Neurocomputing 701 (2026) 134613

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Survey paper

Optical flow for human activity recognition: A systematic review from classical methods to deep learning

Abdorreza Alavi Gharahbagh [a], Vahid Hajihashemi [a], J.J.M. Machado [b], João Manuel R.S. Tavares [b,*]

[a] Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, S/N, 4200-465, Porto, Portugal
[b] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal

HIGHLIGHTS

- Systematic review of 205 optical flow-based human activity recognition (HAR) studies published between 2000 and 2025 across IEEE, Elsevier, and Springer databases.
- Quantitative analysis of the transition from traditional classifiers to deep learning architectures in optical flow-driven HAR systems.
- Comparative evaluation of four dominant optical flow methods (Lucas-Kanade, Horn-Schunck, Farnebäck, and TV-L1) across benchmark datasets and application scenarios.
- Identification of key limitations and a research roadmap addressing computational efficiency, robustness, and task-specific optical flow design for HAR.

ARTICLE INFO

Communicated by Y. Long

Keywords:
Human activity recognition
Optical flow
Deep learning
Convolutional neural networks
Spatiotemporal features

ABSTRACT

Human Activity Recognition (HAR) from video has become increasingly important in applications ranging from surveillance and healthcare monitoring to human–robot interaction. Optical flow, which describes apparent motion between consecutive frames, remains one of the most widely used temporal representations in video-based HAR. Despite this widespread use, the literature still offers limited systematic guidance on how different optical flow methods should be selected and integrated into activity-recognition pipelines. This review addresses that gap through a systematic analysis of 205 studies published between 2000 and 2025 and retrieved from IEEE, Elsevier, and Springer databases. The analysis focuses on four principal knowledge-based optical flow formulations: Lucas–Kanade, Horn–Schunck, Farnebäck, and TV-L1. For each method, the review examines its combination with feature extraction strategies, classifier architectures, and benchmark datasets. The results reveal substantial methodological changes over the review period. Lucas–Kanade dominated early work and appeared in 76 studies, whereas Farnebäck became more prominent in later periods, appearing in 53 studies. A clear shift from traditional machine learning to deep learning classifiers occurred after 2017: neural-network adoption increased from 13% in 2014–2017 to 51% in 2018–2020 and 63% in 2021–2025. Before 2018, histogram-based descriptors were the standard representation, but they were gradually replaced by the direct use of optical flow fields as inputs to convolutional and recurrent neural networks. TV-L1 showed particularly strong compatibility with deep learning architectures, with 15 of the 18 TV-L1 studies using neural-network classifiers. The review identifies persistent challenges, including computational efficiency for embedded deployment, robustness to environmental variation, and the lack of optical flow formulations designed specifically for activity recognition. A comparative analysis of 11 benchmark datasets is also provided to support evaluation under both controlled and real-world conditions.

* Corresponding author.
Email address: tavares@fe.up.pt (J.M.R.S. Tavares).

https://doi.org/10.1016/j.neucom.2026.134613
Received 28 January 2026; Received in revised form 28 June 2026; Accepted 26 July 2026
Available online 27 July 2026
0925-2312/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Nomenclature

List of Acronyms

AAE Average Angular Error
AdaBoost Adaptive Boosting
AdaScan Adaptive Scan Pooling in Deep CNN
AFM Attention Fusion Model
aHOF accumulated Histograms of Optical Flow
ANN Artificial Neural Network
B2C Bidirectional Co-temporal and Cross-Spatial
BoHOF Bag of Histograms of Optical Flow
CASDLSTM Convolutional Attention Sparse Deep LSTM
CNN Convolutional Neural Network
CRF Conditional Random Field
DBN Deep Belief Network
DGTC Differential Geometric Trajectory Cloud
DL Deep Learning
DSA Daily and Sports Activities
DS-ELM Double Similarity constrained ELM
EHOM Enhanced Histograms of Motion Words
ELM Extreme Learning Machine
EMD Earth Mover's Distance
FLFCN Fast Learning Fully Complex valued Neural
GMCP Generalised Maximum Clique Problem
GMM Gaussian Mixture Model
GRNN Generalised Regression Neural Networks
GWOF Gradient Weighted Optical Flow
HAR Human Activity Recognition
HCI Human Computer Interaction
HMM Hidden Markov Model
HMTO Histogram of Motion Trajectory Orientation
HOF Histogram of Optical Flow
HoFG Histogram of Flow and Gradient
HoNOF Histograms of Normalised Optical Flow
HOOF Histogram of Oriented Optical Flow
HOOFD Histograms of Oriented Optical Flows from Depth
HS Horn Schunck
HSGA Histogram of Spatial Gradient of Acceleration
I3D Inflated 3D ConvNet
IGMM Infinite Gaussian Mixture Model
IncSFA Incremental Slow Feature Analysis
KFSENet Key Frame and Skeleton Networks
KL divergence Kullback Leibler divergence
KNN K Nearest Neighbours
L2STM Lattice LSTM (L^2STM)
LBP Local Binary Patterns
LCF Learning Correlation Fusion
LDA Linear Discriminant Analysis
LDCRF Latent Dynamic Conditional Random Field
LFF Limb Flow Fields
LK Lucas Kanade
LR Linear Regression
LSDA Locality Sensitive Discriminant Analysis
LSTM Long Short-Term Memory
LT Net Long-Term Spatiotemporal Features Extraction Network
LVQ Learning Vector Quantisation
ML Machine Learning
MBH Motion Boundary Histograms
McFIS Meta cognitive Neuro Fuzzy Inference System
McRBFN Meta cognitive Radial Basis Function Network
MC STIP Motion Coherence Based Spatiotemporal Interest Point
MIL Multiple Instance Learning
MIO Multiple Independent Observations
MKL Multiple Kernel Learning
MLP Multi-Layer Perceptron
MMA Multi Modal Action
MoCA Multi-view Cooking Actions
MRDFNet Multimodal RGB D Flow Network
MRGCN Multi Region Based Radial Graph Convolutional Network
MSOF Mixed Scale Optical Flow
MSR Microsoft Research Action Data Set II
MTL Multi Task Learning
MVFI Motion Vector Flow Instance
NBMIM Naive Bayes based Mutual Information Maximisation
NMF Non negative Matrix Factorization
NUMA Northwestern UCLA Multi-view Action 3D
OMR map Orientation Magnitude Response map
PaHOF Pyramid of accumulated Histograms of Optical Flow
PCA Principal Component Analysis
PI3D Pose Guided Inflated 3D ConvNet
pLSA probabilistic Latent Semantic Analysis
RBM Restricted Boltzmann Machine
RHOF Refined HOF
RLS Regularised Least Squares
RMTL Robust Multi-Task Learning
RNN Recurrent Neural Network
SIFT Scale Invariant Feature Transform
S LDA Semi Latent Dirichlet Allocation
SRC Sparse Representation based Classifier
STA Spatio Temporal Attention
STCN Spatial Temporal Convolutional Networks
ST GCN Spatial Temporal Graph Convolutional Network
STIP Sparse Temporal Interest Points
ST Net Short Term Spatiotemporal Features Extraction Network
STSP Spatio Temporal Steerable Pyramid
SVM Support Vector Machine
SWTA Sparse Weighted Temporal Attention
TS CNN LSTM Three Stream CNN LSTM
TSI Translation and Scale Invariant
TSN Temporal Segment Networks
TWSVM Twin Support Vector Machines
uulmMAD University of Ulm Multi-perspective Action Dataset
VFAM Video Level Features and Attention Mechanism
VR Visual Rhythm
W KNN Weighted KNN
WSRTL Weakly Supervised Regional and Temporal Learning
3DS HONV 3D Spherical Histogram of Oriented Normal Vectors

1. Introduction

Video-based HAR has become a central topic in pattern recognition and computer vision, driven by increasing demand in surveillance, healthcare monitoring, autonomous systems, and human–machine interaction. Recognising human motion from video requires robust temporal representations capable of capturing how body parts and objects move across consecutive frames. In this context, human motion is usually described in terms of gestures, actions, interactions, and group activities. Gestures are expressive movements of the face, hands, body, or feet; actions are individual movements such as waving, clapping, or running; interactions involve two or more people; and group activities combine gestures, individual actions, interactions, and contextual information [1].

- **Gesture:** A person uses facial, hand, body, or foot movements to express a specific idea or intention. These movements act as nonverbal language and complement verbal communication [2–4].
- **Action:** An individual movement, such as waving, clapping, or running, is typically performed by a single person and is often associated with everyday tasks [2–4].
- **Interaction:** An activity resulting from two people acting together, such as shaking hands, giving a high-five, or hugging [2–4].
- **Group activity:** A combination of gestures, individual actions, and interactions among several people, usually requiring contextual information from the surrounding scene [2–4].

HAR systems may rely on video, audio, inertial signals, or multimodal combinations, depending on the intended application. Sensor-based methods using accelerometers and gyroscopes have achieved strong results with deep recurrent architectures [5]; however, the present review focuses exclusively on video-based recognition. Owing to advances in artificial intelligence, pattern recognition, machine learning, and image/video processing, video-based HAR is now used in a wide range of domains [6–8]. In visual surveillance, it supports the detection of suspicious or abnormal behaviour in public spaces, shopping centres, and airports [8]. In human–computer interaction, it enables users to control devices through body movements and gestures without direct physical input [9,10]. In education, it can help identify signs of distraction or disruptive behaviour and support learning analytics [9,10]. In healthcare, HAR is used to monitor elderly people or patients with dementia, detect risky events, and support rehabilitation and physiotherapy [9,11,12]. Other application domains include robotics, sports-performance analysis, injury prevention, and abnormal-behaviour detection [11–14].

Several challenges still limit the robustness of video-based HAR. Anthropometric variation affects motion appearance because people differ in height, body size, and body structure; common mitigation strategies include training on large and diverse datasets and applying appropriate normalisation [11,12,15,16]. Camera viewpoint is another major factor, since real-world scenes vary in angle, distance, background, and perspective; therefore, training data should include diverse viewpoints, supported when necessary by preprocessing such as perspective correction and image normalisation [16,17]. Ambiguous actions, such as running, walking, and jogging, can be difficult to distinguish because they share similar motion patterns, although deep learning models can help capture subtle discriminative cues [16,17]. Performance is also affected by video quality, occlusion, illumination changes, shadows, camera motion, and adverse weather, all of which can degrade motion estimation and recognition accuracy [15–18].

The first stage of many HAR pipelines is the detection and representation of human motion. This stage becomes particularly challenging when the background is dynamic, the camera moves, visibility is degraded by fog, rain, or snow, video quality is low, or people are partially occluded [19,20]. Errors at this stage propagate through the entire recognition pipeline and can substantially reduce final HAR performance. Motion representations commonly used in HAR include optical flow, Motion History Images, and Motion Energy Images [21]. Earlier work also employed bag-of-visual-words and codebook-based representations for activity classification [22]. Among motion representations, optical flow is particularly important because it estimates both the direction and magnitude of apparent pixel motion between consecutive frames, usually under a brightness-constancy assumption. Motion History Images capture the temporal trace of moving objects, whereas Motion Energy Images indicate where motion has occurred without explicitly encoding its direction.

Optical flow extraction methods are commonly grouped into knowledge-based and learning-based approaches [23]. Knowledge-based methods estimate flow from mathematical constraints applied to consecutive frames. They do not require training data, but their performance can be affected by large object motion, occlusion, illumination changes, noise, and computational cost. Many refinements have therefore been proposed to improve their robustness under challenging conditions. Learning-based methods, in contrast, estimate motion from data and can achieve high accuracy and fast inference once trained, especially when optimised for a specific task. However, their generalisability depends strongly on training data, and they may perform less reliably outside the distribution on which they were trained. Several HAR systems combine these two paradigms to exploit the interpretability and portability of classical flow with the representational capacity of deep learning.

In parallel with optical flow research, video-based HAR has moved rapidly towards deep architectures. Two-stream networks and inflated 3D convolutional networks, such as I3D trained on the Kinetics dataset, substantially improved recognition accuracy over earlier approaches [24], while temporal segment networks established effective strategies for learning from full videos with limited training data [25]. More recently, transformer-based video models have achieved leading results on major benchmarks. TimeSformer learns directly from frame patches using space–time self-attention [26]; ViViT builds classification models from spatiotemporal tokens [27]; the Video Swin Transformer introduces local hierarchical attention for efficiency [28]; and the Multiscale Vision Transformer organises attention as a feature pyramid [29]. Self-supervised learning has also become increasingly important for learning video and motion representations without large labelled datasets, as noted in recent surveys [30,31]. These developments raise an important question: if modern models can learn motion directly from RGB frames, what role remains for explicit optical flow?

Optical flow has not disappeared from modern HAR. The two-stream I3D architecture retains a dedicated optical flow stream alongside the appearance stream [24], and recent transformer-based HAR pipelines still incorporate flow as a motion input, for example by combining a video Swin Transformer with a convolutional LSTM to model long-term dependencies [32]. This continued use motivates the present review. Although optical flow is a widely adopted motion feature, systematic guidance on how to select and integrate different optical flow methods in HAR remains limited. This study addresses that gap by reviewing how optical flow methods have been used in activity-recognition pipelines over the past 25 years, with particular attention to feature extraction, classifier choice, datasets, temporal trends, and methodological transitions.

The analysis is structured around the main components shown in Fig. 1. First, the theoretical foundations of the four most common knowledge-based optical flow methods in HAR are presented. Second, the review examines the feature extraction strategies and classifiers associated with each optical flow method across successive time intervals. Third, the temporal evolution of feature representations, classifier families, and method adoption is analysed quantitatively. Finally, the datasets used in the reviewed studies are compared, and the main challenges, research gaps, and future directions are discussed. Overall, the review considers four optical flow methods, five temporal intervals, six broad classifier families, and two dataset categories.

Fig. 2 summarises the study selection process following the PRISMA approach. The search covered three databases—IEEE Xplore, ScienceDirect, and Springer—and included studies published between 2000 and 2025. The identification stage used the query **Human + (Activity or Action) + Recognition**, which returned 362,432 records: 117,368 from IEEE Xplore, 139,308 from ScienceDirect, and 105,756 from Springer. Duplicate records were removed before screening. Titles, abstracts, and keywords were then screened, leaving 22,304 records after removing duplicates, studies unrelated to video-based activity recognition, and records whose full text was not in English. The eligibility stage focused on studies involving optical flow, using the query **"Optical flow" OR "Optic flow"**, which reduced the pool to 578 records. These were assessed in full against four required reporting items: the optical flow method, the feature representation, the classifier, and the evaluation dataset. Studies that did not provide sufficient

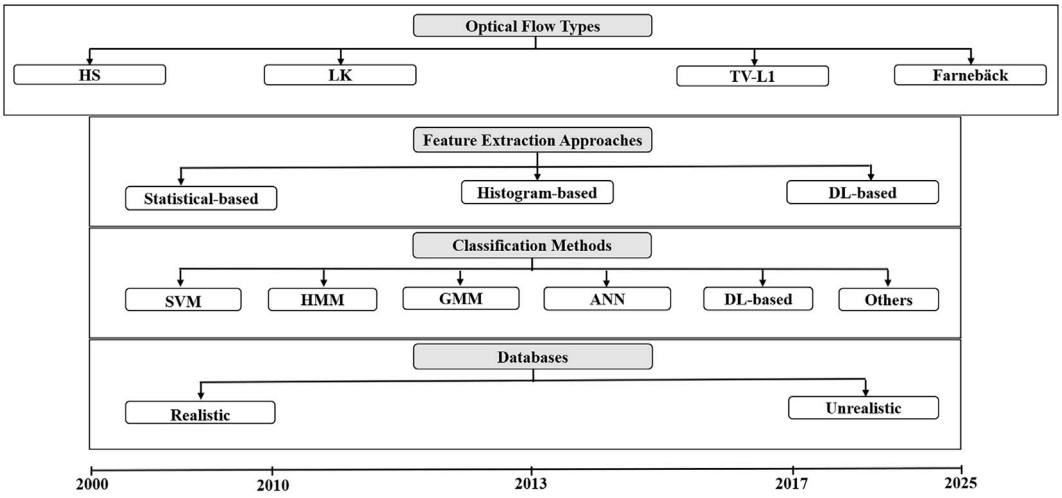


Fig. 1. Main topics addressed in this systematic review.

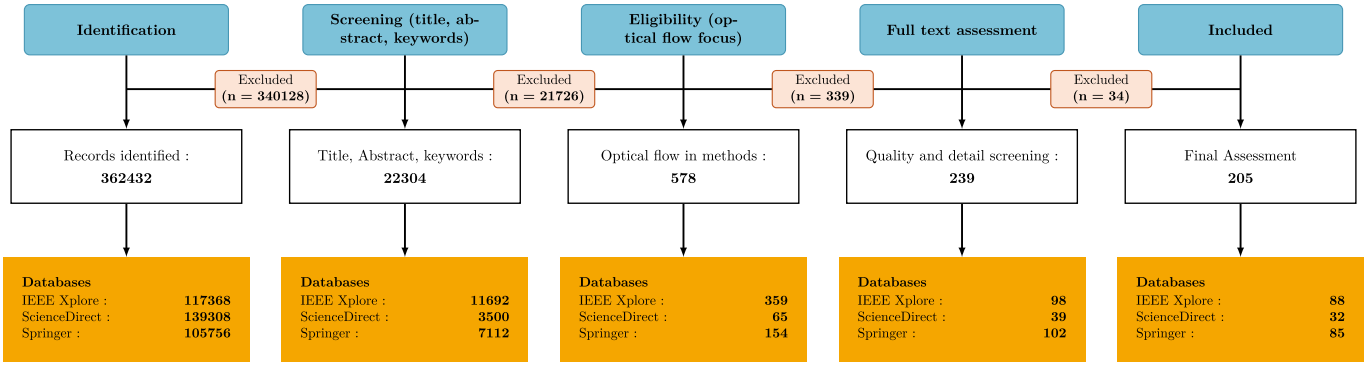


Fig. 2. PRISMA flow diagram of the study selection process.

information on these items were excluded. The final corpus contained 205 studies. The number of records removed at each stage, including the per-database counts, is shown in Fig. 2 to make the selection process transparent and reproducible.

The remainder of this article is organised as follows. Section 2 positions this review in relation to recent surveys of video-based activity recognition and clarifies its contribution. Section 3 presents the theoretical foundations of the four dominant knowledge-based optical flow methods and briefly discusses learning-based alternatives. Section 4 systematically reviews the 205 selected studies, grouped by optical flow type, and examines the corresponding feature extraction approaches, classifier architectures, and temporal usage patterns. Section 5 analyses the characteristics and usage frequency of the public benchmark datasets employed in optical flow-based HAR. Section 6 discusses temporal trends, the transition from traditional classifiers to deep learning, comparative method characteristics, current limitations, and future research directions. Section 7 summarises the main findings.

2. Related surveys

Video-based activity recognition has been surveyed from several directions in recent years, yet existing reviews approach the subject differently from this study. Saleem et al. [33] present a broad taxonomy of HAR that distinguishes between online and offline recognition, unimodal and multimodal inputs, and handcrafted features and learning-based models, and they compare methods based on activity complexity, dataset size, and recognition rate. Bukht et al. [34] examine video-based HAR published between 2018 and 2024 and group the work by the methods used to extract, represent, and classify actions. Both provide a broad overview of the field, but neither addresses the choice of optical flow method as a topic in its own right.

A further group of reviews looks mainly at deep learning. Gong and Luo [30] build their survey around the main deep architectures for video action recognition, namely two-stream networks, 3D CNNs, RNNs, LSTMs, and attention-based models, and report benchmark results; here, optical flow features are only one possible input rather than the organising idea. AlShami et al. [35] introduce the SMART-Vision taxonomy, which shows how two-stream, 3D convolutional, graph convolutional, and transformer designs combine into hybrid systems, and they also consider open-set recognition. Heo et al. [31] take a task-based view and follow six tasks, from action classification through to action anticipation, over the past decade. These reviews map the modern landscape well, but each treats optical flow as one feature among many and does not examine how the separate flow formulations behave within HAR pipelines. A third line of work is tied to a single domain. Yin et al. [36] review action recognition for team sports such as football and basketball, where rapid play and frequent player contact pose unique challenges. This is helpful for that setting, yet it stands apart from a general account of motion features.Recent primary studies also indicate a change in how motion is handled. Giveki and Ahmadi [37] place a masked cross-correlation operator inside a convolutional gated recurrent unit, so that spatial appearance, motion, and temporal change are learned together within a single network instead of through a separate optical flow stage, and they report high accuracy on UCF101, HMDB51, Hollywood2, UCF50, and Something-Something V1. Giveki [38] applies the same idea to engine

block assembly, using a cross-correlation operator within a gated recurrent unit to capture motion while avoiding the cost of explicit optical flow computation. Both works move motion estimation into the learning model itself, offering a useful contrast to the explicit flow methods reviewed here and a point this study returns to when weighing the role of optical flow in modern architectures.

Against this background, the contribution of the present review becomes clear. None of the previous surveys places optical flow at the centre of the analysis. This review is organised around the four knowledge-based optical flow formulations most commonly used in HAR, namely the LK, HS, Farnebäck, and TV-L1 methods. For each method, it systematically maps the feature extraction schemes and classifiers with which it has been combined in studies published between 2000 and 2025. It also quantifies, for each optical flow method, the transition from traditional classifiers to deep learning approaches and synthesises this evidence into practical guidance for selecting a suitable flow method for a given task. This method-centred perspective on the literature is not provided by earlier surveys.

3. Optical flow methods

This section presents the theoretical foundations of the optical flow methods most relevant to HAR. It first sets out the four knowledge-based methods that anchor this review, and then reviews the learning-based estimators and their place in HAR.

3.1. Knowledge-based optical flow methods

Knowledge-based optical flow methods can be classified in various ways [39,40]. Across the studies reviewed here, knowledge-based optical flow methods are the most widely used in HAR systems. The four knowledge-based approaches that appear most often in HAR research are: **HS** [41], **LK** [42], **Farnebäck** [43], and **TV-L1** [44]. These methods are generally designed under the assumption of brightness constancy across sequential frames. The motion of each pixel p after the time dt with velocity $\vec{v}_p$ in the frame I can be represented as:

$$I(p + \vec{v}_p dt), \tag{1}$$

where $\vec{v}_p = u_p i + v_p j$ is the optical flow vector, i.e., velocity, of each pixel p. Using the brightness constancy assumption, Eq. (1) can be written as [41,45,46]:

$$I(p, t) = I(p + \vec{v}_p dt, t + dt). \tag{2}$$

The gradient form of this brightness constancy equation, including spatial and temporal derivatives, becomes:

$$\nabla I(p, t).\vec{v}_p + I_t = I_x u + I_y v + I_t = 0, \tag{3}$$

where $\nabla I(p, t) = I_x \hat{i} + I_y \hat{j}$ represents the two-dimensional spatial gradient, the symbol $(\cdot)$ indicates the dot product, and I_x, I_y, I_t are spatial and temporal derivatives of I as:

$$\begin{aligned} I_x &= \frac{\partial I(p,t)}{\partial x}, \\ I_y &= \frac{\partial I(p,t)}{\partial y}, \\ I_t &= \frac{\partial I(p,t)}{\partial t}. \end{aligned} \tag{4}$$

It can be shown that Eq. (3) generally has an infinite number of solutions for $\vec{v}_p$; thus, solving it requires extra constraints and assumptions. The four knowledge-based methods reviewed here, namely HS, LK, TV-L1 and Farnebäck, each add their own assumptions to resolve this, and their full formulations and pseudocode are collected in Appendix A. Before turning to how these methods have been used in HAR, it helps to see what their output looks like on real frames. Fig. 3 shows the magnitude of the optical flow computed for two sample actions. The LK and HS algorithms extracted the boundary of the moving region, while the TV-L1 and Farnebäck methods produced a blurred motion map.

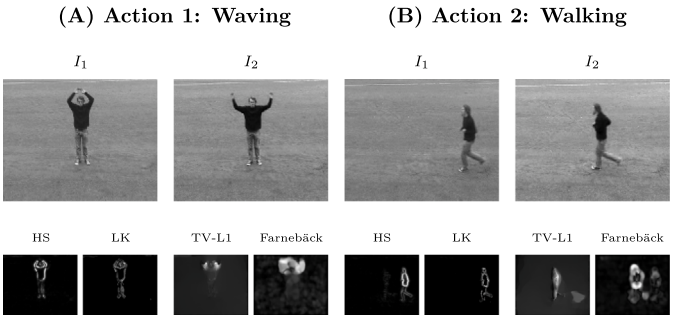


Fig. 3. Comparison of optical flow magnitude maps using Horn-Schunck (HS), Lucas-Kanade (LK), TV-L1, and Farnebäck methods for two human actions.

