

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences



**TECNOLÓGICO
DE MONTERREY®**

Shape reconstruction system based on multi-view photogrammetry

A thesis presented by

Marco Antonio González Barreto

Submitted to the
School of Engineering and Sciences
in partial fulfillment of the requirements for the degree of

**MASTER OF SCIENCE IN
ENGINEERING SCIENCE**

Monterrey Nuevo León, December 2nd, 2021

Dedication

I want to dedicate this thesis to my parents Lorena Barreto and Marco Antonio González for trusting and supporting me throughout my personal and professional life. Also to my sisters Ingrid and Lizett for the unconditional support. Also to my advisors Dr. Joel C. Huegel and Dr. Jorge Luis Pérez González, for the opportunity to work with them, their contributions and above all for sharing their experience and knowledge in the research field.

Acknowledgements

The author wishes to express his appreciation to Tecnológico de Monterrey for the tuition support through the postgraduate programs and postgraduate scholarships given by CONACYT, No. CVU: 1045645.

Abstract

Mexico has a growing need for lower limb prostheses due to a rising number of amputations caused primarily, by diabetic foot. Researchers enable functional and comfortable prostheses through prosthetic design by integrating new technologies applied to the traditional handcrafted method for prosthesis fabrication that is still current. That is why computer vision shows to be a promising tool for the integration of 3D reconstruction that may present to be useful for prosthetic design. This thesis has the objective to design, prototype and test of a functional system to scan plaster cast molds and serve as platform for future technologies for lower limb reconstruction applications. The image capture system is based on 5 stereoscopic RGB-D cameras with 3 DOF, as well as algorithms to calibrate the cameras to a global coordinate reference, to segment a target by removing the background, the conversion of RGB-D images into point clouds for later registration and surface reconstruction. The results for the plaster cast models show the segmentation and registration validations. The segmentation metrics of DC and HD show strong visual similarity with an average similarity percentage of 0.87 and average error of 6.40 *mm* respectively. Moving forward, the system was tested using a known 3D printed model obtained from a CT scan, to which comparison results via HD show an average error of ≤ 1.93 *mm* making the system competitive against the systems reviewed from the state-of-the-art.. The tests and results of the proposed system demonstrate it to be useful for the applications stated in the proposal, applications such as the generation of a digital library for plaster cast molds, serving as a platform for assisting the prosthesis during the manufacturing process, updating external geometries of previous CT scans to avoid radiation exposure and future extraction of viscoelastic properties of external tissue used for deformation analysis.

Keywords: Point Cloud, Iterative Closest Point, Dynamic Reconstruction, Multi-view

List of Figures

1.1	The overall process of conventional socket fabrication is shown. a) Starting with the limb measurements. b) Then, a plaster cast is obtained with the residuum geometry. d) Later, the plaster cast model is post-processed by adding and removing material at key locations. d) Next, lamination and curing take place. e) Finally, the socket gets polished and the cut line is shaped.	2
1.2	A typical process an amputee follows to obtain a prosthesis. 1) Limb measurements and socket fabrication. 2) Check socket. 3) Provisional/Definitive socket. 4) 6-12 months before receiving a definitive socket.	2
1.3	Current socket storage running low and hindering important work areas at a typical prosthetics shop.	10
1.4	The experimental system and the proposal on the four main reasons to address, (a) assisting the manufacturing process, (b) obtaining viscoelastic properties of external tissue, (c) updating external geometries of a CT, (d) creating a digital library from plaster molds.	11
2.1	System pipeline	13
2.2	Realsense models	15
2.3	SR300 cameras turned on independently shows the difference between a color image and a depth image.	16
2.4	Introduction of noise along the edges of the target when multiple SR300 are turned on at once, caused by interference of the structured light.	16
2.5	Simultaneous streaming using 5 D435i cameras. Each indexed image shows on the left side the target with the removed background and the right side shows the depth map.	17
2.6	A vision area of $318.52 \times 178.88 \text{ px}$ at 230 mm from the target using the D435i stereo camera. This results in a resolution of $0.24 \frac{\text{mm}}{\text{px}}$	19
2.7	Mechanical design overview showing (a) Frame and cameras setup with overlapping areas, (b) Frame's overall dimensions.	20

2.8	Example of the user interaction with the system, with a non-contact scheme. Labeled in blue, the rolling chair enables the user to move towards the frame. In orange, previously calibrated cameras aiming at the residual limb. In green, the fixed structure, with a <i>C</i> shape at the top for ease of access.	20
2.9	Calibration setup and the view from all 5 cameras. All sensors look at a common checkerboard as the calibration item. The green box displays the dimensions of the book underneath the checkerboard to validate that the process was completed successfully. This shows the simultaneous view of the calibration phase.	22
2.10	Automatic segmentation, feature extraction, contour smoothing, contour outlier removal, and masking process via the algorithms for one camera image. This shows the automatic segmentation from start to finish. . . .	24
2.11	Comparison of the original scene and the segmented target. This shows the correct working of the segmentation algorithms and PC conversion. . .	25
2.12	Comparison of a) the initial global 5 image alignment and b) the output model after applying a refine ICP registration and Poisson surface reconstruction. This shows the input and the output of the reconstruction algorithms.	26
3.1	The configuration used for the plaster molds show a) 4 cameras mounted on a support that enables rotation and position, b) the workspace of the system being calibrated and c) the model being captured. Showing the camera mounts and the ability to change orientation and position onto the objective.	30
3.2	The configuration used for participants show a) the workspace of the system being calibrated and b) the removable support for target calibration in magenta, c) the interaction with the participant. Showing the camera mounts and the ability to change orientation and position onto the objective. 30	
3.3	Testing of the system with an <i>in-vivo</i> participant starting from a) the general view of the system and the participant, b) the color image from one of the cameras, c) the aligned depth image, d) the segment of the target, e) the mapping of the PC and f) the resulting surface mesh after the 5 images reconstruction. This shows the process from start to finish of the system looked from one camera view.	31
3.4	DC and HD metrics for color images segmentation for four scanned models. In each model, the top row displays the segmented object using manual segmentation as ground truth. Below, the segmentation results using the algorithms. This shows a strong visual comparison.	33

3.5	M1 model measured via HD. a) the original model from the CT, b) the scanned model with the proposed system, c) the quality map between the scanned and original models.	34
3.6	Diameter measurements across all 4 models. Comparison for both SW and physical diameters at 3 different heights. Units are in <i>mm</i>	37
3.7	Participant M6 two poses results measured via HD. a) the residuum model in a relaxed pose, b) the residuum model while co-contracting the muscles, c) Quality map between both poses.	38
3.8	The traditional pipeline with the addition of technology integration. 1) Limb measurements and socket fabrication. 2) Check socket. 3) Provisional/Definitive socket. 4) Scan of current plaster cast mold. 5) 6-12 months before receiving a definitive socket. 6) 3D Re-fabrication of the mold from the digital library.	39

List of Tables

1.1	Comparison of different image acquisition methods used for 3D reconstruction for lower limb amputations.	8
3.1	DC and HD statistics obtained from the segmentation results in Fig. 3.4. Showing a mean of 0.87 out of 1.00 for DC. Also, a mean of 26.67 px for HD which can be translated to 6.40 mm using the system resolution of 0.24 $\frac{mm}{px}$, showing a strong resemblance to the ground truth segmentation. . . .	32
3.2	HD statistics results were obtained from the model comparison. Values in mm	35

Contents

Acknowledgements	v
Abstract	vii
List of Figures	ix
List of Tables	xiii
Contents	xv
1 Introduction	1
1.1 State of the art	3
1.1.1 Cameras	3
1.1.2 Machine vision algorithms	3
1.1.3 Biomedical applications	5
1.1.4 Acquisition methods comparison	7
1.1.5 Previous & current work	9
1.2 Problem statement	9
1.3 Justification & proposal	10
1.4 Objectives	11
2 Methodology	13
2.1 Requirements & specifications	13
2.2 Capture environment	14
2.2.1 Cameras	15
2.2.2 Capture speed	16
2.2.3 Mechanical design	19
2.2.4 User interaction	19
2.3 Calibration	20
2.4 Segmentation	22
2.5 RGB-D alignment & Point Cloud creation	24
2.6 Reconstruction	24

2.7	Validation	26
3	Results & discussion	29
3.1	Prototype	29
3.2	Segmentation results	31
3.3	Performance evaluation	34
3.4	Measurements verification	36
3.5	<i>In-vivo</i> validation	38
4	Conclusions & future developments	41
	Bibliography	43
A	Drawings	47
	Abbreviations	49
	List of Symbols	51
B	Curriculum Vitae	53
C	Publications	57

1 | Introduction

To understand the context, Mexico, according to the Mexican Institute of Social Security, shows that just in 2014, they were near 900 thousand amputees, to which are added 75 amputations per day. This means that each year Mexico is accumulating more than 70 thousand people with lower limb amputations, forming part of the people with disabilities in the country [1]. This generates economic problems and dependency for their caregivers, their families, and the community. This is why prostheses have an important role in recovering a percentage of mobility in amputees, providing autonomy to the person in carrying out activities, as well as complementing the person in a social and cultural aspect, preserving their individual identity [1].

This is why researchers currently focus their efforts to address this issue with all the "Extreme Bionics" projects. The main goal is to seek comfort and functionality for all the people who receive a prosthesis so they can walk again. It is important to mention that the most important part of the prosthesis is the socket, because it is unique for each individual. Also, many of the projects in the laboratory are related to socket design. It forms part of the four sections in a lower limb prosthesis, the socket, knee, pylon, and foot [2]. While several of the parts can be obtained in large quantities, materials, and sizes. The socket is the part where the amputation load rests, so a correct fit must be met. A favorable socket design stimulates muscle growth and relieves pressure in sensitive areas providing comfort to the user [3]. On the other hand, improper fitting may result in skin problems and tissue injury caused by unbalanced weight [4].

That said, the socket requires a manufacturing process that ensures a match with the user measurements. Currently, the design and manufacture continue to be handcrafted since all the fabrications are unique and unrepeatable [4]. The overall socket fabrication can be seen in Fig. 1.1. First, a prosthetist performs an evaluation by taking limb measurements Fig. 1.1a and a cast mold Fig. 1.1b is taken. Subsequently, the mold is filled to obtain a model that describes the geometry of the residual limb. Afterward, post-processing is necessary to rectify, add and remove material, to ensure that the dimensions match with those obtained Fig. 1.1c. Then, the socket is manufactured by laminating the mold and then curing it, usually with polypropylene shown in Fig. 1.1d. Finally the socket get polished in Fig. 1.1e. The time and the iterations invested in the process depend on the experience of the prosthetist, since obtaining a fitting socket consists of trial

and error. Typically the participant will receive a check socket, return for adjustments to get a provisional socket that will last from 6 months to a year before the residuum strengthens and changes in volume. Here again, a check socket must be fabricated, and afterward a definitive socket [5]. This timeline that a participant follows in order to obtain their definitive socket can be seen in Fig. 1.2.

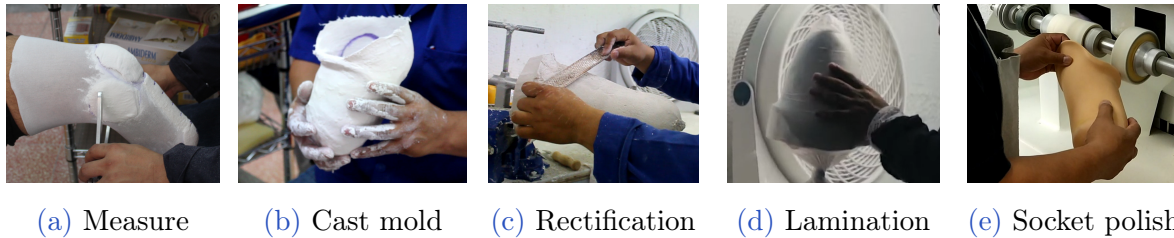


Figure 1.1: The overall process of conventional socket fabrication is shown. a) Starting with the limb measurements. b) Then, a plaster cast is obtained with the residuum geometry. d) Later, the plaster cast model is post-processed by adding and removing material at key locations. d) Next, lamination and curing take place. e) Finally, the socket gets polished and the cut line is shaped.

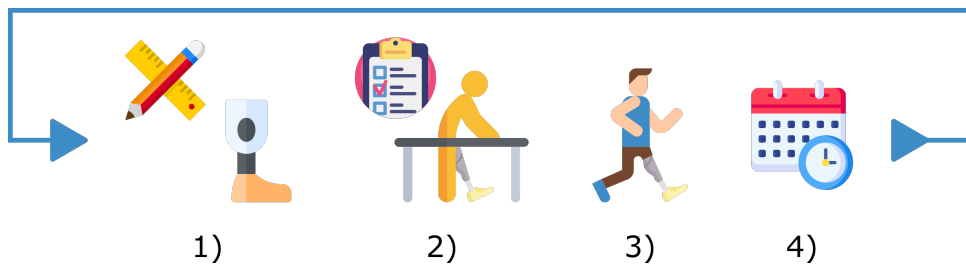


Figure 1.2: A typical process an amputee follows to obtain a prosthesis. 1) Limb measurements and socket fabrication. 2) Check socket. 3) Provisional/Definitive socket. 4) 6-12 months before receiving a definitive socket.

Computer vision shows to be a tool that benefits these processes through digital reconstruction with the use of photogrammetry to obtain a digital model. This technique is used to obtain detailed information from photographs. Then, algorithms take these images and estimate their position to generate a 3D point cloud. This can be done with multi-view stereo cameras as mentioned in [6]. In the literature, the study of lower limb reconstruction has advanced Computer-Aided Design (CAD), assisting in socket manufacturing, digital models allow to iterate at a minimal cost and reducing the time taken for their manufacture. Moreover, digital models allow accurate geometry and additional relevant information for socket fabrication. Existent technologies and methods involve photogrammetry, 3D scanning with structured light, 3D laser scanning, Computer Tomography (CT), and Magnetic Resonance Image (MRI). Nevertheless, there are still challenges, particularly with regard to limitations on current technologies available and

their integration with biomedical applications. These technologies are going to be seen and compared in-depth in section 1.1.

1.1. State of the art

This section covers the types of cameras that exist to obtain color and depth data 1.1.1, several machine vision algorithms used for segmentation, registration, and reconstruction as well as popular implementations 1.1.2. Then the review of certain biomedical applications of interest 1.1.3. Then a comparison of methods and techniques for 3D reconstruction 1.1.4. Last, an overview of the previous and current work 1.1.5.

1.1.1. Cameras

There are different sensors and types of cameras to acquire 3D data. These methods include 3D laser scanning also known as Light Detection And Ranging (LiDAR) which uses a laser beam that bounces onto a surface and measures the traveling light to the objective. It is mainly used in the industry for mapping wide open areas such as cities and autonomous cars as mentioned in [6]. There are also structured light scanning cameras that have Color and Depth sensors (RGB-D), the depth sensors project a coded light, which measures the distance of the light traveled to the objective. An example of this type could be the Intel® Realsense® SR300 camera. They have a direct comparison to other cameras like the Microsoft® Kinect® [7]. Then there is the stereoscopic vision, having two or more sensors. Cameras like the Intel® Realsense® D435 have two image sensors in combination with an infrared projector to estimate depth. Many of these sensors may be used alone or in tandem to cover a wider range as discussed in the next section.

