Designing a Contactless, AI System to Measure the Human Body using a Single Camera for the Clothing and Fashion Industry

Mohammad Montazerian

Department of Computer Science

Goldsmiths, University of London

This dissertation is submitted for the degree of

PhD Computer Science

Goldsmiths, University of London February 2024

I dedicate this thesis to my loving parents, whose unwavering support and endless love have been my guiding lights, grounding me and providing me the strength to reach new heights.

Declaration

I, Mohammad Montazerian confirm that the work presented in this thesis is my own. Where information has been derived from other resources, I confirm that this has been indicated in the work.

Mohammad Montazerian
February 2024

Acknowledgements

I would like to extend my deepest gratitude to my supervisor, Prof. Frederic Fol Leymarie, for his unwavering support, friendship, and expert guidance throughout the course of this project. His vast knowledge, inspiring enthusiasm, and meticulous attention to detail have been pivotal to the realization of this work. His patience and encouragement have made this journey enriching and enlightening, fostering a conducive learning environment and enabling the exploration of new perspectives and ideas.

I am also profoundly thankful to my parents (Ebrahim, Akram) and siblings (Samaneh, Ali, and Saba), whose invaluable support has been my anchor during the ups and downs of this academic pursuit. Their faith in my capabilities and their moral support, along with the unwavering encouragement of my friends, have been unceasing sources of strength and resilience. Together, they have contributed significantly to my motivation and perseverance.

A special word of thanks is due to Nevena Nikolova, an experienced couture designer of women's fashion in London, whose assistance has been instrumental in the progression of my research. Her generosity in providing tape measures and her active involvement in data collection have been indispensable. Her willingness to leverage her customer base has facilitated the acquisition of robust and diverse data, enriching the quality and breadth of my PhD thesis.

I also want to express my heartfelt appreciation to my significant other half (Slaveya), who has been my constant companion and pillar of support throughout this journey. Her unwavering belief in my vision and her relentless support in times of hardship have been

the bedrock upon which this academic endeavor has flourished. The journey would have been significantly more arduous without her love, kindness, and companionship, and for that, I am eternally grateful.

In conclusion, the collective wisdom, support, and generosity of all mentioned have been the guiding lights of my journey, and it is to them that I dedicate the fruits of this endeavor.

Abstract

Using a single RGB camera to obtain accurate body dimensions rather than measuring these manually or via more complex multi-camera or more expensive 3D scanners, has a high application potential for the apparel industry.

In this thesis, a system that estimates upper human body measurements using a set of computer vision and machine learning techniques. The main steps involve: (1) using a portable camera; (2) improving image quality; (3) isolating the human body from the surrounding environment; (4) performing a calibration step; (5) extracting body features from the image; (6) indicating markers on the image; (7) producing refined final results.

In this research, a unique geometric shape is favored, namely the ellipse, to approximate human body main cross sections. We focus on the upper body horizontal slices (i.e. from head to hips) which, we show, can be well represented by varying an ellipse's eccentricity, this per individual. Then, evaluating each fitted ellipse's perimeter allows us to obtain better results than the current state-of-the-art for use in the fashion and online retail industry.

In our study, I selected a set of two equations, out of many other possible choices, to best estimate upper human body horizontal cross sections via perimeters of fitted ellipses. In this study, I experimented with the system on a diverse sample of 78 participants. The results for the upper human body measurements in comparison to the traditional manual method of tape measurements, when used as a reference, show ±1cm average differences, sufficient for many applications, including online retail.

Key Words: *Human Body Measurement, Machine Learning, Computer Vision, Ellipse equations, Mobile Scanner, 3D Scanner, Fashion Technology*

Table of contents

List of Acronyms

ABS	ABSOLUTE Function	xxxvi
AD	AdaBoost	149
AP	Average Precision	162
AR	Augmented Reality	19
ASPP	Atrous Spatial Pyramid Pooling	186
CCD	Couple Charged Device	33
CNN	Convolutional Neural Network	56
DEMs	Digital Elevation Models	36
DL	Deep Learning	134
DLT	Direct Linear Transformation	338
DPM	Deformable Parts Model	54
DSSD	Deconvolutional Single Shot Detector	54
FPN	Feature Pyramid Network	54
HMR	Human Mesh Recovery	122
HOG	Histogram of oriented gradients	xxiv
IBC	Image Based Capture	6
ISO	The International Organization for Standardization	24
M2Det	Multi-level Feature Pyramid Network Detection	54
mAP	mean Average Precision	105
ML	Machine Learning	134
MLP	Multi-Layer Perceptron	59
NMS	Non-Maximum Suppression	61

List of figures

List of tables

Chapter 1

Introduction

The main purpose of this study is to develop a state-of-the-art lightweight measurement system that is capable of extracting anthropometric measurements from multiple-depth images to capture the surface of the upper human body with application in the apparel/-fashion industry.

In this work, we seek to obtain a system which offers a number of key advantages over current off-the-shelf available platforms: simplicity, accuracy, flexibility.

Simplicity is in relation to our way of representing the human body as a series of adaptable elliptic horizontal slices, which although not following all body details, is the main shape favored by fashion designers, and well suited for most day-to-day clothing. Accuracy is in term of the needs in the fashion sector to achieve results at least similar to the main traditional measurements taken directly with tape. Flexibility is in term of designing a system which, while remaining simple in its concepts, can evolve with the progress made in technologies, such as provided by machine learning.

Underlying these goals, we implicitly take advantage of the symmetry of the human body, with respect to a normal standing pose. In this work, the human body is represented by a main vertical axis going through its center from a top highest (approximated) point at the tip of the head (when using a frontal view). Limbs (arms and legs) are assumed symmetric and ellipses only need be fit to one (arm or leg). However, while limbs are very well approximated by elliptic slices (again using a central approximate skeletal axis for each limb), we focus in this work on the more challenging parts of the human body: from the head down to the hips. With such a grounding in symmetry, we will describe a practical system which combines recent machine learning for object and contour localisation in images, together with more traditional computer vision techniques to refine our results, and selection of best equations to calculate elliptic circumferences in two main categories as a function of an ellipse's elongation.

Online shopping platforms have been attracting many customers since they were introduced mainly in the last decade. Customers can purchase any products or goods anytime and anywhere without the need to physically go from store-to-store to find a product or wait in a queue to check out. Furthermore, in the absence of person-to-person interaction (for example due to the COVID-19 pandemic), state-of-the-art, self-service contactless body measuring solutions are needed, enabling a simple way to digitize measurement capture so that made-to-measure businesses can easily operate online.

Existing body measurement systems available on portable devices (such as smartphones or tablets) are not able to satisfy both consumers and retailers due to a lack of accuracy and robustness.

From previous research, a digital circumference anthropometric system was developed using only images (static views). Anthropometric dimension measurement based on images has attracted attention due to its high potential for ease of use, portability, and low-cost [19]. However, there are several drawbacks in existing software solutions. There are significant errors between the direct (tape) measurement results and those based on software, especially for upper human body circumferences such as for the chest and waist. The problem emerges because there have not been a careful study of which mathematical models to use to better represent upper human body circumference anthropometric variables. Most existing approaches based on affordable consumer hardware, such as smartphones, heavily rely on user input. This implies that the user needs to follow specific instructions to get the right measurements, such as performing specific movements in front of the camera or standing still for different poses for specified amounts of time.

1.1 A short Overview of Contactless body measuring solutions

Non-contact human body measurements play an important role in surveillance, virtual fitting, physical healthcare, and online business. Usually, before biometric measurements[1], we must obtain human body models. The three-dimensional (3D) body scanner and laser range scanner[2] are existing biometric measurement technologies in the market that can provide accurate human body reconstructions [85]. Although these technologies provide accurate data, they are still very expensive, with a range of $40,000 to $500,000, and, also, to capture accurate information requires consumers to stand with tight or almost no clothes on in front of cameras [178]. Therefore, they cannot be an everyday choice for consumers in getting their body measurements in a short amount of time. Nowadays, markerless motion capturing

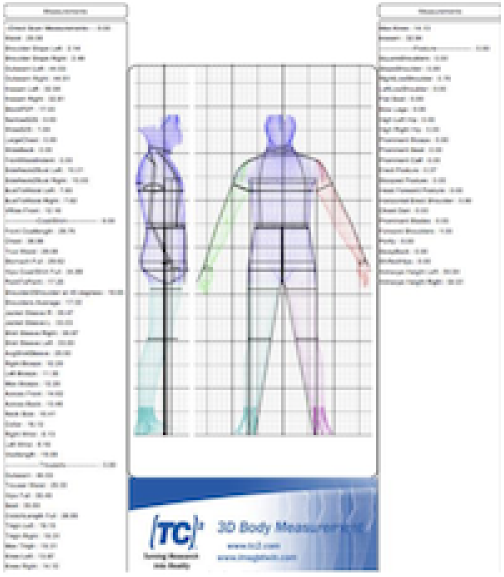


Fig. 1.1 A Sample of 3D body scanner into the (TC)2 company's database, [193, 68]

[1] The technical terms for body measurements and calculations known as biometric measurements. It related to metrics related to human characteristics

[2] The process of analysing an environment or real-world object to collect data on its shape and possibly appearance.

with multi-view systems are more and more capable and proficient at obtaining and acquiring human body models because of the efforts of scientists and researchers who spent their time on finding quicker technology for capturing the human body and make their life easier and save more time. However, this technology requires too much space and is difficult to set up. Recent attempts by researchers provide better solutions for capturing the human body measurements within seconds, by taking only two photos , from the front and side, or utilising a video (360-degree view) of consumers and developing a 3D model based on the measurements. Although this recent attempt still does not provide accurate data in comparison to the 3D body scanner, it is a much easier option for consumers to have with them every time they want. This recent attempt can be installed on any smartphone within seconds and users are able to download it. We can mention some of the applications as well as: 3DLookMe [1], Zalando [9],, TechMed [7], SizeStream (MeThreeSixty) [6], Esenca [3], PreSize [5] and VyoO [8]. Later on in this thesis, we delve into several of these applications to investigate their accuracy and reliability in comparison to state-of-the-art methods for measuring the human body.

1.2 Background and Motivation

Nowadays the young generation is preferring online shopping. A massive amount of clothes is bought on the internet [69, 77, 157]. The huge number of product returns is a major problem for online retailers. Customers often order the same clothing in different sizes to choose the right size and send back the others. This causes significant costs for the retailer and also has a huge negative impact on the environment.

According to National Retail Federation, the total value of return Goods in 2021 only for the US, was a total of 761 billion dollars [82]. The return rate could be dramatically reduced if customers had an easy way of obtaining anthropometric measurements. Of course, there should be better understanding and communication between retailer and customers to have a consistent definition of the anthropometric measurements relevant for selecting the size of the correct garment. Current approaches for analysing the accurate 3D shape of the human body often require expensive hardware. Furthermore, most of our software is not widely available, therefore it is not portable.

Additionally, by reducing the cost of returns, we can have significant beneficial impacts on the environment. The fashion industry is the second most water-intensive industry in the world, consuming approximately 79 billion cubic metres of water per year [113]. Considering that 2,7 billion people are presently affected by water scarcity, this stunning statistic becomes even more alarming [174]. This issue gets worse by the rapid fashion industry's insatiable thirst for water, while billions of people lack access to an adequate water supply. To put this in perspective, it takes an outstanding 2,700 litres of water to produce a typical cotton T-shirt, which is enough to sustain one individual for 900 days. We can contribute to a more sustainable future by addressing the high rate of garment returns and implementing strategies to reduce them.

Most existing approaches based on affordable consumer hardware rely on user support. This means that the user needs to follow specific instructions to get the right measurements, such as performing specific motions in front of the camera or standing still for a specified amount of time. Also, the privacy of such software in keeping the user's data is a big question mark.

With Gen Z - which makes up to 35% of the worldwide population [98] - now entering the workforce, uniform manufacturers and distributors must prepare for changing customer demands. Almost 50% of Gen Z customers want products tailor-made to their needs and taste according to IBM and the National Retail Federation [66]. Also, according to Imran et al. [81], more than 90% believe that companies are responsible to address environmental and social issues. With the help of advanced AI-powered technology, businesses can design, and tailor items based on accurate size, fit, and shape data for the Gen Z market. Consumers will be pleased not just with the garment they receive, but also with the beneficial changes that firms are making not only to the environment, but also a more sustainable industry

Motivated by the current situation, we propose a simple, yet effective technique to provide a state-of-the-art lightweight measurement system, for instance, based on a smartphone, capable of extracting sufficiently accurate anthropometric measurements to capture the relevant upper human body data, with applications in the apparel/fashion sector. The primary objective of this research is to address the inherent challenges in acquiring accurate body dimensions in the context of fashion technology, aiming to offer valuable insights and solutions to the industry by utilising a single RGB camera for precise measurements. This innovative approach, eliminating the need for manual measurements and complex multi-camera or laser-based sensors, holds significant potential for enhancing garment fit and design. Further details and hypotheses related to this research will be elaborated upon in Chapter 3.

1.3 Problems and Solutions

This thesis will address problems and possible solutions such as the Mo-Cap system [4] (including markerless), laser scanner (including 3D body scanner), smartphone body scanning (advance machine learning and computer vision methods), and image-based motion capture[3] to capture human body measurements. The problems of and solutions for constructing a 3D human body from a single pose and motion from images and video will be investigated.

One of the main problems of capturing the human body from a single camera, instead of using a stereo or multi-camera setup, is that the 3D information will not be available directly, in particular, because of the depth inference from a single image, the systems cannot predict the distance of the human body from the camera. However, it greatly

[3] Image Based Capture (IBC) is the process of capturing the human movement by using 2d tracking systems.

simplifies the process of data acquisition, also speeding up the process of human body measurement, and it potentially allows both archived film and video footage and rushes shot directly with the final production camera to be used.

By capturing the information on the human body in motion, we are potentially able to find the missing data that the single image methods were not able to provide. However, the process of estimating human body measurement will be much longer.

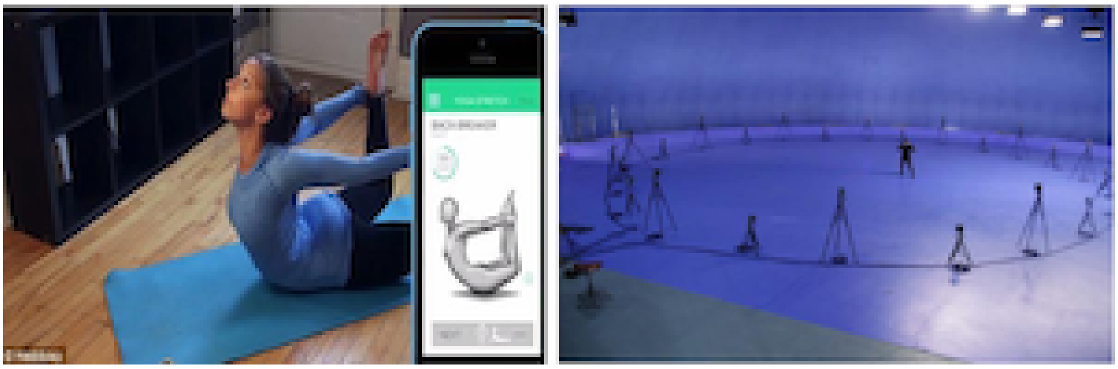


Fig. 1.2 The image on the left handside shows a single camera inorder to capture the human body and the image on the right handside is a multistereo motion capture studio. [116, 109]

Estimating the human pose, with the range of physiques and the variety of clothes that the human wears and the deformability, are all a real challenge. Once the light and camera viewpoints change, shadows and other variations in image appearance will have effects. Therefore, the reliable information that is extracted from image features is irrelevant.

Secondly, the complexity of the human body with the range of multiple degrees of freedom can be another issue in motion in both the rigid and non-rigid parts of the body. The part of the body which is observed under the shoulder girdle, will be able to be captured only in T-pose. Also, as the body moves, with the impact of breathing, the body expands and contracts. section 2.11 addresses these problems in more detail.

In terms of orientation and limb positions to characterise pose, a simplistic skeletal human body model usually requires between 30 and 60 parameters; therefore, the inference should be taken in a high dimensional space of possible 3D configurations, which makes the process yet more difficult [18].

With advanced research on machine learning, and computer vision, we can capture a higher quality image of the human body for measurements. Also, localisation and object detection are two core methods in deep learning which have been used in many existing applications to detect the human body in photos or video footage that was taken with any smartphone on any background. Furthermore, by creating a mathematical models for different human body shape to estimate upper human body circumferences, we can minimize the difference error between direct measurement and a software solution. Therefore, by deeply researching and investigating the machine learning, computer vision methods, and mathematical model we can improve the quality and accuracy of the data captured of human body performance.

1.4 Overview

Chapter 2 – Literature Review: In this literature review chapter, we explore the current state-of-the-art methods and technologies for human body measurement and evaluation, with a focus on garment fit. We begin by tracing the history of human body measurement techniques and the impact they have had on the fashion industry. We discuss garment fit and its challenges, followed by an examination of linear methods, 3D body scanning, motion capture systems, and smartphone body scanning technology. We review related work in human body detection and pose estimation, related technologies, datasets, and challenges associated with body measurement and pose estimation. Overall, this chapter provides an in-depth overview of the subject, setting the stage for the proposed software solution.

Chapter 3 - **Methodology**: This chapter outlines the methodology for developing this software solution to improve garment fit accuracy. It starts by stating the hypotheses and objectives of the research, followed by the selection of relevant human body measurements and techniques. The participants chosen for the research are limited to who self identified as females between the ages of 18-45. The limitations of the research are then discussed, along with the methods for data analysis, which include statistical analysis, t-test and anova-test. The chapter concludes by summarizing the methodology and emphasizing the approach used to achieve the research objectives and develop the proposed software solution. Overall, this chapter provides a clear and well-structured overview of the research methodology.

Chapter 4 - **Experiments**: This chapter examines experiments conducted on existing fashion and entertainment applications for measuring human body measurements. The chapter reviews the datasets used and describes the methods used to measure the human body, including fashion and entertainment applications testing, RMPE, measuring ruler, motion tracking for consoles and smartphones, PIFuHD, AR and deep learning-based automatic human body measurement system, and human body measurements using computer vision. The chapter evaluates each experiment's ability to accurately measure human body measurements and identifies the most effective method. The results of the experiments can be used to inform the development of the software solution proposed in this research project.

Chapter 5 - **Body Detection using MobileNet SSD**: This chapter provides an overview of object detection algorithms for automatic human body detection from a single image, with a focus on the MobileNet SSD method. It discusses the history of object detection, challenges, and popular datasets. The chapter compares several object detection techniques and presents findings on their effectiveness. It explains the region of interest and bounding box extraction methods and proposes the MobileNet SSD method for human body detection. The chapter also discusses the identification of body section names and differentiation of the upper and lower human body. Finally, it concludes with a discussion of results, limitations, and implications for the proposed software solution.

Chapter 6 - **Image Corrections, Image Segmentation, Skin Detection, and Camera Calibrations for Human Body Analysis**: This chapter delves into two main aspects of improving image-based data collection - image correction and camera calibration. Initially, the chapter outlines a software solution that emphasizes image correction to enhance the overall quality and appearance of uploaded images. It investigates two methodologies for image segmentation, comparing their effectiveness. The optimal method chosen is classical computer vision image segmentation. In this context, the specifics of skin detection using various color spaces are explored, followed by the detailed implementation of the skin detection algorithm. Experimental results comparing images with plain and cluttered backgrounds are presented to validate the accuracy of the segmentation process. Subsequently, section 6.3, titled "Camera Calibration", underscores the necessity to ascertain the distance between the camera and the human body for precise data acquisition. This section offers a comprehensive overview of diverse calibration techniques, with a particular emphasis on the algorithm provided by OpenCV. Research findings reveal that a distance ranging from 0.5 to 3 meters yields the most reliable data.

Chapter 7 - **Ellipse Equation**: The Ellipse Equations chapter of the thesis focuses on using ellipse-like approximations to accurately estimate human body circumferences. The chapter evaluates six different elliptical mathematical models, integrates them into the software, and presents the results of the study. The findings highlight the importance of choosing the right mathematical model to improve the accuracy of measurements in the human body. A fully trained system is developed that can choose the best ellipse equation based on the human body shape. The chapter emphasizes the potential for further research in this area and concludes with a summary of the findings and their implications for the proposed software solution.

Chapter 8 - **Results and Discussion**: This Chapter presents the results and discussions of this research, where we aimed to develop a software solution for precise human body measurements, taking into account variations in clothing, lighting, backgrounds, distances, and body forms. The software showed good accuracy, especially with mean errors of just 0.15 cm and 0.14 cm for shoulder and sleeve lengths, respectively. For circumference measurements such as chest, bust, waist, and hips, the accuracy was also strong, though the waist had a slightly higher error possibly due to loose clothes. When compared to existing solutions, our software was better in terms of accuracy, computational efficiency, and user-friendliness. In tests with 78 participants, the system had an average difference of ±1 cm against regular tape measurements. Notably, the software performed well in different backgrounds and lighting conditions. With potential improvements like user tutorials, we hope to reduce the measurement difference to around ±0.5 cm. In the end, our software offers a practical, cost-effective tool for the apparel industry, making processes more efficient and reducing waste and returns.

Chapter 9 - **Conclusion**: This chapter summarizes the methodologies, experiments, and outcomes from this computer science doctorate research, re-evaluating the primary hypotheses. Additionally, it outlines potential avenues for future advancements in this area of study.

Chapter 2

Literature Review

In this chapter, we provide a comprehensive overview of the current state-of-the-art methods and technologies for human body measurement and evaluation, with a particular focus on garment fit. We begin by tracing the history of human body measurement techniques, highlighting the evolution of these methods over time and their impact on the fashion industry.

Following this, we delve into the concept of garment fit and its importance in the fashion industry. We discuss the various factors that contribute to garment fit, including body shape and size, and explore the challenges associated with achieving a good fit. This leads us to examine various technical aspects of body measurement.

We explore linear methods for body measurement, including manual methods such as tape measurement and anthropometric methods. While these methods are widely used, we also discuss the limitations of these methods and their relevance in the context of modern technology.

We then move on to 3D body scanning technologies, which have become increasingly popular in recent years due to their ability to provide highly accurate measurements of the human body. We discuss the various types of 3D scanning technologies available and their respective strengths and weaknesses.

Another technology used in the field is motion capture systems, which are widely used to capture human motion and create realistic 3D shapes. We discuss the various components of a motion capture system and their role in capturing accurate motion data.

We also explore smartphone body scanning technology, which has emerged as a promising alternative to traditional 3D scanning methods. We discuss the various smartphone-based scanning apps available and their respective capabilities.

Moving on to related work in human body detection and pose estimation, we review recent research in this field, including various methods for detecting and estimating human pose and shape from images and videos. We also discuss related technologies that can be used to improve the accuracy and usability of our proposed software solution.

Additionally, we provide an overview of related datasets, including publicly available datasets of human body measurements and pose data. Finally, we examine the various challenges associated with body measurement and pose estimation, including technical challenges and challenges related to data privacy and ethical considerations.

Overall, this chapter provides a comprehensive overview of the current state-of-the-art in human body measurement and evaluation, setting the stage for our proposed software solution, and highlighting key findings and insights gained from our review of the literature.

2.1 A Brief History

The retail and fashion sectors have come a long way to make shopping as easy as it is today, but one area in which things appear to be growing more complicated with the time that has passed is the sizing of clothing items. All of these guidelines, which were made with the intention of simplifying the process of selecting an appropriate outfit, frequently have the opposite effect. So how did we get there? [164].

Measuring tapes as we know them today were initially developed in the early 19th century. They were considered as revolutionary because they allowed tailors to make accurate estimates of proportions between different body measures, which in turn led to the creation of drafting systems like the standard shirt. Ready-to-wear was made possible by the development of cutting equipment and then the sewing machine, both of which were based on machine technology [123].

The rise of machine-made clothing, casual lifestyles, and shopping malls contributed to the decrease of custom and made-to-order clothing in the 19th and 20th centuries. True custom, on the other hand, was an art form that required a significant investment of time and money. With a multitude of fittings that only a few were able to afford to buy.

The explosion of the fashion business over the past two decades has led to the phenomenal expansion of mass-market stores. As consumers began purchasing apparel online in the early 21st century, this business model produced a sizing and fit monster, which resulted in $751 billion in returned items in the United States alone last year, according to the National Retail Federation [82]. As a result of consumers' frustration with the complication of online apparel shopping, fashion companies struggle to comprehend what body types their customers have. Add to this an emphasis on the comprehensiveness of size and the contribution of the industry to environmental waste. Consequently, we are witnessing the return of the made-to-order business model. Therefore, technological progress enables innovative enterprises to create individual outfits based on the fit, body type, and performance of each customer.

Later on in this section, the evolution of body measuring will be discussed. Figure 2.1 represents a timeline of the body measurements' evolution with some key information about the advantages and disadvantages of each of these technologies.

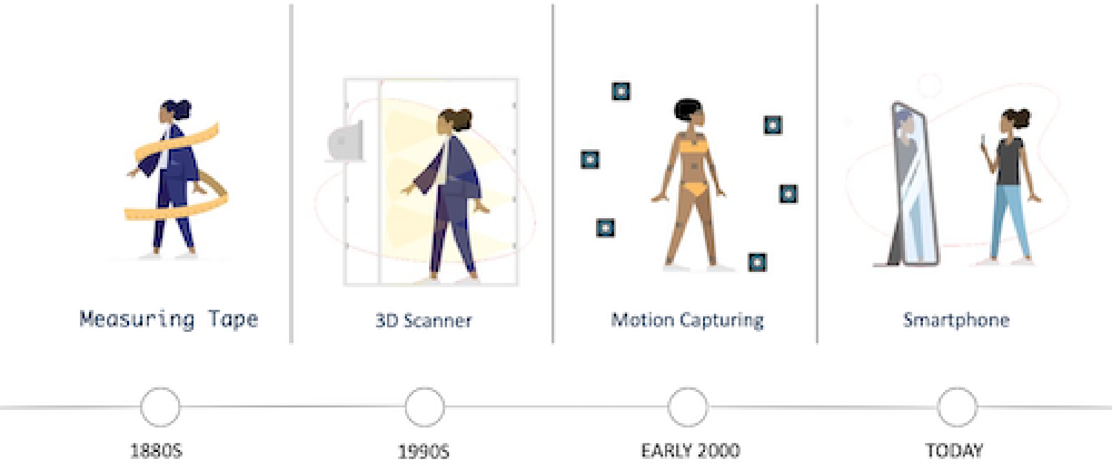


Fig. 2.1 The Evolution of body measurement [11].

2.1.1 Linear Methods

The accuracy of a customer's measurements is solely dependent on the skill of the individual using the measuring tape. It is both time-consuming and expensive to send expert tailors to the homes of customers or to have them work in the store itself. If a company wants to expand this process, they will need to hire a large number of tailors both nationally and internationally. Not only would this be prohibitively expensive, but it will also result in measurement standards that are inconsistent. Even the most experienced professional is unable to consistently achieve perfect precision in their work.

Some businesses demand that clients take their own measurements and have integrated drawn-out video tutorials or written guidance into the purchasing process in order to help them do so. On the other hand, these methods assume that the customer will have access to a measuring tape at home and will be able to gather the knowledge necessary to properly employ it. This results in a significant margin of error and, of course increase the return rates.

Made to measure and bespoke clothing companies are increasingly turning to the use of more advanced technologies in order to help optimise costs, time, and efficiency, and to provide a better experience for their customers. This is in response to the growing demand for custom-fit clothing.

2.1.2 3D Body Scanning Technology

The purpose of a 3D body scanner is to produce an accurate rendering of an individual's entire body in a 3D environment. The end product is a particular three-dimensional model that shows the exact form of the body and includes specific information regarding body measurements, posture analysis, and numerous features of the human body.

These technologies typically look like a closed booth, which is a cabin equipped with a number of cameras and sensors to capture an individual's full body. The customer must remain motionless while holding the same position during the whole scanning process while standing in the centre of the cabin. "Stitching" the photos together is the process that the 3D software does to assemble the final 3D model.

Sophisticated scanners can offer data for hundreds of measurements and are used not only in the fashion industry but also in fitness, healthcare, 3D printers, and many more domains.

Because of their size and the materials they are made of, today's physical 3D body scanners are rather bulky. Furthermore, these scanners still require the client to be physically present, which results in unnecessary travel time and costs. As was indicated in the previous chapters, the cost of a 3D scanner can be extremely expensive. The higher the level of precision a machine can achieve, the higher its price. A minor movement on the part of a human during the scanning process might cause a loss of

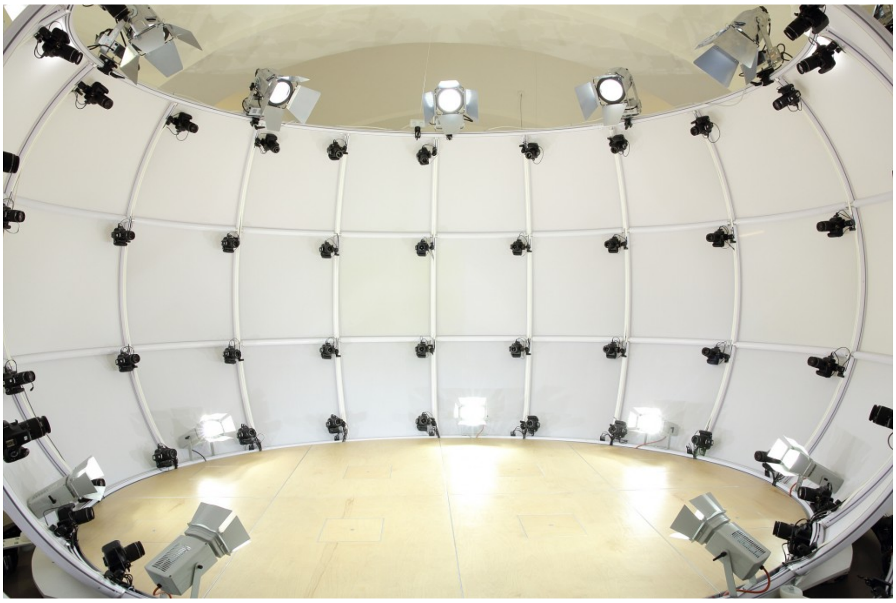


Fig. 2.2 3D Scanning booth [2]

accuracy of several millimetres, making the scan not acceptable. As a result, individuals are required to remain still during the process. The sensors used in 3D body scanners are extremely sensitive and react differently depending on the temperature, lighting, and the duration of time. The challenge of accurately adapting all of these relatively minor distortions is a difficult one [12].

2.1.3 Motion Capture System

Motion capture systems is the process of tracking the human body of live movement, and processing this information either live or by recording. The results is a specific 3D data model that shows the human body movements and provide skeleton with including details data for human body pose estimation.

These technologies require of a studio equipped with approximately 60 motion capture cameras. These cameras will record each and every aspect of the human body's performance and then translate it to the character avatar in real time. In addition, the individual who is going to execute particular movements in front of the camera needs to wear a suit that has approximately 36 markers placed all over the body.

This technology is used in many different fields ranging from military implementation to medical application to entertainment. This technology is been used only in few research studies in fashion industry for the purpose of the human body measurements and is been barely used in real life applications. However, this technology can offer so many valuable data in the fashion industry businesses such as 3D model. Also, this technology can be used in technologies such as Kinect, Virtual Reality (VR) and Augmented Reality (AR).

In the early stages of our research, we considered using motion capture (MoCap) as a potential tool. However, after careful evaluation, we found that while MoCap excels in pose estimation, it may not be the most reliable option for measuring precise human body dimensions. As a result, we decided to rule out this technique and explore alternative methods better suited to our research objectives.

2.1.4 Smartphone Body Scanning

The next generation of mobile scanning solutions has been enabled by the advancement of AI-powered technologies like deep learning and computer vision. In order to determine human body dimensions in seconds or minutes, the majority of body scanning solutions require customers to submit multiple photographs. This new technology offers substantial benefits to clients, including a shorter period of time required to obtain their measurements, more portability, and cheaper costs.

> *"AI and machine learning are transforming the fashion industry, enabling faster, more accurate and efficient clothing design and production. Automated body shape detection and size recognition are helping brands to better*

> *understand customer preferences, and reduce time and cost to market"* - Patrick Schwerdtfeger, Forbes.

Furthermore, the advancement of this technology enables the customer to digitally measure themselves wherever they decide to. Additionally, smartphone scanners can obtain information from the human body and provide dress measurements for the consumers and also create a 3D model of each specific consumer on the basis of the data that was obtained through mobile body scanning.

This innovative technology enables the consumer to digitally measure themselves anywhere they desire, allowing made-to-measure businesses to meet the customer's requirements while saving time and money. Moreover, despite the existence of 3D body scanners, this technology can measure people while they are wearing clothes, which makes it even much more handy and convenient to use than it already is.

The ability to provide instant access to additional, valuable data points for each consumer beyond just their measurements alone is one of the innovations that mobile body scanning technology brings to organisations in the modern fashion industry. This capability is one of the benefits of mobile body scanning technology. It is possible to segment both 3D models (avatars) and body form data in a variety of ways, such as by location, height, gender, race, and age.

With the aid of this new advanced technology, extensive customer profiles will be built based on measurement and shape data, providing made-to-measure firms with fresh customer insights and paving the way for a new age of custom-driven fashion.

2.2 Garment Fit

The theory of garment fit is the understanding of the relationship between the human body and the garment [42, 29]. Apparel fit is the relationship of the human body to the garment.

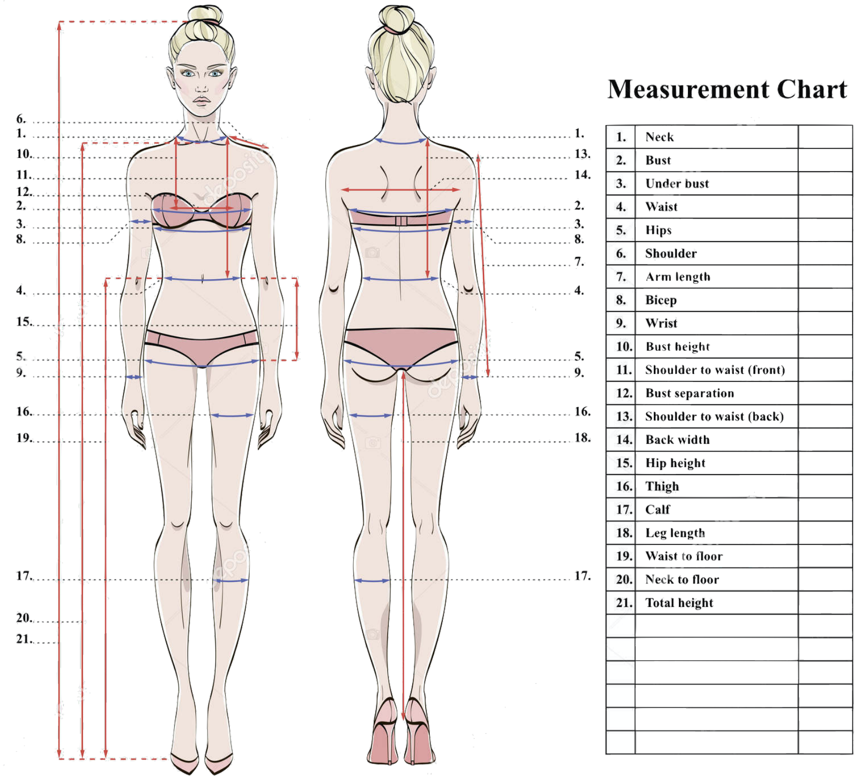
The fit was affected by a number of different elements. The most crucial factor is comfort, followed closely by the way the garment looks. Garment fit analysis is a procedure that evaluates the look, the wearer's feeling of comfort, and the garment's placement on the body. This process is used to determine the relationship that exists between apparel and the human body. It is important to perform fit analysis in order to achieve a good and accurate fit, as well as client satisfaction and acceptable style lines. Researchers have developed a method for visual analysis of clothing on the body, ways to judge the appropriateness of clothing, based on the responses from expert judges on the wearer's subjective perception of the garment and apparel. This method was developed on the basis of responses from expert judges on the wearer's subjective perception of the garment and apparel. Elements of fit include visual fit, physiological comfort, and physical comfort, all of which can be categorised as comfort. The aesthetic fit is evaluated by professional panels according to fit criteria as well as the comfort assessment based on the users' perceptions and preferences.

Comfort is one of the most important factors in clothing design. A pleasant state of physical and physiological harmony between the environment and human being. Cheng et al. [50], Slater [156] provided a definition of comfort, which is elaborated upon in the following three sections: To begin, the aesthetic appearance of the garment plays a significant role in the level of psychological comfort it provides. This includes factors such as appropriate fit, appropriateness for the occasion, flattering garment style, body image cathexis and fashion, and other psychological factors that affect comfort. Second, the physiological comfort of a person is dependent on the way in which the parts of the body and the clothing interact with one another mechanically. The "global feeling" about the garment and "the local feel" of the cloth against the skin are the two most important things to take into consideration [99]. Pressure comfort, which might include a feeling of tightness or heaviness, is related to global sensation. The term "local feel" refers to the sensations experienced by touch, such as roughness and itching. The last factor that contributes to physical comfort is the contact that occurs between a part of the body and the clothes that is worn. While physiological comfort is associated with the skin, physical comfort refers to aspects of the body such as size, posture, movement, and dimensions. According to Ashdown and Dunne [30], the fit of a garment is the result of interactions between a number of different aspects, including as the size of the garment, its dimensions, its drape, and the proportions and posture of the person wearing the garment.

The visual fit is yet another element of the fit. According to Erwin and Kinchen (1964), there are five primary components that make up visual fit: set, line, balance, grain, and ease. A good set is when the garment drapes about the body without creating unattractive wrinkles and the lines of the garment follow the contours of the body to create an attractive shape. A piece of clothing that is balanced both front to back and side to side might be indicated by its symmetrical design. Last but not least, the ease of movement, the ease of use as an additional fabric required for comfort, and a style that extends beyond body measurement should run parallel to the centre of the back and front of the grain. It is essential that a garment have enough fitting ease in order to fulfil the functions of being comfortable and allowing for movement. Defining the fit ease of a garment is especially significant due to the fact that both the wearer's bodily comfort and their desired style have an effect on the garment (visual fit) at the same time [158].

2.2.1 Female body Landmarks and body shape

For basic pattern development to secure accurate measurements, it is clear, according to Bye et al. [43], that consistent body landmarks as well as linear point and volumetric circumferences are essential. Anatomical points are commonly used as landmarks for garment pattern development. Corresponding to the anatomical points there are 21 key landmarks [61]. Figure 2.3 shows all the 21 key critical points over the figure of the human body.



Measurement Chart

1.	Neck	
2.	Bust	
3.	Under bust	
4.	Waist	
5.	Hips	
6.	Shoulder	
7.	Arm length	
8.	Bicep	
9.	Wrist	
10.	Bust height	
11.	Shoulder to waist (front)	
12.	Bust separation	
13.	Shoulder to waist (back)	
14.	Back width	
15.	Hip height	
16.	Thigh	
17.	Calf	
18.	Leg length	
19.	Waist to floor	
20.	Neck to floor	
21.	Total height	

Fig. 2.3 Key body Landmarks and Critical Anatomical Points by Fan et al. [61]. Image courtesy of https://depositphotos.com/171890380/stock-illustration-woman-body-measurement-chart-scheme.html

As discussed in section 2.2, it is essential to understand the relationship between the human body and the garment in order to have good garment fit. Therefore, to have a good garment fit, first, we need to understand the human body landmarks well before

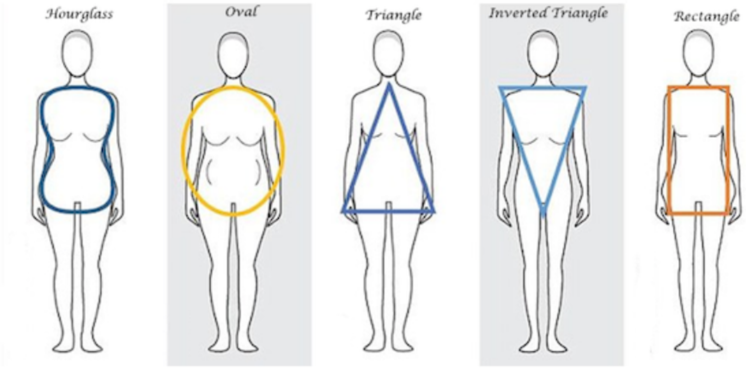


Fig. 2.4 Classifying the female body types. Image courtesy of http://www.michaelajedinak.com

starting the process of garmenting. In the next paragraph, we are investigating the key anatomical points.

2.2.1.1 The International Organization for Standardization

The International Organization for Standardization (ISO) is a worldwide federation of national standards bodies (ISO member bodies). The first of a series which deal with the federation and generation of anthropometric measurements, shape profiles, the creation sizes, and their application in the field of clothing is the international standards [13–15]. This section provides information about anthropometric measurements for females which can be used as a basis for the creation of physical and digital anthropometric databases. The list of measurements in this section is intended to help as a guide for the apparel (fashion) industry and those who are required to apply their knowledge to choose the population market segment and to develop size and shape profiles for the development of patterns of all garment types. Over 70 measurements have been collected. Most of them are from ISO 8559-1, and there is information from human body landmark points, volumetric circumferences, and linear methods. The upper-body information will be investigated for this research. More information about it can be found in section 3.2.

To take anthropometric measurements of the human body, the list will provide a guide on how to take the measurements of the human body, as well as information about fit mannequin manufacturers and clothing product development teams on the principle of measurement and their underlying anatomical and anthropometrical bases. The full list of measurements can be found in Figure C.2, Figure C.3, Figure C.4.

2.2.2 Garment Fit and Body Movement

2.2.2.1 Ease and Movement

Body movement change analysis and body movement analysis are the most prominent topics of research on ease allowance. Body movement analysis is one of the effective and efficient tools for the study of ease. Regarding the construction and elongation of clothing, it is necessary to examine the construction and elongation of each individual body part in order to properly identify the regions where ease is required. Using body movement analysis, one may estimate clothing comfort [80]. On the basis of Huck et al. [80], after analysing the total variations in back length, the required crotch ease in a one-piece protective garment was determined for a variety of active postures. In order to calculate the ease quantities, it is necessary to simultaneously analyse both the changes in body measurement and the changes in body movement. Choi and Ashdown [52] was able to distinguish the pants that provided the ideal level of comfort by applying measurements of changes in the lower body surface. The measurement around the waist grew by eight percent, while the measurement around the hips climbed by seven percent when compared to while the subject was standing.

Joint movements are the most essential element that has been identified in measurement change analysis and body movement. Muscle start to work with the skeleton once the joints move, when there is a movement in the body, there is a significant change in the positions of different joints, thus this will cause body skin around the articulations to be contracted and extended. Therefore, body skin change informs body measurement changes. Between clothing ease, joint movement and body measurements, there is an interaction found by Wang et al. [172]. In the case study that they conducted, they were particularly interested in the differences in human body measurements that were generated by various joint movements. Up to twenty different active postures have been measured as a result of this, including the joint motions of the knees, waist, hips, elbows, and one of the main joints which is shoulders. Because of the movement of the joint, the lengths and circumferences of the body around the joint were dramatically altered. The primary body measurements were determined based on the subjects' inactive postures. A total of 18 measurements were assigned, including the following:

1. Circumference above the waistline: elbow girth, bust girth, forearm girth flexed, and wrist girth
2. Width: horizontal shoulder and back width
3. Length above the waistline: arm length, under-armhole-point-to-waist length and the 7th-cervical-to-waist length
4. Circumference below and with the waistline: knee girth, mid-thigh girth, calf girth, ankle girth, hip girth, and waist girth
5. Length below and with the waistline: front and outside of the leg length, crotch depth

The highest reported increases and decreases, respectively, were in the under armhole, which was 8.4 centimetres, and the shoulder width, which was -15.75 centimetres. As a result, they came to the conclusion that the essential body proportions for comfortable clothing are the width of the back, the circumference of the hip, the circumference of the knee, and the length of the upper side.

2.2.2.2 Ease and Sizing

In the vast majority of research studies on sizing clothing, ease amounts have not been taken into account [42]. Regarding the sizing and grading of garments, LaBat and Schofield prepared an insight into the relationship between the anthropometric data and the pattern data in 2005. However, they did not use ease in their analysis because it was not discussed [150]. It would appear that the same level of ease is applied to each and every size in the range. The amount of ease varied from one size to the next but was always one inch at the waist, two inches at the hips, and three inches at the bust. The level of ease may also vary depending on the size of the body. Petrova and Ashdown [126] has investigated the ease of pants in terms of their dependence on size and shape. Ease quantities were estimated as a percentage of the relevant body measurements after comparing body measurements to garment dimensions and finding a mismatch between the two sets of data. The difference between these two measurements is referred to as the "garment–body percent difference." The distance between a body scan and a clothing item was measured with the help of a 3D scanner and technology for creating a 3D body. According to the findings, increasing the size resulted in a decrease in the proportion of the ease difference; however, there was no dependent form observed. There were no significant changes in the percent differences between the groups, with the exception of the hip circumference for size dependency. Their idea of comfort is solely predicated on how well the garment visually fits the wearer; the user's actual sensations of comfort were not taken into consideration. It is generally agreed that the ease of a garment is one of the most important factors in determining both comfort and appropriateness of fit. As a consequence of this, there are a great deal of questions regarding the relationship between size and comfort that remain unanswered.

2.3 Linear Method

The technique of gathering linear measurements over the body surface with a tape measure and then utilising these measurements to draw the pattern based on approximation and mathematical basis is what has traditionally been involved in pattern drafting in the garment industry [121]. The linear measurements are obtained by calculating the distance between two places on the body. Traditional measuring tools such as tape measures, anthropometers, and callipers are utilised in the process of recording the essential two-dimensional data related to a three-dimensional form.

Linear measures can be broken down into two categories: circumference and length. One example of a point-to-point measurement is the length measurement, which is taken from the shoulder to the wrist or elbow. This measurement is transferred to the design without any adjustments, such as taking away or adding to it, just as it was when it was obtained from the body. This procedure is straightforward and simple. A measuring tape is used to take a circumference measurement by going around the body in a circle, such as around the chest, the hips, the knees, and the waist. When it comes to the measurements of ease, further tolerance is required. A circumferential ease allowance is a component of any garment design system; however, the precise nature of this allowance varies depending on the type of garment being constructed [Hulme, 1954] [158].

Manual methods for measuring the body, such as tape measurements and manual anthropometry, have been shown to have errors ranging from 1cm to 3cm [140, 44, 62]. Despite the fact that the linear technique has been widely used in a variety of research with large sample sizes due to the fact that it is convenient, the linear technique is unable to describe the three-dimensional character of the human body [43]. The length and circumference of an object do not necessarily indicate its shape. For example, despite the fact that all three human models have chests that are exactly the same, their overall shapes are different from one another. As a result, the majority of researchers' efforts are directed toward the scientific investigation of different methodologies for identifying body form.

2.4 3D Body Scanning Technologies

In the fashion industry, the 3D body scanning system has become popular for consumers. There are several significant issues with 3D body scanners and the three main problems are cost, speed, and physical space requirements. There are many applications that have been developed. With this strategy, sellers are able to keep their current consumers by offering a value-added professional service, which in turn allows them to attract new clients who demand a more personalised fit. The numerous varieties of 3D scanners each come with their own set of advantages and qualities. While Table 2.1 provides an overview of the most prominent 3D body scanner manufacturers, Table 2.2 contrasts and compares the various types of 3D body scanners.

Note: 3D scanner information.

Table 2.1 Major 3D scanner manufacturers [61].

Light-based system			**Laser-Based System**			**Microwave-based system**		
Company	*Product*	*Optical Method*	*Company*	*Product*	*Optical Method*	*Company*	*Product*	*Optical Metho*
Fujinon	FM40SC	Moire topography	Cubic	Cubic	Laser Tech	Intelifit	Intelifit	
Hamamatsu	Body Line Scanner C9035-02	Infrared Tech	Cyberware	WB4, WBX	Laser Tech			
DCTA	Automate	Phase Shift	Hamano	Voxelan HEW1800	Laser Tech			
Hokuriku	Conusette	Infrared Tech	Inspeck	Capturor	Photogrammetry			
Loughborough University	LASS	Photogrammetry	Polhemus	FastScan	Laser Tech			
RSI	DigiScan 2000	Phase Shift	TechMath	VitusSmasrt	Laser Tech			
Poly U	CubuCam	Moire topography	Human Solution	Vitus LC Vitus XXL	Laser Tech			
TC2	ImageTwin	Phase Shift	Vitronic	Vitus Pro Vitus Smart Pedus	Laser Tech			
TELMAT	SYMCAD	Phase Shift						
Wicks and Wilson	TriForm	Phase Shift						
Turing	Turing C3D	Phase Shift						

Table 2.2 Comparison between different type of 3D body scanner [61].

Scanning system	Capture time	Weight (kg)	No. of Sensors	Accuracy	Room Condition
Cubic	1 sec	30	2	3mm	Dark
Cyberware WB4	17 sec	450	8	5mm $\times 2mm$	Dark
Hamamatsu BodyLine	10 sec	250	8	$\pm 0.5\%$	Dark
HamanoVoxelan HEW1800	10-60 sec	NA	8	$\pm 2mm$	Dark
Hokuriku Conusette	16 sec	350	6	$\pm 0.5\%$	Dark
Inspeck Capturor	0.3 sec	NA	2	4mm	Dark
Polhemus FASTSCAN	NA	NA	2	1mm	NA
Poly U CubiCAM	1/15000 $\times 3sec$	8.6	1	4mm	Normal
RSI DigiScan2000	1 sec	NA	12	0.2mm	NA
TELMAT SYMCAD	7.2 sec	50	NA	$\pm 2mm$	Dark
TC$^2 ImageTwin$	8 sec	NA	6	5-60mm	Dark
Vitronics VitusPro	8-20 sec	NA	16	1-2mm	Dark
Wicks and Wilson TriForm	12 sec	m1120	28	$\pm 2mm$	Dark

Methods of 3D body scanning made the job of demonstrating easier, provided vital 3D data of the human body, and demonstrated the possibility of enhancing the garment's fit. However, these experimental research were limited to the examination of fundamental garments. In the past two decades, 3D body scanning and its application in garment product creation have received increased attention. Using body scanner technology, a range of shapes and angles, as well as linear measurement data, including length, width, and circumference, may now be generated and enhanced for use in constructing the extraction pattern. To build ideal patterns and ensure a perfect fit, research has also been conducted on the development of promising advanced 3D approaches. The visual image of a 3D scanner can be rotated in order to observe a real human body shape on a computer. The data consisting of point, shape, surface, body volume, and line produces the most comprehensive form of any approach. In order to generate more precise descriptive approaches than conventional linear methods, these new technologies have the advantage of being less invasive and significantly faster than traditional methods [158].

McKinney et al. [112] conduct research to study the relationships between body patterns, body measurements, and the curves that are taken from body scans and used to generate pants blocks. The ease amount has been investigated in order to analyse the form of the pattern and the crotch shape of the body, as well as to research the relationship between the pattern and the body. According to the findings, the ease quantities vary in connection to body shape. As a result, in order to get a suitable fit, it is essential to determine pattern shaping device locations, amounts, and their link to body shape. The use of 3D scanning technology has allowed for an analysis of body shape, which included the prominence of the bust, the location of the acromion, the shoulder slope (angle), and the back curvature [185, 53]. It was determined that women's body shape information had a substantial effect on the fit of their bras when it was discovered that there is a relationship between bra fit and the shape of a woman's body. In a similar vein, research conducted by Shin and Istook [152] discovered that body form influences garment fit issues discovered by Istook and Hwang [84]. In addition, similar issues might arise when it comes to the fit of pants, which can be improved by employing body scan information to compare the dimensions of different demographic body types. They also discovered that even among people of the same type and size, different ethnic groups showed significantly differences in body shape and fit problems. There were distinct variations in physical size among the various ethnic groups [159].

Therefore, software and technologies such as laser scanning are the emerging methods for clothing design and body scanning. The data from a body scan can be converted into a digital format, allowing us to create specific designs for personal-fit garments.

2.4.1 Laser Scanning

Laser Scanning is a technological process based on light-plane and triangulation concepts. This technique uses laser beams to accurately and rapidly capture and measure the three-dimensional surface of the human body. The laser serves as a light source, and a piece of equipment known as the Couple Charged Device (CCD) scans the observed region. This is accomplished by measuring the distance light has travelled on a body. The 3D body scanner is capable of gathering as many as 60,000 points each second [26]. Due to technological advancements, the services provided by these technologies may now accommodate a wide variety of customers and project outcomes.

A laser scanner has a wide variety of applications, which were used to digitise the human body in a recent research studies including anthropometric data collection [103] ergonomic [104] and medical and sport training applications [95, 54] and also garment fitting simulation [51].

Prior to scanning the human body, a laser beam must be projected across the entire area of interest. Depending on the size of the human body part that needs to be captured, the number of laser sources may vary. Utilising the same triangulation technique, a range of commercially available laser scanners can digitally examine the complete human body. The fundamental difference is in the technique in which they capture and store images, as well as the position of their laser beams. Other scanners, such as those manufactured by Hamano Engineering, display vertical stripes over the body using two rotatable mirrors. This is in contrast to the majority of scanners, which utilise the vertical movement of the laser head and lay horizontal strips on the body. Such scanners are manufactured by Creaform, Vitronic, Hamamatsu, and TechMatch, among others.

By blocking the subject's head, we are able to acquire data of a higher quality, which in turn allows us to cut scanning errors by up to half while simultaneously cutting scanning time by thirty percent [56]. The time required for the laser scanning process ranges anywhere from 5 to 20 seconds. The acquisition time for medical scanners is equal to 0.3 seconds, and their accuracy is equal to 0.2 millimetres, which is a clear indication that their performance is better to that required by the clothing industry. The price of body scanners, on the other hand, reflects the fact that the cost of these technologies is not affordable for customers and that they remain quite pricey. The amount of money that participants are required to spend on these technologies is a significant drawback, as this significantly slows down the process of spreading this technology [125].

2.4.2 Structured Light Projection

According to Pribanić et al. [132] and Yu and Xu [183], 3D body scanners are generating an exceptional research effort based on structured light projection. This can be seen in both of these studies. Structured light has an advantage over laser scanning in that it will allow for the capture of the shape in a single step, which will result in a significant savings in the amount of time spent on the process of capturing the human body. Indeed, projection of stripes takes approximately one second, whereas laser scanning can take anywhere from five to twenty seconds. The quickness with which the subject is scanned contributes to the reduction in the total number of errors that are caused by the subject's uncontrolled movement. However, in order for it to function appropriately, you will need to make use of many units, each of which will consist of CCD sensors and a light projector, whenever the area that needs to be scanned is expansive. These scanners compete with laser scanners despite the fact that natural light poses a significantly lower risk to the user than the laser beam does. Additionally, because there are no moving components, the product is simple to use, requires no maintenance, and has a low sensitivity to colour, which are all additional advantages [125].

2.4.3 Image Processing

Image processing techniques are applied to create 3D representations of the human body from digital photographs alone. In this section, we will examine the various image processing application strategies.

2.4.3.1 Silhouettes Extraction

Multiple or single cameras, a single rotatable camera, or a platform can be used to extract distinct silhouettes from a set of photos of a person using a database of captured photographs. The photos are reportedly analysed in order to obtain 2D profiles, which, when combined, result in the generation of the 3D model, as stated by D'Apuzzo [57]. The volumetric representation of the physical body is collected from the junction of the visual cones described by Bottino and Laurentini [38]. This is accomplished by projecting each silhouette onto the viewpoint that corresponds to it. This method can produce disconnected protrusions or volumes that are not related to actual sections of the human body, as described by Percoco [125]. This is due to the extremely complicated geometric shape that is produced as a result of using this method.

2.4.3.2 Photogrammetry

The collection and matching of numerous photographs in photogrammetry enables the development of comprehensive 3D data, making photogrammetry an instantaneous 3D imaging method. Due to their insensitivity to even minute and partial motions of the human body, these methods are ideally suited for the digitisation of data relevant to the human body. In order to develop and construct a model of the human body, the 3D data of the participant can be collected either directly from the participant or by modifying a generic model of a previous case.

According to Simmons [154], photogrammetry is not commonly used to scan the surface of the human body because this duty is normally performed by scanners that employ structured light or lasers. However, the deployment of photogrammetry would result in a large decrease in these prices. The fact that the prices of these technologies are still relatively costly poses a barrier to the proliferation of these useful scanners. Chandler et al. [45] described the precise photogrammetric applications that might be carried out using the participant cameras. On the basis of their findings, they explored the potential of inexpensive digital cameras for close-range surface assessment by applying picture-matching techniques based on image features. In order to accomplish this, they examined the precision of three inexpensive consumer-grade digital cameras and extracted "Digital Elevation Models (DEMs)" Other related research endeavours, notably for medical applications, revolve around the matching techniques created by Ang and Mitchell [24] and the invention of more accurate calibration procedures by Chong et al. [54].

Furthermore, several studies in the body of academic research have been done utilising proprietary software to evaluate the quality of the data that was obtained for comparable applications. One such programme is PhotoModeler, which was developed by Eos System Inc. Larsen et al. [96] has evaluated it with persons who are about the same height, the inter- and intra-observer variability of human body measurements of dressed humans in two different examined and postures, and whether or not the lengths of body parts could be utilised to recognise one another. For example, the height was replicated to within 1.5 centimetres of the original in both the inter- and intra-observer studies. Taşdemir et al. [163] was able to acquire full body measures with the same level of precision by using a camera with a low resolution and by not setting any targets on the body.

According to Deli et al. [58]'s explanation, the use of targets has significantly improved the accuracy of measurements made with PhotoModeler for applications relevant to human facial digitisation. Sub-millimetre accuracy is achieved when comparing laser digitisation to the photogrammetric method using the Minolta Vivid 910i [125].

2.4.3.3 Microwaves Body Scanner

Kim and Forsythe [92] millimeter-wave scanning, which is utilised more commonly in the garment industry [126], is the foundation of more modern systems than the others. This is because these systems provide more precision. This approach uses low-power electromagnetic waves, commonly known as millimetre (mm) waves, to generate a three-dimensional depiction of the human body. Both millimetre waves and microwaves are suited for usage in anthropometric applications due to their potential to be biocompatible even at low power levels. The frequency range for millimetre waves is between 30 and 300 GHz, whereas the frequency range for microwaves is between 1 and 30 GHz. While digitising hundreds of thousands of points, the procedure typically takes no longer than ten seconds and has an accuracy of six millimetres. The biggest limitations are the poor precision and the unwillingness of individuals to be exposed to radio waves. Additionally, some individuals object to being exposed to radio frequencies. In addition, this technology shares difficulties with other laser scanners, and it is now very expensive for individuals to use [125].

2.4.4 Benefit of 3D Scanned Data

Taking measures of the human body with modern body scanning technology offers a number of significant advantages over more traditional methods, such as measuring with a tape. For instance, the amount of time users spend using body scanners is much shorter than the amount of time they spend having their body measures recorded in the typical manner. In addition, body scanners necessitate no physical contact, which eliminates the risk of erroneous, unreliable, and subjective assessment processes. When utilising conventional methods, detailed measurements and body models serve as the foundation for defining and designing one-of-a-kind patterns. As a result, the inherent observer error that can arise when employing these methods is avoided.

The human form can be captured using 3D scanners, which can make it much simpler for customers to purchase clothing that is both flattering and comfortable to wear on their bodies. They will even have the opportunity to test garments that have only been planned but not yet produced. According to this line of thinking, these could improve the level of happiness that customers have with the retail business by reducing the amount of time spent by customers in dressing and fitting rooms. According to the results of a survey conducted for this study fig. B.2, one of the primary reasons that online shoppers are likely to reduce the amount of time they spend shopping online is because they are unsure about their size. As a result, they are forced to travel to the stores in order to try on the clothing, and once there, they frequently change their minds about whether or not they want to purchase particular items of clothing due to the limited quantity that is currently available [26]. According to a separate research by the company 3Dlook [11], over forty percent of all online-purchased clothing is returned to the shop because the consumer was unhappy with the size that was provided.

The aggregation of data from a large number of consumers could offer merchants and manufacturers with the information required to design clothing that will likely fit a greater number of people. With the certainty that the hardware and software acquired by garment designers, manufacturers, and retailers would be suitable, the designers and manufacturers of scanners and the systems that utilise the scanned data may capitalise on wider markets. This confidence is buoyed by the availability of larger markets.

In general, 3D body scanning appears to have unlimited applications. Using 3D form data to construct an avatar for online gaming, performing body dimension analysis on target markets, maintaining health and fitness, and making motion graphics are examples of these uses. However, this concession about the measurement method is limited to a single instance. To enable continuous use of the stored data for new garment orders, for instance. Currently, it is not possible to share the scanned data of a single body scanner with other systems; as a result, the site where the scan was performed increases consumer loyalty to the store where the scan was performed.

2.4.5 Disadvantages of 3D Scanned Data

Several issues might be considered a disadvantage, while body scanning technology has the ability to greatly help the fashion industry and assist participants in acquiring well-fitting apparel when compared to the traditional techniques of garment measurement, part of the disadvantages are addressed by the type of technologies such as light, laser or microwave and how the image is captured and collected.

When employing light-based systems that rely on the interaction between the color of the scanning equipment, skin, and hair, the quality of image collection can be significantly impacted. This influence becomes evident in the final output, primarily driven by the lighting conditions in the scanning environment. In situations where the lighting is insufficient or dim, it becomes more difficult to achieve high-quality scanned images. This is because obtaining a favorable image outcome necessitates ample light reflection or a substantial contrast between the skin and the scanning apparatus. Moreover, various hair colors introduce specific challenges as they may cause issues when reflecting light. Additionally, misplaced or improperly positioned hairs can obscure critical indicators around the shoulder and neck area, affecting the results in both laser and light-based systems.

Due to the fact that both laser and light-based scanning systems capture the surface of the garment on the outside of the body, clothing worn throughout the scanning process will attract insects. Moreover, loose clothing will certainly increase the dimensions of the extracted measures, resulting in more inaccuracies and inaccurate data. Likewise, the opposite is also true. When consumers wear clothing that is too tight, the measurements will be smaller than they should be. Therefore, the recommendation for both systems is clothing that is close-fitting but not restrictive. However, it is anticipated that the microwave system (Intellifit) clothing has no effect on the measured measurements.

There are typically parts of the body that cannot be "seen", such as the shoulder girdle, the top of the head and the bottom of the feet, and the crotch at the junction of the legs. This is because the quantity of data-capturing devices in a system and their positions define which areas of the body can and cannot be caught by the scanning technology's vision devices. In general, a system with a higher level of sophistication will be more expensive. Numerous technologies have developed algorithms that generate "averaged" image data from the obtained data surrounding them in order to produce high-quality data on the mentioned hidden areas. Clearly, this is an oversimplification; the true technique is considerably more complex and sophisticated, and its major goal is to fill the region, such as a hole, with the least amount of effort required by the computer. Some of the missing information may be disregarded or ignored, If it has no importance or relevance to apparel development/fit, it is irrelevant [83].

The importance of 3D body scanning technology lies in the capability of the systems to not only record the 3D image of the human body but also to extract measurements at exact locations on the human body. This ability allows the systems to analyse the human body in a more accurate manner. The uniformity of the measuring procedures used by the scanners is a significant and important concern and issue. According to Simmons and Istook [155], even if the number of scanners that are currently available is increasing, there is a large gap between the methods that each scanner uses to extract or record a particular body measurement. Some systems require the establishment of physical landmarks in advance of the scanning process. This will give the human being the ability to select the appropriate location of significant points of measurement (landmarks), such as the hip point, shoulder, bust point, and so on. After that, the measurement-extracting software for these particular systems uses the previously indicated landmarks as the basis for the extracted values.

2.4.6 Comparison of existing 3D scanning companies

As was previously covered in this section, a large number of companies have developed software for the fashion industry that is based on 3D body scanning technology. As shown in Table 2.3, we are analysing some of the most well-known companies in this industry based on the scanning method they use, the capturing procedure, and, as a last point of comparison, the accuracy level they claim to achieve based on the datasets they utilise.

Note: This sign in table [1] means: Mini External hand scanners

Table 2.3 Comparison of existing 3D body scanners company in fashion industry

Company Name	Technology	Capture time	Accuracy
3DMD	Eye Safe laser technology	5-10 sec	$\pm 2mm$
Avalution (AVAone)	Depth Sensor Technology	1 sec	$\pm 5mm$
Avalution (AVAtwo)	Eye safe laser Technology	10 sec	$\pm 3mm$
Avalution (VITUSbodyscan)	Eye safe laser Technology	8 sec	$\pm 1mm$
Avalution (Tiger 3D)	Eye safe laser Technology (Foot Scanner)	5-10 sec	$\pm 1mm$
IBV	laser scanner (ShapeScan100/IBV - Foot Scanner)	6 sec	$\pm 0.16mm$
myeggO	Photogrametric 3D Scanner	5-10 sec	$\pm 2mm$
REVOPOINT[1]	Blue Light 3D Scanner	min of 2 mins	$\pm 0.5mm$
TECHMED3D[1]	Variety of Sensor Laser Scanners	min of 2 mins	$\pm 5mm$
ZOZO	3D laser scanners + ZOZOSUIT	2-5 mins	$\pm 3.7mm$

2.5 Motion Capture System

Researchers have been able to estimate the extent to which the human body expands or contracts during activities and over time as a result of advancements in motion capture technologies over the past several decades. Mattmann et al. [111] has proved that using an optical Mo-Cap system, body posture can be calculated. This method employs body sensors that are incorporated into a close-fitting clothing. They have measured the length of his back while sitting in various postures and with his clothing extended out to obtain the most accurate reading. Markers were positioned at a distance of every 5 centimetres, for a total of 90 markers. By analysing the distance that existed between each marker, the researchers were able to determine the elongation of the piece of clothing. Calculating the amount of stretch that clothes has allowed for can be a useful approach for determining how the body moves and where it is positioned. This conclusion was reached on the basis of the findings of the research study. Since it is not possible to make an assumption about the elongations of the skin surface area to the same extent as the elongation of the clothing and, during the activities, changes in body postures could be observed, the data that was provided cannot be translated into measurements of the body's surface area. The findings, on the other hand, indicate that it would be possible to use the Mo-Cap system to monitor the dynamic changes in the body. According to Zhou et al. [191], the Mo-Cap system was used to record the directional breast movements of women in order to inform the creation of women's sports bras. Reflective markers were used to record these motions. Although it was not quantified by Okabe and Kurokawa [119], the particular 3D breast movement path helps to improve the design of the sports bra when hiking or running. Although the findings of a number of studies have demonstrated a connection between body movement and the way clothing should fit, there is still only a small amount of research that has been done to measure the reconstruction of the human body for the purpose of applying it to clothing. In the past, researchers have focused mostly on variations to body surface measures that occurred during a variety of active postures as well as standard anthropometric positions that were tightly controlled. The natural motions that occur when people participate in real-world activities are more complex than the controlled postures utilised in previous studies. When employing Mo-Cap technology, a more natural, continuous motion may be considered. This enables for the measurement of the body's genuine dynamic changes.

The motion capture technology is employed extensively in the entertainment industry, as well as sports, medical applications, and the validation of computer vision. Nevertheless, with the aid of this technology, it is possible to calibrate the room to measure the distance between customers and the camera point of view, develop a 3D model based on the shape of the human body, and many other things in addition to measuring the human body for the fashion industry. This technology itself is not a good option for measuring the human body for the fashion industry.

2.5.1 Benefit of Motion Capture System

In the fashion industry, the ability to capture the human body through the use of a motion capturing equipment has a variety of significant advantages. For instance, (i) it provides immediate and real-time results, (ii) it enables numerous tests to be conducted with different styles or delivery methods, (iii) it is capable of recreating complex movement and realistic physical interactions in a physically accurate manner, (iv) it is capable of calibrating a room, (v) it provides information such as the distance between users and their surrounding environments, and many others.

2.5.2 Disadvantages of Motion Capture System

When it comes to motion capture systems, there are a number of issues that could be deemed drawbacks. For example, in order to generate data that is unavailable to a large number of individuals, this technique necessitates the use of certain hardware and a specialised software application. It is possible that the capturing system has specific requirements for the environment in which it is operated. When it comes to capturing the human body with a stereo[1], not a lot of people have access to or are able to make the effort to purchase such a system. Additionally, a motion capture system primarily provides information such as human body movements and data as the skeleton. This process requires specialised skills and expertise, which not a lot of people possess. Capturing and accurately interpreting human body movements using a motion capture system involves setting up the equipment, calibrating the system, ensuring proper marker placement on the body, and effectively capturing the movement data. Moreover, processing and interpreting the captured data to extract meaningful insights, such as skeletal information, also require specific knowledge and experience. While some researchers have been able to measure the human body with the assistance of this technology [158], it is important to note that utilising motion capture for measuring the human body in industries such as fashion is not yet straightforward. The technology is still in development, and it requires a level of expertise that is not commonly found among the general population. Furthermore, motion capture systems can be quite expensive, which adds to the challenges of widespread adoption in certain industries.

2.6 Smartphone Body Scanning technologies

The use of smartphone body scanning technology has been increasingly commonplace in the apparel and fashion industry over the past few of years. This technology is a promising solution powered by artificial intelligence that enables body measurements to be taken with a high level of accuracy. In most cases, these solutions involve the utilisation of smartphone devices in order to provide customers with correct measurements regardless of where they may be. These days, there is a variety of choices for

[1] Stereo refers to the use of two or more cameras placed at different positions to capture the motion of an object or a person from different viewpoints. By having multiple camera views, it becomes easier to track and reconstruct the 3D movement of the subject. This technique allows for more accurate and detailed motion capture data.

mobile body scanning solutions on the market, and the procedure for scanning differs depending on the company that provides the equipment. AI body measurement tools typically require the customer to take only a few photos (front/side) or record a video (360 views) of their body in order to get scanned. This is in contrast to manual measuring or hardware scanners (i.e. 3D scanners, motion capture systems), which require customers to undergo a lengthy and uncomfortable process. These innovative solutions use a combination of advanced technologies, such as computer vision, deep learning, and 3D matching, to analyse the photos, detect key points, generate accurate body measurements, and provide a 3D model of the human body in a matter of minutes based on the detected key points.

Using computer vision technology for the analysis of images capturing and processing human bodies is facilitated. Advanced computer vision algorithms enable the recognition of a human body in a snapshot, regardless of the background or the type of smartphone used to capture it. Deep learning is used to find neural networks, identify key points, and build probability maps for each key point. And Finally, 3D Matching, which enables apps to precisely capture human body measurements by creating a unique 3D model of each scanned customer based on the main points identified during the scanning process.

To calculate human body measurements using current technologies, there are four key steps: scanning, detecting key points, parameterizing, and finally, processing body measurements.

1. Scanning: The first step is where the customer snaps a few photos from front and side or record video (360 views) of their body.
2. Detecting key points: the second step involves determining exactly where the person is located in the photo. There are different methods to achieve this, either by creating a segmentation mask that can point to a person's presence, pixel by pixel. Or another common solution for this is to use neural networks for the determination

of the location of a person on screen. Most sophisticated models allow getting segmentation masks for separate body parts. Neural networks require a lot of computation, therefore, it is the main downside of such an approach. However, many modern smartphones have embedded neural network acceleration and even embedded solutions for a popular machine learning task, which allowed the current software to implement all the necessary computations on the device. Due to privacy and policies concerns, many applications opt not to utilise cloud-based systems for image analysis, as it goes against the preferences of users.

3. Parameterizing: The third step, based on detecting key points from pictures/videos, develop a 3D model.

4. Processing body measurements: this is the final step where the software calculates human body measurements.

2.6.1 Benefits of Smartphone Scanning

Convenience, portability, and lower costs are just some of the key benefits that may be delivered by smartphone body scanning systems to online shopping platforms and their respective users. This technology is being widely adopted by the apparel industry, which has encouraged the development of unique applications for measuring the human body.

These technologies offer a contact-free alternative to the traditional fitting procedure via an unique cross-platform widget that gathers body measurements from a few photographs or video recordings of the entire human body (360-degree view). A front and side shot of the customer in tight or underwear clothing, typically taken against a clean background, can be utilised to build an accurate 3D body model and several points of measure that are analysed in real time. Smartphone body scanning technology has produced a more effective fitting procedure that cuts time, productivity, and revenue costs for uniform suppliers and their clients by digitising the measurement process for the online shopping apparel industry.

Employees of businesses do not all collaborate with one another, nor do they act in unison. The most convenient solution for one employee could not be the most convenient for another. Similarly, a tailor who has been despatched to measure employees may not have the full day to wait for the conclusion of that important meeting. Due to smartphone scanning technology, staff fitting can be simplified to a simple one-minute operation, hence accelerating activities throughout the uniform production supply chain. Employees are free to take their own actions wherever and whenever it is most convenient for them, without assistance or support from anybody else.

With the help of these solutions, businesses are able to better optimise their planning and production processes, which in turn reduces the production of products that are not suitable. Additionally, businesses can ensure that their products meet the exact size, shape, and fit requirements of their customers, thereby lowering the risk of product returns. This is because such a system is able to rapidly process human body measurement data and present it to a web-based dashboard in real time. This information is segmented and analysed by the dashboard to assist you in applying size and fit intelligence to the specific pattern and grading system utilised by each individual consumer. By mapping human body data to product data, the solution's algorithms may also provide size and fit suggestions to businesses that sell their uniforms online. This is accomplished by mapping the data on people bodies to the data on the products. The forecasting, inventory management, and distribution processes can all be improved with the help of this intelligence system.

2.6.2 Disadvantages of Smartphone Scanning

There are a number of concerns that could be categorised as the most significant drawbacks of smartphone scanning solutions (AI-powered technologies). When the human body is captured by a single camera instead of using stereo, one of the most difficult aspects of this technology is that the 3D information will not be directly available. This is particularly due to the depth interference that occurs when a single image is used, which means that the systems are unable to accurately predict how far away the human body is from the camera. Additionally, the technology is unable to collect 3D information within a single image. However, this method makes the process of data acquisition significantly easier, and it also makes the process of measuring the human body go much more quickly.

Although the actual scanning may only take a few seconds for the majority of AI-powered technologies, the processing time can take anywhere from two minutes to five minutes per individual before providing results for body measurements. If there is a problem with the images that are being input, this process may take even longer, or users may need to retake the photos until the software is able to calculate their measurements. This issue can be very frustrating for many users.

Cluttered background, is another main challenge of such a system which can significantly reduce the accuracy of human body measurements. Most of the existing software, requests users to snap photos from themselves where there is a clean background as this can affect the final results. This is because if the image is not well diffused, part of the background can be detected as part of the body sections, therefore, it can reduce the accuracy of the measurements. More advanced computer vision solutions can greatly reduce the average differences in terms of accuracy, which yet is not achieved by any of the existing applications.

There is a significant impact on the accuracy of the human body measurements if there is any change in lighting and camera viewpoints as this could cause effects like shadows and variations in image appearance.

Consumers may not be willing to do things such as pose for a few photos in front of a camera while wearing specific clothing, remaining motionless for a few minutes, not moving, and adopting a clear point of view; they simply do not want to. Certain types of clothing, such as tight clothing or underwear, may make consumers feel uncomfortable in front of the camera. The impact of breathing can cause expansion and contraction of the body. The wrong pose can observe part of the body that is not able to be captured by the camera especially the part of the body observed under the shoulder griddle. All of these factors can influence the final body measurement estimations.

Capturing the part of the body that is covered by hair is another of the system's main challenge. Since the software is unable to recognise the part of the body that is obscured by hair, it cannot calculate the human body's dimensions accurately.

Typically, these solutions also provide a customer-centric solution that classifies each person into one of categories: small, medium, large and etc. You will be able to consistently exceed your customers' expectations if you collect a wider range of precise measurements and shape data. This will allow you to provide products that fit the unique body shapes of your customers and satisfy their style preferences. This cannot be accomplished with any of the current off the shelf available solutions.

2.6.3 Comparison of existing Smartphone scanning companies

Different types of applications with different approaches have been created to measure the human body for the purpose of garment fit. In this section, we plan to review some of them. These technologies have been chosen on the basis of the accuracy of the data that was provided by them. Chapter 4 investigates each of these methods and approaches in more detail, and some of them have been tested, and the results have been provided.

In the last couple of years, a few of of applications that calculate human body measurements for the fashion industry were developed using AI-powered technologies (computer vision and machine learning). In the same amount of time, they produce results that are nearly similar to those produced by 3D scanners; however, the price and cost to customers is more reasonable than those produced by 3D scanners. Although these applications contain errors comparable to those of the 3D scanner, consumers are more inclined to utilise them. This is due to the lower cost or availability of these applications, as well as their greater accessibility. In Table 2.4, we will compare several of these new applications based on the technology they employ, the scanning process, the capturing time as well as the processing of the measurements, and lastly, the level of accuracy that the company claims to achieve based on their datasets. Please refer to subsection 4.2.1 for additional information regarding the functionality of these applications.

Table 2.4 Comparison of existing smartphone scanning solutions in the fashion industry. Please note that the accuracy mentioned here is as stated by the companies on their respective websites.

Company Name	Technology	Scanning Process	3D Model	Capture time	Accuracy
3DLOOKME	AI-Powered Technology	2 Images	Yes	2 min	$\pm 0.5cm$
ESENCA	AI-Powered Technology	2 Images	No	2 min	$\pm 1cm$
Zalando	AI-Powered Technology	2 Images	N/A	2 min	$\pm 1cm$
Bodygram	AI-Powered Technology	2 Images	Yes	2 min	$\pm 1cm$
IBV	AI-Powered Technology	2 Images	Yes	2 min	$\pm 2cm$
MeThreeSixty	AI-Powered Technology	2 Images	Yes	2 min	$\pm 2cm$
Mirrorsize	AI-Powered Technology	2 Images	No	3 min	$\pm 2cm$
VYOO	AI-Powered Technology	2 Images	No	2 min	$\pm 0.5cm$
SizeStream	AI-Powered Technology + 3D Scanner (Booth)	Multiple Images	Yes	2 min	$\pm 5mm$
SWAN	AI-Powered Technology	Video	No	1 min	$\pm 1cm$
Presize.ai (Find-Size)	AI-Powered Technology	Video	Yes	2-5 min	$\pm 2cm$
Prismlabs	AI-Powered Technology	Video	Yes	2 min	$\pm 2cm$

2.7 Evolution of Object Detectors & Related Works

Well known Computer Vision approaches for object detection include Histogram of Oriented Gradient (HOG) is one of the most popular and classic methods used in object detection. This method works by extracting the gradients from the image and creating a histogram which is then used to detect the objects. The main advantage of this method is that it is fast and has good accuracy. However, the downside is that it requires a lot of manual feature engineering and is not very robust. Haar Cascade is another popular method of object detection. This method is based on the Viola-Jones algorithm and works by using a series of “Haar-like” features which are applied to an image in order to detect objects. The main advantage of this method is that it is very fast and efficient. However, the downside is that it is not very accurate and is not very robust.