3.2. DL-based optical flow methods

This subsection reviews learning-based optical flow estimators and their relevance to HAR. FlowNet [47] introduced end-to-end learning for optical flow estimation through two architectures: FlowNet-Simple, which processes two consecutive frames with a standard encoder–decoder network, and FlowNet-Correlation, which explicitly models frame-to-frame correspondence through a correlation layer. These architectures showed that optical flow could be learned directly from data rather than computed only through hand-designed mathematical constraints. FlowNet has been used in several HAR systems [48–50]. FlowNet 2.0 [51] improved the original model by stacking networks, warping the second frame according to the current flow estimate, and adding a specialised subnetwork for small motions. These changes substantially reduced endpoint error while preserving fast inference.

SPyNet [52] estimates optical flow in a coarse-to-fine spatial pyramid. This design reduces computational cost while maintaining useful accuracy, making SPyNet attractive for deployment-oriented systems. Conceptually, its pyramidal processing resembles traditional coarse-to-fine methods such as Farnebäck (Algorithm 4), but replaces hand-crafted operations with learned convolutions. SPyNet has been used as a feature-extraction step in HAR methods [50,53]. PWC-Net [54] combines pyramid feature extraction, warping, and cost-volume construction, achieving strong accuracy with a smaller model than FlowNet. It has also been incorporated into HAR systems [55,56].

RAFT [57] differs from earlier single-pass architectures by maintaining a flow field that is iteratively refined. At each iteration, a recurrent unit uses correlation features to update the current flow estimate. This iterative strategy has achieved high accuracy and has been adopted in HAR-related systems [58,59]. Although RAFT shares the idea of iterative refinement with classical methods such as TV-L1 (Algorithm 3), the refinement mechanism is learned rather than derived from a variational formulation. More recent models have moved beyond the cost volume and long update chain used by RAFT. GMFlow [60] formulates optical flow as a global matching problem and uses a transformer to compare features across the entire image, which improves performance under large displacements. SEA-RAFT [61] retains the RAFT structure but adds a new loss, a direct initial flow estimate, and rigid-motion pretraining, improving accuracy and generalisation while reducing runtime.

Learning-based and knowledge-based optical flow methods differ mainly in training cost, memory requirements, inference behaviour, and generalisation. Knowledge-based methods such as LK and HS require no training and have a small memory footprint, since they store only image gradients and flow fields. Learning-based methods require network weights and substantial training resources. For example, PWC-Net contains approximately 9 million parameters and RAFT approximately 5 million, which makes their memory requirements much higher than those of classical methods. Training also requires large datasets of image pairs with ground-truth flow and typically depends on high-end GPUs.

Table 1
Comparison of knowledge-based and learning-based optical flow methods.

Aspect	Knowledge-based methods (HS, LK, Farnebäck, TV-L1)	Learning-based methods (FlowNet, FlowNet 2.0, SPyNet, PWC-Net, RAFT, GMFlow, SEA-RAFT)
Accuracy on flow benchmarks	Generally below the learned methods, with known sensitivity to noise and large motion [39]	State of the art; PWC-Net was the first learned model to surpass classical methods, and RAFT, GMFlow and SEA-RAFT lead current benchmarks [60–62]
Computational cost	No training needed; LK and HS are fast, while TV-L1 is slow	Heavy training on GPUs, but fast at inference once trained, for example, FlowNet 2.0 up to 140 frames per second [51]
Memory	Small footprint, holding only image gradients and flow fields	Must load network weights, around 9 million parameters for PWC-Net and 5 million for RAFT
Generalisation	Work on any footage without retraining	Depend on training data and can transfer poorly across domains, although rigid motion pre-training improves behaviour across datasets [61,62]
Use in HAR	Used in the great majority of the reviewed studies (Tables 2 to 12)	Used as a motion front end in some studies; flow benchmark error is not well matched to recognition, where accuracy at boundaries and small motions matters most [50]

Once trained, however, learning-based methods can be very fast at inference. FlowNet-Simple, for example, can process a frame pair in tens of milliseconds on modern GPUs, making it considerably faster than computationally heavier classical methods such as TV-L1.

The relative accuracy of learned and classical methods has also changed over time. The first FlowNet models did not consistently outperform well-tuned variational methods, especially for small motions and real video footage [51]. PWC-Net was among the first learned models to clearly surpass classical approaches, and RAFT later reduced benchmark error rates substantially relative to previous strong baselines [62]. Subsequent work has shown that training data and training schedules can be as important as network architecture, with careful retraining substantially improving older models such as PWC-Net [62]. For HAR, however, optical flow benchmark accuracy is not sufficient. Sevilla-Lara et al. [50] showed that endpoint error does not align perfectly with recognition accuracy and that activity recognition benefits especially from accurate motion boundaries and small-displacement estimates. They also found that training the flow network to minimise classification error, rather than endpoint error, improves recognition performance. Thus, a method that performs well on optical flow benchmarks is not automatically the best choice for activity recognition.

Overall, learning-based optical flow has matured quickly and is now accurate, fast, and increasingly relevant as a motion front end in HAR. Nevertheless, the systematic part of this review focuses on four knowledge-based methods because most HAR studies published between 2000 and 2025 still rely on them rather than on learned flow. Learning-based approaches are therefore treated as an important and growing direction, and their role in HAR is revisited in the discussion. Table 1 compares the two families across the main dimensions discussed above.

4. HAR methods based on optical flow

HAR methods that use optical flow can be analysed from several perspectives, including the optical flow formulation, the feature representation derived from the flow, the classifier, and the evaluation dataset. This review organises the studies primarily by optical flow type, which allows the methodological evolution of each formulation to be examined independently. Before discussing each method, it is important to clarify how the selected studies were assessed. Each article was checked for four reporting items that are essential for interpreting an optical flow-based HAR pipeline: the optical flow method, the feature representation, the classifier, and the dataset. Most studies reported all four items and therefore form the core of the tables that follow.

Reporting was not always complete. A sizeable group of 42 studies did not state which optical flow method was used. These studies were not discarded, because they still provide useful evidence about feature and classifier trends; instead, they are reported separately in Tables 11 and 12. A smaller number of studies did not specify the feature representation or classifier, which is indicated by a dash in the corresponding table cells. Studies that lacked sufficient information on the four reporting items were excluded during the final selection stage. This assessment does not rank studies by performance, since heterogeneous datasets and protocols make direct comparison unreliable, but it provides a transparent view of how completely the field reports its methods and explains the inclusion of an unspecified-flow category.

4.1. HAR methods based on LK optical flow

According to the reviewed literature, LK is the most frequently used optical flow method in HAR. Its use is analysed across four time intervals selected to provide comparable numbers of studies. In the early period, from 2000 to 2010, 19 studies used LK. Usage peaked in 2011–2013, with 22 studies, then decreased to 18 studies in 2014–2017 and 17 studies in 2018–2025. Although LK remains present in recent HAR research, with 9 studies in 2018–2020 and 8 in 2021–2025 (Fig. 6), its relative use has gradually declined as denser flow methods and deep learning pipelines have become more common.

Tables 2–5 list the LK-based HAR studies by time interval. Histogram-based representations dominate this group, especially HOF, used in 22 studies, and HOOF, used in 10 studies. In the 2018–2025 period, six deep learning studies used LK output directly as network input. Thus, in traditional pipelines, LK was most often converted into HOF or HOOF descriptors, whereas in recent deep learning pipelines it was more commonly used as a direct motion input. The classifier landscape is diverse: SVM is the most frequent classifier, appearing in 23 studies, followed by KNN, HMM, ANN, and CNN-based methods. Tables 2–5 therefore show both the historical dominance of LK and the gradual diversification of classifier choices over time.

Fig. 4 shows how classifier use changed across the LK time intervals. HMM was prominent in 2000–2010 but declined sharply afterwards. SVM became the dominant traditional classifier in 2011–2013 and 2014–2017, while KNN and ANN appeared less consistently. From 2018 onwards, deep learning classifiers, especially CNN-based architectures, expanded rapidly and reduced the relative presence of traditional classifiers (Fig. 4b).

4.2. HAR methods based on farnebäck optical flow

Farnebäck is the second most frequently used optical flow method in the reviewed HAR literature. Its temporal pattern differs from that of LK and is analysed across three intervals: before 2018, 2018–2020, and 2021–2025. Before 2018, 23 studies used Farnebäck optical flow; this decreased to 16 studies in 2018–2020 and 14 studies in 2021–2025. Although Farnebäck remained more common than LK after 2018, its overall trajectory also indicates a gradual decline.

Tables 6–8 summarise the Farnebäck-based HAR studies. Histogram-based features again dominate: 32 of the 53 studies used histogram representations, showing that Farnebäck, like LK, has often been integrated

Table 2
LK optical flow-based HAR studies from 2000 to 2010.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[63]	2000	Handshape features	HMM	Not mentioned
[64]	2006	The average value of the optical flow, PCA	HMM	KU gesture, manually captured data
[65]	2006	The average value of the optical flow, PCA	HMM	Korea University gesture
[66]	2007	k-medoids clustering	S-LDA	KTH, Soccer
[67]	2008	HOOF	SVM	KTH
[68]	2008	HOOF	HMM, CRF	KTH, Weizmann
[69]	2009	Several optical flow vectors weighted by correlation	KNN, ANN	user-generated
[70]	2009	Position dependent histograms	Gaussian kernel density estimation	KTH, Weizmann
[71]	2009	HOOF	Nonlinear Dynamical Systems	Weizmann
[72]	2009	HOOF	HMM	KTH
[73]	2009	HOOF	SVM	Soccer
[74]	2009	Position dependent histograms	SVM	KTH, Weizmann
[75]	2009	HOOF	KNN	KTH, Weizmann
[76]	2009	Handmade features	AdaBoost	Weizmann
[77]	2010	Handmade features	HMM	user-generated
[78]	2010	Features as in [79]	SVM	Kathmandu University Gesture, KTH
[80]	2010	Optical flow correlation	Template matching	CMU
[81]	2010	EHOM	SVM	KTH, Weizmann
[82]	2010	HOF	KNN, Bayesian Network	IXMAS

Table 3
LK optical flow-based HAR studies from 2011 to 2013.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[83]	2011	Handmade features	ANN	Weizmann
[84]	2011	Handmade features	KNN	Weizmann, IXMAS
[85]	2011	HOF	SVM	Weizmann
[86]	2011	Features extracted as in [87]	KNN	KTH, Weizmann
[88]	2012	HOOF	SVM	KTH, Weizmann
[89]	2012	Patch motion vectors	FLFCN classifier	KTH, Weizmann
[90]	2012	HOOF	ANN	Weizmann
[91]	2012	HOOF	ANN	KTH, Weizmann
[92]	2012	HOF	pLSA [93]	KTH, Weizmann
[94]	2012	HOF	HMM	Shotokan karate katas
[95]	2012	Features extracted as in [96]	McFIS	KTH, Weizmann
[97]	2012	HOF	KNN	Weizmann, CAVIAR
[98]	2012	HOF	Radius-based clustering	Weizmann, CAVIAR, MICC-UNIFI, Surveillance, Hollywood2
[99]	2012	HOF	SVM	KTH
[100]	2012	GMM	K-means	KTH, Weizmann, IXMAS, UCF Sports, Kecklab Gesture, UCF YouTube
[101]	2013	HOF	SVM	HMDB 51, UCF 50, UCF 101
[102]	2013	HOF-LBP	GRNN [103]	Weizmann
[104]	2013	3DS-HONV	SVM	MSR Action 3D, MSR-Gesture 3D
[105]	2013	HOF	SVM	KTH, MSR, Coffee & Cigarette, CMU human-object interaction, UCLA
[106]	2013	Features extracted as in [107]	SVM	KTH, IXMAS, HMDB 51
[108]	2013	Features extracted as in [83]	McRBFN	Video Analytics Lab (VAL), MHAD
[109]	2013	—	LDA	user-generated

into handcrafted feature pipelines. Among deep learning approaches, nine studies used Farnebäck optical flow directly as input, and five studies used deep networks as feature extractors before the final classifier. This contrasts with LK, for which no reviewed study used a pretrained deep network specifically for feature extraction from LK flow. Farnebäck therefore appears better suited to pretrained feature-extraction pipelines. In classification, Farnebäck-based studies show less diversity than LK-based studies. SVM remains the most common classifier, with 18 studies, while deep learning classifiers form the only other large group.

The Farnebäck studies also show a clear transition from traditional classifiers to deep learning. SVM usage declined more sharply for Farnebäck than for LK: in 2021–2025, SVM appeared in only one study, whereas deep learning methods appeared in seven (Fig. 5). By contrast, before 2018 there were 10 SVM-based studies and only one deep learning study. This pattern confirms the growing role of deep learning in Farnebäck-based HAR pipelines (Fig. 5b).

4.3. HAR methods based on TV-L1 optical flow

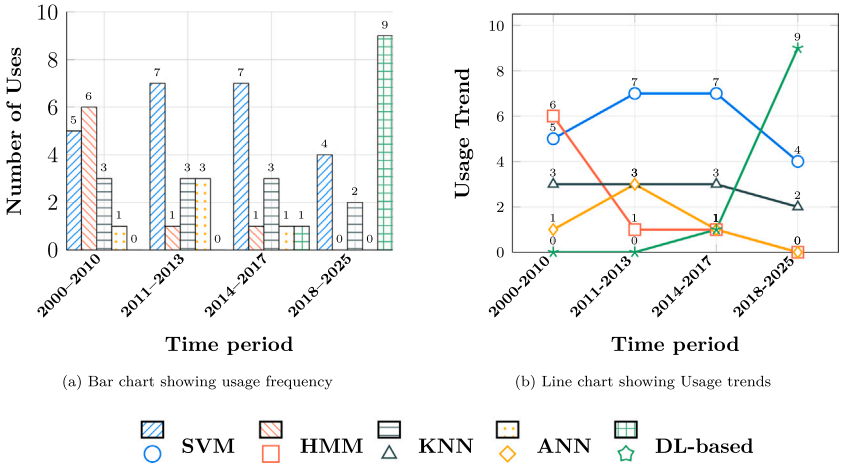
TV-L1 is the third most frequently used optical flow method in the reviewed HAR studies. Although it generally provides high-quality dense flow and preserves motion boundaries well, its higher computational cost limited its early adoption. As hardware improved and deep learning

Table 4
LK optical flow-based HAR studies from 2014 to 2017.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[110]	2014	Features extracted as in [89]	McFIS	user-generated
[111]	2014	Handmade features	KNN	Weizmann, IXMAS
[112]	2014	HOF	Bayesian framework	BEHAVE
[113]	2014	STSP	GMM	KTH, UCF Sports, HMDB 51
[114]	2014	HOOF	K-means	KTH, Weizmann
[115]	2014	HOF	SVM	KTH, Weizmann
[116]	2014	HOF	NBMIM	YouTube
[117]	2015	HOF	KNN	Weizmann
[118]	2015	DGTC features	Fuzzy cloud KNN	KTH, UCF Youtube, HMDB 51
[119]	2015	GWOF	SVM	LIRIS, UT-Interaction
[120]	2016	HOF	SVM	KTH
[121]	2016	HOF	SVM	UTD-MHAD
[122]	2016	MBH [123]	SVM	YouTube
[124]	2016	HMTO	SVM	KTH, UCF 101, HMDB 51
[125]	2016	HOF	HMM	IXMAS
[126]	2017	Features extracted as in [127]	ANN	KTH
[128]	2017	3D HOF	SVM	Huawei/3DLife
DL-based classifiers				
[129]	2017	HOF	3D CNN	KTH, UCF 101

Table 5
LK optical flow-based HAR studies from 2018 to 2025.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[130]	2018	—	—	KTH
[131]	2018	HOF	Random Forest	KTH, Weizmann
[132]	2019	Velocity vector	SVM	KTH, Weizmann
[133]	2019	HOF	SVM	KTH, Weizmann, UCF Sports
[134]	2020	Energy of flow pixels	KNN	KTH, Weizmann, Ballet, UT-Interaction
[135]	2021	HOF	KNN-SVM	UT-Interaction
[136]	2024	Optical flow in corners	SVM	UEX action
DL-based classifiers				
[137]	2019	Handmade features	3D CNN	IXMAS, NewIXMAS
[138]	2019	—	AlexNet	KTH, Weizmann
[139]	2019	VR	CNN	UCF 101, HMDB 51
[140]	2020	—	Two-stream CNN	KTH, UCF 101
[141]	2021	—	Two-stream CNN	HUMAN 3.6 M
[142]	2021	—	CNN	KTH, UCF
[143]	2021	HOF	Two-stream CNN	KTH, UCF 101
[144]	2023	—	TS-CNN-LSTM	NTU RGB + D
[145]	2023	OMR map	CNN	UTD-MHAD, MSR-Daily activity
[146]	2024	—	Two-Stream CNN	UMN, PETS2009



(a) Bar chart showing usage frequency

(b) Line chart showing Usage trends

Fig. 4. Classifier frequency in LK optical flow-based HAR studies across temporal periods.

Table 6
HAR methods based on the Farnebäck optical flow found until 2017.

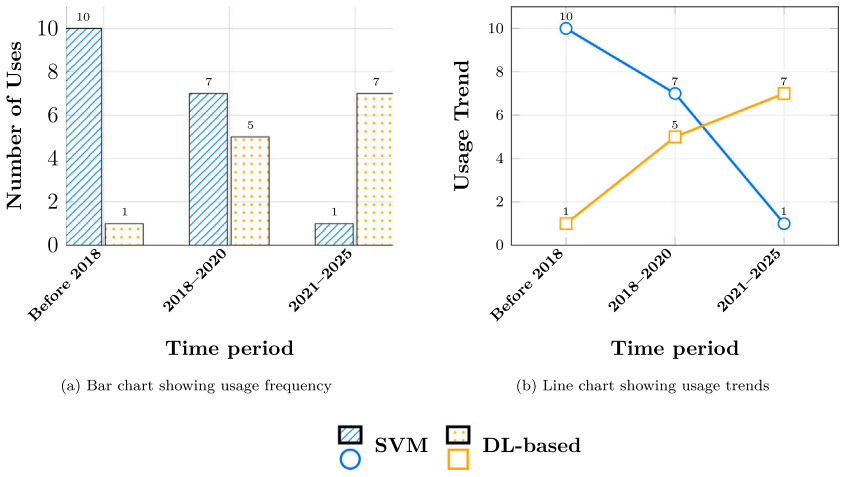
Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[75]	2009	Orientation histograms	KNN	KTH, Weizmann
[147]	2010	aHOF	RBM	Weizmann, ViHASi
[148]	2011	aHOF	RBM	KTH, Weizmann
[149]	2012	MVFI	KNN	self-made
[88]	2012	Orientation histograms	SVM	KTH, Weizmann
[150]	2014	HOF	SVM	KTH, UCF sports, TV human interaction
[151]	2014	HOF	ELM	Hollywood 3D, Olympic Sports, Hollywood2
[152]	2015	HOF	SRC	HMDB51, Hollywood2, UCF101
[153]	2015	HOF	SVM	Hollywood2, Olympic Sports, UCF50, HMDB51
[154]	2015	NMF	Bayesian framework	H3D, VOC2011
[155]	2015	HOF	SVM	HMDB51, UCF50, UCF101
[156]	2015	HOF	SVM	UCF-Sports, YouTube, Olympic Sports, Hollywood2, HMDB51
[157]	2015	HOF	SVM	Hollywood2, UCF50
[158]	2015	HOF	SVM	KTH, ADL
[159]	2016	HOF	SVM	KTH, Olympic
[160]	2016	HOF	SVM	Behave, CAVIAR, Hollywood2, SDHA(2010), UCF101, MediaEval Violent Scenes, Hockey fights, ViolentFlows
[161]	2016	—	SVM	KTH, Weizmann, UCF
[162]	2017	HOF	GMM	HMDB51, Hollywood2, UCF50
[163]	2017	Frobenius norm	GMCP	UCF-Sports, sub-JHMDB, THUMOS13
[164]	2017	HOF	SRC	HMDB51, Olympic Sports, Hollywood2
[165]	2017	HOF	MTL	IXMAS, Northwestern UCLA
[166]	2017	HOF	LCF-LR	HMDB51, UCF101
DL-based classifiers				
[167]	2017	—	Two-stream CNN	Weizmann, UCF Sports

Table 7
HAR methods based on the Farnebäck optical flow found from 2018 to 2020.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[168]	2018	HOF	SVM	UCF50, HMDB51
[169]	2018	HOF	SVM	OlycSpos, HMDB51, UCF50
[170]	2018	HOF	MKL	KTH, UCF sports, YouTube
[171]	2018	HOF	RMTL	MMA
[172]	2018	HOF	SVM	UCF50
[173]	2019	HOF, IGMM	SVM	UCF50, HMDB51
[174]	2019	HOF	GMM	Hollywood 3D
[175]	2020	K-means	KL-divergence	Weizmann, KTH
[176]	2020	HOF	SVM	KTH, YouTube Action
[177]	2020	HOF	SVM	KTH, JHMDB, HMDB51
[178]	2020	1D-CNN	SVM	UTD-MHAD
DL-based classifiers				
[179]	2018	2D ResNet50	3D CNN	JHMDB, PennAction
[180]	2018	—	TSN	NTU RGB + D
[181]	2019	—	Temporal ConvNet	UCF101, HMDB51
[182]	2019	HOF	LSTM	J-HMDB, MPII Cooking
[183]	2020	—	Inception-ResNetv2	UCF101, HMDB51

Table 8
HAR methods based on the Farnebäck optical flow found from 2021 to 2025.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[184]	2021	Divergence and Curl	SVM	UCF Sports, UCF101, HMDB51, JHMDB, KTH
[185]	2021	Gradient of optical flow	KFSENet	UCF Sport, self-made
[186]	2021	HOF	KNN, DL	UTKinect-Action 3D
[187]	2022	Temporal Network	MLP	UCF11, UCF50, HMDB51
[188]	2022	—	Overlap index	SAMM, CASME, CAS(ME)2, CAS(ME)3
[189]	2022	HOF	LSDA	MCAD, IXMAS
[190]	2023	K-means	RLS	MoCA
DL-based classifiers				
[191]	2021	TSN	VFAM	HMDB51
[192]	2021	HOF	Mask-RCNN	UCF101
[193]	2022	—	DL	NTU RGB + D 120
[194]	2023	The Dual-flow CNN	Decision tree	KTH, UCF Sports
[195]	2023	—	1D CNN	Okutama, Drone Action, MOD20
[196]	2024	—	3D CNN	Toyota Smarthome
[197]	2024	HOF	DS-ELM	Olympic Sports, UCF11, Hollywood2, UCF101, self-made



(a) Bar chart showing usage frequency

(b) Line chart showing usage trends

Fig. 5. SVM and DL classifier frequency in Farnebäck optical flow-based HAR studies across temporal periods.

Table 9
HAR methods based on the TV-L1 optical flow approach found in this review.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[198]	2008	Flow map	SVM	Weizmann, KTH
[199]	2017	Temporal ConvNet	SVM	HMDB51, UCF50, UCF101
[178]	2020	1D-CNN	SVM	UTD-MHAD
DL-based classifiers				
[200]	2016	—	Two-stream CNN	UCF101
[201]	2017	—	AdaScan	UCF101, HMDB51
[202]	2017	LSTM	L^2STM	UCF101, HMDB51
[203]	2017	LT-Net	Weighted fusion	UCF101, HMDB51
[204]	2018	Two-stream C3D network	RNN	UCF101, HMDB51
[205]	2018	Dual-stream CNN	LSTM	Place205, UCF101, HMDB51, TRECVID, Open Video Project (OVP), YouTube
[206]	2020	—	Two-stream CNN	UCF101, HMDB51
[207]	2021	PI3D	Score fusion	HMDB-51, SYSU 3D, JHMDB, Sub-JHMDB
[208]	2021	ConvNet	Weighted fusion	NTU RGB + D, SYSU 3D HOI, UWA3D II
[209]	2021	ResNet152	CASDLSTM	UCF-101, HMDB-51
[210]	2022	—	WSRTL	BP4D, DISFA
[211]	2023	DBN	Ensemble classifier	NTU-RGB + D, NTU-RGB + D 120, Northwestern, UCLA, Penn Action
[212]	2023	ResNet-34	MRDFNet	REMOD-ES, THU-READ
[213]	2023	I3D + STA	Weighted fusion	KTH, HMDB51, Weizmann, UCF-101, UCF-Crime, NPDI
[214]	2024	—	Two-stream net [215]	HMDB-51, YUP++, MPII Cooking Activities, Charades

pipelines became more common, its use increased. Table 9 shows that 6 studies used TV-L1 before 2018, 4 studies used it in 2018–2020, and 8 studies used it in 2021–2025. This pattern differs from the decline observed for several other classical methods and suggests that TV-L1 has become more attractive in recent HAR research.

