Kinect Based System for Breast 3D Reconstruction

Pedro Miguel Ferro da Costa

Supervisor: Hélder Filipe Pinto de Oliveira, PhD

Co-Supervisor: Jaime dos Santos Cardoso, PhD

Integrated Master in Bioengineering

June, 2014

Faculdade de Engenharia da Universidade do Porto

Kinect Based System for Breast 3D Reconstruction

Pedro Miguel Ferro da Costa

Dissertation submitted for the

Integrated Master in Bioengineering

June, 2014

Abstract

Breast cancer is, nowadays, one of the most severe diseases affecting women worldwide. The Breast Cancer Conservative Treatment (BCCT) appeared in the past few years as a reliable substitute for mastectomy as the preferred treatment: instead of removing completely the breast, the tumor is excised as well as some living healthy tissue. Some medical procedures associated with BCCT required volumetric and shape information about the breast. Such necessity motivates the usage of tridimensional (3D) reconstruction tools for 3D breast modeling.

The commonly used reconstruction systems, although having high resolution, often require complex and costly hardware and expertise staff, not being easily implemented on the medical daily routine. This created the need for a practical tool that is easy to use, being able to reconstruct accurately the body structures.

This work aims at designing a simple and affordable Kinect-based system for the 3D reconstruction of the breast. The RGB-D data obtained using the Kinect is used to generate point clouds, that are registered to obtain the reconstructed models. The proposed system relies on a Tessellation-based coarse registration methodology, that uses the Delaunay Triangulation principle and a color-validation stage to select robust keypoints and pre-align the point clouds. In addition, two newly designed evaluation metrics are proposed: one based on the polar representation of the reconstructed and the reference models; the other based on an histogram representation of the distances between the keypoints of the point clouds.

The obtained results are promising and show that the tool here described is capable of performing a good reconstruction of the female breast in feasible time. The relevant and distinctive structures of the breast were correctly enhanced and reconstructed and a good performance was found even when dealing with low-overlapping and highly misaligned views. It is expected the implementation of this reconstruction system in an aesthetics outcome evaluation and/or surgery planning tool in a near future.

Keywords: Breast Cancer; 3D Modeling; Microsoft Kinect; Tessellation; Delaunay Triangulation; Point Cloud.

Resumo

O cancro da mama é atualmente uma das doenças mais severas associadas ao sexo feminino. O tratamento conservativo do cancro da mama estabeleceu-se nos últimos anos como um substituto confiável à tradicional mastectomia: em vez de remover completamente a mama doente, apenas o tumor é excisado, em conjunto com algum tecido saudável que o rodeia. Os procedimentos médicos associados a este tratamento requerem informação acerca da forma da mama e do seu volume, motivando a utilização de ferramentas de reconstrução 3D.

Os sistemas utilizados de forma mais comum, apesar de terem alta-resolução, requerem hardware caro e complexo, bem como staff especializado. Isto faz com que a sua implementação não seja fácil num contexto médico diário. Isto criou a necessidade de uma ferramenta prática, fácil de usar e que seja capaz de realizar uma boa reconstrução as partes do corpo em estudo.

Este trabalho tem como objectivo o desenvolvimento de um sistema simples e de baixo custo para reconstrução 3D da mama, utilizando o Microsoft Kinect. A informação RGB-D adquirida utilizando o Kinect é convertida em nuvens de pontos que são, por sua vez, registadas e alinhadas para obter os modelos reconstruídos. O sistema aqui proposto baseia-se num método de *coarse registration* que foi denominado Tessellation e que usa o princípio da Triangulação de Delaunay e um critério de cor para selecionar pontos de interesse na nuvem de pontos e, com base neles, estabelecer o seu alinhamento. Além disso, duas novas métricas de avaliação são propostas: uma é baseada em representações polares das nuvens de pontos de referência e reconstruída; a outra é baseada na construção de histogramas de distâncias representativos dos modelos em estudo.

Os resultados obtidos são promissores e demonstram que a ferramenta aqui descrita é capaz de reconstruir adequadamente a mama num tempo aceitável. As estruturas mais distintivas e relevantes da mama foram corretamente reconstruídas e uma boa performance foi observada, mesmo lidando com nuvens de pontos pouco sobrepostas ou marcadamente desalinhadas entre si. Num futuro próximo, é expecável a implementação deste sistema de reconstrução num sistema de avaliação do resultado estético do tratamento conservativo ou num sistema de planeamento cirúrgico.

Keywords: Cancro da Mama; Modelação 3D; Microsoft Kinect; Tessellation; Triangulação de Delaunay; Nuvem de Pontos.

Agradecimentos

Escrevo estas linhas a olhar para a minha pasta, onde as fitas cor de engenharia têm palavras dirigidas por todos quanto gosto. É muito gratificante sentir-me rodeado de tantas pessoas importantes para mim e que sentem que, de uma forma ou outra, sou importante para elas. Desde sempre me convenci de que somos mais do que os nossos sonhos, ambições ou desejos. Também somos um bocadinho aquilo que as nossas experiências e os nossos amigos ou conhecidos fazem de nós. E num fim de caminhada como este, não posso deixar de me sentir para sempre agradecido a todos aqueles que contribuíram para que eu seja, sem sombra de dúvida, uma pessoa feliz.

Este trabalho não teria sido possível sem o apoio e a contribuição de diversas pessoas, às quais dirijo desde já o meu obrigado: ao Professor Jaime Cardoso por me ter permitido realizar este trabalho no seio grupo VCMI do INESC-TEC, e também pelo acompanhamento científico prestado e por fazer (sempre) a pergunta difícil, aquela que me fazia sentir que ainda podemos fazer sempre mais e melhor. Um especial e fraterno agradecimento ao Hélder Oliveira, que foi um orientador excecional e que me fez sentir desde o primeiro dia que acreditava em mim e que seria capaz de realizar este trabalho. Este é um obrigado de aluno para mestre, mas também de amigo para amigo. Obrigado por tudo. Ao João Pedro Monteiro e ao Hooshiar Zolfagharnasab porque, sem eles e sem o trabalho de equipa que desenvolmentos, este trabalho estaria, certamente, mais pobre, quer em conteúdo quer em ideias. Finalmente, ao projeto PICTURE, especialmente à Fundação Champalimaud e ao University College London, pela disponibilização de dados[1].

Quando dizem que não fazemos um curso sozinhos, não se enganam. A vocês, Mario, Catalina, Hugo, Fred, Liliana, Ivo, Rui, Dinis, André Costa, André Silva e Ricardo Lé, quer estejam agora mais presentes ou ausentes, mais perto ou longe, um enorme obrigado e o abraço ou beijinho respetivos. Fizeram dos meus anos de estudante univeristário os mais divertidos, sonhadores e de aprendizem que já tive. Num registo semelhante, ao João Castelo, por tudo o que ensinou nos últimos tempos e por toda a confiança depositada nas minhas capacidades.

À equipa de natação do Sporting Clube de Espinho. Aos que estiveram, aos que estão e a todos os quantos cruzaram o meu caminho. Passei 12 anos da minha vida com vocês, 11 meses por ano, 6 dias por semana. Foi aí que aprendi a vencer, a perder, onde fui campeão e onde também percebi que nem tudo é como queremos. Por me terem ensinado a ser guerreiro, persistente e alguém com objetivos maiores do que o esperado, o meu obrigado.

Aos amigos que a vida vai fazendo ficar para sempre, especialmente à Ivana e ao Zé, à Marta, ao Vítor, à Mariana, ao Gonçalo, à Sof, à Lili ao Alex e ao Rui, ao João Paulo, ao Luís, à Xana, ao Lemos, à Catarina Pimenta, à Rita Ribeiro, Leandro ao Cardoso e à Liseta, por serem essas pessoas. Não imaginam o quão feliz sou por sentir o quanto confiam em mim, por saber que

[1]The work leading to these results has received funding from the European Community's Seventh Framework Programme [grant number FP7600948]. This work was also financed by the European Regional Development Fund (ERDF) through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the Fundac¸ ~ao para a Ci^encia e Tecnologia (FCT) - Portugal within project PTDC/SAU-ENB/114951/2009. We acknowledge Professor Keshtgar of the Royal Free Hospital in London for providing data.

nunca faltarão abraços, gargalhadas, praia e aventura e brincadeiras. Por serem os meus amigos e alguém de quem nunca estarei separado, por muito longe que esteja.

À minha família, em especial aos meus primos Rui e Matilde e aos meus avós. Os primeiros porque são únicos e cresceram juntos comigo, confiando em mim, olhando para mim com um exemplo e um amigo. E aos meus avós, por me darem mais do que aquilo que precisavam e por todo o orgulho que sei que sentem por mim. Não tenho palavras capazes de vos agradecer o tanto que fazem por mim. Aos meus padrinhos, por estarem sempre presentes e por me terem dado a mão em todos os passos do meu crescimento.

Ao meu irmão, por ser um miúdo incrível e que sempre olhou para mim como uma referência e um exemplo, embora eu esteja longe de o ser para alguém. Porque tal carinho me enche o coração e me faz sentir todos os dias responsável por estar do teu lado e te apoiar em tudo. E aos meus pais, para os quais todas a palavras são poucas. São exemplos para mim, especialmente por tudo aquilo de que abdicam para não nos faltar nada, pela boa disposição, pelo ocasional raspanete e pelo aconselhamento. Aos três, amo-vos do fundo do coração.

À Sara. Por me ter ensinado que também podemos trabalhar sem enrugar a testa. E principalmente por todo o amor, pelos sorrisos e pelos planos. Obrigado.

Pedro Costa

"The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can,
Pursuing it with weary feet,
Until it joins some larger way,
Where many paths and errands meet.
And whither then? I cannot say."

J.R.R. Tolkien

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Bidimensional
3D	Tridimensional
4PCS	4-Points Congruent Sets
BCCT	Breast Cancer Conservative Treatment
CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide Semiconductor
fps	Frames per Second
DT	Delaunay Triangulation
EMD	Earth Movers Distance
FOV	Field of View
ICP	Iterative Closest Point
IEEE	Institute of Electrical and Electronics Engineers
INESC	Instituto de Engenharia de Sistemas e Computadores
IR	Infrared
ISBI	International Symposium on Biomedical Imaging
NITE	Natural Interaction Technology for End-User
PCA	Principal Component Analysis
RANSAC	Random Sample Consensus
ROI	Region of Interest
SDK	Software Development Kit
SR	Super Resolution
TOF	Time of Flight
VCMI	Visual Computing and Machine Intelligence

Chapter 1

Introduction

Computer vision techniques, especially tridimensional (3D) modeling and reconstruction of scenes and objects have become an important tool for several fields of research (Jarvis, 1983; Sansoni et al., 2009) such as robotics, navigation, automated cartography, quality control and, more recently, medical diagnosis (Blais, 2004). Before the 1970's, the 3D assessment of an object structure was obtained using contact probes (Besl, 1988), mainly for quality control purposes (Sansoni et al., 2009). However, the development of software and the appearance of 3D sensors allowed 3D modeling to gain its space in the aforementioned applications.

3D sensors are capable of creating depth-maps by acquiring information from a surrounding scene. These sensors started as research curiosities. Jarvis (1983) analyzed the first methods for image range sensing towards a first comprehension of their applicability. Later, Besl (1988) surveyed the existing active optical range sensors. In the following years, these works were complemented by other researches (Kanade and Asada, 1981; Tiziani, 1997; Chen et al., 2000) that gave relevant insight on the development and understanding on how it is possible to sense the world around us. The area grew so much that, nowadays, 3D modeling is only a fraction of the myriad of ways the sensing of an environment can be done (Blais, 2004).

At this time, accurate and high-resolution outcomes can be obtained using the established techniques for 3D sensing of scenes and objects. Nevertheless, the common techniques are being adapted and applied for new problems, creating a constant and challenging pressure for innovation. This is especially true in the Biomedical Engineering field. The versatility of the new techniques and the accuracy of computer-aided diagnosis based on these tools is growing (Way et al., 2006; Chen et al., 2010; Iwami and Umeda, 2011; Oliveira et al., 2014). 3D sensors provide range data containing the necessary properties of an object or scene for the generation of a mesh. After that, data is processed and structured to create a consistent polygonal surface that realistically represent the modeled scene (Remondino and El Hakim, 2006). Afterwards, color and texture information can be added to the surface by other image processing techniques.

1.1 Motivation

Annually, more than a million women are diagnosed with breast cancer, 14% of which resulting in death. Breast cancer presents high incidence and prevalence rates, being a relevant disease and considered a matter of public health (Jemal et al., 2011). Considering surgical intervention, two different treatment procedures are possible: mastectomy or BCCT. While the mastectomy procedure entirely removes the breast, Breast Cancer Conservative Treatment (BCCT) excises the tumor together with a margin of cancer free breast tissue (Oliveira et al., 2013). This conservative approach allows a local intervention on the disease and provides similar survival rates to mastectomy. Moreover, best aesthetic results are achieved (Cardoso et al., 2014).

The medical procedures related with the BCCT have evolved towards the usage of affordable and practical tools, along with the recent inclusion of volumetric information of the breast. Creating a richer three-dimensional (3D) model of the female torso can be associated to many advantages: the breast can be viewed from a multitude of angles, it is possible to estimate the volume/volume deficit between breasts and surgery planning. The development of a deformable, patient-fitted 3D model of the breast would allow physicians to quantitatively analyse the characteristics of the breast and interactively plan the surgery towards a more aesthetic outcome.

The commonly used methodologies to assess the 3D shape of the breast demand the usage of costly and operationally complex systems and require expertise hardware and staff. In addition, as the breast is a bare body part (nearly featureless), there is the need to searching for an affordable and simple methodology, capable of providing reliable 3D representations of a scene. To overcome that necessity, some low-cost RGB-D sensors, like the Microsoft Kinect, can being used. Their capacity to store color and distance information can be used to perform the whole body or body part 3D reconstruction (Newcombe et al., 2011; Oliveira et al., 2014; Sun et al., 2012; Böhm, 2012; Izadi et al., 2012; Wang et al., 2012; Cui and Stricker, 2013; Han et al., 2013), as well as other applications (see Newcombe et al. (2011); Cruz et al. (2012); Izadi et al. (2012); Han et al. (2013)). Textured point clouds can be obtained from the RGB-D data and its registration can be used to reconstruct the female torso and the breast.

The standard registration algorithms perform well under controlled acquisition procedures. However, when operating under less-controlled conditions (high translations and rotations between views), the quality of the reconstructions start to diverge. A practical tool specifically designed to be used in a medical environment is intended to correctly compensate this misalignments, leading to better reconstructions. This motivates the development of a registration pipeline especially capable of dealing with these uncontrolled conditions.

On the other hand, the reconstructed and the *ground-truth* 3D models are often mapped into different coordinate systems. This causes the common distance-based reconstruction evaluation methodologies to fail. A pose-invariant metric that allows the computation of the similarity between two 3D models is needed, in order to correctly evaluate the quality of the reconstructions.

1.2 Objectives

The final goal of this research is to develop a simple and affordable system for the creation of a 3D model of the breast using the Microsoft Kinect. The acquired raw RGB-D data should be used to create point clouds from the object of interest. Then, the proposed system must be capable of extracting robust keypoints from such point clouds in order to perform their registration in the same coordinate basis. A smaller goal is to design a pose-invariant evaluation methodology that answers to the limitations associated to the common evaluation approaches.

1.3 Contributions

The contributions of the thesis are listed below. In this thesis:

1. A Tessellation-based coarse registration method that uses both color and depth information was proposed, in contrast with the state of the art depth-based methodologies that are commonly used.

2. Two new metrics for the evaluation of point cloud registration were developed.

The work related with this thesis resulted in two international conference submissions:

- Pedro Costa, Hooshiar Zolfagharnasab, João P. Monteiro, Jaime S. Cardoso, Hélder P. Oliveira, "'3D Reconstruction of body parts using RGB-D sensors: Challenges from a biomedical perspective"', in the Proceedings of the *5th International Conference on 3D Body Scanning Technologies* (accepted); 2014.[1]

- Pedro Costa, João P. Monteiro, Hooshiar Zolfagharnasab, Jaime S. Cardoso, Hélder P. Oliveira, "'Tessellation-based Coarse Registration Method for 3D Reconstruction of the Female Torso"', in the Proceedings of the *IEEE Internacional Conference on Bioinformatics and Biomedicine* (submitted); 2014.[2]

1.4 Document Structure

This thesis is organized in seven major chapters, each one related to a specific part of the research. The remainder of this section presents the main motivations, objective and thesis contributions.

In Chapter 2, an overview of the 3D Imaging Techniques is presented. The chapter starts by presenting each of the most relevant techniques, stating its advantages and drawbacks. Then, a technique comparison is made, based on some features that show how simple and affordable a system that uses each of the specific techniques can be.

[1] http://www.3dbodyscanning.org/
[2] http://scm.ulster.ac.uk/ bibm/2014/

Chapter 3 summarizes the main 3D body reconstruction systems and techniques, then focusing on the RGB-D cameras, especially the Kinect. It finishes by presenting the major developments that have been done in the past few years using this device.

Chapter 4 starts by presenting the state-of-the-art methodologies on Point Cloud Registration. Then, in Chapter 5, the whole pipeline of the developed Kinect Based System for 3D Modeling is shown and explained, from the acquisition and calibration until the registration. The obtained results and the respective discussion are found in Chapter 6.

Finally, conclusions and some insight on future work and research are provided in Chapter 7.

Chapter 2

Overview of 3D Imaging Techniques

The main objective of 3D modeling is to reconstruct the shape of an object or scene from a set of bidimensional (2D) images. The output may be point coordinates, including its depth or, in general, the 3D shape (as characterized by the orientation of each point) of an object. The image data that represents a 3D object are also referred to as range images, depth-maps or 2.5-D images, due to the inclusion of depth information for each point coordinate (Haralick and Shapiro, 1991).

Currently, the non-contact systems that allow the generation of 3D models use light-wave-based systems, in particular active or passive 3D sensors (Remondino and El Hakim, 2006), as seen in Figure 2.1.

Active sensors act by projecting light onto the objects and recording the reflected energy (Sansoni et al., 2009). This information is then used for the reconstruction. In addition, they allow to directly representing an object or scene. These sensors keep the properties of invariance in rotation and translation of 3D objects, being able to solve most of the problems found by 3D scanning systems (Beraldin et al., 2003). On the other hand, passive sensors only acquire information from the scene using information acquired by simple cameras (Remondino and El Hakim, 2006; Sansoni et al., 2009). This classification is not rigid. Depending on the problem on hand, mixed approaches can be used, as well as adaptations of each of the techniques (Blais, 2004).

In this section, a general overview of the existent techniques is presented, as well as a short review of their physical principles, for a comparative evaluation of each one within the scope of this work.

Besides from laser triangulation, structured light and time of flight techniques, no other active sensoring methods represent a major contribution for the 3D modeling field of research (Moons et al., 2010). The same applies for passive techniques, with exception to the traditional stereo vision methodologies.

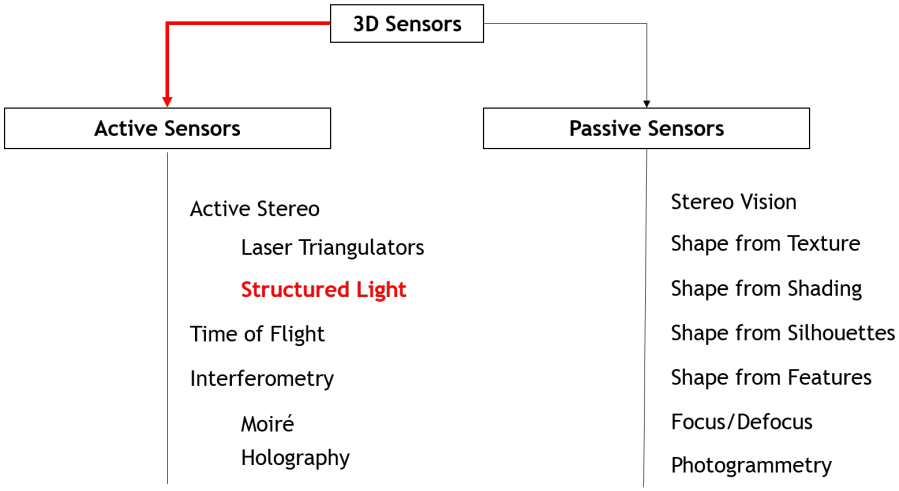


Figure 2.1: Tree representing the different types of 3D active and passive sensors. Highlighted branch describes the chosen approach for this work (please see section 2.2.4). Adapted from Sansoni et al. (2009); Remondino and El Hakim (2006); Mada et al. (2003).

2.1 Stereo Vision

Stereo vision is a passive technique for the reconstruction of a scene, based on images observed from multiple viewpoints (Poggio and Poggio, 1984; Blais, 2004; Moons et al., 2010). Consider two images, taken from the same scene from different perspectives. The principle behind the reconstruction is the following: given two projections of the same pixels of the scene onto the two viewpoint images, its 3D position lays in the intersection of the two projection rays (Moons et al., 2010) as seen in Figure 2.2.

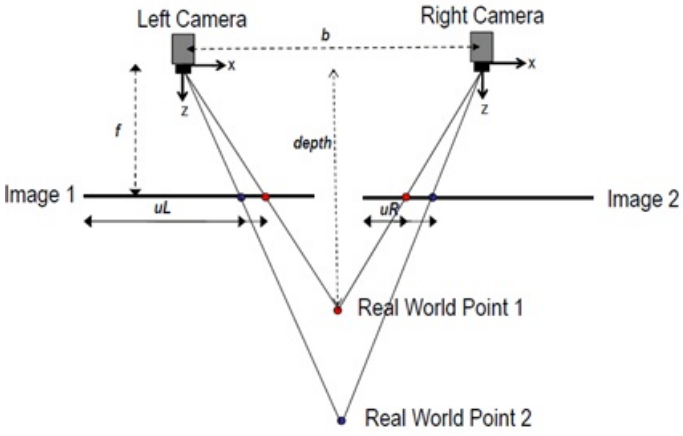


Figure 2.2: The principle of passive stereo vision. Given two images of an object, the position of each point is calculated as the intersection of two projections of the point in each image. This process is also called 'triangulation. [1]

[1] http://www.ni.com/white-paper/14103/pt/

The process is repeated for several pixels to get the whole shape of the object. This *triangulation* requires previous knowledge about the acquisition system: the relative position of the cameras, their orientation and other settings like lens curvature or focal length (Poggio and Poggio, 1984). This information can be gathered on the calibration matrix K presented below:

$$K = \begin{bmatrix} f & s & c_x \\ 0 & a_f & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.1}$$

where a is the aspect ratio of the image, f is the focal length, c_x and c_y are the pixel coordinates of the center of the image plane and s is the skew factor. The calibration matrix yields an approximation to the situation in which the physical imaging plane is not perfectly perpendicular to the optical axis of the lens of objective. This matrix is also inversely proportional to the tangent of the angle between the X and Y-axis of the camera-centered reference frame (Moons et al., 2010).

This knowledge allows the user to directly estimate the world coordinates (X,Y,Z) of the scene point M based on its pixel coordinates $m(X,Y)$ (Moons et al., 2010; Szeliski, 2011):

$$m = P \cdot M \tag{2.2}$$

where P is the relative information between viewpoints, and is function of a rotation matrix R and a translation vector t:

$$P = K \cdot [R|t] \tag{2.3}$$

However, if these settings are unknown, stereo vision is not able to deal with the point correspondence problem. The answer to that problem can rely on a camera calibration stage based on the content of the image, that can be performed using epipolar geometry (Szeliski, 2011; Moons et al., 2010).

2.2 Active Reconstruction

2.2.1 Laser Triangulation

A laser triangulator projects either well-defined (coherent) laser points or laser stripes. The emitted light interacts with the objects on the scanned scene, being backscattered to a sensor, usually a CCD (charge-coupled device). The distance between each point on the scene and the sensor is then calculated based on the active triangulation principle (Beraldin et al., 2003; Blais et al., 1988), as seen in Figure 2.3.

One of the main advantages of using laser light is its very large depth of focus for large baselines. In addition, the CCD and lens configuration guarantee that the sensor always receives a focused image (Blais, 2004). In the conventional configuration, a compromise between the Field of View (FOV), the resolution and the shadow effects when large inclinations are used is needed

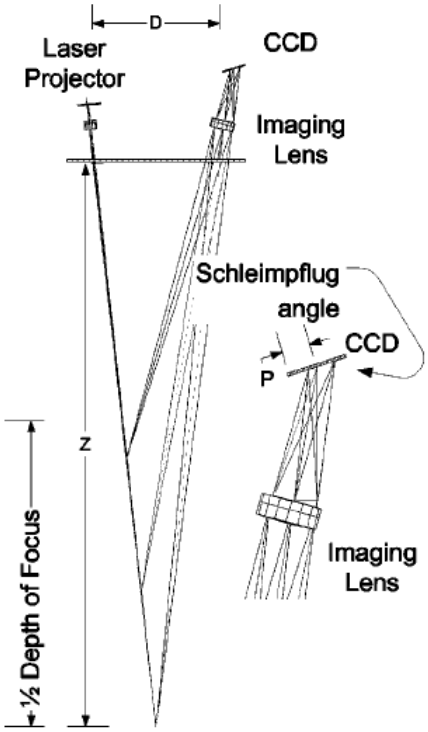


Figure 2.3: The basic triangulation principle. The range z has a direct relation with the position of the image in the CCD. From Blais (2004).

