Rui Xu, and Dong Ming. The role of surface electromyography in data fusion with inertial sensors to enhance locomotion recognition and prediction. *Sensors*, 21(18):6291, September 2021. ISSN 1424-8220. doi: 10.3390/s21186291.
- Benjamin F. Mentiplay, Luke G. Perraton, Kelly J. Bower, Yong-Hao Pua, Rebekah McGaw, Sophie Heywood, and Ross A. Clark. Gait assessment using the microsoft xbox one kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables. *Journal of Biomechanics*, 48(10):2166–2170, July 2015. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2015.05.021.
- Benjamin F. Mentiplay, Gavin Williams, Dawn Tan, Brooke Adair, Yong-Hao Pua, Chek Wai Bok, Kelly J. Bower, Michael H. Cole, Yee Sien Ng, Lek Syn Lim, and Ross A. Clark. Gait velocity and joint power generation after stroke: Contribution of strength and balance. *American Journal of Physical Medicine and Rehabilitation*, 98(10):841–849, December 2018. ISSN 0894-9115. doi: 10.1097/phm.0000000000001122.
- Hylton B. Menz, Mark D. Latt, Anne Tiedemann, Marcella Mun San Kwan, and Stephen R. Lord. Reliability of the GAITRite walkway system for the quantification of temporo-spatial parameters of gait in young and older people. *Gait & Posture*, 20(1):20–25, August 2004. ISSN 0966-6362. doi: 10.1016/s0966-6362(03)00068-7.
- Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, October 2016. doi: 10.1109/3dv.2016.79.
- Arnold Milstein and Eric J. Topol. Computer vision’s potential to improve health care. *The Lancet*, 395(10236):1537, May 2020. ISSN 0140-6736. doi: 10.1016/s0140-6736(20)31090-4.
- Anat Mirelman, Paolo Bonato, Richard Camicioli, Terry D. Ellis, Nir Giladi, Jamie L. Hamilton, Chris J. Hass, Jeffrey M. Hausdorff, Elisa Pelosin, and Quincy J. Almeida. Gait impairments in Parkinson’s disease. *The Lancet Neurology*, 18(7):697–708, July 2019. ISSN 1474-4422. doi: 10.1016/s1474-4422(19)30044-4.
- Shiwei Mo and Daniel H. K. Chow. Accuracy of three methods in gait event detection during overground running. *Gait & Posture*, 59:93–98, jan 2018. doi: 10.1016/j.gaitpost.2017.10.009.
- Rolf Moe-Nilssen and Jorunn L. Helbostad. Spatiotemporal gait parameters for older adults – an interactive model adjusting reference data for gender, age, and body height. *Gait & Posture*, 82: 220–226, October 2020. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2020.09.009.

- Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2):90–126, 2006. ISSN 1077-3142. doi: 10.1016/j.cviu.2006.08.002.
- Jashila Nair Mogan, Chin Poo Lee, Kian Ming Lim, and Kalaiarasi Sonai Muthu. Gait-ViT: Gait recognition with vision transformer. *Sensors*, 22(19):7362, September 2022. ISSN 1424-8220. doi: 10.3390/s22197362.
- Noraini Mohamed, Mumtaz Begum Mustafa, and Nazean Jomhari. A review of the hand gesture recognition system: Current progress and future directions. *IEEE Access*, 9:157422–157436, 2021. ISSN 2169-3536. doi: 10.1109/access.2021.3129650.
- Mohamed Mohandes, Mohamed Deriche, and Salihu O. Aliyu. Classifiers combination techniques: A comprehensive review. *IEEE Access*, 6:19626–19639, 2018. ISSN 2169-3536. doi: 10.1109/access.2018.2813079.
- Rozhin Molavian, Ali Fatahi, Hamed Abbasi, and Davood Khezri. Artificial intelligence approach in biomechanics of gait and sport: A systematic literature review. *Journal of Biomedical Physics and Engineering*, 13(5), 2023. ISSN 2251-7200. doi: 10.31661/jbpe.v0i0.2305-1621.
- Andrew S. Monaghan, Jessie M. Huisinga, and Daniel S. Peterson. The application of principal component analysis to characterize gait and its association with falls in multiple sclerosis. *Scientific Reports*, 11(1), June 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-92353-2.
- Gyeongsik Moon, Hongsuk Choi, and Kyoung Mu Lee. Accurate 3d hand pose estimation for whole-body 3d human mesh estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2022. doi: 10.1109/cvprw56347.2022.00257.
- Atsushi Mori, Yasushi Makihara, and Yasushi Yagi. Gait recognition using period-based phase synchronization for low frame-rate videos. In *Proceedings of the 20th International Conference on Pattern Recognition*, pages 2194–2197. IEEE, August 2010. doi: 10.1109/icpr.2010.537.
- Matteo Moro, Giorgia Marchesi, Filip Hesse, Francesca Odone, and Maura Casadio. Markerless vs. marker-based gait analysis: A proof of concept study. *Sensors*, 22(5):2011, March 2022. ISSN 1424-8220. doi: 10.3390/s22052011.
- Meg E. Morris, Jennifer McGinley, Frances Huxham, Janice Collier, and Robert Iansek. Constraints on the kinetic, kinematic and spatiotemporal parameters of gait in Parkinson’s disease. *Human Movement Science*, 18(2–3):461–483, June 1999. ISSN 0167-9457. doi: 10.1016/s0167-9457(99)00020-2.
- Quentin Mourcou, Anthony Fleury, Céline Franco, Frédéric Klopčič, and Nicolas Vuillerme. Performance evaluation of smartphone inertial sensors measurement for range of motion. *Sensors*, 15(9):23168–23187, September 2015. ISSN 1424-8220. doi: 10.3390/s150923168.
- Zihao Mu, Francisco M. Castro, Manuel J. Marin-Jimenez, Nicolas Guil, Yan-Ran Li, and Shiqi Yu. ReSGait: The real-scene gait dataset. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, August 2021. doi: 10.1109/ijcb52358.2021.9484347.
- Marion Mundt, Arnd Koeppe, Sina David, Franz Bamer, Wolfgang Potthast, and Bernd Markert. Prediction of ground reaction force and joint moments based on optical motion capture data during gait. *Medical Engineering & Physics*, 86:29–34, December 2020. ISSN 1350-4533. doi: 10.1016/j.medengphy.2020.10.001.

- Marion Mundt, Steffi Colyer, Logan Wade, Laurie Needham, Murray Evans, Emma Millett, and Jacqueline Alderson. Automating video-based two-dimensional motion analysis in sport? implications for gait event detection, pose estimation, and performance parameter analysis. *Scandinavian Journal of Medicine & Science in Sports*, 34(7), July 2024. ISSN 1600-0838. doi: 10.1111/sms.14693.
- Alvaro Muro-de-la Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, 14(2):3362–3394, February 2014. ISSN 1424-8220. doi: 10.3390/s140203362.
- Mary Patricia Murray. Gait as a total pattern of movement. *American Journal of Physical Medicine & Rehabilitation*, 6(1):290–328, 1967.
- Melissa P Murray. Walking patterns of normal woman. *Archives of Physical Medicine and Rehabilitation*, 51:637–650, 1970.
- Eadweard Muybridge. *The Human Figure in Motion*. Dover Publications Inc., 1955. ISBN 978-0486202044.
- Yang Na. Human body gait movement simulation based on adams. *Mechanical Engineering & Automation*, 2015.
- Laurie Needham, Murray Evans, Darren P. Cosker, Logan Wade, Polly M. McGuigan, James L. Bilzon, and Steffi L. Colyer. The accuracy of several pose estimation methods for 3d joint centre localisation. *Scientific Reports*, 11(1), October 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-00212-x.
- Laurie Needham, Murray Evans, Logan Wade, Darren P. Cosker, Miranda Polly McGuigan, James L. Bilzon, and Steffi L. Colyer. The development and evaluation of a fully automated markerless motion capture workflow. *Journal of Biomechanics*, 144:111338, nov 2022. doi: 10.1016/j.jbiomech.2022.111338.
- Richard R. Neptune, Steven A. Kautz, and Felix E. Zajac. Contributions of the individual ankle plantar flexors to support, forward progression and swing initiation during walking. *Journal of Biomechanics*, 34(11):1387–1398, 2001. doi: 10.1016/s0021-9290(01)00105-1.
- Donald A. Neumann. *Kinesiology of the musculoskeletal system*. Elsevier, St. Louis, Missouri, third edition edition, 2017. ISBN 9780323287531. Includes bibliographical references and index.
- Thanh Trung Ngo, Yasushi Makiyara, Hajime Nagahara, Yasuhiro Mukaigawa, and Yasushi Yagi. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, 47(1):228–237, January 2014. ISSN 0031-3203. doi: 10.1016/j.patcog.2013.06.028.
- Trong-Nguyen Nguyen, Huu-Hung Huynh, and Jean Meunier. 3D reconstruction with time-of-flight depth camera and multiple mirrors. *IEEE Access*, 6:38106–38114, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2854262.
- Robert C. Nickerson, Upkar Varshney, and Jan Muntermann. A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3): 336–359, May 2013. ISSN 1476-9344. doi: 10.1057/ejis.2012.26.

- Mario Nieto-Hidalgo, Francisco Javier Ferrández-Pastor, Rafael J. Valdivieso-Sarabia, Jerónimo Mora-Pascual, and Juan Manuel García-Chamizo. A vision based proposal for classification of normal and abnormal gait using RGB camera. *Journal of Biomedical Informatics*, 63:82–89, October 2016. ISSN 1532-0464. doi: 10.1016/j.jbi.2016.08.003.
- Simon Niklaus and Feng Liu. Softmax splatting for video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5436–5445. IEEE, June 2020. doi: 10.1109/cvpr42600.2020.00548.
- Michael Nirenberg, Wesley Vernon, and Ivan Birch. A review of the historical use and criticisms of gait analysis evidence. *Science & Justice*, 58(4):292–298, July 2018. ISSN 1355-0306. doi: 10.1016/j.scijus.2018.03.002.
- Sharon Nissimov, Jacob Goldberger, and Victor Alchanatis. Obstacle detection in a greenhouse environment using the kinect sensor. *Computers and Electronics in Agriculture*, 113:104–115, apr 2015. doi: 10.1016/j.compag.2015.02.001.
- Karen J. Nolan, Kiran K. Karunakaran, Kathleen Chervin, Michael R. Monfett, Radhika K. Bapineedu, Neil N. Jasey, and Mooyeon Oh-Park. Robotic exoskeleton gait training during acute stroke inpatient rehabilitation. *Frontiers in Neurobotics*, 14, October 2020. ISSN 1662-5218. doi: 10.3389/fnbot.2020.581815.
- Nor Muaza Nor Adnan, Mohd Nor Azmi Ab Patar, Hokyoo Lee, Shin-Ichiroh Yamamoto, Lee Jong-Young, and Jamaluddin Mahmud. Biomechanical analysis using kinovea for sports application. *IOP Conference Series: Materials Science and Engineering*, 342:012097, April 2018. ISSN 1757-899X. doi: 10.1088/1757-899x/342/1/012097.
- Tom F. Novacheck. The biomechanics of running. *Gait & Posture*, 7(1):77–95, 1998. doi: 10.1016/s0966-6362(97)00038-6.
- João F. Nunes, Pedro M. Moreira, and João M. R. S. Tavares. Human motion analysis and simulation tools: A survey. In Francisco Miranda and Carlos Abreu, editors, *Handbook of Research on Computational Simulation and Modeling in Engineering*, pages 359–388. IGI Global, Hershey, PA, USA, 2016. doi: 10.4018/978-1-4666-8823-0.ch012.
- João Ferreira Nunes. Binary images clustering with k-means. In *Proceedings of the 5th Doctoral Symposium in Informatics Engineering*, DSIE, pages 191–202, 2010.
- João Ferreira Nunes, Pedro Miguel Moreira, and João Manuel R. S. Tavares. Shape based image retrieval and classification. In *Proceedings of the 5th Iberian Conference on Information Systems and Technologies*, CISTI, pages 433–438, 2010.
- João Ferreira Nunes, Pedro Miguel Moreira, and João M. R. S. Tavares. Fast and simple shape based image retrieval - web based tool, 2011a. URL <http://shapes.estg.ipv.pt>.
- João Ferreira Nunes, Pedro Miguel Moreira, and João Manuel R. S. Tavares. Simple and fast shape based image retrieval. In *Proceedings of the Computational Vision and Medical Image Processing*, VipIMAGE, pages 73–78. Taylor & Francis Group, 2011b. ISBN 9780415683951.
- João Ferreira Nunes, Pedro Miguel Moreira, and João M. R. S. Tavares. A survey on human motion analysis and simulation tools. In *Proceedings of the International Conference on Computational and Experimental Biomedical Sciences*, ICCEBS, 2013.