TV-L1 also differs from the other optical flow methods in how it is integrated into HAR pipelines. It is most often used as a direct input to deep learning models rather than as the basis for handcrafted descriptors. Traditional feature extraction methods, such as HOF, appear only rarely with TV-L1. Similarly, while SVM dominates several other optical flow categories, it appears in only three TV-L1 studies, the latest of which was published in 2020. These patterns indicate strong compatibility between TV-L1 flow fields and deep learning classifiers.

4.4. HAR methods based on HS optical flow

HS appears much less frequently than LK, Farnebäck, or TV-L1 in optical flow-based HAR, which makes it difficult to define detailed temporal usage periods. Only 12 reviewed studies used HS (Table 10). Within this small group, SVM was the most frequent classifier and HOF was the most frequent feature representation. Only one study used HS optical flow as input to a deep neural network. The limited adoption of HS is consistent with its known sensitivity to noise and its tendency to oversmooth motion boundaries, both of which reduce its suitability for complex HAR scenarios.

4.5. HAR methods based on unspecified optical flow

This subsection covers the 42 studies that used optical flow but did not specify the particular method. These studies are reported in Tables 11 and 12, separated into pre-2018 and post-2018 intervals. In the first group (Table 11), the pattern resembles the LK, HS, and Farnebäck groups: 18 studies used histogram-based features, and SVM was the most common classifier, appearing in 13 studies. After 2018 (Table 12), 13 of the 21 unspecified-flow studies used deep learning approaches, mainly dual-stream or 3D CNN classifiers. Among the remaining non-deep-learning studies, SVM was again the most frequent classifier and HOF was the most frequent feature representation.

5. Datasets

Datasets are central to the evaluation and comparison of HAR methods. In optical flow-based HAR, datasets must provide video sequences containing diverse actions under sufficiently varied recording conditions. Because HAR is applied in domains such as urban management, security, transportation, sports, and healthcare, datasets differ substantially in scale, environment, viewpoint, modality, and realism.

Table 10
HAR studies that use the HS optical flow approach found in this review.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[94]	2012	HOF	HMM	self-made
[216]	2013	HOF	LDCRF	Weizmann, UT-Interaction
[217]	2013	KNN	SVM	KTH, UCF sports
[218]	2014	MHI	GMM	Weizmann
[219]	2014	HOOFD	SVM	MSR Action
[220]	2015	AAE	SVM	Weizmann, IXMAS
[221]	2016	Divergence and Curl	SVM	KTH, Weizmann, UCF-11, UCF-50, HMDB-51
[222]	2017	HOF	SVM	UCF50, UCF101, HMDB51
[223]	2017	HOF	SVM	Hollywood & Bollywood, UT-interaction
[224]	2020	Handmade	W-KNN	Weizmann, MuHAVi, WUV, MSR, KTH, UCF sports
[225]	2021	—	SVM	KTH, Weizmann, UCF Sports, UCF YouTube
DL-based classifiers				
[226]	2021	—	CNN	KTH, UCF101, HMDB51

Table 11
HAR methods based on an unspecified optical flow before 2018.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[227]	2008	Kinematic Features	MIL	KTH, Weizmann
[228]	2010	Covariance Matrix	EMD	KTH, Weizmann
[229]	2010	HOF	SVM	KTH
[230]	2010	MC-STIP	SVM	KTH, Weizmann, UCF Sports
[231]	2011	HOF	TWSVM	KTH
[232]	2011	HOF, SIFT	SVM	YouTube, HOHA-2
[233]	2011	HOF	SVM	UT-Interaction
[234]	2012	HOF	HMM-MIO	KTH
[235]	2013	PaHOF	HMM	IXMAS
[236]	2014	HOF	SVM	KTH, Weizmann, IXMAS
[237]	2014	HOF	Generative TSI-pLSA	KTH, Weizmann
[238]	2014	HOOF	SVM	KTH
[239]	2014	HOF	Bayesian framework, SVM	HMDB, Web actions, PPMI
[240]	2015	HSGA	K-means	KTH, YouTube, UCF50
[241]	2015	HOF	SVM	uulmMAD
[242]	2015	HOF	SVM	YouTube, UCF Sports
[243]	2015	HOOF	KNN, SVM	Weizmann
[244]	2015	HOF	GMM	YouTube, UCF 50
[245]	2016	HOOF	SVM	UCF Sports, TV-Human Interaction, KTH
[246]	2016	HOF	SVM	IXMAS, WVU
[247]	2017	HOF	ELM	UTD-MHAD

Table 12
HAR methods based on an unspecified optical flow after 2018.

Ref	Year	Features	Classifier	Dataset
Traditional ML classifiers				
[248]	2018	Motion pathway	IncSFA	KTH, Weizmann
[249]	2018	HOF	GMM	KTH, UCF101, HMDB51
[250]	2018	HOF	SVM	DSA, IXMAS, NUMA
[251]	2018	STIP	LDA-Kmeans	KTH, UCF-11, UCF Sports, Hollywood2 action
[252]	2019	BoHOF	SVM	KTH
[253]	2019	RHOF	SVM	KTH, Weizmann
[254]	2021	ResNeXt	SVM	MSR Action 3D, UCF101, HMDB51
[255]	2022	—	LVQ	KTH
DL-based classifiers				
[256]	2019	One-stream CNN	3D CNN	UCF101, HMDB 51
[257]	2019	—	3D CNN	UCF101
[258]	2020	BN-Inception network: MSOF	LSTM	UCF101, HMDB51
[259]	2020	3D CNN	Two-Stream I3D	self-made
[260]	2020	—	CNN, LSTM	UCF101
[261]	2021	Temporal features	STCN	NTU-RGB + D
[262]	2021	ST-GCN	Dual-stream CNN	UCF 101
[263]	2021	HOF	ST-GCN	NTU RGB + D
[264]	2022	HoFG	MRGCN	NTU RGB + D
[265]	2023	LFF	B2C-AFM	Kinetics, NTU-RGBD
[266]	2023	ST-Net	CNN, LSTM	UCF 101, HMDB 51
[267]	2023	—	SWTA Inception-v3	Okutama, DroneAction, MOD20
[268]	2024	—	Two-stream CNN	UCF 101

Table 13
Characteristics and usage frequency of benchmark datasets in the reviewed optical flow-based HAR studies.

Dataset	# Actions	# Videos	Resolution	Type	Environment	Modality	Usage frequency
KTH [271]	6	600	160 × 120	Controlled	Controlled, static camera, uniform background	RGB	58 studies
Weizmann [272]	10	90	180 × 144	Controlled	Controlled, static camera, uniform background	RGB	42 studies
IXMAS [273]	11	1650	390 × 291	Controlled	Controlled, 5 synchronized cameras	RGB multi-view	15 studies
UCF Sports [274]	10	150	720 × 480	In-the-wild	Sports videos, camera motion, varied lighting	RGB	18 studies
Hollywood2 [275]	12	3669	Variable	In-the-wild	Movie scenes, extreme variation in conditions	RGB	16 studies
UT-Interaction [276,277]	6	20	720 × 480	Controlled	Controlled, focuses on human interactions	RGB	9 studies
HMDB51 [278]	51	6766	320 × 240	In-the-wild	Diverse sources, large-scale, varied conditions	RGB	32 studies
UCF50 [279]	50	6681	320 × 240	In-the-wild	YouTube videos, varied quality and conditions	RGB	13 studies
UCF101 [280]	101	13,320	320 × 240	In-the-wild	YouTube videos, large-scale, diverse actions	RGB	35 studies
UTD-MHAD [281]	27	861	512 × 424	Controlled	Controlled, multimodal capture	RGB + D, skeleton	6 studies
NTU RGB + D [282]	60	56,880	Variable	Controlled	Controlled, large-scale, multi-view	RGB + D, skeleton	8 studies

Table 14
Further characteristics of the benchmark datasets relevant to optical flow.

Dataset	Frame rate	Setting	Viewpoint	Multi person activity	Occlusion and camera motion
KTH [271]	25 fps	Indoor and outdoor, plain backgrounds	Single static camera	No, one actor at a time	Little occlusion, static camera
Weizmann [272]	25 fps (50 fps for the robustness set)	Outdoor, plain background	Single static camera	No, one actor at a time	A separate set adds partial occlusion and a moving camera
IXMAS [273]	Not stated	Indoor	Five calibrated cameras, multi-view	No, one actor at a time	Self occlusion across views, fixed cameras
UCF Sports [274]	Not stated	Mostly outdoor sports broadcasts	Varied scenes and viewpoints	Some scenes with other people present	Camera motion and varied lighting
Hollywood2 [275]	Not stated	Indoor and outdoor film scenes	Varied scale, viewpoint and background	Yes, several people in many clips	Strong occlusion and camera motion
UT-Interaction [276,277]	Not stated	Outdoor, fixed setting	Single fixed camera	Yes, two-person interactions	Camera jitter and changing background
HMDB51 [278]	30 fps for all clips	Mixed, from films and the web	Varied, front and back views	Yes, in many clips	Camera motion, varied quality and occlusion
UCF50 [279]	Not stated	Mixed, web videos indoor and outdoor	Varied viewpoints	Yes, in some classes	Camera motion, cluttered background, occlusion
UCF101 [280]	25 fps	Mixed, web videos	Varied, arbitrary views	Yes, including human to human classes	Camera motion, cluttered background, partial occlusion
UTD-MHAD [281]	About 30 fps	Indoor, controlled capture	Single front-facing sensor	No, one subject at a time	Little occlusion, fixed sensor
NTU RGB + D [282]	Not stated	Indoor only	Three cameras at different horizontal angles	Yes, includes mutual actions between two people	Cross view self occlusion, fixed cameras

This section examines the public datasets most frequently used in the reviewed optical flow-based HAR studies, as reported in Tables 2–12.

5.1. Dataset characteristics

HAR datasets can be broadly divided into controlled and realistic datasets. Controlled datasets contain actions recorded under managed conditions, usually with fixed cameras, simple backgrounds, and predefined lighting. Some controlled datasets add artificial noise, lighting variation, or occlusion to approximate more realistic conditions. Realistic datasets, by contrast, are collected from real-world videos and typically include variable lighting, occlusion, camera motion, dynamic backgrounds, and heterogeneous scene content. In controlled datasets, preprocessing is often less demanding and the main emphasis is placed on feature representation and classification. In realistic datasets, however, preprocessing becomes more important because camera motion, background clutter, and environmental variation can strongly affect optical flow estimation [269,270].

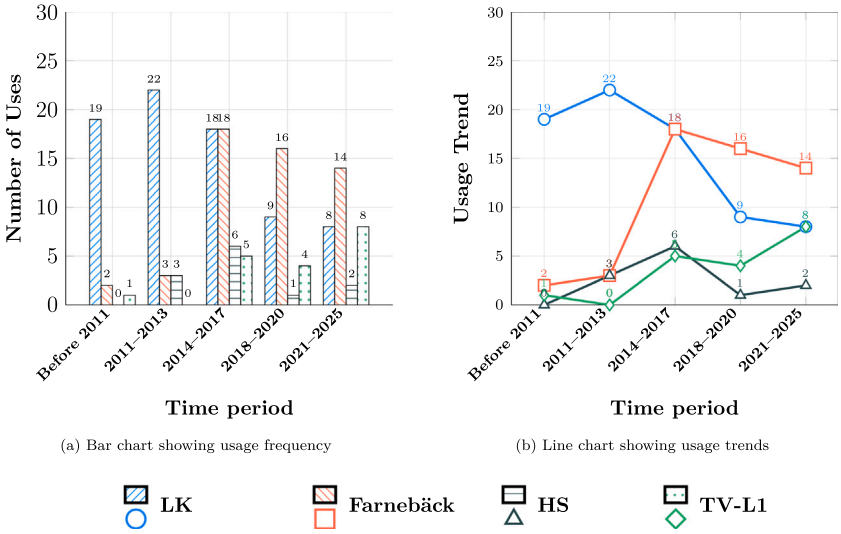
Table 13 compares the datasets that appear most frequently in Tables 2–12, including their characteristics and usage frequency. Table 14 reports additional properties relevant to optical flow, such as frame rate, viewpoint, multi-person activity, occlusion, and camera motion. Studies using multiple datasets were counted once for each dataset. KTH is the most frequently used dataset, with 58 appearances, followed by Weizmann with 42, UCF101 with 35, and HMDB51 with 32. This distribution shows that optical flow-based HAR has historically relied heavily on controlled benchmark datasets, although more recent studies increasingly use larger and more realistic datasets.

5.2. Dataset analysis

KTH and Weizmann are the most frequently used datasets (Table 13) and provide controlled evaluation environments. Their static cameras, uniform backgrounds, and consistent lighting enable isolation of optical flow algorithm performance from environmental confounding factors. The predominance of these datasets in early studies reflects their value for establishing baseline performance and comparing optical flow methods under ideal conditions. However, the simplicity of these datasets limits their ability to evaluate the robustness of HAR systems in real-world conditions. Optical flow methods that perform well on the KTH dataset may fail in scenarios with camera motion or dynamic backgrounds, which are common in realistic deployments.

The low image resolution (160 × 120 for KTH) also masks computational efficiency differences between methods, as processing time remains minimal even for computationally expensive algorithms, such as TV-L1. IXMAS provides controlled conditions while adding multi-view complexity. The synchronised five-camera setup enables evaluation of view-invariant recognition and multi-view fusion strategies. For optical flow evaluation, IXMAS enables assessing whether optical flow features generalise across viewpoints, an important consideration for surveillance applications. UCF101 and HMDB51 represent the most widely adopted realistic datasets in the reviewed studies.

Their diversity in action types, environmental conditions, and video sources enables a more comprehensive evaluation of optical flow robustness than the KTH, Weizmann, and IXMAS datasets. Camera motion, lighting changes, and background clutter in these datasets challenge optical flow algorithms beyond the capabilities of controlled datasets.



(a) Bar chart showing usage frequency

(b) Line chart showing usage trends

Fig. 6. Distribution of HAR studies by optical flow methods across temporal periods.

Hollywood2 introduces additional challenges through movie footage featuring dramatic lighting, intentional camera effects, and editing cuts. The 3669 video clips, drawn from diverse movie scenes, create extreme variation that reveals the drawbacks of optical flow methods. Studies using Hollywood2 often focus on robustness rather than accuracy. UCF Sports provides an intermediate evaluation between controlled and unconstrained datasets. Sports videos feature camera motion and varied outdoor lighting, while focusing on well-defined action categories within relatively structured scenarios. This dataset is a good choice for evaluating optical flow performance under camera motion.

NTU RGB + D and UTD-MHAD, featured in recent studies, facilitate the evaluation of multimodal fusion approaches. These datasets provide synchronised RGB, depth, and skeletal data, enabling comparisons of optical flow with alternative motion representations and the evaluation of fusion strategies that combine optical flow with pose estimation or depth information. The scale of NTU RGB + D (56,880 videos) makes it particularly valuable for training and evaluating DL approaches, which dominate recent work (63% of studies from 2021 to 2025, per Table 17). However, the controlled capture conditions limit the evaluation of robustness to environmental variations despite the dataset's scale.

Pulling these points together, the makeup of a dataset has a direct bearing on which optical flow method performs well and what can be measured fairly. On low-resolution controlled sets such as KTH, every method runs quickly, so the real cost gap between a cheap method like LK and an expensive one like TV-L1 barely shows, and a study can pick the heavier method without paying for it. Camera motion is the single biggest strain on optical flow, and it is largely absent from the controlled sets but common in the realistic ones, such as UCF101, HMDB51 and Hollywood2, which is why methods that look strong on KTH can fall away on this footage. Multi-view sets, such as IXMAS, test whether flow features still hold up across different camera angles, a question that single-view sets cannot ask. The upshot is that no single dataset stresses all the conditions optical flow has to cope with, so a method is best judged on data that matches its intended use rather than on a controlled set alone.

There is also a clear bias in how these datasets have been used, and it shapes what the field thinks it knows. The reviewed studies lean heavily on a few controlled sets, with KTH appearing in 58 studies and Weizmann in 42, far ahead of any realistic set (Table 13). Because these sets are simple, with a static camera, plain background and steady lighting, accuracy on them is high and tightly bunched, often above 95%, which leaves little room to tell methods apart and tends to flatter whatever is used. Progress measured mainly on KTH and Weizmann can therefore appear stronger than it really is when the same method is tested on real footage. The realistic sets carry their own bias, since UCF101, UCF50, and HMDB51 are drawn from YouTube and Hollywood2, respectively, from films (Table 13), so they reflect the kinds of videos that end up online or in cinema rather than everyday scenes many HAR systems are meant for. On top of this, none of the eleven sets was built specifically to test optical flow for activity recognition, with controlled occlusion, varied frame rates or graded camera motion, so the conditions that matter most for flow are rarely isolated. Taken together, these biases mean that headline results should be read with care, and that broader, purpose-built datasets would provide a fairer test of optical flow-based HAR.

6. Discussion

This section presents a comparative analysis of optical flow methods in HAR across different temporal periods. Classification techniques are analysed separately for each optical flow method. The section concludes with the identification of challenges, research gaps and future directions.

6.1. Temporal analysis

Fig. 6 shows the distribution of optical flow methods across the temporal periods considered in this review. Overall, LK and Farnebäck are the most frequently used methods, whereas HS appears least often. Recent years show a decline for most classical methods, with TV-L1 being the main exception. These shifts can be explained by the type of motion representation each method produces and by the changing requirements of the classifiers used in HAR. LK estimates motion sparsely at strong corners, which suited early histogram-based and computationally light pipelines. This also explains its continued use in real-time and edge applications. However, sparse motion gives deep networks less spatial information than dense flow fields, contributing to the decline of LK as learned classifiers became dominant. Farnebäck produces dense flow and can support both histogram descriptors and pretrained network features, which explains its strong presence in the middle of the review period. TV-L1 has remained comparatively strong because it produces dense flow with sharper motion boundaries and improved robustness to illumination variation [44], properties that make it suitable for convolutional architectures. HS has remained marginal because its global smoothness constraint can oversmooth motion boundaries and makes it sensitive to noise [39], particularly in cluttered or realistic videos.

Table 15
Strengths and weaknesses of the four knowledge-based optical flow methods in HAR.

Method	Output	Strengths	Weaknesses and failure cases	Best suited to
HS [41]	Dense field, global smoothness	Smooth whole body motion, low computing cost	Sensitive to noise and poor in changing light, blurs motion boundaries through global smoothing [39]	Stable, controlled scenes; least used in the review (Table 10)
LK [42]	Sparse, at strong corners	Fast and light, well suited to real-time and edge use	Sparse output gives a deep network less to work with, and small windows miss large motion	Real time and resource-limited HAR (Tables 2 to 5)
Farnebäck [43]	Dense field from polynomial fitting	Stable dense flow across a range of motion sizes, fits both histograms and pretrained networks	Heavier iterative computation, and the dense map can blur fine detail	Mixed fine and gross movements (Tables 6 to 8)
TV-L1 [44]	Dense, sharp at boundaries	Keeps motion boundaries and copes with changing light, a strong match for deep networks	The slowest of the four, which long held back its use	Boundary sensitive and deep learning HAR, growing in recent years (Table 9)

6.2. Comparative characteristics of optical flow methods

Trends in optical flow use reflect trade-offs among computational efficiency, accuracy, motion-detail preservation, and robustness to environmental variation. HS produces dense flow fields through a global smoothness constraint, which can be useful for broad whole-body motion analysis [41]. However, its sensitivity to noise and tendency to blur motion boundaries limit its applicability in realistic HAR scenarios [39]. Its limited presence in recent studies (Table 10) suggests that low computational cost alone is no longer sufficient to justify its use when more robust alternatives are available.

LK is computationally efficient and has therefore been widely adopted in real-time HAR systems. Studies using LK generally seek a practical balance between speed and recognition performance [70,85]. TV-L1 addresses several limitations of earlier methods by better preserving motion discontinuities and handling illumination changes [44]. Its higher computational cost is often justified in applications requiring accurate motion boundaries, especially under occlusion or when dense motion fields are used as deep learning inputs. The strong association between TV-L1 and deep learning in the reviewed studies (Table 9), where 15 of 18 studies use deep classifiers, supports this interpretation.

Farnebäck estimates dense flow through polynomial expansion and captures motion at multiple scales [43]. Although its iterative refinement is computationally heavier than LK, it produces stable flow estimates across different motion magnitudes. Farnebäck is therefore well suited to activities involving both fine and large-scale movements [153,166]. Table 15 summarises the strengths, weaknesses, and most suitable HAR settings for the four knowledge-based optical flow methods.

Performance values must be interpreted with caution, because they are as dependent on datasets and protocols as on optical flow methods. The reviewed studies most often report accuracy and confusion matrices, but these values cannot be directly aligned across studies because datasets, train/test splits, augmentation strategies, and evaluation protocols differ substantially. The same optical flow method may exceed 95% accuracy on a controlled dataset such as KTH and perform much less strongly on a realistic dataset such as UCF101. A single accuracy value therefore often reflects dataset difficulty as much as method quality.

There is a deeper point too: the error measure used to rank flow methods on flow benchmarks does not track recognition accuracy, and what helps recognition most is correct motion at boundaries and at small displacements [50]. For these reasons, this review does not present a single league table of accuracy or F1 across the studies, since it would mislead more than it would inform. The safer reading is the one drawn out above, matching the qualities of each method to the demands of the task. Drawing the strengths, costs, and usage patterns together, Table 16 provides a brief decision guide, linking common HAR settings to a sensible first choice of optical flow method.

Table 16
A practical guide to selecting an optical flow method for HAR by setting.

Setting or main constraint	Suggested first choice	Reasoning
Real-time use on limited or edge hardware	LK	Fast and light, and the most common choice for such work in the reviewed studies (Tables 2 to 5)
Accuracy with a deep network on a large dataset	TV-L1	Sharp dense flow that keeps boundaries and feeds deep networks well, the strongest pairing with deep learning (Table 9)
Mixed fine and gross movements	Farnebäck	Stable dense flow across motion sizes, and it also suits pretrained feature extraction (Tables 6 to 8)
Stable, controlled scenes with steady light	HS or LK	Low cost methods are enough when the scene is simple, so the extra cost of dense methods is hard to justify
Changing light or strong occlusion	TV-L1	Copes better with changing light and keeps motion boundaries under occlusion [44]
Very large datasets or end-to-end training	Learned or embedded motion	Avoids the cost of precomputing flow for every frame, the direction the newest work is taking [37,283]

6.3. Feature extraction and classification

HOF and HOOF are the most frequently used optical-flow-derived feature representations in the reviewed literature; no other handcrafted feature family appears with comparable frequency. TV-L1 is the main exception to this pattern, because the reviewed TV-L1 studies rarely use traditional feature extraction and instead employ the flow field mainly as input to deep learning models.

At the classification stage, two groups dominate: SVM and deep learning-based methods. HMM, KNN, and ANN also appear, but less consistently. Fig. 7 shows the number of classifier uses associated with each optical flow method across time intervals. The figure highlights the growth of deep learning approaches and the decline of HMM and ANN in more recent periods.

6.4. Quantitative analysis of methodological transitions

All ML and DL classifiers reported in Tables 2–12 were analysed to quantify the transition from traditional ML to deep learning. Table 17 summarises this transition over time. Traditional ML dominated until 2017, accounting for 114 of 122 classifier uses (93.4%) before 2018. After 2018, deep learning methods accounted for 49 of 84 classifier uses

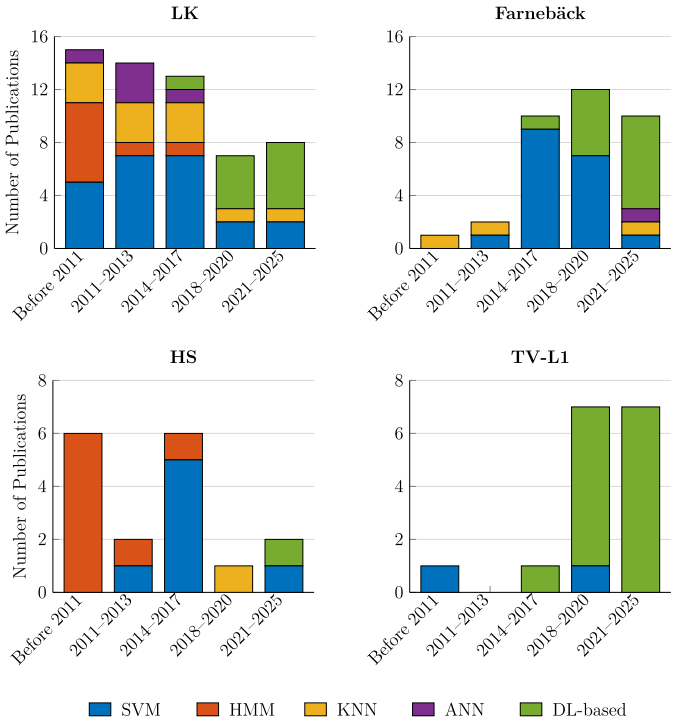


Fig. 7. Number of classifiers used with each optical flow method across different time intervals.

