FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

A Rigid 3D registration framework of women body RGB-D images

António Bastos Pintor

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Hélder P. Oliveira

Co-Supervisor: João P. Monteiro

July 29, 2016

© António Bastos Pintor, 2016

Abstract

Breast cancer is one of the most aggressive diseases, affecting thousands of women every year. Usually, this kind of patients require invasive treatments, such as surgery and radiation therapy, leading to undesired physical and psychological secondary effects. At an aesthetic level, there is a demand for solutions that help processing clinic evaluations before and after the procedures, since the commonly used subjective approach is based on photography. This approach may lead to non-optimal decisions and consequently unpleasant results. For this reason, 3D body model reconstruction has been shown as a practical solution in order to facilitate the doctor's work on observing a patient torso. The existent equipments in the market, used for 3D modeling, exhibit high-costs and normally require special knowledge behind them, making the task of applying them in a regular basis for physicians, a very complex job. Over the last 6 years, since the release of the Microsoft Kinect, there has been a growth of studies concerning low-cost RGB-D cameras. Because of this, developers have seen the opportunity to implement inexpensive and simpler solutions for 3D modeling in order to use them in a myriad of applications, specially in medicine. The creation of 3D models from human bodies demands algorithms to get minimal errors in the registration of the point clouds. These errors are a recurring problem, due to the non-rigidity of the human body that is captured during the acquisition process.

This thesis relies on improving and automatizing the framework that has been developed for the project PICTURE from the VCMI group of INESC-TEC, with the purpose of creating an inexpensive and easy to use 3D model system, for medical analysis on breast cancer patients, without any special knowledge. This work proposes a few improvements for the given framework, such as: automatic body pose selection, automatic segmentation process of rigid body parts, 3D data processing for noise reduction and a study of rigid registration methods for multiple clouds points is done.

The results have shown some optimistic improvements from the previous framework, where the reconstruction of the patient's 3D models with data from low-cost RGB-D cameras were achieved with low distortion errors in comparison with models from an high-end 3D modeling System such as the 3dMD. Additionally the framework automation was accomplished for the selection of the patient's pose and the extraction of rigid body parts.

Keywords: 3D Modeling and Reconstruction; Rigid Registration; Low-cost cameras; RGB-D sensors; Breast Cancer.

Resumo

O cancro da mama é das doenças mais agressivas que afeta milhares de mulheres todos os anos. Geralmente, este tipo de doente necessita de tratamentos agressivos, como cirurgia e radioterapia levando a efeitos secundários indesejáveis tanto físicos como psicológicos. A nível estético existe uma procura recorrente de soluções que ajudem a processar avaliações clínicas, antes e depois, de procedimentos comuns como abordagem de avaliação subjetiva baseada em fotografia. Esta abordagem pode conduzir a decisões menos exatas e consequentemente a resultados desagradáveis. Por esta razão o modelação 3D para reconstrução do corpo humano tem sido apresentada como uma solução prática, que facilita o trabalho dos médicos durante a observação do torso da paciente. Os equipamentos existentes no mercado para modelação 3D, têm custos elevados e, normalmente, exigem conhecimentos prévios sobre estes. Tornando a sua aplicação regular uma tarefa complexa para o médico. Nos últimos 6 anos, desde o lançamento do Microsoft Kinect, houve um aumento dos estudos sobre câmaras RGB-D de baixo custo. Devido a isto, investigadores têm visto a oportunidade para implementar soluções menos dispendiosas e mais simples em modelação 3D para as usa em inúmeras aplicações, tal como na medicina. A criação de modelos 3D do corpo humano exige algoritmos para minimizar a obtenção de erros no registo das nuvens de pontos. Estes erros são um problema recorrente devido à falta de rigidez do corpo humano que são observados durante o processo de aquisição.

Esta tese assente na melhoria e automatização de uma ferramenta que tem sido desenvolvida para o projeto PICTURE do grupo VCMI, do INESC-TEC, com o objetivo de criar um sistema de modelação reconstrução 3D para a análise médica de cancro da mama de pacientes sem necessidade de conhecimentos adicionais. Este trabalho propõe alguns melhoramentos da ferramenta desenvolvida como: seleção automática da posição do corpo, processo automático de segmentação das partes rígidas do corpo, processamento de dados 3D para redução do ruído e um estudo sobre métodos de registo rígido de múltiplas nuvens de pontos.

Os resultados mostraram alguns melhoramentos em comparação com a ferramenta anterior, onde a reconstrução dos modelos 3D das pacientes, a partir de dados adquiridos das câmaras RGB-D de baixo custo, apresentam erros baixos de distorção, em comparação com modelos de uma sistema topo como o 3dMD. Adicionalmente a automatização da ferramenta foi conseguida na seleção da posição do paciente e a extração de partes rígidas do corpo.

Palavras-chave: Modelação e reconstrução 3D; Registo Rígido; Câmeras de baixo-custo; Sensores RGB-D; Cancro da mama.

Agradecimentos

Ao longo destes cinco anos de curso cruzei-me com pessoas espectaculares que contribuiram para o desenvolvimento das minhas competências e para uma vida académica mais agradável. Obrigado Miguel, Hugo F., André C., Ricardo, Isabel, Júnio, Francisco, Vilhena, Rita, Diogo, Rafa, Zé Pedro, André R., Hugo S. e Álvaro, agradeço por todas as experências que partilhamos.

Quero agradecer a todo o grupo do VCMI pela amizade, pela demonstração do grande espirito de grupo que possuem e por estarem sempre prontos para ajudar. Obrigado Ricardo, João T., Pedro, Kelwin, Eduardo, Ana, Hooshiar e aos meus orientadores João Monteiro e Hélder Oliveira.

Por fim, para a minha família, não existe maneira de agradecer o suficiente por todo o apoio que me deram. Ninguém me conhece melhor que os meus pais, irmãos e avós e desde que comecei a estudar eles nunca duvidaram das minhas capacidades. Graças a isto, não custou tanto chegar até aqui.

Desejo dedicar este meu trabalho aos meus pais. Ao meu pai, pois sem ele nunca iría para engenharia pelo apoio incondicional que demonstrou desde sempre. Fez-me ver o mundo de uma forma diferente e ensinou-me importantes lições de vida. À minha mãe, agradeço por todo o seu esforço por me proporcionar uma vida feliz e tornar-me naquilo que sou hoje.

A todos o meu Muito Obrigado!

António Pintor

"In the end it is not the years in your life that count,
it's the life in your years."

Abraham Lincoln

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Bi-dimensional
3D	Tri-dimensional
API	Application Programming Interface
BCCT	Breast Cancer Conservative Treatment
CMOS	Complementary Metal-Oxide Semiconductor
FPS	Frames Per Second
HD	High Definition
Hz	Hertz
ICL	Iterative Closest Line
ICP	Iterative Closest Point
ICT	Iterative Closest Triangle patch
IMF	Infra-Mammary Fold
INESC-TEC	INstituto de Engenharia de Sistemas e Computadores - Tecnologia e Ciência
IR	Infrared
NITE	Natural Interaction Technology for End-User
OEM	Original Equipment Manufacturer
PCA	Principal Component Analysis
PCL	Point Cloud Library
PNG	Portable Network Graphics
RANSAC	Random Sample Consensus
RGB-D	Red Green Blue and Depth
SDK	Software Development Kit
TOF	Time Of Flight
USB	Universal Serial Bus
VGA	Video Graphics Array

Chapter 1

Introduction

1.1 Background

In the age of technological evolution, particularly in computer vision, there has been a huge search for the development of tools for 3D modeling and reconstruction of objects and scenarios, with perspectives of great potential. In the level of applicability, we can find several areas of research such as: robotics, military, quality control, 3D printing, virtual reality, multimedia, computational animation and medicine [1].

In the world of medicine there is a constant demand for new techniques and tools, with the requirements of being more efficient, non expensive and with an easy applicability in many different medical specialties, going from the detection of pathologies to surgery.

Breast cancer is a disease with a large incidence on women, with a high aesthetic and psychological impact. Its aggressiveness implies treatments that lead, in many cases, to surgery such as mastectomy or *Breast Cancer Conservative Treatment* (BCCT), in other words, the local removal of the tumor [2].

In clinical breast cancer evaluations, parameters such as aesthetic, color, geometry, volume, profile and symmetry before and after treatment or surgery, traditionally resumed to photography, imply a subjective evaluation from one or more observers.

The introduction of new technologies like 3D cameras, to create complete 3D models of the torso of the woman, have allowed to estimate and do a quantitative analysis of the characteristics of the breast, helping the doctors planning their treatments, such as surgery for the patient [3]. Clinical routine needs practical tools. Much of the equipment that exists in the market has high costs, lack of portability, a complexity in its use due to the requirement of special knowledge and an increase of expenditure of hiring personnel [4].

3D modelling is considered as a field in development. Diverse organizations have presented solutions to the production of software tools and explore new techniques. Due to these organizations it is possible to obtain inexpensive 3D cameras, providing free software tools with a wide shared documentation within a community that is interested in the look for innovation in the branch of computer vision [5].

1.2 Motivation

Currently, there are scenarios where a surgeon would benefit from a tool which allows to share information with the medical team and show how some parameters could affect the aesthetic outcome of the surgery. From the point of view of the patient, it would be more interesting to get a more comprehensible and illustrative way to understand the conditions of a surgery, before and after, in order to observe its impact.

The existent solutions available on the market have become expensive, require specialized personnel and only operate in controlled environments. The motivation for this research arises from the need of finding solutions for the breast cancer area with the introduction of modeling technologies and 3D reconstruction. Those solutions should comprise simple and low-cost systems, to be used in aesthetic evaluation and surgery planning, by any healthcare professional without requiring any extra knowledge.

Taking into account that it is intended to use a non usual equipment in health, it shows how this is a major challenge to develop algorithms that build realistic 3D models and complete from several 'views' shot of a non-rigid object, that varies its shape in time. The human body is prone to natural involuntary movements such as breathing, leading to errors in data readings very often, making the whole process harder for these systems.

1.3 Objectives

The VCMI group from the INESC-TEC, has been looking for solutions, as part of the European research project PICTURE [6], for the development of a framework to be integrated in 3D model reconstruction system, for breast cancer analysis and treatment, with the goal of using simple and low-cost RGB-D sensors. This sensors work by collecting data through a RGB camera, in order to get several photographs and simultaneously with a depth sensor capable of getting information on the distances from the surfaces of an object of interest. After mapping and processing all the values from the collected data, this will finally result in point clouds that will be used for the registration and generation of unique complete 3D model of a woman torso.

The main objectives of this thesis focus on improving the given framework and solve its main issues as for the automation, accuracy and efficiency, in order to make it more independent of the acquisition technology and user input.

1.4 Contributions

The main contributions that have resulted in this thesis are the following:

- Body Pose Selection - Automatize the process of selecting the main views of frontal, left and right poses;
- Segmentation of rigid body parts - Automatize the segmentation of the rigid body parts to improve the rigid registration process;

- Preliminary processing - Implement a 3D data processing module, to improve the appearance of the point clouds by smoothing the surface and remove outlier points which result from the noise of the acquisition data with the Microsoft Kinect;
- Rigid Registration - Improve the accuracy of the algorithm so it can handle the difficulties of non rigidity from the human body and its variability in time;

1.5 Document Structure

In addition to the introduction, this dissertation contains five more chapters. For the chapter 2, it is made a bibliographic revision and an overview on related works. In the chapter 3, some context about the previous work done for the project 'PICTURE' is given, in order to explain the purpose of this thesis and where it is necessary to intervene in the framework. In chapter 4, it is explained in detail the methodology followed for each step of the framework that was developed. In the following chapter 5 results and discussion according with the validation. Finally, in chapter 6 a conclusion of this work is made.

Chapter 2

Literature Revision

In this chapter, it is discussed the state of the art related with the purpose of the thesis. It reviews what exists today and the different approaches that can be found on 3D imaging techniques, including some examples of devices which are available on the market for consumer use. Also, a few studies on their applicability in 3D modeling and body pose selection and identification are made. Next, a review about rigid registration on point clouds is done, followed by some new techniques that have been recently proposed.

2.1 3D imaging techniques

A lot of new techniques for 3D imaging have been developed over the past few decades. Taking into account the variability and complexity of different situations, not every sensor and technique may be appropriated for every kind of application. At the moment it can be said that, for all kinds of objects and scenarios, there is no single modelling technique able to satisfy all the requirements of high geometric accuracy, portability, full automation, photo-realism, low-cost, flexibility and efficiency [7].

A good part of current sensors operate primarily in two forms. First the active form, which act by projecting controlled light onto objects followed by recording the respective reflections, which will contain the information about the shape. The second is known as the passive form, where the main point is acquiring energy from an object that is being transmitted [8]. In both forms the outcome is dependent of the surrounding environment [9].

In general 3D imaging techniques can be grouped into three types of technological methods as follows: optical triangulation, time delay and the use of monocular images. For the measurements it is possible to get them with direct techniques, where they result in range data, creating a relation with distances between the surface and the sensor. Indirect measures are obtained from techniques based on monocular images or by the usage of prior knowledge about the surfaces properties [8].

The next table 2.1, from the 2009 Sansoni's review [8], classifies the variety of non-contact and optical methods used by nowadays hardware.

Table 2.1: A classification of 3D imaging techniques (from [8]) .

	Triangulation	Time delay	Monocular Images	Passive	Active	Direct	Indirect	Range	Surface Orientation
Laser triangulators	X				X	X		X	
Structured light	X				X	X		X	
Stereo vision	X			X		X		X	
Photogrammetry	X			X		X		X	
Time of Flight		X			X	X		X	
Interferometry		X			X	X		X	
Moiré fringe range contours			X		X		X	X	
Shape from focusing			X	X	X		X	X	
Shape from shadows			X		X		X	X	
Texture gradients			X	X			X		X
Shape from shading			X		X		X		X
Shape from photometry			X		X		X		X

In the next sections, it is presented an overview of three very common techniques that had a major impact in several of today's well known 3D information acquiring equipment [10].

2.1.1 Structured light

The structured light sensors project bi-dimensional patterns of non-coherent light to the scene. As the structure of the projected pattern is known, the object depth map can be reconstructed by looking at the differences between the projected and the recorded pattern. The projected pattern is deformed by the object and can be used to describe its structure orientation or texture [11] [9].

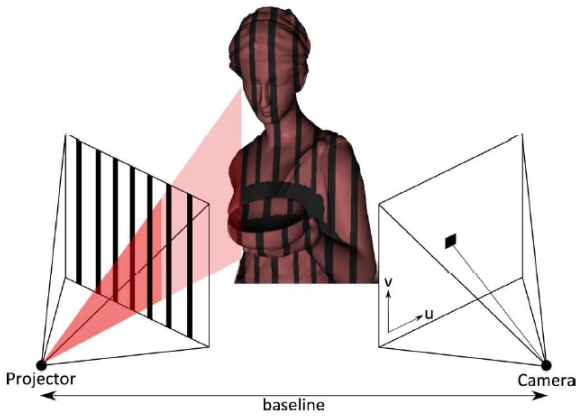


Figure 2.1: Structured light principle, with vertical slits projection pattern example (from [11]).

The projection of grid patterns, dot patterns, multiple vertical slits (as on Figure 2.1) of multicolor projection patterns have been extensively studied. There is a wide variety of patterns and decoding algorithms. However, they share a set of common characteristics or required steps that every algorithm must follow, and can be categorized in: camera projector calibration, pattern generation, projection and recovery, finding correspondences, triangulation, and surface creation [12].

Since it is considered a simple process, some low-cost sensors have arisen in the markets with high data acquisition rate and with resolution quality [13] [14]. Some of the weaknesses are found in situations with missing data due the presence of occlusions and shadows [11].

In more recent developments (mobile 3d system, broadway scanner) of structured light systems, the main goal has been to increase the speed of projections into multiple patterns, in order to enable the real-time acquisition, with a special look to motion and human body acquisitions [8].

Some known examples of this technique are described as it follows.

2.1.1.1 Microsoft Kinect 1.0

In November 2010, Microsoft introduced the Microsoft Kinect for the Xbox 360 video game console. Designed to work along with a video display while tracking a player body and hand movements in 3D space, it allows the user to interact with the console [15].

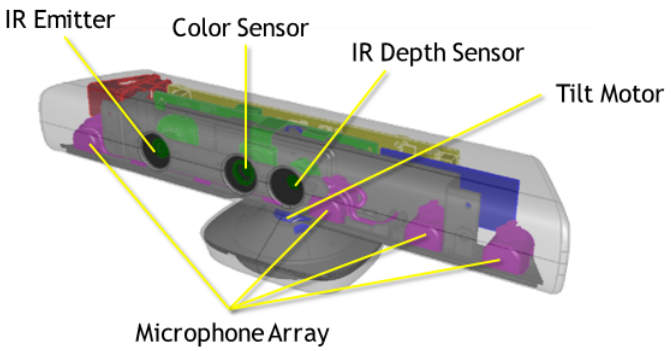


Figure 2.2: Microsoft Kinect hardware[1].

As Figure 2.2 suggests, the Microsoft Kinect contains a RGB VGA camera, a depth sensor, an infrared (IR) light source projector, a three-axis accelerometer, a multi-array microphone (to get the direction of the audio source) and a supporting hardware which allows to send information coming from the sensor to an external device via USB. It is a low weight and small dimension sensor, with an angular field of view of 57º horizontally and 43º vertically [16].

The principle of depth imaging is structured light, where the IR projector projects an IR speckle dot pattern on the object and the IR pass filtered CMOS camera captures the reflected light. The depth is calculated (by triangulation) from the deformation of the known irregular pattern, caused by object distance. In other words, the amount of disparity observed, will correspond to the necessary shift in order to match the pattern captured by the IR camera with the reference

[1] https://msdn.microsoft.com/en-us/library/jj131033.aspx

model [11]. The depth images are provided by a frame rate around 30 Hz and have a spatial resolution of 640×480 pixels with 11 bits, which provides 2048 levels of depth [17]. The range of operation is *0.8 to 3.5 meters* and the average resolution is *1.31 centimeters* at *0.8 meters* [18].

Some of the reasons that explain why the Microsoft Kinect has reached a high level of popularity within the research communities are its price, which is around 100€, and the amount of documentation available online freely [10]. It comes with a versatile Software Development Kit (SDK) by Microsoft and several open-source drivers, allowing any user to connect the device to his/her own personal computer and start developing [16]. The community has never stopped working to evolve and bring a lot of new libraries and open source drivers such as OpenNI, NITE and OpenKinect [1]. All have emerged for various kind of applications in order to create tools for scene perception and analysis with the advantage of being able to be produced in many different programming languages like C#, C++, Visual Basic, Java, Python and ActionScript [11]. On March 2013, a new library called Kinect Fusion was released which allowed to reconstruct 3D scenes in real time just by holding in hand the device and moving across space [19].

2.1.1.2 Orbbec Astra

Figure 2.3: Orbbec Astra[2].

More recently, Orbbec released the Astra camera (Figure 2.3) and it claims to be a powerful and reliable standalone 3D camera. It is Optimized for long and wide range distances for scenarios including gesture control, robotics and 3D scanning. It was developed to be highly compatible with existing OpenNI applications, which is ideal for pre-existing software on the market. It also comes with the Orbbec Astra SDK software for developing along side the OpenNI framework in very known operating systems such as Windows, Linux and Android [20]. For the price of $149.99 we can get a small, light-weight device, with a depth sensor, using the structured light technique, on a resolution of 640×480 pixels of 16bit at the rate of 30Hz. It presents a field of view of 60^{o} horizontally and 49.5^{o} vertically with ranges between *0.4 to 8 meters* [20].

[2]https://www.linkedin.com/company/orbbec-3d-technology-international-inc-

2.1.1.3 Intel RealSense R200

Intel is very well know for their production of Original Equipment Manufacturer (OEM) products. With this in mind, Intel RealSense R200 camera (Figure 2.4) is meant to be integrated into the back of a tablet or laptop display in a rear-facing topology, but also to be used in installation projects by providing toolkits for developers. It offers capabilities for applications on Scene perception, face tracking and recognition, 3D scanning objects and bodies, "Depth enabled Photo & Video" and speech recognition, with the help of a microphone. Being considered good for medium-long range indoor applications, it was developed in a variety of frameworks supporting C++, C#, Javascript, Processing, Unity and Cinder, thanks to the robust RealSense SDK [21].

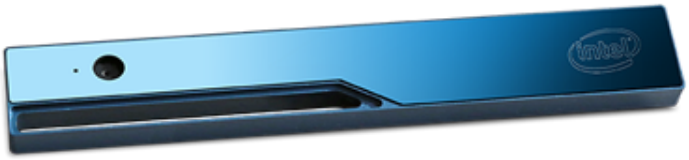


Figure 2.4: Intel RealSense R200[3].

This camera brings a Full HD 1080p (1920 × 1280 pixels) RGB video resolution at the rate of 30Hz, while the depth sensor is at 640 × 480 pixels for a rate of 60Hz. Since it also uses the structured light, it possesses a laser projector companion with the sensor for field of view of 90º both horizontally and vertically for ranges from *0.5 to 3.5 meters*. It was released in the September of 2015 and is currently available on the market for a price of $99 [21].

2.1.2 Time of flight

This active method produces a depth image from the object's surface information in real time [22]. The object is hit with a light signal, featured as a modulated amplitude cosine wave, with a frequency generated by the sensor's laser pulse emitter. As seen in Figure 2.5, the sensor's receiver will detect the reflection and measure the phase difference between the emitted and received lights so we can get the depth of the object at short ranges [10] [11].

This calculations can be achieved by the formula 2.1, where c is the velocity of light, d the distance travelled by it, $fmod$ is the modulation frequency and Δ represents the phase shift [10].

[3] https://software.intel.com/sites/default/files/managed/89/d4/realsense-r200-camera-375x295.png

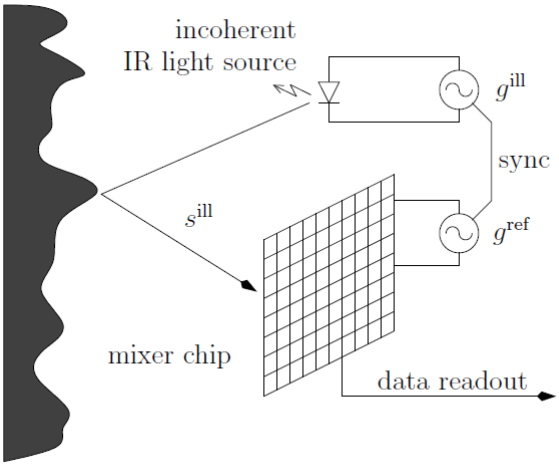


Figure 2.5: Time of flight principle from [11].

$$d = \frac{c}{2}\frac{\Delta\Phi}{2\pi f mod} \tag{2.1}$$

Another different approach for the time of flight (easy to find in LiDAR sensors) is to send pulses with a rapid laser at the object, such that it becomes possible to calculate the time that the pulse took to travel and get back to the sensor. For close range objects, about one meter, using time of flight gets harder because the time differences get shorter. This will require high speed timing circuitry. Although the accuracy drops sharply at close measurement ranges, it is possible to get good results with medium large ranges, from *15 to 100 meters* [8].

Sensors that use this technology normally face a few problems with shiny surfaces, which will not reflect so well unless they are perpendicularly oriented to the line of sight. Some key advantages are found as the good acquisition rate and the performance independently to the ambient light [8].

Next, a few examples of sensors using this technique are described.

2.1.2.1 Microsoft Kinect 2.0

The second generation of the Microsoft Kinect sensor (Figure 2.6) was released along with the new Xbox One console by the summer of 2014. Comparing this new camera with the first generation of Microsoft Kinect, the main difference is found in the technical part with the switch of 3D imaging technique, going from structured light to the approach of Time of flight. It promised to be more precise, since it can create a series of different output images independently of ambient light [23]. Those series can be acquired at multiple modulation frequencies, which can help eliminating the ambiguity of depth measurements [11].

[4]http://www.xbox.com/pt-PT/xbox-one/accessories/kinect-for-xbox-one#fbid=DWcy0kOhflT

Figure 2.6: Kinect second generation[4].

Also a few of other technical specifications have been improved to solve some drawbacks of the first Microsoft Kinect version, such as, the low geometric quality, the poor quality from the RGB camera and problems with the structured light approach, for not being always robust enough to provide complete framed scenes. Some extracted information would come with some missing parts and very noisy [11][4]. However the new camera brings an High Definition color camera (resolution of 1920×1080 pixels) and a new depth camera (resolution of 512×424 pixels) with much better field of view 70^{o} horizontally and 60^{o} vertically [24][23]. The new released SDK 2.0 offers tools to track six full skeleton, including the information of the position, direction and rotation of 26 skeleton joint for each detected body, with a good accuracy at the ranges of 5 meters [25]. The external hardware interface USB 3.0 makes this device able to transmit data at the rate of 30 Hz but with a new improvement on the quality of its images [24]. Although this sensor offers a new quality standard for its price, around $150, for being so recent, there is not much research and documentation available. For that reason, researchers from the computer vision communities still prefer the original Microsoft Kinect 1.0 for their own applications.

2.1.2.2 Softkinectic DepthSense 325

This camera follows the time-of-flight technique for consumer and industrial close range applications [26]. As soon as Softkinetic launched this pocket camera (Figure 2.7), it alleged that it would be the most accurate depth camera at the market. It was built to provide precise finger and hand tracking into a wide range of applications in different platforms. The depth sensor delivers real time 3D distance data in order to create depth map images with a resolution of 320×240 pixels from a rate of 25Hz up to 60Hz. It works in a range distance between 0.15 and 1 meter, with a field of view of 74^{o} horizontally and 58^{o} vertically. The noise that can be found in the data normally is less than 1.4 centimeters at 1 meter distance from the object to the sensor. The RGB camera allows to record video with a resolution of 1280×720 pixels with a field of view 63.2^{o}

[5]http://www.theverge.com/2012/6/5/3065706/softkinetic-ds325-worlds-smallest-gesture-camera

Figure 2.7: Softkinectic DepthSense 325[5].

horizontally and 49.3^{o} vertically. With the dual microphones also integrated it allows audio-based interaction. Although production has been discontinued, it is still available for the price of $259 [26].

2.1.2.3 PMD[vision] CamCube 2.0

CamCube 2.0 is an optical sensor created by PMD[vision], considered a high-end product at the time of its release due to its performance (Figure 2.8). It presents a depth resolution of 204×204 pixels, in an average frame rate of 25Hz for measurement ranges between 0.3 to 7 meters and with a field of view of 40^{o} horizontally and 40^{o} vertically [27].

Figure 2.8: PMD[vision] CamCube 2.0[6].

Because of the use of time-of-flight technology, the key features of this device are the following: being flexible in measurement ranges using modular light sources, multi-camera operation using different frequency channels, flexible readout with programmable region-of-interest (ROI), suitable for indoor and outdoor environments and the use of Suppression of background Illumination (SBI) technology. The toolkits and API available are mainly for programming interfaces in C and MATLAB [27].

2.1.3 Stereo Vision

This method uses two or more cameras that capture the same scene simultaneously in displaced space (see Figure 2.9). No special equipment or projections are needed [10].

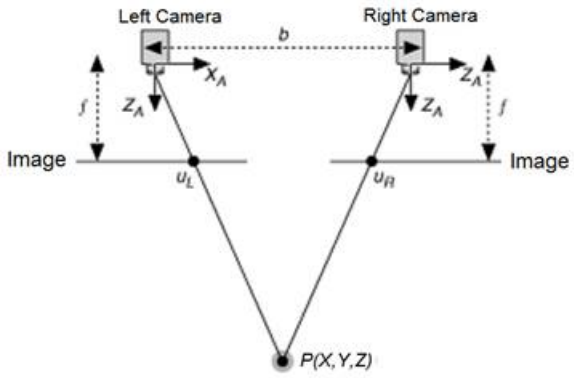


Figure 2.9: Simplified Stereo Vision System[7].

The acquired images will present a small displacement, which allows to get depth information of the scene points from that divergence. The depth values for each point are calculated by means of the triangulation with the corresponding points found within a pair of images. After the calculations are made, it is possible to reconstruct the 3D scene [10].