You Only Look Once (YOLO) is a newer method for object detection which is based on convolutional neural networks. It works by using a single network to detect multiple objects in an image. The main advantage of this method is that it is very accurate and is very robust. The downside is that it is computationally expensive and requires a lot of data for training. MobileNet-SSD is a deep learning-based object detector which is designed specifically for mobile devices. It works by using a MobileNet-based convolutional neural network to detect objects in an image. The main advantage of this method is that it is fast and efficient and has good accuracy. The downside is that it requires a lot of data for training and is not very robust. EfficientNet is an innovative deep learning-based technique for object detection that is also remarkable. Utilising a compound scaling technique that optimises model architecture and computational resources, EfficientNet is known for its efficiency and high accuracy. The benefits of EfficientNet consist of its superior performance in terms of precision and efficacy. Similar to other deep learning models, it requires a large quantity of training data and can be computationally intensive. OpenPose is another famous object detection technique. It is mainly dealing with human pose estimation, accurately detecting and monitoring human body keypoints in images and videos. OpenPose's capacity to provide detailed and accurate information about human poses enables applications such as gesture recognition and motion analysis. OpenPose's computational complexity could affect its real-time performance on devices with limited resources.Figure 2.5 shows the progress of significant object detection techniques throughout time [60].

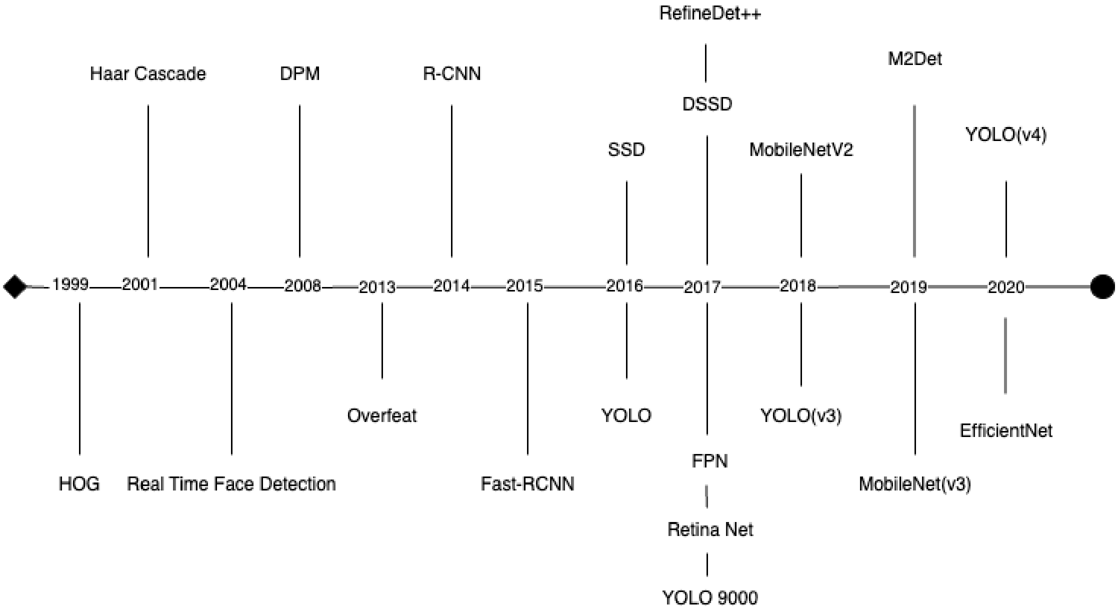


Fig. 2.5 Year wise evolution of object detection algorithms[2].

Not only can deep learning do well in object detection, but also in other fields. Deep learning offers a variety of models to effectively manage fashion industry data. From the beginning of the Covid-19 pandemic, numerous research projects have been conducted to limit human-to-human contact. This is so designers may take customer measurements from anywhere, at any time. Some of these machine learning approaches will be examined in greater detail in section 5.2.

[2] Deformable Parts Model (DPM), Region-based Convolutional Neural Network (R-CNN), Deconvolutional Single Shot Detector (DSSD), Feature Pyramid Network (FPN), Multi-level Feature Pyramid Network Detection (M2Det), Single Shot MultiBox Detector (SSD) and You Only Look Once (YOLO)

2.8 Related Work in Human Body Detection and Pose Estimation

There are many different kinds of technologies and methods that have been developed for the purpose of estimating human pose. In the following paragraphs, we will look into several of these topics. The accuracy of the information that these technologies delivered was a primary consideration in their selection. In Chapter 4, as the first stage of our experiment, we investigated and tested a number of these methods and approaches in greater detail. Later in this section, we will focus primarily on machine learning approaches for estimating the human body pose.

2.8.1 Human Body Performance Capture

Using a simpler model, according to Wu et al. [175], can improve the quality of the human body performance capture. The use of kinematic skeletons, human parametric models, the facilitation of single-view reconstruction, and even the segmentation of human subjects in motion have been proposed as additional improvements. Currently, multi-view depth-based methods are being intensively researched in order to achieve even higher levels of precision and greater robustness. Wang et al. [170] utilised RGB inputs and sparse depth sensors for the purpose of measuring the textured surfaces of moving bodies. Orts-Escolano et al. [120] should retain the services of an active studio equipped with high-quality cameras and specialised acquisition devices in order to perform real-time, high-quality motion capture.

Further investigation concentrated primarily on photometric stereo and the folding patterns of clothing, deriving shape from shading and attempting to capture dynamic details. In many of these approaches, Pons-Moll et al. [130] came up with the idea of a multi-cloth 3D model in order to reconstruct both the clothes and the body from 4D scan sequences; to evaluate an unclothed body shape by using and tracking the clothing over time; and to lessen the computational expense for the opposite rendering problem. In addition, other methods required the actor to dress in normal clothes; however, in order to create a template skeleton and mesh, it is necessary for each actor to be captured in advance with this idea. An additional research approach that has not been investigated before Huang et al. [78]'s work is to find a method that can create the mesh in a fully automatic manner. This can be accomplished without making use of any ready-made template models.

2.8.2 Multi-View 3D Deep Learning

In order to acquire deep learning for a wide variety of tasks, such as shape segmentation, correspondence matching, object classification and identification, and unique view synthesis as mentioned by Kalogerakis et al. [90], the Multi-View Convolutional Neural Network (CNN) has been introduced. According to Arsalan Soltani et al. [28], a number of earlier publications use multi-view CNNs to 3D reconstruction problems. These challenges can be approached directly using supervised or unsupervised methods to generate the final geometry, or indirectly using normal maps, silhouettes, or colour images. Others have, according to Huang et al. [79], adjusted feature projection and ray compatibility in a way that is easily recognisable by making use of this formula within a finite network to predict volumetric representations of a three-dimensional object. This formula was inspired by multi-view stereo constraints.

Moreover, Hartmann et al. [74] came up with the idea of a deep learning-based approach to anticipate the likeness between image patches across multiple views, which enables 3D reconstruction using stereo. In this research study, we mainly focus on a novel idea, which is the different and more challenging task of finding the pre-point possibility of lying on the reconstructed surface and directly connecting the 3D volume base to 2D projections on the image plane. A research study closer to our work is a cross-model network that measures parametric body surface shape from a single silhouette image [59]. There are many disadvantages to this method, such as that it cannot predict the depth of the human from only one single image and it requires a naked body shaped in neutral poses to predict the results, while our plan is to generalise dynamic clothed bodies in extreme poses. However, recently there is a new technology called Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitisation [145] that can predict extremely detailed 3D model from a single image, which the following paragraph will describe in more details about this project.

2.8.2.1 Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitisation

Recovering extremely detailed 3D model from a single 2D image is one of the most challenging task in automated human digitisation with computer vision which has a wide array of applications to the digital media industry.

Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization (PIFuHD) [145] present 3D human reconstruction framework, that recovers faithful details of the person in 3D at 1K resolution from a single image fully automatically. Many human reconstruction model-based rely on the parametric regression model, however, the prediction is typically for the almost naked body (tight clothes or underwear) and all the personal details such as hairstyles and clothing are ignored. On the other hand, model-free regression approaches have demonstrated higher flexibility and expressiveness, as a result, it can recover details present in the input image more accurately. However, the previous methods rely on explicit shape representation such as box-cells which is limiting the resolution due to high memory footprint.

To address resolution limitation, the author proposed Pixel-Aligned Implicit Function (PIFu) [143] a year before this project was released which I will briefly explain how it works and point out that it is a limitation for a more precise human reconstruction which is the main focus of this project. PIFu used neural impresses function for shape representation

3D shapes are represented by the level set or by occupancy fields[3] parameterize by a neural network since this formulation does not involve any discretization which can be effectively model 3D shapes at an arbitrary resolution.

$$f(X,I) = \begin{cases} 1 & \text{if X inside mesh surface} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Implicit Shape Representation

$$f(X,I) = 0 \in \mathbf{R} \tag{2.2}$$

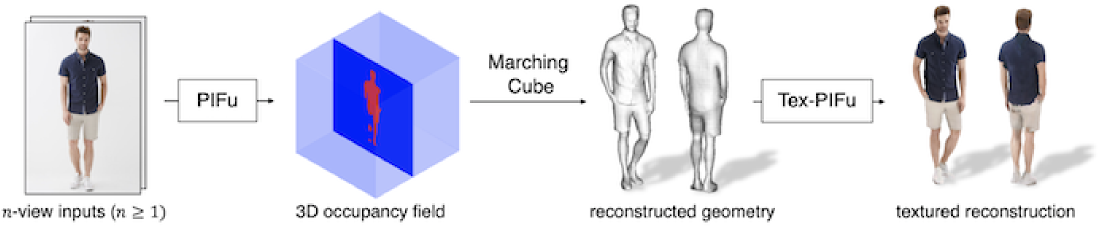


Fig. 2.6 Overview of clothed human digitisation pipeline [143].

All in works including occupancy network and code as input image as a global descriptor to reconstruct corresponding 3D shapes. However, this way the network does not leverage the special relationship between query 3D points and the input image. To address this issue the author introduced PIFu where they utilise free conventional image features to associate queries 3D point and the projected pixel coordinate given the pixel align in the code and query to depth value Z, they can inverse 3D occupancy fields in a pixel aligned manner. However, they found that the reconstruction quality is still bounded by the feature resolution of an image encoder. the original PIFu implementation utilises an image encoder that has done sampling to keep the spatial resolution small for the following reasons, first small feature resolution allows for holistic reasoning as 3D

[3] Introduced a deep learning approach based on a volumetric occupancy field that can capture dynamic clothed human performances using sparse viewpoints as input [78], [70]

reconstruction requires the absolute location of where the surface is, holistic reasoning is indispensable. Moreover, the small resolution allows the author to train the model with a reasonable amount of time the downside is limiting the expressiveness up to the small resolution resulting in fewer details reconstruction than the original input image. What about utilising a shallow but high-resolution image encoder although the result becomes sharper and more detailed the reconstruction is a pawn to add facts due to the lack of global information. In this project, the following are addressed as disadvantages of both architectures by a coarse to fine multi-level approach.

Several human reconstruction methods utilise a coarse-to-fine strategy. However, Saito et al. [143] argues that since existing methods have multiple limitations, the existing coarse model and detailed inference from the surface normal but the base geometry tend to be low-fidelity and detail hallucination[4] on top of inaccurate base geometry does not provide a basis for high-resolution geometry. Thus, they have introduced a latent multi-level representation. First, they have trained the coarse PIFu as in the original PIFu implementation and reviewed intermediate latent features of the Multi-Layer Perceptron (MLP) as a 3D embedding, and the fine image encoder takes a 3D embedding from the coarse PIFu instead of querying the depth value. This way, the final module can easily perform at least as well as coarse modules as it provides all 3D information used to make the coarse prediction.

Moreover, in their pursuit of capturing fully focused geometric details, the coarse module surpasses the capabilities of existing coarse-to-fine approaches by directly predicting high-resolution 3D geometry without the need for additional post-processing steps. In essence, their latent multi-level representation achieves exceptionally detailed 3D reconstruction while being memory-efficient, as demonstrated in their work [145]. For a more in-depth exploration of their experiments, please refer to subsection 4.2.5. The results, as shown in Figure 2.7, reveal their ability to recover intricate geometry

[4] "Detail hallucination" is when reconstruction methods add false, high-resolution details to an inaccurate base model, leading to inaccuracies.

from self-captured video frames, not just on the front side but also on the backside. It's worth noting that these are the raw outputs from PIFuHD without any post-processing. Additionally, they have enhanced the fidelity of the backside geometry by conditioning the network with the inference surface normal for the backsides [144].

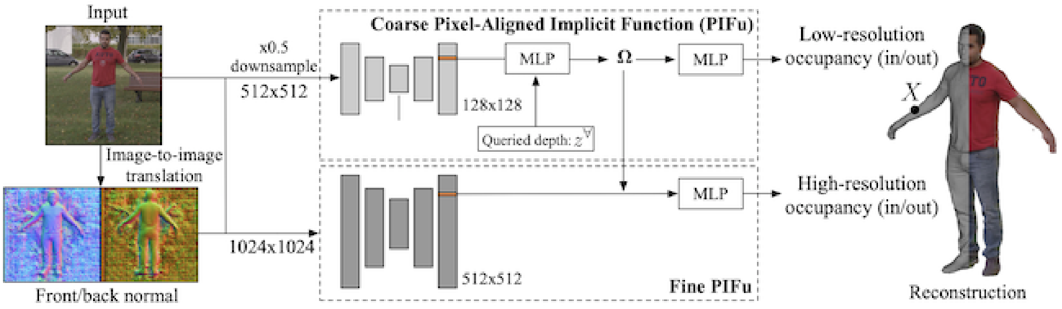


Fig. 2.7 Overview of PIFuHD framework. Two levels of pixel-aligned predictors produce high-resolution 3D reconstructions. The coarse level (top) captures global 3D structure, while high-resolution detail is added by the fine level [145].

2.8.3 Human Body Data Capture Software

Different approaches have been considered to detect human body measurements and for estimation. In this section, we will explain briefly what the Human Pose Estimation is, and, in Chapter 4, we will investigate and test the methods of human pose estimations to detect human body measurements.

2.8.3.1 Human Pose Estimation

For computer vision, human pose estimation is defined as the problems of localisation of human joints in images or video. In practice, calculating the pose of one person is way easier than calculating the pose of numerous bodies in a video. Despite many years of research, however, human pose estimation remains a very difficult and still largely unsolved problem. To approach this problem, recent attempts are to use either a part-based or two-step framework. The part-based framework discovers human body parts independently first and then connects and assembles the discovered and detected body segments to form multiple human poses, while the two-step framework first draws a bounding box around the human body and then calculates the pose within each box independently. To date, none of those approaches can produce satisfactory results in general; for instance, when two or more people are too close together, the assembled human poses are ambiguous. Moreover, a part-based framework loses the capability to recognise body parts from a global pose view, which is due to the mere utilisation of second-order human body parts dependence, while, in the two-step framework, to capture accurate body pose estimation, the quality of the discovered bounding boxes is very important [64, 177, 192].

Although state-of-the-art human detectors have demonstrated good performance, some errors in localisation and recognition are inevitable in multi-person pose estimation. Therefore, these errors can create failures for a SPPE, especially for methods that depend on human detection results. Regional Multi-person Pose Estimation (RMPE) is a new top-down framework to facilitate pose estimation in the presence of inaccurate human bounding boxes. This framework consists of three main components: SSTN, Pose-Guided Proposals Generator (PGPG), and Non-Maximum Suppression (NMS). Figure 2.8 demonstrates a pipeline of an RMPE framework with three major components. This method increases up to 17 % in mAP over the state-of-the-art method on the MPII (multi-person) dataset. This method often relies on the accuracy of the person detector in order to generate correct results, due to the fact that posture estimation is performed on the region where the person is located. Inaccuracies in the duplicate bounding box and localisation prediction can therefore result in suboptimal performance for the pose extraction technique. This method proposes SSTN as a solution to this challenge in order to extract a high-quality single-person zone from an inaccurate bounding box. This individual's human postural skeleton has been estimated via SPPE. SDTN is utilised to remap the estimated human pose to the image coordinate system in which it was initially saved. As a solution to the problem of redundant pose detection, NMS has been implemented. Therefore, we have investigated this method in more detail, and some of the methods have been used as part of Chapter 4 [134].

Part of the algorithms of this method have been used in subsection 4.2.3. More information about the algorithms can be found in subsection 4.2.2.

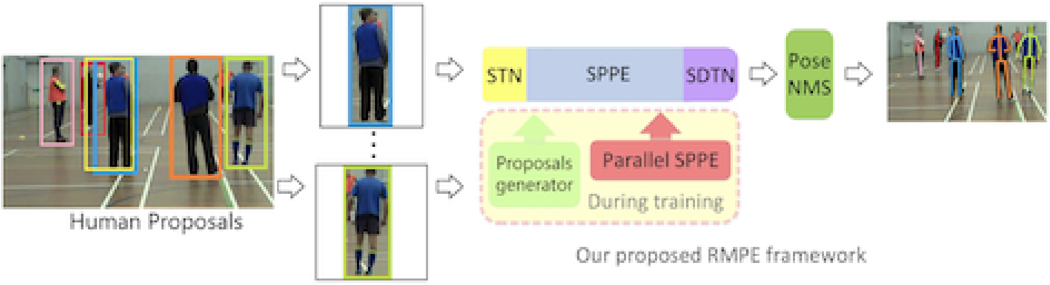


Fig. 2.8 RMPE framework pipeline demonstrate three main factors which are; **1 SSTN and Parallel SPPE 2 Parametric Pose NMS 3 Pose-guided Proposals Generator**. Symmetric STN contains of SDTN and STN which are connected after and before the SPPE. The human proposal will receive by STN and the pose proposal will generate by SDTN. During the exercising phase, the aligned SPPE acts as an extra regularizer. In the end, to remove unnecessary pose estimation, the parametric Pose NMS (p-Pose NMS) is carried out. The SPPE + SSTN module with pictures generated by PGPG, unlike traditional way [64].

2.8.4 Volumetric Representation

Volumetric representation is a method that shows a captured body scan using voxels[5] that span the bounds of the scan's volume. A body scan in this representation, therefore, consists of a voxel inside, outside and on the surface of the scan. A correspondence is established in this case by using the same number of voxels for all body scans in a sample. Therefore, there would be a fixed number of voxels along with height, width and depth of a scan. Each voxel for one scan would then correspond to a voxel in the same

[5] Individual elements of three-dimensional space are called volume elements or voxels

grid location in the representation of another scan. Voxel representation (also referred to as volumetric representation below) can then be directly compared and used to build statistical models of human body shape. Examples of the method being used to model human shape are described below.

Huang et al. [79] present a deep learning-based volumetric approach, as an example of volumetric representations being used for performance capture using a passive sparse multi-view capture system. Their method aims to estimate a dense 3D field that encodes the probabilistic distribution of the reconstructed surface by giving multiple views and their corresponding camera calibration parameter as input. They formulated the probability prediction as a classification problem; therefore, at the high level, their method resembles the spirit of the shape-from-silhouette method: to obtain 3D points staying inside the reconstructed object, reconstructing the surface according to the consensus from the sequence of multi-view images. However, instead of directly using silhouettes, which only contain a limited amount of data, they leverage the deep features learnt from a multi-view convolution neural network. Thus, the method projects the query point onto the multi-view image planes using the input camera parameters for each query point in the 3D space. Figure 2.9 demonstrates the network architecture.

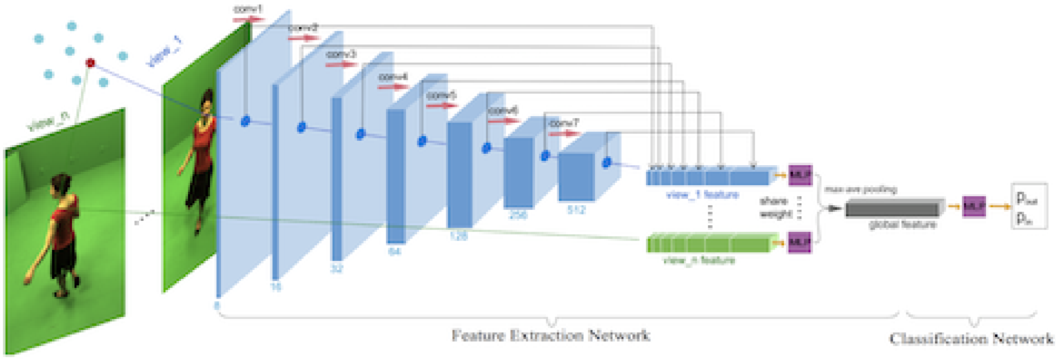


Fig. 2.9 Network architecture by Huang et al. [79]

Afterwards, for the query point to get the final global features, it starts to obtain the multi-scale convolutional neural network features learnt at each projected location and aggregate them through a pooling layer. The feature of each point is later fed to a classification network to most likely result in lying inside and outside the reconstructed object.

2.9 Technologies related to our proposed Software

In this section, we present the research efforts that focus on estimating human body size by using similar techniques to our project.

2.9.1 Accurate 3D Body Shape Regression using Metric and Semantic Attributes

The SHAPY framework, introduced by Choutas et al. [55], revolutionises the field of 3D body shape estimation by providing a novel approach that achieves accurate results from images without relying on explicit 3D shape supervision. By leveraging linguistic shape attributes and anthropometric measurements as proxy annotations, SHAPY enables the training of a regressor that can predict body shape with enhanced precision.

To evaluate the effectiveness of SHAPY, the researchers utilised various datasets, including the "Human Bodies in the Wild" (HBW) dataset. This dataset offers a unique set of challenges as it contains images of individuals in natural settings, showcasing diverse body shapes and clothing variations. The HBW dataset also provides ground-truth 3D shape data obtained from body scans, allowing for a comprehensive evaluation of the framework's performance.

The accuracy of SHAPY was evaluated using multiple metrics tailored for body shape estimation. The framework demonstrated remarkable results, surpassing existing methods in terms of accuracy and precision on the HBW dataset. It showcased superior performance in estimating critical body measurements such as height, chest circumference, waist circumference, and hip circumference, with mean absolute errors ranging from 5.8 to 6.2 mm across different gender and attribute combinations. The accuracy of anthropometric measurements, such as height and weight, was also evaluated, with mean absolute errors computed between the ground-truth and estimated measurements.

The SHAPY framework offers several advantages over traditional approaches. Its ability to handle diverse body shapes and clothing variations makes it suitable for real-world applications. By reducing the cost and time associated with obtaining anthropometric measurements, SHAPY has the potential to revolutionise industries such as virtual reality, computer animation, and the clothing industry. Furthermore, the framework's utilisation of linguistic attributes and anthropometric measurements as proxy annotations enhances accuracy and enables the generation of highly realistic digital human models.

However, there are limitations to consider. The model-agency training dataset used in SHAPY may not represent the entire human population comprehensively, which could impact the accuracy of predicting larger body shapes. Additionally, the computational complexity involved in estimating anthropometric measurements from dense and accurate 3D data might pose challenges for practical deployment.

In conclusion, the SHAPY framework introduces a significant advancement in the field of 3D body shape estimation. Its utilisation of linguistic attributes and anthropometric measurements as proxy annotations results in improved accuracy and precision, as demonstrated by its superior performance on challenging datasets like HBW. While limitations exist, including the need for more diverse training data and addressing computational complexities, the potential applications and promising results make SHAPY a compelling tool for industries seeking accurate 3D body shape estimation.

2.9.2 Automatic human body feature extraction and personal size measurement

The paper by Xiaohui et al. [176], presents an automatic approach for extracting feature points[6] and measuring garments on 3D human bodies. The authors claim that the average error of feature point extractions is 0.0617 cm, and the average errors of shoulder width and girth are 1.332 cm and 0.7635 cm, respectively.

First, the proposed method demonstrates high accuracy in extracting feature points on 3D human bodies. The precise localisation of feature points is achieved, with an average error of 0.0617 cm, indicating accurate and reliable results. Second, the approach enables automatic measurement of various body dimensions, including shoulder width, bust, hips, and waist girth. These measurements are crucial for garment fitting and customisation, providing valuable information for personalised clothing design. Third, the method is designed to be invariant to isometric deformations, ensuring robustness in capturing variations in body shape and posture. This feature allows the approach to accommodate diverse body types and poses, making it suitable for a wide range of applications. Fourth, the scalability of the method is notable, as it can handle a large number of body shapes and sizes. This scalability makes it applicable for mass customisation and virtual try-on applications, enhancing the clothing shopping experience in various settings such as shopping malls and home environments.

However, there are certain limitations to consider. Firstly, the proposed approach relies on a depth-sensing platform, such as a Kinect sensor, which may limit its accessibility as such devices are not readily available for most online shoppers. Secondly, multiple poses need to be captured for training purposes, which adds complexity and time requirements to the data acquisition process. While the exact number of poses used in the research is not specified in the paper, it is generally recognised that capturing

[6] refer to specific anatomical or characteristic points on a 3D human body model that are essential for measuring various aspects of the body size.

multiple poses improves the accuracy of the method. Moreover, the approach involves randomly selecting markers around the human body to compute feature points. While this randomness provides flexibility, it may introduce some data errors and inconsistencies due to the distribution of markers not always accurately representing the desired features.

In terms of accuracy, the authors report that the average errors of shoulder width and girth are 1.332 cm and 0.7635 cm, respectively. These results indicate a reasonable level of accuracy in estimating body measurements. However, specific accuracy levels for other feature points or dimensions are not provided in the summary.

In conclusion, the proposed method offers an automatic solution for feature extraction and size measurement on 3D human bodies. It demonstrates high accuracy in locating feature points and provides reliable body measurements. However, the dependency on a depth-sensing platform, the requirement for multiple pose capturing, and the random selection of markers are important considerations. Future work may focus on addressing these limitations and improving the computational efficiency of the method.

2.9.3 A dynamic fitting room based on Microsoft kinect and augmented reality technologies

Chang et al. [47] proposed a dynamic fitting room that combines Microsoft Kinect and augmented reality technologies to allow individuals to visualise themselves in real-time while trying on different virtual garments. The system utilises two Kinect cameras, one for capturing the front view and the other for the side view, to estimate the user's body height based on head/foot joints and depth data. The system achieves sufficiently accurate results with an error of less than a centimeter in estimating body height.

The proposed system offers several advantages. Firstly, it provides a more immersive and interactive fitting experience by enabling users to see themselves wearing virtual clothes in real-time. This enhances the shopping experience for customers in clothing stores, e-commerce platforms, and personal use, reducing the need for physical garment try-ons and saving time. Secondly, the system leverages the capabilities of Microsoft Kinect and augmented reality to accurately track users' body movements and project virtual garments onto their live video feed, creating a realistic representation of how the selected clothes would look on their bodies.

The automated size estimation feature is another advantage of the system. By analysing the joint data and depth information captured by the Kinect cameras, the system calculates the user's body size and provides size suggestions. This eliminates the manual size selection process and reduces the chances of ordering ill-fitting clothes.

However, the system has certain limitations. One drawback is the requirement for users to garment themselves with Kinect, which may be inconvenient or uncomfortable for some individuals, potentially hindering the system's adoption. Additionally, the system's size classification is limited to coarse ranges commonly used in retail, such as "small, medium, large, XL, XXL." This may not provide the level of accuracy needed for precise fitting, especially in cases where more nuanced size options are required. The system's size classification limitations are primarily attributed to algorithmic constraints rather than Kinect's capabilities. The algorithm's reliance on coarse size ranges commonly found in retail restricts its ability to offer more precise sizing options, potentially impacting its suitability for applications requiring finer size distinctions.

In terms of accuracy, the system achieves sufficiently accurate results in estimating body height, with an error of less than a centimeter. However, it is important to note that the accuracy level may not be suitable for all practical use cases. While the system's size evaluation closely aligns with users' claimed sizes in the conducted experiment, the coarse size ranges used may not meet the requirements of applications that demand more precise sizing information.

In conclusion, dynamic fitting room system offers advantages such as real-time visualisation, accurate tracking, and automated size estimation. However, it also faces limitations related to the requirement for Kinect garmenting and the coarse size classification used. The reported accuracy, although sufficient for certain use cases, may not meet the demands of all practical applications.

2.9.4 Single camera body tracking for virtual fitting room application

Chandra et al. [46] proposed a system which utilises portable cameras commonly found in smartphones to capture images of the user's body and tracks the body outlines to estimate relevant measurements. The system begins by detecting the presence of a human body within its field of view. Once a body is detected, the system tracks the outlines of the body to determine its shape and proportions. The authors specifically focus on identifying the face, neck, and shoulders as key markers for the upper body measurements.

The accuracy of the system in estimating body measurements is not explicitly mentioned in the summary. However, it can be inferred that the system aims to provide reasonably accurate estimations based on the tracking of body outlines and key markers. The article likely includes detailed discussions on the accuracy of the system, which would provide more insights.

While the proposed approach offers the convenience of using a single camera for body tracking, it has certain limitations. The system requires specific environmental conditions to be set, including lighting, background, and the type and contrast of clothes worn by the user. These conditions must be controlled to ensure accurate tracking and measurement estimations. Moreover, users need to repeat the measurements to improve the accuracy, suggesting that the initial estimations might not be completely reliable.

Expanding on the summary, it would be beneficial to provide more details about the working mechanism of the system. For example, it would be interesting to explore the algorithms or techniques used for body detection, outline tracking, and marker identification. This information would give readers a better understanding of the technical aspects of the proposed approach.

Furthermore, it would be useful to discuss the potential advantages and disadvantages of using a single camera for body tracking in virtual fitting rooms. Some advantages might include cost-effectiveness and accessibility since most individuals already possess smartphones with built-in cameras. On the other hand, potential disadvantages could involve limitations in tracking accuracy compared to more advanced tracking systems that utilise multiple cameras or depth sensors.

To provide a comprehensive summary, it would be beneficial to include information about the accuracy of the system in estimating body measurements. This could be achieved by referencing specific accuracy metrics or results mentioned in the article.

In conclusion, the article presents an approach for body tracking using a single camera in the context of virtual fitting rooms. The system detects the presence of a human body, tracks body outlines, and identifies key markers for estimating body measurements. While the proposed approach offers convenience, it requires specific environmental conditions to be set and might require repeated measurements for improved accuracy. Further details regarding the system's working mechanism, pros and cons, and accuracy metrics would enhance the understanding of the proposed approach.

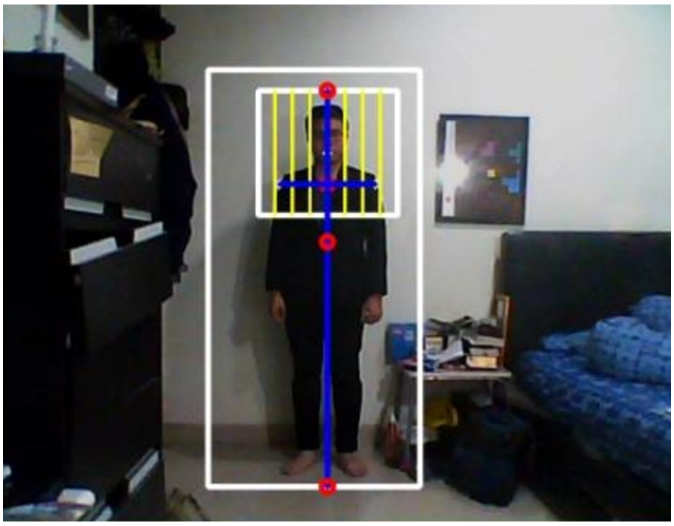


Fig. 2.10 displays the body tracking results from [46] software.

2.9.5 Fitme: Body measurement estimations using machine learning method

Ashmawi et al. [31] proposed an approach to estimate human body measurements through the use of smartphone cameras. The system utilises Haar-based detectors to detect various body areas, including the upper, lower, and full body.

The working principle of the system involves several steps. First, the human body is detected in the images captured by the smartphone camera. Next, features of the body are extracted from the image, followed by the determination of focal points within the body. Finally, body measurements are calculated by computing the difference between these focal points.

However, it should be noted that the system's ability to estimate body measurements is limited to providing coarse ranges, such as small, medium, large, XL, XXL commonly used in retail classification. The accuracy achieved by the system is not sufficient for practical use, as the paper highlights the need for further improvements in this regard.

In addition to using Haar-based detectors to estimate human body measurements through smartphone cameras, the system also incorporates Support Vector Machine (SVM) classifiers for predicting clothing sizes. SVM is a powerful supervised machine learning algorithm that excels in classification tasks. Multiple SVM models are employed to predict standard sizes (e.g., XS, S, M) for different clothing categories (upper, lower, and full pieces), employing a one-versus-all multiclass classification approach. These predictions determine the recommended clothing size for the shopper.

One advantage of the system is its utilisation of widely accessible smartphone cameras, making it convenient for users to estimate their body measurements without the need for specialised equipment. Additionally, the system's use of Haar-based detectors allows for real-time object detection, enhancing the efficiency of the process. The extraction of focal points and subsequent calculation of body measurements provide a structured approach to estimate sizes based on image analysis.

However, the system also has certain disadvantages. The coarse ranges provided by the system may not meet the precise requirements of individuals, especially when it comes to purchasing clothing that requires accurate sizing. The paper acknowledges that the achieved accuracy falls short of practical usability, indicating the need for further development and refinement of the system.

In terms of accuracy, the paper does not explicitly state the specific accuracy level achieved by the system. However, it highlights that the obtained accuracy is not sufficient for practical use, indicating a relatively low level of accuracy. This suggests that further improvements are necessary to enhance the accuracy and reliability of the system's size estimation capabilities.

In conclusion, this system utilises smartphone cameras and Haar-based detectors to estimate human body measurements. While the system offers convenience and real-time object detection, it currently provides only coarse ranges of sizes and falls short of practical usability due to its accuracy limitations. Future research and development efforts are needed to enhance the accuracy and make the system more suitable for practical applications in online shopping and size estimation.

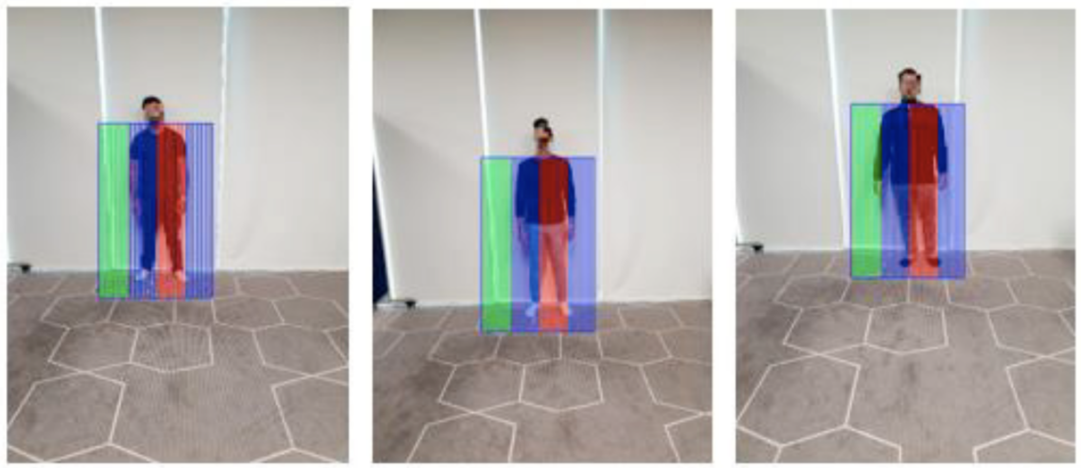


Fig. 2.11 displays features extraction using segmented the detected image into parts by FITME: [31]

2.9.6 Anet: A Deep Neural Network for Automatic 3D Anthropometric Measurement Extraction

This paper presents the development of Anet, a deep neural network designed for automatic extraction of 3D anthropometric measurements from 3D human body scans which proposed by Nourbakhsh Kaashki et al. [118]. Anet consists of two main components: a feature extraction network and an anthropometric measurement extraction network. The feature extraction network captures important features from the 3D scan, including body posture, shape, and proportions. These features are then used by the anthropometric measurement extraction network to automatically extract 3D anthropometric measurements.

The authors conducted comprehensive evaluations of Anet on various datasets and compared its performance to existing methods. The results demonstrated that Anet achieves highly accurate anthropometric measurement extraction, with a mean absolute error of less than one centimeter. This significant improvement over existing methods indicates the potential of Anet for a wide range of anthropometric measurement tasks.

One advantage of Anet is its ability to reduce the cost and time associated with obtaining anthropometric measurements. By automating the measurement extraction process, Anet eliminates the need for manual measurements, which can be time-consuming and labor-intensive. This has implications for industries that require precise body measurements, such as fashion, ergonomics, and healthcare.

Furthermore, Anet can streamline the generation of digital human models for applications like virtual reality and computer animation. With its accurate measurement extraction capability, Anet can create realistic and customisable digital representations of human bodies, enabling more realistic and personalised virtual experiences.

Despite its advantages, there are some limitations to consider. Anet's performance may be affected by the quality of the input scans, particularly when dealing with scans captured by certain devices that introduce high levels of noise. Improving the robustness of Anet to handle noisy input scans could be a focus for future research. Additionally, while Anet demonstrates high accuracy, further validation and testing on larger and more diverse datasets would be beneficial to establish its generalizability and performance across various populations.

In conclusion, this paper presents Anet, a deep neural network capable of automatically extracting accurate 3D anthropometric measurements from 3D human body scans. The evaluation results highlight its superior performance compared to existing methods. The potential benefits of Anet include cost and time reduction in obtaining anthropometric measurements, as well as enabling automated generation of digital human models. Further research could focus on improving robustness to handle noisy scans and expanding validation on diverse populations. With its advancements in anthropometric measurement extraction, Anet opens new possibilities for various industries and applications requiring precise body measurements.

2.10 Related Datasets

To the best of our knowledge, none of the currently used datasets for human body data were collected specifically to explore the human body measurement task. Related work can be found in body modeling, human body measurement, and computer graphics. However, only a few of the many published works provide a dataset with a significant number of subjects. Traditionally, there have been many models to represent the human body. From 1D structures living in 3D space, skeleton based to 2D and 3D models. Below we discuss existing datasets that have been used for human body data and focus on a specific set of attributes that are important for human pose estimation and scanned body reconstruction. Table 2.5 shows the comparison between all the top human body datasets.

CAESAR [180] One of the most powerful resources to date is the CAESAR dataset. This dataset was created to categorise human body models in a variety of poses and shapes. The datasets contain 1,500 registered male and female meshes with point-to-point correspondences, with 25K facets and 12.5K vertices for each mesh.

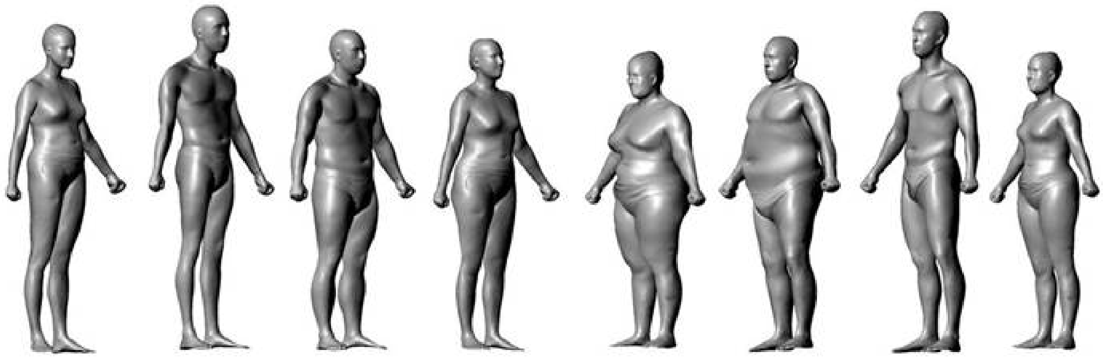


Fig. 2.12 displays the example of CAESAR dataset

CAESAR3D [139] Another extensive 3D database, this one originating from the civilian American and European Surface Anthropometry Resource Project and including measurements from the entire North American population sample of 2400 male and female subjects, aged between 18 and 65 years old, along with demographic information. This is the first database of its kind, and it includes typical 1-D measurements as well as scans of 3D models. There is a wide number of poses to choose from, such as standing, coverage, and relaxed seating positions. In addition, there are 40 conventional 1-D anthropometric measurements in the database. These measurements were taken with a tape measure and a calliper. This dataset is currently the most comprehensive 3D dataset derived from real scans that is available. This dataset is not freely available, and human shape analysis may not capture all of the significant variety that exists in the population of shapes.

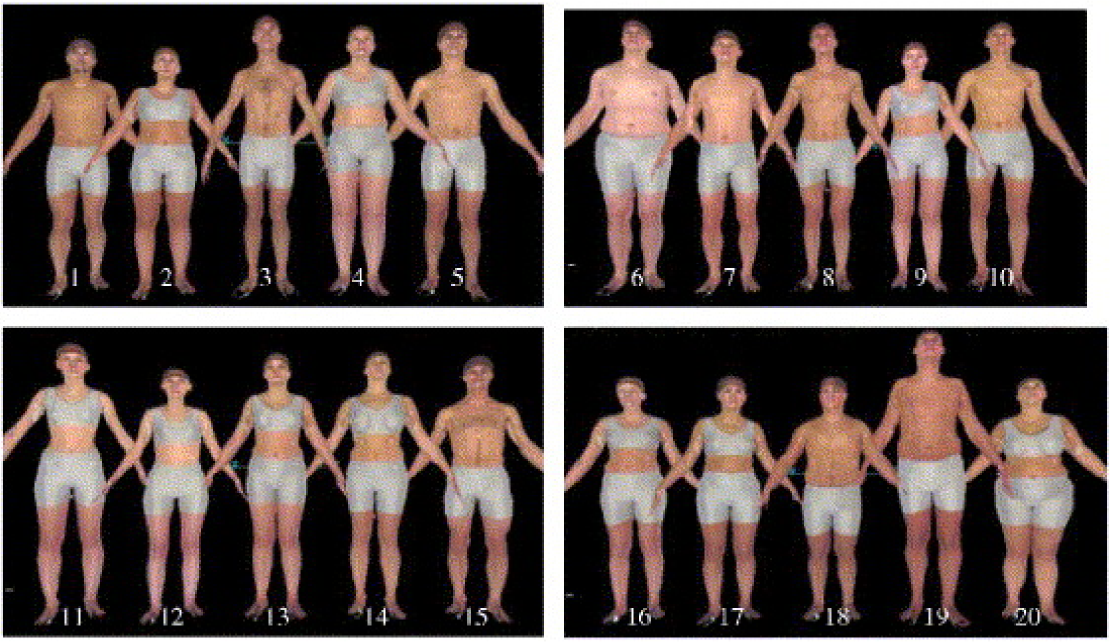


Fig. 2.13 displays the example of CAESAR3D dataset

SCAPE [25] was originally designed for capturing animated motion. Muscle deformations are induced automatically by the space created and the body's deformations. In addition, one of the benefits of this collection is that it has extremely precise data for every single human body pose. The collection consists of 71 registered meshes of a specific individual in various positions. As morphs of the template models, the scape completion of each scan and the correspondences between the template and each scan with the original dataset have been published. Although the number of patients in the dataset is not statistically significant, the mesh description for human body shape analysis is quite precise.

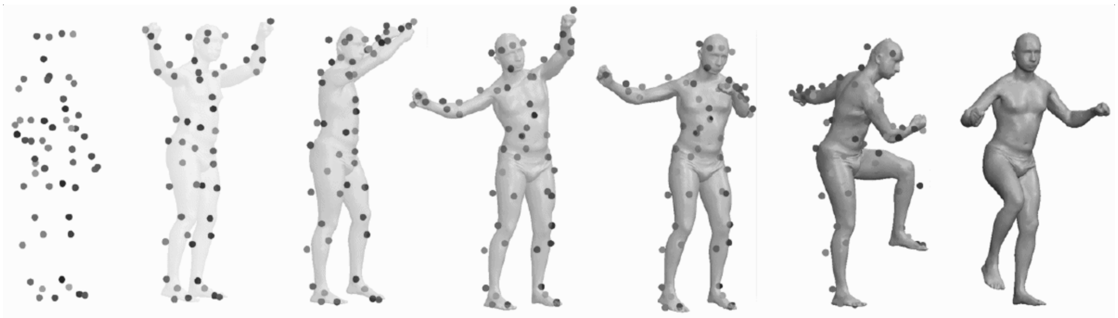


Fig. 2.14 displays the example of SCAPE dataset

FAUST [36] new dataset is comprised of 300 scans of 10 individuals, displaying a wide range of ages and body compositions in a variety of wide-ranging poses. The dataset provides very high-resolution photos in addition to displaying all of the data that was previously unavailable. The surface registration is going to be the primary focus of this effort. Generally speaking, the registration process is difficult for non-rigid and articulated things such as human bodies. The authors handle the registration problem using an innovative approach to mesh registration in order to achieve high-quality alignment. This allows them to incorporate information about the 3D shape and look of the object. Despite the fact that this collection contains information on real persons depicted in a variety of poses, the number and types of subjects included in it are still quite restricted.

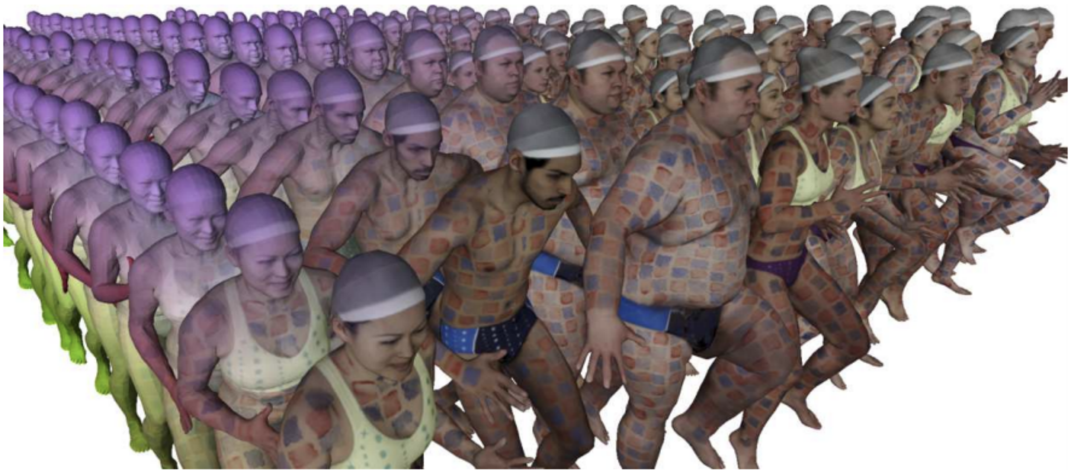


Fig. 2.15 displays the example of FAUST dataset

MPII Human Shape [23] consists of about 25,000 images depicting over 40,000 individuals and their annotated joints. The dataset contains information on 410 distinct actual human actions, and the photographs come with highly-detailed descriptions.

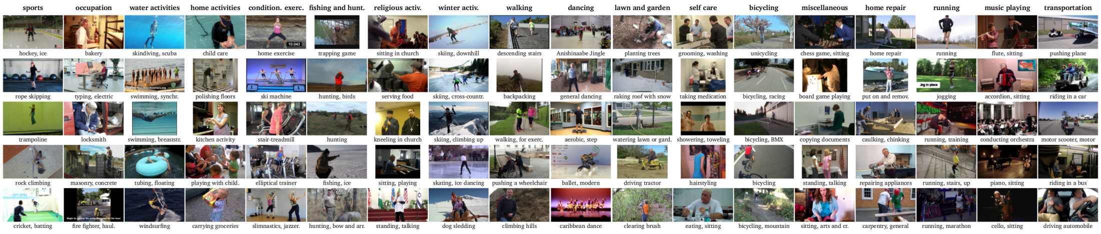


Fig. 2.16 displays the example of MPII Human Shape dataset

Fashion-Gen dataset [141] contains 293,000 high-definition (1360 x 1360 pixels) fashion images paired with item descriptions provided by professional stylists. These datasets contain multiple angles and pose for each item. The main purpose of this dataset is to explore the text-images synthesis task.

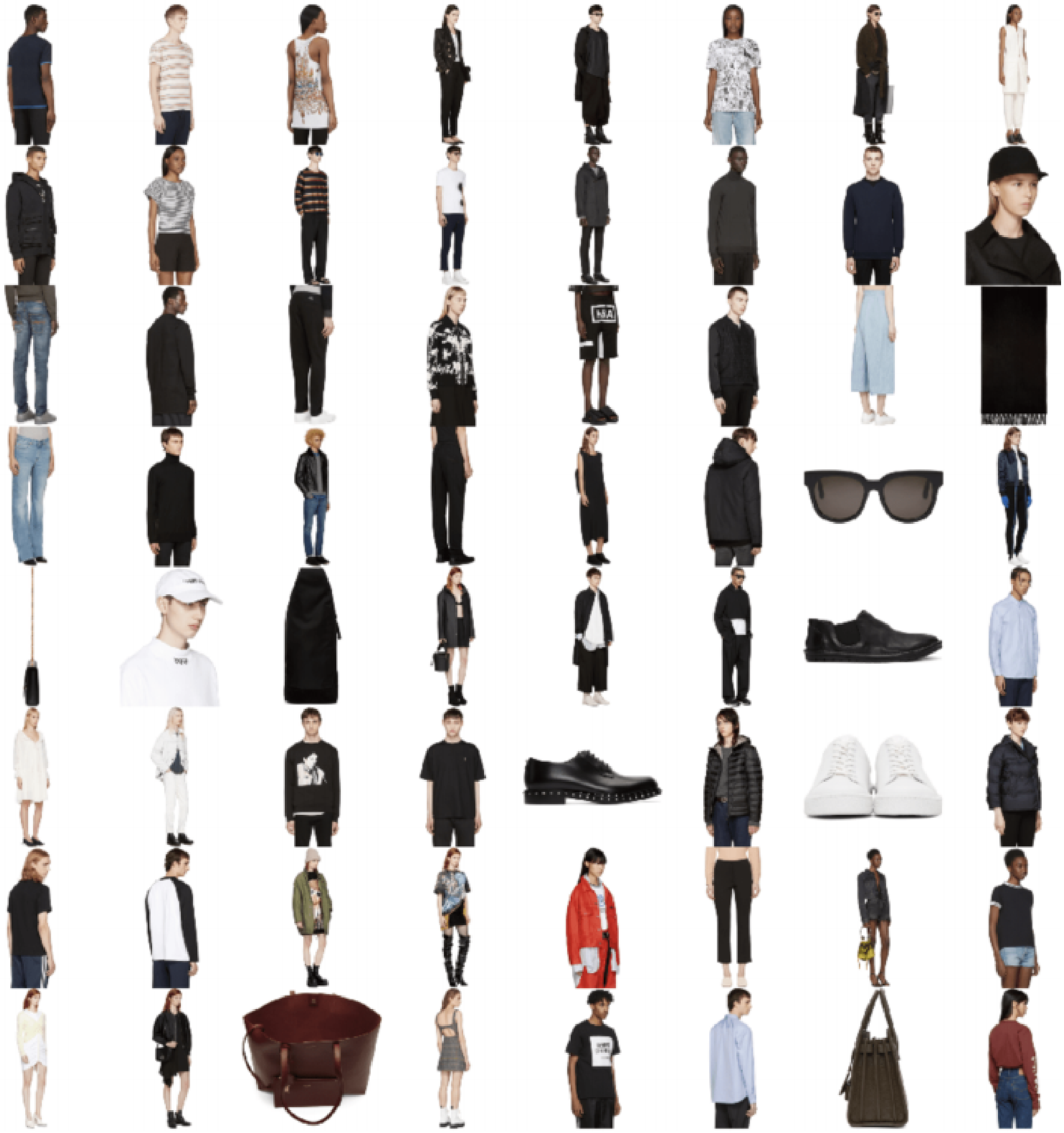


Fig. 2.17 displays the example of Fashion-Gen dataset

TOSCA [40] created a dataset that was designed exclusively for the recovery of 3D shapes. In shape retrieval, the emphasis is placed on the creation of shape descriptors or signatures that accurately capture the one-of-a-kind characteristics of the shape and remain unchanged regardless of the type of transformation being applied. The rotation and translation transformations are the ones that are utilised the most frequently in rigid shape analysis. When it comes to challenges involving shape recovery, the number of possible transformations is significantly higher. These include scale, missing parts, varied sampling, and triangulation. The database includes a total of 41 different animals, such as 11 cats, 9 dogs, 8 horses, 6 centaurs, 4 gorillas, and 3 wolves. Additionally, there are 14 human bodies in the database, with 12 distinct female figures and 2 different male figures, each with 7 and 20 different positions. Due to the fact that this dataset does not focus solely on the human body, it is not suitable for conducting a study of the human body's shape.

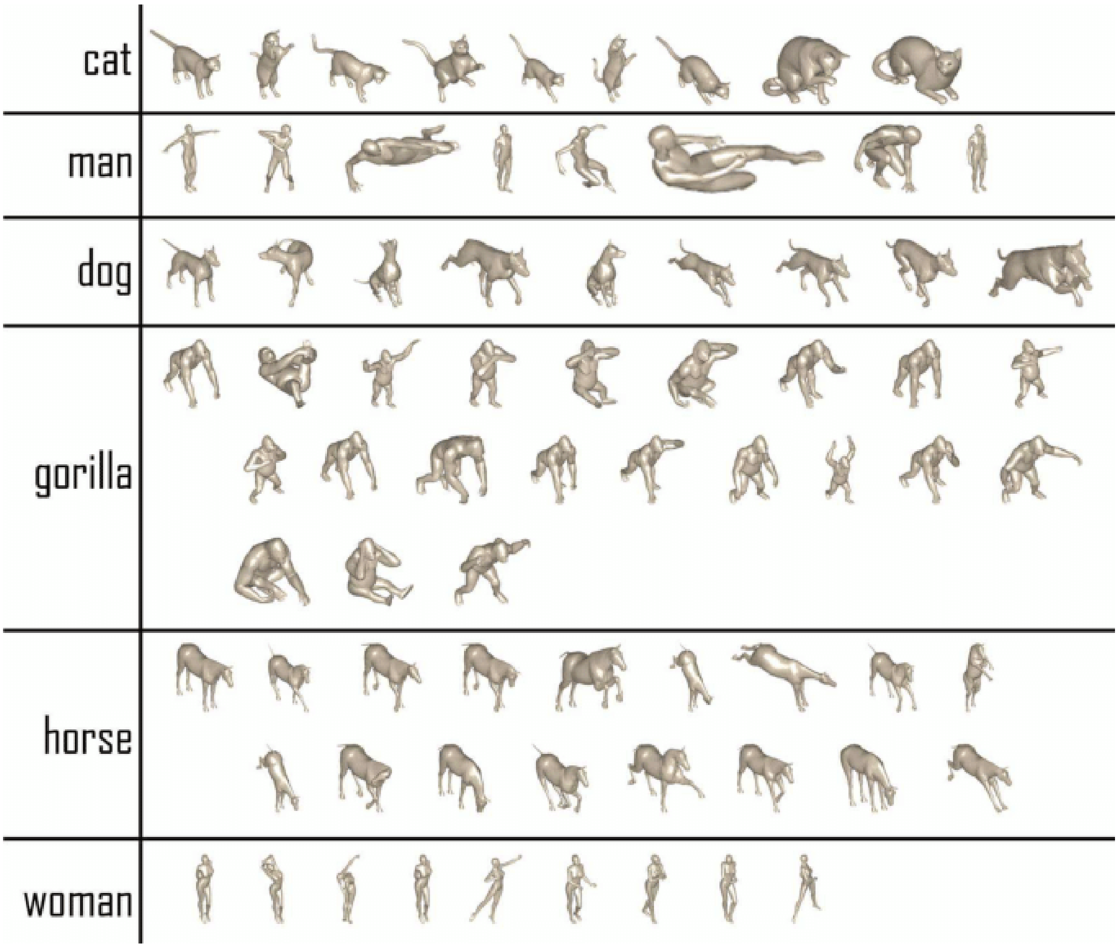


Fig. 2.18 displays the example of TOSCA dataset

SHREC’10 [39] The TOSCA dataset was expanded thanks to the contributions made by SHREC’10, which included more reliable large-scale retrieval, correspondence, and feature recognition and description. This database grew from 80 to 148 objects during this expansion. Unfortunately , the number of human subjects only increased slightly, with additional 12 female subjects being the only ones to volunteer for the study.

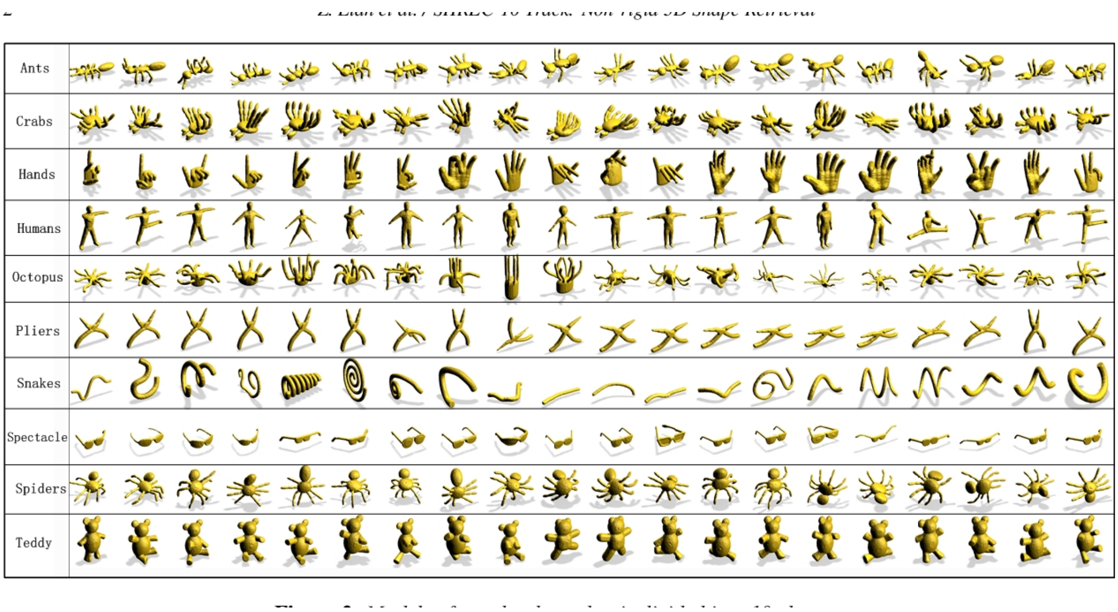


Fig. 2.19 displays the example of SHREC10 dataset

NHANES [89] The National Health and Nutrition Examination Survey, also known as NHANES, is a programme of research that aims to evaluate the health as well as the nutritional status of children and adults living in the United States. The findings from this survey will be used to identify the prevalence of major diseases as well as the risk factors for diseases. This survey is one of a kind since it combines interviews and physical examinations to get at its findings. Although it does not include any 3D data or photographs of the participants, this dataset does give an enormous quantity of information on the subjects’ habits, health, and morphology in the form of anthropometric measures. These measurements may be found in the dataset. In addition, this dataset is not very valuable when considered from the perspective of computer vision. However, it is highly beneficial when considered from the perspective of statistical analysis of the population as well as the generation of virtual subjects.

In Chapter 4 we will focus more on these datasets and how some of them helped us to obtain quality data from the applications and software that already exists in the store.

Table 2.5 Comparison of different type of human body dataset

Datasets	Real Human	Number of images meshes	Resolution	Description	Poses	Number of items
CAESAR	1500	300K	Varying size	No	multiple	37k
CAESAR3D	2400	Unknown	Varying size	No	multiple	Unknown
SCAPE	71	125K	Varying size	No	multiple	Unknown
FAUST	10	300	612 x 512	No	multiple	Unknown
MPII Human Shape	410	25K	Varying Size	yes	multiple	40k
Fashion-Gen Dataset	Unknown	325K	1360 x 1360	yes	multiple	78k
TOSCA	14	Unknown	Varying Size	No	multiple	Unknown
SHREC’10	26	Unknown	Varying Size	No	multiple	Unknown
NHANES	5000 each year	Unknown	Varying Size	yes	multiple	Unknown

Note: Data collected directly from each paper or original website of the following datasets.

2.11 Challenges

Among the most significant challenges are (1) complexity of human body, especially non-rigid parts of the body, (2) variability of the human physique, (3) complexity of human skeletal structure, (4) the impact of breathing, as the body expands and contracts (5) variability in lighting condition, causing shadows (6) depth, i.e. the loss of 3D data that results from observing the pose from 2D planar image projections, and (7) complexity in capturing parts of the human body that are covered by loose clothes.

One of the main challenges in this research study is the complexity of the human body with the range of multiple degrees of freedom. As the body moves, with the impact of breathing, the body expands and contracts. To accommodate the expansion and contraction of the body the garment may be distorted; thus, the human body changes that occur during movement could affect the fit of the garment.

Depth is another main issue because some of the 3D information is not available directly, in particular the depth inference from image sequences. Obtaining information about the depth of the human using a single camera is very challenging, considering the multiple degrees of freedom of the human body described in the previous paragraph.

Changes in lighting and camera viewpoints cause effects like shadows and other variations in image appearance. Also, shadows are cast on areas of the subject's body which are covered by hair, Those are all image-based problems. All of these elements make it impossible to reliably identify image features that can be relied on to extract relevant information.

Wearing specific clothes, posing with a specific posture in front of the camera is something that the consumers may not be willing to do, thus all these are creating other issues. Most of the applications that already exist in the store require the consumer to take some steps, such as wearing tight clothes or posing in front of the camera, which makes them uncomfortable with using these applications.

2.12 Summary: Literature Review

To summarise, this study have exposed several topics starting from the evolution of human body measurements, going into the number of different 3D and smartphone body scanning technology followed by a review of existing companies' approaches for the calculations of human body measurements for the fashion industry.

Through a better understanding and investigation the interaction between the human body, garment, and pattern, this study is able to improve garment fit. In reaching a good garment fit, body measurement is critical. For producing a better garment, researchers have developed new methods and software to capture a higher quality of data from the human body, such as with motion capture system, 3D body scanning technology, smartphone body scanning. The next frontier for body measurement methods can be to measure the human body surface from one single depth-camera. The significant element that affects the body measurement changes is measuring and evaluating the human body data in motion, which is related to wearing-ease amounts. The changes in body surface measurements between various dynamic postures and standard anthropometric measurements were mainly focused on previous research studies. By using technologies such as computer vision and machine learning methods, we can consider more natural continuous motion during real movement; however, there is still limited research on the body measurement changes during natural movement and the difference in body measurement changes across body sizes.

Moreover, by emphasis on shape in subsection 2.4.6 and subsection 2.6.3 different software and methods have been investigated in more detail in order to find out the most reliable option to capture human body measurements in static view or motion within seconds. Challenges were considered, including reviewing the human body landmarks for the purpose of finding the most accurate garment from the human body surface.

This study also understand that capturing the human body to find out the size of the consumers provides several benefits and becomes necessary under several scenarios; however, owing to the complexity of finding accurate measurements and the various parts of the body, there are still great hurdles to overcome before reliable and effective technology becomes available for practical applications. Among the 3D body scanner and a laser scanner that can be used, the combination of computer vision and machine learning methods has the biggest potential, as it requires minimum pre-setting, can be installed on any smartphone, has no need for subject cooperation and the price is fairer for daily use. However, existing applications are still not a perfect option for clients to use, as the captured data are not providing sufficiently information.

In section 2.10, some useful datasets with a focus on body modeling, human body measurement, and computer graphics. These datasets are the basis for the representation that will be partially used in Chapter 4 as part of the initial stage of our experiments. Also, taking advantage of the provided datasets and understanding the lack of them to restrict our attention to human body measurements, we have developed our own dataset including human body models in a more specific variety of poses and shapes according to our research studies.

Chapter 3

Method

In this chapter, we describe the methodology used in my research, which is aimed at developing a software solution for improving the accuracy of garment fit. We begin by outlining the hypotheses of my PhD, which includes the objectives of this research.

Next, we discuss the selection of measurements, focusing on the human body measurements that are relevant for achieving accurate garment fit. We describe the various measurement techniques that we will use and explain why they are important for our research.

Moving on to the participant selection and limitations, we explain that we have chosen to focus on female participants, as they offer greater variability in shape but also because they represent the majority of the online shopping sector in the fashion industry. Acknowledging the potential challenges and constraints that may affect our results. We discuss the limitations associated with participant selection, measurement techniques, and data analysis.

Next, we describe the methods that we will use for data analysis, including statistical analysis and machine learning algorithms. We explain how these methods will be used to identify patterns and relationships in the data and to develop predictive models for improving garment fit.

Finally, we summarise the chapter, highlighting the key elements of our methodology and the approach we will use for our research. Overall, this chapter provides a clear and detailed overview of the methods that we will use to achieve our research objectives and to develop our proposed software solution.

We have developed a state-of-the-art lightweight measurement system that is capable of extracting anthropometric measurements from two 2D images to capture the surface of the upper human body for application in the apparel/fashion sectors.

We have explored different approaches to achieve improvements in state-of-the-art human body measurements with the help of machine learning, computer vision, and ellipse mathematical equations.

We have compared the final stage of the prototype that has been generated with the tape measurement methods. The final stage of the prototype has been generated with the help of machine learning and computer vision methods such as; MobileNet SSD, image corrections, chessboard detection (camera calibration), photogrammetry, and image segmentation. These techniques helped me to improve the state-of-the-art in terms of accuracy to $\pm$ 1cm average error (in comparison to the tape measurement methods) on **78** actual human participants (who self-identified as female).

The following chief approaches are:

1. To capture a user-friendly way the upper body parts of human participant's
2. To illustrate the collected data to the fashion designer who we are in contacts.
3. To compare the data with the tape measurements methods.
4. To develop user-friendly software to capture the upper body parts of human subjects (at this point the final stage of the prototype is running on the terminal).

As the habit of shopping online for clothing becomes more popular, so is the tendency to return garments with poor fit, often due to a mismatch between what is visually experienced online and the real-life experience of the purchaser. The high rate of returns (for some items up to and over 50% [105, 108]) represents a major challenge for the industry. Better capturing and understanding body measurements are known to be some of the major factors limiting progress on this issue. Furthermore, we also conducted a state-of-the-art survey of the younger generation and found that they are more willing to do online shopping than before. Therefore, the demand to find a better garment fit in digital stores is rapidly growing.

In the absence of person-to-person interaction (for example due to the COVID-19 pandemic), state-of-the-art, self-service contactless body measuring solutions are needed, enabling a simple way to digitise measurement capture so that made-to-measure businesses can easily operate online.

Moreover, many human reconstructions are model-based, relying on the parametric regression model; however, the prediction is typically for the almost naked body (tight clothes or underwear) and all the personal details such as clothing are ignored.

Also, due to the lack of technologies based on our phase 1 experiments (please refer to Chapter 4 for more details), the average errors of the current software and applications for every individual part of the human body are up to $\pm$ 5cm.

3.1 Hypotheses

It is possible to develop a measurement system that is capable of extracting anthropometric measurements from multiple images to capture the human body surface, with improvements to the state-of-the-art in terms of accuracy-computational expenses. Thus, our approach will be feasible for performing anthropometric measurements on the new-gen of smartphone devices.

The study has the following chief objectives:

1. Capturing human body reconstruction with the help of new AI-powered technologies (machine learning, computer vision, and 3D matching) with any backgrounds. The proposed system aims to revolutionize the field of human body measurement by offering the following capabilities: (i) Estimation of the human pose of the person on the screen and providing accurate body measurements for the selected areas Table 3.1. This includes capturing key joint positions, limb lengths, and torso dimensions, enabling a detailed understanding of the human body's physical characteristics. (ii) Enabling precise body measurements of the selected regions, regardless of the background. The system utilises advanced computer vision algorithms to separate the human subject from the surrounding environment, ensuring accurate and reliable measurements even in complex backgrounds or cluttered scenes (iii) and, Facilitating body measurement from varying distances from the camera viewpoint. By leveraging machine learning techniques, the system can adapt to different camera distances and angles, allowing for flexible data acquisition. Whether the subject is close to the camera or positioned farther away, the system can adjust and provide accurate body measurements consistently.

2. Capturing the human body reconstruction from multiple images at least two (where 3D information is not available directly, because of the depth inference from a single image, the systems cannot predict the distance of the human body from the camera).

3. Measuring the human body circumferences with the help of ellipse formulas. A fully trained system that can choose the best ellipse equation according to the human body shape to calculate human body circumferences.

The proposed objectives imply a number of hypotheses, namely:

1. There is a need from both *industry* and the *public* for a better technical solution to measure the surface of the human body for industrial applications.

2. As the demand is rapidly growing in the young generation to have better fitting garments, this is even more desirable, so more and more people are interacting digitally in the apparel/fashion industry.

3. There is technology to achieve this: with the help of recent machine learning and computer vision to measure the human body surface with new-gen of smartphone devices.

4. It is possible to develop a measurement system that is capable of extracting human body measurements in movements, as well as body changes, that impact the experience of wearing a piece of clothing.

5. It is possible to improve the state-of-the-art in terms of accuracy (± 2cm average error).

6. Given Hypothesis 3 (together with Objective 2) and a supervised learning approach, it should be possible to extract 3D information directly from multiple images to collect the lost 3D data. The system also can predict the distance of the human body from the camera.

7. It is possible to predict all of the personal details, such as clothing. Also, to demonstrate higher flexibility and expressiveness, to recover details presents in the input image more accurately [145].

Hypotheses 1, 2, 3, 4, 5, 6 (together with Objective 1) are based on several successful attempts to develop a measurement system for human body reconstructions [145, 11]. The results will be evaluated (i) qualitatively, based on the opinion of the designer that we are collaborating with and (ii) quantitatively based on the series of user studies. These will be conducted at Goldsmiths University of London and will be generally aimed at developing better measurement systems. Hypothesis 7 is grounded on sociological evidence that participants prefer to wear appropriate clothing in front of the camera rather than being subjected to specific clothes (tight clothes or underwear).

3.2 Selection of Measurements

As an initial further constraint, we plan to focus our research and development on the upper body parts. Our initial research shows that some of the most serious garment distortions occur in the upper arm, elbow, back, and armhole areas when extended body postures are performed.

Table 3.1 demonstrates the selected joints of the human body that are going to be investigated in this project. This table represents a list of all anthropometric measurements that will be used obtained from the participants.

Anthropometric Measurement	Identifier
1. Upperarm Circumference (UA C)	measure/measure-upperarm-circ-incr|decr
2. Upperarm Length (UA L)	measure/measure-upperarm-length-incr|decr
3. Lowerarm Length (LA L)	measure/measure-lowerarm-length-incr|decr
4. Wrist Circumference (Wr C)	measure/measure-wrist-circ-incr|decr
5. Bust Circumference (B C)	measure/measure-bust-circ–incr|decr
6. Upperbust Circumference (UB C)	measure/measure-upperbust-circ-incr|decr
7. Back width (B W)	measure/measure-back-width–incr|decr
8. Waist Circumference (WA C)	measure/measure-waist-circ-incr|decr
9. Hip Circumference (H C)	measure/measure-hips-circ-incr|decr
10. Neck Circumference (N C)	measure/measure-neck-circ-incr|decr
11. Neck Height (N H)	measure/measure-neck-height-incr|decr
12. Shoulder Distance (S D)	measure/measure-shoulder-dist-incr|decr

Table 3.1 A list of all anthropometric measurements will be used to obtain from the participants including a list of identifiers.

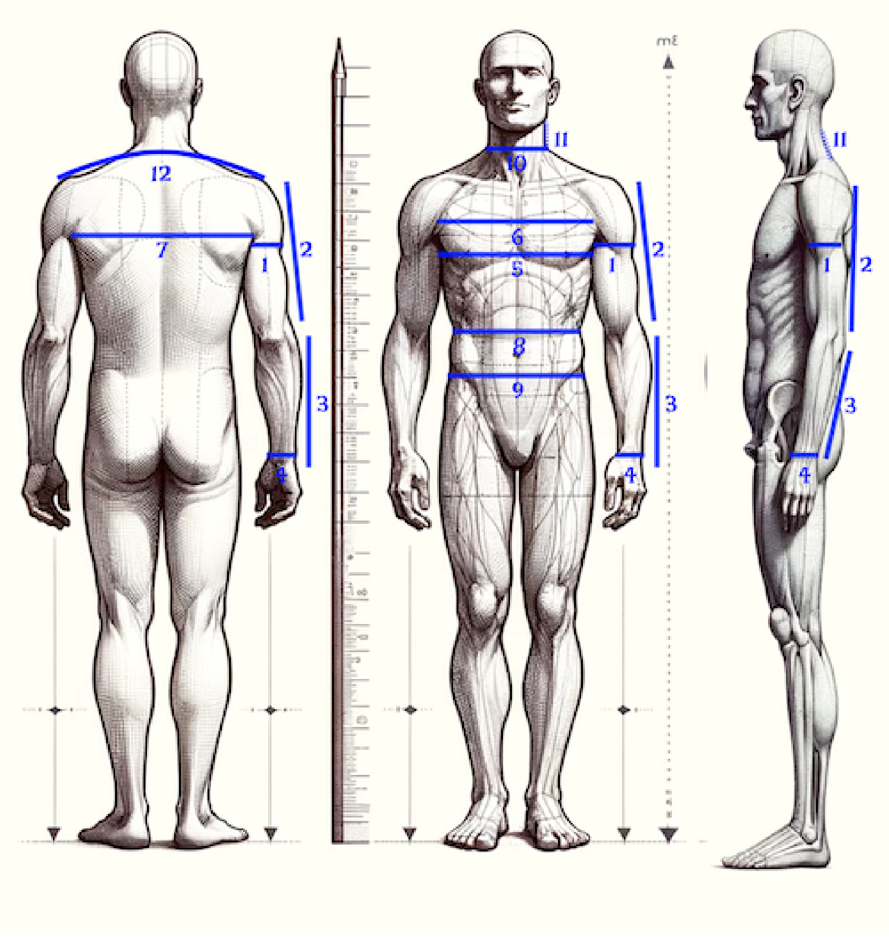


Fig. 3.1 Provide a compilation of all the anthropometric measurements based on the table 3.1

The standard chest, bust, waist, and hip measurements will be the main focus of the results for comparison with previously published research and validation of technique.

3.3 Choosing the participants

The participant will be divided into two groups. The first group is the volunteer students from Goldsmiths, University of London, who will be recruited through posters or via email. The volunteer participants will be paid a small compensation. Please look at the **Ethical Approval Forms** that attached to the end of this thesis Appendix G.

The second group of our dataset consisted of volunteer customers associated with, **Nevena Nikolova**, an established fashion designer both in the UK and Bulgaria. These participants were sourced owing to Ms. Nikolova's willingness to collaborate and contribute to this research study.

3.4 Limitation

The participants will be limited to the age range (18-45) and women. According to my initial studied and surveys, women are willing to buy more clothes than men, which is why we have decided to limit the participants to the group. Also, the younger generation is more likely to employ modern technology to discover the best fit for their clothing. Furthermore, on the basis of research and experiments we have been through, the average error of different applications is higher for women than for men (Please see Figures 4.2), this is because of the awkward positioning of certain measurement areas. Therefore, we found it more challenging to continue our research based on this.

In addition, Ms. Nikolova's clients are all women so based on her needs we found that it is better to limit our participant to the women.

3.5 Analysis of the data

In order to answer the research questions, quantitative data will be collected from the proposed applications. The quantitative data will be analysed using a descriptive statistic, an independent t-test, Anova test and Post-hoc[1]. The quantitative data will be reviewed and compared at every different stage, such as in terms of accuracy and reliability of each method and technique. Quality data will be collected as well based on the opinion of the designer that we are collaborating with.

[1] Post hoc ("after this" in Latin) tests are used to uncover specific differences between three or more group means when an analysis of variance (ANOVA) F test is significant.

Our data collection will be divided into four main areas: (i) comparison of the ellipse equations, (ii) comparison of the different background images - plain/cluttered, (iii) comparison of different human body postures (relax posture, t-pose, A-pose), and finally (iv) analysis of the data based on the human body's distance from the camera's point of view.

3.6 Summary: Research Proposal

The exploratory research will be developed on the basis of the framework for micro and macro levels of fit. To examine body surface measurements using the new generation of smartphone camera (which acting as a scanner) with the help of machine learning approaches and computer vision the methods for this study will be developed. The upper body surface in relaxed and natural posture in relation to the chest, back, shoulder and hips will be chosen for this research. Measurements for this study will be selected on the basis of the application to pattern development. The preparations for capturing the opponents will be prepared well before the participants' visit. The collected data from each participant will be compared once the scanning process is completed using relaxed and natural posture. The process of scanning the participants will be repeated at least twice to compare the results with one and each other.

For the standing posture, the average value of the two trials from the proposed application will be used to compare to the measurements from the other existing technology (tape measurements). From the 12 selected measurements during the relaxing and natural posture test, the mean value of minimum decrease and the maximum increase will be used to examine the changes in body measurement. The quantitative data from the proposed application and other existing technologies (possibly old traditional tape measurements) will be analysed using descriptive statistics.

Answering these research hypotheses in unison implies searching for a series of instructions, a plausible set of processes, that when executed by a computer lead to the generation of human body reconstructions. The identification of such processes is intrinsically equivalent to the inference of a theory of human body measurements which is generative as opposed to solely descriptive. In addition, the investigation of these technologies is likely to contribute to active areas of research in the field of Computer Science and Fashion Technology and represent a further step towards the development of human body measurement systems for garment fit.

Chapter 4

Human Body Measurement Experiments and Dataset

This chapter focuses on the experiments conducted on existing fashion and entertainment applications to measure human body measurements. The chapter begins with a review of the datasets used for the experiments. Then, it delves into the various methods used to measure the human body, which are summarised in the chapter. These methods include fashion and entertainment applications testing, RMPE (Regional Multi-Person Pose Estimation), measuring ruler, motion tracking for consoles and smartphones, PI-FuHD, AR and deep learning-based automatic human body measurement system for babies, and human body measurements using computer vision.