- João Ferreira Nunes, Luís Paulo Reis, Rosaldo J. F. Rossetti, Pedro Miguel Moreira, and João Manuel R. S. Tavares. Dynamic simulation of human motion. In A. Carvalho Brito, J. da Silva Tavares, and C. Bragança, editors, *Proceedings of the European Simulation and Modelling Conference*, ESM, pages 395–397, 2014.
- João Ferreira Nunes, Pedro Miguel Moreira, and João M. R. S. Tavares. GRIDDS - a gait recognition image and depth dataset, 2018. URL <http://gridds.ipvc.pt>.
- João Ferreira Nunes, Pedro Miguel Moreira, and João Manuel R. S. Tavares. GRIDDS - a gait recognition image and depth dataset. In *VipIMAGE 2019*, pages 343–352. Springer International Publishing, 2019a. doi: 10.1007/978-3-030-32040-9_36.
- João Ferreira Nunes, Pedro Miguel Moreira, and João Manuel R. S. Tavares. *Benchmark RGB-D Gait Datasets: A Systematic Review*, pages 366–372. Springer International Publishing, 2019b. ISBN 9783030320409. doi: 10.1007/978-3-030-32040-9_38.
- Tommy Oberg, Alek Karsznia, and Kurt Oberg. Basic gait parameters: reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation research and development*, 30(2):210–223, 1993.
- Tommy Öberg, Alek Karsznia, and Kurt Öberg. Basic gait parameters: Reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation research and development*, 30(2):210–223, 1993. ISSN 0748-7711.
- Inseok Oh, Seungeun Rho, Sangbin Moon, Seongho Son, Hyoil Lee, and Jinyun Chung. Creating pro-level ai for a real-time fighting game using deep reinforcement learning. *IEEE Transactions on Games*, 14(2):212–220, June 2022. ISSN 2475-1510. doi: 10.1109/tg.2021.3049539.
- Jeonghoon Oh, Christopher Kuenze, Marco Jacopetti, Joseph F. Signorile, and Moataz Eltoukhy. Validity of the Microsoft Kinect™ in assessing spatiotemporal and lower extremity kinematics during stair ascent and descent in healthy young individuals. *Medical Engineering & Physics*, 60:70–76, October 2018. ISSN 1350-4533. doi: 10.1016/j.medengphy.2018.07.011.
- Temitayo Olugbade, Marta Bieńkiewicz, Giulia Barbareschi, Vincenzo D’amato, Luca Oneto, Antonio Camurri, Catherine Holloway, Mårten Björkman, Peter Keller, Martin Clayton, Amanda C De C Williams, Nicolas Gold, Cristina Becchio, Benoît Bardy, and Nadia Bianchi-Berthouze. Human movement datasets: An interdisciplinary scoping review. *ACM Computing Surveys*, 55(6):1–29, December 2022. ISSN 1557-7341. doi: 10.1145/3534970.
- Javier Ortells, María Trinidad Herrero-Ezquerro, and Ramón A. Mollineda. Vision-based gait impairment analysis for aided diagnosis. *Medical & Biological Engineering & Computing*, 56(9):1553–1564, February 2018. ISSN 1741-0444. doi: 10.1007/s11517-018-1795-2.
- Vanessa E. Ortiz-Padilla, Mauricio A. Ramírez-Moreno, Gerardo Presbítero-Espinosa, Ricardo A. Ramírez-Mendoza, and Jorge de J. Lozoya-Santos. Survey on video-based biomechanics and biometry tools for fracture and injury assessment in sports. *Applied Sciences*, 12(8):3981, April 2022. ISSN 2076-3417. doi: 10.3390/app12083981.
- Muyinat Y. Osoba, Ashwini K. Rao, Sunil K. Agrawal, and Anil K. Lalwani. Balance and gait in the elderly: A contemporary review. *Laryngoscope Investigative Otolaryngology*, 4(1):143–153, February 2019. ISSN 2378-8038. doi: 10.1002/lio2.252.

- Laura Maria Oudenhoven, Adam Thomas Crawford Booth, Annemieke Irene Buizer, Jaap Harlaar, and Marjolein Margaretha van der Krogt. How normal is normal: Consequences of stride to stride variability, treadmill walking and age when using normative paediatric gait data. *Gait & Posture*, 70:289–297, May 2019. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2019.03.011.
- Adeline Paiement, Lili Tao, Massimo Camplani, Sion Hannuna, Dima Damen, and Majid Mirmehdi. Online quality assessment of human motion from skeleton data. In *Proceedings of the British Machine Vision Conference*, BMVC 2014, pages 79.1–79.12. British Machine Vision Association, 2014. doi: 10.5244/c.28.79.
- Cristina Palmero, Albert Clapés, Chris Bahnsen, Andreas Møgelmoose, Thomas B. Moeslund, and Sergio Escalera. Multi-modal rgb–depth–thermal human body segmentation. *International Journal of Computer Vision*, 118(2):217–239, April 2016. ISSN 1573-1405. doi: 10.1007/s11263-016-0901-x.
- Leila Panahi and Vahid Ghods. Human fall detection using machine vision techniques on rgb–d images. *Biomedical Signal Processing and Control*, 44:146–153, July 2018. ISSN 1746-8094. doi: 10.1016/j.bspc.2018.04.014.
- A. Pantelopoulos and N. G. Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):1–12, January 2010. ISSN 1558-2442. doi: 10.1109/tsmcc.2009.2032660.
- Georgios Th. Papadopoulos, Apostolos Axenopoulos, and Petros Daras. Real-time skeleton-tracking-based human action recognition using kinect data. In *Proceedings of the 20th Anniversary International Conference on MultiMedia Modeling - Volume 8325*, MMM 2014, pages 473–483, New York, NY, USA, 2014. Springer International Publishing. ISBN 978-3-319-04113-1. doi: 10.1007/978-3-319-04114-8_40.
- Eirini Papageorgiou, Angela Nieuwenhuys, Ines Vandekerckhove, Anja Van Campenhout, Els Ortibus, and Kaat Desloovere. Systematic review on gait classifications in children with cerebral palsy: An update. *Gait & Posture*, 69:209–223, March 2019. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2019.01.038.
- Georgios I. Papagiannis, Athanasios I. Triantafyllou, Ilias M. Roumpelakis, Frantzeska Zampeli, Pinioti Garyfallia Eleni, Panayiotis Koulouvaris, Elias C. Papadopoulos, Panayiotis J. Papagelopoulos, and George C. Babis. Methodology of surface electromyography in gait analysis: review of the literature. *Journal of Medical Engineering & Technology*, 43(1):59–65, January 2019. ISSN 1464-522X. doi: 10.1080/03091902.2019.1609610.
- Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, 6(1), October 2019. ISSN 2052-4463. doi: 10.1038/s41597-019-0247-7.
- Anubha Parashar, Rajveer Singh Shekhawat, Weiping Ding, and Imad Rida. Intra-class variations with deep learning-based gait analysis: A comprehensive survey of covariates and methods. *Neurocomputing*, 505:315–338, September 2022. ISSN 0925-2312. doi: 10.1016/j.neucom.2022.07.002.

- Anubha Parashar, Apoorva Parashar, Andrea F. Abate, Rajveer Singh Shekhawat, and Imad Rida. Real-time gait biometrics for surveillance applications: A review. *Image and Vision Computing*, 138:104784, October 2023a. ISSN 0262-8856. doi: 10.1016/j.imavis.2023.104784.
- Anubha Parashar, Apoorva Parashar, Weiping Ding, Rajveer S. Shekhawat, and Imad Rida. Deep learning pipelines for recognition of gait biometrics with covariates: a comprehensive review. *Artificial Intelligence Review*, 56(8):8889–8953, January 2023b. ISSN 1573-7462. doi: 10.1007/s10462-022-10365-4.
- Monica Parati, Emilia Ambrosini, Beatrice De Maria, Matteo Gallotta, Laura A. Dalla Vecchia, Giorgio Ferriero, and Simona Ferrante. The reliability of gait parameters captured via instrumented walkways: a systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine*, 58(3), June 2022. ISSN 1973-9095. doi: 10.23736/s1973-9087.22.07037-x.
- Ji Su Park and Choong Hyun Kim. Ground-reaction-force-based gait analysis and its application to gait disorder assessment: New indices for quantifying walking behavior. *Sensors*, 22(19): 7558, October 2022. ISSN 1424-8220. doi: 10.3390/s22197558.
- Young-Hoon Park, Chang-Hong Youm, and Kook-Woong Seo. Accuracy and consistency of three-dimensional motion analysis system. *Korean Journal of Applied Biomechanics*, 15(2):83–92, 2005.
- Rashmi Pathak, Puja S. Prasad, Vinit Kumar Gunjan, and Vijender Kumar Solanki. *Normalization Techniques in Multi Modal Biometric*, pages 425–431. Springer Singapore, August 2019. ISBN 9789811387159. doi: 10.1007/978-981-13-8715-9_51.
- Thanyamon Pattanapisont, Prarinya Siritanawan, Kazunori Kotani, Toshiaki Kondo, and Jessada Karnjana. Gait image analysis based on human body parts model. In *Proceedings of the IEEE International Conference on Agents (ICA)*, pages 24–27. IEEE, December 2023. doi: 10.1109/ica58824.2023.00014.
- Amol Patwardhan and Gerald Knapp. *Aggressive actions and anger detection from multiple modalities using Kinect*, 2016.
- Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11):100336, November 2021. ISSN 2666-3899. doi: 10.1016/j.patter.2021.100336.
- F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 724–732. IEEE, June 2016. doi: 10.1109/cvpr.2016.85.
- Chamalka Kenneth Perera, Zakia Hussain, Min Khant, Alpha Agape Gopalai, Darwin Gouwanda, and Siti Anom Ahmad. A motion capture dataset on human sitting to walking transitions. *Scientific Data*, 11(1), August 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03740-z.
- Jacquelin Perry. *Gait analysis: normal and pathological function*. SLACK, Incorporated, Thorofare, 2nd ed. edition, 2010. ISBN 9781617117862.