Stereo vision, among all the passive sensors techniques, is the one that has become more popular for applications in robotics and computer vision, where the interpretation of the scenario is more important than the quality of the data, which depends with the surface texture [28].

This technique can be found in the sensors which will be next described.

2.1.3.1 StereoLabs ZED Stereo Camera

The Zed Stereo Camera (Figure 2.10) is a depth sensor based on the passive stereo vision technique. It is able to produce a high resolution side-by-side video on a USB 3.0 interface containing two synchronized left and right video streams. With the use of ZED SDK, a host machine is able

[6] https://upload.wikimedia.org/wikipedia/commons/thumb/9/98/PMDvision_CamCube.jpg/640px-PMDvision_CamCube.jpg

[7] http://www.ni.com/cms/images/devzone/tut/lrlwhxvh4348425072667506924.jpg

[8] https://www.stereolabs.com/img/about/zed_face.jpg

Figure 2.10: ZED Stereo Camera[8].

to process the data and compute the depth map from the side-by-side video in real-time [29]. It captures video in different quality levels, being the best at 2.2K resolution (4416×1242 pixels) with a frame rate of 15Hz. Despite others modes provide less resolution quality, they are able to operate at faster frame rates up to 120Hz. The Depth data provide at the resolution of the video with 32-bits depth resolution per pixel, for distances ranging between 1 to 15 meters. It uses wide angle lens with reduced distortion leading to a field of view with a maximum of 110^{o} diagonally. This camera is targeted for Windows and Linux environments with its own SDK provided for developers and compatibility with OpenCV toolbox. Although the great performance and powerful hardware, the provided SDK is limited to just capturing the depth data stream, without any further processing or high level interpretation, which means that the developers are responsible for the implementation of applications, such as tracking objects and scanning scenarios. This device was released in May 2015 and can be found at the price of $449 at StereoLabs online shop [29].

2.1.3.2 DUO3D DUO MLX

The DUO MLX is a compact depth sensor (Figure 2.11), intended for the use in research areas such as robotics, inspections, microscopy and human computer interaction. It is based in the stereo vision technique combined with high power infrared LEDs and filters allowing to precisely control lighting environment for both indoor and outdoor usage [30].

It offers a configurable stereo resolution in a set of different modes, from 752×480 pixels at the rate of 56Hz up to 360Hz but for a lower resolution of 320×120 pixels. It operates in a field of view of 170^{o} wide angle and with distance ranges between 0.3 to 2.4 meters with low distortion. This specifications work for both RGB and Depth data that is processed by the provided basic SDK for the creation of depth maps although there is no higher-level interpretation. The SDK is

[9]https://duo3d.com/public/media/products/web_00-4.png

Figure 2.11: DUO MLX[9].

available for almost every operating system including even ARM-based systems. Since its release in May 2013, it can be bought for the price of $695 [30].

2.1.3.3 Ensenso N10 Stereo 3D camera

Figure 2.12: Ensenso N10 Stereo 3D camera from IDS[10].

Imaging Development Systems (IDS) introduced the 3D stereo camera Ensenso N10 (Figure 2.12), which works by using the stereo vision principle combined with a projected light technique. It owns two global shutter CMOS sensors and a pattern projector, which is used to project a random light pattern on the object. Through triangulation, stereo images are matched with the help of the projected patterns which presents good results even in unstructured surfaces [31] [28].

It is a very light weight compact device, compared to other sensors. Besides, it provides images with a resolution of 752×480 pixels and operates in a distance between *280-1400 mm*, at a frame rate up to 30 Hz. The depth sensor presents a resolution around 0.1mm and 1.6mm depending of the object distance. For development, IDS provides supporting software freely, which includes MVTec HALCON interface and an object-oriented API (in C++ language) [10]. Due to these specifications, this equipment may be considered as an high-end device.

2.1.4 Summary

After this overview concerning the three most common 3D imaging techniques, a summary on the pros and cons is done, with the following simplified Table 2.2.

Table 2.2: 3D imaging Technique Strengths and Weaknesses.

3D imaging Technique	Strengths	Weaknesses
Structured Light Pattern	- High data acquisition rate; - Not so much computation demanding; - Devices using this technique are easy to find for a low costing price;	- Prone to interference of the light conditions; - Missing data in correspondence with occlusions and shadows;
Time of Flight	- Good data acquisition rate; - Great accuracy ; - Performance generally independent of ambient light;	- Average cost is higher ; - Complex build and circuitry; - Multiple reflections of the emitted laser;
Stereo Vision	- High accuracy on well defined targets; - For the acquisition equipment, only need 2 cameras and no special electromagnetic emitter and sensor receiver are required; - Depth resolution as good as the RGB cameras;	- More computation demanding; - Sparse data covering; - Limited to well defined scenes; - Low data acquisition rate; - Equipment cost;

Concerning the objectives of this thesis on getting a relatively affordable, easy to use and portable system, by the analysis of the Table 2.2 it is possible to conclude that the most preferred technique would be the Structured Light pattern due to the fact that it has a good combination of strengths that the others do not have. In other words, this is a technique where it is easy to find very affordable sensors, with a high data acquisition rate and does not demand so much computational resources.

[10] https://en.ids-imaging.com/store/media/catalog/product/cache/2/image/795x795/9df78eab33525d08d6e5fb8d27136e95/c/a/camera-usb2-ensenso-3d-1_6.jpg

Within the preferred technique, it is presented a summarized analysis (see Table 2.3) of the main advantages and disadvantages between the previous selected and described examples of sensors.

Table 2.3: RGB-D Strengths and Weaknesses comparison table.

RGB-D Sensors	Strengths	Weaknesses
Microsoft Kinect 1.0	- Huge amount of documation and software toolkits built within the computer vision's community available online targeting this device; - Good frame rate for both depth and color sensors; - Low-Cost price; - Simplicity of usage;	- Accuracy dependent on lighting conditions; - The field of view's angles are not the best with this camera; - Lower depth resolution compared with more recent sensors;
Intel RealSense R200	- Small dimensions, good for portability; - Good frame rate at 60Hz; - Cost;	- Lack of documentation support; - Depth range distances are shorter compared with others; - No Skeleton tracking;
Orbbec Astra	- Good pixel depth resolution at 16 bits; - Large distance ranges; - Well made system, with a fast performance processor;	- Very few supporting toolkits; - Lack of documentation support; - Low frame rate for the RGB camera;

It is fair to conclude that, from the available sensors in the market, the one that calls more attention would be the first generation of the Microsoft Kinect. Despite having sensors with better hardware specifications, they turn out to always be more expensive, difficult to use and generally present a lack of information with software. Besides, Microsoft was the first to release an affordable RGB-D with great support by the computer vision community dedicated to 3D modeling. This has encouraged developers and groups of research to build the most part of the documentation and software, that are found today and available freely.

2.2 3D modeling applications with Microsoft Kinect

As previously stated, since the release of the Microsoft Kinect sensor, several researchers have given new and different perspectives in the field of computer vision, contributing specially to 3D modeling. In this section, a few projects involving the use of Microsoft Kinect sensor for 3D body modeling will be shown.

One of the first approaches in this area was presented in 2011 by the research group Kinect Fusion from Microsoft [19]. Their project contributed with a system that would allow a user to pickup a standard Microsoft Kinect camera and move around a room in order to reconstruct a very precise 3D model of the scene, with a great quality. Basically, to obtain this, the system would continually track the 6 degrees-of-freedom from the pose of the camera and fuse with the live depth data from the camera into a single three-dimensional model. As long as the user gets

more new views of the physical scene, more data is available and revealed to be fused into the same reconstruction model, so that in the end it gets more complete and refined. A few of the uses of KinectFusion goes for Low-cost Handheld Scanning (Figure 2.13), Object segmentation through direct interaction (Figure 2.14) and Geometry-Aware Augmented Reality (Figure 2.15) [32]. For the implementation, they have designed algorithms for real-time camera tracking and surface reconstruction, working altogether for parallel execution on the Graphics Processing Unit (GPU) pipeline [32]. This pipeline consist on four main stages:

- **Depth Map Conversion** converts image coordinates into 3D points and normals in the coordinate space, to get a live depth map.
- **Camera Tracking** in this phase, the goal is to align the current oriented points with the previous frame with a rigid transformation based on the 6 degrees-of-freedom.
- **Volume Integration** they use volumetric surface representation instead of creating a mesh by estimating the physical surface from the conversion of the oriented point into global coordinates based on the global pose of the camera.
- **Raycasting** In the end, the volume is raycast to extract the views of the implicit surface. This raycasted view of the volume will represent an estimation of a synthetic depth map that can be used as a less noisy and consistent reference frame for the alignment of the frames.

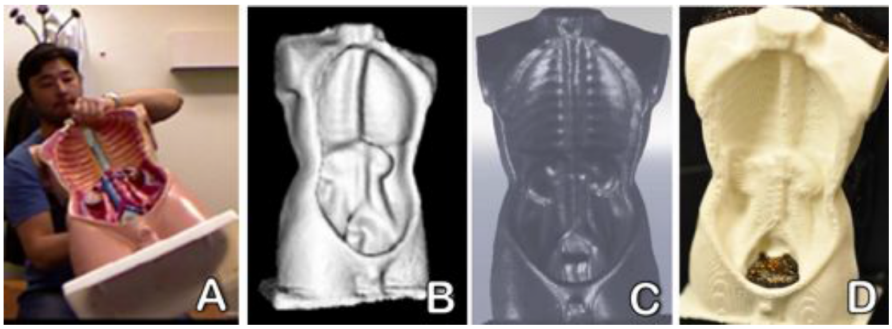


Figure 2.13: The user rotates an object in front of a fixed Microsoft Kinect to allow a 360º view 3D reconstruction and printout the outcome 3D model (from [32]).

In 2011 a new method was presented by Alexander Weiss for human shape reconstruction from noisy single image from one Microsoft Kinect sensor [33]. It combines low-resolution image silhouettes with coarse range data to estimate a parametric model of the body. A SCAPE body model was used, which factors a consistent 3D body shape and poses variations [34]. With a simple method it is possible to estimate standard body measurements from the recovered SCAPE model and show how the accuracy can be nearly the same as high cost commercial body scanning systems.

Guanglin Zhang published in 2014 a new technique for human body 3D modeling with the use of a single Microsoft Kinect camera, where the models are reconstructed by using the tools

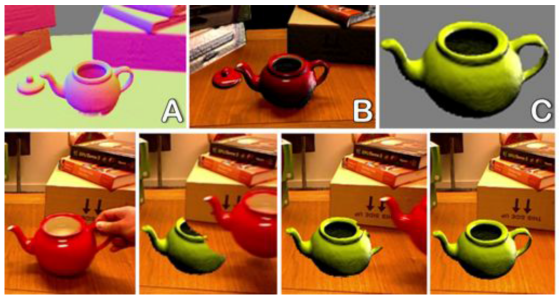


Figure 2.14: After scanning an entire scene including the object of interest, the 3D reconstruction shows the surface normals and textures mapped model. Allowing the system to monitor in real-time changes and for example color yellow the reconstruction of the segmented object that has changed position(from [32]).

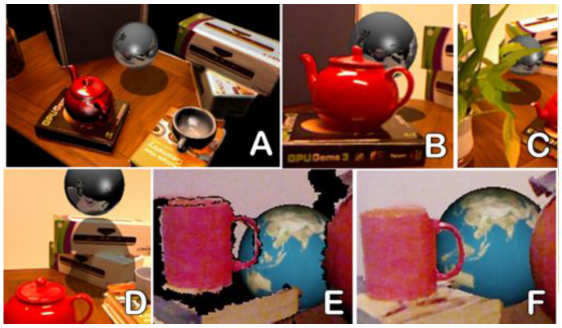


Figure 2.15: Virtual sphere composited onto texture mapped 3D model and calibrated live Microsoft Kinect. Real-time 3D model used to handle precise occlusions of virtual by complex physical geometries(from [32]).

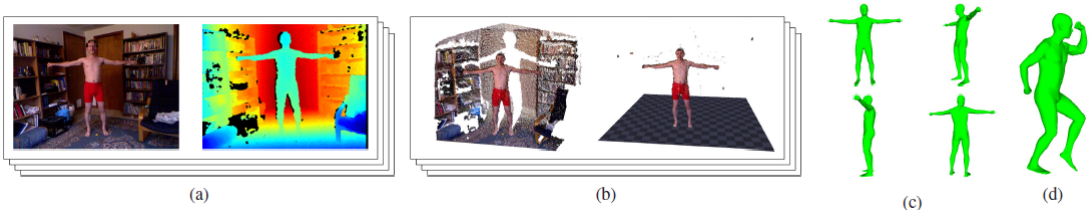
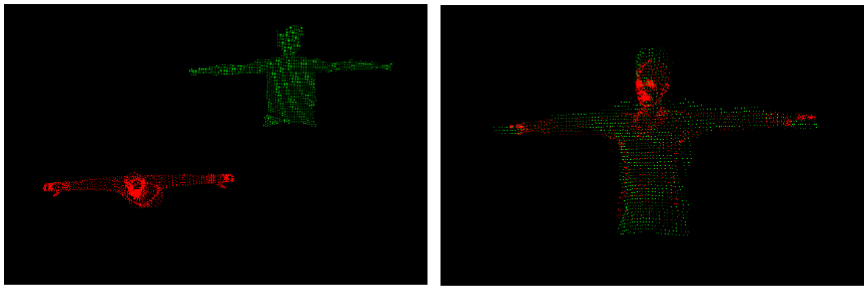


Figure 2.16: An overview of the Weiss method proposal. (2a) Four views of the body in different poses are captured from a single Microsoft Kinect. (2b) 3D point cloud and segmented 3D point cloud with ground plane for four frames (one shown). (2c) Recovered pose and shape (4 frames). (2d) Recovered shape in new pose. (from [33]).

of Processing and Point Cloud Library (PCL) [35]. In order to achieve the reconstruction, it was adopted Iterative Closest Point (ICP) algorithm for registering the captured upper human body 3D point cloud data, with the standard reference human body data. The 3D data should be converted to an appropriated format so that it can be viewed in the PCL. Zhang concluded that the number of points of a human body is too much to register precisely and that may lead to a little matching error, although the use of the Kinect showed (see Figure 2.17) to be enough for gathering data of the upper human body.



(a) Before transformation (b) After transformation

Figure 2.17: This Figure represents the reference model and the target model. A transform is needed to be applied to make sure that these sets can be gathered in the same coordinate plane and coincident after a few iterations of the algorithm [35].

Another approach was presented by Zhenbao Liu [36] in 2014 with the idea of using multiple low-cost depth cameras with particular interest of using Microsoft Kinects for 3D real human reconstruction. The cameras are positioned in the form of a polygonal mesh, helping the user to enter a virtual and immersive environment. First, the system removes the static background to obtain a 3D partial view of the human, such that later every two neighboring partial views can be registered. The whole human model is registered by using all the partial views in order to obtain a single clean 3D body point cloud. The 3D mesh model is obtained from the point cloud by implementing Delaunay triangulation and Poisson surface reconstruction. This strategy has found some limitations mainly due to the involuntary motion of the users appearing in the different partial views, which leads to the lost of depth values, leading to the failure of the reconstruction process. Another observed aspect that complicates the process, is when the overlapping region between two views is almost flat, making more difficult to get an accurate registration. Because of the lack of features in the overlap, it becomes harder to find the right correspondence given two views.

Figure 2.18: Environment of the experiment, where six depth cameras were placed around the user(from [36]).

Figure 2.19: This Figure shows progression from two partial 3D views without any correspondence, that were aligned by initial feature correspondence, resulting after a few iterations a final registration between the two (from [36]).

Figure 2.20: An example of a failure case of registration with two views resulting in a defective registration because of the variation position of the body (from [36]).

2.3 Rigid Registration

Point cloud registration is the process of overlaying two or more point clouds of the same scene taken at different times, from different viewpoints. It will align two point clouds geometrically, based on the differences between the pair due to different conditions. This process is important to achieve a complete model of the object, because of the incomplete and noisy data from just a single 3D viewpoint [37]. In medical applications, it is usual to face problems due to the non-rigid scenes causing difficulties in registration [4]. The involuntary and unpredictable movements of the object of interest and the lack of a priori knowledge about the poses and views, are a few of the challenges that researchers have been looking to overcome [38].

The 3D Registration methods can be subdivided in Coarse Registration and Fine Registration. Next a brief description for each group of methods is done, followed by some examples.

2.3.1 Coarse Registration Methods

In coarse registration, the goal is to compute an initial estimation of the rigid motion between two point clouds using correspondences between both. In order to compute it, distances between correspondences are minimized. These methods can be described by: Registration strategy, Robustness, Motion estimation and kind of correspondence [39].

For the group of coarse registration, it is possible to classify on shape features or matching methods [40]. Shape features methods use neighborhood information with the goal of finding correspondences along the search for point's characteristics. On the other hand, Matching methods focus on finding points from a pair of surfaces to be associated [39] [38].

2.3.1.1 Spin Images

This method is a 2D image characterization of a point belonging to a surface. Considering a given point, its tangent plane is computed by using the position of its neighboring points. After that, a region around the given point is considered in which two distances are computed to determine the spin image. The distance between each point to the normal vector by the tangent plane and the distance between this point to the tangent plane are defined. Finally, a table is generated with the values of the distances, where each cell contains the number of points that corresponds to a certain region. Spin images are computed in the first cloud and then, for each one, the best correspondences are sought in the second view. The transformation can be applied after finding the best correspondence [39] [41].

2.3.1.2 Principal component analysis - PCA

The idea for this method is to use the direction of the main axis of the volume given by the cloud of points to align the sequence of range images between them. In a given moment, the overlapping region will become large enough (about 50%), such that both main axes should be almost coincident and related to a rigid motion so that registration successes. In other words, the principle is to apply a single transformation which will align both axes. This method can be considered very fast but it will only be effective if there is a sufficient number of points. The most challenging detail about this method is in operation of surfaces that contain symmetries [39] [1].

2.3.2 Fine Registration Methods

In the case of fine registration, the main principle is to obtain the most accurate solution as possible. Using an initial estimation of the motion to represent all range images with respect to a reference system, the transformation matrix will then be refined by getting the best minimization in the distances between the correspondences. This is an iterative process that will try to converge to a more accurate output. Usually, these methods require a lot of processing to decide which is the closest point. The important aspects to characterize fine registration methods are: registration

principle, use of an efficient search method, robustness and minimization of the distances strategy [39].

Robustness can be seen as how well the methods deal with noise and false correspondences because of the non-overlapping regions. This is important, specially in medical images where there are always real images and the misalignments may be imminent due to the non-rigidity of objects or bodies.

2.3.2.1 Iterative Closest Point - ICP

The ICP method was presented by Besl and McKay in 1992 [42]. The main goal is to get an accurate result by minimizing the distance between points with correspondences, known as closest points. It is also known for being a Pair-wise registration where only a pair range image is registered in every execution [39]. The algorithm can be described by the following steps, with the input of a reference point cloud followed with a second cloud:

1. Pre-processing - clean data using an outlier filter and make an inliner selection of source points from both clouds, with the options of using all points, Uniform sub-sampling, Random Sampling or Normal sampling;
2. Matching - Associate points from reference to data by finding correspondence with the use of neighbor search and/or search of features;
3. Weighting - Change importance of pairs, which depends on: distance between point-correspondences, compatibility of normals and the uncertainty of the covariance matrix;
4. Rejection - When a cloud is not a subset of the following in the sequence, some correspondences are outliers and those pairs must be discarded;
5. Error computation - Compute the error of each pair, which in this case is point-to-point;
6. Minimization - Find the best transformation to minimize the errors, to apply on the second cloud, which can be the combinations of translations and rotations;
7. Go back to step 2 and repeat process until convergence.

For the Matching step, there are two possible approaches using nearest neighbor search:

- Linear search, exhaustive but good for really high dimension or low number of points;
- Space partitioning, K-dimensional Tree, more complex but helps to speed processing.

The output of this process is going to be a transformation between the the data input and reference. A few criteria to consider convergence and stop the algorithm's iteration can be defined as the following:

- Number of iterations has reached to a selected maximum;

- The transformation epsilon difference between the previous transformation (translations and rotations) and the current estimated transformation is smaller than an imposed value;
- The sum of Euclidean squared errors is smaller than a defined threshold.

Iterative closest point is considered as one of the basis for 3D registration and since its disclosure, a lot of variants have emerged with the goals of improving the errors and the computational cost. Commonly, researchers look for alternatives to get better matching strategies or error estimation functions and metrics or minimization transformation algorithms [43].

2.3.2.2 Chen and Medioni Method

After Besl and McKay proposal, Chen and Medioni proposed a similar alternative to the original ICP algorithm that same year, with the standard ICP algorithm [44], with the idea of using the minimization of the distance between points and planes. The minimization function worked with the distances between point in the first point cloud with respect to tangent planes in the second. Considering a point in the first cloud, the intersection of the normal vector at that point with the second surface determines a second point at which the tangent plane is computed. At the time, they proposed a new algorithm to find intersections between lines and point clouds, which is a process that demands computational power. Although this method is more robust, with good result tests by Salvi's review [39] against others registration methods, it has the drawback for the lack of sensibility in the presence of non-overlapping regions. Chen's principle normally requires less iterations than the ICP approach, but due to the computational power demand, it takes more time in the in the overall process.

2.3.2.3 Iterative Closest Point Non-Linear

The Iterative Closest Point Non-linear is considered a variant of the original ICP method, with the difference of the metric used for the step of computing the error. While the original idea was the sum of squared distances between corresponding points (or point-to-plane in the Chen method), the alternative proposed was the usage of the Levenverg-Marquardt Algorithm [45] to solve non linear least-squares equations, which is seen as an optimization technique that can allow to find more generic minimization functions rather then just the sum of euclidean distances. Some results for the 3D registration were presented in the work of A. Fitzgibbon in his paper of 2003 [46].

2.3.2.4 Generalized-ICP

Aleksandr V. Segal in 2009 released a paper [47] proposing a new idea combine the ICP point-to-point with the point-to-plane approach into a single probabilistic framework. The author claims that it can be seen as a plane-to-plane approach, since the framework models locally planar surface structures from both point clouds, which would be more robust to incorrect correspondences and making it easier to tune the maximum match distance parameter present in a big part of the ICP variants. Additionally, the proposed method allows for more expressive probabilistic models while

keeping the performance and simplicity of ICP framework and allows the addition of outlier terms, measurement noise and others techniques in order to get robustness.

2.3.2.5 Other approaches

Some studies came with another approaches instead of the ICP, such as the Iterative Closest Line (ICL) [48] which is similar to the ICP but with the difference of matching lines instead of points, from each model. It uses the Hough transformation for edge detection in order to extract lines to be used in point clouds by projection. Also in some situations it is used the Iterative closest Triangle (ICT) [49], where the search is for sets of three points to form triangles. This means that large triangles represent points that are far from each other, and for the opposite, small triangles can be easily found in curved surfaces [4].

In 2010 Zexiao Xie released a paper [50] with a proposal of a high-accuracy method for fine registration of overlapping point clouds. The algorithm is basically a variation of the Method of Chen, with the approach of establishing the original correspondences from two point clouds by adopting a dual surface fitting using a B-spline interpolation. The combined constraint uses global rigid motion in conjunction with local geometric invariant to reject unreliable correspondences, in order to estimate transformation parameters in an efficient way. This method takes in account the surface shape and geometric properties of the object, claiming to be less likely influenced by the quality of the original data sets. His experimental results (see Figure 2.21) demonstrated high registration accuracy in the overlapping region and uniform error distribution.

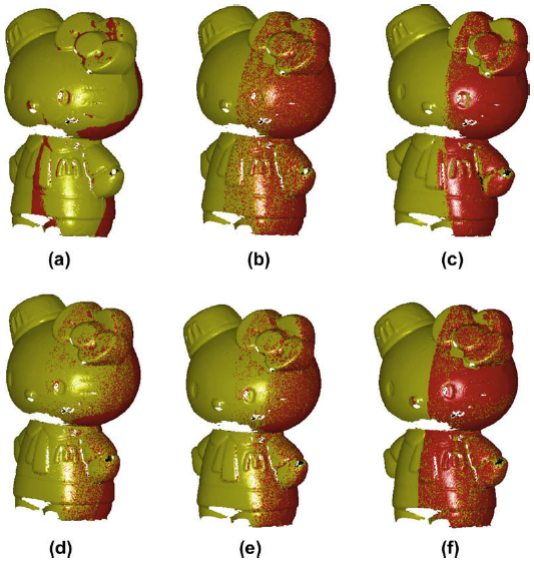


Figure 2.21: Experimental results from [50] with a comparison of two sectional point clouds from a cat toy before and after registration using different approaches: (a) before registration; (b) after registration using proposed method; (c) ICP approach; (d) Chen's approach; (e) proposed method with OAPC and (f) proposed method with OCC.

A more recent study published in 2014 by Y. Guo [51] proposed an algorithm for pairwise and multi-view range image registration. The first step of the algorithm extracts a set of Rotational

Projection Statistics (RoPS) features, from pairs of range images and performs feature matching for each group. Next, the two range images are registered using a variation of the ICP based on the pairwise registration algorithm, by means a shape growing based multi-view registration algorithm. With the initialization of the seed shape, sequentially it will update with pairwise registration between itself and the pair. Then, all input range images are registered iteratively during the growing process. The results from the comparative experimental tests (see Figure 2.22) lead to conclude that the algorithms for pairwise and for multi-view range image registrations, have shown good accuracy and robustness. In different resolutions for depth images, the experiments have shown the reconstructed 3D models to be complete and accurate.

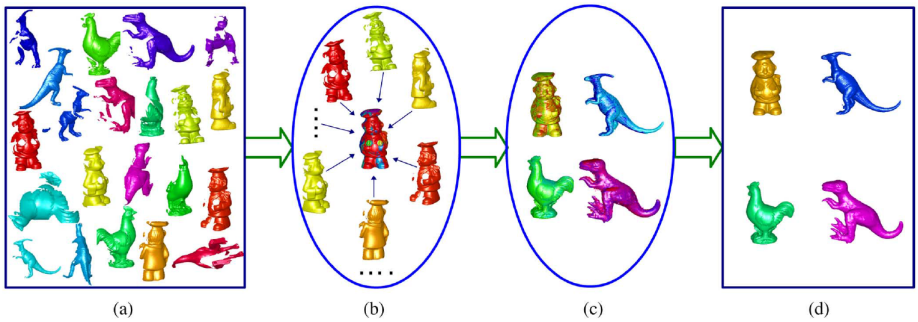


Figure 2.22: This Figure illustrates the proposed 3D object model method, going trough the steps of: (a) input meshes, (b) shape growing, (c) Multi-view Registration and (d) final result of the 3D models (from [51]).

In the middle of 2015, a proposal from Jun Xie et al. [52] claimed of a new fine registration method with superior accuracy and, at the same time, maintaining the computational power. The ICP has shown some limitations to produce good results in challenging scenarios involving objects that suffer from the lack of features due to structural ambiguity. The proposed approach introduces a new cost function with dynamic weights for the ICP algorithm, by balancing the significance of structural and photometric features with dynamically adjusted weights to improve the error minimization. Additionally it is included a novel outlier rejection method, which is adapted according to a defined threshold in every ICP iteration, while using the structural information of the object and the spatial distances of sparse SIFT (Scale-invariant feature transform) feature pairs. The paper shows a comparison between the proposed solution against other approaches and presents good results when it comes to RMS error in situations with symmetrical objects, views with less overlap regions and cases with distinctive geometric structures, while having good processing times.

2.4 Human Body Parts and Pose Selection and Segmentation from RGB-D sensors

A considerable amount of research as been done towards the body parts identification from depth and RGB data in the last decade. This fields of research are useful for such applications in computer vision for different scenarios such as medical, sports, security or military applications. In

this Section it will be shown a few works, which focus towards the human body pose recognition from depth images.

One of the most recognized works on this field is by Jamie Shotton *et. al* [5] in 2011, where he proposes an approach to build a randomized decision tree to find an approximate pose of body parts from single depth frames in Real-time application. He used a large training dataset to allow the classifier estimate body parts regardless the pose or body shape. Next it is generated confidence scored 3D proposal of several body joints in order to obtain the whole skeleton and determine the body pose. Although this approach requires several amounts of training with a considerable large dataset, the authors results have shown to get fast and accurate predictions of the body joints. This method is currently available in the Microsoft Kinect SDK for Real-Time applications environment [5] .