The chapter aims to examine which experiment can obtain the most accurate information from the human body based on macro and micro data. The experiments are evaluated based on their ability to accurately measure human body measurements. The chapter discusses the results of each experiment and compares them to one another to identify the most effective method for measuring human body measurements.

In conclusion, this chapter presents an in-depth analysis of the experiments conducted on existing fashion and entertainment applications for measuring human body measurements. The chapter provides valuable insights into the various methods used for measuring the human body and helps to identify the most effective methods for achieving accurate measurements. The results of the experiments can be used to inform the development of the software solution proposed in this research project.

4.1 Dataset

A key to a successful deep learning model is a good training set of sufficient size. To analyse the population of 3D human bodies, we had to collect data from multiple databases. Numerous publicly accessible research datasets permit the examination of posture and form variations simultaneously; however, they provide data on no more than 100 individuals [25, 37, 75], which restricts the range of shape variations. We came across the CAESAR database [127] with more than 4,500 subjects in a standard pose, which is the largest commercially available dataset to date that contains 3D scans, and it represents a much richer sample of the human physique. Multi-person datasets (MPII) and MSCOCO datasets have also been reviewed as part of the Regional Multi-Person Pose Estimation (RMPE) project. As part of our experiments subsection 4.2.1, we have acquired a subset of 30 human body measurements, comprising 15 male and 15 female bodies.

4.2 Measuring the Human Body

4.2.1 Experiment 1: Fashion and Entertainment Application, Testing

As part of our experiments, we have acquired a subset of 30 human body measurements, comprising 15 male and 15 female bodies.

How do garment applications work?

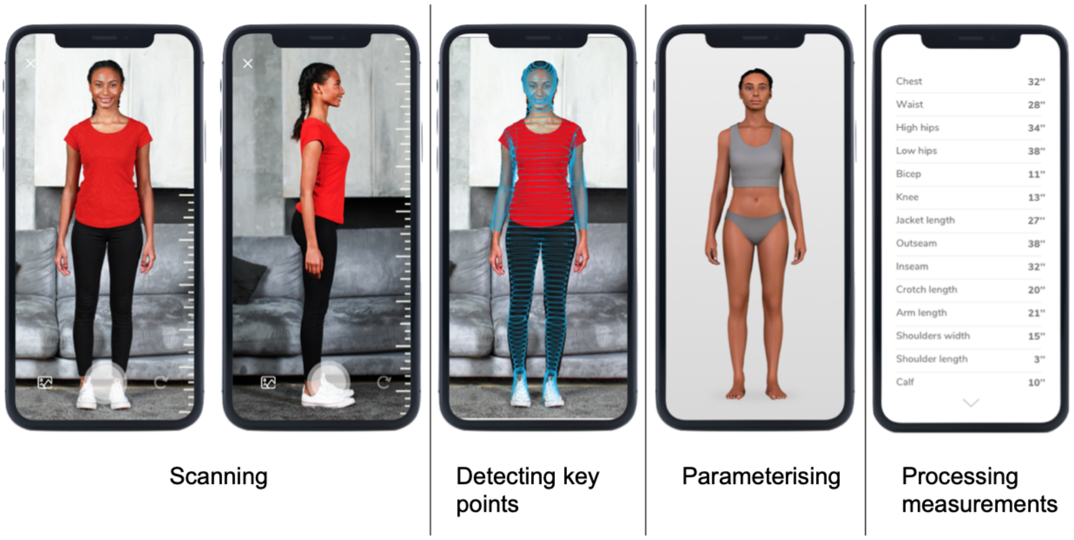
Off the shelf applications in the store which are created for human body measurement, such as Zalando, 3DLookME, VyoO, SizeStream (methreesixty), MyFiziq, Nettelo 3D, etc. All these applications have the same concept and work very similarly to one another. The process of capturing the human body starts with users entering some personal details such as height and weight in the application and afterwards taking two pictures from the front and side of themselves and uploading it to the app. The applications generate a 3D avatar representation of the model with a rough circumference measurement. In the following paragraph, we look at the different steps to reaching human body measurements.

Computer vision, machine learning, and 3D matching are the key techniques used by these applications to capture the human body from only two photos.

What are the steps to get the size of clothes?

The existing apps have four steps to calculating human body measurements including scanning the actual body (front and side view), detecting key points, developing a 3D avatar (virtual model on the surface) and, finally, processing the measurement and handing it over to the user. Figure 4.1 shows all the different steps to measuring the human body.

Although the actual scanning only takes a few seconds, the processing time can take around one to two minutes per person to get the results, and occasionally there can be problem which forced the users to conduct the scanning process again and again in order to get the final findings. The scanning process starts with a very straightforward stage that involves using a smartphone to take pictures from the front and the side, followed by the detection of a scan of the intended human body using computer vision. The contour of the human body that was collected is immediately sent to be analysed by the algorithm of the neural networks. After that, the application, determines and detects



Scanning

Detecting key points

Parameterising

Processing measurements

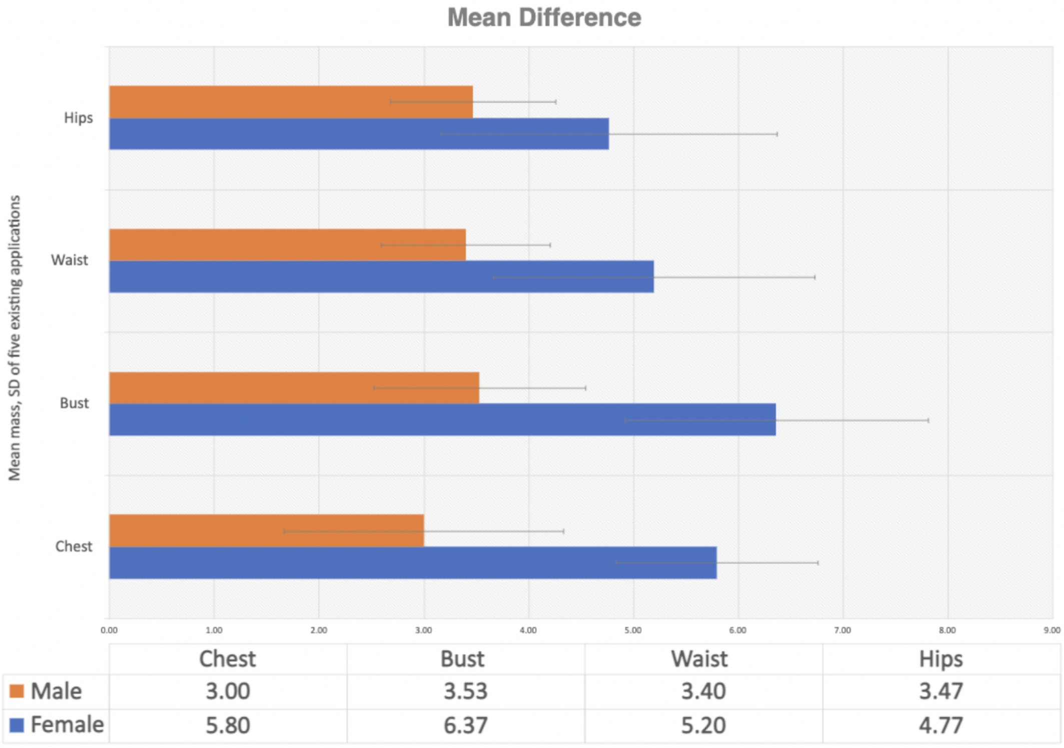
Chest	32"
Waist	28"
High hips	34"
Low hips	38"
Bicep	11"
Knee	13"
Jacket length	27"
Outseam	38"
Inseam	32"
Crotch length	20"
Arm length	21"
Shoulders width	15"
Shoulder length	3"
Calf	10"

Fig. 4.1 shows different steps of 3DLookMe application to get the accurate measurement. Image courtesy of [11]

key points and creates a set of probability maps for each key point (the measurement key points could be customised, which allowed specific profiles for selected garment types and a standardised size table to be automatically generated, so the users were able to move the key points to the correctly located place on the images). These maps go through continuous processing and are mixed with various specified filters at each level. On the photos, important spots were determined in order to begin the process of initialising the 2D contour matching.

A virtual camera is utilised in order to develop each of the 2D contour models into parameterised 3D human character projections onto an image plane. This is done by projecting the models onto the image plane. By matching virtual 2D with the real 3D human character, it ensures that there are no errors, or at least just a tiny error, in the scanning of the human body assessment and processing of measurements. The process of 3D matching is also utilised in the construction of a 3D model of a human body. The inaccuracy of these applications that are available in the store is demonstrated in more detail in Figure 4.2.In the following sentence, we will provide an explanation of how we organised our tests across all of these applications in greater depth.

The following statistics represent the average error of different sections of the human body in different applications. Some of the participants did the capturing process a couple of times to obtain the most accurate data from the apps. As can be seen from the statistics below, to obtain data from the human body, users are divided by gender on the basis of the different anatomy of the male and female bodies. Users are asked to wear appropriate or tight-fitting clothing to capture accurate information, otherwise the looseness of the clothes might be observed as part of the human body. Each user stood approximately 10 times in front of the camera, and the best data are selected for these statistics Figure 4.2. Please refer to the Excel files to assess the outcomes for each of these application measurements.



	Chest	Bust	Waist	Hips
🟧 Male	3.00	3.53	3.40	3.47
🟦 Female	5.80	6.37	5.20	4.77

Fig. 4.2 displays mean difference (cm) and standard division of five existing applications in the store (3DLookMe, SizeStream, Esenca, PreSize and TechMed). This data is been taken from 15 (who self-identified) as female/male participants in the same room conditions from five different applications that exist in the store. **This experiment took place in December 2022.**

Challenges in existing Applications

According to the collected data, it is not as easy as we think to find and determine total and accurate data for clothing provision. With the traditional, manual way, precise location of body landmarks and how these are handled can be subjective. Also, there are many more issues of ethical consideration that are due to the awkward positioning of certain parts of the body and measurement areas. Most up-to-date technologies are including serious issues as well as the old traditional way in landmarking. Furthermore, wearing undergarments in front of the camera, to capture human body measurements, is another main issue of such a system. The other problem that we can address in this thesis is the impact of breathing. If the part of the body which is observed is covered by hairs, shadows are cast on areas of the subject's body. Those are all image-based problems, which means we could face us some errors in particular places.

4.2.2 Experiment 2: RMPE: Regional Multi-Person Pose Estimation

The RMPE project has been tested, to estimate the human pose estimation. RMPE is one of the most popular top-down methods of pose estimation. Accuracy of the person detector is very important for top-down methods, as pose estimation is performed on the region where the person is located. The pipeline of the proposed RMPE is illustrated in Figure 2.8. The RMPE framework represents three major components: **1** SSTN and Parallel SPPE **2** Parametric Pose NMS **3** Pose-guided Proposals Generator PGPG.

Evaluation Datasets, Implementation and testing results

By utilising the terminal and PyTorch on both Windows 10 and Mac OS Mojave, we were able to successfully implement the algorithm that was suggested in a paper published by RMPE and that was evaluated both qualitatively and quantitatively on two typical multi-person datasets: MPII and MSCOCO. The minimum required GPU for this project is 6[1], otherwise, the operation will not be successful and will fail.

They have used the VGG[2] based SSD-512[3] as the human detector (we have used the same object), as it performs object detection effectively and efficiently. The detected human proposal is extended by 30 % in both directions, in order to make sure that the entire person region will be extracted.

In addition, the libraries Torch, TensorFlow, and Caffe are necessary for the operation of this project. We were finally able to get this project up and running when the prerequisites had all been satisfied, including the installation of the necessary libraries. It is important to keep in mind that one can speed up the process of the posture estimation stage by using multi-GPU testing for large batch testing. This is possible if one possesses a large quantity of GPU memory or numerous GPUs.

By putting the results of both datasets next to each other (MPII and MSCOCO), the author found that this can accurately predict pose in multi-person images. Figure 4.3 displays the results of RMPEE for both MPII and MSCOCO datasets.

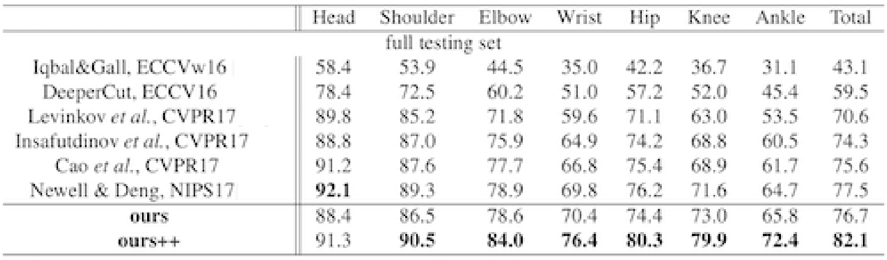
The quantitative findings from the comprehensive test set are presented in the following table. An average accuracy of 72 mean Average Precision (mAP)[4] was attained by the author Fang et al. [64] when detecting problematic joints such ankles, wrists, knees, and elbows. This is 3.3 mAP greater than the previous state-of-the-art result. They are able to arrive at a final accuracy of 70.4 mAP for the wrist and 73 mAP for the knee. We are able to significantly improve our results and attain 82.1 mAP by employing a more robust human detector and posture estimator. This puts us 4.6 mAP ahead of our previous best result.

[1] For the successful operation of the implemented algorithm requires a minimum GPU compute capability of 6

[2] VGG is a convolutional neural network that is trained on more than a million images from the ImageNet database.

[3] SSD is designed for object detection in real-time which is composed of 2 parts: 1. Extract feature maps 2. Apply convolution filters to detect objects

[4] Please refer to Equation 5.6 for a validation of the Mean Average Precision (mAP) formula.



	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
	full testing set							
Iqbal&Gall, ECCVw16	58.4	53.9	44.5	35.0	42.2	36.7	31.1	43.1
DeeperCut, ECCV16	78.4	72.5	60.2	51.0	57.2	52.0	45.4	59.5
Levinkov *et al.*, CVPR17	89.8	85.2	71.8	59.6	71.1	63.0	53.5	70.6
Insafutdinov *et al.*, CVPR17	88.8	87.0	75.9	64.9	74.2	68.8	60.5	74.3
Cao *et al.*, CVPR17	91.2	87.6	77.7	66.8	75.4	68.9	61.7	75.6
Newell & Deng, NIPS17	**92.1**	89.3	78.9	69.8	76.2	71.6	64.7	77.5
ours	88.4	86.5	78.6	70.4	74.4	73.0	65.8	76.7
ours++	91.3	**90.5**	**84.0**	**76.4**	**80.3**	**79.9**	**72.4**	**82.1**

Fig. 4.3 Results on the MPII multi-person test set (mAP). “++” denotes using faster-rcnn with softnms [35] as human detector, PyraNet [179] with input size 320x256 as pose estimator [64].

Applications

Pose Estimation has applications in myriad fields, some of which are listed below.

1. Motion Capture and Augmented Reality

   The use of human pose estimation in CGI applications is one of the more fascinating applications of this technique. It is possible to superimpose graphics, fancy upgrades, styles, artwork, and equipment on a person if the human pose can be predicted. As the individual moves, the displayed images can "*naturally fit*" the person by sensing the variety in human position and adapting themselves accordingly.

2. Capturing the human body in 3D with Kinects and Smartphones

   Tracking the human in motion is another interesting application of pose estimation. To track the motion of the human and to use it to render the actions of virtual characters (by using IR sensor data), Kinects use 3D pose estimation.

4.2.3 Experiment 3: Measuring Ruler

The measuring of the 2D and 3D surface by using Apple devices camera has been analysed as part of this research study. We have developed two different measuring applications with the help of ARKit documents in this project to test the accuracy of such systems to measure the surface and human body.

4.2.3.1 Experiment 3.1: Measuring Distance (Smartphone)

We have developed a basic Ruler application on iOS as one of our experiments for this project. This application allows users to measure a 2D object from point-to-point (x1 to x2). To achieve these results, we have used the ARKit[5] framework. Figures 4.4 demonstrates how this application and the measuring process works.

[5] ARKit is a framework in iOS 11 which allows us to blend digital objects with the real world.

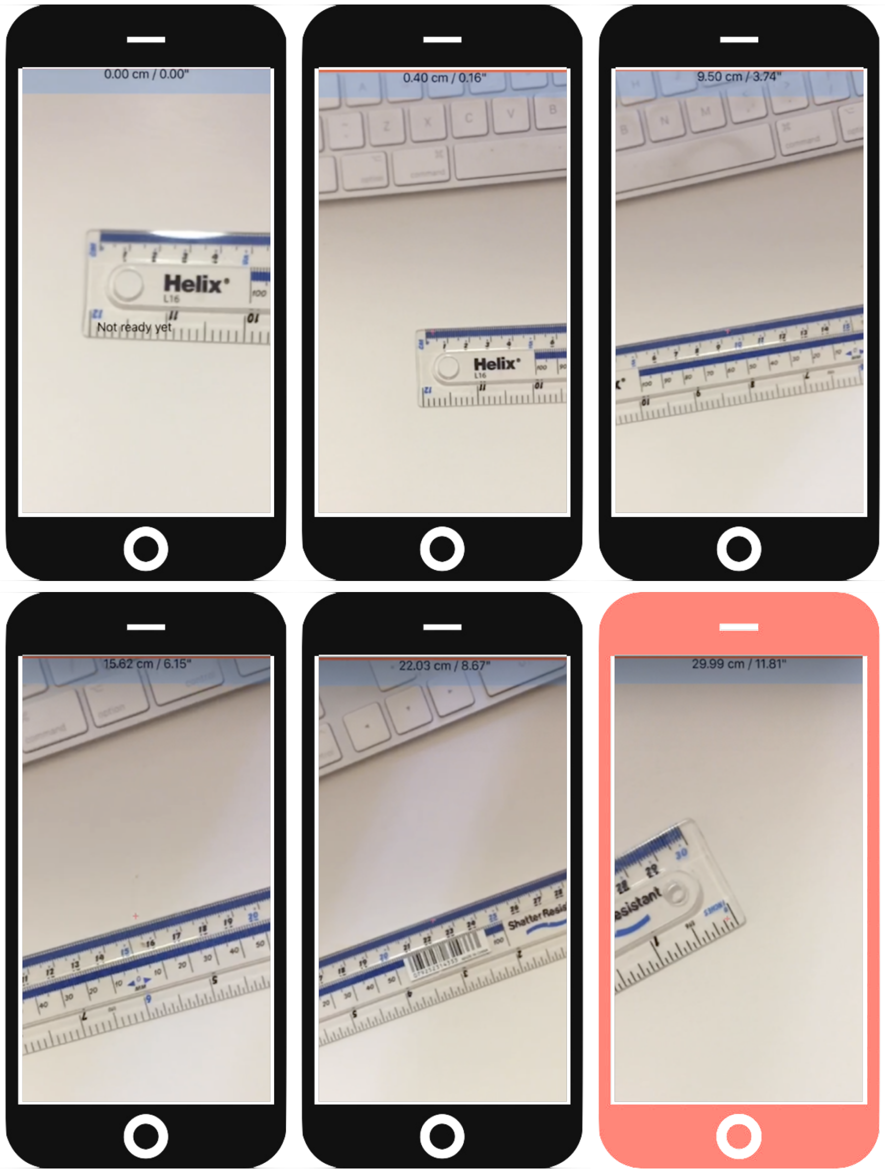


Fig. 4.4 During the initial phase of this project, an experiment was conducted to calculate 2D measurements with the assistance of ARKit.. Results reconstructed from six views

Using XCode, we created this application just for iPhone and iPad users. The macOS Mojave operating system and an iPhone 7s running iOS 11 or a later version were used for all of the experimental testing that was done. While the user is dragging the plus (+) icon from one location to another on the screen, this application is able to output the data simultaneously.

This application is speed-sensitive since it needs to calculate the things that are on the surface. In addition to this, it was not developed to compute 3D data information such as depth information. The distance to the object to be measured that has been selected is also very crucial; without it, the system will provide the users with inaccurate information.

One of the most significant benefits of using such tools is that the results may be processed very quickly, whereas the majority of the currently available apps require a processing time of no more than one to two minutes.

4.2.3.2 Experiment 3.2: Measuring Objects (Smartphones)

4.2.3.2.1 Augmented Reality The term "augmented reality" AR refers to a category of technologies that superimpose digital information on physical items or locations in the real world in order to improve the user experience. AR is not the same VR, which refers to the technology that places users in a computer-generated or digital environment. The ability to merge information from the real world and digital sources is currently the subject of research and development in many different industries, including marketing, museums, science, fashion, and many more. AR has a wide range of possible applications in both the commercial and academic research sectors [34].

4.2.3.2.2 ARToolKit To achieve calculating the human body measurement reconstruction, this framework plays an essential role: the C language software library, which was created for the creation of AR reality. This framework uses computer vision methods in order to compute the orientation and real camera position relative to physical markers in real-time, which allows one to overlay virtual objects. subsection 4.2.3 shows the experiments and testing that we went through with this technology [147].

- Acquisition of images from cameras
- Recognition of patterns and detection of markers

- Measurement of the orientation of markers and three-dimensional locations
- Display of composite three-dimensional images and real images

A method for determining the size of any object by seeing its surroundings through the camera of an Apple smartphone. Frameworks such as ARKit, SceneKit, and OpenGLES were utilised during the development of such an application. This application will function properly on mobile devices running iOS 11 or later. In order to give such data, it is necessary to put on a few essential classes. These classes are required to be added on. The primary operation that ARKit is responsible for was commanded and controlled by a *ARSession* object. These activities involve reading data from the device's motion sensing hardware, managing the built-in camera of devices, and doing image analysis on collected camera images. In addition, these processes may include reading data from external sources. As a consequence of this, the session brings together all of these findings to produce a connection between the virtual space and the space in the actual world. Therefore, by making use of *ARWorldTrackingConfiguration* in the beginning, users will be able to monitor and augment the world that is in front of the position and orientation of their iOS device. This class monitors the motion of the device using six degrees of freedom (6DOF) to measure movement along three translation axes (movement in x, y, and z) and three rotation axes (roll, pitch, and yaw). Secondly, by calculating the width and height, in pixels, of the captured camera image with the help of *CGSize*, we are able to estimate the measurements [27].

The Simultaneous Location And Mapping (SLAM) technology is what makes the iOS measure tool possible; it is a hybrid of photogrammetry and computer vision. The programme processes the frames that were taken by the camera, and computer vision analyses the collected image to locate and follow certain points of interest within the region that was captured. These tracked points are referred to as anchors, and they are represented by the yellow dots that can be seen in the figure 4.5 shown above. After the software has determined where the anchors are located on the screen, it will request that

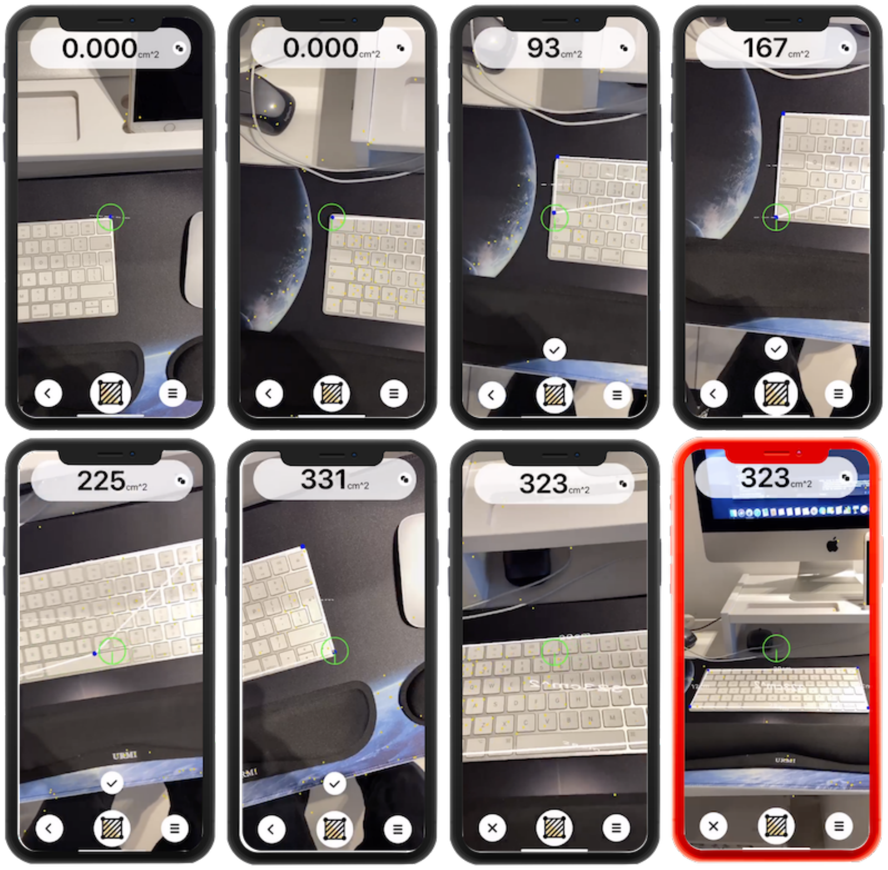


Fig. 4.5 During the initial phase of this project, an experiment was conducted to calculate 2D measurements with the assistance of ARTKit, SceneKit and OpenGLES. demonstrates few steps to get measurements from an object

the user move the device from one location to another. This move will update the new location of the anchors by shifting the perspective from which the cameras are viewing the action. This method makes adjustments to hundreds of thousands of points every single second. The programme is able to make an estimation of the distances between each point with a level of accuracy that is considered to be satisfactory. This estimation is based on the distinction that exists between the anchors in the various captures, the location of the camera, and to some extent, geometric shapes. Photogrammetry, which is the study of obtaining measurements based on photographs, can assist in this process so that it can accomplish its goals. An image captured by a camera is nothing more than a projection of three-dimensional data onto a two-dimensional plane. One loses a dimension of information while the image is being generated, thus this must be accepted (the depth). However, by capturing an image from different angles, as shown in Figure 4.6, we can retrieve the data that was previously lost.

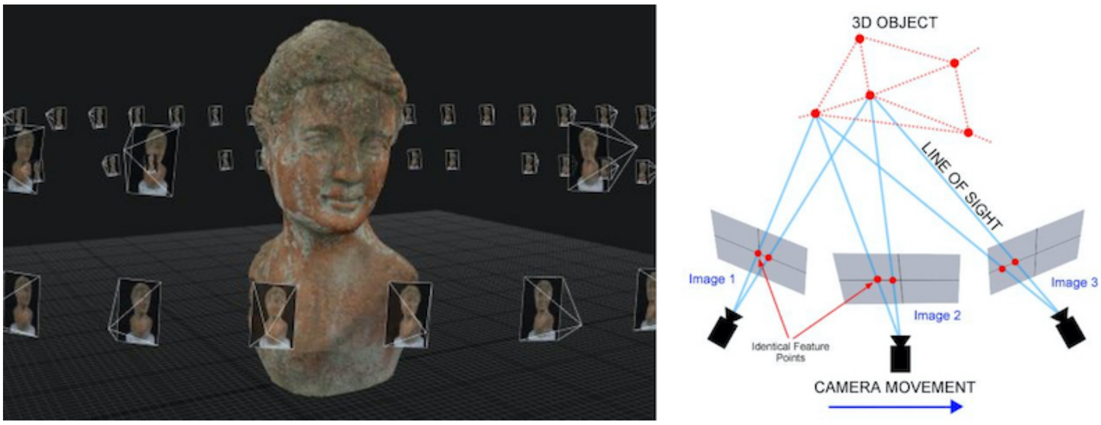


Fig. 4.6 The image on the left shows capturing 3d object from different angles [100] , The image on the right shows the concepts behind the photogrametry and how this technique works to capture lost data [110]

This is related to how our vision works. Our two eyes produce 2D information, in different positions, which is enough for the brain to reconstruct the depth information. Also, this is a similar principle to how we measure the distance of the stars, by comparing the images obtained from different places.

4.2.4 Experiment 4: Motion Tracking for Consoles and Smartphones

As part of the experiments that we conducted for this research study, we analysed the level of accuracy that can be achieved when capturing the motion of a human body using a Kinect console or a smartphone. The implementation of these strategies will be discussed in further detail in the following sections. Experiment 4.1 was developed in the past, and all that we did was test and compare the data from that experiment with the data from Experiment 4.2, which we developed with the assistance of ARKit documentation.

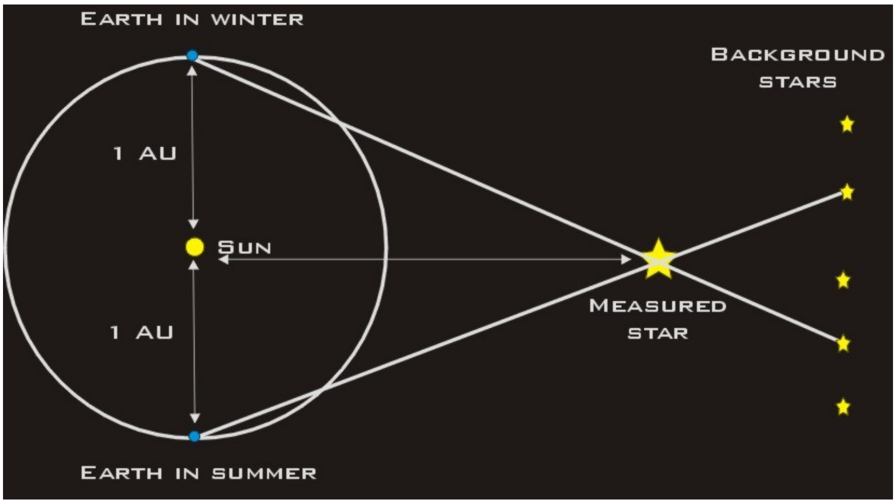


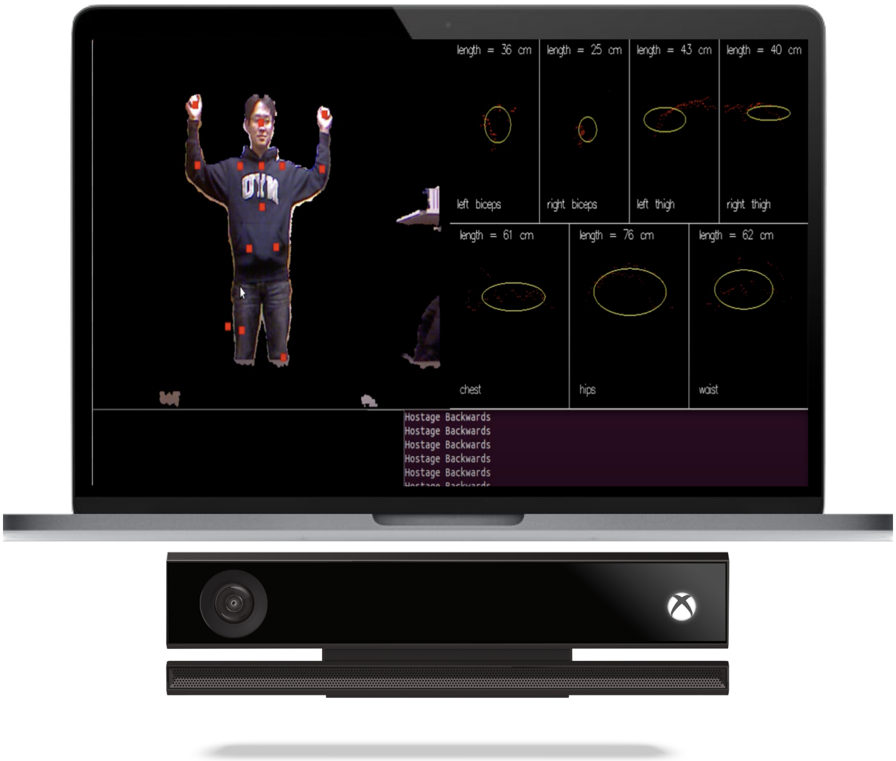
Fig. 4.7 Using Parallax to measure a star as seen from earth 6 months apart. Credit: ESA Science Technology [128]

4.2.4.1 Experiment 4.1: Kinect Sensors (Consoles)

The Microsoft Kinect is an example of a device that falls into the category known as depth cameras. It is the combination of an RGB camera and a depth camera, and it has the ability to measure simultaneously the distance to thousands of points inside a scene. The accuracy of the depth data obtained by a Kinect is much lower than that obtained by 3D body scanning; however, depth cameras such as Kinect are a cheaper system, and as a result, can be an easier solution for customers to purchase. In comparison to conventional 3D body scanning devices, depth cameras such as Kinect are a cheaper option for the system. During the course of our investigation, we came across a project called *"Skeleton Tracking and Body Measurements using Kinect"*, which was carried out at New York University using the Kinect. For this project, many Kinects were employed, yielding exceptionally accurate results in comparison to other approaches considered. Their findings motivated us to conduct our own experiments based on this project so that we could evaluate its accuracy in comparison to the various other approaches that we have considered. You may locate this piece of work on Github.

4.2.4.1.1 Skeleton Tracking and Body Measurements with Kinect Owing to the lack of resources, the reliability of these methods is difficult to consider. All the data have been reviewed and investigated from a research paper by Yuri Boykov [184].

This paper presents an implementation of the proposed method that is built with Visual C++ and the Microsoft Kinect Software Development Kit 2.0 running on the Windows 8 operating system. Each iteration of the test was run on a PC that featured a Core i5 CPU running at 1.8 GHz and 4 gigabytes of random access memory. The Microsoft Kinect SDK 2.0 is capable of extracting skeleton data at a rate of roughly 30 frames per second (fps). The results of the proposed skeleton augmentation algorithm are generated at a pace of 10 milliseconds (msec) each frame. Because of this, the final tracking was done at a rate of about 20 to 25 frames per second, which enabled the extraction of skeletal data in real time.



```
Hostage Backwards
Hostage Backwards
Hostage Backwards
Hostage Backwards
Hostage Backwards
```

Fig. 4.8 demonstrates a mock-up of calculating human body segments from the Kinect sensors on the PC machine by [87]

4.2.4.2 Experiment 4.2: Apple IOS devices - Body Detection (Smartphones)

The ARkit-3 framework, which has been made available to us, can be utilised in order to integrate actual people into augmented reality settings, which is one method for recording the movement of people. This brand-new function is not available on any other devices than those made with Apple A12 processor or later, and they must also be iOS-based. This innovative technique makes it possible to record all of a person's motions and then apply those recordings to the animation of a digital character, giving the character the ability to mimic the person's movements exactly. In the initial step of the process, an estimate of the pose of the person displayed on the screen is generated using machine learning technology. After that, that position is used to generate a fully-fledged skeleton with a very high level of fidelity. In the end, we give the user the completed character by combining this skeleton with a mesh that was provided by the user. Consequently, ARkit was utilised in the creation of the user interface. This technology is fully compatible with RealityKit[6], and as a result, we are able to control animated figures and render them on the screens. Machine learning is the engine that drives it, and Apple Neural Engine ensures a smooth experience.

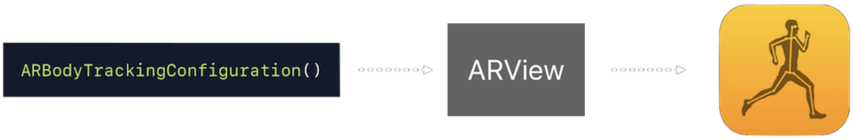


Fig. 4.9 Motion Capture in RealityKit [160]

Using this technology, we are able to extract data from both two-dimensional and three-dimensional objects. A use case diagram encompassing each stage of the process of extracting data from objects in 2D and 3D views is presented in the following graphics.

[6] Simulate and render 3D content for use in your augmented reality apps.

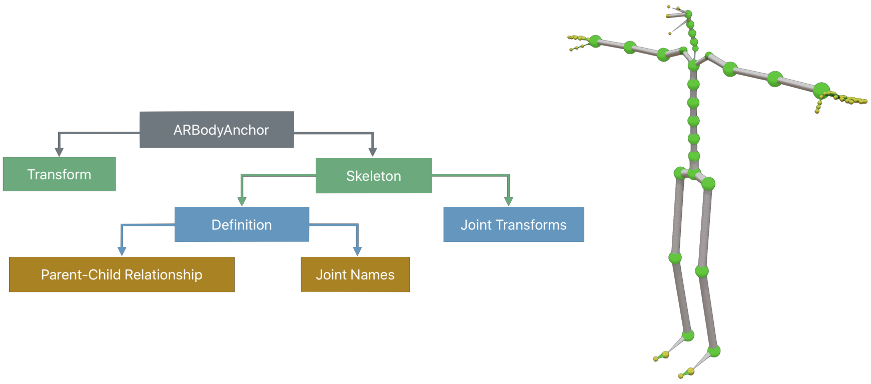


Fig. 4.10 Extracting data from 3D skeleton [160]

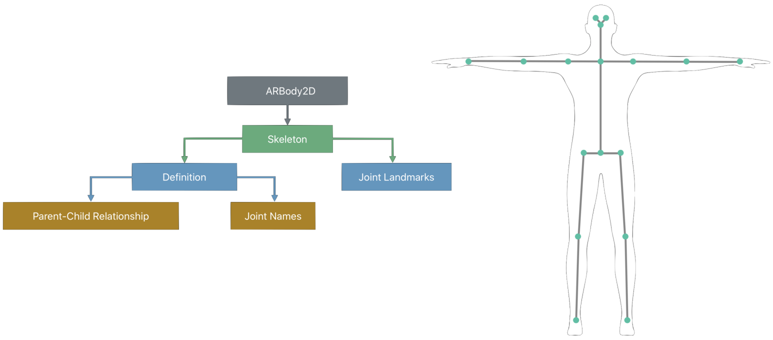


Fig. 4.11 Extracting data from 2D skeleton [160]

Figure 4.12 depicts a human body that was taken using the Body-Detection software that was offered by Apple developers to catch the human in motion. This app was used to take the picture that is seen in the figure. In order to gain further insight into the processes and procedures utilised by the application, we re-created it as part of our research and testing.

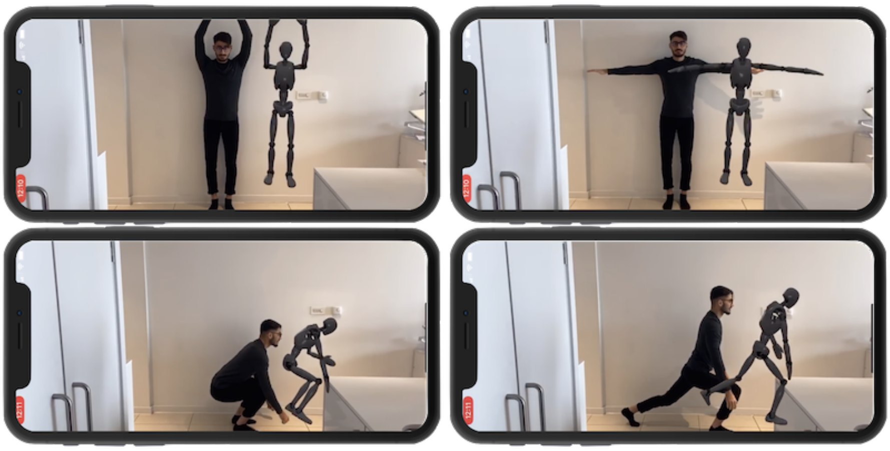


Fig. 4.12 demonstrates capturing and detecting human body in this experiment

4.2.5 Experiment 5: PIFuHD

PIFuHD is a new technology that was released by Facebook. It is capable of and does create 3D models of humans from one single photo. PIFuHD works by employing a multi-level architecture that is based on a deep neural network. This architecture enables users to create highly detailed 3D models using only a photograph. The process of automating picture digitisation is applicable to a variety of fields, including medical imaging and virtual reality, thanks to this technology. Additionally, consumers have the ability to put this technology through its paces on Google Colab. PIFuHD utilises a trainable end-to-end coarse-to-fine framework for an implicit surface, with a target image resolution of 1K. This allows for high-resolution 3D human reconstruction. This technology first uses image-to-image translation to obtain front and back perspective, and then it delivers the 3D figure in high-resolution by processing the input image through coarse-PIFu at a downsampled resolution so that it obtains global 3D structure at a low resolution. In this way, the technology is able to obtain global 3D structure at a lower resolution. The primary contribution of this article is the addition of finer-level details to this at a higher resolution, which, as a result, provides a reconstruction of the input image that is accurate down to the pixel level [145].

For optimal performance, the PIFuHD project must be run in a setting that contains at least 8 GPU[7]. The demonstration can also be executed on Google Colab for users of this project who do not have access to the environment described above. The process of putting PIFuhd through its paces on Google Colab shouldn't take more than five to ten minutes to complete. This technology can convert any photo into a 3D gaming avatar in a matter of minutes, making it one of the most cutting-edge features available in this sector. The end outputs are delivered to the consumers in three different file formats:.png,.obj, and.mp4 respectively [144].

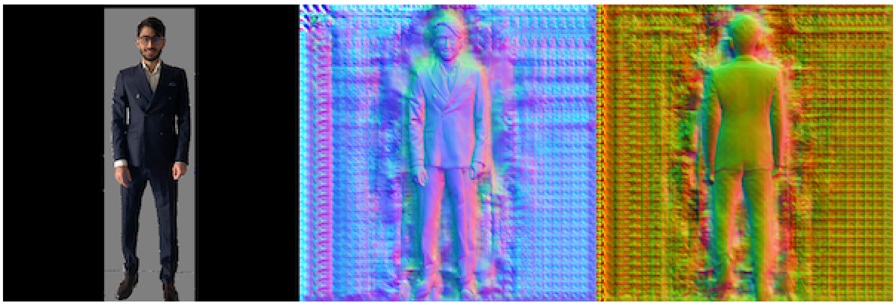


Fig. 4.13 demonstrates the .png results from Google Colab

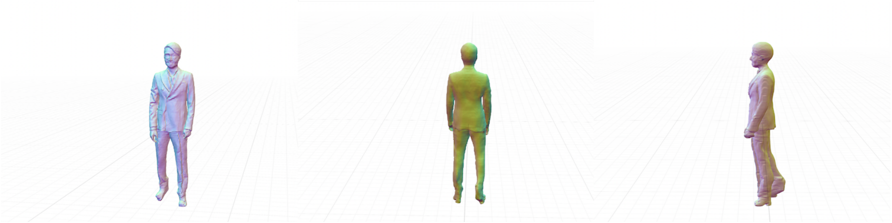


Fig. 4.14 demonstrates the .obj results from Google Colab

4.2.6 Experiment 6: AR and Deep Learning based Automatic Human Body Measurement System for babies

During the course of the experiments that we carried out for the purpose of this research study, we came across a project that aimed to provide a dependable answer for parents

[7] For the successful operation of the implemented algorithm requires a minimum GPU compute capability of 8.

to know the correct baby's body measurements in order to get the appropriate size diaper and clothing and to get a future size prediction with the assistance and support of AR and machine learning techniques [131]. The purpose of these experiments was to provide a reliable answer for parents to know the correct baby's body measurements in order to get the appropriate size diaper and clothing and to get. The author was able to determine the distance between two key locations on the body, which is an essential component of AR measuring, with the use of technology known as ARkit and/or ARCore. Pose estimation and edge keypoint recognition, both based on machine learning, are utilised in the process of locating keypoints on a server that is located on the other side of the network. These keypoints are then sent to the device that is located on the server side, where measurements are then made after they have been received. TensorFlow is going to be used for machine learning and to train models on the web server, as the author has made this decision because it is faster and easier to configure.

PoseNet, the engine that drives this programme, has the capability of estimating a single posture or many poses simultaneously. The single person posture detector is not only quicker and easier to use, but it also requires that the image contain only one subject. Because they only need to measure one body at a time, they only employ a single pose estimate for body measurements. This allows them to measure more accurately. On a broad scale, posture estimation can be broken down into two stages: First, A convolutional neural network is used to process an RGB image that serves as the input. Decoding poses, pose confidence scores, keypoint placements, and keypoint confidence scores from the model outputs can be done with either a single-pose or multi-pose decoding algorithm. Secondly, PoseNet will provide a pose object as its final output. This object will have a list of keypoints as well as an instance-level confidence score for each individual that was detected. A keypoint position is a pair of x and y coordinates in the original input image that have been determined to be the location of a keypoint.

[131] did not have access to an image dataset consisting of baby photographs, thus in order to validate the method, [131] used a dataset that consisted of pictures of the human body in general. In addition to this, we have evaluated their program based on the comprehensive human body dataset.

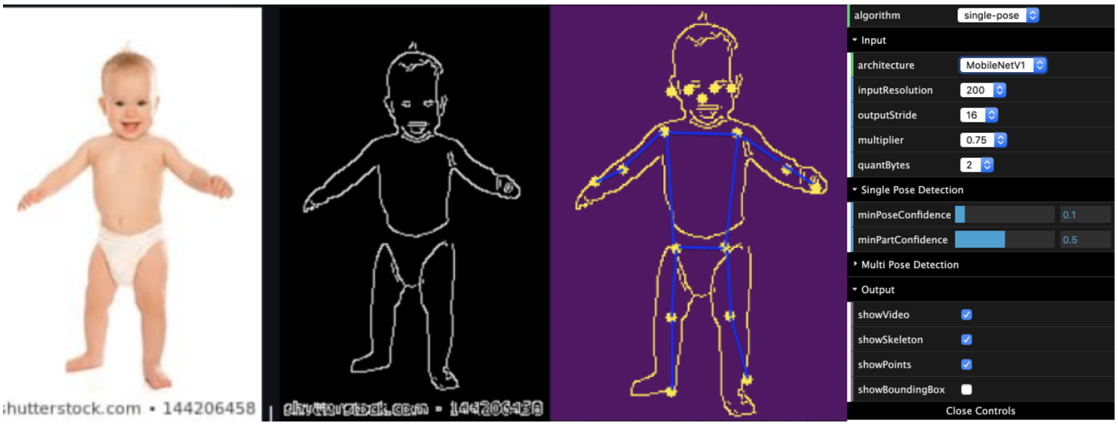
This piece of software was built on the foundation of six fundamental sections: architecture, outputStride, inputResolution, multiplier, quantBytes, and modelUrl.

- **Architecture**: both MobileNetV1 and ResNet50 are available to users as potential networks for use by the user. It decides which version of PoseNet's architecture should be loaded based on the current state of the network.
- **OutputStride**[8]: Users have the option of picking between 8, 16, and 32. (Stride 16, 32 are supported for the ResNet architecture and stride 8, 16, 32 are supported for the MobileNetV1 architecture). It outlines the stride that will be produced by the PoseNet model. The lower the value, the higher the output resolution, and the more precise the model will be; however, this will come at the expense of speed. If you want more speed but are willing to sacrifice some accuracy, increase this amount.
- **InputResolution**: A number or an object with the "width: number, height: number" type property set. The default value is 257. Before being input into the PoseNet model, the image will be scaled and padded to this size, which is specified by this parameter. When this value is increased, the model's accuracy improves, albeit at the expense of its speed. You can boost the speed of this at the expense of its accuracy by lowering this setting. If a number is specified, the image will be resized and padded such that it has the same width and height but is in the shape

[8] OutputStride is a feature in PoseNet that regulates the relationship between the resolution of the input image and the scale of the model's output. Lower values provide higher accuracy but slower processing, while higher values sacrifice some accuracy for faster performance.

of a square. If width and height are specified, the image will be scaled and padded to match the width and height values that have been provided.

- **Multiplier**: User's can choose any value between 1.01 and 1.0, 0.75 and 0.50. (The value is used only by the MobileNetV1 architecture and not by the ResNet architecture). It is the float multiplier for the depth, which is the number of channels, for each and every convolution operation. When this value is increased, the amount of the model's layers also increases, leading to a more accurate representation at the expense of speed. You can boost the speed of this at the expense of its accuracy by lowering this setting.
- **QuantBytes**: The bytes that are utilised for weight quantization are under the control of this argument.
- **ModelUrl**: A string that, if present, identifies the model's custom url. This parameter is optional. This is useful for local development or countries that don't have access to the model hosted on GCP.



algorithm single-pose
architecture MobileNetV1
inputResolution 200
outputStride 16
multiplier 0.75
quantBytes 2
minPoseConfidence 0.1
minPartConfidence 0.5
showVideo ☑
showSkeleton ☑
showPoints ☑
showBoundingBox □

Fig. 4.15 demonstrates the calculation of the baby's body posture step-by-step. image courtesy of: https://github.com/AI-Machine-Vision-Lab/body-measure

This research demonstrates a creative edge keypoint recognition based on the keypoints that are produced by the single pose estimate approach. These edge keypoints are going to be necessary in order to collect measurements of the baby's body. Using this procedure, the keypoints on the edge of the body are moved. They work with the image that has the contours detected. Because the location of the posture keypoint is already known, the nearest detected contour may be computed based on that information. The nearest contour is marked as one keypoint, and for another keypoint, a vector that is at an angle that is opposite to the current vector is indicated. The second keypoint will be the nearest contour that is detected on that vector. The length of the thigh will need to be measured using these two important areas.

Despite the fact that this software provides several features and fast single pose estimations, based on this research, the findings were not very promising, and numerous bugs were discovered while using this software. Also, this program cannot be considered a viable solution for measuring the human body.

4.2.7 Experiement 7: Body Measurements with Computer Vision

Another experiment that we carried out as part of these research studies involved determining how to calculate human body measurements with the help of computer vision, which was proposed by Faraz [65]. This piece of software requires only a single snapshot to evaluate the human body and generate a three-dimensional representation of the subject based on the evaluation results. After mapping a single input image onto a three-dimensional model, this piece of software then extracts body measures like waist, chest, and so on. Human Mesh Recovery (HMR) is utilised in order to complete the 3D reconstruction. TensorFlow version 1.13.1 was used for testing.

The Human Mesh Recovery (HMR) system is a framework that can reconstruct a whole 3D mesh of a human body using just a single RGB image of the subject. This framework generates a mesh representation that is richer and more usable than the bulk of the approaches that are currently in use, which compute 2D or 3D joint locations. The framework does this by parameterising the mesh with form and 3D joint angles. The major objective is to lessen the amount of keypoints that are lost due to reprojection. This will make it possible for the model to be trained utilising photographs taken in the wild with only ground-truth 2D annotations. However, reprojection loss by itself has a significant amount of room for for improvement. Training in HMR can be done with or without the assistance of paired 2D-to-3D supervision. Without resorting to an intermediate method of locating 2D keypoints, this method infers 3D location and shape characteristics directly from the picture's individual pixels. All models are executed in real time when provided with a bounding box that contains a person. This framework proves its methodology on various images taken in their natural environments, beats earlier optimization-based systems that build 3D meshes, and obtains competitive results on tasks such as 3D joint position prediction and component segmentation [91].

This software use a pre-trained COCO datasets [101] model to calculate the reconstruction of human body measurements. The authors were able to calculate the human body with the use of these datasets and the computer vision technologies. However, coco's pre-trained model was not developed specifically for their software's needs; hence, the result obtained is insufficient.

Despite the fact that this program is capable of measuring a significant number of human body sections (23 in total), the results are not very precise, with an average difference of 6.5 cm based on our testing. This indicates that the tool is not yet accurate enough to be used.

4.3 Summary: Experiments

According to the data obtained in this chapter, the existing fashion and entertainment applications will not be a solution to satisfy consumers' needs, mainly due to the lack of accuracy. Awkward positioning in front of the camera, as well as standing without moving for a certain amount of time, wearing tight clothes, adjusting the camera in the right place to cover all the body, all cause issues that do not encourage consumers to use this new technology. However, minimising the error and finding a way to make users more comfortable while using the applications and creating a user-friendly experience, such applications based on image processing have the highest potential in comparison to the other existing software and scanners. Therefore, these applications grab our attention because of the potential they have.

Also, results on the RMPE and Microsoft Kinect experiments show that such methods are a powerful way to capture pose estimation and human body in motion. However, these methods required powerful systems to compile data, which is not accessible for many users. Also the accuracy of the data is not good enough to satisfy retailers, compared to a 3D body scanner.

The Measuring Ruler application provided very accurate data in comparison to the other systems and apps by achieving less than 1 cm average errors. Various techniques have been combined in these applications to provide such accurate data, including deep learning, computer vision and 3D matching. By investigating more about such applications and adding or adapting machine learning and computer vision techniques, we can calculate human body information during activities.

Computer Vision and Machine Learning Approaches for Calculating the Human Body Measurements

"Machine learning is revolutionizing the fashion industry, providing new insights into body shape and size, body measurements, and fabric detection with unprecedented accuracy." - Nancy Wang, Forbes

In the following chapters, we utilised computer vision and machine learning methods, to estimate human upper body measurements and predict their body size from 2D images (front-side) taken from regular smartphones based on our findings in the previous chapter. To calculate human body measurements, in our approach, (i) once a body is detected (using a smartphone camera and a deep learning model), we then (ii) improve the image quality, (iii) use region of interest (ROI) to discard irrelevant or unwanted part of an image (iv) use calibration to determine the depth of field (distance from the camera focal point to the human body), (v) extract body features from the image, (vi) semi-automatically set a small number of markers, and (vii) by computing difference between markers, we estimate human upper body measurements.

The overall pipeline is illustrated in Figure 4.17. As part of the data capture process users can wear casual clothing. Main constraints are that they should wear clothing which is tight enough (does not mask) around the waistline, the shoulder line and the neck. Note that this is much less restrictive than many other proposed approaches which require to be naked or wearing only tight clothes over the entire body. In terms of poses, our system requires a T-pose/A-pose or relax-pose for the frontal view capture, while there are no restrictions for the side views (examples to follow). The T-pose/A-pose is only needed for situating markers for the shoulders. Again, this is more flexible than what can be found currently in available applications or published work using only a smartphone camera to provide (only) image data.

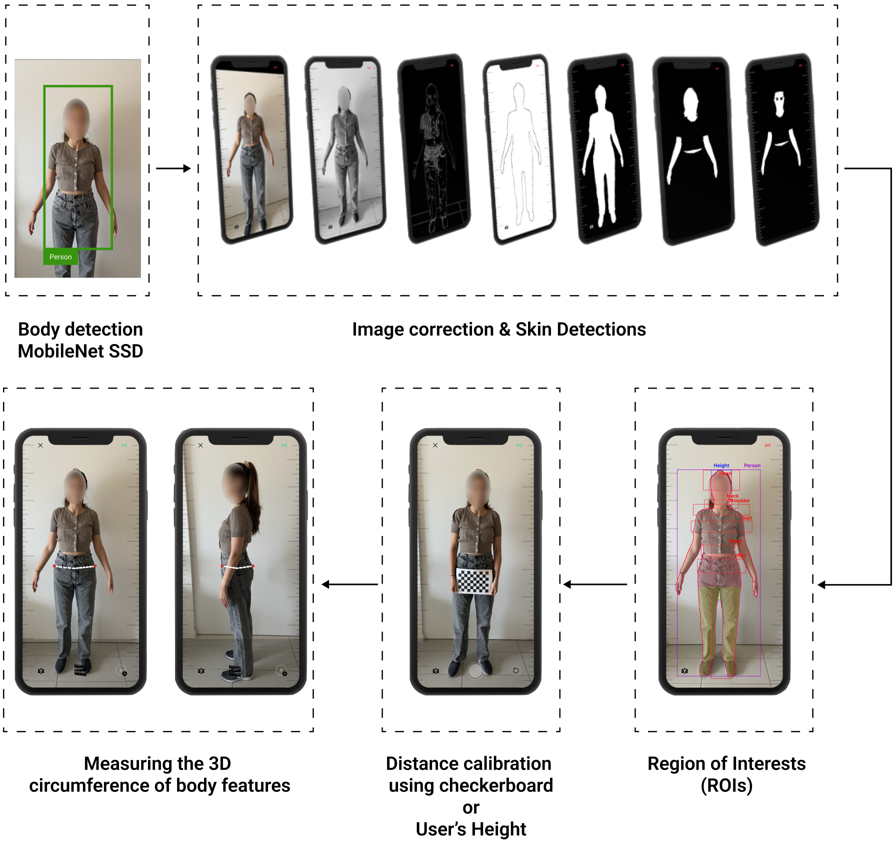


Fig. 4.16 Pipeline - The following flowchart display every stage to calculate the human upper body measurements

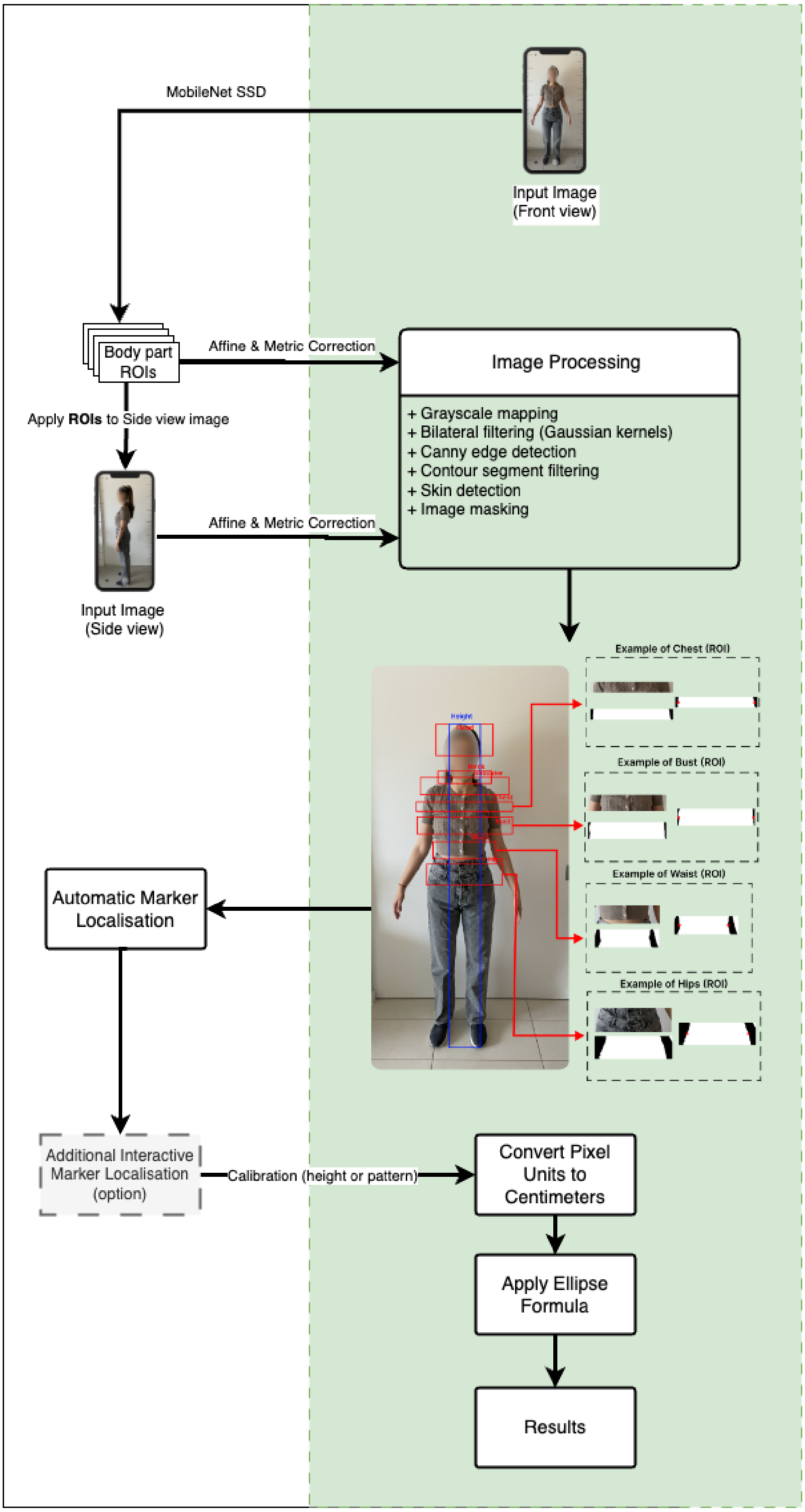

MobileNet SSD
Input Image (Front view)
Body part ROIs
Affine & Metric Correction
Image Processing
+ Grayscale mapping
+ Bilateral filtering (Gaussian kernels)
+ Canny edge detection
+ Contour segment filtering
+ Skin detection
+ Image masking
Apply ROIs to Side view image
Affine & Metric Correction
Input Image (Side view)
Example of Chest (ROI)
Example of Bust (ROI)
Example of Waist (ROI)
Example of Hips (ROI)
Automatic Marker Localisation
Additional Interactive Marker Localisation (option)
Calibration (height or pattern)
Convert Pixel Units to Centimeters
Apply Ellipse Formula
Results

Fig. 4.17 Proposed method as a diagram view of a pipeline of steps and processes

All images of participants' faces have been obscured to maintain anonymity in the following chapters. This is in line with our ethical approval commitment, which ensures the confidentiality of participant information in any published papers or thesis.

The following algorithms represents the process of our proposed software from the beginning to the end.

Algorithm 4.1: Semi Automatic - Human Upper Body Measurements

```
Result: Set of Measurements M, where $t^{th}$ row is the measurement around joint $j_t$
while Full Body Detected do
    if Region of Interests (ROIs) Detected then
        Crop each ROI - Isolate Human body from Surrounding Environments (ROI) ;
        Affine & Metrics Correction;
        Execute Image Processing for each ROIs;
        (i) Convert image to grayscale;
        (ii) Bilateral Filtering (Gaussian Kernels);
        (iii) Execute Canny Edge follow by findContours();
        (iv) Color Space Conversion (HSV & YCrCb) + thresholding;
        (v) Masking image ;
        Compute Mid-Point markers;
        Apply to Markers at the edge of each body contour;
        Optional - (Request user to Draw Markers);
        Get head point;
        Calibrate the camera using the provided height data;
        if User calibrated then
            Save calibrated pose;
            Call Calibration;
            while Calibration=Success do
                Call getDistance;
                for Joint label t = 1: T do
                    if $J_t$ is available then
                        Set automated point around body;
                        Request semi=auto marker from user - Optional;
                        Convert pixel unit to cm;
                        Compute intersection points;
                        Ellipse fitting to obtain $m_t$;
                        if Semi-Major axes is not 3 times longer than Semi-Minor axes then
                            Use Equation 7.2;
                            if $m_t$ is available for all T joints then
                                Save M;
                                Break;
                        if Semi-Major axes is 3 times longer than Semi-Minor axes then
                            Use Equation 7.3;
                            if $m_t$ is available for all T joints then
                                Save M;
                                Break;
Print Perimeter;
```

Algorithm 4.2: Manual - Human Upper Body Measurements

```
Result: Set of Measurements M, where $t^{th}$ row is the measurement around joint $j_t$
while Full Body Detected do
    if Region of Interests (ROIs) is NOT Detected then
        Affine & Metrics Correction;
        Execute Image Processing for Full Body;
        (i) Convert image to grayscale;
        (ii) Bilateral Filtering (Gaussian Kernels);
        (iii) Execute Canny Edge follow by findContours();
        (iv) Color Space Conversion (HSV & YCrCb) + thresholding;
        (v) Masking image ;
        Request user to draw markers;
        Get head point;
        Calibrate the camera using the provided height data;
        if User calibrated then
            Save calibrated pose;
            Call Calibration;
            while Calibration=Success do
                Call getDistance;
                for Joint label t = 1: T do
                    if $J_t$ is available then
                        Convert pixel unit to cm;
                        Compute intersection points;
                        Ellipse fitting to obtain $m_t$;
                        if Semi-Major axes is not 3 times longer than Semi-Minor
                          axes then
                            Use Equation 7.2;
                            if $m_t$ is available for all T joints then
                                Save M;
                                Break;
                        if Semi-Major axes is 3 times longer than Semi-Minor axes
                          then
                            Use Equation 7.3;
                            if $m_t$ is available for all T joints then
                                Save M;
                                Break;
Print Perimeter;
```

Chapter 5

Body Detection using MobileNet SSD

This chapter provides an overview of object detection algorithms for fully automatic human body detection from a single image, with a focus on the MobileNet SSD method. The chapter starts with a brief history of object detection, followed by a discussion of machine learning and computer vision, object classification and localisation, challenges in object detection, and popular datasets and characteristics. The chapter then compares several object detection techniques, including Haar-Cascade, Histogram of Oriented Gradients, MobileNet SSD, YOLO (You Only Look Once, and EfficientNet and presents findings on their effectiveness. The inclusion of these object detectors in our comparison was motivated by several factors. Firstly, previous research has utilised some of these techniques, such as Haar Cascade, which was employed in Ashmawi et al. [31]'s paper. Additionally, these detectors were chosen due to their relevance, high accuracy, and being the latest techniques available.

The chapter proceeds to explain the region of interest and bounding box extraction methods used to detect the human body from an image. The proposed method for this chapter is MobileNet SSD, which is chosen based on the datasets used and the findings from the comparison of object detection techniques. The chapter then discusses the identification of body section names, such as head, chest, bust, waist, hips, and height, as well as the differentiation of upper and lower human body.

Finally, the chapter includes a discussion on the results of the proposed method and its potential limitations. It concludes with a summary of the chapter's findings and implications for the development of the proposed software solution. Overall, this chapter provides a comprehensive overview of object detection algorithms and their application to human body detection, with a focus on the MobileNet SSD method.

5.1 A brief history of object detection

Object detection is a computer vision task that involves detecting an object in an image or a video. Many Machine Learning (ML) and Deep Learning (DL) models are used to improve the performance of object detection and associated tasks. Historically, two-stage object detectors were extremely popular and effective. In compared to the majority of two-stage object detectors, single-stage object detection and its underlying algorithms have substantially improved with recent advancements. Object detection is the process of locating an object in an image or video, and outlining its boundaries. It may also include the identification of the object, such as the type of object or its class.

Object detection stands as a foundational challenge in the field of computer vision. It functions as a foundation for a variety of other computer vision tasks, including instance and image segmentation, image captioning, and object tracking, among others. Object detection applications cover a broad range of tasks, including identifying pedestrians, detecting animals, recognising vehicles, counting people, detecting features, extracting text, and estimating poses. It is also used in areas such as medical imaging and fashion industry. In the fashion industry, object detection is used to detect body parts in images of people wearing clothes. This can help designers to understand how the clothes fit the body, and how they look on different body types. Object detection algorithms have come a long way since their early days, and they are continuing to improve.

The history of object detection can be traced back to the early days of computer vision research. Early object detection algorithms used hand-crafted features to detect objects in images. These algorithms relied on manual feature extraction and were limited in their accuracy.

In the late 2000s, deep learning algorithms began to be used for object detection. These algorithms use CNN to automatically extract features from images and videos. This allowed for more accurate object detection, as the CNNs are able to learn the features that are important for detecting objects.

One of the significant advantages of object detectors is their remarkable versatility and adaptability. These detectors can be programmed to perform a wide variety of responsibilities and accommodate specialised applications. Through the automated recognition of objects, individuals, and scenes, it is possible to extract valuable information, thereby facilitating the automation of tasks at various phases of business value chains. This capability enables businesses to streamline processes such as enumeration, inspection, and verification, resulting in greater productivity and efficiency.

However, a significant disadvantage of object detectors is their high computational requirements and substantial processing power requirements. This can present difficulties, especially when deploying large-scale object detection models. The operational costs associated with operating these models can accumulate rapidly, posing a threat to the economic viability of business use cases. The need for extensive processing resources and the attendant costs may prevent organisations from adopting object detection technology, particularly in circumstances where cost-effectiveness is of paramount importance.

5.1.1 Machine Learning and computer vision

Body detection for the fashion industry is the process of using computer vision and machine learning algorithms to detect the shape, size and proportions of a person's body. This data can then be used to make recommendations for clothing and other fashion items that would fit and look good on them.

Computer vision is a field of artificial intelligence that deals with analysing and understanding images and videos. It's used to identify objects, identify patterns, and even reconstruct 3D models of the world. Computer vision algorithms are used to detect and analyse the human body in order to extract meaningful information such as body shape, size, and proportions.

Machine learning is a subset of artificial intelligence that focuses on the development of computer programs that can learn from data and make decisions with minimal human intervention. Machine learning algorithms can be used to identify patterns in data and make predictions about future outcomes. In the fashion industry, machine learning algorithms are used to analyse customer data and make recommendations for clothing that would fit and look good on them.

The use of computer vision and machine learning for body detection in the fashion industry has a number of advantages. First, it can reduce the amount of time and effort required to measure someone's body in order to determine their size and shape. Second, it can provide more accurate measurements than manual methods, resulting in better fitting clothes. Third, it can provide personalised recommendations to customers based on their body shape, size and proportions. Finally, it can help to reduce returns and improve customer satisfaction by providing accurate size information.

In order to use computer vision and machine learning for body detection, a number of steps need to be taken. First, a data set of images and videos of people of all body shapes and sizes needs to be collected. This data can then be used to train the computer vision and machine learning algorithms. Once the algorithms are trained, they can then be used to detect the shape, size and proportions of a person's body from a single image or video.

The use of computer vision and machine learning for body detection in the fashion industry is still in its early stages. However, it has the potential to revolutionise the way people shop for clothes and make recommendations for clothing that would fit and look good on them. With continued development and refinement of computer vision and machine learning algorithms, body detection could soon become an integral part of the fashion industry.

This chapter focused on object detection and its subfields, such as object localisation and segmentation, as one of the most essential and widely used tasks in computer vision. Convolution Neural Network (CNN), Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM), Restricted Boltzmann Machines (RBM), and Stacked Autoencoders are examples of standard deep learning models that can be used to any computer vision job [168].

5.1.2 Object classification and localisation

Object classification and localisation is a method of computer vision used to identify and localise objects in an image or video. It is a type of supervised learning, in which algorithms are trained to recognize objects from a given set of images. The algorithm learns to recognize an object based on its features, such as colour, shape, texture, size, etc.

Image Classification is the classification of an image or object within an image into one of the established categories. Typically, this problem is tackled using supervised machine learning or deep learning techniques in which the model is trained using a large labelled dataset. ANN, SVM, Decision trees, and KNN are a few of the most often used machine learning models for this job [165]. CNNs and its architectural successors and modifications dominate other deep models for categorising photos and related works on the deep learning side. In addition to well-defined machine learning and deep learning models, additional approaches such as fuzzy logic and genetic algorithms are also used for the aforementioned objectives [67].

Object localisation is the process of determining the position of a single object or numerous objects in an image or frame using a rectangular box known as a bounding box. Image segmentation is the process of dividing an image into several segments, with each segment including either an entire object or a portion of an object. Image segmentation is often used to identify the objects, lines, and curves that define the segment or object boundaries in an image. In general, pixels inside a segment share a set of attributes, such as brightness, texture, etc. The primary objective of image segmentation is to provide a meaningful representation of the image. Object detection is also a combination of categorisation, localisation, and segmentation. It is the task of accurately classifying and effectively localising single or many objects in a picture, typically using supervised algorithms with a suitably big labelled training set. In the context of object detection, Figure 5.1 provides a clear grasp of classification, localisation, and segmentation for single and numerous items in an image.

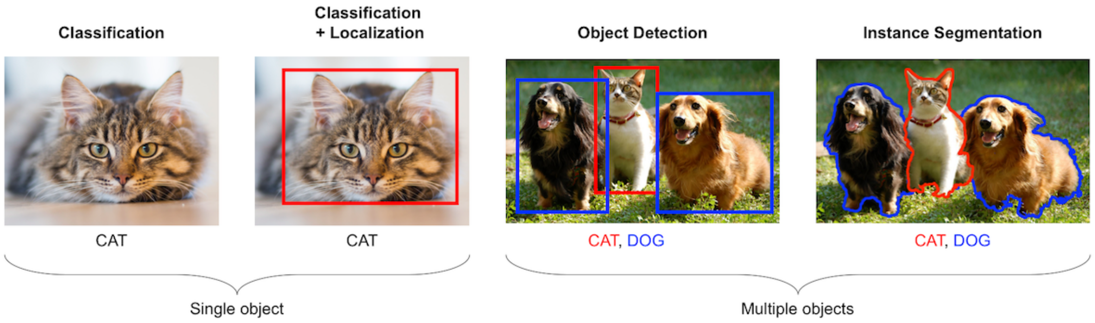


Fig. 5.1 Detection or localisation and segmentation for single and multiple objects image from [135]

In the fashion industry, object classification and localisation can be used to identify body parts for body detection. Body detection is the process of detecting human body parts in images or videos. It is used to improve the accuracy and speed of clothing selection for online shopping. By recognizing body parts, the system can identify the best-fitting clothes for a person.

Body detection is an important task for the fashion industry, as it can help reduce the time and effort required for customers to find clothes that fit them. It can also reduce the chances of customers buying clothes that do not fit them.

Body detection can be used in combination with object classification and localisation to identify body parts in images. The system can learn to recognize body parts by training on a set of images with labeled body parts. The algorithm can then be used to detect body parts in images or videos. This can be used to accurately identify the size and shape of the body, which can then be used to recommend the best-fitting clothes.

Object classification and localisation can also be used to detect the pose of a person in an image. This can be used to recommend clothes that are suitable for different activities, such as running or yoga.

In overall, object classification and localisation is an important tool for the fashion industry. It can be used to accurately detect body parts, identify the pose of a person, and recommend the best-fitting clothes. This can help improve the accuracy and speed of online shopping, and reduce the chances of customers buying clothes that do not fit them.

5.1.3 Challenges in object detection

The primary challenges in object detection include (i) the occupancy of an object in an image has an inherent variation, such that objects in an image may occupy the majority of the pixels (70–80%) or very few pixels (10–20%), (ii) processing of low-resolution visual contents, (iii) dealing with multiple objects of varying sizes within a single image, (iv) Availability to labelled data, and (v) dealing with overlapping objects in visual content.

Object detection is a challenging task for computer vision and machine learning applications, particularly in the fashion industry. As the fashion industry is becoming increasingly digitized, object detection algorithms are being used to help identify clothing items in images and videos. Object detection involves using an algorithm to detect and classify objects in an image or video, such as clothing items, people, and objects in a scene.

The main challenges in object detection involve recognizing objects in an image or video and correctly classifying them. This requires the algorithm to be able to accurately identify objects and differentiate them from other objects in the image or video. Additionally, object detection algorithms must be able to deal with varying levels of occlusion, scale, and pose. For example, in the fashion industry, an algorithm must be able to detect and identify clothing items regardless of whether the item is partially or fully occluded, or at different angles or distances.

Another major challenge in object detection is dealing with different lighting conditions. When dealing with images or videos from the fashion industry, the lighting conditions can vary greatly depending on the environment. For example, an image/scene that been shot outdoor/indoor will have different lighting conditions. Additionally, the lighting conditions can also vary based on the time of day, the weather outside, and the type of lighting used. An object detection algorithm must be able to detect and classify clothing items, human body segments regardless of the lighting conditions.

Finally, object detection algorithms must be able to deal with a variety of different body types. In the fashion industry, clothing items are designed to fit a variety of body types, and an object detection algorithm must be able to accurately detect clothing items on different body types. This requires the algorithm to be able to recognize and differentiate different body types, as well as accurately detect clothing items on those body types.

The following is a summary of the problems that most object detectors have when trying to solve them using deep learning methods.

- **Multi-scale training**: The majority of object detectors are trained for a particular resolution of input. In general, these detectors underperform for inputs with varying scales or resolutions.
- **Foreground-background class imbalance**: Imbalance or disproportion among instances of different categories might have a significant impact on the performance of the model.
- **Detection of relatively smaller objects**: If the model is trained on larger items, all object detection techniques will tend to perform well when detecting larger things. Unfortunately, these models perform poorly with respect to smaller-sized objects.
- **Necessity of large datasets and computational power**: Deep learning object detection techniques require larger datasets for computing, labor-intensive approaches for annotations, and significant computer resources for processing [102]. Owing to the exponential growth of generated data from multiple sources, annotating each and every object in the visual contents has become a laborious and time-consuming operation [102, 169, 187].
- **Smaller sized datasets**: Deep learning models display poorer performance when evaluating on datasets with fewer occurrences, despite the fact that they outperform classical machine learning algorithms by a significant margin and have a significant advantage in this regard.
- **Inaccurate localisation during predictions**: Bounding boxes are approximations of the actual ground conditions. Typically, background pixels are also incorporated

in algorithmic forecasts, which reduces its accuracy. localisation errors are typically caused by the presence of background in the predictions or the detection of comparable items.

5.1.4 Popular dataset and characteristics

MSCOCO[101] (Microsoft Common Objects in Context) and Pascal VOC (Visual Object Classes) are two of the most commonly used datasets for object detection tasks. MSCOCO is a large-scale dataset that contains over 330,000 images with more than 2 million labeled object instances, while Pascal VOC is a smaller dataset consisting of roughly 10,000 images with 20 object classes. Both datasets have been widely used for benchmarking object detection algorithms and for developing new approaches for object detection.

MSCOCO is a well-known and widely used dataset for object detection tasks. It contains a large number of images from different domains, including both indoor and outdoor scenes. The images are annotated with a variety of objects such as people, animals, vehicles, and furniture. Additionally, MSCOCO contains a large amount of semantic information, such as scene categories and captions. This makes it ideal for training deep learning models that require a large amount of data. Furthermore, MSCOCO contains a variety of tasks, such as object detection, instance segmentation, and caption generation, which makes it suitable for a wide range of applications.

Pascal VOC is another commonly used dataset for object detection tasks. It consists of 10,000 images with 20 different object classes. It contains both indoor and outdoor scenes, with objects such as cars, animals, and furniture. Unlike MSCOCO, Pascal VOC does not contain any semantic information. However, it is smaller in size and easier to use for training deep learning models. Also, it contains a variety of tasks, such as object detection, semantic segmentation, and image classification.

In terms of advantages and disadvantages, MSCOCO is larger in size and contains more semantic information, making it ideal for training deep learning models that require a large amount of data. However, it is more difficult to use and can be time-consuming to annotate. On the other hand, Pascal VOC is smaller in size and easier to use. However, it does not contain any semantic information, making it less suitable for deep learning tasks.

In conclusion, both MSCOCO and Pascal VOC are two of the most commonly used datasets for object detection tasks. MSCOCO is larger in size and contains more semantic information, making it ideal for training deep learning models. However, it is more difficult to use and can be time-consuming to annotate. On the other hand, Pascal VOC is smaller in size and easier to use, but does not contain any semantic information.

5.1.5 Object detection algorithms

Recent object detection techniques can be divided broadly into two types *two stage object detectors* and *Single stage object detector*. In the former, the first stage is responsible for creating Regions of Interest (RoI) using the Region Proposal Network (RPN), while the second stage predicts the items and bounding boxes for the proposed regions. In this section, we will explore both object detectors, as well as their applications in diverse sectors. Figure 5.2 displays different example of single and two stage object detectors.