1.1.2. Machine vision algorithms

Multi-view systems are based on photogrammetry, which assumes parts from having multiple repeated elements of the same object in multiple images. This is so to obtain information about physical objects and environment by capturing a sequence and processing it to extract data such as properties and measurements as mentioned in [6, 8]. Most of the algorithms used are oriented to 3D reconstruction. Such that, a triangulation of points can be made from images and the world. So a point cloud can be made from this point triangulation. This can be approached with a single camera having different views or with an array of cameras by covering a whole object and having overlapping areas of said object [6]. Two of the most significant pieces of researches make use of photogrammetry and multi-view for 3D reconstruction [9, 10].

Different approaches have been explored, one of the most common ones is Simultaneous Localization and Mapping (SLAM) [11]. This technique's goal is to create a map and determine the location of an object inside an environment. So basically this map is a sequence of images matched together by common regions, with enough images a 3D map may be represented. One of the challenges SLAM faces is the acquisition and integration of multiple scans that have dynamic changes that update in the scene. Various implementations of this technique can be seen, such as the Kinect Fusion [12] and Dynamic Fusion [13] being the most popular ones. The idea is the same, but, these implementations are focused in punctual object or scene reconstruction rather than area mapping. What stands out about these algorithms is their response to dynamic changes for rigid and non-rigid transformations. It is more clear in the Dynamic Fusion were facial expressions are scanned and updated in real-time [13].

These above-mentioned papers and many computer vision works depend on classic algorithms for camera calibration, segmentation, registration, and surface reconstruction. The first and most necessary step is calibrating the cameras to obtain intrinsic properties, related to internal camera properties, and extrinsic properties, related to the physical position of the camera in the world coordinate system. This calibration is performed employing matching key elements from the images sequence. In the literature, one of the most common algorithms for calibration is Zhang's Methods using a checkerboard as a target [11, 14]. Zhang's algorithm works both with single and multi-view setups. This algorithm is limited to RGB sensors which do not utilize the capabilities of RGB-D cameras.

The next step is the segmentation which takes place to separate a portion of interest from the rest of an image. Popular algorithms like Otsu's method binarize an image by setting up a threshold where parts of an image remain white while others black make a contrast or mask of the original image [15]. Using Canny's algorithm for edge detection, like binarization, the resulting image has white edges that surround the objects and black for the rest [15]. Another technique is the watershed for region growing, this algorithm resembles the flooding of a valley, as the water rises it fills the terrain, in this case, the gradients of the image, the bigger the water becomes would represent the elements on the image [15]. Also, these techniques aren't used all alone, in practice, a combination of them is required in addition to applying filters to the image, like blurring and sharpening [15]. Their use will depend on the case needed for the application. So, as a general idea, segmentation creates a binary mask of the desired object and when applied to the original image, a region of interest can be extracted.

Moving on, the third step is registration charged with the alignment of two pairs of coordinate points sets, commonly known as Point Clouds (PCs), whether 2D or 3D. Registration is still a challenge to find the best transformation for PCs that properly preserve

the morphology of the object. These transformations involve the use of rigid transformations that translate and rotate the object, without scaling or deforming the object. The most common algorithms are the Random Sample Consensus (RANSAC) [6], which randomly tries to fit a line over a set of points, the most amount of points that lie over the line are considered in-liers. Another technique, is the Iterative Closest Point (ICP), that matches point to point given a threshold distance [6]. While RANSAC is usually used for a first coarse alignment, ICP is used as a fine one. These pairwise alignments represent rigid and non-rigid transformations. A prior work reports the performance of different registration techniques, including Artificial Neuronal Networks (ANN), Coherent Point Drift (CPD), RANSAC, and ICP. The study consisted of using these mentioned methods to align 2D PCs. Results demonstrate that RANSAC had the least accumulated error, followed by ICP, then ANN, and finally CPD [16]. Variations in results are determined in part by the morphology of the target.

Once the PCs have been properly aligned when the desired threshold is met, a surface reconstruction is completed to create a mesh that can be exported as a 3D model. In the literature, there are several methods such as Ball pivoting and Poisson surface reconstruction. The BPA, Ball Pivoting Algorithm [17], is a surface reconstruction method that consists in creating triangles in between points and by pivoting a mathematical ball with a specific radius around the points without falling through the points a triangle is created, a drawback for this algorithm is that points must be uniformly distributed and the distance between three points must be smaller than the selected ball radius, otherwise the resulting mesh would be left out with undesired holes. Another widely used is the Poisson surface reconstruction [18], it solves the optimization problems to obtain a smooth surface and even extrapolates in areas to achieve it, which is preferred over the BPA, as the surface is not smooth due to the resulting low polygon density from the PC.

These algorithms are well known when speaking of a standard 3D reconstruction pipeline. Various biomedical applications make use of these techniques to create 3D models. Mainly by using photogrammetry, structured light, LiDAR, CT, and MRI, which is seen in more detail in the different biomedical applications of interest for this study in the next section.

1.1.3. Biomedical applications

Previous research shows the use of several medical image techniques for 3D scanning. Shuxian et al. presented a 3D reconstruction for residual limb with the use of a CT. From it, bone structure and skin surface are reconstructed and then aligned to represent the inner and the external part of the residuum. This work utilizes the compound model to be submitted to a Finite Element model (FE-model) simulation [3]. Sengeh et al. focus on the in-contact patient *in-vivo* indentation system, which allows tissue displacement, along

with MRI, combined with an FE model to evaluate the mechanical and loading conditions over the skin [4]. Works like Bonacini et al. and Colombo et al. conducted a detailed study on a non-contact method for obtaining external and internal 3D reconstruction through the use of laser scanners, CT, and MRI and compares these morphology acquisition techniques [5, 19]. Their results show that CT has good quality for bone structure and external shapes, whereas MRI provides detailed representations of both internal and external. But, for both CT and MRI surfaces resemble the skin very little. While laser-scanned helps as a reference for aligning the models [5] and it has important features such as cicatrix and marks left from the amputation having a better correspondence to the real skin surface [19]. These studies demonstrate that bone structure and skin surface are essential when performing a 3D reconstruction for prosthetic design along with an FE-model to understand the loading conditions for the socket.

After the first socket completion, multiple iterations might be necessary for the user, there is no socket that can adjust to the residuum's volume change [1]. Soft tissue like skin, muscle, and adipose are susceptible to increasing or decreasing their volume, while inner geometries like bone can be considered to remain without morphological changes over long periods of time and under varying loads. As these changes affect the external structure, photogrammetry shows to be a useful tool for external surface reconstruction, by helping to update previous models during the iterative process of the socket while remaining with a single CT capture containing the bone structure. A similar process might be followed to update the external surface scan and align it to the original CT and performed by Bonacini et al. So, to address the radiation problem generated by CT and MRI, radiation exposure may be minimized by updating the external surface only through photogrammetry.

Several studies on skin reconstruction have taken place. Taqriban et al. conducted a study on close-range photogrammetry by using a digital reflex camera and Autodesk® Recap Photo software to obtain a 3D model. The residuum of a participant was scanned 360° with two different shooting angles, a 3D model of the external surface was obtained with Recap and then a manual rectification was performed. Its worth mentioning that the manipulation process towards the socket manufacture is still digitally handcrafted with this method [20]. A soft tissue reconstruction was conducted by Lacher et al., who proposed a nonrigid reconstruction of 3D breast surfaces with RGB-D. The work mentions the importance of a non-invasive way for the evaluation. As well, the work addresses the model deformation caused by the involuntary movement of the body during data acquisition. This work finds that multi-view is important when analyzing body parts, to minimize the involuntary movement of the body [21]. Another study by Wu et al. proposed a system with six RGB-D cameras in an array that identifies landmarks and measurements from a foot. The results showed a 3D capture system able to scan a

foot, obtain multiple PCs, and merge them using ICP. Also, twelve anatomical landmarks of the foot can be detected automatically without manual intervention [22]. A similar study using multi-view by Solav et al. presents a multi-camera array to analyze muscular deformation through Digital Image Correlation (DIC) by using 12 RGB cameras, which demonstrates the non-contact multi-view *in-vivo* capabilities [23]. Contrasting to the in-contact presented by Sengeh et al. with the skin indentation [4].

So, these studies demonstrate the importance of multi-view photogrammetry to complement with additional information from the skin surface, as to update previous models and needed for an FE-model [5, 19]. Moreover, methods like the 3D breast soft tissue reconstruction highlight the importance of systems that minimize inaccurate data from the involuntary movement of the body, and the presented DIC to analyze muscular deformation shows the capabilities of a non-contact *in-vivo* analysis [21, 23].

All of these studies have important attributes that make strong research in their field, the use of different approaches or a combination of these can be useful for this work to move forward the socket design. This is why a comparison is made in the next part denoting the benefits and disadvantages of the available technologies.

1.1.4. Acquisition methods comparison

Table 1.1 compares the above-mentioned technologies and techniques used for 3D reconstruction involving their measurement accuracy, speed capture, geometries, benefits, and disadvantages.

For photogrammetry there are two approaches to obtain corresponding features, the first one is by using a single moving camera around the target as the SLAM algorithm and the second is by using multiple fixed cameras as described in the study to obtain foot measurements. Both of them rely on multi-views. Also, this method allows any commercial digital camera to be used. In the case of a stereo camera, it follows the principle of photogrammetry, as two sensors are fixed to the frame of the camera. There are models such as the D435i RGB-D that have this capability.

For 3D structured light scanning, there are commercial options such as the SR300 RGB-D or the Kinect®), the accuracy comparison are shown respectively in Table(1.1). On one hand, coded light cameras and LiDAR cannot, because the light from one camera interferes with another camera aiming at the same target [22, 24]. Also, these two technologies struggle with reflective surfaces [24]. On the other hand, stereo cameras have the main advantage on multi-view setups over the coded light depth cameras. While having an array of stereoscopic cameras, they may be turned on all at once.

Recent research pipelines involve the use of CT and MRI to optimize socket design

through the use of CAD [3–5, 19]. These techniques show to be the most accurate and comfortable for socket design. These, have the advantage to obtain internal and external geometries needed to locate bone structure and soft tissue. Results show a better quality socket with less involvement of the participant attending the laboratory for limb measurements [5]. As mentioned above, the participant will have external volume changes [1]. So the amputee will have to return in the future for a new socket with fresh CT or MRI images for this pipeline. Resulting in high expenses per scan for the individual and the drawback of radiation exposure for the CT.

Another important aspect is the capture speed. This affects the quality of the images. The longer the time for a capture session, the more prominent body stuttering will become, resulting in blurry images. For this application, photogrammetry demonstrates to minimize the capture speed [21]. By using simultaneous cameras, the speed could be narrowed down to the manufacturer specification of 30 *fps*. In the case of structured light and LiDAR, they also perform in the time set by the manufacturer, usually 30 *fps*. But, for a multiple sensor setup, these need multiplexing as further discussed in the cameras section in 2.2. The multiplexing issue on these technologies can drop significantly the *fps*. In previous work, the use of multiple structured light cameras delivered a capture speed of $\frac{1}{6}$ *fps* due to multiplexing [24]. Capture speeds for CT have been recorded to be average to 1-2 *min* taken from the trials performed with the laboratory’s participants. A similar time was recorded for MRI [19]. These times depend on many factors that could be explained mainly by image resolutions, slices thickness, and the number of images. Protocols can be adjusted to increase or decrease these capture speeds. Still, CT and MRI show to be the least effective in this category. Knowing the current methods and technologies enables the planning and proposal of this work, which has a strong basis on previous work as discussed next.

Table 1.1: Comparison of different image acquisition methods used for 3D reconstruction for lower limb amputations.

Method	Measurement Accuracy	Capture Speed	Geometries	Benefits	Disadvantages
Photogrammetry	Relatively 1–10%. [6] For Stereo cameras 1.25% of range [6]	15- 120 <i>fps</i>	External	Use of multiple cameras at a time. Allows capturing textures. [6] May use any digital camera [20]. Used for DIC.[23]	Often, not a plug and play solution.
3D Structured light scanning	2-5 <i>mm</i> [7] Relatively 1–6% [6]	$\frac{1}{6}$ <i>fps</i> [24]	External	Often used on commercial scanners with plug and play solution. May capture textures if an RGB sensor is integrated. [6]	Use of one camera at a time. [22, 24] Has trouble with reflective surfaces. [24]
LiDAR	1–10 <i>mm</i> [6]	-	External	High speed scanning. Able to scan large objects.[6]	Often used on vehicles, rarely used on biomedical applications. Has trouble with reflective surfaces.
CT	0.3 ± 0.4 <i>mm</i> [25]	1-2 <i>min</i>	Inner and external	Allows medical diagnostic.	Radiation exposure No resemblance of skin surface. [19]
MRI	For 1 <i>mm</i> slice, relatively 4-15% [26]	1:50 <i>min</i> [19]	Inner and external	Allows medical diagnostic.	Cannot be used in presence of inner metallic supports. [19] No resemblance of the skin surface. [19]

1.1.5. Previous & current work

In prior work, I presented a system capable of capturing a whole target with four fixed cameras that surround it, to minimize the involuntary movement of the body. A hand was used to demonstrate its capabilities for non-static and deformable targets [24]. Said research served as a preliminary methods for the current work, as it establishes the type of cameras needed, as the amount and their position for a full reconstruction. Moreover, it provides equations that relate the distance between the target and the camera, and the resolution. Sections 2.1 and 2.2 rely heavily on these aspects which, determine the cameras used in this work, the mounting frame, and the computational resources needed to meet the application. Also, it has an influence on the ICP and Poisson surface reconstruction algorithms selected in the section 2.6.

This thesis focuses on the construction of algorithms that can provide automated lower limb external reconstruction with the use of photogrammetry using consumer-grade RGB-D cameras, requiring a single capture from all different cameras. This study introduces and tests a five-camera array system, by fixing cameras to a frame to facilitate the calibration and capture of the system. While the integration of range images is not novel, the particular setup and application are novel with the following particular issues. The scanned PCs have an initial rough alignment from the calibration and then a fine one using Iterative Closest Points (ICP). This produces a 3D model that can be exported for further processing. At this stage, due to the iterative nature of the work involving programming, participants were not involved, but for validation purposes only. Instead, another issue was addressed to help Proactible, a local prosthetics clinic, in building a digital library for plaster cast models. This helped the development by having a controlled environment, while having replicas of transtibial, below the knee, and transfemoral, above the knee, lower limbs. This writing is organized as follows, the methodology presents the pipeline and the experimental setup, along with the segmentation and registration algorithms. Then results display the performance of the said used algorithms with the molds mentioned and the validation with an *in-vivo* participant, also discussion precedes the conclusions.

1.2. Problem statement

The problem that the Proactible laboratory is currently facing is that they are running out of physical storage as shown in Fig. 1.3 and a digital library is desired. This would enable the backup of the plaster cast model for later re-fabrication if needed. The scope of this thesis only includes the digitization of physical models leaving the fabrication and re-fabrication of molds to future work.