Table 17
Transition from traditional ML to DL classifiers in HAR studies.

Time period	Traditional ML	DL-based	Total	DL percentage
Before 2011	29	0	29	0%
2011 to 2013	33	0	33	0%
2014 to 2017	52	8	60	13%
2018 to 2020	17	18	35	51%
2021 to 2025	18	31	49	63%
Total	149	57	206	28%

Table 18
Distribution of the reviewed studies by the optical flow method, extracted from Tables 2 through Table 12.

Optical flow type	Total Studies	Peak Period	Peak Count
LK (Tables 2 to 5)	76	2011 to 2013	22 studies
Farnebäck (Tables 6 to 8)	53	2014 to 2017	18 studies
TV-L1 (Table 9)	18	2021 to 2025	8 studies
HS (Table 10)	12	2014 to 2017	6 studies
Unspecified (Tables 11 and 12)	42	Before 2018	21 studies
Total	201	—	—

(58.3%), indicating a clear methodological shift. Tables 2–12, together with Section 3.2, contain 209 entries representing 205 unique studies. The difference results from four studies that used more than one optical flow method and therefore appear in more than one table, for example studies combining Lucas–Kanade and Farnebäck.

In Tables 17 and 18, which report how often each classifier and each optical flow method appears, a study that used more than one method is counted in more than one row, and a study that used a deep network for feature extraction together with a traditional classifier is counted in both the Traditional ML and the DL columns. Fig. 8 shows this shift over time, with deep learning methods taking over after 2018.

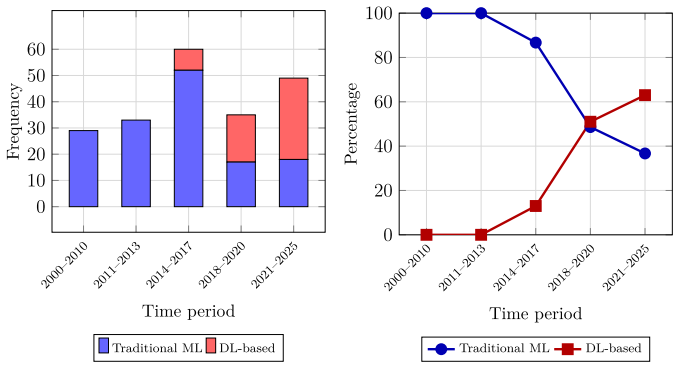


Fig. 8. Temporal transition from traditional ML to DL approaches in HAR studies.

This transition pattern varies substantially by optical flow method. TV-L1 shows the strongest association with deep learning, with 15 of its 18 studies using neural architectures. Farnebäck shifted from SVM dominance to deep learning after 2020, whereas LK retained greater classifier diversity while still showing increased use of deep learning, rising from one study in 2014–2017 to nine studies in 2018–2025. The transition also extends to feature extraction. Before 2018, researchers typically extracted HOF or HOOF descriptors from optical flow and passed them to traditional classifiers; these histogram-based features appear in 72 reviewed studies. After 2018, optical flow was increasingly used directly as input to deep learning classifiers, eliminating the intermediate handcrafted feature step. This direct-input strategy appeared in 20 studies after 2018: 6 using LK, 6 using Farnebäck, 4 using unspecified methods, 3 using TV-L1, and 1 using HS. This change in feature representation is illustrated in Fig. A.13 in Appendix A.

Accuracy and confusion matrices are the most commonly reported performance metrics, but direct comparison across studies would be misleading because datasets and protocols differ. Controlled datasets such as KTH often yield accuracies above 95%, whereas more complex datasets such as UCF101 usually produce lower values. Cross-validation strategies, train/test splits, and augmentation procedures also vary widely. For this reason, the review emphasises usage patterns, methodological combinations, and qualitative suitability rather than a single quantitative performance ranking.

The quantitative patterns in Table 18, Fig. 6, and 7 reveal three connected shifts: movement from LK and HS towards Farnebäck and TV-L1, replacement of traditional ML classifiers by deep learning after 2017, and decline of histogram-based features in favour of direct optical flow input to neural networks. These shifts are linked. Large labelled video datasets, improved hardware, and the attraction of end-to-end learning enabled deep networks to use dense flow fields directly. Dense methods such as Farnebäck and TV-L1 therefore became more suitable for newer pipelines than sparse LK flow, particularly because TV-L1 preserves motion boundaries. Dataset choice followed the same trend: early traditional studies relied heavily on controlled datasets such as KTH and Weizmann, whereas recent deep learning studies increasingly use larger and more realistic datasets such as UCF101, HMDB51, and NTU RGB + D. Thus, the transition is not only a change in classifier type; it is a coordinated shift across dataset scale, motion representation, feature extraction, and classification.

6.5. *Proposed taxonomy of optical flow-based HAR pipelines*

The reviewed studies vary along several connected dimensions rather than along a single methodological axis. To make these relationships explicit, this review proposes a taxonomy that organises optical flow-based HAR pipelines around three axes. The first axis is the way optical flow is computed: either by a knowledge-based method, such as HS,

Table 19
Proposed taxonomy of optical flow-based HAR pipelines.

Family	Motion representation route	Typical flow method	Typical classifier	Example studies
A. Handcrafted motion route (mainly before 2018)	Handcrafted descriptors such as HOF and HOOF	Knowledge-based, mostly LK and Farnebäck	Traditional ML (SVM, HMM, KNN, ANN)	Tables 2 to 4, 6, 10, 11
B. Direct dense flow route (mainly after 2018)	Dense flow field used as direct input	Knowledge-based, mostly Farnebäck, TV-L1 and LK	Deep networks (CNN, two-stream, 3D CNN, LSTM)	Tables 5, 7–9, 12
C. Pretrained feature route (2018 onwards)	Flow passed through a pre-trained network for feature extraction	Knowledge-based, Farnebäck	Deep features with a traditional or deep classifier	Five Farnebäck studies in Tables 7 and 8
D. Embedded motion route (recent and emerging)	Motion learned inside the network, with no separate flow	Learning based, or none precomputed	Deep networks (recurrent, two-stream)	[37,38,283,284]

LK, Farnebäck, or TV-L1, or by a learning-based estimator, such as the FlowNet family and later deep models. The second axis is the way the flow is represented for recognition: handcrafted descriptors such as HOF and HOOF, dense flow fields used directly as neural-network input, flow passed through a pretrained network for feature extraction, or motion embedded within the recognition network so that no separate flow computation is required. The third axis is the classifier family: traditional classifiers such as SVM, HMM, KNN, and ANN, or deep neural architectures.

Reading the reviewed literature through these axes yields four main pipeline families, summarised in Table 19. Family A, the handcrafted motion route, combines descriptors such as HOF and HOOF with traditional classifiers and covers most early studies, including the 72 studies that used histogram features before 2018. Family B, the direct dense flow route, feeds dense flow fields directly into deep networks and becomes prominent after 2018, accounting for the 20 studies that used flow as a direct neural input. Family C, the pretrained feature route, passes optical flow through a pretrained network before final classification; in this review, it appears mainly in Farnebäck-based studies. Family D, the embedded motion route, learns motion inside the recognition model and removes the separate precomputed flow stage, as in hidden two-stream, representation-flow, and cross-correlation models. Although Family D lies partly outside the four knowledge-based methods that anchor the systematic analysis, it represents an important emerging direction. This taxonomy helps identify well-explored combinations, such as LK with histogram features and SVM, as well as underexplored combinations, such as TV-L1 outside deep learning.

6.6. Synthesis

The evolution of optical flow use in HAR reflects both technological progress and changing recognition architectures. Early HAR systems prioritised computational efficiency, which explains the dominance of LK, and to a lesser extent HS, during 2000–2013 (Tables 2 and 3) and the continued but limited use of HS across the review period (Table 10). As hardware improved, researchers increasingly adopted more computationally demanding methods that provide denser and more informative motion fields.

The shift towards Farnebäck and TV-L1 after 2018 coincides with the rise of deep learning in HAR. These methods produce dense flow fields that neural networks can process directly, reducing reliance on intermediate histogram descriptors and preserving more spatial motion information. Application requirements also shape method choice. Healthcare and rehabilitation settings, which require precise motion tracking, increasingly favour TV-L1 despite its computational cost. Surveillance systems with strict real-time constraints continue to use LK, often with traditional classifiers. The strong pairing between TV-L1 and deep learning (Table 9) suggests that TV-L1 provides motion representations well suited to convolutional processing, particularly because it preserves motion boundaries and is comparatively robust to illumination changes.

The overall evidence indicates that no single optical flow method is optimal for all HAR scenarios. Method selection should consider motion characteristics, environmental conditions, computational constraints, and the way the flow representation will be integrated with the classifier.

6.7. Is optical flow still needed in modern HAR?

A central question for modern HAR is whether explicit optical flow is still needed when strong models can learn directly from raw RGB frames. Recent video architectures provide an important context for this question. Pure RGB transformer models, including Video Swin Transformer, TimeSformer, ViViT, and the Multiscale Vision Transformer, achieve strong benchmark performance without a separate optical flow stream [26–29]. A related line of work preserves the idea of motion modelling but learns it inside the recognition network rather than computing flow in advance. Hidden two-stream networks predict actions from raw frames while learning motion on the fly, reducing the cost of the conventional two-stage pipeline that precomputes flow [283]. Representation-flow layers embed an optical-flow-like operation inside the network and train it end-to-end for recognition [284]. The cross-correlation models discussed earlier follow the same direction by incorporating motion estimation into recurrent units [37,38]. The main motivation behind this shift is computational and memory cost, since computing and storing optical flow for every frame of a large dataset is expensive [283].

Nevertheless, optical flow has not been retired. The two-stream form of I3D still benefits from a dedicated flow stream alongside its appearance stream [24], and recent transformer-based HAR work continues to incorporate flow as an explicit input [32]. One reason is that optical flow is relatively insensitive to appearance, which can improve robustness when clothing, lighting, and background vary across clips [50]. Precomputed flow also remains useful in practical settings such as small or controlled datasets and real-time systems with limited hardware, where the reviewed studies show that classical methods such as LK are still used (Tables 2–5). Thus, optical flow is no longer the only way to represent motion in HAR, and on large labelled datasets RGB-only models or networks that learn motion internally can match or exceed older two-stream pipelines. Even so, explicit flow remains valuable when appearance invariance, limited data, interpretability, or constrained computation are central concerns. For this reason, a systematic account of optical flow methods and their appropriate use cases remains relevant to HAR.

6.8. Application domains

This subsection summarises the main application domains in which optical flow-based HAR is used, with emphasis on surveillance, monitoring, healthcare, and rehabilitation. In surveillance, optical flow-based HAR supports threat assessment, abnormal-behaviour detection, and automated alert generation. Studies using datasets such as CAVIAR [97], BEHAVE [112], UCF-Crime [213], and PETS2009 [146] address scenarios in which systems must detect unusual behaviours, assess risk, and support timely responses. When several cameras observe the same scene,

decision fusion becomes important. Studies using multi-view datasets such as IXMAS [84,137] and Northwestern UCLA [165,211] illustrate how optical flow can be computed from different viewpoints and combined to improve robustness to occlusion and viewpoint variation.

In smart healthcare and rehabilitation, optical flow-based HAR can contribute to patient monitoring, fall detection, movement assessment, and clinical decision support. Studies using datasets such as UTD-MHAD [121,145,178], MHAD [108], and MoCA [190] focus on scenarios requiring automated assessment of patient movements. Continuous monitoring systems aim to detect falls, unusual patterns, or activities that may indicate medical distress. Rehabilitation systems compare patient movements with reference patterns to identify deviations from correct technique. These applications require optical flow estimates that remain robust under varied lighting, camera placement, and partial occlusion.

6.9. Challenges, research gaps, and future directions

Systematic analysis of the studies presented in Tables 2 through Table 12 reveals several challenges and research gaps in optical flow-based HAR systems. This section presents current limitations observed across the reviewed literature and suggests research directions to address these gaps.

6.9.1. Current limitations

Three main limitations identified in the reviewed studies affect the practical deployment of optical flow-based HAR systems.

Computational efficiency. The computational cost of optical flow extraction remains a significant barrier to real-time and edge deployment, as discussed in Section 3. TV-L1 (Table 9) preserves motion boundaries more effectively than several alternatives, but it requires substantially more computation than LK (Tables 2–5). This explains the continued use of LK in recent real-time and edge applications despite the representational advantages of TV-L1. Studies targeting resource-constrained hardware, including edge devices, mobile platforms, and embedded systems, predominantly selected LK over TV-L1 or learning-based optical flow. This indicates that computational and memory constraints still strongly influence method selection in practical HAR systems.

- **Robustness to environmental variations.** Performance degradation under camera motion, occlusion, and dynamic backgrounds appeared across studies in Tables 2 through Table 12. These environmental challenges necessitate additional preprocessing, which imposes an extra computational burden and impacts system response time. Sensitivity to lighting variations, mentioned in Section 1 as a fundamental challenge, remains inadequately addressed. Although TV-L1 demonstrates some robustness to illumination changes (discussed in Section 6.6), no single optical flow method proved uniformly robust across all environmental conditions documented in the datasets (Table 13).
- **Method selection and integration.** HAR performance sensitivity to optical flow methods, addressed through the comparative analysis in Section 6.2, indicates insufficient understanding of method-task matching. Patterns in Tables 2 through Table 12 demonstrate that researchers often select optical flow methods based on computational constraints or prior work rather than systematic sensitivity analysis. The transition from traditional classifiers to DL approaches (Tables 17 and Fig. 8) has occurred largely independently of the optical flow method evolution, indicating opportunities for enhancing optical flow efficiency, in both computational cost and accuracy, in HAR systems.

6.9.2. Research directions

Based on the above limitations and the trends identified throughout this review, eight research directions are proposed. Table 20 summarises these directions, linking each to current limitations.

Table 20
Research roadmap linking current limitations to proposed future directions.

Current limitation	Proposed direction
Computational efficiency	Design Optical Flow Algorithms for HAR Systems, with emphasis on computational efficiency, memory consumption, and accuracy.
Robustness to environmental variations	Enhancing robustness of current optical flow methods by proposing new representations or adding pre- and post-processing steps, while considering computational efficiency, memory consumption, and accuracy.
Method selection and integration	Incorporation of sensitivity analysis or ablation studies to identify optimal optical flow approaches in HAR systems.
Energy efficiency	Design of energy-efficient optical flow methods specifically for HAR systems.
Condition-dependent effectiveness	Designing HAR systems with the ability to adapt optical flow type based on input video characteristics to increase efficiency.
No DL-based optical flow for HAR	Development and training of DL-based optical flow specifically for HAR systems.
Single-modality limitations	Employing multimodal data and fusion-based architectures. Studies using RGB + D datasets (NTU RGB + D in Tables 5, 8 and 9) demonstrate superior robustness compared to RGB-only approaches.
Labelled data requirements	Employing unsupervised and self-supervised learning approaches.

Efficient preprocessing approaches, including optimal keyframe selection strategies [20], can reduce computational burden while preserving essential motion features for recognition.

6.10. Limitations of this review

Several limitations of this review should be stated explicitly. The search covered IEEE, Elsevier, and Springer, which include a large share of the relevant literature but not all possible venues; consequently, studies published only in other databases, workshops, repositories, or preprint servers may have been missed. The review was also restricted to English-language articles, which may exclude relevant work published in other languages. Because the focus is optical flow, studies that perform activity recognition using other motion cues alone, such as skeletons or depth data without optical flow, were outside the intended scope. The counts and trends reported here depend on how each study described its method, and several articles did not specify the optical flow formulation used; these were therefore placed in the unspecified category. Finally, as discussed above, reported accuracies could not be compared directly across studies because datasets, protocols, and evaluation settings differ substantially. The review therefore emphasises usage patterns, methodological combinations, and method suitability rather than a single ranking of recognition performance. These limitations define the scope within which the findings should be interpreted, but they do not alter the main trends identified.

7. Conclusion

This systematic review examined how optical flow methods have been used in video-based HAR over the past 25 years. By analysing 205 studies retrieved from IEEE, Elsevier, and Springer databases, the review traced the evolution from early pipelines dominated by LK and traditional classifiers towards more recent systems using Farnebäck, TV-L1, and deep learning architectures. The results show a clear methodological transition after 2017: traditional classifiers such as SVM and HMM progressively gave way to deep neural networks, and histogram-based representations such as HOF and HOOF were increasingly replaced by the direct use of optical flow fields as neural-network inputs.

The comparative analysis shows that no single optical flow method is best for all HAR scenarios. LK remains attractive when speed, simplicity, and resource efficiency are critical, especially in real-time or edge settings. Farnebäck provides dense and stable flow across different motion scales and supports both traditional and deep feature-extraction

pipelines. TV-L1 is computationally heavier but is particularly compatible with deep learning because it produces dense flow fields with well-preserved motion boundaries. HS appears less frequently because its global smoothness assumption can over-smooth motion and reduce robustness in realistic scenes.

Several challenges remain open, including the design of optical flow methods tailored specifically to activity recognition, improved robustness under camera motion, occlusion, illumination changes, and dynamic backgrounds, as well as more efficient implementations for embedded and edge deployment. The proposed taxonomy and research roadmap provide a structured basis for comparing existing pipelines and for guiding future work on optical flow-based HAR.

CRediT authorship contribution statement

Abdorreza Alavi Gharahbagh: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Vahid Hajihashemi:** Writing – original draft, Visualization, Investigation, Formal analysis. **J.J.M. Machado:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis. **João Manuel R.S. Tavares:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This article is partially a result of the project “Sensitive Industry”, with reference SIFN-01-9999-FN-182852, co-funded by the European Regional Development Fund (ERDF), through the Operational Programme for Competitiveness and Internationalisation (COMPETE 2020), under the PORTUGAL 2020 Partnership Agreement. The second author would like to thank “Fundação para a Ciência e a Tecnologia (FCT)”, in Portugal, for his PhD grant, reference 2021.08660.BD.

Appendix A. Optical flow formulations and pseudocodes

This appendix gathers the detailed formulations and the pseudocode for the four knowledge-based optical flow methods reviewed in Section 3. The main text keeps the shared starting point, namely the brightness constancy assumption and its gradient form in Eqs. (1) to (4), while the method specific derivations and the algorithms are collected here so as not to interrupt the flow of the review. Each method below begins from Eq. (3), the brightness constancy constraint, which on its own has infinitely many solutions and therefore requires the extra assumptions that set the four methods apart. The equation and algorithm numbers are continuous with the main text, so a reference such as Eq. (A.5) or Algorithm 2 points to the same item wherever it is cited.

Horn Schunck method

In the HS method, a global smoothness constraint is added to Eq. (3), which gives:

$$\int\int\left(\left(I_x u+I_y v+I_t\right)^2+\lambda^2\left(\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2\right)\right)dxdy, \tag{A.1}$$

where λ represents the smoothing term, which is assumed to be 0.5 in [41]. To find the solution, the above expression must be minimised. Using the definition of derivatives, u and v can be calculated iteratively as:

$$\begin{aligned} u_p^{k+1} &= \bar{u}_p^k-\frac{I_x[I_x\bar{u}_p^k+I_y\bar{v}_p^k+I_t]}{\lambda^2+I_x^2+I_y^2},\\ v_p^{k+1} &= \bar{v}_p^k-\frac{I_y[I_x\bar{u}_p^k+I_y\bar{v}_p^k+I_t]}{\lambda^2+I_x^2+I_y^2}, \end{aligned} \tag{A.2}$$

where k is the iteration index, $[u_p^k \quad v_p^k]$ are the optical flow vector estimates for the pixel p in iteration k, and $[\bar{u}_p^k \quad \bar{v}_p^k]$ are the neighbourhood averages of $[u_p^k \quad v_p^k]$.

The pseudocode of the HS Optical Flow method is given in Algorithm 1, and its flowchart is shown in Fig. A.9.

Lucas Kanade method

In the LK method, the most commonly used approach in HAR systems, Eq. (3) is minimised using a cost function of the form:

$$\int\int W^2\left(I_x u+I_y v+I_t\right)^2 dxdy. \tag{A.3}$$

In a digital image, Eq. (A.3) is converted to [42,286–290]:

$$\sum_{p\in \text{Image}} W_p^2\left(I_x(p)u_p+I_y(p)v_p+I_t(p)\right)^2, \tag{A.4}$$

where W represents the window function. It can be shown that, for minimising Eq. (A.4), the following condition must be satisfied [286,291]:

$$\begin{bmatrix} \sum W^2 I_x^2 & \sum W^2 I_x I_y \\ \sum W^2 I_y I_x & \sum W^2 I_y^2 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} \sum W^2 I_x I_t \\ \sum W^2 I_y I_t \end{bmatrix}, \tag{A.5}$$

Algorithm 1: HS Optical Flow algorithm [285].

Input: Two consecutive frames I_1 and I_2
Output: Optical flow $\vec{v}$
Data: Number of iterations N_{iter}; Smoothing term λ; Initial flow $\vec{v}^0$
Compute I_x, I_y from I_1 using the Sobel operator and its transpose [285]
Compute $I_t \leftarrow I_2 - I_1$
for $k \leftarrow 0$ **to** $N_{iter} - 1$ **do**
  if $k = 0$ **then**
    Initialize u and v:
    $[u^0 \quad v^0] \leftarrow \vec{v}^0$
  Compute local averages over each pixel's 4-neighborhood $\mathcal{N}(p)$:
  $\bar{u}_p^k \leftarrow \frac{1}{4}\sum_{q\in\mathcal{N}(p)} u_q^k, \quad \bar{v}_p^k \leftarrow \frac{1}{4}\sum_{q\in\mathcal{N}(p)} v_q^k$
  Compute $d \leftarrow I_x\bar{u}^k + I_y\bar{v}^k + I_t$
  foreach *pixel p* **do**
    Update flow:
    $u_p^{k+1} \leftarrow \bar{u}_p^k - \frac{I_x(p)\,d_p}{\lambda^2 + I_x(p)^2 + I_y(p)^2}$
    $v_p^{k+1} \leftarrow \bar{v}_p^k - \frac{I_y(p)\,d_p}{\lambda^2 + I_x(p)^2 + I_y(p)^2}$
$\vec{v} \leftarrow [u \quad v]$
return *Optical flow* $\vec{v}$

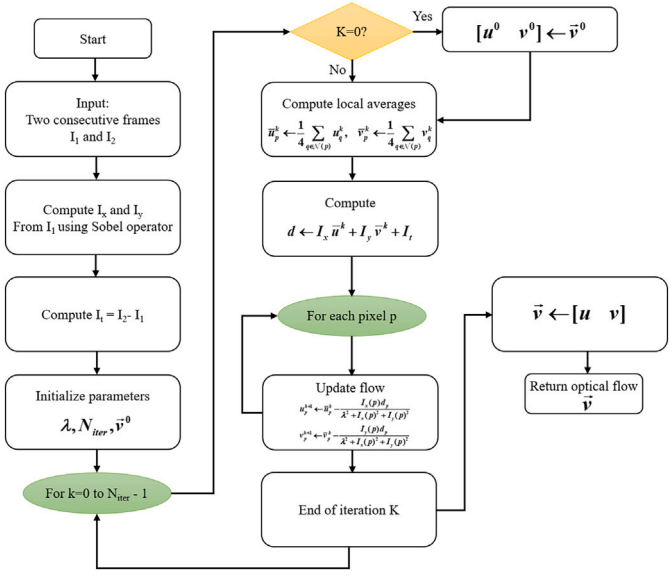


Fig. A.9. Flowchart of the HS optical flow method.

where I_x, I_y, I_t are spatial and temporal derivatives of I defined in Eq. (4). Eq. (A.5) must be solved for each pixel to obtain u and v. For simplicity, the matrix A is defined as:

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \sum W^2 I_x^2 & \sum W^2 I_x I_y \\ \sum W^2 I_y I_x & \sum W^2 I_y^2 \end{bmatrix}. \tag{A.6}$$

The eigenvalues of A can be derived as follows:

$$\lambda_1 = \frac{a+c}{2} + \frac{\sqrt{4b^2+(a-c)^2}}{2}, \quad \lambda_2 = \frac{a+c}{2} - \frac{\sqrt{4b^2+(a-c)^2}}{2}, \tag{A.7}$$

which are always real values and $\lambda_1 > \lambda_2$. The eigenvalues are compared with a noise reduction threshold τ. According to [39], λ_1 and λ_2 are important variables in computing LK Optical Flow.