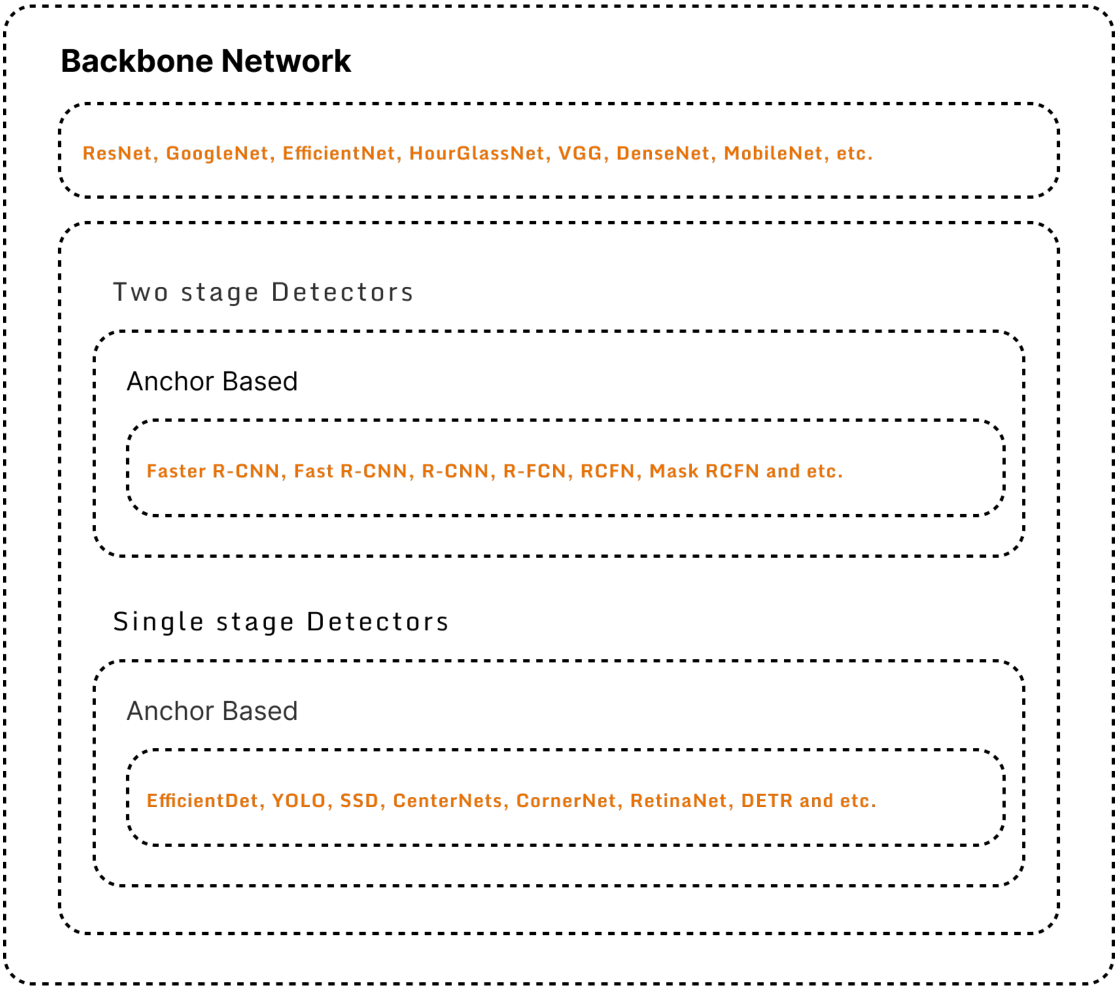

Backbone Network
ResNet, GoogleNet, EfficientNet, HourGlassNet, VGG, DenseNet, MobileNet, etc.
Two stage Detectors
Anchor Based
Faster R-CNN, Fast R-CNN, R-CNN, R-FCN, RCFN, Mask RCFN and etc.
Single stage Detectors
Anchor Based
EfficientDet, YOLO, SSD, CenterNets, CornerNet, RetinaNet, DETR and etc.

Fig. 5.2 Different type of object detetion algorithms

Single stage detectors, while faster, typically sacrifice some accuracy in comparison to two-stage detectors.

5.1.5.1 Two Stage Detector

In the case of two-stage detection, the first stage makes use of regional design networks to generate high-probability regions of interest. Object detection is the second stage, and it is responsible for the final bounding box regression and classification. Among the many two-stage detectors are RCNN, Fast RCNN, SPPNET, Faster RCNN, etc.

5.1.5.2 Single Stage Detector

Where object detection is a simple regression problem that learns probability classes and bounding box coordinates from incoming data. EfficientDet, YOLO, SSD, and RetinaNet, among others, fall under the same phase detector. Object detection is an advanced kind of imaging classification in which a neural network predicts the presence of objects in an image and highlights them via bounding boxes.

5.2 Comparison of the object detection techniques

Now that we have a better understanding of how object detections function, let's examine how we can detect human body segments using existing technologies. Detecting a human body in an image or video can be challenging since humans can perform a number of poses, each with a different shape.

In order to specify the properties of the upper human body, we studied numerous object detection and segmentation techniques. Table 5.1 compares the available object detection and segmentation techniques. Five relevant machine learning algorithms (including Haar-Cascade Classifier, HOG, MobileNet SSD, YOLO, and EfficientNet) were selected to detect the human torso, and they will be examined and studied in greater depth in this section.

5.2.1 Haar-Cascade

Haar Cascade was the first approach that we researched in OpenCV[1]. Haar Cascade uses machine learning to accurately identify every object on which it has been trained. It operates by training using a huge number of images that are positive examples of whatever you are attempting to detect, as well as a large number of images that are negative examples of whatever you are attempting to detect. Then, it creates Haar

[1] OpenCV (Open Source Computer Vision Library: http://opencv.org) is an open-source BSD-licensed library that includes several hundreds of computer vision algorithms

Table 5.1 Comparison of the object detection and segmentation

Object detection and segmentation
RCNN family: RCNN, Fast RCNN, Mask RCNN 2015-18 *multiple iterations over an image* robust, but slow
YOLO family: YOLO, YOLO 9000, YOLOv3, YOLOv4, YOLO-LITE 2015-20 *single shot for multiple objects* robust and fast
CrowdDet 2020 *one proposal, multiple predictions* earth mover's distance and Set NMS
EfficientDet 2020 *bidirectonal feature pyramid* uniform scaling of resolution, depth and width of backbone, feature network and prediction networks
UniverseNET 2021 *built on RetinaNet baseline* detects both small and large objects, both in and out of natural image domains
YOLOR 2021 *combining implicit and explicit knowledge* multi-task learning
MobileNet V3 2019 *tuned to mobile CPUs using hardware-aware network architecture search* NetAdapt algorithm
HRNet-OCR 2020 *attention-based approach to combining multi-scale predictions* hierarchical attention mechanism – memory efficiency
CDCL+Pascal 2020 *combine real and synthetic data* use skeleton representation of human bodies to bridge the gap between real and synthetic domains

Features that are an ideal match for the positive set. In order for the Haar Cascade method to recognise the torso of a human body from an image, the following algorithm steps must be executed in order: (1) Calculating Haar Features (2) Creating Integral Images (3) Using Adaboost (4) Implementing Cascading Classifiers.

The algorithm will examine each rectangular portion of the image in an effort to identify any (Haar features) that it identifies as being consistent with a human body.

$$f_i = \begin{cases} +1 & v_i \geq t_1 \\ -1 & v_i \leq t_1 \end{cases} \tag{5.1}$$

In this context, "v_i" is the response of a weak classifier to a specific Haar-like feature in an image's region of interest, showing how well the feature aligns with the desired pattern. The threshold "t_1" determines the presence of this feature. If "v_i" is greater than or equals "t_1", the classifier deems the feature present, outputting "+1". If it's less than or equal, it outputs "-1", indicating the feature's absence. The Haar cascade method amalgamates numerous weak classifiers, each associated with distinct Haar-like features and thresholds. This combination creates a powerful classifier. By applying these weak classifiers in a cascade manner, the technique effectively and precisely detects intricate objects in images.

Different kinds of Haar-like features are illustrated in Figure 5.3. When dealing with large images, it may be more difficult to identify these features. Therefore, integral images play a significant role in this context, which helps to reduce the total number of operations. Integral images is an algorithm for quickly and efficiently computing the sum of values in a rectangle subset of a grid. It is also known as a summed area table, which is another name for this type of table.

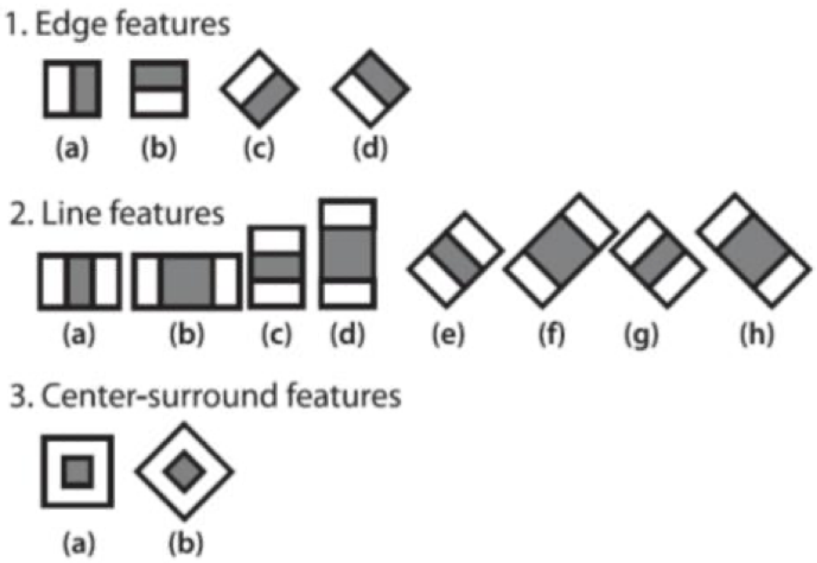


Fig. 5.3 Types of Haar-like features [86]

In these features, the colours white and black don't literally refer to colours in an image but instead represent regions of relative intensity. The white region signifies a positive weight while the black indicates a negative weight. In essence, the Haar-like feature calculations involve subtracting the sum of pixel intensities in black regions from the sum in white regions. This difference can highlight contrasts in intensity patterns, enabling the detector to recognize specific structural features in images, like edges or contrasting regions. These contrasts are pivotal for object recognition processes, such as detecting facial structures.

Instead of computing each pixel individually, it creates sub-rectangles and array references for each sub-rectangle to speed up the process. This is done to save time. Using these, the Haar features are subsequently reconstructed.

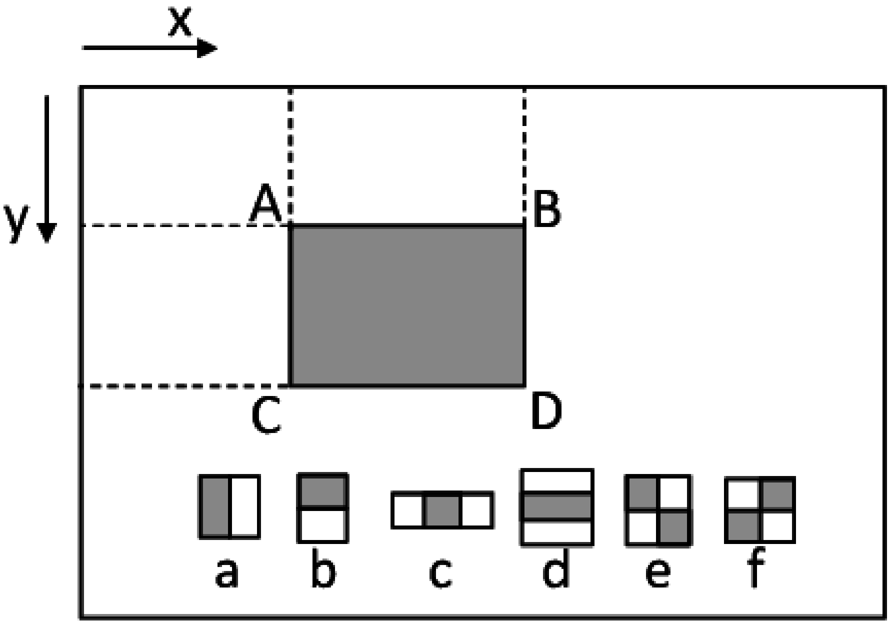


Fig. 5.4 Illustration of the integral image and Haar-like rectangle features (a-f) [186].

The integral image and Haar-like rectangle features are foundational in rapid object detection, as used in the Viola-Jones face detection algorithm. When you see an illustration with points A, B, C, and D, these typically represent the four corners of a rectangular region in the image. The concept of an integral image is to simplify the calculation of the sum of pixel values within a rectangular region. For any point (x,y), the value at the integral image gives the sum of all pixel values above and to the left of that point, inclusive.

For the Haar-like features (a-f): (a) represent vertical edges, (b) detect horizontal edges, (c) recognizes vertical lines, (d) recognizes horizontal lines, (e) A 2x2 checker-board pattern which is sensitive to diagonal features, and (f) it is the inverse of "e".

When it comes to object detection, virtually all of the Haar features will be rendered irrelevant because the only features that matter are those that are possessed by the object itself. Consequently, this is where the AdaBoost (AD) come into play.

The AD is a method that combines a number of hypotheses that are considered to be "weak classifiers" in order to come up with highly accurate hypotheses that are considered to be "strong classifiers." The best features are selected by the AD learning algorithm, and the classifiers are then trained to use these features. The process of creating weak learners involves moving a window across an input image and calculating Haar features for each subsection of the image. This results in the creation of weak learners. A learned threshold is used to differentiate between non-objects and objects, and this difference is compared to that threshold. Because these are "weak classifiers," a significant number of Haar features' attributes are necessary to achieve the desired level of accuracy when constructing a robust classifier.

In conclusion, the cascade classifiers turn the weak learners into strong learners by combining them. Boosting is a method of training weak learners that enables the construction of a highly accurate classifier based on the average prediction of all of the weak learners. As a result, the classifier will either decide to mark the object as known (a positive decision) or not indicate the object as known (a negative decision), after which it will proceed to the next stage. Since the bulk of the windows do not depict the human torso, the negative samples were created to move on as quickly as possible in order to hasten the procedure. This was done in order to save time (upper human body).

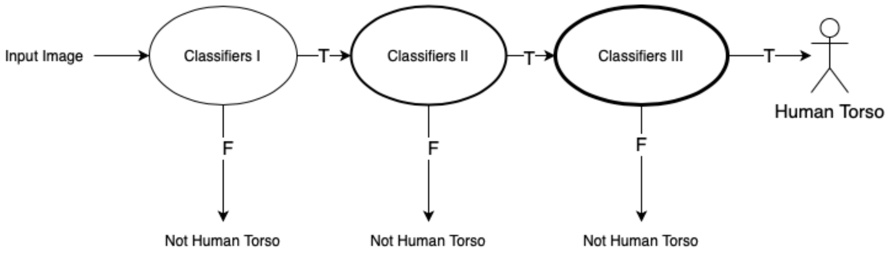


Fig. 5.5 Cascade classifier [133]

If the user's legs or torso were not entirely in frame, we would be able to tell thanks to Haar Cascade's ability to differentiate between upper and lower bodies as well as full bodies. This would allow us to direct the user to move backwards or adjust their phone to the appropriate angles. Following a significant amount of testing carried out in a variety of settings, we came to the conclusion that we needed to adopt a new strategy. Haar Cascade was capable of locating bodies, but the results were not very reliable. Based on the findings, it was determined that the precision of the Haar approaches was only 0.52. For further details, please refer to Table 5.4. In addition to the challenges we encountered when attempting to apply the algorithms to side views, it was necessary to have adequate lighting and an extremely clean background.

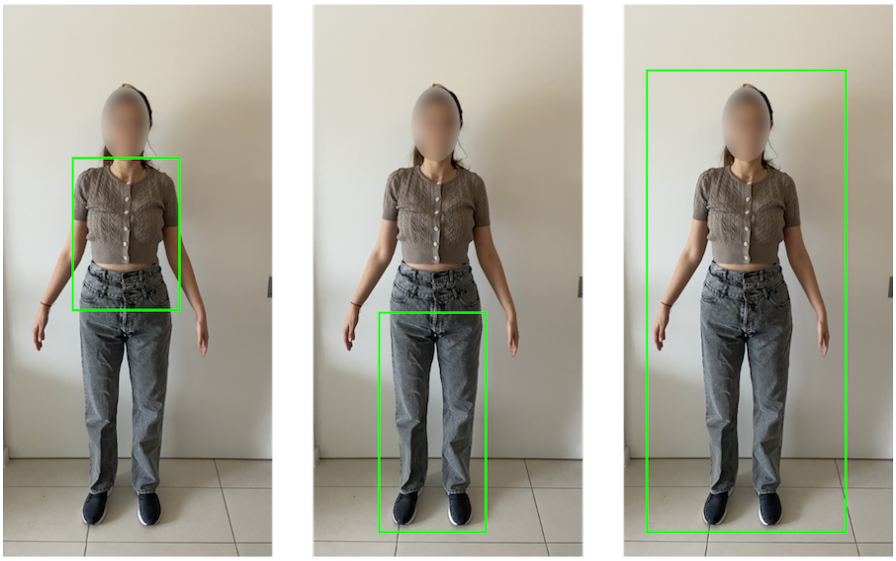


Fig. 5.6 A demonstration of the haar-cascade tegnique: Upper detector, lower detector and full detector from front image

5.2.2 Histogram of Oriented gradients (HOG)

The HOG method operates in a manner that is similar to that of the Haar method; however, rather than detecting blocks of dark and light, it identifies the angles of gradients. The HOG is designed to extract features into a vector, which is then fed into a classification algorithm such as a support vector machine (SVM), which determines whether or not a human body (or any other object that the software has been trained to recognise) is present in a certain region.

The HOG algorithm to detect human body can be explained in the following four stages: (1) Compute the Gradient Images (2) Compute the HOG (3) Block normalisation (4) Feeding the vector into SVM.

Firstly, the algorithm needs to compute the (x, y) gradients of the image, by filtering the image with the kernels. The kernel used for computing gradients in the HOG algorithm consists of two filters: the horizontal filter [-1, 0, 1] and the vertical filter [-1, 0, 1]. Each value in the filter represents a weight or coefficient. The "-1" and "1" coefficients emphasize horizontal and vertical edges, respectively, when the kernel is applied to an image region with intensity transitions. The "0" coefficient has no effect on the convolution operation and acts as a neutral element. Convolution involves multiplying the pixel intensities with the corresponding weights and summing up the results, highlighting specific patterns in the image. These gradients are vital for computing gradient magnitudes and orientations in the subsequent stages of the HOG algorithm.

It is important that the input image have a fixed aspect ration in HOG. The gradient image will remove a lot of non-essential information such as coloured background, and in return will highlights outlines. The image still can be identified and at each pixel, the gradients has a magnitude and a direction.

After computing the HOG, the image is divided into 8x8 cells to provide a compact representation and make the HOG more noise-resistant. The use of 8x8 cells in the HOG computation instead of 16x16 enhances the algorithm's sensitivity to local image details and helps capture finer-grained gradient information, resulting in improved object detection performance. Smaller cells allow for a more fine-grained representation of gradient directions, contributing to the HOG's noise resistance and discriminative power in object recognition tasks. The HOG will thereafter be computed for each of these cells. To determine the gradient direction inside a region, we simply produce a histogram of 64 (8x8) values. Each bin corresponds to a gradient direction (0-180 degrees, 9 bins per cell) (20 degree for each bin). Consequently, this will decrease the values of the 64 vectors to only 9 values.

A block with dimensions of 16 by 16 can be added to the image in order to normalise it and make it independent of the illumination. By reducing the value of each pixel by a factor of two (16x16), we are making the image darker; thus, the magnitude of the gradient will change by half, and the values of the histogram will also change by half. Block normalisation refers to this particular procedure.

In conclusion, a block with dimensions of 16 by 16 contains four histograms, each of which can be chained together to produce a vector with 36 elements and one element vector. After that, the window is shifted to the left by 8 pixels, a normalised 36x1 vector is produced over this window, and the operation is carried out once again. This approach can train a soft SVM and make predictions about the human torso (or any other object been trained for).

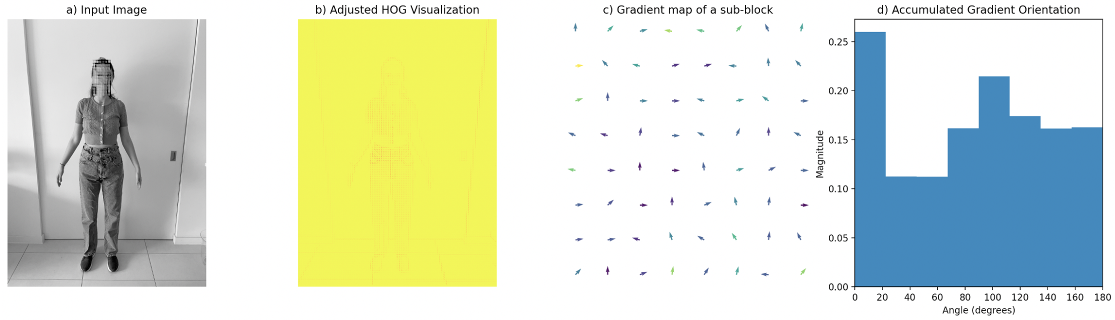


Fig. 5.7 A demonstration of the HOG feature extraction method: a) the input image; b) HOG visualisation emphasizing structural patterns; c) Gradient map of a sub-block; d) Accumulated Gradient Orientation

Gradient map of a sub-block: A detailed view of a small part (sub-block) of the input image, showing the gradient magnitude and direction. The arrows (quivers) point in the direction of the gradient and their colour and length represent the magnitude of the gradient. This visualisation gives an idea of how gradient information is extracted from the image at a micro level.

Accumulated Gradient Orientation: A histogram showing how often specific gradient orientations (angles) occur in the selected sub-block, weighted by their magnitudes. This gives an aggregate view of the dominant directions of edges or transitions within the sub-block.

Despite the fact that we used the Haar technique, HOG was able to function correctly once we implemented the code, even in poorly light conditions, and it had no difficulty identifying textured backgrounds. However, the human subject must be located within a "perfect" area on the screen for the HOG technique to work well. If you get too close (less than 0.5m), it won't be able to see the person. When the distance is too far (more than 1.5m), it will not detect the human. In addition, one of the most major disadvantages of utilising this approach is the fact that it will need a significant amount of time in order to be operational. Based on the findings, it was determined that the Inference time of the HOG approaches was 38.3. For further details, please refer to Table 5.4.

5.2.3 MobileNet SSD

MobilenetSSD is a model for object detection that computes the object's bounding box and category from an input image. This Single Shot Detector (SSD) object detection approach utilises Mobilenet as its backbone to deliver mobile device-optimized, rapid object detection. It is a lightweight, low power convolutional neural network (CNN) designed for efficient object detection. It is based on a combination of depthwise separable convolutions and pointwise convolutions. This makes it much faster and more efficient than regular CNNs.

The goal of MobileNet SSD is to provide a fast, accurate and efficient object detection model that can be deployed on mobile devices. It was developed by Google and has been used in several mobile applications such as Google Photos, Google Play Store and Google Maps.

The MobileNet SSD architecture is based on a series of convolutional layers, which are used to extract features from the input image. This layer is followed by a series of depthwise separable convolutions, which are used to reduce the computational complexity of the network. The output of this series of convolutions is then passed through a 1x1 pointwise convolution layer, which is used to reduce the number of parameters. The output from the pointwise convolution layer is then fed into a series of fully connected layers, which are used to classify the objects in the image.

One of the main innovations in MobileNet is the use of depthwise separable convolutions to reduce the number of parameters and computations. Let's break down how MobileNet does this:

1 .**Standard Convolution**: In a standard convolution, if you're processing an input feature map with D_f channels using D_k filters, each of size $K \times K$, the number of computations would be roughly: $D_f \times D_k \times K \times K$

2. Depthwise Separable Convolution: This type of convolution splits the standard convolution into two parts: a depthwise convolution and a pointwise convolution.

(a) Depthwise Convolution: This involves applying a single filter per input channel. The number of computations for this would be: $D_f \times K \times K$

(b) Pointwise Convolution: After the depthwise convolution, you apply a 1×1 convolution. This is the "pointwise" step that combines the output channels from the depthwise step. The number of computations here would be: $D_f \times D_k$ Adding the computations from the depthwise and pointwise steps, you get: $(D_f \times K \times K) + (D_f \times D\_k)$ This is much less than the computations in the standard convolution if $K \times K$ is much larger than 1 (e.g., $K = 3$ for a 3×3 filter).

By replacing standard convolutions with depthwise separable convolutions, MobileNet significantly reduces the number of parameters and computations, making it efficient.

The MobileNet SSD architecture also uses the non-maximum suppression (NMS) algorithm for post-processing. This algorithm is used to identify the most likely object in the image and suppress the other less likely objects in the scene. This algorithm helps to reduce false positives, which increases the accuracy of the model.

The MobileNet SSD model is trained using a transfer learning approach. This means that the model is pre-trained on large datasets such as ImageNet, and then fine-tuned for the specific task. This helps to reduce the time and computational complexity required to train the model from scratch.

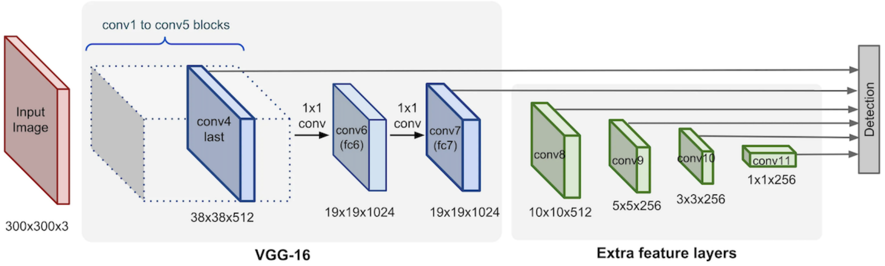


Fig. 5.8 SSD Mobilenet Layered Architecture, image coutesy of: https://lilianweng.github.io/posts/2018-12-27-object-recognition-part-4/

When compared to approaches that are based on regional proposal networks (RPN), such as the R-CNN series, which require two shots, one for generating region proposals and one for detecting the object of each proposal, SSD only requires us to take a single shot in order to detect multiple objects within an image.

5.2.4 YOLO

Redmon et al. [136] introduced the YOLO model in 2016, which was designed for real-time processing. By utilising a novel technique, YOLO revolutionised object detection. Rather than segmenting an image into smaller sections, YOLO applies a single neural network to the entire image and then divides it into smaller regions to simultaneously predict probabilities and bounding boxes [115]. This distinguishing feature of YOLO enables it to utilise a single-stage deep learning algorithm based on convolutional neural networks [117].

YOLO has gone through significant development over the years, culminating in multiple versions including YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, and most recently YOLOv8 [117]. Each version features unique enhancements and modifications. Notably, YOLOv3 employs a distinct feature extractor known as Darknet-53, which utilises 3x3 and 1x1 convolutional networks [137]. In contrast, YOLOv4 and YOLOv5 use CSPDarknet53 as their primary feature extractor. In terms of class predictions, YOLOv3 and YOLOv4 use binary cross-entropy loss [117], [137], [88], whereas YOLOv5 employs binary cross-entropy and logits loss functions [117], [88].

At the core of the YOLO architecture is the YOLO model's deep neural network, which is typically based on a CNN. The network takes an image as input and processes it through several convolutional and pooling layers, gradually capturing higher-level features. The final layers of the network then make predictions for object classes and bounding boxes.

To improve accuracy, YOLO assigns each bounding box prediction to the cell with the highest overlap, or Intersection over Union (IoU), between the predicted box and the ground truth box. This approach ensures that each object in the image is assigned to the most relevant cell, reducing the chances of multiple cells detecting the same object.

The YOLO architecture introduces anchor boxes to handle objects of different scales and aspect ratios. Anchor boxes are pre-defined bounding boxes of various sizes and shapes. Each anchor box is associated with specific aspect ratios and scales, allowing the model to adapt to different object dimensions. The model adjusts the predictions based on the anchor boxes to accurately localise objects in the image.

One of the significant advantages of the YOLO architecture is its speed. By applying the neural network to the entire image at once, YOLO avoids the need for sliding windows or region proposal networks, significantly reducing computation time. The single-pass processing approach makes YOLO highly efficient and well-suited for real-time applications where fast and accurate object detection is essential.

Furthermore, YOLO achieves a good balance between accuracy and speed. Although it may not match the detection accuracy of slower, two-stage models like Faster R-CNN, YOLO provides a practical solution for many real-world scenarios where real-time processing is crucial. Its real-time capabilities make it suitable for applications such as video surveillance, autonomous driving, and robotics.

However, the YOLO architecture does have some limitations. One of the primary challenges is detecting small objects. Since the network operates on a grid scale, small objects may not be accurately captured. YOLO struggles with objects that are significantly smaller than the grid cells, leading to lower detection performance for tiny objects.

Another drawback is the loss of fine-grained spatial information. Since the entire image is divided into a grid, YOLO may struggle with precise localisation, especially for objects with complex shapes or occlusions. This limitation can lead to slightly less accurate object boundaries compared to two-stage models that refine object proposals.

Furthermore, YOLO may have difficulty detecting densely packed objects. When objects overlap or are tightly packed, the model can struggle to separate and accurately predict individual objects. This limitation can affect the detection performance in crowded scenes.

5.2.5 EffiecientNet

EfficientNet and EfficientDet are state-of-the-art models in the field of computer vision, specifically designed for efficient and accurate object detection tasks. The EfficientNet architecture was introduced by Tan and Le [161] in 2020, while EfficientDet was developed as an extension of EfficientNet. These models have achieved remarkable performance by carefully balancing network depth, width, and resolution, leading to improved accuracy and efficiency.

The EfficientNet architecture is based on a scaling method that uniformly scales all dimensions of depth, width, and resolution using a compound coefficient. The key idea is to balance these dimensions to achieve optimal performance. By scaling up the network, EfficientNet achieves better accuracy without sacrificing efficiency. The baseline network, EfficientNet-B0, is obtained through a multi-objective neural architecture search that optimizes both accuracy and FLOPS (floating-point operations per second). The main building block of EfficientNet is the mobile inverted bottleneck MBConv, which is augmented with the squeeze-and-excitation optimization technique.

To further enhance the performance, EfficientNet undergoes a two-step scaling process. In the first step, the scaling coefficients, denoted as α, β and γ, are determined through a grid search, assuming twice the available resources. These coefficients are chosen based on the constraint $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$. In the second step, the baseline network is scaled up with different scaling factors, denoted as , to obtain models from EfficientNet-B1 to EfficientNet-B7. The scaling factors are determined using the previously fixed α, β and γ coefficients. This approach allows for efficient scaling of the network while maintaining the desired accuracy and resource efficiency.

EfficientNet has demonstrated outstanding performance on various benchmark datasets. For instance, EfficientNet-B7 achieved state-of-the-art 84.3% top-1 accuracy on ImageNet, outperforming previous ConvNets while being significantly smaller and faster in terms of inference time [161]. The EfficientNet models also generalize well to other datasets, achieving state-of-the-art accuracy on CIFAR-100, Flowers, and other transfer learning datasets, with significantly fewer parameters.

EfficientDet, on the other hand, builds upon the EfficientNet backbone and extends it to object detection tasks. It combines the EfficientNet feature extractor with a custom detection and classification network. EfficientDet offers highly efficient and accurate object detection capabilities, particularly suited for real-time applications.

EfficientDet has achieved state-of-the-art performance on the COCO dataset, surpassing models such as YOLOv3 [161]. The architecture of EfficientDet varies according to different versions, ranging from EfficientDet D0 to D7. These versions correspond to models with increasing size and capacity, providing a trade-off between accuracy and resource requirements.

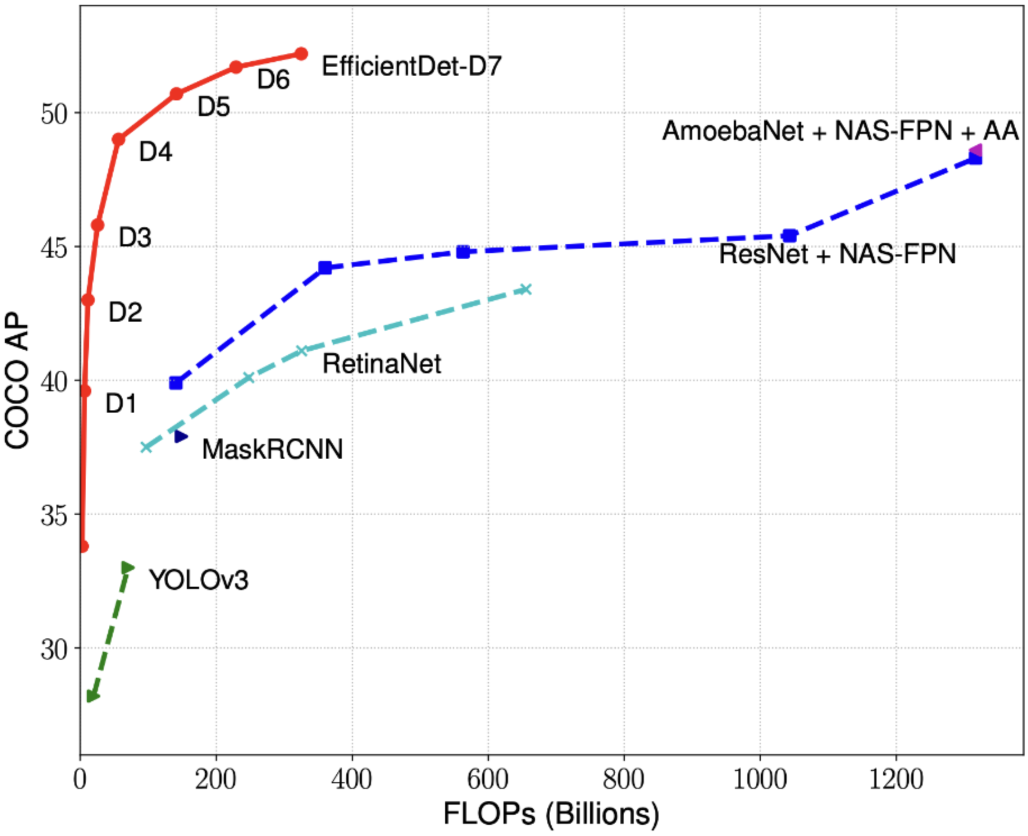


Fig. 5.9 : Comparison of Model Size and ImageNet Accuracy of various object detectors from [161].

EfficientDet-D0, the smallest version of EfficientDet, contains only 4 million weight parameters and achieves impressive performance. It can run inference in just 30ms and requires 17 megabytes of storage, making it both fast and compact.

5.2.6 Dataset Description and Splitting Strategy

The dataset utilised in this study consists of images obtained from 78 who self identified as female participants during the measurement process. Specifically, a collection of 312 frontal view images was gathered from these participants, and each image was annotated accordingly. Table 1 provides an overview of the class distribution and the corresponding balances within the dataset.

Table 5.2 Class Balance

Classes	Amount
Head	312
Chest	312
Bust	312
Waist	312
Hips	312

To facilitate a comparative analysis of different object detection algorithms, the experiment was conducted using the ementioned dataset. The dataset was divided into two different sets of data, namely the training and test data, with proportions of 80% and 20% respectively.

Table 5.3 Class Balance

Classes	Amount
Train Data	249
Test data	63

Initial image pre-processing for the MobileNet-SSDv2, YOLO, and EfficientNet models involves resizing and cropping the images based on the input layer sizes while preserving a specified aspect ratio. After resizing the dataset, data augmentation techniques are implemented. This process permits the application of various geometric and colour transformations to the images, including scaling, transformation, and colour adjustment. During the comparison, the latest version of YOLO available was YOLOv5. Additionally, we included SSD MobileNetv2 FPN-Lite and EfficientDet D0 512x512 in our testing.

5.2.7 Performance Metrics

Several performance metrics were used to evaluate each algorithm. Precision, recall, F1-score, Average Precision (AP), mAP, and inference time were chosen in this research.

$$Precision = \frac{TruePositive(TP)}{TruePositive(TP) + FalsePositive(FP)} \tag{5.2}$$

$$Recall = \frac{TruePositive(TP)}{TruePositive(TP) + FalseNegative(FN)} \tag{5.3}$$

$$F1 - Score = 2 \times \frac{Precision.Recall}{Precision + Recall} \tag{5.4}$$

$$AP = \int_0^1 p(r)dr \tag{5.5}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5.6}$$

Equation 5.2, Precision quantifies the accuracy of prediction results by calculating the percentage of correct predictions in relation to both false positives and true positives. It assesses the level of correctness in the predictions made. Equation 5.3, Recall determines the effectiveness of an algorithm in identifying all positive cases. It measures how well the algorithm captures and includes all instances of positive cases in its predictions. Equation 5.4, The F1-Score represents a balanced measure between precision and recall, with values ranging from 0 to 1. A higher F1-Score indicates a better balance between precision and recall, indicating that the algorithm achieves both high accuracy in its predictions (precision) and effectively captures a significant portion of positive cases (recall). Equation 5.5, AP has become the standard measure of assessment for Object detection. This is calculated as the average accuracy of detection at multiple recall levels and is assessed individually for each class Equation 5.6, To compare the performance of all the classes, the mAP of each class is taken and used as the final measure for evaluation in Object detection and related areas.

In this study, various Intersection over Union (IoU) values were used to determine the mean average precision (mAP). The IoU measures the proportion of overlap between the predicted and ground-truth bounding boxes. mAP@.5 indicates that the IoU threshold for calculating the mean average precision has been set to 0.5. In addition, mAP@.5:95 is utilised, which represents the average mAP across different IoU thresholds varying from 0.5 to 0.95. Another important metric is inference time, which indicates the rate at which the algorithm makes predictions in real-time. Real-time applications would benefit from a reduced inference time. In addition, we considered the resource requirements, including GPU utilisation and the viability of implementation on mobile devices, favouring methods that require fewer resources and are more compatible with mobile devices.

5.2.8 Findings

Our investigations were divided into two sections. During the initial stage, we investigated conventional computer vision techniques, such as HOG and Haar-cascade. These techniques are extensively employed for object detection, but we wanted to evaluate their performance using real-world data. Therefore, we moved on to the next phase, where we trained more sophisticated models capable of detecting human body parts and labelling them into five distinct regions: head, chest, bust, waist, and hips (as shown in Figure 5.10. We conducted a comparative analysis of three cutting-edge models for this purpose: MobileNet SSD v2, YOLOv5, and EfficientDet-D0. The selection of these models was based on extensive prior research that evaluated their efficacy in a variety of studies, particularly in the field of human body detection on smartphone devices.

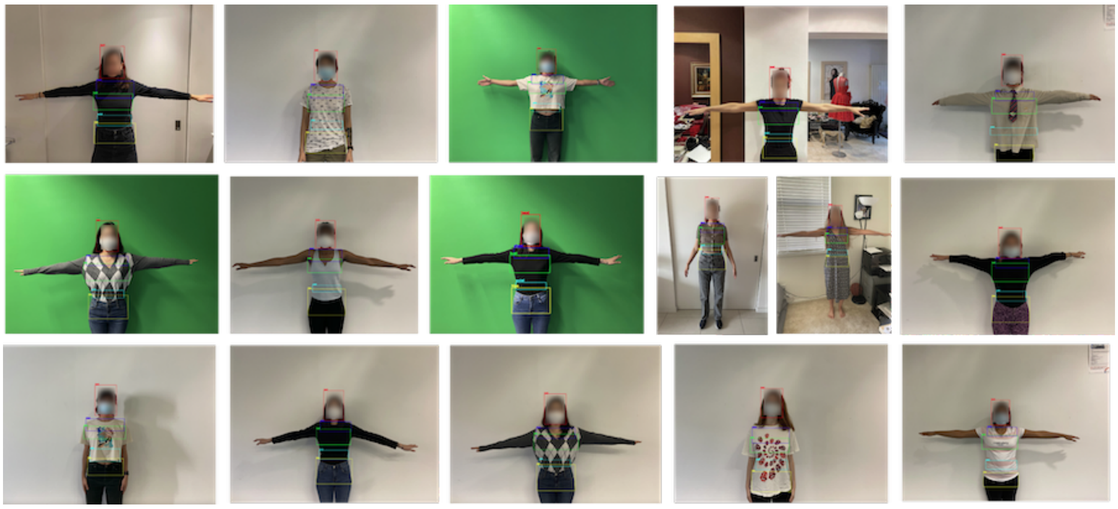


Fig. 5.10 : Labelling our datasets

MobileNet SSD v2, YOLOv5, and EfficientDet-D0 were selected due to a number of technical factors demonstrating their suitability for our particular human body detection task.

MobileNet SSD v2 is renowned for its efficiency and precision in mobile device object detection. It utilises depth-wise separable convolutions, which enables it to attain high performance at low computational cost. MobileNet SSD v2 was chosen due to its balance between accuracy and real-time processing, which aligns well with our goal of efficient human body detection on mobile devices.

YOLOv5 models are well-known for their real-time object detection abilities. The most recent version of YOLO at the time of our study, YOLOv5, features several improvements over previous versions. Its architecture is optimised for speed and precision, making it an attractive option for detecting human body parts on mobile devices. Incorporating YOLOv5 into our comparative analysis was a no-brainer due to its ability to process frames quickly and robust detection performance.

EfficientDet-D0 is a member of the EfficientDet model family, which provides a balance between efficiency and precision. EfficientDet-D0, the series' smallest variant, is ideally adapted for environments with limited resources, such as smartphones. It provides satisfactory object detection results while preserving a lightweight architecture, which makes it an attractive candidate for our human body detection task.

By selecting these models, we ensured that our comparative analysis included a variety of cutting-edge approaches that have demonstrated efficacy in object detection and human body detection in particular. We evaluated the technical advantages of each model, such as their optimised architectures, efficient processing capabilities, and ability to operate efficiently on mobile devices. These factors were crucial in determining the suitability of these models for our research and allowed us to make informed comparisons of their performance in human body detection.

Given the inherent differences between computer vision techniques and deep learning models, it may not be ideal to compare the results directly. To ensure a fair comparison and determine the most suitable option for our software, we made every effort to ensure that all models had similar settings. In Table 5.4, we present the testing outcomes of our experiments involving a combination of computer vision techniques (HOG and Haar-Cascade) and machine learning models (MobileNet SSD v2, YOLOv5, and EfficientDet-D0). By comparing the performance metrics of each object detection algorithm, we hoped to obtain a better understanding of their capabilities and select the most suitable one for our software.

Table 5.4 Comparison of the various computer vision and machine learning techniques for Body Detection

Measure	HOG	Haar-Cascade	MobileNet SSD v2	YOLOv51	EfficientDet-D0
Precision	0.68	0.52	0.79	0.82	0.85
Recall	0.26	0.29	0.54	0.54	0.58
F1-Score	0.22	0.3	0.61	0.64	0.61
mAP@.5	N/A	N/A	0.70	0.69	0.73
mAP@.5:95	N/A	N/A	0.55	0.53	0.6
Inference Time (ms)	38.3	35.41	8.4	24.28	21.34

Note that metrics such as mAP (mean Average Precision) or mAP at different IoU thresholds (mAP@.5, mAP@.5:95) are not typically calculated for HOG and Haar-cascade.

On the basis of the performance metrics, EfficientDet-D0 and YOLOv5 achieve high precision values of 0.85 and 0.82, respectively, when identifying bodies. EfficientDet-D0 also excels in recall, with a value of 0.58, followed by YOLOv5's 0.54. In terms of F1-score, mAP@.5, and mAP@.5:95, both techniques outperform others, with EfficientDet-D0 achieving F1-scores of 0.61 and mAP@.5 values of 0.73. MobileNet SSD v2 and YOLOv5 also perform fairly well with respect to these metrics. MobileNet SSD v2 has the fastest inference time at 8.4 milliseconds, followed by EfficientDet-D0 and YOLOv5 at 21.34 and 24.28 milliseconds, respectively. HOG and Haar-Cascade have significantly longer inference times than the other techniques.

Based on a comparison of mAP@.5 and inference time, EfficientDet-D0 offers the highest accuracy with a mAP@.5 of 0.73, surpassing MobileNet SSD v2 and YOLOv5. MobileNet SSD v2 has a lower inference time of 8.4 milliseconds, which indicates quicker processing and makes it suitable for real-time applications. While EfficientDet-D0 offers greater precision and YOLOv5 performs better in F1-Score, MobileNet SSD v2 was selected for our software due to its balanced performance, lower GPU requirements, and mobile-optimized design.

5.3 Method

Several steps are required to use the mobileNet SSD (computer vision and machine learning) model for body detection. First, a collection of photos of people with varying body types must be compiled. This information can subsequently be used to train computer vision and machine learning systems. Once the algorithms have been trained, a single photograph can be used to detect the size, sections and proportions of a person's body.

Our primary focus in this chapter does not merely lie in extracting only the human body from the image; we are also interested in extracting the upper human body, as well as classifying each part of the body into categories such as head, chest, bust, waist, hips, shoulder, and height. This will enable us to speed up the process of measuring each body area more precisely.

To achieve this, we utilise LabelImage [166], an open-source mobile computer vision and machine learning software library, It uses a MobileNet SSD model. This functionality will allow us to train our image databases and appropriately label them. The model can then be utilised to identify things inside an image. This is achieved by applying the image to the model. The model will then produce the detected object's bounding box and label. By utilizing these labels, we can automatically extract the regions of interest from each section of the body and utilize the relevant area of the image to calculate the various components of the human body. Let's move on to the approach of this part and explore in further detail how we were able to detect the human body using mobileNet SSD and what its dependencies are.

5.3.1 Identify body section names

To train a custom human body detector, we must divide our project into 2 different phases, each with its own sub-steps (as shown in Figure 5.11):

1. **Training:** Here, we'll concentrate on loading a human body detection dataset from disc, training a model on this dataset using TensorFlow, and then serialising the human body detector to disc. Serializing a human body detector entails saving the trained model's architecture and parameters to a file, enabling easy reuse, sharing, and deployment in various applications.
2. **Deployment:** Once the human detector is trained, we can then move on to loading the human detector, performing upper body detection, and then classifying each upper body as head, chest, bust, waist, hips and height

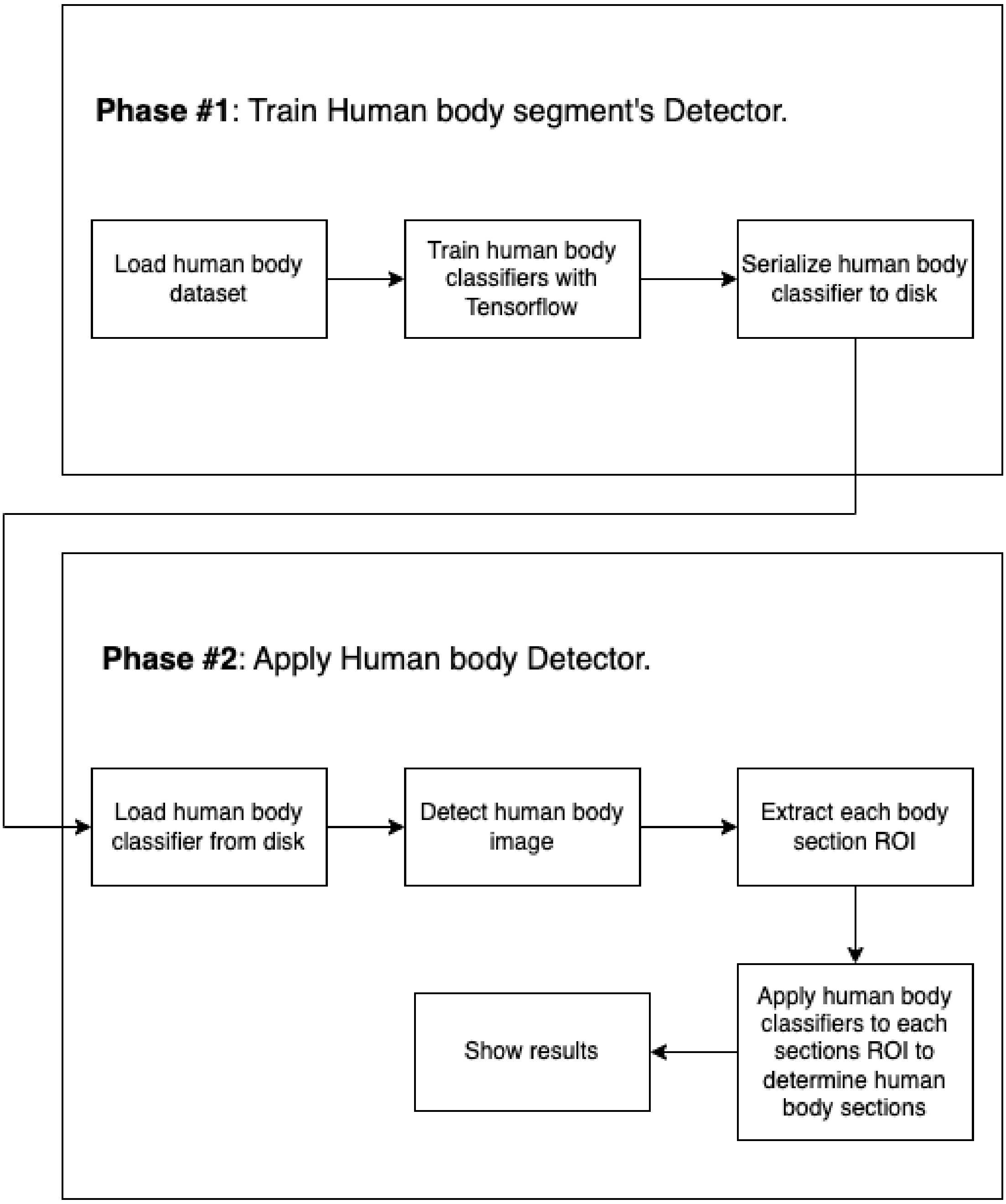

Phase #1: Train Human body segment's Detector.
Load human body dataset
Train human body classifiers with Tensorflow
Serialize human body classifier to disk
Phase #2: Apply Human body Detector.
Load human body classifier from disk
Detect human body image
Extract each body section ROI
Apply human body classifiers to each sections ROI to determine human body sections
Show results

Fig. 5.11 : Phases and individual steps for building a human body detector with computer vision and deep learning using Python, OpenCV, and TensorFlow/.

5.3.1.1 Phase 1: Train Human Body Segment's Detector

The initial phase involves training a classifier to detect and segment the human body.

1. **Load Human Body Dataset**: The first step involved the collection and preparation of our own unique dataset. It was observed that publicly available datasets such as COCO or MPII Human Pose were not ideally suited for our specific requirements. Consequently, we compiled our own custom dataset, featuring a total of 312 images from a total of 78 who self-identified as female participants. This was intended to provide our model with a diverse set of data and improve its ability to recognize and segment the human body.

2. **Train Human Body Classifier with TensorFlow**: After the dataset was prepared, the next step was to train our classifier using this data. TensorFlow is an excellent library for building and training deep learning models. In this case, the model of choice was MobileNet SSD (Single Shot MultiBox Detector), an architecture well-known for its efficiency and effectiveness in real-time processing tasks, even on less powerful hardware. Training the model involved feeding it our images, adjusting the internal weights and biases of the MobileNet SSD via backpropagation, and optimizing this process in an iterative manner to minimize the difference between the model's prediction and the actual labeled data (loss function).

3. **Serialize Human Body Classifier to Disk**: Post-training, it was essential to save the model for future use, a process known as model serialization. TensorFlow provides this functionality through its model.save() function. This enables the model's architecture, weight values, and training configuration to be stored, facilitating its loading from disk for subsequent use or further analysis.

5.3.1.2 Phase 2: Apply Human Body Detector

This second phase involved applying our trained classifier to detect and label different sections of the human body in new, unseen images.

1. **Load Human Body Classifier from Disk**: I began by loading our serialized model from disk. This was made possible through TensorFlow's tf.keras.models.load_model() function. Armed with the trained model, our software was now capable of detecting human body sections.

2. **Detect Human Body in Image**: The next step involved processing new images and detecting human bodies using our trained model. This generally required pre-processing activities such as image normalization, resizing to match the model's input shape, and converting the image to an array. The processed image was then passed through the model for inference, resulting in a set of bounding boxes and confidence scores for each detected human body part.

3. **Extract Each Body Section ROI**: After detection, the software pinpointed Regions of Interest (ROI) subsection 5.3.2 - areas containing each body part. These ROIs were then extracted for further analysis. OpenCV, a powerful library for image processing and computer vision tasks, was essential in this process.

4. **Apply Human Body Classifier to Each Section's ROI**: The next step was to apply the classifier to each extracted body section's ROI. The aim was to determine the specific type of each body part - an head, shoulder, chest, etc. At this point, the labeling of each body section occurred, based on the detections made by our MobileNet SSD classifier.

5. **Show Results**: Finally, results are provided. This was typically done by drawing bounding boxes and labels around the detected body parts on the original image and then displaying it as shown in Figure 5.12.

5.3.2 Region of Interests

The utilization of the Region of Interest (ROI) in our project forms a core aspect of the Single Shot MultiBox Detector (SSD) architecture we are employing, playing an invaluable role in defining the focus of our image processing. Before moving into the technicalities of the RoI application, let's consider why we chose to use RoI in the first place.

The RoI method serves as a viable solution to isolating the entire human body from its surrounding environments, an essential step that significantly aids our second and third pipeline stages—Chapter 6. This task of removing a background scene to separate a human body from a scene has posed a substantial challenge, mainly due to the rapid change of the background environment. Although multiple experiments were conducted, we concluded that the image correction steps by themselves cannot entirely separate the human body from the background. This recurring issue hinged on the fact that extraction of human contours is vital for the development of vision-based non-contact human body measurement and modelling systems. The efficiency of the RoI arises from its ability to focus on specific areas of an image and discard irrelevant or unwanted parts.

Given this backdrop, the RoI becomes a crucial component in our system. It directly affects the accuracy of subsequent measures of body size, making the accurate extraction of human outlines from photos a priority. However, the varying background environment has a direct effect on the accuracy of human identification segmentation techniques. Our solution to this, therefore, is a method that enables us to accurately extract human body contours from a complex background environment.

The RoI technique, in its essence, is a digital image manipulation technique. Users can isolate specific parts of an image or video frame for further processing, with a focus on the chosen areas of an image, discarding the irrelevant sections. This process is achieved by selecting a rectangular area, or region, of an image, and then only processing the pixels in that region.

In our software, this concept is extended to define regions of the human body in relation to the bounded rectangle of the mask image. These regions encompass the head, neck, hands (shoulder width for side views), bust/chest, waist, hips, and height. This division allows the software to treat each region as an independent unit of interest for more precise measurements (as can be seen in Figure 5.12). To accomplish this, we use a unique approach. We employ a different colour space to differentiate between the upper and lower regions of the human body.

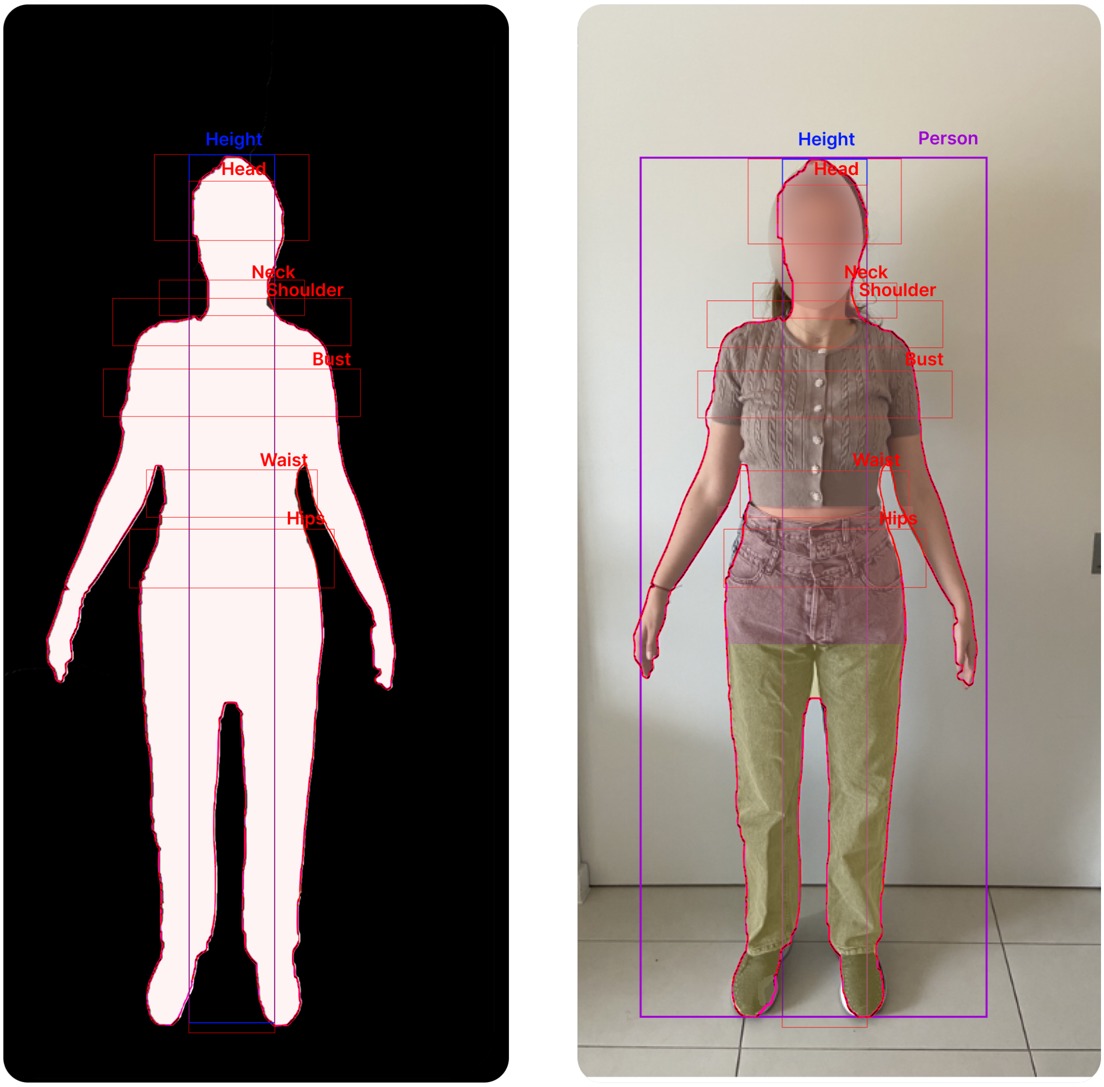


Fig. 5.12 : Image obtained by adding masks from a set of training images

This feature is beneficial because our software primarily focuses on measuring the upper body. While data from the lower body is collected, it isn't used in the automatic computations, ensuring that the software's focus remains on the upper body (see Figure 5.13). After the initial differentiation of the upper and lower sections of the human body by the SSD, the software proceeds to apply unique colour spaces to each of these sections, enhancing the distinguishing characteristics of each segment of the body. This step allows for better isolation and analysis of each part, improving the detection and measurement processes.

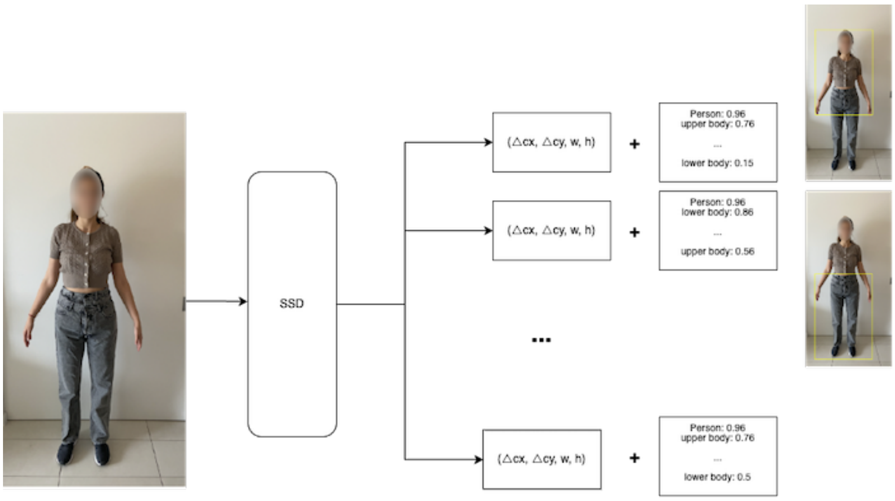


Fig. 5.13 : SSD predicts two distinct classes from a single image by selecting the class with the greatest score for the bounded object and assigning the class '0' to non-bounded objects. CX and CY in MobileNet SSD are the x and y coordinates of the center of the bounding box. These coordinates are used to determine the location of the detected object within the image.

One significant attribute of the RoI within the SSD framework is its ability to efficiently manage the high-dimensionality problem. While SSD can detect objects at different scales due to the varying sizes of its feature maps, this flexibility also introduces an array of bounding box aspect ratios, leading to a significant number of RoIs to consider. Here, the RoI acts as a strategic filter to pinpoint the areas worth analysing, significantly reducing the number of candidate boxes.

Our implementation leverages the multi-layer architecture of MobileNet SSD, using different feature map layers for detecting objects of different sizes. This unique ability, in the context of our project, allows for the segmentation of the human body into upper and lower sections. The SSD can extract RoIs from various feature map layers, enabling the differentiation of various body sections based on their size and spatial location in the image.

Finally, the RoI pooling layer, a part of the Fast R-CNN structure embedded in SSD, enables the model to handle RoIs of varying sizes and create fixed-size feature maps. This ensures that, regardless of the human body's size in the image, we can create consistent and fixed-size feature maps of the upper and lower body sections, ensuring the uniformity of the data and the resulting measurements.

5.4 Summary

In this chapter, we began by providing a brief history of object detection and specifically focused on detecting the human body. We highlighted the challenges faced in human body detection, particularly emphasizing the impact of backgrounds on the accuracy of detection results.

To address these challenges, we explored several machine learning methods for object detection, including Haarcascade, HOG, YOLO, MobileNet SSD and EfficientNet. While EfficientDet-D0 exhibited impressive accuracy with an mAP@.5 value of 0.73, it had a slightly higher inference time of 21.34 milliseconds. After careful examination and analysis, we concluded that MobileNet SSD is the most suitable choice for our application. Our decision was based on key factors such as mean Average Precision (mAP) and Inference Time. MobileNet SSD demonstrated superior performance with an mAP value of 0.7 and the ability to process up to 8.4 milliseconds, making it a compelling option for real-time object detection in our specific scenario.

We also considered the future deployment of our software on mobile devices, and MobileNet SSD stood out as it has been specifically designed for such systems. It combines accuracy and efficiency, making it an ideal choice for real-time applications.

Following the selection of MobileNet SSD, we delved into the implementation details of our method, which consists of two phases: training and deployment. In the training phase, we loaded a human body dataset and trained a human body classifier using TensorFlow. The resulting classifier was serialized and stored on disk for later use.

In the deployment phase, we loaded the human body classifier from disk and applied it to detect human bodies in images. We employed the concept of Region of Interest (RoI) to isolate each section of the body, leveraging the capabilities of MobileNet SSD. RoI proved to be a viable solution for separating the human body from its surroundings, allowing for better isolation and analysis of each body part.

Furthermore, our system prioritizes the upper body by differentiating it from the lower body during the initial stages of detection. Although data from the lower body is collected, it is not utilised in the automatic computations, ensuring that the software's focus remains on the upper body. Additionally, we applied unique colour spaces to each section of the body, enhancing their distinguishing characteristics. This step further improves the detection and measurement processes, facilitating more accurate analysis.

Chapter 6

Image Corrections, Image Segmentation, Skin Detection, and Camera Calibrations for Human Body Analysis

This chapter of our proposed software solution aims to enhance the overall quality and appearance of uploaded images through image correction. The chapter begins by discussing the challenges posed by cluttered backgrounds, noise, and shadows, which can impact the accuracy of collected data. We explore two potential approaches for image segmentation, classical computer vision image segmentation and DeeplabV3, and compare their effectiveness in improving image quality. Based on our findings, we select classical computer vision image segmentation as the optimal method for our software solution.

Next, we discuss skin detection, which involves identifying the human body's skin color using different color spaces such as RGB, HSV, and YCbCr. We explain the skin detection algorithm and how we implement it into our software solution. Finally, we present the results of our experiments, which compare plain and cluttered backgrounds to ensure the effectiveness of our image correction and skin detection methods in accurately collecting data from the human body in two images.

In what follows, we also give details on how we used camera calibration based on OpenCV's functions, which allows to accurately convert pixel measurements to corresponding centimeters. We also summarise ours findings which identify the range of 0.5 to 3 meters as the most accurate distance for data collection, given the architecture of our designed system.

Furthermore, this chapter provides insights into the automatic and semi-automatic methods for detecting and localizing landmarks on the body. These methods contribute to the overall precision of the software solution, enabling efficient data collection and analysis.

The chapter concludes with a summary of the findings and implications for the development of the software solution. By utilising classical computer vision image segmentation and skin detection algorithms, it becomes possible to enhance the quality of uploaded images and gather precise data from the human body. The significance of image correction in elevating the overall image quality and ensuring accurate data collection is underscored in this chapter. Detailed results and discussions are presented in Chapter 8.

6.1 Image Corrections

Before launching into the methods of this chapter, it's useful to review some of the problems that need to be fixed in advance. Two general categories of problems as follow.

1. An image needs Improvement

2. Low-level features must be detected

When an image needs improvement, it is often due to low-resolution, poor lighting, an incorrect color balance, or the use of an incorrect color space. Low-resolution images can be improved by up-sampling or interpolation, which can increase the detail and definition of the image. Poor lighting can be improved by using a flash or using a light box to create even lighting. Color balance can be adjusted using color correction tools, such as curves or levels. Color spaces can also be adjusted to match the intended output. On the other hand, Low-level features must be detected in an image to perform more advanced image processing algorithms, such as object recognition and classification. This can be done using edge detection algorithms, which can detect lines, corners and curves in an image. In addition, texture detection algorithms can be used to detect different characteristics of an image. Edge detection is then applied based on the classic Canny edge detection algorithm. We conduct edge contour tracking via hysteresis to detect body part contour segments by suppressing weak pixels not connected to strong ones as highlighted by the Canny operator (Additional details on this topic can be located within the section D.1.). These algorithms can be used to detect regions of interest, or to identify different objects within an image. By detecting these low-level features, more advanced image processing algorithms can be applied to the image.

In the context of image processing, two key categories of issues that need to be addressed: image enhancement and low-level feature detection. Image improvement involves enhancing image quality by addressing issues like low resolution, poor lighting, incorrect color balance, and color space discrepancies. Techniques such as up-sampling, flash usage, and color correction tools are employed for this purpose. On the other hand, low-level features pertain to fundamental elements within an image, including lines, corners, curves, and textures. Detecting these features is essential for advanced image processing tasks like object recognition and classification, facilitating the identification of regions of interest and objects within images.

In order to detect low-level features, improve the image quality; we must purge the image of various types of noise, improve the contrast between adjacent regions and features, and simplify the image through selective smoothing and the elimination of features at small scales while maintaining other features at desired scales - [107], we have considered two different image segmentation's methods;

1. Classical computer vision image segmentation
2. DeeplabV3

6.1.1 Classical computer vision image segmentation

In what follows, we provide a quick overview of the various stages of image improvement and low-level feature recognition using traditional computer graphics segmentation. We proceed as follows; **(i)** metric correction of an affine camera model, **(ii)** grey level mapping of an RGB input [151] [17], **(iii)** removal of small image regions and image smoothing (blurring) [151] [17], **(iv)** detecting edges [107], **(v)** mask operations on matrices and **(vi)** thresholding [17].

6.1.1.1 Affine and Metrics Correction

With an affine and metric correction step, we may increase the visibility and geometry of the many regions into which an image can be segmented, as well as the clarity of image features inside these regions. Affine and Metrics Correction in image segmentation is a process which is used to rotate an image and make it 90 degrees in order to better analyse or segment the image. It involves adjusting the scaling, rotation, and translation of an image so that the original shape of the image is maintained.

An affine transformation is a linear mapping between two coordinate systems. It involves three parameters, a scaling factor, a rotation angle, and a translation vector. Metrics Correction is the process of adjusting an image to better fit the desired shape or size. This involves changing the aspect ratio of the image, or making changes to the coordinates of the image. Both before and after the conversion, parallel lines will stay parallel. An image's affine transformation can be represented by a mapping that looks like this: x $|\rightarrow$ Mx+b, where M is a linear transform (matrix) and b is an offset vector. This mapping applies to every pixel in the image.

The process of the affine and matrices corrections is as follows: (i) Read the color image, convert it into grayscale, and obtain the grayscale image shape (ii) scale the image based on the result of the camera calibration (iii) Rotate the image by the results that is obtained on camera calibration counter-clockwise. It is a composite operation first, need to shift/center of the image, apply rotation, and then apply inverse shift(iv) Apply reflection and transition transform (along the x-axis) (v) Apply shear transform to the image (x,y axis)

This will helps to accurately measure the distance and area of the objects and also helps in object detection and image segmentation.

6.1.1.2 Grey level mapping

Grey level mapping in image segmentation is a process used to convert an image from its original color format to a grey scale version. This can be used for a variety of tasks such as image segmentation, object recognition, and edge detection. Because it is common practise to extend the grey values of an image that is too dark onto the complete set of grey values that are available, the process of remapping the grey value is frequently referred to as "*stretching*". We have altered the intensity values of the pixels in order to improve the images. Gamma correction is used to scale the image pixel intensities from the range of [0, 255] to [0, 1.0]. The following equation shows the output gamma-corrected image:

$$I_{out} = I_{in}^{(1/G)}$$

Where I_{in} is our input image and G is the gamma value. This level mapping, helps to reduce problems due to wearing light-colored clothing (such as white or off-white), as well as those affecting people whose input photographs contained shadows and reflections.

In our case we converted the image to greyscale by calling **cv2.cvtColor** and providing the image and **cv2.COLOR_BGR2GRAY** flag. The cv2.COLOR_BGR2GRAY function takes the three color channels (Red, Green, and Blue) and calculates the average of them, which is then used as the grey level of the pixel.

Once the image has been converted to a grey scale version, it can then be used for image segmentation. A variety of thresholding methods can be used to divide the image into two groups: background and foreground. The most common thresholding methods are Otsu's thresholding and Adaptive thresholding;

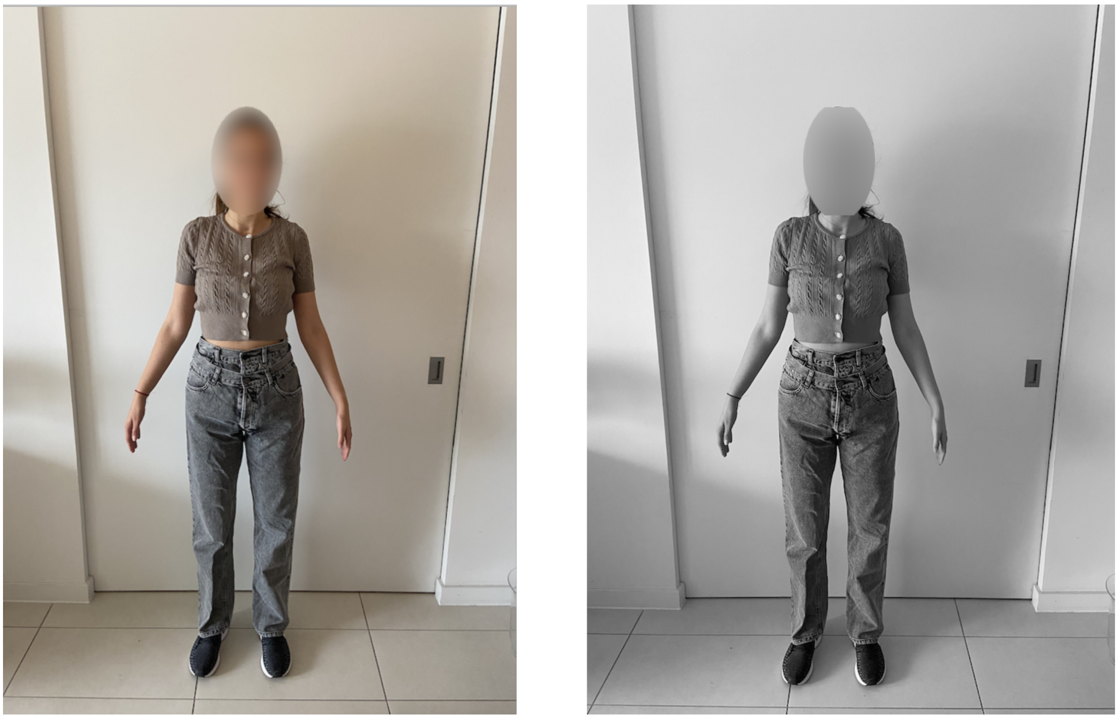


Fig. 6.1 Display the original image on the left and gray-scale image on the right

6.1.1.3 Removal of Small Image Regions, Image Smoothing and Blurring

There are occasions when cropping out little sections of an image can be quite helpful. It's possible that this is just the consequence of noise, but it could also reflect low-level detail that has to be omitted from the image description that's being generated. It is possible to get rid of small regions either by modifying individual pixels or by eliminating components once a region retrieved [151].

Often, we want to detect and describe some underlying ideal structure from the composed image, together with some random noise or artefact which we want to remove. For instance, we can observe that blurring is utilised during the construction of a simple body scanner, and that smoothing is utilised to assist us in locating our marker throughout the process of determining the distance between an object and our camera (Appendix E.1.3). As a result, in this particular illustration, the image's precise points have been simplified, but the majority of the picture's structural elements have been left unchanged.

According to the findings of the research that I have conducted, a great deal of the image processing and computer vision functions, such as the thresholding described in Appendix D.1.2.3 and the edge detection described in subsubsection 6.1.1.4, produce superior outcomes if the image is first blurred or smoothed as it reduce the potential effects of noise or tiny high frequency features.

Although this may appear at first counter-intuitive, reducing the number of details in an image allows us to more easily locate the objects of interest in an image that (i.e. the upper human body). For additional details on this topic, please refer to Appendix D.1.

On the basis of our findings, we have decided to employ Gaussian blur to eliminate noises and smooth the images in our software. It keeps the majority of the image's edges, which is essential for the next phase in our image optimization phase. The amount of blurring is controlled by the blur value variable, which is a tuple describing the blurring kernel size. It is also very efficient to compute with a pair of 1D convolutions along the horizontal and vertical directions.

Next, by detecting edges using masks, we then replace the background pixels of the image with the blurred image's pixels. In the following part, we will elaborate on the various types of edge detection utilising masks.

6.1.1.4 Detecting Edges using Masks

When we speak of image edges, we are referring to abrupt changes in an image's intensity. Changes in intensity typically correspond to physical changes such as differences in texture, object boundaries, reflectance and depth or orientation discontinuities or differences in the objects' lighting when it comes to certain properties of the surfaces of imaged three-dimensional objects. Edge detection is generally assumed important for higher-level vision tasks, and it can also lead to some inferences about the physical features of the 3D world.

Filtration can use either a linear or nonlinear approach. Additionally, the enhancement may either be directional or isotropic in nature. Algorithms for edge detection contain the following steps:

- **Noise Filtering**: Reduce noise as much as possible without damaging the actual edges. Filtering is often used to enhance the effectiveness of an edge detector relative to noise. There is a tradeoff, however, between edge strength and noise reduction. More filtering to eliminate noise can diminish edge strength.
- **Enhancement**: Applying a filter to improve the overall quality of the image's edges.
- **Detection**: Only the points with strong edge content require our attention at this time. However, there are a lot of locations in an image that have a value that's not zero for the gradient, and our software shall not consider all of those points to be edges.

- **Localisation**: determine the location of an edge is possibly with *sub-pixel* resolution.

For additional details on this topic, please refer to Appendix D.1.2.

6.1.2 DeeplabV3

With the aid of traditional computer vision, our software is able to handle challenges such as shadows, noise, lighting, thresholding, etc. As an example of more sophisticated neural network methods, we considered DeepLabV3, a popular semantic segmentation algorithm. Hence, it has been decided to analyse this technology and compare the results to those of traditional computer vision in order to determine which methods are superior for our use-case.

DeepLabV3 is a state-of-the-art human body detection algorithm developed by researchers at Google. It is based on convolutional neural networks (CNNs), a type of deep learning algorithm that is becoming increasingly popular for image segmentation tasks. DeepLabV3 uses an encoder-decoder architecture, where the encoder is a powerful CNN used for feature extraction, and the decoder is a simple CNN used for pixel-level prediction after training. It takes an input image and produces a pixel-level segmentation map that labels each pixel in the image with one of the classes of interest. The algorithm has been shown to have superior performance compared to other methods, and is currently one of the leading methods for human body detection.

DeepLabV3 utilizes a fully convolutional network (FCN) architecture, which is composed of an encoder network and a decoder network. The encoder network consists of several convolutional and pooling upsampling layers, which are used for extracting features from the input image. These features are then passed to the decoder network, which is composed of several convolutional and upsampling layers. The upsampling is used to reconstruct the pixel-level segmentation map from the features extracted by the encoder network. In addition to the FCN architecture, DeepLabV3 also utilizes an Atrous Spatial Pyramid Pooling (ASPP) module, which is used to capture richer context information from the input image. ASPP consists of multiple parallel atrous convolutional layers with different receptive field sizes, which are used to capture different scales of context information from the input image. This helps the algorithm to better detect objects at different scales.

The algorithm is trained using a combination of supervised and unsupervised techniques. For supervised learning, it is trained using a large dataset, containing images labeled with the different classes of interest. For unsupervised learning, it uses a self-supervised learning technique called image inpainting, which is used to fill in missing pixel-level information in the training images.

6.1.2.1 Atrous Separable Convolution

Atrous Convolution (also known as dilated convolution) is an algorithm used in deep learning and computer vision to increase the receptive field of convolutional neural networks. It is used in the DeepLab v3 architecture to capture more contextual information from the input images, than with more traditional CNN architecture.