(Blais, 2004; Sansoni et al., 2009). To overcome this problem, a dual view principle was suggested (Rioux, 1984; Rioux and Blais, 1986; Blais, 2004; Sansoni et al., 2009). Using this approach, the accuracy of the sensor is improved due to the presence of redundancies on the measurements (Blais, 2004).

Laser triangulation is mostly used for volumetric descriptions, geometric calculations and dislocation measurements.

2.2.2 Structured Light

Structured light systems project a 2D pattern of non-coherent light (emitted photons have different wave frequencies and oscillate in different directions) over an object (Sansoni et al., 2009). The projected pattern is deformed by the object and can be used to describe its structure (Yang et al., 1984), orientation or texture (Batlle et al., 1998).

In other words, the active triangulation principle mentioned above is used, not for a single point or line, but for a specific pattern. It is possible to obtain the range information for a set of points at the same time in a single video frame (Blais, 2004; Sansoni et al., 2009). Each point on the sensor is directly associated to a point on the scene, without the need of geometrical constraints (Batlle et al., 1998).

Almost any kind of pattern can be created with different coding schemes: grids and dots (Bickel et al., 1985) or lines (Blais and Rioux, 1987; Vuylsteke and Oosterlinck, 1987) are some

of the most elegant solutions found, but there are also some intricate checkerboard-like patterns (Moons et al., 2010). The most popular methods use a temporal coding scheme (Blais, 2004), which results in the projection of a binary line pattern, as seen in Figure 2.4.

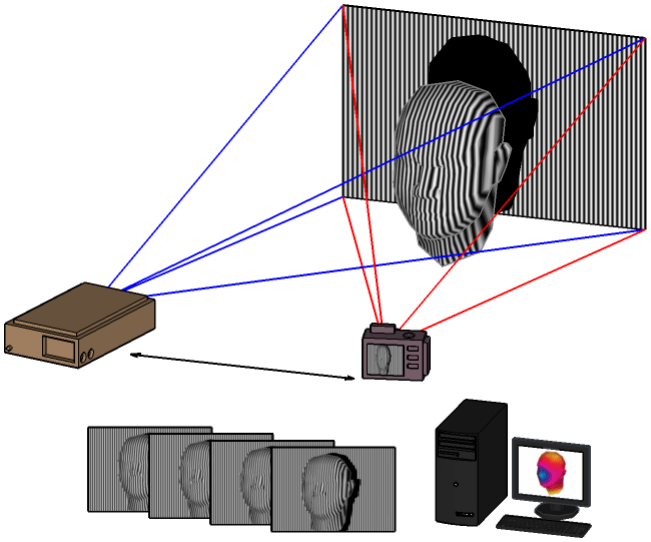


Figure 2.4: Fundamentals of a structured light-based 3D scanning system using a temporal coding scheme. [2]

2.2.3 Time of Flight Sensors

Time of Flight (TOF) sensors solve the distances based on the amount of time light takes to travel from the system to the object and back (Sansoni et al., 2009; Moons et al., 2010). A short light pulse is emitted towards a surface and each pixel counts time until the returning signal is detected (Blais, 2004)), as seen in Figure 2.5.

The measured time is proportional to the distance from the object (Moons et al., 2010). This concept is in the origin of RADAR (sound waves instead of light) or LADAR (laser-based) technology, for example.

A TOF sensor can also use light pulses modulated either by frequency or amplitude. The proposed methods for an amplitude modulation of the optical carrier measure range from the phase variation between the transmitted and received signals. Frequency modulation techniques obtain results with an excellent resolution (around 1 mm).

Frequency modulation methods often use formulations based on sinusoidal signals (Lange and Seitz, 2001; Oggier et al., 2004; Weingarten et al., 2004; Kolb et al., 2008) for the demodulation and range measurement. However, other types of periodic signals can be used. Every pixel on the sensor measures the intensity of light reflected by the object four times, in equal sampling periods (m_0, m_1, m_2 and m_3) as seen in Figure 2.5 (Foix et al., 2011). This TOF measurement by phase demodulations is usually referred to as four-bucket sampling.

[2] http://main.alcestech.com/2012/01/3d-depth-capture-and-sensing.html

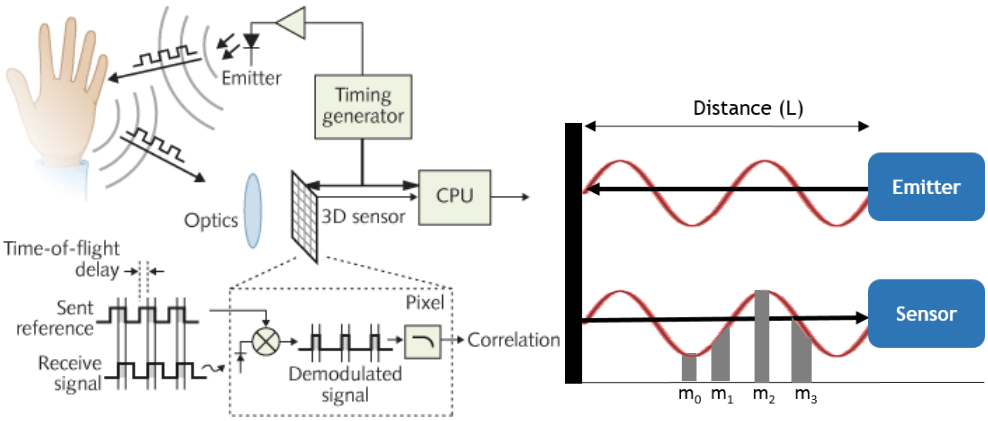


Figure 2.5: (Left) Basic principle of TOF 3D sensors. (Right) Distance (L) measurement using phase offset. Each m_k is a period of acquisition (and depends on the sampling frequency). Adapted from Foix et al. (2011).

The measured phase, ϕ, is given by:

$$\phi = arctan\left(\frac{m_3 - m_1}{m_0 - m_2}\right) \tag{2.4}$$

The target depth D can be easily calculated using the measured phase:

$$D = L\left(\frac{\phi}{2\pi}\right) \tag{2.5}$$

Where L is the ambiguity-free distance range of the sensor, determined by the modulation frequency (f_m) and the speed of light in the vacuum (c).

$$L = \frac{c}{2f_m} \tag{2.6}$$

To achieve better results, the use of multiple modulated frequencies or lowering the modulation frequency, for a higher unambiguous metric range, were proposed by Foix et al. (2011). Although facing great difficulties with shiny surfaces (which reflect little back-scattered light unless the sensor is oriented perpendicularly to the object being scanned (Sansoni et al., 2009)), TOF is, by far, the preferred choice for long range (up to 100 m) or high volume measurements (Blais, 2004; Moons et al., 2010). Sometimes, a TOF sensor is incorporated in other scanning devices for a more realistic scene representation (Sansoni et al., 2009).

2.2.4 Conclusions

The accuracy and quality of the results depend greatly on the correct choice of the image acquisition technique. Depending on the type of object or scene being scanned, a specific method is preferred or provides better results. In Table 2.1 a comparison between the studied techniques is presented.

Table 2.1: Comparison between the major 3D sensing techniques based on their performance based on the referred parameters. Grades from excellent (+++) to major drawback (- - -) Adapted from Blais (2004) and Sansoni et al. (2009).

Parameter	Stereo Vision	Laser Triangulation	Structured Light	TOF
Accuracy	++	+++	+++	++
Usability	- -	++	+	- -
Versatility	++	+	+++	++
Cost	++	- -	- -[3]	+
Range Limit	++	-	-	+++
Acquisition Rate	- -	++	++	++

A sensing system for medical purposes must combine accuracy and simplicity, while being affordable and safe to use, which automatically discards laser triangulation and strengthens the importance of a low-cost approach. Although stereo vision, TOF and structured light low-cost sensors are available, the wide variety of patterns that can be used using structured light techniques and the possibility of a dynamic pattern change confers an important flexibility to the analysis. As the proposed techniques can provide similar accuracies, the versatility of the structured light systems is desirable.

Nowadays, several structured-light-based, low-cost 3D sensors can be found in the market. In the following chapter, those systems are presented after a brief overview on the 3D Body Acquisition Techniques.

[3]The majority of the solutions are expensive, although some low-cost structured light 3D sensors, such as the Microsoft Kinect, have appeared in the past few years.

Chapter 3

3D Body Acquisition Techniques

There is a myriad of available systems for the reconstruction of the human body. The most accurate, however, present great drawbacks concerning their usage for medical purposes: they are expensive, complex and require trained and specialized staff. Such problems open the door for the recently implemented low-cost RGB-D cameras, which can provide the based for the development of practical systems for reconstruction of body parts, especially the breast.

In this chapter, a general overview about depth-map based solutions for body reconstruction is presented, with special focus on the Microsoft Kinect and its applicability on the reconstruction of body parts.

3.1 Depth-Map Based Techniques

RGB-D cameras are sensors that capture RGB (color) images along with per-pixel depth information (Henry et al., 2012). In order to achieve this, either active stereo[1] [2] or TOF[3] can be used. Although being now a widely used resource for computer vision purposes by research groups, the key drive for the success of these cameras was the gaming industry. Some gaming consoles are now equipped with low-cost RGB-D cameras whose potential can have multiple applications, such as scene reconstruction and 3D modeling.

As seen in Figure 3.1 RGB-D cameras estimate the depth information for the pixel on an image, with the exception of the points described as having unknown depth data – white pixels in Figure 3.1. This may happen because they are too far or too close to the sensor – the sensor can only provide depth information up to a certain distance (Konolige, 2010), or due to occlusions or a bad surface angle or composition (Henry et al., 2010). To overcome that, the user can either estimate or define a specific value for those pixels, or simply not consider them on the analysis.

These sensors combine the capture of rich visual information – color stream, which make them useful for loop-closure detection (Snavely et al., 2006; Newman et al., 2009) with depth data

[1] http://www.xbox.com/en-US/kinect (10th December 2013)
[2] http://www.asus.com/Multimedia/Xtion_PRO/ (10th December 2013)
[3] http://www.softkinetic.com/Solutions/DepthSensecameras.aspx (10th December 2013)

Figure 3.1: Depth information estimation from a scene captured with a RGB-D camera. White spots in the image represent unknown distance values (Henry et al., 2012)).

extraction, used for frame-to-frame alignment and dense 3D reconstruction. This combination makes RGB-D cameras well suited to address a 3D modeling challenge, as their versatility allows the user to choose different and complementary approaches to solve the problem on hand. In addition, the currently available RGB-D cameras, such as the Microsoft Kinect, are particularly suited for biomedical applications: they are practical, transportable, transport and install and allow data streaming into a computer while recording.

Some low-cost RGB-D cameras use structured light techniques to create depth-maps of a specific scene being scanned. These cameras have a reduced field of view (Henry et al., 2012) and generate low-resolution and noisy data (Oliveira, 2013). A processing step is needed to improve the resolution of disparity information and to calibrate the sensor, once the color and depth images are not aligned by default. Nevertheless, depth and appearance information can be retrieved from images at a high rate (30 *fps*).

The challenge presented by low-cost RGB-D cameras is the accurate modeling of an object or a scene as good as other pre-established, complex and costly solutions.

3.1.1 Microsoft Kinect

The Microsoft's Kinect was released as non-contact motion sensor in 2010 for gaming applications (Microsoft XBOX 360). Nevertheless, it has been recently applied to robotics and health application with good results (Chang et al., 2011; Huang, 2011; Henry et al., 2012; Oliveira, 2013).

When compared to other sensors, Kinect presents drivers with higher quality, more stability with USB controllers and its position can be controlled remotely. On the other hand, its size (30 cm x 7.6 cm x 6.4 cm) and the need of ACDC power supply are the main drawbacks. In addition, a stable and versatile Software Development Kit (SDK) is freely available. An overview of Kinect's requirements is shown on Table 3.1.

3.1.1.1 Hardware

PrimeSense[4] developed a sensor implemented on most of the RGB-D cameras available, including Kinect. This device has three major optical components, all shown in Figure 3.2: a laser based near infrared (IR) projector, and IR PS1080 CMOS (complementary metal-oxide semiconductor) camera/detector, which is used to receive the reflected light, and a conventional color camera. The incorporated RGB camera with a maximum resolution of 1600 x 1200 pixels to match the depth data. In addition, a four-sensor microphone array (to separate sounds from different directions) and an accelerometer are available. Table 3.1 presents the Kinect requirements overview.

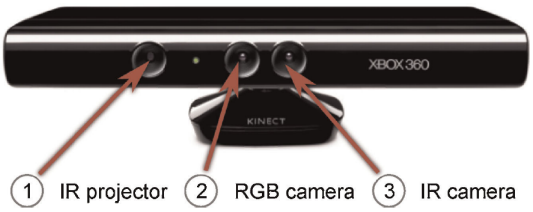


Figure 3.2: The Kinect hardware. It is possible to observe the RGB camera and the depth sensor consisting on an infrared projector and a monochrome CMOS sensor (Zollhöfer et al., 2011)).

Concerning the near IR projector, it emits a speckle pattern towards the object that repeats itself after 211 horizontal spots and 165 vertical spots, creating several blocks, each one with its own bright center point (Andersen et al., 2012; Cruz et al., 2012; Izadi et al., 2012; Oliveira, 2013). As a whole, the projected pattern consists on a 3x3 repetition of the mentioned blocks, resulting on 633 x 495 spots. This allows an easy extraction of features and depth information (Oliveira, 2013).

The depth image from the IR camera has a maximum resolution of 640 x 480 pixels with 11-bits, which provides 2048 levels or sensitivity (Cruz et al., 2012). At 2 m from the sensor, it is able to resolve down to 3 mm for height and 1 cm for depth, and operates for ranges between 0.8 and 3.5 m (Oliveira, 2013).

The quality of the depth-map image can be affected by light conditions. One of the recurrent problem is related with the creation of shadows due to the distance between the illuminator and the IR camera illuminated objects (Andersen et al., 2012). The presence of this shadow does not allow the sensor to estimate the depth and therefore the pixels in that particular area are not assigned to any specific value.

The recently released Kinect for Windows[5] deals with the problem associated to the depth data, once it provides real values for distance measurements. Similar to the Microsoft Kinect, also has some other improved features, such as low-distance tracking (the minimum operating range is now 0.4 m). However, although being licensed for commercial use, its SDK only works in Windows Operating Systems.

[4] http://www.primesense.com/
[5] http://www.microsoft.com/en-us/kinectforwindows/

Table 3.1: Kinect requirements overview

External Power Required	Yes
Angular Field of View	Horizontal 57º | Vertical 43º
Sensorization	RGB and Depth Sensors, Microphone Array, Accelerometers
Resolution	RGB and Depth: 640 x 480 pixels @ 30 Hz
Platform	At least dual-core 2.66-GHz processor, 32/64 bit, 2GB RAM
Interface	Dedicated USB 2.0
Software	Microsoft Kinect SDK , OpenKinect, OpenNI and OpenCV
Programming Language	C#, C++, Visual Basic, Java, Python, ActionScript
Dimensions	30 cm x 7.6 cm x 6.4 cm
Notes	Motor tilt from -27º to 27º

3.1.1.2 Software

The Kinect users can collect and process the recorded images from the sensor using one of the free available libraries and drivers. Six days after the official release of Kinect, the OpenKinect[6] community was born, when Hector Martin released his own open-source driver for the device. The evolution of the community never stopped and a driver called *libfreenect* under Apache 2.0 or GPLv2 license was released, allowing users to connect the device with a personal computer. In a joint effort, PrimeSense and other companies launched a non-profit organization called OpenNI. The goal was to achieve an industry standard framework for the interoperability of natural interaction devices. For that, the OpenNI driver and the NITE (Natural Interaction Technology for End-user) Middleware Software were supplied with tools for scene perception and analysis. These libraries can be applied for applications written in different programming languages (see Table 3.1). Both *libfreenect* and OpenNI projects work on Windows, Linux (Ubuntu) and Mac OS X.

3.1.2 ASUS XTION

The ASUS Xtion Pro[7] is based on the same sensor that inspired Microsoft Kinect, developed by PrimeSense (Böhm, 2012)). Because of that, both devices are quite similar in their functionalities. However, although ASUS Xtion can capture depth data the same way Kinect does, it does not record color information. Depth-maps are generated with 60 frame per second acquisition rate, with a nominal range from 0.8 to 2.5 m and 1 280 x 1 024 pixels. A 30 frame per second acquisition rate is possible, with a resolution similar to the one obtained using Kinect. The ASUS Xtion was described as being a good tool for body shape reconstruction and movement tracking (Böhm, 2012). However, the camera needs to be horizontally aligned to the scene being scanned, because the technology embedded within the sensor is based on the predictability of the system setup. Figure 3.3 shows the ASUS Xtion Pro camera and Table 3.2 presents the ASUS Xtion requirements.

[6] http://openkinect.org/wiki/Main_Page
[7] http://www.asus.com/Multimedia/Xtion_PRO/

Figure 3.3: The ASUS XTION Pro camera consists in a dot pattern projector using a laser diode operating in the IR domain, and an IR sensor (Böhm, 2012). Range is estimated using the information retrieved from the sensor.

Table 3.2: ASUS Xtion Pro requirements overview

External Power Required	Yes
Angular Field of View	Horizontal 58º | Vertical 45º | Diagonal 70º
Sensorization	Depth Sensor, 2 Microphones
Resolution	SXGA: 1280 x 1024 pixels @ 60 fps
Frame Rate	VGA: 30 fps | QVGA: 60 fps
Platform	WIN 32/64: XP, Vista, 7, 8 | Ubuntu 10.10+ | Android
Interface	Intel X86 & USB 2.0
Software	OpenNI SDK bundled
Programming Language	C++, C#, JAVA
Dimensions	18 cm x 3.5 cm x 5 cm

3.1.3 SoftKinetic DepthSense Cameras

SoftKinetic[8] is a reference brand in natural gesture recognition, which develop TOF sensors for the creation of real time 3D RGB, grayscale confidence and depth-map images, at a 60 Hz rate. SoftKinetic developed two different devices: DS311 and DS325 (shown in Figure 3.4), applied, respectively, for long-range and close-range detections. DS325 is the most accurate depth sensor in the market, to be used in close analysis (such as finger and hand tracking and detection). Technical specifications are presented in Table 3.3.

Figure 3.4: The DepthSense DS325 and DS311 cameras from SoftKinetic.

3.1.4 Final Considerations

Low-cost sensors have been gaining importance in the medical field, especially the Microsoft Kinect. This device presents itself as a reliable tool for the proposed work. The Kinect is capable of operating under any interior lighting conditions and is compact and light. Resolutions of

[8] http://www.softkinetic.com/

Table 3.3: DepthSense cameras requirements overview

	DS311	**DS325**
External Power Required	Yes	Yes
RGB Field of View	Horizontal 57.3º Vertical 42º Diagonal 73.8º	Horizontal 63.2º Vertical 49.3º Diagonal 75.2º
Sensorization	3 axis accelerometer and 2 microphones	2 microphones
Resolution	RGB: VGA 640x480 pixels Depth: QQVGA 160x120 pixels	RGB: HD 720p Depth: QVGA 320x240 pixels
Frame Rate	QQVGA: 25-60 fps	QVGA: 25-60 fps
Operating Range	0.15-4.5 m	0.15-1.0 m
Depth Noise	< 3cm at 3m	< 1.4 cm at 1 m
Connectivity	USB 2.0	USB 2.0
Dimensions	10.5 cm x 3 cm x 2.3 cm	24 cm x 4 cm x 5 cm

1.3 mm/pixel have been described using this device (Oliveira, 2013). This is a value within the physiological range of size changes of body parts, according to specialists and rather good for object modeling. The availability of free software for image processing, easiness of use, its versatility and the possibility of use both color and depth data makes the Kinect an appropriate tool for 3D modeling purposes. In fact, this sensor has been used in the past few years to address 3D body modeling problems, as to be presented in the following section, with good results.

In this thesis, a Microsoft Kinect-based system for 3D Modeling purposes will be developed. This RGB-D camera will be used to capture depth and color data from the object whose 3D model will be build. In addition, the Kinect Fusion algorithm available in the Microsoft Kinect SDK will be used as the state of the art solution for 3D Modeling using this device.

3.2 The Kinect Sensor as a Tool for 3D Modeling

The use of the Kinect sensor for 3D modeling purposes has been studied since its release. The Microsoft research group Kinect Fusion presented, in 2011, the state of the art homonym algorithm that allows users to scan a scene using hand-held movements using the Kinect (Newcombe et al., 2011; Izadi et al., 2012). The main contributions of this project were dense surface mapping and camera tracking algorithms. Briefly, during the scan and for every viewpoint available, the algorithm fuses the new depth information to the 3D model being build, creating high-resolution outputs, as seen in Figure 3.5.

The first stage is depth-map conversion, which takes the raw depth from Kinect and converts it into floating point depth in meters. Then, an optional conversion to an oriented point cloud of 3D points/vertices in the camera coordinate system, and the surface normals (orientation of the surface) at these points are computed. The global/world camera pose (its location and orientation)

is estimated in relation with the initial pose and the final stage consists in the fusion the depth data into the volumetric representation of the scene.

Figure 3.5: Performance of the Kinect Fusion algorithm. Normal maps (color) and the rendered 3D model (greyscale) are shown. The output is smooth and complete, in contrast with the incomplete and noisy input data from the Kinect (Newcombe et al., 2011).

Furthermore, the Kinect Fusion is robust to the presence of moving targets on a scene. This means that any dynamic structure that moves inside the algorithm's field of view is removed from the final model. In addition, it means that we cannot reconstruct a scene by simply walking in while holding a Kinect.

A method for body scanning by aligning depth and color images using a single Kinect was proposed by Cui and Stricker (2013). They achieve a 360° scan by moving the sensor around the object and, then, data is processed using a 3D Super Resolution (SR) algorithm followed by a loop-closing method, specific for Kinect data. Finally, a non-rigid registration stage is performed to correct residual errors of movements, resulting in a complete, smooth and detailed 3D model, as seen in Figure 3.6.

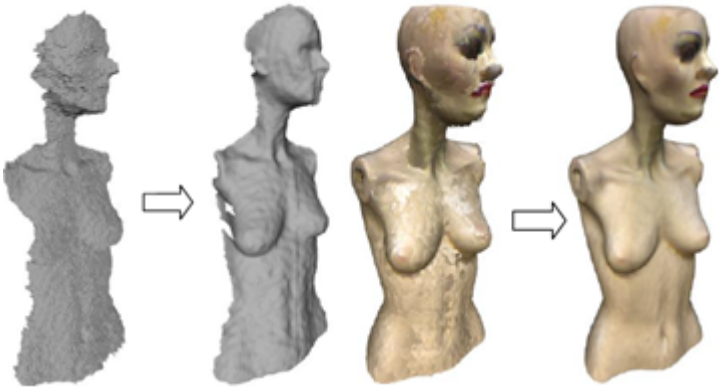


Figure 3.6: The results obtained by Cui and Stricker (2013). From left to right, the raw data point cloud, the result after the SR algorithm implementation, the results after adding the color information and the final model are presented.

Zollhöfer et al. (2011) proposed a method for high-quality personalized avatar creation for home use, using a combination of depth and color images. The proposed algorithm combines robust non-rigid registration and fitting of a morphable face model to generate a high-quality reconstruction of the facial geometry and texture, as seen in Figure 3.7.

Figure 3.7: High-quality personalized avatar creation using the Kinect, by Zollhöfer et al. (2011).

Weiss et al. (2012) described a new method for human shape reconstruction, combining image silhouettes with coarse range data to estimate a parametric model of a human body. The 3D shape estimation is obtained combining multiple views of a subject moving in front of the sensor. To achieve consistent results while allowing pose to vary, the authors used a SCAPE body model (Anguelov et al., 2005) to fit the processed data (Figure 3.8).