Algorithm 2: LK Optical Flow algorithm.

Input: Two consecutive frames I_1 and I_2
Output: Optical flow $\vec{v}$
Data: Mask $W = \begin{bmatrix} \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16} \end{bmatrix}$, Noise threshold τ
Compute I_x, I_y from I_1 using following 1D formula:
$\frac{\partial f}{\partial k} = \frac{-f(k-2)+8f(k-1)-8f(k+1)+f(k+2)}{12}$
Compute $I_t \leftarrow I_2 - I_1$
foreach *pixel p* **do**
  Compute a, b, c using Eq. (A.6)
  Compute λ_1, λ_2 using Eq. (A.7)
  if $\lambda_1 > \lambda_2 \geq \tau$ **then**
    Solve Eq. (A.5) using Cramer's Rule to find u_p and v_p
  else if $\lambda_1 \geq \tau$ ***and*** $\lambda_2 < \tau$ **then**
    The gradient flow is normalised to find u_p and v_p
  else
    Set $u_p = 0$ and $v_p = 0$
  end
end
$\vec{v} \leftarrow [u \quad v]$
return *Optical flow* $\vec{v}$

The pseudocode and flowchart of LK Optical Flow are given in Algorithm 2 and Fig. A.10, respectively.

TV-L1 method

In the TV-L1 method, Eq. (3) is minimised using a cost function of the form [44]:

$$\begin{aligned} &\int_\Omega [\lambda \phi(I_1(p) - I_2(p + u dt)) + \psi(u, \nabla u, \ldots)] dx, \\ &\int_\Omega [\lambda \phi(I_1(p) - I_2(p + v dt)) + \psi(v, \nabla v, \ldots)] dy, \end{aligned} \tag{A.8}$$

where I_1 and I_2 are two consecutive frames with a time difference dt, and u, v represent the optical flow vector $\vec{v}$ elements. The function ϕ is defined as the image data fidelity, ψ corresponds to the regularisation force, and λ controls the balance between the fidelity term and the regularisation force. By choosing $\phi = (\nabla u)^2$, Eq. (A.8) leads to the HS method. In the TV-L1 method, ϕ is || (absolute function), and Eq. (A.8) simplifies to:

$$\begin{aligned} E &= \int_\Omega \left(\lambda |I_1(p) - I_2(p + u dt)| + |\nabla u|\right) dx, \\ E &= \int_\Omega \left(\lambda |I_1(p) - I_2(p + v dt)| + |\nabla v|\right) dy. \end{aligned} \tag{A.9}$$

Although this equation appears very simple, the absolute function is not always differentiable, so an iterative algorithm of the following form is required to compute $\vec{v}$ [292]. The pseudocode of TV–$L1$ Optical Flow is given in Algorithm 3.

The function TH is defined as:

$$\mathrm{TH}(\vec{v}, \vec{v}^0) = \begin{cases} \lambda\, \theta\, \nabla I_2\left(p + \vec{v}^0 dt\right), & \text{if } \rho(\vec{v}) < -\lambda\, \theta \left|\nabla I_2\left(p + \vec{v}^0 dt\right)\right|^2, \\ -\lambda\, \theta\, \nabla I_2\left(p + \vec{v}^0 dt\right), & \text{if } \rho(\vec{v}) > \lambda\, \theta \left|\nabla I_2\left(p + \vec{v}^0 dt\right)\right|^2, \\ \dfrac{\rho(\vec{v})\, \nabla I_2\left(p + \vec{v}^0 dt\right)}{\left|\nabla I_2\left(p + \vec{v}^0 dt\right)\right|^2} & \text{otherwise,} \end{cases} \tag{A.10}$$

where:

$$\rho(\vec{v}) = \nabla I_2\left(p + \vec{v}^0 dt\right) \cdot \left(\vec{v} - \vec{v}^0\right) + I_2\left(p + \vec{v}^0 dt\right) - I_1(p), \tag{A.11}$$

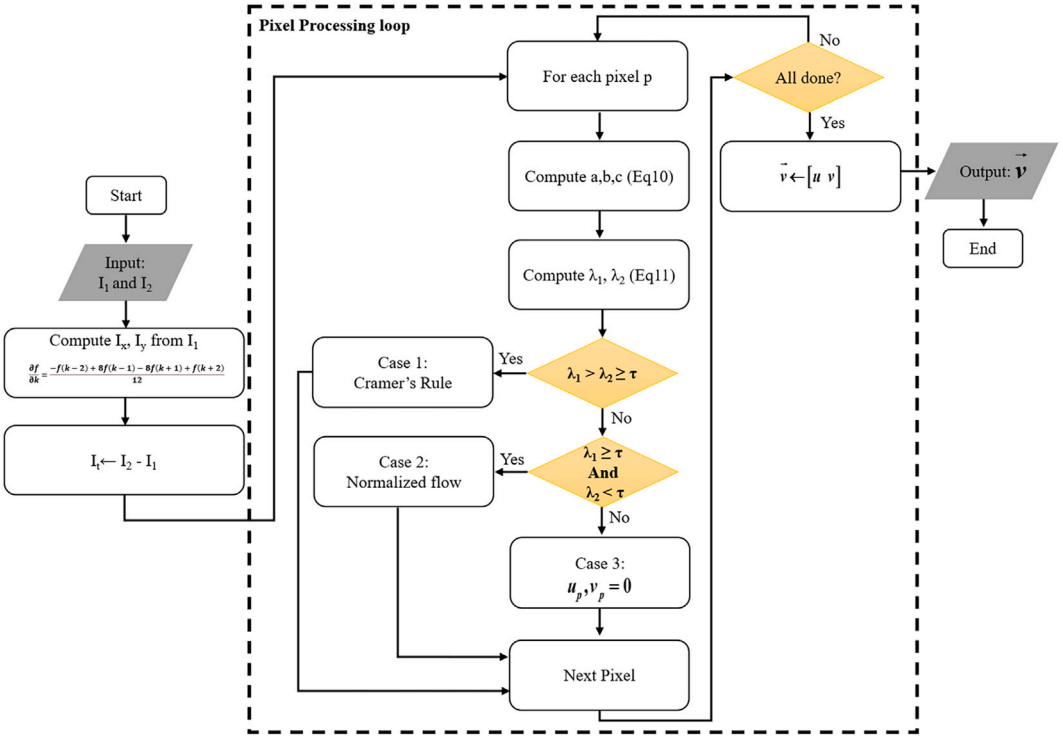


Fig. A.10. The flowchart of the LK optical flow method.

Algorithm 3: TV-L1 Optical Flow algorithm.

Input: Two consecutive frames I_1 and I_2
Output: Optical flow $\vec{v}$
Data: λ, θ, τ: algorithm parameters; N_{warps}, N_{iters}: iteration counts; Initial flow $\vec{v}^0$
Initialize: $P \leftarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$; Compute $V_2 \leftarrow \nabla I_2$
for $w \leftarrow 0$ **to** N_{warps} **do**
  Compute $I_2(p + \vec{v}^0 dt)$ and $V_2(p + \vec{v}^0 dt)$
  for $k \leftarrow 0$ **to** N_{iters} **do**
    $\vec{A} \leftarrow \vec{v} + \mathrm{TH}(\vec{v}, \vec{v}^0)$
    Compute $\nabla \cdot P$
    $\vec{v} \leftarrow \vec{A} + \theta \nabla \cdot P$
    [u v] $\leftarrow \vec{v}$
    Compute ∇u, ∇v
    $P_x \leftarrow \frac{P_x + \frac{\tau}{\theta}\nabla u}{1 + \frac{\tau}{\theta}|\nabla u|}$
    $P_y \leftarrow \frac{P_y + \frac{\tau}{\theta}\nabla v}{1 + \frac{\tau}{\theta}|\nabla v|}$
  end
end
return *Optical flow* $\vec{v}$

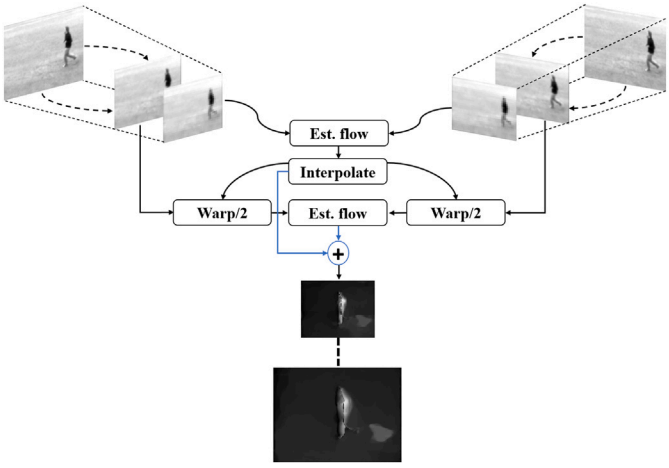


Fig. A.11. Block diagram of the TV-L1 optical flow method.

and λ is also an input parameter for the algorithm. The block diagram of the TV-L1 optical flow method is shown in Fig. A.11.

Gunnar Farnebäck method

In the Gunnar Farnebäck method, which is henceforth called Farnebäck for simplicity, optical flow is first estimated using two frames and then adjusted based on background motion. In this method, the brightness of each pixel is approximated by fitting a second-degree polynomial to its neighbourhood [43,293].

Specifically, the intensity at position x and time t is modelled as:

$$I_p(t) = \mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + \mathbf{c}, \tag{A.12}$$

where A is a symmetric matrix, b is a vector, and c is a scalar. A weighted least-squares method is used to obtain this polynomial approximation fit over the pixel neighbourhood. The local quadratic approximation in the first frame is denoted by:

$$I_{1p}(t) = \mathbf{x}^{\mathbf{T}}\mathbf{A}_1\mathbf{x} + \mathbf{b}_1^T\mathbf{x} + \mathbf{c}_1. \tag{A.13}$$

Then, after a time Δt, the following is obtained:

$$\left.\begin{array}{l} I_{2p}(t+\Delta t) = I_{1p}\left(t,\ x - \vec{v}\Delta t\right) \\ d \leftarrow x - \vec{v}\Delta t \end{array}\right\} I_{1p}(t,\ x-d) = (x-d)^{\mathbf{T}}\mathbf{A}_1(x-d)$$
$$+\mathbf{b}_1^T(x-d) + \mathbf{c}_1 = x^T A_1 x + \left(b_1 - 2A_1 d\right)^T x + d^T A_1 d - \mathbf{b}_1^T d + c_1. \tag{A.14}$$

Equating the quadratic model of the second frame with Eq. (A.14) yields:

$$\mathbf{A}_2 = A_1, \tag{A.15}$$

$$b_2 = b_1 - 2A_1 d, \tag{A.16}$$

$$\mathbf{c}_2 = d^T A_1 d - \mathbf{b}_1^T d + c_1. \tag{A.17}$$

To obtain d, if the corresponding matrix system is solvable, either Eqs. (A.16) or (A.17) can be used. For example, to solve Eq. (A.16), the following is considered:

$$A_1 d = -\frac{1}{2}(b_2 - b_1), \tag{A.18}$$

which involves fewer computations and is solvable if $\det(A_1) \neq 0$. Equivalently, it can be defined:

$$\Delta b = -\frac{1}{2}(b_2 - b_1), \tag{A.19}$$

so that

$$A_1 d = \Delta b, \tag{A.20}$$

where d is the optical flow vector. To reduce the impact of noise and improve accuracy, Eq. (A.20) is typically evaluated over a neighbourhood around each pixel, and the objective function is defined as:

$$\sum_{q \in N_p} w_q \left\| A_q d_q - \Delta b_q \right\|^2, \tag{A.21}$$

where w_q is the weight assigned to neighbour q. The d_p that minimizes Eq. (A.21) can be defined as:

$$d_p = \left(\sum w A^T A\right)^{-1} \sum w A^T \Delta b. \tag{A.22}$$

Since, in a real case, A_1 and A_2 are never exactly, an iterative scheme is used to improve accuracy and suppress possible errors when solving Eq. (A.22):

$$\begin{array}{l} d = 0 \\ for \quad i = 1 : N_{iter} \\ \quad \tilde{x} = x + d \\ \quad A = \frac{1}{2}(A_1(x) + A_2(\tilde{x})) \\ \quad \Delta b = -\frac{1}{2}(b_2(\tilde{x}) - b_1(x)) + Ad \\ end \end{array} \tag{A.23}$$

Incorporating simple motion models or assuming a two-dimensional motion pattern can further accelerate the calculation and help mitigate residual errors. Farnebäck's Optical Flow pseudocode is given in Algorithm 4. The flowchart of the Farnebäck method with complete details is given in Fig. A.12. The three nested iteration loops in this flowchart make it more complex than previous methods.

The shift in feature representation over the review period can also be drawn together in one place. Before 2018, histogram features such as HOF and HOOF were the usual choice and appear in 72 studies. After 2018, optical flow was fed directly into deep networks instead, in 20 studies, made up of LK in 6, Farnebäck in 6, unspecified methods in 4, TV-L1 in 3 and HS in 1. Fig. A.13 sets these two periods side by side.

Algorithm 4: Farnebäck Optical Flow algorithm.

Input: Two consecutive frames I_1 and I_2
Output: Optical flow $\vec{v}$
Data: Levels L; Polynomial window parameters PW_ρ; Number of iterations N_{iter}; Neighborhood $\mathcal{N}(p)$; Weight function W; Model $\in$ {pointwise, constant, affine}
$\vec{v}_0^{\,L} \leftarrow \mathbf{0}$
$\{I_1^l\}_{l=1}^{L} \leftarrow$ Build_Pyramid(I_1, L)
$\{I_2^l\}_{l=1}^{L} \leftarrow$ Build_Pyramid(I_2, L)
for $l \leftarrow L$ ***downto*** 1 **do**
  foreach *pixel* p *in* I_1^l **do**
    $(A_1, b_1, \epsilon_1) \leftarrow$ Polynomial_Expansion$(I_1^l(p), PW_\rho)$
    for $i \leftarrow 1$ **to** N_{iter} **do**
      $\tilde{I}_2^l \leftarrow I_2^l\big(p + \vec{v}_0^l(p)\,dt\big)$
      $(A_2, b_2, \epsilon_2) \leftarrow$ Polynomial_Expansion$\big(\tilde{I}_2^l(p), PW_\rho\big)$
      $A \leftarrow \frac{A_2 + A_1}{2}$
      $\Delta b \leftarrow -\frac{(b_2 - b_1)}{2} + A\,\vec{v}_0^l(p)$
      Solve $A\,\vec{v}(p) = \Delta b$ using W, Model, $\mathcal{N}(p)$
      $\vec{v}_0^l(p) \leftarrow \vec{v}(p)$
    end
  end
  if $l > 1$ **then**
    $\vec{v}_0^{l-1} \leftarrow$ Upsample$\big(\vec{v}_0^l,\ \text{size}(I^{l-1})\big)$
  end
end
return *Optical flow* $\vec{v}$

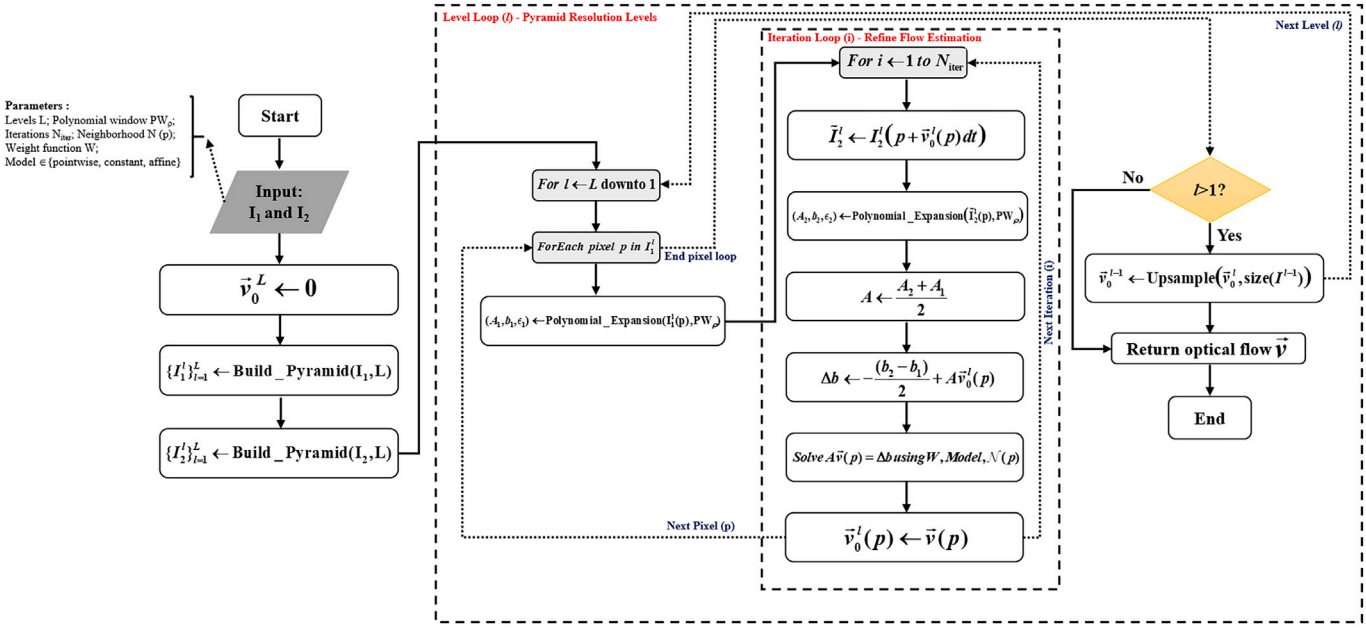


Fig. A.12. Flowchart of the Farnebäck optical flow method.

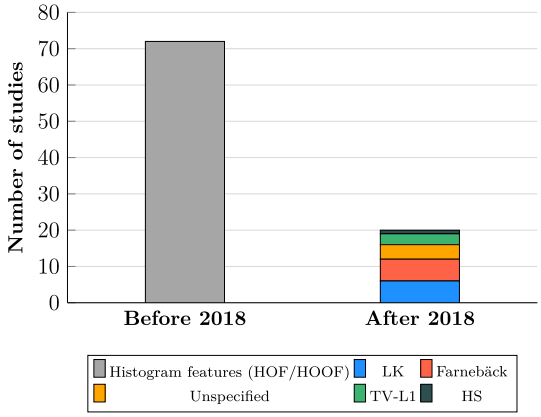


Fig. A.13. Shift in feature representation in optical flow-based HAR.

Data availability

No data was used for the research described in the article.

References

[1] S. Khan, M.A. Khan, M. Alhaisoni, U. Tariq, H.-S. Yong, A. Armghan, F. Alenezi, Human action recognition: a paradigm of best deep learning features selection and serial based extended fusion, Sensors 21 (23) (2021) 7941, https://doi.org/10.3390/s21237941

[2] L. Guo, Z. Lu, L. Yao, Human-machine interaction sensing technology based on hand gesture recognition: a review, IEEE Trans. Hum.-Mach. Syst. 51 (4) (2021) 300–309, https://doi.org/10.1109/THMS.2021.3086003

[3] L. Fiorini, F.G.C. Loizzo, A. Sorrentino, J. Kim, E. Rovini, A. Di Nuovo, F. Cavallo, Daily gesture recognition during human-robot interaction combining vision and wearable systems, IEEE Sens. J. 21 (20) (2021) 23568–23577, https://doi.org/10.1109/JSEN.2021.3108011

[4] M. Alsawadi, Advancing Action Recognition Through Artificial Intelligence: A Comprehensive Approach for Home Safety Monitoring Using Skeleton Data and Spatial Temporal Graph Convolutional Neural Networks (Ph.D. thesis), UCL (University College London), 2024, https://discovery.ucl.ac.uk/id/eprint/10185865.

[5] V. Hajihashemi, A.A. Gharahbagh11, B.J.M.R. Tavares, Automated human activity recognition by a modified hybrid recurrent neural network, in: 16th Doctoral Symposium in Informatics Engineering, 2021, pp. 129.

[6] N.G. Nia, E. Kaplanoglu, A. Nasab, H. Qin, Human activity recognition using machine learning algorithms based on IMU data, in: 2023 5th International Conference on Bio-Engineering for Smart Technologies (BioSMART), IEEE, 2023, pp. 1–8, https://doi.org/10.1109/BioSMART58455.2023.10162095

[7] S. Sakka, V. Liagkou, C. Stylios, Exploiting security issues in human activity recognition systems (harss), Information 14 (6) (2023) 315, https://doi.org/10.3390/info14060315

[8] A.T. Raheem, T.A. Al-Assadi, A review for video summarization based on trajectories features, 2024, https://doi.org/10.1063/5.0199628

[9] D.R. Beddiar, B. Nini, M. Sabokrou, A. Hadid, Vision-based human activity recognition: a survey, Multimed. Tools Appl. 79 (41) (2020) 30509–30555, https://doi.org/10.1007/s11042-020-09004-3

[10] A. Manaf, S. Singh, Computer vision-based survey on human activity recognition system, challenges and applications, in: 2021 3rd International Conference on Signal Processing and Communication (ICPSC), IEEE, 2021, pp. 110–114, https://doi.org/10.1109/ICSPC51351.2021.9451736

[11] S. Ramasamy Ramamurthy, N. Roy, Recent trends in machine learning for human activity recognition—a survey, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (4) (2018) e1254, https://doi.org/10.1002/widm.1254

[12] L. Alpoim, A.F. da Silva, C.P. Santos, Human activity recognition systems: state of art, in: 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG), IEEE, 2019, pp. 1–4, https://doi.org/10.1109/ENBENG.2019.8692468

[13] Y. Zheng, Z. Wang, K. Gu, L. Wu, Z. Li, Y. Xiang, Multi-scale motion-based relational reasoning for group activity recognition, Eng. Appl. Artif. Intell. 139 (2025) 109570, https://doi.org/10.1016/j.engappai.2024.109570

[14] A.A. Gharahbagh, V. Hajihashemi, M.C. Ferreira, J.J. Machado, J.M.R. Tavares, Abnormal action recognition in social media clips using deep learning to analyze behavioral change, in: World Conference on Information Systems and Technologies, Springer, 2024, pp. 359–370, https://doi.org/10.1007/978-3-031-60328-0_36

[15] I. Paliwal, S. Gochhait, Classification of machine learning and power requirements for HAI (human activity identification), in: 2023 International Conference on Inventive Computation Technologies (ICICT), IEEE, 2023, pp. 199–204, https://doi.org/10.1109/ICICT57646.2023.10134470

[16] A.R. Kumar, D. Sumathi, A review of multidata human activity recognition, in: 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), IEEE, 2023, pp. 745–751, https://doi.org/10.1109/ICSCDS56580.2023.10104780

[17] X. Qin, J. Wang, S. Ma, W. Lu, Y. Zhu, X. Xie, Y. Chen, Generalizable low-resource activity recognition with diverse and discriminative representation learning, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 1943–1953, https://doi.org/10.1145/3580305.3599360

[18] A.A. Gharahbagh, V. Hajihashemi, J.J. Machado, J.M.R. Tavares, Camera movement cancellation in video using phase congruency and an FFT-based technique, in: World Conference on Information Systems and Technologies, Springer, 2023, pp. 172–181, https://doi.org/10.1007/978-3-031-45651-0_18

[19] M. Pandey, R. Mishra, A. Khare, V. Rathore, Keyframe extraction for salient human activity recognition using a canny edge detector, SN Comput. Sci. 6 (4) (2025) 380, https://doi.org/10.1007/s42979-025-03911-8

[20] A.A. Gharahbagh, V. Hajihashemi, M.C. Ferreira, J.J. Machado, J.M.R. Tavares, Best frame selection to enhance training step efficiency in video-based human action recognition, Appl. Sci. 12 (4) (2022) 1830, https://doi.org/10.3390/app12041830

[21] P. Pareek, A. Thakkar, A survey on video-based human action recognition: recent updates, datasets, challenges, and applications, Artif. Intell. Rev. 54 (3) (2021) 2259–2322, https://doi.org/10.1007/s10462-020-09904-8

[22] V. Hajihashemi, E. Pakizeh, Human activity recognition in videos based on a two levels K-means and hierarchical codebooks, Int. J. Mechatron. Electr. Comput. Technol 6 (2016) 3152–3159.