Atrous convolution works by applying convolutions with larger than normal filters to the input data, while preserving the spatial resolution of the data. By increasing the size of the filter, it is possible to capture more context from the data. This can be used to extract features from the images, such as edges and objects, with a much higher accuracy than standard convolutional neural networks.

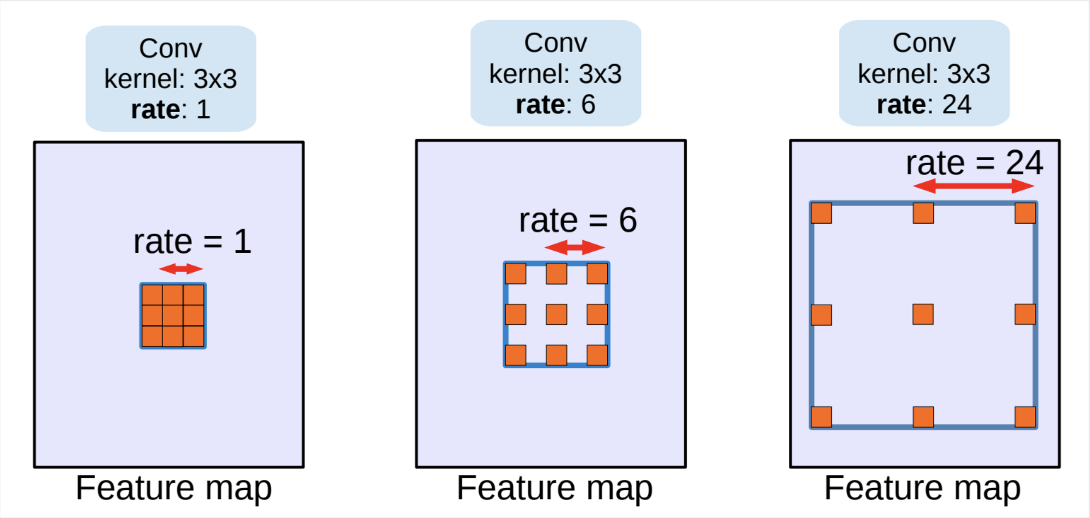


Fig. 6.2 Variable-rate atrous convolution with a 3 3 kernel size. Conventional convolution corresponds to atrous convolution with a rate of 1. Utilizing a high atrous rate enlarges the model's field of view, enabling the encoding of objects of various sizes - Chen et al. [48].

$$y[i] = \sum^{K} x[i + r.k]w[k] \tag{6.1}$$

For each position **i** at the output **y** and filter **w**, the atrous convolution is performed on the input characteristic map **x**, where the atrous rate **r** corresponds to the sampling step of the input signal. This is the same as inputting **x** deviation with sampled filters, which are constructed by inserting zeros **r-1** zero between two consecutive filter values along each spatial dimension. With **r = 1**, it is conventional convolution. By altering **r**, we can adaptively modify the filter's field of vision. Also known as dilated convolution (DilatedNet) or the Hole Algorithm. For additional details on this topic, please refer to Appendix D.1.3.

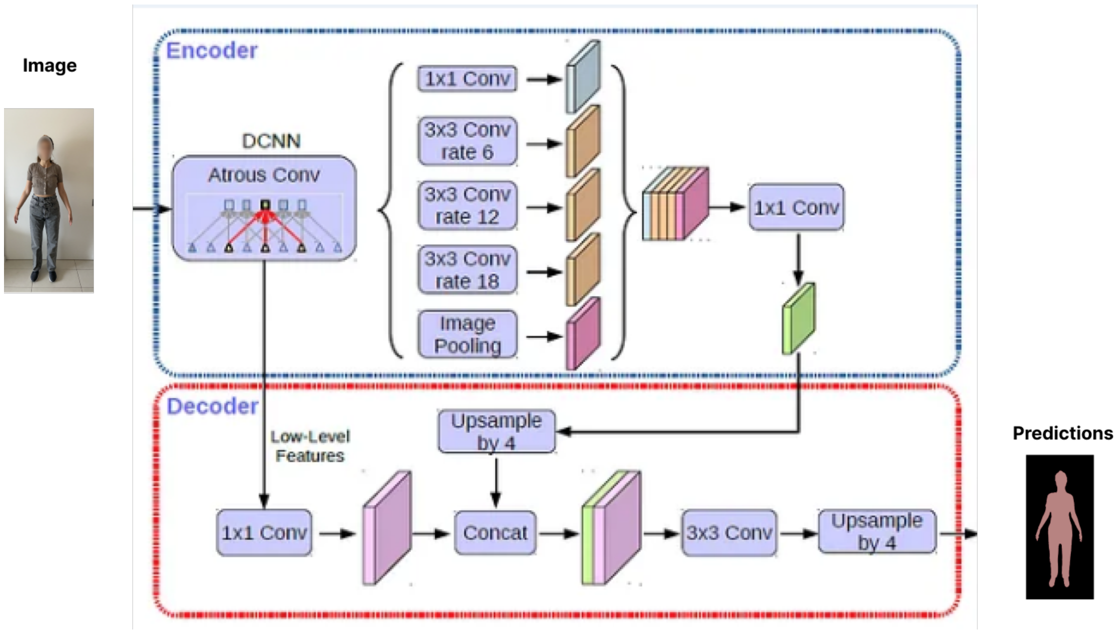


Fig. 6.3 Encoder-Decoder Architecture (with the sample of our software predictions) [153]

In terms of simplicity and GPU usage, classical image segmentation is a better choice for simple software that needs less GPU power. However, if accuracy is a priority, then Deeplab v3 is a better option. Deeplab v3 can achieve much better accuracy than classical image segmentation methods, but it requires more GPU power.

DeepLabV3 is also able to capture multi-scale contextual information, which allows it to accurately segment objects in complex scenes. Yet, this semantic segmentation is better suited to video recording or complicated image frames. As a result, while we are just attempting to improve a small number of frames (two photos), we have decided to employ traditional computer vision, as it greatly aids the software's processing speed and the results are sufficient for measuring human body circumferences[1]. In addition, as mentioned in the previous chapters, our human body subjects will be detected by mobileNet SSD, including parts of their environments. This will greatly simplify the image

[1] A bounding box will be positioned arounf the subject

segmentation, so traditional computer vision methods can be used in certain cases for further analysis.

In overall, Classical image segmentation is better than Deeplab v3 for software that needs less or no GPU because classical image segmentation algorithms are generally more computationally efficient.

6.2 Skin Detection

The objective here is to classify regions within an image that resemble human skin. The process often begins by transforming the given image from the standard RGB (Red, Green, Blue) color space to a color space that separates luminance from chrominance, like the YCbCr or HSV (Hue, Saturation, Value) color spaces. This separation is important as it enables more effective discrimination between skin and non-skin regions, given that skin tones, across diverse ethnicities, mainly differ in chrominance rather than luminance.

The core of the skin detection mechanism often relies on a set of threshold values in the selected color space. For instance, in the YCbCr color space, typical skin color values might be restricted within certain bounds for the Cb and Cr channels. Once the image has been transformed into the chosen color space, every pixel's chrominance values are evaluated against these thresholds to determine whether they fall within the skin color range.

Mathematically, a simple thresholding algorithm can be expressed as:

$$\text{skin}(x,y) = \begin{cases} 1, & \text{if } Cb_{\text{min}} \leq Cb(x,y) \leq Cb_{\text{max}} \text{ and } Cr_{\text{min}} \leq Cr(x,y) \leq Cr_{\text{max}} \\ 0, & \text{otherwise} \end{cases}$$

Where skin(x,y) is a binary mask indicating the presence (1) or absence (0) of skin at pixel location (x,y). $Cb(x,y)$ and $Cr(x,y)$ are the chrominance values of the pixel at (x,y), and Cb_{min}, Cb_{max}, Cr_{min}, and Cr_{max} are the predefined threshold values for skin detection.

However, relying solely on chrominance values can sometimes lead to false positives, as many objects might share color characteristics with human skin. To enhance the accuracy, machine learning approaches, such as support vector machines (SVM) or neural networks, can be employed. These models are trained using labeled datasets that contain both skin and non-skin samples, enabling them to learn a more complex decision boundary that goes beyond simple thresholding.

It's essential to remember that while these methods can be effective, they are not flawless. Factors such as lighting conditions, shadows, or individual variations in skin tones can challenge the accuracy of skin detection algorithms. Regularly updating the training dataset and fine-tuning the model can help mitigate some of these challenges. As described in the previous section, proactive image pre-processing techniques are used in response to these challenges. These adjustments include gray-level mapping, noise reduction, and color calibration. These measures are designed strategically to reduce possible challenges, thus improving the overall resilience and effectiveness of the skin detection procedure.

As part of our seventh objective, we mentioned that it is possible to predict all of the personal characteristics of a human body from two images, such as clothing. Consequently, the primary objective of this section is to first distinguish skin from clothing and then, with the help of ROI as described in the preceding chapter (subsection 5.3.2), improve the image correction procedures. Before we evaluate our region of interest, we proceed as follows:

1. Convert input image after image correction to HSV Color Space

2. Convert input image after image correction to YCbCr Color Space

At this point, the primary objective of the RGB-HSV-YCbCr color space is to identify the human body skin, separate it from clothing, and then isolate the entire body from its surrounding environments. Once these processes are completed, we will train our system to extract the region of interest from the human body. The proposed body contour extraction flowchart is displayed in Figure 6.4. This section focuses on the properties of the human body and how to efficiently and effectively extract the human body.

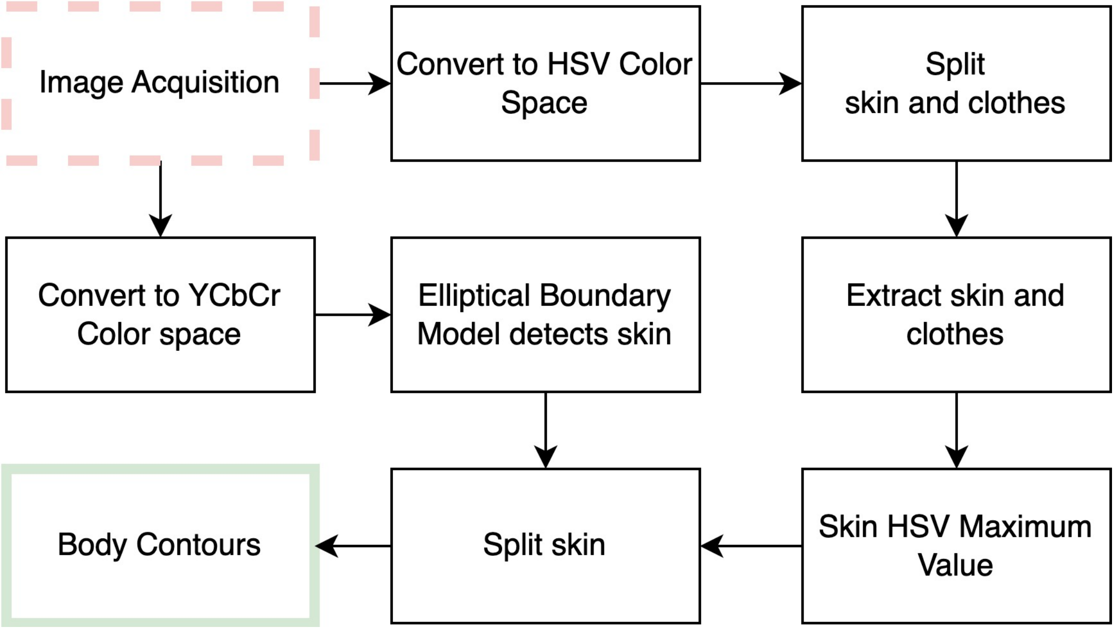


Fig. 6.4 : Body contour extraction flow chart

Our method was inspired by the influential study on the detection of human skin with a combined RGB-HSV-YCbCr color space model [94]. We initiated an effort to enhance and broaden the scope of this field of research. The foundation of their study was based on the application of a threshold color space model for the purpose of detecting human facial skin. The selected methodology employed a process that transformed an image into a two-dimensional matrix, where each element of the matrix corresponds to a pixel in the image. Upon careful examination, the ARGB[2] color of each individual pixel was determined. The 32-bit ARGB value was carefully adjusted to obtain distinct sub-values including red, green, blue, and alpha. The significance of the alpha channel was underscored, highlighting the difference in opacity levels for individual pixels. Afterwards, a system of threshold values was developed in order to determine whether a pixel corresponds to human skin.

However, the paper's primary focus was the facial aspect of human skin detection. Beyond facial analysis, our methodology seeks to conduct a thorough examination of the entire human body. In doing so, our primary objective is not only to distinguish skin from its surroundings, but also to distinguish it from apparel.

Our approach expands the horizons of this study by transitioning from a localised facial analysis to an exhaustive examination of the entire human body. We move into the domain of whole-body skin detection by separating the human body's skin from clothing materials. In the subsection 6.2.4, we shall go deeper, explaining the processing of our skin detection algorithm. But first, let's review the color space in greater detail.

The color space is a mathematical representation of color information as three or four different color components. For various applications, such as computer graphics, image processing, TV broadcasting, and computer vision, several color spaces (models) are utilised. There are various color spaces accessible for skin detection. They are RGB-based color spaces (RGB), Hue-based color spaces (HSI, HSV, and HSL), and Luminance-based color spaces (YCbCr, YIQ, and YUV), [167]. These models are explained in the succeeding sections. The selection of the color space is the first step in skin color modelling and classification. One or more color spaces can provide the best threshold value for skin detection in a picture. The choice of color space is frequently affected by the skin detection technique and the application. We utilise the following

[2] ARGB represents a color model that includes Alpha, Red, Green, and Blue channels. "A" stands for Alpha, which determines the opacity or transparency level of a pixel.

color spaces for skin pixel recognition: RGB Color Space, HSV Color Space, YCbCr Color Space. For more details on colorimetry refer to the book by Koenderink [93].

6.2.1 RGB Color Space

RGB is an additive color model in which red, green, and blue light are combined in a variety of ways to create an extensive color palette. RGB color is utilised in the creation of digital images and videos because it is the most efficient method of representing the entire color spectrum. Additionally, it is the most popular color space for digital photography, computer graphics, and television.

In the early days of television, when engineers needed a way to display color images on a black-and-white television, they developed the RGB color model. Engineers were able to create a spectrum of colors by combining red, green, and blue lights of differing intensities. Since then, the computer industry has widely adopted this color model, which is now used in virtually all digital images and videos.

The RGB color model uses three primary colors—red, green, and blue—and assigns each of these primary colors a numerical value as shown in Figure 6.5. This value is used to create a spectrum of colors ranging from the darkest black to the brightest white. For instance, if the numerical value of all three primary colors is 0, the resulting color is black. The result is white if all three primary colors have a numerical value of 255 (the maximum value that can be assigned to each primary color). By adjusting the numerical values of each primary color, it is possible to create a variety of hues.

Depending on how much is taken from each base color, any color can be created. Reversing this technique, a specific color can be broken down into its red, blue and green components as shown in equation Equation 6.2

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \tag{6.2}$$

RGB color model has its strengths and weaknesses. Its strengths include its direct representation, as it signifies the primary colors used in digital imaging directly, making it easy to extract and work with. Another advantage is its ubiquity; most digital images are in RGB by default, and the primary color values are straightforward to obtain without any conversions. However, it has its downsides. One significant weakness is its sensitivity to lighting variations, where changes in the environment can result in considerable variations in RGB values. Another drawback is its lack of human-centric perception, as our eyes perceive color and intensity differently from the linear progression of RGB values.

6.2.2 HSV Color Space

The HSV color model is close to the human color-aware simulation model: H stands for chromaticity (a measure of the composition of the color spectrum), S for saturation (the pure wavelength ratio in the main wavelength) as shown in Figure 6.5, indicating the degree to which two colors are the same brightness, and V for value (relative to the brightness of white light). As shown in Equation 6.3, the RGB color space can be converted into the HSV color space.

$$V = \max(R, G, B)$$

$$S = \begin{cases} \frac{V - \min(R,G,B)}{V} & , \text{ if } V \neq 0 \\ 0 & , \text{ if } V = 0 \end{cases}$$

$$H = \begin{cases} \frac{60\times(G-B)}{V-\min(R,G,B)} & , \text{ if } V = R \text{ and } G \geq B \\ \frac{60\times(G-B)}{V-\min(R,G,B)} + 360 & , \text{ if } V = R \text{ and } G < B \\ 120 + \frac{60\times(B-R)}{V-\min(R,G,B)} & , \text{ if } V = G \\ 240 + \frac{60\times(R-G)}{V-\min(R,G,B)} & , \text{ if } V = B \end{cases}$$

$$H = H + 360 \ , \quad H < 0 \tag{6.3}$$

Examining the image's HSV channels as well as its histogram is the first step in separating the various parts of the human body from one another. The use of HSV channels has a number of advantages, the most prominent of which are that these channels are great for describing images that are difficult to automatically segment and do not need for the spatial positions of objects to be taken into consideration. It can be very difficult to directly find the threshold of the human body based on the histograms. It is possible to utilise H as a color comparison tool since H is essentially unaffected by changes in the lighting; this enables objects to be differentiated from their colorful backgrounds.

HSV color model one of its primary strengths is its human-centric perception, as HSV provides a color representation more aligned with how humans interpret color. It distinguishes the color aspect (hue) from its intensity (value) and vividness (saturation). Another strength lies in its ability to handle lighting variations. Due to the separation of hue from value, changes in brightness, such as from a shadow, might modify the value but not notably impact the hue. However, HSV isn't without its challenges. One main weakness is associated with saturation extremes. At very high or low saturations, the hue can become unreliable. Specifically, in regions of very low saturation, colors might tend to shift toward grayscale.

6.2.3 YCbCr Color Space

The YCbCr color space is ofthen used to identify objects of interest within an image. YCbCr stands for luminance (Y), chrominance blue (Cb), and chrominance red (Cr) as shown in Figure 6.5. The YCbCr color space is commonly used in digital video and image processing, as it allows for the separation of luminance and chrominance signals, which allows for the identification of objects within an image, as well as the removal of unwanted objects.

In order to select a region of interest in a mask image YCbCr, a threshold must be set in order to determine which pixels will be classified as part of the region of interest. This is done by selecting a region of interest in the YCbCr color space, such as a specific range of luminance and chrominance values. Any pixels within this range are then considered to be part of the region of interest.

Once the region of interest has been selected, a mask can then be created. This mask will be used to remove any unwanted objects from the image. The mask will contain only the pixels that have been selected as part of the region of interest. All other pixels will be ignored.

YCbCr color model most notable strength lies in the luminance-chrominance separation. The color space uniquely divides the brightness of a color (luminance) from its color particulars (chrominance). Such a distinction becomes especially beneficial in situations where one aims to detect features independent of the lighting conditions. Additionally, due to YCbCr's approach of handling color and brightness as separate entities, it can yield a more consistent color representation under varying illuminations.

Despite these advantages, there's a challenge associated with YCbCr. Specifically, its weakness revolves around complexity. The process of converting to and understanding YCbCr values tends to be somewhat more intricate than RGB.

The integration of RGB, HSV, and YCbCr allows for a much more comprehensive approach to skin detection.

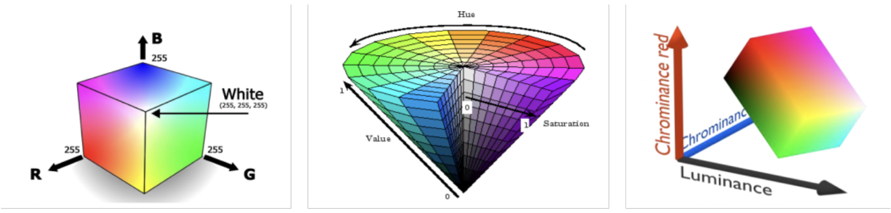
Complementing Information: Each color space provides a unique view of the image. While RGB offers a raw view of the colors, HSV provides a human-centric perspective, and YCbCr gives a separation of luminance and chrominance. Combining the three ensures that the diverse facets of color and brightness information are considered.

Mitigating Weaknesses: The shortcomings of one color space can be offset by the strengths of the others. For example, while RGB might struggle with lighting variations, the hue in HSV remains relatively consistent, and YCbCr's separation further aids in maintaining color consistency.

Enhanced Precision: By simultaneously analysing the same pixel across multiple color spaces, the algorithm can make more informed decisions. If two spaces indicate a pixel is likely skin while one does not, the algorithm can weigh this information and make a better-informed decision than if relying on a single space.

Versatility: Different images might have different quirks. Some might be more affected by lighting, some might have colors that are hard to differentiate in one space but not the others. By using all three spaces, the algorithm ensures it is equipped to handle a wide variety of image challenges.

In essence, while each color space can offer valuable insights individually, their combined strength provides a more robust and adaptable framework for skin detection across diverse scenarios and challenges.



(a) RGB (b) HSV (c) YCbCr

Fig. 6.5 : Color mode [94]

6.2.4 Skin Detection Algorithm

Many skin detection techniques have been proposed over the years. By analysing various studies and techniques [162], [94], [21], [122], certain similarities and patterns regarding effective thresholds in the HSV and YCbCr color spaces become evident. The initial values for our thresholds were informed by these findings, giving us a reliable starting point, which will be covered further in this section. To validate and refine these initial values, a diverse dataset comprising images of individuals of varying skin tones, under different lighting conditions, and from multiple sources was assembled. This dataset ensured our thresholds would be robust across varied scenarios.

Using our dataset, we began the process of experimental testing. The idea was to adjust these threshold values and observe the performance of the skin detection technique. Metrics like accuracy, false positives, and false negatives were used to measure effectiveness. Firstly, for each color space channel (H, S, V, Y, Cb, Cr), we adjusted the thresholds iteratively, optimizing for the best balance between precision (minimizing false positives) and recall (minimizing false negatives), and afterwards, visual inspections were also conducted to ensure the results were in line with human perception.

Based on the performance metrics, the thresholds were established this involved fine-tuning the value to enhance the balance between detecting actual skin regions and reducing the inclusion of non-skin regions. We began with broader ranges and, using feedback loops, systematically narrowed down the values that produced the most accurate and consistent results. Based on our experimental findings, we identified optimal values for detecting human skin. These values will be detailed in the following paragraphs.

The core function of skin detection algorithm takes in an image in the BGR color space and a key skin value in the YCbCr space. The image is first converted to the HSV color space, then to BGR, and finally to YCbCr. A mask is generated based on the Euclidean distance between the pixels in the Cb and Cr channels and the key skin value. This mask represents the detected skin regions.

The RGB image is first converted into two separate color spaces: HSV and YCbCr. The conversion from RGB to HSV is given by:

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B > G \end{cases} \tag{6.4}$$

Where $\theta = \arccos\left(\frac{1}{2}\left(\frac{R-G}{R-B}\right)\right)$.

Saturation S and Value V are calculated as:

$$S = 1 - \frac{3}{R+G+B}\min(R,G,B)$$
$$V = \frac{1}{3}(R+G+B)$$

For YCbCr, it separates the luminance (Y) from the chrominance (Cb and Cr). The conversion is primarily a linear transformation from RGB:

$$Y = 0.299R + 0.587G + 0.114B$$
$$Cb = -0.169R - 0.331G + 0.500B$$
$$Cr = 0.500R - 0.419G - 0.081B$$

Each pixel's HSV and YCbCr values are then tested against empirically determined thresholds to check if they likely belong to skin or not. The criteria are:

$$\begin{aligned}
80 \leq & H \leq 255 \\
50 \leq & S \leq 255 \\
0 \leq & V \leq 255 \\
0 \leq & Y \leq 255 \\
80 \leq & Cb \leq 120 \\
133 \leq & Cr \leq 177
\end{aligned}$$

The skin detection step involves checking every pixel of the image, resulting in a time complexity of $O(n \times m)$, where n and m are the height and width of the image, respectively.

Afterwards, a binary mask is generated where pixels that satisfy the above conditions are set to white (255) and others remain black. This results in a preliminary mask where white regions indicate detected skin areas. To refine the mask and reduce noise, morphological operations are employed:

- **Opening**: An erosion followed by dilation. This operation aims to eliminate small white noise present in the mask. Mathematically, the operation can be denoted for an image I and a kernel K as:

$$\text{Opening}(I) = \text{Dilation}(\text{Erosion}(I,K),K) \tag{6.5}$$

- **Closing**: A dilation followed by erosion, aiming to close small holes in the detected skin regions:

$$\text{Closing}(I) = \text{Erosion}(\text{Dilation}(I,K),K) \tag{6.6}$$

The final processed mask is displayed, where the white regions represent detected skin, offering a visual verification of the skin detection process. This method presents a straightforward yet efficient technique for skin detection. By leveraging the discriminative power of the HSV and YCbCr color spaces and refining the results through morphological operations, it offers a robust solution for our software. The algorithm's steps are shown in flowchart form in Figure 6.6, which may be seen here.

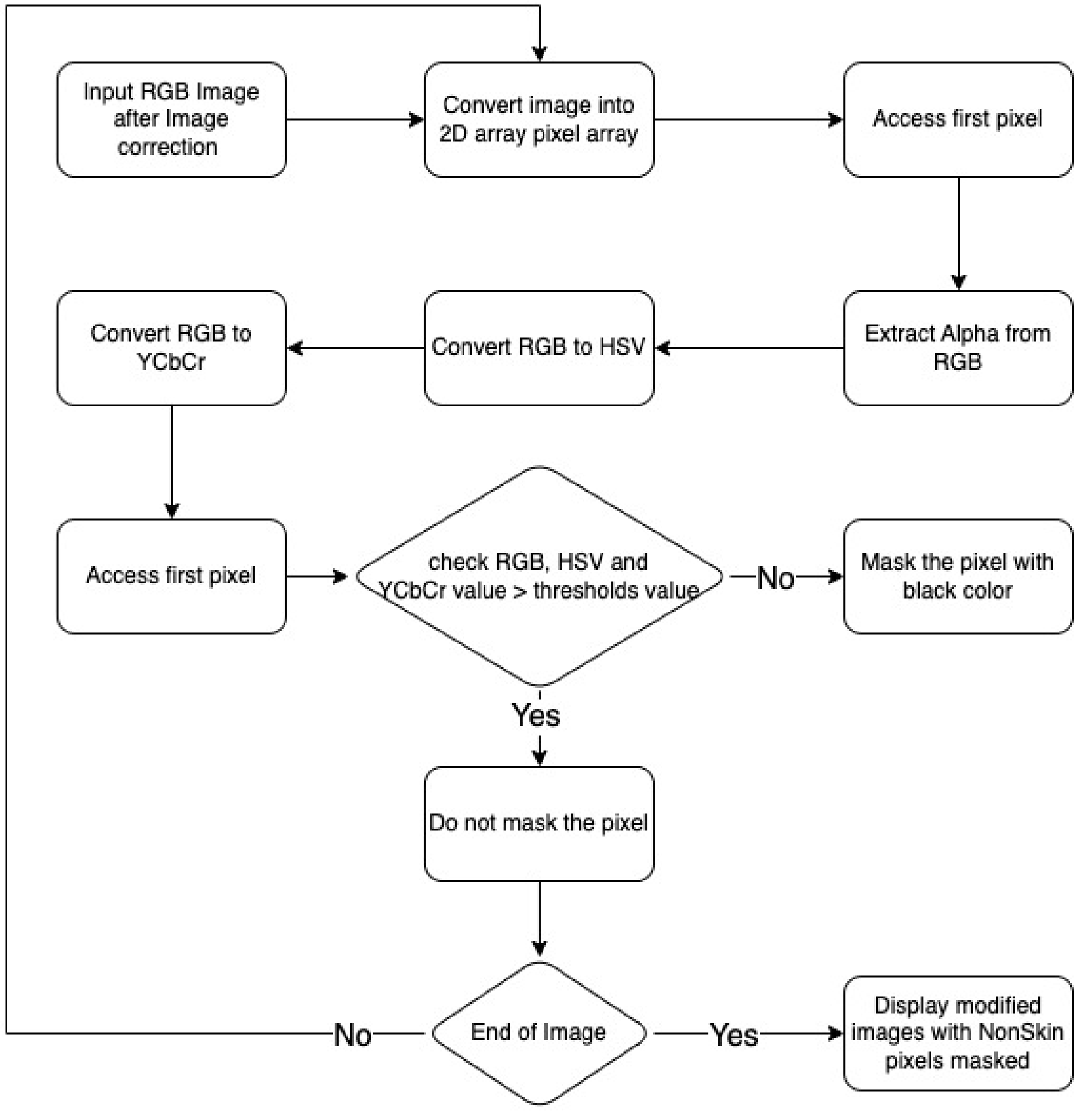
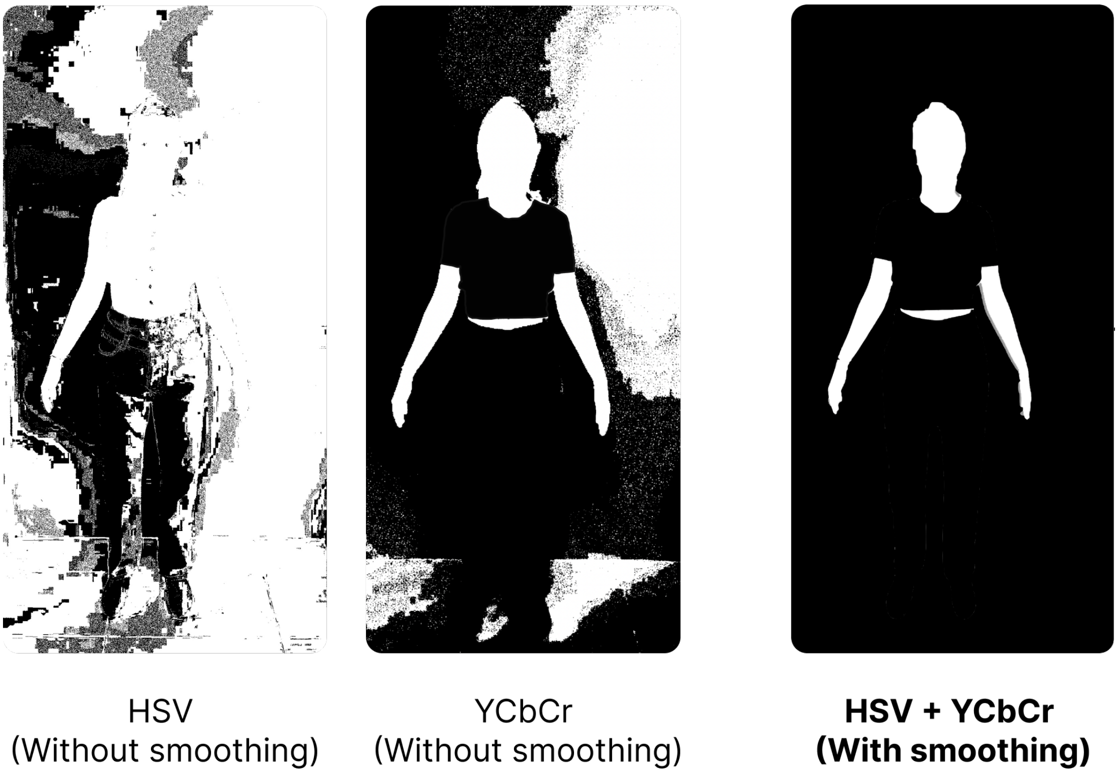

Input RGB Image after Image correction
Convert image into 2D array pixel array
Access first pixel
Extract Alpha from RGB
Convert RGB to HSV
Convert RGB to YCbCr
Access first pixel
check RGB, HSV and YCbCr value > thresholds value
No
Mask the pixel with black color
Yes
Do not mask the pixel
End of Image
No
Yes
Display modified images with NonSkin pixels masked

Fig. 6.6 Flow chart of the skin detection



HSV
(Without smoothing)

YCbCr
(Without smoothing)

HSV + YCbCr
(With smoothing)

Fig. 6.7 is divided into three parts. The leftmost image captioned "HSV skin detections without any smoothing" shows HSV skin detections without any smoothing. In the middle, we have the YCbCr representation itself without any smoothing, and on the right-hand side, the caption reads "YCbCr + HSV with image smoothing," representing the combination of YCbCr and HSV with the application of image smoothing.

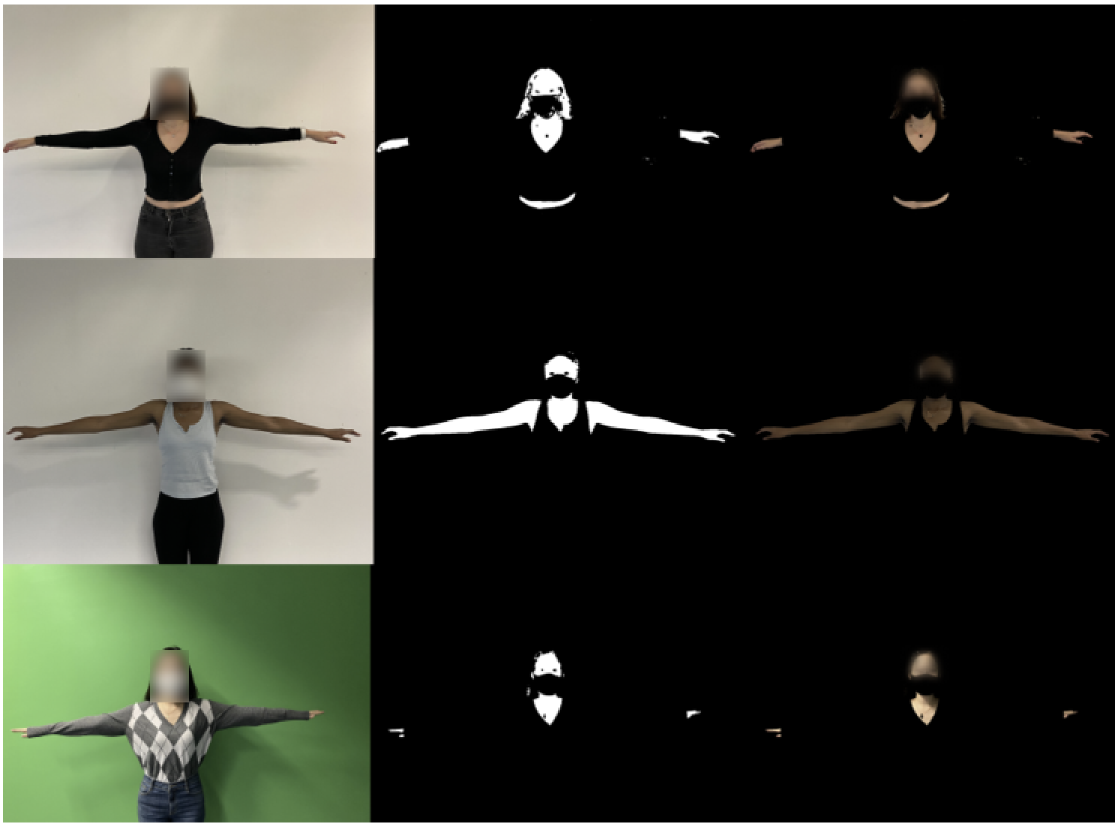


Fig. 6.8 Experimental results on sample images. The code first converts the input image to masks in both YCbCr and HSV colorspaces using predefined threshold values. Then, it combines these masks to identify potential human skin regions. Finally, it refines the skin detection by applying the Watershed algorithm and morphological operations [73], resulting in a final skin mask that can be overlaid onto the original image to highlight the skin regions.

While many traditional methods focus mainly on face or hand detection, our method pivots towards the apparel industry. We utilise skin detection for precise human body measurements. This assists in efficient image processing, enabling superior segmentation by eliminating disruptions like noise, lighting conditions, and shadows. Our method processes HSV and YCbCr simultaneously and combines the results. This synchronised technique saves processing overhead and increases detection accuracy by using both colour fields, While the other techniques move between different colour spaces.

Unlike some recent techniques that necessitate high computational power or GPUs, our method is designed for speed. This translates into faster processing times, making it more feasible for real-time applications. Our current limitation is the size of our dataset. However, we recognise this challenge and are working on to collect a more extensive dataset. This will make the way for potential training and refinement, thereby improving our system's robustness. Our method is built around being open to all ethnicity. We have tried our skin detection algorithm on people of many different races, and it has always given us accurate results, which shows that it can be used by anyone.

6.3 Camera Calibration and Distance Estimation

Accurate image measurements require camera calibration. The focal length, principal point, and distortion coefficients are its core. Calibration usually involves taking multiple images from different angles and distances. Software libraries such as OpenCV and SciPy are prominent for their calibration utilities. The OpenCV library, in particular, provides functions that calculate both a camera's intrinsic parameters and the affine transformation between pair of images, which is essential for 3D reconstructions and measurements.

During our initial research phase, multiple calibration techniques were evaluated including OpenCV's Camera Calibration Algorithm, Zhang's Method, and the Direct Linear Transformation (DLT) [173, 41, 32]. Out of these, the OpenCV Camera Calibration Algorithm proved most effective, mainly due to its combination of feature detection, optimization, and calibration techniques. It also stands as the most accurate method. Calibration is needed for determining the camera parameters, including the intrinsic 3x3 matrix K, the 3x3 rotation matrix R, and the 3x1 translation vector t using a set of known 3D points and their corresponding image coordinates. The camera is considered calibrated when both intrinsic and extrinsic parameters have been precisely estimated.

Intrinsic parameters are determined by assessing the image size in pixels and the size of the object within the image. We leverage the checkerboard pattern for its orthogonal and regular geometry, which makes it recognizable for both users and computer vision systems. Modern calibration systems display a live feed of the camera's perspective, allowing users to adjust the checkerboard's orientation towards the camera. The calibration software automatically detects the checkerboard pattern in the captured image. If recognized, it confirms that the pattern is visible and approximately aligned with the camera. Else, the user may be prompted to adjust it. Furthermore, these systems are designed with a built-in tolerance, ensuring accurate calibrations even if the checkerboard isn't perfectly aligned.

While the checkerboard is effective, its practicality for everyday users is limited. Recognizing this, we transitioned to using the user's height as a calibration reference. Just as the checkerboard's corners serve as identifiable points, the user's height provides a tangible reference for conversion from pixel units to real-world measurements. However, our research indicates that around 70% of users may not know their exact height, introducing potential calibration inaccuracies. Hence, we champion alternative, standardized reference points to enhance precision. To address this, we're exploring using universally accessible and standard-sized cards, like credit or bank cards, as calibration tools.

Our literature review reveals that many applications now use the user's height for calibration. However, they often overlook potential inaccuracies stemming from users' inaccurate height knowledge. In our experiments, accurate height inputs align well with checkerboard-based measurements. Deviations in height inputs, however, can reduce accuracy. Additionally, it's worth noting that some existing applications employ a two-step calibration process. They not only rely on user-provided height but also instruct users to stand within a specific distance from the camera's point of view. Users are directed to move forward and backward within the app until they are positioned at their ideal distance from the camera. These distance measurements are then used in conjunction with the user's height input to convert pixel units to centimeters, further contributing to the overall calibration process. For a comprehensive discussion on the calibration methods and detailed insights from the research and the reasons for selecting this method, please refer to the Appendix E.1.1.

6.4 Selecting Markers

The primary objective of our feature extraction phase is to identify specific points of interest — we call markers — from the detected body outline segments for each body part. The 3D measurements of various horizontal body regions, including the waist, chest, hips, and shoulders, can then be estimated based on these markers. We identify markers in pairs, along horizontal body "slices". To accomplish this, we first find the best vertical body line to utilise as an approximate mirror-like splitting axis. Then, we search for the right and left pairs of body extremities in each ROI, i.e., pixel locations at the body's edge.

Initially, the top central head point—head tip—is identified from the "height" ROI, serving as the reference to center the vertical splitting line. Subsequently, one can walk along a body contour segment, for each ROI, on one side (left or right) and monitor the slope. When this slope goes through a large change in value, such as for the neck and shoulder ROIs, a potential marker location is determined. Alternatively, a locus mid-way along a contour segment is selected, which proves useful for other ROIs (see Figure 6.9.).

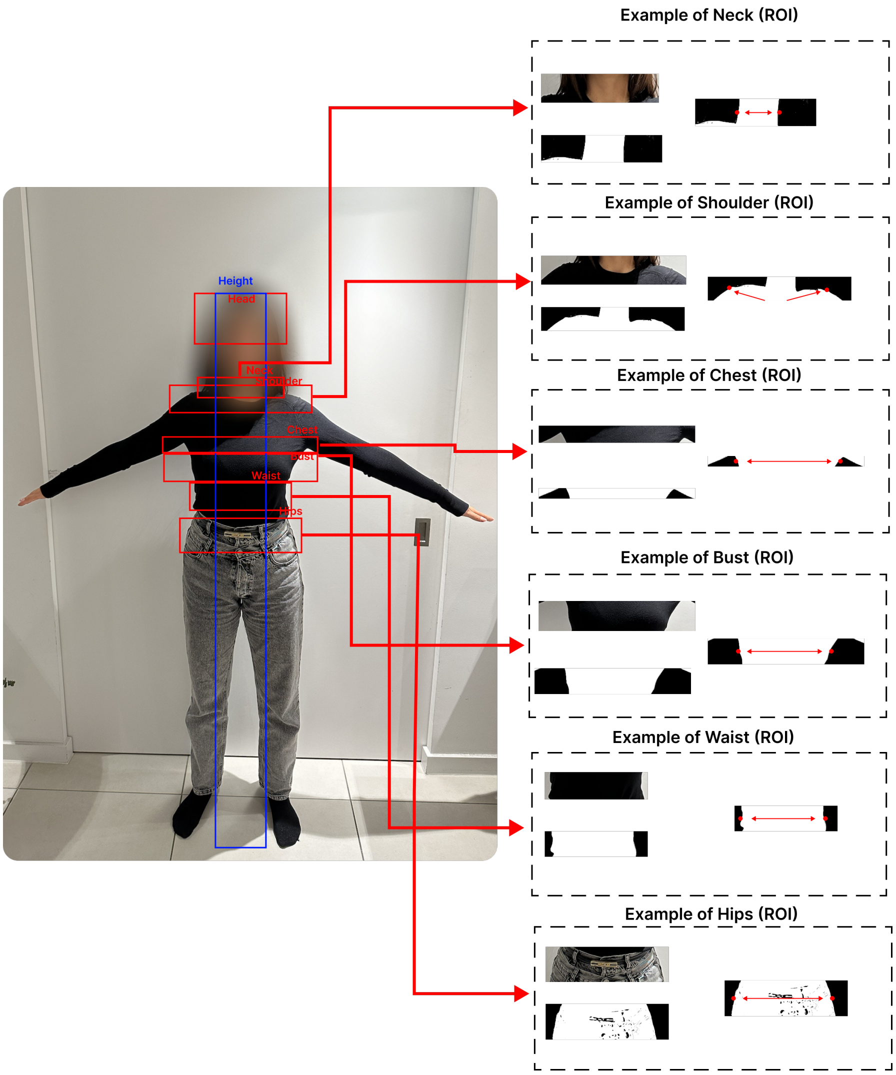


Fig. 6.9 Illustration of the procedure for *automatic* marker detection, applied to individual ROIs. A proper A-pose leads to a useful automatic marker localisation (red dots, also indicated by arrows for greater visibility).

A collection of measurements from the ISO 8559 [13] standard, which prescribes the location of anthropometric measurements used in the production of physical and digital anthropometric databases, has been selected in order to compare the suggested method with other state-of-the-art approaches.

Currently, in the system, there's the provision for a semi-automatic method where a user can move the proposed (detected) markers of one ROI either horizontally, along the current line joining markers, or by first moving vertically that same line along the main body axis. This proves useful, especially when the automatic method may fail, such as when the A-pose is too “weak”, e.g., when the arms of the user are kept too close to their chest, such that an armpit is not clearly visible (see Figures 6.10 and 7.7).

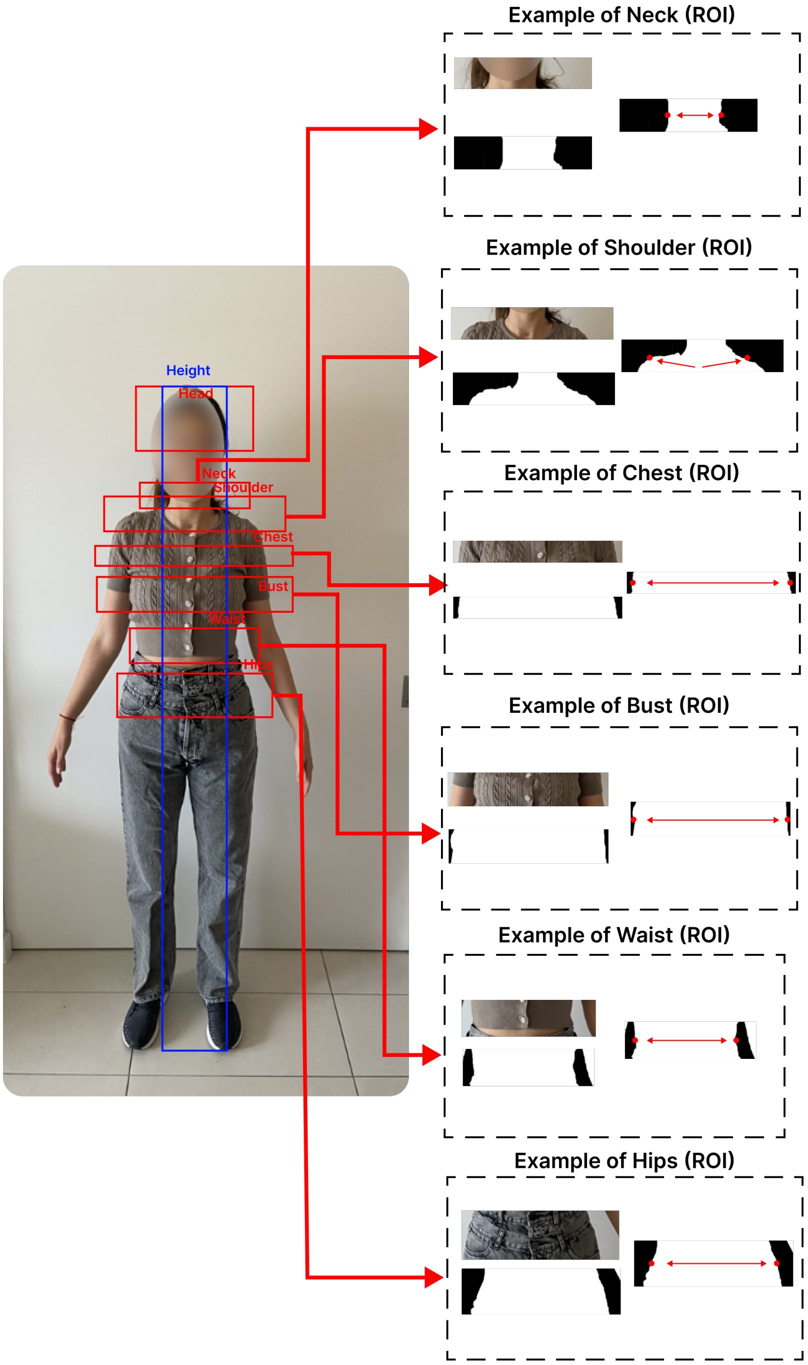
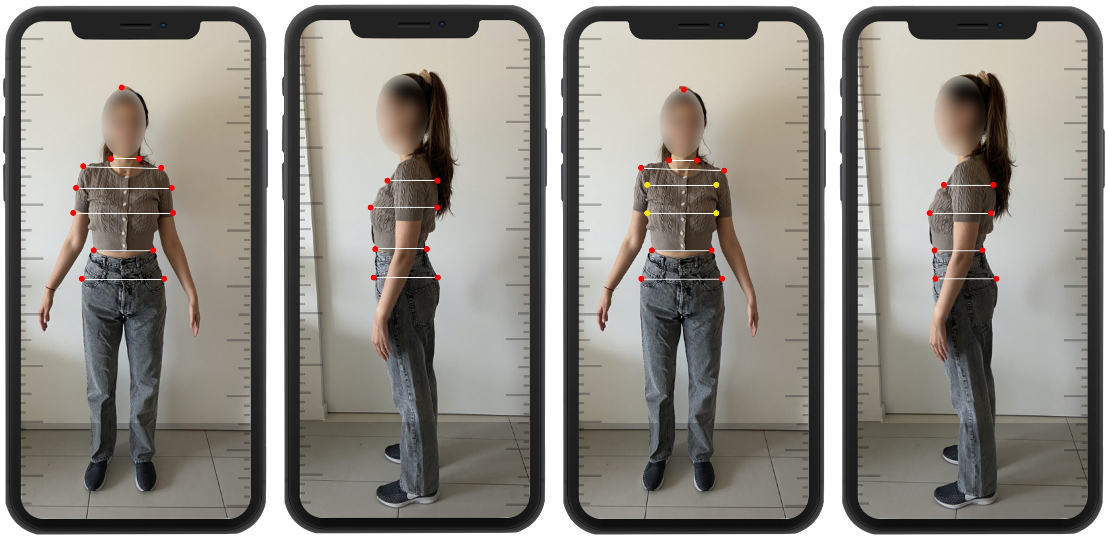


Fig. 6.10 Illustration of the procedure for automatic marker detection (red dots), which leads to incorrect localisation in some cases. An incorrect A-pose (with the arms kept too close to the upper body) results here in the inaccurate automatic detection of some markers, near the chest and bust areas.



Automatic marker detection

Automatic marker detection

User moves some markers

User accepts provided markers

Fig. 6.11 Automatic and interactive localisation of detected markers. The first two images on the left illustrate the markers automatically detected (red dots) for the front and side views. The other two images on the right correspond to the final marker localisation after the user has modified some markers. Only the third image (front view) sees two pair of markers having been moved horizontally (yellow dots); other markers are judged fine by the user and kept as provided by the system.

Note that this step could also be performed automatically in adverse cases, by using an ML-based method; however, it would require additional training data, which has yet to produce. In contrast, having a semi-automatic method for the selection of markers permits the user to take back some control on the system, which is often seen as a desirable feature, particularly by fashion designers or tailors.

After obtaining the set of marker pairs, the circumferences of the upper human body for each horizontal slice can be approximated by fitting an ellipse to the data points, using both the frontal and side views. The idea of using an ellipse as a useful geometric model of human horizontal body contours has been previously validated [181]. By evaluating the semi-axes of the ellipse from the two images (front and side), the circumference of a human body "slice" with respect to the selected marker pair can be estimated with sufficient accuracy (e.g., waist circumference).

Note that for our domain of application in the fashion sector, using an ellipse-like form tightly fit to a body horizontal contour (slice) is what is needed in most cases: i.e., the goal is not to have a piece of clothing follow too closely all the body contour details and deformations when considering or designing a pattern for cutting the fabric to be assembled into a wearable piece. For a different application domain, such as the medical realm, recovering detailed contour measurements might be desirable and the ellipse fitting process may only provide an initial approximation. Therefore, in the upcoming chapter, the focus will be on computing the perimeter of fitted ellipses and understanding its importance within the context of human body circumference measurements.

In conclusion, this chapter has focused on improving the quality and appearance of uploaded images for data collection purposes. The challenges posed by cluttered backgrounds, noise, and shadows were addressed, and two image segmentation approaches were explored: classical computer vision image segmentation and DeeplabV3. The results and discussions have highlighted classical computer vision image segmentation as the optimal choice for the software solution. Additionally, skin detection techniques using various color spaces were discussed, along with the outcomes of experiments comparing plain and cluttered backgrounds. The camera calibration process using OpenCV's functions was detailed, and the ideal distance range of 0.5 to 3 meters for accurate data collection was identified. Please refer to section 8.1 for a comprehensive review of the summary, results, and discussion in this chapter.

Chapter 7

Evaluation of using Ellipses for Human Body Measurements

This chapter focuses on finding accurate measurements for human body horizontally aligned perimeters through the fitting of ellipses. We begin by evaluating six different elliptical mathematical models and discussing the findings. We then describe how the selected equations are integrated into our system. Finally, we present results, comparing the six elliptic equations and identify a pair of best equations for estimating human body horizontal perimeters.

Based on these results, a fully trained system is developed that can choose the best ellipse equation based on the human body shape. We conclude with a summary of the findings and their implications for the development of the proposed software solution. By using ellipse equations that are well fitted to the body shape, we are able to accurately estimate horizontal perimeters. Overall, this chapter emphasizes the importance of choosing the right mathematical model to improve the accuracy of measurements in the human body.

Two types of dimensions can be calculated in our proposed method, one type includes shoulder and sleeve lengths. The second type includes the perimeters of the chest, bust, waist, hip, and other useful horizontal “slices”, which can be well approximated by ellipses of varying eccentricity Yao et al. [181]. However, the real perimeters of human subjects are only approximately elliptical (Figure 7.1). The challenge is to discover the best fit. Unlike a circle, there is no closed form solution for exactly calculating the ellipse’s perimeter; only approximations, more or less complex, exist.

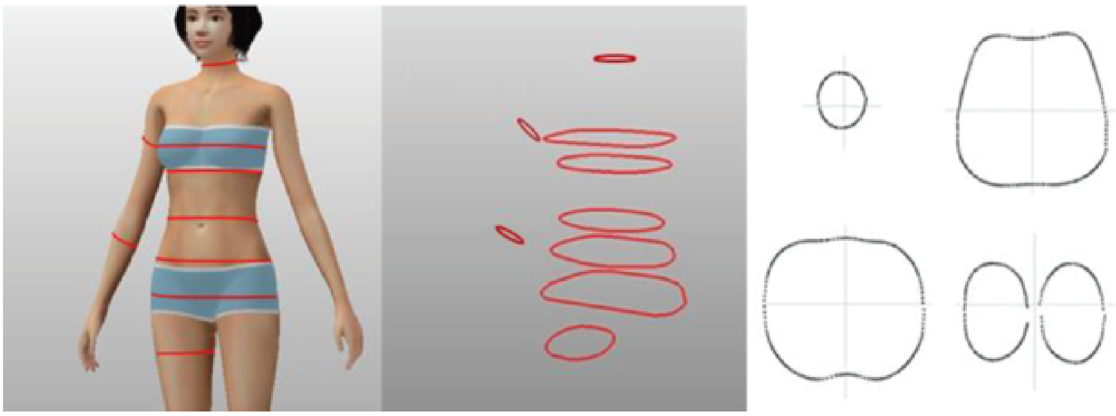


Fig. 7.1 Female template chosen from ISO, image 20947 [10]. Examples of virtual body horizontal cross sections.

When it comes to the challenging task of accurately capturing the human body’s horizontal perimeters, various curves can be considered, such as Bezier curves and splines, cardioids, ovals, and more. Our comprehensive research led us to prioritize the use of ellipse equations. There are several reasons for this decision, as we discuss next.

Firstly, the ellipse, in its essence, is a straightforward curve. Only the major and minor axes are needed to define it, thereby streamlining the measurement process. This inherent simplicity becomes indispensable, especially when processing 2D image analyses; by reducing complexities, we can achieve faster and more efficient results.

Afterwards, we recognized that while an ellipse might not perfectly represent every section of the human body, it has shown commendable accuracy in a wide array of scenarios. For instance, in our examination of the breast area, known for its significant variability among individuals, we found the cardioid formula might fall short in some cases. The ellipse, in comparison, provided a more general fit.

Furthermore, our choice of the ellipse does not stop one from exploring other geometrical shapes or fitting models. The ellipse can be used as a basis or reference that can be complemented or refined if needed by more advanced image analysis techniques. As we gather a larger and more varied dataset, we envision delving into other geometric formulas, from complex curves to splines and other geometrical primitives. Such explorations — beyond the scope of this thesis — hold the promise of more finely capturing the human body to address other potential application domains, such as in bra design and fit, of health purposes.

Lastly, an often overlooked factor in this discussion is the innate elliptic (and symmetric) shape evident in many clothing designs when viewed slice-wise. Given that clothing serves as a primary application point for these measurements, this observation naturally supports our preference towards ellipse equations.

As we already mentioned, perimeters of ellipses cannot be calculated exactly using a standard (aka closed form) formula. However, numerous approximation formulas have been proposed in the past. In order to calculate the upper human body circumferences in this study, we investigated six different formulas to approximate the perimeter of an ellipse, with the goal to identify the most accurate ones in comparison to tape measurements (used a ground truth). Next, we summarise how we identified the two best formulae to use for our application domain. Additional details can be found in Appendix F.1.1.

7.1 Selection of best formulae to compute elliptic perimeters

First, we notice the large range of human body shape variations which is apparent when considering horizontal slices from accurate 3D scans (Figure 7.2). Also, we re-emphasise that the human upper body (from the hips and up) can be well approximated by mixing ellipses with varying eccentricities. We also know that the differences between each shape can impact the final estimation (Figure 7.5).

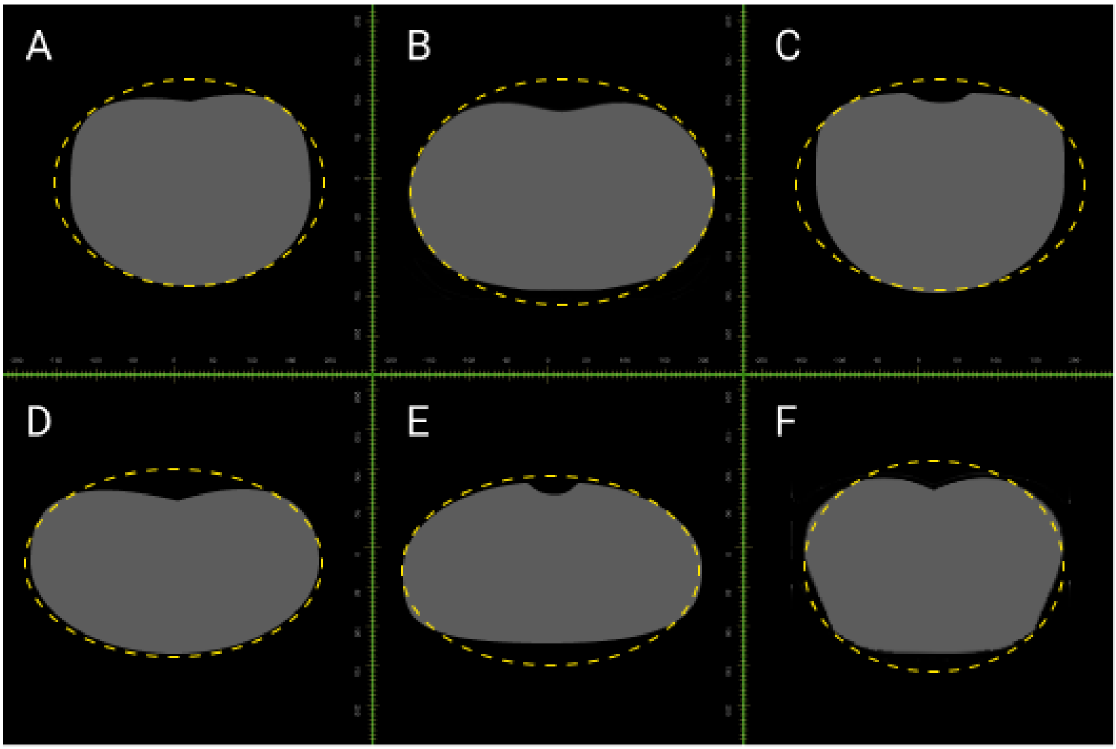
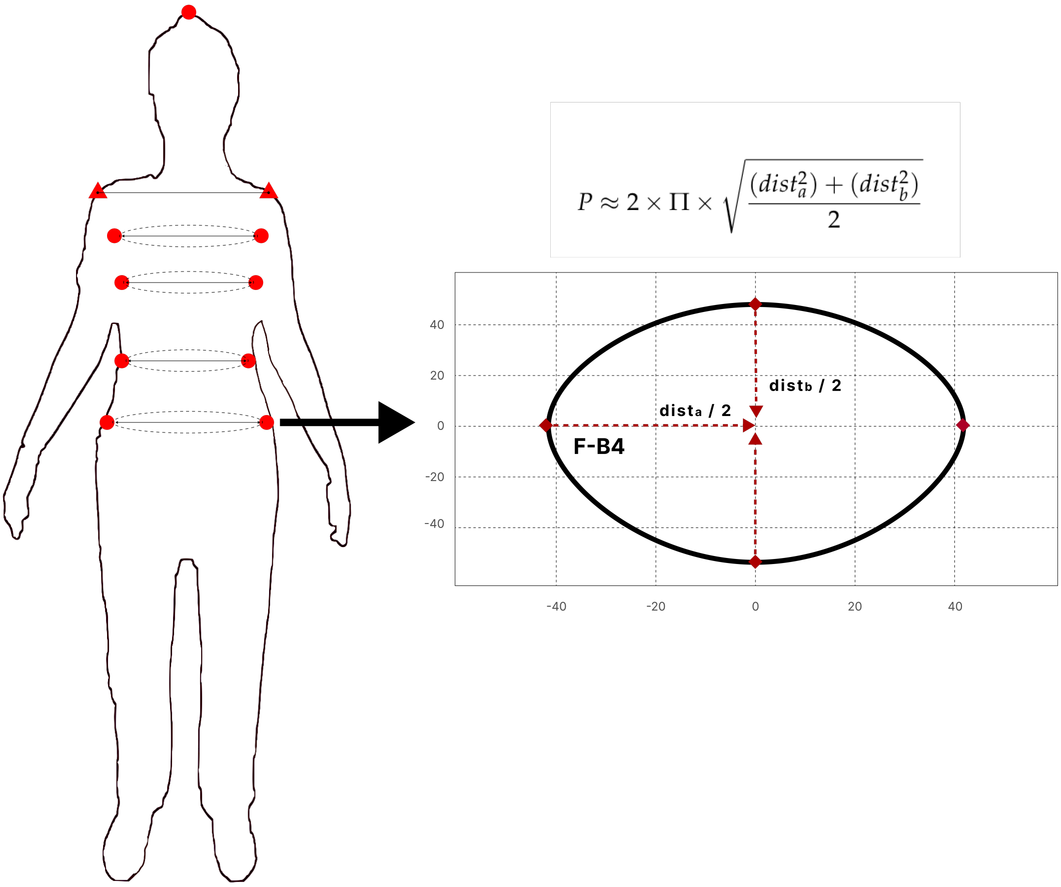


Fig. 7.2 shows the participant shapes for which We collected data during our capture procedure. We used PifuHD (for more information please refer to subsection 4.2.5) to turn the photos into a 3D model (.fbx file), and then we used Autodesk Maya to separate the body parts from the 3D model.



$$P \approx 2 \times \Pi \times \sqrt{\frac{(dist_a^2) + (dist_b^2)}{2}}$$

Fig. 7.3 For individuals with a body shape resembling a circle from the hips upward, we use this equation.

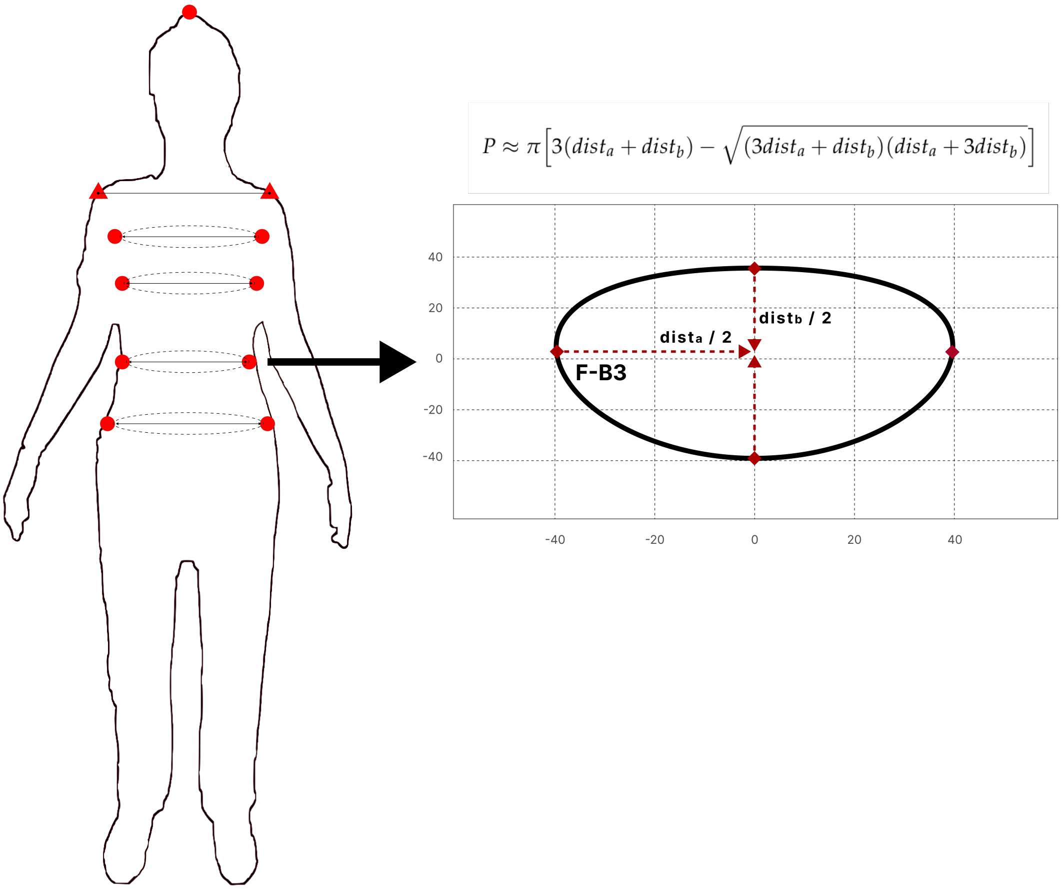

$P \approx \pi \left[3(dist_a + dist_b) - \sqrt{(3dist_a + dist_b)(dist_a + 3dist_b)} \right]$
40
20
0
-20
-40
-40
-20
0
20
40
distb / 2
dista / 2
F-B3

Fig. 7.4 For those with a more distinctly elongated elliptical shape from the hips up, we use this equation.

In our comparison presented in Figure 7.5, Ramanujan's formula, Equation 7.3 [20, 33], stands out alongside another favorite of ours, Equation 7.2, for its simplicity and high accuracy in representing different human body shapes as detailed in subsection 7.2.1 (Figure 7.9). This figure charts various ellipse equation approximations with the x-axis depicting the ellipse's major axis and the y-axis showing the error rate. The shape of the ellipse greatly influences the approximation's accuracy. For instance, Equation F.1 is more accurate for shapes resembling circles. However, its accuracy reduce as the ellipse's major axis lengthens. In addition, Equation F.6 and Equation F.5a consistently offer the best accuracy among all equations in the figure. Some methods for approximating ellipses involve never-ending series of calculations which, while slightly more accurate, can slow down software and don't offer a significant advantage, as shown in Figure 8.9. We prioritize methods that are both fast and nearly as precise.

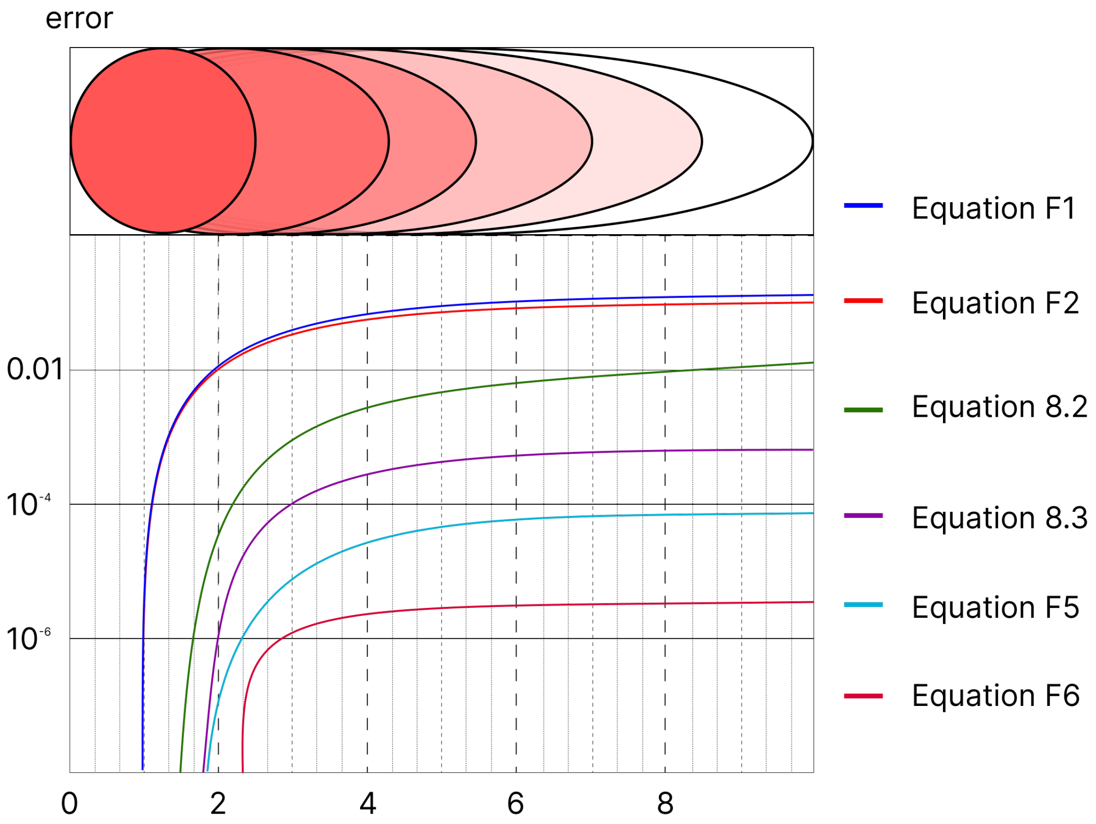


Fig. 7.5 The y-axis represents the error rate, while the x-axis refers to the major axis of the ellipse. As we move from left to right, the error rate rises, indicating that a longer major-axis corresponds to a higher error rate. Image from [138]

Once our proposed software detect the contour of the human body, we require two orthogonal views to calculate 10 landmarks, labeled such that the person is facing the camera in one image and is seen from a side in another (aka front and side views).

The anthropometric dimensions are obtained from landmarks extracted in the front and side views. As shown in Figure 7.6, the dimensions obtained include two lengths (F-L1:F-L2), eight perimeters (F-B1:F-B4, S-D1:S-D4). The definition of the dimensions for the upper human body according to the ISO 8559-1 [13], 8559-2 [14], 8559-3 [15] can be found in Table 7.1. While state-of-the-art methodologies may include a broader set of landmarks for entire human body measurements, our research intentionally focuses its attention to the selected landmarks which prove sufficient.

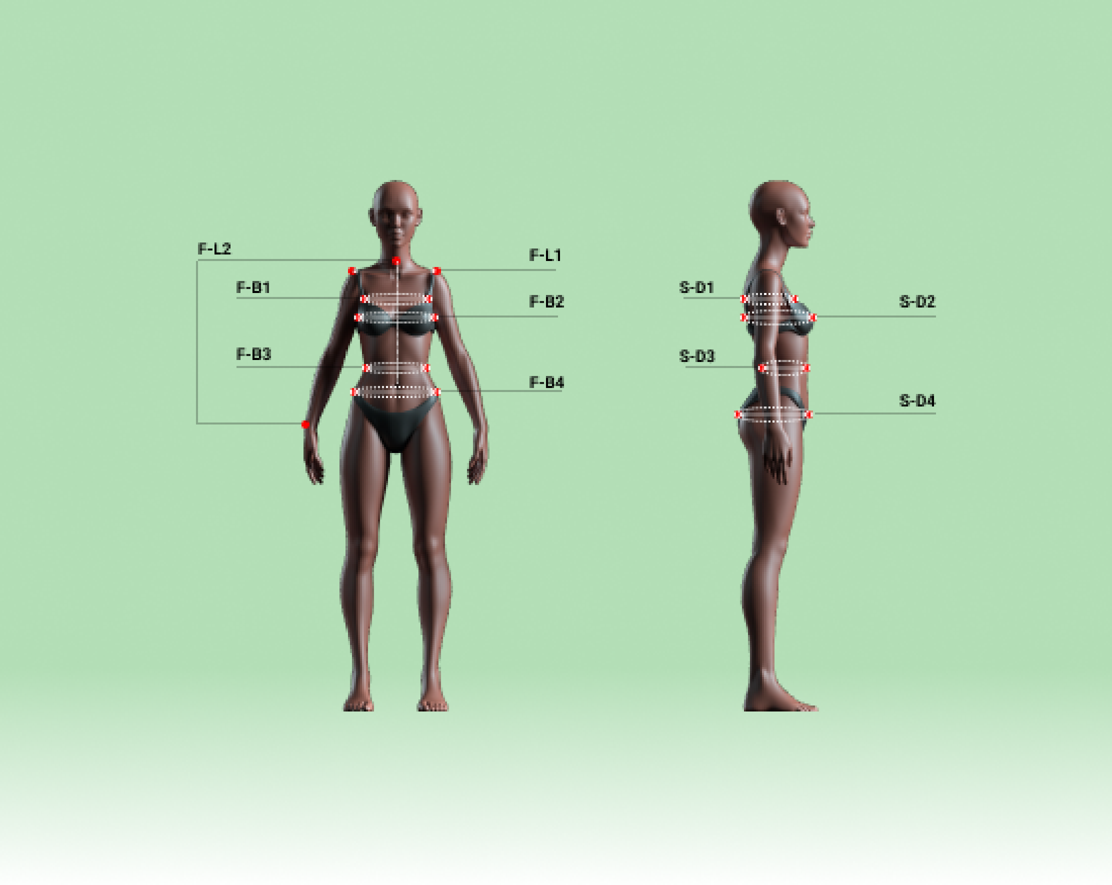


Fig. 7.6 The anthropometric dimensions (a) front image (b) side image - 3d model image from: https://www.turbosquid.com/3d-models/female-basemesh-realistic-body-3d-142544

Shoulder and sleeve lengths are one of two types of measurements that can be determined with the help of our proposed method. In the context of these measurements, it is not necessary to employ two images; instead, the length can be determined by calculation using a single image. The second group consists of various body measurements such as the bust, waist, and hips circumferences.

Table 7.1 The definition of dimensions

Dimensions	Definition	Dimensions	Definition
F-L1	Shoulder Lengths	F-B4	Hip Breadth
F-L2	Sleeve Lengths	S-D1	Chest Depth
F-B1	Chest Breadth	S-D2	Bust Depth
F-B2	Bust Breadth	S-D3	Waist Depth
F-B3	Waist Breadth	S-D4	Hip Depth

The length between each point of the front view corresponds to the major axis and the length between each point of the side view corresponds to the minor axis (as shown in Figure 7.7). Hence, by computing such data from the images, we can calculate both the length and perimeter of (ellipses approximating) the human torso.

Once the points of interest are extracted, the Euclidean distance between pairs of points (x_1, y_1) and (x_2, y_2) is given by the usual equation:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{7.1}$$

This gives pixel distances between the points. These need to be converted to real-world measurements (like centimeters or inches), using a reference or scale. For example,

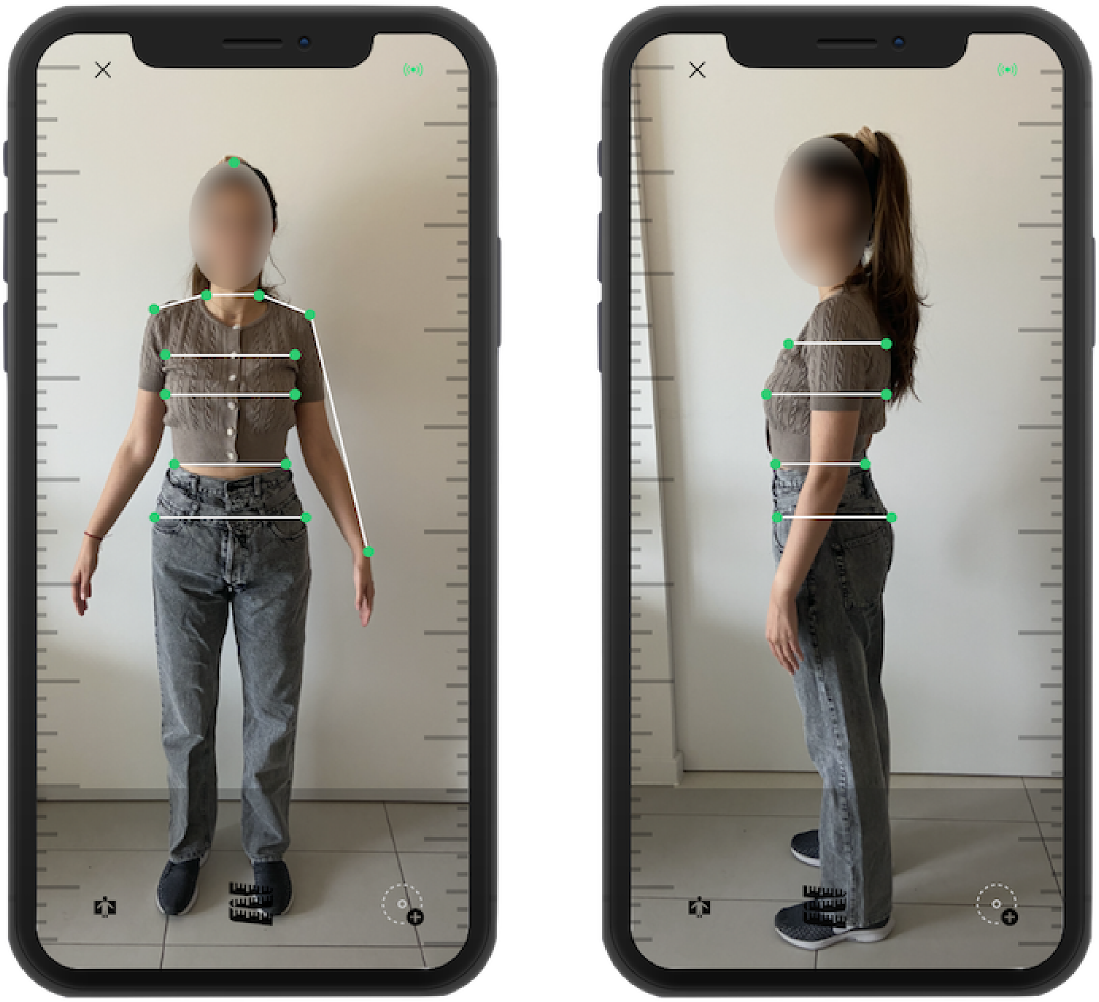


Fig. 7.7 demonstrates how these markers were selected around each body landmark using our software.

if we know that a certain object in the image is exactly 10 cm in the real world and it measures 100 pixels in the image, then the conversion factor is 0.1 cm per pixel. Using such a conversion factor, you can then translate all pixel measurements to real-world units. To automate this step, the conversion factor is obtained through a camera calibration process previously mentioned.