Furthermore, the problem that the prosthetic researchers and designers are facing today is that no digitally-based system that makes reconstruction for transtibial lower limb amputations on one run. Current systems are comprised by hand-held scanners, moving cameras, or a turning table to perform scanning. By using only one camera and moving it around the target, captured data becomes inaccurate due to the involuntary movement of the body, as noise is carried out through the reconstruction process [21]. Similarly, dimensions do matter in the manufacturing process, large tolerances could lead to an uncomfortable and unusable socket [22] in the dimensions requiring sufficiently accurate digitally-based reconstructions to ensure accuracy. Moreover, the system is sought to serve as a platform that enables the integration of future technology such as the use of CT scans and deformation analysis which needs synchronous, real time capture for the socket design.



Figure 1.3: Current socket storage running low and hindering important work areas at a typical prosthetics shop.

1.3. Justification & proposal

There are four main reasons to solve the aforementioned problems. The first, to assist the prosthetist during the manufacturing process by obtaining dimensions and skin-related information due to surgery Fig. 1.4a. Second, extract the viscoelastic properties of external tissues to later feed an FE-model for an automated socket conformation as shown in Fig. 1.4b. Third, to update the external surface from previous CT scans Fig. 1.4c. Fourth, assist Proactible in the creation of a digital library to backup plaster cast models for future re-fabrication Fig. 1.4d.

The proposal is to use multiple stereoscopic RGB-D cameras to obtain a full reconstruction with a minimum of 5 cameras, 4 that would cover 360° in a circular pattern, and a fifth camera that would look at the end view of the residuum. This way cameras would not need to move around the residuum and a full scan could be performed in a single multi-view capture. Have a frame to hold all cameras fixed in place. Also, have a segmentation algorithm that cleans the pictures and an algorithm for the reconstruction process. This reconstruction should give us enough accuracy to have an error of $\leq 2 \text{ mm}$ for validation.

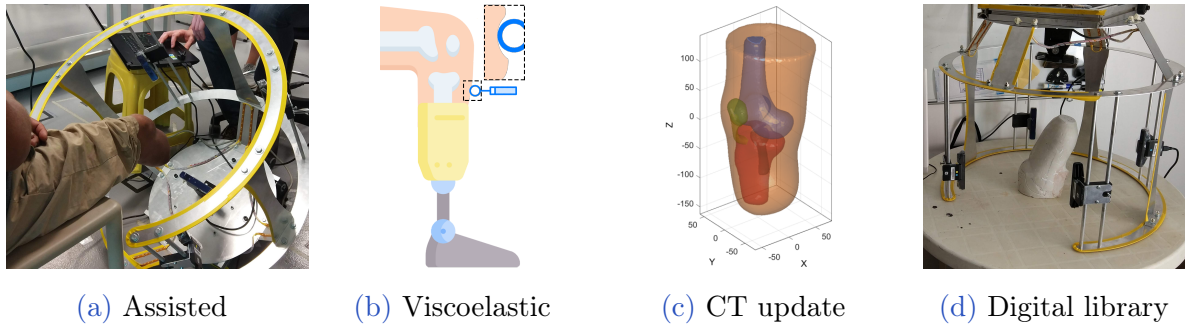


Figure 1.4: The experimental system and the proposal on the four main reasons to address, (a) assisting the manufacturing process, (b) obtaining viscoelastic properties of external tissue, (c) updating external geometries of a CT, (d) creating a digital library from plaster molds.

1.4. Objectives

The general objective is to prototype and test an automated multi stereoscopic view acquisition and reconstruction system. In other words, a multipurpose scanner for lower limb amputations.

The specific objectives consist of the following and are presented in the same order throughout the methodology.

- To design and manufacture a functional frame with a radial array able to support 5 mounted cameras to cover an objective around 360° and capture simultaneously across all cameras. Also, to integrate camera movement through orientation and position for calibration.
- To develop a segmentation algorithm, allowing to keep the desired object, by removing the background and unwanted elements in the scene. Moreover, compare the algorithms versus a based truth segmentation via Dice Coefficient (DC) and Hausdorff Distances (HD).
- To implement a reconstruction algorithm that receive RGB-D images as input and outputs a 3D triangle mesh of the target. Furthermore, compare the resulting mesh with an original digital model via HD to seek an error of $\leq 2 \text{ mm}$.
- To display the result of using the system by scan 4 different plaster cast models and verify their diameter measurements through software and physically with the use of a caliper.
- To demonstrate the final use scenario with an *in-vivo* participant and its capabilities for skin deformation changes through HD.

2 | Methodology

This section hereby presents the stages of the system. In Fig. 2.1 a pipeline of the system can be found, it contains six main parts. The capture system involves the cameras used and the frame where cameras are fixed. Then a calibration takes place where cameras are located within a global coordinate system by using a checkerboard. Next, a segmentation takes place by making a selection of the target using the color image. From a set of RGB-D images, a Point Cloud (PC) is created using the intrinsic parameters. Moving on, all five PCs are loaded and globally aligned from calibration parameters and reconstruction takes place. As a result, a 3D model is obtained from the system and validation takes place to ensure proper dimensions by using Dice coefficient (DC) and Hausdorff Distance (HD) metrics.

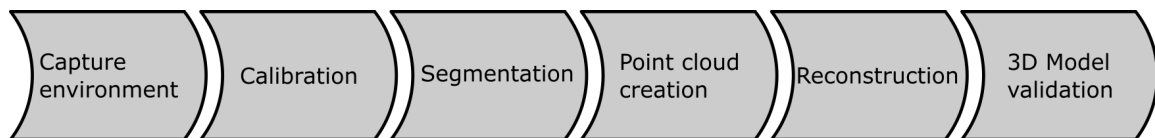


Figure 2.1: System pipeline

2.1. Requirements & specifications

The system needs to meet the 4 different applications stated in section 1.3. The following list of requirements ensures a proper implementation:

- Contain a wide enough field of view to cover the target,
- Capture simultaneously and stream visualization from all cameras,
- Duration of the capture should last no longer than 1 second to avoid body movement,
- Resistant frame and light weighted, made out of a resistant material,
- Have rotation and position mechanisms to adjust cameras for calibration phase,
- Consist of an interface to be easily approached towards the user,
- Incorporate a certain inclination so the user can be seated in a comfortable position,
- Capacity for residuum with a circumference up to 550 *mm*,

- Comprise of an automatic reconstruction pipeline with little user intervention.

First, the vision and its integration. To address the first 3 elements of the requirement list, Table 1.1 shows the comparison between different technologies available. Photogrammetry shows to be the best option for this work for capturing external geometries, as it is possible to use multiple cameras at the same time. Also, it allows the capture of color images and enables the use of DIC. In addition, it has the lowest time capture from all technologies. For this reason, stereo cameras show promise for this work, multiple sensors can be combined to capture simultaneously a scene. This work needs an inward-facing configuration for a full capture of the target. By requiring multiple sensors, there are several considerations when connected to a computer. The number of cameras connected depends on data bandwidth and power consumption, preferring connecting multiple cameras straight to the computer and choosing lower resolutions to limit the raw data transmission below 1200 *Mbps*. With a specific configuration, a camera streaming depth and color data at 15 *fps* can translate to an average of 220 *Mbps* with a power consumption of 2 *W*. Also, multiple cameras will rely on processing power from the CPU, for the RealSense® cameras, the tests performed by the manufacturer shows that an Intel® i7-6700HQ can support 4 simultaneous cameras [27]. From these requirements and considerations, the next list presents the engineering specifications for the application:

- Stereo camera,
- Restriction design of $\leq 0.5 \frac{mm}{px}$,
- Accuracy of $\leq 2 mm$,
- Data transmission below 1200 *Mbps*,
- Independent 3.0 USB controllers,
- i7 Quad-Core processor.

Second, the frame represents the physical space where cameras are mounted and the interaction with the user. Points 4 through 8 include the manufacturing process, the material used, and the size of the structure. Also, the installed rails and tripod heads are used for adjusting the cameras for calibration. Also, discussed in sections 2.2.3 and 2.2.4.

The last point of the requirements emphasizes the software in charge of performing the segmentation registration and surface reconstruction seen in-depth in the next sections.

2.2. Capture environment

This section deepens into the camera type used in this thesis, the capture speeds, system resolution, mechanical design, and user interaction.

2.2.1. Cameras

The capture system uses the Intel® RealSense® D435i RGB-D camera in Fig. 2.2b. This component meets the requirements, by capturing depth and color images. In addition, this model provides a stereoscopic solution as it has two sensors, which through homography, it matches pair key points to obtain depth information. The model has an infrared projector that lies down a pattern covering the object, this additional feedback results in more accurate depth data. The manufacture specifications state that the camera has an accuracy of $\leq 2\%$ at 2 m , which is enough for the 2 mm error for our validation. It is worth mentioning that this projector does not interfere with other cameras in the setup, giving the ability to turn on all cameras at the same time.

The cameras depth and color configuration set to $640 \times 480\text{px}$ and $1280 \times 720\text{px}$ respectively at 15 fps ensure a data transmission below 1200 Mbps . This study uses a computer with 5 independent 3.0 USB ports with an Intel® Core i7 4770K, 16 Gb RAM, and a GTX 760 GPU that handles the specifications above mentioned.

In contrast to previous work [24], It uses the SR300 model in Fig. 2.2a, which consists of structured light for the depth images, but this caused problems with the setup. The light coming from one camera added noise to the other cameras in the array if all cameras are turned on at the same time. So, the SR300 model needed software multiplexing to turn on and off one camera at a time for each capture, until a full swipe was complete.

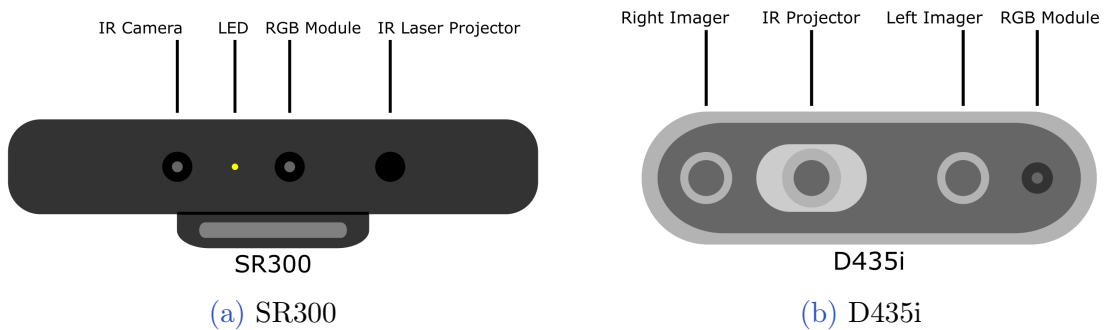


Figure 2.2: Realsense models

To demonstrate this issue on how multiple SR300 cameras cannot be turned on at the same time, Fig. 2.3 shows how depth data would normally operate by multiplexing each camera independently. On the other hand, Fig. 2.4 displays how noise is introduced if multiple cameras are turned on at once. For this study, the SR300 is not a viable option. Making the D435i superior for multi-view setups.

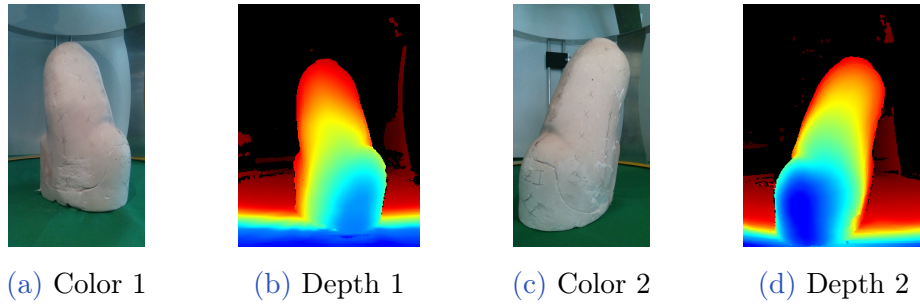


Figure 2.3: SR300 cameras turned on independently shows the difference between a color image and a depth image.

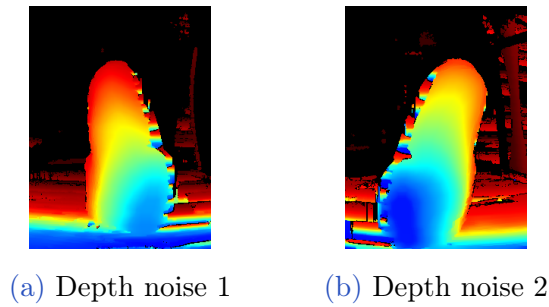


Figure 2.4: Introduction of noise along the edges of the target when multiple SR300 are turned on at once, caused by interference of the structured light.

2.2.2. Capture speed

The capture speed determines the accuracy of the reconstruction. As stated before in section 1.1.4 and Table 1.1, with the use of photogrammetry, systems can achieve the same capture speeds for each camera like the ones stated by the manufacturer. The use of five D435i stereo cameras in this work, allows simultaneous streaming and capture as shown in Fig. 2.5 with all 5 cameras. Each indexed image displays on the left, the background removal with a given threshold, and on the right, the depth map. These images are obtained and saved directly from the programmed visualizer, showing that there is no presence of interference caused by the other cameras. These saved images re-appear in the segmentation section 2.4. For this study, the capture speed of the system is set so all cameras capture at 15 *fps*, this may be changed up to 90 *fps*. Moreover, this allows the system to get a single capture in 1 *s* for static analysis. In contrast, previous work delivers a capture speed of $\frac{1}{6}$ *fps* and could only work for static modeling [24].

Resolution

In this part, I calculate the optimal distance from the camera to target and the resolution that ensures a resolution less than $0.5 \frac{mm}{px}$ considering the range distance suggested by the manufacturer. The distance from the cameras to the objective and the resolution, share

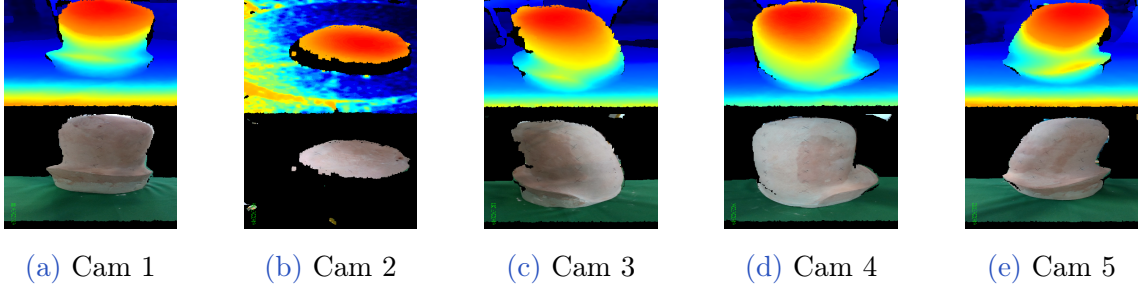


Figure 2.5: Simultaneous streaming using 5 D435i cameras. Each indexed image shows on the left side the target with the removed background and the right side shows the depth map.

a common relationship through the focal distance in Eq. 2.1, obtained from prior work [24],

$$f = \frac{B}{G}(l) \quad (2.1)$$

where f is the focal distance, B is the image size, G is the target size and l is the distance to the target.