[23] M. Zhai, X. Xiang, N. Lv, X. Kong, Optical flow and scene flow estimation: a survey, Pattern Recognit. 114 (2021) 107861, https://doi.org/10.1016/j.patcog.2021.107861

[24] J. Carreira, A. Zisserman, Quo vadis, action recognition? A new model and the kinetics dataset, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 4724–4733, https://doi.org/10.1109/CVPR.2017.502

[25] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, L. Van Gool, Temporal segment networks: towards good practices for deep action recognition, in: European Conference on Computer Vision, Springer, 2016, pp. 20–36, https://doi.org/10.1007/978-3-319-46484-8_2

[26] G. Bertasius, H. Wang, L. Torresani, Is space-time attention all you need for video understanding? arXiv:2102.05095, 2021, https://doi.org/10.48550/arXiv.2102.05095

[27] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, C. Schmid, ViViT: a video vision transformer, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 6816–6826, https://doi.org/10.1109/ICCV48922.2021.00676

[28] Z. Liu, J. Ning, Y. Cao, Y. Wei, Z. Zhang, S. Lin, H. Hu, Video swin transformer, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022, pp. 3192–3201, https://doi.org/10.1109/CVPR52688.2022.00320

[29] H. Fan, B. Xiong, K. Mangalam, Y. Li, Z. Yan, J. Malik, C. Feichtenhofer, Multiscale vision transformers, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 6804–6815, https://doi.org/10.1109/ICCV48922.2021.00675

[30] P. Gong, X. Luo, A survey of video action recognition based on deep learning, Knowl.-Based Syst. 320 (2025) 113594, https://doi.org/10.1016/j.knosys.2025.113594

[31] S. Heo, J. Moon, S.K. Jung, Action Recognition: A Comprehensive Survey of Tasks, Methods, and Challenges, ICT Express (2025, https://doi.org/10.1016/j.icte.2025.11.015

[32] J. Qin, S. Chen, Z. Ye, J. Liu, Z. Liu, Video swin-CLSTM transformer: enhancing human action recognition with optical flow and long-term dependencies, PLoS One 20 (7) (2025) e0327717, https://doi.org/10.1371/journal.pone.0327717

[33] G. Saleem, U.I. Bajwa, R.H. Raza, Toward human activity recognition: a survey, Neural Comput. Appl. 35 (5) (2023) 4145–4182, https://doi.org/10.1007/s00521-022-07937-4

[34] T.F.N. Bukht, H. Rahman, M. Shaheen, A. Algarni, N.A. Almujally, A. Jalal, A review of video-based human activity recognition: theory, methods and applications, Multimed. Tools Appl. 84 (17) (2025) 18499–18545, https://doi.org/10.1007/s11042-024-19711-w

[35] A.K. AlShami, R. Rabinowitz, K. Lam, Y. Shleibik, M. Mersha, T. Boult, J. Kalita, Smart-vision: survey of modern action recognition techniques in vision, Multimed. Tools Appl. 84 (27) (2025) 32705–32776, https://doi.org/10.1007/s11042-024-20484-5

[36] H. Yin, R.O. Sinnott, G.T. Jayaputera, A survey of video-based human action recognition in team sports, Artif. Intell. Rev. 57 (11) (2024), https://doi.org/10.1007/s10462-024-10934-9

[37] D. Giveki, S. Ahmadi, Motion-inclusive skip connection gated recurrent unit: an end-to-end deep neural network for action recognition, Mach. Learn. Appl. (2026) 100853, https://doi.org/10.1016/j.mlwa.2026.100853

[38] D. Giveki, Proposing a cross-correlational-gated recurrent unit neural network for engine block assembly action recognition, J. Intell. Manuf. 36 (8) (2025) 5629–5672, https://doi.org/10.1007/s10845-024-02518-9

[39] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of optical flow techniques, Int. J. Comput. Vis. 12 (1994) 43–77, https://doi.org/10.1007/BF01420984

[40] S.T.H. Shah, X. Xuezhi, Traditional and modern strategies for optical flow: an investigation, SN Appl. Sci. 3 (3) (2021) 289, https://doi.org/10.1007/s42452-021-04227-x

[41] B.K. Horn, B.G. Schunck, Determining optical flow, Artif. Intell. 17 (1–3) (1981) 185–203, https://doi.org/10.1016/0004-3702(81)90024-2

[42] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: IJCAI'81: 7th International Joint Conference on

Artificial Intelligence, 2, 1981, pp. 674–679, https://dl.acm.org/doi/10.5555/1623264.1623280.
[43] G. Farnebäck, Two-frame motion estimation based on polynomial expansion, in: Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13, Springer, 2003, pp. 363–370, https://doi.org/10.1007/3-540-45103-X_50
[44] C. Zach, T. Pock, H. Bischof, A duality based approach for realtime tv-l 1 optical flow, in: Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12–14, 2007. Proceedings 29, Springer, 2007, pp. 214–223, https://doi.org/10.1007/978-3-540-74936-3_22
[45] H.-H. Nagel, Displacement vectors derived from second-order intensity variations in image sequences, Comput. Vis. Graph. Image Process. 21 (1) (1983) 85–117, https://doi.org/10.1016/S0734-189X(83)80030-9
[46] C.L. Fennema, W.B. Thompson, Velocity determination in scenes containing several moving objects, Comput. graph. image process. 9 (4) (1979) 301–315, https://doi.org/10.1016/0146-664X(79)90097-2
[47] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, T. Brox, Flownet: learning optical flow with convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2758–2766, https://doi.org/10.1109/ICCV.2015.316
[48] C. Peng, H. Huang, A.-C. Tsoi, S.-L. Lo, Y. Liu, Z.-Y. Yang, Motion boundary emphasised optical flow method for human action recognition, IET Comput. Vis. 14 (6) (2020) 378–390, https://doi.org/10.1049/iet-cvi.2018.5556
[49] G. Kani, P. Geetha, Adaptive weighted flow net algorithm for human activity recognition using depth learned features, Comput. Syst. Sci. Eng. 46 (2) (2023), https://doi.org/10.32604/csse.2023.035969
[50] L. Sevilla-Lara, Y. Liao, F. Güney, V. Jampani, A. Geiger, M.J. Black, On the integration of optical flow and action recognition, in: German Conference on Pattern Recognition, Springer, 2018, pp. 281–297, https://doi.org/10.1007/978-3-030-12939-2_20
[51] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, FlowNet 2.0: evolution of optical flow estimation with deep networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 1647–1655, https://doi.org/10.1109/CVPR.2017.179
[52] A. Ranjan, M.J. Black, Optical flow estimation using a spatial pyramid network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4161–4170, https://doi.org/10.1109/CVPR.2017.291
[53] G. Cong, G. Domeniconi, J. Shapiro, C.-C. Yang, B. Chen, Video action recognition with an additional end-to-end trained temporal stream, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 51–60, https://doi.org/10.1109/WACV.2019.00013
[54] D. Sun, X. Yang, M.-Y. Liu, J. Kautz, Pwc-net: CNNS for optical flow using pyramid, warping, and cost volume, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8934–8943, https://doi.org/10.1109/CVPR.2018.00931
[55] S.J. Berlin, M. John, Spiking neural network based on joint entropy of optical flow features for human action recognition, Vis. Comput. 38 (1) (2022) 223–237, https://doi.org/10.1007/s00371-020-02012-2
[56] B. Rui, H. Wu, ADL: anomaly detection and localization in crowded scenes using hybrid methods, 2023, https://doi.org/10.20944/preprints202305.1617.v1
[57] Z. Teed, J. Deng, RAFT: recurrent all-pairs field transforms for optical flow, in: European Conference on Computer Vision, Springer, 2020, pp. 402–419, https://doi.org/10.1007/978-3-030-58536-5_24
[58] X. Wang, S. Zhang, Z. Qing, C. Gao, Y. Zhang, D. Zhao, N. Sang, Molo: motion-augmented long-short contrastive learning for few-shot action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 18011–18021, https://doi.org/10.1109/CVPR52729.2023.01727
[59] R. Wang, Z. Wang, P. Gao, M. Li, J. Jeong, Y. Xu, Y. Lee, C.M. Baum, L.T. Connor, C. Lu, Real-time human action recognition on embedded platforms, arXiv preprint arXiv:2409.05662, 2024, https://doi.org/10.48550/arXiv.2409.05662
[60] H. Xu, J. Zhang, J. Cai, H. Rezatofighi, D. Tao, GMFlow: learning optical flow via global matching, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022, pp. 8111–8120, https://doi.org/10.1109/CVPR52688.2022.00795
[61] Y. Wang, L. Lipson, J. Deng, Sea-raft: simple, efficient, accurate RAFT for optical flow, in: European Conference on Computer Vision, Springer, 2024, pp. 36–54, https://doi.org/10.1007/978-3-031-72667-5_3
[62] D. Sun, C. Herrmann, F. Reda, M. Rubinstein, D.J. Fleet, W.T. Freeman, Disentangling architecture and training for optical flow, in: European Conference on Computer Vision, Springer, 2022, pp. 165–182, https://doi.org/10.1007/978-3-031-20047-2_10
[63] J.-J.-J. Lien, T. Kanade, J.F. Cohn, C.-C. Li, Detection, tracking, and classification of action units in facial expression, Robot. Auton. Syst. 31 (3) (2000) 131–146, https://doi.org/10.1016/S0921-8890(99)00103-7
[64] M. Ahmad, S.-W. Lee, HMM-based human action recognition using multiview image sequences, in: 18th International Conference on Pattern Recognition (ICPR'06), 1, IEEE, 2006, pp. 263–266, https://doi.org/10.1109/ICPR.2006.630
[65] M. Ahmad, S.-W. Lee, Human action recognition using multi-view image sequences, in: 7th International Conference on Automatic Face and Gesture Recognition (FGR06), IEEE, 2006, pp. 523–528, https://doi.org/10.1109/FGR.2006.65
[66] Y. Wang, P. Sabzmeydani, G. Mori, Semi-latent dirichlet allocation: a hierarchical model for human action recognition, in: Workshop on Human Motion, Springer, 2007, pp. 240–254, https://doi.org/10.1007/978-3-540-75703-0_17
[67] N. Ikizler, R.G. Cinbis, P. Duygulu, Human action recognition with line and flow histograms, in: 2008 19th International Conference on Pattern Recognition, IEEE, 2008, pp. 1–4, https://doi.org/10.1109/ICPR.2008.4761434
[68] M. Mendoza, N. Pérez De La Blanca, Applying space state models in human action recognition: a comparative study, in: International Conference on Articulated Motion and Deformable Objects, Springer, 2008, pp. 53–62, https://doi.org/10.1007/978-3-540-70517-8_6
[69] S.A. Etemad, P. Payeur, A. Arya, Automatic temporal location and classification of human actions based on optical features, in: 2009 2nd International Congress on Image and Signal Processing, IEEE, 2009, pp. 1–5, https://doi.org/10.1109/CISP.2009.5304683
[70] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, G. Serra, Effective codebooks for human action categorization, in: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, IEEE, 2009, pp. 506–513, https://doi.org/10.1109/ICCVW.2009.5457658
[71] R. Chaudhry, A. Ravichandran, G. Hager, R. Vidal, Histograms of oriented optical flow and binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 1932–1939, https://doi.org/10.1109/CVPR.2009.5206821
[72] M.A. Mendoza, N.P. de la Blanca, M.J. Marin-Jimenez, PoHMM-based human action recognition, in: 2009 10th Workshop on Image Analysis for Multimedia Interactive Services, IEEE, 2009, pp. 85–88, https://doi.org/10.1109/WIAMIS.2009.5031438
[73] C.-C. Chen, J. Aggarwal, Recognizing human action from a far field of view, in: 2009 Workshop on Motion and Video Computing (WMVC), IEEE, 2009, pp. 1–7, https://doi.org/10.1109/WMVC.2009.5399231
[74] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, G. Serra, Recognizing human actions by fusing spatio-temporal appearance and motion descriptors, in: 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 3569–3572, https://doi.org/10.1109/ICIP.2009.5414332
[75] M. Lucena, N.P. de la Blanca, J.M. Fuertes, M.J. Marín-Jiménez, Human action recognition using optical flow accumulated local histograms, in: Pattern Recognition and Image Analysis: 4th Iberian Conference, IbPRIA 2009 Póvoa de Varzim, Portugal, June 10–12, 2009 Proceedings 4, Springer, 2009, pp. 32–39, https://doi.org/10.1007/978-3-642-02172-5_6
[76] N. Sawant, K. Biswas, Human action recognition based on spatio-temporal features, in: Pattern Recognition and Machine Intelligence: Third International Conference, PReMI 2009 New Delhi, India, December 16–20, 2009 Proceedings 3, Springer, 2009, pp. 357–362, https://doi.org/10.1007/978-3-642-11164-8_58
[77] A. Ramadass, M. Suk, B. Prabhakaran, Feature extraction method for video based human action recognitions: extended optical flow algorithm, in: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2010, pp. 1106–1109, https://doi.org/10.1109/ICASSP.2010.5495352
[78] I. Parvin, M. Ahmad, Global and local motion descriptions for human action recognition, in: International Conference on Electrical & Computer Engineering (ICECE 2010), IEEE, 2010, pp. 678–681, https://doi.org/10.1109/ICELCE.2010.5700784
[79] M. Ahmad, S.-W. Lee, Human action recognition using shape and CLG-motion flow from multi-view image sequences, Pattern Recognit. 41 (7) (2008) 2237–2252, https://doi.org/10.1016/j.patcog.2007.12.008
[80] P. Guo, Z. Miao, H.-D. Cheng, Masks based human action detection in crowded videos, in: 2010 IEEE International Conference on Image Processing, IEEE, 2010, pp. 693–696, https://doi.org/10.1109/ICIP.2010.5651209
[81] S. Liu, J. Liu, T. Zhang, H. Lu, Human action recognition in videos using hybrid motion features, in: Advances in Multimedia Modeling: 16th International Multimedia Modeling Conference, MMM 2010, Chongqing, China, January 6–8, 2010. Proceedings 16, Springer, 2010, pp. 411–421, https://doi.org/10.1007/978-3-642-11301-7_42
[82] R. Cilla, M.A. Patricio, A. Berlanga, J.M. Molina, Fusion of single view soft k-NN classifiers for multicamera human action recognition, in: Hybrid Artificial Intelligence Systems: 5th International Conference, HAIS 2010, San Sebastián, Spain, June 23–25, 2010. Proceedings, Part II 5, Springer, 2010, pp. 436–443, https://doi.org/10.1007/978-3-642-13803-4_54
[83] R.V. Babu, S. Suresh, Fully complex-valued ELM classifiers for human action recognition, in: The 2011 International Joint Conference on Neural Networks, IEEE, 2011, pp. 2803–2808, https://doi.org/10.1109/IJCNN.2011.6033588
[84] M.A. Naiel, M.M. Bdelwahab, M. El-Saban, W. Mikhael, Highly efficient human action recognition using compact 2DPCA-based descriptors in the spatial and transform domains, in: 2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS), IEEE, 2011, pp. 1–4, https://doi.org/10.1109/MWSCAS.2011.6026502
[85] K. Lertniphonphan, S. Aramvith, T.H. Chalidabhongse, Human action recognition using direction histograms of optical flow, in: 2011 11th International Symposium on Communications & Information Technologies (ISCIT), IEEE, 2011, pp. 574–579, https://doi.org/10.1109/ISCIT.2011.6089701
[86] W. Yang, Y. Wang, G. Mori, Learning transferable distance functions for human action recognition, in: Machine Learning for Vision-Based Motion Analysis: Theory and Techniques, Springer, 2011, pp. 349–370, https://doi.org/10.1007/978-0-85729-057-1_13
[87] Efros, Berg, Mori, Malik, Recognizing action at a distance, in: Proceedings Ninth IEEE International Conference on Computer Vision, IEEE, 2003, pp. 726–733, https://doi.org/10.1109/ICCV.2003.1238420
[88] M. Lucena, N. Pérez de la Blanca, J.M. Fuertes, Human action recognition based on aggregated local motion estimates, Mach. Vis. Appl. 23 (2012) 135–150, https://doi.org/10.1007/s00138-010-0305-9

[89] R.V. Babu, S. Suresh, R. Savitha, Human action recognition using a fast learning fully complex-valued classifier, Neurocomputing 89 (2012) 202–212, https://doi.org/10.1016/j.neucom.2012.03.003

[90] T. da Rocha, F. de Barros Vidal, A.R.S. Romariz, A proposal for human action classification based on motion analysis and artificial neural networks, in: The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE, 2012, pp. 1–6, https://doi.org/10.1109/IJCNN.2012.6252612

[91] V. Vo, N. Ly, An effective approach for human actions recognition based on optical flow and edge features, in: 2012 International Conference on Control, Automation and Information Sciences (ICCAIS), IEEE, 2012, pp. 24–29, https://doi.org/10.1109/ICCAIS.2012.6466597

[92] X. Zhang, Z. Miao, L. Wan, Human action categories using motion descriptors, in: 2012 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 1381–1384, https://doi.org/10.1109/ICIP.2012.6467126

[93] T. Hofmann, Probabilistic latent semantic indexing, 1999, https://doi.org/10.1145/312624.312649

[94] S. Stasinopoulos, P. Maragos, Human action recognition using histographic methods and hidden Markov models for visual martial arts applications, in: 2012 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 745–748, https://doi.org/10.1109/ICIP.2012.6466967

[95] K. Subramanian, S. Suresh, Human action recognition using meta-cognitive neuro-fuzzy inference system, Int. J. Neural Syst. 22 (6) (2012) 1250028, https://doi.org/10.1142/S0129065712500281

[96] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2247–2253, https://doi.org/10.1109/TPAMI.2007.70711

[97] K. Lertniphonphan, S. Aramvith, T.H. Chalidabhongse, The region-based distance of oriented gradient and motion direction for human action classification, in: 2012 International Symposium on Communications and Information Technologies (ISCIT), IEEE, 2012, pp. 615–619, https://doi.org/10.1109/ISCIT.2012.6380974

[98] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, G. Serra, Effective codebooks for human action representation and classification in unconstrained videos, IEEE Trans. Multimed. 14 (4) (2012) 1234–1245, https://doi.org/10.1109/TMM.2012.2191268

[99] H. Tabia, M. Gouiffes, L. Lacassagne, Motion histogram quantification for human action recognition, in: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE, 2012, pp. 2404–2407.

[100] Y. Yang, I. Saleemi, M. Shah, Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2012) 1635–1648, https://doi.org/10.1109/TPAMI.2012.253

[101] O. Murthy, R. Goecke, Ordered trajectories for large scale human action recognition, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 412–419, https://doi.org/10.1109/ICCVW.2013.61

[102] B.M. Nair, V.K. Asari, Regression based learning of human actions from video using Hof-LBP flow patterns, in: 2013 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 2013, pp. 4342–4347, https://doi.org/10.1109/SMC.2013.741

[103] D.F. Specht, et al., A general regression neural network, IEEE Trans. Neural Netw. 2 (6) (1991) 568–576, https://doi.org/10.1109/72.97934

[104] Q.D. Tran, N.Q. Ly, Sparse spatio-temporal representation of joint shape-motion cues for human action recognition in depth sequences, in: The 2013 RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future (RIVF), IEEE, 2013, pp. 253–258, https://doi.org/10.1109/RIVF.2013.6719903

[105] B.Z. Yao, B.X. Nie, Z. Liu, S.-C. Zhu, Animated pose templates for modeling and detecting human actions, IEEE Trans. Pattern Anal. Mach. Intell. 36 (3) (2013) 436–452, https://doi.org/10.1109/TPAMI.2013.144

[106] L. Liu, L. Shao, P. Rockett, Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition, Pattern Recognit. 46 (7) (2013) 1810–1818.

[107] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vis. 42 (2001) 145–175, https://doi.org/10.1023/A:1011139631724

[108] R.V. Babu, R. Savitha, S. Suresh, B. Agarwal, Subject independent human action recognition using spatio-depth information and meta-cognitive RBF network, Eng. Appl. Artif. Intell. 26 (9) (2013) 2010–2021, https://doi.org/10.1016/j.engappai.2013.07.008

[109] M. Mackay, R.G. Fenton, B. Benhabib, Pipeline-architecture based real-time active-vision for human-action recognition, J. Intell. Robot. Syst. 72 (2013) 385–407, https://doi.org/10.1007/s10846-012-9810-6

[110] K. Subramanian, V.B. Radhakrishnan, S. Sundaram, An optical flow feature and McFIS based approach for 3-dimensional human action recognition, in: 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), IEEE, 2014, pp. 1–6, https://doi.org/10.1109/ISSNIP.2014.6827689

[111] M.A. Abdelwahab, M.M. Abdelwahab, Human action recognition based on two-view optical flow in the transformed domain, in: 2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS), IEEE, 2014, pp. 805–808, https://doi.org/10.1109/MWSCAS.2014.6908537

[112] D. Gerónimo, H. Kjellström, Unsupervised surveillance video retrieval based on human action and appearance, in: 2014 22nd International Conference on Pattern Recognition, IEEE, 2014, pp. 4630–4635, https://doi.org/10.1109/ICPR.2014.792

[113] X. Zhen, L. Shao, X. Li, Action recognition by spatio-temporal oriented energies, Inf. Sci. 281 (2014) 295–309, https://doi.org/10.1016/j.ins.2014.05.021

[114] J. Song, F. Hu, Human action recognition with block-based model and flow histograms, in: Proceedings of the 3rd International Conference on Multimedia Technology (ICMT 2013), Springer, 2014, pp. 407–418, https://doi.org/10.1007/978-3-642-41407-7_40

[115] R. Benmokhtar, Robust human action recognition scheme based on high-level feature fusion, Multimed. Tools Appl. 69 (2014) 253–275, https://doi.org/10.1007/s11042-012-1022-3

[116] D.-J. Duh, C.-C. Kan, S.-Y. Chen, C.-M. Lu, Human action recognition in video under clutter and moving background, in: Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2014 International Workshops: DANTH, BDM, MobiSocial, BigEC, CloudSD, MSMV-MBI, SDA, DMDA-Health, ALSIP, SocNet, DMBIH, BigPMA, Tainan, Taiwan, May 13–16, 2014. Revised Selected Papers 18, Springer, 2014, pp. 722–734, https://doi.org/10.1007/978-3-319-13186-3_64

[117] A. Ladjailia, I. Bouchrika, H.F. Merouani, N. Harrati, Automated detection of similar human actions using motion descriptors, in: 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), IEEE, 2015, pp. 398–403, https://doi.org/10.1109/STA.2015.7505099

[118] I. Gomez-Conde, D.N. Olivieri, A KPCA spatio-temporal differential geometric trajectory cloud classifier for recognizing human actions in a CBVR system, Expert Syst. Appl. 42 (13) (2015) 5472–5490, https://doi.org/10.1016/j.eswa.2015.03.010

[119] S. Mukherjee, Human action recognition using dominant pose duplet, in: Computer Vision Systems: 10th International Conference, ICVS 2015, Copenhagen, Denmark, July 6–9, 2015, Proceedings 10, Springer, 2015, pp. 488–497, https://doi.org/10.1007/978-3-319-20904-3_44

[120] B. Jagadeesh, C.M. Patil, Video based action detection and recognition human using optical flow and SVM classifier, in: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, 2016, pp. 1761–1765, https://doi.org/10.1109/RTEICT.2016.7808136

[121] A.B. Mahjoub, M. Atri, Human action recognition using RGB data, in: 2016 11th International Design & Test Symposium (IDT), IEEE, 2016, pp. 83–87, https://doi.org/10.1109/IDT.2016.7843019

[122] H.-B. Zhang, Y. Ma, F. Guo, M. Zhang, Y. Lin, Motion difference histogram for recognizing human action in video, in: 2016 International Conference on Applied System Innovation (ICASI), IEEE, 2016, pp. 1–4, https://doi.org/10.1109/ICASI.2016.7539764

[123] H. Wang, C. Schmid, Action recognition with improved trajectories, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3551–3558, https://doi.org/10.1109/ICCV.2013.441