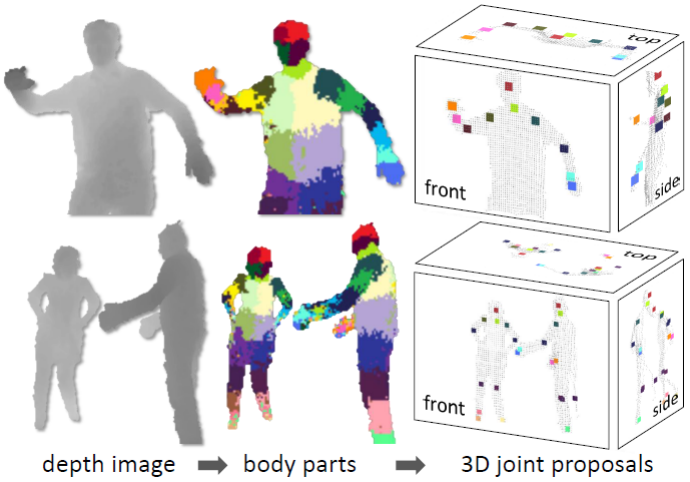


Figure 2.23: This Figure illustrates an overview of this approach by Shotton (from [5]), with a single input depth image, a per-pixel body part distribution is done (colors refer the part labels at each pixel and corresponding joint proposals). This approach tries to estimate proposals for the locations of body joints in 3D space, even for multiple users.

In 2011 Daniel L. Ly presents a new method [53] for pose information from a single depth image given an arbitrary kinematic structured without a priori beliefs or pre-trained models. Using an evolutionary algorithm to obtain the optimal kinematic configuration which better applies to the observed image. Figure 2.24 shows an example of its use.

James Charles and Mark Everingham [54] proposed a method for learning shape models to estimate articulate human poses, tested with depth images from the Microsoft Kinect device. Their proposal uses for each limb a 2D shape models in form of a mixture over probabilistic masks, by using the depth images and explore them with automatic segmentation. They claim that using their 'Pictorial structure model' based framework, has improved the accuracy on their pose estimations, because of the improvements in the fidelity of the models to the observed silhouettes (see Figure 2.25).

Before the release of the kinect system, in 2006 Ankur Agarwal and Bill Triggs [55] proposed a learning-based method to obtain the 3D body pose from singles images and sequences. Without

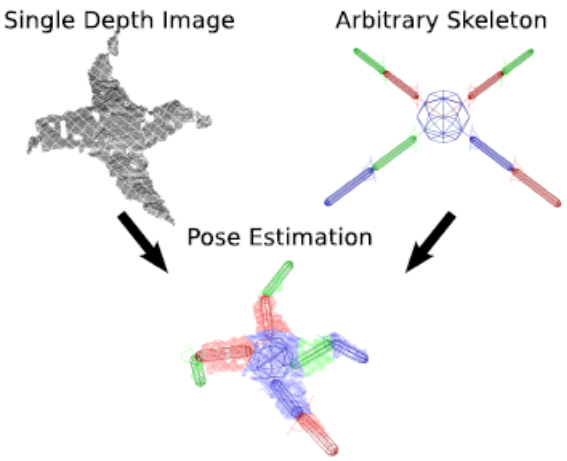


Figure 2.24: This Figure illustrates the process of estimating the pose from a single depth image using an arbitray skeleton. With a depth image as input (top left) and a parameterizable skeleton (top right), the algorithm will set of the parameters in order to get a skeleton which better fits with the data(from [53]).

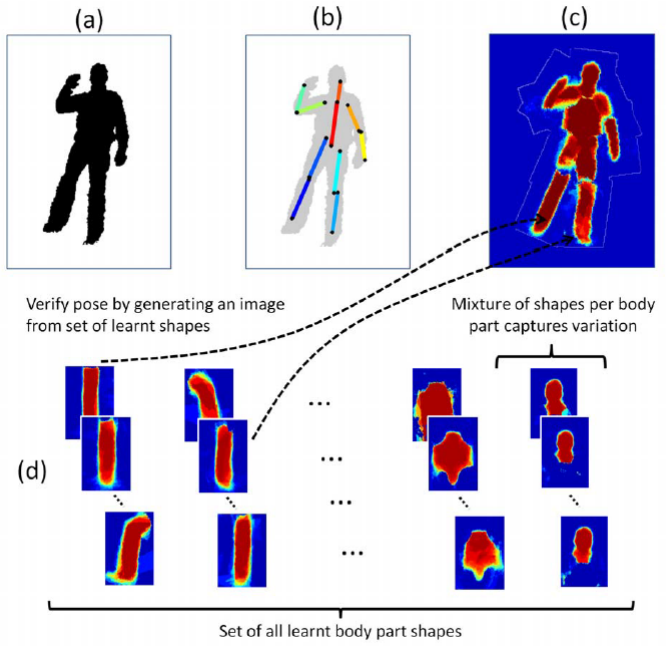


Figure 2.25: This Figure illustrates the proposed method: From a the silhoutte (a), they infer 2D human pose (b) using the models of shape (c). The mixture models of probabilistic shape templates for each limb are learnt with the depth images from the Kinect by deducing the segmentation of the limbs from the silhouette (from [54]).

requiring body models or prior labeling of body parts. Their alternative was to recover the pose by direct nonlinear regression of joint angles against shape descriptors extracted automatically from image silhouettes. To handle the loss of depth and limb labeling information they propose

a regressive tracking framework, to estimate a learned regression value to disambiguate the pose. For testing they train the regressors with a wide range of view-points, with the author claiming results of mean angular errors of 4-6°for a variety of walking motions.

2.5 Depth maps Filters

For RGB-D cameras, during the acquisition process, it is impossible to avoid the existence of noise coming from different sources. The depth information output may be affected by some of the following variables: the interference of the ambient light, due to the technology involved of using structured light, incorrect calibration between the cameras, the delay of capture between a depth frame and the corresponding RGB color frame and also with the inaccuracy of the depth camera's resolution [56]. One way of addressing these issues is by using filters for noise reduction and smoothing on depth maps. In this section, two known filters will be shown for these kind of scenarios.

2.5.1 Bilateral Filters

In 1998, C. Tomasi and R. Manduchi [57] proposed the non iterative bilateral filter for removing noise from images with a smoothing effect, while preserving edges with a non-linear combination of nearby image values. They had defined range and domain filtering, while the former will average image values with weights that decay with dissimilarity, the latter domain filtering concerns about enforcing closeness by weighing pixel values with coefficients that fall off with the distance. The spatial locality is still an important concept, due to the fact that range filtering by itself would just distort the image map. It was the combination of the two filtering insights that made this interesting process be denoted as bilateral filtering.

More recently, in a conference paper by Li Chen in 2012 [58], he proposed an approach where Region Growing and Bilateral Filter are used to counter the poor accuracy of depth image captured by the Microsoft Kinect caused by invalid pixels, noise and unmatched edges. The important role of the bilateral filter was to fill the holes with the estimated values of invalid pixels from the region growing technique and also smooth the surface considering the special noise property of the Kinect's depth sensor.

2.5.2 Outlier Remover Filters

Rusu, R. B. et al in 2008 published a work with the purpose of investigating the problem acquiring 3D object maps of indoor household environments [59]. A new approach was introduced for mapping the point cloud data, with sophisticated interpretation methods, including statistical analysis, to eliminate noise and resample the data without deleting important details, in terms of planes and 3D geometrical shapes (see Figure 2.26).

[11] http://www.pointclouds.org/assets/images/contents/documentation/filters_statistical_noise.png

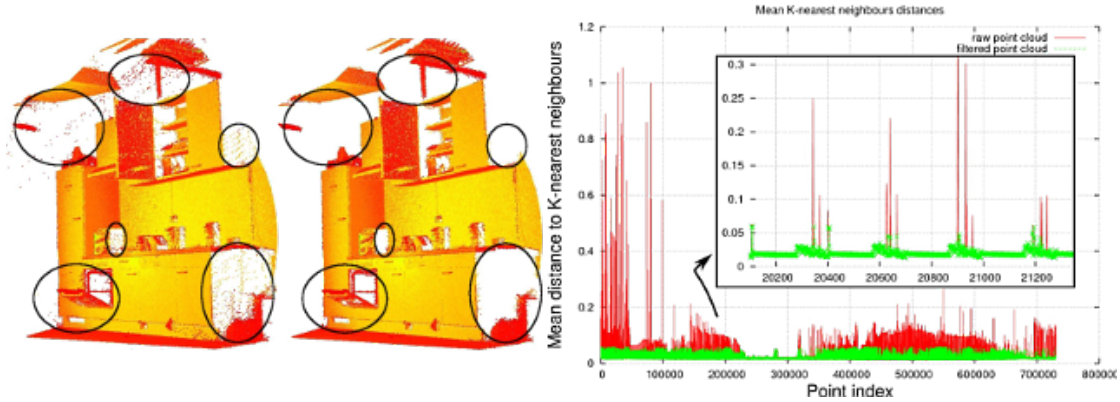


Figure 2.26: Left: raw scan; middle: scan after applying outlier removal; right: mean distances to k = 30 neighbors before and after removal [11].

The measurement errors on depth sensors typically lead to sparse outliers, which corrupt the results and becomes harder for local point characteristics estimations such as surfaces. These irregularities can be corrected by computing the mean distances between a determined number of nearest neighbors and standard deviation, with the goal of removing the points that fall outside of a defined threshold. For the selection of the nearest points, an approach combining spatial decomposition techniques and Euclidean distances calculations is commonly used. This search can be done until a desired number of points are found or all points within a bounding sphere with a defined radius are obtained [59].

2.6 3D Model Reconstructing tools available on the market

Back in 2010, Gladilin E. [60] published a paper on the possibility of making use of 3D optical body scanning of patients breast to achieve more success full aesthetic results, with customized surgery planning with visual reliability on 3d shapes in order to a get 3D photo-realistic appearance of the breast to simulate different surgical scenarios.

Following that same motivation, there is a demand by the medicine field to provide new ways of examining the patients and get more reliable information about them [61]. In that perspective, this section presents a few of the solutions which are currently available on the market for both breast cancer patients and plastic surgery planning by using 3D models generated from high-end equipment.

2.6.1 Crisalix 3D

Crisalix is a pioneer company on developing Web-Enabled 3D Consultation Tool for Breast plastic Surgery along with researchers from the Institute for Surgical Technology and Biomechanics, University of Bern in Switzerland [62]. The solution does not need any kind of special hardware or training, being mainly a software that takes a few 2D color pictures of the woman torso as an input and then generates a 3D model, which is available for visualization by using a browser (Figure

2.27). With the possibility to interact with the tool to perform simulations on dedicated servers, allowing the surgeon, together with the patient, pre-visualize the impact of the surgery beforehand. They claim to have good feedback, with appealing results from physicians and patients, with rates of 96% of the clients satisfied with the 3D simulation, while 53% believe that 3D was the decisive factor to the surgery. The least expensive license for plastic surgery clinics goes for 3490€ annually, with some features in the solution including 3D imaging for the face [63].

More recently, the company has been working to incorporate virtual reality technology for an immersive and greater experience of visualizing the simulation results [64].

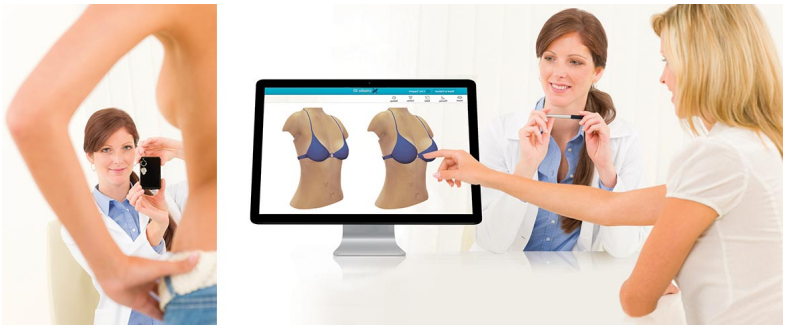


Figure 2.27: Crisalix 3D application [12].

2.6.2 Vectra XT

Vectra XT is a product provided by the American company, Canfield, which focus on developing ultra-high resolution 3D simulation systems for face, breast and body imaging. This product also delivers software to work along side the machine (see Figure 2.28). Some of the specifications of the capture system are described as the 1.2mm geometry resolution, 3.5 milliseconds capture time, stereophotogrammetry technology with an on-board computer and flat panel display. To see the results a computer, with a recent graphics card and at least 8 Gigabytes of memory, will be required [65].

Figure 2.28: Vectra XT [13].

[12] http://parksidecosmetic.com.au/wp-content/uploads/crisalix-3d-modelling-feat.jpg
[13] http://www.canfieldsci.com/common/images/products/9/title/product-title.gif

Some of the software features that are included are: Automated volumetric measurements, automatic stitching of patient's the views into a single image, dynamic soft tissue modeling technology to generate 3D models of the breast implants and visualize the expectations of the simulations of adding or removing volume in the body. Although there is not a reference about the pricing for the hardware, for installation and training on-site by Canfield it is available at $1750 for the first day and $1000 for each additional day. In alternative there are live Webinar sessions available for $250 each [65].

2.6.3 3dMD

3dMD has developed a 3D capture technology incorporating different camera viewpoints (Figure 2.29) with the goal of achieving ultra-fast capture speeds which are required to track human subjects. It is based on an approach with high-precision 3D surface image of a patient's face, head, torso, limb, thorax or even full body. Thanks to this system, hospitals and researchers are able to accurately obtain images in a non-invasive procedure. The manufacturer claims to be simple to use and a reliable enough system to handle the pressure of acquisition in a high-throughput environment and keep the accuracy and speed [66].

Figure 2.29: 3dMDtorso System [14].

For the specifications on the 3dMDtorso System, it features a capture speed of approximately 1.5 milliseconds, with a wide capture angle for the torso and breast area, working along with a configuration of 12 synchronized cameras, with a geometry accuracy of 0.2-0.5mm RMS and the generation of a continuous 3D polygon surface mesh with color mapping without any image stitching [66].

2.6.4 Axis Three

Axis Three has the motivation to offer a solution for cosmetic and plastic surgery, with a simulation system to show patients, during the consultation process, a more accurate view of the surgical outcome prior to surgery [67].

[14] http://www.3dmd.com/wp-content/uploads/2012/03/Torso-system-shot-300x242.jpg

This solution offers a robust, low complexity 3D image capture, with multiple cameras (see Figure 2.30) and projector pairs capture, based on the Color Coded Triangulation [68] (CTT) technology patented by Siemens, and later combining the images to provide an 180 degree view in a single scan. The 3D simulation software know as "Tissue Behavior Simulation" that is provided for Axis Three responds to the patient's body attributes so that when an implant is placed, the breast tissue reacts similarly as in reality. Besides, tissue elasticity adjustment is included, in order to provide more realistic and accurate outcome compared with the patient's body characteristics. This is done by using actual clinical data to enhance physics based on tissue typing [67].

Figure 2.30: Axis Three system [15].

2.7 Summary

This revision has shown different approaches in 3D imaging techniques which organizations have followed to create low-cost RGB-D sensors. RGB-D sensors, and have attained a lot of attention in the areas of 3D model reconstruction. Some sensors have shown technical specifications that are enough for applications in 3D body model reconstruction, such as in the breast cancer research for the women torso.

It is also important to conclude on the registration methods, that there are a number of different approaches, which happen to be variations of the same basis, implemented specially for a concrete application, defined by the developer. Usually, the criteria for choosing an algorithm depends on the required level of computing power or the time required to complete the process. In this case, there is a special need for a robust algorithm based on rigid transformations in scenarios of modeling human bodies, which are known for their observable non-rigidity over time in the different collected views.

Acquiring bad views and not doing segmentation of the body's rigid parts, will make the rigid registration fail, with the output models showing undesired results with lot of noise and artifacts, as seen in Pedro Costa work [1]. This effects may be attenuated with the use of techniques such as filtering for depth maps, smoothing the surface, filling the missing data and removing outlier points without compromising the edges and the color information.

[15] http://www.axisthree.com/phpthumb/phpThumb.php?src=../uploads/jpg/1390816495–ax3.jpg

In the end, the main goal is to get quality in 3D body models with low-cost and simple to use equipment like the Microsoft Kinect. In section 2.6, a few of the solutions available on the market are shown. Although they are really reliable on accuracy, they demand high-costs and special knowledge to operate them, which justifies the motivation for this thesis to attempt revolutionize this field of computer vision in Breast Cancer applications.

Chapter 3

Previous Work

The research group VCMI from the INESC-TEC, has been working to build solutions, as part of the European research project PICTURE [6], where their focus has been into the development of a framework to be used with low-cost acquisition technology for the creation of 3D models of women's torso. The final product should be used as a supporting tool for the analysis of the breast cancer patient's physiology for both aesthetic quantification and surgical planning.

This framework is divided in two main modules, the Surface Reconstruction Module and the Texture Mapping Module. The Surface Reconstruction module appertains to the attainment of patient-specific surface data, from readily available off-the-shelf imaging devices, in a setting that could be widely adopted in clinical practice. An overview of the method is presented in the Figure 3.1. Briefly, given a sequence of RGB-D images of a patient in upright position turning about the longitudinal axis of the body, a set of poses are selected manually and the corresponding point clouds are generated and registered using a two-step ICP-based method.

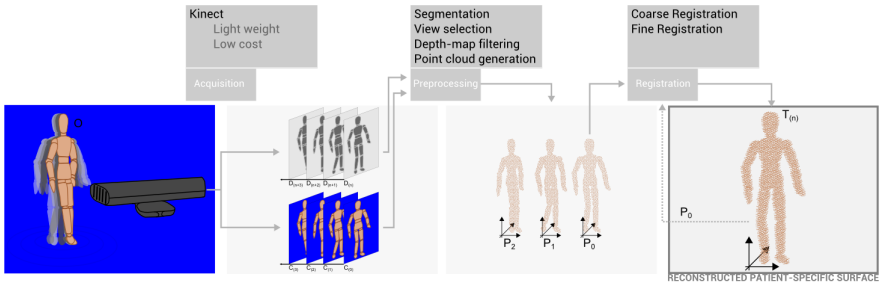


Figure 3.1: Overview of the proposed surface reconstruction approach.

Moreover, the steps that precede the point cloud generation are:

1. **Segmentation of the human silhouette in the depth images**, performed via a discontinuity based approach, using Gabor filters followed by the Otsu's algorithm;

2. **View selection**, carried through a rule-based approach using pose features (centre of mass, geometrical centre, shoulder line angle to camera, normalized width) extracted from the segmented silhouette;

3. **Depth map filtering** [69], using the segmentation of the color image obtained with GrabCut algorithm.

Although the selected depth maps provide some insight on the 3D shape of the female torso, a single view does not yield a complete characterisation of the surface of interest. Thus, richer 3D models are be obtained by registering the selected views of the patient in a common coordinate basis. Coarse registration is the first stage and aims to give an initial estimation of the rigid motion between the views. The three point clouds are pre-aligned based on the centre of mass values of each view, taking the frontal view as reference. Follows a strategy based on the concept of Tessellation surfaces in order to select some keypoints with the purpose of obtaining high descriptive feature points, as well as providing a representative sampling of the original clouds. These keypoints are automatically selected using the Delaunay Triangulation (DT) principle. An example of the key point selection stage using a female torso model is shown in Figure 3.2.

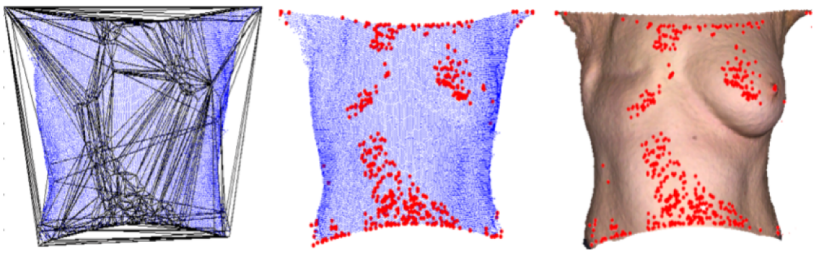


Figure 3.2: Visualization of the vertices of the free boundary triangles of the Delaunay Triangulation of an example point cloud.

The second step, or fine registration, searches for the most accurate solution possible by performing an iteratively refined alignment. The Iterative Closest Point algorithm (ICP) is used taking the coordinate of frontal view as its target. The algorithm performs the registration of each lateral view with the corresponding point cloud subset of the frontal view. Performing the registration of lateral views with each corresponding half of the frontal view, provides that the non-overlap part of frontal view is not considered for the registration. The fine registration stage is finished either when: (1) The point clouds remain almost unchanged after one iteration (the mean square error between consecutive poses of the point clouds is below a predefined threshold defined successively); or, if the first condition is not met, (2) a maximum number of iterations is reached.

The Texture Mapping module refers to the mapping of color information onto the reconstructed patient-specific external surface. For this purpose, Kinect's color sensing and mapping tools are used.

Although the color provided by the Kinect has enough quality for a myriad of applications, due to different acquisition light condition when patient is rotating, the color of the reconstructed patient-specific surface presents artifacts. To suppress these artifacts, a three-stage color inconsistency correction method is used. An overview of the method is presented in the Figure 3.3.

In the first step, the closest point correspondences between each lateral set of points and the frontal pose point cloud are established using the points' xyz-coordinates as features. Each lateral

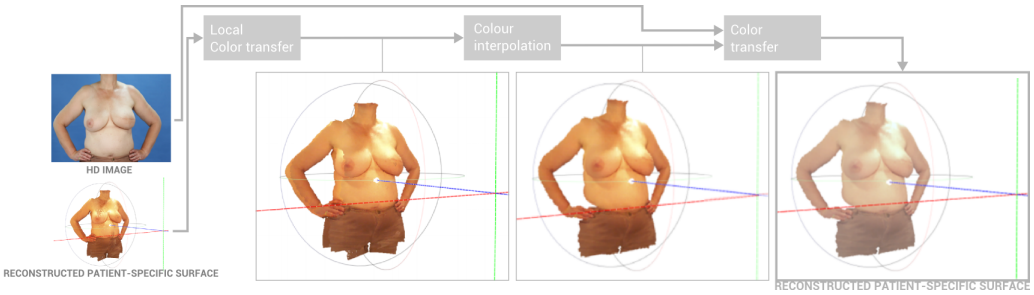


Figure 3.3: Overview of the proposed color inconsistency correction approach.

point color is then replaced by for the corresponding frontal point information considering a small neighbourhood threshold.

Secondly, all remaining points in the oblique views, which preserved the original color, are re-colored based on an iterative approach, using spherical coordinates. The origin of the spherical coordinates is the nipple in each side. The new color of the points in the oblique views is found based on an interpolation approach, taking as reference the points from the frontal views in the same radius, as illustrated in Figure 3.4.

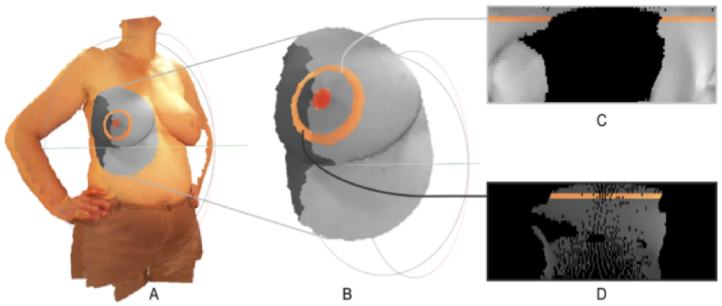


Figure 3.4: color interpolation using spherical coordinates. A) Point cloud after initial local color transfer. B) Region of interest for color interpolation, centred in the nipple – Frontal view (light grey) – Oblique view to be interpolated (dark grey). C) Polar image centred in the nipple with information from frontal view. D) Polar image centred in the nipple with information from oblique view.

In the final step, the information from the 2D HD color images is used as reference to improve the color appearance of the final point cloud, using the following color transfer approach [70]:

$$fR(x) = (\frac{\sigma_{2D}}{\sigma_K}) \times (x - \mu_K) + \mu_{2D} \tag{3.1}$$

Where μ_K, σ_K are respectively the mean and standard deviation of the source image (Kinect data) and μ_{2D}, σ_{2D} the same for the target image (2D HD data).

The next Figure 3.5 represents an high level block diagram of the described framework.

Through the development of this framework, some problems were identified and demanded an improvement of the used methods or the use of new approaches.

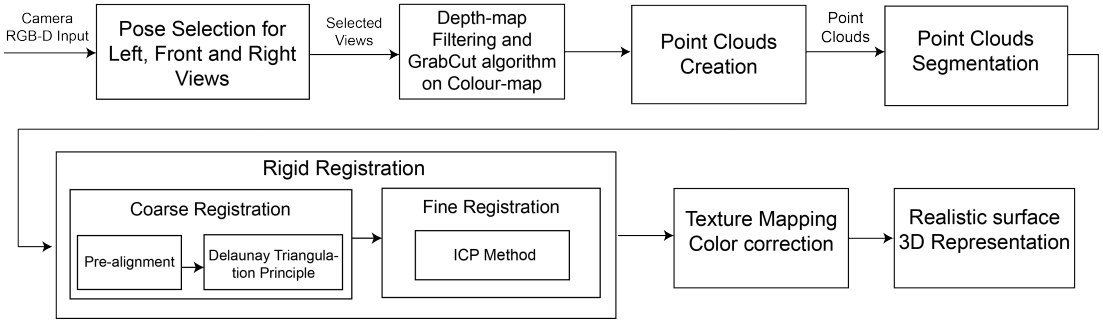


Figure 3.5: High-level block-diagram of the framework.

The purpose of this thesis involves the improvement of certain aspects of the existing algorithm, in order to become more objective, automatic (without any kind of intervention in the middle of the process, for example parameters correction), obtain more realistic results and make it more independent of the acquisition technology and user input. In order to accomplish this, these are the crucial points of the algorithm that need an improvement:

- Body Pose Selection - Automatize the process of selecting the main views of frontal, left and right poses;
- Segmentation of rigid body parts - Automatize the segmentation of the rigid body parts to improve the rigid registration process;
- Preliminary processing - Implement a 3D data processing module, to improve the appearance of the point clouds by smoothing the surface and remove outlier points which result from the noise of the acquisition data with the Microsoft Kinect;
- Rigid Registration - Improve the accuracy of the algorithm so it can handle the difficulties of non rigidity from the human body and its variability in time;

Chapter 4

3D Model Reconstruction framework for Breast Cancer Patients

In this chapter, the proposed methodologies and procedures used to solve the objectives defined in the Introduction (see chapter 1) will be described. Their goal is to improve the framework developed by the VCMI group for 'PICTURE' project as explained in the Previous work (see chapter 3), with dedicated sections for each step of the pipeline.

Looking at the Figure 4.1 as a reference, in Section 4.2 a pose selection algorithm is proposed to work with the raw RGB-D input data, in order to get automatically the main views of the scene, for later registration. In Section 4.3, it is shown a procedure to automatically segment rigid parts of the body, with the purpose of avoiding registration errors caused by the non-rigidity of the human body through time in moving actions scenarios.

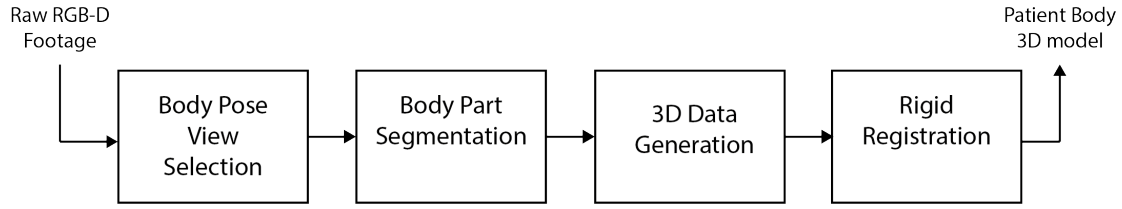


Figure 4.1: Main steps of the framework for improvement.