From Eq. 2.1, B holds the information of the image width i and height j , while G is the size in mm of the objective for width and height denoted by w and h . Then, the resolution of the image can be calculated with Eq. 2.2, obtained from previous work [24].

$$res = \frac{G}{B} = \frac{(i, j)}{(x, y)} \quad (2.2)$$

where w and h are the real width and height in mm captured by the camera, i and j is the number of pixels regarding the camera's width and height, res is the resulting resolution in terms of $\frac{mm}{px}$. Since pixels are squared, the resolution can be calculated with only one axis component.

The camera's manufacturer establishes that the ideal usage range is from 200 mm to 3 m for a proper depth estimation for the D435i, but may go up to 10 m . Given these specifications, I propose the dimensions of the frame to have a diameter of 610 mm . Next, I calculate the distance from the cameras to the objective given these parameters. An experimental measurement is taken from the circumference C of a healthy man's leg, with a 1.85 m height and a body mass of 73 kg to find a radius r in Eq. 2.6:

$$c = 470mm \quad (2.3)$$

$$c = \pi(d) \quad (2.4)$$

$$d = \frac{470mm}{\pi} = 149.6mm \quad (2.5)$$

$$r = \frac{d}{2} = 74.8mm \approx 75mm \quad (2.6)$$

where c is the circumference, d is the diameter, r is the radius.

Then in Eq. 2.7 the distance l from the frame circumference to the target is obtained:

$$l = 305mm - 75mm = 230mm \quad (2.7)$$

the resulting distance $l = 230mm$ lies within the ideal usage range with an additional $30mm$ gap to avoid the $200mm$ lower limit.

In this setup, the D435i uses a color resolution of $1280 \times 720px$, with a field of view of 69.4×42.5 , horizontal and vertical respectively, provided by the manufacturer [28]. At this distance, the horizontal range is found with Eq. 2.10, similarly, the height range is obtained in Eq. 2.10 and it can be seen in Fig. 2.2. The camera is placed in a vertical stance to cover the most length possible of the target. Then, the resolution is calculated with Eq. 2.12:

$$\frac{w}{2} = \tan(34.7)(230mm) = 159.26mm \quad (2.8)$$

$$w \approx 318.52mm \quad (2.9)$$

$$\frac{h}{2} = \tan(21.25)(230mm) = 89.44mm \quad (2.10)$$

$$h \approx 178.88mm \quad (2.11)$$

$$res = \frac{318.52mm}{1280px} = 0.24 \frac{mm}{px} \quad (2.12)$$

where w and h are the width and height length of the field of view.

Finding a maximum resolution of $0.24 \frac{mm}{px}$ for this setup at a distance of $230mm$, for a lower value, the higher the resolution. The resolution obtained of $0.24 \frac{mm}{px}$ is less than the $0.5 \frac{mm}{px}$ specification, meaning a better image resolution for the application. The cameras may be adjusted on the rails to increase or decrease the distance to the target from a minimum of $230mm$ up to $400mm$ maximum. allowing flexibility for larger objects.

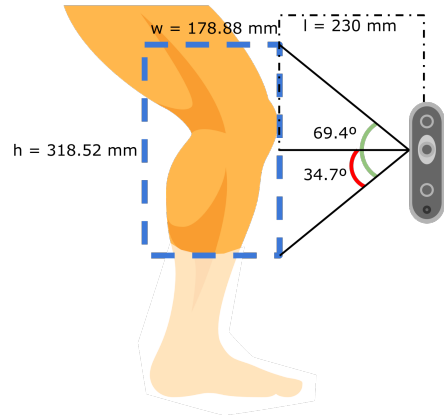


Figure 2.6: A vision area of $318.52 \times 178.88 \text{ px}$ at 230 mm from the target using the D435i stereo camera. This results in a resolution of $0.24 \frac{\text{mm}}{\text{px}}$.

2.2.3. Mechanical design

The frame is designed and manufactured to support and hold the cameras in place, it is made out of laser-cut and bend aluminum sheets capable to support the system components. Fig. 2.7a shows a top view of the frame, where cameras are placed spaced every 90° showing to be equidistantly using 4 cameras and the calculated radius obtained in Eq. 2.6 for the target. The number of cameras and their radial position was obtained through trial and error. Since plaster cast molds and residuums have convex shapes, results show no occlusion, thus the use of 4 cameras at 90° . These locations are marked in blue dots. Also, it displays the overlapping angles between cameras. The structure has an overall dimension of 534.50 mm height and an inner diameter of 610 mm where the camera's lenses are placed, as shown in Fig. 2.7b. As well as the adjustable locator with the ball head mount for all 4 cameras and the one end camera that captures the end view of the residuum. Previous work utilizes only 4 radial cameras, demonstrating a robust full capture of the target [24]. This study reveals the need for a fifth camera directly pointing at the head of the residuum, as it contains important skin information post-surgery [19]. So, with a total of 5 cameras, 4 radially placed every 90° and 1 camera at the base capturing the end view of the residuum. This, represents the mechanical design. Furthermore, mechanical drawings are available in Anex A.

2.2.4. User interaction

The structure interacts with the amputee when scanning. This is why, the system works in a non-contact scheme, where the design in C and the system can be accessed from a wheelchair or other sitting position. This prevents accidents caused by bumps or scrapes. Fig. 2.8 displays the example scenario of this interaction, highlighting three main parts, the roller chair with the amputee, the fixed frame, and the cameras.

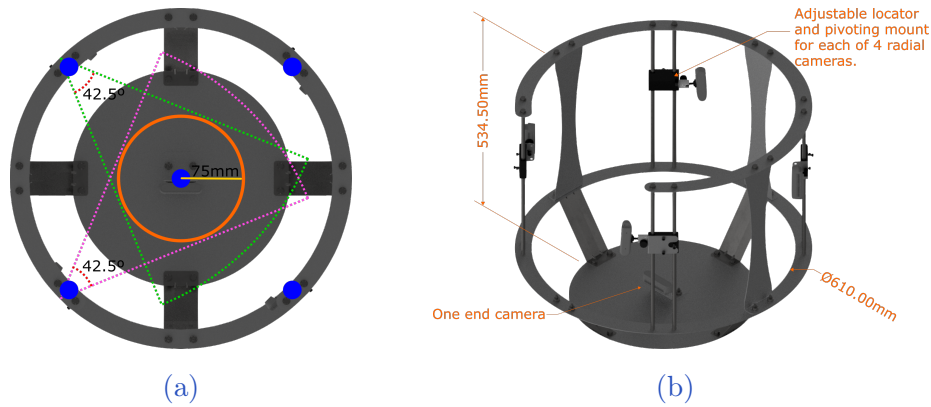


Figure 2.7: Mechanical design overview showing (a) Frame and cameras setup with overlapping areas, (b) Frame's overall dimensions.

For the placement procedure, the participant is seated at the edge of a raised surface, in this case, a roller chair. The amputation rests flying in a relaxation state, then the roller chair is slid towards the fixed structure by the operator. While gently, the residual limb is centered within the camera's angles. The opening in C allows the system to move comfortably to enclose the amputation, keeping it in the center of the structure. The residuum is kept at rest during the scan process. In the end, the operator removes the roller chair from the structure in the same way.

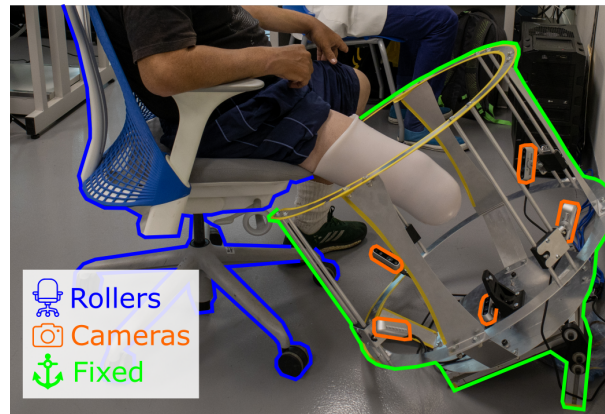


Figure 2.8: Example of the user interaction with the system, with a non-contact scheme. Labeled in blue, the rolling chair enables the user to move towards the frame. In orange, previously calibrated cameras aiming at the residual limb. In green, the fixed structure, with a C shape at the top for ease of access.

2.3. Calibration

Calibration is used to obtain intrinsic and extrinsic parameters. Intrinsic parameters describe the position of the internal components, in this case, the distance from one color sensor to the other and the position of the projector, depth scaling factor, and the

distortion parameters. These are needed to map out RGB-D images into PCs. While extrinsic parameters describe the rigid transformations of a global coordinate system, in other words, where are the cameras located in the physical space. The transformations are useful for global registration, helping as a first rough alignment between PCs.

Since RealSense cameras are used, most of our implementations rely on base code provided by the manufacturer from the official repository [29]. Intrinsic parameters are given by the manufacturer and may be accessed through the camera's firmware. On the other hand, extrinsic parameters need to be calculated. Intel® already provides a calibration demo¹ for the D435i cameras that are used in this project [29]. All cameras must be facing inwards, viewing a common 6×9 checkerboard as the calibration item as shown in Fig. 2.9. The calibration is performed by using Kabsch Algorithm [30], this algorithm finds the optimal rotation matrix between two sets of points. It superimposes both sets of points and minimizes the Root Mean Square Deviation (RMSD). This is realized in three stages.

First, the centroid of two point sets \mathbf{P} and \mathbf{Q} are calculated (2.13), which is the average position overall points and a translation of the centroid to the origin of the coordinate system is performed (2.14), this is obtained by subtracting the centroid from the point coordinates:

$$\mathbf{C}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \left(\frac{\sum_{k=1}^n x_k}{n}, \frac{\sum_{k=1}^n y_k}{n}, \frac{\sum_{k=1}^n z_k}{n} \right) \quad (2.13)$$

$$\mathbf{t}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = (x - x_c, y - y_c, z - z_c) \quad (2.14)$$

where \mathbf{C} is the resulting centroid, \mathbf{t} is the translation, x , y and z are point coordinates in a 3D space.

Second, the covariance matrix \mathbf{H} is computed in Eq. 2.15:

$$\mathbf{H} = \mathbf{P}^T \mathbf{Q} \quad (2.15)$$

where \mathbf{P} is the reference matrix and \mathbf{Q} is the matrix to be rotated.

Third, a Singular Value Decomposition (SVD), is computed from covariance matrix \mathbf{H} in Eq. 2.16. Then, a check sign takes place to ensure a proper rotation in Eq. 2.17. Finally, the rotation matrix \mathbf{R} is calculated in Eq. 2.18. This returns a superimposed matrix that "fits" the reference matrix by minimizing the RMSD in Eq. 2.19:

¹https://github.com/IntelRealSense/librealsense/tree/master/wrappers/python/examples/box_dimensioner_multicam

$$\mathbf{H}_{n \times m} = \mathbf{V}_{n \times n} \mathbf{S}_{n \times m} \mathbf{W}_{m \times m}^T \quad (2.16)$$

$$d = \text{sign}\{\det(\mathbf{W}\mathbf{V}^T)\} \quad (2.17)$$

$$\mathbf{R} = \mathbf{W} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{V}^T \quad (2.18)$$

$$\text{RMSD}(p,q) = \sqrt{\frac{1}{n} \sum_{k=1}^n |p_k - q_k|^2} \quad (2.19)$$

where \mathbf{H} is an $n \times m$ covariance matrix, \mathbf{V} and \mathbf{W} are orthogonal, and \mathbf{S} is diagonal. p and q are the k -th points of matrices \mathbf{P} and \mathbf{Q} respectively.

Once the calibration is complete, the program saves the intrinsic and extrinsic parameters. Fig. 2.9 displays the view from all 5 cameras looking at a common checkerboard target serving as the calibration object. Also, all the images display a green bounding box corresponding to the book underneath the checkerboard, correct orientation and dimensions validate a successful calibration of the cameras.

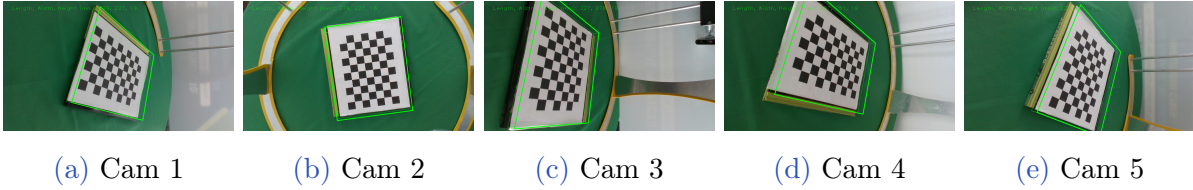


Figure 2.9: Calibration setup and the view from all 5 cameras. All sensors look at a common checkerboard as the calibration item. The green box displays the dimensions of the book underneath the checkerboard to validate that the process was completed successfully. This shows the simultaneous view of the calibration phase.

2.4. Segmentation

By default the cameras take RGB and depth images of the whole scene, this means that background, surfaces, and the target are present, which would add noise to the reconstruction. This is why segmentation is needed. It allows keeping only the desired object by removing unnecessary elements. OpenCV² [31] and Scikit-image³ [32] libraries are used for their segmentation algorithms. First, the original RGB image is taken in Fig. 2.10a. Then the background gets removed by using a depth threshold of 500 mm Fig. 2.10b, which leaves the target and the table previously covered with a green sheet.

²https://github.com/opencv/opencv_contrib

³<https://github.com/scikit-image/scikit-image>

Then, a color segmentation is employed by setting a color range filter U with an upper and lower bound $U = x_k - x_1$, the Hue, Saturation, and Value (HSV) color space is used to control by selecting darker and brighter values of the same color. This extracts the pixels corresponding to any green color on the image belonging to the sheet that falls into the selected color range (2.10c). This operation creates a binary image called "mask" that shows the pixels corresponding to that color range Fig. 2.10e.

Consequently, an algorithm checks an area with the biggest connected elements in the image, assuming the desired object is the largest on the remaining scene. To do so, Otsu's Thresholding algorithm is computed to find the intensity threshold value that separates pixels into two classes determined by maximizing its variance in Eq. 2.20:

$$\sigma_{\omega}^2(th) = \omega_0(th)\sigma_0^2(th) + \omega_1(th)\sigma_1^2(th) \quad (2.20)$$

where ω are the probabilities of the classes, th is the threshold, and σ are the variances. The resulting threshold is used to binarize the image.

Next, connected pixels of the same value represents an area. So, for every region pixels are counted, the region with the highest amount of pixels is assumed to be the target. In Fig. 2.10d an example is shown of the biggest element tagged with color and enclosed within a red box.

Afterward, median and erosion filters are applied to the "mask" to smooth out the edges of the remaining element, The median filter selects the median value pixel from a 15×15 squared region named kernel and replaces each entry with this value. This process is repeated 10 times swiping the whole image with the kernel represented in Eq. 2.21:

$$median(k) = \frac{k_{(n+1)}}{2} \quad (2.21)$$

where median is the resulting filter, k are the values inside the kernel. This filter is useful to smooth the edges from the previous colored segmentation as shown in Fig. 2.10f.