To calculate the pixel unit to centimetre conversion, first, we must know the distance between the camera and the object (review section 6.3). This can be accomplished by measuring the size of chessboard squares or any known object size in the image of camera calibration and dividing by the distance between the camera and the chessboard. This value represents the pixel size in centimetres. The equation for converting pixels units to centimeters is then simply:

$$\textit{Size of object in cm = Number of pixels in objectx} \frac{Size\ of\ chessboard\ square\ in\ pixels}{Distance\ between\ camera\ and\ chessboard\ in\ cm}$$

We compared six popular and relatively simple (i.e. not involving an integral or infinite series) ellipse equations at the beginning of our study, which we then reduced to the following two due to the better accuracy level they permit (details in section 7.2), the second being one of the famous equations discovered by Ramanujan:

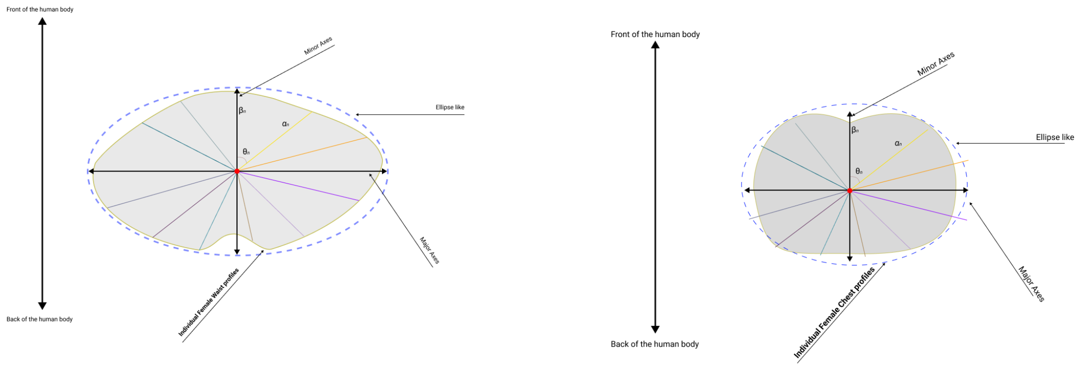
$$P \approx 2 \times \Pi \times \sqrt{\frac{(dist_a^2)+(dist_b^2)}{2}} \tag{7.2}$$

$$P \approx \pi \left[3(dist_a + dist_b) - \sqrt{(3dist_a + dist_b)(dist_a + 3dist_b)}\right] \tag{7.3}$$

Where dista = Semi-Major axis, distb = Semi-Minor axis.

Using the distances calculated from the points on the human body circumferences from both front/side image, we can estimate the major and minor axes of the ellipse, and then use the formula above to compute the approximate perimeter.

Based on human images from multiple angles, a multi-ellipse model has been developed in the polar coordinate system, which is a general model for calculating the dimension of an ellipse. Due to the special structure of the human torso, the curvature of different segments of the same part is different. As shown in Figure 7.8, the calculated dimensions are the cumulative sum of the elliptical sections with different curvatures.



(a) Individual female waist profile

(b) Individual female Chest profile

Fig. 7.8

Using the major and minor axes calculated from the front and side images, the algorithm will automatically select Equation 7.2 for those sections of the body where the major and minor axes are no more than three times longer, and Equation 7.3 for those sections where the major and minor axes are more than three times longer, leading to better accuracy.

7.2 Results and Discussion

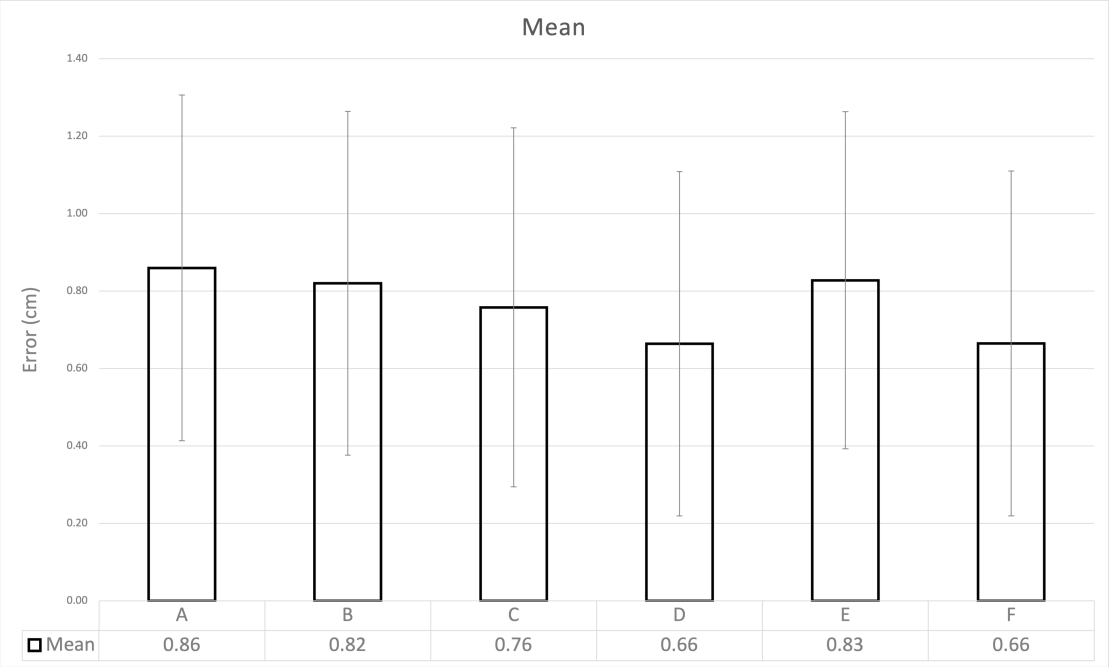
This result and discussion contain two parts: In the first part of this section, we analyse the similarities and differences between six different elliptical equations; in the second part, we dive deep into three of these equations by evaluating how they apply to individuals with different body shapes.

In this case study, 78 female torsos and 234 images were used to investigate various elliptical equations. All the images in this area have plain background, without any lighting issues (such as shadow), with a perfect distance (0.5cm-2cm) between the participants and the camera to identify the best possible equations. As noted previously traditionally, the following female body types (i.e. chest, bust, waist and hips circumferences) have been chosen to assess our data in comparison to the tape measurements for ground truth comparison. In our data, we are analysing mean value, minimum and maximum accuracy and standard deviations. We are also looking at a plot that displays the average error correlations values of each separate body segment in detail for all the chosen 78 female body participants for this case study. In addition, we are examining into female body torso in more details to understand how each varied body shape influences on the accuracy of the chosen equations. Lastly, in the summary of this section, we are discussing about the optimal equation or equations for our proposed software.

7.2.1 Comparison of six elliptic equations

Considering the findings from the previous sections, we were able to calculate the human body horizontal perimeters by including each equation into the software that we proposed. Also, we were able to discover the variance between each equation and the tape measurement method so that we could make a comparison to the ground truth (i.e. tape measurements).

In summary, Formula A (Equation F.1) has a mean difference and standard deviation of (m=0.86cm, sd=0.446), Formula B (Equation F.2), (m=0.82cm, sd=0.443), Formula C (Equation 7.2), (m=0.76cm, sd=0.464), Formula D (Equation 7.3), (m=0.66, sd=0.445), Formula E (Equation F.5) (m=0.83, sd=0.435) and finally, Formula F (Equation F.6) (m=0.66, sd=0.445).



	A	B	C	D	E	F
□ Mean	0.86	0.82	0.76	0.66	0.83	0.66

Fig. 7.9 Comparison of Ellipse equations in our propose software

Hence, Equation 7.3 and Equation F.6 offered the highest level of accuracy compared to the other equations. However, a closer examination of the data reveals that, in some instances, Equation 7.2 provides greater accuracy than the other formulae. As illustrated in Figure 7.10 and discover that, Equation 7.2 is more accurate than all other equations for around 35% of participants, despite the fact that, it provides less overall accuracy in compare to Equation 7.3 and Equation F.6.

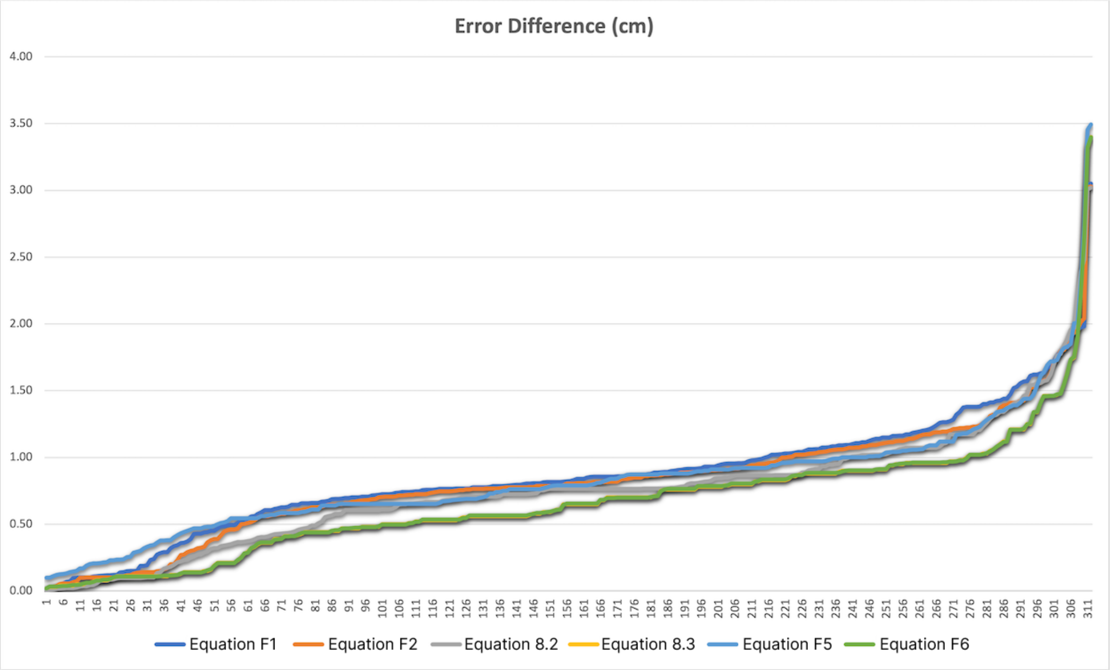


Fig. 7.10 Error Difference (cm) for each and every Ellipse formula. This diagram illustrates the error variance for 78 subjects' chest, bust, waist, and hip circumferences.

In the next section, we examine the three best elliptic equations based on the data obtained we will evaluate and the results for each equation based on various human body shapes.

7.2.2 Comparison of the three best elliptic equations for human body shapes

The quantitative data for the 78 female participants were analysed by using a descriptive statistic, an ANOVA single factor, and Post-hoc (t-test with two samples assuming equal variances).

Data description has been used throughout all phases of our data collection. The dependability of measured data is compared between Equation 7.2, Equation 7.3 and Equation F.6 using an ANOVA test. Lastly, we utilised a t-test: Two-sample comparison assuming equal variances to compare individual equations to ensure that we were able to improve the accuracy of the state-of-the-art. At each stage, the quantitative data were analysed and compared in terms of the accuracy and reliability of each method and technique. The following tables report on the mean, median, minimum and maximum differences as well as the standard deviation according to different body shape.

We have divided our Anova analysis according to different categories of measurements we have, which include Chest circumference, bust circumference, waist circumference, and hip circumference.

These results indicate that, as a function of variations in body shape (horizontal slices), one formula is more accurate than others in minimizing the difference between direct measurement and software solutions. For body shapes with semi-major axes that are not more than three times longer than their semi-minor axes (— in other words, ellipses that are not overly elongated), Equation 7.2 provides more accurate results, whereas Equation 7.3 or Equation F.6 provide more accurate results for body shapes with semi-major axes that are more than three times longer than their semi-minor axes (squashed ellipses). Due to the complexity of Equation F.6, we have depended in practice just on the first two formulas. Equation 7.3 and Equation F.6 provide remarkably similar accuracy, thus we have relied on only the Equation 7.2 and Equation 7.3.

By knowing the human body shape and ellipse equation better, we were able to reduce the mean difference. Using Equation 7.2 for those whose major and minor axes are less than three times longer and Equation 7.3 for those whose major and minor axes are more than three times longer allowed us to reduce the mean difference to less than 0.5cm for this case study. As previously stated, Equation 7.2 has a mean and standard deviation of (m=0.76cm, sd=0.47) while Equation 7.3 has a mean and standard deviation of (m=0.67cm, sd=0.45). Figure 7.11 illustrates the overall outcome.

Additionally, in this research study, we have undertaken extra analysis (post-hoc) to support our findings. it was discovered that Equation 7.2 significantly differs from Equation 7.3 and Equation F.6. This finding aligns with the ANOVA results.

Body Type	Chest	Bust	Waist	Hips
Mean (cm)	0.727	0.701	0.850	0.751
Median (cm)	0.765	0.705	0.769	0.764
Max Differences (cm)	2.386	2.480	3.009	3.022
Min Differences (cm)	0.006	0.002	0.017	0.015
Standard Deviation	0.462	0.399	0.474	0.506

(a) Equation 7.2

Body Type	Chest	Bust	Waist	Hips
Mean (cm)	0.621	0.708	0.613	0.714
Median (cm)	0.619	0.697	0.574	0.654
Max Differences (cm)	1.733	2.208	1.754	3.398
Min Differences (cm)	0.038	0.018	0.039	0.032
Standard Deviation	0.355	0.432	0.363	0.588

(b) Equation 7.3

Body Type	Chest	Bust	Waist	Hips
Mean (cm)	0.618	0.707	0.612	0.715
Median (cm)	0.624	0.695	0.569	0.664
Max Differences (cm)	1.732	2.202	1.754	3.398
Min Differences (cm)	0.038	0.022	0.039	0.029
Standard Deviation	0.357	0.431	0.362	0.590

(c) Equation F.6

Table 7.2 Summary of the differences between our software and tape-measured data for the 78 participants included in this study. All participants captured in the same room conditions; lighting, backgrounds, and also, they have been told to wear tight clothes to discover the best possible equations. The best, worst, mean (ABS) and standard deviation of the accuracy of all the three different ellipse equations on female body torso for comparison with the tape measurements.

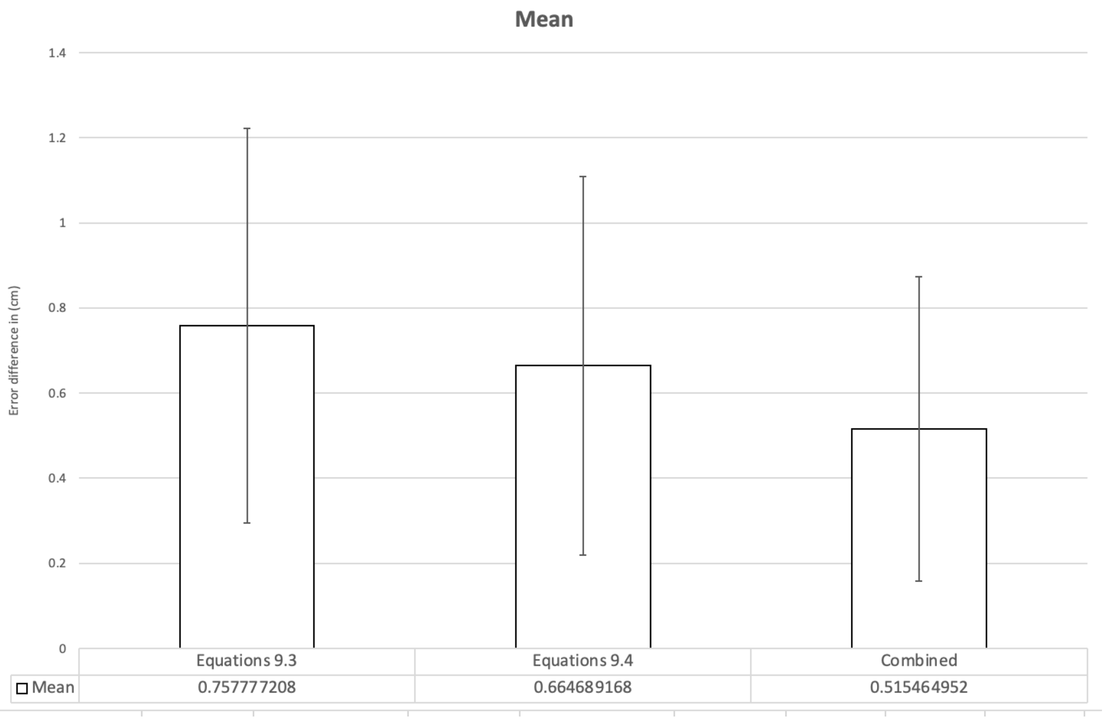
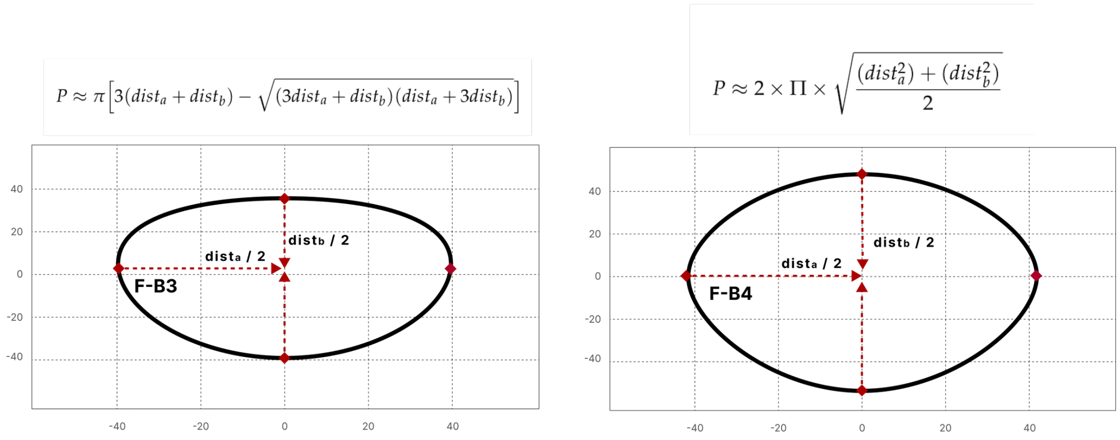


Fig. 7.11 displays the mean difference between Equation 7.2, Equation 7.3, as well as our proposed programme based on the application of both equations.



$$P \approx \pi\left[3(dist_a + dist_b) - \sqrt{(3dist_a + dist_b)(dist_a + 3dist_b)}\right]$$

$$P \approx 2 \times \Pi \times \sqrt{\frac{(dist_a^2) + (dist_b^2)}{2}}$$

Fig. 7.12 Image right: (F-B3 = waist circumference, dist$_a$ = Major axes, dist$_b$ = Minor axes). Image left: (F-B4 = hips circumference, dist$_a$ = Major axes, dist$_b$ = Minor axes)

However, there is no statistically significant difference between Equation 7.3 and Equation F.6, implying that their means are statistically similar. Hence, the selection of various multi-ellipse equations was made in accordance with different human body shapes, as illustrated in the Figure 7.12. The comprehensive details of our ANOVA test and post hoc test can be found in Appendix F.1.2.1.

Chapter 8

Results and Discussion

This research focused on developing a software solution for accurately measuring various body dimensions without the need for tight or specific clothing, lighting conditions, background variations, distance between the camera and participants, and different body postures and shapes. Overall, the software demonstrated promising results, showcasing its potential for accurate body measurements.

For length measurements such as shoulder and sleeve length, the software achieved excellent accuracy, with mean errors of only 0.15 cm and 0.14 cm, respectively. These results highlight the reliability and precision of the software in capturing length-related body parameters.

In terms of circumference measurements, including chest, bust, waist, and hips, the software performed exceptionally well compared to existing solutions in the market. The mean errors for these measurements ranged from 0.78 cm to 0.96 cm, indicating a high level of accuracy. However, the waist area exhibited the highest mean error, which can be attributed to challenges posed by loose-fitting clothes and occasional inclusion of hands in the measurement.

In the comparative analysis conducted, the software exhibited superior performance compared to other existing methods with regard to accuracy, computational efficiency, and user-friendliness. This highlights the potential of this approach to surpass current state-of-the-art standards in the field.

In conclusion, the system designed for estimating upper human body measurements through the application of computer vision and machine learning techniques demonstrates promising results. The study involving 78 participants reveals an average difference of ±1 cm compared to tape-measured data, with minimal deviations observed across various body circumferences. The system's robustness against cluttered backgrounds and varying lighting conditions, in addition to its simplicity and accuracy, positions it as a valuable tool for the apparel industry. Potential enhancements, such as the provision of clear user instructions and tutorials, hold the capacity to reduce the mean difference in software measurements to approximately ±0.5 cm. Overall, the system presents a convenient and cost-effective solution for precisely capturing human body dimensions, thereby contributing to increased efficiency in the apparel industry, while also reducing waste and returns.

In this chapter, the evaluation of the results and discussion will begin with an individual examination of the image corrections and camera calibrations. Subsequently, the overall performance and results of our proposed software, focusing on addressing the challenges mentioned in the section 2.11. These challenges include issues related to different lighting conditions, background variations, distance between the human body and the camera, and different body postures and shapes.

In the previous chapters, we presented data related to specific measurements such as chest, bust, waist, and hips circumferences, based on our hypothesis and validation techniques. However, in this chapter, we will examine the overall performance of the software, including length measurements such as shoulder and sleeve length, which were also collected from each participant during data collection.

It is important to note that we faced several challenges throughout this research study. The complexity of the human body, especially non-rigid parts, variability in human physique, complexity of the skeletal structure, the impact of breathing on body shape, lighting variability causing shadows, loss of 3D data in 2D image projections, and difficulty capturing body parts covered by loose clothing were among the significant challenges we encountered.

The complexity of the human body with multiple degrees of freedom poses a challenge as the body moves and undergoes changes during breathing. This can affect the fit of garments and distort their shape.

Depth inference from image sequences is another challenging aspect as obtaining accurate depth information from a single camera is difficult, especially considering the multiple degrees of freedom in human body movements.

Changes in lighting conditions and camera viewpoints introduce variations in image appearance, including the presence of shadows. Shadows cast on areas covered by hair further complicate the extraction of relevant information from images.

Additionally, the requirement for participants to wear specific clothes or pose in front of the camera creates discomfort for consumers, making it necessary to develop a solution that does not rely on such constraints.

Based on our hypothesis, it is feasible to develop a measurement system that can extract anthropometric measurements from multiple depth images, thereby capturing the surface of the human body. This system aims to improve accuracy while minimizing computational expenses, making it suitable for implementation on new-generation smartphone devices.

Considering our primary objectives, the software needs to be capable of capturing the human body accurately regardless of the background, whether it is plain or cluttered. It should be able to measure various human body dimensions at any distance between the camera and the participants. Additionally, the software should be able to accommodate different body postures and shapes during measurement. Furthermore, it should have the capability to predict personal details such as clothing and provide accurate measurements based on that information. Of course, it is important to note that our system is not an X-ray technology, and it is designed to work specifically with certain types of clothing, such as shirts, t-shirts, and dresses. It is not intended for measuring garments like jackets and similar items. However, even with these limitations, the software still offers significant advantages over existing solutions that require participants to wear tight clothes or even underwear in order to obtain accurate measurements.

The limitations of the software have implications for its practicality and usability in real-world scenarios. One limitation is the inability to accurately measure garments like jackets. This can affect the applicability of the software in scenarios where users are interested in obtaining measurements for garments that cover the body extensively. While it is generally understood that the accuracy level may be impacted by such clothing due to their volume and potential obscuring of body contours, the software offers more flexibility by allowing users to align the measurement points around their body if needed. This feature provides an opportunity to capture more accurate results even with clothing variations.

In comparison to other applications in the market, where users may not have the option to adjust the measurement points, the software's approach enhances user experience and addresses the limitations associated with garment measurements. Users can align the points around their body to account for clothing variations and improve the accuracy of measurements. It is important to note that these limitations are inherent in most body measurement applications, and the software offers a more user-friendly approach by empowering users to participate in the measurement process.

Considering these implications, the software remains a practical and usable solution for capturing body measurements in real-world scenarios. By providing users with control over the alignment of measurement points and accommodating different clothing types, the software offers enhanced versatility and convenience. While certain limitations persist across all applications, the software's user-adjustable points feature provides an added advantage in achieving accurate measurements, thereby improving the overall usability and user satisfaction

To assess the effectiveness of the software, we gathered a dataset consisting of 78 female bodies, which encompassed a total of 1560 images captured from front and side views including a total of 312 data. This dataset enabled us to thoroughly evaluate and analyse the accuracy of the software across various scenarios. The collected images illustrated static human body shapes, obtained from scans taken in a similar standing stance, although not precisely identical. The dataset comprised diverse poses and variations in limb and whole-body shapes. The images also featured different backgrounds, ranging from plain to cluttered, and were captured at varying distances between the participants and the camera. Furthermore, we accounted for different camera orientations, including portrait and landscape modes, as well as different lighting conditions.

Participants who were over 18 years old. To ensure a diverse representation, we included participants with different body types based on the somatotype system. The somatotype system categorises individuals into distinct body types, namely ectomorph, mesomorph, and endomorph, based on variations in body proportions, fat distribution, and overall body composition. By including participants with different body types, we aimed to capture a range of body shapes and sizes, enhancing the representativeness of our dataset. This approach allows us to evaluate the performance of the software across different body types, considering potential variations in body proportions and fat distribution that may affect measurement accuracy. The inclusion of participants with diverse body types ensures that the software is designed to provide to a broader user base and is applicable to a wide range of body shapes and sizes.

Therefore, we present a comprehensive view of our datasets, including results from Chapter 6, Chapter 7 collectively. We assess the performance of the software across various stages, encompassing factors such as high/low resolution, varying lighting conditions, correct/incorrect color balance, plain/cluttered background, a range of 0 to 3 meters distance between the camera and participant, A-poses/T-poses and relax posture, different body shapes, as well as tight and loose clothing.

In the sections that follow, descriptive statistics derived from our results are presented first. These statistics provide a comprehensive overview of the data and insight into the software's efficiency. To determine the efficiency of our strategy, we evaluate multiple metrics, including accuracy, precision, and recall. By analysing these statistics, we gain a better understanding of the capabilities and limitations of the software.

In addition, we conduct a thorough comparison between the software and existing methods in the field. This comparative analysis enables us to determine if the software outperforms cutting-edge alternatives. To determine the superiority of our approach, we consider multiple factors, such as measurement accuracy, computational efficiency, and user-friendliness.

Through this comprehensive evaluation, we aim to determine if the software has achieved a higher level of accuracy and reliability compared to other methods available in the market. By leveraging cutting-edge techniques and addressing limitations observed in previous approaches, we strive to provide an advanced software solution that surpasses the state-of-the-art standards in the field.

8.1 Image Corrections and Camera Calibrations Evaluation

As discussed in Chapter 6, we have employed classical computer vision image segmentation, Skin detection and Camera calibration techniques to enhance the image quality, isolate the human body from its surrounding environment and estimate the distance in order to convert pixel units to centimeters. It is important to note that ROI assisted us in enhancing our final algorithms. Our findings and subsequent discussions are divided into two parts. Initially, we will showcase the results for image corrections and skin detections. In the second segment, we will display outcomes related to camera calibration. This calibration was important in measuring the distance between the camera's viewpoint and the human body, facilitating the conversion from pixel measurements to centimeters.

8.1.1 Image Corrections and Skin Detection

According to our hypothesis 1 (together with objectives 5 and 7), it is crucial to ensure that the software can offer accurate data in any environment, including images with poor diffusion, a noisy background, and shadows. Also, it is possible to correctly measure the human body on many types of apparel, like T-shirts, shirts, skirts, among others. Furthermore, garments that are either too tight or too loose. Therefore, in this case study, 78 female subjects and 312 images were utilised to examine how our program performs in variety of backgrounds situations. Images contain plain/cluttered background, high/low resolution, good/poor lighting and correct/incorrect color balance for validation of technique. In the following paragraphs, we will examine the findings that the software has generated.

At all phases of our data collecting, data description has been utilised. Using a t-test: two-sample assuming unequal variance to compare individual situations in order to enhance the accuracy of the state-of-the-art, the reliability of measurement data is evaluated between plain and cluttered backgrounds. At each stage, the quantitative data for each environment and type of clothing were analysed and compared in terms of accuracy and reliability.

Table 8.1, Table 8.2 display the results for all of our subjects who were photographed in various environments, including those with good/poor lighting and shadows, correct/incorrect color balance, and tight/loose clothing.

Body Type	Chest	Bust	Waist	Hips
Mean (cm)	0.48	0.5	0.56	0.52
Median (cm)	0.48	0.56	0.56	0.56
Best (cm)	0.01	0.01	0.02	0.02
Worst (cm)	1.73	1.25	1.75	3.02
Standard Deviation	0.35	0.30	0.35	0.42

Table 8.1 The best, worst, mean (ABS), median and standard deviation of the accuracy of female body torso for comparison with the tape measurements on a plain background

Body Type	Chest	Bust	Waist	Hips
Mean (cm)	0.57	0.56	0.65	0.59
Median (cm)	0.59	0.54	0.66	0.58
Best (cm)	0.03	0.02	0.06	0.03
Worst (cm)	1.74	1.26	1.92	3.03
Standard Deviation	0.38	0.30	0.38	0.43

Table 8.2 The best, worst, mean (ABS), median and standard deviation of the accuracy of female body torso for comparison with the tape measurements on a cluttered background

As can be seen in the above tables, subjects photographed against a plain background with better image quality and lighting circumstances provide greater accuracy; yet, the difference between these findings and those obtained with a cluttered background are not large. The average difference for a plain background is 0.52 cm, whereas the average difference for a cluttered background is 0.59 cm. Using t-test, we conducted more research on our results to ensure that the software can process input images with any type of background.

Due to the fact that the human body is composed of various body portions, it is essential that we compare and study each body section against a plain and cluttered background. This is owing to the fact that some parts of the body are simpler to capture than others. For each participant, we will analyse chest, bust, waist, and hips against a plain and cluttered background. For example, Participant 1, who was photographed in two different environments, will be analysed as Chest Plain background versus Chest Cluttered background, etc. In addition to the presented data, further analysis was conducted using t-tests and descriptive statistics for each section of the human body torso.

T-tests were conducted to examine whether background conditions (plain vs. cluttered) had a significant impact on chest, bust, waist, and hip circumference measurements. The null hypothesis for all tests was that there is no difference between measurements taken in different backgrounds.

For chest circumferences, the mean measurements were 0.48 (plain) and 0.57 (cluttered), with n = 78 in both cases. T-test results with 153 degrees of freedom yielded a p-value of 0.07 for a one-tailed test and 0.13 for a two-tailed test. Fail to reject the null hypothesis.

Bust circumferences showed mean measurements of 0.50 (plain) and 0.56 (cluttered), with n = 78 in both cases. T-test results with 154 degrees of freedom gave a p-value of 0.12 for a one-tailed test and 0.24 for a two-tailed test. Fail to reject the null hypothesis.

Waist circumferences displayed mean measurements of 0.56 (plain) and 0.65 (cluttered), with n = 78 in both cases. T-test results with 153 degrees of freedom yielded a p-value of 0.05 for a one-tailed test and 0.11 for a two-tailed test. Fail to reject the null hypothesis.

Hip circumferences indicated mean measurements of 0.52 (plain) and 0.59 (cluttered), with n = 78 in both cases. T-test results with 154 degrees of freedom yielded a p-value of 0.16 for a one-tailed test and 0.32 for a two-tailed test. Fail to reject the null hypothesis.

Overall, measurements taken in plain and cluttered backgrounds did not show statistically significant differences. Error correlations between background conditions were strong (ranging from 0.90 to 0.96), suggesting consistent measurement trends. You can find more details about the t-test and descriptive statistics in the Appendix D.2.1.

The results obtained demonstrate that the software provides consistent and accurate measurements, even in challenging environments such as cluttered backgrounds, poor lighting, and variations in clothing tightness. This sets the software apart from existing technologies in the market, which often require specific conditions like a white background and optimal lighting for accurate measurements.

The performance of the software was evaluated by comparing the measurements taken by the software with tape measurements for various body circumferences, including chest, bust, waist, and hips. The statistical analysis conducted using t-tests and error correlation provided insights into the accuracy and reliability of the software in different backgrounds. Despite minor variations in accuracy, the results showed that the software consistently provided accurate measurements in both plain and cluttered backgrounds.

In terms of accuracy, the majority of measurements obtained from the software had an error difference of less than 1 cm, indicating a high level of precision. This demonstrates the effectiveness of our image processing and correction techniques, which contribute to the accurate detection and isolation of the human body from the background and clothing.

Comparing the software to existing technologies in the market, we can emphasize the superiority of our approach. Many current technologies heavily rely on controlled environments, requiring users to stand against a white background with good lighting conditions to ensure accurate measurements. In contrast, the software surpasses these limitations by providing accurate results irrespective of the background environment. This versatility and robustness make the software more practical and user-friendly, as users are not constrained by specific setup requirements.

The success of the software can be attributed to the careful implementation of classical computer vision techniques and skin detection algorithms. Initially, using just MobileNet SSD-v2, we achieved a mean Average Precision (mAP) of 70% at an Intersection over Union (IoU) threshold of 0.5 for isolating the human body (please review the information presented in Table 5.4). We then further enhanced our approach by leveraging advanced methods like image corrections, which significantly improved our body detection performance. Consequently, with these corrections, we achieved a remarkable mAP of 98% at an IoU threshold of 0.5 for isolating the human body from its surroundings, according to our datasets.

Moreover, the software demonstrates consistency in performance across different body circumferences. Whether measuring chest, bust, waist, or hips, the software provided accurate results, with minimal variations observed between plain and cluttered backgrounds. This consistency further highlights the reliability of the software in capturing accurate measurements across various environmental conditions.

Lastly, the results of the test showed that the software worked well with both plain and cluttered backgrounds. The majority of the measurements had an error difference of less than 1 cm, and the software's accuracy differed little between plain and cluttered backgrounds. As a result, the software can be used to accurately measure human torso circumferences in both environments.

To ensure the widespread usability and convenience of the software, we acknowledge the need to address certain limitations, such as the software's accuracy may be influenced by variations in clothing types and styles. While we have tested the software on different types of apparel, including T-shirts, shirts, and skirts, there may still be limitations when it comes to measuring body circumferences on certain clothing materials or designs. For example, loose or flowing garments may introduce additional challenges in accurately capturing body contours. Therefore, users should consider the clothing characteristics and their potential impact on measurement accuracy.

8.1.2 Camera Calibration and Distance Estimation

OpenCV's camera calibration algorithm was used to first measure the distance between the camera point and the human body and then use this value to convert every pixel unit to centimetres. We collected a set of data from 78 female participants (including 312 images) for this case study. From the camera's viewpoint, images were captured at various distances from the human body, ranging from 0 to more than 3 metres. This is to ensure that the software can accurately convert pixel units to centimetres at practical useful distances from the human body.

The datasets, created for this study, were divided into two different groups based on participants' distance to the camera. The first group includes individuals standing within a range of 0.5 to 3 metres from the camera's viewpoint, while the second group comprises those standing within a range of 0 to 0.5 metres or more than 3 metres from the camera's viewpoint. To establish the reliability of the method employed in this study, all the participants were photographed under consistent environmental conditions, which entailed a uniform background, evenly distributed lighting, and the absence of shadow effects.

A t-test with two-sample assuming unequal variance to compare individual situations in order to enhance the accuracy and reliability of measurement data, evaluated between each group of data. The results for all participants who were photographed from different distances relative to the camera's perspective are displayed in Table 8.3 (range: 0.5m – 3m) and Table 8.4 (ranges 0 – 0.5m and more than 3m).

Body Type	**Chest**	**Bust**	**Waist**	**Hips**
Mean (cm)	0.48	0.5	0.56	0.52
Median (cm)	0.48	0.56	0.56	0.56
Best (cm)	0.01	0.01	0.02	0.02
Worst (cm)	1.73	1.25	1.75	3.02
Standard Deviation	0.35	0.30	0.35	0.42

Table 8.3 The best, worst, mean (ABS), median and standard deviation of the accuracy of female body torso for comparison with the tape measurements within the range of (0.5m – 3m).

Body Type	**Chest**	**Bust**	**Waist**	**Hips**
Mean (cm)	1.02	1.02	1.29	0.99
Median (cm)	0.99	1.01	1.10	0.98
Best (cm)	0.02	0.02	0.13	0.02
Worst (cm)	3.13	2.64	3.16	3.81
Standard Deviation	0.58	0.53	0.62	0.60

Table 8.4 The best, worst, mean (ABS), median and standard deviation of the accuracy of female body torso for comparison with the tape measurements within the range of (0 – 0.5m and more than 3m).

In Table 8.3, the mean accuracy values for the body parts range from 0.48cm to 0.56cm, with the waist having the highest mean accuracy. The median accuracy values range from 0.48cm to 0.56cm, with the bust and waist having the highest median accuracy. The best accuracy values range from 0.01cm to 0.02cm, while the worst accuracy values range from 1.73cm to 3.02cm. The standard deviation values range from 0.30cm to 0.42cm, indicating a moderate level of variability in the measurements.

In Table 8.4, the mean accuracy values for the body parts range from 1.02cm to 1.29cm, with the waist having the highest mean accuracy. The median accuracy values range from 0.98cm to 1.10cm, with the waist having the highest median accuracy. The best accuracy values range from 0.02cm to 0.13cm, while the worst accuracy values range from 2.64cm to 3.81cm. The standard deviation values range from 0.53cm to 0.62cm, indicating a slightly higher level of variability compared to Table 1.

Overall, the comparison of the two tables suggests that the software performs better for measurements within the range of 0.5m to 2m compared to measurements within the range of less than 0.5m or more than 3m. The mean accuracy, median accuracy, and standard deviation values are all better for the former range. However, it is important to note that the accuracy results for both ranges are still relatively consistent across the 78 participants. Therefore, the proposed software may still be useful for body measurements within the range of less than 0.5m or between 0.5m to 2m, but further improvements may be necessary to enhance the accuracy of the measurements.

In addition to the presented data, further analysis was conducted using t-tests and descriptive statistics for each section of the human body torso. These additional analyses provide more detailed insights into the accuracy measurements.

A t-test was conducted to assess whether significant differences exist between chest circumferences measured at distances between 0.5 to 3 meters and those measured at distances less than 0.5 meters or more than 3 meters. The data suggests that measurements taken at a distance less than 0.5 meters or more than 3 meters exhibit higher stress levels compared to those taken at a distance between 0.5 to 3 meters (t(127) = -7.14, p < 0.001, one-tail). Similar results were observed for bust circumferences, where measurements at distances less than 0.5 meters or more than 3 meters showed higher stress levels compared to those taken between 0.5 to 3 meters (t(154) = -7.42, p < 0.001, one-tail). For waist circumferences, the measurements taken at a distance less than 0.5 meters or more than 3 meters displayed higher stress levels compared to those at distances between 0.5 to 3 meters (t(122) = -9.13, p < 0.001, one-tail). Lastly, the analysis of hips circumferences revealed that measurements taken at distances less than 0.5 meters or more than 3 meters showed higher stress levels compared to those taken at distances between 0.5 to 3 meters (t(138) = -5.56, p < 0.001, one-tail).

Overall, the findings consistently indicate that circumferential measurements taken at closer or farther distances from the camera tend to exhibit higher stress levels, although the software's accuracy remains satisfactory in both scenarios. Additional information concerning the t-test and descriptive statistics can be referenced in the Appendix E.3.1.

The results presented in the previous paragraphs demonstrate the effectiveness of our method in measuring body dimensions from images captured by a camera. However, the accuracy of the measurements varied depending on the distance between the camera and the human body. In this discussion, we will delve into the implications of these findings and provide insights into the potential applications and limitations of the software.

Firstly, the results revealed that the software performed better for measurements taken within the range of 0.5 meters to 3 meters compared to measurements taken at distances less than 0.5 meters or more than 3 meters. The mean accuracy, median accuracy, and standard deviation values consistently favored the former range. This suggests that the software is more reliable when the subject is positioned at a moderate distance from the camera. The accuracy of the measurements decreases as the distance increases, likely due to the decrease in pixel resolution and the influence of other factors such as lighting conditions.

Nevertheless, it is important to note that the overall accuracy of the software remained acceptable even for measurements outside the optimal range. The mean accuracy values for the body dimensions in both ranges ranged from 0.48 cm to 1.29 cm, indicating a reasonably accurate estimation of body measurements. While the standard deviation values ranged from 0.30 cm to 0.62 cm, indicating a moderate level of variability, the measurements were still within an acceptable range for most applications.

One significant implication of our findings is the potential utility of the software for body measurements in various domains. The ability to accurately measure body dimensions from images can have a wide range of applications, including custom clothing design, virtual try-on systems, and fitness tracking. For instance, fashion retailers can utilise the software to provide personalised sizing recommendations to customers, reducing the need for physical try-ons and improving the overall shopping experience. Fitness tracking applications can leverage the software to accurately monitor body changes over time, enabling users to track their progress and make informed decisions about their health and fitness goals.

Furthermore, the software has the potential to address a common challenge faced by existing body measurement solutions: the need for users to know their exact height. Our research revealed that more than 70% of participants were unsure about their precise height. This lack of accurate height information can significantly affect the accuracy of body measurements. However, by using computer vision techniques and the OpenCV camera calibration algorithm, the software eliminates the need for users to input their height manually. Instead, it calculates the distance between the camera and the human body, enabling accurate measurements without relying on (potentially inaccurate) user-provided data.

Despite these promising outcomes, there are certain limitations and areas for improvement that should be considered. Our current approach requires users to have a chessboard for calibration purposes in order to capture accurate results. This constraint may pose practical challenges, as not all users may have access to a chessboard or be familiar with its usage. To address this issue, we have devised an alternative method that leverages the user's accurate height to achieve comparable accuracy in body measurements. We can then estimate the camera's intrinsic and extrinsic parameters using only the user's height information. This enables us to align the measurements obtained from the camera with real-world distances, almost replicating the precision achieved through traditional circumference measurements.

By offering this option, we ensure that a broader range of users can benefit from our advanced AI-powered technology without the need for specialized calibration patterns.

However, it is worth noting that this constraint (of calibration via some external input, be it a pattern or a precise measure like height) can be addressed and improved in future iterations of the software. One possible solution is to replace the need for a chessboard with a known common object size in the image, such as a credit or business card (or some object of standard size). By incorporating an object of known dimensions, the software can calibrate the camera and accurately estimate distances without relying on the presence of a specific unusual calibration pattern. This approach would enhance the usability and accessibility of the software, as credit cards or similar objects are more commonly available to users.

By removing the requirement for a chessboard and introducing a more readily available reference object, we can further streamline the process and increase the practicality of using the software for body measurements. This improvement would contribute to a more user-friendly and inclusive experience, enabling a wider range of individuals to benefit from the accurate measurement capabilities of the software.

Additionally, the analysis of error differences revealed that a small percentage of measurements had larger errors, especially in the range of less than 0.5 meters or above 3 meters. These outliers could be attributed to factors such as occlusion, subject movement, or difficulties in accurately estimating depth at extreme distances. Future iterations of the software should focus on addressing these challenges to improve the overall accuracy and reliability of the measurements.

8.2 Final Results

The quantitative data for the 78 female participants were analysed using descriptive statistics. We captured 2D images of each participant based on the following scenarios:

1. With a plain background;

2. With a cluttered or textured background;

3. Using a pair of different body postures (A-pose and relax pose);

4. Using different distances from the camera;

5. With clothing that has visible creases;

6. Under different lighting conditions;

7. Using devices with different camera specifications:

    (a) Apple devices;

    (b) Samsung devices.

With 78 participants, this has resulted in an initial dataset with 312 data points.

In the following table, we will examine the complete descriptive statistics for specific body measurements, including chest, bust, waist, and hip circumferences, as well as shoulder and sleeve lengths. These statistics will be analysed based on the different stages previously mentioned, encompassing factors such as lighting conditions, background variations, distance between the camera and participant, body posture, body shape, and clothing type.

Body Type	Chest	Bust	Waist	Hips	shoulder	sleeve
Mean (cm)	0.78	0.78	0.96	0.78	0.15	0.14
Median (cm)	0.76	0.76	0.82	0.76	0.15	0.15
Best (cm)	0.01	0.01	0.02	0.02	0.001	0.001
Worst (cm)	3.14	2.66	3.19	3.82	0.33	0.31
Standard Deviation	0.55	0.50	0.62	0.56	0.07	0.07

Table 8.5 The best, worst, mean (ABS), median and standard deviation to assess the accuracy of female body torso measurements compared to tape measurements.

The results indicated promising outcomes for the software. In terms of length measurements, specifically shoulder and sleeve length, we achieved excellent accuracy, with the best mean error measuring only 0.15 cm and 0.14 cm, respectively. These findings demonstrate the reliability and precision of the software in capturing length-related body parameters.

For the circumference measurements, which included chest, bust, waist, and hips, the results were generally acceptable and showcased superior performance compared to other existing software in the market. The mean error values for these areas ranged from 0.78 cm to 0.96 cm, indicating a high level of accuracy in capturing the circumference measurements. However, it is worth noting that the waist area exhibited the highest mean error among all parameters. Further investigation revealed that the inaccuracy in waist measurements could be attributed to the participants' clothing. Loose-fitting clothes worn by some participants caused significant variations in the captured data, resulting in less accurate waist circumference measurements. Additionally, in the relax posture where the hands were placed next to the body, parts of the hands were occasionally captured as part of the waist, potentially contributing to increased inaccuracies in this specific region. Despite these challenges, the overall performance in the circumference measurements remained satisfactory.

Analysing the best and worst results across all parameters, we observed remarkable precision for the best case scenario, with mean errors as low as 0.01 cm for chest, bust, and as low as 0.02 cm for waist and hips. These exceptional outcomes indicate that the software can consistently provide highly accurate measurements. On the other hand, the worst-case scenarios revealed some outliers, with mean errors reaching 3.14 cm for waist, 2.66 cm for bust, 3.19 cm for hips, and 3.82 cm for shoulder. These deviations can be attributed to various factors, including participant posture variations, clothing constraints, and occasional challenges in accurately identifying specific body landmarks. Nonetheless, even with these outliers, the software outperformed competing solutions available in the market.

The overall performance was further validated through statistical analysis, where the median values closely mirrored the mean errors, confirming the reliability of the software's measurements. Additionally, the standard deviation values for all parameters were relatively low, ranging from 0.50 cm to 0.62 cm. This indicates the consistency and stability of the software's performance, as lower standard deviations suggest less variation in the measurements.

A comprehensive comparative analysis between the software and several popular applications available in the market was conducted as part of this study. The selected applications, namely 3Dlook, PreSize, TechMed, SizeStream (methreesixty), and Esenca, employ similar techniques as the software, utilising computer vision, machine learning, and 3D matching for body measurement purposes. These applications were chosen based on various factors such as accuracy, reliability, popularity, availability, and relevance to our research topic.

Through this experiments and evaluation, we compared the performance of these existing applications with the software using a dataset collected in 2022. The dataset comprised measurements of 30 participants across different body dimensions. The analysis focused on key body circumferences, namely chest, bust, waist, and hips.

The average differences in measurement errors, expressed in centimeters, were calculated for each application. When comparing the software's performance against the selected applications, our results showcased exceptional accuracy and outperformed the other applications across various body measurements. For chest circumferences, the software exhibited an average error of 0.78 cm, significantly lower than the average error of 5.80 cm reported by 3Dlook, PreSize, TechMed, SizeStream, and Esenca. Similarly, for bust circumferences, the software achieved an average error of 0.78 cm, surpassing the average error of 6.37 cm reported by the other applications. In terms of waist circumferences, the software demonstrated an average error of 0.96 cm, outperforming the average error of 5.2 cm from the comparison group. Finally, for hip circumferences, the software showcased an average error of 0.78 cm, which was substantially lower than the average error of 4.77 cm reported by the other applications.

This comprehensive comparative analysis highlights the strengths of the software in terms of measurement accuracy and reliability when compared to existing solutions. While each application has its own merits, the software consistently demonstrated superior performance across multiple body dimensions.

It is important to note that the comparative analysis was based on data collected in 2022, and advancements or updates to the existing applications may have occurred since then. However, at the time of our research, the software surpassed the other applications in terms of accuracy and reliability

In the following figures, we will look into a detailed analysis of the average error differences for each specific body section. For a comprehensive understanding, we will present individual plots that highlight the error differences in greater detail for each parameter, providing a more in-depth perspective on the performance of the software in capturing accurate measurements for different areas of the body.

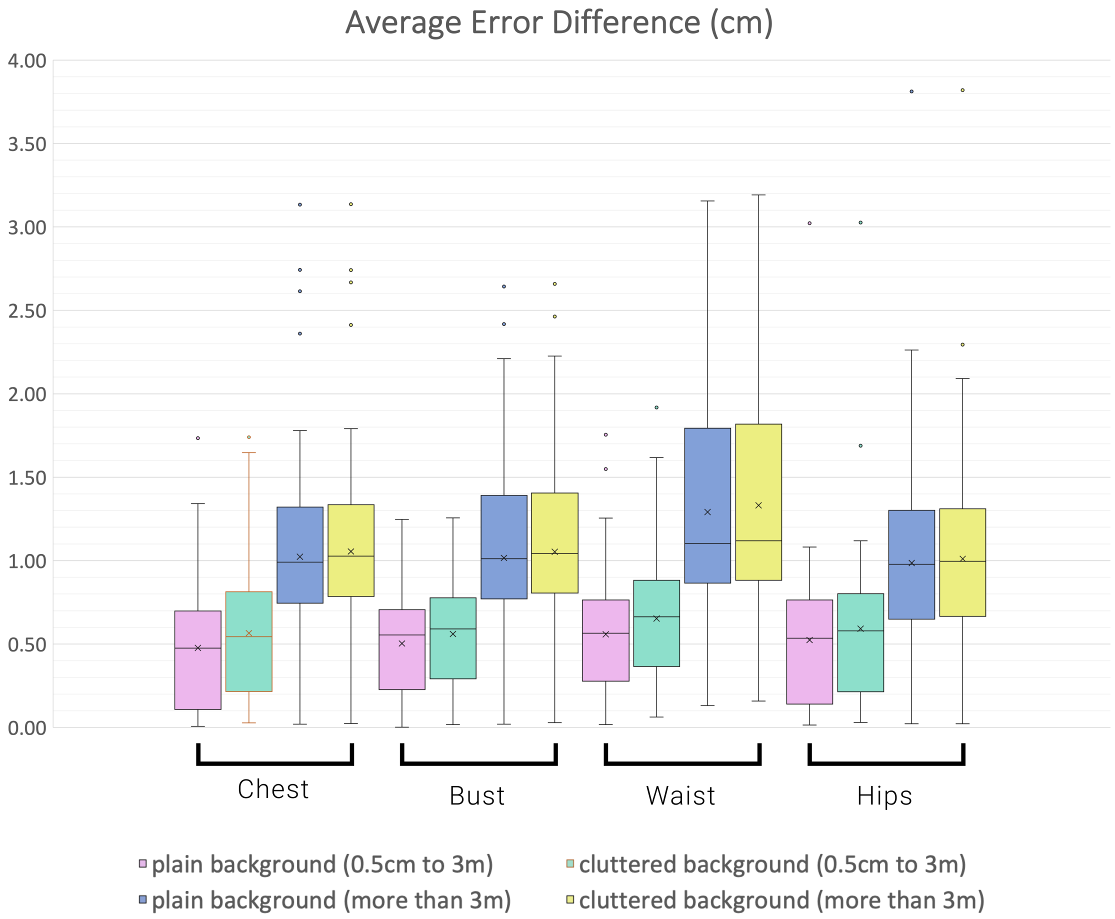


Fig. 8.1 Average margin of error for all circumference measurements.

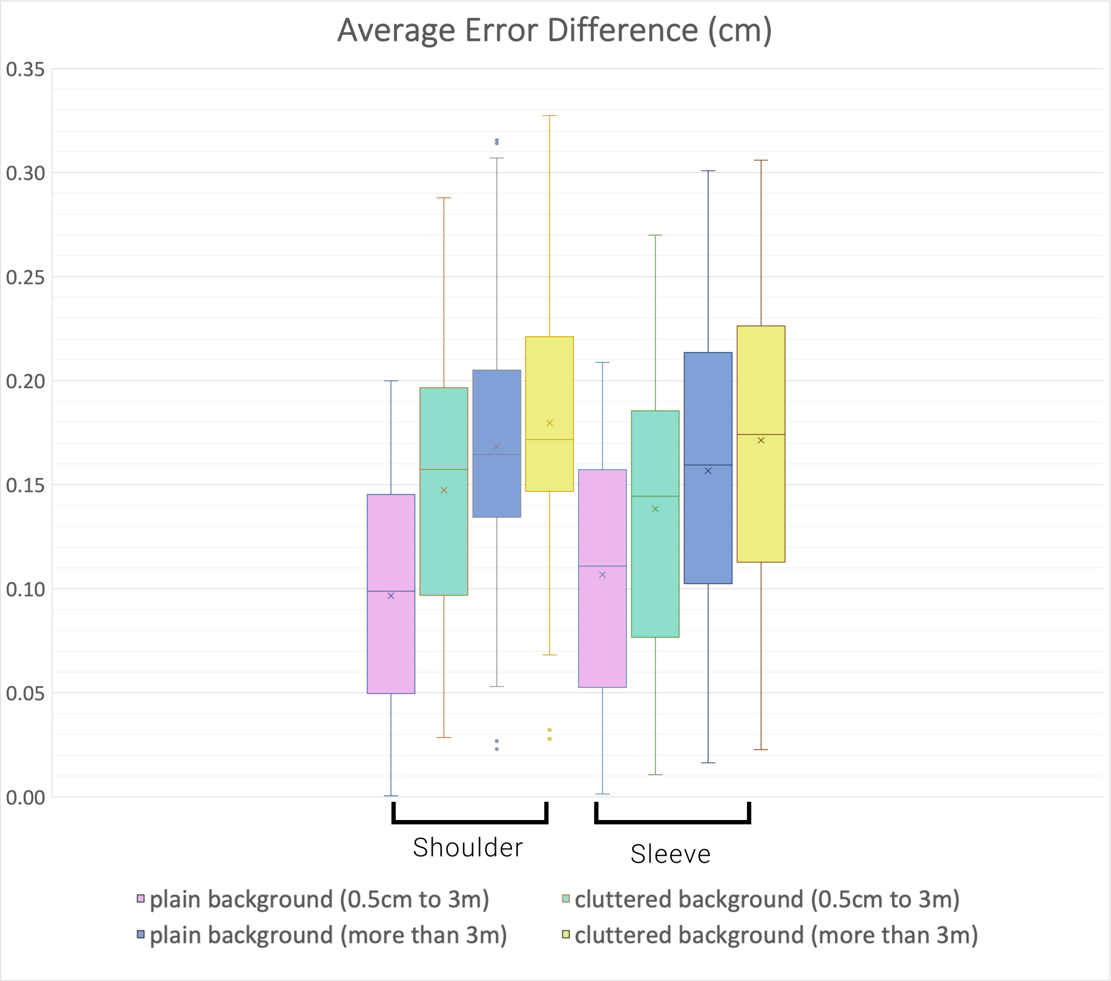


Fig. 8.2 Average margin of error for all length measurements.

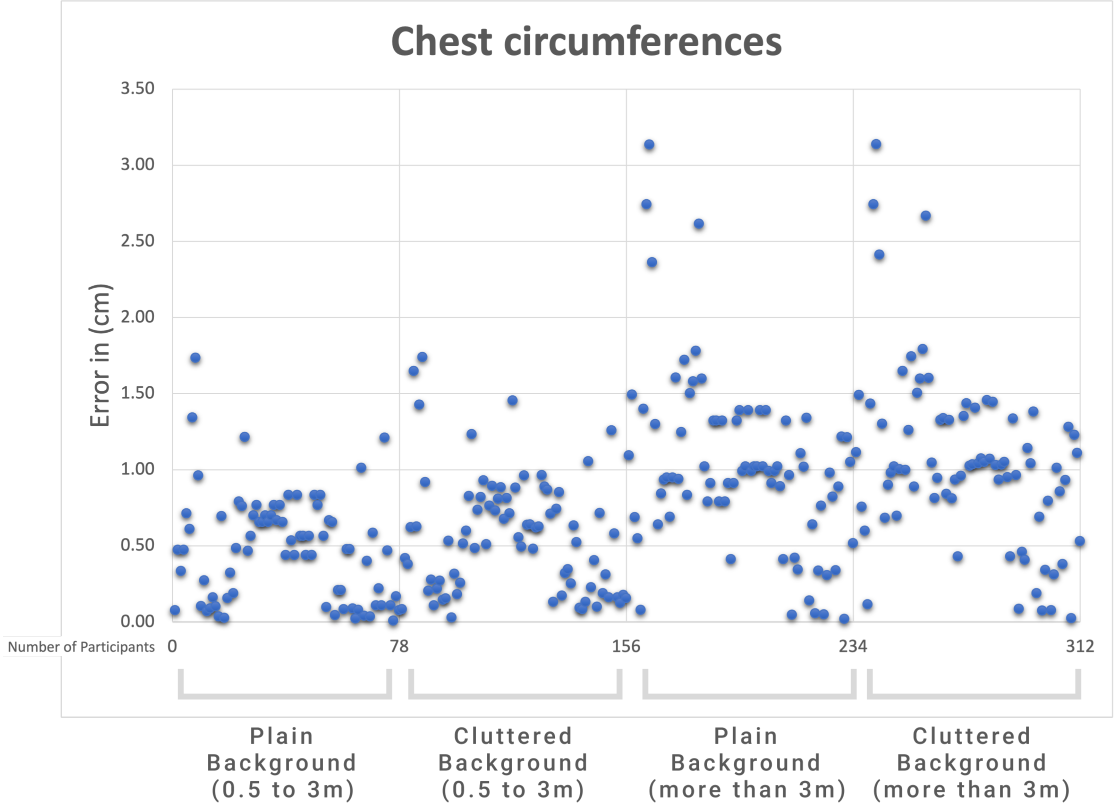


Fig. 8.3 Error Difference Analysis: This plot showcases the error difference (in centimeters) between software and tape measurements for chest circumference across various conditions and scenarios. The analysis encompassed 78 participants, resulting in a total of 312 data points. The evaluated factors included high and low resolutions, varying lighting conditions, correct and incorrect color balance, plain and cluttered backgrounds, distances ranging from 0 to 3 meters between the camera and participant, A-poses and T-poses and relax posture and relax posture, different body shapes, as well as tight and loose clothing. The plot provides a comprehensive overview of the software's performance under diverse conditions, offering valuable insights into its accuracy and potential areas for improvement.

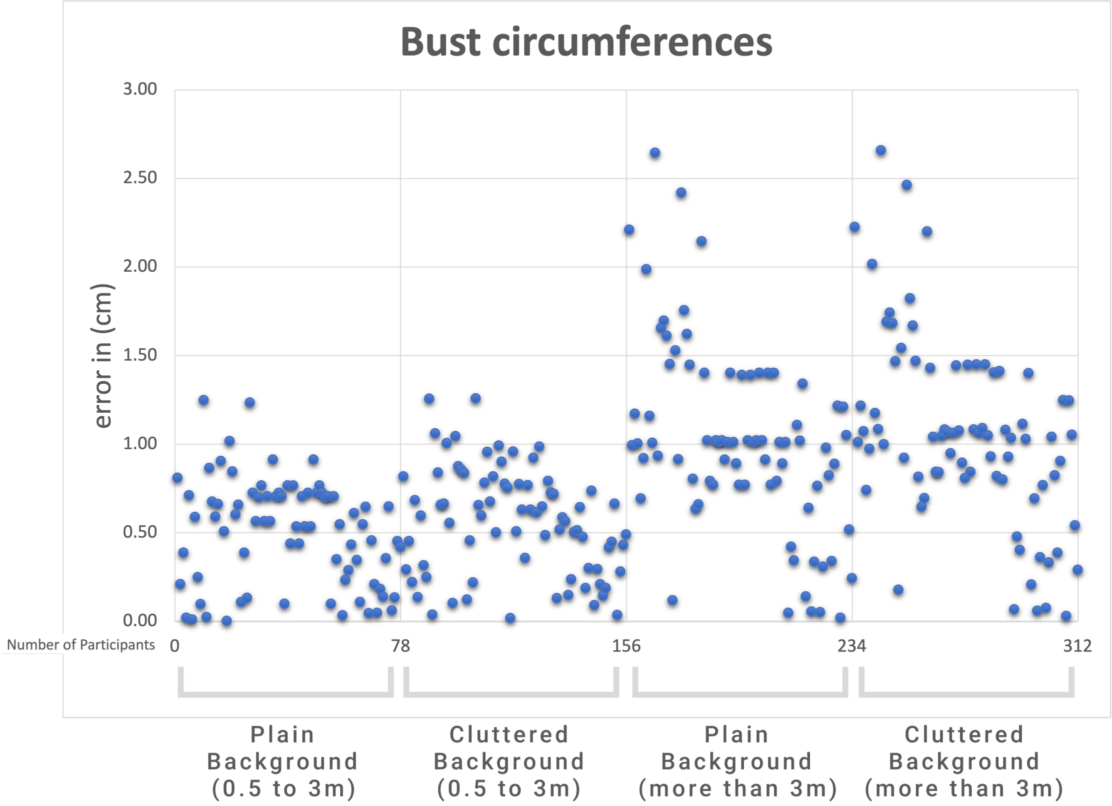


Fig. 8.4 Error Difference Analysis : This plot showcases the error difference (in centimeters) between software and tape measurements for bust circumference across various conditions and scenarios. The analysis encompassed 78 participants, resulting in a total of 312 data points. The evaluated factors included high and low resolutions, varying lighting conditions, correct and incorrect color balance, plain and cluttered backgrounds, distances ranging from 0 to 3 meters between the camera and participant, A-poses and T-poses and relax posture and relax posture, different body shapes, as well as tight and loose clothing. The plot provides a comprehensive overview of the software's performance under diverse conditions, offering valuable insights into its accuracy and potential areas for improvement.

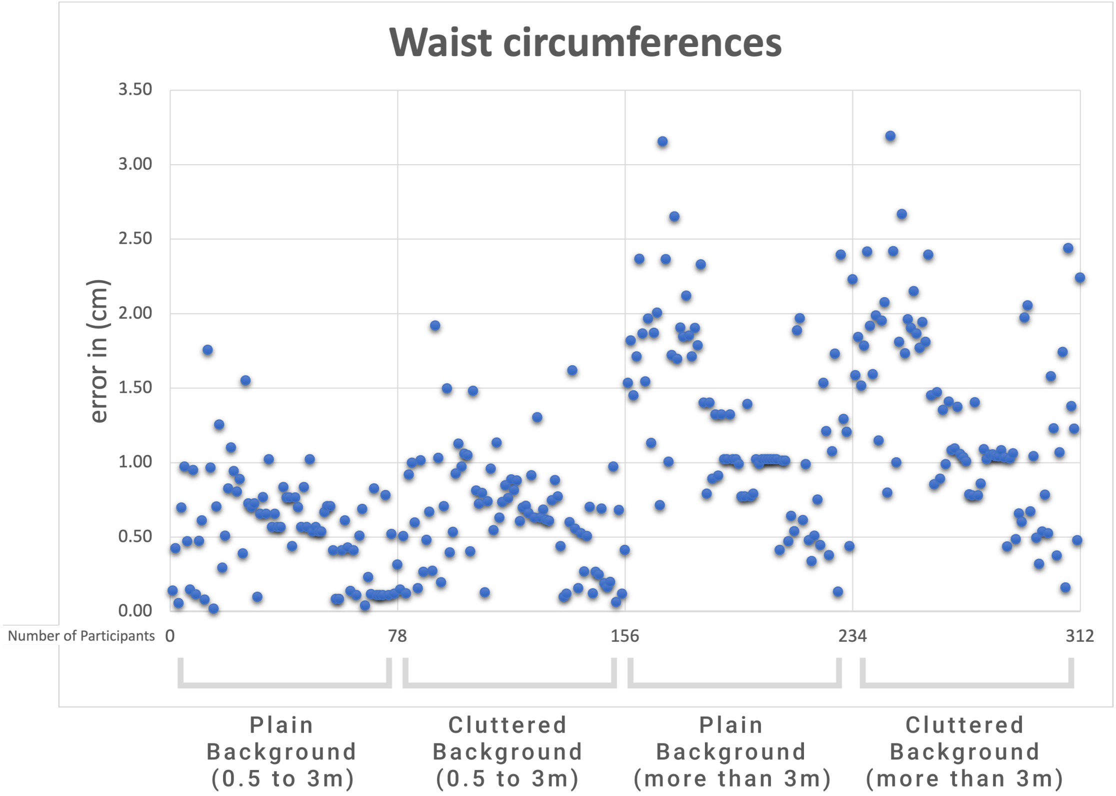


Fig. 8.5 Error Difference Analysis : This plot showcases the error difference (in centimeters) between software and tape measurements for waist circumference across various conditions and scenarios. The analysis encompassed 78 participants, resulting in a total of 312 data points. The evaluated factors included high and low resolutions, varying lighting conditions, correct and incorrect color balance, plain and cluttered backgrounds, distances ranging from 0 to 3 meters between the camera and participant, A-poses and T-poses and relax posture and relax posture, different body shapes, as well as tight and loose clothing. The plot provides a comprehensive overview of the software's performance under diverse conditions, offering valuable insights into its accuracy and potential areas for improvement.

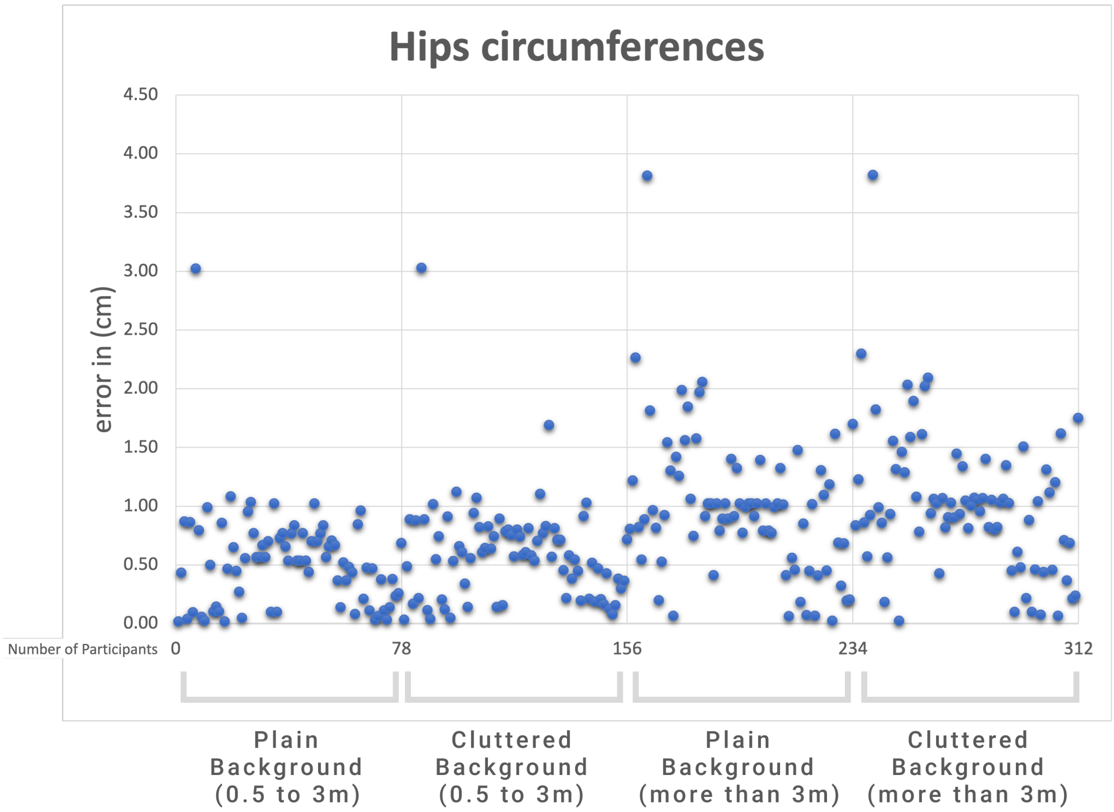


Fig. 8.6 Error Difference Analysis : This plot showcases the error difference (in centimeters) between software and tape measurements for hips circumference across various conditions and scenarios. The analysis encompassed 78 participants, resulting in a total of 312 data points. The evaluated factors included high and low resolutions, varying lighting conditions, correct and incorrect color balance, plain and cluttered backgrounds, distances ranging from 0 to 3 meters between the camera and participant, A-poses and T-poses and relax posture and relax posture, different body shapes, as well as tight and loose clothing. The plot provides a comprehensive overview of the software's performance under diverse conditions, offering valuable insights into its accuracy and potential areas for improvement.

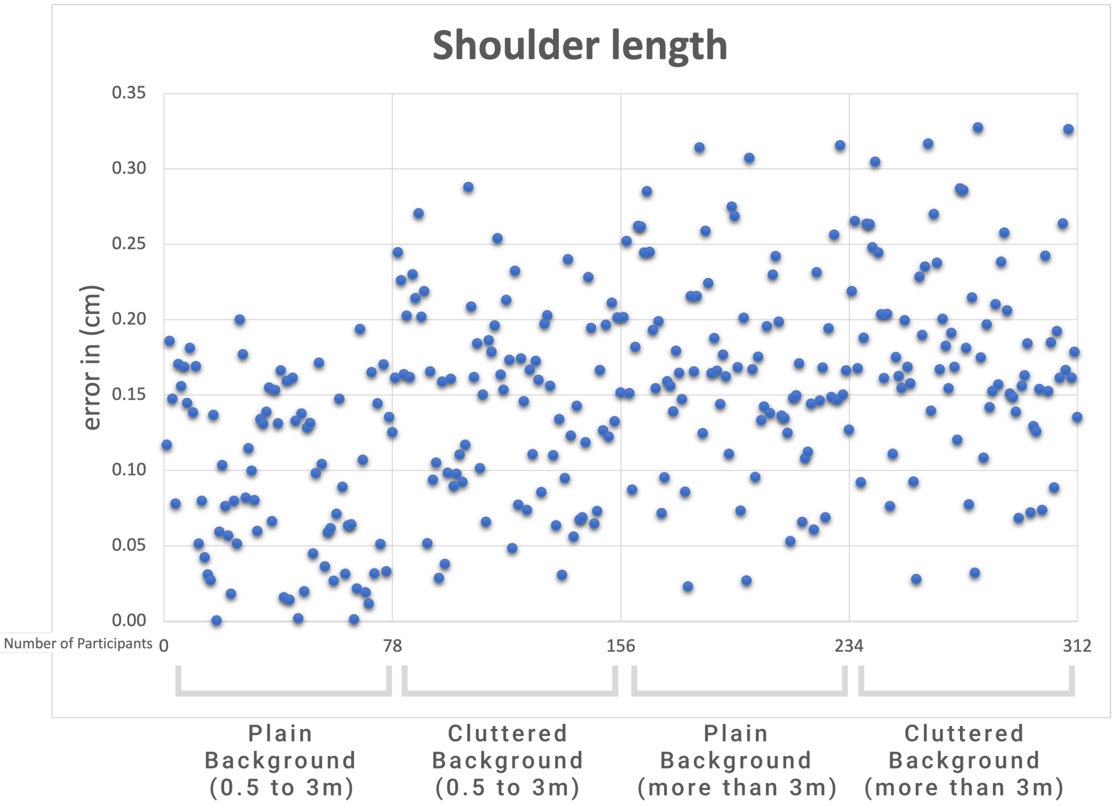


Fig. 8.7 Error Difference Analysis : This plot showcases the error difference (in centimeters) between software and tape measurements for shoulder length across various conditions and scenarios. The analysis encompassed 78 participants, resulting in a total of 312 data points. The evaluated factors included high and low resolutions, varying lighting conditions, correct and incorrect color balance, plain and cluttered backgrounds, distances ranging from 0 to 3 meters between the camera and participant, A-poses and T-poses and relax posture and relax posture, different body shapes, as well as tight and loose clothing. The plot provides a comprehensive overview of the software's performance under diverse conditions, offering valuable insights into its accuracy and potential areas for improvement.

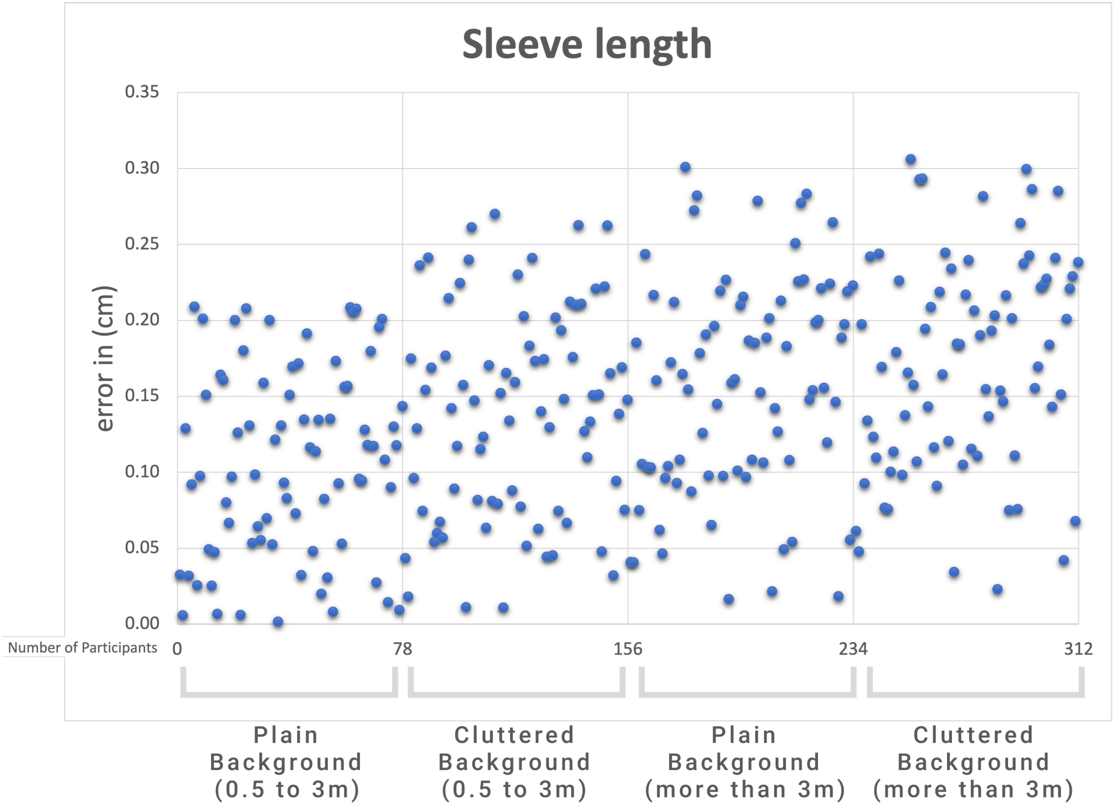


Fig. 8.8 Error Difference Analysis : This plot showcases the error difference (in centimeters) between software and tape measurements for sleeve length across various conditions and scenarios. The analysis encompassed 78 participants, resulting in a total of 312 data points. The evaluated factors included high and low resolutions, varying lighting conditions, correct and incorrect color balance, plain and cluttered backgrounds, distances ranging from 0 to 3 meters between the camera and participant, A-poses and T-poses and relax posture and relax posture, different body shapes, as well as tight and loose clothing. The plot provides a comprehensive overview of the software's performance under diverse conditions, offering valuable insights into its accuracy and potential areas for improvement.

In conclusion, this research has demonstrated the effectiveness and potential of the software in accurately capturing body measurements. However, there are several avenues for future work and improvement. Firstly, we can continue to refine and optimize the algorithms and techniques used in the software to enhance its accuracy and robustness, especially in challenging scenarios such as capturing measurements with voluminous garments or in complex lighting conditions. Secondly, expanding the range of measurement parameters to include additional body dimensions will further enhance the versatility and utility of the software in various applications.

Chapter 9

Conclusion

This chapter conclude the works that has been carried out in this computer science doctorate research. The methodology, experiments and results of the different aspects of this research are reviewed and used the evaluate the main hypothesis of this thesis. Future work is then outlined and discusses possible advances to this research.

9.1 Objective of this Research and the Main Hypothesis

Our research began with a clear objective: to investigate the possibility of developing a measurement system capable of extracting anthropometric measurements from multiple in-depth images that capture the surface of the human body. We aimed to improve the state-of-the-art in terms of accuracy-computational expenses, focusing on creating a software that would be compatible with the next generation of smartphone devices.

Our study's primary objectives can be divided into three broad categories. First, we attempted to recreate the human body using AI technologies, such as machine learning, computer vision, and 3D matching. Our objective was to estimate the human pose of the person displayed on the screen and provide body measurements for specific areas.

The second objective centred on reconstructing the human body from multiple depth images, particularly in situations where direct 3D data was unavailable. We acknowledged that systems using a single image have difficulty estimating the distance between the camera and the human body. Consequently, our solution aimed at avoiding this limitation.

Thirdly, we attempted to measure the circumferences of the human body using ellipse formulas. To accurately calculate body circumferences, it was necessary to design a system capable of selecting the optimal ellipse equation based on human body shape.

Our research was supported by several hypotheses. One significant assumption was that the public and industry required a more advanced technical solution to measure the surface of the human body, especially for industrial applications. We also hypothesised that the demand for this technology among younger generations seeking better-fitting garments is growing rapidly. This demand is a result of the growing digitisation of the apparel/fashion industry and the need for more precise, personalised fittings.

We also assumed that, with the help of recent advances in machine learning and computer vision, it is possible to measure the surface of the human body using next-generation smartphones. We hypothesised that the development of a measurement system capable of extracting human body measurements during movement, as well as body changes that affect the experience of wearing a garment, was not simply a possibility but an approaching reality.

In addition, we anticipated an increase in the precision of the state of the art. As opposed to the current margin of error of 2cm, we set our sights on achieving an average error of less than 1cm. Finally, we hypothesised that, using Objective 2 and a supervised learning approach, it should be possible to directly extract 3D information from multiple depth images in order to recover lost 3D data. We hypothesised that our system would accurately determine the distance between the human body and the camera.