As seen earlier in the literature revision, low-cost cameras like the Microsoft Kinect can introduce noise and artifacts from various sources. To work around this problem, in Section 4.4, it is proposed a processing module, for 3D data Generation, with a process to smooth the surface, fill the invalid gaps and remove outliers points, such that the final results have a more realistic appearance and, possibly, the registration step will benefit from this. The last step of the pipeline is the Rigid Registration, where the point clouds of the single views are aligned and combined to obtain a model of the full body. To perform it is used the open-source library Point Cloud Library (PCL), which provides various implementations of fine rigid transforms estimations such as ICP and other variations. For this Section 4.5 this registration methods are test in different conditions

such as manual or automatic segmentations and using the '3D data Generation' module, in order to find how their perform under diverse scenarios.

4.1 Acquisition challenges and conditions

Due to the nature of the equipment, the acquisition protocol and the subject's body, there are several challenges that strongly compromise the whole framework pipeline in the different modules.

4.1.1 Microsoft Kinnect depth camera noise

The video footage of each camera is composed of sequential frames with a configured rate of 15 Hz, where each frame is saved as a PNG file associated with an identification number and a timestamp in milliseconds. Also, it has a spatial resolution of 640×480 pixels, where for the color image there are 8 bits for each RGB color component, while the depth image is saved with 16 bits but just 11 of them are in fact used by the Microsoft Kinect to store the depth intensity values, ranging between 0 and 2047.

As for the operation of the device, since the capture of the depth and color frames are from different cameras, it is difficult to get both frames at the same instant, normally getting a variable delay between them. Additionally the depth frame are vulnerable from different sources of noise, as reported by the complete review about 'Noise in Kinect Depth Images' from Mallick *et. al* in 2014 [71], where he categorizes four types of spatial noise that may affect the acquisition of a depth frame.

- Object Distance, refers to the limited ranges of the Kinect and axial noise, which makes the accuracy of the depth measure decrease while the distance of the object increases.

- Imaging Geometry, where the imaging deals with the geometric structure of the objects, getting noises defined as Shadow noise, Lateral noise and the effect of background objects, meaning the difficulties of detecting and quantify the edge pixels of an object, which depends of its position with the background [71].

- Surface or Medium Property, the noise due to the IR emitter in the Kinect, which may be affected by a reflective or transparent/absorbing surface, causing the speckle pattern to diffuse or not reflecting back to the sensor.

- Sensor technology, the Kinect may have problems of interpreting the Structured Light Coding due to two specific types of noise. Where band noise is related to the windowing effect of block correlation, used for calculating the disparity. Then, there is the Structural Noise which reports depth variances, with formations of wavy to circular ripple patterns verified by the author with a plane surface [71].

4.1.2 Acquisition Protocol

To acquire the data, a protocol was created to state the rules of the acquisition for the 'PICTURE' project (see Appendix A) in order to guide the patients while the capture of video footage is done, composed of depth and color information. A fixed RGB-D camera is settled, where the subject must follow a constant and stable 180°degrees rotation around its own vertical axis between the full lateral views, so every possible view of the torso may be caught.

For the given acquisition protocol,it is not possible to guarantee that is performed perfectly. In situations such as, the patient having motion difficulties due to health conditions may lead to situations that may condition the later processing. Normally the biggest issue is for not rotating properly in its own axis, by changing the center of mass too much through the footage. Also rotating too fast and positioning the arms incorrectly will certainly cause obstructions and body's posture changes as tilting and twisting the torso. Furthermore, if the head and legs are not visible enough for detection, the Microsoft Kinect's SDK skeleton joint tracking system will be impossible to use.

4.1.3 Diversity of patients characteristics

About the patients characteristics, there are several different body types, which leads to the important point of implementing procedures to accommodate all kinds of unpredictable features that may be lacking or are very prominent in the body. Such as, the size of the woman abdomen, breast changes after surgery, arms thickness, hips width and the person's height.

4.2 Pose Selection Algorithm

The introduction of a step to select the best views based on the pose, comes with the purpose of simplifying the work for 3D modeling, since three unique views are enough to cover the body's torso frontal view and minimize the computational power for 3D registration, as shown in Pedro Costa work [1] and used in Crisalix System [63]. Although in the previous framework, the selection of pose step was done manually to provide reliable selections for further steps of the framework, it lacked for automation, increasing the user input and time consumed.

The body poses to look for (see Figure 4.3) are the full frontal view to the camera 4.3b, and the right 4.3c and left 4.3a sides of the body, with enough rotation to see the breast's Infra-Mammary Fold (IMF) [72] (see Figure 4.2) and get both lateral and frontal view's common features, to help later with the matching step in the registration process.

As it is explicit in the diagram's first block of the Figure 4.4, it is necessary to perform a pairing between the depth and the RGB images, which is done by finding the minimal difference between frame's timestamps from each camera and reject frames which do not find a pair within 70 milliseconds. Since the camera was configured with a capture rating of 15 Hz, which means 1/15 of a second between consecutive frames, approximately 66.667 milliseconds. Additionally, it

[1] https://en.wikipedia.org/wiki/Inframammary_fold#/media/File:Imframammary_fold.jpg

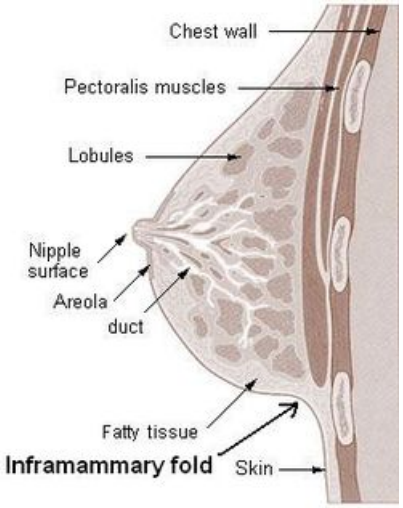


Figure 4.2: Cross section of the breast of an adult, female human[1].

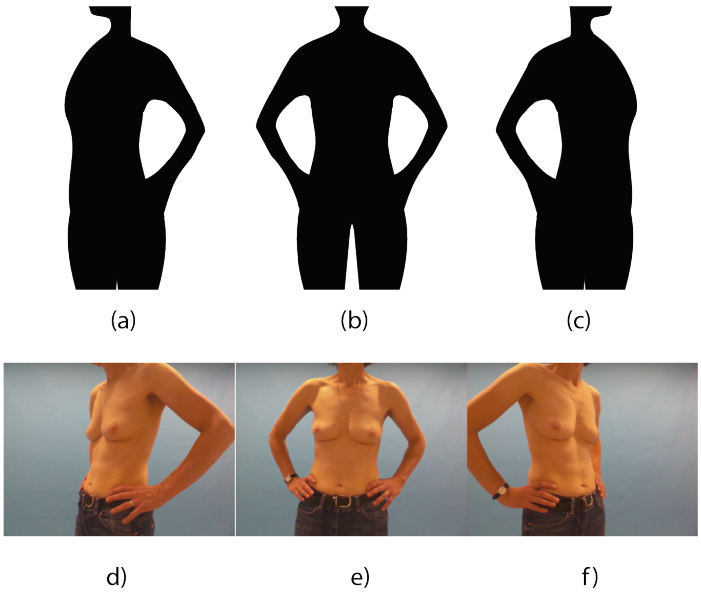


Figure 4.3: Three main poses to select, (a) Left pose view (b) Frontal pose view (c) Right pose view. (d) (e) and (f) exemplar cases from the PICTURE dataset of the aforementioned poses.

is also necessary to confirm if each pair, of depth and color image, has unique images and remove any repetitions between matched pairs.

4.2.1 Segmentation using K-means quantization

As seen in Shotton's [5] work, the huge color and texture variability induced in the scene and the data being reduced to just 2D silhouettes. Only depth images will be used to analyze the patients in this proposal, since they allow to find more features about body surface's curvature besides the human figure.

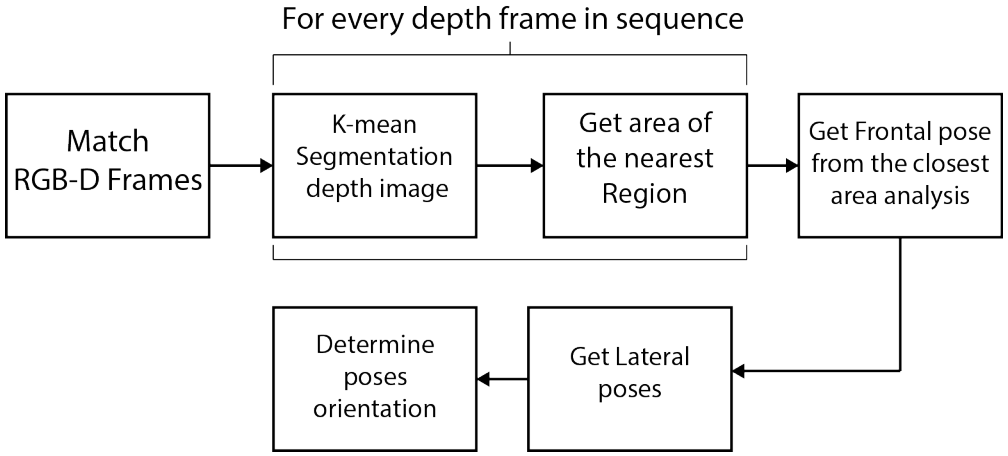


Figure 4.4: An high-level diagram block of the pose selection algorithm.

The patient is found in a certain distance in front of the wall during the acquisition. In order to extract the foreground for each depth frame in the sequence, a threshold limit is applied by using the Otsu method [73] as shown in Figure the input 4.5a) and the output result with the background removed 4.5b).

Figure 4.5: Background Removal. a) Examples of input normalized depth maps; b) normalized depth maps after applying otsu threshold and background removed.

In the James Charles [54] work, on learning shapes models, the k-means technique is used to segment the human body in Kinect depth images. Segments of the body appear as different regions that were classified into depth classes. The subject's body pose will be more frontal to the camera when the closest region gets a larger number of pixels within the same range of depth, due to the large and flat area of pixels from the chest and abdomen being more equally distant. In contrast, for a lateral view, the closest body part will usually be the arm, such that the area will get much fewer points and therefore be smaller.

Following this principle, the K-mean Clustering method [74] is used for every frame, such that

the area of the nearest region's area gets measured. The number of classes used was experimentally tuned, with the concern of segmenting an acceptable number of regions to preserve the major body parts individually, as can be seen in the Figure 4.6.

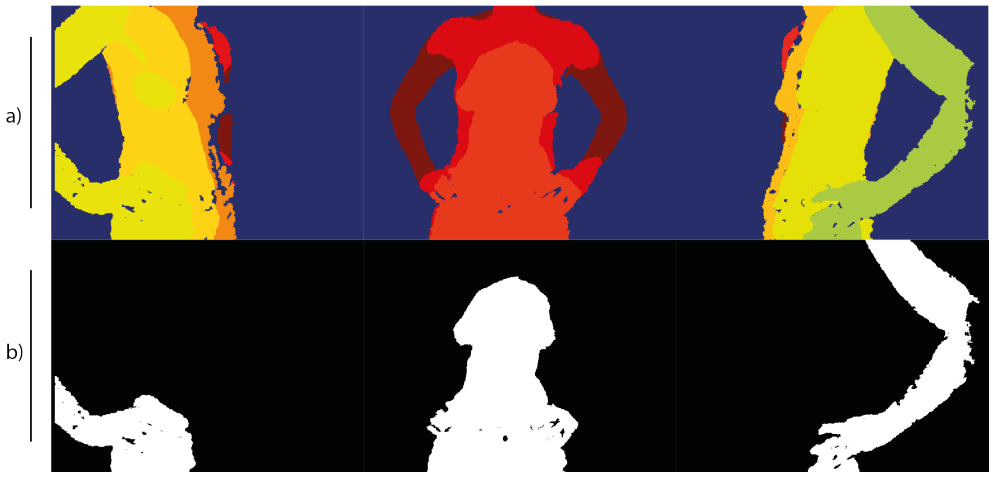


Figure 4.6: Closest Region Segmentation a) Results after applying the K-means Clustering method for segmentation, with different colors for labeling different regions; b) The closest region is segmented and isolated to measurement its area.

4.2.2 Closest Region Analysis

Figure 4.7 illustrates the measurements along the frame sequence, showing a notorious presence of noise and outlier measurements, introduced from the irregularity of the previous segmentations along each frame. In order to smooth these values for better evaluation (see Figure 4.7), a Gaussian filter is applied, with a sigma of 10 being enough for this kind of range values, which are related with the camera's spatial resolution and the size of the footage.

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{4.1}$$

The estimation of the possible frontal pose is done by computing the mean value of the areas and identify the frames which are near this calculation. This results in two frames from each side of the whole sequence, as can be seen in Figure 4.7. Finally, the frame with the frontal pose is obtained by getting the mean position between these identified two frames.

4.2.3 Lateral Poses

Looking at the previously identified frames that were found near the average area measurements, they usually occur in moments of transition between the footage endpoints and the frontal position. The observation of this heuristic approach allowed to find these frames leading to acceptable body poses being selected for right and left side views.

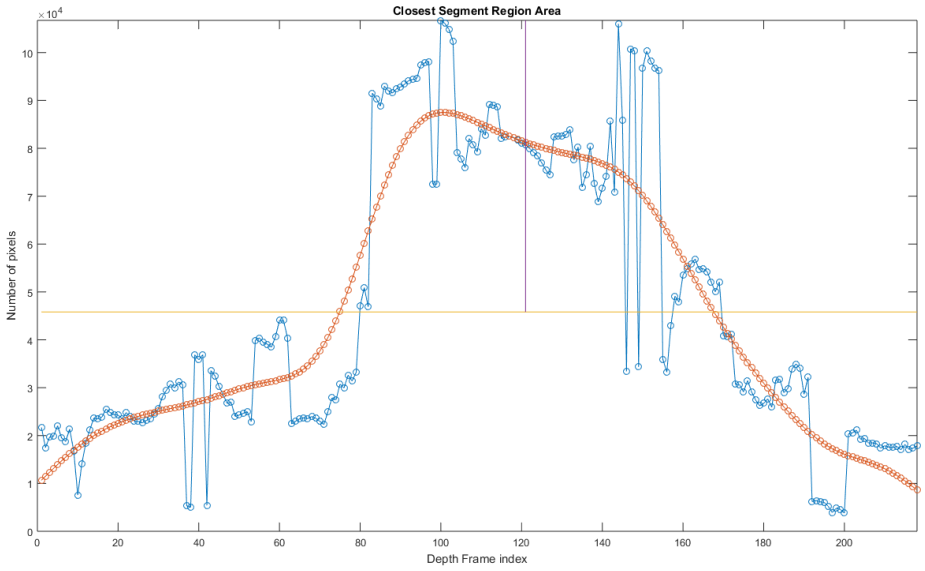


Figure 4.7: Area Measurements, of the nearest region for every frame in the sequence, represented by the blue line; Area measurements after smoothing, drawn by the red line; Orange line tracing the Mean area value; Purple line indicating the selected frame for frontal pose.

During the rotation of the subject, its center of mass will inevitably change within the image frame due to the size of the body and its own position. Taking advantage of this fact, it is possible to determine from the previously selected lateral frames which pose they represent, left or right, relatively to the frontal pose. This is done with a comparative evaluation of the calculated centroids from each body's silhouette binary masks as can be seen at the Figure 4.8, being noticeable the relative difference of both centroid's horizontal coordinate. Using the Figures 4.3 as a reference for the body pose to look for, and, Figure 4.8 for the relative centroids distance, the view which has the closest centroid to the right limiter of the frame is selected as the Left pose view. Given that selection result, the other frame gets identified as the Right pose view.

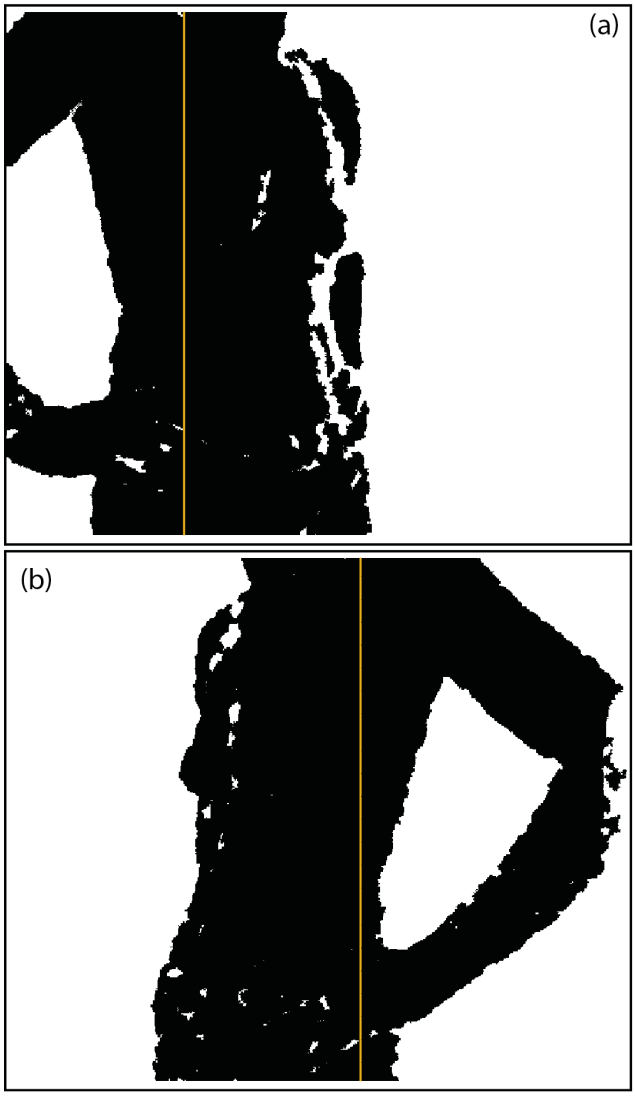


Figure 4.8: Centroid of two selected lateral views with the orange line indicated horizontal coordinate. (a) and (b) are right and left, respectively.

4.3 Body Part Segmentation

As described in Literature Revision (see chapter 2), the Rigid Registration is a procedure that matches different views of the same scene or object from different perspectives, into a single 3D model. In order to avoid misalignment and distortion, the object of interest shall not change its shape and keep its position within the scene during the views acquisition. Otherwise, the different views will not be similar and the difficulty of matching common features during the registration is increased.

For this work, the patients will perform a rotation around their center of mass, while following the Protocol (see appendix A). As explained in the Section 4.1, some parts of the human body may

move involuntarily and change their posture during this rotation, such as legs, arms, head and the abdomen. To avoid this non-rigidity, the rigid parts have to be segmented and used to estimate the best transformation to apply to the original full point cloud.

This Section will focus on obtaining segments for each kind of point cloud, from the previous selected poses. Figure 4.9, illustrates the proposal to segment each pose, where the green segments represent the regions to be preserved. The proposed approach for this module is to create two different kind of segmentations for each pose. For the point clouds used to estimate the registration, hereafter denoted as rPC_v $v \in \{L : Left, F : Frontal, R : Right\}$, the main focus is in the torso area, ignoring any non-rigid visible parts such as legs, arms, abdomen and head, as described in Figure 4.9a. This segmentation will be performed mainly in the 2D depth images, before the generation of the point clouds to avoid higher computational power that may be required in the 3D space.

In the other hand, the full point clouds, which are going to be later transformed with computed estimations from the rPC, are hereafter denoted as wPC_v $v \in \{L : Left, F : Frontal, R : Right\}$. Common body parts that are vulnerable to the rotation movements, e.g. the arms, may appear in different positions if their articulation or shape vary between different captured instants. To avoid these visual overlap conflicts in the final output, the arms in the wPC_F need to be removed, the same way it was done for rPC_F in 2D depth map, see Figure 4.9b. Meanwhile,for the poses wPC_R and wPC_L (Figures 4.9c and 4.9d), a vertical cut is performed in order to remove furthest points, which are subjected to a greater error. This vertical cut is the only operation for segmentation done in the 3D space.

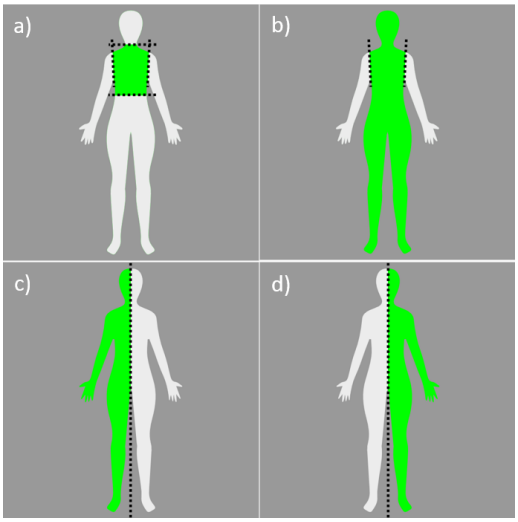


Figure 4.9: Body segmentation approach: a) Torso segmentation to be applied in every pose rPC_v for rigid registration, b) Segmentation for the wPC_F pose, c) Vertical cut for wPC_R, d) Vertical cut for wPC_L.

The pipeline of this module will follow the diagram from the Figure 4.10. Starting with the frontal segmentation rPC_F, the torso limits (Top Limit (TL) and Bottom Limit (BL)) are found and everything above the shoulders and below the Infra-mammary Fold (IMF) is removed. After isolating the torso region, the arms are removed by finding the torso edges. Next the same principle

is applied for the lateral rPC_R and rPC_L views, by applying the torso limits found previously and then remove the closest arm. For the wPC_F, just the arms will be removed, while the last step will be for rPC_R and rPC_L by applying a vertical cut.

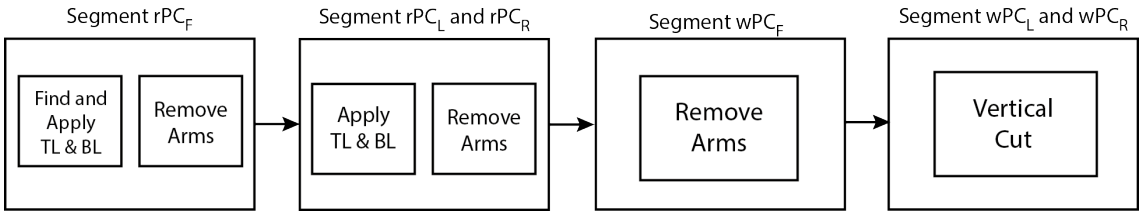


Figure 4.10: Body segmentation pipeline.

4.3.1 Segment of Point Cloud for registration Frontal rPC_F

4.3.1.1 Finding Torso Limits

To remove the non-rigid parts such as the head, abdomen and legs, in this Section it is proposed to find the top and bottom limits (see Figure 4.11) which will segment just the torso region.

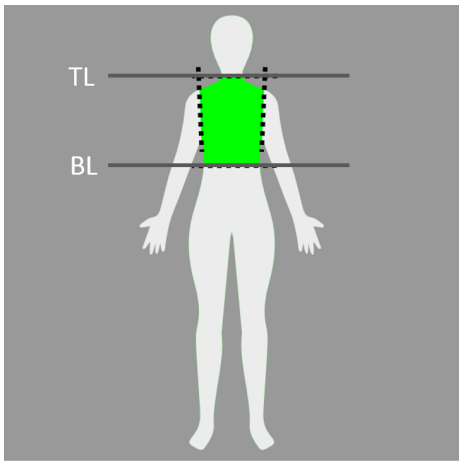


Figure 4.11: Torso segmentation defined with lines limits TL- top limit and BT- bottom limit.

In the frontal pose view, by looking at the upper-body's from the top to the bottom, it is noticeable the presence of a characteristic silhouette, existing a small width in the head area and increasing when going down through the body, due to the shoulders, the wide torso and the arms getting thicker. With this in mind, it is possible to analyze the variance of the body silhouette's width in the frame by applying an horizontal projection with the lines of pixels. This will result in a pattern, presented in Figure 4.12, with a peak value being at the height where the body appears the most in a single line of pixels, with the large torso and arms in a still position, while the hands are touching the hips. In contrast, the lowest values can be found at the top with the presence of the head and neck.

The desirable position for the Top line limit (Figure 4.11 line TL) is within the transition of the neck and shoulders, which can be found by looking at the previously calculated projection

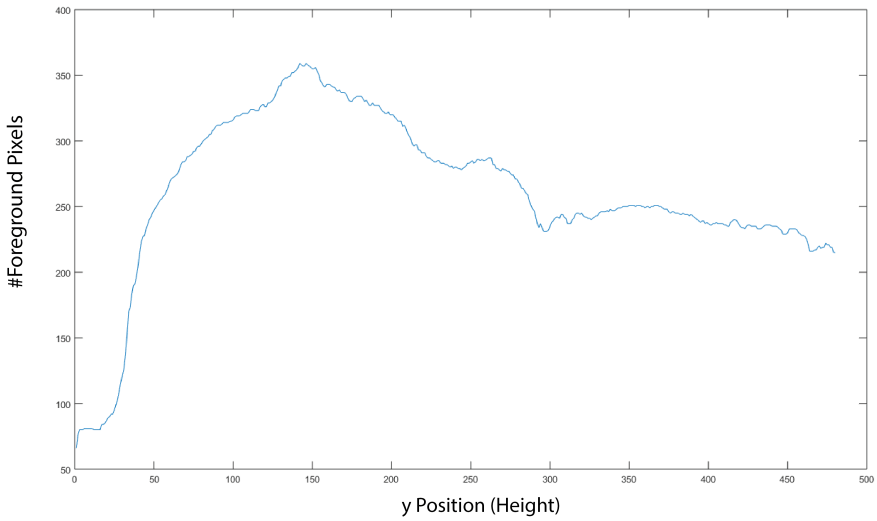


Figure 4.12: Depth map horizontal projection.

(in Figure 4.12), somewhere above the position maximum projection value. In Figure 4.13, an overlap of the depth image, and the horizontal projection is done to demonstrate the progression of the lines of pixels along the body.

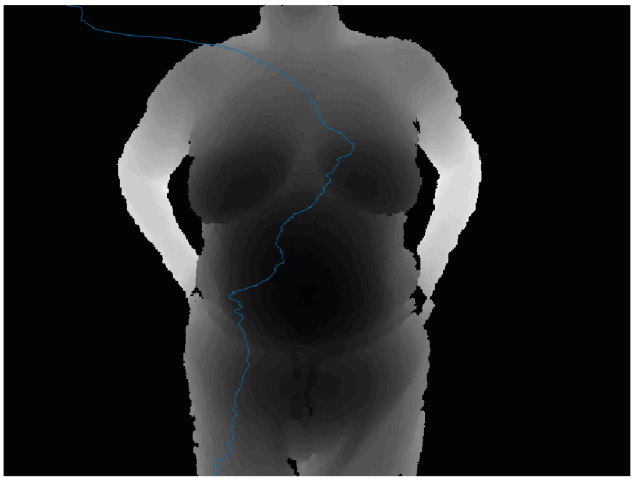


Figure 4.13: Depth map Normalized with the horizontal projection from Figure 4.12 with vertical orientation and overlapped in blue.

With its observation, it is possible to notice that when the shoulders start to appear, the projection will show closely half of pixels presence than the peak value which is around breast area and the arms thickness. In order to find the desired value, the algorithm 1 is applied. With the principle of looking for the first line of pixels which meets the condition to be selected as the preferred height in order to make the cut. Even for situations where the neck does not appear inside

the frame, the cycle will break in the first iteration and accept the first line of pixels as it should.

Data: Depth image and the horizontal projection.
Result: TL, Top line which first meets the condition.
for $line = 1$ **to** $numberOfLines(image)$ **do**
  if $sum(image(line)) >= 0.5 \times max(projection)$ **then**
    $TL \leftarrow line$;
    break;
  end
end

Algorithm 1: Find the Top line bounding torso.

This Top line will be used to segment the body with a cut in the depth map, removing the neck and head while keeping the shoulders and the rest to find the next limit. The Figure 4.14 shows the results of cut with a overlap of the previously calculated projection to evidence the intuition to find this top limit.