- Jacquelin Perry and Judith Burnfield. *Gait Analysis: Normal and Pathological Function*. CRC Press, June 2024. ISBN 9781003525592. doi: 10.1201/9781003525592.
- Tigran Petrosyan, Arayik Dunoyan, and Hasmik Mkrtychyan. Application of motion capture systems in ergonomic analysis. *Armenian Journal of Special Education*, 4(2):107–117, February 2020. ISSN 2579-2881. doi: 10.24234/se.2020.1.1.157.
- Ekkasit Pinyoanuntapong, Ayman Ali, Pu Wang, Minwoo Lee, and Chen Chen. Gaitmixer: Skeleton-based gait representation learning via wide-spectrum multi-axial mixer. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, June 2023. doi: 10.1109/icassp49357.2023.10096917.
- C. Pizzolato, M. Reggiani, L. Modenese, and D. G. Lloyd. Real-time inverse kinematics and inverse dynamics for lower limb applications using opensim. *Computer Methods in Biomechanics and Biomedical Engineering*, 20(4):436–445, October 2016. ISSN 1476-8259. doi: 10.1080/10255842.2016.1240789.
- Ronald Poppe. Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*, 108(1-2):4–18, 2007. doi: 10.1016/j.cviu.2006.10.016.
- Ronald Poppe. A survey on vision-based human action recognition. *Image Vision Computing*, 28(6):976–990, 2010. ISSN 0262-8856. doi: 10.1016/j.imavis.2009.11.014.
- Douglas Powell, A. Joseph Threlkeld, Xiang Fang, Anburaj Muthumani, and Ruiping Xia. Amplitude- and velocity-dependency of rigidity measured at the wrist in parkinson’s disease. *Clinical Neurophysiology*, 123(4):764–773, April 2012. ISSN 1388-2457. doi: 10.1016/j.clinph.2011.08.004.
- Preena Prasad, J. Anitha, and Divapriya Anil. A systematic review of noise types, denoising methods, and evaluation metrics in images. In *Proceedings of the IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, pages 1–9. IEEE, November 2023. doi: 10.1109/rasse60029.2023.10363591.
- Hari Prasanth, Miroslav Caban, Urs Keller, Grégoire Courtine, Auke Ijspeert, Heike Vallery, and Joachim von Zitzewitz. Wearable sensor-based real-time gait detection: A systematic review. *Sensors*, 21(8):2727, April 2021. ISSN 1424-8220. doi: 10.3390/s21082727.
- Carina Price, Timo Schmeltzpfenning, Chris J. Nester, and Torsten Brauner. *Foot and footwear biomechanics and gait*, pages 79–103. Elsevier, 2021. ISBN 9780128216064. doi: 10.1016/b978-0-12-821606-4.00013-2.
- Yongbin Qi, Cheong Soh, Erry Gunawan, Kay-Soon Low, and Rijil Thomas. Estimation of spatial-temporal gait parameters using a low-cost ultrasonic motion analysis system. *Sensors*, 14(8): 15434–15457, August 2014. ISSN 1424-8220. doi: 10.3390/s140815434.
- Dhanesh Ramachandram and Graham W. Taylor. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6):96–108, November 2017. ISSN 1053-5888. doi: 10.1109/msp.2017.2738401.
- Erica M. Ramirez, Kathrin Ebinger, Denis Nam, Christopher Ferrigno, and Markus A. Wimmer. A machine learning approach using gait parameters to cluster tka subjects into stable and unstable joints for discovery analysis. *The Knee*, 54:167–177, June 2025. ISSN 0968-0160. doi: 10.1016/j.knee.2025.02.018.

- Heilym Ramirez, Sergio A. Velastin, Ignacio Meza, Ernesto Fabregas, Dimitrios Makris, and Gonzalo Farias. Fall detection and activity recognition using human skeleton features. *IEEE Access*, 9:33532–33542, 2021. ISSN 2169-3536. doi: 10.1109/access.2021.3061626.
- Veenu Rani and Munish Kumar. Human gait recognition: A systematic review. *Multimedia Tools and Applications*, 82(24):37003–37037, March 2023. ISSN 1573-7721. doi: 10.1007/s11042-023-15079-5.
- Rahm Ranjan, David Ahmedt-Aristizabal, Mohammad Ali Armin, and Juno Kim. Computer vision for clinical gait analysis: A gait abnormality video dataset. *IEEE Access*, 13:45321–45339, 2025. ISSN 2169-3536. doi: 10.1109/access.2025.3545787.
- John Rasmussen. *The AnyBody Modeling System*, pages 85–96. Elsevier, 2019. ISBN 9780128167137. doi: 10.1016/b978-0-12-816713-7.00008-8.
- Rabab A. Rasool. Feature-level vs. score-level fusion in the human identification system. *Applied Computational Intelligence and Soft Computing*, 2021:1–10, June 2021. ISSN 1687-9724. doi: 10.1155/2021/6621772.
- Paula Rego, Pedro Miguel Moreira, and Luis Paulo Reis. Serious games for rehabilitation: A survey and a classification towards a taxonomy. In *Proceedings of the 5th Iberian conference on information systems and technologies*, pages 1–6. IEEE, 2010.
- Paula Alexandra Rego, Rui Rocha, Brígida Mónica Faria, Luís Paulo Reis, and Pedro Miguel Moreira. A serious games platform for cognitive rehabilitation with preliminary evaluation. *Journal of Medical Systems*, 41(1), nov 2016. doi: 10.1007/s10916-016-0656-5.
- Francesco Ricciardi and Lucio Tommaso De Paolis. A comprehensive review of serious games in health professions. *International Journal of Computer Games Technology*, 2014:1–11, 2014. doi: 10.1155/2014/787968.
- Jim Richards, David Levine, and Michael W. Whittle. *Whittle’s gait analysis*. Churchill Livingstone/Elsevier, Edinburgh, 6th edition edition, 2022. ISBN 9780702084973. Includes bibliographical references and index.
- C. Rodrigues, M. V. Correia, J. M. C. S. Abrantes, J. Nadal, and M. A. B. Rodrigues. *Validation of Whole-Body COM Movement from 3D Anthropometric Image with Dynamic Data at Different Human Standard MVJ*, pages 353–365. Springer International Publishing, 2019. ISBN 9783030320409. doi: 10.1007/978-3-030-32040-9_37.
- S. Roeles, P. J. Rowe, S. M. Bruijn, C. R. Childs, G. D. Tarfali, F. Steenbrink, and M. Pijnappels. Gait stability in response to platform, belt, and sensory perturbations in young and older adults. *Medical & Biological Engineering & Computing*, 56(12):2325–2335, June 2018. ISSN 1741-0444. doi: 10.1007/s11517-018-1855-7.
- Daniel Roetenberg, Henk Luinge, Per Slycke, et al. Xsens mvn: Full 6dof human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV, Tech. Rep.*, 1(2009):1–7, 2009.
- Michael Ronthal. Gait disorders and falls in the elderly. *Medical Clinics of North America*, 103(2):203–213, March 2019. ISSN 0025-7125. doi: 10.1016/j.mcna.2018.10.010.
- C. Rougier, E. Auvinet, J. Meunier, M. Mignotte, and J. A. de Guise. Depth energy image for gait symmetry quantification. In *Proceedings of the Annual International Conference of the*

- IEEE Engineering in Medicine and Biology Society*. IEEE, aug 2011. doi: 10.1109/iembs.2011.6091272.
- Erynne Rowe, Marla K. Beauchamp, and Janie Astephen Wilson. Age and sex differences in normative gait patterns. *Gait & Posture*, 88:109–115, July 2021. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2021.05.014.
- Aditi Roy, Shamik Sural, and Jayanta Mukherjee. Gait recognition using pose kinematics and pose energy image. *Signal Processing*, 92(3):780–792, 2012. doi: 10.1016/j.sigpro.2011.09.022.
- Andrey Rudenko, Tomasz P. Kucner, Chittaranjan S. Swaminathan, Ravi T. Chadalavada, Kai O. Arras, and Achim J. Lilienthal. THÖR: Human-robot navigation data collection and accurate motion trajectories dataset. *IEEE Robotics and Automation Letters*, 5(2):676–682, April 2020a. ISSN 2377-3774. doi: 10.1109/lra.2020.2965416.
- Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Darius M Gavrila, and Kai O Arras. Human motion trajectory prediction: a survey. *The International Journal of Robotics Research*, 39(8):895–935, June 2020b. ISSN 1741-3176. doi: 10.1177/0278364920917446.
- Daniel Rueckert, Paul Aljabar, Rolf A. Heckemann, Joseph V. Hajnal, and Alexander Hammers. *Diffeomorphic Registration Using B-Splines*, pages 702–709. Springer Berlin Heidelberg, 2006. ISBN 9783540447283. doi: 10.1007/11866763_86.
- Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (pcl). In *2011 IEEE International Conference on Robotics and Automation*. IEEE, May 2011. doi: 10.1109/icra.2011.5980567.
- Roland Rössler, Jonathan Wagner, Raphael Knaier, Nikki Rommers, Reto W. Kressig, Arno Schmidt-Trucksäss, and Timo Hinrichs. Spatiotemporal gait characteristics across the adult lifespan: Reference values from a healthy population – analysis of the complete cohort study. *Gait & Posture*, 109:101–108, March 2024. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2024.01.005.
- Nasrin Sadeghzadehyazdi, Tamal Batabyal, and Scott T. Acton. Modeling spatiotemporal patterns of gait anomaly with a CNN-LSTM deep neural network. *Expert Systems with Applications*, 185:115582, December 2021. ISSN 0957-4174. doi: 10.1016/j.eswa.2021.115582.
- F. Salis, S. Bertuletti, T. Bonci, U. Della Croce, Mazzà C., and A. Cereatti. A method for gait events detection based on low spatial resolution pressure insoles data. *Journal of Biomechanics*, 127:110687, October 2021. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2021.110687.
- Nagwan Abdel Samee, Mohammed A. Al-masni, Eman N. Marzban, Abobakr Khalil Al-Shamiri, Mugahed A. Al-antari, Maali Ibrahim Alabdulhafith, Noha F. Mahmoud, and Yasser M. Kadah. New gait representation maps for enhanced recognition in clinical gait analysis. *Bioengineering*, 12(10):1130, October 2025. ISSN 2306-5354. doi: 10.3390/bioengineering12101130.
- M. M. Samson, A. Crowe, P. L. de Vreede, J. A. G. Dessens, S. A. Duursma, and H. J. J. Verhaar. Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and body weight. *Aging Clinical and Experimental Research*, 13(1):16–21, February 2001. ISSN 1720-8319. doi: 10.1007/bf03351489.

- Martin Sandau, Henrik Koblauch, Thomas B. Moeslund, Henrik Aanæs, Tine Alkjær, and Erik B. Simonsen. Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane. *Medical Engineering & Physics*, 36(9):1168–1175, September 2014. ISSN 1350-4533. doi: 10.1016/j.medengphy.2014.07.007.
- Catarina B. Santiago, Armando Sousa, and Luis Paulo Reis. Vision system for tracking handball players using fuzzy color processing. *Machine Vision and Applications*, 24(5):1055–1074, dec 2012. doi: 10.1007/s00138-012-0471-z.
- Nikolaos Sarafianos, Bogdan Boteanu, Bogdan Ionescu, and Ioannis A. Kakadiaris. 3d human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, 152:1–20, November 2016. ISSN 1077-3142. doi: 10.1016/j.cviu.2016.09.002.
- Renan Sarcinelli, Rânik Guidolini, Vinicius B. Cardoso, Thiago M. Paixão, Rodrigo F. Berriel, Pedro Azevedo, Alberto F. De Souza, Claudine Badue, and Thiago Oliveira-Santos. Handling pedestrians in self-driving cars using image tracking and alternative path generation with frenet frames. *Computers & Graphics*, 84:173–184, 2019. doi: 10.1016/j.cag.2019.08.004.
- J. B. Saunders, Verne T. Inman, and Howard D. Eberhart. The major determinants in normal and pathological gait. *The Journal of Bone & Joint Surgery*, 35:543–558, 1953.
- Rene Schuster, Oliver Wasenmuller, Christian Unger, and Didier Stricker. Ssgp: Sparse spatial guided propagation for robust and generic interpolation. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 197–206. IEEE, January 2021. doi: 10.1109/wacv48630.2021.00024.
- Johanna Seifert, Orsolya Friedrich, and Sebastian Schleidgen. Imitating the human. new human–machine interactions in social robots. *NanoEthics*, 16(2):181–192, August 2022. ISSN 1871-4765. doi: 10.1007/s11569-022-00418-x.
- Clara Selves, Gaëtan Stoquart, and Thierry Lejeune. Gait rehabilitation after stroke: review of the evidence of predictors, clinical outcomes and timing for interventions. *Acta Neurologica Belgica*, 120(4):783–790, March 2020. ISSN 2240-2993. doi: 10.1007/s13760-020-01320-7.
- Numan Senel, Klaus Kefferpütz, Kristina Doycheva, and Gordon Elger. Multi-sensor data fusion for real-time multi-object tracking. *Processes*, 11(2):501, February 2023. ISSN 2227-9717. doi: 10.3390/pr11020501.
- Alireza Sepas-Moghaddam and Ali Etemad. View-invariant gait recognition with attentive recurrent learning of partial representations. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(1):124–137, January 2021. ISSN 2637-6407. doi: 10.1109/tbiom.2020.3031470.
- Alireza Sepas-Moghaddam and Ali Etemad. Deep gait recognition: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):264–284, January 2023. ISSN 1939-3539. doi: 10.1109/tpami.2022.3151865.
- Dian Shao, Yue Zhao, Bo Dai, and Dahua Lin. FineGym: A hierarchical video dataset for fine-grained action understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020a. doi: 10.1109/cvpr42600.2020.00269.
- Dian Shao, Yue ZhaoBo, and DaiDahua Lin. Finegym. <https://sdolivia.github.io/FineGym/>, 2020b.