Our research confirmed these hypotheses and exceeded our expectations. However, there were a few restrictions. Our system was accurate within a range of 0.5cm to 3m, but accuracy started declining at distances of less than 0.5cm and greater than 3m. In addition, our current system is partially automated. There is room for improvement in order to fully automate the system, and this can be accomplished with larger datasets.

The implications of our research for the real world are immense. We've proposed a technique that significantly improves the E-Commerce apparel/fashion industry by enabling users to estimate their body measurements using a pair of 2D smartphone images in a straightforward and accurate manner. It is a revolutionary tool for the modern world, where digital shopping is becoming routine and personalised, well-fitted clothing is in high demand.

In this study, 78 self-identified female participants were used to test our method. Compared to traditional tape measurement methods, we discovered that our method improved the accuracy of the state-of-the-art by a maximum of 1 cm. This represents a significant step forward in achieving our goal of an error of less than two centimetre.

9.2 Future Work

In considering future directions, there are several opportunities for further enhancing our software and advancing the field of body measurement technology. While currently, our system requires either a checkerboard or the user's height for calibration, our goal is to develop mechanisms that eliminate or minimize the need for these calibration references, streamlining the measurement process and enhancing user experience. Developing a larger database to test our system on a more diverse population will also be an objective of this future work. By adding additional images and postures to the training set, we aim to improve the consistency and precision of body shape extractions.

Having access to larger datasets will also allow us to accurately label more body parts. With more data, the system's ability to identify and label body shapes and parts in new images will improve. In turn, this will reduce the inaccuracy of the multi-ellipse model, maximising the overall efficiency of the system.

As we delve into larger datasets, we intend to improve our skin detection capabilities, aiming to distinguish human skin from non-skin regions with increased precision. A promising avenue of research is to investigate if such refined detection can help in labeling apparel and potentially estimating clothing thickness. Determining the boundary between clothing and actual body surface could be pivotal in reducing measurement errors. With the expanded datasets and by exploring the potential of our dual HSV and YCbCr technique, we hope to ascertain the feasibility of accurately estimating the body measurements, factoring in the influence of varied apparel thickness.

Another avenue for future improvement lies in exploring new measurement parameters beyond the ones considered in this study. While we focused on key body circumferences and length measurements, there are other important body dimensions that could be incorporated into the software (lower body), such as leg length. Expanding the range of measurement parameters will enhance the software's versatility and usefulness in various domains, including fashion, fitness, and healthcare.

As we continue to expand and improve our software, it is crucial that we ensure its inclusivity and relevance to diverse user groups. One important direction of our future research will be to adapt and optimize our body measurement technology for individuals with mobility challenges. Capturing accurate measurements for this particular group is essential, as they often face difficulties in obtaining well-fitted garments and assistive devices due to non-standard postures or physical conditions. By prioritizing this inclusion, we hope to bridge the gap in measurement precision, providing a comprehensive solution that caters to all, regardless of their mobility status. This endeavor not only enhances the universality of our software but also demonstrates our commitment to fostering inclusivity in technological advancements.

Furthermore, considering applications in different domains can also be a promising future direction. Our software has demonstrated its potential for accurate body measurements, and its application can extend beyond personal use. For instance, it could be integrated into clinical settings for monitoring body changes in medical interventions.

Another promising direction for our software's evolution is the integration of a "Virtual Try-On" feature. After our system has accurately captured a user's measurements, it would allow them to virtually "try on" apparel items from a digital catalogue. Not only does this ensure the fit and feel are right, but it also provides an immersive shopping experience from the comfort of one's home. This feature holds great potential for the e-commerce industry, eliminating the need for size guesswork and reducing product returns due to misfit

By pursuing these future directions, we can continue to push the boundaries of body measurement technology, improve the accuracy and usability of our software, and contribute to advancements in fields such as fashion, healthcare, sports, and body-related research. Continued research and development efforts in these areas will open up new possibilities for enhancing the accuracy, practicality, and impact of body measurement software in both consumer and professional contexts

In conclusion, our research sets the foundation for future applications in a variety of industries, including garment customization, virtual fitting, 3D human modelling, and related applications. The developed system satisfies the increasing demands of the fashion industry and establishes a new standard for anthropometric measurement precision. While there is still room for improvement, our research provides a promising foundation for advancing this technology and ultimately enhancing the digital fashion industry's user experience.

References

[1] (2023a). 3dlook.me. Available at https://3dlook.me/.

[2] (2023b). 3dprintingindustry. Available at https://3dprintingindustry.com/news/3dcopysystems-introducing-3-full-body-scanning-booth-77854/.

[3] (2023c). Esenca. Available at https://www.esenca.app/.

[4] (2023d). Optitrack. Available at https://optitrack.com/.

[5] (2023e). Presize. Available at www.presize.ai/. Available online.

[6] (2023f). Sizestream. Available at https://www.sizestream.com/.

[7] (2023g). Techmed3d. Available at https://techmed3d.com/.

[8] (2023h). Vyoo. Available at https://www.vyoo.ai/.

[9] (2023i). Zoolando. Available at https://www.zalando.co.uk/.

[10] 20947, I. (2021). Performance evaluation protocol for digital fitting systems — part 1: Accuracy of virtual human body representation.

[11] 3DLook (2019). The evolution of body measuring: How advances in technology are driving the future of made to measure fashion.

[12] 3DLookMe (2019). The evolution of body measuring: How advances in technology are driving the future of made to measure fashion. *3DLOOK.me*.

[13] 8559-1, I. (2017). Size designation of clothes–part 1: Anthropometric definitions for body measurement.

[14] 8559-2, I. (2017). Size designation of clothes — part 2: Primary and secondary dimension indicators.

[15] 8559-3, I. (2018). Size designation of clothes — part 3: Methodology for the creation of body measurement tables and intervals.

[16] Adrian, R. (2016). Measuring size of objects in an image with opencv. [Accessed: 21/06/2021]. [Online]. Available: https://www.pyimagesearch.com/2016/03/28/measuring-size-of-objects-in-an-image-with-opencv/.

[17] Adrian, R. (2019). Opencv tutorial: A guide to learn opencv. [Accessed: 21/06/2021]. [Online]. Available: https://www.pyimagesearch.com/2018/07/19/opencv-tutorial-a-guide-to-learn-opencv/.

[18] Agarwal, A. (2006). *Machine Learning for Image Based Motion Capture*. Theses, Institut National Polytechnique de Grenoble - INPG.

[19] Ahn, S. H., Park, J., Nam, Y.-J., and Yun, M. H. (2015). 1c1-2 analysis and usability testing of the 3d scanning method for anthropometric measurement of the elderly. *The Japanese Journal of Ergonomics*, 51(Supplement):S394–S397.

[20] Aiyangar, S. and Berndt, B. (1985). *Ramanujan's Notebooks: Part I*. Ramanujan's Notebooks. Springer New York.

[21] Albiol, A., Torres, L., and Delp, E. (2001). Optimum color spaces for skin detection. In *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, volume 1, pages 122–124 vol.1.

[22] Aldrich, W., Smith, B., and Dong, F. (1998). Obtaining repeatability of natural extended upper body positions: its use in comparisons of the functional comfort of garments. *Journal of Fashion Marketing and Management: An International Journal*, 2(4):329–351.

[23] Andriluka, M., Pishchulin, L., Gehler, P., and Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693.

[24] Ang, K.-S. and Mitchell, H. L. (2010). Non-rigid surface matching and its application to scoliosis modelling. *The Photogrammetric Record*, 25(130):105–118.

[25] Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., and Davis, J. (2005). Scape: shape completion and animation of people. In *ACM transactions on graphics (TOG)*, volume 24, pages 408–416. ACM.

[26] Apeagyei, P. R. (2010). Application of 3d body scanning technology to human measurement for clothing fit. *International Journal of Digital Content Technology and its Applications*, 4(7):58–68.

[27] Apple-Developer (2020). Arkit developer. Available at https://developer.apple.com/documentation/arkit (12/03/2020).

[28] Arsalan Soltani, A., Huang, H., Wu, J., Kulkarni, T. D., and Tenenbaum, J. B. (2017). Synthesizing 3d shapes via modeling multi-view depth maps and silhouettes with deep generative networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1511–1519.

[29] Ashdown, S. P. and DeLong, M. (1995). Perception testing of apparel ease variation. *Applied Ergonomics*, 26(1):47–54.

[30] Ashdown, S. P. and Dunne, L. (2006). A study of automated custom fit: Readiness of the technology for the apparel industry. *Clothing and Textiles Research Journal*, 24(2):121–136.

[31] Ashmawi, S., Alharbi, M., Almaghrabi, A., and Alhothali, A. (2019). Fitme: Body measurement estimations using machine learning method. *Procedia Computer Science*, 163:209–217.

[32] Barone, F., Marrazzo, M., and Oton, C. (2020). Camera calibration with weighted direct linear transformation and anisotropic uncertainties of image control points. *Sensors (Basel)*, 20(4):1175.

[33] Berndt, B. C. (2012). *Ramanujan's notebooks: Part III*. Springer Science & Business Media.

[34] Berryman, D. R. (2012). Augmented reality: a review. *Medical reference services quarterly*, 31(2):212–218.

[35] Bodla, N., Singh, B., Chellappa, R., and Davis, L. S. (2017). Soft-nms–improving object detection with one line of code. In *Proceedings of the IEEE international conference on computer vision*, pages 5561–5569.

[36] Bogo, F., Romero, J., Loper, M., and Black, M. J. (2014a). FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA. IEEE.

[37] Bogo, F., Romero, J., Loper, M., and Black, M. J. (2014b). Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3794–3801.

[38] Bottino, A. and Laurentini, A. (2001). A silhouette based technique for the reconstruction of human movement. *Computer Vision and Image Understanding*, 83(1):79–95.

[39] Boyer, E., Bronstein, A. M., Bronstein, M. M., Bustos, B., Darom, T., Horaud, R., Hotz, I., Keller, Y., Keustermans, J., Kovnatsky, A., et al. (2011). Shrec 2011: robust feature detection and description benchmark. *arXiv preprint arXiv:1102.4258*.

[40] Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2008). *Numerical geometry of non-rigid shapes*. Springer Science & Business Media.

[41] Burger, W. (2016). Zhang's camera calibration algorithm: In-depth tutorial and implementation. Technical Report HGB16-05, University of Applied Sciences Upper Austria, School of Informatics, Communications and Media, Softwarepark 11, 4232 Hagenberg, Austria. Revised in April 2020.

[42] Bye, E., LaBat, K., McKinney, E., and Kim, D.-E. (2008). Optimized pattern grading. *International Journal of Clothing Science and Technology*, 20(2):79–92.

[43] Bye, E., Labat, K. L., and Delong, M. R. (2006). Analysis of body measurement systems for apparel. *Clothing and Textiles Research Journal*, 24(2):66–79.

[44] Cai, X., Fan, Y., Chen, S., and Wang, Z. (2018). A study on measuring the body sizes of chinese females based on 3d anthropometry. *Applied Sciences*, 8(11):2245.

[45] Chandler, J. H., Fryer, J. G., and Jack, A. (2005). Metric capabilities of low-cost digital cameras for close range surface measurement. *The Photogrammetric Record*, 20(109):12–26.

[46] Chandra, R. N., Febriyan, F., and Rochadiani, T. H. (2018). Single camera body tracking for virtual fitting room application. In *Proceedings of the 2018 10th International Conference on Computer and Automation Engineering*, pages 17–21.

[47] Chang, H.-T., Li, Y.-W., Chen, H.-T., Feng, S.-Y., and Chien, T.-T. (2013). A dynamic fitting room based on microsoft kinect and augmented reality technologies. In *International Conference on Human-Computer Interaction*, pages 177–185. Springer.

[48] Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

[49] Chen, X., Li, J., Zhang, J., and Li, Y. (2019). Real-time non-planar calibration using a fisheye lens for camera networks. *IEEE Transactions on Image Processing*, 28(6):2968–2980.

[50] Cheng, Z., Ligouri, A., Fogle, R., and Webb, T. (2015). Capturing human motion in natural environments. *Procedia Manufacturing*, 3:3828–3835.

[51] Cho, C.-S., Park, J.-Y., Boeing, A., and Hingston, P. (2010). An implementation of a garment-fitting simulation system using laser scanned 3d body data. *Computers in Industry*, 61(6):550–558.

[52] Choi, S. and Ashdown, S. P. (2011). 3d body scan analysis of dimensional change in lower body measurements for active body positions. *Textile Research Journal*, 81(1):81–93.

[53] Choi, S.-Y. and Ashdown, S. P. (2010). Application of lower body girth change analysis using 3d body scanning to pants patterns. *Journal of the Korean Society of Clothing and Textiles*, 34(6):955–968.

[54] Chong, A. K., Milburn, P., Newsham-West, R., and Voert, M. (2009). High-accuracy photogrammetric technique for human spine measurement. *The Photogrammetric Record*, 24(127):264–279.

[55] Choutas, V., Muller, L., Huang, C.-H. P., Tang, S., Tzionas, D., and Black, M. J. (2022). Accurate 3d body shape regression using metric and semantic attributes.

[56] Daanen, H. A., Brunsman, M. A., and Robinette, K. M. (1997). Reducing movement artifacts in whole body scanning. In *Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No. 97TB100134)*, pages 262–265. IEEE.

[57] D'Apuzzo, N. (2002). Surface measurement and tracking of human body parts from multi-image video sequences. *ISPRS journal of Photogrammetry and Remote Sensing*, 56(5-6):360–375.

[58] Deli, R., Di Gioia, E., Galantucci, L. M., and Percoco, G. (2010). Automated landmark extraction for orthodontic measurement of faces using the 3-camera photogrammetry methodology. *Journal of Craniofacial Surgery*, 21(1):87–93.

[59] Dibra, E., Jain, H., Oztireli, C., Ziegler, R., and Gross, M. (2017). Human shape from silhouettes using generative hks descriptors and cross-modal neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4826–4836.

[60] Diwan, T., Anirudh, G., and Tembhurne, J. V. (2023). Object detection using yolo: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, 82(6):9243–9275.

[61] Fan, J., Yu, W., and Hunter, L. (2004). *Clothing appearance and fit: Science and technology*. Elsevier.

[62] Fan, Y., Yan, X., Chen, S., Liu, W., and Wang, Z. (2017). 3d body scanning and anthropometric study of chinese children and adolescents. *Applied Sciences*, 7(11):1155.

[63] Fang, A. C. and Pollard, N. S. (2003). Efficient synthesis of physically valid human motion. *ACM Transactions on Graphics (TOG)*, 22(3):417–426.

[64] Fang, H.-S., Xie, S., Tai, Y.-W., and Lu, C. (2017). Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2334–2343.

[65] Faraz, B. (2021). Human body measurements using computer vision. Available at https://github.com/farazBhatti/Human-Body-Measurements-using-Computer-Vision (30/10/2022).

[66] Federation, N. R. (2018). What do gen z shoppers really want? Available at https://nrf.com/research/what-do-gen-z-shoppers-really-want (30/10/2022).

[67] Gavali, P. and Banu, J. S. (2019). Deep convolutional neural network for image classification on cuda platform. In *Deep learning and parallel computing environment for bioengineering systems*, pages 99–122. Elsevier.

[68] Gayomali, C. (2014). Here's what it's like to step into a 3-d body scanner for a custommade suit. *Fast Company*, 3.

[69] Gazzola, P., Pavione, E., Pezzetti, R., and Grechi, D. (2020). Trends in the fashion industry. the perception of sustainability and circular economy: A gender/generation quantitative approach. *Sustainability*, 12(7).

[70] Gilbert, A., Volino, M., Collomosse, J., and Hilton, A. (2018). Volumetric performance capture from minimal camera viewpoints. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 566–581.

[71] Gonzalez-Sosa, E., Vera-Rodriguez, R., Fierrez, J., and Ortega-Garcia, J. (2014). Comparison of body shape descriptors for biometric recognition using mmw images. In *2014 22nd International Conference on Pattern Recognition*, pages 124–129.

[72] Gudla, N. K. and Challa, S. (2018). An efficient approach for camera calibration using distorted planar objects. *IEEE Transactions on Instrumentation and Measurement*, 67(7):1654–1664.

[73] Hajraoui, A. and Sabri, M. (2014). Face detection algorithm based on skin detection, watershed method and gabor filters. *International Journal of Computer Applications*, 94(6):33–39.

[74] Hartmann, W., Galliani, S., Havlena, M., Van Gool, L., and Schindler, K. (2017). Learned multi-patch similarity. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1586–1594.

[75] Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., and Seidel, H.-P. (2009). A statistical model of human pose and body shape. In *Computer graphics forum*, volume 28, pages 337–346. Wiley Online Library.

[76] Heikkila, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1106–1112.

[77] Hines, T. and Bruce, M. (2007). *Fashion marketing: contemporary issues*. Routledge.

[78] Huang, Z., Li, T., Chen, W., Zhao, Y., Xing, J., LeGendre, C., Luo, L., Ma, C., and Li, H. (2018a). Deep volumetric video from very sparse multi-view performance capture. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 336–354.

[79] Huang, Z., Li, T., Chen, W., Zhao, Y., Xing, J., LeGendre, C., Ma, C., Luo, L., and Li, H. (2018b). Deep volumetric video from very sparse multi-view performance capture. In *European Conference on Computer Vision (ECCV)*.

[80] Huck, J., Maganga, O., and Kim, Y. (1997). Protective overalls: evaluation of garment design and fit. *International Journal of Clothing Science and Technology*, 9(1):45–61.

[81] Imran, A., Anita, B., Marco, B., Achim, B., Saskia, H., and Felix, R. (2019). What do gen z shoppers really want? Available at https://www.mckinsey.com/industries/retail/our-insights/the-influence-of-woke-consumers-on-fashion (30/10/2022).

[82] Inman, D. (2022). Retail returns increased to $761 billion in 2021 as a result of overall sales growth gradients(hog). *National Retail Federation*.

[83] Istook, C. (2008). Three-dimensional body scanning to improve fit. *Advances in apparel production*, pages 94–116.

[84] Istook, C. L. and Hwang, S.-J. (2001). 3d body scanning systems with application to the apparel industry. *Journal of Fashion Marketing and Management: An International Journal*, 5(2):120–132.

[85] Ivan, B. (2016). 3-d body scanners in stores can help you find apparel in your true size.

[86] Jain, H. P., Subramanian, A., Das, S., and Mittal, A. (2011). Real-time upper-body human pose estimation using a depth camera. In *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, pages 227–238. Springer.

[87] Jang, J. (2012). Skeleton tracking and body measurements with kinecs. Available at https://www.youtube.com/watch?v=e4HTlmEA6Ec (12/03/2020).

[88] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., TaoXie, Fang, J., imyhxy, Michael, K., Lorna, V, A., Montes, D., Nadar, J., Laughing, tkianai, yxNONG, Skalski, P., Wang, Z., Hogan, A., Fati, C., Mammana, L., AlexWang1900,

Patel, D., Yiwei, D., You, F., Hajek, J., Diaconu, L., and Minh, M. T. (2022). ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference.

[89] Johnson, C. L., Paulose-Ram, R., Ogden, C. L., Carroll, M. D., Kruszan-Moran, D., Dohrmann, S. M., and Curtin, L. R. (2013). National health and nutrition examination survey. analytic guidelines, 1999-2010.

[90] Kalogerakis, E., Averkiou, M., Maji, S., and Chaudhuri, S. (2017). 3d shape segmentation with projective convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3779–3788.

[91] Kanazawa, A., Black, M. J., Jacobs, D. W., and Malik, J. (2018). End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Regognition (CVPR)*.

[92] Kim, J. and Forsythe, S. (2008). Adoption of virtual try-on technology for online apparel shopping. *Journal of Interactive Marketing*, 22(2):45–59.

[93] Koenderink, J. J. (2010). *Color for the Sciences*. MIT press.

[94] Kolkur, S., Kalbande, D., Shimpi, P., Bapat, C., and Jatakia, J. (2017). Human skin detection using rgb, hsv and ycbcr color models. *arXiv preprint arXiv:1708.02694*.

[95] Kovacs, L., Zimmermann, A., Brockmann, G., Gühring, M., Baurecht, H., Papadopulos, N., Schwenzer-Zimmerer, K., Sader, R., Biemer, E., and Zeilhofer, H. (2006). Three-dimensional recording of the human face with a 3d laser scanner. *Journal of plastic, reconstructive & aesthetic surgery*, 59(11):1193–1202.

[96] Larsen, P. K., Hansen, L., Simonsen, E. B., and Lynnerup, N. (2008). Variability of bodily measures of normally dressed people using photomodeler® pro 5. *Journal of forensic sciences*, 53(6):1393–1399.

[97] Lee, J., Chai, J., Reitsma, P. S., Hodgins, J. K., and Pollard, N. S. (2002). Interactive control of avatars animated with human motion data. In *ACM Transactions on Graphics (ToG)*, volume 21, pages 491–500. ACM.

[98] Lee j, M. and Wei, L. (2018). Gen z is set to outnumber millennials within a year. Available at https://www.bloomberg.com/news/articles/2018-08-20/gen-z-to-outnumber-millennials-within-a-year-demographic-trends?leadSource=uverify%20wall (30/10/2022).

[99] Li, Y. and Dai, D. X. (2006). *Biomechanical engineering of textiles and clothing*. Woodhead Publishing.

[100] Lievendag, N. (2018). Updated for 2018: Realitycapture review.

[101] Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.

[102] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., and Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision*, 128:261–318.

[103] Lu, J.-M. and Wang, M.-J. J. (2008). Automated anthropometric data collection using 3d whole body scanners. *Expert Systems with Applications*, 35(1-2):407–414.

[104] Lu, J.-M., Wang, M.-J. J., and Mollard, R. (2010). The effect of arm posture on the scan-derived measurements. *Applied Ergonomics*, 41(2):236–241.

[105] Madeline, M. (2020). Consumers to return half of online clothing purchases this holiday season. Available at https://www.prnewswire.com/news-releases/consumers-to-return-half-of-online-clothing-purchases-this-holiday-season-300760466.html (03/09/2020).

[106] Mansurov, N. (2010). What is focal length in photography? *Photography Life*.

[107] Maragos, P. (2005). Morphological filtering for image enhancement and feature detection. *The Image and Video Processing Handbook*, pages 135–156.

[108] Mark, H. (2020). Study: Half of online clothing purchases get returned. Available at https://www.consumeraffairs.com/news/study-half-of-online-clothing-purchases-get-returned-120618.html (03/09/2020).

[109] Mark, P. (2019). Motion capture gym kit means you hit the perfect yoga pose every time.

[110] Mason, A. (2020). Making 3d models with photogrammetry.

[111] Mattmann, C., Clemens, F., and Tröster, G. (2008). Sensor for measuring strain in textile. *Sensors*, 8(6):3719–3732.

[112] McKinney, E. C., Bye, E., and LaBat, K. (2012). Building patternmaking theory: a case study of published patternmaking practices for pants. *International Journal of Fashion Design, Technology and Education*, 5(3):153–167.

[113] Mogavero, T. (2020). Clothed in conservation: Fashion & water. Sustainable Campus, Florida State University. Accessed on: July 23, 2020.

[114] Mohammad, M. (2019). How to measure human body clothing by using mocap systems.

[115] Morera, , Sánchez, , Moreno, A. B., Sappa, A. D., and Vélez, J. F. (2020). Ssd vs. yolo for detection of outdoor urban advertising panels under multiple variabilities. *Sensors (Basel)*, 20(16):4587.

[116] N, A. (2019). Motion capture | alverina studios.

[117] Nepal, U. and Eslamiat, H. (2022). Comparing yolov3, yolov4 and yolov5 for autonomous landing spot detection in faulty uavs. *Sensors*, 22(2):464.

[118] Nourbakhsh Kaashki, N., Hu, P., and Munteanu, A. (2021). Anet: A deep neural network for automatic 3d anthropometric measurement extraction. *IEEE Transactions on Multimedia*, pages 1–1.

[119] Okabe, K. and Kurokawa, T. (2006). A study of the relationships between breast vibration, clothing pressure and dislocation under running condition for designing sports brassiere. *Descente Sports Sci*, 27:75–85.

[120] Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., et al. (2016). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 741–754. ACM.

[121] Paek, K.-J. (2009). Fit analysis for mens bodice pattern using 3d scans-compared to traditional fit evaluation. *Journal of the Korean Society of Clothing and Textiles*, 33(1):139–148.

[122] Patacchiola, M. and Cangelosi, A. (2017). Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognition*, 71:132–143.

[123] Paul, W. (2018). A history of measuring. *MOOD SEWCIETY*.

[124] Paul-Louis, P. (2017). An introduction to different types of convolutions in deep learning. *medium*.

[125] Percoco, G. (2011). Digital close range photogrammetry for 3d body scanning for custom-made garments. *The Photogrammetric Record*, 26(133):73–90.

[126] Petrova, A. and Ashdown, S. P. (2008). Three-dimensional body scan data analysis: Body size and shape dependence of ease values for pants' fit. *Clothing and Textiles Research Journal*, 26(3):227–252.

[127] Pishchulin, L., Wuhrer, S., Helten, T., Theobalt, C., and Schiele, B. (2017). Building statistical shape spaces for 3d human modeling. *Pattern Recognition*.

[128] PlanetarySciences (2018). Astronomical distances.

[129] Pollefeys, M. (2004). Simultaneous estimation of principal point, focal length, and camera pose for omnidirectional cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1380–1393.

[130] Pons-Moll, G., Pujades, S., Hu, S., and Black, M. J. (2017). Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics (TOG)*, 36(4):73.

[131] Prasad, k. (2020). Ar and deep learning based automatic human body measurement system. Available at https://github.com/KNU-Machine-Vision-Lab/body-measure (30/10/2022).

[132] Pribanić, T., Mrvoš, S., and Salvi, J. (2010). Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning. *Image and Vision Computing*, 28(8):1255–1266.

[133] Rahmad, C., Asmara, R., Putra, D., Dharma, I., Darmono, H., and Muhiqqin, I. (2020). Comparison of viola-jones haar cascade classifier and histogram of oriented gradients (hog) for face detection. In *IOP conference series: materials science and engineering*, volume 732, page 012038. IOP Publishing.

[134] Raj, B. (2019). An overview of human pose estimation with deep learning. *A Medium Corporation*.

[135] Rajalingappaa, S. (2022). Detection or localization and segmentation. Pages 145–205. Available at https://www.oreilly.com/library/view/deep-learning-for/9781788295628/4fe36c40-7612-44b8-8846-43c0c4e64157.xhtml (21/04/2022).

[136] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection.

[137] Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement.

[138] Roberts, M. (2019). A formula for the perimeter of an ellipse.

[139] Robinette, K. M., Daanen, H., and Paquet, E. (1999). The caesar project: a 3-d surface anthropometry survey. In *Second international conference on 3-D digital imaging and modeling (cat. No. PR00062)*, pages 380–386. IEEE.

[140] Ross, W. and Shepherd, J. (1991). Body size and shape assessment in young adult females using three-dimensional photonic scanning. *Ergonomics*, 34(7):877–886.

[141] Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., and Pal, C. (2018). Fashion-gen: The generative fashion dataset and challenge. *arXiv preprint arXiv:1806.08317*.

[142] Safonova, A., Hodgins, J. K., and Pollard, N. S. (2004). Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. In *ACM Transactions on Graphics (ToG)*, volume 23, pages 514–521. ACM.

[143] Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., and Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2304–2314.

[144] Saito, S., Simon, T., Saragih, J., and Joo, H. (2020a). Blender tutorial - convert your photo to 3d mesh - pifuhd full tutorial. Available at https://www.youtube.com/watch?v=LWDGR5v3-3o&feature=emb_logo (03/09/2020).

[145] Saito, S., Simon, T., Saragih, J., and Joo, H. (2020b). Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[146] Salvi, J. and Moreno-Noguer, F. M. (2004). Automatic estimation of principal point position and focal length from single views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1355–1360.

[147] Sato, H. and Cohen, M. (2010). Using motion capture for real-time augmented reality scenes. In *Proceedings of the 13th International Conference on Humans and Computers*, pages 58–62. University of Aizu Press.

[148] Satya, M. (2020). Geometry of image formation. [Accessed: 21/06/2021]. [Online]. Available: https://learnopencv.com/geometry-of-image-formation/.

[149] Satya, M. and Kaustubh, S. (2020). Camera calibration using opencv. [Accessed: 21/06/2021]. [Online]. Available: https://learnopencv.com/camera-calibration-using-opencv/.

[150] Schofield, N. A. and LaBat, K. L. (2005). Exploring the relationships of grading, sizing, and anthropometric data. *Clothing and Textiles Research Journal*, 23(1):13–27.

[151] Shapiro, L. and Stockman, G. (2000). Computer vision. Pages 145–205. Available at http://lmvicente.com/ee7716/docs/art4ShapiroComputerVision.pdf (21/04/2021).

[152] Shin, S.-J. H. and Istook, C. L. (2007). The importance of understanding the shape of diverse ethnic female consumers for developing jeans sizing systems. *International Journal of Consumer Studies*, 31(2):135–143.

[153] Sik-Ho, T. (2019). Review: Deeplabv3+ — atrous separable convolution (semantic segmentation). *medium*.

[154] Simmons, K. P. (2001). Body measurement techniques: a comparison of three-dimensional body scanning and physical anthropometric methods. *Unpublished A1 paper, North Carolina State University, Raleigh*, 23.

[155] Simmons, K. P. and Istook, C. L. (2003). Body measurement techniques: Comparing 3d body-scanning and anthropometric methods for apparel applications. *Journal of Fashion Marketing and Management: An International Journal*, 7(3):306–332.

[156] Slater, K. (1977). Comfort properties of textiles. *Textile progress*, 9(4):1–70.

[157] Smith, P. (2023). Clothing: Quarterly sales volume index in great britain 2015-2022. *Statista*. Published on Feb 22, 2023.

[158] Sohn, M. (2012). Analysis of upper body measurement change using motion capture.

[159] Sohn, M. and Bye, E. (2014). Exploratory study on developing a body measurement method using motion capture. *Clothing and Textiles Research Journal*, 32(3):170–185.

[160] Tamnay, B. (2019). Bringing people into ar. Available at https://developer.apple.com/videos/play/wwdc2019/607?time=1365 (12/03/2020).

[161] Tan, M. and Le, Q. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks.

[162] Tan, W. R., Chan, C. S., Yogarajah, P., and Condell, J. (2012). A fusion approach for efficient human skin detection. *IEEE Transactions on Industrial Informatics*, 8(1):138–147.

[163] Taşdemir, Ş., Yakar, M., Ürkmez, A., and İnal, Ş. (2008). Determination of body measurements of a cow by image analysis. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, page 70. ACM.

[164] Team, S. (2019). A brief history of sizing systems. *Sizolution Team*.

[165] Thai, L. H., Hai, T. S., and Thuy, N. T. (2012). Image classification using support vector machine and artificial neural network. *International Journal of Information Technology and Computer Science*, 4(5):32–38.

[166] Tzutalin (2015). Labelimg. Free Software: MIT License.

[167] Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. Graphicon*, volume 3, pages 85–92. Moscow, Russia.

[168] Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., et al. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.

[169] Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. (2020). Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 390–391.

[170] Wang, R., Wei, L., Vouga, E., Huang, Q., Ceylan, D., Medioni, G., and Li, H. (2016). Capturing dynamic textured surfaces of moving targets. In *European Conference on Computer Vision*, pages 271–288. Springer.

[171] Wang, X. and Huang, T. (2010). A novel method for estimating the principal point of a camera. *IEEE Transactions on Image Processing*, 19(8):2065–2070.

[172] Wang, Y.-J., Mok, P.-Y., Li, Y., and Kwok, Y.-L. (2009). Effects of joint movements on body measurement and clothing ease design.

[173] Wang, Y. M., Li, Y., and Zheng, J. B. (2010). A camera calibration technique based on opencv. In *The 3rd International Conference on Information Sciences and Interaction Sciences*, pages 403–406.

[174] World Wildlife Fund (2023). Threats: Water scarcity. World Wildlife Fund. Accessed on: July 23, 2020.

[175] Wu, C., Varanasi, K., and Theobalt, C. (2012). Full body performance capture under uncontrolled and varying illumination: A shading-based approach. In *European Conference on Computer Vision*, pages 757–770. Springer.

[176] Xiaohui, T., Xiaoyu, P., Liwen, L., and Qing, X. (2018). Automatic human body feature extraction and personal size measurement. *Journal of Visual Languages & Computing*, 47:9–18.

[177] Xiu, Y., Li, J., Wang, H., Fang, Y., and Lu, C. (2018). Pose Flow: Efficient online pose tracking. In *BMVC*.

[178] Xu, H., Yu, Y., Zhou, Y., Li, Y., and Du, S. (2013). Measuring accurate body parameters of dressed humans with large-scale motion using a kinect sensor. *Sensors*, 13(9):11362–11384.

[179] Yang, W., Li, S., Ouyang, W., Li, H., and Wang, X. (2017). Learning feature pyramids for human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1281–1290.

[180] Yang, Y., Yu, Y., Zhou, Y., Du, S., Davis, J., and Yang, R. (2014). Semantic parametric reshaping of human body models. In *2014 2nd International Conference on 3D Vision*, volume 2, pages 41–48. IEEE.

[181] Yao, J., Zhang, H., Zhang, H., and Chen, Q. (2008). R&d of a parameterized method for 3d virtual human body based on anthropometry. *Age*, 20(30):30–40.

[182] Yoon, S. and Kim, H. (2011). A new efficient method for camera calibration using only the principal point. *IEEE Transactions on Instrumentation and Measurement*, 60(6):2027–2035.

[183] Yu, W. and Xu, B. (2010). A portable stereo vision system for whole body surface imaging. *Image and vision computing*, 28(4):605–613.

[184] Yuri Boykov, V. K. (2019). Skeleton tracking and body measurements with kinecs.

[185] Zeng, X., Koehl, L., Chen, Y., Happiette, M., Bruniaux, P., Ng, R., and Yu, W. (2008). A new method of ease allowance generation for personalization of garment design. *International journal of clothing science and technology*.

[186] Zhang, C. and Zhang, Z. (2010). *A Survey of Recent Advances in Face Detection*.

[187] Zhang, H. and Deng, Q. (2019). Deep learning based fossil-fuel power plant monitoring in high resolution remote sensing images: A comparative study. *Remote Sensing*, 11(9):1117.

[188] Zhang, Z. (2000a). Camera calibration using regularized radial basis function networks. *IEEE Transactions on Instrumentation and Measurement*, 49(3):650–654.

[189] Zhang, Z. (2000b). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.

[190] Zhang, Z. (2004). Camera calibration with one-dimensional objects. *IEEE transactions on pattern analysis and machine intelligence*, 26(7):892–899.

[191] Zhou, J., Yu, W., and Ng, S.-P. (2011). Methods of studying breast motion in sports bras: a review. *Textile research journal*, 81(12):1234–1248.

[192] Zhou, X., Zhu, M., Pavlakos, G., Leonardos, S., Derpanis, K. G., and Daniilidis, K. (2018). Monocap: Monocular human motion capture using a cnn coupled with a geometric prior. *IEEE transactions on pattern analysis and machine intelligence*, 41(4):901–914.

[193] Zwane, P. E., Sithole, M., and Hunter, L. (2010). A preliminary comparative analysis of 3d body scanner, manually taken girth body measurements and size chart measurements. *International journal of consumer Studies*, 34(3):265–271.

Appendix A

Public Dimensions

A.1 Supplementary Materials

Research Website: https://www.montazerian.com/research/ Mohammad [114]

ResearchGate Profile : https://www.researchgate.net/profile/Mohammad_Montazerian2

A.2 Published Papers

1. Research Workshop 2: Technical and conceptual innovations (EVA London 2020) - How to Measure Human Body When Wearing Clothes by Using Camera Systems

2. Conference Paper: Measuring the Human Body from a Single Camera, with Applications to the Clothing and Fashion Industry (Proceedings of 3DBODY.TECH 2022 13th Int. Conference and Exhibition)

3. Conference Paper: Adaptive Body Circumference Measurement Technique Using Ellipse Formula (Proceedings of 3DBODY.TECH 2022 13th Int. Conference and Exhibition)

4. Article: Simple Hybrid Camera-Based System Using Two Views for Three-Dimensional Body Measurements (Symmetry (mdpi) - 2024)

5. In Process of Publishing this paper: A Survey Paper on the Evolution of Human Body Measurement (2024)

Appendix B

Research Survey

Research Questions(Survey) https://www.surveymonkey.co.uk/r/PHYQXH6

The specific research questions are:

1. Age:

2. Gender:

3. Have you ever had Online clothing shopping ?

   (a) Yes

   (b) No

   - If No why ?

4. How often do you purchase new item of clothing ?

   (a) 1-5times a year

   (b) 5-10times a year

   (c) Once a month

   (d) 2/3 times a month

   (e) 4+ per month

5. What are your 2 main criteria when purchasing clothing?
   (a) Quality
   (b) Price
   (c) Comfort
   (d) Style
   (e) Size
   (f) Material
6. Have you felt any problem while conducting online purchase ?
   (a) Yes
   (b) No
       - If yes what kind of problems ?
7. Please tell us if you agree or disagree with each of the following statement?
   (a) Not being able to physically inspect the goods before purchase
   (b) Online shopping is convenient to shop
   (c) Online shopping saves time and energy
   (d) Not too sure about my size
8. Have you ever used any application to measure your body ?
   (a) Yes
   (b) No
       - If yes please tell us the name of the application
       - If No why
9. Which one of the method do you prefer to garment your body ?

(a) Traditional way

(b) Application

- Why ?

10. How many times did you return the product because the size of the items & products it was not your good fit ?

    (a) 0

    (b) 1-3

    (c) 3-6

    (d) 6+

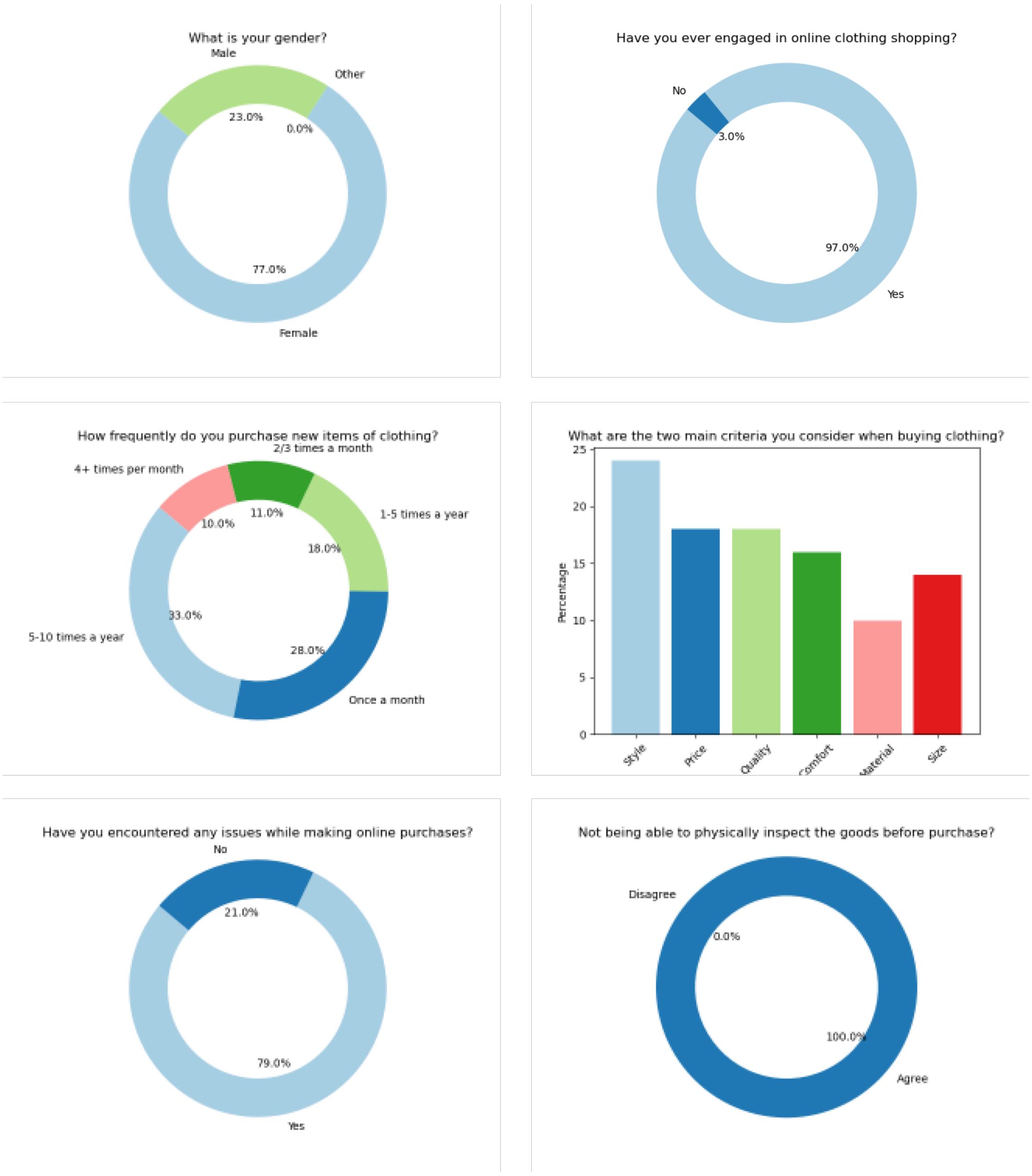


Fig. B.1 A survey was carried out with a sample size of 100 individuals, consisting of 77 female participants and 23 male participants.

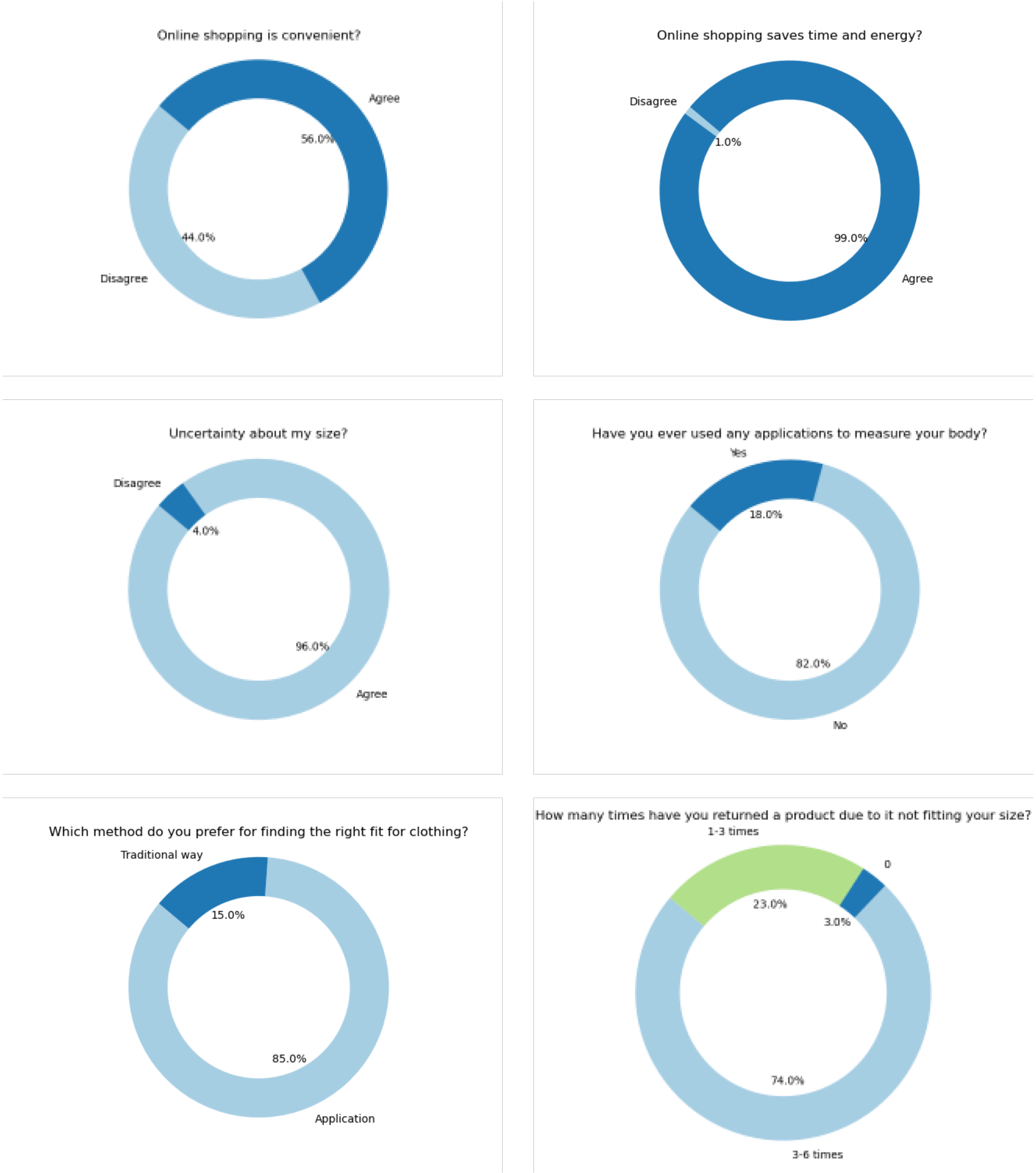


Fig. B.2 A survey was carried out with a sample size of 100 individuals, consisting of 77 female participants and 23 male participants.

Appendix C

Appendices[1]

Phase 1 Studies The research method is considered in the following order:

1. Research design
2. Choosing the participants
3. Limitations
4. Dataset and data collection
5. Analysis of the data

C.1 Research Design

Research Questions?

1. Which methods are the most reliable option to capture and measure the human body reconstruction in the fashion industry?
   (a) Mo-Cap System
   (b) Machine Learning approaches

   (c) 3D Scanner

   (d) Mobile Scanner approaches

   (e) Photogrammetry approaches

2. Is the new generation of smartphone cameras a reliable method to measure the human body reconstruction in motion?

   (a) How accurate is the new smartphone scanner in comparison to the other methods?

   (b) How are the new smartphone scanners recovering lost data (i.e. depth) during capturing for human body reconstruction?

3. For the human body in motion are there any changes to the human body surface?

   (a) For the body scanner and the Mo-Cap system, are there any changes between the body measurements extracted?

   (b) Are there any changes in other parts of the body during the shoulder girdle movement?

   (c) What are the maximum and minimum changes of the elected upper body measurements during the shoulder movement?

4. Among participants with different body sizes, is there any difference in measurement changes?

C.2 Machine Learning and the Mo-Cap System

Collections of samples are used for the process of deriving successful mathematical models for the process of machine learning. The model is useful in instances where the process is difficult to model exactly and it is often statistical. Even when partial models are available, in verifying different models and approximating their parameters by fitting them to the data, the learning technique can play an important role.

In markerless motion capture, by learning how to predict 3D configuration directly from images that are observed, we can escape the need for human body rendering and accurate data modelling. As for turning data, these need a set of images and their associated configurations. One way to create such information and data is using motion capture, which is one of the technologies that exists today. We are able to record the movements of humans or any subjects performing a range of different activities. As the details of representation vary between systems, such methods can typically provide the human body configuration as a set of angles between limbs at the main body joints. The corresponding images can be found by providing an artificial image based on the motion capture data, because it is used to synthesise motion in graphics, or by using a synchronised conventional video camera with the camera of the motion capture systems. Therefore, the data will be compressed into mathematical models that can be used to predict the human pose from new images by using machine learning methods.

How to show the image content is one of the challenges. Usually, for input in the learning process, into a vector to create a feature space representation some low-level visual attributes are gathered. As mentioned in section 1.3, to recognise the suitable set of features for measuring human body appearance, there are multiple elements creating difficulty. However, by using a set of pictures as training data, we can learn about effective representations, by creating statistics from the pictures and the main responses featuring on them.

Dynamic methods are other features of human motion analysis that can be useful for learning-based methods. The development of human body 3D models (avatars) in motion needs a clear understanding of biomechanics and complicated interdependencies of various parts of the body, which in the next chapter we review in more detail. In contrast, it is much easier to capture the correlations and the temporal dependencies between the movement of multiple parts of the body through statistical models. To synthesise natural-looking motion for human motion analysis, machine learning technology has begun to be used more and more Safonova et al. [142], Fang and Pollard [63], Lee et al. [97].

C.2.1 Selection of technologies

To measure the human body surface in motion, different techniques have been considered such as computer vision, machine learning and 3D matching. Such technologies are key parameters to record and measure the human body measurements and also analyse to provide accurate numerical data simultaneously on new smartphones. New smartphones contain a maximum of three cameras. Each camera can capture the human body (x, y and z coordinates of each marker) over time. With the help of computer vision, first to analyse photos or video from which we acquire and process the human body. Secondly, by using a neural network to detect and determine key-points and produce a set of probability maps for each key point (machine learning). Finally, by using statistical modelling and 3D geometry to create a 3D model of the human body, based on the detected key-points, which allows to accurately obtain human body measurements (3D matching). Recent advance in AI-powered technologies on smartphone scanners, make in it possible, to see the 3D data from any orientation since it can convert the 2D data to 3D. This system can provide measurement change over time, instead of a measurement at a certain stage of motion, this is one of the main benefits of the new mobile generation.

Note: Selection of Motion Capture system has been also considered for our phase 1 studies, however, owing to the lack of ability of such system to capture the human body mass, we have decided to not to continue our work based on this system. Please refer to section C.4 for more information

C.3 Dataset and Data Collection

Preparation of the application, as well as subject registration and recording process, are the three steps of the data collection procedure.

C.3.1 Preparation of the application

Preparation of the application will be done in the following order: preparing the camera phone and working area, testing, taking few photos from multiple perspective or a short video (360 view). iPhone 11 and later devices will be chosen (is that these devices contain minimum of 3 cameras).

Before initiating the scanning process, a test has to determine where the participants needs to perform a natural and relaxed posture, standing still, looking forward, and breathing normally during the scanning.

C.3.2 Subject Registration

The research process will take place at Goldsmiths, University of London, where we can access to our tools and system. Once participants are selected through the screening process, they will be asked to schedule a 30-minute appointment with the researcher.

C.3.3 Recording Process

Participants will be asked to stand in front of the camera twice to perform the motion. The reason for asking participants to perform their motion twice is that we want to compare the first results with the second results, for which participants will do some stretching exercises to warm up the muscles (5 minutes). The participants will perform the motion several times according to a reference video and the researcher's guidance, for this study.

Participants will be asked to take the natural and relaxed posture, standing still, looking forward, and breathing normally during the scanning process, in order to compare data on the maximum and minimum increases and decreases of the body in motion. When the participants are ready, they will be asked to perform the selected movements. As the shoulder is one of the joints of the body that every other part of the upper body is connected to, the participants will be asked to perform an arm rotation in order to compare the maximum and minimum expansion and contraction of the muscles that are connected to the shoulder.

C.4 Selection of Motion Capture System

To measure the human body surface in motion, different systems have been considered such as an optical motion tracking system and body motion capturing in 3D with the help of ARKits framework on new smartphones. Such technologies can record and measure the human body measurements and also analyse to provide accurate numerical data simultaneously. An optical motion capture tracking system contains several cameras (minimum six) to capture various reflective markers on the moving object, and new smartphones contain a maximum of three cameras. Each camera is able to capture the human body (x, y and z coordinates of each marker) over time. Thirty-six markers will be attached to each participant to capture the full human body shape in optical Mo-Cap-system while capturing the body motion in 3D by using a new generation of smartphones, with the help of machine learning, first to estimate the human pose of the person on the screen and afterwards to create a full-fledged, high-fidelity skeleton for every individual limb by using the pose. Mo-Cap systems have the ability to see the 3D data from any orientation since it is able to convert the 2D data to 3D. This system can provide measurement change over time, instead of a measurement at a certain stage of motion, this is one of the main benefits of the motion capture system.

C.5 Dataset and Data Collection

Preparation of the Mo-Cap system, as well as subject registration, placement of marker process, 3D scanning process and the Mo-Cap process, are the five steps of the data collection procedure.

C.5.1 Preparation of the Mo-Cap System

Preparation of the motion capture system will be done in the following order: preparing the camera and working area, calibration, and pilot testing. The 12 cameras will be placed around the Goldsmiths studio and the motive system is turned on in order to aim the cameras precisely and focus the lens at the centre of the studio.

Calibration is necessary for using the two hinged axes and wand provided with the system after positioning the cameras. To store calibration parameters for each participant, it is necessary to create a new calibration file before starting to capture each individual body in the studio. Calibration included six essentials steps:

1. Prepare and optimise the capture setup
2. Clear existing masks.
3. Use Mask Visible to mask extraneous reflections that cannot be removed from capture volume.
4. Wanding
5. Calculate and review the calibration results.
6. Set the Ground Plane.

Before initiating the capture process, a pilot test has to determine where the markers need to be attached to the human body and how many markers are necessary. It is necessary to examine and refine the marker positioning during the pilot test. To measurement placement, attaching the markers is fundamental.

C.5.2 Subject Registration

The research process will take place in the Mo-Cap Studio at Goldsmiths, University of London, where the motion tracking system and body scanner is located. Once participants are selected through the screening process, they will be asked to schedule a 30-minute appointment with the researcher.

C.5.3 Placement of Marker Process

Markers are the essentials part of the motion capture tracking system and by placing them on a wrong part of the body or missing them, data will be distorted; therefore, the markers need to be placed at the correct location on the body corresponding to the selected measurements.

Markers are around 15 mm in diameter and are attached to the participant's body with double-sided sticky tape, which is made for this system. The markers will be placed at the two end-points of each measurement.

Anatomical landmarks related to pattern development helped to determine the exact location of the markers. The markers will be attached to the human body where the anatomical landmarks, such as the acromion and the cervical, are obvious. The motion software has a feature to show an avatar of a person with the markers attached, which needs to be followed, as the software and the program will then be ready to capture the human body.

C.5.4 3D Scanning Process

As part of this research study, scanning the body plays an important role in getting high-quality information and quite accurate data from every joint of the human body. The body scanner is not a good option to measure the human body in motion; however, because it is a natural and relaxed posture that participants take, the data that we can collect from them are significantly better than any other technology that already exists in the market.

One of the advantages of the laser scanner compared to motion capturing technology is that the participants do not need to attach any sensors or wear any specific suits in order to capture their body. So once the participants were ready, the process of scanning can start and they should stand in front of the laser scanner with a natural and relaxed posture, standing still, looking forward, and breathing normally during the scanning.

To obtain a high-quality measurement from the human body, we came across software that already exist in the store (Structure Sensor Mark II [1]). This single laser camera will attach to the iPhone's or iPad's camera and users need to download the Tech-Med application (free apps in the app store) to start the process of scanning the human body.

Once the application is opened and the camera is connected to the iPad or iPhone, the process of scanning will start. Thus, the basic standing posture can be described: the participants will be asked to stand upright but naturally and relaxed, looking straight ahead, arms hanging relaxed at the sides with feet together Aldrich et al. [22]. Once the scanning process is finished, the data will be checked for quality and, if needed, the participants will be asked to participate in the process of scanning once again. Afterwards, the participants will be asked to get ready and move to the Mo-Cap Studio lab for motion capturing.

[1] Structure Sensor Mark II is powerful laser scanner that will be mounted on an iPad or iPhone camera, which requires no cables or PC nearby and offers you real freedom to scan

C.5.4.1 TechMed3D

As part of this research study, one of the main goals is to gather as much data as possible to create a dataset that can be used in future for the implementation and final testing. One of the software that is considered is TeckMed3D, which controls and optimises 3D scanning with the Structure Sensor Mark II. This app can obtain very accurate data from the human body shape structure for individual segments of the body.

C.5.5 Mo-Cap Process

Participants will be asked to stand in the motion capture studio twice to perform the motion. The reason for asking participants to perform their motion twice is that we want to compare the first results with the second results, for which participants will do some stretching exercises to warm up the muscles (5 minutes). The participants will perform the motion several times according to a reference video and the researcher's guidance, for this study.

With 12 cameras surrounding the participants, they will stand in the optical motion tracking space. participants will be asked to take the same posture as in the body scanning process, in order to compare data on the maximum and minimum increases and decreases of the body in motion. When the participants are ready, they will be asked to perform the selected movements. As the shoulder is one of the joints of the body that every other part of the upper body is connected to, the participants will be asked to perform an arm rotation in order to compare the maximum and minimum expansion and contraction of the muscles that are connected to the shoulder.

C.6 Measuring App with ARKit

ARKit is a framework in IOS 11 that allows us to blend digital objects in the real world. It is been used in Pokemon Go or IKEA app and etc. ARKit runs on the Apple A9 and A10 processors.

C.6.1 ARuler app

Lets start with coding in Xcode. first thing we need to do is to create emtpy project that supports ARkit which Xcode 9 has a separate template for this.

I have created very simple UI for this demo, to represent some labels which inform us about the final outcome and results. Figure 4.4 shows a view from the application

So, what it is next, we need to setup the session and set the **sessionConfig**. As part of using ARkit we are able to detect only horizontal planes but it is not a problem in our app, as we will relate to the features points.

Based on Features points - points identified by the framework as part of the surface, we are able to detect things

```
func setupScene() {
        sceneView.delegate = self
        sceneView.session = session

        session.run(sessionConfig, options: [.resetTracking, .removeExistingAnchors])

        resetValues()
}
```

For better detection, we need a function for detecting feature points, then to calculate the distance between two points we need to trasnlate it into 3D coordinates. The code below shows an extension of the **ARSCNView**:

The next step that I did was to implement the method to **updateAtTime** and after that call our detection function.

```
func realWorldVector(screenPos: CGPoint) -> SCNVector3? {
        let planeTestResults = self.hitTest(screenPos, types: [.
    featurePoint])
        if let result = planeTestResults.first {
            return SCNVector3.positionFromTransform(result.worldTransform
    )
        }

        return nil
}
```

```
 func renderer(_ renderer: SCNSceneRenderer, updateAtTime time:
    TimeInterval) {
        DispatchQueue.main.async {
            self.detectObjects()
        }
}
```

Finally, I have developed a function to save the beginning and end point that we want to measure. Therefore, we need to calculate the distance between the first and end points which is called **typeSCNVector3**. Please refer to the code below to see how this section is done:

```
 func distance(from vector: SCNVector3) -> Float {
        let distanceX = self.x - vector.x
        let distanceY = self.y - vector.y
        let distanceZ = self.z - vector.z

        return sqrtf((distanceX * distanceX) + (distanceY * distanceY) +
        (distanceZ * distanceZ))
 }
```

To conclude, the results that I have achieved was surprised me a lot as it was quite accurate, however, because of **featurePoits** we get some strange results sometimes and this can be easily fixed by moving the device and find the best spot or place to start measuring.

C.7 Key Landmark by ISO

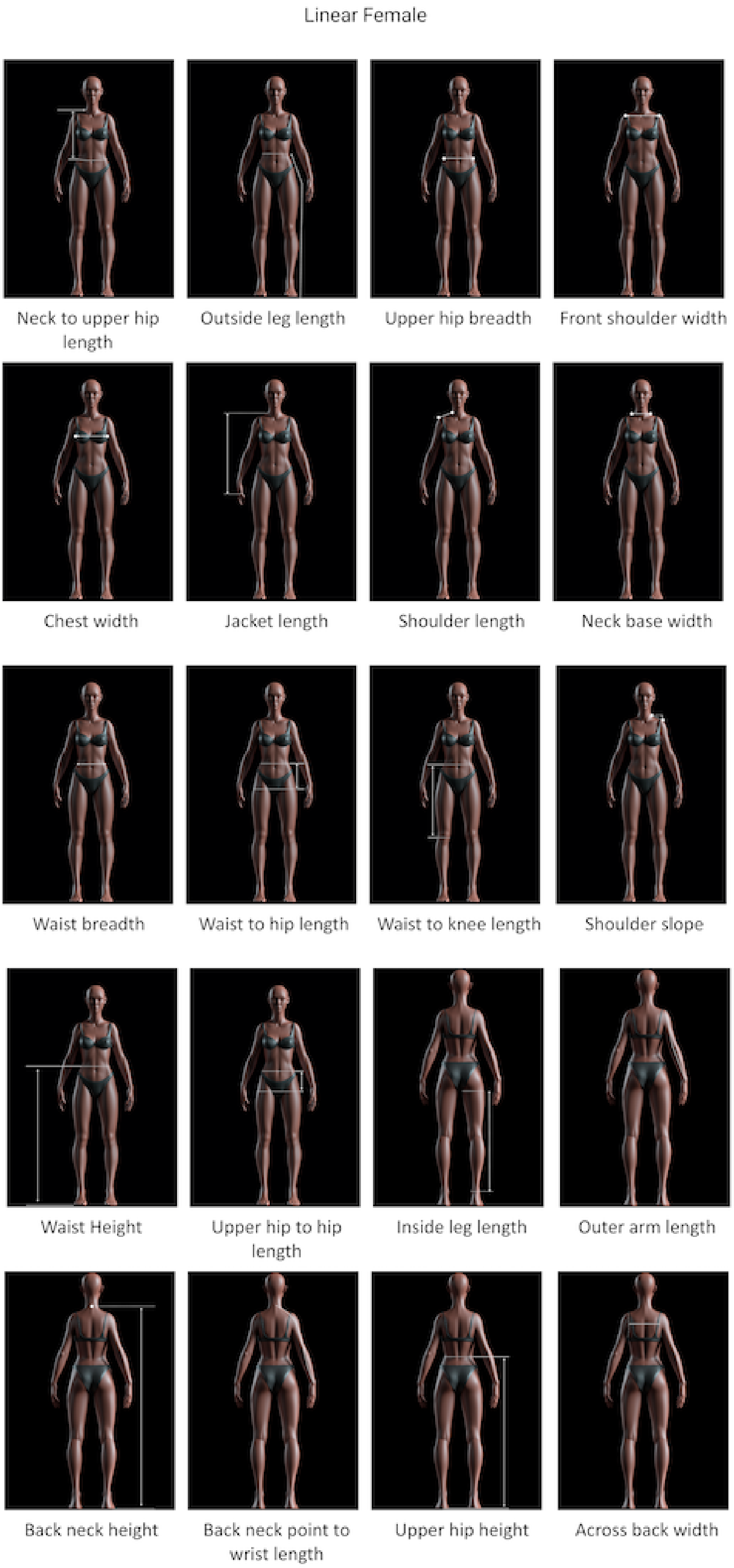


Fig. C.1 Key body Linear by 8559-1 [13] 8559-2 [14] 8559-3 [15]

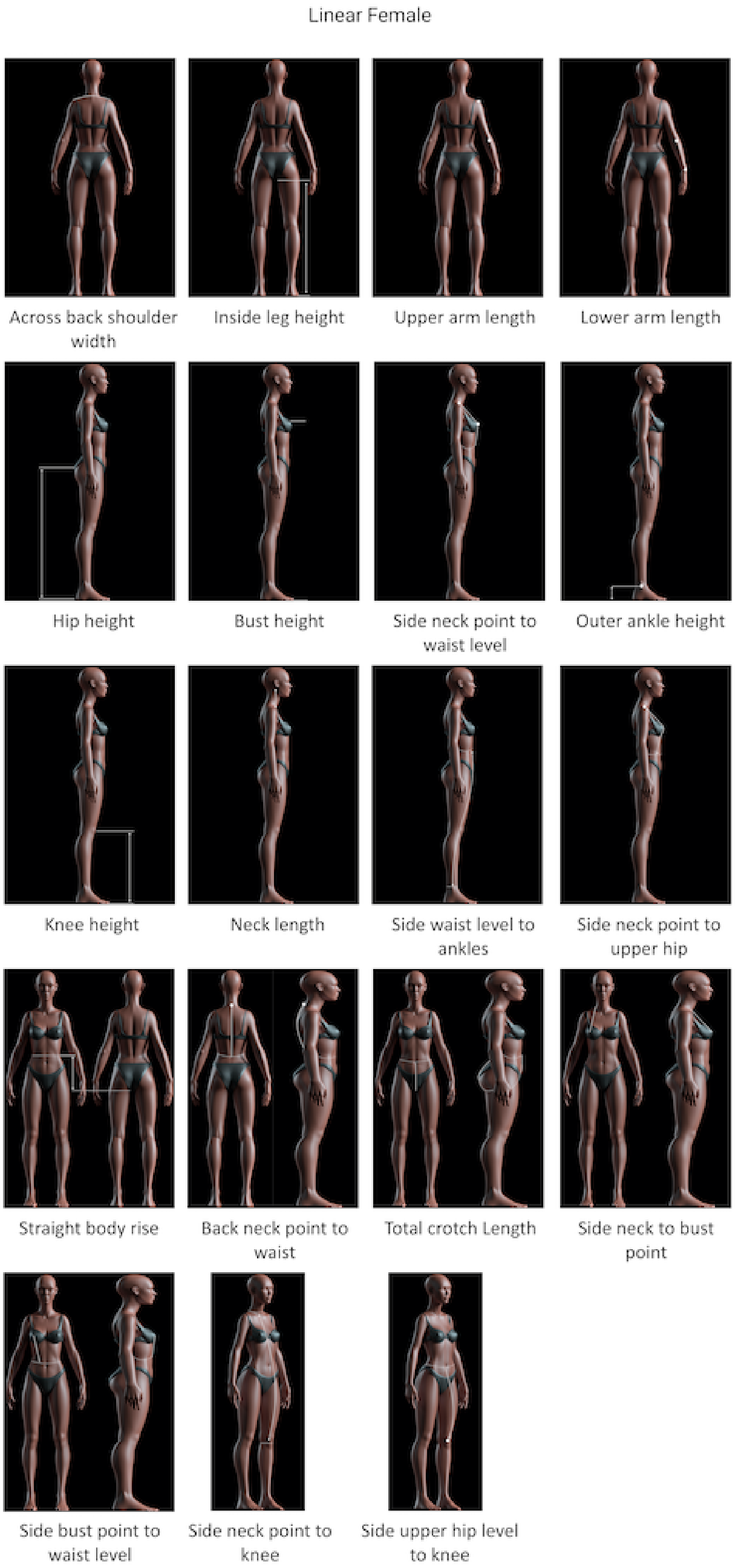


Fig. C.2 Key body Linear by 8559-1 [13] 8559-2 [14] 8559-3 [15]

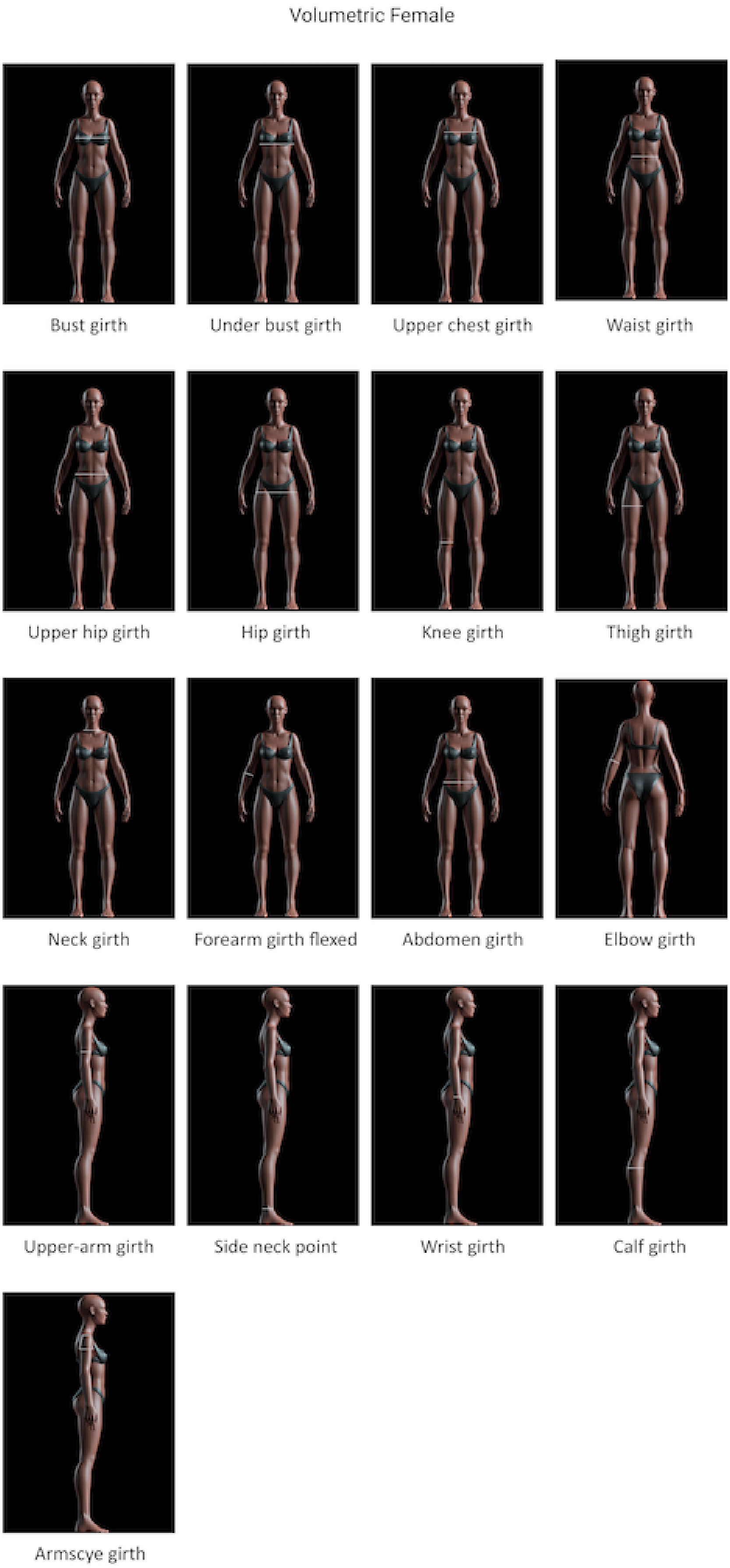


Fig. C.3 Key body Volumetric circumferences by 8559-1 [13] 8559-2 [14] 8559-3 [15]

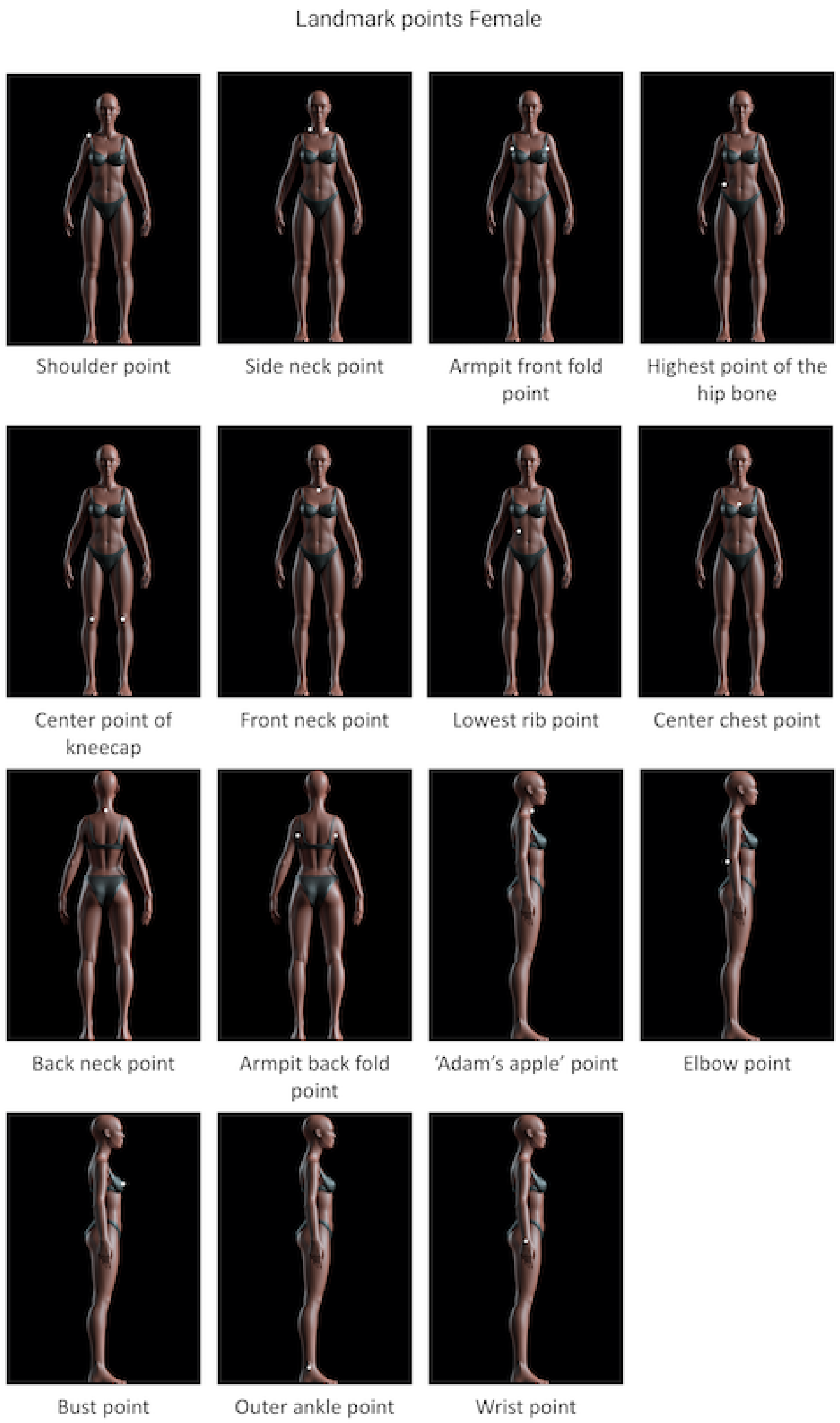


Fig. C.4 Key body Landmark points by 8559-1 [13] 8559-2 [14] 8559-3 [15]

Appendix D

Chapter-6*

D.1 Removal of Small Image Regions, Image Smoothing and Blurring

In the following few paragraphs, we are going to talk about the four primary smoothing and blurring options that we had, as well as the one that we ultimately decided was the best option for this project.

D.1.1 Image Blurring

1) Average Blurring

An average filter takes a region of pixels surrounding a central pixel, calculates the average of these pixels, and replaces the central pixel with this average. By averaging the region surrounding a pixel, we smooth it out and replace it with the neighborhood's local value. By focusing on the average, we are able to lower noise and detail levels.

To achieve the average blur, we actually wrap our image with a normal M $\times$ N filter, both M and N are odd integers.

This kernel shifts its position within our input image from left to right and top to bottom for each every pixel (Human body image - front view). The pixel that is located in the middle of the kernel has its value set to be the same as the average of the other pixels that are located nearby. Let's begin by defining an average 3 × 3 kernel that can be used to blur the central pixel with a radius of 3 pixels:

$$K = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{D.1}$$

If you have noticed, we are assigning the same amount of weight to each pixel that is a part of the kernel, therefore each element in the kernel matrix is weighted uniformly. When the size of the kernel rises, we are able to cover a larger part of the image, which results in a more blurred picture overall.

2) Gaussian Blurring

The Gaussian blurring technique is quite comparable to the medium blurring technique; however, rather than utilising a basic mean, we employ a weighted mean. This means that neighbouring pixels that are closer to the centre pixel contribute more "weight" to the mean. It is a very common effect in graphics software, and its primary purpose is to lessen the appearance of image noise while maintaining the level of detail.

Compared to the average procedure outlined in the previous section, our image is less blurred and seems more "naturally." Moreover, compared to average smoothing, we can preserve the majority of image edges using this weighting. When doing Gaussian blurring, the kernel size is M × N, just as it is when performing average blurring; however, both M and N are odd integers. Because we will be weighting pixels according to their distance from the central pixels, we will require an equation in order to generate our kernel.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{D.2}$$

It is the sum of two Gaussian functions, one in each dimension, in two dimensions::

$$G(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{D.3}$$

Where x and y represent the distances away from the origin in the horizontal and vertical centres of the kernel, respectively, and where X is the standard deviation of the Gaussian kernel.

Gaussian blurring is commonly used with edge detection that are are sensitive to noise; the 2-D Laplacian which is highly sensitive to noisy environment (Please refer to subsubsection 6.1.1.4 for more information). Also lower-end digital cameras, including many smartphone cameras, commonly use Gaussian smoothing to cover up image noise caused by higher ISO light sensitivities.

Since we need to detect the edges of the human body in the image, we have utilised this blurring technique as part of our image smoothing.

The computation of the weighted mean, rather than allowing all pixels in the kernel neighbourhood to have the same weight, is what gives the Gaussian Blur its more natural appearance when compared to the averaging approach. This is the reason why the Gaussian Blur is preferred. When the value is higher, there is a greater amount of blur, and vice versa.

3) Median Filtering

The operation known as "Median blur" is very much like the other ways of averaging. The median of all the pixels in the kernel area is substituted for the element that was previously located in the centre of the image. During this operation, the noise is also removed while the edges are being processed. The median blur technique is often considered to be the most efficient way to get rid of the salt-and-pepper noise.

When applying for the median blur, the first thing we do is define the size of the kernel. Then, in the same way that we did with the approach of averaging and blurring, we consider all of the pixels in the surrounding area of size k × K, where K is an odd integer, because there are no true middle pixels otherwise.

It is imperative that the kernel size for the median blurring be square, in contrast to the average and Gaussian blurring, in which the kernel size is rectangular. Unlike to the average blurring, which replaces the central pixels with the neighborhood's average value, this method will instead replace the central pixels with the median value.

Because each central pixel in median blurring is always replaced with the pixel intensity that already exists in the image, this technique is more effective than others at removing the salt-and-pepper type of noise that can be found in an image. Since the median is unaffected by outliers more than other statistical methods, such as the average, the salt-and-pepper noise will have less of an impact on the median.

4) Bilateral Filtering

A bilateral blurring filter is one option we have that will both reduce noise and preserve edge. The effect of bilateral blurring can be achieved by using two Gaussian distributions in the calculation. The first Gaussian consists of pixels in the image's (*x, y*)-coordinate space that are located in close distnce into one and another. The second Gaussian transform modifies the pixel by substituting the intensity value of each pixel with a weighted average of the intensity values of neighbouring pixels.

Intuitively, this is logical. If the pixels in a (small) neighbourhood all have the same value for that pixel, then they most likely represent the same object. However, if two adjacent pixels in a neighbourhood have values that are different from one another, we are able to analyse the edge or boundary of an item, and we are able to preserve this edge.