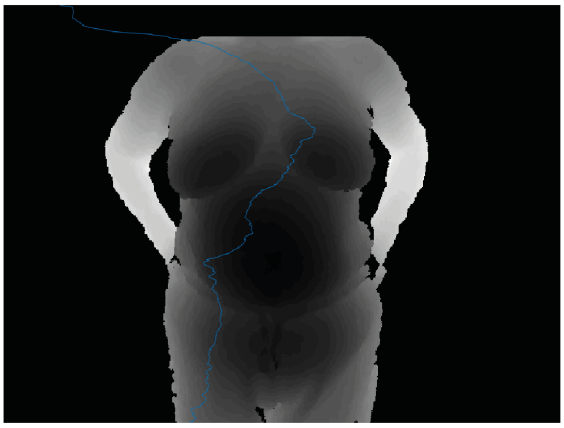


Figure 4.14: Depth map with Top limit applied and the previous calculated horizontal projection overlapping the image.

4.3.1.2 Bottom line Limit

Given the fact that the subject has to use its legs to perform the rotation, this will make them modify their position and shape between views. In addition the breathing action will cause the body's diaphragm move and consequently change the abdomen and stomach region's shape. The proposed approach finds the preferred position, to define a bottom line, close to the Infra-Mammary Fold (IMF), so it limits over the line, the rigid area of the torso, keeping the breast features and remove the critical regions for better registration. The chest area is characterized for the vertical deviation due to the existence of edge from the IMF. In order to find the IMF edge position, a vertical gradient filter with a size 21 is applied in the raw depth map image, in order to find natural transitions between the breasts and the body. This filter is large enough to properly detect the

slope of the breast. But before that, in order to avoid the appearance of slopes due to shoulders and arms, a threshold is applied to the depth map. The threshold value is obtained from the mean depth values of the foreground. This is done by taking in account that in a frontal view the torso is the closer comparatively to the arms.

In Figure 4.15, the detected edges by the gradient filter can be seen, where the most weighted values are mainly the edges between the body and the background. In order to separate body features from the body silhouette edges, a binarization is done with a threshold. Figure 4.15b shows the resulting binary image.

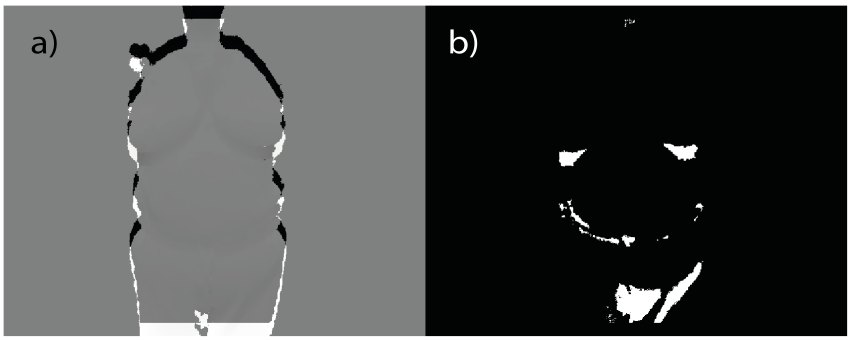


Figure 4.15: a) Gradient Filter with vertical orientation; b) Binary mask with the strong edges of body with the background filtered out.

Assuming that the conditions of acquisition do not change among the patients, the patient's chest will always show at the top half of the frame, which means that is enough to just consider the top half lines of the image to find the Infra-Mammary Fold line. This is done with a horizontal projection with just the top half lines of the binary image 4.15b, resulting the plot in the Figure 4.16. The value to look for will be the maximum peak value, which is the point where the IMF region is localed and it can be used to define the desired BL line limit from the Figure 4.11.

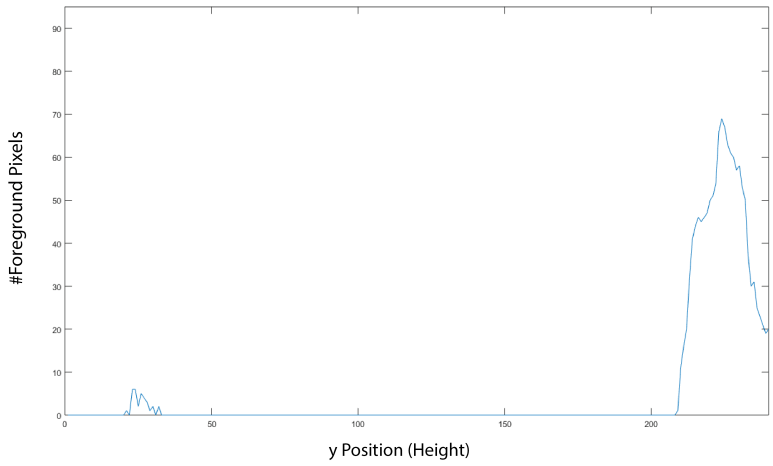


Figure 4.16: Horizontal projection from the binary mask in Figure 4.15b.

Next, a crop is done to the depth image, by removing everything below the estimated Bottom Line (BL), resulting in the depth map of the Figure 4.17.

Figure 4.17: Resulting image after using the computed torso's limits.

4.3.1.3 Arms Removal

The arms are body parts that can easily move and change their shape during the body's movements, which gives them non-rigidity properties and makes them body parts that must be removed. As defined in the acquisition protocol, the patient hands must stand still at the hips, which will make the arms open into an arc shape giving some space off the torso. This way, the body's edges with the background may be identified by the application of an horizontal gradient filter.

After its application, the vertical edges, which become more salient and have more weight, are kept due to the transitions with the background. The inner edges belong to the torso, such that if they are identified, it is possible to eliminate the arms pixels. This assumes that the patient followed the protocol and kept the arms steady.

Since both arms appear in the frontal view, it is preferable to separate them into two binary masks to analyse each torso edge. To perform this separation of the body, the centroid of the torso area is used, as exemplified in Figure 4.18.

It is noticeable that there is also present in the depth maps the edge of the arm in each side of the body, since it is a part which can be ignored, a way of isolating the torso's edges is by finding just one of its pixels and use it as a seed for the Region Growing technique, as described in the book [75] by Richard E. Woods. These pixels used as seeds, can be found by doing a search, by starting from the center of the body, with the centroid's coordinates calculated earlier, and go right and left so it finds both side of torso's edge.

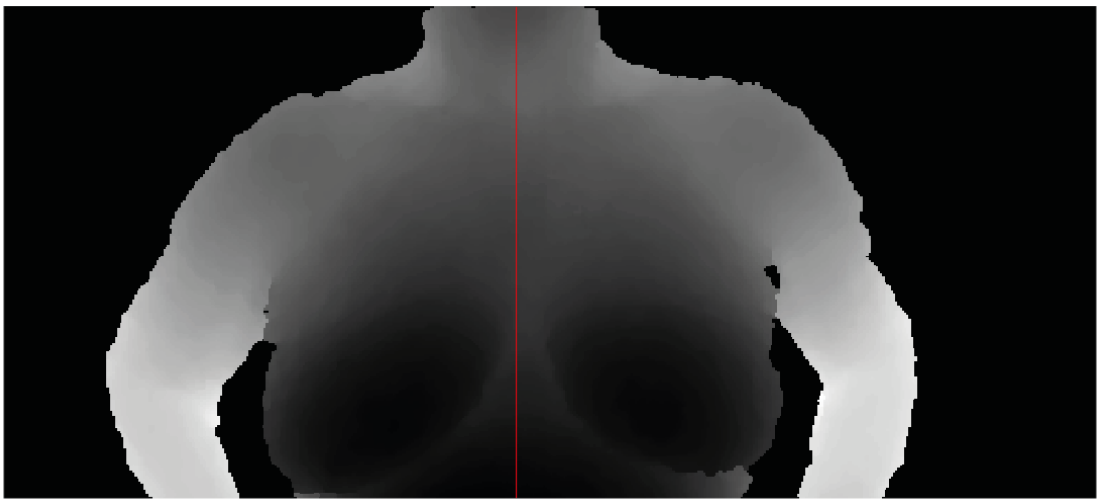


Figure 4.18: Depth map of the torso segment with a red line representing the calculated coordinate x of the centroid.

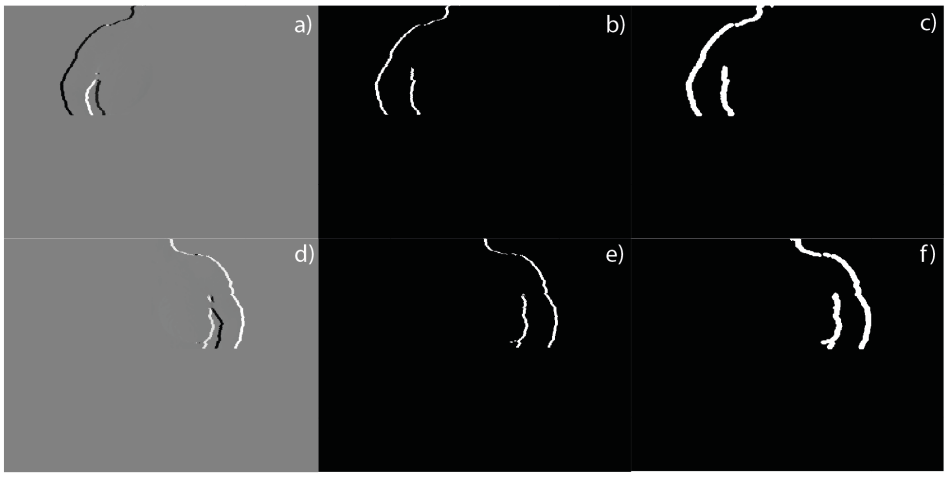


Figure 4.19: Top: a) Left Half of the body after gradient filter, b) binary mask with weighted negative transitions from gradient, c) After dilating edges on the binary image; Bottom: d) Right Half of the body after gradient filter, e) binary mask with weighted positive transitions from gradient, f) After dilating edges on the binary image.

After identifying and isolating the edges with the Region Growing technique, as seen in Figure 4.20b, now it is possible to start removing the arms from the depth maps, by using the edges to know coordinates of each side as the reference of the lines to remove pixels from those start points until the border of the frame. Thanks to this approach, it is possible to obtain a clean cut just for the arms and preserve the integrity for the rest of the body, demonstrated in Figure 4.20c.

Finally in order to remove the rest of the arm above the torso edge but preserve a portion of the shoulder, which can be considered rigid, a cut with a 45°degrees orientation is performed. Using the edges, found in Figure 4.20b, top endpoints as a starting point and remove everything in the opposite direction of the body, as can be seen in Figure 4.20d with the final segmentation result

for the rPC_F.

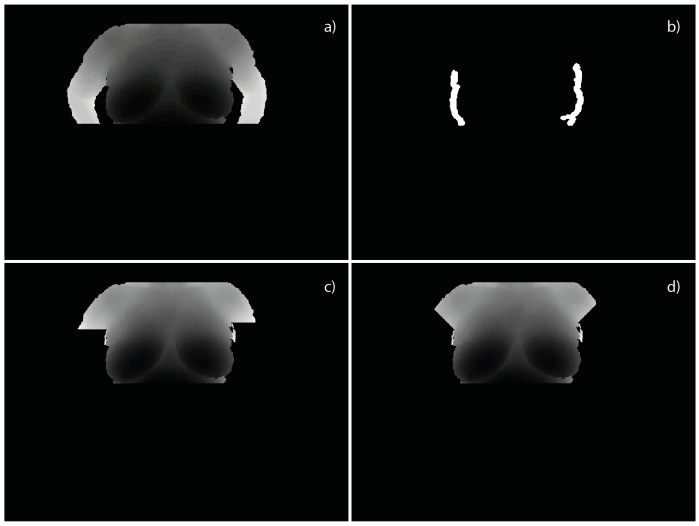


Figure 4.20: Top: a) Torso segment Depth map, b) binary mask of the identified and isolated torso edges; Bottom: c) Both arms erased from the Depth map, d) Applied the 45°degrees oblique cut for the shoulders.

4.3.2 Segment of Point Cloud for registration Left rPC_L and Right rPC_R

As for the lateral poses segmentations used for registration, the process will be similar to the rPC_F. The torso limits TL and BL will be used to segment the torso region, since the patient will not change her height during the rotation. In other words, the shoulders and the IMF limit will not change their vertical position enough to go off the limits.

Finally, while for the rPC_F, both arms need to be removed, in the lateral views (see Figure 4.21a and 4.21c) normally just one of the arms appears on the view, so it is just required to find one edge of the torso, from the side of the closest arm to the camera, and eliminate each pixel from that edge to the closest frame limits.

4.3.3 Segment of whole Point Cloud Frontal wPC_F

As explained before in the beginning of this Section, this segmentation will be used just for the final model. With this in mind, in order to avoid faulty overlaps of the arms from both registered lateral views, they must be removed to avoid this visual effect. To proceed this, the exact same principle for the rPC_F in 4.3.1.3 is used, to find the torso edges (see Figure 4.22b) and remove both arms (Figure 4.22c). Additionally, the shoulders are also trimmed with a 45°degrees oblique cut, to obtain the final depth map segmented in Figure 4.22d.

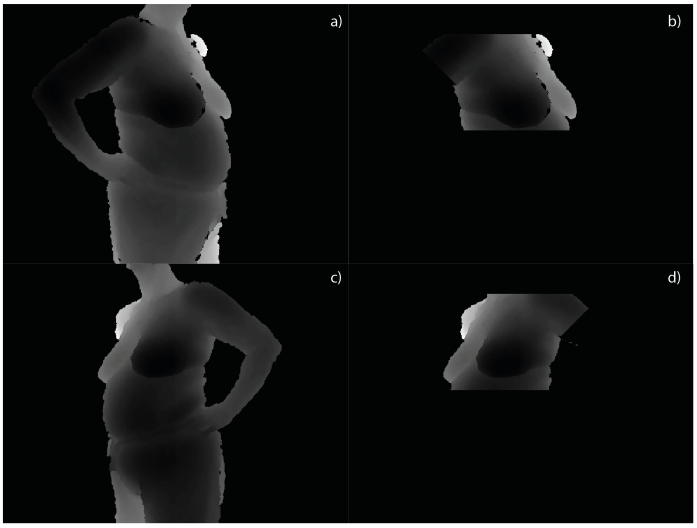


Figure 4.21: The proposed algorithm applied in the lateral views Top: a) Right pose full before, b) Right pose Segmented; Bottom: c) Left pose full before, d) Left pose Segment.

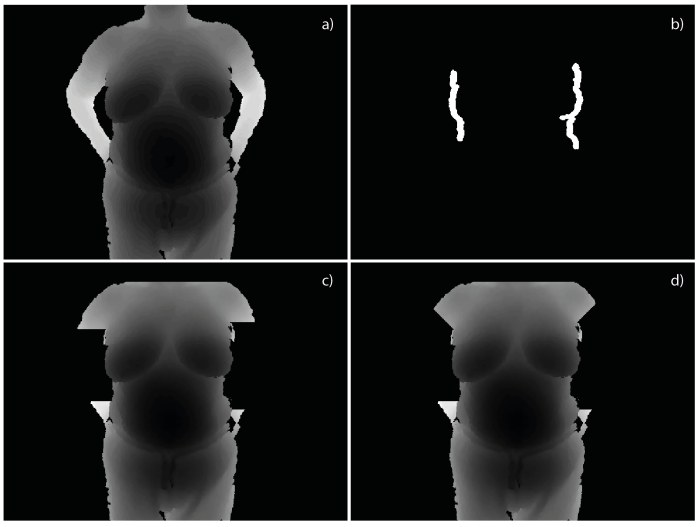


Figure 4.22: Top: a) Depth map input, b) binary mask of the identified and isolated torso edges; Bottom: c) Both arms erased from the Depth map, d) Applied the 45°degrees oblique cut for the shoulders.

4.3.4 Segment of whole Point Cloud Left wPC_L and Right wPC_R

4.3.4.1 Vertical Cut

After the application of the estimated transforms in the full lateral point clouds, the overlap of this lateral views common regions may cause some visible artifacts, mainly caused by imperfect registration and the camera difficulties on capturing this surface portions that are smaller and further away.

As explained later in the Rigid Registration framework's step (see Section 4.5), the frontal view is used as reference for the lateral Point Clouds, which means that the final model will preserve frontal view orientation, as shown in Figure 4.23. If the frontal view is facing well to the camera, calculating its center of mass, would give a close approximation of the body's center.

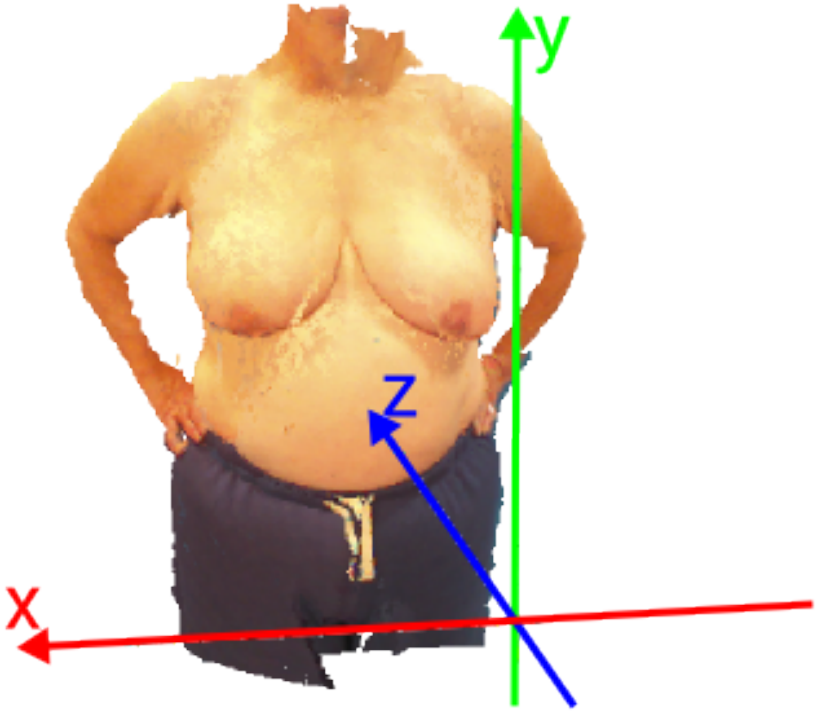


Figure 4.23: Patient model after registration with axis for reference in 3D space, axis X in red, axis Y in green and axis Z in blue.

Using the axis at the Figure 4.23 as reference, the frontal point cloud mean value of the x coordinate, will be used as the center position for the vertical limit to apply the cut. Its application, is done for each wPC_{LR} individually, after the applied rigid transform, by removing all the points which show an offset, in the x coordinate, from the calculated limit to the inverse direction of respective lateral view, has suggested previously in the Figure 4.9. The resulting cuts are visible in the Figures 4.24 for wPC_L and 4.25 for wPC_R.

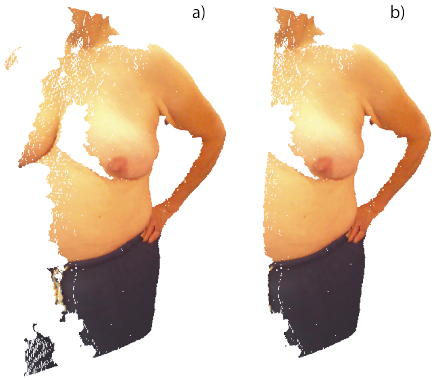


Figure 4.24: Left view Point cloud before a) and after b) the vertical cut.

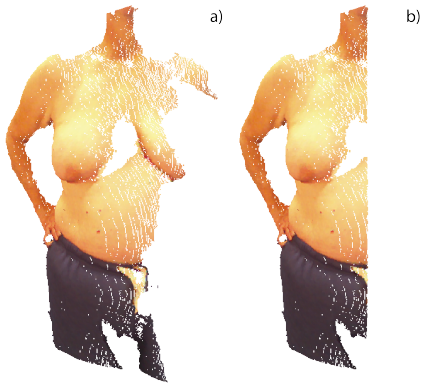


Figure 4.25: Right view Point cloud before a) and after b) the vertical cut.

4.4 3D Data Generation for each View

Applications using low-cost RGB-D cameras for 3D modeling have been increasing in popularity in the recent years, since it is possible to obtain acceptable models in 3D space using cheap equipment. Even then, as previously reported in Section 4.1, different acquisition problems may compromise the results, leading to lower success rates for certain applications.

Following what the literature revision contains about filtering for 3D modeling (see Section 2.5), the proposed approach for this step will be as shown in Figure 4.26. During the Preliminary Processing, a bilateral filter is applied to the previously segmented depth images in order to smooth the body surface. Following, a closing operation is performed to fill some small gaps. Meanwhile, a depth image filtering scheme aiming at removing untrustworthy noisy points at the edge of the foreground silhouette is applied. Therefor the color image is aligned with the depth map according to the camera calibration setup given by the used equipment. The aligned color image is then

segmented by the grabcut algorithm, initialized with a binary mask from depth map. The result of the grabcut is then applied to the depth map.

Finally, the point cloud will be generated and an outlier removal filter will be used to detect and remove the points which are too isolated from the body.

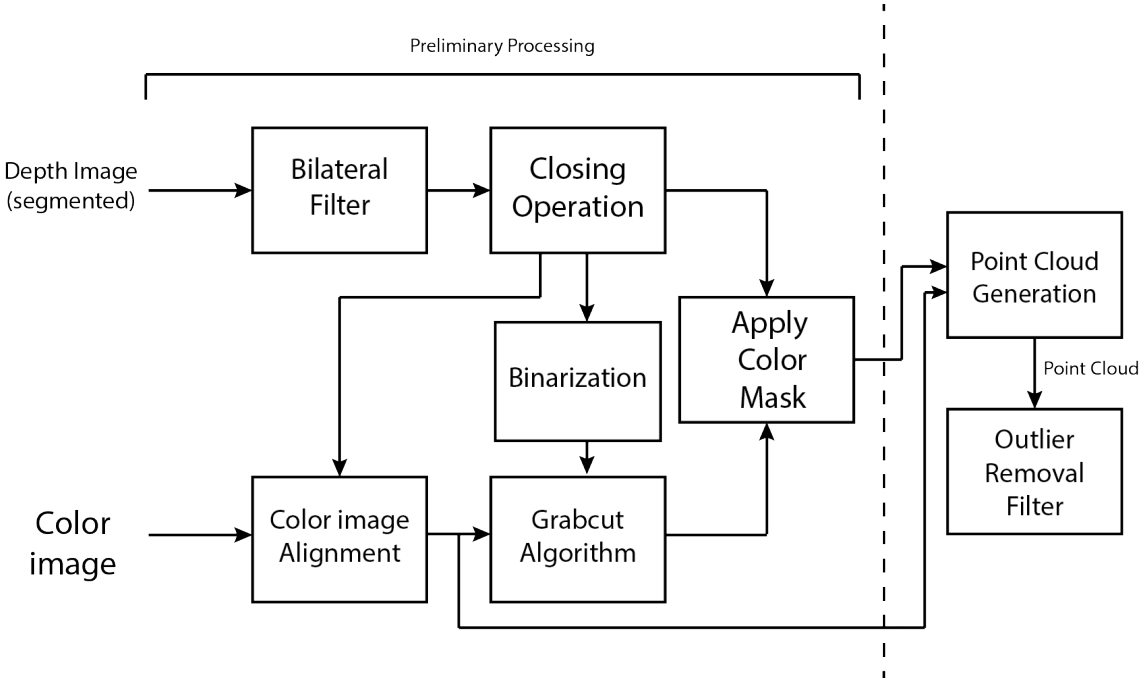


Figure 4.26: 3D data Generation pipeline.

4.4.1 Preliminary Processing

Due to its smoothing surfaces and edge conservative properties, the bilateral filter was chosen to work with the 2D depth maps captured by the Kinect, see Figure 4.27 . It was implemented in MATLAB using the mathworks scripts available with this filter functions, following the original process developed by C. Tomasi and R. Manduchi [57].

Since it is a technique which cares about keeping the edges, in order to find the invalid pixel values, a closing morphology operation [76] is done with a disk of radius 5. This configuration was found empirically to fill invalid regions which may appear due to depth estimation failures from the Kinect's depth camera and also avoid the introduction of points and artifacts that were removed previously in the segmentation process, as seen the example in Figure 4.28.

Since for this thesis, the equipment of choice has been the Microsoft Kinect, the Kinect Software Development Kit (SDK) must to be used to manage the calibration setup which is necessary to rectify the alignment between the depth and RGB data. Using the depth map as the reference, is aligned such that each color pixel corresponds to the same point in the real world.

The aligned color image is then used for a process of segmentation with the Grabcut algorithm, published by Carsten Rother *et. al* in 2004 [77], to extract the foreground and generate a mask.

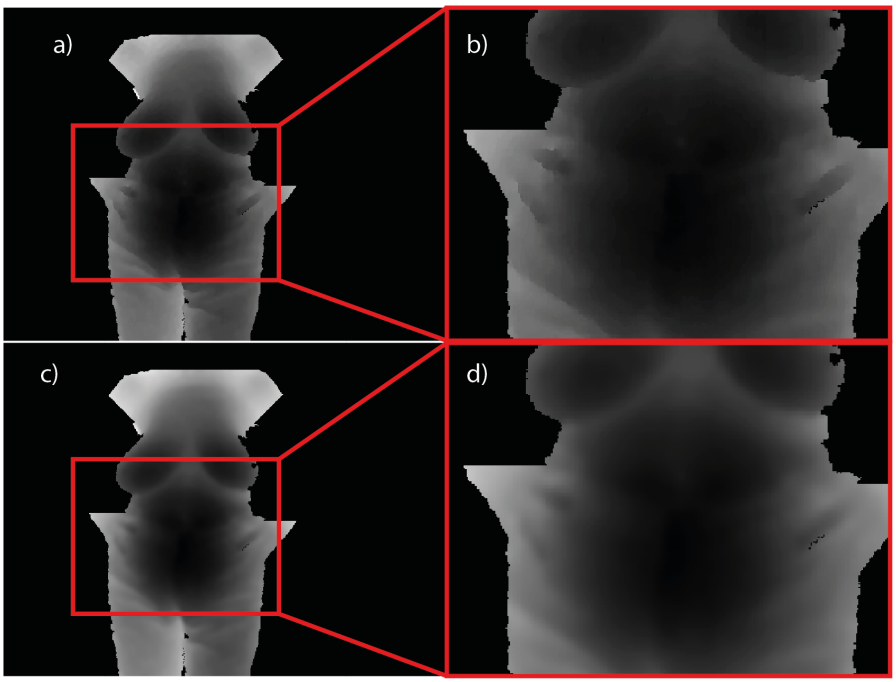


Figure 4.27: The depth images in the (b) and (d) are the same in (a) and (c) zoomed, respectively. The images (a) and (b) are the input and the depth images, in (c) and (d) are results after the Bilateral Filter.

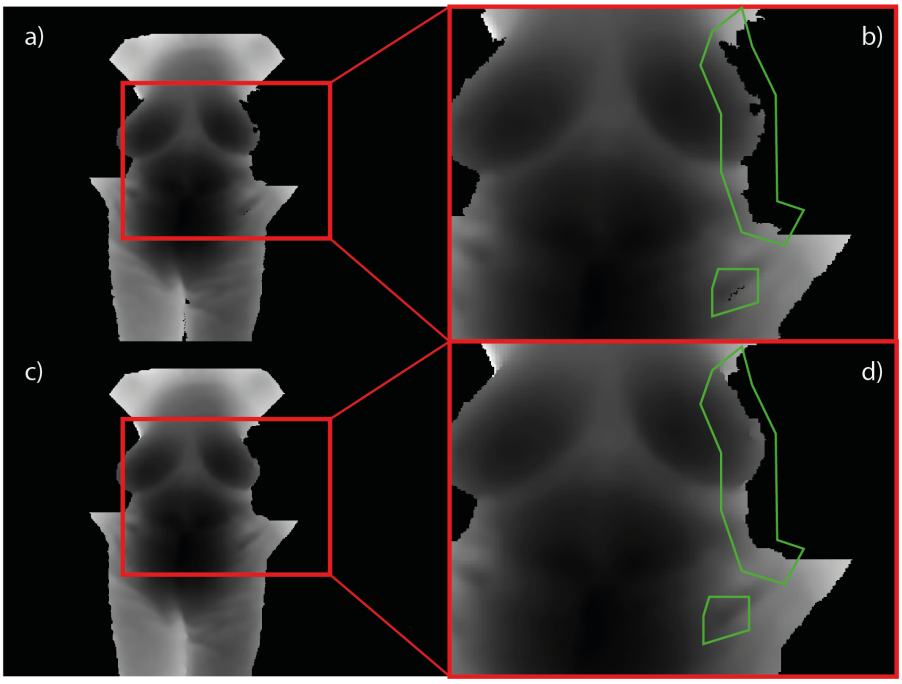


Figure 4.28: The depth images in the (b) and (d) are the same in (a) and (c) zoomed, respectively. The images (a) and (b) are results after the Bilateral Filter in (c) and (d) are results after the Closing Operation.