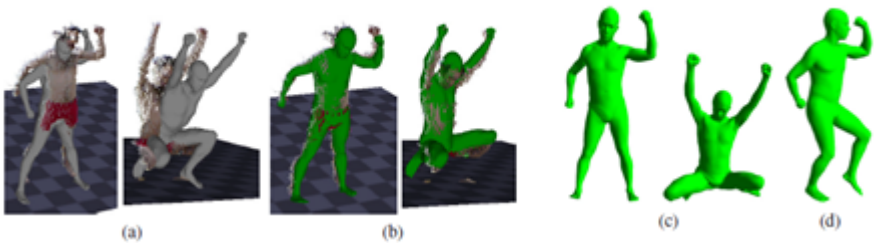


Figure 3.8: Varying poses from Weiss et al. (2012) solution. In (a), the point cloud initialization is obtained, (b) represents the fitting of the raw data with the SCAPE body models, represented in (c) and (d).

For scene scanning purposes in robotics, Henry et al. (2012) proposed a method for 3D indoor dense map creation. The authors used a Kinect-style RGB-D camera, employed a novel joint optimization algorithm and combined visual features with shape-based alignment stages. Visual and depth information are also combined for a view-based loop-closure detection, followed by pose optimization to achieve globally consistent maps.

Recently, a system for 3D full body shape scanning using multiple Kinects was proposed by Tong et al. (2012), one Kinect for each part of the body. That allowed each camera to be closer to the body, increasing the quality of the recorded data. The recorded frames were non-rigidly aligned in a template created from the first frame, dealing with success with loop closure with detail. The obtained results are shown in Figure 3.9. The authors faced some problems regarding complex occlusions that created some misaligned structures.

Wang et al. (2012) proposed a body modeling system with a single fixed Kinect. The authors estimated an initial model using a registration method for four key poses (front, back and two profiles) out of the entire depth video sequence. Then, an articulated part-based cylindrical body

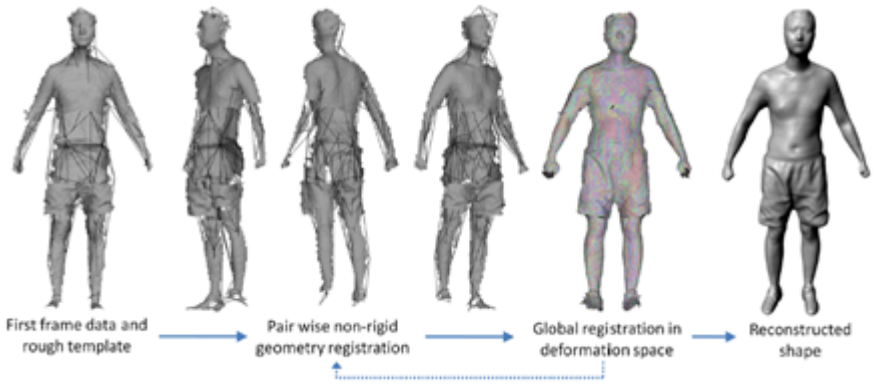


Figure 3.9: The main framework of the reconstruction algorithm described by Tong et al. (2012).

model is used to process the rough and noisy previous registration to obtain a more accurate 3D model, as seen in Figure 3.10.

Figure 3.10: On the left, the approximation of the body by cylindrical structures is shown. On the left, a 3D model of a subject is shown in four different poses (Wang et al., 2012)

Sturm et al. (2013) described Kfusion_ROS, a novel approach to create 3D miniatures of persons using a Kinect sensor and a 3D color printer, based on the Kinect Fusion algorithm. Both color and depth data is acquired while the subject is rotating on a chair and the model is represented using a signed distance function. The approach proposes a novel weighting function to perform an automatic filling of small holes created by small-occlusions and a method for turn detection. The authors described a significant upgrade on accuracy when compared to the Kinect Fusion algorithm. Only 300 frames are needed to create a model that can be manufactured using a conventional 3D printer, as shown in Figure 3.11.

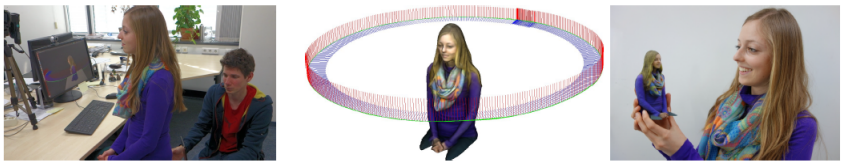


Figure 3.11: The acquisition process from Sturm et al. (2013). The method runs in real-time and displays a live view of the high accurate reconstruction model on the scene. The model can be printed in 3D.

Li et al. (2013) also developed a pipeline that allows ordinary users to capture complete and fully textured 3D models of themselves using a single Kinect sensor in non-controlled conditions, and without the need for a second operator. The user is asked to rotate on the same pose for eight

different views (around 45 degrees) to cover the full body. The system merges all the scanned frames using a rigid registration method, which is then refined using a multi-view non-rigid registration procedure. This is done by minimizing an energy, thus avoiding the deformation of the model. To ensure consistent texturing, the authors use the Poisson blending method. The model can then be 3D printed to obtain a physical representation of the scan. The whole pipeline is shown in Figure 3.12.

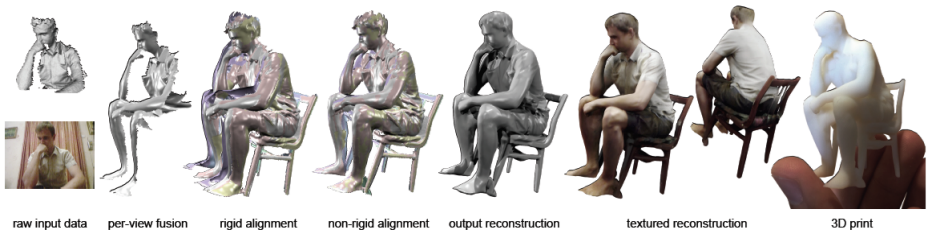


Figure 3.12: The 3D modeling pipeline of Li et al. (2013). The algorithm takes as input 150 frames of raw depth-maps and textures for each of the eight captured views. These frames are fused and segmented to yield per-view fused surfaces that are, first, registered using a rigid alignment, which is then refined using a non-rigid registration in a global optimization. The final model can be 3D printed

3.2.1 Conclusions

Since 2005, several research groups have studied the feasibility of assessing the 3D breast shape using eith active or passive lighting techniques. On the other hand, some work concerning 3D body modeling using depth-maps using the Kinect has already been developed in the past few years. The Kinect Fusion is the state of the art methodology for 3D reconstruction. It is optimized for static scene analysis, having some associated problems concerning the 3D reconstruction of objects: dynamic objects are removed from the final model, works poorly on the point cloud alignment stage for symmetric objects and the depth estimation fails in some situations, thus leading to the presence of background information in the objects of interest. In addition, its parametrization requires expertise by the user, which is not always guaranteed in a medical environment. The commonly used methodologies for breast 3D modeling are complex and/or costly, requiring very controlled acquisition sets and are not easy to implement in a daily basis on a medical environment.

The technique overview presented in this chapter suggests that a new, affordable and practical system using the Kinect is required. Moreover, it must be able to robustly acquire keypoints from the female torso and work under feasible time, while maintaining the quality of the reconstruction.

Chapter 4

Point Cloud Registration

For 3D modeling purposes, a single 3D view is often incomplete and noisy, which is not accurate enough for specific medical application. Thus, different approaches have been proposed in the past few years to achieve such requirements. One of these approaches is called point cloud registration. Point cloud registration methods take a set of range images in different positions and the depth information is used to create point clouds that are fused towards a single model. Biomedical applications of 3D modeling often deal with non-rigid registration problems, like the movement of the object of interest and the lack of previous knowledge about the pose of the views (Salvi et al., 2007). Conversely, non-moving objects lead to rigid registration problems.

The point cloud registration stage can be done in several ways, as depicted in Figure 4.1:

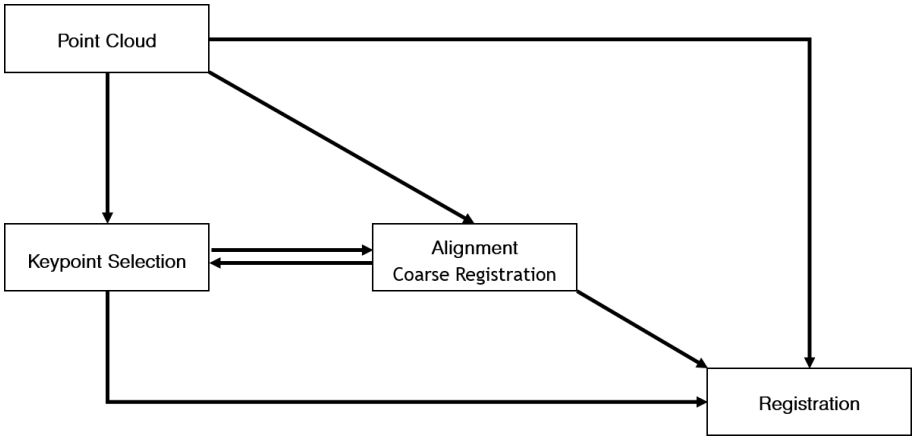


Figure 4.1: The different possible approaches of point cloud registration.

1. By merging the views directly, using registration algorithms without any pre-processing.
2. By detecting robust keypoints which provide a representative description of the point clouds and aligning them. The transformation that matches the views is assumed to be the same that aligns the keypoints.
3. By first aligning the views and then performing the registration. The alignment stage, or coarse registration, aims at computing an initial estimation of the rigid motion between two

point clouds, either by detecting keypoints or not. In such approach, as the registration stage searches for the most accurate solution possible, it is named fine registration(Salvi et al., 2007). This is the standard registration approach.

At any stage of the registration pipeline, inlier selection methodologies can be applied. They look for the most representative subset of points, eliminating those wich can compromise the result of the registration. In this chapter, the standard inlier selection methodology is studied, followed by an overview about the different stages on the registration pipeline. Regarding the inlier and keypoint selection stages, some preliminary results and conclusions are presented.

4.1 Inlier Selection: Random Sample Consensus

RANSAC (*Random Sample Consensus*) (Fischler and Bolles, 1981) is the most commonly used inlier selection method used for 3D reconstruction purposes (Salvi et al., 2007; Yang et al., 2010; Henry et al., 2010, 2012; Castellani and Bartoli, 2012). The RANSAC is a parameter estimation methodology that tries to maximize the set of inliers under a given model and error measure. It has four major steps:

1. Random selection of a subset;
2. Build of the model based on that subset;
3. Determination of the inliers and outliers, considering the model;
4. Repetition of the algorithm a certain number of times or until the set of inliers is big enough;

This approach tries to avoid the presence of outliers in static or rigid scenes and allows the suggestion of a representative set of points for a fast and accurate convergence of registration algorithms (Rusch et al., 2006). However, for non-rigid problems (common in the medical field), with non-static objects of interest, the performance is not good.

The RANSAC inlier selection was tested using different point clouds. It was observed that for views with different information or with irregular shapes, it is very difficult to correctly define a sub-surface of inliers that correctly describes the model while being present in consecutive views. This makes the registration more prone to failure. On the other hand, the algorithm is not appropriate when time is critical because a high computational cost is required for a good performance (Salvi et al., 2007).

4.2 Keypoint Selection

The registration techniques demand a previous detection of correspondent features between both point clouds and use them to estimate the motion between the views. Features provide unique or highly-descriptive 3D information and can be either a point, a curve or a sub-surface. However,

by choosing feature points (or keypoints) it is possible to establish point-to-point correspondences between clouds and directly estimate the motion between them.

In this section, the tested keypoint selection algorithms are presented. Besides from those methods, there are some other approaches that have been proposed in the past few years: the **4-Point Congruent Sets** (4PCS) algorithm (Aiger et al., 2008), the **Fast Persistent Feature Histograms** (Rusu et al., 2009) and the **CIRCON descriptors** (Torre-Ferrero et al., 2012).

4.2.1 Spin Images

The Spin Images were proposed by Johnson (1997) as a new method to represent 3D point clouds. Initially proposed for image recognition, spin images provide an invariant local shape description of the point clouds. The main advantage of spin images is that they are object centered. This means that surfaces can be compared without being aligned, and the point correspondences are direct. The image generation process can be visualized as a sheet spinning about the normal of a point p belonging to a 3D mesh, which needs to be created prior to the coarse registration phase. The process is exemplified in Figure 4.2. A spin-image is a 2D table that relates the values of two invariant descriptors, α and β, for each point on the neighborhood of a specific point p, as seen in Equations 4.1 and 4.2: (1) α is the distance between each point to the normal vector defined by the tangent plane P; (2) β is the distance between the point p to the tangent plane.

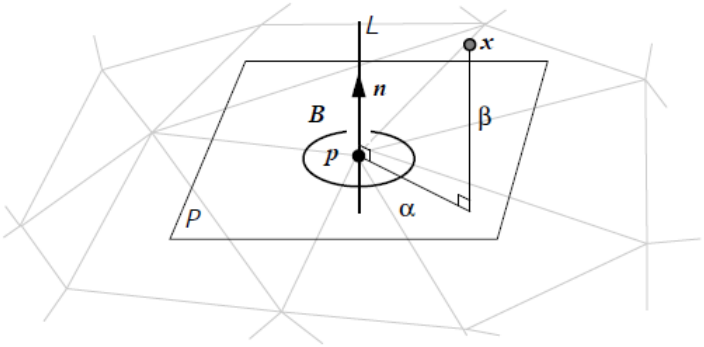


Figure 4.2: Schematic showing how to calculate the spin image invariant descriptors, α and β, respectively the radial distance to the surface normal line L and the axial distance above the tangent plane P. (Johnson, 1997)

$$\alpha = \sqrt{||x-p||^2 - (n(x-p))^2} \tag{4.1}$$

$$\beta = n(x-p) \tag{4.2}$$

where p is the given point, n is the normal vector at this point and x a point of the set of neighbors used to generate the spin image. As said previously, these distances are used to generate a 2D table representing α on the x-axis and β on the y-axis.

Some spin-images are computed in the first view and then, for each one, the potential correspondences are search in the second view (Salvi et al., 2007)). Then, a rigid transformation is finally computed using the candidates.

4.2.2 Curvedness

Point clouds are also 3D surfaces, meaning that they share some properties, including curvature. High-curvature points are good features because they often correspond to relevant and distinctive regions of the surface, and can be used for a coarse alignment of point clouds. Some **Curvature-Based Methods** have been proposed in the past few years, based on the maximum and minimum curvatures, respectively k_1 and k_2. These curvatures can be calculated as follows:

$$k_1 = H + \sqrt{H^2 - K} \tag{4.3}$$

$$k_2 = H - \sqrt{H^2 - K} \tag{4.4}$$

where H and K stand, respectively, for the Mean and Gaussian curvatures for each point. These values measure the maximum and minimum bending of a surface in a specific point, as seen in Figure 4.3. The curvature is positive if the surface turns in the same direction as the normal vector (Ho, 2009).

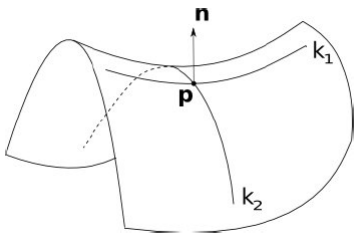
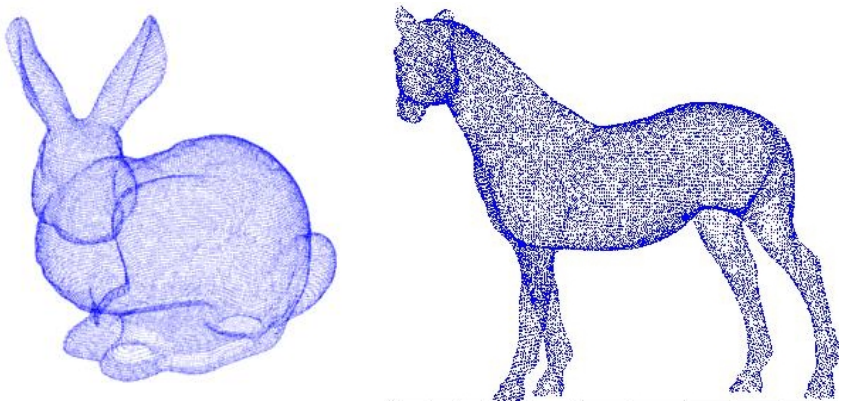


Figure 4.3: Principal curvatures at a point P on a surface (Ho, 2009).

The curvature values can also be used to indicate the bending energy of the point cloud at each point: its curvedness (see Equation 4.5).

$$C_p = \sqrt{\frac{k_1^2 + k_2^2}{2}} \tag{4.5}$$

In other words, the curvedness measures how highly or gently curved a surface is (Ho, 2009). High energy points often belong to heterogeneous regions of a point cloud, being good candidates to be feature points. A candidate is considered as a feature when it is a top curvedness point for three different neighborhood sizes. As for spin-images, the results are improved when higher resolutions are used —lower neighborhood scales— but increasing the computational cost in such a way that it is not appropriate to be used under the scope of this problem.



(a) The Stanford Bunny 3D Model (b) The Stanford Horse 3D Model

Figure 4.4: The Bunny and Horse 3D Models from the Stanford Digital Michelangelo Project.

4.2.3 Preliminary Tests

The previously studied coarse registration methodologies were tested in the scope of this thesis. These tests were conducted using two reference 3D models from the Stanford Digital Michelangelo Project[1], and consisted simply in the registration of two consecutive views. Figure 4.4 shows the whole aspect of both models. Regarding the ICP, its results and performance are shown in the following chapters.

The **Spin Image** algorithm was tested and, although good results can be found, they strongly rely on the resolution of the method (Johnson, 1997). This means that, in order to have a good performance, the preliminary stage in which a triangular mesh is created needs to have a great number of elements, thus making the response time of the method not appropriate to be used under feasible time.

Concerning the usage of **Curvedness** for the feature detection, the work of Feldmar and Ayache (1996) was implemented, with the results shown in Figure 4.5.

First, the Gaussian and Mean curvatures for each point were calculated using the Moving Least Squares Algorithm, as implemented by Yang and Qian (2007). Then, the maximum and minimum k_1 and k_2 curvatures were used to compute the curvedness at each point. A candidate was considered as a feature when it is a top curvedness point for three different neighborhood scales: 90, 120 and 150 neighbors. The results show that, although the represented views are consecutive and high-overlapping, the candidate point distribution is not similar, thus leading to wrong correspondences.

These preliminary results suggest that a more robust keypoint selection methodology is needed. It should be based on the morphology of the point cloud to determine coherently distributed feature points, to allow the establishment of wrong correspondences.

[1] https://graphics.stanford.edu/data/3Dscanrep/

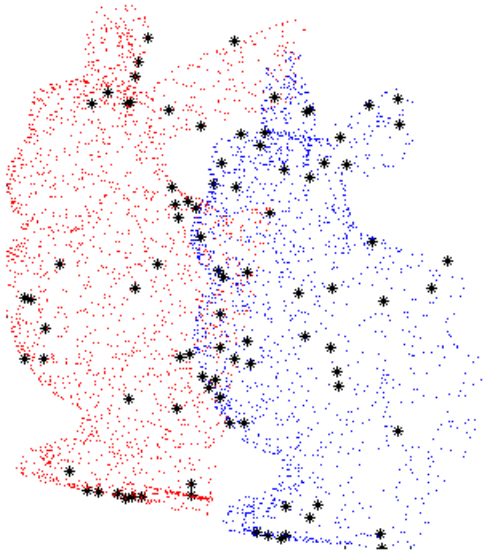


Figure 4.5: Feature point selection for the Stanford Bunny using the curvedness descriptor. It is possible to observe that, although similar views are used, the feature point selection is not optimal.

4.3 Alignment

The main goal of the alignment (also designated as coarse registration) is to search for an initial estimation of the transformation between the 3D point clouds being registered. They perform a rough alignment of clouds, guaranteeing a better convergence of the registration algorithms towards the optimal solution. This stage is especially important when high levels of noise are found. In other words, such process is desirable when two consecutive 3D views are highly rotated and/or translated, in order to reduce the probability of a bad reconstruction and computational cost of the task (Rusch et al., 2006; Salvi et al., 2007).

4.3.1 Centroid Alignment

The simplest way of aligning two point clouds is by overlapping their centroids (or centers of mass). Such approach is not able to compensate the rotation between views, but provides a fair minimization of their translational misalignment. When dealing with very different or low-overlapping point clouds, such approach is not expected to provide good initial alignments, since the centers of mass will still be highly misaligned.

4.3.2 Keypoint Correspondence Establishment

As the name suggest, such alignment strategy looks for point-to-point correspondences between the keypoints (which can be determined as previously presented) in each views. Then, the corresponding point are used to estimate a rigid transformation that can be extrapolated to align the views. The correspondence establishment can be done by:

- performing a nearest neighbor search.
- computing descriptors of a point cloud at each point and looking for similar descriptors on the other point cloud.

The main drawback associated to this approach is that, if some correspondences are missed, the alignment does not provide an initial pose good enough to help the registration to converge.

4.3.3 Principal Component Analysis (PCA)

Proposed by Chung et al. (1998), this method used the orientation of the three main axis of a point cloud to align a sequence of views. Briefly, the coarse alignment is the one that best aligns both main axes using a rigid transformation.

Salvi et al. (2007) presented a simple and fast framework for the coarse alignment of point clouds. The first stage of the method involves calculating the covariance matrix of each view:

$$\text{Cov} = \frac{1}{N}\sum_{i=0}^{N-1}(p_i - \bar{p})(p_i - \bar{p})^T \tag{4.6}$$

where N is the number of points, $\bar{p}$ is the center of mass of the point clouds and p_i is the ith point of the view. By single value decomposition is, then, easy to calculate U_i, the orientation of the main axes:

$$\text{Cov}_i = U_i D_i {U_i}^T \tag{4.7}$$

Then, the rotation and translation can be calculated, respectively, by the product of the eigenvector matrices (Equation 4.8) and the translation is simply the distance between the centers of mass, μ_1 and μ_2, of both point clouds, with respect to the same axis (Equation 4.9):

$$R = U_1 {U_2}^{-1} \tag{4.8}$$

$$t = \bar{\mu_2} - R\bar{\mu_1} \tag{4.9}$$

Best results are found for high overlapping point clouds and when a sufficient number of points is used. Because of its low computational cost and capacity of dealing well with rotational misalignments between high-overlapping point clouds, PCA is suitable to be used in biomedical applications which aim a reconstruction of body parts in feasible time, being taken as the standard methodology in the scope of this thesis.

4.4 Registration

The registration stage looks for the best alignment between the point clouds being merged. In order to solve this problem, a distance function is minimized (Besl and McKay, 1992). Either

rigid and non-rigid registration approaches are possible. Rigid registration consists on applying affine transformations to the views being merged, such as rotation, translation and scaling. Considering a rigid registration problem where scaling is not relevant, the computed homogeneous transformation matrix, T, describes motion that best matches the point sets as follows (Salvi et al., 2007):

$$T = \begin{bmatrix} R & \vec{t} \\ 0^3 & 1 \end{bmatrix} \tag{4.10}$$

where $\vec{t}$ is the translation between the keypoints and $R = R_x(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma)$ is a 3x3 matrix that describes the rotational motion around the three main axis:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) \\ 0 & sin(\alpha) & cos(\alpha) \end{bmatrix} \cdot \begin{bmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{bmatrix} \cdot \begin{bmatrix} cos(\gamma) & -sin(\gamma) & 0 \\ sin(\gamma) & cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.11}$$

On the other hand, non-rigid registration approaches look for transformations that deform the point clouds being registered towards the optimal alignment. However, in the scope of this thesis, such type of transformation will not be considered.

4.4.1 Iterative Closest Point

Proposed in 1992, the ICP algorithm allows the registration of surfaces to each other (Besl and McKay, 1992; Zhang, 1994). The main goal is to find iteratively the rigid transformation that best aligns one point cloud to another, by minimizing the mean squared distance between the aligned views. This guarantees a monotonic convergence to a minimum (Rusch et al., 2006). An overview on the ICP method is shown in Figure 4.6.

Consider the point clouds of the reference and the source viewpoints, $P = \{p_1, \ldots, p_n\}$ and $Q = \{q_1, \ldots, q_n\}$, respectively. The error function is chosen as:

$$E_{ICP}(a, P, Q) = \frac{1}{n} \sum_{i=1}^{n} \| p_1 - Rq_1 - t \|^2 \tag{4.12}$$

Where R is the rotation and t the translation that describe the transformation a between both point clouds, P and Q. The algorithm can fall in local minima, so it is important to provide a good initialization. In other words, the two views must be close to each other (Castellani and Bartoli, 2012). On the other hand, it is required that the two views almost fully overlap or that the source point cloud must be a subset of the target point cloud.

The search for the closets points can be fastened by using *k-d trees* (Castellani and Bartoli, 2012). A k-d tree is a binary tree in which each node corresponds to one point of the cloud. This partition of the 3D space allows the algorithm to search faster for the closest points (Ho, 2009). Some other modifications can be done to improve the performance of the ICP, such as using a

weighted feature distance (Sharp et al., 2002), using a variable number of points at each iteration (Jost and Hugli, 2003) or using different minimization functions.