The erosion filter is a morphological operator, it computes the local minimum within a 3×3 kernel. This erosion takes place on the "mask" to slim the target and iterates 10 times. This helps to remove possible outliers that neither the depth threshold nor color segmentation could not resolve. The erosion of a binary image can be described in Eq. 2.22:

$$\mathbf{A} \ominus \mathbf{B} = \left\{ \begin{array}{l} \min \mathbf{A}(u) \\ u \in \mathbf{B} \cap \mathbf{A} \end{array} \right\} \quad (2.22)$$

where \mathbf{A} is the matrix image and \mathbf{B} is the kernel Fig. 2.10g.

The image in Fig. 2.10h is the result of doing a bit-wise operation of the "mask" over the original image.

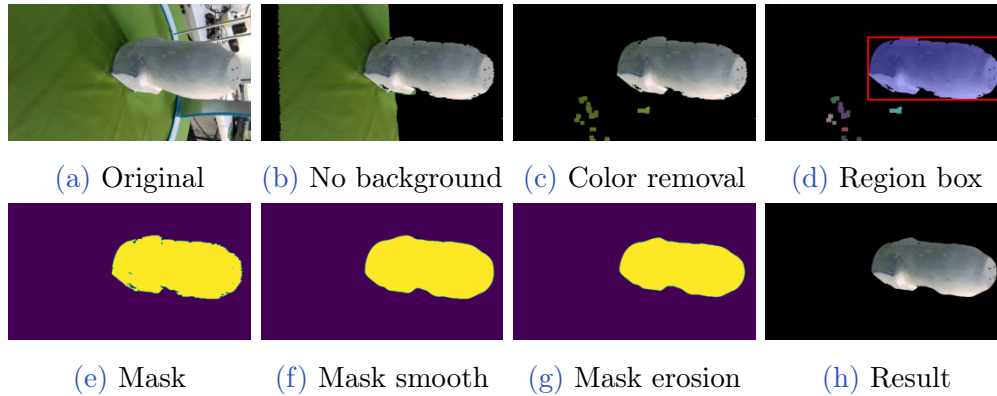


Figure 2.10: Automatic segmentation, feature extraction, contour smoothing, contour outlier removal, and masking process via the algorithms for one camera image. This shows the automatic segmentation from start to finish.

2.5. RGB-D alignment & Point Cloud creation

After the segmentation process completes, RGB and depth images must be aligned before converting them into a Point Cloud (PC). To do so, epipolar geometry is needed to find a correspondence between the depth image and the color image. This is also the case for stereovision with the D435i cameras, for estimating depth. The results of the epipolar geometry are the extrinsic and intrinsic parameters. From the factory, the D435i has these parameters to align the images. Following, a program⁴ converts aligned RGB-D images into a PC by mapping the depth information using the intrinsic parameters from each camera. This is performed by using this piece of code as a base and modifying it to meet the needs of this project [33]. Now that the color image is segmented, only valid color data is processed with the depth information, meaning that everything out of the segmented region is dropped out. This results in a clean PC in Fig. 2.11b, in contrast with a PC from the original scene as shown in Fig. 2.11a.

2.6. Reconstruction

All the reconstruction process is implemented using the Open3D⁵ library [34], as it can deal with 3D data, and contains alignment algorithms, surface reconstructions, filters, and an integrated visualizer. The process starts by importing the five PCs and apply-

⁴<https://gist.github.com/Shreeyak/9a4948891541cb32b501d058db227fff>

⁵<https://github.com/isl-org/Open3D>

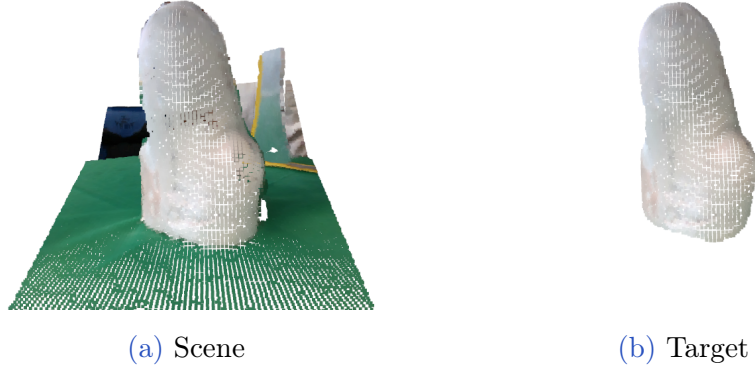


Figure 2.11: Comparison of the original scene and the segmented target. This shows the correct working of the segmentation algorithms and PC conversion.

ing the transformation obtained from the extrinsic parameters. Each PC suffers a rigid transformation by a 4×4 homogeneous transformation matrix \mathbf{T} in Eq. 2.27. This transformation corresponds to the initial global alignment obtained from the extrinsic parameters during the calibration phase and can be seen in Fig. 2.12a.

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.23)$$

$$\mathbf{R}_y = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.24)$$

$$\mathbf{R}_z = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.25)$$

$$\mathbf{t}(x, y, z) = \begin{bmatrix} 1 & 0 & 0 & \mathbf{t}_x \\ 0 & 1 & 0 & \mathbf{t}_y \\ 0 & 0 & 1 & \mathbf{t}_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.26)$$

$$\mathbf{T} = R_x R_y R_z \mathbf{t}(x, y, z) \quad (2.27)$$

Then, PCs are down-sampled by a voxel size of 0.002 for a faster processing time, being 0 the original size of the PC. Afterward, a refine alignment is performed with a pairwise registration for each acquired projection by using the Iterative Closest Points (ICP) algorithm, through a point to plane estimation with the objective function found

in Eq. 2.28. The ICP algorithm was preferred, as demonstrated to have the second best performance for aligning 2D PCs in previous work and since there is already an initial global alignment, ICP is best suited as a refine alignment [16].

$$E(\mathbf{T}) = \sum_{(p,q) \in \kappa} ((p - \mathbf{T}q)n_p)^2 \quad (2.28)$$

where E is the result from the objective function, κ is the correspondence set of points (p, q) from a target PC \mathbf{P} , and a reference PC \mathbf{Q} , n_p is the normal of point p . The pairwise registration is performed in the same order in which the cameras are physically placed on the frame. Once all of them are aligned, a Poisson surface reconstruction is performed [18]. This algorithm solves an optimization problem to obtain a smooth surface. Then, the resulting surface needs to be closed to be converted into a volume using MeshFix [35]. Finally, it can be exported as .PLY or any other mesh-supported format, the comparison of the resulting mesh and the initial alignment can be seen in Fig. 2.12.

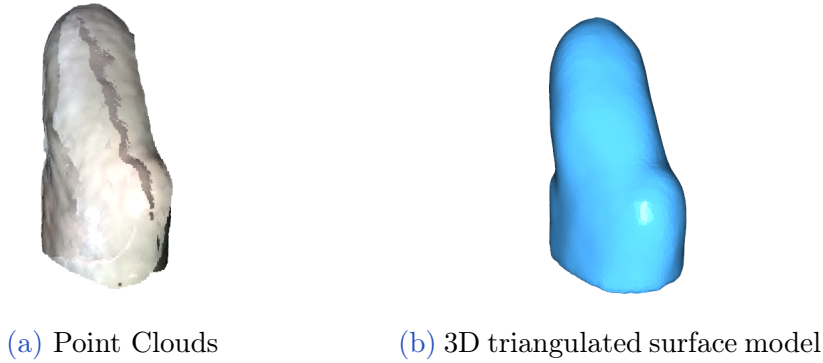


Figure 2.12: Comparison of a) the initial global 5 image alignment and b) the output model after applying a refine ICP registration and Poisson surface reconstruction. This shows the input and the output of the reconstruction algorithms.

These previously mentioned algorithms that integrate the reconstruction pipeline of the system are then validated by evaluating the segmentation and reconstruction phases using region and distance based metrics to quantify the performance of the algorithms used. These metrics are further described in the next section.

2.7. Validation

Validation is split into 5 blocks allowing different methods to demonstrate similarity, measurements, and soft tissue deformation capabilities. These are listed as follow,

- The functional and tested prototype.
- Segmentation similarity.

- Comparison of a CT model vs the scan with the proposed system.
- Diameter measurement verification with 4 plaster cast molds.
- *In-vivo* use demonstration with skin deformation.

For these comparison tests, Dice Coefficient (DC) and Hausdorff Distances (HD) are proposed to measure the error. On one hand, DC calculates the similarity between two samples with Eq. 2.29, where 1 is identical and 0 is no similarity.

$$\text{dice}(\mathbf{A}, \mathbf{B}) = \frac{2|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A}| + |\mathbf{B}|} \quad (2.29)$$

where \mathbf{A} and \mathbf{B} are images to be compared. The order does not matter, as the result would be the same.

On the other hand, HD metric measures the Euclidean distance between two sets with Eq. 2.30 and returns the worst-case scenario, where 0 is the best and ≥ 0 is the worst outcome.

$$d_{HD}(\mathbf{A}, \mathbf{B}) = \max\left\{\sup_{p \in \mathbf{A}} \inf_{q \in \mathbf{B}} d(p, q), \sup_{p \in \mathbf{B}} \inf_{q \in \mathbf{A}} d(p, q)\right\} \quad (2.30)$$

where \mathbf{A} and \mathbf{B} are images to be compared, **sup** and **inf** determine the least upper bound and the greatest lower bound, p and q are points.

First, a total of four molds were scanned. The first mold corresponds to a 3D printed model to which its digital model was obtained from a CT image from the transtibial amputee participant M1 from Proactible. The remaining three models correspond to plaster cast models obtained from Proactible's orthosis and prosthesis laboratory. Model 2 belongs to a transtibial amputee, while models 3 and 4 belong to transfemoral amputees. The segmentation results obtained in this phase were submitted to DC and HD metrics, obtaining a percentage value for DC and pixel count for HD, which can later be related using the system resolution obtained in section 2.2.2. Results for the segmentation phase can be seen in Fig. 3.4. Top row displays a manual segmentation using photo editing software taken as the ground truth and the bottom row shows the segmentation employed with the implemented algorithms. Both segmentations result in binary "masks" which are compared via DC and HD metrics, each model comprises 5 images.

Second, The M1 transtibial 3D printed model gets scanned with the proposed system. The 3D printed model was printed using HP's Multi Jet Fusion technology with a $\pm 0.3\%$ (with lower limit on $\pm 0.2 \text{ mm}$) accuracy stated by the manufacturer [36] and this error would be expected for validation. Segmentation results can be seen in 3.4a as stated

above. Then, the reconstructed model obtained with the system is compared via HD only. As the system outputs a surface mesh, HD best suits the application.

Third, the diameter of all four scanned models gets measured at the same view angle presented in Fig. 3.6 at three different heights. The bottom parts of the models were not considered to be included, as they serve as bases for the cutlines and are not important for the limb measurements.

Last, the system validates its use with a transtibial *in-vivo* participant M6 from Proactible. This scenario shows the capabilities of the non-contact automatic reconstruction of the proposed system. Two scans at different poses were obtained during this phase. The first pose in a relaxed state was obtained. A second one by co-contracting the residuum. A comparison between these two poses was computed through HD to visualize muscle contraction with the heat map. This color map demonstrates skin deformation and geometries displacements in Fig. 3.7.

The results for these four tests are displayed and discussed in the Results section, as well as the advantages of using this system and improvement areas for future development.

3 | Results & discussion

This section presents the results to demonstrate the objectives previously presented. The proposed system shows to accomplish the mount of 5 cameras and their calibration. It also can simultaneously capture across all 5 cameras, segment, and reconstruct automatically a 3D surface mesh. Then, segmentation comparison using DC and HD takes place demonstrating a strong resemblance to the ground truth segmentation. Next, the 3D printed model obtained from a CT scan gets compared to the results of the proposed system using HD metrics. Then, all four scanned models with the system get compared by diameter measurement via software and physically. Finally, the demonstration closes with an *in-vivo* participant.

3.1. Prototype

Hereby, I present the prototype and test of a functional capture system. The system was used in two different configurations for the capture and digitization of shapes. The first configuration is used for use with plaster molds as shown in Fig. 3.1, the structure is placed on a table covered by a green tablecloth. While the second configuration is used adjacently for use with participants as shown in Fig. 3.2. Where it shows the workspace, the calibration support and the user interaction. The frame has a "C" shape cut out at the top so the participant can be moved closer to the frame. A 45° position of the frame was chose so participants could easily place the residuum at the same orientation of the frame while sitting on a chair. Regardless of configurations, the frame demonstrates the camera mounting, the workspace, and the calibration support. This shows the frame holding 5 cameras with the ability to change orientation and position to direct the field of view to the target and calibration item. The frame can hold four cameras radially every 90° with 4 degrees of freedom (DOF), by using a ball and socket joint combined with a rail axis seen in Fig. 3.2a, and a fifth fixed camera capturing the end view of the residuum. Also, all the cameras capture simultaneously the desired object in place. An image of the workspace can be seen in Fig. 3.2b displaying the setup of the frame before being used by a participant. Looking at the top, the frame counts with removable support to hold the calibration target as displayed in Fig. 3.2c. Then, Fig. 3.3 displays the *in-vivo* non-contact application for external geometry reconstruction where a general view shows

the system and the participant. Then the following images demonstrate the automatic process that the system follows from the captured images up to the surface reconstruction resulting in the output mesh of the system.

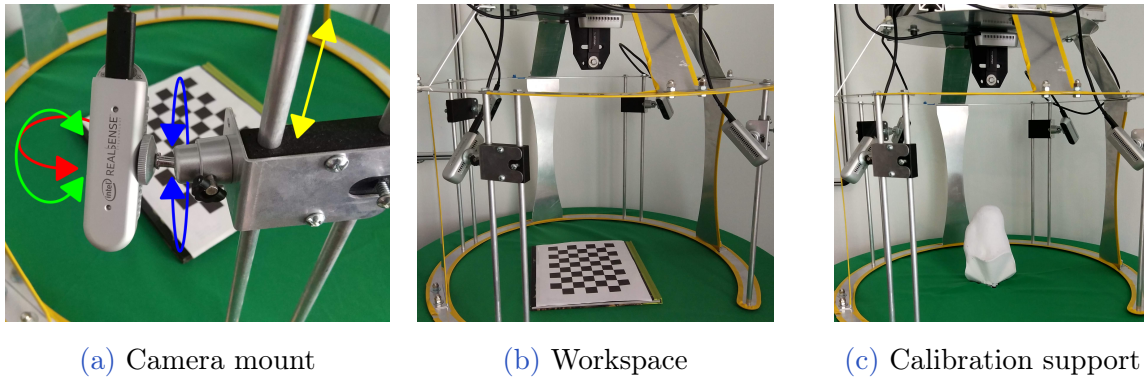


Figure 3.1: The configuration used for the plaster molds show a) 4 cameras mounted on a support that enables rotation and position, b) the workspace of the system being calibrated and c) the model being captured. Showing the camera mounts and the ability to change orientation and position onto the objective.