- Shubham Sharma, Shubhankar Verma, Mohit Kumar, and Lavanya Sharma. Use of motion capture in 3d animation: Motion capture systems, challenges, and recent trends. In *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 289–294. IEEE, February 2019a. doi: 10.1109/comitcon.2019.8862448.
- Sneha Sharma, Rinki Gupta, and Arun Kumar. On the use of multi-modal sensing in sign language classification. In *Proceedings of the 6th International Conference on Signal Processing and Integrated Networks*. IEEE, mar 2019b. doi: 10.1109/spin.2019.8711702.
- Allam Shehata, Francisco M. Castro, Nicolás Guil, Manuel J. Marín-Jiménez, and Yasushi Yagi. OUMVLP-OF: Multi-view large population gait database with dense optical flow and its performance evaluation. *IEEE Access*, 13:87100–87111, 2025. ISSN 2169-3536. doi: 10.1109/access.2025.3570280.
- Chuanfu Shen, Fan Chao, Wei Wu, Rui Wang, George Q. Huang, and Shiqi Yu. Lidargait: Benchmarking 3d gait recognition with point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1054–1063. IEEE, June 2023. doi: 10.1109/cvpr52729.2023.00108.
- Zhihao Shi, Xiangyu Xu, Xiaohong Liu, Jun Chen, and Ming-Hsuan Yang. Video frame interpolation transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17482–17491, June 2022.
- Jungpil Shin, Akitaka Matsuoka, Md. Al Mehedi Hasan, and Azmain Yakin Srizon. American sign language alphabet recognition by extracting feature from hand pose estimation. *Sensors*, 21(17):5856, August 2021. ISSN 1424-8220. doi: 10.3390/s21175856.
- Kohei Shiraga, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. GEINet: View-invariant gait recognition using a convolutional neural network. In *Proceedings of the International Conference on Biometrics (ICB)*, pages 1–8. IEEE, June 2016. doi: 10.1109/icb.2016.7550060.
- Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 1297–1304. IEEE Computer Society, June 2011. ISBN 978-1-4577-0394-2. doi: 10.1109/cvpr.2011.5995316.
- J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. *On a Large Sequence-Based Human Gait Database*, pages 339–346. Springer Berlin Heidelberg, 2004. ISBN 9783540452409. doi: 10.1007/978-3-540-45240-9_46.
- Tasriva Sikandar, Mohammad F. Rabbi, Kamarul H. Ghazali, Omar Altwijri, Mahdi Alqahtani, Mohammed Almijalli, Saleh Altayyar, and Nizam U. Ahamed. Using a deep learning method and data from two-dimensional (2d) marker-less video-based images for walking speed classification. *Sensors*, 21(8):2836, April 2021. ISSN 1424-8220. doi: 10.3390/s21082836.
- Luis M. Silva and Nick Stergiou. *The basics of gait analysis*, pages 225–250. Elsevier, 2020. ISBN 9780128133729. doi: 10.1016/b978-0-12-813372-9.00007-5.
- Vinícius Silva, Filomena Soares, Celina P. Leão, João Sena Esteves, and Gianni Vercelli. Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network. *Sensors*, 21(13):4342, June 2021. ISSN 1424-8220. doi: 10.3390/s21134342.

- Arnaud Simonet, Arnaud Delafontaine, Paul Fourcade, and Eric Yiu. Vertical center-of-mass braking and motor performance during gait initiation in young healthy adults, elderly healthy adults, and patients with parkinson's disease: A comparison of force-plate and markerless motion capture systems. *Sensors*, 24(4):1302, February 2024. ISSN 1424-8220. doi: 10.3390/s24041302.
- Dalwinder Singh and Birmohan Singh. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97:105524, December 2020. ISSN 1568-4946. doi: 10.1016/j.asoc.2019.105524.
- Jasvinder Pal Singh, Sanjeev Jain, Sakshi Arora, and Uday Pratap Singh. Vision-based gait recognition: A survey. *IEEE Access*, 6:70497–70527, 2018. ISSN 2169-3536. doi: 10.1109/access.2018.2879896.
- Jasvinder Pal Singh, Sakshi Arora, Sanjeev Jain, and Uday Pratap Singh SoM. A multi-gait dataset for human recognition under occlusion scenario. In *Proceedings of the International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, pages 1–6. IEEE, September 2019. doi: 10.1109/iciict46931.2019.8977673.
- Sabesan Sivapalan, Daniel Chen, Simon Denman, Sridha Sridharan, and Clinton Fookes. Gait energy volumes and frontal gait recognition using depth images. In *Proceedings of the International Joint Conference on Biometrics*, pages 1–6. IEEE, October 2011. doi: 10.1109/ijcb.2011.6117504.
- Sabesan Sivapalan, Daniel Chen, Simon Denman, Sridha Sridharan, and Clinton Fookes. Histogram of weighted local directions for gait recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 125–130. IEEE, June 2013. doi: 10.1109/cvprw.2013.26.
- Li Siyao, Shiyu Zhao, Weijiang Yu, Wenxiu Sun, Dimitris Metaxas, Chen Change Loy, and Ziwei Liu. Deep animation video interpolation in the wild. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6583–6591. IEEE, June 2021. doi: 10.1109/cvpr46437.2021.00652.
- Robert Socha and Bogusław Kogut. Urban video surveillance as a tool to improve security in public spaces. *Sustainability*, 12(15):6210, August 2020. ISSN 2071-1050. doi: 10.3390/su12156210.
- Markus D. Solbach and John K. Tsotsos. Vision-based fallen person detection for the elderly. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*. IEEE, oct 2017. doi: 10.1109/iccvw.2017.170.
- Chunfeng Song, Yongzhen Huang, Weining Wang, and Liang Wang. Casia-e: A large comprehensive dataset for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–16, 2022. ISSN 1939-3539. doi: 10.1109/tpami.2022.3183288.
- Hui Song, Xu Zhai, Zhongyang Gao, Teng Lu, Qian Tian, Haopeng Li, and Xijing He. Reliability and validity of a coda motion 3-d analysis system for measuring cervical range of motion in healthy subjects. *Journal of Electromyography and Kinesiology*, 38:56–66, February 2018. ISSN 1050-6411. doi: 10.1016/j.jelekin.2017.11.008.
- Shmuel Springer and Galit Yogeve Seligmann. Validity of the kinect for gait assessment: A focused review. *Sensors*, 16(2):194, February 2016. ISSN 1424-8220. doi: 10.3390/s16020194.

- Tim Steinebach, Eric H. Grosse, Christoph H. Glock, Jurij Wakula, and Alexander Lunin. Accuracy evaluation of two markerless motion capture systems for measurement of upper extremities: Kinect V2 and Captiv. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 30(4):291–302, March 2020. ISSN 1520-6564. doi: 10.1002/hfm.20840.
- Jan Stenum, Melody M. Hsu, Alexander Y. Pantelyat, and Ryan T. Roemmich. Clinical gait analysis using video-based pose estimation: Multiple perspectives, clinical populations, and measuring change. *PLOS Digital Health*, 3(3):e0000467, March 2024. ISSN 2767-3170. doi: 10.1371/journal.pdig.0000467.
- Erik E. Stone and Marjorie Skubic. Fall detection in homes of older adults using the microsoft kinect. *IEEE Journal of Biomedical and Health Informatics*, 19(1):290–301, jan 2015. doi: 10.1109/jbhi.2014.2312180.
- Fabio A. Storm, Christopher J. Buckley, and Claudia Mazzà. Gait event detection in laboratory and real life settings: Accuracy of ankle and waist sensor based methods. *Gait & Posture*, 50: 42–46, October 2016. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2016.08.012.
- Andreas Stotz, Daniel Hamacher, and Astrid Zech. Relationship between muscle strength and gait parameters in healthy older women and men. *International Journal of Environmental Research and Public Health*, 20(7):5362, April 2023. ISSN 1660-4601. doi: 10.3390/ijerph20075362.
- Jesus Suarez and Robin R. Murphy. Hand gesture recognition with depth images: A review. In *Proceedings of the IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, sep 2012. doi: 10.1109/roman.2012.6343787.
- Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2018. doi: 10.1109/cvpr.2018.00678.
- Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5686–5696. IEEE, June 2019. doi: 10.1109/cvpr.2019.00584.
- David H. Sutherland. The evolution of clinical gait analysis: Part I - kinesiological EMG. *Gait & Posture*, 14(1):61–70, July 2001. ISSN 0966-6362. doi: 10.1016/s0966-6362(01)00100-x.
- David H. Sutherland. The evolution of clinical gait analysis: Part ii - kinematics. *Gait & Posture*, 16(2):159–179, October 2002. ISSN 0966-6362. doi: 10.1016/s0966-6362(02)00004-8.
- David H. Sutherland. The evolution of clinical gait analysis: Part iii - kinetics and energy assessment. *Gait & Posture*, 21(4):447–461, 2005. ISSN 0966-6362. doi: <http://dx.doi.org/10.1016/j.gaitpost.2004.07.008>.
- David H. Sutherland and John L. Hagy. Measurement of gait movements from motion picture film. *The Journal of Bone & Joint Surgery*, 54(4):787–797, 1972. ISSN 1535-1386.
- Adam D. Sylvester, Steven G. Lautzenheiser, and Patricia Ann Kramer. A review of musculoskeletal modelling of human locomotion. *Interface Focus*, 11(5):20200060, August 2021. ISSN 2042-8901. doi: 10.1098/rsfs.2020.0060.