[124] S. Megrhi, M. Jmal, W. Souidene, A. Beghdadi, Spatio-temporal action localization and detection for human action recognition in big dataset, J. Vis. Commun. Image Represent. 41 (2016) 375–390, https://doi.org/10.1016/j.jvcir.2016.10.016

[125] X. Ji, Z. Ju, C. Wang, C. Wang, Multi-view transition HMMs based view-invariant human action recognition method, Multimed. Tools Appl. 75 (2016) 11847–11864, https://doi.org/10.1007/s11042-015-2661-y

[126] J. Talukdar, B. Mehta, Human action recognition system using good features and multilayer perceptron network, in: 2017 International Conference on Communication and Signal Processing (ICCSP), IEEE, 2017, pp. 0317–0323, https://doi.org/10.1109/ICCSP.2017.8286369

[127] J. Shi, et al., Good features to track, in: 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 1994, pp. 593–600, https://doi.org/10.1109/CVPR.1994.323794

[128] G.T. Papadopoulos, P. Daras, Global flow and temporal-shape descriptors for human action recognition from 3D reconstruction data, in: Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15–20, 2017, Proceedings 13, Springer, 2017, pp. 47–62, https://doi.org/10.1007/978-3-319-62416-7_4

[129] Y. Qin, L. Mo, B. Xie, Feature fusion for human action recognition based on classical descriptors and 3D convolutional networks, in: 2017 Eleventh International Conference on Sensing Technology (ICST), IEEE, 2017, pp. 1–5, https://doi.org/10.1109/ICSensT.2017.8304460

[130] S. Sen, M. Dhar, S. Banerjee, Implementation of human action recognition using image parsing techniques, in: 2018 Emerging Trends in Electronic Devices and Computational Techniques (EDCT), IEEE, 2018, pp. 1–6, https://doi.org/10.1109/EDCT.2018.8405091

[131] U. Srinivasu, S.P. Sahoo, S. Ari, Semi-supervised learning in random forest classifier for human action recognition, in: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, 2018, pp. 237–241, https://doi.org/10.1109/ICOEI.2018.8553899

[132] D.K. Singh, Human action recognition in video, in: Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2, Springer, 2019, pp. 54–66, https://doi.org/10.1007/978-981-13-3140-4_6

[133] H. Abrishami Moghaddam, A. Zare, Spatiotemporal wavelet correlogram for human action recognition, Int. J. Multimed. Inf. Retr. 8 (2019) 167–180, https://doi.org/10.1007/s13735-018-00167-2

[134] D.K. Vishwakarma, A two-fold transformation model for human action recognition using decisive pose, Cogn. Syst. Res. 61 (2020) 1–13, https://doi.org/10.1016/j.cogsys.2019.12.004

[135] M. Ghalan, R.K. Aggarwal, Multifold classification for human action recognition, in: 2021 IEEE Bombay Section Signature Conference (IBSSC), IEEE, 2021, pp. 1–6, https://doi.org/10.1109/IBSSC53889.2021.9673392

[136] Y. Chen, L. Wang, G. Wang, S. Yang, Y. Wang, M. Xiang, X. Zhang, H. Chen, D. Hu, H. Cheng, Spatio-temporal features for fast early warning of unplanned self-extubation in ICU, Eng. Appl. Artif. Intell. 127 (2024) 107294, https://doi.org/10.1016/j.engappai.2023.107294

[137] V.A. Chenarlogh, F. Razzazi, N. Mohammadyahya, A multi-view human action recognition system in limited data case using multi-stream CNN, in: 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), IEEE, 2019, pp. 1–11, https://doi.org/10.1109/ICSPIS48872.2019.9066079

[138] N.M. Zamri, G.F. Ling, P.Y. Han, O.S. Yin, Vision-based human action recognition on pre-trained AlexNet, in: 2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), IEEE, 2019, pp. 1–5, https://doi.org/10.1109/ICCSCE47578.2019.9068586

[139] H. Tacon, A.S. Brito, H.L. Chaves, M.B. Vieira, S.M. Villela, H. de Almeida Maia, D.T. Concha, H. Pedrini, Human action recognition using convolutional neural networks with symmetric time extension of visual rhythms, in: International Conference on Computational Science and Its Applications, Springer, 2019, pp. 351–366, https://doi.org/10.1007/978-3-030-24289-3_26

[140] Q. Xiong, J. Zhang, P. Wang, D. Liu, R.X. Gao, Transferable two-stream convolutional neural network for human action recognition, J. Manuf. Syst. 56 (2020) 605–614, https://doi.org/10.1016/j.jmsy.2020.04.007

[141] Y. Peng, H. Lee, T. Shu, H. Lu, Exploring biological motion perception in two-stream convolutional neural networks, Vis. Res. 178 (2021) 28–40, https://doi.org/10.1016/j.visres.2020.09.005

[142] S.C.B. Dash, S.R. Mishra, K. Srujan Raju, L. Narasimha Prasad, Human action recognition using a hybrid deep learning heuristic, Soft Comput. 25 (20) (2021) 13079–13092, https://doi.org/10.1007/s00500-021-06149-7

[143] S.K. Park, J.H. Chung, T.K. Kang, M.T. Lim, Binary dense SIFT flow based two stream CNN for human action recognition, Multimed. Tools Appl. 80 (28) (2021) 35697–35720, https://doi.org/10.1007/s11042-021-10795-2

[144] H. Zan, G. Zhao, Human action recognition research based on fusion TS-CNN and LSTM networks, Arab. J. Sci. Eng. 48 (2) (2023) 2331–2345, https://doi.org/10.1007/s13369-022-07236-z

[145] F. Khezerlou, A. Baradarani, M.A. Balafar, R.G. Maev, Multi-stream CNNs with orientation-magnitude response maps and weighted inception module for human action recognition, in: 2023 3rd International Conference on Artificial Intelligence and Signal Processing (AISP), IEEE, 2023, pp. 1–5, https://doi.org/10.1109/AISP57993.2023.10134817

[146] V. Vinothina, A. George, et al., Recognizing abnormal behavior in heterogeneous crowd using two stream CNN, in: 2024 Asia Pacific Conference on Innovation in Technology (APCIT), IEEE, 2024, pp. 1–5, https://doi.org/10.1109/APCIT62007.2024.10673589

[147] M.J. Marín-Jiménez, N.P. de la Blanca, M.A. Mendoza, RBM-based silhouette encoding for human action modelling, in: 2010 20th International Conference on Pattern Recognition, IEEE, 2010, pp. 979–982, https://doi.org/10.1109/ICPR.2010.245

[148] M.J. Marín-Jiménez, N. Pérez de la Blanca, M. Mendoza, Learning features for human action recognition using multilayer architectures, in: Iberian Conference on Pattern Recognition and Image Analysis, Springer, 2011, pp. 338–346, https://doi.org/10.1007/978-3-642-21257-4_42

[149] D.N. Olivieri, I.G. Conde, X.A.V. Sobrino, Eigenspace-based fall detection and activity recognition from motion templates and machine learning, Expert Syst. Appl. 39 (5) (2012) 5935–5945, https://doi.org/10.1016/j.eswa.2011.11.109

[150] J. Yu, M. Jeon, W. Pedrycz, Weighted feature trajectories and concatenated bag-of-features for action recognition, Neurocomputing 131 (2014) 200–207, https://doi.org/10.1016/j.neucom.2013.10.024

[151] A. Iosifidis, A. Tefas, I. Pitas, Multi-view regularized extreme learning machine for human action recognition, in: Artificial Intelligence: Methods and Applications: 8th Hellenic Conference on AI, SETN 2014, Ioannina, Greece, May 15–17, 2014. Proceedings 8, Springer, 2014, pp. 84–94, https://doi.org/10.1007/978-3-319-07064-3_7

[152] J.-J. Seo, J. Son, H.-I. Kim, W. De Neve, M.R. Yong, Efficient and effective human action recognition in video through motion boundary description with a compact set of trajectories, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1, IEEE, 2015, pp. 1–6, https://doi.org/10.1109/FG.2015.7163123

[153] H. Wang, Y. Yi, Tracking salient keypoints for human action recognition, in: 2015 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 2015, pp. 3048–3053, https://doi.org/10.1109/SMC.2015.530

[154] A. Eweiwi, M.S. Cheema, C. Bauckhage, Action recognition in still images by learning spatial interest regions from videos, Pattern Recognit. Lett. 51 (2015) 8–15, https://doi.org/10.1016/j.patrec.2014.07.017

[155] O.R. Murthy, R. Goecke, Ordered trajectories for human action recognition with large number of classes, Image Vis. Comput. 42 (2015) 22–34, https://doi.org/10.1016/j.imavis.2015.06.009

[156] S.-L. Lo, A.-C. Tsoi, Motion boundary trajectory for human action recognition, in: Computer Vision-ACCV 2014 Workshops: Singapore, Singapore, November 1–2, 2014, Revised Selected Papers, Part I 12, Springer, 2015, pp. 85–98, https://doi.org/10.1007/978-3-319-16628-5_7

[157] M. Sekma, M. Mejdoub, C. Ben Amar, Bag of graphs with geometric relationships among trajectories for better human action recognition, in: Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7–11, 2015, Proceedings, Part I 18, Springer, 2015, pp. 85–96, https://doi.org/10.1007/978-3-319-23231-7_8

[158] Y. Cheng, Y. Yi, Human action recognition in videos based on dense trajectory selection, in: 2015 8th International Congress on Image and Signal Processing (CISP), IEEE, 2015, pp. 30–34, https://doi.org/10.1109/CISP.2015.7407845

[159] L. Yao, Y. Liu, S. Huang, Spatio-temporal information for human action recognition, EURASIP J. Image Video Process. 2016 (2016) 1–9, https://doi.org/10.1186/s13640-016-0145-2

[160] P.C. Ribeiro, R. Audigier, Q.C. Pham, RIMOC, a feature to discriminate unstructured motions: application to violence detection for video-surveillance, Comput. Vis. Image Underst. 144 (2016) 121–143, https://doi.org/10.1016/j.cviu.2015.11.001

[161] H. Chen, J. Chen, R. Hu, Structural imosift for human action recognition, Wuhan Univ. J. Nat. Sci. 21 (3) (2016) 262–266, https://doi.org/10.1007/s11859-016-1169-2

[162] Z. Xu, R. Hu, J. Chen, C. Chen, H. Chen, H. Li, Q. Sun, Action recognition by saliency-based dense sampling, Neurocomputing 236 (2017) 82–92, https://doi.org/10.1016/j.neucom.2016.09.106

[163] W. Sultani, M. Shah, Automatic action annotation in weakly labeled videos, Comput. Vis. Image Underst. 161 (2017) 77–86, https://doi.org/10.1016/j.cviu.2017.05.005

[164] J.-J. Seo, H.-I. Kim, W. De Neve, Y.M. Ro, Effective and efficient human action recognition using dynamic frame skipping and trajectory rejection, Image Vis. Comput. 58 (2017) 76–85, https://doi.org/10.1016/j.imavis.2016.06.002

[165] Z. Gao, S. Li, G. Zhang, Y. Zhu, C. Wang, H. Zhang, Evaluation of regularized multi-task leaning algorithms for single/multi-view human action recognition, Multimed. Tools Appl. 76 (2017) 20125–20148, https://doi.org/10.1007/s11042-017-4384-8

[166] Y. Yi, H. Wang, B. Zhang, Learning correlations for human action recognition in videos, Multimed. Tools Appl. 76 (2017) 18891–18913, https://doi.org/10.1007/s11042-017-4416-4

[167] V. de Oliveira Silva, F. de Barros Vidal, A.R.S. Romariz, Human action recognition based on a two-stream convolutional network classifier, in: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2017, pp. 774–778, https://doi.org/10.1109/ICMLA.2017.00-64

[168] M. Fan, Q. Han, X. Zhang, Y. Liu, H. Chen, Y. Hu, Human action recognition based on dense sampling of motion boundary and histogram of motion gradient, in: 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), IEEE, 2018, pp. 1033–1038, https://doi.org/10.1109/DDCLS.2018.8515970

[169] Y. Yi, H. Wang, Motion keypoint trajectory and covariance descriptor for human action recognition, Vis. Comput. 34 (2018) 391–403, https://doi.org/10.1007/s00371-016-1345-6

[170] Y. Yi, P. Hu, X. Deng, Human action recognition with salient trajectories and multiple kernel learning, Multimed. Tools Appl. 77 (2018) 17709–17730, https://doi.org/10.1007/s11042-017-5209-5

[171] Z. Gao, T.-T. Han, H. Zhang, Y.-B. Xue, G.-P. Xu, MMA: a multi-view and multi-modality benchmark dataset for human action recognition, Multimed. Tools Appl. 77 (22) (2018) 29383–29404, https://doi.org/10.1007/s11042-018-5833-8

[172] J. Fernández-Ramírez, A. Álvarez-Meza, A. Orozco-Gutiérrez, Video-based human action recognition using kernel relevance analysis, in: Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13, Springer, 2018, pp. 116–125, https://doi.org/10.1007/978-3-030-03801-4_11

[173] J.L. Fernández-Ramírez, A.M. Álvarez-Meza, .A. Orozco-Gutiérrez, J.D. Echeverry-Correa, Infinite Gaussian fisher vector to support video-based human action recognition, in: Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7–9, 2019, Proceedings, Part II 14, Springer, 2019, pp. 38–49, https://doi.org/10.1007/978-3-030-33723-0_4

[174] P. Habashi, B. Boufama, I.S. Ahmad, Human action recognition using stereo trajectories, in: Mediterranean Conference on Pattern Recognition and Artificial Intelligence, Springer, 2019, pp. 94–105, https://doi.org/10.1007/978-3-030-37548-5_8

[175] M. Haddad, V.K. Ghassab, F. Najar, N. Bouguila, Instance-based learning for human action recognition, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2020, pp. 147–153, https://doi.org/10.1109/SMC42975.2020.9283020

[176] S.S. Mohith, S. Vijay, V. Sanjana, N. Krupa, Trajectory based human action recognition using centre symmetric local binary pattern descriptors, in: 2020 IEEE 17th India Council International Conference (INDICON), IEEE, 2020, pp. 1–6, https://doi.org/10.1109/INDICON49873.2020.9342248

[177] Y. Yi, A. Li, X. Zhou, Human action recognition based on action relevance weighted encoding, Signal Process. Image Commun. 80 (2020) 115640, https://doi.org/10.1016/j.image.2019.115640

[178] J. Imran, B. Raman, Evaluating fusion of RGB-D and inertial sensors for multimodal human action recognition, J. Ambient Intell. Humaniz. Comput. 11 (1) (2020) 189–208, https://doi.org/10.1007/s12652-019-01239-9

[179] Y. Huang, S.-H. Lai, S.-H. Tai, Human action recognition based on temporal pose CNN and multi-dimensional fusion, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, https://doi.org/10.1007/978-3-030-11012-3_33

[180] H. Jiang, Y. Li, S. Song, J. Liu, Rethinking fusion baselines for multi-modal human action recognition, in: Pacific Rim Conference on Multimedia, Springer, 2018, pp. 178–187, https://doi.org/10.1007/978-3-030-00764-5_17

[181] S. Zhu, Z. Fang, Y. Wang, J. Yu, J. Du, Multimodal activity recognition with local block CNN and attention-based spatial weighted CNN, J. Vis. Commun. Image Represent. 60 (2019) 38–43, https://doi.org/10.1016/j.jvcir.2018.12.026

[182] K. Yamada, S. Ito, N. Kaneko, K. Sumi, Human action recognition via body part region segmented dense trajectories, in: Computer Vision–ACCV 2018 Workshops: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers 14, Springer, 2019, pp. 64–72, https://doi.org/10.1007/978-3-030-21074-8_6

[183] S.A. Khowaja, S.-L. Lee, Semantic image networks for human action recognition, Int. J. Comput. Vis. 128 (2) (2020) 393–419, https://doi.org/10.1007/s11263-019-01248-3

[184] J. Wang, L. Xia, W. Ma, Human action recognition based on motion feature and manifold learning, IEEE Access 9 (2021) 89287–89299, https://doi.org/10.1109/ACCESS.2021.3088155

[185] H.-H. Phan, T.T. Nguyen, N.H. Phuc, N.H. Nhan, C.T. Tran, B.N. Vi, et al., Key frame and skeleton extraction for deep learning-based human action recognition, in: 2021 RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, 2021, pp. 1–6, https://doi.org/10.1109/RIVF51545.2021.9642132

[186] Y.-T. Lee, T. Pengying, S.Y. Yayilgan, O. Elezaj, Data-driven machine learning approach for human action recognition using skeleton and optical flow, in: Intelligent Technologies and Applications: Third International Conference, INTAP 2020, Grimstad, Norway, September 28–30, 2020, Revised Selected Papers 3, Springer, 2021, pp. 163–175, https://doi.org/10.1007/978-3-030-71711-7_14

[187] R. Huang, C. Chen, R. Cheng, Y. Zhang, J. Zhu, Human action recognition based on three-stream network with frame sequence features, in: 2022 7th International Conference on Image, Vision and Computing (ICIVC), IEEE, 2022, pp. 37–44, https://doi.org/10.1109/ICIVC55077.2022.9887162

[188] Y. He, Z. Xu, L. Ma, H. Li, Micro-expression spotting based on optical flow features, Pattern Recognit. Lett. 163 (2022) 57–64, https://doi.org/10.1016/j.patrec.2022.09.009

[189] M.R. Samsudin, S.A. Abu-Bakar, M.M. Mokji, An improved open-view human action recognition with unsupervised domain adaptation, Multimed. Tools Appl. 81 (20) (2022) 28479–28507, https://doi.org/10.1007/s11042-022-12822-2

[190] V. Nair, P. Hemeren, A. Vignolo, N. Noceti, E. Nicora, A. Sciutti, F. Rea, E. Billing, M. Bhatt, F. Odone, et al., Kinematic primitives in action similarity judgments: a human-centered computational model, IEEE Trans. Cogn. Dev. Syst. 15 (4) (2023) 1981–1992, https://doi.org/10.1109/TCDS.2023.3240302

[191] Q. Cai, J. Yan, H. Li, Y. Deng, Human action recognition method based on video-level features and attention mechanism, in: Proceedings of 2020 Chinese Intelligent Systems Conference: Volume I, Springer, 2021, pp. 225–233, https://doi.org/10.1007/978-981-15-8450-3_24

[192] A. Burgos-Madrigal, L. Altamirano-Robles, Video and image complexity in human action recognition, in: Progress in Artificial Intelligence and Pattern Recognition: 7th International Workshop on Artificial Intelligence and Pattern Recognition, IWAIPR 2021, Havana, Cuba, October 5–7, 2021, Proceedings 7, Springer, 2021, pp. 349–359, https://doi.org/10.1007/978-3-030-89691-1_34

[193] M. Kohli, A.K. Kar, V.G. Prakash, A. Prathosh, Deep learning-based human action recognition framework to assess children on the risk of autism or developmental delays, in: International Conference on Neural Information Processing, Springer, 2022, pp. 459–470, https://doi.org/10.1007/978-981-99-1648-1_38

[194] T. Feng, A new double-flow network via decision tree for human action recognition, in: 2023 5th International Conference on Decision Science & Management (ICDSM), IEEE, 2023, pp. 29–33, https://doi.org/10.1109/ICDSM59373.2023.00017

[195] S.K. Yadav, A. Luthra, E. Pahwa, K. Tiwari, H. Rathore, H.M. Pandey, P. Corcoran, DroneAttention: sparse weighted temporal attention for drone-camera based activity recognition, Neural Netw. 159 (2023) 57–69, https://doi.org/10.1016/j.neunet.2022.12.005

[196] L.-J. Marhuenda, F. Gomez-Donoso, M. Cazorla, HARTO: human activity recognition through optical flow architecture, in: 2024 International Joint Conference on Neural Networks (IJCNN), IEEE, 2024, pp. 1–7, https://doi.org/10.1109/IJCNN60899.2024.10650156

[197] C. Wu, Y. Gao, G. Li, C. Shi, Encoding learning network combined with feature similarity constraints for human action recognition, Multimed. Tools Appl. 83 (16) (2024) 48631–48658, https://doi.org/10.1007/s11042-023-17424-0

[198] K. Schindler, L. Van Gool, Combining densely sampled form and motion for human action recognition, in: Pattern Recognition: 30th DAGM Symposium Munich, Germany, June 10–13, 2008 Proceedings 30, Springer, 2008, pp. 122–131, https://doi.org/10.1007/978-3-540-69321-5_13

[199] H. Wang, Y. Yang, E. Yang, C. Deng, Exploring hybrid spatio-temporal convolutional networks for human action recognition, Multimed. Tools Appl. 76 (2017) 15065–15081, https://doi.org/10.1007/s11042-017-4514-3

[200] Y. He, S. Shirakabe, Y. Satoh, H. Kataoka, Human action recognition without human, in: Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 And 15-16, 2016, Proceedings, Part III 14, Springer, 2016, pp. 11–17, https://doi.org/10.1007/978-3-319-49409-8_2

[201] A. Kar, N. Rai, K. Sikka, G. Sharma, Adascan: adaptive scan pooling in deep convolutional neural networks for human action recognition in videos, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3376–3385, https://doi.org/10.1109/CVPR.2017.604

[202] L. Sun, K. Jia, K. Chen, D.-Y. Yeung, B.E. Shi, S. Savarese, Lattice long short-term memory for human action recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2147–2156, https://doi.org/10.1109/ICCV.2017.236

[203] I.C. Duta, B. Ionescu, K. Aizawa, N. Sebe, Spatio-temporal vlad encoding for human action recognition in videos, in: MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4–6, 2017, Proceedings, Part I 23, Springer, 2017, pp. 365–378, https://doi.org/10.1007/978-3-319-51811-4_30

[204] Y. Yuan, Y. Zhao, Q. Wang, Action recognition using spatial-optical data organization and sequential learning framework, Neurocomputing 315 (2018) 221–233, https://doi.org/10.1016/j.neucom.2018.06.071

[205] H. Zhang, M. Xin, S. Wang, Y. Yang, L. Zhang, H. Wang, End-to-end temporal attention extraction and human action recognition, Mach. Vis. Appl. 29 (7) (2018) 1127–1142, https://doi.org/10.1007/s00138-018-0956-5

[206] Y. Cai, Y. Shang, Y. Tan, Z. Tang, B. Zhao, Human action recognition based on deep learning, in: International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019: Applications and Techniques in Cyber Intelligence 7, Springer, 2020, pp. 1595–1600, https://doi.org/10.1007/978-3-030-25128-4_198

[207] Q. Wu, A. Zhu, R. Cui, T. Wang, F. Hu, Y. Bao, H. Snoussi, Pose-guided inflated 3D ConvNet for action recognition in videos, Signal Process. Image Commun. 91 (2021) 116098, https://doi.org/10.1016/j.image.2020.116098

[208] Z. Ren, Q. Zhang, X. Gao, P. Hao, J. Cheng, Multi-modality learning for human action recognition, Multimed. Tools Appl. 80 (11) (2021) 16185–16203, https://doi.org/10.1007/s11042-019-08576-z

[209] A. Aghaei, A. Nazari, M.E. Moghaddam, Sparse deep LSTMs with convolutional attention for human action recognition, SN Comput. Sci. 2 (3) (2021) 151, https://doi.org/10.1007/s42979-021-00576-x

[210] J. Yan, J. Wang, Q. Li, C. Wang, S. Pu, Weakly supervised regional and temporal learning for facial action unit recognition, IEEE Trans. Multimed. 25 (2022) 1760–1772, https://doi.org/10.1109/TMM.2022.3160061

[211] H. Zheng, B. Zhang, J. Lin, D. Zhao, G. Shang, A two-stage modality fusion approach for recognizing human actions, IEEE Sens. J. (2023), https://doi.org/10.1109/JSEN.2023.3316137

[212] Y. Yao, X. Wang, G. Zhou, Q. Wang, Hand action recognition from RGB-D egocentric videos in substations operations and maintenance, in: 2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG), IEEE, 2023, pp. 1–6, https://doi.org/10.1109/ETFG55873.2023.10408532

[213] R. Gowada, D. Pawar, B. Barman, Unethical human action recognition using deep learning based hybrid model for video forensics, Multimed. Tools Appl. 82 (19) (2023) 28713–28738, https://doi.org/10.1007/s11042-023-14508-9

[214] L. Wang, P. Koniusz, Flow dynamics correction for action recognition, in: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2024, pp. 3795–3799, https://doi.org/10.1109/ICASSP48485.2024.10446223