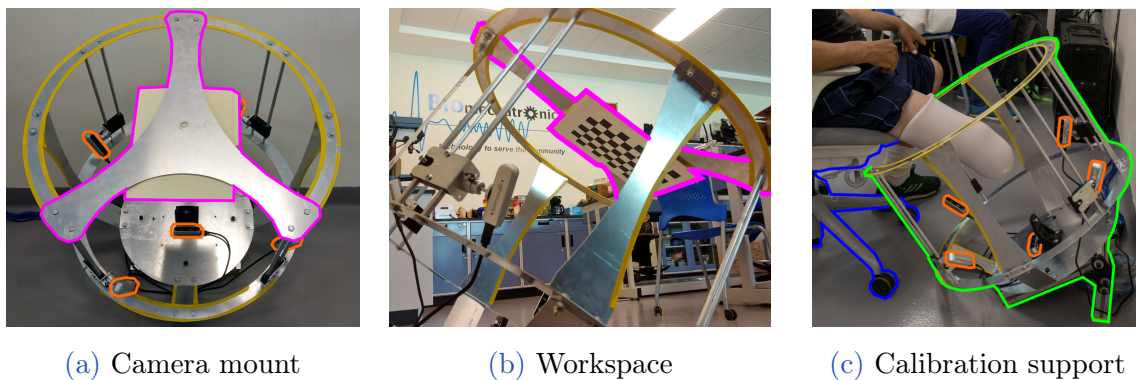


Figure 3.2: The configuration used for participants show a) the workspace of the system being calibrated and b) the removable support for target calibration in magenta, c) the interaction with the participant. Showing the camera mounts and the ability to change orientation and position onto the objective.

Fig. 3.3 shows the process from start to finish with the system previously calibrated. In the time span of 5 minutes, a full reconstruction from start to finish was accomplished with little intervention from the system's operator. During this time-lapse, it includes setting up the parameters of the camera, creating a folder for the current workspace. Also, executing the parts of the program that include the calibration, image capture, segmentation, PC conversion, and reconstruction. The system's operator intervention includes only inputting the data for the workspace creation, the placement, and removal of the calibration target, saving the desired images in the capture phase, and selecting

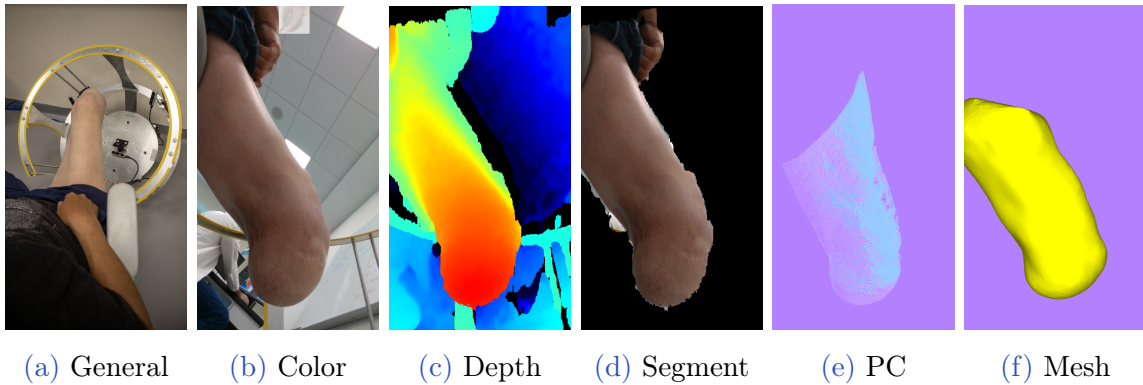


Figure 3.3: Testing of the system with an *in-vivo* participant starting from a) the general view of the system and the participant, b) the color image from one of the cameras, c) the aligned depth image, d) the segment of the target, e) the mapping of the PC and f) the resulting surface mesh after the 5 images reconstruction. This shows the process from start to finish of the system looked from one camera view.

the camera order in which reconstruction takes place. These may be performed previously as the setup stage, except the capture phase. Once concluded, the system is ready instantaneously for a full automatic reconstruction.

Wu et al. reported the average processing time for the computer with their system to be around 19 seconds including calibration, scanning, registration, and reconstruction. Furthermore, they mention that they used 6 structured light cameras, which did not allow the system for simultaneous capture and make use of multiplexing the cameras on/off which could increase the overall execution time for the system.

Simultaneous streaming was an issue in previous work using structured light. With the use of stereoscopic cameras, the issue was solved in this work. Although the system is configured for the simultaneous stream, cameras can be synchronized via hardware to work at the same clock. This feature would enable synchronous video recording to record continuous tissue deformation.

3.2. Segmentation results

Fig. 3.4 shows on the top part the ground truth segmentation, on the bottom the use of algorithms, at the top of each comparison displays the serial number for each camera. The footnote presents the numerical results corresponding to the Dice coefficient and HD. While DC presents a percentage, where digits closer to 1 means better. On the other hand, HD digits closer to 0 represent a better segmentation. Also, HD is measured in pixel count, error in mm may be obtained with the $0.24 \frac{mm}{px}$ resolution previously calculated in section 2.2.2. From the results presented in Fig. 3.4 a summary of the stats can be

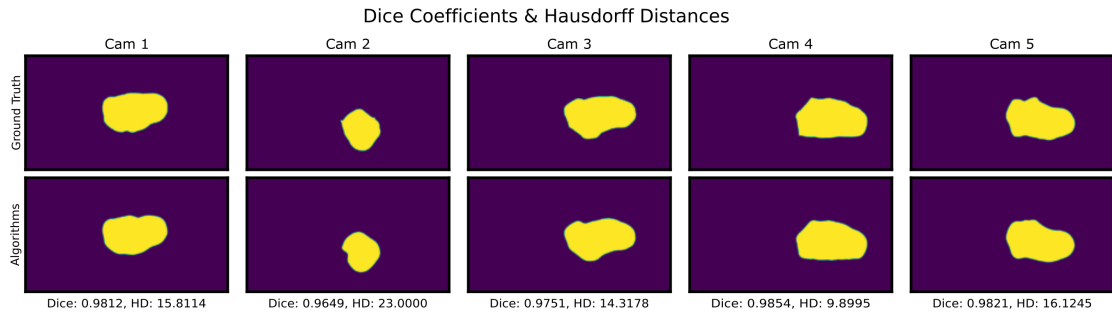
seen in Table 3.1 showing a mean of 0.87 for DC and 26.67 px or 6.40 mm for HD using the segmentation algorithms. This determines a strong resemblance to the ground truth segmentation. Other stats are available at the table such as the upper (3rd Qu.) and lower (1st. Qu) quartiles, the Interquartile range (IQR) that shows the difference between the 3rd Qu. and 1st Qu., and the root mean squared (RMS) calculated from the DC and HD data itself. It is important to notice that different volume molds were used. For the segmentation phase it can be seen in 3.4 that models are arranged from smallest to largest. Since focal lengths and the cameras themselves are fixed smaller objects would occupy a smaller portion of the image resolution. For smaller objects this could mean lower object resolution that could affect the error for the segmentation phase. Also, there are errors due to the average and erosion filters that remove a portion of the edge. However, the areas of interest for reconstruction are the overlapping areas of the shape, not the edges. So these errors do not present a problem for reconstruction.

Table 3.1: DC and HD statistics obtained from the segmentation results in Fig. 3.4. Showing a mean of 0.87 out of 1.00 for DC. Also, a mean of 26.67 px for HD which can be translated to 6.40 mm using the system resolution of 0.24 $\frac{mm}{px}$, showing a strong resemblance to the ground truth segmentation.

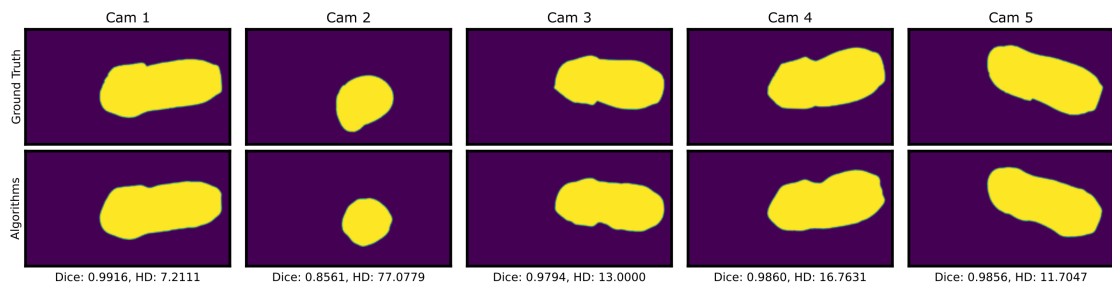
Type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	IQR	RMS
DC(%):	0.47	0.80	0.98	0.87	0.99	0.99	0.19	0.89
HD(px):	7.21	13.17	15.97	26.67	22.68	138.76	9.51	40.30
HD(mm):	1.73	3.16	3.83	6.40	5.44	33.30	2.28	9.67

For the validation of the results, it was sought to have a comparison of both regions and distance, so I proposed the use of DC and HD metrics. Segmentation shows a strong visual correspondence between the ground truth and algorithms.

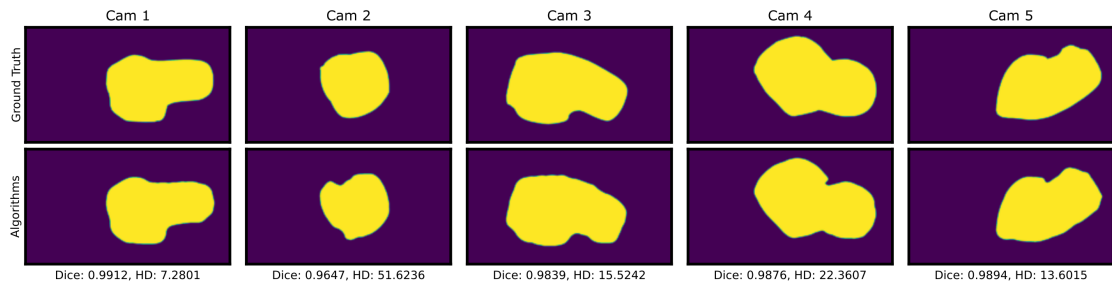
These tests show good results, in this part of the segmentation we see errors due to the averaging and erosion filters that remove part of the edges. In this case the sections of interest are those overlapping of the shape, not the edges. So these errors do not present a problem for reconstruction. But the algorithms can be further improved. Right now the algorithms employed to calculate the segmented region with the biggest element in the image make use of the Otsu's method to calculate areas within the image. Otsu's algorithm may calculate wrong areas if lightning conditions are poor or over-exposed. As well, the colors in the image. This specific algorithm works well with plaster cast models since they are white, because brighter areas get marked in white and darker areas get marked in black. During the *in-vivo* test, 2 out of 5 of the segmented images had trouble with the region selection due to this mentioned issue and manual intervention was needed. Since cameras can be automatically adjusted for light exposure, no additional light source



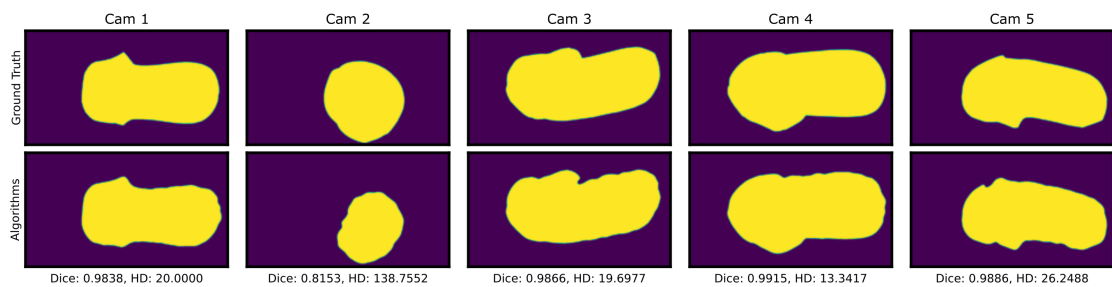
(a) M1



(b) Model 2



(c) Model 3



(d) Model 4

Figure 3.4: DC and HD metrics for color images segmentation for four scanned models. In each model, the top row displays the segmented object using manual segmentation as ground truth. Below, the segmentation results using the algorithms. This shows a strong visual comparison.

was needed. But the use of an external light source could result in smoother and uniform light along the target to minimize the shadows if the environment needs it.

3.3. Performance evaluation

Next, Fig. 3.5c presents the model comparison through HD for the M1 model. Where the original model in light blue was obtained with a CT scan, and the scan with the proposed system in yellow. A histogram of the HD results is shown at the left of the model, from the minimum value 0 *mm* up to the maximum value 10.7 *mm* going from blue through red respectively. Segmentation results for M1 can be seen in Fig. 3.4a above. Then, Table 3.2 shows the stats summary from the HD histogram presented, showing an error mean of 1.93 *mm* and a median of 1.27 *mm* which is ≤ 2 *mm* error proposed in the objectives. The median is preferred over other stats as it better describes the central tendency.

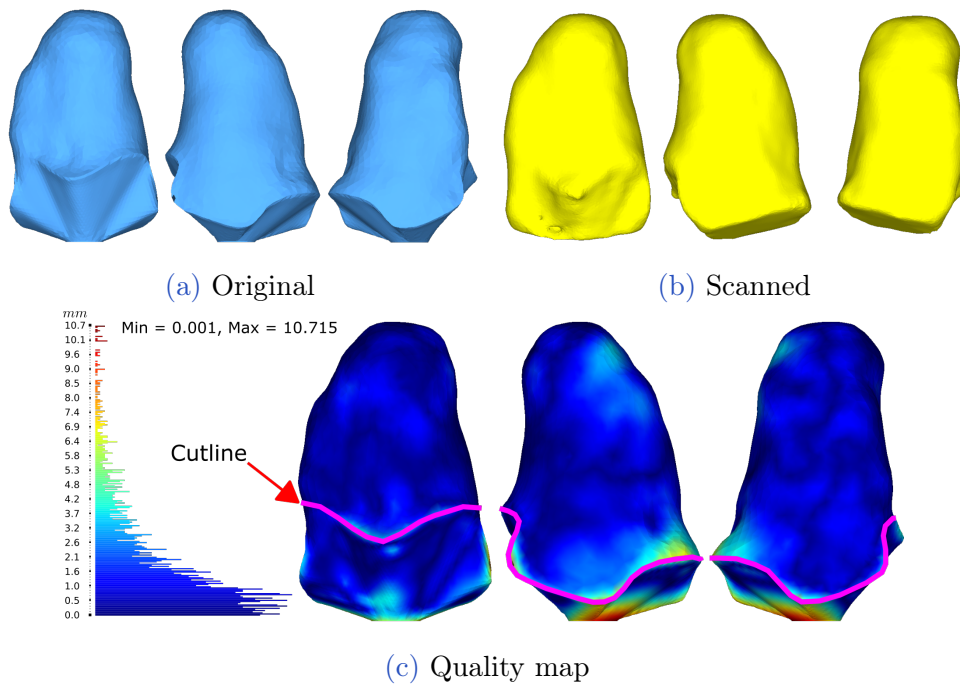


Figure 3.5: M1 model measured via HD. a) the original model from the CT, b) the scanned model with the proposed system, c) the quality map between the scanned and original models.