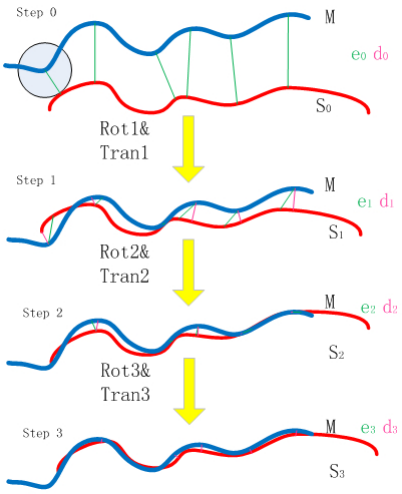


Figure 4.6: Schematic demonstration of ICP algorithm. At each iteration, the algorithm estimates the transformation between the point sets via mean square error minimization.

Because of its feasible response time and the flexibility for user-defined modifications (Salvi et al., 2007)), the ICP is the standard methodology for the fine registration of point clouds.

4.4.2 Chen and Medioni Method

In the same year, Chen and Medioni (1992) proposed an alternative version of the ICP, which minimizes the distance between points and planes. When registering two views, at each iteration, the algorithm minimizes the distance between points on the first cloud with respect to tangent planes in the second(Salvi et al., 2007). Such tangent plane is computed on the intersection between the normal vector on the considered point and the second surface.

Although requiring less iterations than the standard ICP approach, the point-to-plane distance computation is a computationally demanding and more difficult task to perform when compared to point-to-point distance. Moreover, Salvi et al. (2007) state that this method lacks sensibility when non-overlapping clouds are being registered.

4.4.3 Matching Signed Distance Fields

Masuda (2001) presented a registration algorithm based on matching signed distance fields. First, the initial estimation of the values for the rotation and translation are computed. Then, a set of artificial keypoints is generated on a 3D grid. Finally, new motion parameters are computed based on found correspondences between each point on the cloud and each artificial keypoint. This is done iteratively and once the solution converges, it is possible to compute signed distance fields: each point and corresponding normal vector (Salvi et al., 2007). The method allows to register all

views at the same time, avoiding the error propagation effect associated to registering each view at a time. However, it presents a non-feasible response time in terms of biomedical applications.

4.4.4 Genetic Algorithm

More recently, Chow et al. (2004) used genetic algorithms to align point clouds. The core of the solution is based on finding the three components of the translation vector and the three rotation angles. As any genetic algorithm, a fitness function must be defined. The authors chose to minimize the median of the registration error, as follows:

$$F(T) = Median(E_i) \tag{4.13}$$

$$E_i = min|T \cdot p - q| \tag{4.14}$$

being T the transformation matrix, p and q the points on both views and $F(T)$ the fitness function.

Commonly, genetic algorithms provide good results and avoid local minima (Salvi et al., 2007), however computationally demanding. Other relevant problem is associated with the computation of the registration error. The aforementioned error, E_i calculates the difference between the clouds being registered, instead of relying on corresponding points. This means that two successfully registered views, if different enough, can have a high and wrong registration error. When non-overlapping point clouds are being merged, this can compromise the convergence towards a good alignment.

4.5 Conclusions

Among all the presented point cloud registration possibilities, a pipeline based on keypoint extraction for a coarse registration, followed by a fine registration stage seems to be the most robust solution in a biomedical context. This has to do with the possibility of dealing with non-overlapping point clouds with high misalignments between them.

The standard coarse registration methodologies here presented are not suitable for an implementation in a biomedical context, since they do not meet the requirements for an affordable yet precise reconstruction system: they present high computational cost, require user-defined parameters and experienced staff and are not able to estimate repeatable and robust keypoints. As the fine registration methodologies often converge well to a good solution when the views being merged are well approximated, coarse registration plays an important role on the reconstruction pipeline. These characteristics motivate the development of a coarse registration methodology which bases the keypoint selection and the point-to-point correspondences on the morphology of the point cloud. As the work is focused on the reconstruction of the breast, the determination of distinctive body structures such as the breast contour and the nipple is essential. Additionally, the introduction of color information could improve the keypoint correspondence stage.

On the other hand, the ICP proves itself to be a reliable registration approach, because of its feasible response time and higher capability on dealing with low-overlapping point clouds. Over the past few years, several versions of this algorithm have been proposed with good results.

Chapter 5

A Kinect Based System for 3D Modeling Purposes

One of the main objectives of this thesis is the development of an affordable and simple point cloud registration system. Figure 5.1 describes the developed Kinect Based System for 3D Modeling. A set of range and color images is captured using a Microsoft Kinect and point clouds are generated, based on the camera calibration, depth and color information. Then, the reconstructed model is build by registering each view iteratively.

Instead of using a *tree-like* approach, which consists in the pairwise registration of views until the whole model is reconstructed, in this thesis the final model is iteratively build. Although this approach can cause the propagation of errors, it mimics well a real-time 3D modeling system, in which each new view contributes for the refinement of the final reconstructed model.

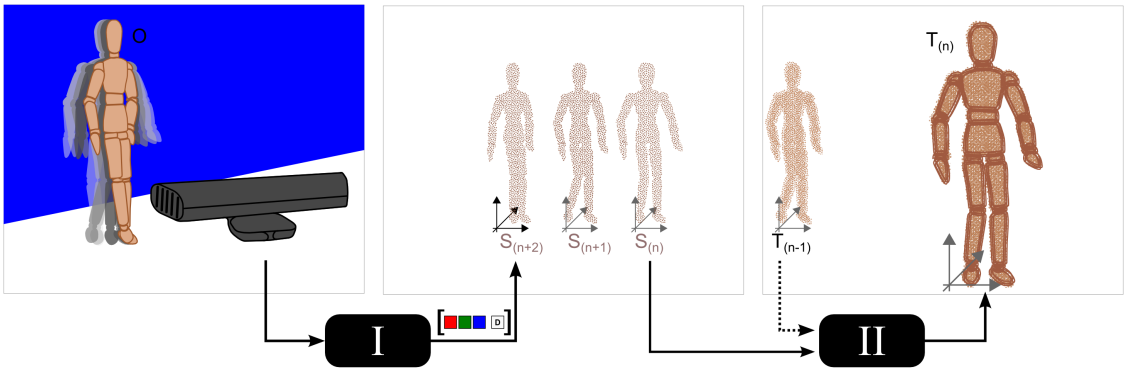


Figure 5.1: The general pipeline implemented in this work. Multiple views of the scene (O) are acquired using a Microsoft Kinect. The depth and color data are used to create source point clouds, S, of the model (**I**). At each iteration, the incoming source point clouds, S_n and the global model being built, T_{n-1} are registered (**II**) to create the new model T_n.

The general pipeline of the proposed system is presented in the flowchart of Figure 5.2. Three main processing stages are presented. The first one uses the RGB-D data acquired using the

Microsoft Kinect to create point clouds of the object of interest. The following two describe the point cloud registration stage.

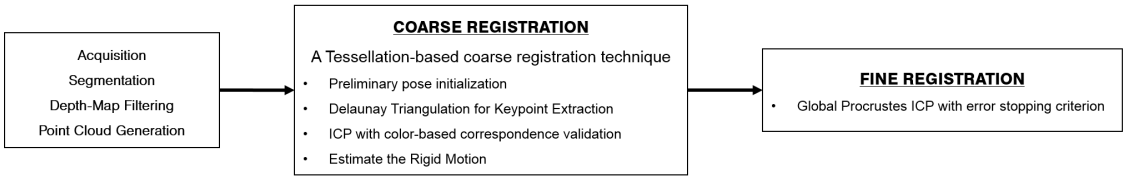


Figure 5.2: Flowchart describing the three major processing stages of the developed system.

The registration phase is a two-stage approach. The coarse registration is done using a proposed Tessellation-based method that uses color and depth data. Then, the Global Procrustes ICP (Toldo et al., 2010) is used for the fine registration stage. The minimization of the cost function is performed by the Generalized Procrustes Analysis technique (Gower, 1975)), in which all views are considered. The fine registration algorithm was slightly modified to have an error-based stopping criterion rather than the number of iterations. The Global Procrustes ICP finishes the fine registration stage when:

1. The point clouds remain almost unchanged after an iteration (the mean square error between consecutive poses of the point clouds is below 1×10^{-5} m);

Or, if the first condition is not met:

2. A maximum number of iterations is reached

5.1 Calibration, Acquisition and Point Cloud Generation

In this section the used acquisition protocol and the point cloud generation stage using raw RGB-D data are described. This last stage was implemented using the version 1.8 of Microsoft Kinect Software Developers Kit (SDK) and OpenCV library (version 2.4.8) and comprehends two preprocessing steps: segmentation and a depth-map filtering stage. For the tasks of mapping between different coordinate spaces it was used the device specific manufacturing calibration data provided by the Kinect SDK.[1].

The acquisition of raw RGB-D data from the Kinect follows the protocol established within the context of the PICTURE Project[2], described in Appendix A. The Kinect is placed on a tripod at 90 cm from the subject being scanned. The images are taken consecutively at 15 frames per second while the subject performs a 180 rotation between lateral views, with hands on the hips, in front of an homogeneous blue background panel.

[1] This part of the work was developed by the Visual Computing and Machine Intelligence (VCMI) group of INESC-TEC, formerly INESC-Porto.

[2] http://vph-picture.eu/

Arises from the acquisition protocol that the scene observed by the Kinect comprises elements with different apparent motions: the patient (closer to Kinect) rotating around itself and the surrounding environment. The raw depth images were segmented to retain only information respective to the patient. Given that simple global threshold methods have been proved to fail(Oliveira et al., 2014), a discontinuity based approach was taken. Firstly, the gradient of the depth image was obtained using Gabor filters. Then, the Otsu's segmentation method was applied over that result, followed by a connected components computation stage. Subsequently, the closest region with respect to the camera was labeled as foreground.

Kinect depth maps present different sorts of noise (Mallick et al., 2014). Noisy points are notably observed at the edge of objects in the depth image, as schematized in Figure 5.3. Such errors usually create irregular edges in a depth-map representation of objects whose silhouettes are, in reality, smooth. It also originates along the contour of a given foreground the presence of background information and/or the other way around. Such effect is designated as lateral noise (Mallick et al., 2014).

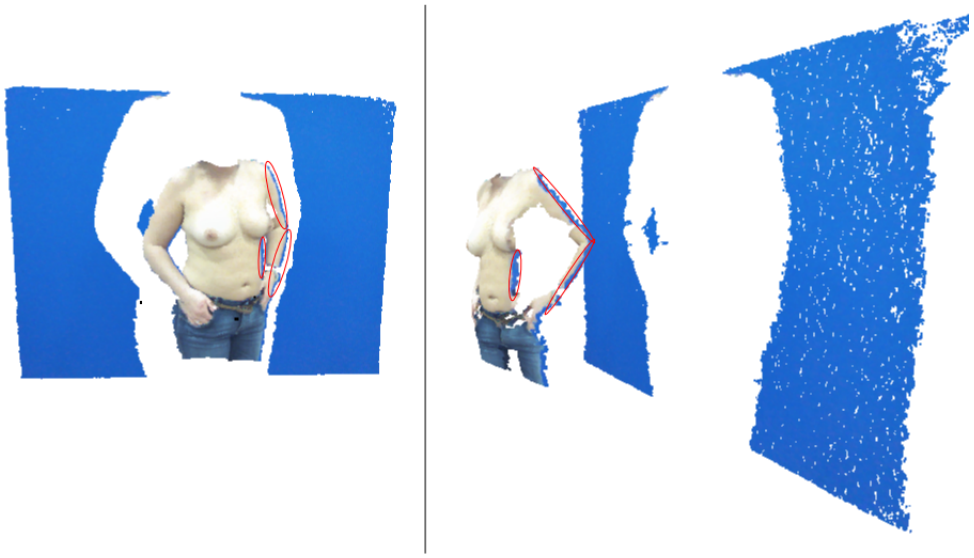


Figure 5.3: The presence of lateral noise on the Kinect depth-maps creates noisy edges on the point clouds, with information from the background appearing on the foreground. These noisy parts of the point cloud are signaled using red ellipsoids.

Looking for a solution to the still open problem of lateral noise, the depth image filtering stage aims at removing untrustworthy noisy points at the edge of the foreground silhouette. Similarly to Schmeing and Jiang (2014), it was considered that given a single object of interest both color images are a set (S) of two disjoint regions, defined as foreground and background (F and B respectively). Given the aforementioned acquisition protocol, intra-similarity of such disjoints sets both in the depth and color spaces was assumed. Additionally, it was observed (Schmeing and Jiang, 2014) that if the edges in the color image and in the depth-map perfectly align, then the two sets S_{depth} and S_{color} must be the same. On the other hand, we also assumed that, since depth-maps are more noise prone than RGB images, the correct edge information can be found on the color

frame. Hereupon, depth image filtering step is performed by:

1. The color image is mapped to the depth space;

2. The binary mask obtained for depth image segmentation is used as the initial foreground/background estimate for the GrabCut Algorithm Rother et al. (2004), which performs the segmentation of the RGB data;

3. Every F_{depth} element that belongs to B_{color} is label as invalid .

The outcome of this procedure, which results in smoother borders, enhanced features and the absence of background noise on the point cloud, is exemplified in Figure 5.4. Such approach guarantees that we discard from the depth-map all the potential foreground pixels that, in the color image, are labeled as background. Apparently, we lose information about the object of interest on this process. Nevertheless, by using a multiview approach, we guarantee that, at least in one of the subsequent views, such information will appear on the foreground. Other approaches include, for instance, the interpolation of unsolved depth values. However, as these pixels are found on the inherently noisy borders of the objects, it is not trivial to find a continuity criterion. Thus, the chosen approach was to discard the untrusted information, rather than taking the risk of introducing more noise into the depth data.

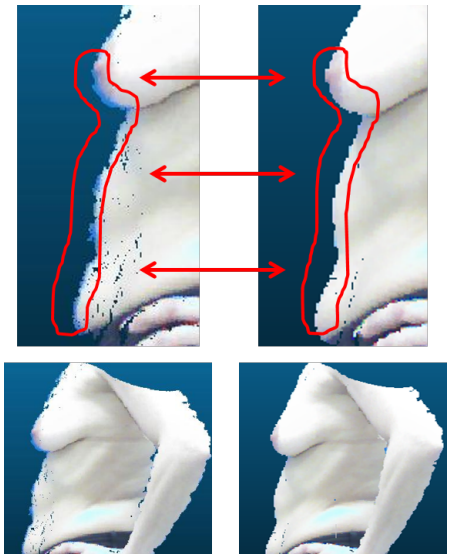


Figure 5.4: An output of the denoising methodology developed to remove the background from the generated point clouds. Smoother borders are obtained on the point cloud as well as some relevant features (for instance, the nipple) are more visible than before.

Finally, the Microsoft Kinect SDK was used to provide the integration of the RGB-D data into a colored point cloud. Briefly, the valid pixels from each pair of filtered-depth and color images are mapped into a common world space using the aforementioned calibration data.

5.2 A Tessellation-based methodology for coarse registration of point clouds

As previously said, the state of the art methodologies for the coarse registration of point clouds perform their best under controlled conditions, meaning that the results start to diverge when some rotation and translation affect increase the misalignment between views. This opens the door for the development of a new method, which performs between under these conditions. In this thesis, a Tessellation-based coarse registration method that uses depth and color information is proposed.

A tessellation is a surface composed by at least one geometric shape, with no overlaps or gaps. The simplest polygon that we can use to create a surface is a triangle. Consider now that an artificial surface of independent triangles is created to roughly describe a point set: it means that we created a tessellation over that point cloud.

The flowchart in Figure 5.5 describes the created Tessellation-based coarse registration method and its relationship with the fine registration stage. Briefly, a subset of the point cloud is computed using the Delaunay Triangulation (DT) principle, and the vertices of this surface are selected as keypoints for the registration. Then, the rigid transformation that best approximates the two views with each other is estimated using the ICP equipped with a color-validation stage, which evaluates each of the possible point-to-point correspondences.

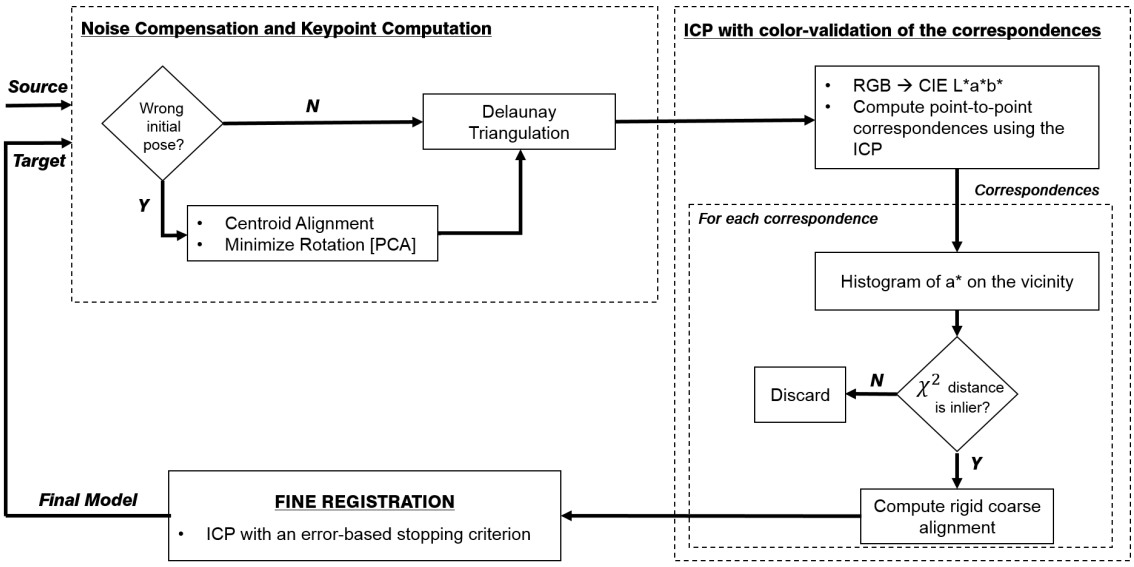


Figure 5.5: Flowchart describing the main processing stages of the Tessellation-based coarse registration technique. The method takes as input each new incoming view and the whole model being reconstructed. Then a coarse alignment of the two is calculated as the initial guess for the following fine registration steps.

The described preliminary pose initialization stage tries to alleviate the effects of high rotational and/or translational noise. It is a simple preliminary pose adjustment stage. The clouds pre-aligned by their centroids and rotated by aligning their eigenvectors, obtained by PCA. The

main goal is to minimize the random noise in such a way that the point clouds being registered are well positioned for a convergence towards an optimal reconstruction.

5.2.1 Keypoint Selection

Starting from the premise that keypoints must reliably describe the point clouds, the main goal of this stage was to downsample them based on their morphology, rather than randomly. In order to keep the computational simplicity, the main idea was to create a tessellation of triangles over the clouds. In other words, a triangulation of the surface was created using the Delaunay Triangulation principle, and certain vertices belonging to that triangulation were chosen as keypoints.

5.2.1.1 2D Delaunay Triangulation

Based on the work of Delaunay (1934), and commonly defined in 2D, the Delaunay Triangulation principle states that for a set P of 2D points, exists a triangulation $DT(P)$ such that no point in P is inside a circumcircle defined by any triangle in $DT(P)$. The main applications of the Delaunay Triangulation are related to nearest neighbor searches, meshing and path planning (George and Borouchaki, 1998) and inspired its use in the context of this thesis.

Gallier (2008) described clearly a direct relationship between Delaunay Triangulations and convex hulls, first found by Edelsbrunner and Seidel (1986). This relationship states that given a set of P points in the Euclidean space $\Re^n$, such points can be lifted to a paraboloid defined in $\Re^{n+1}$. Then, the Delaunay Triangulation of P is the back-projection to $\Re^n$ of the lowermost faces of the convex hull of the set of lifted points, as depicted in Figure 5.6.

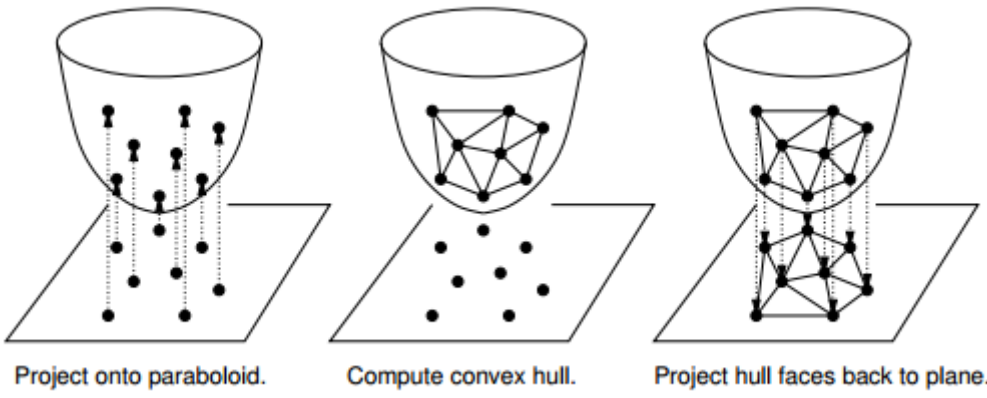


Figure 5.6: Relationship between Delaunay Triangulation and 3D Convex Hull (Mount, 2005).

Consider a set of points $P = p,q,r$ on the plane $\Re^2$, being $P' = p',q',r'$ their respective projections on the paraboloid $z = x^2 + y^2$. As mention before, the triangle defined by the points which belong to P is contained on the Delaunay Triangulation if and only no point s is not on its interior.

Extrapolating to $\Re^3$, its projection, s', must not be on the lower side of the plane passing trough the points on P'. This is called the Delaunay Lemma.

To prove the lemma, take an arbitrary non-vertical plane, which is tangent to the paraboloid above some point (a,b) on the plane. The equation of such plane can be taken from the partial derivatives of the paraboloid:

$$\frac{\partial z}{\partial x} = 2x, \qquad \frac{\partial z}{\partial y} = 2y \tag{5.1}$$

At the considered point (a,b), the partial derivatives are evaluated to $2a$ and $2b$, meaning that the plane has as equation:

$$z = 2ax + 2by + \gamma \tag{5.2}$$

To solve for γ it should be considered that the plane passes through (a,b,a^2+b^2), which implies that:

$$a^2+b^2 = 2a\cdot a + 2b\cdot b + \gamma \Longleftrightarrow \gamma = -(a^2+b^2) \Longrightarrow z = 2ax + 2by - (a^2+b^2) \tag{5.3}$$

The intersection of the plane with the paraboloid can be obtained by shifting the plane a positive amount r^2 and taking the paraboloid equation to replace z:

$$x^2+y^2 = 2ax + 2by - (a^2+b^2) \Longrightarrow (x-a)^2 + (y-b)^2 = r^2 \tag{5.4}$$

which is just a circle. Then, it is proved that the intersection of an arbitrary plane, which passes through the lifted points, with the paraboloid is an ellipse, whose projection on $\Re^2$ is a circumcircle of p,q and r, centered in (a,b), as shown in Figure 5.7. Furthermore, the point s lies within this circumcircle, if and only if its projection s' onto the paraboloid stands within the lower half-space of the plane passing through p,q and r (Mount, 2005).

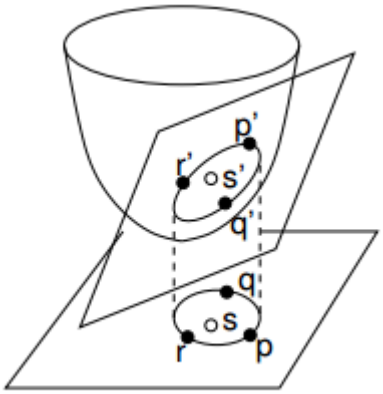


Figure 5.7: A specific point is inside the circumcircle defined by p,q and r if an only if its projection of the 3D-space is below the plane that contains the projections p',q' and r' (Mount, 2005).

To conclude about the relative position of s and the circumcircle, two determinants, Δ and Γ, must be computed. If they have opposite signs, s is inside the circle. The determinants are calculated as follows:

$$\Delta = \begin{bmatrix} 1 & p & p' & p^2+p'^2 \\ 1 & q & q' & q^2+q'^2 \\ 1 & r & r' & r^2+r'^2 \\ 1 & s & s' & s^2+s'^2 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} 1 & p & p' \\ 1 & q & q' \\ 1 & r & r' \end{bmatrix} \tag{5.5}$$

The Delaunay Triangulation algorithms try to define the triangles between adjacent points that do not have any other point on its interior, i.e. that guarantee that the aforementioned determinants have the same sign.