The aligned color mask is applied into the depth map image, to remove any foreground depth pixels which belong to the background in the color map.

4.4.2 Point Cloud Generation

The resulting depth map is converted into the 3D space, with each point receiving its corresponding color from the color map. Based on the Kinect calibration setup, from the used camera and saved during the acquisition.

4.4.3 Outlier Removal Filter

The principle of this filter, as described in Section 2.5 in the chapter 2, can be seen in the following pseudo-code (see algorithm 2). In a nutshell, a point is seen as an outlier if the mean distance of the k nearest neighbors is higher than a certain threshold.

input : PCin Point cloud, k neighbors, Thresh threshold
output: PCout, new filtered Point cloud

$std \leftarrow$ `getStandardDeviation`$(PCin)$;
foreach *Point p of the Point cloud PCin* **do**
  $neighbors \leftarrow$ `findNeighbors`$(PCin[p], k)$;
  $meanDistance \leftarrow$ `mean`$(PCin[p], PCin[neighbors])$;
  $normalizedDistance \leftarrow meanDistance/std$;
  if $normalizedDistance < Thresh$ **then**
    `// Save Point for the Output Point cloud`
    $PCout[p] \leftarrow$ `addPoint`$(PCin[p])$;
  end
end

Algorithm 2: Point Cloud Outlier Removal.

To look for the starting parameters in this implementation, experiments with a dataset of point clouds were done, finding some variables which must be taken into account, such as, the conditions of the acquisition, distance from the body to the camera, the density and size of point cloud achieved with the Kinect. The conclusion was that, for larger clouds, more closer neighbors should be used to achieve better results. For full size view point clouds (with about hundred thousand points in average) 100 nearest neighbors were used with a threshold of 0.05, while for the view segments, which represent nearly a quarter of the full sized clouds, proportionally just 25 neighbors were used.

This filter will remove points which appear isolated from the main group, caused by measurement errors, and also clean the irregular edges of cloud as described in R. Radu's work [59]. Additionally for this application, the filter may help correct some minor unwanted faults from previous steps of the framework's pipeline, such as the segmentation of rigid parts and the close operation from the Preliminary Processing 4.4.1.

4.4.4 Summary

The main goal of this step was to achieve point clouds with smooth surfaces, less noise, get invalid values filled with new points and remove outliers.

Figures 4.29 and 4.30 show the difference between the old framework point clouds output (Top row) against the ones that went through the new proposed approach (Bottom row), whose results have better appearance than the original, the gaps were almost completely fulfilled, edges of the breasts and torso are smoother with low artifacts and the points density in skin got better distributed, which makes it look cleaner and colors match better with their neighbors.

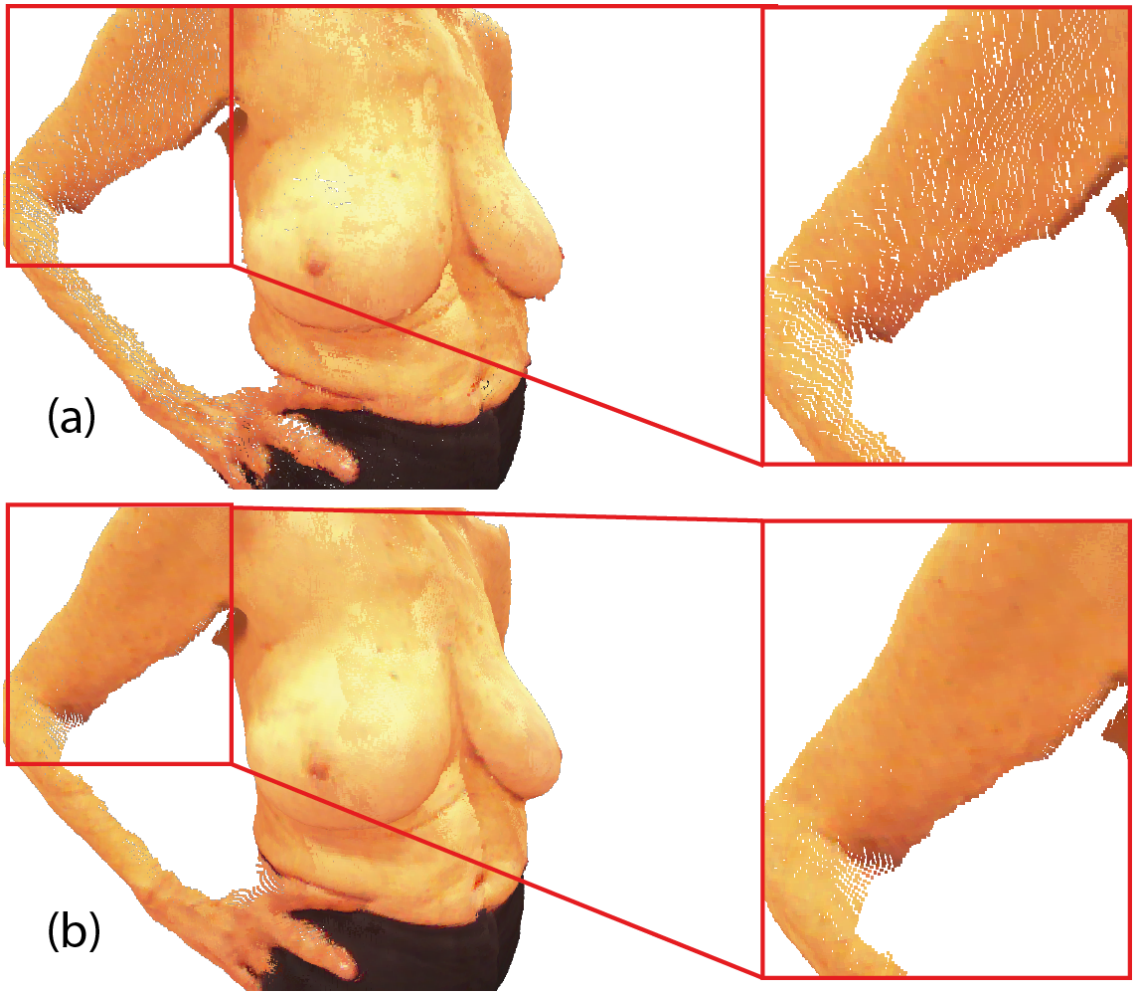


Figure 4.29: Comparative Analysis of the patient models, point distribution: (a) old framework model result without filters in preliminary processing; (b) Model results with new approach for preliminary processing.

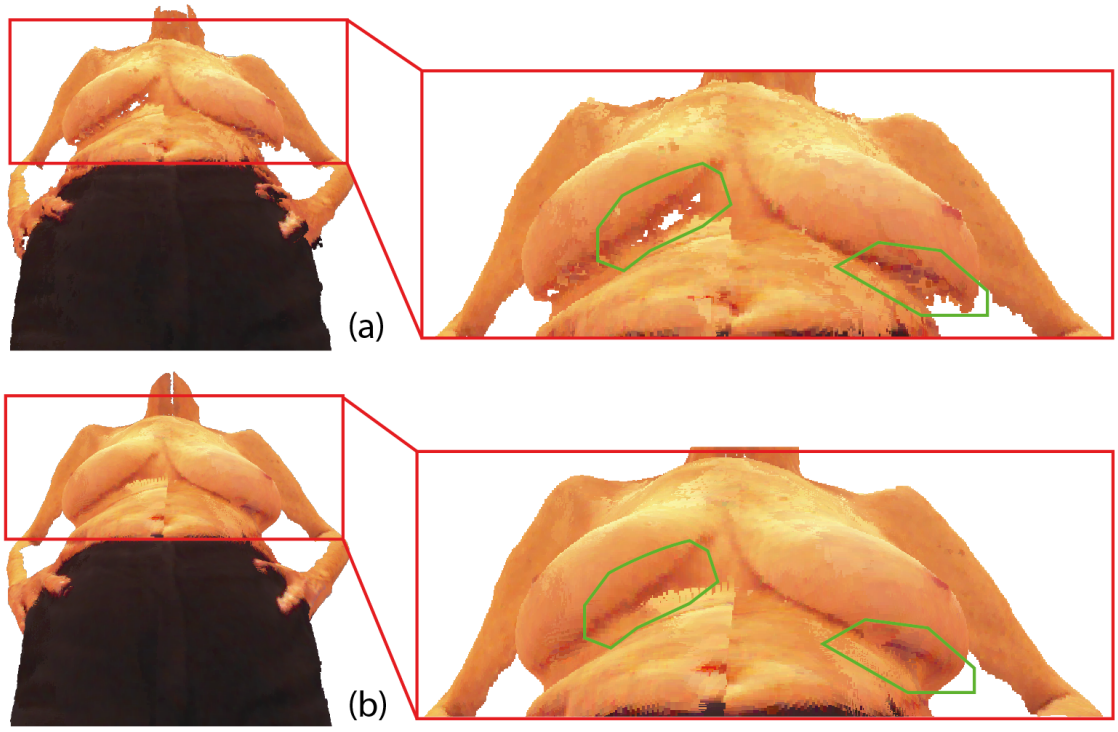


Figure 4.30: Comparative Analysis of the patient models, filling gaps: (a) old framework model result without filters in preliminary processing; (b) Model results with new approach for preliminary processing.

4.5 Rigid Registration

Registration is the method used to combine different perspectives from the object of interest with acquired 3D data into a single model. This thesis will focus only on rigid registration methods instead of non-rigid methods, given the prevalence of studies for rigid registration in computer vision, the availability of public libraries and the expected higher computational demand for non-rigid methods. Rigid registration means using just translations and rotations (while non-rigid methods includes scale and skew transforms) to determine the best transformation, which maximizes the matching between two different 3D samples.

Reviewing the previous work on the old framework (see chapter 3), the registration process can be decomposed in three main components, as described by the diagram in Figure 4.31.

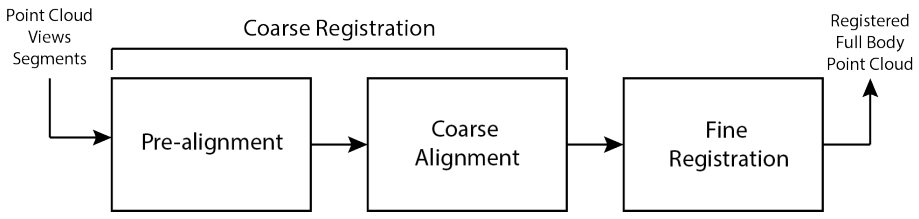


Figure 4.31: A high-level diagram of the registration process.

As already explained briefly in the Section 4.3, in order to achieve more successful registrations, non-rigid body part have to be removed. The previously defined point clouds denoted as rPC are the point clouds which are going to be used for registration, while the denoted wPC will receive the estimated transforms that were found with the rPC. The main goal for the registration in this application is to match both lateral poses with the reference frontal pose.

Since the point clouds refer to different instants of capture from a fixed camera, they will appear overlapped with different orientations as shown in Figure 4.32a). In order to help the registration and maximize the matching, an angle of 25°degrees (average rotation difference between lateral and frontal poses) is used to perform a rotation transform for each lateral view's point clouds, like in Figure 4.32b, to make them aim in the same direction as the frontal view, which is always used as reference.

Then, a coarse alignment is followed, by doing a registration with the Iterative closest Point method, using the frontal view as a reference and a down-sampled version of every view point cloud obtained from the Delaunay Triangulation principle as suggested by Pedro Costa work [1] (Figure 4.32c).

Finally, for the fine registration, the last transforms estimations are computed with the point clouds which resulted from the torso segmentation step, after going through the coarse registration. The Rigid Registration method chosen was the Iterative Closest Point (ICP) using the frontal view as reference to find the best transform for each one of the lateral point clouds (Figure 4.32d).

The computed transforms are applied to the full lateral view point clouds versions and combined together with the frontal view into the single 3D model reconstructed in the Figure 4.32e.

The Point Cloud Library (PCL) is an open project library [78] for 2D/3D image and point cloud processing, containing several state-of-the art algorithms for filtering, feature estimation, registration, model fitting and others. Its development has been done from a large number of different organizations around the world and supported by well known technology companies, such as, Toyota, Nvidia, Google, Leica and Intel [79].

For the framework's Registration module, this thesis will focus on analyzing different approaches for the fine rigid registration block in the framework's diagram Figure 4.31. The methods provided in the Point Cloud Library (PCL) were used and their performance was compared, focusing the execution time and the mean and Hausdorff distance errors. The discussion about the findings will be done in Chapter 5.

The different Fine Registration methods used for comparison were the following:

- Iterative Closest Point - Original Point-to-Point [42];
- Iterative Closest Point - Non-Linear [46];
- Iterative Closest Point - Point-to-Plane [44];
- Iterative Closest Point - Generalized [47];
- Iterative Closest Point - Point-to-Plane Estimation with Levenberg Marquardt [45];

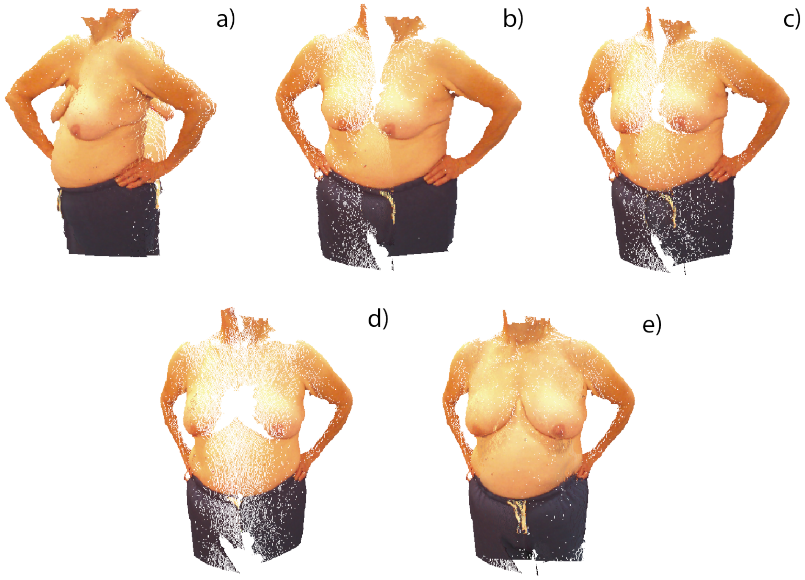


Figure 4.32: Framework Registration procedure, Top: a) Input Lateral views point clouds, b) Pre-alignment, c) Coarse alignment; Bottom: d) Fine Registration, e) Patient Complete Point cloud model after adding the frontal view.

For the reconstruction of the models in every scenario and methods, the same stop criteria options were applied for all patients:

- Transform Epsilon: 1e-25, The epsilon (difference) between the previous transformation and the current estimated transformation is smaller than an user imposed value;
- Euclidean distance between two Point Clouds: 1e-9, The sum of Euclidean squared errors is smaller than a user defined threshold;
- Max iterations: 1000000, Number of iterations which has reached the maximum user imposed number of iterations;

With the new set of methods proposals for this framework, a way to analyse everything was established by testing every registration method in all patients in four different scenarios with a combination of conditions, in order to find which method/scenario provides the overall best performance for this kind of application. The results and their discussion is present in Chapter 5.

This four scenarios can be described as:

- Original old framework configuration, with manual segmentation and no filters being applied in any point cloud.

- Automatic segmentation with all proposed methods for the module '3D data generation' used just for wPC_{LFR}.
- Automatic segmentation with all proposed methods in the module '3D data generation' used for wPC_{LFR}, while for the rPC_{LFR} just the preliminary processing proposals are used, which includes Bilateral filter and Closing Operation without the outlier removal filter.
- Automatic segmentation with all proposed methods in the module '3D data generation' applied for both wPC_{LFR} and rPC_{LFR}, including outlier removal filter.

4.6 Final Summary

New approaches were proposed to solve the main issues of the old framework. In order to reduce the need for user input, an algorithm was developed to automatically select the patient's main poses. After choosing the frontal and lateral views, it was done an implementation of a procedure to extract, from the depth maps, the rigid body parts to help and assure the rigid registration module.

Looking at problems described in the Section 4.1 to encounter the noise introduced in the data by the camera, new processing steps for the 3D data generation were proposed. This proposal intends to implement a module in the framework to: smooth the surface with a bilateral filter, fill the gaps using the closing operation and after generating the point cloud an outlier removal filter is applied.

Finally, a test was implemented with different scenarios of using manual or automatic segmentation, using Processing for 3D data and apply different fine registration methods.

Chapter 5

Results and Discussion

In this chapter, the results from the algorithms and methodologies proposed in Chapter 4 are presented and discussed. The tests were made to evaluate the performance of the automation in selection of the poses and the segmentation of rigid body parts. Additionally, it is done a discussion about the different proposed scenarios for the Fine Registration methods in combination with the '3D data Generation for each View' module described in Section 4.4.

For the evaluation of the timings performances, the tests were done in an Intel Core i7-3770K CPU @ 3.50GHz, 16GB RAM (64-bit) computer.

5.1 Pose Selection

The automatic pose selection, proposed in Section 4.2, is an important step of the pipeline for picking the three main views, which will be crucial for the model's reconstruction. In order to test the algorithm, a dataset was used, which is composed of video depth and color frames recorded from 23 different patients, following the acquisition protocol (see Appendix A). The validation is based on the pose boundaries annotations, in which the images are associated with a given pose according to manually selected frames. For example, given all the frames, certain interval ranges can be considered as the left, frontal and right pose classes. The automatic selection is considered successful if the selected frames for each pose are contained in the corresponding intervals above defined.

In table 5.1, it is displayed the algorithm's success rate for each pose selection and an average percentage error distance from the missed selections to the respective pose limits, given the footage size from the respective patient. The full results for each patient are available in the Appendix B.

Regarding the execution time, this algorithm took in average $25.10 seconds$ for a given patient with a standard deviation $12.43 seconds$, which is proportional to the number of frames acquired during a rotation.

The results have showed a solid performance on finding the frontal pose, while the lateral poses despite their good ratings, were not so robust possibly because of the irregularities from the K-means segmentation, which affected the determination of the nearest region. The patients may

Table 5.1: Pose selection Success Rate results in percentage and an average percentage distance error from missed selections.

Pose	Success Rate (%)	Average Distance Error (%)
Frontal	100.0	0.0 ± 0.0
Right	87.0	13.4 ± 10.4
Left	78.3	7.9 ± 8.5

not always followed the protocol correctly, for example by not rotating in a constant speed, their arms not be positioned correctly and unwanted body movements. These inconsistencies may lead to less reliable closest area measurements and consequently selecting lateral poses incorrectly.

5.2 Body Part Segmentation

As explained in the Section 4.3, it is important to remove the non-rigid body parts, in order to help the rigid registration step. To evaluate the performance of the segmentation algorithm, segmentations for the previous selected 23 patients were done manually for each pose in the 3D space by selecting the points to be removed. Then, thanks to the Microsoft Kinect SDK, this point cloud is converted into depth-mask for comparison with the generated results from the proposed algorithm. The validation task was performed with a similarity test between the automatic and manually binary masks segments by using the Jaccard Index [80] with equation 5.1 and the Sørensen–Dice index [81] with equation 5.2. Additionally, an error metric is computed based on the percentage of manual segmentation pixels missing in the automatic segmentation mask.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{5.1}$$

$$D(A,B) = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{5.2}$$

$$MissingArea(A,B) = (1 - \frac{|A \cap B|}{|A|}) * 100 \tag{5.3}$$

Table 5.2 presents the results of the average indexes (Jaccard and Sørensen–Dice), their standard deviation and the missing pixels error for each pose. The tests were only done for the wPC_F depth mask and the three rPC depth masks that were used for the registration, since they were the only segmentations done in the depth maps. The full results for each patient are available in the Appendix B.

From the results in table 5.2, it is possible to see that the segmentation algorithm is able to achieve good levels of similarity to the reference. As for the lateral poses(rPC_L and rPC_R), they seem to give worse segmentations, possibly due to the patient inconsistent rotation, leading to body movements such as tilting or not keeping an erect position which may change their height. Consequently, the torso gets out of applied top and bottom limits that were found earlier in the

Table 5.2: Similarity Indexes Averages and Error for each Segment Results.

Views Segments	Jaccard Index	Sørensen–Dice Index	Missing area (%)
wPC_F depth map	0.89 ± 0.03	0.94 ± 0.02	2 ± 1.40
rPC_F depth map	0.77 ± 0.07	0.87 ± 0.05	15 ± 7.55
rPC_L depth map	0.72 ± 0.07	0.84 ± 0.05	17 ± 9.54
rPC_R depth map	0.68 ± 0.07	0.81 ± 0.05	20 ± 9.55

frontal image. Although some of the segments results show Missing areas between 15% and 20%, it will be shown in the next section that it is still possible to perform registration correctly. This comes with no surprise, taking into account that the manual segmentation used for this test was done only by one annotator. Looking at [82], it shows how much variation the results may get from different annotators, which leads to the conclusion that these results cannot provide the true performance of the algorithm but just an approximation evaluation. This same principle may be applied to the selection Pose algorithm and its tests results.

5.3 Rigid Registration and Filtering

In order to obtain the 3D models, the registration procedure has to be done such that all the views are transformed into a single model. As previously said in the Section 4.5, four scenarios were tested in order to perform a comparative analysis between the different registration methods and observe the impact of using the '3D data Generation' module from Section 4.4 for rPC_{LFR} (point clouds used for registration estimation) and wPC_{LFR} (point clouds to apply the estimated transformation) from manual and automatic segmentation. This evaluation is done by using the resulting registered models, from a dataset of 7 patients, against the *Ground Truth* models obtained from the high precision sensor 3dMD, see section 2.6.3. These tests will include the mean euclidean distance error of the model's points in both directions, Hausdorff Distance and execution time comparison.

Table 5.3 describes how each scenario is composed from the possible different conditions, with associated labels, used for the evaluation of Rigid registration methods and Point Cloud Processing:

These are the different Registration Methods, with denoted labels, as will be seen hereafter for the evaluation results:

- M1 - Iterative Closest Point - Original Point-to-Point [42];
- M2 - Iterative Closest Point - Non-Linear [46];
- M3 - Iterative Closest Point - Point-to-Plane [44];
- M4 - Iterative Closest Point - Generalized [47];

Table 5.3: Combination of conditions for each scenario.

Conditions \Labels	S1	S2	S3	S4
Manual Segmentation	√			
Automatic Segmentation		√	√	√
Preliminary Processing + Outlier Removal for wPC		√	√	√
Preliminary Processing for rPC			√	√
Outlier Removal for rPC				√

- M5 - Iterative Closest Point - Point-to-Plane Estimation with Levenberg Marquardt algorithm [45];

From these described labels, it is relevant to notice that the combination S1 M1 refers to the old framework configuration, where the Registration Method is the Iterative Closest Point, Original Point-to-Point approach, with manual segmentations and no use of Processing for the 3D data.

Figure 5.1 presents a graph of the average execution time, for each registration method different scenario.

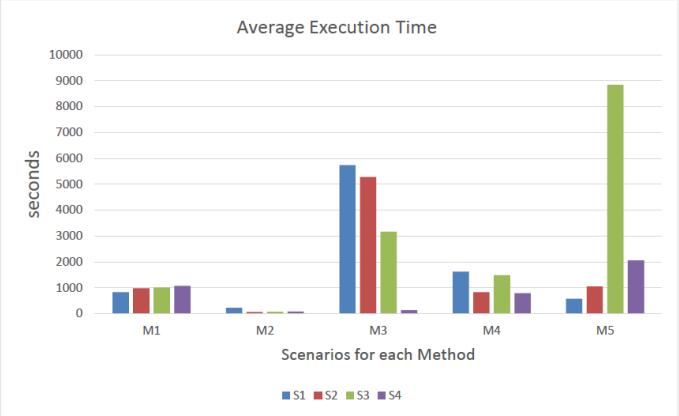


Figure 5.1: The average execution time in seconds for each method in the different scenarios.

The Standard deviation (stdev) for the Figure 5.1 is not represented, due to fact that the stdev for one of the patients was 17811 seconds what would yield an unintelligible plot. Nonewithstanding, the values may be seen in the Appendix B.

In the next tables, the results values were computed with the free software CloudCompare to calculate the error distances, using a matching function with 3dMD model as reference to all other generated models for comparison. The full results for each patient are available in the Appendix B.

Figure 5.2 presents the mean distance error obtained when testing the methods from the different scenarios in the 7 patients.

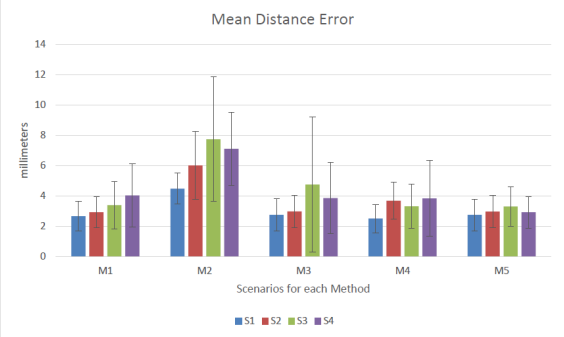


Figure 5.2: The mean euclidean distances error (in millimeters) for each method and scenario, using the direction from the 3dMD model to the Microsoft Kinect model.

From Figure 5.3, it is possible to observe that the error in the direction of the Microsoft Kinect model to the 3dMD model is homogeneous among the different scenarios and methods. This happened because the former model has a much larger point cloud than the latter.

Figure 5.3: The average mean euclidean distances error (in millimeters) for each method and scenario, using the direction from the Microsoft Kinect model to the 3dMD model.

As for the Hausdorff Distance, Figure 5.4 contains the results from the 3dMD model to the Microsoft Kinect Models, it is a metric used to measure how far they are from each other, providing the longest distance from a given point of the reference point cloud to the nearest point from another point cloud [83]. Doing an overview of the results, they follow the tendencies of the mean distance errors, this is, for cases where the latter is higher, the former is also higher.

From the results, several conclusions can be taken. From all the scenarios, S1 is the one which has showed the lowest mean distance error, since the other 3 scenarios were all obtained with the automatic segmentation depth maps and filtering treatments for the model's point clouds. The accuracy of segmentation algorithm's results and the possible unnoticeable distortion introduced by the filters explain this behavior, despite achieving better looking models.

For the first scenario, the method with the lowest errors was the M4, which corresponds to the Generalized-ICP, being able to step ahead of the standard ICP method used in the old framework. However, by looking at the execution time, it took about the double of the time to achieve this improvement which is near to a decimal of a millimeter, concluding that its use is not justified if

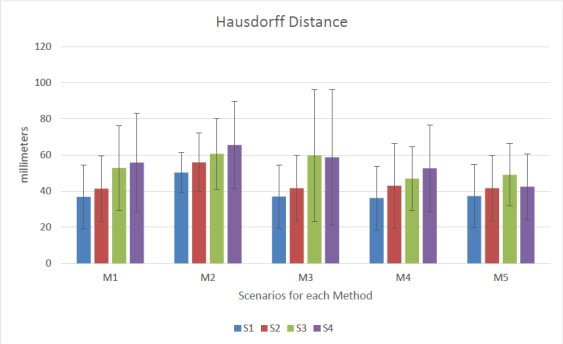


Figure 5.4: The average Hausdorff Distance (in millimeters) for each method and scenario, using the direction from the 3dMD model to the Microsoft Kinect model.

time is an important factor.