- Juri Taborri, Justin Keogh, Anton Kos, Alessandro Santuz, Anton Umek, Caryn Urbanczyk, Eline van der Kruk, and Stefano Rossi. Sport biomechanics applications using inertial, force, and EMG sensors: A literature overview. *Applied Bionics and Biomechanics*, 2020:1–18, June 2020. ISSN 1754-2103. doi: 10.1155/2020/2041549.
- Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, 10(4):1–14, February 2018. ISSN 1882-6695. doi: 10.1186/s41074-018-0039-6.
- Hotaka Takizawa, Shotaro Yamaguchi, Mayumi Aoyagi, Nobuo Ezaki, and Shinji Mizuno. Kinect cane: An assistive system for the visually impaired based on three-dimensional object recognition. In *Proceedings of the IEEE/SICE International Symposium on System Integration (SII)*. IEEE, dec 2012. doi: 10.1109/sii.2012.6426936.
- Daoliang Tan, Kaiqi Huang, Shiqi Yu, and Tieniu Tan. Efficient night gait recognition based on template matching. In *Proceedings of the 18th International Conference on Pattern Recognition*. IEEE, 2006. doi: 10.1109/icpr.2006.478.
- Chenyu Tang, Wentian Yi, Sanjeev Kumar, Gurvinder Singh Virk, and Luigi G. Occhipinti. Emg-based human motion analysis: A novel approach using towel electrodes and transfer learning. *IEEE Sensors Journal*, 24(6):9115–9123, March 2024. ISSN 2379-9153. doi: 10.1109/jsen.2024.3354307.
- Weijun Tao, Tao Liu, Rencheng Zheng, and Hutian Feng. Gait analysis using wearable sensors. *Sensors*, 12(12):2255–2283, February 2012. ISSN 1424-8220. doi: 10.3390/s120202255.
- Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow (extended abstract). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-2021*, pages 4839–4843. International Joint Conferences on Artificial Intelligence Organization, August 2021. doi: 10.24963/ijcai.2021/662.
- Torben Teepe, Ali Khan, Johannes Gilg, Fabian Herzog, Stefan Hormann, and Gerhard Rigoll. Gaitgraph: Graph convolutional network for skeleton-based gait recognition. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2314–2318. IEEE, September 2021. doi: 10.1109/icip42928.2021.9506717.
- Torben Teepe, Johannes Gilg, Fabian Herzog, Stefan Hörmann, and Gerhard Rigoll. Towards a deeper understanding of skeleton-based gait recognition, 2022.
- Hongjing Teng, Genning Zhang, Hui Zhou, S M Atoar Rahman, and Antonio Frisoli. Estimation of elbow joint angle from emg and imu measurements based on graph convolution neural network. In *Proceedings of the 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, pages 1–6. IEEE, October 2024. doi: 10.1109/m2vip62491.2024.10746173.
- Luigi Tesio and Viviana Rota. The motion of body center of mass during walking: A review oriented to clinical applications. *Frontiers in Neurology*, 10, September 2019. ISSN 1664-2295. doi: 10.3389/fneur.2019.00999.
- Daksh Thapar, Aditya Nigam, Divyansh Aggarwal, and Punjal Agarwal. VGR-net: A view invariant gait recognition network. In *Proceedings of the IEEE 4th International Conference on Identity, Security, and Behavior Analysis*, pages 1–8. IEEE, January 2018. doi: 10.1109/isba.2018.8311475.

- John S. Theodoropoulos, Jeremy Bettle, and Jonathan D. Kosy. The use of GPS and inertial devices for player monitoring in team sports: A review of current and future applications. *Orthopedic Reviews*, 12(1), April 2020. ISSN 2035-8237. doi: 10.4081/or.2020.7863.
- Haoyu Tian, Xin Ma, Hanbo Wu, and Yibin Li. Skeleton-based abnormal gait recognition with spatio-temporal attention enhanced gait-structural graph convolutional networks. *Neurocomputing*, 473:116–126, February 2022. ISSN 0925-2312. doi: 10.1016/j.neucom.2021.12.004.
- Luke K. Topham, Wasir Khan, Dhiya Al-Jumeily, and Abir Hussain. Human body pose estimation for gait identification: A comprehensive survey of datasets and models. *ACM Computing Surveys*, 55(6):1–42, December 2022. ISSN 1557-7341. doi: 10.1145/3533384.
- Luke K. Topham, Wasir Khan, Dhiya Al-Jumeily, Atif Waraich, and Abir J. Hussain. A diverse and multi-modal gait dataset of indoor and outdoor walks acquired using multiple cameras and sensors. *Scientific Data*, 10(1), May 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-02161-8.
- Alexander Toshev and Christian Szegedy. DeepPose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014. doi: 10.1109/cvpr.2014.214.
- Lam Tran, Thang Hoang, Thuc Nguyen, Hyunil Kim, and Deokjai Choi. Multi-model long short-term memory network for gait recognition using window-based data segment. *IEEE Access*, 9: 23826–23839, 2021. ISSN 2169-3536. doi: 10.1109/access.2021.3056880.
- Nicholas A. Trasolini, Kristen F. Nicholson, Joseph Mylott, Garrett S. Bullock, Tessa C. Hulburt, and Brian R. Waterman. Biomechanical analysis of the throwing athlete and its impact on return to sport. *Arthroscopy, Sports Medicine, and Rehabilitation*, 4(1):e83–e91, Jan 2022. doi: 10.1016/j.asmr.2021.09.027.
- Ursula Trinler, Hermann Schwameder, Richard Baker, and Nathalie Alexander. Muscle force estimation in clinical gait analysis using anybody and opensim. *Journal of Biomechanics*, 86: 55–63, March 2019. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2019.01.045.
- Rajesh Kumar Tripathi, Anand Singh Jalal, and Subhash Chand Agrawal. Abandoned or removed object detection from visual surveillance: a review. *Multimedia Tools and Applications*, 78(6): 7585–7620, August 2018. ISSN 1573-7721. doi: 10.1007/s11042-018-6472-9.
- Akira Tsuji, Yasushi Makihara, and Yasushi Yagi. Silhouette transformation based on walking speed for gait identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 717–722. IEEE, June 2010. doi: 10.1109/cvpr.2010.5540144.
- Chenxi Tu, Eijiro Takeuchi, Alexander Carballo, Chiyomi Miyajima, and Kazuya Takeda. Motion analysis and performance improved method for 3d lidar sensor data compression. *IEEE Transactions on Intelligent Transportation Systems*, 22(1):243–256, January 2021. ISSN 1558-0016. doi: 10.1109/tits.2019.2956066.
- Jeffrey A. Turner, Courtney R. Chaaban, and Darin A. Padua. Validation of opencap: A low-cost markerless motion capture system for lower-extremity kinematics during return-to-sport tasks. *Journal of Biomechanics*, 171:112200, June 2024. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2024.112200.

- Michal Tölgyessy, Martin Dekan, and L'uboš Chovanec. Skeleton tracking accuracy and precision evaluation of kinect v1, kinect v2, and the azure kinect. *Applied Sciences*, 11(12):5756, June 2021. ISSN 2076-3417. doi: 10.3390/app11125756.
- M. Zasim Uddin, Thanh Trung Ngo, Yasushi Makihara, Noriko Takemura, Xiang Li, Daigo Muramatsu, and Yasushi Yagi. The OU-ISIR large population gait database with real-life carried object and its performance evaluation. *IPSJ Transactions on Computer Vision and Applications*, 10(1):5, 2018. ISSN 1882-6695. doi: 10.1186/s41074-018-0041-z. URL <https://doi.org/10.1186/s41074-018-0041-z>.
- Md. Zasim Uddin, Daigo Muramatsu, Noriko Takemura, Md. Atiqur Rahman Ahad, and Yasushi Yagi. Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion. *IPSJ Transactions on Computer Vision and Applications*, 11(1), November 2019. ISSN 1882-6695. doi: 10.1186/s41074-019-0061-3.
- Scott D. Uhlich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S. Chaudhari, Jennifer L. Hicks, and Scott L. Delp. Opencap: Human movement dynamics from smartphone videos. *PLOS Computational Biology*, 19(10):e1011462, October 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011462.
- Mike Uschold and Michael Gruninger. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 11(2):93–136, June 1996. ISSN 1469-8005. doi: 10.1017/s0269888900007797.
- Somasundaram Vadivelu, Sudakshin Ganesan, O. V. Ramana Murthy, and Abhinav Dhall. Thermal imaging based elderly fall detection. In *Computer Vision – ACCV 2016 Workshops*, pages 541–553. Springer International Publishing, 2017. doi: 10.1007/978-3-319-54526-4_40.
- Eline van der Kruk and Marco M. Reijne. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*, 18(6):806–819, May 2018. ISSN 1536-7290. doi: 10.1080/17461391.2018.1463397.
- Sander van Hove, Bernard Leenstra, Paul Willems, Martijn Poeze, and Kenneth Meijer. The effect of age and speed on foot and ankle kinematics assessed using a 4-segment foot model. *Medicine*, 96(35):e7907, September 2017. ISSN 0025-7974. doi: 10.1097/md.0000000000007907.
- Nina M. van Mastrigt, Kevin Celie, Arjan L. Mieremet, Arnout C. C. Ruifrok, and Zeno Geradts. Critical review of the use and scientific basis of forensic gait analysis. *Forensic Sciences Research*, 3(3):183–193, July 2018. ISSN 2471-1411. doi: 10.1080/20961790.2018.1503579.
- Gül Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. *BodyNet: Volumetric Inference of 3D Human Body Shapes*, pages 20–38. Springer International Publishing, 2018. ISBN 9783030012342. doi: 10.1007/978-3-030-01234-2_2.
- Vani Vasudevan and Mohan S. Gounder. *A Systematic Review on Machine Learning-Based Sports Video Summarization Techniques*, pages 1–34. Springer International Publishing, 2023. ISBN 9783031205415. doi: 10.1007/978-3-031-20541-5_1.
- Christopher L. Vaughan. Theories of bipedal walking: an odyssey. *Journal of Biomechanics*, 36(4):513–523, 2003. doi: 10.1016/s0021-9290(02)00419-0.
- Christopher L. Vaughan, Brian L. Davis, and Jeremy C. O'Connor. *Dynamics of Human Gait*. Number v. 2 in Dynamics of Human Gait. Human Kinetics Publishers, Champaign, Ill., 2nd edition, 1999. ISBN 9780873223683.

- Ruben Vera-Rodriguez, John S.D. Mason, Julian Fierrez, and Javier Ortega-Garcia. Comparative analysis and fusion of spatiotemporal information for footstep recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):823–834, April 2013. ISSN 2160-9292. doi: 10.1109/tpami.2012.164.
- Tanmay Tulsidas Verlekar, Luís Ducla Soares, and Paulo Lobato Correia. Gait recognition in the wild using shadow silhouettes. *Image and Vision Computing*, 76:1–13, August 2018. ISSN 0262-8856. doi: 10.1016/j.imavis.2018.05.002.
- B. Vidya and Sasikumar P. Gait based parkinson’s disease diagnosis and severity rating using multi-class support vector machine. *Applied Soft Computing*, 113:107939, December 2021. ISSN 1568-4946. doi: 10.1016/j.asoc.2021.107939.
- Ashley Chey Vincent, Haley Furman, Rebecca C. Slepian, Kaitlyn R. Ammann, Carson Di Maria, Jung Hung Chien, Ka-Chun Siu, and Marvin J. Slepian. Smart phone-based motion capture and analysis: Importance of operating envelope definition and application to clinical use. *Applied Sciences*, 12(12):6173, June 2022. ISSN 2076-3417. doi: 10.3390/app12126173.
- Aditya Viswakumar, Venkateswaran Rajagopalan, Tathagata Ray, and Chandu Parimi. Human gait analysis using OpenPose. In *Proceedings of the 5th International Conference on Image Information Processing (ICIIP)*, pages 310–314. IEEE, November 2019. doi: 10.1109/iciip47207.2019.8985781.
- Slavka Viteckova, Patrik Kutilek, Zdenek Svoboda, Radim Krupicka, Jan Kauler, and Zoltan Szabo. Gait symmetry measures: A review of current and prospective methods. *Biomedical Signal Processing and Control*, 42:89–100, April 2018. ISSN 1746-8094. doi: 10.1016/j.bspc.2018.01.013.
- Jan P. Vox, Anika Weber, Karen Insa Wolf, Krzysztof Izdebski, Thomas Schüler, Peter König, Frank Wallhoff, and Daniel Friemert. An evaluation of motion trackers with virtual reality sensor technology in comparison to a marker-based motion capture system based on joint angles for ergonomic risk assessment. *Sensors*, 21(9):3145, May 2021. ISSN 1424-8220. doi: 10.3390/s21093145.
- Logan Wade, Laurie Needham, Polly McGuigan, and James Bilzon. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. *PeerJ*, 10:e12995, February 2022. ISSN 2167-8359. doi: 10.7717/peerj.12995.
- Changsheng Wan, Li Wang, and Vir V. Phoha. A survey on gait recognition. *ACM Computing Surveys*, 51(5):1–35, August 2018. ISSN 1557-7341. doi: 10.1145/3230633.
- Chen Wang, Junping Zhang, Jian Pu, Xiaoru Yuan, and Liang Wang. Chrono-gait image: A novel temporal template for gait recognition. In *Computer Vision – ECCV 2010*, pages 257–270. Springer Berlin Heidelberg, 2010a. doi: 10.1007/978-3-642-15549-9_19.
- Chenye Wang, Saihui Hou, Aoqi Li, Qingyuan Cai, and Yongzhen Huang. Ra-gar: A richly annotated benchmark for gait attribute recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(7):7591–7599, April 2025. ISSN 2159-5399. doi: 10.1609/aaai.v39i7.32817.
- Hsin-Huan Wang, Wei-Chi Tsai, Chi-Yao Chang, Min-Hao Hung, Jui-Hung Tu, Ti Wu, and Chia-Hsiang Chen. Effect of load carriage lifestyle on kinematics and kinetics of gait. *Applied Bionics and Biomechanics*, 2023:1–7, February 2023a. ISSN 1176-2322. doi: 10.1155/2023/8022635.

- Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. (IEEE), 2012. doi: 10.1109/cvpr.2012.6247813.
- Jin Wang, Mary She, Saeid Nahavandi, and Abbas Kouzani. A review of vision-based gait recognition methods for human identification. In *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, DICTA*, pages 320–327. IEEE, December 2010b. doi: 10.1109/DICTA.2010.62.
- Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*, 210: 103225, September 2021. ISSN 1077-3142. doi: 10.1016/j.cviu.2021.103225.
- Lei Wang, Du Q. Huynh, and Piotr Koniusz. A comparative review of recent kinect-based action recognition algorithms. *IEEE Transactions on Image Processing*, 29:15–28, 2020. ISSN 1941-0042. doi: 10.1109/tip.2019.2925285.
- Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, 2003a. ISSN 0031-3203. doi: [http://dx.doi.org/10.1016/S0031-3203\(02\)00100-0](http://dx.doi.org/10.1016/S0031-3203(02)00100-0).
- Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003b. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1251144.
- Rui Wang, Chuanfu Shen, Chao Fan, George Q. Huang, and Shiqi Yu. Pointgait: Boosting end-to-end 3d gait recognition with point clouds via spatiotemporal modeling. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, September 2023b. doi: 10.1109/ijcb57857.2023.10448817.
- Sheng Wang. Biomechanical analysis of the human knee joint. *Journal of Healthcare Engineering*, 2022:1–8, March 2022. ISSN 2040-2295. doi: 10.1155/2022/9365362.
- Shengke Wang, Long Chen, Zixi Zhou, Xin Sun, and Junyu Dong. Human fall detection in surveillance video based on PCANet. *Multimedia Tools and Applications*, 75(19):11603–11613, June 2015. ISSN 1573-7721. doi: 10.1007/s11042-015-2698-y.
- Tian Wang, Meina Qiao, Zhiwei Lin, Ce Li, Hichem Snoussi, Zhe Liu, and Chang Choi. Generative neural networks for anomaly detection in crowded scenes. *IEEE Transactions on Information Forensics and Security*, 14(5):1390–1399, may 2019. doi: 10.1109/tifs.2018.2878538.
- Xiaoming Wang, Hongliu Yu, Søren Kold, Ole Rahbek, and Shaoping Bai. Wearable sensors for activity monitoring and motion control: A review. *Biomimetic Intelligence and Robotics*, 3(1): 100089, March 2023c. ISSN 2667-3797. doi: 10.1016/j.birob.2023.100089.
- Vaishnavi Warutkar, Ragini Dadgal, and Utkarsha R Mangulkar. Use of robotics in gait rehabilitation following stroke: A review. *Cureus*, November 2022. ISSN 2168-8184. doi: 10.7759/cureus.31075.
- Fraja Watson, Peter C. Fino, Matthew Thornton, Constantinos Heracleous, Rui Loureiro, and Julian J. H. Leong. Use of the margin of stability to quantify stability in pathologic gait – a qualitative systematic review. *BMC Musculoskeletal Disorders*, 22(1), June 2021. ISSN 1471-2474. doi: 10.1186/s12891-021-04466-4.

- Wilhelm Weber and Eduard Weber. *Mechanics of the Human Walking Apparatus*. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG, 1992.
- Yating Wei. Deep-learning-based motion capture technology in film and television animation production. *Security and Communication Networks*, 2022:1–9, February 2022. ISSN 1939-0114. doi: 10.1155/2022/6040371.
- Brandon Welsh. *Making public places safer: surveillance and crime prevention*. Studies in Crime and Public Policy Ser. Oxford University Press, Incorporated, Cary, 2009. ISBN 9780195326215. Description based on publisher supplied metadata and other sources.
- Angela R. Weston, Keith R. Lohse, Andrew Kittelson, Laurie A. King, Patty Carlson-Kuhta, Leland E. Dibble, and Martina Mancini. Turning speed as a more responsive metric of age-related decline in mobility: A comparative study with gait speed. *Clinical Biomechanics*, 113:106196, March 2024. ISSN 0268-0033. doi: 10.1016/j.clinbiomech.2024.106196.
- M. Whittle. *Gait Analysis: An Introduction*. Butterworth-Heinemann, 4th edition, 2007. ISBN 9780750688833.
- Klaus Widhalm, Sebastian Durstberger, Andrea Greisberger, Brigitte Wolf, and Peter Putz. Validity of assessing level walking with the 2d motion analysis software templo and reliability of 3d marker application. *Scientific Reports*, 14(1), January 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-52053-z.
- C. R. Winby, P. Gerus, T. B. Kirk, and D. G. Lloyd. Correlation between emg-based co-activation measures and medial and lateral compartment loads of the knee during gait. *Clinical Biomechanics*, 28(9–10):1014–1019, November 2013. ISSN 0268-0033. doi: 10.1016/j.clinbiomech.2013.09.006.
- D. A. Winter. Human balance and posture control during standing and walking. *Gait & Posture*, 3(4):193–214, December 1995. ISSN 0966-6362. doi: 10.1016/0966-6362(96)82849-9.
- David A. Winter. *Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological*. University of Waterloo Press, 1991. ISBN 0-88898-105-8.
- David A. Winter. *Biomechanics and motor control of human movement*. Wiley, Hoboken, N.J, 4th edition, 2009. ISBN 9780470398180.
- David A. Winter, Arthur O. Quanbury, Douglas A. Hobson, H. Grant Sidwall, Gary Reimer, Brian G. Trenholm, Thomas Steinke, and Henry Shlosser. Kinematics of normal locomotion – a statistical study based on t.v. data. *Journal of Biomechanics*, 7(6):479–486, November 1974. ISSN 0021-9290. doi: 10.1016/0021-9290(74)90082-7.
- David J. Wright, Donald J. DeSanto, Michelle H. McGarry, Thay Q. Lee, and John A. Scolaro. Supplemental fixation of supracondylar distal femur fractures: A biomechanical comparison of dual-plate and plate-nail constructs. *Journal of Orthopaedic Trauma*, 34(8):434–440, 2020. doi: 10.1097/bot.0000000000001749.
- David J. Wright, Donald J. DeSanto, Michelle H. McGarry, Thay Q. Lee, and John A. Scolaro. Nail diameter significantly impacts stability in combined plate-nail constructs used for fixation of supracondylar distal femur fractures. *OTA International: The Open Access Journal of Orthopaedic Trauma*, 5(1):e174, February 2022. ISSN 2574-2167. doi: 10.1097/oi9.0000000000000174.

- Liyu Wu, Can Zhang, and Yuexian Zou. Spatiotemporal focus for skeleton-based action recognition. *Pattern Recognition*, 136:109231, April 2023. ISSN 0031-3203. doi: 10.1016/j.patcog.2022.109231.
- Zifeng Wu, Yongzhen Huang, Liang Wang, Xiaogang Wang, and Tieniu Tan. A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, February 2017. ISSN 2160-9292. doi: 10.1109/tpami.2016.2545669.
- Chi Xu, Yasushi Makihara, Xiang Li, Yasushi Yagi, and Jianfeng Lu. *Speed Invariance vs. Stability: Cross-Speed Gait Recognition Using Single-Support Gait Energy Image*, pages 52–67. Springer International Publishing, 2017a. ISBN 9783319541846. doi: 10.1007/978-3-319-54184-6_4.
- Chi Xu, Yasushi Makihara, Gakuto Ogi, Xiang Li, Yasushi Yagi, and Jianfeng Lu. The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation. *IPSJ Transactions on Computer Vision and Applications*, 9(24):1–14, December 2017b. ISSN 1882-6695. doi: 10.1186/s41074-017-0035-2.
- Shihao Xu, Jing Fang, Xiping Hu, Edith Ngai, Wei Wang, Yi Guo, and Victor C. M. Leung. Emotion recognition from gait analyses: Current research and future directions. *IEEE Transactions on Computational Social Systems*, 11(1):363–377, February 2024. ISSN 2373-7476. doi: 10.1109/tcss.2022.3223251.
- Zhiwei Xu, Zirong Luo, Lei Ren, Kunyang Wang, and Dan Hu. Three-dimensional human modeling and dynamics simulation based on ADAMS. In *Proceedings of the International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. IEEE, 2017c. doi: 10.1109/ihmsc.2017.191.
- Kentaro Yagi, Yuta Sugiura, Kunihiro Hasegawa, and Hideo Saito. Gait measurement at home using a single RGB camera. *Gait & Posture*, 76:136–140, February 2020. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2019.10.006.
- Muhammad Yahya, Jawad Ali Shah, Kushsairy Abdul Kadir, Zulkhairi M. Yusof, Sheroz Khan, and Arif Warsi. Motion capture sensing techniques used in human upper limb motion: a review. *Sensor Review*, 39(4):504–511, July 2019. ISSN 0260-2288. doi: 10.1108/sr-10-2018-0270.
- Hiroyuki Yamada, Jeongho Ahn, Oscar Martinez Mozos, Yumi Iwashita, and Ryo Kurazume. Gait-based person identification using 3D LiDAR and long short-term memory deep networks. *Advanced Robotics*, 34(18):1201–1211, July 2020. ISSN 1568-5535. doi: 10.1080/01691864.2020.1793812.
- Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), April 2018. ISSN 2159-5399. doi: 10.1609/aaai.v32i1.12328.
- Zaoming Yan, Pengcheng Lei, Tingting Wang, Faming Fang, Junkang Zhang, Yaomin Huang, and Haichuan Song. Explicit depth-aware blurry video frame interpolation guided by differential curves. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1994–2004. IEEE, June 2025. doi: 10.1109/cvpr52734.2025.00192.