5.2.1.2 3D Delaunay Triangulation

The presented definition of the Delaunay Triangulation can be extrapolated for higher dimensions. Especially for sets of points in $\Re^3$, they must be lifted towards a paraboloid-like representation in $\Re^4$ whose re-projection into the original dimension creates a set of adjacent tetrahedrons, whose circumscribed spheres have empty interiors. The intersection of each tetrahedron is either an empty set or a common face or edge (Cignoni et al., 1995). By applying the Delaunay Triangulation, a set of nearly-regular tetrahedrons is obtained. The obtained triangular faces are almost equilateral, since the since the smallest angle such triangles is maximized.

5.2.1.3 Choosing keypoints from a 3D Delaunay Triangulation

The vertices of the free boundary triangular faces are chosen as keypoints. A free boundary triangle is a face of a tetrahedron on the 3D Delaunay Triangulation which faces the outside of the triangulation and belongs to only one tetrahedron. An example of the keypoint selection stage using a female torso model is shown in Figure 5.8. Based on these conditions, it is possible to observe that the Delaunay Triangulation method, and the correspondent tessellation, can be used to get some global information about a point set. Especially some distinctive points of the point cloud and its convex hull can always be obtained. In the following section, it will be demonstrated how this information can be used to perform a coarse alignment of two 3D views of the same object.

5.2.2 Correspondence estimation and validation

Using the Delaunay Triangulation, a subspace of the point clouds being registered is created, i.e, the views are downsampled. Then, assuming that the tessellations describe correctly the point clouds, these keypoints are used by the ICP algorithm to coarsely align them:

1. For each point on the source tessellation, a nearest neighborhood search is done on the target tessellation, and *vice-versa*.

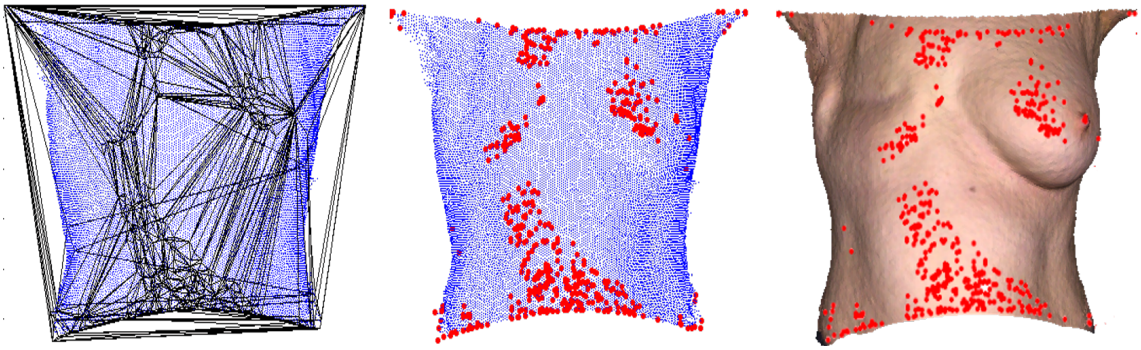


Figure 5.8: The keypoint extraction for the Stranford Bunny model using the Delaunay Triangulation principle. The keypoints are shown in red, on the right. It is possible to see that the distinctive structures of the model are correctly detected.

2. Point-to-point correspondences are determined based on that neighborhood information obtained using a *kd-tree*.

3. The found correspondences are used to estimate a rigid motion that best aligns the tessellations.

4. The transformation is extrapolated to the original point clouds being registered.

The Tessellation-based coarse registration algorithm proposed in this thesis performs a color-based validation of these correspondences. These stage is presented in Figure 5.9 and has the following processing pipeline:

1. The keypoints are the tessellations build using the Delaunay Triangulation.

2. Calculation of the ICP correspondences using those selected keypoints.

Then, for a user defined neighborhood size and for each correspondence between keypoints:

3. Take the RGB information on a neighborhood with size 500 for each keypoint *on the original clouds*, and transform it to the CIE L*a*b* color space.

4. Compute the histogram of the a* values on the selected neighborhood. Each histogram corresponds to the neighborhood of each corresponding point.

5. Calculate the Chi-Square Distance between the histograms of the points that are possible correspondences.

6. Considering the Chi-Square Distance distribution, the inliers are accepted using the criterion presented in Equation 5.6.

7. Estimate the motion based on the accepted correspondences.

The CIE L*a*b* color space was created to mimic the human vision: L^* defines the lightness of the image, a^* denotes the red/green value and b^* the yellow/blue value (Hunter, 1958). As the probability of finding some red features on the human skin is higher than finding yellow or blue ones, the algorithm discards the $b*$ value. In addition, because of the shadows and occlusions that exist in non-controlled acquisition environments, the luminance value, L^*, was also discarded. Obviously, this color-space transformation is specifically directed for body reconstruction purposes. For other applications, the color-validation stage should be performed using the original RGB information, or the color data must be transformed into another color-space.

As stated previously, the histogram computation stage is done for all the candidate corresponding points, and the motion is estimated using the inliers. An inlier is defined as follows:

$$\chi^2(H(C_T), H(C_S)) < \mu + \sigma \tag{5.6}$$

where $H(C_T)$ and $H(C_S)$ stand, respectively, for the histogram of a* of the analyzed neighborhood in the target and the source point clouds; μ is the mean chi-square distance and σ is the standard deviation of the chi-square distance.

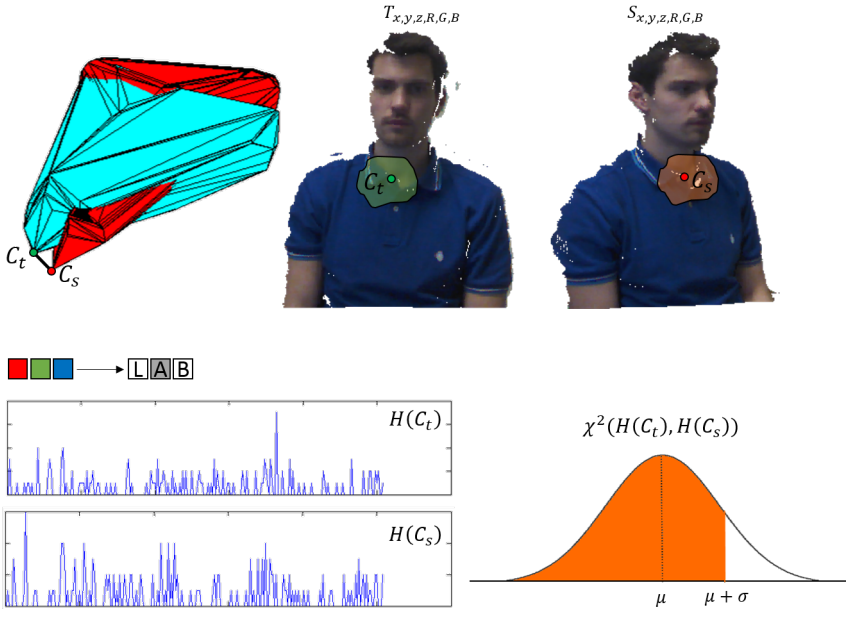


Figure 5.9: Schematic describing how the color information is used to validate the correspondences during the coarse registration stage. The potential correspondences are estimated from the Delaunay Triangulation vertices of the input point clouds S and T. Then, a correspondence-validation stage is performed. The RGB values of the vertices that are inside a user-defined neighborhood are converted to the CIE L*a*b* color space and histograms of $a*$ are calculated. A potential keypoint correspondence is accepted if the Chi-square distance between the computed histograms is below the average distance plus the standard deviation.

5.2.3 Final comments on the proposed coarse registration methodology

The proposed coarse registration technique guarantees that only the closest correspondences are used to compute the motion between the views, discarding the outliers. The estimated transformation is then applied to the original point cloud, because it is supposed to provide a good initial guess for the alignment of the two clouds. The color-validation stage rejects an average of 11% of outlier correspondences.

This contribution can be of great value, especially for the registration of non-rigid data. For instance, for two point clouds acquired from a subject using the Kinect, the coarse registration stage performs well, especially on the estimation of the rotation between the point clouds (Figure 5.10). Registering two highly rotated point clouds can be a very demanding task. By being able to perform this rotation adjustment, this Tessellation-based coarse registration technique is expected to be capable of, not only perform well in easy registration problems, but also to reconstruct objects of interest in less-controlled scenes.

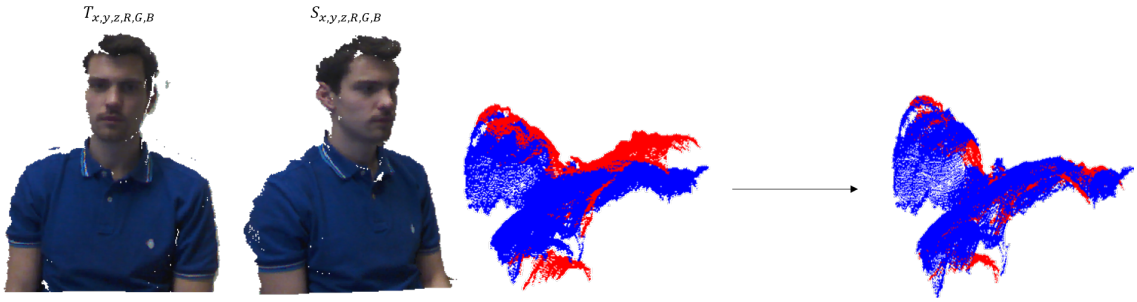


Figure 5.10: The proposed Tessellation-based coarse registration algorithm is able to estimate well the rotation between the point clouds that are being registered. On the left, the original views before the registration are seen. On the right, the result of the coarse registration phase show that the rotation was correctly compensated.

Some work on this topic was developed by Godin et al. (1994), which used color and curvature as a matching constraint for feature detection. However, the author only tested the performance of the algorithm using simple objects, without shadows, occlusions or luminance changes (Salvi et al., 2007). This means that this contribution can lead to more *ground-breaking* improvements on the registration stages.

5.3 Evaluation Metrics

After performing a 3D reconstruction, the obtained result must be evaluated. The most intuitive approach is to calculate some kind of Euclidean Distance between the reference and the reconstructed point clouds, for instance the Hausdorff Distance. This metric evaluates the mismatch between two point sets, $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$, by measuring the distance from the point of the first set that is farthest from any point in the second set, and *vice-versa*

(Huttenlocher et al., 1993). In other words, this metric maximizes the minimum distance found between the point clouds, as follows:

$$H(A,B) = max\{min(||A-B||), min(||B-A||)\} \tag{5.7}$$

The Hausdorff Distance provides a good evaluation of the reconstruction if the reconstructed and the reference models are on the same axis and completely aligned. However, this is not always possible, especially considering the methodology proposed in this thesis: the registration is not performed by aligning a moving point cloud with a static one, but by moving iteratively both point clouds to align them. This changes the pose of the final model and justifies why the Hausdorff Distance is not suitable for the evaluation in this context. The only possible solution is to guarantee that both point clouds are in the axis by registering them by either relying on the proposed methodology or some state of the art technique, thus influencing the evaluation metric.

Approaches based on the Mean Square Error have been proposed. These methodologies perform an optimal alignment between the point clouds, compute point-to-point correspondences, P' and Q', and measure how close each point p_i and its corresponding point q_i are (Gelfand et al., 2005) using different metrics, such as the Coordinate Root Mean Square Error:

$$cRMS^2(A,B) = cRMS^2(P',Q') = min\frac{1}{N}\{\sum_{i=1}^{N}||Rp_i + t - q_i||^2\} \tag{5.8}$$

where R and t are, respectively, the rotation matrix and the translation vector that determine the optimal alignment between the clouds, and N stands for the number of correspondences found between the point clouds.

Conversely, the similarity between the clouds can also be evaluated by comparing all the internal pairwise distances possible between points with correspondence, and using the Distance Root Mean Square Error formulation for that purpose:

$$dRMS^2(A,B) = dRMS^2(P',Q') = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(||p_i - p_j|| - ||q_i - q_j||)^2 \tag{5.9}$$

This Distance Root Mean Square is used, for instance, to evaluate the similarity between protein structures (Koehl, 2001). Although robust, the Mean Square Error-based methodologies have the same associated problems that the Euclidean Metrics: a registration stage or, at least, a correspondence determination phase are required in order to guarantee the accuracy of the evaluation.

Mian et al. (2006) performed the evaluation of the reconstruction using a different approach: firstly they broke the reference model into 26 different subsets and introduced known artificial transformations that misaligned each subset from the first view, saving the respective rotation matrix and the translation vector. Then, they registered each view iteratively and compared the determined transformation the introduced noise, thus obtaining the rotation and translation error for the registration. The main drawback of the proposed methodology is that different transformations can lead to reconstructions with similar quality, for instance, when the registration is performed using a different corresponding point set.

As a whole, the robustness of the above mentioned evaluation metrics relies on a previous optimal alignment between the clouds, creating the need for an evaluation metric that is independent of the final position of the reconstructed model. In this thesis, two new evaluation metrics are proposed: the first is based on the computation of an histogram of distances and the second is based on the difference between the polar representations of the reference and the reconstructed model.

5.3.1 An histogram-based evaluation metric

Figure 5.11 shows how the histogram of distances was calculated. Firstly, the Delaunay Triangulation, T of both the reference and the reconstructed model is calculated. Then, a distance matrix, $D_{nm} \ \forall \ n,m \in T_{1...N}$, is computed: each element of this matrix stores the distance ρ_{nm} between the respective two vertices of the point cloud.

Then, it is possible to build a normalized histogram of distances. Similar models are expected to have similar histograms. Thus, by using simple histogram comparison techniques, it is possible to evaluate the quality of the reconstruction. After building the normalized histogram of distances, the quality of the reconstruction can be evaluated using three different distance measures:

1. Chi-Square Distance
2. Earth Movers Distance
3. Cross-Correlation

In addition, the Bhattacharyya Distance and the Histogram Intersection were evaluated, however with inconsistent results.

5.3.1.1 Chi-Square Distance

The **Chi-Square Distance**, often represented as χ^2, is a distance measure developed for correspondence analysis and ordination techniques (Minchin, 1987). More recently, it has been used for object categories classification, local descriptors matching, shape classification (Belongie et al., 2002) and boundary detection (Martin et al., 2004).

The Chi-Square distance gives more importance to the difference between small bins than to the distance between large bins. It is defined as follows:

$$\chi^2(H_1,H_2) = \frac{1}{2}\sum_I \frac{(H_1(I)-H_2(I))^2}{H_1(I)+H_2(I)} \tag{5.10}$$

where H_1 and H_2 are the histograms being compared. Higher Chi-Square Distances correspond to more different histograms.

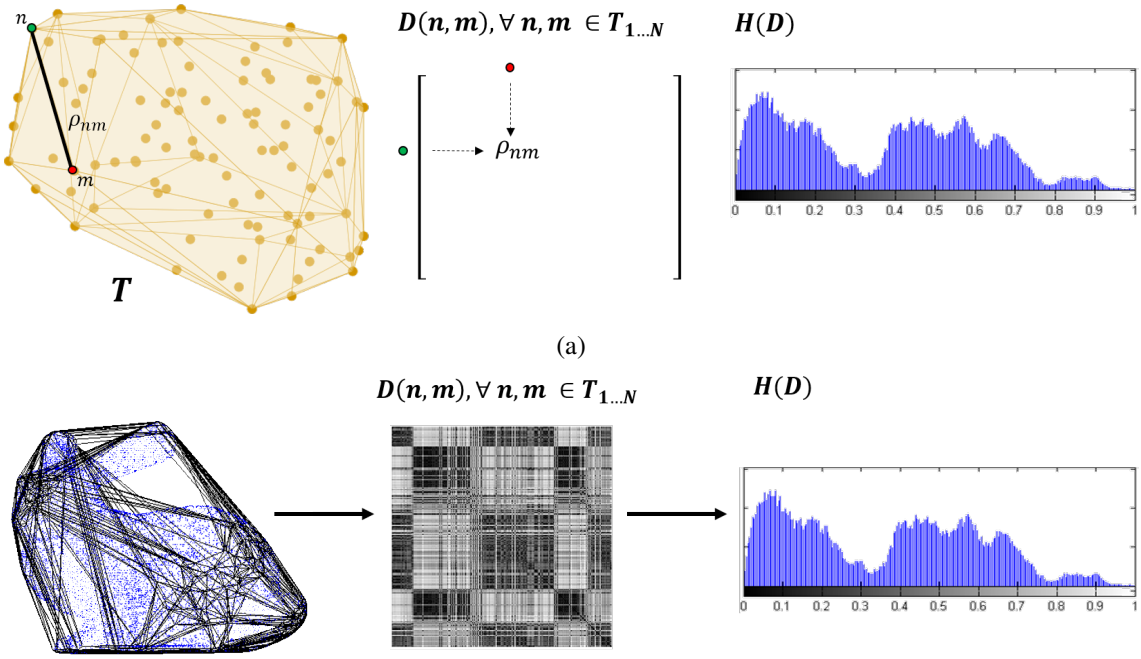


Figure 5.11: (a) Schematic representation on how to extract the normalized histogram of distances from a reconstructed (or the reference) model. The histogram describes the distances between the vertices of the Delaunay Triangulation of the input point cloud and (b) a practical example.

5.3.1.2 Earth Movers Distance

The **Earth Movers Distance (EMD)** was first proposed by Rubner et al. (2000) for image retrieval purposes. It is a cross-bin distance measure that calculates the cost that must be paid to transform an histogram H_1 into another H_2 (Pele and Werman, 2009), as seen in Equation 5.11. The EMD has the advantage to be fast and to perceive distances as humans do. In the past few years, it has been used successfully for edge and corner detection, keypoint matching, texture classification and contour matching purposes, among others (Pele and Werman, 2010).

Given two histograms P and Q, the EMD is:

$$\mathrm{EMD}(P,Q) = \min \frac{\sum_{ij} f_{ij} d_{ij}}{\sum_{ij} f_{ij}}, \qquad \sum_j f_{ij} \leq P_i, \sum_i f_{ij} \leq Q_j \tag{5.11}$$

$$\sum_{ij} f_{ij} d_{ij} = \min \left(\sum_i P_i, \sum_j Q_j \right) \tag{5.12}$$

where $\{f_{ij}\}$ denotes the flows. Each element of the flow represents the amount transported from the ith supply to the jth demand (Pele and Werman, 2009). On the other hand, d_{ij} is the ground distance between the bins i and j in the histograms. Figure 5.12 shows that one histogram can be interpreted as the supply and the other one as the demand. The EMD is simply the cost of moving the bins of P to Q (or the other way around). In other words, how much it costs to

transform all the demand in supply. As for Chi-Square Distance, a higher EMD corresponds to a worse reconstruction.

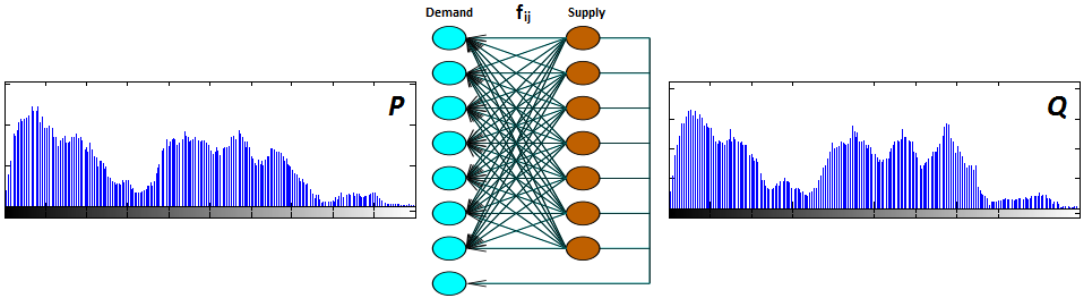


Figure 5.12: The EMD histogram distance measure can be interpreted as a simple supply and demand problem, in which we calculate the cost of transforming the histogram Q in the histogram P. Adapted from Pele and Werman (2009).

5.3.1.3 Cross-Correlation

Finally, we can measure the **cross-correlation** between two histograms, H_1 and H_2, as defined in Equations 5.13 and 5.14. For correlation, a high score represents a better match. A perfect match returns 1 and a complete mismatch -1. For a correlation of 0, either no correlation between the histograms is found or they are randomly associated.

$$Corr(H_1,H_2) = \frac{\sum_I (H_1(I)-\bar{H_1})(H_2(I)-\bar{H_2})}{\sqrt{\sum_{1I}(H_1(I)-\bar{H_1})^2 \sum_I (H_2(I)-\bar{H_2})^2}} \tag{5.13}$$

where

$$\bar{H_k} = \frac{1}{N}\sum_j H_k(J) \tag{5.14}$$

5.3.2 Integral of a polar curve difference

The second developed metric is based on polar curves that summarize the shape of an object of interest, as shown in Figure 5.13. Consider a point cloud, whose centroid, C, has the normal vector $\vec{n}$. The vector $\vec{n}$ is the resulting vector obtained from the vectors that describe the three main orientations of the point clouds (the three principal components or eigenvectors).

Each point P on the cloud can be described by one angle and one distance: θ is the angle between the normal vector and $\vec{PC}$; ρ is the distance between the point and the center of mass.

With these two pose invariant descriptors, it is possible to create the polar curves $M(\theta,\rho)$ and $R(\theta,\rho)$ that describe, respectively, the reconstructed model and the reference model. Then, the integral of the difference between these polar functions, τ, is calculated as follows:

$$\tau(\text{m.rad}) = \int_0^{\theta_{\max}} |R(\theta,\rho) - M(\theta,\rho)| \; d\theta \tag{5.15}$$

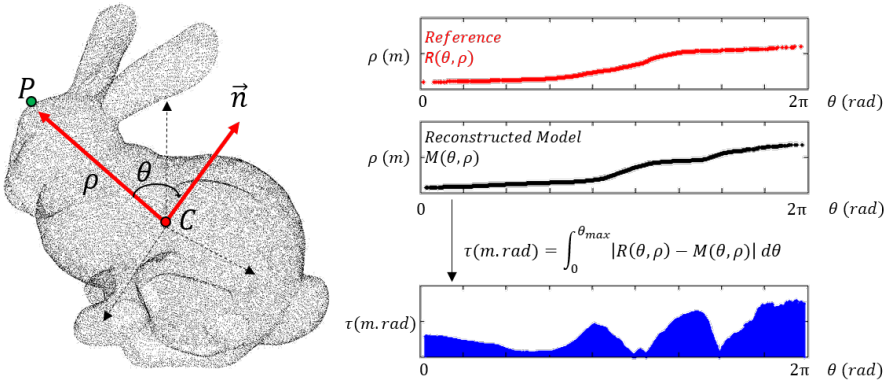


Figure 5.13: Computation of the integral of a polar difference curve. For both the reconstructed and the reference model, a pose invariant curve descriptor is computed: each point P can be described by θ, the angle between $\vec{n}$ and the vector $\vec{PC}$ (C is the center of mass of the point cloud); and ρ the distance between the point and the center of mass. Then, the integral of the difference between the two polar descriptors is calculated.

When measuring the integral distance between two point clouds with a different number of vertices, this methodology interpolates the smallest curve in order to obtain curves with similar size.

Good reconstructions correspond to smaller values of τ, because similar point clouds can be described by similar curves. As said previously, this is an invariant descriptor function: each model has an unique set of pairs (θ,ρ), regardless of the spatial orientation. The numeric value of this metric is expected to increase as the number of point increases.

5.3.3 Noise sensitivity

One experiment was performed to evaluate the noise sensitivity of the proposed metrics to the existence of noise. Two equal clouds were taken and, while the first one remained unchanged, increasing levels of noise were applied to the second view and the output of the metrics was observed.

Firstly, the maximum amount of noise was set to 5 mm. Then, for each noise level, each point P belonging to the second cloud was *noisified* using the following relation:

$$P_i = P_{i-1} + N_i, \qquad i \in \{1,\ldots,100\} \tag{5.16}$$

where

$$N_i(m) = \frac{i}{100} \cdot 0.05 \cdot K \tag{5.17}$$

Here, i stands for the upper level noise that can be applied to that point in a specific iteration; K is a vector with size 1×3 with a random number between 0 and 1 at each position. This

randomization guarantees that, for each iteration, a noise level between 0 and the maximum noise percentage allowed in that iteration is applied to the point.

Figures 5.14 and 5.15 show the response of the developed metrics to increasing levels of noise. The observed tendency on the results behaves as expected.

The distance between increasingly different point clouds is correctly described by the proposed metrics. Although random noise is introduced, the captured tendency suggests that the output developed evaluation metrics is coherent: the higher is the introduced noise, the higher the measured distances and the lower the correlation.