[215] C. Feichtenhofer, A. Pinz, A. Zisserman, Convolutional two-stream network fusion for video action recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1933–1941, https://doi.org/10.1109/CVPR.2016.213

[216] C. Chen, J. Zhang, Z. Gan, Human action recognition based on latent-dynamic conditional random field, in: 2013 International Conference on Wireless Communications and Signal Processing, IEEE, 2013, pp. 1–5, https://doi.org/10.1109/WCSP.2013.6677263

[217] S. Megrhi, W. Souidene, A. Beghdadi, Spatio-temporal SURF for human action recognition, in: Advances in Multimedia Information Processing–PCM 2013: 14th Pacific-Rim Conference on Multimedia, Nanjing, China, December 13–16, 2013. Proceedings 14, Springer, 2013, pp. 505–516, https://doi.org/10.1007/978-3-319-03731-8_47

[218] S. Audin, S. Nath, S. Basak, F. Rahman, R. Nath, S. Fattah, A human action recognition scheme based on spatio-temporal variation of region of interest in horizontal and vertical direction, in: 2014 International Conference on Informatics, Electronics & Vision (ICIEV), IEEE, 2014, pp. 1–4, https://doi.org/10.1109/ICIEV.2014.6850809

[219] B.C. Ustundag, M. Unel, Human action recognition using histograms of oriented optical flows from depth, in: International Symposium on Visual Computing, Springer, 2014, pp. 629–638, https://doi.org/10.1007/978-3-319-14249-4_60

[220] M. Sivarathinabala, S. Abirami, R. Baskaran, View invariant human action recognition using improved motion descriptor, in: Computational Intelligence in Data Mining-Volume 3: Proceedings of the International Conference on CIDM, 20–21 December 2014, Springer, 2015, pp. 545–554, https://doi.org/10.1007/978-81-322-2202-6_49

[221] C. Beaudry, R. Péteri, L. Mascarilla, An efficient and sparse approach for large scale human action recognition in videos, Mach. Vis. Appl. 27 (2016) 529–543, https://doi.org/10.1007/s00138-016-0760-z

[222] I.C. Duta, J.R.R. Uijlings, B. Ionescu, K. Aizawa, A.G. Hauptmann, N. Sebe, Efficient human action recognition using histograms of motion gradients and VLAD with descriptor shape information, Multimed. Tools Appl. 76 (2017) 22445–22472, https://doi.org/10.1007/s11042-017-4795-6

[223] C. Dhamsania, T. Ratanpara, Human action recognition using trajectory-based spatiotemporal descriptors, in: Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications: FICTA 2016, vol. 1, Springer, 2017, pp. 1–9, https://doi.org/10.1007/978-981-10-3153-3_1

[224] M. Sharif, M.A. Khan, F. Zahid, J.H. Shah, T. Akram, Human action recognition: a framework of statistical weighted segmentation and rank correlation-based selection, Pattern anal. appl. 23 (2020) 281–294, https://doi.org/10.1007/s10044-019-00789-0

[225] F. Afza, M.A. Khan, M. Sharif, S. Kadry, G. Manogaran, T. Saba, I. Ashraf, R. Damaševičius, A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection, Image Vis. Comput. 106 (2021) 104090, https://doi.org/10.1016/j.imavis.2020.104090

[226] P. Kalaivani, S. Mohamed Mansoor Roomi, M. Maheesha, V. Subathraa, Novel CNN architecture for human action recognition, in: Advances in Automation, Signal Processing, Instrumentation, and Control: Select Proceedings of i-CASIC 2020, Springer, 2021, pp. 2323–2333, https://doi.org/10.1007/978-981-15-8221-9_217

[227] S. Ali, M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2008) 288–303, https://doi.org/10.1109/TPAMI.2008.284

[228] C. Yuan, W. Hu, X. Li, S. Maybank, G. Luo, Human action recognition under log-euclidean riemannian metric, in: Computer Vision–ACCV 2009: 9th Asian Conference on Computer Vision, Xi'an, September 23–27, 2009, Revised Selected Papers, Part I 9, Springer, 2010, pp. 343–353, https://doi.org/10.1007/978-3-642-12307-8_32

[229] F. Monti, C.S. Regazzoni, Human action recognition using the motion of interest points, in: 2010 IEEE International Conference on Image Processing, IEEE, 2010, pp. 709–712, https://doi.org/10.1109/ICIP.2010.5651011

[230] Z. Zhang, J. Huang, Z.-N. Li, Discovering motion patterns for human action recognition, in: Advances in Multimedia Information Processing-PCM 2010: 11th Pacific Rim Conference on Multimedia, Shanghai, China, September 2010, Proceedings, Part II 11, Springer, 2010, pp. 716–727, https://doi.org/10.1007/978-3-642-15696-0_66

[231] K. Mozafari, J.A. Nasiri, N.M. Charkari, S. Jalili, Hierarchical least square twin support vector machines based framework for human action recognition, in: 2011 7th Iranian Conference on Machine Vision and Image Processing, IEEE, 2011, pp. 1–5, https://doi.org/10.1109/IranianMVIP.2011.6121601

[232] Q. Wu, S. Lu, Z. Wang, F. Deng, W. Kang, D.D. Feng, Structure context of local features in realistic human action recognition, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, 2011, pp. 1496–1501, https://doi.org/10.1109/ICCVW.2011.6130427

[233] J.-C. Nebel, M. Lewandowski, J. Thévenon, F. Martínez, S. Velastin, Are current monocular computer vision systems for human action recognition suitable for visual surveillance applications? in: Advances in Visual Computing: 7th International Symposium, ISVC 2011, Las Vegas, NV, USA, September 26–28, 2011. Proceedings, Part II 7, Springer, 2011, pp. 290–299, https://doi.org/10.1007/978-3-642-24031-7_29

[234] E. Zare Borzeshi, O. Perez Concha, M. Piccardi, Human action recognition in video by fusion of structural and spatio-temporal features, in: Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, SSPR&SPR 2012, Hiroshima, Japan, November 7–9, 2012. Proceedings, Springer, 2012, pp. 474–482, https://doi.org/10.1007/978-3-642-34166-3_52

[235] C. Orrite, P. Monforte, M. Rodriguez, E. Herrero, Human action recognition under partial occlusions, in: Pattern Recognition and Image Analysis: 6th Iberian Conference, IbPRIA 2013, Funchal, Madeira, Portugal, June 5–7, 2013. Proceedings 6, Springer, 2013, pp. 398–405, https://doi.org/10.1007/978-3-642-38628-2_47

[236] M.J. Marín-Jiménez, N. Pérez De La Blanca, M. Mendoza, Human action recognition from simple feature pooling, Pattern anal. appl. 17 (2014) 17–36, https://doi.org/10.1007/s10044-012-0292-8

[237] P. Guo, Z. Miao, Y. Shen, W. Xu, D. Zhang, Continuous human action recognition in real time, Multimed. Tools Appl. 68 (2014) 827–844, https://doi.org/10.1007/s11042-012-1084-2

[238] R. Hemati, S. Mirzakuchaki, Using local-based harris-phog features in a combination framework for human action recognition, Arab. J. Sci. Eng. 39 (2014) 903–912, https://doi.org/10.1007/s13369-013-0816-6

[239] M.S. Cheema, A. Eweiwi, C. Bauckhage, A stochastic late fusion approach to human action recognition in unconstrained images and videos, in: Pattern Recognition: 36th German Conference, GCPR 2014, Münster, Germany, September 2–5, 2014, Proceedings 36, Springer, 2014, pp. 616–628, https://doi.org/10.1007/978-3-319-11752-2_51

[240] A. Edison, C. Jiji, Hsga: a novel acceleration descriptor for human action recognition, in: 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), IEEE, 2015, pp. 1–4, https://doi.org/10.1109/NCVPRIPG.2015.7489944

[241] M. Glodek, G. Layher, F. Heilemann, F. Gawrilowicz, G. Palm, F. Schwenker, H. Neumann, uulmMAD–a human action recognition dataset for ground-truth evaluation and investigation of view invariances, in: Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction: Third IAPR TC3 Workshop, MPRSS 2014, Stockholm, Sweden, August 24, 2014, Revised Selected Papers 3, Springer, 2015, pp. 77–91, https://doi.org/10.1007/978-3-319-14899-1_8

[242] Z. Zhang, S. Liu, S. Liu, L. Han, Y. Shao, W. Zhou, Human action recognition using salient region detection in complex scenes, in: The Proceedings of the Third International Conference on Communications, Signal Processing, and Systems, Springer, 2015, pp. 565–572, https://doi.org/10.1007/978-3-319-08991-1_58

[243] S. Al-Ali, M. Milanova, H. Al-Rizzo, V.L. Fox, Human action recognition: contour-based and silhouette-based approaches, Comput. Vis. Control Syst.-2: Innov. Pract. (2015) 11–47, https://doi.org/10.1007/978-3-319-11430-9_2

[244] X. Xiao, L. Liu, H. Hu, Discriminative feature fusion with spectral method for human action recognition, in: Biometric Recognition: 10th Chinese Conference, CCBR 2015, Tianjin, China, November 13–15, 2015, Proceedings 10, Springer, 2015, pp. 641–648, https://doi.org/10.1007/978-3-319-25417-3_75

[245] A. Abdulmunem, Y.-K. Lai, X. Sun, 3D GLOH features for human action recognition, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 805–810, https://doi.org/10.1109/ICPR.2016.7899734

[246] B. Saghafi, D. Rajan, W. Li, Efficient 2D viewpoint combination for human action recognition, Pattern anal. appl. 19 (2016) 563–577, https://doi.org/10.1007/s10044-016-0537-z

[247] F. Khezeli, F. Razzazi, Real-time human action recognition with extreme learning machine, in: 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP), IEEE, 2017, pp. 159–163, https://doi.org/10.1109/IranianMVIP.2017.8342341

[248] B. Yousefi, C.K. Loo, A dual fast and slow feature interaction in biologically inspired visual recognition of human action, Appl. Soft Comput. 62 (2018) 57–72, https://doi.org/10.1016/j.asoc.2017.10.021

[249] J.M. Carmona, J. Climent, Human action recognition by means of subtensor projections and dense trajectories, Pattern Recognit. 81 (2018) 443–455, https://doi.org/10.1016/j.patcog.2018.04.015

[250] S. Mambou, O. Krejcar, K. Kuca, A. Selamat, novel human action recognition in RGB-D videos based on powerful view invariant features technique, Mod. Approaches Intell. Inf. Database Syst. (2018) 343–353, https://doi.org/10.1007/978-3-319-76081-0_29

[251] Q. Lei, H. Zhang, M. Xin, Y. Cai, A hierarchical representation for human action recognition in realistic scenes, Multimed. Tools Appl. 77 (2018) 11403–11423, https://doi.org/10.1007/s11042-018-5626-0

[252] S.P. Sahoo, R. Silambarasi, S. Ari, Fusion of histogram based features for human action recognition, in: 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), IEEE, 2019, pp. 1012–1016, https://doi.org/10.1109/ICACCS.2019.8728473

[253] K. Akila, S. Chitrakala, Highly refined human action recognition model to handle intraclass variability & interclass similarity, Multimed. Tools Appl. 78 (2019) 20877–20894, https://doi.org/10.1007/s11042-019-7392-z

[254] Y. Liu, Q. Zhang, W. Chen, Massive-scale complicated human action recognition: theory and applications, Future Gener. Comput. Syst. 125 (2021) 806–811, https://doi.org/10.1016/j.future.2021.06.060

[255] S.A.L. Nair, R.K. Megalingam, Fusion of bag of visual words with neural network for human action recognition, in: 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2022, pp. 14–19, https://doi.org/10.1109/Confluence52989.2022.9734221

[256] J. Chen, Y. Xu, C. Zhang, Z. Xu, X. Meng, J. Wang, An improved two-stream 3D convolutional neural network for human action recognition, in: 2019 25th International Conference on Automation and Computing (ICAC), IEEE, 2019, pp. 1–6, https://doi.org/10.23919/IConAC.2019.8894962

[257] J. Li, Z. Xu, J. Li, J. Wang, An improved human action recognition method based on 3D convolutional neural network, in: Advanced Hybrid Information Processing: Second EAI International Conference, ADHIP 2018, Yiyang, China, October 5–6, 2018, Proceedings 2, Springer, 2019, pp. 37–46, https://doi.org/10.1007/978-3-030-19086-6_5

[258] X. Zhao, Y. Yi, Z. Qiu, Q. Zeng, Feature retrieving for human action recognition by mixed scale deep feature combined with attention model, in: 2020 5th International Conference on Computer and Communication Systems (ICCCS), IEEE, 2020, pp. 235–239, https://doi.org/10.1109/ICCCS49078.2020.9118516

[259] P.C. Law, Y.L. Wang, L.C. Poon, A.W. Chung, K.W. Lai, Smart prison-video analysis for human action detection, in: IECON 2020 the 46th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2020, pp. 513–516, https://doi.org/10.1109/IECON43393.2020.9255402

[260] Y. Zhao, K.L. Man, J. Smith, K. Siddique, S.-U. Guan, Improved two-stream model for human action recognition, EURASIP J. Image Video Process. 2020 (2020) 1–9, https://doi.org/10.1186/s13640-020-00501-x

[261] H. Wang, Deeply-learned and spatial–temporal feature engineering for human action understanding, Future Gener. Comput. Syst. 123 (2021) 257–262, https://doi.org/10.1016/j.future.2021.04.021

[262] Q. Xu, W. Zheng, Y. Song, C. Zhang, X. Yuan, Y. Li, Scene image and human skeleton-based dual-stream human action recognition, Pattern Recognit. Lett. 148 (2021) 136–145, https://doi.org/10.1016/j.patrec.2021.06.003

[263] H.-B. Jang, C.-W. Lee, ST-GCN based human action recognition with abstracted three features of optical flow and image gradient, in: Frontiers of Computer Vision: 27th International Workshop, IW-FCV 2021, Daegu, South Korea, February 22–23, 2021, Revised Selected Papers 27, Springer, 2021, pp. 203–217, https://doi.org/10.1007/978-3-030-81638-4_17

[264] H.-B. Jang, C.-W. Lee, Multi-region based radial GCN algorithm for human action recognition, in: International Workshop on Frontiers of Computer Vision, Springer, 2022, pp. 325–342, https://doi.org/10.1007/978-3-031-06381-7_23

[265] F. Guo, T. Jin, S. Zhu, X. Xi, W. Wang, Q. Meng, W. Song, J. Zhu, B2c-AFM: bi-directional co-temporal and cross-spatial attention fusion model for human action recognition, IEEE Trans. Image Process. (2023), https://doi.org/10.1109/TIP.2023.3308750

[266] V.C. Bhagawat, C. Kalaiarasan, R. Chinnaiyan, Motion-classification using CNN and LSTM, in: 2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS), 1, IEEE, 2023, pp. 1–6, https://doi.org/10.1109/ICCAMS60113.2023.10526025

[267] S.K. Yadav, E. Pahwa, A. Luthra, K. Tiwari, H.M. Pandey, SWTA: sparse weighted temporal attention for drone-based activity recognition, in: 2023 International Joint Conference on Neural Networks (IJCNN), IEEE, 2023, pp. 01–08, https://doi.org/10.1109/IJCNN54540.2023.10191750

[268] J. Liu, T. Mao, Y. Huang, D. He, An action recognition algorithm based on two-stream deep learning for metaverse applications, in: 2024 International Wireless Communications and Mobile Computing (IWCMC), IEEE, 2024, pp. 639–642, https://doi.org/10.1109/IWCMC61514.2024.10592362

[269] A. Alavigharahbagh, V. Hajihashemi, J.J. Machado, J.M.R. Tavares, Deep learning approach for human action recognition using a time saliency map based on motion features considering camera movement and shot in video image sequences, Information 14 (11) (2023) 616, https://doi.org/10.3390/info14110616

[270] A.A. Gharahbagh, V. Hajihashemi, M.C. Ferreira, J. Machado, J.M.R. Tavares, Hybrid time-spatial video saliency detection method to enhance human action recognition systems, Multimed. Tools Appl. 83 (30) (2024) 74053–74073, https://doi.org/10.1007/s11042-024-18126-x

[271] I. Laptev, T. Lindeberg, Local descriptors for spatio-temporal recognition, in: International Workshop on Spatial Coherence for Visual Motion Analysis, Springer, 2004, pp. 91–103, https://doi.org/10.1007/11676959_8

[272] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2, IEEE, 2005, pp. 1395–1402, https://doi.org/10.1109/ICCV.2005.28

[273] D. Weinland, R. Ronfard, E. Boyer, Free viewpoint action recognition using motion history volumes, Comput. Vis. Image Underst. 104 (2–3) (2006) 249–257, https://doi.org/10.1016/j.cviu.2006.07.013

[274] M.D. Rodriguez, J. Ahmed, M. Shah, Action Mach a spatio-temporal maximum average correlation height filter for action recognition, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8, https://doi.org/10.1109/CVPR.2008.4587727

[275] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8, https://doi.org/10.1109/CVPR.2008.4587756
[276] M.S. Ryoo, J.K. Aggarwal, UT-interaction dataset, ICPR contest on semantic description of human activities (SDHA), 2010, http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html.
[277] M.S. Ryoo, J.K. Aggarwal, Spatio-temporal relationship match: video structure comparison for recognition of complex human activities, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 1593–1600, https://doi.org/10.1109/ICCV.2009.5459361
[278] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, T. Serre, HMDB: a large video database for human motion recognition, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 2556–2563, https://doi.org/10.1109/ICCV.2011.6126543
[279] K.K. Reddy, M. Shah, Recognizing 50 human action categories of web videos, Mach. Vis. Appl. 24 (5) (2013) 971–981, https://doi.org/10.1007/s00138-012-0450-4
[280] K. Soomro, A.R. Zamir, M. Shah, A dataset of 101 human action classes from videos in the wild, Cent. Res. Comput. Vis. 2 (11) (2012) 1–7.
[281] C. Chen, R. Jafari, N. Kehtarnavaz, UTD-MHAD: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 168–172, https://doi.org/10.1109/ICIP.2015.7350781
[282] A. Shahroudy, J. Liu, T.-T. Ng, G. Wang, NTU rgb+ d: a large scale dataset for 3D human activity analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1010–1019, https://doi.org/10.1109/CVPR.2016.115
[283] Y. Zhu, Z. Lan, S. Newsam, A. Hauptmann, Hidden two-stream convolutional networks for action recognition, in: Asian Conference on Computer Vision, Springer, 2018, pp. 363–378, https://doi.org/10.1007/978-3-030-20893-6_23
[284] A. Piergiovanni, M.S. Ryoo, Representation flow for action recognition, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 9937–9945, https://doi.org/10.1109/CVPR.2019.01018
[285] J.W. Woods, Chapter 7 - image enhancement and analysis, in: Multidimensional Signal, Image, and Video Processing and Coding, (Second Edition), Second ed, Academic Press, Boston, 2012, pp. 223–256, https://doi.org/10.1016/B978-0-12-381420-3.00007-2
[286] B.D. Lucas, Generalized Image Matching by the Method of Differences, Carnegie Mellon University, 1985, https://dl.acm.org/doi/abs/10.5555/912172.
[287] E.H. Adelson, The extraction of spatio-temporal energy in human and machine vision, in: IEEE Workshop on Motion: Representation and Analysis, 1986, pp. 151–155, https://cir.nii.ac.jp/crid/1572824499448438016.
[288] J.K. Kearney, W.B. Thompson, D.L. Boley, Optical flow estimation: an error analysis of gradient-based methods with local optimization, 1987, https://doi.org/10.1109/TPAMI.1987.4767897
[289] E.P. Simoncelli, Distributed Representation and Analysis of Visual Motion (Ph.D. thesis), Massachusetts Institute of Technology, 1993, https://dspace.mit.edu/bitstream/handle/1721.1/12590/28232726-MIT.pdf?sequence=2.
[290] E.P. Simoncelli, E.H. Adelson, D.J. Heeger, Probability distributions of optical flow, in: CVPR, 91, 1991, pp. 310–315, https://doi.org/10.1109/CVPR.1991.139707
[291] J.-Y. Bouguet, et al., Pyramidal implementation of the affine Lucas kanade feature tracker description of the algorithm, Intel corp. 5 (1–10) (2001) 4.
[292] T. Romera, A. Petreto, F. Lemaitre, M. Bouyer, Q. Meunier, L. Lacassagne, D. Etiemble, Optical flow algorithms optimized for speed, energy and accuracy on embedded GPUs, J. Real-Time Image Process. 20 (2) (2023) 32, https://doi.org/10.1007/s11554-023-01288-6
[293] R. Radhakrishnan, D. Sharma, V. Murthy, A review on particle image velocimetry and optical flow methods in riverine environment, 2017, https://www.researchgate.net/publication/320908264_A_Review_On_Particle_Image_Velocimetry_And_Optical_Flow_Methods_In_Riverine_Environment.

Author biography



Abdorreza Alavi Gharahbagh is a researcher at the Faculdade de Engenharia da Universidade do Porto, Portugal, whose work applies machine learning, optimisation, and mathematical modelling across an unusually wide range of fields. His current research at Porto centres on video and image processing and human action recognition, including video saliency and the handling of camera motion. A defining thread across his work is the use of evolutionary and metaheuristic optimisation alongside deep learning to address practical problems. In signal processing he has also worked on acoustic and sound event detection, acoustic scene classification, and fault detection from sound, with a recurring interest in combining audio and visual information. His research further spans medical image analysis, natural language processing for Persian text, and satellite and hyperspectral image classification, alongside a substantial and highly cited body of work on environmental and energy systems, including marine oil spill response, urban stormwater management, and renewable energy optimisation. His methods are implemented across MATLAB, Python,R, and mathematical modelling software.



Vahid Hajihashemi is a researcher at the Faculdade de Engenharia da Universidade do Porto, Portugal. His research spans signal processing and machine learning, with a focus on audio analysis, including acoustic and sound event detection, acoustic scene classification, and processing in noisy real world conditions, alongside video and image work such as human action recognition and visual saliency. A recurring thread is the integration of audio and visual information into multimodal systems. He has also contributed to natural language processing, medical and satellite image analysis, and 3D holographic compression, and his work extends to mechanical domains such as flywheel energy storage and fault detection, where mechanical principles are combined with acoustic event detection. His methods rest on a strong mathematical and probabilistic foundation, drawing on statistical modelling, time-frequency analysis, and optimisation, and are implemented across Python, MATLAB, C, C++, R, and Visual Studio and Visual Studio Code. His research connects rigorous theory with practical problems in smart cities, healthcare, and multimedia.



José J. M. Machado received the master's degree in mechanical engineering and the Ph.D. degree in mechanical engineering, with a focus on adhesively bonded automotive structures, from the Univer- sity of Porto, in 2012 and 2019, respectively. His doctoral research was conducted in collaboration with Nagase ChemteX and Aston Martin Lagonda. He is an Assistant Professor with the Department of Me- chanical Engineering (DEMec), Faculty of Engineering, University of Porto (FEUP), where he has been a Faculty Member since October 2021. From 2012 to 2016, he was a Mechanical Engineer, specialising in electric motors for hazardous areas within the Oil and Gas industry, and in the maintenance of mining equipment. He held a postdoctoral position with Tokyo Institute of Technology from 2019 to 2020. He has co supervised several M.Sc. theses and is the co author of more than 60 articles published in national and international peer reviewed journals. His primary research interests lie in design and manufactur- ing, computational vision, and new product development.



João Manuel R. S. Tavares received the degree in mechanical en- gineering, the M.Sc. and Ph.D. degrees in electrical and computer engineering, and the Habilitation degree in mechanical engineering from the Universidade do Porto (UP), Portugal, in 1992, 1995, 2001, and 2015, respectively. He is a Senior Researcher with the Institute of Science and Innovation in Mechanical and Industrial Engineering (INEGI) and a Full Professor with the Department of Mechanical Engineering (DEMec), Faculdade de Engenharia da Universidade do Porto (FEUP). Since June 2023, he has been the Head of DEMec. He has made significant contributions to his field, serving as a co editor for over 95 books and co authoring more than 55 book chapters and 700 articles in international and national journals and conferences. He also holds three international and three national patents. His re- search focuses on computational vision, medical imaging, biomechan- ics, biomedical engineering, and new product development. He is the Co Founder and Co Editor of the "Lecture Notes in Computational Vision and Biomechanics" series (Springer), the Founder and Edi- tor in Chief of Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization (Taylor & Francis), and the Editor in Chief of Computer Methods in Biomechanics and Biomedi- cal Engineering (Taylor & Francis).