The bottom of the model colored in red represents the highest disparity for two reasons. First, the bottom part of the original CT model was modified for a proper 3d printing. Second, due to the position of the model placed over the workspace. Adding to this, the system algorithms creates a flat surface at the bottom where it touches the table. This region, although it shows the greater error is not important because it is below the cutline marked in magenta in Fig. 3.5, meaning that is not part of the socket and is

Table 3.2: HD statistics results were obtained from the model comparison. Values in *mm*

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	IQR	RMS
0.00	0.59	1.27	1.93	2.64	10.72	2.06	2.73

not desired to be compared. The important region resides above the cutline showing a central tendency with a median of 1.27 *mm* of error. By comparing results with Wu et al. with their foot scanner. They report two accuracy validations by using a sphere with a real circumference of 942 *mm* with an average measured value of 940 *mm* for the scanned model. Also, they compared their system against a commercial laser foot scanner reporting the largest error of 5 *mm* and 80 % of their data with ≤ 1.5 *mm* of error with a mean error of 1.42 *mm* [22]. Accuracy stats from previous work reported the use of a freeform and prismatic shape. The scale in that study was updated to meet the same units in this thesis for a proper comparison. The freeform showed a max error of 8.3 *mm* with a mean of 1.3 *mm* and the tangram showed a max error of 4.8 *mm* with a mean of 1.3 *mm* [24]. The characterization study for the SR300, demonstrated the multi-view performance using freeform and prismatic shapes with a max error of 2.5 *mm* [7]. Figs. 17-18 [7] show a lower and upper bound for their compared meshes, with no mean errors reported. The work of Solav et al. presented errors of 0.06 *mm* evaluating a cylinder with random patterns as the calibration target.

These mentioned studies related to the use of multi-view systems for reconstruction and the use of Euclidian distances for measuring error. The system presented here showed a median error of 1.27 *mm* which is less than the reported with some of the compared studies. The system proposed uses 5 stereoscopic cameras, while my previous work used 5 structured light cameras. Comparatively, Wu et al. used 6 structured light cameras and Solav et al. used 12 RGB cameras of 8 *Mpx*. The calibration procedure in this work utilizes a checkerboard using the Kabsch algorithm, previous work lacked a calibration target. Wu et al. use a T-shaped checkerboard pattern using a variation of Zhang’s method algorithm. Solav et al. use a cylinder with random dotted patterns. Another main difference regarding the calibration phase is that in this work and Wu et al. systems get calibrated previously during the setup stage, after that the checkerboards get removed before capturing phase. While Solav et al. paints a similar pattern over the subject that serves as markers to later correlate the images, resulting in less error reported from the discussed studies, but with the drawback of having to paint the test subject.

With the current system, the global alignment error can be reduced if the cameras setup is taken care of by obtaining values close to 0 during calibration with the Kabsch

algorithm. In a practical way, I have noticed that placing the camera as close to being perpendicular to the board gets better results. However, the system also makes use of the ICP algorithm that depends on a threshold, which influences the final error of the model. So there is still room of improvement with the ICP algorithm to minimize this error.

These results allow us to discuss the ideal tolerances that should exist between the socket and the residuum. Works in the literature focus their efforts mainly on the reconstruction and manufacture tolerances, but no interface tolerances have been reported today. This could be caused due to the skin's elastic properties and ability to adapt to unconventional socket shapes. Traditional methods work on a trial-and-error basis in adding and removing material from the mold for socket manufacturing. Recent techniques with the use of medical images are very accurate and can have resolutions of 0.3 *mm* previously mentioned in Table 1.1. However, the use of FE-models that optimize the shape by deforming the 3d model will depend on each participant, which will have variations for each socket. This would result in different tolerances for every socket. This is still an area of opportunity that needs further research. Current methods try to approach these tolerances by using pressure sensors that can evaluate the interface pressure to understand the loading conditions and possible pain thresholds [37–42]. The resulting mesh from the proposed capture system can be further processed to meet desired diameters which would be well suited for socket tolerances and manufacture.

3.4. Measurements verification

Furthermore, the scanned models get compared with the real diameter dimensions of the plaster cast models at the same view angle presented. Fig. 3.6 demonstrates all four models with their corresponding measurements. It shows the diameters at three different heights obtained via software (SW) of the scanned model and the physical counterpart. Additionally, M1 has the model obtained from the CT and the scanned model. The heights are proportionally distributed with respect to each model, meaning that the distances used vary and are not the same. This helps to verify that the measurements obtained from the scanned model resemble the physical model. Moreover, the different heights selected are considered to be below the cutline in the case of M1, models 2, and 3. For model 2, heights were considered to be below the knee, considering that the models are upside down. Also, the molds were arranged by size from small to large to demonstrate the dimension scan capacity of the system. Demonstrating that the system allows the scanning of small and large objects by adjusting the cameras with their respective 4 DOF. The largest mold tested has a diameter of ≈ 150 *mm*, as calculated in Eq. 2.6 to ensure a proper field of view of the target.

The obtained measurements of the diameters enable a fast verification of measurements

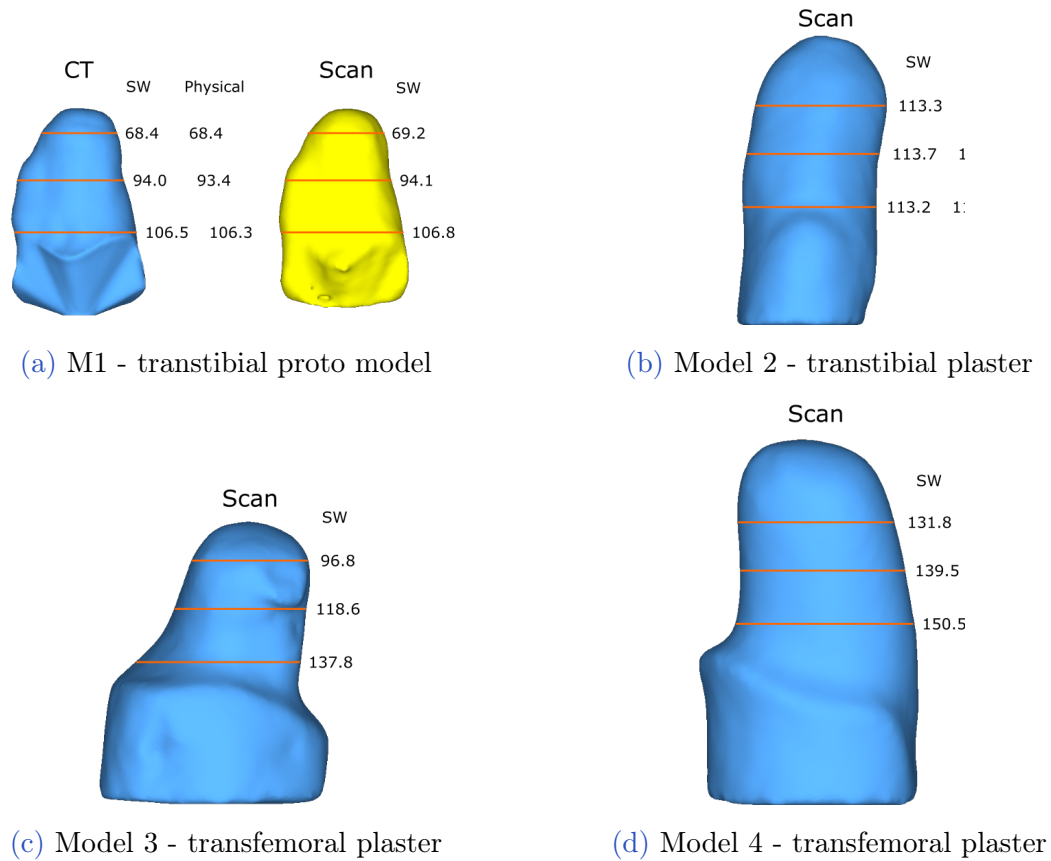


Figure 3.6: Diameter measurements across all 4 models. Comparison for both SW and physical diameters at 3 different heights. Units are in *mm*

using a caliper, by taking measuring the diameter at 3 different heights demonstrates that overall, objects follow the same dimensions. As mentioned above in the study of Wu et al. they compared a real sphere vs the scanned one having a difference of 2 *mm* for the circumference. The use of diameters from a single viewpoint was proposed due to the irregular shape of the amputation models which cannot be considered completely cylindrical. Initially, the measurement of the circumference was proposed but software limitations could not return the circumference, making the measurement of diameters at specific heights more viable. Overall error measurements remain ≤ 1.6 *mm*, having a strong resemblance with the real model. This proves to be useful for the assistance of the prosthesis in the prostheses manufacture. Facilitating the comparison of measurements of physical and digital models.

3.5. *In-vivo* validation

The results in Fig. 3.7 show the comparison between two different poses. The first one is in a relaxed state and the second one performs muscle co-contraction. The quality map shows the differences and changes from one state to another showing great differences in 3 main parts, the knee, vastus lateralis, being the external part of the quadriceps, and the residuum. The knee and residuum differences relate to the change in position, while the quadriceps region was caused by muscle contraction. Also, an artifact appears at the same spot in both poses at the top part, making a sink. This could possibly be occasioned by the reconstruction algorithm while stitching the mesh.

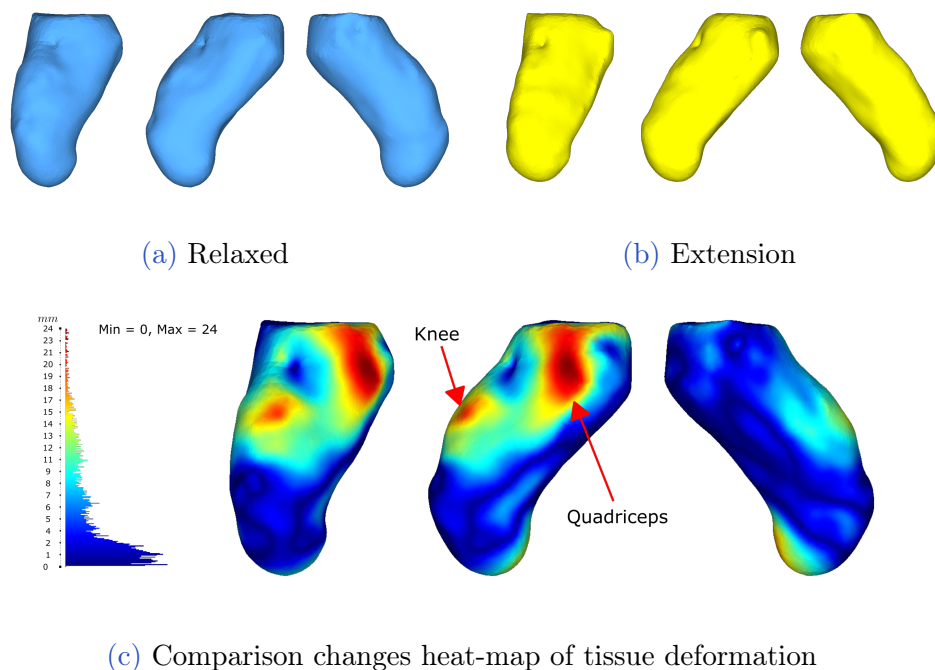


Figure 3.7: Participant M6 two poses results measured via HD. a) the residuum model in a relaxed pose, b) the residuum model while co-contracting the muscles, c) Quality map between both poses.

Overall this new functional and tested prototype meets the requirements stated in the proposal section 1.3. The proposed system serves as a useful tool to scan and create a digital library for plaster cast models. Also, the diameters measurements verification demonstrates the use scenario to assist the prosthetist for residuum dimensions and manufacturing process. Looking back into the traditional pipeline in Fig. 1.2, the integration of this system would be a hallmark for any local prosthetics clinic to be able to create dossiers and see patients' progress. The system would mean an update in the pipeline with the integration of the new technology as shown in Fig. 3.8. This integration means the scan of the current plaster model to re-fabricate it for the definitive socket after 6-12 months have passed. Another possible change in future pipelines could be the replacement

of the plaster cast models and directly scanning the amputee, resulting in a reduction of using materials and faster model iterations.

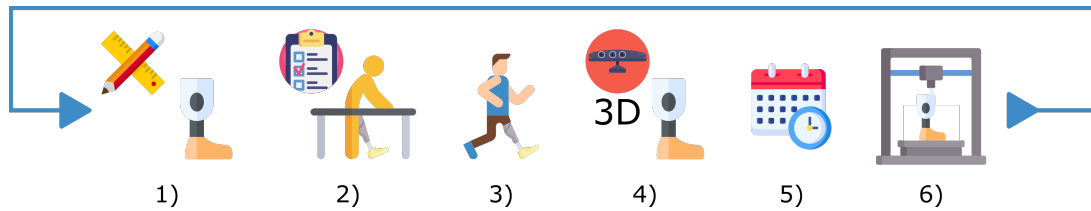


Figure 3.8: The traditional pipeline with the addition of technology integration. 1) Limb measurements and socket fabrication. 2) Check socket. 3) Provisional/Definitive socket. 4) Scan of current plaster cast mold. 5) 6-12 months before receiving a definitive socket. 6) 3D Re-fabrication of the mold from the digital library.

Furthermore, through the literature review, there is already a new socket design paradigm using CT scans, which enables the location of external and internal geometries that can be used along an FE-model for an automatic socket design. The proposed system can be used to update the external geometries of previous CT models to avoid radiation exposure. More on that, future work will involve the use of the proposed system along with multi-view DIC to properly extract the viscoelastic tissue properties to feed an FE-model.

Right now the reconstruction algorithm needs to input the camera order in which the reconstruction has to be carried out. This process needs the fixing of the first PC and making a pair-wise registration of the next PC and so on until all PCs have been applied the ICP algorithm. While it works adequately, it is not ideal. Algorithms can be further improved so key features can be detected and correspondences between images can be achieved to automatically determine the order in which registration should take place. Also, the use of body markers could be implemented to facilitate the detection of correspondences. This could also involve the calibration and the analyzing method used, which enables the possibility to implement a multi-view DIC to detect corresponding points across images.

4 | Conclusions & future developments

The work presented addresses the design, prototype, and testing of a new fully automatic capture system. The prototype was designed and manufactured out of laser cutting and bending of aluminum sheets that can support the mounting of 5 stereoscopic cameras with 3 DOF for position and orientation towards the target. The frame has a "C" shape cut out for easy interaction with the participant. The system can perform an automatic 360° reconstruction including the end view of a target in less than 5 *min*. The system presents the advantage of simultaneously capturing across all RGB-D cameras. Adding to this, it is calibrated to a global coordinate system via a checkerboard by implementing the Kabsch algorithm. Then the system segments the images to keep the target, which is then mapped out into PCs with an initial alignment. Afterward, the registration algorithms align and reconstruct the surface of these PCs, resulting in a closed mesh.