- J. F. Yang and D. A. Winter. Electromyographic amplitude normalization methods: improving their sensitivity as diagnostic tools in gait analysis. *Archives of physical medicine and rehabilitation*, 65(9):517–521, 1984.
- Dingqiang Ye, Chao Fan, Jingzhe Ma, Xiaoming Liu, and Shiqi Yu. Biggait: Learning gait representation you want by large vision models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 200–210. IEEE, June 2024. doi: 10.1109/cvpr52733.2024.00027.
- Hongyi Ye, Tanfeng Sun, and Ke Xu. Gait recognition based on gait optical flow network with inherent feature pyramid. *Applied Sciences*, 13(19):10975, October 2023. ISSN 2076-3417. doi: 10.3390/app131910975.
- Jing Ye, Hongde Wu, Lishan Wu, Jianjun Long, Yuling Zhang, Gong Chen, Chunbao Wang, Xun Luo, Qinghua Hou, and Yi Xu. An adaptive method for gait event detection of gait rehabilitation robots. *Frontiers in Neurorobotics*, 14, jul 2020. ISSN 1662-5218. doi: 10.3389/fnbot.2020.00038.
- TzeWei Yeoh, Hernan E. Aguirre, and Kiyoshi Tanaka. Clothing-invariant gait recognition using convolutional neural network. In *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–5. IEEE, October 2016. doi: 10.1109/ispacs.2016.7824728.
- Xiangyu Yin, Boyuan Yang, Weichen Liu, Qiyao Xue, Abrar Alamri, Goeran Fiedler, and Wei Gao. ProGait: A multi-purpose video dataset and benchmark for transfemoral prosthesis users, 2025.
- Xiao-Lei Yin, Dong-Xue Liang, Lu Wang, Jing Qiu, Zhi-Yun Yang, Jun-Hui Xing, Jian-Zeng Dong, and Zhao-Yuan Ma. Reducing the x-ray radiation exposure frequency in cardio-angiography via deep-learning based video interpolation, 2020.
- Yilong Yin, Lili Liu, and Xiwei Sun. *SDUMLA-HMT: A Multimodal Biometric Database*, pages 260–268. Springer Berlin Heidelberg, 2011. ISBN 9783642254499. doi: 10.1007/978-3-642-25449-9_33.
- Fletcher Young, Alexander J. Hunt, Hillel J. Chiel, and Roger D. Quinn. *Using Animatlab for Neuromechanical Analysis: Linear Hill Parameter Calculation*, pages 402–414. Springer International Publishing, 2020. ISBN 9783030643133. doi: 10.1007/978-3-030-64313-3_38.
- Reem N. Yousef, Abeer T. Khalil, Ahmed S. Samra, and Mohamed Maher Ata. Model-based and model-free deep features fusion for high performed human gait recognition. *The Journal of Supercomputing*, 79(12):12815–12852, March 2023. ISSN 1573-0484. doi: 10.1007/s11227-023-05156-9.
- Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proceedings of the 18th International Conference on Pattern Recognition, ICPR*. IEEE, 2006a. doi: 10.1109/icpr.2006.67.
- Shiqi Yu, Qing Wang, and Yongzhen Huang. A large RGB-D gait dataset and the baseline algorithm. In *Biometric Recognition*, pages 417–424. Springer International Publishing, 2013. doi: 10.1007/978-3-319-02961-0_52.

- Shiqi Yu, Haifeng Chen, Qing Wang, Linlin Shen, and Yongzhen Huang. Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing*, 239:81–93, May 2017. ISSN 0925-2312. doi: 10.1016/j.neucom.2017.02.006.
- Shiqi Yu, Rijun Liao, Weizhi An, Haifeng Chen, Edel B. García, Yongzhen Huang, and Norman Poh. GaitGANv2: Invariant gait feature extraction using generative adversarial networks. *Pattern Recognition*, 87:179–189, March 2019. ISSN 0031-3203. doi: 10.1016/j.patcog.2018.10.019.
- Shiqi Yu, Yongzhen Huang, Liang Wang, Yasushi Makihara, Shengjin Wang, Md Atiqur Rahman Ahad, and Mark Nixon. Hid 2022: The 3rd international competition on human identification at a distance. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, volume 44, pages 1–9. IEEE, October 2022a. doi: 10.1109/ijcb54206.2022.10007993.
- Xinguo Yu, H.-W. Leong, Changsheng Xu, and Qi Tian. Trajectory-based ball detection and tracking in broadcast soccer video. *IEEE Transactions on Multimedia*, 8(6):1164–1178, 2006b. ISSN 1520-9210. doi: 10.1109/TMM.2006.884621.
- Zhiyang Yu, Yu Zhang, Xujie Xiang, Dongqing Zou, Xijun Chen, and Jimmy S. Ren. *Deep Bayesian Video Frame Interpolation*, pages 144–160. Springer Nature Switzerland, 2022b. ISBN 9783031197840. doi: 10.1007/978-3-031-19784-0_9.
- Jorge Zafra-Palma, Nuria Marín-Jiménez, José Castro-Piñero, Magdalena Cuenca-García, Rafael Muñoz-Salinas, and Manuel J. Marín-Jiménez. Health & gait: a dataset for gait-based analysis. *Scientific Data*, 12(1), January 2025. ISSN 2052-4463. doi: 10.1038/s41597-024-04327-4.
- Ana Paula Janner Zanardi, Edson Soares da Silva, Rochelle Rocha Costa, Elren Passos-Monteiro, Ivan Oliveira dos Santos, Luiz Fernando Martins Kruel, and Leonardo Alexandre Peyré-Tartaruga. Gait parameters of Parkinson’s disease compared with healthy controls: a systematic review and meta-analysis. *Scientific Reports*, 11(1), January 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-80768-2.
- Karl E. Zelik and Eric C. Honert. Ankle and foot power in gait analysis: Implications for science, technology and clinical assessment. *Journal of Biomechanics*, 75:1–12, June 2018. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2018.04.017.
- Cun Zhang, Xing-Peng Chen, Guo-Qiang Han, and Xiang-Jie Liu. Spatial transformer network on skeleton-based gait recognition. *Expert Systems*, 40(6):e13244, January 2023. ISSN 1468-0394. doi: 10.1111/exsy.13244.
- Erhu Zhang, Yongwei Zhao, and Wei Xiong. Active energy image plus 2DLPP for gait recognition. *Signal Processing*, 90(7):2295–2302, 2010. doi: 10.1016/j.sigpro.2010.01.024.
- He Zhang, Shenghao Ren, Haolei Yuan, Jianhui Zhao, Fan Li, Shuangpeng Sun, Zhenghao Liang, Tao Yu, Qiu Shen, and Xun Cao. Mmvp: A multimodal mocap dataset with vision and pressure sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21842–21852. IEEE, June 2024. doi: 10.1109/cvpr52733.2024.02063.
- Jing Zhang, Wanqing Li, Philip O. Ogunbona, Pichao Wang, and Chang Tang. RGB-D-based action recognition datasets: A survey. *Pattern Recognition*, 60:86–105, December 2016. ISSN 0031-3203. doi: 10.1016/j.patcog.2016.05.019.

- Ziyuan Zhang, Luan Tran, Xi Yin, Yousef Atoum, Xiaoming Liu, Jian Wan, and Nanxin Wang. Gait recognition via disentangled representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4705–4714. IEEE, June 2019. doi: 10.1109/cvpr.2019.00484.
- Ziyuan Zhang, Luan Tran, Feng Liu, and Xiaoming Liu. On learning disentangled representations for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 345–360, January 2022. ISSN 1939-3539. doi: 10.1109/tpami.2020.2998790.
- Peijun Zhao, Moises Alencastre-Miranda, Zhan Shen, Ciaran O’Neill, David Whiteman, Javier Gervas-Arruga, and Hermano Igo Krebs. Computer vision for gait assessment in cerebral palsy: Metric learning and confidence estimation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:2336–2345, 2024. ISSN 1558-0210. doi: 10.1109/tnsre.2024.3416159.
- Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 56(1):1–37, August 2023a. ISSN 1557-7341. doi: 10.1145/3603618.
- Jinkai Zheng, Xinchun Liu, Wu Liu, Lingxiao He, Chenggang Yan, and Tao Mei. Gait recognition in the wild with dense 3D representations and a benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20196–20205. IEEE, June 2022. doi: 10.1109/cvpr52688.2022.01959.
- Jinkai Zheng, Xinchun Liu, Shuai Wang, Lihao Wang, Chenggang Yan, and Wu Liu. Parsing is all you need for accurate gait recognition in the wild. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM ’23, pages 116–124. ACM, October 2023b. doi: 10.1145/3581783.3612052.
- Shuai Zheng, Junge Zhang, Kaiqi Huang, Ran He, and Tieniu Tan. Robust view transformation model for gait recognition. In *Proceedings of the 18th IEEE International Conference on Image Processing*. IEEE, September 2011. doi: 10.1109/icip.2011.6115889.
- Gao Zhongyang, Song Hui, Ren Fenggang, Li Yuhuan, Wang Dong, and He Xijing. Reliability and validity of coda motion analysis system for measuring cervical range of motion in patients with cervical spondylosis and anterior cervical fusion. *Experimental and Therapeutic Medicine*, September 2017. ISSN 1792-1015. doi: 10.3892/etm.2017.5239.
- Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2871–2878, 2012. doi: 10.1109/CVPR.2012.6248013.
- Huiyu Zhou, Datao Xu, Wenjing Quan, Minjun Liang, Ukadike Chris Ugbole, Julien S. Baker, and Yaodong Gu. A pilot study of muscle force between normal shoes and bionic shoes during men walking and running stance phase using opensim. *Actuators*, 10(10):274, October 2021. ISSN 2076-0825. doi: 10.3390/act10100274.
- Lin Zhou, Eric Fischer, Can Tunca, Clemens Markus Brahms, Cem Ersoy, Urs Granacher, and Bert Arnrich. How we found our imu: Guidelines to imu selection and a comparison of seven imus for pervasive healthcare applications. *Sensors*, 20(15):4090, July 2020. ISSN 1424-8220. doi: 10.3390/s20154090.

- Zirui Zhou, Junhao Liang, Zizhao Peng, Chao Fan, Fengwei An, and Shiqi Yu. Gait patterns as biomarkers: A video-based approach for classifying scoliosis, 2024.
- Wentao Zhu, Xiaoxuan Ma, Dongwoo Ro, Hai Ci, Jinlu Zhang, Jiaxin Shi, Feng Gao, Qi Tian, and Yizhou Wang. Human motion generation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2430–2449, April 2024. ISSN 1939-3539. doi: 10.1109/tpami.2023.3330935.
- Xiaopeng Zhu, Lijun Yun, Feiyan Cheng, and Chunjie Zhang. LFN: Based on the convolutional neural network of gait recognition method. *Journal of Physics: Conference Series*, 1650(3): 032075, October 2020. ISSN 1742-6596. doi: 10.1088/1742-6596/1650/3/032075.
- Zheng Zhu, Xianda Guo, Tian Yang, Junjie Huang, Jiankang Deng, Guan Huang, Dalong Du, Jiwen Lu, and Jie Zhou. Gait recognition in the wild: A benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14789–14799, oct 2021.
- Shinan Zou, Chao Fan, Jianbo Xiong, Chuanfu Shen, Shiqi Yu, and Jin Tang. Cross-covariate gait recognition: A benchmark. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.
- Burakhan Çubukçu, Uğur Yüzgeç, Raif Zileli, and Ahu Zileli. Reliability and validity analyzes of kinect v2 based measurement system for shoulder motions. *Medical Engineering & Physics*, 76:20–31, February 2020. ISSN 1350-4533. doi: 10.1016/j.medengphy.2019.10.017.
- Pınar Güner Şahan, Suhap Şahin, and Fidan Kaya Gülağız. A survey of appearance-based approaches for human gait recognition: techniques, challenges, and future directions. *The Journal of Supercomputing*, 80(13):18392–18429, May 2024. ISSN 1573-0484. doi: 10.1007/s11227-024-06172-z.

Appendix A

GRIDDS Joint-angle Trajectories

The following three full-page figures present **per-subject mean joint-angle trajectories** across all recorded strides (one panel per subject, 35 subjects). Angles are calculated from joint positions in the camera-space and normalized in time to the 0–100% gait cycle. For each joint, the red line shows the mean *left* angle with a red shaded band indicating ± 1 SD across strides; the blue line shows the mean *right* angle with a blue shaded band indicating ± 1 SD.