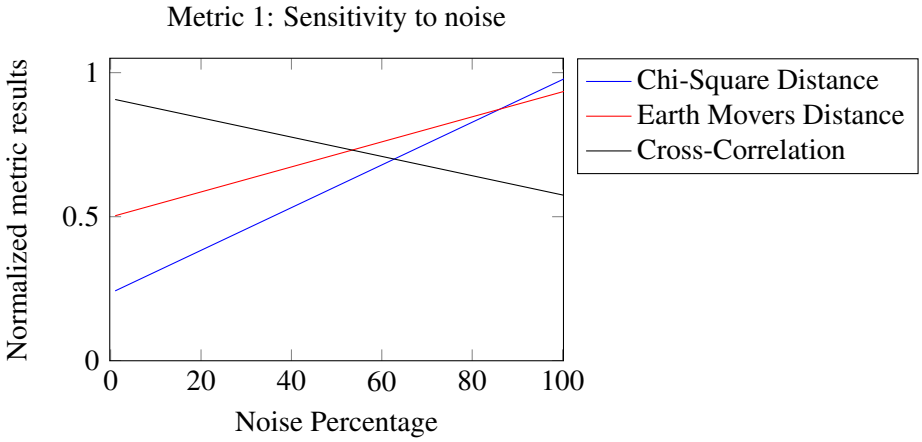


Figure 5.14: The evolution of the histogram-based evaluation metric to noise. The Chi-Square Distance, the EMD and the Cross-Correlation normalized responses are presented, respectively, in the first, second and third row.

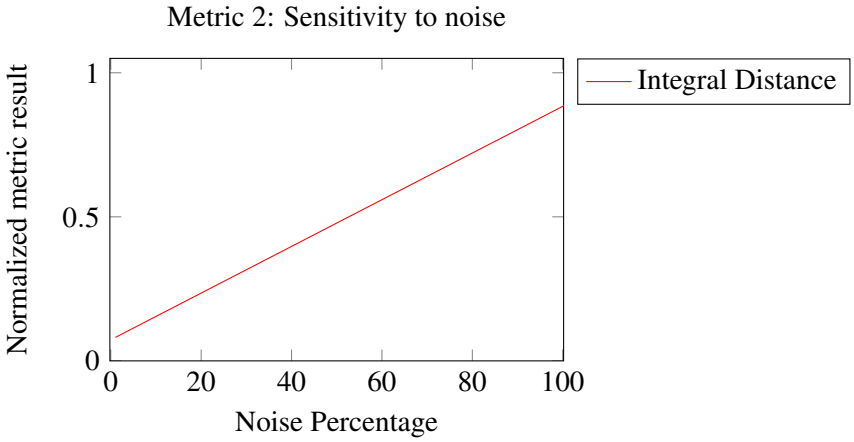


Figure 5.15: The evolution of the metric based on the integral of a polar curve difference to noise.

5.3.4 Final Considerations

The proposed evaluation metric seems to capture correctly the presence of increasing levels of noise in the point clouds. This, in addition to their pose invariance, suggest their robustness in the

description of the shape of point clouds. Nevertheless, these solutions present some drawbacks.

The first problem is common to both metrics: although similar point clouds have similar histograms of distances or similar polar curves, different point clouds can also provide similar descriptors. In other words, the metrics do not avoid redundancies: different point clouds can lead to the same result. However, in the context of this thesis, this should not be a problem because the metrics will be used to compare the performance of different reconstruction methodologies. As the initial conditions and the point clouds being registered are the same, this problem is not expected to occur.

The second problem is specific for the metric based on polar curves, which intend to summarize the shape of both the reference and the reconstructed point clouds. As previously said, each point is described by a pair (θ, ρ), which is not unique. In other words, it is possible to have points that share the same angle, or even both descriptors. This is because one angle is not sufficient to uniquely describe the relative position of points in 3D. Some work is being currently done to minimize this drawback. The initial thought on how to solve this problem include the calculation of the angle, not only with respect to the vector $\vec{n}$, but to two of the principal components of the point cloud. The task will be transformed into the calculation of a volume: the integral of the difference between two 3D surfaces defined by ρ and two angles.

Finally, the proposed evaluation metrics do not provide any insight on the resolution of the reconstruction. It would be interesting to modify the algorithms to incorporate this feature, as well as to solve the above mentioned problems.

5.4 Conclusions

A new methodology for the coarse registration of point clouds is proposed. This methodology uses the Delaunay Triangulation principle to create a tessellation of the surfaces, whose vertices correspond to distinctive structures or regions on the point clouds which are chosen as keypoints. This methodology is designed to downsample the point clouds and is desired to compensate well high rotation and/translation misalignments between the point clouds. One improved feature of this methodology is its color-based validation stage, which guarantees the coarse alignment between clouds is performed using reliable corresponding points.

In addition, two new reconstruction evaluation metrics were developed. These metrics are designed to be pose-invariant, thus being capable of correctly describe the point clouds independently of their reference coordinates and spatial orientation. It would be interesting to evaluate how these metrics can be incorporated in the coarse and fine registration methodologies as stopping criterion, or even to guide the reconstruction stage. Although the proposed metrics correctly describe the increasing distance between increasingly different point clouds, they have some associated problems and further research must be done towards the minimization of their negative effect concerning the quality of the evaluation.

Chapter 6

Results and Discussion

In this chapter, the performance of the methodology proposed in Chapter 5 is presented and discussed. The main goal is to show how the proposed Tessellation-based coarse registration technique improves the global performance of the registration pipeline, especially when dealing with non-overlapping and noisy data.

In order to evaluate such performance, two types of analysis were performed in a Intel Core i7-2600 CPU @ 3.40 Ghz, 8GB RAM (64-bit) computer:

1. **Rigid Registration:** Stanford Bunny and Stanford Horse.
2. **Non-Rigid Registration:** Male Head (with *ground truth*) and the Female Torso of three breast cancer patients (without *ground truth*).

6.1 Rigid Registration

As previously said, in a biomedical context, the point cloud registration methodologies always deal with non-rigid data. Nevertheless, testing the methodologies in a rigid registration context can be essential to understand in which cases the commonly used algorithms fail. In addition, such type of models allows the user to control the amount of noise on the processing environment.

Both the Stanford Bunny and the Stanford Horse were reconstructed, using 5 views, under different initial poses, 5 times for each condition.

- **Rotation around z-axis:** 0, 5, 10, 15, 20, 35, 30 and 35 degrees.
- **Random shift factor:** random translation with magnitude up to 1/10, 1/50 or 1/100, respectively 10%, 50% and 100% of the maximum length found in the point cloud.
- **Sampling:** 10% and 50% of the points were used.
- **Maximum number of iterations:** 200.

As the models did not have any color information, the color-validation stage of the coarse registration was skipped.

6.1.1 Results obtained with the histogram-based evaluation metric

This is a metric composed by three descriptors: the Chi-Square Distance, the EMD and the Cross-Correlation. Each of the metrics contributes with relevant and distinct information about the performance.

Figures 6.2 and 6.3 show the Chi-Square Distance between the normalized histograms of distances that describe both the reconstructed and the reference models. From this figure, three main conclusions can be drawn:

1. The global performance of the proposed methodology —which uses the Tessellation-based coarse registration algorithm— is better than the performance of the standard counterparts. This is especially true when dealing with high rotation and translational noise. The magnitude of the error remains stable and low up to a certain threshold (divergence) point.

2. The algorithm starts to diverge from the optimal solution when the rotational noise reaches a value around 20 degrees. This is a major improvement when compared with the standard methods because the threshold point for them is reached before. The result of the reconstruction for each methodology at their divergence point can be found in Figure 6.1.

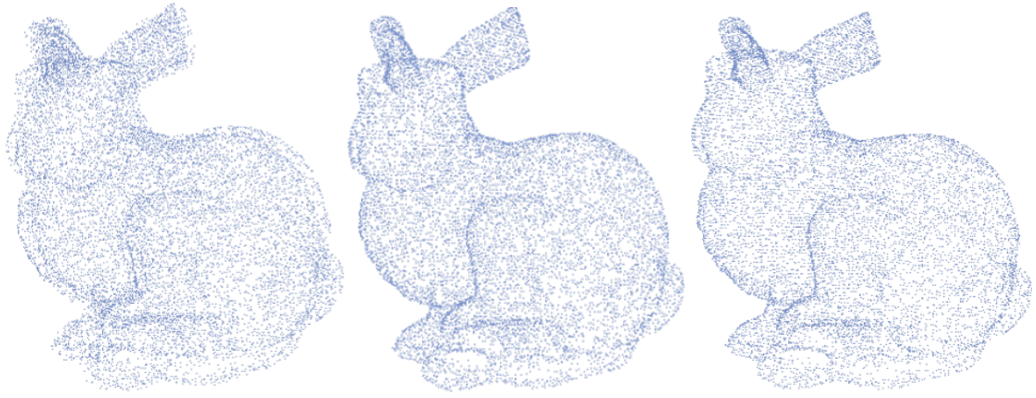
3. The standard deviation for the proposed methodology is very small (close to 0) for shift factors of 1/10 and 1/50, which correspond to higher translational noise. On the other hand, when dealing with low translational noise, the standard deviation increases as function of the rotational noise. Regarding the standard methodologies, the standard deviation of the error remains approximately constant.

4. Counter-intuitively, using more points on the registration does not mean that best results are obtained: the amount of associated error seems to increase when more points are used.

The low standard deviation values found for high translational data can be explained by how the proposed methodology deals with high initial misalignment between views. As previously said, when high translation and rotation are detected, the algorithm pre-aligns the point clouds being registered using their centroids and uses PCA information to pre-rotate them. This means that, although random noise is applied, the proposed approach always pre-processes data to start the registration stage from a *comfortable* initial guess.

This is rather important for biomedical applications if the acquisition protocol is not performed correctly, thus leading to the existence of high translated and rotated consecutive point clouds. If a good initial guess is obtained, the algorithm converges better towards the optimal solution.

Concerning item 4, the quality of the reconstructions is worse when more points are used. During the fine registration stage, the Global Procrustes ICP finds correspondences between the views being reconstructed and estimates the motion based on that correspondences. If some of them are incorrect, which is easy when dealing with complex shapes, the error propagates more easily. This justifies the need to down-sample the data at an appropriate rate, which diminishes the



(a) ICP
R:0°; SF:1/100

(b) PCA+ICP
R:10°; SF:1/100

(c) Tess+ICP
R:15°; SF:1/10

(d) ICP
R:0°; SF:1/100

(e) PCA+ICP
R:10°; SF:1/100

(f) Tess+ICP
R:15°; SF:1/10

Figure 6.1: The reconstructed models obtained using each of the tested methodologies at theirs divergence point. If more rotational or translational noise is applied, the results become less acceptable. It is possible to observe that the *Tessellation+ICP* methodology diverges later than the other methods and returns noisy reconstructed models. [1]

processing time and avoids this error propagation effect, while still being able to correctly describe the surfaces being registered.

Considering Figures 6.4 and 6.5, it describes the EMD between the normalized histograms of distances. These results have an expected behavior: when translation and rotation increases, the error does the same. Finally, Figures 6.6 and 6.7 shows how correlated the evaluated histograms are. The main conclusion drawn from these results reinforces what was stated previously in item 1: not only the proposed methodology leads to a better global performance when high translational noise conditions are found, but also achieves a higher correlation between the reconstructed and the reference models.

Obviously, this correlation is degraded when higher rotational noise is found. Nonetheless, the performance is still similar or better than the standard techniques.

[1] All reconstructed models available in: http://www.inescporto.pt/ jpsm/bibm14/rigid/

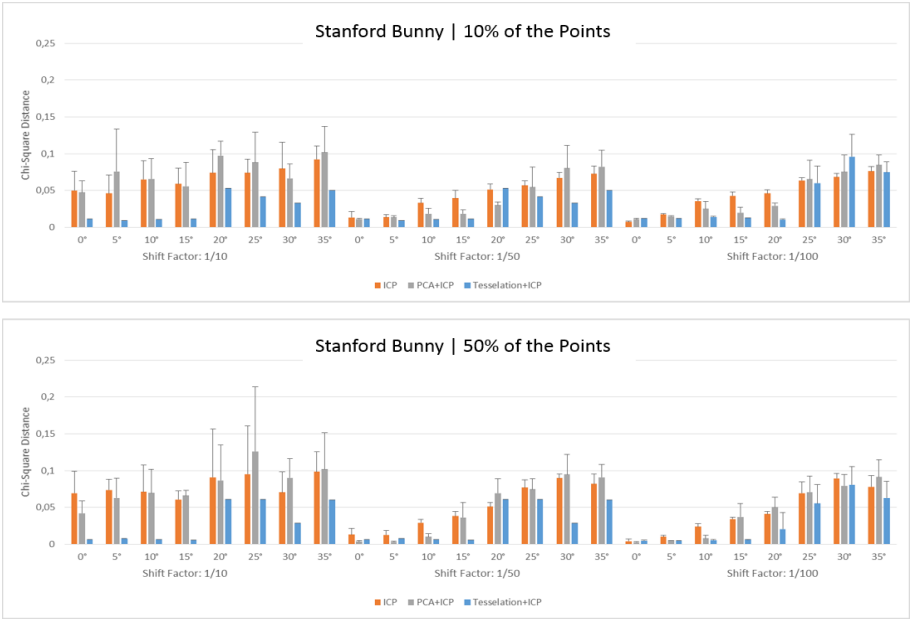


Figure 6.2: The Chi-Square Distance values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the normalized histograms of distances that describe the reconstructed and the reference Bunnny Models. A higher Chi-Square Distance corresponds to a worse reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

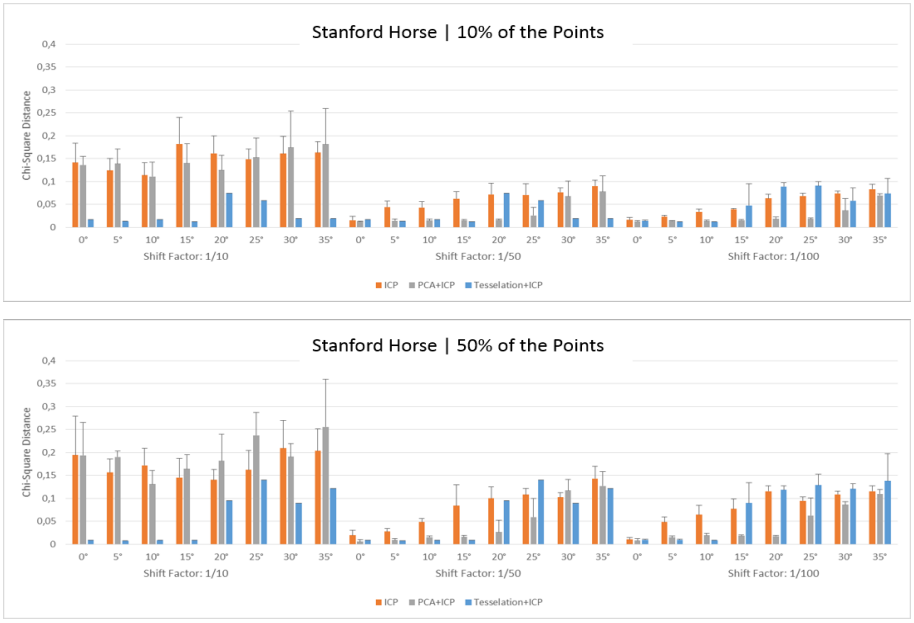


Figure 6.3: The Chi-Square Distance values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the normalized histograms of distances that describe the reconstructed and the reference Horse Models. A higher Chi-Square Distance corresponds to a worse reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

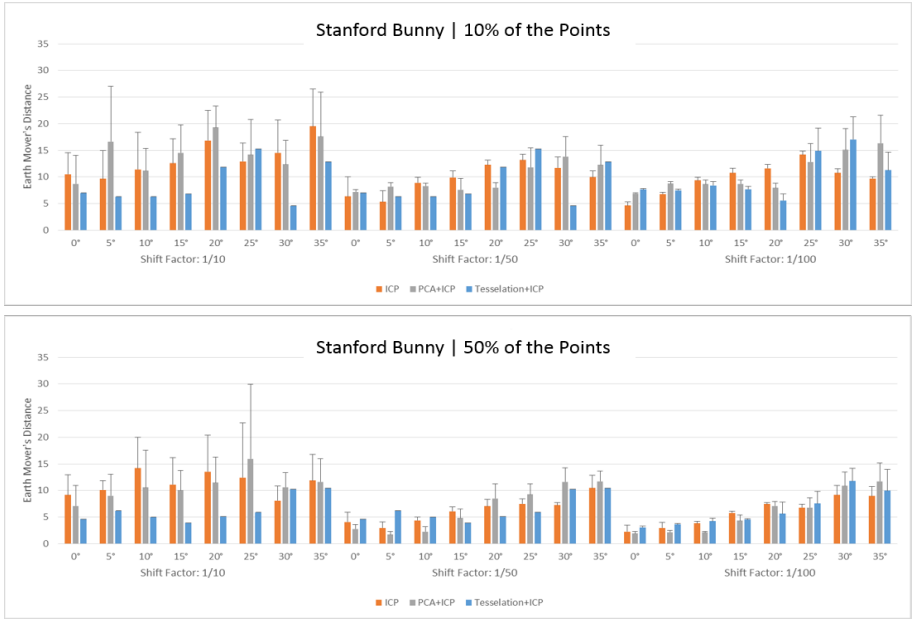


Figure 6.4: The Earth Movers Distance values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the normalized histograms of distances that describe the reconstructed and the reference Bunny Models. A higher Earth Movers Distance corresponds to a worse reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

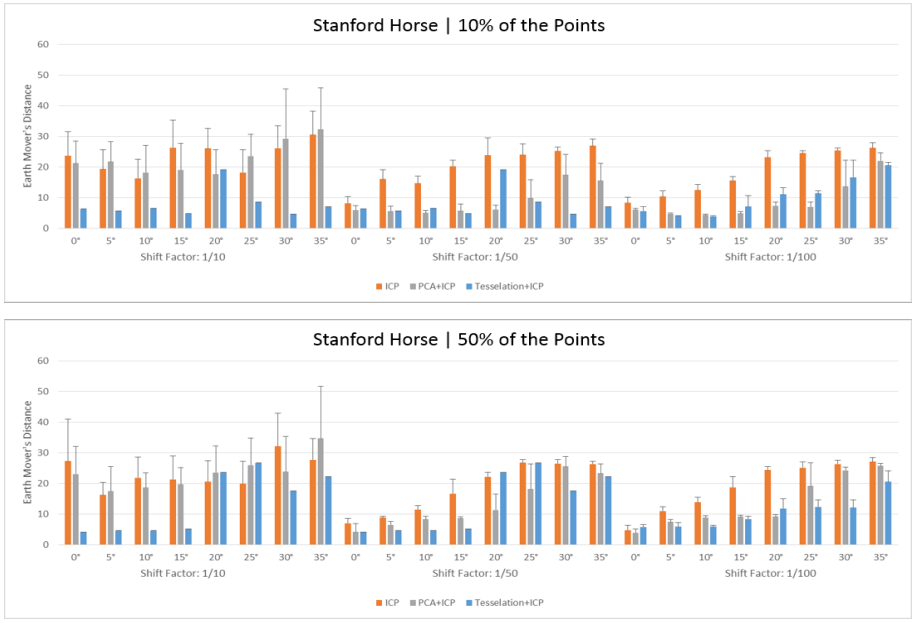


Figure 6.5: The Earth Movers Distance values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the normalized histograms of distances that describe the reconstructed and the reference Horse Models. A higher Earth Movers Distance corresponds to a worse reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

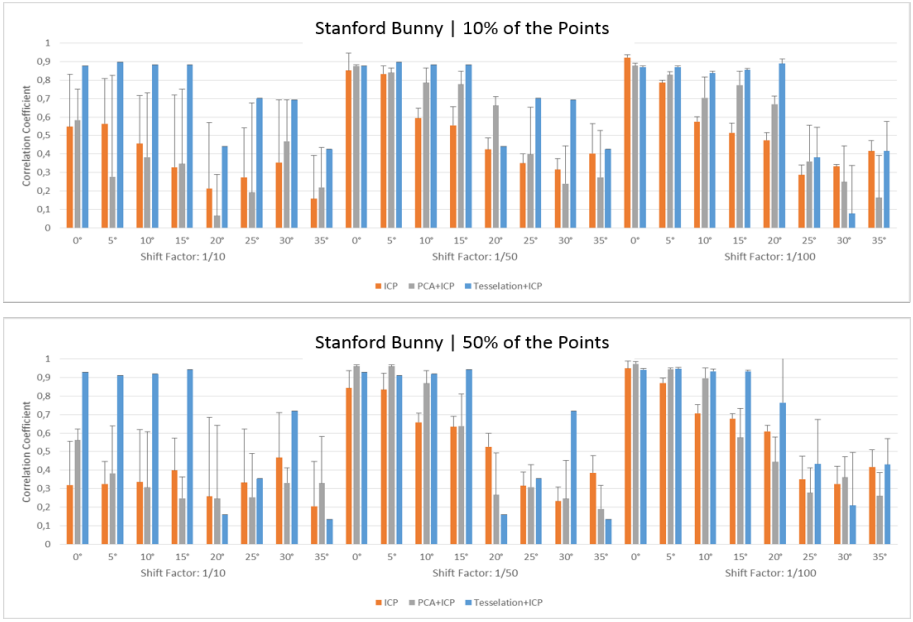


Figure 6.6: The Cross-Correlation values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the normalized histograms of distances that describe the reconstructed and the reference Bunny Models. A higher correlation corresponds to a better reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

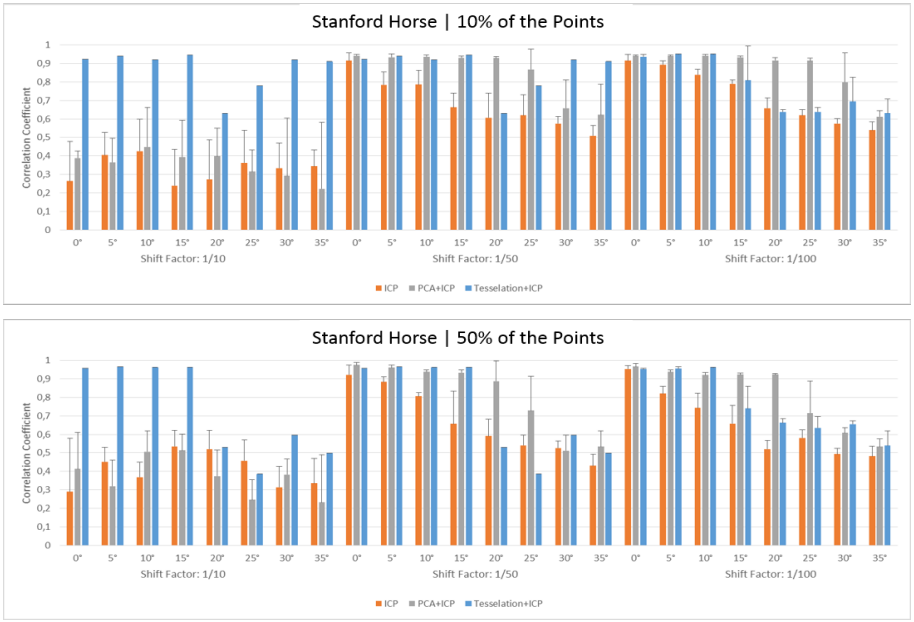


Figure 6.7: The Cross-Correlation values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the normalized histograms of distances that describe the reconstructed and the reference Horse Models. A higher correlation corresponds to a better reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

6.1.2 Results obtained using the polar curve difference metric

The results obtained using this metric are shown in Figures 6.8 and 6.9. Two main conclusions can be drawn from these results:

1. The more noisy the initial conditions of the registration stage are, the worse the performance. Nonetheless, when the Tessellation-based coarse registration algorithm is used, the quality of the reconstructions under noisy conditions seems to be as good as it for low-noise point cloud registration problems.

2. There are relevant differences on the error distribution between different models. This means that the evaluation metrics and the results are model dependent. This also means that it is not possible to compare the robustness of the algorithms based only on its performance data for different models. The proposed metrics only give some insight on the relative performance of the algorithms for the reconstruction of a specific object of interest.

6.1.3 Performance evaluation

Table 6.1 presents the global performance overview concerning rigid registration. It is possible to observe that, using the proposed Tessellation-based coarse registration methodology, both the fine registration time and the number of iterations needed to register each view decrease. Using this method, not only best reconstructions are possible, but also with a lower computational cost. In addition, although the number of iterations per view remains close to the maximum (200), a relevant decrease on the fine registration time is achieved. This is especially true when a high number of points is used. Regarding the coarse registration stage, it takes slightly longer to finish the processing when compared to the standard methods. Nevertheless, the *trade-off* is advantageous because a small increase on the complexity of the coarse alignment stage reflects on a faster, but still accurate, fine registration. The unexpectedly high standard deviation values obtained may be related with the randomization of the initial pose of the point clouds being registered.

Table 6.1: The performance evaluation of the different tested methodologies on the registration of rigid models. The values presented are the average values for 5 runs under all specified conditions.