The method M2, which refers to the ICP non-linear, despite having really short time performances, the mean distances are the highest and that is noticeable by looking at the resulting models, confirming the registration failures on matching the views with almost every scenario and patients.

The scenarios S1 and S2 are the best in the error distances metric for the registration methods with the point-to-point approach. Meanwhile, from the scenario S3 to scenario S4 it is observable an improvement, mainly for point-to-plane method approaches, for both time and error metrics. It is very likely that this occurred due to the use of the outlier removal filter, which is able to remove the distortion from lateral rPC, introduced with Preliminary Treatment and therefore improve the normals estimation to be used in the point-to-plane approach.

Although the dataset used for this evaluation was limited to 7 patients, it is possible to address its diversity of cases. By looking at the Appendix B in the mean distance from the 3dMD model to the Microsoft Kinect model tables, it is observable, for example, in the last two scenarios the patient PX_018_026_N had bigger errors due to some failures in registration, which also contributed for bigger time consumption. However, from the median execution time results (see the tables in Appendix B) and mean distance errors, without considering the outlier's results which have great influence for methods such as M5, the performance reaches levels of error similar to the old framework, while being faster and providing better appearance models.

If the determining factor is the timing associated with good visual results, then the preferred option is the use of ICP point-to-plane (M3) in the last scenario. Against the old framework (S1 scenario with method M1) it is able to achieve an acceptable average mean distance error of just 1 millimeter in difference, performing four times faster and gives a model with more pleasant visuals due to the Preliminary Treatment and Outlier Removal Filter. As reported in the Literature Revision, see section 2.3, the point-to-plane approaches demand higher computational power, which is noticeable with large time consumption in data with a lot of noise as in S1 and S2. Although here is presented a situation where is possible to attenuate that impact, by smoothing the object of interest's surface. With its normals calculated for each point, these will have a better

similarity and distribution due to the noise reduction. This way, the method will be able to estimate the transform to apply in less iterations.

When looking for the model with the least mean distance error and at the same time having good visual appearance, the best scenario is S2 with the ICP standard method M1, although M3 and M5 are in average just 0.05 millimeters away, which is barely unnoticeable in the real world.

Figure 5.5 presents the visual outcome of the reconstructed 3D models of the 7 breast cancer patients with the combination of the scenario S2 with method M5. The overview of the models present very promising visual appearance results, with noise reduction and more complete surfaces.

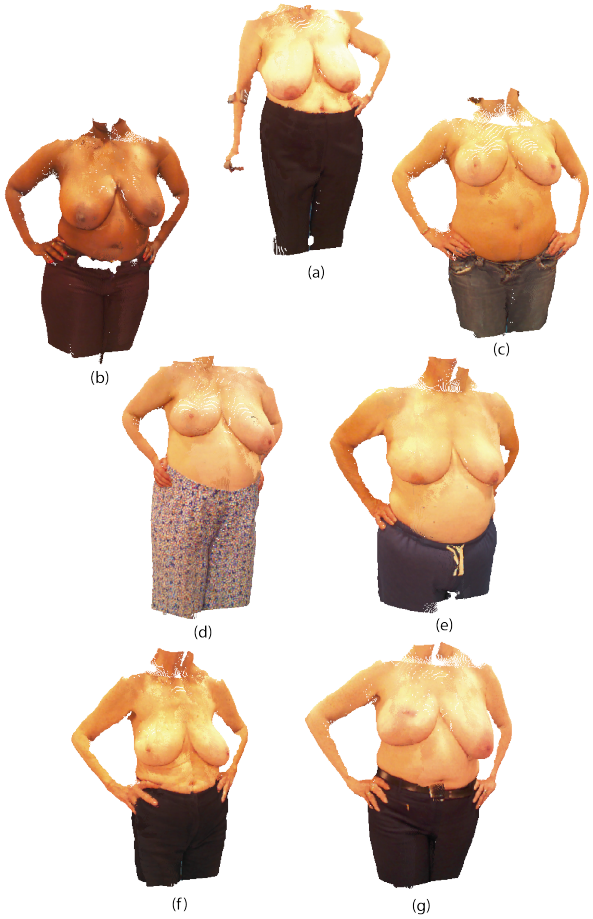


Figure 5.5: The complete 3D models of 7 patients in the scenario S2 with method M5.

The point clouds of every patient for every combination of scenario and registration method are available in the dissertation webpage[1] for download.

[1] https://paginas.fe.up.pt/~ee11098/Tese/PointClouds.html

5.4 Color Correction

Although the color correction was not a defined objective for this thesis, a simple test was done to verify the Texture Mapping module from the old framework, as referred in chapter 3. The goal of this task was to see the results between the registered point clouds which went through the '3D data Generation' module (see Section 4.4) and after the color transfer module step with 2D HD photos, against the output of the old framework with the whole Texture Mapping module. This comparison can be seen in the Figure 5.6, with 2 patients models from two different views side-by-side to see the differences.

Looking at the frontal views, although the original models look more uniform, they show visual artifacts which affect their photo-realism. Those artifacts become particularly relevant in the lateral views given the interpolation color step from the Texture Mapping Module. On the other hand, the new results of the proposed methods, despite showing some artifacts from the reconstructed person's rotation, present some ambitious results. The proposed method allows to lose less high frequency information and fine detail, being possible to conclude that it may have potential in order to avoid or simplify additional processing steps.

5.5 Conclusions

The methodologies proposed have showed good performance in terms of automation, registration and modelling. Solid results were achieved when considering the pose selection (mainly the frontal pose) despite the environment conditions, and the segmentation in terms of removing the non-rigid regions. Several results were obtained by combining the different methods and scenarios proposed. To decide which is the best scenario to use in the framework, the choice should take into account three main factors: processing time, mean distance error and visual results. If the main goal is to reduce the time required, the best models came from the ICP point-to-plane model (M3) in scenario S4. When the visual results have higher priority, the best scenario was S2 since it uses filters in the wPC point clouds and shows the best compromise between time of execution and mean distance error. However, if the most important factor is the mean distance error, than the first scenario (S1) using the Generalized-ICP (M4), without any point cloud visual enhancement processing, seems to be the most adequate.

Due to the fact of using heuristics and empirical approaches for the algorithms in the automatic pose selection and segmentation modules, an additional independent dataset of 23 patients was used, without any algorithm adjustment, in order to prove its robustness for other different patients. Tables 5.4 (Pose Selection algorithm) and 5.5 (Body Part Segmentation algorithm) show the results which validate the proposed algorithms to automatize the framework's initial procedures since they are similar with the results to the previous dataset.

Additionally, these new 23 patients were used to verify the performance of the rigid registration with the combination of method M5 in scenario S4. The new results have showed an average execution time of 2471 ± 3284 seconds and an average distance error of 2.81 ± 0.82 millimeters

with the direction from the 3dMD model to the Kinect models. In comparison with the previous results, for the same combination of method and scenario, there are no significant changes between both datasets, which demonstrates its robustness given different patients and their known non-rigid properties. The overall results of this new dataset are available at Appendix B .

Table 5.4: Pose selection results for the independent dataset, showing the Success Rate results in percentage and an average percentage of distance error between selected frames and the closest respective pose limits, given the footage size from the respective patient.

Pose	Success Rate (%)	Average Distance Error (%)
Frontal	95.7	3.8 ± 0.0
Right	78.3	19.7 ± 12
Left	78.3	20.5 ± 13

Table 5.5: Segmentation algorithm results for the independent dataset, showing the Similarity Indexes Averages and Error for each Segment Results.

Views Segments	Jaccard Index	Sørensen–Dice Index	Missing area (%)
wPC_F depth map	0.89 ± 0.02	0.94 ± 0.01	1 ± 0.60
rPC_F depth map	0.84 ± 0.05	0.91 ± 0.03	9 ± 4.99
rPC_L depth map	0.72 ± 0.06	0.84 ± 0.04	18 ± 7.86
rPC_R depth map	0.75 ± 0.06	0.86 ± 0.04	19 ± 6.69

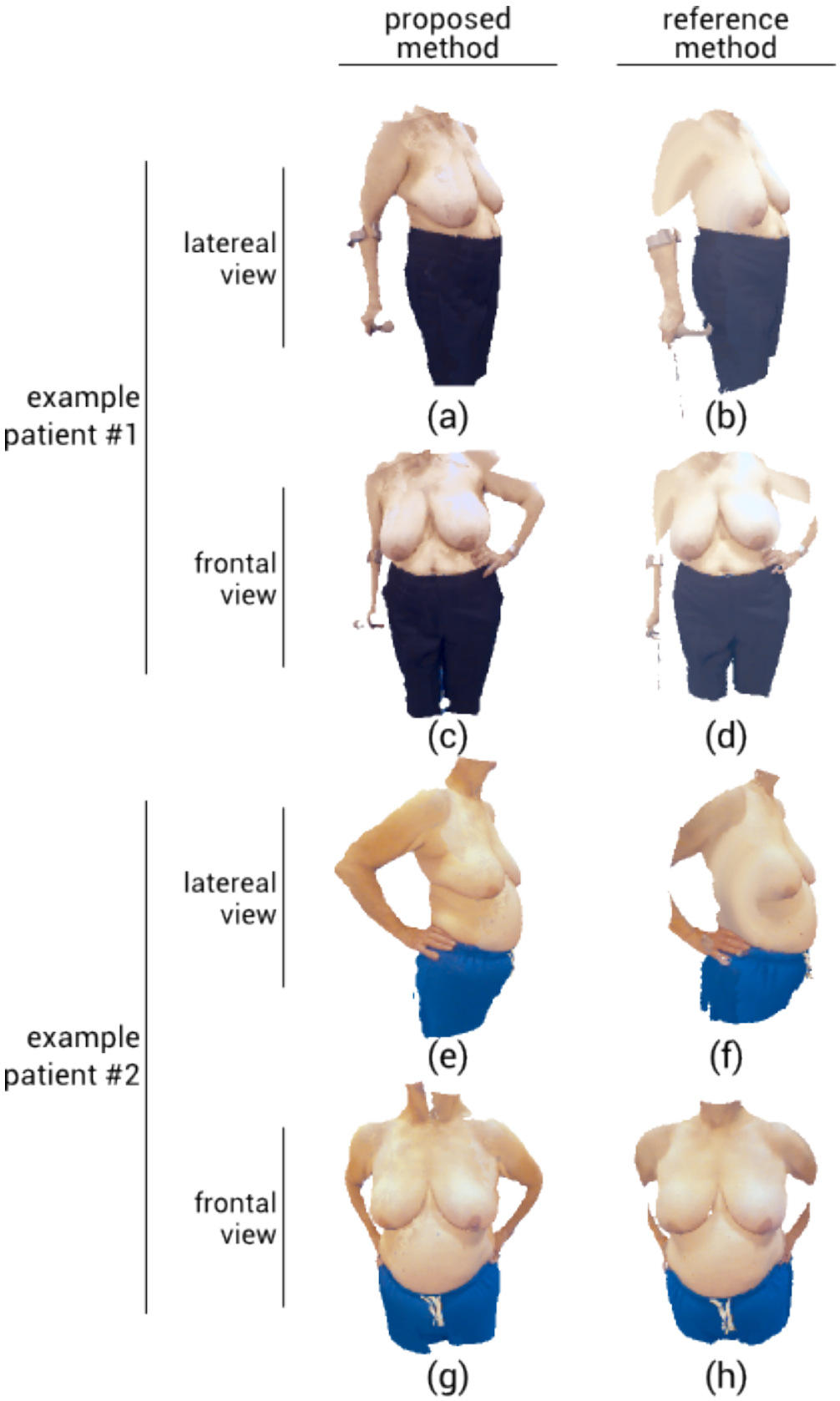


Figure 5.6: The 3D models with the proposed method (a),(c),(e) and (g) in comparison with the models done with the reference method (b),(d),(f) and (h).

Chapter 6

Conclusions

Breast Cancer is a disease which affects a huge number of women across the world and known for both physical and psychological impact. Its treatment requires an intensive monitoring and recurrent analysis to keep track of the evolution. To face this medical demand, the evolution of technology in the last years, such as the introduction of 3D modeling, has enabled the arising of tools for the acquisition and sharing information between physiologists and patients.

3D modeling has gained popularity over the last few years thanks to the development of low-cost RGB-D cameras. This equipment is able to produce 3D models with acceptable accuracy for a relative reduced cost, in comparison with traditional systems available on the market which are known to be expensive, complex to operate and require special knowledge.

A medical analysis application was developed by the VCMI group from INESC-TEC, a framework which reconstructs a 3D model of the patient's torso with the data acquired from the low-cost RGB-D device Microsoft Kinect.

This kind of medical applications for 3D modeling faces many challenges, in order to assure the correct registration of multiple views captured during a rotation. As for the human, it is known for its non-rigidity properties such as involuntary movements, change of shape and limbs articulation. Additionally, it is not possible to guarantee the correct execution of the acquisition protocol by the patients and an adequate environment, since the lighting conditions and the space may affect the camera.

This thesis proposes new approaches to improve the framework and overcome its major flaws in the four main steps of the processing pipeline: (a) pose selection, (b) rigid body parts segmentations, (c) 3D data Generation and (d) Rigid Registration. For (a) and (b) the purpose was to find a more automatic procedure to reduce user input, by selecting the patients poses from video footage and extract the rigid body parts from the views before generating the point cloud. In (c) and (d) a new 3D data Generation for each view approach is proposed and used for the conducted analysis of different scenarios, in combination with various rigid registration methods, in order to find the best configuration in terms of time execution, visual appearance and level of distortion.

The results achieved by the proposed techniques were successful at obtaining an automatic system and an improved aesthetic outcome, good enough for comparison with the high-end system

3dMD. Furthermore, due to the fact that heuristics and empirical approaches were used for these algorithms, automatic pose selection and segmentation modules, an additional independent dataset of 23 patients was tested, without adjusting the implementations, to verify its performance. The final tests showed similar results to the previous dataset, which validates the robustness for the proposed algorithms in order to automatize the framework's initial procedures.

6.1 Future Work

Regarding future work, the proposed approaches should be tested on larger and different datasets, in order to test the methods in a higher variety of patients and find more unexpected conditions, which may compromise the generation of the point clouds. Overcoming these aspects is important to pursue a more robust framework, capable of retrieving successful registrations for different patient scenarios.

Although the results have showed some improvements against the old framework, there is still room for upgrades in some modules, such as the Texture Mapping Module, where the colors provided by the Microsoft Kinect do not match the real patient's color skin due to different light acquisition condition. This module needs automation and rework in order to fill gaps and assign acceptable color for these new points. After obtaining the best point cloud possible, the mesh has to be generated for a complete surface of the model.

Finally, this system needs to have the framework able to work independently of the acquisition technology.

Appendix A

Acquisition Protocol

For this section it is presented the image acquisition protocol used to obtain the data from patients. This process works with the patient spinning smoothly and constantly in its own vertical axis along 180 degrees, while the 3D camera is records the depth and RGB images, such as the Microsoft Kinect.

A.1 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 ~90 centimeters 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 180°rotation between lateral views, performed as smoothly as the patient is able (from left to right and left to right), see Figure A.1.

Specifications

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

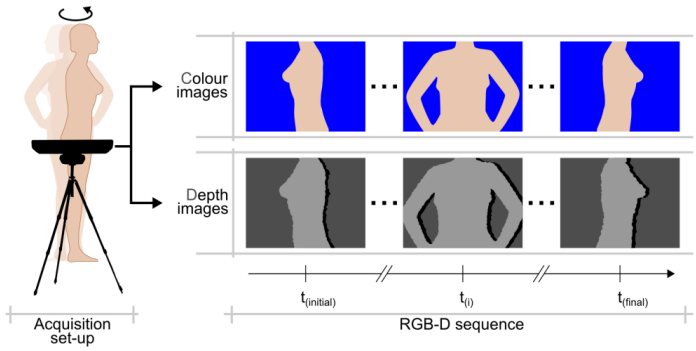


Figure A.1: RGB-D image acquisition protocol using the Microsoft Kinect.

Appendix B

Full Results

In this section the full results for each proposed improvement of the framework are presented.

B.1 Pose Selection

Table B.1: Pose Selection with each patient results.

Patients	From	Left	Frontal	Right	Number of frames	Processing Time (s)	
PX_044_001_U	Lisbon				204	15,51	
PX_044_002_Z	Lisbon				186	15,07	
PX_044_003_E	Lisbon	56		-61	232	18,82	
PX_044_004_K	Lisbon				179	13,25	
PX_044_007_B	Lisbon				672	66,41	
PX_044_012_D	Lisbon				473	42,24	
PX_044_016_A	Lisbon				222	17,29	
PX_044_024_K	Lisbon	-4			366	30,18	
PX_044_034_Q	Lisbon				311	24,22	
PX_045_006_R	Leiden				218	16,21	
PX_045_011_T	Leiden				93	6,36	
PX_045_017_A	Leiden				175	11,79	
PX_045_019_M	Leiden				129	8,57	
PX_045_037_A	Leiden	19		-41	314	24,33	
PX_045_043_Z	Leiden	19			263	19,73	
PX_045_046_R	Leiden				148	10,50	
PX_018_025_G	London	-6		4	539	44,01	
PX_018_026_N	London				515	40,60	
PX_018_029_D	London				502	40,93	
PX_018_033_A	London				512	42,07	
PX_018_060_P	London				633	65,42	
PX_018_077_X	London				334	26,65	
PX_018_098_C	London				396	33,44	
Total Patients		23	23	23	331,13	27,55	Average
Poses Found		18	23	20			
Success Rate		78,3%	100,0%	87,0%			

Table B.2: Pose Selection with each patient results, for the independent dataset.

Patients	From	Left	Frontal	Right	Number of frames	Processing Time (s)	
PX_018_028_Y	London	146		-144	483	22,25	
PX_018_032_V	London	120		-125	422	15,99	
PX_018_048_Y	London				433	22,51	
PX_018_051_Q	London				404	20,33	
PX_018_052_V	London				598	26,38	
PX_018_055_M	London	86		-69	250	12,00	
PX_018_057_X	London				517	25,68	
PX_018_059_H	London	-25			408	19,50	
PX_018_062_R	London				514	25,19	
PX_018_071_Q	London				490	23,34	
PX_018_079_H	London				385	17,11	
PX_018_082_R	London				390	17,54	
PX_018_084_B	London				440	20,20	
PX_018_085_E	London				537	24,59	
PX_018_086_N	London			-11	423	20,50	
PX_018_087_T	London		-17	-40	448	22,15	
PX_018_088_Y	London				467	22,60	
PX_018_090_J	London	-14			449	24,56	
PX_018_091_Q	London				542	21,54	
PX_018_092_V	London				577	28,33	
PX_018_095_M	London				542	26,32	
PX_018_096_S	London				669	37,48	
PX_018_100_P	London				499	25,11	
Total Patients		23	23	23	473,35	22,66	Average
Poses Found		18	22	18			
Success Rate		78,3%	95,7%	78,3%			

B.2 Body Segmentation

Table B.3: Segmentations Similarity Jaccard Index

Patients	From	Jaccard Index			
		wPC_f	rPC_f	rPC_R	rPC_L
PX_044_001_U	Lisbon	0,94	0,72	0,65	0,61
PX_044_002_Z	Lisbon	0,96	0,87	0,73	0,66
PX_044_003_E	Lisbon	0,89	0,80	0,78	0,74
PX_044_004_K	Lisbon	0,92	0,68	0,58	0,56
PX_044_007_B	Lisbon	0,91	0,74	0,67	0,56
PX_044_012_D	Lisbon	0,84	0,60	0,72	0,71
PX_044_016_A	Lisbon	0,82	0,74	0,79	0,80
PX_044_024_K	Lisbon	0,90	0,77	0,76	0,70
PX_044_034_Q	Lisbon	0,88	0,81	0,76	0,66
PX_045_006_R	Leiden	0,89	0,79	0,77	0,65
PX_045_011_T	Leiden	0,89	0,79	0,78	0,77
PX_045_017_A	Leiden	0,89	0,82	0,77	0,76
PX_045_019_M	Leiden	0,91	0,79	0,73	0,73
PX_045_037_A	Leiden	0,90	0,71	0,63	0,69
PX_045_043_Z	Leiden	0,90	0,80	0,76	0,66
PX_045_046_R	Leiden	0,82	0,63	0,55	0,57
PX_018_025_G	London	0,85	0,86	0,80	0,77
PX_018_026_N	London	0,87	0,76	0,71	0,63
PX_018_029_D	London	0,87	0,80	0,80	0,80
PX_018_033_A	London	0,89	0,84	0,78	0,70
PX_018_060_P	London	0,90	0,84	0,73	0,68
PX_018_077_X	London	0,90	0,79	0,72	0,66
PX_018_098_C	London	0,87	0,70	0,70	0,59
Average		0,89	0,77	0,72	0,68
Median		0,89	0,79	0,73	0,68
Standard deviation		0,03	0,07	0,07	0,07

Table B.4: Segmentations Similarity Dice Index

Patients	From	Sørensen–Dice Index			
		wPC_f	rPC_f	rPC_R	rPC_L
PX_044_001_U	Lisbon	0,97	0,84	0,79	0,76
PX_044_002_Z	Lisbon	0,98	0,93	0,84	0,79
PX_044_003_E	Lisbon	0,94	0,89	0,88	0,85
PX_044_004_K	Lisbon	0,96	0,81	0,74	0,72
PX_044_007_B	Lisbon	0,95	0,85	0,80	0,72
PX_044_012_D	Lisbon	0,91	0,75	0,84	0,83
PX_044_016_A	Lisbon	0,90	0,85	0,88	0,89
PX_044_024_K	Lisbon	0,95	0,87	0,86	0,83
PX_044_034_Q	Lisbon	0,94	0,89	0,87	0,79
PX_045_006_R	Leiden	0,94	0,88	0,87	0,79
PX_045_011_T	Leiden	0,94	0,88	0,87	0,87
PX_045_017_A	Leiden	0,94	0,90	0,87	0,86
PX_045_019_M	Leiden	0,95	0,88	0,85	0,84
PX_045_037_A	Leiden	0,95	0,83	0,77	0,82
PX_045_043_Z	Leiden	0,95	0,89	0,86	0,86
PX_045_046_R	Leiden	0,90	0,77	0,71	0,72
PX_018_025_G	London	0,91	0,92	0,89	0,87
PX_018_026_N	London	0,93	0,86	0,83	0,77
PX_018_029_D	London	0,93	0,89	0,89	0,89
PX_018_033_A	London	0,94	0,91	0,88	0,83
PX_018_060_P	London	0,94	0,91	0,84	0,81
PX_018_077_X	London	0,95	0,88	0,84	0,80
PX_018_098_C	London	0,93	0,82	0,82	0,74
Average		0,94	0,87	0,84	0,81
Median		0,94	0,88	0,85	0,82
Standard deviation		0,02	0,05	0,05	0,05

Table B.5: Segmentations Missing Pixels Error

Patients	From	Missing Pixels			
		wPC_f	rPC_f	rPC_R	rPC_L
PX_044_001_U	Lisbon	0,16%	24,38%	33,56%	36,54%
PX_044_002_Z	Lisbon	1,68%	2,06%	1,03%	4,47%
PX_044_003_E	Lisbon	2,66%	10,34%	13,14%	16,99%
PX_044_004_K	Lisbon	0,31%	25,47%	35,83%	34,91%
PX_044_007_B	Lisbon	0,09%	20,70%	28,48%	28,27%
PX_044_012_D	Lisbon	0,05%	23,25%	18,84%	20,50%
PX_044_016_A	Lisbon	0,02%	19,15%	13,23%	11,27%
PX_044_024_K	Lisbon	0,22%	16,58%	10,80%	16,13%
PX_044_034_Q	Lisbon	0,60%	12,93%	17,73%	26,06%
PX_045_006_R	Leiden	1,92%	13,54%	12,37%	25,63%
PX_045_011_T	Leiden	1,01%	13,28%	9,57%	9,79%
PX_045_017_A	Leiden	2,83%	7,40%	12,84%	17,77%
PX_045_019_M	Leiden	1,07%	14,12%	20,04%	14,73%
PX_045_037_A	Leiden	2,85%	22,15%	30,60%	23,90%
PX_045_043_Z	Leiden	0,89%	8,13%	12,08%	15,01%
PX_045_046_R	Leiden	4,50%	32,04%	33,50%	35,65%
PX_018_025_G	London	3,33%	6,40%	10,88%	9,76%
PX_018_026_N	London	4,26%	17,90%	20,46%	25,05%
PX_018_029_D	London	3,16%	8,60%	5,61%	6,90%
PX_018_033_A	London	2,83%	8,84%	4,10%	14,14%
PX_018_060_P	London	0,23%	6,93%	14,22%	7,37%
PX_018_077_X	London	0,07%	12,14%	13,98%	16,17%
PX_018_098_C	London	3,06%	25,74%	24,72%	33,97%
Average		2,00%	15,00%	17,00%	20,00%
Median		1,07%	13,54%	13,98%	16,99%
Standard deviation		1,44%	7,55%	9,54%	9,55%

Table B.6: Segmentations Similarity Jaccard Index, for the independent dataset

Patients	From	Jaccard Index			
		wPC_f	rPC_f	rPC_R	rPC_L
PX_018_028_Y	London	0,90	0,88	0,77	0,71
PX_018_032_V	London	0,87	0,78	0,74	0,68
PX_018_048_Y	London	0,90	0,83	0,76	0,72
PX_018_051_Q	London	0,90	0,83	0,69	0,83
PX_018_052_V	London	0,92	0,89	0,74	0,77
PX_018_055_M	London	0,91	0,79	0,62	0,58
PX_018_057_X	London	0,86	0,86	0,74	0,75
PX_018_059_H	London	0,82	0,89	0,73	0,70
PX_018_062_R	London	0,91	0,93	0,82	0,78
PX_018_071_Q	London	0,91	0,82	0,77	0,75
PX_018_079_H	London	0,90	0,85	0,83	0,78
PX_018_082_R	London	0,88	0,74	0,71	0,64
PX_018_084_B	London	0,90	0,80	0,73	0,72
PX_018_085_E	London	0,90	0,77	0,76	0,75
PX_018_086_N	London	0,90	0,83	0,79	0,70
PX_018_087_T	London	0,91	0,90	0,77	0,76
PX_018_088_Y	London	0,89	0,90	0,79	0,72
PX_018_090_J	London	0,90	0,90	0,75	0,73
PX_018_091_Q	London	0,89	0,84	0,78	0,68
PX_018_092_V	London	0,90	0,86	0,75	0,75
PX_018_095_M	London	0,88	0,81	0,55	0,61
PX_018_096_S	London	0,92	0,84	0,80	0,82
PX_018_100_P	London	0,89	0,80	0,84	0,69
Average		0,89	0,84	0,75	0,72
Median		0,90	0,84	0,76	0,72
Standard deviation		0,02	0,05	0,06	0,06

Table B.7: Segmentations Similarity Dice Index, for the independent dataset

Patients	From	Sørensen–Dice Index			
		wPC_f	rPC_f	rPC_R	rPC_L
PX_018_028_Y	London	0,94	0,94	0,87	0,83
PX_018_032_V	London	0,93	0,88	0,85	0,81
PX_018_048_Y	London	0,95	0,91	0,86	0,84
PX_018_051_Q	London	0,95	0,91	0,82	0,91
PX_018_052_V	London	0,96	0,94	0,85	0,87
PX_018_055_M	London	0,95	0,89	0,77	0,73
PX_018_057_X	London	0,93	0,93	0,85	0,86
PX_018_059_H	London	0,90	0,94	0,85	0,82
PX_018_062_R	London	0,95	0,96	0,90	0,88
PX_018_071_Q	London	0,95	0,90	0,87	0,86
PX_018_079_H	London	0,94	0,92	0,90	0,88
PX_018_082_R	London	0,94	0,85	0,83	0,78
PX_018_084_B	London	0,95	0,89	0,84	0,84
PX_018_085_E	London	0,95	0,87	0,86	0,86
PX_018_086_N	London	0,95	0,91	0,88	0,82
PX_018_087_T	London	0,95	0,95	0,87	0,86
PX_018_088_Y	London	0,94	0,95	0,88	0,83
PX_018_090_J	London	0,95	0,95	0,86	0,84
PX_018_091_Q	London	0,94	0,91	0,88	0,81
PX_018_092_V	London	0,95	0,93	0,86	0,85
PX_018_095_M	London	0,94	0,90	0,71	0,76
PX_018_096_S	London	0,96	0,91	0,89	0,90
PX_018_100_P	London	0,94	0,89	0,91	0,82
Average		0,94	0,91	0,86	0,84
Median		0,95	0,91	0,86	0,84
Standard deviation		0,01	0,03	0,04	0,04

Table B.8: Segmentations Missing Pixels Error, for the independent dataset

Patients	From	Missing Pixels			
		wPC_f	rPC_f	rPC_R	rPC_L
PX_018_028_Y	London	0,08%	6,27%	11,57%	16,30%
PX_018_032_V	London	0,07%	12,08%	20,05%	22,80%
PX_018_048_Y	London	0,68%	12,90%	17,29%	22,12%
PX_018_051_Q	London	0,70%	5,94%	24,70%	7,87%
PX_018_052_V	London	0,03%	4,05%	13,32%	16,33%
PX_018_055_M	London	0,73%	15,57%	32,51%	37,34%
PX_018_057_X	London	0,05%	4,92%	18,95%	14,99%
PX_018_059_H	London	0,02%	2,76%	18,03%	18,77%
PX_018_062_R	London	0,04%	1,83%	8,30%	9,47%
PX_018_071_Q	London	0,27%	11,65%	16,77%	15,30%
PX_018_079_H	London	0,48%	11,29%	7,21%	13,34%
PX_018_082_R	London	1,16%	18,38%	24,96%	22,50%
PX_018_084_B	London	0,20%	16,07%	21,12%	21,55%
PX_018_085_E	London	1,19%	18,68%	17,87%	18,98%
PX_018_086_N	London	0,41%	12,03%	14,84%	23,04%
PX_018_087_T	London	0,43%	4,85%	12,31%	16,67%
PX_018_088_Y	London	1,31%	4,64%	15,62%	12,41%
PX_018_090_J	London	0,57%	3,31%	15,44%	17,68%
PX_018_091_Q	London	0,49%	10,41%	13,97%	27,25%
PX_018_092_V	London	0,02%	4,93%	17,63%	11,38%
PX_018_095_M	London	0,40%	12,47%	43,01%	30,75%
PX_018_096_S	London	2,72%	11,46%	15,69%	14,21%
PX_018_100_P	London	0,14%	7,91%	6,80%	21,57%
Average		1,00%	9,00%	18,00%	19,00%
Median		1,44%	13,54%	13,98%	16,99%
Standard deviation		0,60%	4,99%	7,86%	6,69%

B.3 Rigid Registration and Filtering

B.3.1 Times of execution

Table B.9: Times of execution for each method in scenario S1.