The results for the plaster cast models show the segmentation and registration validations. The segmentation metrics of DC and HD show strong visual similarity. Moving forward, the system was tested using a known prototype model obtained from a CT scan, to which comparison results via HD show an average error of $\leq 1.93 \text{ mm}$ making the system competitive against the state-of-the-art reviewed systems with only 5 stereoscopic cameras. The different diameters obtained of the four models through SW present a reasonably close resemblance to the physical models. These results from the objectives show that the system is useful for the generation of a digital library to create a patient history to see the shape progression over time, as well as assist the prosthesisist in the manufacturing process, the saved model can serve to update external geometries of previous CT scans, and future extraction of viscoelastic properties of the skin.

A full reconstruction was possible with only 5 images, since no occlusion areas were observed with the convex shape of the amputations including the molds, having no issue during the surface reconstruction with this application. Right now registration takes place by selecting the order in which cameras were placed. While results were promising, problems may arise if selecting the order of the wrong cameras. Image correlation, key features, or even color ICP may be implemented to overcome this issue. Additionally, calibration

via a traditional checkerboard method is sensitive to external disturbances and gets easily decalibrated, adding error to the initial global alignment. A cylindrical calibration target may be implemented along with the use of body markers to resolve this problem. The results obtained from the participant demonstrate that the system is capable for *in-vivo* non-contact reconstruction and analysis. Also, cameras can be synchronized via hardware to ensure an instantaneous capture of all sensors, which would enable the possibility of synchronous video recording to obtain an overtime shape deformation useful for the extraction of viscoelastic properties of the skin.

This thesis accomplishes to design, prototype, and test a new fully automatic capture system to scan plaster molds and serve as a platform for future technologies in lower limb reconstruction applications. The challenges involved the selection and integration of technologies that could be useful to comply with the application of developing an automatic capture system. In less than 5 minutes the reconstruction of an object can be acquired with little intervention with the proposed system. Current commercial and industrial solutions for 3D reconstruction consist of large equipment for which a dedicated area is required for its operation, such as full-body scanning systems. This thesis addresses this issue by scaling down a multi-view system to scan small and deformable objects over time, and minimize involuntary movements of the body.

Bibliography

- [1] E. Vázquez, “Los amputados y su rehabilitación,” tech. rep., Academia Nacional de Medicina de México, Mexico City, 2016.
- [2] P. K. Kumar, M. Charan, and S. Kanagaraj, “Trends and Challenges in Lower Limb Prosthesis,” *IEEE Potentials*, vol. 36, no. 1, pp. 19–23, 2017.
- [3] Z. Shuxian, Z. Wanhua, and L. Bingheng, “3D reconstruction of the structure of a residual limb for customising the design of a prosthetic socket,” *Medical Engineering and Physics*, vol. 27, no. 1, pp. 67–74, 2005.
- [4] D. M. Sengeh, K. M. Moerman, A. Petron, and H. Herr, “Multi-material 3-D viscoelastic model of a transtibial residuum from in-vivo indentation and MRI data,” *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 59, pp. 379–392, 6 2016.
- [5] G. Colombo, S. Filippi, C. Rizzi, and F. Rotini, “A new design paradigm for the development of custom-fit soft sockets for lower limb prostheses,” *Computers in Industry*, vol. 61, no. 6, pp. 513–523, 2010.
- [6] Q. Wang, Y. Tan, and Z. Mei, “Computational Methods of Acquisition and Processing of 3D Point Cloud Data for Construction Applications,” *Archives of Computational Methods in Engineering*, vol. 27, no. 2, pp. 479–499, 2020.
- [7] M. Carfagni, R. Furferi, L. Governi, M. Servi, F. Uccheddu, and Y. Volpe, “On the Performance of the Intel {SR}300 Depth Camera: Metrological and Critical Characterization,” *{IEEE} Sensors Journal*, vol. 17, pp. 4508–4519, 7 2017.
- [8] S. V. Riel, “Exploring the use of 3D GIS as an analytical tool in archaeological excavation practice,” no. June, 2016.
- [9] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR’06)*, vol. 1, pp. 519–528, IEEE.
- [10] S. Fuhrmann, F. Langguth, and M. Goesele, “MVE - A Multi-View reconstruction

- environment,” *2014 Eurographics Workshop on Graphics and Cultural Heritage, GCH 2014*, pp. 11–18, 2014.
- [11] R. Szeliski, *Computer Vision*. Texts in Computer Science, London: Springer London, 2011.
- [12] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, *KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera**. No. 11.
- [13] R. A. Newcombe, D. Fox, and S. M. Seitz, “DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time,” tech. rep., 2015.
- [14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press, 2nd ed., 3 2004.
- [15] S. Dey, “Hands-On Image Processing with Python,” pp. 471–489, Packt Publishing, 2018.
- [16] B. S. Dennis, M. Solís, and B. S. H. Aranda, “Registro de Nubes de Puntos 2D : Comparación entre Métodos Iterativos y una Red Neuronal Artificial,” *Revista Aristas*, vol. 8, no. 16, pp. 301–305, 2021.
- [17] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, “The ball-pivoting algorithm for surface reconstruction,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 5, no. 4, pp. 349–359, 1999.
- [18] M. Kazhdan and H. Hoppe, “Screened poisson surface reconstruction,” *ACM Transactions on Graphics*, vol. 32, no. 3, pp. 1–13, 2013.
- [19] D. Bonacini, C. Corradini, and G. Magrassi, “3D Digital Models Reconstruction: Residual Limb Analysis To Improve Prosthesis Design,” 2002.
- [20] R. B. Taqriban, R. Ismail, M. Ariyanto, and A. F. Y. S. Putra, “3D Model of Photogrammetry Technique for Transtibial Prosthetic Socket Design Development,” in *2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019*, pp. 456–461, Institute of Electrical and Electronics Engineers Inc., 12 2019.
- [21] R. M. Lacher, F. Vasconcelos, N. R. Williams, G. Rindermann, J. Hipwell, D. Hawkes, and D. Stoyanov, “Nonrigid reconstruction of 3D breast surfaces with a low-cost RGBD camera for surgical planning and aesthetic evaluation,” *Medical Image Analysis*, vol. 53, pp. 11–25, 2019.
- [22] G. Wu, D. Li, P. Hu, Y. Zhong, and N. Pan, “Automatic foot scanning and mea-

- surement based on multiple RGB-depth cameras,” *Textile Research Journal*, vol. 88, no. 2, pp. 167–181, 2018.
- [23] D. Solav, K. M. Moerman, A. M. Jaeger, K. Genovese, and H. M. Herr, “MultiDIC: An open-source toolbox for multi-view 3D digital image correlation,” *IEEE Access*, vol. 6, pp. 30520–30535, 2018.
- [24] M. A. Barreto, J. C. Huegel, R. Q. Fuentes, and J. Perez, “Non-static object reconstruction system based on multiple RGB-D cameras,” in *Twelfth International Conference on Machine Vision (ICMV 2019)* (W. Osten and D. P. Nikolaev, eds.), vol. 11433, p. 4, SPIE, 1 2020.
- [25] K. E. Smith, P. K. Commean, D. D. Robertson, T. Pilgram, and M. J. Mueller, “Precision and accuracy of computed tomography foot measurements,” *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 7, pp. 925–929, 2001.
- [26] P. Gadeberg, H. J. Gundersen, and F. Tågehøj, “How accurate are measurements on mri? A study on multiple sclerosis using reliable 3d stereological methods,” *Journal of Magnetic Resonance Imaging*, vol. 10, no. 1, pp. 72–79, 1999.
- [27] A. Grunnet-Jepsen, P. Winer, A. Takagi, J. Sweetser, K. Zhao, T. Khuong, D. Nie, and J. Woodfill, “Using the Intel [®] RealSense TM Depth cameras D4xx in Multi-Camera Configurations,” 2018.
- [28] Intel Corporation, “Intel[®]RealSense - Product Family D400 Series: Datasheet,” no. June, pp. 1–121, 2020.
- [29] Intel Corporation, “Intel RealSense SDK 2.0 [Source code],” 2017.
- [30] W. Kabsch, “A solution for the best rotation to relate two sets of vectors,” *Acta Crystallographica Section A*, vol. 32, pp. 922–923, 9 1976.
- [31] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [32] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Guillard, T. Yu, and the scikit-image contributors, “scikit-image: image processing in {P}ython,” *PeerJ*, vol. 2, p. e453, 2014.
- [33] Shreeyak, “create_pointCloud.py [code],” 2019.
- [34] Q.-Y. Zhou, J. Park, and V. Koltun, “{Open3D}: {A} Modern Library for {3D} Data Processing,” *arXiv:1801.09847*, 2018.
- [35] M. Attene, “A lightweight approach to repairing digitized polygon meshes,” *The Visual Computer*, vol. 26, pp. 1393–1406, 11 2010.
- [36] B. Alkaios, “HP MJF vs. SLS: A 3D printing technology comparison,” 2013.

- [37] S. I. Aguila, G. J. Sánchez, E. E. Sauvain, B. Alemon, R. Q. Fuentes-Aguilar, and J. C. Huegel, “Interface pressure system to compare the functional performance of prosthetic sockets during the gait in people with trans-tibial amputation,” *Sensors (Switzerland)*, vol. 20, no. 24, pp. 1–20, 2020.
- [38] A. L. Lenz and T. R. Bush, “Evaluating shear and normal force with the use of an instrumented transtibial socket: A case study,” *Medical Engineering and Physics*, vol. 71, pp. 102–107, 2019.
- [39] Z. Meng, D. Wai-Chi Wong, M. Zhang, and A. Kam-Lun Leung, “Analysis of compression/release stabilized transfemoral prosthetic socket by finite element modelling method,” *Medical Engineering and Physics*, vol. 83, pp. 123–129, 2020.
- [40] A. Ballit, I. Mougharbel, H. Ghaziri, and T.-T. Dao, “Fast Soft Tissue Deformation and Stump-Socket Interaction Toward a Computer-Aided Design System for Lower Limb Prostheses,” *IRBM*, vol. 41, pp. 276–285, 2020.
- [41] P. Sewell, S. Noroozi, J. Vinney, R. Amali, and S. Andrews, “Static and dynamic pressure prediction for prosthetic socket fitting assessment utilising an inverse problem approach,” *Artificial Intelligence in Medicine*, vol. 54, pp. 29–41, 2012.
- [42] M. C. Faustini, R. R. Neptune, and R. H. Crawford, “The quasi-static response of compliant prosthetic sockets for transtibial amputees using finite element methods,” *Medical Engineering & Physics*, vol. 28, pp. 114–121, 2006.

Abbreviations

	Description
CAD	Computer-Aided Design
CT	Computer Tomography
MRI	Magnetic Resonance Image
LiDAR	Light Detection And Ranging
RGB	Color sensors
RGB-D	Color and Depth sensors
SLAM	Simultaneous Localization and Mapping
PC	Point Cloud
RANSAC	Random Sample Consensus
ICP	Iterative Closest Point
DIC	Digital Image Correlation
HD	Hausdorff Distances
CPU	Central Processing Unit
RAM	Random Acces Memory
RMSD	Root Mean Square Deviation
SVD	Singular Value Decomposition
HSV	Hue, Saturation, and Value

List of Symbols

Variable	Description	unit
f	focal distance	mm
B	image size	px
G	target size	mm
l	distance to the target	mm
res	resolution	$\frac{mm}{px}$
i	image width	px
j	image height	px
c	circumference	mm
d	diameter	mm
r	radius	mm
w	width field of view	mm
h	height field of view	mm
C	centroid matrix	px
t	translation vector	px
P	point set	px
Q	point set	px
x	point coordinate	px
y	point coordinate	px
z	point coordinate	px
H	covariance matrix	px
R	rotation matrix	px
V	an orthogonal matrix	px
W	an orthogonal matrix	px
S	a diagonal matrix	px
p	k-th points of P	px
q	k-th points of Q	px
U	range filter	px

ω	probabilities of the classes	<i>px</i>
th	threshold	<i>px</i>
σ	variances	<i>px</i>
k	values	<i>px</i>
<i>median</i>	median filter	<i>px</i>
u	element from a matrix	<i>px</i>
A	an image matrix	<i>px</i>
B	an image matrix	<i>px</i>
T	transformation matrix	<i>px</i>
E	ICP objective function	px
κ	correspondence set of points	
n_p	normal of point p	
dice	dice coefficient	%
d_{HD}	hausdorff distances	<i>px</i>
sup	least upper bound	<i>px</i>
inf	greatest lower bound	<i>px</i>

B | Curriculum Vitae & Publications

Marco A. Barreto

Mechatronics Engineer

Mail: A01226069[at]itesm.mx

Phone: +52 1 3338 2047 22

[Reaserchgate](#), [Google Scholar](#)

Education

2021 MSc degree (Graduation in December)

Major: Machine Vision at Tecnológico de Monterrey

2019 Bachelor's degree

Major: Mechatronics Engineering at Tecnológico de Monterrey

Languages

- English - TOEFL Score: 91 - October 2019
- Spanish - Native
- French - B2

Experience

2020 - Current Biomechatronics Lab. at ITESM

Guadalajara, Jalisco

Master's Student and Researcher

The Lab. researches technology oriented to biomechatronical applications to serve the community. Currently, the Lab. has a research partnership with MIT's Media Lab involving the Extreme Bionics project.

I am a full time student and researcher in the biomechatronical field, engaged in the development of 3D reconstruction algorithms for non-rigid object and deformable soft tissues for the applications in prosthetic design. My research involves machine vision, machine learning and biomedical applications.

Lab website : gda.itesm.mx/biomechatronics

2019 - 2020 Catena Machinery Business

Guadalajara, Jalisco

Mechanical Design

Catena is a manufacturing company specialized in industrial machinery.

I performed as a designer, component selection, manufacturing and machinery testing.

- Welding self aligning pipe rotators 10Ton.
- Heat tube sealing machine for pharmaceutic industry.

Publications

2021

“Registro de Nubes de Puntos 2D: Comparación entre Métodos Iterativos y una Red Neuronal Artificial”

D. Mendoza, H. Aranda, Y. Fleitas, M. A. Barreto, J. Bravo, N. Hevia, J. Perez-Gonzalez
Iterativos y una Red Neuronal Artificial”

4th Conference on Computer Science and Computer Engineering (CoCSCE '21)

http://revistaaristas.tij.uabc.mx/index.php/revista_aristas/article/view/121

2020

“Non-static object reconstruction system based on multiple RGB-D cameras”

M. A. Barreto, J.C Huegel, R. Q. Fuentes, J. Perez-Gonzalez.

Proc. SPIE, Twelfth International Conference on Machine Vision (ICMV 2019).

<http://dx.doi.org/10.1117/12.2556357>

Technical Skills

Coding

3yrs Python

1yr C/C++, C#

Software

3yrs Inventor, Meshlab, Meshmixer, Matlab, Octave

2yrs GIT

1yr nTopology, SolidWorks

Other

Latex - Sphinx - Pycharm - Jupyter - Linux - Docker - Realsense - Open3D - OpenCV