Points	**Method**	**Time (s)** $\mu(\pm\sigma)$**/view**		**Iterations** $\mu(\pm\sigma)$**/view**
		Coarse Reg.	*Fine Reg.*	*Fine Reg.*
10%	ICP	-	12.568(± 0.889)	199.613(± 1.878)
	PCA+ICP	0.017(± 4×10^{-4})	10.933(± 1.977)	199.613(± 1.878)
	Tessellation + ICP	0.693(± 0.032)	10.485(± 2.575)	167.632(± 41.255)
50%	ICP	-	46.133(± 22.767)	115.134(± 65.89)
	PCA+ICP	0.079(± 0.004)	48.284(± 21.701)	120.501(± 64.867)
	Tessellation + ICP	1.939(± 0.146)	28.369(± 20.006)	103.633(± 69.137)

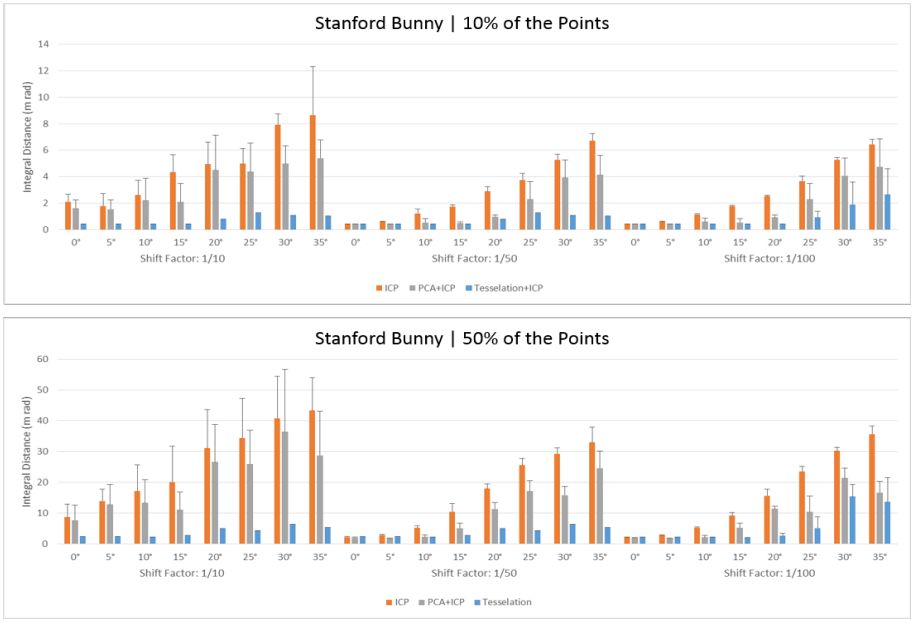


Figure 6.8: The Integral Distance values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the polar curves that describe the reconstructed and the reference Bunny Models. A higher Integral Distance corresponds to a worse reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

Figure 6.9: The Integral Distance values [● ICP; ● PCA+ICP; ● Tessellation+ICP] between the polar curves that describe the reconstructed and the reference Horse Models. A higher Integral Distance corresponds to a worse reconstruction. The translational noise decreases from left to right and, for each shift factor, rotational noise from 0 to 35 degrees is shown.

6.2 Non-Rigid Registration

The registration of non-rigid data is a common problem in the Biomedical Engineering field. In this section, the applicability of the proposed Tessellation-based coarse registration methodology to the 3D breast reconstruction is studied. As only one ground-truth was available, the reconstruction of the human head was also studied, for a better characterization.

The reconstruction of such models was performed under different conditions, again 5 times for each condition, in the same Intel Core i7-3632QM CPU @ 2.20 Ghz, 8GB RAM (64-bit) computer:

- **Randomness on the initial pose:** shift factor of 1/10, which is the most severe translational noise. The rotational noise is intrinsic to the acquisition protocol (see Chapter A).
- **Maximum number of iterations:** 300
- **Number of views:** 3 views. The original clouds were subdivided into three sets, respectively, for frontal, left and right view. A point cloud from each subset was randomly chosen for the registration.
- **Registration order:** Frontal view, left view and right view.
- **Downsampling:** Male Head and Female Torso with *ground-truth*: 100% of the points; Female Torso without *ground-truth*: 8% of the points.

The *ground truth* for the Head Model was acquired using the Kinect Fusion Algorithm (Newcombe et al., 2011), available in the Microsoft Kinect SDK. In addition, the reconstruction of such structure for 8 breast cancer patients was performed. The *ground-truth* for one of them was acquired using a high-resolution commercial 3D scanner, while the other 7 results can only be evaluated visually.

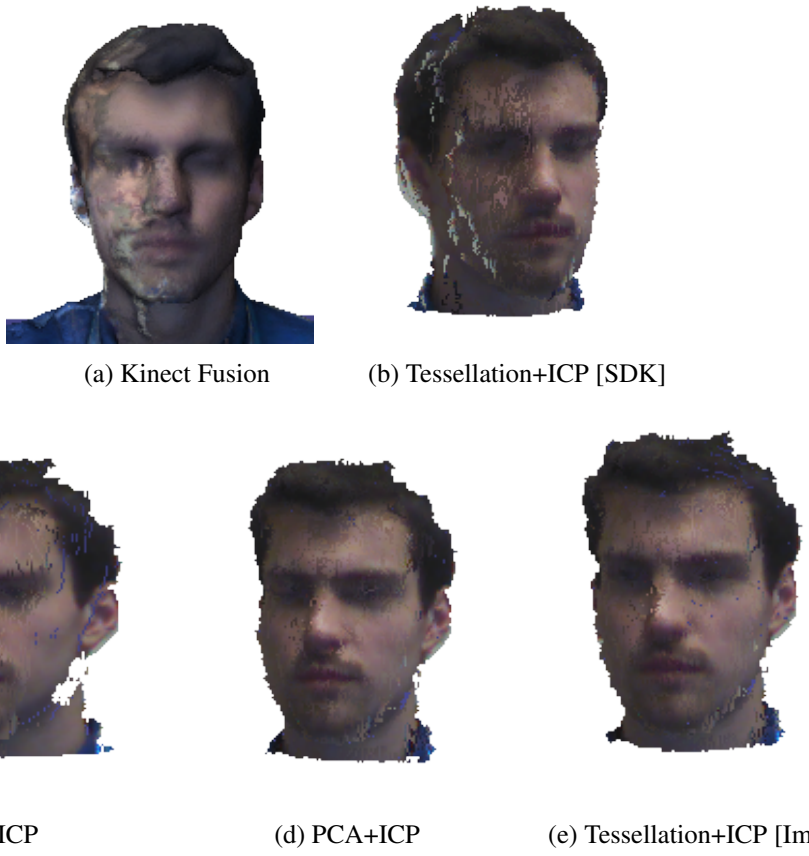
6.2.1 Male Head Reconstruction

The reconstructed models of the male head obtained from each of the tested methodologies are shown in Figure 6.10[2]. The obtained results were, also, evaluated using the previously presented metrics, and the results are shown in Figure 6.11. A global analysis of the results suggests that using the Tessellation-based coarse registration algorithm improves the quality of the reconstructions.

Some important conclusions can be drawn by analyzing the results:

1. When no translational noise is added (no shift factor), the quality of the results of the PCA+ICP and the Tessellation+ICP approaches is similar. Nevertheless, the latter one still has a better numeric and visual performance on the reconstruction of the male head. When

[2]Non-rigid registration results available for observation in: http://www.inescporto.pt/ jpsm/bibm14/torso/



(a) Kinect Fusion (b) Tessellation+ICP [SDK]

(c) ICP (d) PCA+ICP (e) Tessellation+ICP [Imp.]

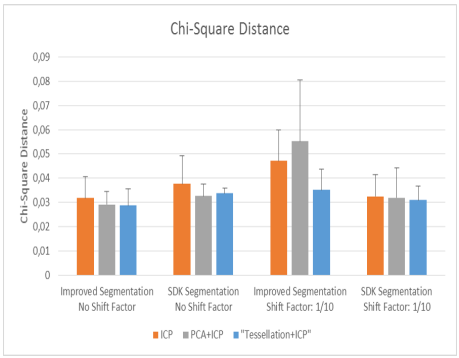
Figure 6.10: The reconstructed models of a male head using 3 different views. Although the reconstructions from the PCA+ICP and the Tessellation+ICP are visually similar, the latter one presents less deformation on the nose and is globally less noisy. Using the improved (Imp.) version of the segmentation algorithm during the point cloud generation stage, instead of the SDK version, better results are obtained.

the shift factor increases to 1/10, the gap between the quality of the methods increases. In other words, similarly to what happened for rigid registration, the Tessellation-based coarse registration algorithm leads to a much better quality of the reconstruction of models with high translational noise, when compared to the commonly used methods.

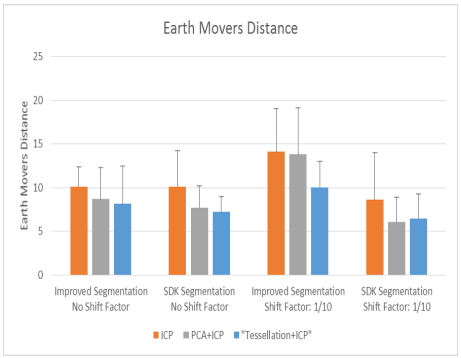
2. The improvements made to the point cloud generation stage improve the quality of the reconstructed models. Firstly, because the artifacts that are created by the SDK segmentation algorithm (see Figure 6.10b) almost disappear when the improved version of the algorithm is used. Secondly, the error on the reconstruction expressed by the integral distance is significantly lower.

Regarding the visual aspect, the Tessellation+ICP approach returns a reconstructed model with uniform texture. The number of the color artifacts created is significantly lower when compared to other methodologies, and none of them seems to compromise the global quality of the reconstruction.

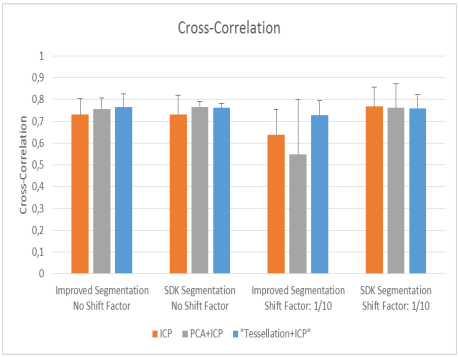
Only 3 views were used for the reconstruction. Theoretically, registering a high number of views means that the reconstructed model is more detailed and contains more information. That is not necessarily true. When 6 or more views were registered, the reconstructed models were significantly more degraded because of the accumulation of noise. Thus, a compromise between the number of views and the amount of information must be defined, being the quality of the reconstruction a decisive factor. In order to avoid this, some improvements should be done in the proposed methodology, especially the implementation of a denoising stage after the registration of each view. It would also be interesting to incorporate non-rigid transformations, capable of performing a local deformation of the point clouds for more flexibility on the alignment.



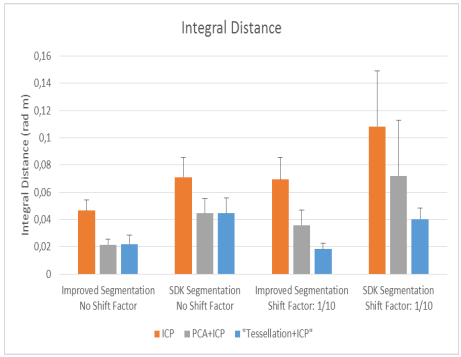
(a) Chi-Square Distance



(b) Earth Movers Distance



(c) Cross-Correlation



(d) Integral Distance

Figure 6.11: The results [● ICP; ● PCA+ICP; ● Tessellation+ICP] for the Non-Rigid Registration of the Head Model: (a), (b) and (c) correspond, respectively, to the results obtained by applying the histogram-based evaluation metrics; (d) corresponds to the integral distance between the polar curves associated to the models. **For each bar plot**, from left to right: *Improved Segmentation; No Shift Factor*, *SDK Segmentation; No Shift Factor*, *Improved Segmentation; Shift Factor: 1/10* and *SDK Segmentation; Shift Factor: 1/10*. The Tessellation-based coarse registration methodology seems to perform better than the standard algorithms, especially when the point clouds are generated using the improved segmentation algorithm.

6.2.2 Female Torso Reconstruction

The 3D Modeling of the female torso is the main topic of this thesis. In order to understand the applicability of the proposed methodology, two different experiments were conducted using data obtained for the PICTURE project and following the protocol available in the Appendix A.

The first experiment consisted in the reconstruction of the torso of a single-breasted female patient, whose *ground-truth* was obtained using a high-resolution commercial 3D scanner. An example of the obtained reconstructions is shown in Figure 6.12[3], and the quantitative evaluation of the reconstructions is available in Figure 6.13. In this experiment, a region of interest of the torso that contained the breast was selected by an artificial segmentation stage.

The obtained results suggest some relevant conclusions:

1. Only the Tessellation+ICP approach provides high-quality in the reconstruction. Although the reconstructed model has some color artifacts, thus creating a non-uniform texture, the shape of the reconstructed model is as expected. The important features on the breast, such as the nipple and the breast contour, are correctly enhanced e replicated. Also the visual aspect of the region where the removed breast existed is correctly reconstructed.

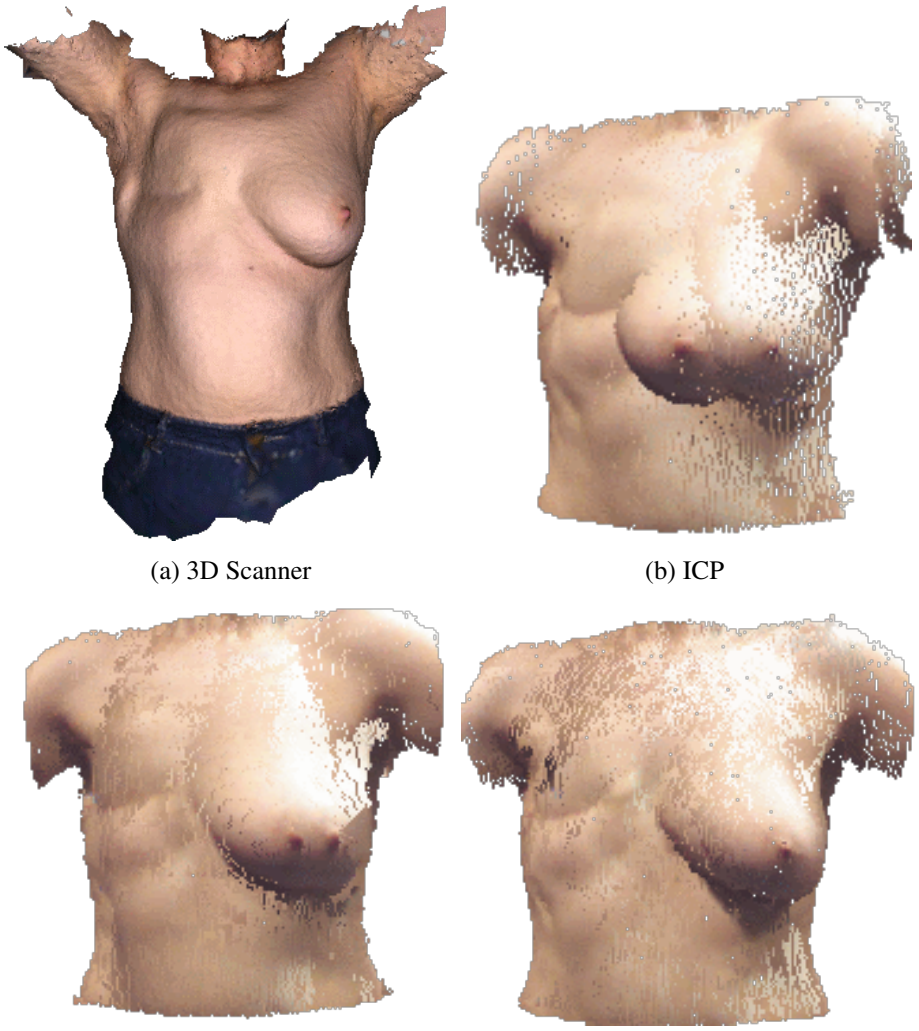
2. The ICP and the PCA+ICP methodologies were not able to perform a good reconstruction of the female torso. This may be caused by the absence of one breast, which causes a decrease in the number and the quality of the feature regions. This also true for the Tessellation+ICP approach: as previously said, the breasts are distinctive structures of the point clouds and a source of high-quality keypoints. This means that we are dealing with a more complex registration problem, in which the commonly used methodologies fail. The Tessellation-based coarse registration methodology showed a high capacity in dealing with heterogeneous shapes of the torso and with a low number of keypoints.

3. Regarding the evaluation metrics, the proposed methodology had lower Chi-Square and Earth Movers Distance and a higher Cross-Correlation, considering the histogram-based evaluation metric; in addition, a lower integral distance, measured by the polar curve-based metric, was obtained. This results prove the higher quality of the reconstructions obtained by applying the Tessellation+ICP approach.

4. The segmentation of the breast region of interest allowed the registration algorithm to focus more on the details of the breast, guaranteeing a good quality of the reconstruction. The results also suggest that the standard methodologies fail, as expected, when registering almost featureless surfaces and would need more information in order to perform better.

For the second experiment, the reconstruction of the torso of 7 breast cancer patients, using only 3 views, was performed. The visual outcome of the reconstruction for each patient is shown

[3]Non-rigid registration results available for observation in: http://www.inescporto.pt/ jpsm/bibm14/torso/



(a) 3D Scanner (b) ICP

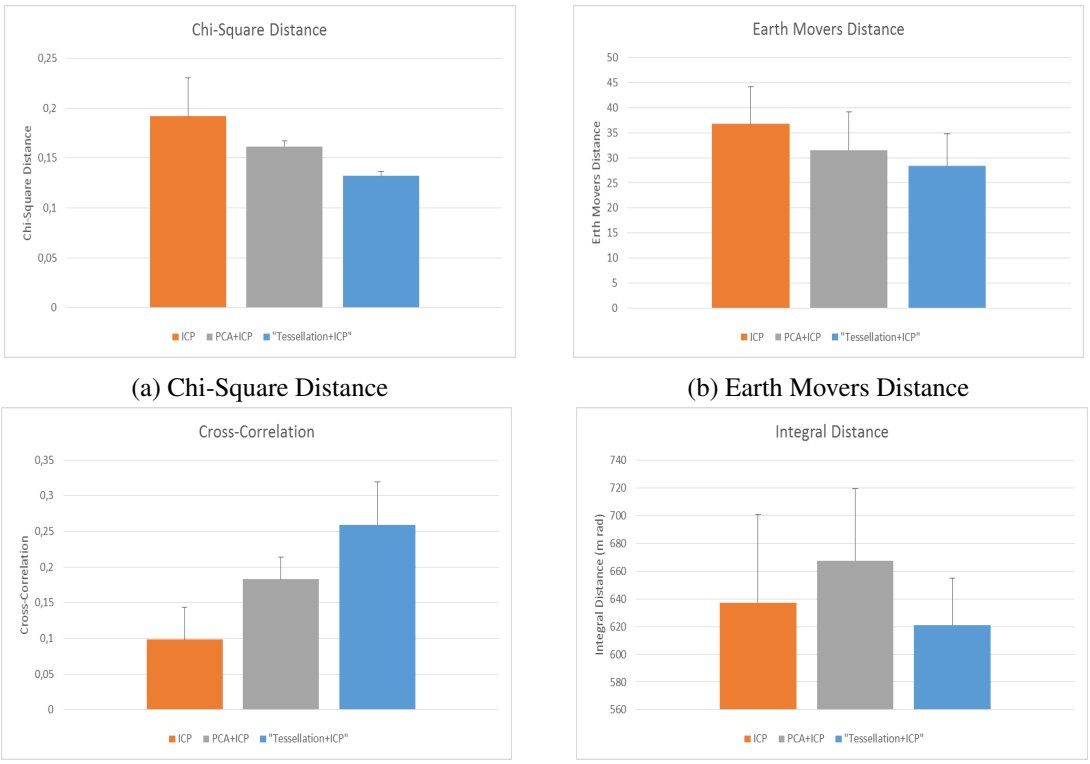
(c) PCA+ICP (d) Tessellation+ICP

Figure 6.12: The reconstructed models of the torso of a single-breasted female patient using 3 different views. A reconstruction with more quality is obtained when the Tessellation-based coarse registration methodology is used.

in Figure 6.14[4]. The reconstructed models present a high-level of detail on the breast, as well as a smooth texture. In other words, no relevant noise and artifacts are found on the models. Although performed using only 8% of the original points, the reconstruction provided good visual results, suggesting that the proposed methodology deals well with low-dense point clouds and can have relevant application on the breast-care field.

The downsampling allowed a decrease on the processing time, while maintaining the quality of the reconstructions. In addition, the algorithm deals well with heterogeneous shapes: independently on the breast format, using only 3 downsampled views, well-detailed reconstructions are created. Nevertheless, a higher level of detail can be obtained if the breast region of the point

[4]Non-rigid registration results available for observation in: http://www.inescporto.pt/ jpsm/bibm14/torso/



(a) Chi-Square Distance

(b) Earth Movers Distance

(c) Cross-Correlation

(d) Integral Distance

Figure 6.13: The results [● ICP; ● PCA+ICP; ● Tessellation+ICP] for the Non-Rigid Registration of the Female Torso: (a), (b) and (c) correspond, respectively, to the results obtained by applying the histogram-based evaluation metrics; (d) corresponds to the integral distance between the polar curves associated to the models. The Tessellation-based coarse registration methodology has results more close to the reference model than the studied standard algorithms.

clouds being registered is segmented. The definition of a region of interest will allow the registration algorithm to focus more on the little details of the breast, thus leading to higher accuracy on the reconstruction.

Some less detailed structures are found on the upper and lower parts of the reconstructed model. These artifacts are caused by the acquisition protocol. As the subject is asked to rotate in front of the camera, it is not always possible to minimize small and undesired motions that occur. This random motion cause the local misalignment between the point clouds being registered. It is not a severe problem, because the structures of interest, in this case the breasts, are not affected.

As previously said, the Delaunay Triangulation chooses distinctive feature points of the point clouds as keypoints for the registration. As the breasts are prominent structures on the female torso, they will have a lot of keypoints, meaning that a correct registration of those structures is obtained.

As a whole, the results obtained from the two experiments suggest the proposed methodology is capable of obtaining high-quality reconstructions of an object of interest using only 3 views. The fine details on the point clouds are correctly enhanced and reconstructed, in feasible time.

(a) Patient #1; 203.6 s (b) Patient #2; 103.3 s (c) Patient #3; 119.9 s)

(d) Patient #4; 94.5 s (e) Patient #5; 100.8 s (f) Patient #6; 113.7 s

(g) Patient #7; 135.0 s

Figure 6.14: The reconstruction of the female torso of seven breast cancer patients. The obtained results are very promising and show that the developed system can be of great interest in this field of research. Good reconstructions are obtained for high downsampling rates (here 8% of the points) and small processing times, meaning that the methodology is robust when registering less dense point clouds, decreasing the processing time while maintaining the quality of the reconstruction.

Nonetheless, more accuracy could be obtained if this effect is minimized. That can be done by introducing some other rigid transformations, besides from rotation and translation, when registering the point clouds, especially for freely-held cameras:

- **scale:** to compensate random movements that change the distance between the subject and the sensor;
- **skew**, to compensate oblique movements of the subjects —when the torso is not parallel to the sensor.

In this thesis, the scale and skew transformations were not considered because the distance between the subject and the sensor was kept constant. Although, for a future development of a free-motion acquisition system, these type of motions need to be well detected and compensated. It would be possible for physicians to simple move the camera around the structure of interest and get a complete and precise 3D model. Then, the reconstructed models can be refined by applying non-rigid transformations, which deform the point clouds being registered towards the optimal solution.

Finally, a module responsible for the creation of a mesh from the registered model could be implemented. However, that should not be a problem because the number of points belonging to each point cloud being registered is high (and the clouds so dense) that by now, a simple triangulation would reconstruct the mesh with good resolution.

6.3 Performance Analysis

The performance evaluation of the methodologies on the registration of the male head and the female torso is shown in Table 6.2. In both cases, the proposed methodology takes significantly longer to coarsely align the views being registered. The increased processing time is caused by the higher number of points that exist in the non-rigid models, when compared to the rigid ones.

Table 6.2: The performance evaluation of the different tested methodologies on the registration of the male head.