Patients	From	Registration Methods timings				
		M1	M2	M3	M4	M5
PX_018_025_G	London	668,87	153,11	5839,87	813,86	512,62
PX_018_026_N	London	581,84	234,91	8662,84	1089,84	223,45
PX_018_029_D	London	819,78	141	9070,12	1103,41	317,69
PX_018_029_A	London	1030,1	284,05	3454,4	2342,5	267,35
PX_018_060_P	London	870,14	194,84	9197,3	2913,58	2004,03
PX_018_077_X	London	722,07	273,13	1540,07	976,27	161
PX_018_098_C	London	1089,09	343,57	2410,77	2095,14	564,61
Average		825,98	232,09	5739,34	1619,23	578,68
Median		819,78	234,91	5839,87	1103,41	317,69
Standard deviation		172,36	68,39	3059,94	759,07	597,84

Table B.10: Times of execution for each method in scenario S2.

Patients	From	Registration Methods timings				
		M1	M2	M3	M4	M5
PX_018_025_G	London	803,52	49,17	1678,75	533,25	826,23
PX_018_026_N	London	1073,02	59,33	11015,68	408,00	258,18
PX_018_029_D	London	626,95	87,39	5307,80	1455,44	1634,10
PX_018_029_A	London	1464,39	70,48	1727,13	1719,93	311,90
PX_018_060_P	London	767,16	60,00	1266,52	172,02	190,80
PX_018_077_X	London	977,43	63,48	5158,63	273,47	157,72
PX_018_098_C	London	1166,93	55,83	10850,38	1248,09	4005,92
Average		982,77	63,67	5286,41	830,03	1054,98
Median		977,43	60,00	5158,63	533,25	311,90
Standard deviation		261,66	11,42	3884,81	581,47	1300,26

Table B.11: Times of execution for each method in scenario S3.

Patients	From	Registration Methods timings				
		M1	M2	M3	M4	M5
PX_018_025_G	London	1198,20	56,05	159,79	1516,49	5203,45
PX_018_026_N	London	804,50	46,77	4116,09	1955,79	52277,32
PX_018_029_D	London	739,12	95,14	4708,55	1114,02	294,13
PX_018_029_A	London	1906,51	78,64	1290,06	1296,22	602,40
PX_018_060_P	London	449,09	69,05	2927,46	588,43	124,52
PX_018_077_X	London	731,50	58,02	338,99	1953,91	2705,74
PX_018_098_C	London	1242,75	111,52	8635,19	1958,99	707,70
Average		1010,24	73,6	3168,02	1483,41	8845,04
Median		804,50	69,05	2927,46	1516,49	707,70
Standard deviation		447,39	21,44	2772,30	484,66	17811,06

Table B.12: Times of execution for each method in scenario S4.

Patients	From	Registration Methods timings				
		M1	M2	M3	M4	M5
PX_018_025_G	London	1566,99	50,25	94,60	1521,32	324,90
PX_018_026_N	London	1773,20	50,59	176,51	210,14	7070,45
PX_018_029_D	London	1118,02	86,94	90,65	1064,82	2073,72
PX_018_029_A	London	990,47	110,76	131,40	789,54	269,41
PX_018_060_P	London	485,30	79,52	83,23	323,72	512,49
PX_018_077_X	London	860,29	73,29	140,12	530,73	576,16
PX_018_098_C	London	694,11	80,19	219,43	1087,12	3618,02
Average		1069,77	75,93	133,71	789,63	2063,59
Median		990,47	79,52	131,40	789,54	576,16
Standard deviation		427,16	19,54	46,53	434,17	2340,50

B.3.2 Mean Distance Error from 3dMD Model to Microsoft Kinect Model

Table B.13: Mean distance error from 3dMD Model to Microsoft Kinect Model in scenario S1

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	2,716	3,85	2,812	2,16	2,82
PX_018_026_N	London	2,155	5,531	2,147	1,998	2,167
PX_018_029_D	London	2,091	3,848	2,094	2,661	2,104
PX_018_029_A	London	1,422	2,848	1,402	1,558	1,4
PX_018_060_P	London	4,518	4,471	4,71	4,628	4,573
PX_018_077_X	London	3,602	6,11	3,858	2,457	3,929
PX_018_098_C	London	2,125	4,649	2,163	2,031	2,173
Average		2,66	4,47	2,74	2,5	2,74
Standard Deviation		0,9813	1,0177	1,0709	0,9282	1,0440
Median		2,155	4,471	2,163	2,16	2,173

Table B.14: Mean distance error from 3dMD Model to Microsoft Kinect Model in scenario S2

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	3,286	5,634	3,453	3,471	3,433
PX_018_026_N	London	2,566	8,052	2,568	6,018	2,573
PX_018_029_D	London	2,21	4,101	2,202	2,745	2,209
PX_018_029_A	London	1,455	4,151	1,462	1,96	1,453
PX_018_060_P	London	4,706	5,416	4,751	4,49	4,707
PX_018_077_X	London	3,824	10,498	4,062	3,991	4,023
PX_018_098_C	London	2,299	4,179	2,299	3,105	2,309
Average		2,91	6	2,97	3,68	2,96
Standard Deviation		1,0213	2,2443	1,0739	1,2204	1,0566
Median		2,566	5,416	2,568	3,471	2,573

Table B.15: Mean distance error from 3dMD Model to Microsoft Kinect Model in scenario S3

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	3,252	5,959	3,32	3,519	3,323
PX_018_026_N	London	6,402	11,34	15,346	6,134	5,354
PX_018_029_D	London	2,134	5,059	2,216	2,293	2,212
PX_018_029_A	London	1,45	4,722	1,454	1,493	1,446
PX_018_060_P	London	4,533	5,738	4,634	4,47	4,582
PX_018_077_X	London	3,648	16,38	3,865	2,907	3,854
PX_018_098_C	London	2,207	4,955	2,316	2,324	2,313
Average		3,38	7,74	4,74	3,31	3,3
Standard Deviation		1,5677	4,1204	4,4453	1,4558	1,2964
Median		3,252	5,738	3,32	2,907	3,323

Table B.16: Mean distance error from 3dMD Model to Microsoft Kinect Model in scenario S4

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	3,306	5,484	3,386	3,414	3,398
PX_018_026_N	London	8,69	10,919	9	9,609	2,375
PX_018_029_D	London	2,202	5,089	2,194	2,444	2,219
PX_018_029_A	London	2,4	5,292	1,445	1,43	1,44
PX_018_060_P	London	4,679	6,491	4,589	4,483	4,622
PX_018_077_X	London	4,201	10,8	4,053	3,07	4,022
PX_018_098_C	London	2,71	5,642	2,314	2,387	2,296
Average		4,03	7,1	3,85	3,83	2,91
Standard Deviation		2,0850	2,4112	2,3364	2,5167	1,0493
Median		3,306	5,642	3,386	3,07	2,375

B.3.3 Mean Distance Error from Microsoft Kinect Model to 3dMD Model

Table B.17: Mean distance error from Microsoft Kinect Model to 3dMD Model in scenario S1

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	125,124	123,328	125,477	125,219	125,413
PX_018_026_N	London	160,887	158,707	161,106	161,321	161,097
PX_018_029_D	London	107,57	104,53	107,607	106,263	107,61
PX_018_029_A	London	103,012	101,379	102,921	103,207	102,906
PX_018_060_P	London	163,601	161,282	163,397	162,967	163,694
PX_018_077_X	London	203,832	191,44	203,958	203,157	204,088
PX_018_098_C	London	117,721	109,713	117,872	117,09	117,952
Average		140,25	135,77	140,33	139,89	140,39
Standard Deviation		34,2362	32,2237	34,2417	34,2388	34,3025
Median		125,124	123,328	125,477	125,219	125,413

Table B.18: Mean distance error from Microsoft Kinect Model to 3dMD Model in scenario S2

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	126,929	125,292	126,9	126,944	126,813
PX_018_026_N	London	163,887	165,316	163,573	169,872	163,602
PX_018_029_D	London	108,915	108,414	107,865	110,928	107,83
PX_018_029_A	London	102,024	96,47	102,006	102,277	102,01
PX_018_060_P	London	162,945	159,982	163,047	161,503	163,273
PX_018_077_X	London	201,189	186,966	200,943	205,251	200,572
PX_018_098_C	London	118,306	116,277	118,403	120,913	118,414
Average		140,6	136,96	140,39	142,53	140,36
Standard Deviation		33,5699	31,2888	33,6241	34,5805	33,5618
Median		126,929	125,292	126,9	126,944	126,813

Table B.19: Mean distance error from Microsoft Kinect Model to 3dMD Model in scenario S3

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	127,258	127,337	126,983	127,533	126,779
PX_018_026_N	London	167,335	158,736	163,709	172,218	162,04
PX_018_029_D	London	107,667	103,835	107,677	108,576	107,641
PX_018_029_A	London	101,997	96,293	102,006	101,964	101,996
PX_018_060_P	London	163,069	156,595	157,02	161,925	162,879
PX_018_077_X	London	200,922	170,406	200,794	203,44	200,408
PX_018_098_C	London	117,723	113,359	117,731	120,847	117,807
Average		140,85	132,37	139,42	142,36	139,94
Standard Deviation		34,0882	27,3373	33,1641	34,7664	33,4174
Median		127,258	127,337	126,983	127,533	126,779

Table B.20: Mean distance error from Microsoft Kinect Model to 3dMD Model in scenario S4

Patients	From	Mean Distance Error				
		M1	M2	M3	M4	M5
PX_018_025_G	London	126,959	122,768	126,961	128,352	126,957
PX_018_026_N	London	156,26	167,117	162,071	160,363	163,115
PX_018_029_D	London	107,677	105,381	107,707	109,308	107,759
PX_018_029_A	London	101,433	95,998	101,954	101,621	101,968
PX_018_060_P	London	162,769	148,064	163,037	161,691	162,915
PX_018_077_X	London	200,713	189,596	200,952	201,793	200,779
PX_018_098_C	London	120,14	114,952	118,07	120,419	118,092
Average		139,42	134,84	140,11	140,51	140,23
Standard Deviation		32,8703	31,9211	33,5393	33,0276	33,5711
Median		126,959	122,768	126,961	128,352	126,957

B.3.4 Hausdorff distance from 3dMD Model to Microsoft Kinect Model

Table B.21: Hausdorff distance from 3dMD Model to Microsoft Kinect Model in scenario S1

Patients	From	Hausdorff Distance				
		M1	M2	M3	M4	M5
PX_018_025_G	London	35,1388	49,8948	35,8045	35,7904	35,7988
PX_018_026_N	London	17,0025	52,6356	19,6337	15,5171	19,6347
PX_018_029_D	London	77,1242	74,6761	77,142	74,8612	77,1419
PX_018_029_A	London	29,5551	46,6823	29,5475	25,3519	29,5442
PX_018_060_P	London	39,0405	42,4541	36,957	42,1871	39,0351
PX_018_077_X	London	26,2916	36,2524	26,3232	26,1296	26,3216
PX_018_098_C	London	33,1779	49,1699	33,1841	33,1982	33,1802
Average		36,76	50,25	36,94	36,15	37,24
Standard Deviation		17,7353	11,1679	17,3142	17,6739	17,3301
Median		33,1779	49,1699	33,1841	33,1982	33,1802

Table B.22: Hausdorff distance from 3dMD Model to Microsoft Kinect Model in scenario S2

Patients	From	Hausdorff Distance				
		M1	M2	M3	M4	M5
PX_018_025_G	London	36,3876	31,6268	36,41	36,065	36,4004
PX_018_026_N	London	21,2431	83,0211	22,2329	75,2946	22,2357
PX_018_029_D	London	83,5456	59,4777	84,2193	75,4853	84,212
PX_018_029_A	London	38,6938	68,0125	38,6983	31,4251	38,6964
PX_018_060_P	London	41,581	47,849	41,5742	44,594	41,5856
PX_018_077_X	London	32,7276	61,7279	32,7738	33,7831	32,7693
PX_018_098_C	London	35,3852	39,6681	35,4032	4,3657	35,4146
Average		41,37	55,91	41,62	43	41,62
Standard Deviation		18,2303	16,2415	18,2958	23,4838	18,2931
Median		36,3876	59,4777	36,41	36,065	36,4004

Table B.23: Hausdorff distance from 3dMD Model to Microsoft Kinect Model in scenario S3

Patients	From	Hausdorff Distance				
		M1	M2	M3	M4	M5
PX_018_025_G	London	41,6907	38,0269	41,73	39,2619	41,7322
PX_018_026_N	London	96,9124	98,0814	140,417	68,1559	65,4268
PX_018_029_D	London	80,5685	74,0299	84,2134	79,5254	84,2069
PX_018_029_A	London	38,6924	62,5357	38,6922	34,3801	38,6908
PX_018_060_P	London	41,6005	46,5365	41,5699	40,8915	41,5841
PX_018_077_X	London	36,3652	64,4973	36,3899	28,6417	36,3982
PX_018_098_C	London	33,5177	40,862	34,764	36,9819	34,7692
Average		52,76	60,65	59,68	46,83	48,97
Standard Deviation		23,3200	19,6434	36,5948	17,7246	17,2576
Median		41,6005	62,5357	41,5699	39,2619	41,5841

Table B.24: Hausdorff distance from 3dMD Model to Microsoft Kinect Model in scenario S4

Patients	From	Hausdorff Distance				
		M1	M2	M3	M4	M5
PX_018_025_G	London	41,6871	46,1786	39,2445	42,6753	39,2469
PX_018_026_N	London	112,821	119,698	140,785	93,8742	21,1592
PX_018_029_D	London	79,4354	70,8874	84,2083	84,8646	84,2172
PX_018_029_A	London	42,22804	69,3235	38,6913	38,6944	38,6904
PX_018_060_P	London	41,5743	46,67407	40,9609	43,3635	41,5857
PX_018_077_X	London	34,6181	53,0753	33,577	27,5812	36,3702
PX_018_098_C	London	38,3659	52,3805	33,594	36,6239	35,4172
Average		55,82	65,46	58,72	52,53	42,38
Standard Deviation		27,1198	24,0098	37,3088	23,9132	18,1633
Median		41,6871	53,0753	39,2445	42,6753	38,6904

References

[1] Pedro Costa. Kinect based system for breast 3d reconstruction. diploma thesis, Faculdade de Engenharia da Universidade do Porto, June 2014.

[2] M. J Cardoso, H. Oliveira, and J. Cardoso. Assessing cosmetic results after breast conserving surgery. *Journal of surgical oncology*, 110(1):37–44, 2014.

[3] 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*, 40(2):115–126, 2007.

[4] Mário Aguiar. 3d reconstruction from multiple rgb-d images with different perspectives. diploma thesis, Faculdade de Engenharia da Universidade do Porto, 2015.

[5] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. volume 56, pages 116–124. ACM, 2013.

[6] Picture project webpage. http://medicalresearch.inescporto.pt/breastresearch/index.html. Accessed: July 25, 2016.

[7] Fabio Remondino and Sabry El-Hakim. Image-based 3d modelling: A review. *The Photogrammetric Record*, 21(115):269–291, 2006.

[8] Giovanna Sansoni, Marco Trebeschi, and Franco Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009.

[9] Theo Moons, Luc Van Gool, and Maarten Vergauwen. *3D Reconstruction from Multiple Images Part 1: Principles*, volume 4. Now Publishers Inc., 2010.

[10] Dhanya S Pankaj, Rama Rao Nidamanuri, and P Bhanu Prasad. 3-d imaging techniques and review of products. In *Proceedings of International Conference on" Innovations in Computer Science and Engineering*, 2013.

[11] Hamed Sarbolandi, Damien Lefloch, and Andreas Kolb. Kinect range sensing: Structured-light versus time-of-flight kinect. *Computer Vision and Image Understanding*, 139:1–20, 2015.

[12] Yongyan YU and Yuqin SHU. 3d measurement and reconstruction based on structured light. 2015.

[13] Krystof Litomisky. Consumer rgb-d cameras and their applications. *Rapport technique, University of California*, page 20, 2012.

[14] Different rgb-d sensors. https://stimulant.com/depth-sensor-shootout-2/. Accessed: July 25, 2016.

[15] Jorge Ballester and Chuck Pheatt. Using the xbox kinect sensor for positional data acquisition. *American journal of Physics*, 81(1):71–77, 2013.

[16] Kinect tools and resources. https://dev.windows.com/en-us/kinect/tools. Accessed: July 25, 2016.

[17] Kinect for windows sensor components and specifications. https://msdn.microsoft.com/pt-pt/library/jj131033.aspx?f=255&MSPPError=-2147217396. Accessed: July 25, 2016.

[18] Hélder de Oliveira. *An affordable and practical 3d solution for the aesthetic evaluation of breast cancer conservative treatment*. PhD thesis, 2013.

[19] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.

[20] Orbbec astra. https://orbbec3d.com/product-astra/. Accessed: July 25, 2016.

[21] Intel realsense r200. https://software.intel.com/en-us/RealSense/R200Camera. Accessed: July 25, 2016.

[22] Andreas Kolb, Erhardt Barth, Reinhard Koch, and Rasmus Larsen. Time-of-flight sensors in computer graphics. In *Proc. Eurographics (State-of-the-Art Report)*, volume 6, 2009.

[23] Diana Pagliari and Livio Pinto. Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors. *Sensors*, 15(11):27569–27589, 2015.

[24] Kinect 2.0 hardware. https://dev.windows.com/en-us/kinect/hardware. Accessed: July 25, 2016.

[25] Kinect sdk 2.0. https://www.microsoft.com/en-us/download/details.aspx?id=44561. Accessed: July 25, 2016.

[26] Depthsense cameras. http://www.softkinetic.com/Products/DepthSenseCameras. Accessed: July 25, 2016.

[27] Pmd[vision] camcube. http://docslide.us/documents/datasheet-camcube.html. Accessed: July 25, 2016.

[28] KOKWEE YAP. Designing depth sensor with stereo vision. 2014.

[29] Stereolabs zed stereo camera. https://www.stereolabs.com/zed/specs/. Accessed: July 25, 2016.

[30] Duo3d duo mlx. https://duo3d.com/product/duo-minilx-lv1#tab=specs. Accessed: July 25, 2016.

[31] Ensenso n10 stereo 3d camera. https://en.ids-imaging.com/store/ensenso-n10.html. Accessed: July 25, 2016.

[32] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 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*, pages 559–568. ACM, 2011.

[33] Alexander Weiss, David Hirshberg, and Michael J Black. Home 3d body scans from noisy image and range data. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1951–1958. IEEE, 2011.

[34] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 408–416. ACM, 2005.

[35] Guanglin Zhang, Jiping Li, Jianjun Peng, Hao Pang, and Xulun Jiao. 3d human body modeling based on single kinect. In *Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on*, pages 100–104. IEEE, 2014.

[36] Zhenbao Liu, Hongliang Qin, Shuhui Bu, Meng Yan, Jinxin Huang, Xiaojun Tang, and Junwei Han. 3d real human reconstruction via multiple low-cost depth cameras. *Signal Processing*, 112:162–179, 2015.

[37] Primoz Markelj, Dejan Tomaževič, Bostjan Likar, and Franjo Pernuš. A review of 3d/2d registration methods for image-guided interventions. *Medical image analysis*, 16(3):642–661, 2012.

[38] Gary KL Tam, Zhi-Quan Cheng, Yu-Kun Lai, Frank C Langbein, Yonghuai Liu, David Marshall, Ralph R Martin, Xian-Fang Sun, and Paul L Rosin. Registration of 3d point clouds and meshes: a survey from rigid to nonrigid. *Visualization and Computer Graphics, IEEE Transactions on*, 19(7):1199–1217, 2013.

[39] Joaquim Salvi, Carles Matabosch, David Fofi, and Josep Forest. A review of recent range image registration methods with accuracy evaluation. *Image and Vision computing*, 25(5):578–596, 2007.

[40] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

[41] Andrew E Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5):433–449, 1999.

[42] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[43] François Pomerleau. *Applied Registration for Robotics: Methodology and Tools for ICP-like Algorithms*. PhD thesis, ETH, 2013.

[44] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

[45] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[46] Andrew W Fitzgibbon. Robust registration of 2d and 3d point sets. *Image and Vision Computing*, 21(13):1145–1153, 2003.

[47] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: Science and Systems*, volume 2, 2009.

[48] Majd Alshawa. lcl: Iterative closest line a novel point cloud registration algorithm based on linear features. *Ekscentar*, (10):53–59, 2007.

[49] Qingde Li and JG Griffiths. Iterative closest geometric objects registration. *Computers & mathematics with applications*, 40(10):1171–1188, 2000.

[50] Zexiao Xie, Shang Xu, and Xuyong Li. A high-accuracy method for fine registration of overlapping point clouds. *Image and Vision Computing*, 28(4):563–570, 2010.

[51] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Jianwei Wan, and Min Lu. An accurate and robust range image registration algorithm for 3d object modeling. *Multimedia, IEEE Transactions on*, 16(5):1377–1390, 2014.

[52] Jun Xie, Yu-Feng Hsu, Rogerio Schmidt Feris, and Ming-Ting Sun. Fine registration of 3d point clouds fusing structural and photometric information using an rgb-d camera. *Journal of Visual Communication and Image Representation*, 32:194–204, 2015.

[53] Daniel L Ly, Ashutosh Saxena, and Hod Lipson. Pose estimation from a single depth image for arbitrary kinematic skeletons. *arXiv preprint arXiv:1106.5341*, 2011.

[54] James Charles and Mark Everingham. Learning shape models for monocular human pose estimation from the microsoft xbox kinect. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1202–1208. IEEE, 2011.

[55] Ankur Agarwal and Bill Triggs. Recovering 3d human pose from monocular images. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):44–58, 2006.

[56] Chuong V Nguyen, Shahram Izadi, and David Lovell. Modeling kinect sensor noise for improved 3d reconstruction and tracking. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 524–530. IEEE, 2012.

[57] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[58] Li Chen, Hui Lin, and Shutao Li. Depth image enhancement for kinect using region growing and bilateral filter. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3070–3073. IEEE, 2012.

[59] Radu Bogdan Rusu, Zoltan Csaba Marton, Nico Blodow, Mihai Dolha, and Michael Beetz. Towards 3d point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56(11):927–941, 2008.

[60] Evgeny Gladilin, Barbora Gabrielova, Paolo Montemurro, and Per Hedén. Customized planning of augmentation mammaplasty with silicon implants using three-dimensional optical body scans and biomechanical modeling of soft tissue outcome. *Aesthetic plastic surgery*, 35(4):494–501, 2011.

[61] Michael P Chae, David J Hunter-Smith, Robert T Spychal, and Warren Matthew Rozen. 3d volumetric analysis for planning breast reconstructive surgery. *Breast cancer research and treatment*, 146(2):457–460, 2014.

[62] Pablo de Heras Ciechomski, Mihai Constantinescu, Jaime Garcia, Radu Olariu, Irving Dindoyal, Serge Le Huu, and Mauricio Reyes. Development and implementation of a web-enabled 3d consultation tool for breast augmentation surgery based on 3d-image reconstruction of 2d pictures. *Journal of medical Internet research*, 14(1), 2012.

[63] Crisalix 3d. http://www.crisalix.com/en/doctor. Accessed: July 25, 2016.

[64] Crisalix 3d virtual aesthetics. https://s3.amazonaws.com/media_crisalix/marketing_kit/brochures/Brochure_new_Sensor_2016.pdf. Accessed: July 25, 2016.

[65] Vectra xt. http://www.canfieldsci.com/imaging-systems/vectra-xt-3d-imaging-system/. Accessed: July 25, 2016.

[66] 3dmd system. http://www.3dmd.com/3dmd-systems/#torso. Accessed: July 25, 2016.

[67] Axis three. http://www.axisthree.com/professionals/products/breast-surgery-simulation. Accessed: July 25, 2016.

[68] Gerd Häusler and Dieter Ritter. Parallel three-dimensional sensing by color-coded triangulation. *Appl. Opt.*, 32(35):7164–7169, Dec 1993. URL: http://ao.osa.org/abstract.cfm?URI=ao-32-35-7164, doi:10.1364/AO.32.007164.

[69] Michael Schmeing and Xiaoyi Jiang. Edge-aware depth image filtering using color segmentation. *Pattern Recognition Letters*, 50:63–71, 2014.

[70] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, (5):34–41, 2001.

[71] Tanwi Mallick, Partha Pratim Das, and Arun Kumar Majumdar. Characterizations of noise in kinect depth images: a review. *IEEE Sensors Journal*, 14(6):1731–1740, 2014.

[72] Charles D Muntan, Michael J Sundine, Richard D Rink, and Robert D Acland. Inframammary fold: a histologic reappraisal. *Plastic and reconstructive surgery*, 105(2):549–556, 2000.

[73] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[74] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[75] Rafael C Gonzalez and E Richard. Woods, digital image processing. *ed: Prentice Hall Press, ISBN 0-201-18075-8*, 2002.

[76] Jean Serra. *Image analysis and mathematical morphology, v. 1*. Academic press, 1982.

[77] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

[78] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.

[79] Point cloud library project webpage. http://pointclouds.org/. Accessed: July 25, 2016.

[80] Paul Jaccard. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz, 1901.

[81] Thorvald Sørensen. {A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons}. *Biol. Skr.*, 5:1–34, 1948.

[82] Stefanie Nowak and Stefan Rüger. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval*, pages 557–566. ACM, 2010.

[83] Oswin Aichholzer, Helmut Alt, and Günter Rote. *Matching shapes with a reference point*. Freie Univ., Fachbereich Mathematik, 1994.