GRIDDs: Knee Flexion/Extension

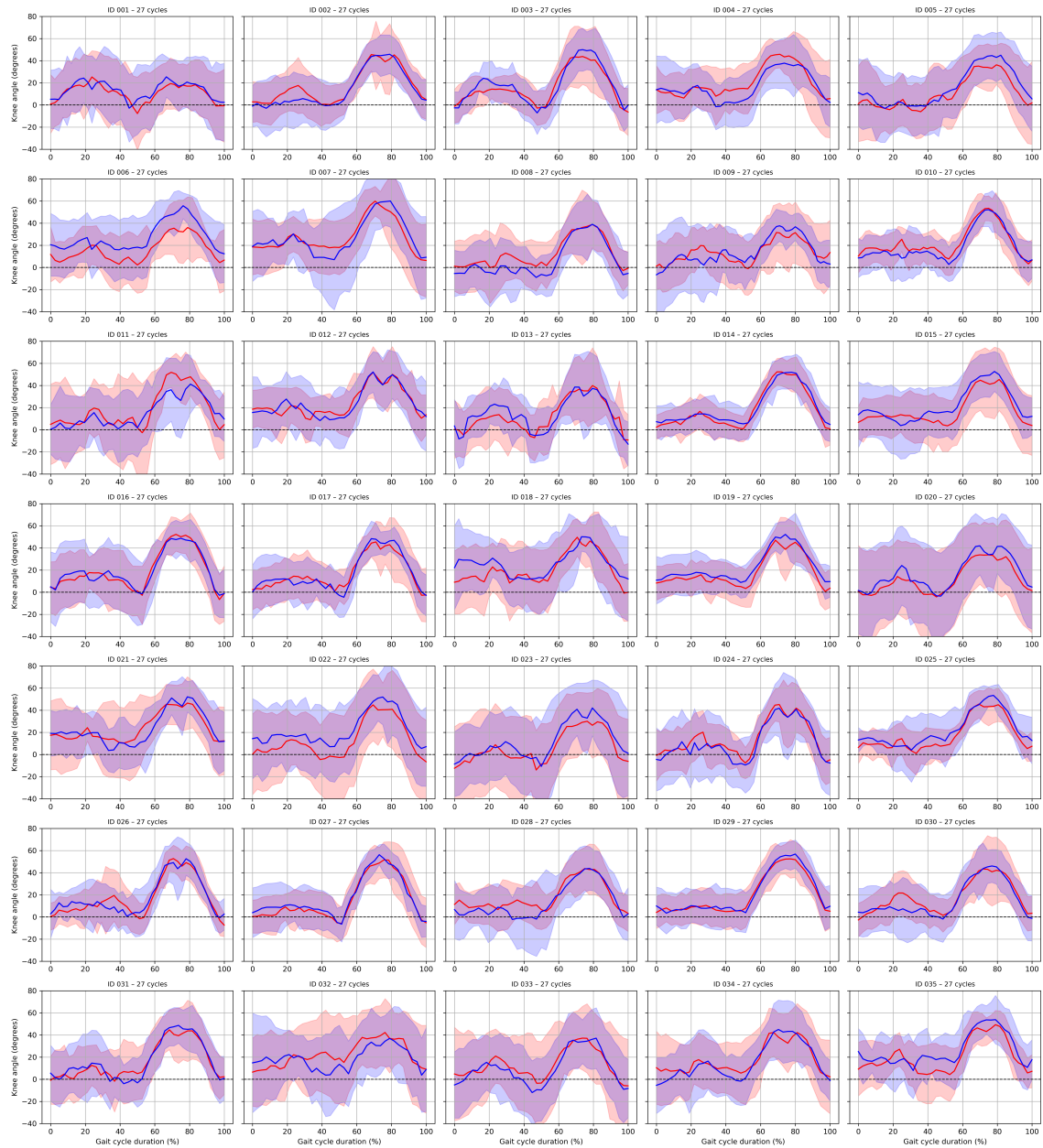


Figure A.1: Per-subject knee-angle trajectories (all subjects). Red/blue lines show mean left/right knee angles; shaded envelopes indicate ± 1 SD across strides. Each panel corresponds to an individual subject (IDs as titled).

GRIDDs: Hip Flexion/Extension

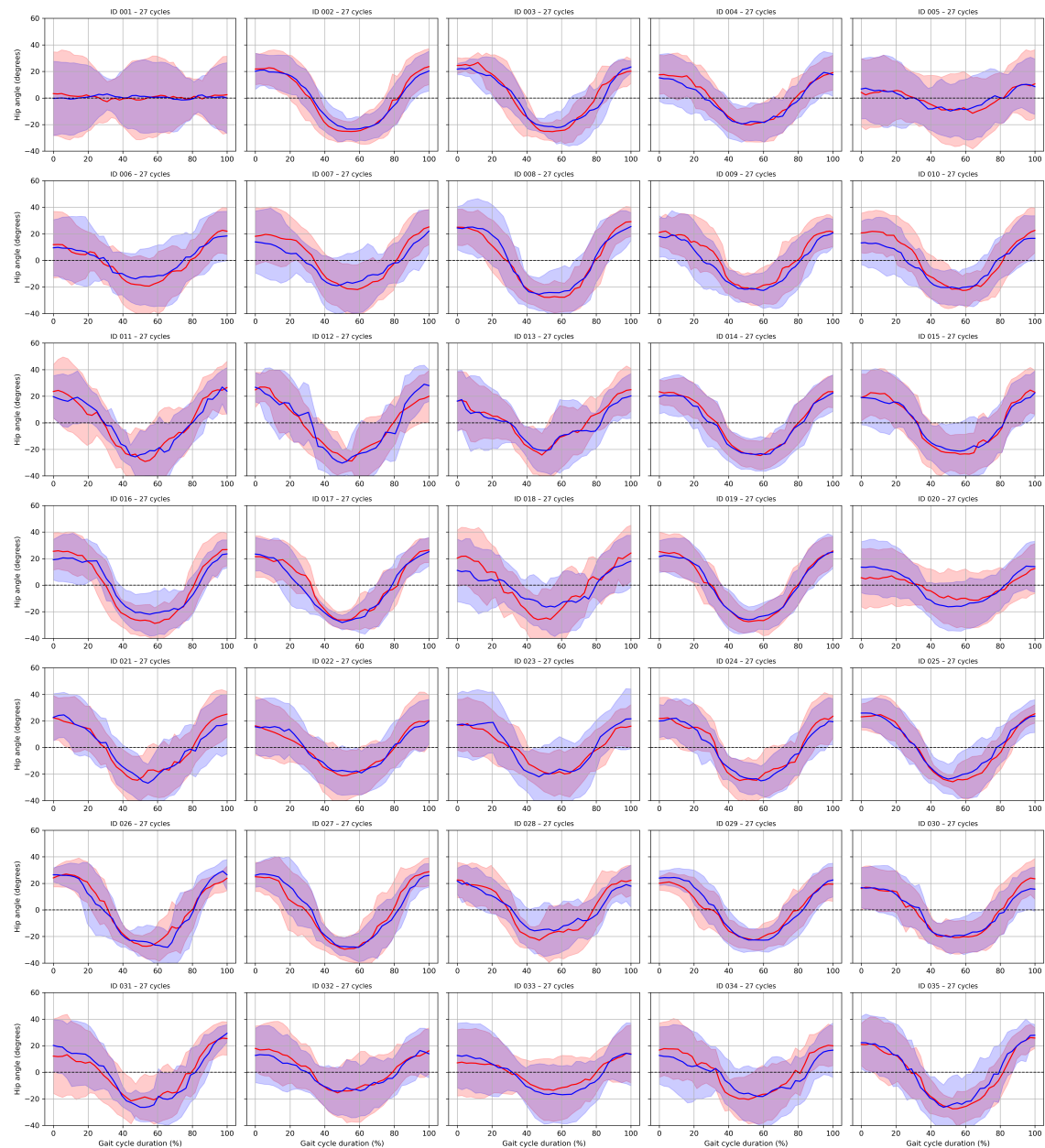


Figure A.2: **Per-subject hip-angle trajectories (all subjects).** Angles computed using the *trunk–thigh included angle* and the *thigh–vertical angle*. Red/blue lines show mean left/right hip angles; shaded envelopes indicate ± 1 SD across strides. Each panel corresponds to an individual subject (IDs as titled).

GRIDDs: Elbow Flexion/Extension

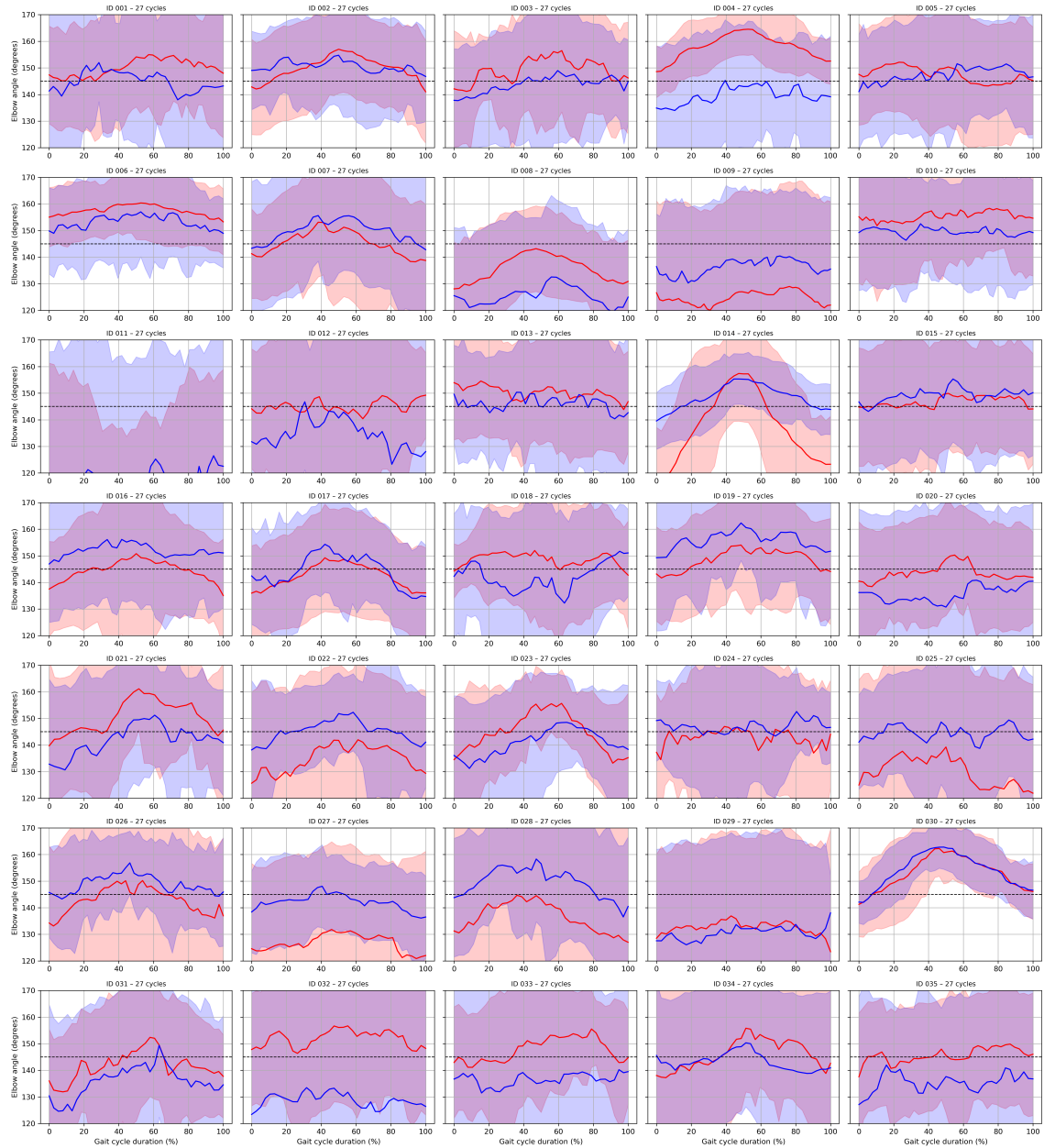


Figure A.3: Per-subject elbow-angle trajectories (all subjects). Red/blue lines show mean left-/right elbow angles; shaded envelopes indicate ± 1 SD across strides. Each panel corresponds to an individual subject (IDs as titled).