Model	**Method**	**Time(s)** $\mu(\pm\sigma)$**/view**		**Iterations** $\mu(\pm\sigma)$**/view**
		Coarse Reg.	*Fine Reg.*	*Fine Reg.*
Male Head	ICP	-	72.956(±13.166)	300(±0)
	PCA+ICP	0.068(±0.011)	74.794(±13.341)	300(±0)
	Tessellation + ICP	21.278(±3.890)	76.486(±13.304)	200(±4.874)
Female Torso	ICP	-	153.253(±4.080)	300(±0)
	PCA+ICP	0.142(±0.003)	154.404(±4.309)	300(±0)
	Tessellation + ICP	33.485(±1.086)	155.430(±2.703)	300(±0)

However taking much time, this coarse registration stage provides a good initial estimation for the fine registration steps, thus making the reconstruction easier and more accurate. The fine registration stage needs a lower number of iterations to finish the reconstruction. In fact, the necessary trade-off between computational time and quality of the reconstruction is acceptable: a high-quality reconstruction is performed in feasible time (less than 3 minutes).

6.4 Final Considerations

The developed methodology is highly versatile, capable of performing a good registration of both the rigid data and different kinds of non-rigid data, such as the male head and the female torso. Although some improvements are still required, the system provides good-quality reconstructions using information obtained using a low-cost RGB-D camera, even under noisy conditions. Using 3 views and spending less than 3 minutes, it was possible to obtain good reconstructions of body parts.

Regarding the proposed evaluation metrics, the most consistent results were obtained for the integral of a polar difference function, which proved to be, not only pose invariant, but also model invariant. The metrics used to evaluate the histograms of distances obtained from the Delaunay Triangulation of both the reconstructed and the reference models are model dependent and need to be analyzed as a whole. This is because some inconsistencies can be found: for instance, sometimes very bad reconstructions lead to good numeric results for one of the three descriptors. Nonetheless, if analyzed together, they describe fairly good the quality of the reconstruction.

The major drawback of the developed metrics is that they do not provide any insight on the accuracy and resolution of the reconstruction methodologies. Nevertheless, they can be used to perform a robust relative comparison between the registration methodologies. In addition, it would be interesting to study if these metrics can be used to determine the stopping criterion for the fine registration algorithms or even guide the reconstruction algorithms.

Chapter 7

Conclusions

The 3D reconstruction of the female breast can be used to provide volumetric information to evaluate the outcome of BCCT. Its main goal is to characterize body structures to help the breast characterization and the medical intervention. The available tools for this purpose are either expensive or complex, which avoids their use by physicians in a medical environment. This created the need for the development of a practical, low-cost, easy to use and accurate system.

The developed application must rely on a computer vision algorithm that uses images taken from different perspectives of the object/body by a 3D sensing system. It must combine accuracy and simplicity, while being affordable and safe to use. As structured light sensors allow the use of a wide variety of patterns and the possibility of a dynamic pattern change, they are especially versatile and flexible, being a desirable tool on the acquisition step. Low-cost sensors have been gaining importance in the medical field. Some of them are also referred as being RGB-D cameras because they provide both color and depth information for the reconstruction. These cameras are often easy to use and transport, which is a relevant condition for its use in a medical environment.

The Microsoft Kinect is a RGB-D camera capable of operating under any interior lighting conditions and is compact and light. Resolutions of almost 1 mm are possible, it has available free software for Kinect data processing and has been used in the past few years with great success in different modeling applications.

In this thesis, a Kinect Based System for 3D Breast Modeling was developed. This system has three major stages of processing: (a) RGB-D data acquisition and point cloud generation; (b) Coarse Registration; (c) Fine Registration.

Regarding the topic (b), a new method that uses the Delaunay Triangulation principle as the keypoint selection criterion is here proposed. The keypoints were defined as the most prominent parts of the point clouds, as expected. The Tessellation-based coarse registration algorithm proved to be capable of performing a good initial estimation of the alignment between the point clouds. This causes the fine registration stage to converge better towards the best solution possible. In addition, the number of iterations needed to reach the reconstructed model decreases.

This good coarse alignment between clouds allowed the generation of high-quality reconstructed models of the body parts of interest, using only 3 views and taking less than 3 minutes.

Such results prove that the proposed methodology can have great relevance for 3D breast reconstruction purposes, especially because distinctive structures such as the breast peak, the nipple or the breast contour were correctly enhanced and reconstructed using an easy registration pipeline.

Two new evaluation metrics are proposed in this thesis. They provide reliable relative comparison between the performance of different algorithms on the reconstruction. However, they fail to provide some insight about the accuracy of the reconstruction. Some improvements must be done to characterize the resolution of the proposed registration pipeline and to increase the descriptive power of the metrics.

7.1 Future Work

Despite being good, the obtained results are not optimal. This creates the need for further developments, which can be used, not only to refine the reconstructed models, but also to create an intuitive registration tool for physicians.

As said before, the future work must be related with the compensation of the random motion of the subject during the acquisition stage. In other words, the fine registration algorithm must be modified to incorporate scale and skew transformations while registering the coarsely aligned views. Then, the development of new non-rigid registration methodologies is essential. These methods should introduce some shape knowledge into the registration pipeline. In other words, what we know about the body structures being reconstructed can be used to guide the registration algorithm.

The registration of each view must be followed by a denoising stage. If the effect of the accumulated noise is minimized, it would be possible to perform the registration of non-rigid data using a higher number of views, thus improving the detail of the reconstructed models.

Some other improvement can be implemented, aiming the best operational flexibility possible. Two of them are: (a) a module that suggests which views should be used for the registration; (b) a module that allows the user to segment a region of interest on the point clouds being registered.

The first one can be done by partitioning the set of views that contains relevant and new information, and taking one view from each partition. Regarding the second one, it would allow the registration pipeline to focus more on the details and less on the global shape of the point clouds, as proved by the second experiment on the reconstruction of the female torso.

Bibliography

D. Aiger, N. J. Mitra, and D. Cohen-Or, "4-points congruent sets for robust pairwise surface registration," in *ACM Transactions on Graphics*, vol. 27, no. 3. ACM, 2008, p. 85.

M. R. Andersen, T. Jensen, P. Lisouski, A. K. Mortensen, M. K. Hansen, T. Gregersen, and P. Ahrendt, "Kinect depth sensor evaluation for computer vision applications," Århus Universitet, Report, 2012.

D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "Scape: shape completion and animation of people," in *ACM Transactions on Graphics (TOG)*, vol. 24. ACM, 2005, Conference Proceedings, pp. 408–416.

J. Batlle, E. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," *Pattern recognition*, vol. 31, no. 7, pp. 963–982, 1998.

S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

J.-A. Beraldin, F. Blais, L. Cournoyer, G. Godin, M. Rioux, and J. Taylor, "Active 3d sensing," 2003.

P. J. Besl, "Active, optical range imaging sensors," *Machine vision and applications*, vol. 1, no. 2, pp. 127–152, 1988.

P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Robotics-DL tentative*. International Society for Optics and Photonics, 1992, Conference Proceedings, pp. 586–606.

J. Böhm, "Natural user interface sensors for human body measurement," *International Archive of Photogrammetry Remote Sensing and Spatial Information Science*, vol. 39, p. B3, 2012.

G. Bickel, G. Hausler, and M. Maul, "Triangulation with expanded range of depth," *Optical Engineering*, vol. 24, no. 6, 1985.

F. Blais and M. Rioux, "Biris: a simple 3-d sensor," in *Cambridge Symposium on Intelligent Robotics Systems*. International Society for Optics and Photonics, 1987, Conference Proceedings, pp. 235–242.

F. Blais, M. Rioux, and J.-A. Beraldin, "Practical considerations for a design of a high precision 3-d laser scanner system," in *Dearborn Symposium*. International Society for Optics and Photonics, 1988, Conference Proceedings, pp. 225–246.

F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, vol. 13, no. 1, 2004.

M. A. Brunsman, H. M. Daanen, and K. M. Robinette, "Optimal postures and positioning for human body scanning," in *Proceedings of the International Conference on Recent Advances in3-D Digital Imaging and Modeling*. IEEE, 1997, Conference Proceedings, pp. 266–273.

J. S. Cardoso and M. J. Cardoso, "Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment," *Artificial Intelligence in Medicine*, vol. 40, no. 2, pp. 115–126, 2007.

M. J. Cardoso, H. Oliveira, and J. Cardoso, "Assessing cosmetic results after breast conserving surgery," *Journal of surgical oncology*, vol. 110, no. 1, pp. 37–44, 2014.

U. Castellani and A. Bartoli, *3D Shape Registration*. Springer, 2012, pp. 221–264.

Y.-J. Chang, S.-F. Chen, and J.-D. Huang, "A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities," *Research in developmental disabilities*, vol. 32, no. 6, pp. 2566–2570, 2011.

C.-I. Chen, D. Sargent, and Y.-F. Wang, "Modeling tumor/polyp/lesion structure in 3d for computer-aided diagnosis in colonoscopy," in *Proceedings of the SPIE*, vol. 7625, 2010, Conference Proceedings, pp. 76 252F–1.

F. Chen, G. M. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," *Optical Engineering*, vol. 39, no. 1, pp. 10–22, 2000.

Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.

C. K. Chow, H. T. Tsui, and T. Lee, "Surface registration using a dynamic genetic algorithm," *Pattern recognition*, vol. 37, no. 1, pp. 105–117, 2004.

C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 63–85, 1997.

D. H. Chung, I. D. Yun, and S. U. Lee, "Registration of multiple-range views using the reverse-calibration technique," *Pattern Recognition*, vol. 31, no. 4, pp. 457–464, 1998.

P. Cignoni, D. Laforenza, R. Perego, R. Scopigno, and C. Montani, "Evaluation of parallelization strategies for an incremental delaunay triangulator in e3," *Concurrency: Practice and Experience*, vol. 7, no. 1, pp. 61–80, 1995.

L. Cruz, D. Lucio, and L. Velho, "Kinect and rgbd images: Challenges and applications," in *25th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*. IEEE, 2012, Conference Proceedings, pp. 36–49.

Y. Cui and D. Stricker, "3d shape scanning with a kinect," in *ACM SIGGRAPH 2011 Posters*. ACM, 2013, Conference Proceedings, p. 57.

B. Delaunay, "Sur la sphere vide," *Izvestiya Akademia Journal of Mathematical and Natural Sciences*, vol. 7, no. 793-800, pp. 1–2, 1934.

H. Edelsbrunner and R. Seidel, "Voronoi diagrams and arrangements," *Discrete & Computational Geometry*, vol. 1, no. 1, pp. 25–44, 1986.

J. Feldmar and N. Ayache, "Rigid, affine and locally affine registration of free-form surfaces," *International journal of computer vision*, vol. 18, no. 2, pp. 99–119, 1996.

M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (tof) cameras: a survey," *Sensors Journal, IEEE*, vol. 11, no. 9, pp. 1917–1926, 2011.

J. Gallier, "Notes on convex sets, polytopes, polyhedra, combinatorial topology, voronoi diagrams and delaunay triangulations," 2008.

N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust global registration." in *in the Proceedings of the Symposium on geometry processing*, vol. 2, no. 3, 2005, p. 5.

P.-L. George and H. Borouchaki, "Delaunay triangulation and meshing: application to finite elements," 1998.

G. Godin, M. Rioux, and R. Baribeau, "Three-dimensional registration using range and intensity information," in *Photonics for Industrial Applications*. International Society for Optics and Photonics, 1994, pp. 279–290.

J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Transactions on Cybernetics*, 2013.

R. M. Haralick and L. G. Shapiro, *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc., 1991.

P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments," in *Proceedings of the 12th International Symposium on Experimental Robotics*, vol. 20, 2010, Conference Proceedings, pp. 22–25.

——, "Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.

H. T. Ho, "3d surface matching from range images using multiscale local features," Ph.D. dissertation, The University of Adelaide Australia, 2009.

B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987.

J.-D. Huang, "Kinerehab: a kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities," in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2011, Conference Proceedings, pp. 319–320.

R. S. Hunter, "Photoelectric color difference meter," *Journal of The Optical Society of America A*, vol. 48, no. 12, pp. 985–993, 1958.

D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

K. Iwami and N. Umeda, "5. rapid prototyping in biomedical engineering," 2011.

S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, and A. Davison, "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2012, Conference Proceedings, pp. 559–568.

R. Jarvis, "A perspective on range finding techniques for computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 122–139, 1983.

A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: a cancer journal for clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

A. E. Johnson, "Spin-images: a representation for 3-d surface matching," Ph.D. dissertation, Citeseer, 1997.

T. Jost and H. Hugli, "A multi-resolution icp with heuristic closest point search for fast and robust 3d registration of range images," in *in the proceedings of the 4th International Conference on3D Digital Imaging and Modeling*. IEEE, 2003, pp. 427–433.

T. Kanade and H. Asada, "Noncontact visual three-dimensional ranging devices," in *1981 Technical Symposium East*. International Society for Optics and Photonics, 1981, Conference Proceedings, pp. 48–55.

C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," *ACM Transactions in Graphics (TOG)*, 2013.

P. Koehl, "Protein structure similarities," *Current opinion in structural biology*, vol. 11, no. 3, pp. 348–353, 2001.

A. Kolb, E. Barth, and R. Koch, "Tof-sensors: New dimensions for realism and interactivity," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2008, Conference Proceedings, pp. 1–6.

K. Konolige, "Projected texture stereo," in *IEEE International Conference on Robotics and Automation*. IEEE, 2010, Conference Proceedings, pp. 148–155.

R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *Quantum Electronics, IEEE Journal of*, vol. 37, no. 3, pp. 390–397, 2001.

H. Li, E. Vouga, A. Gudym, L. Luo, J. T. Barron, and G. Gusev, "3d self-portraits," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 187, 2013.

S. K. Mada, M. L. Smith, L. N. Smith, and P. S. Midha, "Overview of passive and active vision techniques for hand-held 3d data acquistion," in *Opto Ireland*. International Society for Optics and Photonics, 2003, Conference Proceedings, pp. 16–27.

T. Mallick, P. Das, and A. Majumdar, "Characterizations of noise in kinect depth images," *IEEE Sensors*, vol. 14, no. 6, pp. 1731–1740, 2014.

D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, 2004.

T. Masuda, "Generation of geometric model by registration and integration of multiple range images," in *in the proceedings of the 3rd International Conference on 3D Digital Imaging and Modeling*. IEEE, 2001, pp. 254–261.

A. S. Mian, M. Bennamoun, and R. A. Owens, "A novel representation and feature matching algorithm for automatic pairwise registration of range images," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 19–40, 2006.

P. R. Minchin, "An evaluation of the relative robustness of techniques for ecological ordination," in *Theory and models in vegetation science*. Springer, 1987, pp. 89–107.

T. Moons, L. Van Gool, and M. Vergauwen, *3d Reconstruction from Multiple Images: Part 1: Principles*. Now Publishers Inc, 2010.

D. M. Mount, "Computational geometry notes," 2005.

R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *10th IEEE international symposium on Mixed and augmented reality (ISMAR)*. IEEE, 2011, Conference Proceedings.

P. Newman, G. Sibley, M. Smith, M. Cummins, A. Harrison, C. Mei, I. Posner, R. Shade, D. Schroeter, and L. Murphy, "Navigating, recognizing and describing urban spaces with vision and lasers," *The International Journal of Robotics Research*, vol. 28, no. 11-12, pp. 1406–1433, 2009.

T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc, "An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution (swissranger)," in *Optical Systems Design*. International Society for Optics and Photonics, 2004, Conference Proceedings, pp. 534–545.

H. Oliveira, "An affordable and practical 3d solution for the aesthetic evaluation of breast cancer conservative treatment," Ph.D. dissertation, Faculty of Engineering of the University of Porto, Portugal, 2013.

H. P. Oliveira, J. S. Cardoso, A. Magalhães, and M. J. Cardoso, "Methods for the aesthetic evaluation of breast cancer conservation treatment: a technological review," *Current Medical Imaging Reviews*, vol. 9, no. 1, pp. 32–46, 2013.

H. P. Oliveira, J. S. Cardoso, A. T. Magalhães, and M. J. Cardoso, "A 3d low-cost solution for the aesthetic evaluation of breast cancer conservative treatment," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 2, no. 2, pp. 90–106, 2014.

H. P. Oliveira, P. Patete, G. Baroni, and J. S. Cardoso, "Development of a bcct quantitative 3d evaluation system through low-cost solutions," in *Proceedings of the 2nd International Conference on 3D body Scanning Technologies*, 2011, Conference Proceedings, pp. 16–27.

O. Pele and M. Werman, "Fast and robust earth mover's distances," in *12th IEEE International Conference on Computer vision*. IEEE, 2009, pp. 460–467.

——, "The quadratic-chi histogram distance family," in *European Conference on Computer Vision*. Springer, 2010, pp. 749–762.

G. F. Poggio and T. Poggio, "The analysis of stereopsis," *Annual review of neuroscience*, vol. 7, no. 1, pp. 379–412, 1984.

F. Remondino and S. El Hakim, "Image based 3d modelling a review," *The Photogrammetric Record*, vol. 21, no. 115, pp. 269–291, 2006.

M. Rioux, "Laser range finder based on synchronized scanners," *Applied Optics*, vol. 23, no. 21, pp. 3837–3844, 1984.

M. Rioux and F. Blais, "Compact three-dimensional camera for robotic applications," *Journal of the Optical Society of America A*, vol. 3, no. 9, pp. 1518–1521, 1986.

E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *10th IEEE International Conference on Computer Vision, 2005. ICCV 2005*, vol. 2. IEEE, 2005, Conference Proceedings, pp. 1508–1515.

C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM Transactions on Graphics*, vol. 23, no. 3. ACM, 2004, pp. 309–314.

Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

O. Rusch, C. Ruwwe, and U. Zolzer, "An evaluation of feature matching algorithms for maritime images," in *Proceedings of the 6th IASTED International Conference on Visualization, Imaging, And Image Processing*, 2006, Conference Proceedings.

R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 3212–3217.

J. Salvi, C. Matabosch, D. Fofi, and J. Forest, "A review of recent range image registration methods with accuracy evaluation," *Image and Vision Computing*, vol. 25, no. 5, pp. 578–596, 2007.

G. Sansoni, M. Trebeschi, and F. Docchio, "State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation," *Sensors*, vol. 9, no. 1, pp. 568–601, 2009.

J. Santamaría, O. Cordón, and S. Damas, "A comparative study of state-of-the-art evolutionary image registration methods for 3d modeling," *Computer Vision and Image Understanding*, vol. 115, no. 9, pp. 1340–1354, 2011.

M. Schmeing and X. Jiang, "Edge-aware depth image filtering using color segmentation," *Pattern Recognition Letters: Special Issue on Depth Image Analysis (to appear).*, 2014.

G. C. Sharp, S. W. Lee, and D. K. Wehe, "Icp registration using invariant features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 90–102, 2002.

N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," in *ACM transactions on graphics (TOG)*, vol. 25. ACM, 2006, Conference Proceedings, pp. 835–846.

J. Sturm, E. Bylow, F. Kahl, and D. Cremers, *CopyMe3D: Scanning and Printing Persons in 3D*. Springer, 2013, pp. 405–414.

Q. Sun, Y. Tang, P. Hu, and J. Peng, "Kinect-based automatic 3d high-resolution face modeling," in *International Conference on Image Analysis and Signal Processing (IASP)*. IEEE, 2012, Conference Proceedings, pp. 1–4.

R. Szeliski, *Computer vision: algorithms and applications*. Springer, 2011.

J.-P. Tarel, H. Civi, and D. B. Cooper, "Pose estimation of free-form 3d objects without point matching using algebraic surface models," in *IEEE Workshop Model Based 3D Image Analysis*, 1998, pp. 13–21.

H. Tiziani, "Optical metrology of engineering surfaces-scope and trends," *Optical measurement techniques and applications*, pp. 15–49, 1997.

R. Toldo, A. Beinat, and F. Crosilla, "Global registration of multiple point clouds embedding the generalized procrustes analysis into an icp framework," in *in the Proceedings of the 3DPTV Conference*, 2010.

J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3d full human bodies using kinects," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 643–650, 2012.

C. Torre-Ferrero, J. R. Llata, L. Alonso, S. Robla, and E. G. Sarabia, "3d point cloud registration based on a purpose-designed similarity measure," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–15, 2012.

P. Vuylsteke and A. Oosterlinck, "3-d perception with a single binary coded illumination pattern," in *Cambridge Symposium on Intelligent Robotics Systems*. International Society for Optics and Photonics, 1987, Conference Proceedings, pp. 195–202.

R. Wang, J. Choi, and G. Medioni, "Accurate full body scanning from a single fixed 3d camera," in *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*. IEEE, 2012, Conference Proceedings, pp. 432–439.

T. W. Way, L. M. Hadjiiski, B. Sahiner, H.-P. Chan, P. N. Cascade, E. A. Kazerooni, N. Bogot, and C. Zhou, "Computer-aided diagnosis of pulmonary nodules on ct scans: segmentation and classification using 3d active contours," *Medical Physics*, vol. 33, p. 2323, 2006.

J. W. Weingarten, G. Gruener, and R. Siegwart, "A state-of-the-art 3d sensor for robot navigation," in *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3. IEEE, 2004, Conference Proceedings, pp. 2155–2160.

A. Weiss, D. Hirshberg, and M. J. Black, "Home 3d body scans from noisy image and range data," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2012, Conference Proceedings, pp. 1951–1958.

P. Xi, W.-S. Lee, and C. Shu, "Analysis of segmented human body scans," in *Proceedings of Graphics Interface 2007*. ACM, Conference Proceedings, pp. 19–26.

H. Yang, K. Boyer, and A. Kak, "Range data extraction and interpretation by structured light," in *Proceedings of the 1st IEEE Conference on Artificial Intelligence Applications, Denver, Colorado*, 1984, Conference Proceedings, pp. 199–205.

P. Yang and X. Qian, "Direct computing of surface curvatures for point-set surfaces." in *in the Proceedings of the Eurographics Symposium on Point-Based Graphics*, 2007, pp. 29–36.

S.-W. Yang, C.-C. Wang, and C.-H. Chang, "Ransac matching: Simultaneous registration and segmentation," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2010, Conference Proceedings, pp. 1905–1912.

Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *International journal of computer vision*, vol. 13, no. 2, pp. 119–152, 1994.

X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sensing of Environment*, vol. 114, no. 11, pp. 2610–2623, 2010.

B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

M. Zollhöfer, M. Martinek, G. Greiner, M. Stamminger, and J. Süßmuth, "Automatic reconstruction of personalized avatars from 3d face scans," *Computer Animation and Virtual Worlds*, vol. 22, no. 3, pp. 195–202, 2011.

Appendix A

Acquisition Protocol

This section presents the image acquisition protocol that was used to obtain data from female patients. Briefly, the subject being recorded is asked to perform a 180 degree rotation in the most stable way possible while color and depth images are captured using the Microsoft Kinect.

PICTURE – IMAGE ACQUISITION PROTOCOL
MICROSOFT KINECT – 3DMK

Background

- A **neutral background** should be used to prevent reflections from influencing the patient's skin colour (**Light blue**).

Camera Mount

- Camera should be mounted on a tripod at **~90cm** from the subject.
- **Camera height**: mounted to prevent patient identification (below the neck).

Patient Positioning

- The subject positioned **without jewellery or clothing**.
- **Hands on hips** to prevent obstruction of the lateral view.

Image Acquisition Layout

- Images will be acquired **continuously for a full 180º rotation** between lateral views, performed as **smoothly** as the patient is able (**from left to right and left to right**). (see Figure – blue feet).

Specifications

- Computer Windows 7 or higher.
- 8GB Ram
- Hard Disk with 6000 rpm

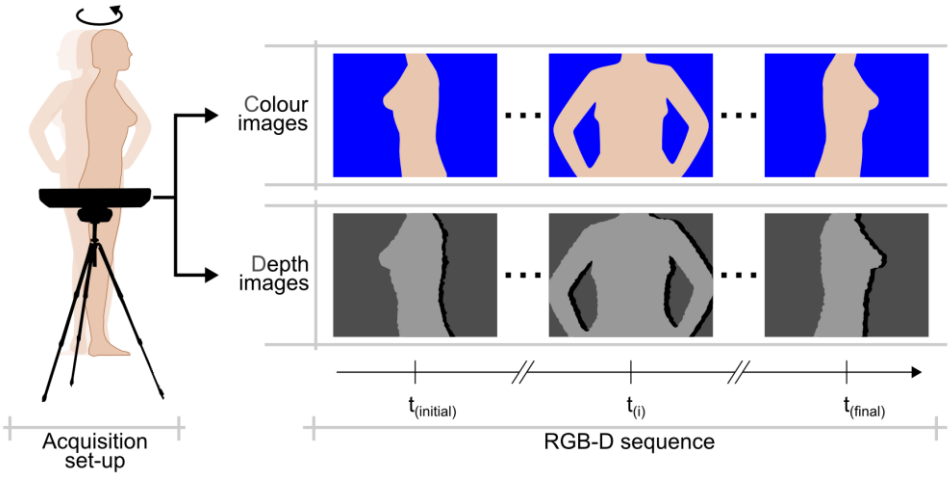


Figure A.1: RGB-D image acquisition protocol using the Microsoft Kinect.