This technique can preserve the edges of an image while also minimising the amount of noise in the picture. The fact that it is so much slower than the other approaches that were covered in this section is by far the most serious disadvantage associated with using this strategy.

D.1.2 Classical Computer Vision Image Segmentation

D.1.2.1 Detecting Edges Using Masks

D.1.2.1.1 Linear Edge Operators In linear edge detection, filtering and enhancement are both accomplished through the use of linear convolutions. These two steps can be combined into one using the derivative of the smoothing kernel in the single stage of the convolution process. There are three well-known methods for edge detection, and they are as follows:

- Convolution with Edge Templates
- ZeroCrossings of Laplacian-of-Gaussian Convolution
- ZeroCrossings of Directional Derivatives of Smoothed Image

D.1.2.1.2 Morphological Edge Detection Morphological transformations are some simple operations that are performed on an image based on its shape. Binary images are the typical targets of this operation. Only needs two inputs: the first is the original image, and the second is what's known as the "*structuring element*" or the "*kernel*," which determines the type of operation to be performed. Both "**Erosion**" and "**Dilation**" are well-known morphological operators; nevertheless, for the purposes of this study, the "Erosion" procedures were utilised.

"**Erosion**" works in the same way that soil erosion does; it slowly dissolves at the edges of the foreground object (by keeping the foreground in white). So what does it accomplish? The kernel can be seen moving through the image (as in 2D convolution). Only if every pixel that falls under the kernel of the original image has a value of 1, would a pixel in the original image, which may be either 1 or 0, be regarded to have a value of 1. (which is made to 0 as black).

After then, depending on the size of the kernel, all of the pixels that are close to the upper human body boundary will be removed from consideration. As a result, the foreground object's thickness or size, as well as the overall amount of white region in the image, both decrease.

Figure D.1 displays an experimental comparison of the LoG and the morphological second derivative in detecting edges by Maragos [107].

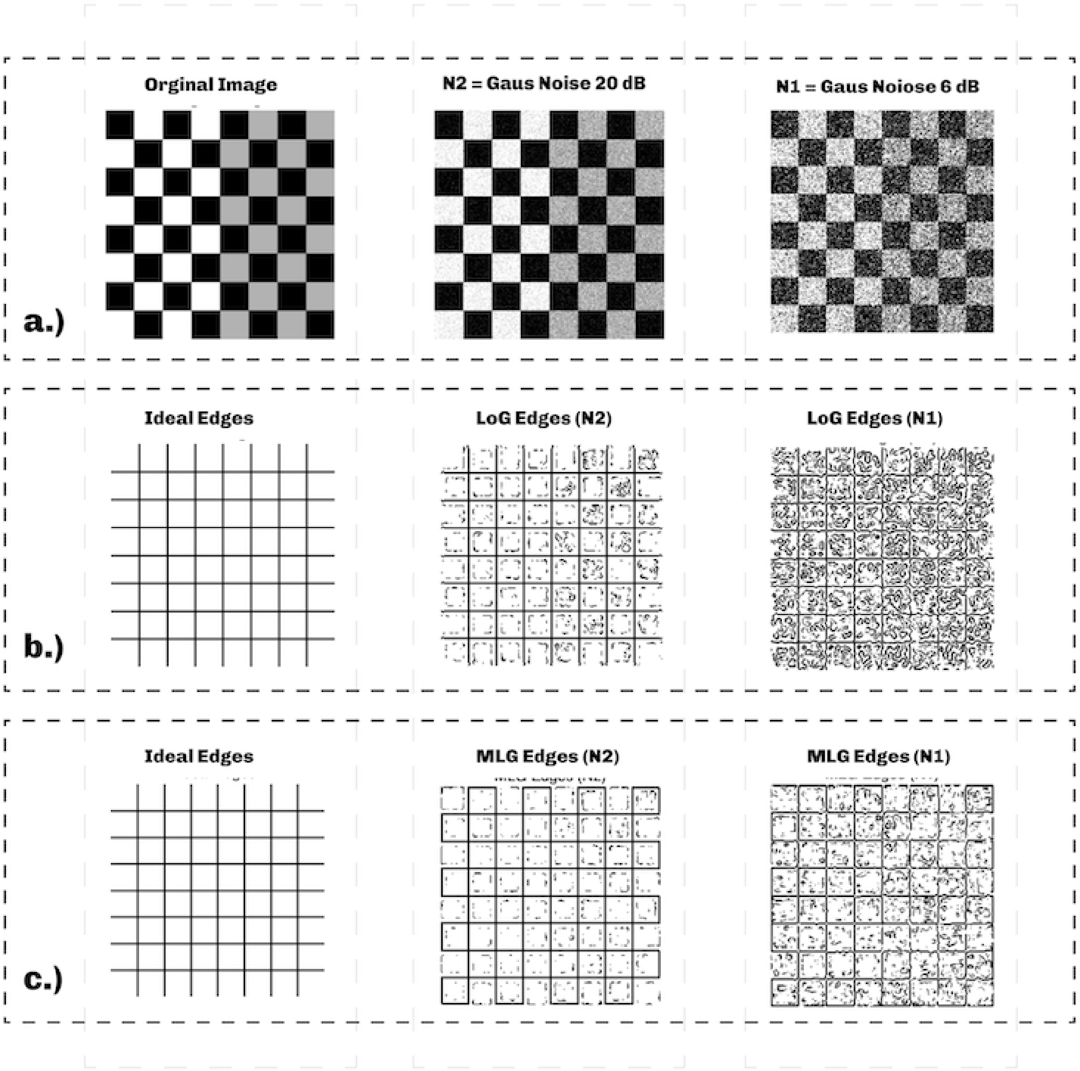


Fig. D.1 (a) Display the original image along with two versions affected by additive Gaussian noise with a signal-to-noise ratio (SNR) of 20 and 6 dB. (b) Show ideal edges as well as edges derived from zero-crossings of the Laplacian-of-Gaussian for each of the two noisy images. (c) Ideal edges, as well as edges derived from zero-crossings of the 2D morphological second derivative (nonlinear Laplacian) of the two noisy images, after some Gaussian pre-smoothing has been applied [107].

D.1.2.2 Mask operations on matrices

By computing the differences in intensity across local image regions, it is possible to identify image spots with a strong contrast. In most cases, such points originate from the boundary between two distinct items or parts of a scene; nevertheless, in the case of our experiments, it is the edge extractions of the human upper body. Each pixel value in an image has been recalculated by me in accordance with the mask matrix (also known as the kernel). This mask contains values that allow the amount of influence that surrounding pixels (as well as the value of the current pixel) have on the new pixel value to be adjusted. This indicates that with the assistance of this method, we are able to isolate significant characteristics from the edges of an image (e.g., corners, lines, curves). These features are utilised by computer vision algorithms operating at a higher level. In essence, what we want to do is apply the following formula to each and every pixel in the image:

$$I\,(i,\,j) = 5 * I\,(i,j) \quad [\; I\,(i1,\,j) + I\,(i{+}1,\,j) + I\,(i,\,j1) + I\,(i,\,j{+}1)\;]$$

$$I\,(i,j) * M, \text{ where } M = \begin{pmatrix} i\backslash j & -1 & 0 & +1 \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 5 & -1 \\ +1 & 0 & -1 & 0 \end{pmatrix}$$

The first notation involves the use of a formula, and the second notation involves the use of a mask to create a quick summary of the first. We made use of the mask by positioning the centre of the mask matrix (in the upper case, this was indicated by the zero-zero index) over the pixel whose value we wanted to calculate, and then we counted the value of that pixel multiplied by the values of the overlapping matrix.

D.1.2.3 Thresholding

Thresholding is a segmentation technique used to distinguish the foreground and background of an image. All pixels over the threshold are assigned to one region, while all pixels below the threshold are assigned to another. This is a sort of supervised machine learning algorithm since the threshold value is determined by the user. Here, the values of pixels are assigned based on the thresholds you define. In computer vision, grayscale pictures are thresholded via thresholding. In a variety of image processing procedures, thresholding is an important intermediary function. Using these tools, we can mask of undesired details such as dark or light areas and outlines. Thresholding is an effective method for extracting important visual elements that are obscured by noise in an image.

- The foreground of the image is represented by pixel values that are set to be between 225 and 255, with white being the highest value (i.e, Human upper body).
- Collecting every pixel in the grey image that is greater than 225 and assigning it to 0 (black), which is the colour that represents the background of the image.

As part of our research, we have researched a variety of thresholding methods in order to determine the best alternative for our suggested programme. The following is the number of thresholding techniques that we have examined:

1. Binary thresholding: is the most straightforward and prevalent type of thresholding. It is utilised to segment images using a single threshold value. It sets all pixels in an image to either 0 (black) or 255 (white), based on whether their intensity value is less than or greater than the specified threshold value.
2. Adaptive Thresholding: is one of the most frequently employed thresholding techniques. Image segmentation based on multiple threshold values can be accomplished with its help. When dealing with images that have non-uniform lighting, this thresholding method is utilised so that the problem can be solved. In order for it to work, it must first determine the mean of the area surrounding a pixel, and then choose a threshold value based on the mean.
3. Otsu's Thresholding: is an automatic thresholding method that maximises the inter-class variance. This method is especially useful when the image's background and foreground have varying intensities. It functions by calculating the image's histogram and then identifying the optimal threshold value that maximises the inter-class variance.

4. Triangle thresholding: is a form of image thresholding method that is used to turn a grayscale or colour image into a binary image. It is based on the idea of determining the ideal threshold value by maximising the resemblance of the generated image histogram to a triangle. The triangle thresholding method is implemented by setting the threshold value to the histogram's peak, which is placed in the middle of the range of grey level values. The grey level values below the threshold are set to zero, and those above it are set to one. This method can be used to recognise and segment objects in images and has been proven to be useful in many applications, such as medical imaging and document analysis.

5. Mask thresholding: is a sort of image segmentation that divides an image into two sections, a foreground and a background, using a specified threshold value. It operates by comparing each pixel in an image to a specified threshold value. If the pixel is over the threshold, it is deemed to be part of the foreground; otherwise, it is deemed to be part of the background. In image processing applications such as object detection, image segmentation, and image analysis, mask thresholding is an useful technique.

Based on our findings, we have chosen to employ mask thresholding because: 1. it is relatively straightforward and requires minimal user input. 2. It's also fast, as there is no need to search for the optimal threshold value. 3. it is also gives better results than global thresholding, especially when working with photographs with non-uniform backgrounds. 4. it can also be used to detect things in a more complex background.

Thresholding by itself is not a particularly effective approach, but when combined with other strategies, as we will see later, it may be quite valuable. Figure D.2 shows every phase of image corrections in our software, from affine and metrics corrections to skin detection.

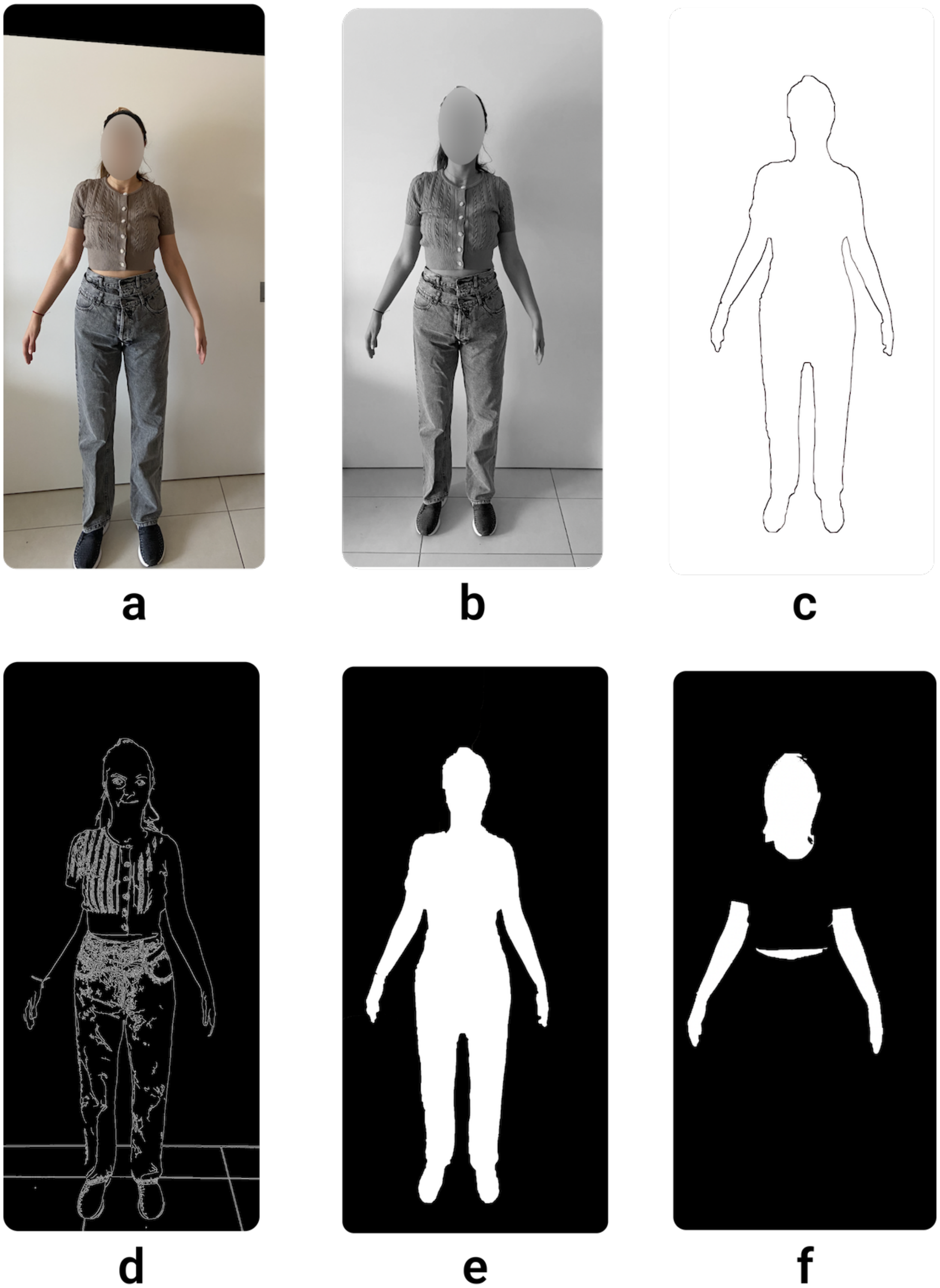


Fig. D.2 demonstrated every step of the image adjustments, beginning with (a) affine modifications, (b) grayscale conversion, (c) edge detection, (d) canny edge identification, (e) threshold application, and (f) skin detection.

D.1.3 DeeplabV3

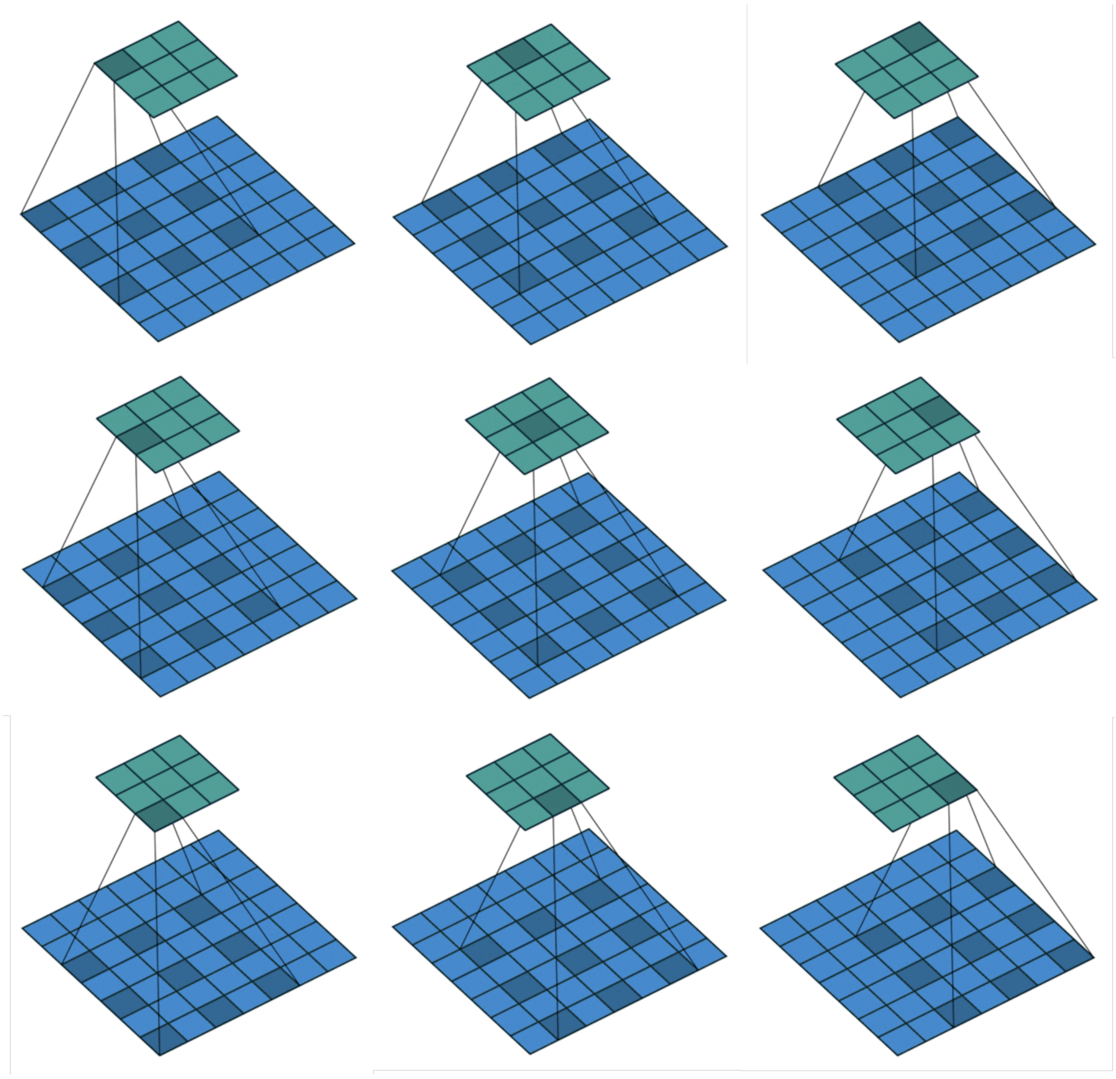


Fig. D.3 An atruous 2D convolution, with a 3 kernel, a 2 atrous rate, and no padding - Paul-Louis [124]

Atrous Separable Convolution

Lets go deeper with atrous convolution using Multi-Grid.

- (a) Without Atrous Convolution: Standard convolution and integration are performed, which increases the output stride, for instance, the output feature map becomes smaller as it deepens. However, because location/spatial data is losing at deeper layers, therefore, consecutive striding is harmful for semantic segmentation.

- (b) With Atrous Convolution: With a larger field-of-view without increasing the number of parameters or the calculation, we can keep the stride constant. Finally, we can have a larger output feature map which is good for semantic segmentation.

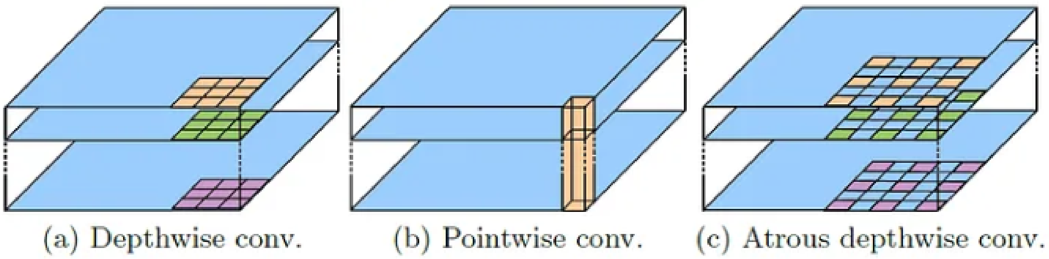


Fig. D.4 Depthwise Separable Convolution Using Atrous Convolution [153]

Figure D.4 is introduces in MobileNetV1 where;

- (a) and (b), Depthwise Separable Convolution: It transforms a conventional convolution into a depthwise convolution followed by a point-wise convolution (i.e., an 1x1 convolution), which substantially decreases processing complexity.
- (c) Atrocious Depthwise Convolution: The depthwise convolution supports atrous convolution. And it is discovered that it greatly decreases the computational complexity of the suggested model while preserving (or improving) performance.
- Combining with point-wise convolution, it is Atrous Separable Convolution.

D.1.3.1 DeepLabv3 as Encoder and Decoder

The Encoder-Decoder architecture is a deep learning architecture used in DeepLabV3 and other computer vision tasks. It consists of two parts: the encoder and the decoder.

The encoder is responsible for extracting features from the input image, while the decoder is responsible for mapping the features to a higher level representation.

The encoder part of the architecture is typically a convolutional neural network (CNN). The network takes the input image and passes it through a series of convolutional layers, each of which extracts features from the image, and produces a feature map. This feature map is then passed to the decoder part of the architecture, which takes the feature map and produces the desired output.

The decoder part of the architecture typically consists of a series of upsampling layers, which increase the size of the feature map, and a set of convolutional layers which map the feature map to the final output. In DeepLabV3, the decoder part of the architecture also includes a “context module”, which helps the model to better distinguish between objects in the image.

Usually, the spatial resolution of the final feature maps for image classification is 32 times lower than the resolution of the original image, hence the output step is 32. It is too tiny to segment semantically. By adopting output stride = 16 (or 8) and applying the Athros complexity to the stride from the last block’s one (or two), a denser feature can be extracted. In addition, DeepLabv3 improves the Atrous Spatial Pyramid Pooling module, which evaluates convolution qualities at various scales by applying atrous convolution at varying rates on image-level features.

The encoder features are first bilinearly upsampled with a coefficient of 4 and then linked to the matching low-level feature. There is a 1 x 1 complexity to low-level features prior to accession to minimise the number of channels since the matching low-level feature typically has a large number of channels (i.e. 512 or 1024) that may exceed the significance of the rich encode features. To fix some characteristics, we use 3 x 3 convolutions after concatenation and then do a basic bilinear upsampling by a factor of 4.

D.2 Results

D.2.1 Statistical Analysis and Comparison of Chest Circumference Measurements: Plain Background vs. Cluttered Background."

A t-test was conducted to determine whether there is a statistically significant difference between the plain-background and cluttered-background measurements of chest circumferences. The hypothesis was that the mean chest circumference in the cluttered background condition would be greater than in the plain background condition.

The plain-background measurements had a mean of 0.48 (SE = 0.04), while the cluttered-background measurements had a mean of 0.57 (SE = 0.04). The variances of the two samples were 0.12 and 0.15 for the plain and cluttered backgrounds, respectively. The sample sizes for both conditions were 78.

The t-statistic for the two-sample t-test assuming unequal variances was calculated to be -1.51. The degrees of freedom (df) for the test were 153. For a one-tailed test, the p-value (P(T<=t)) was found to be 0.07, which is greater than the significance level of 0.05. The critical value for a one-tailed test at a 0.05 significance level (alpha) with 153 degrees of freedom is 1.65.

For a two-tailed test, the p-value (P(T<=t)) was calculated as 0.13, also greater than the significance level of 0.05. The critical value for a two-tailed test at a 0.05 significance level with 153 degrees of freedom is 1.98.

Based on the t-test results, we fail to reject the null hypothesis. There is no statistically significant difference between the chest circumferences measured in the plain and cluttered backgrounds. The two data sets are likely similar.

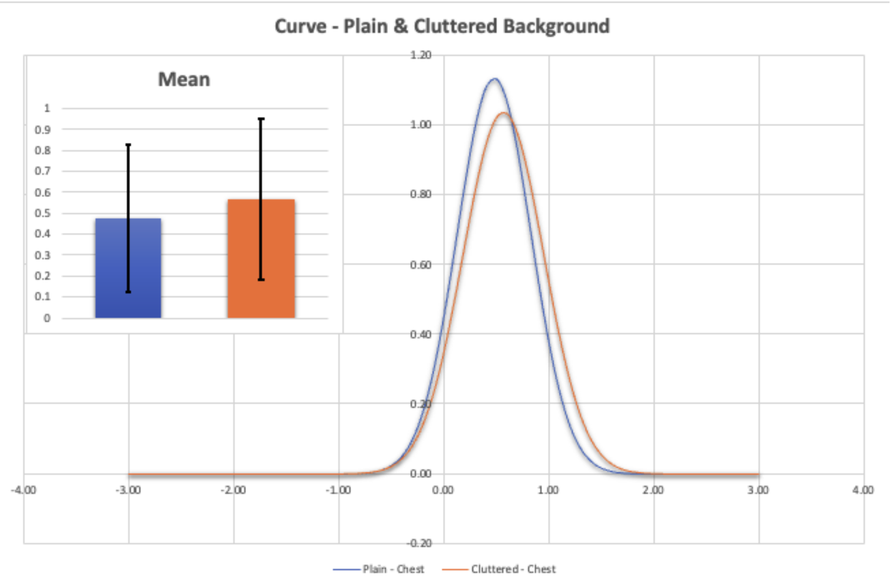


Fig. D.5 : demonstrate the each background's stress level on the bell curve.

For the plain background measurements, the mean chest circumference was M = 0.48 with a standard error of SE = 0.04. The median was 0.48, and the mode was not identified. The standard deviation and sample variance were SD = 0.35 and variance = 0.12, respectively. The data had a positive skewness of 0.79 and a kurtosis of 1.03. The range of measurements was 1.73, with the minimum value being 0.01 and the maximum value being 1.73. The sum of measurements was 37.17, and the sample size was n = 78.

For the cluttered background measurements, the mean chest circumference was M = 0.57 with a standard error of SE = 0.04. The median was 0.54, and there was no mode identified. The standard deviation and sample variance were SD = 0.38 and variance = 0.15, respectively. The data had a positive skewness of 0.85 and a kurtosis of 0.64. The range of measurements was 1.71, with the minimum value being 0.03 and the maximum value being 1.74. The sum of measurements was 44.13, and the sample size was n = 78.

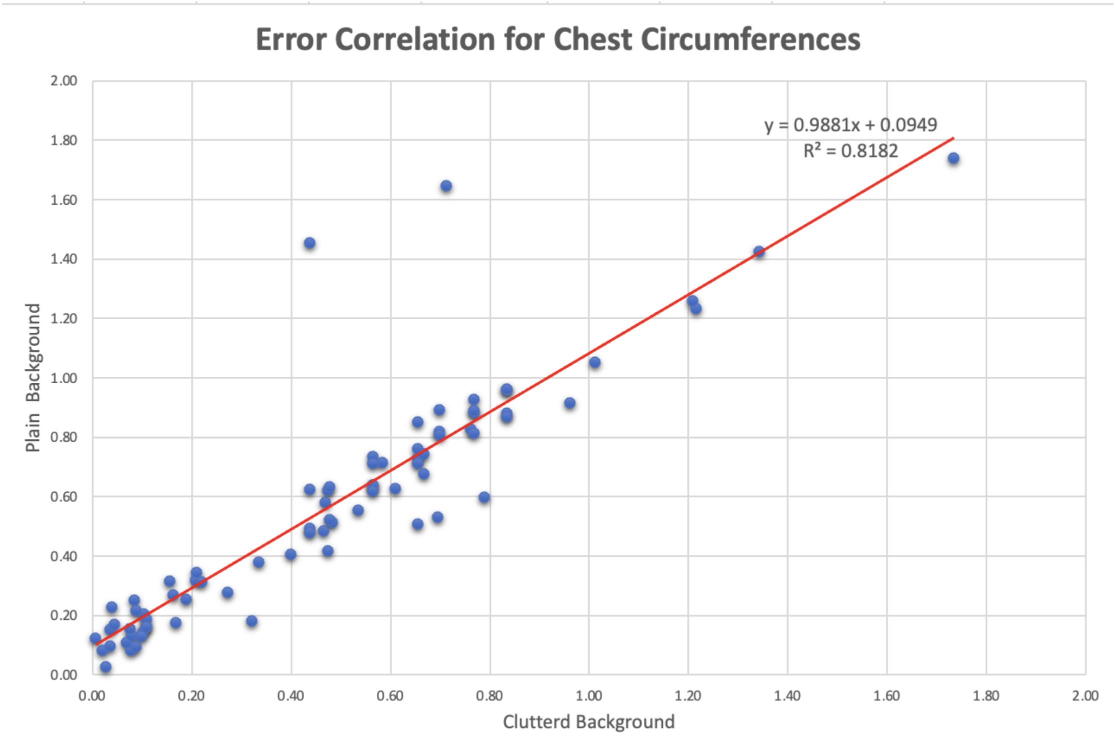


Fig. D.6 : displays the error correlation for the chest circumferences measurements

The error correlation between the plain-background and cluttered-background measurements was found to be 0.90. This indicates a strong positive correlation between the errors in the two sets of measurements.

Additionally, when examining the chest circumference error measurements, it was found that for the plain background measurements, 53.85% of the data fell within the range of 0 to 0.5, 39.74% fell within the range of 0.5 to 1, 5.13% fell within the range of 1 to 1.5, and 1.28% fell within the range of 1.5 to 2. Similarly, for the cluttered background measurements, the percentages were as follows: 53.85% in the range of 0 to 0.5, 39.74% in the range of 0.5 to 1, 5.13% in the range of 1 to 1.5, and 1.28% in the range of 1.5 to 2. These results provide insights into the distribution of errors in both backgrounds.

Overall, the results of the chest circumference error measurements conducted on both the plain and cluttered backgrounds indicate that the software is accurate and reliable, as the majority of measurements had an error difference of less than 1 cm, and only a small minority had an error difference of greater than 1.5 cm. However, it is important to note that the accuracy of the software was slightly lower on the cluttered background than on the plain background, with a greater percentage of measurements having an error difference greater than 1.5 cm on the cluttered background.

The accuracy of the software on both plain and cluttered backgrounds is within acceptable parameters, however it is important to note that the software is slightly more accurate on the plain background than on the cluttered background. In order to ensure consistent accuracy, it is recommended that the software be used on plain backgrounds whenever possible. Additionally, it is recommended that users of the software take extra care when measuring on cluttered backgrounds, as the accuracy of the software is slightly lower in that scenario.

D.2.2 Statistical Analysis and Comparison of Bust Circumference Measurements: Plain Background vs. Cluttered Background."

A t-test was conducted to determine whether there is a statistically significant difference between the plain-background and cluttered-background measurements of bust circumferences. The hypothesis was that the mean bust circumference in the cluttered background condition would be greater than in the plain background condition.

The plain-background measurements had a mean of 0.50 (SE = 0.03), while the cluttered-background measurements had a mean of 0.56 (SE = 0.03). The variances of the two samples were 0.09 and 0.09 for the plain and cluttered backgrounds, respectively. The sample sizes for both conditions were 78.

The t-statistic for the two-sample t-test assuming unequal variances was calculated to be -1.19. The degrees of freedom (df) for the test were 154. For a one-tailed test, the p-value (P(T<=t)) was found to be 0.12, which is greater than the significance level of 0.05. The critical value for a one-tailed test at a 0.05 significance level (alpha) with 154 degrees of freedom is 1.65. For a two-tailed test, the p-value (P(T<=t)) was calculated as 0.24, also greater than the significance level of 0.05. The critical value for a two-tailed test at a 0.05 significance level with 154 degrees of freedom is 1.98.

Based on the t-test results, we fail to reject the null hypothesis. There is no statistically significant difference between the bust circumferences measured in the plain and cluttered backgrounds. The two data sets are likely similar.

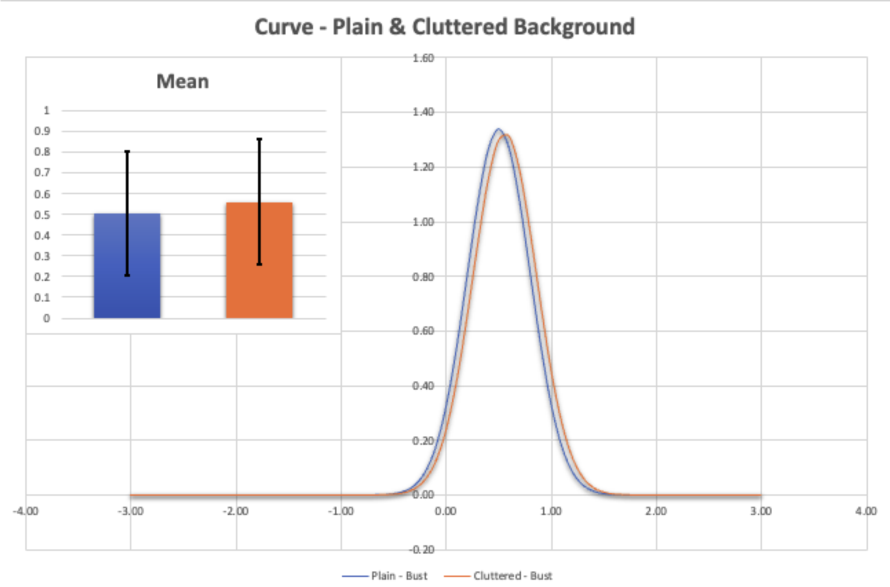


Fig. D.7 : demonstrate the each background’s stress level on the bell curve.

For the plain background measurements, the mean bust circumference was M = 0.50 with a standard error of SE = 0.03. The median was 0.56, and the mode was not identified. The standard deviation and sample variance were SD = 0.30 and variance = 0.09, respectively. The data had a negative skewness of -0.01 and a kurtosis of -0.52. The range of measurements was 1.25, with the minimum value being 0.00 and the maximum value being 1.25. The sum of measurements was 39.28, and the sample size was n = 78.

For the cluttered background measurements, the mean bust circumference was M = 0.56 with a standard error of SE = 0.03. The median was 0.59, and there was no mode identified. The standard deviation and sample variance were SD = 0.30 and variance = 0.09, respectively. The data had a positive skewness of 0.07 and a kurtosis of -0.67. The range of measurements was 1.24, with the minimum value being 0.02 and the maximum value being 1.26. The sum of measurements was 43.72, and the sample size was n = 78.

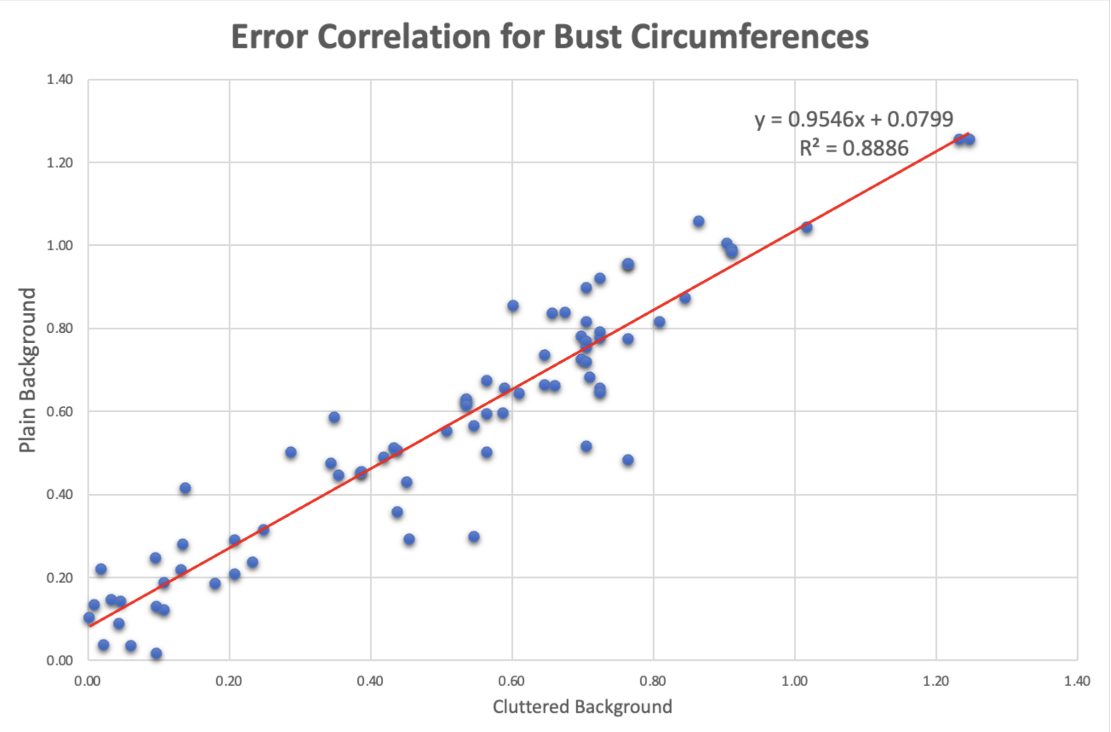


Fig. D.8 : displays the error correlation for the bust circumferences measurements

The error correlation between the plain-background and cluttered-background measurements was found to be 0.94. This indicates a strong positive correlation between the errors in the two sets of measurements.

Additionally, when examining the bust circumference error measurements (as seen in Figure 6.19), it was found that for the plain background measurements, 42.31% of the data fell within the range of 0 to 0.5, 53.85% fell within the range of 0.5 to 1, and 3.85% fell within the range of 1 to 1.5. On the other hand, for the cluttered background measurements, the percentages were as follows: 39.74% in the range of 0 to 0.5, 53.85% in the range of 0.5 to 1, and 6.41% in the range of 1 to 1.5. These findings provide insights into the distribution of errors in both backgrounds.

Overall, these results indicate that there is a slightly higher error rate when taking measurements in a cluttered background than in a plain background.

D.2.3 Statistical Analysis and Comparison of Waist Circumference Measurements: Plain Background vs. Cluttered Background."

A t-test was conducted to determine whether there is a statistically significant difference between the plain-background and cluttered-background measurements of waist circumferences. The hypothesis was that the mean waist circumference in the cluttered background condition would be greater than in the plain background condition.

The plain-background measurements had a mean of 0.56 (SE = 0.04), while the cluttered-background measurements had a mean of 0.65 (SE = 0.04). The variances of the two samples were 0.12 and 0.15 for the plain and cluttered backgrounds, respectively. The sample sizes for both conditions were 78.

The t-statistic for the two-sample t-test assuming unequal variances was calculated to be -1.62. The degrees of freedom (df) for the test were 153. For a one-tailed test, the p-value (P(T<=t)) was found to be 0.05, which is greater than the significance level of 0.05. The critical value for a one-tailed test at a 0.05 significance level (alpha) with 153 degrees of freedom is 1.65. For a two-tailed test, the p-value (P(T<=t)) was calculated as 0.11, also greater than the significance level of 0.05. The critical value for a two-tailed test at a 0.05 significance level with 153 degrees of freedom is 1.98.

Based on the t-test results, we fail to reject the null hypothesis. There is no statistically significant difference between the waist circumferences measured in the plain and cluttered backgrounds. The two data sets are likely similar.

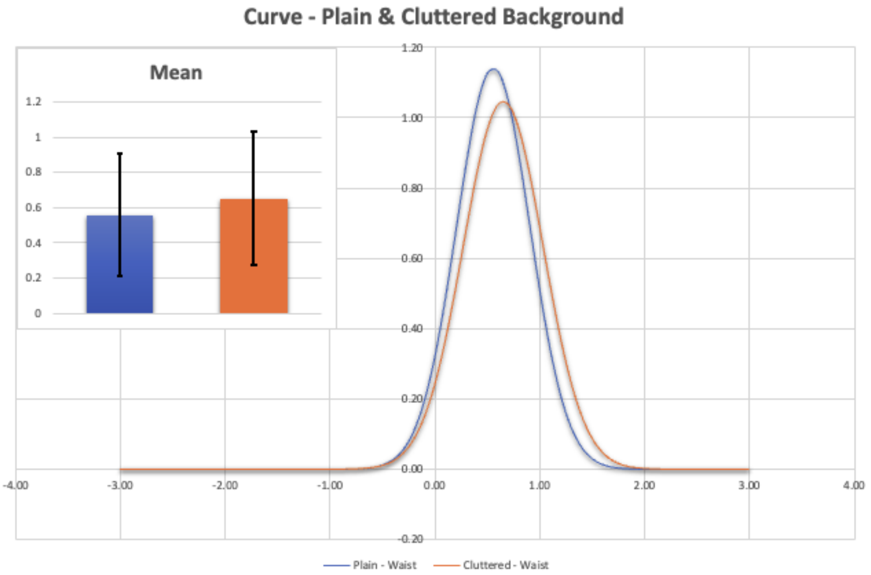


Fig. D.9 : demonstrate the each background's stress level on the bell curve.

For the plain background measurements, the mean waist circumference was M = 0.56 with a standard error of SE = 0.04. The median was 0.56, and the mode was not identified. The standard deviation and sample variance were SD = 0.35 and variance = 0.12, respectively. The data had a positive skewness of 0.61 and a kurtosis of 1.05. The range of measurements was 1.74, with the minimum value being 0.02 and the maximum value being 1.75. The sum of measurements was 43.53, and the sample size was n = 78.

For the cluttered background measurements, the mean waist circumference was M = 0.65 with a standard error of SE = 0.04. The median was 0.66, and there was no mode identified. The standard deviation and sample variance were SD = 0.38 and variance =

0.15, respectively. The data had a positive skewness of 0.67 and a kurtosis of 0.85. The range of measurements was 1.86, with the minimum value being 0.06 and the maximum value being 1.92. The sum of measurements was 50.92, and the sample size was n = 78.

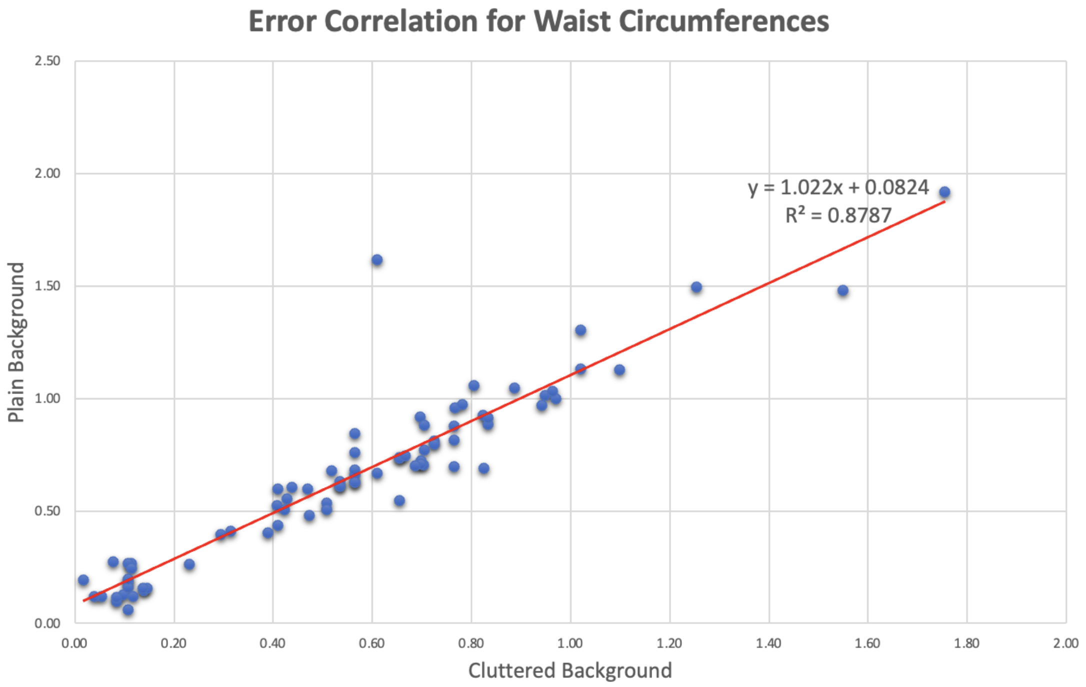


Fig. D.10 : displays the error correlation for the waist circumferences measurements

The error correlation between the plain-background and cluttered-background measurements was found to be 0.94. This indicates a strong positive correlation between the errors in the two sets of measurements.

Additionally, when examining the waist circumference error measurements (as seen in Figure 6.19), it was found that for the plain background measurements, 38.46% of the data fell within the range of 0 to 0.5, 53.85% fell within the range of 0.5 to 1, 5.13% fell within the range of 1 to 1.5, and 2.56% fell within the range of 1.5 to 2. On the other hand, for the cluttered background measurements, the percentages were as follows: 30.77% in the range of 0 to 0.5, 53.85% in the range of 0.5 to 1, 11.54% in the range of 1 to 1.5, and 3.85% in the range of 1.5 to 2. These findings provide insights into the distribution of errors in both backgrounds.

Overall, it appears that taking measurements in a cluttered background resulted in slightly lower accuracy than taking measurements in a plain background. This could be due to the fact that the participants may have been wearing loose clothing around their waists, making it more difficult to accurately measure their waist circumference.

To further improve the accuracy of waist circumference measurements, it may be beneficial to make sure that the participant is wearing tighter fitting clothing so that there is less room for error. It may also be beneficial to move the measurements to a controlled environment to reduce the amount of clutter in the background.

In conclusion, it appears that taking waist circumference measurements in a cluttered background results in slightly lower accuracy than measurements taken in a plain background.

D.2.4 Statistical Analysis and Comparison of Hip Circumference Measurements: Plain Background vs. Cluttered Background."

A t-test was conducted to determine whether there is a statistically significant difference between the plain-background and cluttered-background measurements of hip circumferences. The hypothesis was that the mean hip circumference in the cluttered background condition would be greater than in the plain background condition.

The plain-background measurements had a mean of 0.52 (SE = 0.05), while the cluttered-background measurements had a mean of 0.59 (SE = 0.05). The variances of the two samples were 0.18 and 0.18 for the plain and cluttered backgrounds, respectively. The sample sizes for both conditions were 78.

The t-statistic for the two-sample t-test assuming unequal variances was calculated to be -1.00. The degrees of freedom (df) for the test were 154. For a one-tailed test, the p-value (P(T<=t)) was found to be 0.16, which is greater than the significance level of 0.05. The critical value for a one-tailed test at a 0.05 significance level (alpha) with 154 degrees of freedom is 1.65. For a two-tailed test, the p-value (P(T<=t)) was calculated as 0.32, also greater than the significance level of 0.05. The critical value for a two-tailed test at a 0.05 significance level with 154 degrees of freedom is 1.98.

Based on the t-test results, we fail to reject the null hypothesis. There is no statistically significant difference between the hip circumferences measured in the plain and cluttered backgrounds. The two data sets are likely similar.

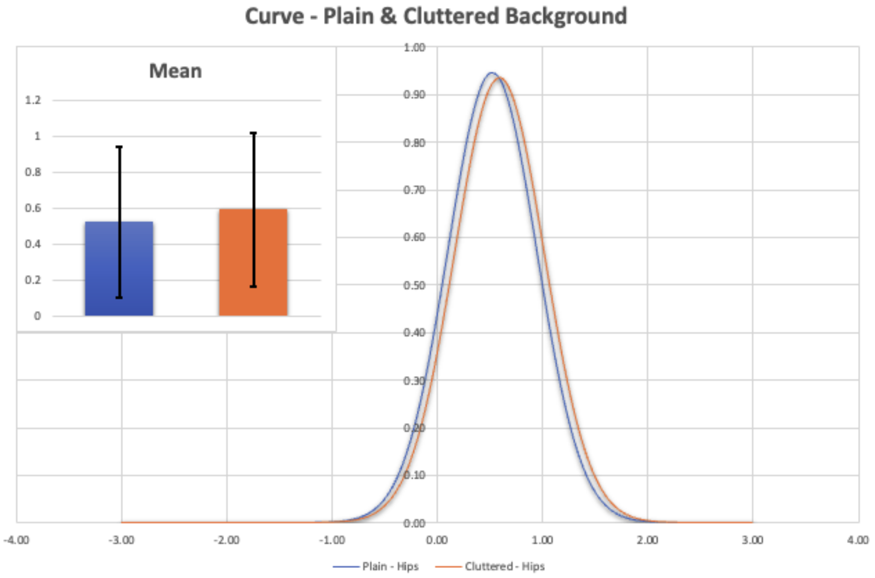


Fig. D.11 : demonstrate the each background's stress level on the bell curve.

For the plain background measurements, the mean hip circumference was M = 0.52 with a standard error of SE = 0.05. The median was 0.53, and the mode was not identified. The standard deviation and sample variance were SD = 0.42 and variance = 0.18, respectively. The data had a positive skewness of 2.63 and a kurtosis of 14.80. The range of measurements was 3.01, with the minimum value being 0.02 and the maximum value being 3.02. The sum of measurements was 40.85, and the sample size was n = 78.

For the cluttered background measurements, the mean hip circumference was M = 0.59 with a standard error of SE = 0.05. The median was 0.58, and there was no mode identified. The standard deviation and sample variance were SD = 0.43 and variance = 0.18, respectively. The data had a positive skewness of 2.52 and a kurtosis of 12.97. The range of measurements was 2.99, with the minimum value being 0.03 and the maximum value being 3.03. The sum of measurements was 46.16, and the sample size was n = 78.

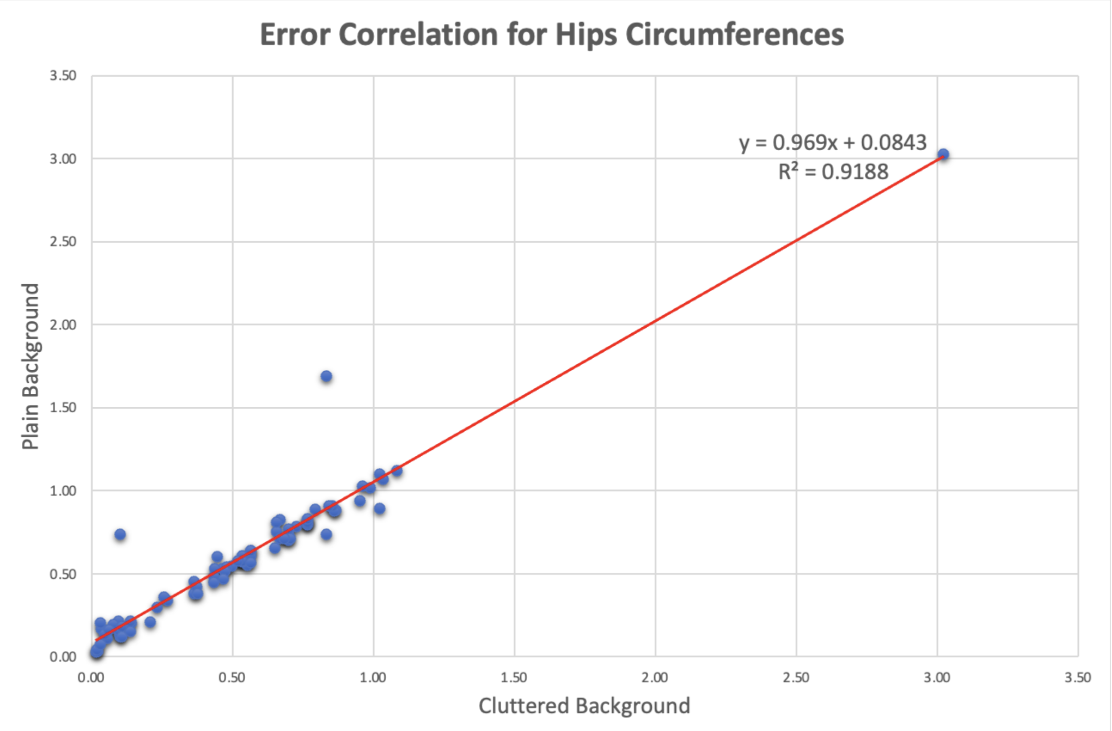


Fig. D.12 : displays the error correlation for the hips circumferences measurements

The error correlation between the plain-background and cluttered-background measurements was found to be 0.96. This indicates a strong positive correlation between the errors in the two sets of measurements.

Additionally, when examining the hip circumference error measurements (as seen in Figure 6.19), it was found that for the plain background measurements, 46.15% of the data fell within the range of 0 to 0.5, 47.44% fell within the range of 0.5 to 1, 5.13%

fell within the range of 1 to 1.5, and 1.28% fell within the range of 3 to 3.5. On the other hand, for the cluttered background measurements, the percentages were slightly different: 38.46% in the range of 0 to 0.5, 52.56% in the range of 0.5 to 1, 6.41% in the range of 1 to 1.5, 1.28% in the range of 1.5 to 2, and 1.28% in the range of 3 to 3.5. These findings provide insights into the distribution of errors in both backgrounds.

Overall, the results of measuring human upper body circumferences using our software in comparison to tape measurements for human body circumferences showed that the accuracy of our software is good for both plain and cluttered backgrounds. The results showed that the majority of the measurements had an error difference of less than 1 cm, and the differences between the accuracy of the software in plain and cluttered backgrounds were minimal. Therefore, our software can be used with confidence for measuring human torso circumferences in both plain and cluttered backgrounds.

Appendix E

Chapter-6.3*

E.1 Camera Calibration and Distance Estimation using OpenCV

E.1.1 Introduction

Calibration of a camera is the process of identifying the parameters of an imaging system, such as the focal length, principal point, and distortion coefficients, to permit the accurate measurement of item sizes in an image. It is crucial for any application that demands exact measurements of objects in images, including robots, surveillance, autonomous driving, and computer vision.

Camera calibration is performed by capturing a sequence of images of an object or scene from different angles and distances. The calibration process can be automated using software, such as OpenCV, or with the help of a specialized hardware device.

To measure the size of a human body in an image, it is necessary to understand the camera's parameters. Without calibration, reliable measurements are impossible. For applications like 3D reconstruction, scene comprehension, and augmented reality, calibration is also crucial.

Using Python, camera calibration can be performed using various libraries such as OpenCV and SciPy. The most common approach is to use the camera calibration library from OpenCV. This library provides functions for calculating the camera's intrinsic parameters, such as the focal length, principal point, and distortion coefficients. It also provides functions for calculating the rotation and translation vectors between two images, which are essential for 3D reconstruction.

In addition to OpenCV, the SciPy library also contains useful functions for camera calibration. Using the camera calibration technique from the SciPy library is the most popular method. This algorithm is based on the Algebraic Distance Minimization (ADM) algorithm, which is used to compute the intrinsic parameters of the camera.

Camera calibration is an essential step for any application that requires precise measurements of objects in images. By using Python, the camera calibration process can be automated and the calibration parameters can be determined with high accuracy. With the use of these parameters, we were able to design a system that can estimate the human body's measurements with precision.

We utilised this method to estimate the distance between the human body and the camera and to convert pixel units to centimetres. When using 3D imaging and computer vision applications, camera calibration is required.

It is now established that to determine the size of an object, such as the upper human body in an image, calibration is necessary. Prior to getting into the details of the methods employed in this section, a survey of several camera calibration techniques is presented. In the course of my research, I examined ten techniques which are described below, that were utilized to calibrate the camera with an image.

1. **Single Point Calibration**: is a method used to calibrate a camera's focus and exposure settings. It allows you to manually adjust a single point in the image to have the desired effect. This allows for more accurate and consistent results

across different lens types and shooting conditions. This method is often used in applications that require precise control over the focus and exposure settings.

2. **Two-Point Calibration**: is a method used to calibrate a camera's focus and exposure settings. It involves adjusting two different points in the image to have the desired effect. This method is often used for applications that require more control over the focus and exposure settings than what is available with single point calibration.

3. **Multi-Point Calibration**: is a method used to calibrate a camera's focus and exposure settings. It involves adjusting multiple points in the image to have the desired effect. This method is often used for applications that require greater control over focus and exposure settings, or when dealing with complex scenes.

4. **Planar Calibration**: is a method used to calibrate a camera's focus and exposure settings when shooting a planar surface, such as a wall or a table. This method is often used for applications that require a greater level of precision when shooting a planar surface.

5. **Radial Distortion Calibration**: is a method used to calibrate a camera's focus and exposure settings when shooting with a wide-angle lens. This method is often used to correct for the distortion caused by wide-angle lenses, which can cause objects in the image to appear distorted or stretched.

6. **3D Calibration**: is a method used to calibrate a camera's focus and exposure settings when shooting in three-dimensional space. This method is often used for applications that require accurate 3D measurements, such as in augmented reality or virtual reality applications.

7. **Autocalibration**: is a method used to automatically calibrate a camera's focus and exposure settings. This method is often used for applications that require frequent calibration of the camera's settings, such as in video surveillance systems.

8. **Zhang's Method**: is a method used to calibrate a camera's focus and exposure settings. This method is based on a mathematical model of the camera's optics and uses a set of images to determine the camera's internal parameter values. This method is often used for applications that require precise control over the focus and exposure settings.

9. **Direct Linear Transformation (DLT)**: is a method used to calibrate a camera's focus and exposure settings. This method is based on a mathematical model of the camera's optics and uses a set of images to determine the camera's internal parameter values. This method is often used for applications that require precise control over the focus and exposure settings.

10. **OpenCV's Camera Calibration Algorithm**: is a method used to calibrate a camera's focus and exposure settings. This method is based on a mathematical model of the camera and its optics, and is often used for applications that require accurate calibration of the camera's settings. It is used in a wide variety of applications, such as video surveillance, medical imaging, and robotics.

Based on the research conducted for this chapter [71, 189, 76], the OpenCV Camera Calibration Algorithm is the most effective method because it combines techniques for feature detection, optimisation, and camera calibration. This algorithm is considered the most accurate and trustworthy method available. OpenCV's Camera Calibration Algorithm is the most precise technique. Zhang's Method is the second most precise technique. Combining camera calibration and optimisation techniques, this method calibrates cameras precisely and precisely. Direct Linear Transformation (DLT), the third technique on the list, is a camera calibration method that estimates parameters using a mathematical methodology. Both Radial Distortion Calibration and Planar Calibration can be used to calibrate cameras, but they are less precise than the preceding three techniques. Two-Point Calibration and Multi-Point Calibration are more reliable, but less precise, than the preceding two methods. Both 3D Calibration and Autocalibration can be used to calibrate cameras, but they are less accurate than the preceding methods. Single Point Calibration, which involves manually calibrating a camera using a single point, is the least accurate of the described procedures. OpenCV's Camera Calibration Algorithm is the most precise method, followed by Zhang's Method, Direct Linear Transformation (DLT), Radial Distortion Calibration, Planar Calibration, Two-Point Calibration, Multi-Point Calibration, 3D Calibration, Autocalibration, and Single Point Calibration. Therefore, we chose to calibrate the images using the Camera Calibration Algorithm of OpenCV.

E.1.2 Camera Calibration Parameters

As previously stated, the primary objective of camera calibration is to identify the focal length, principal point, and distortion coefficients in order to provide accurate measurements from objects inside an image. Before diving into our methodology, we 'll review the meaning of each of these terms in the following paragraphs

Focal length is the distance between the optical centre of a lens and the point where parallel light rays converge. It is commonly measured in millimetres (mm) and is one of the most important lens selection and image quality parameters. Field of view, magnification, and depth of field are all determined by a lens' focal length. This makes it a crucial consideration when choosing how to capture an image. A shorter focal length produces a wider field of view, whereas a longer focal length produces a narrower field of view. This is due to the fact that the shorter the focal length, the more of the scene is contained within the frame. Longer focal lengths are utilised to magnify distant objects, whereas shorter focal lengths are utilised to capture a broader perspective. The focal length also influences the depth of field, which is the portion of an image that appears in focus. A shorter focal length will result in a greater depth of field, allowing a greater portion of the image to be in focus. This makes it ideal for architectural and landscape photography. In contrast, a longer focal length will have a shallower depth of field, making it ideal for portraiture by blurring the background. Lastly, focal length also affects an image's magnification. A shorter focal length will result in a lower magnification, whereas a longer focal length will result in a higher magnification. For this reason, telephoto lenses are used to photograph distant subjects, such as wildlife and sporting events. In conclusion, focal length is an essential consideration when selecting a lens and taking photographs. A shorter focal length will result in a wider field of view and greater depth of field, while a longer focal length will result in a narrower field of view and shallower depth of field. It also affects an image's magnification, making it a crucial parameter for any photographer [190, 106].

The *principal point* is the image's centre, and it is where all incoming light rays converge to form a single image. It is an essential camera calibration parameter used to determine the image's centre and the camera's orientation. Typically, the principal point is measured in pixels and is used to calculate the coordinates of other points within an image. The focal point is determined by the optics and geometry of the camera. It is determined by locating the optical centre of the lens, which is the point where light rays converge. This point is then employed to determine the coordinates of the primary point. The focal point is crucial in determining the orientation and distortion of the camera. It is utilised to calculate the angle of view, which is the camera's field of view. It is also utilised to calculate radial distortion, which is lens-induced distortion. Additionally, the principal point is used to determine the position of the camera in relation to other images. It is possible to determine the relative position of the camera in each image by determining the coordinates of the principal point in each image. This is essential for applications such as stereo vision and three-dimensional reconstruction. In conclusion, the principal point is an essential camera calibration parameter used to determine the image's centre and camera orientation. Calculated from the optical centre of the lens, it is used to determine the camera's field of view, radial distortion, and relative position.

For any application requiring accurate measurements of objects in images, it is essential to comprehend the main point [146, 182, 129, 171].

Distortion coefficients are parameters that used to model the distortion of an optical system. However, they can also be used to represent telescopes and microscopes. The coefficients are used to calculate the image's distortion, which is then applied to correct the lens' aberrations. Frequently, the distortion of a lens is represented by a polynomial equation. This equation is used to calculate the distortion coefficients, which are parameters that describe the lens' distortion. Coefficients are typically represented as k1, k2, and k3, with higher order coefficients added as needed. Using the distortion coefficients, image distortion can be calculated. This information can then be used to correct lens-induced aberrations. The coefficients can be used, for example, to compute the distortion of a fisheye lens, which can then be corrected through image processing techniques. The distortion coefficients are also crucial for camera calibration and applications involving robotic vision. They are used to determine the distortion of a camera lens, which is then used to calculate camera parameters such as focal length and centre of gravity. This information is then used to precisely measure the size of objects in an image. In conclusion, the coefficients of optical system distortion are model parameters. However, they can also be used to simulate other optical systems. The coefficients are used to calculate the image's distortion, which is then applied to correct the lens' aberrations. In addition to determining the camera's parameters, they are required for camera calibration and robotic vision [188, 72, 49].

E.1.3 pixel per metric "ratio"

One of the very first and common approaches in the existing software/applications to determine size of an upper human body in an image was to perform a "calibration" (not to be confused with intrinsic/extrinsic calibration) using a reference object Adrian [16]. Our reference object should have two important properties;

1. We should know the dimension of one of the object in the image (in terms of width or height) in a measurable unit (such as inches, millimeters and etc.)
2. Our reference object should be visible and easily found in an image, either based on the appearances (like being a distinctive color or shape , unique and different from all other objects in the image) or via the placement of the object (such as the reference object always being placed in the bottom-right corner of an image). In either case, our reference should be uniquely identifiable in some manner.

Most of the current applications by asking the height of the participants the use that information as a reference to calculate the pixels per metric. We take this a bit further by using chessboard paper as our reference object in central of the image with the known width.

By guaranteeing the chessboard paper is in the central of the image, we can sort our upper body contours from top to bottom, and use it to define our pixel-per-metric, which we define as;

$$pixels\_per\_metric = object\_width/know\_width \tag{E.1}$$

However, this approach is not very accurate and stable for the following reasons;

1. First, as the photos are taken with the smartphones (such as; iPhone, Android), the angle is most certainly not a perfect 90-degree angle at the object. Without a perfect 90-degree view (or as close to it as possible), the dimensions of the objects can be appear distorted.
2. Second, We did not calibrate our smartphones using the intrinsic and extrinsic parameters of the camera which caused us having prone to radial and tangential lens distortion on the photos.
3. Third, by having a wrong value (width) we will recieve completely different results.

Performing an extra calibration step to find these parameters can “un-distort” our photos and lead to a better upper human body size approximation.

E.1.4 Zhang Method

This method uses a set of known object points to estimate the intrinsic and extrinsic parameters of the camera. The intrinsic parameters are the focal length, principal point, and skew, while the extrinsic parameters are the position and orientation of the camera.

The code first loads the image and then uses the cv2.findChessboardCorners() function to find the corners of the chessboard. The corners are then refined using the cv2.cornerSubPix() function. The code then uses the cv2.calibrateCamera() function to estimate the intrinsic and extrinsic parameters of the camera.

Once the camera parameters are estimated, the code can be used to convert pixel units to centimeter. To do this, the code first calculates the pixel size of the image. This is done by dividing the width of the image by the number of pixels in the width. The code then uses the camera parameters to calculate the physical size of the chessboard in centimeter. This is done by multiplying the pixel size of the image by the focal length of the camera. The code then uses the physical size of the chessboard to calculate the physical size of each pixel in centimeter. This is done by dividing the physical size of the chessboard by the number of pixels in the chessboard.

The code then uses the physical size of each pixel to convert pixel units to centimeter. To do this, the code first finds the coordinates of the point in the image that corresponds to the top left corner of the chessboard. The code then uses the physical size of each pixel to calculate the physical coordinates of the top left corner of the chessboard. The code then uses the physical coordinates of the top left corner of the chessboard to calculate the physical coordinates of any point in the image.

The following is an example of how the code can be used to convert pixel units to centimeter:

```python
import cv2

# Load the image
image = cv2.imread(’image.jpg’)

# Find the corners of the chessboard
corners = cv2.findChessboardCorners(image, (9, 6))

# Refine the corners
corners = cv2.cornerSubPix(image, corners, (5, 5), (-1, -1), criteria)

# Estimate the intrinsic and extrinsic parameters of the camera
camera_matrix, dist_coeffs, rvecs, tvecs = cv2.calibrateCamera(corners,
    image.shape[1::-1], None, None, flags=cv2.CALIB_RATIONAL_MODEL)

# Calculate the pixel size of the image
pixel_size = image.shape[1] / float(corners.shape[0])

# Calculate the physical size of the chessboard
chessboard_size = 6 * pixel_size

# Calculate the physical size of each pixel
pixel_size_centimeter = chessboard_size / float(corners.shape[0])

# Convert pixel units to centimeter
for i in range(corners.shape[0]):
    for j in range(corners.shape[1]):
        corners[i, j] = corners[i, j] * pixel_size_centimeter

# Display the image with the converted coordinates
cv2.imshow(’Image’, image)
cv2.waitKey(0)
```

Listing E.1 Grayscale code in python

The code first loads the image and then uses the cv2.findChessboardCorners() function to find the corners of the chessboard. The corners are then refined using the cv2.cornerSubPix() function. The code then uses the cv2.calibrateCamera() function to estimate the intrinsic and extrinsic parameters of the camera.

The cv2.calibrateCamera() function takes the following parameters:

- corners: A list of 2D points that correspond to the corners of the chessboard in the image.
- image_size: A tuple that contains the width and height of the image.
- object_points: A list of 3D points that correspond to the corners of the chessboard in the real world.
- camera_matrix: A 3x3 matrix that contains the intrinsic parameters of the camera.
- dist_coeffs: A 5-element vector that contains the distortion coefficients of the camera.
- rvecs: A 3x1 vector that contains the rotation vectors of the camera.
- tvecs: A 3x1 vector that contains the translation vectors of the camera.

The cv2.calibrateCamera() function returns the following values:

- camera_matrix: A 3x3 matrix that contains the intrinsic parameters of the camera.
- dist_coeffs:

E.1.5 OpenCV's Camera Calibration Algorithm

OpenCV is an open source computer vision library which has a wide range of functions for processing and analysing digital images. One of the most important functions of OpenCV is camera calibration, which helps to correct the distortions in the captured images caused by the camera's optics.

Collecting a collection of images with a specified calibration pattern is the initial stage in camera calibration. This design, which is often a checkerboard pattern, can be made digitally or printed on paper or cardboard. For instance, the OpenCV library offers a collection of functions for producing a checkerboard pattern. For each image in the calibration set, the checkerboard pattern must be positioned in the camera's field of view.

After capturing the calibration images, the following step is to identify the calibration pattern in each image. This is accomplished by recognising the corners of the checkerboard pattern in the image. OpenCV includes functions for recognising the pattern's corners and computing the transformation between the pattern and the picture.

After the location of the calibration pattern is known, the internal parameters of the camera may be computed. These parameters characterise the optics of the camera, including its focal length, lens distortion, and other features. OpenCV includes functions for computing these parameters from the identified corners of the calibration pattern.

The next step is to fix the distortions in the images using the computed parameters. For carrying out this task, OpenCV offers a collection of functions. These functions can be used to rectify lens distortion and other optically-induced distortions in cameras.

The corrected images can then be utilised to generate a three-dimensional representation of the calibration pattern. OpenCV includes functions for computing the 3D coordinates of the calibration pattern from the corrected images. This 3D model can be used to calibrate additional cameras or make models of other items.

E.1.6 Distance calibrating using checkerboard

E.1.6.1 Geometry of Image Formation

Before beginning the calibration procedure and dive into our method, we must first understand the geometric shape of the image. To facilitate comprehension, let's imagine a camera in a room. Given the 3D point P in the room, the pixel coordinates (u, v) of this

3D point in the image captured by the smartphone camera must be determined. In this configuration, three coordinate systems are in play. including (i) the World Coordinate System, (ii) the Camera Coordinate System, and (iii) the Image Coordinate System [148].

(i). World Coordinate System

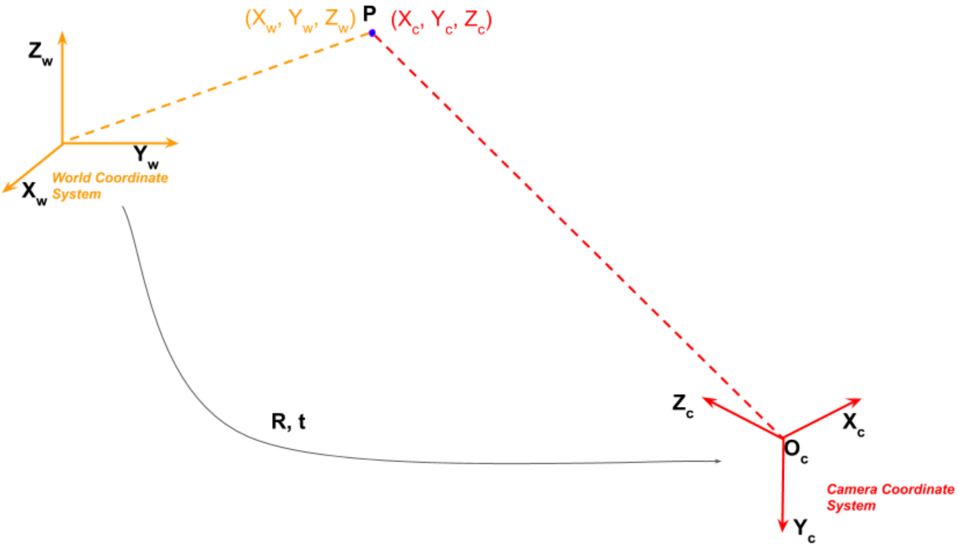


Fig. E.1 The World Coordinate and the Camera Coordinate system are relates by six parameters (3 for rotation, and 3 for translation) which are called the extrinsic parameters of a camera [148].

To establish the location of points in this room, we must first describe its coordinate system, which requires two things: 1. **The room's origin** 2. **X, Y, Z axes**. It is now possible to determine the three-dimensional coordinates of any point in this room by measuring its distance from the origin along the X, Y, and Z axes.

This coordinate system attached to the room is called the **World Coordinate System** shown as a orange colored axes in Figure E.1. We will use (') i.e. $X_w{}'$ to show the axis, and regular font to show a coordinate of the point (i.e. X_w). Let's imagine a point called P in this room. The coordinates of P are given in the world coordinate system by $(X_w, Y_w$

and Z_w). By measuring the distance of this point along the three axes from the origin, we may get the coordinates of this point(X_w, Y_w and Z_w). Now let's move on the next step and put the camera in this room.

(ii). Camera Coordinate System

The results of placing the camera at the origin of the room and aligning it so that its X, Y, and Z axes correspond to the $X_w{}'$, $Y_w{}'$, and $Z_w{}'$ axes of the room will be identical. Nevertheless, we are limited because the camera cannot be placed anywhere in the room. In this instance, it is necessary to determine the relationship between the 3D room (e.g., world coordinates) and the 3D camera coordinates.

Let's say that our camera is situated at arbitrary location (t_x, t_y and t_z) in the room. In technical words, (t_x, t_y and t_z) converts the camera coordinates to world coordinates. In other words, the camera can view in any direction because it rotates to the world coordinate system.

In 3D rotation, three characteristics, yaw, pitch, and roll, are collected. Moreover, we can view it as a three-dimensional axis (two parameters) and a single angular rotation about that axis (one parameter). The world and the camera coordinate are related to the rotation matrix **R** and a 3 elements translation vector **t**. This means that point P is different for the X_w, Y_w and Z_w coordinate values in world coordinates and X_c, Y_c and Z_c coordinate value in the camera coordinate system. In Figure E.1, the camera coordinate system is shown as red. According to the following equations, the two coordinate values are related.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{t} \tag{E.2}$$

Importantly, displaying rotation as a matrix allowed us to perform the rotation by multiplying the simple matrix, as opposed to laboriously manipulating the symbols required by other displays, such as yaw, pitch, and roll. I hope this clarifies why rotations are represented as matrices. Sometimes, the mentioned equations are stated in a more compact form. The 3x1 translation vector is added as a column to the 3x3 rotation matrix to produce a 3x4 matrix known as **Extrinsic Matrix**.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{[R \mid t]} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{E.3}$$

where, the Extrinsic Matrix **P** is given by;

$$\mathbf{P} = \mathbf{[R \mid t]} \tag{E.4}$$

As you may have noticed, Equation E.4 uses homogeneous coordinates, which means that homogeneous coordinates have an additional dimension called **W** that scales the X, Y, and Z dimensions. When W = 1, the X, Y, and Z values are deemed "correct." A point in coordinates (X, Y, Z) can be expressed as (X, Y, Z, 1) in homogeneous coordinates. When W = 0, homogeneous coordinates provide the representation of infinite quantities with finite numbers.

(iii). Image Coordinate System

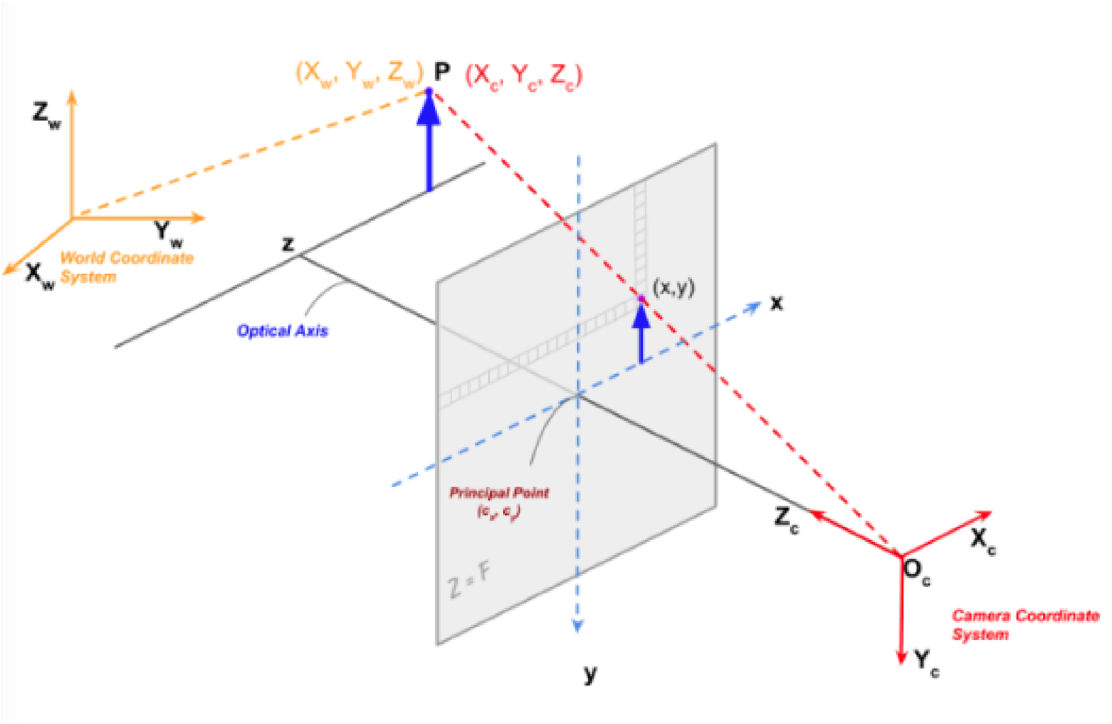


Fig. E.2 displays the projection of point P onto the image [148].

Once we have a point in the camera's 3D coordinate system, we can project it onto the picture plane to determine its location in the image by applying a rotation and translation to the point's world coordinate. Figure E.3 illustrates a point P whose camera coordinates are (X_c, Y_c and Z_c). We can still transform the global coordinate system using the Extrinsic Matrix. Since we do not know the coordinates of this location in the camera's coordinate system, we can determine the world's coordinates using Equation E.3.

The projection of a simple optical centre (pin hole) camera is shown in Figure E.3. The pinhole is represented by O_c. In the real world, a point on the screen produces an inverted image. For mathematical convenience, we perform all the computations as though the picture plane is in front of the optical centre, as the image read from the sensor can be rotated somewhat by 180 degrees to compensate for the inversion. According to Olaf Peters, this is not required in the actual world - "Even simpler, a real camera's sensor simply reads from the lowest row in reverse (from right to left) and then from bottom to top for each row. By using this procedure, the image is automatically created with the correct left-to-right orientation. So, it is no longer necessary to rotate the image in practise ". The distance f (focal length) separates the image plane from the optical centre.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{E.5}$$

We may display the projected picture (x, y) of the 3D point (X_c, Y_c and Z_c) using the simplified matrix equations shown in Equation E.5.

$$x = \frac{X_c}{Z_c} y = \frac{Y_c}{Z_c} \tag{E.6}$$

The K matrix following Equation E.5 is known as the **Intrinsic Matrix** and comprises the intrinsic properties of the camera.

$$\mathbf{K} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{E.7}$$

As the pixels in the image sensor may not be square, we may have two distinct focal lengths, f_x and f_y, despite the fact that the matrix K displays only the focal length. The optical centre (c_x, c_y) of the camera may not match with the image coordinate system's centre. In addition, the x and y axes of the camera sensor may have a slight variation around. With the previous information in mind, the camera matrix may be rewritten as;

$$\mathbf{K} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{E.8}$$

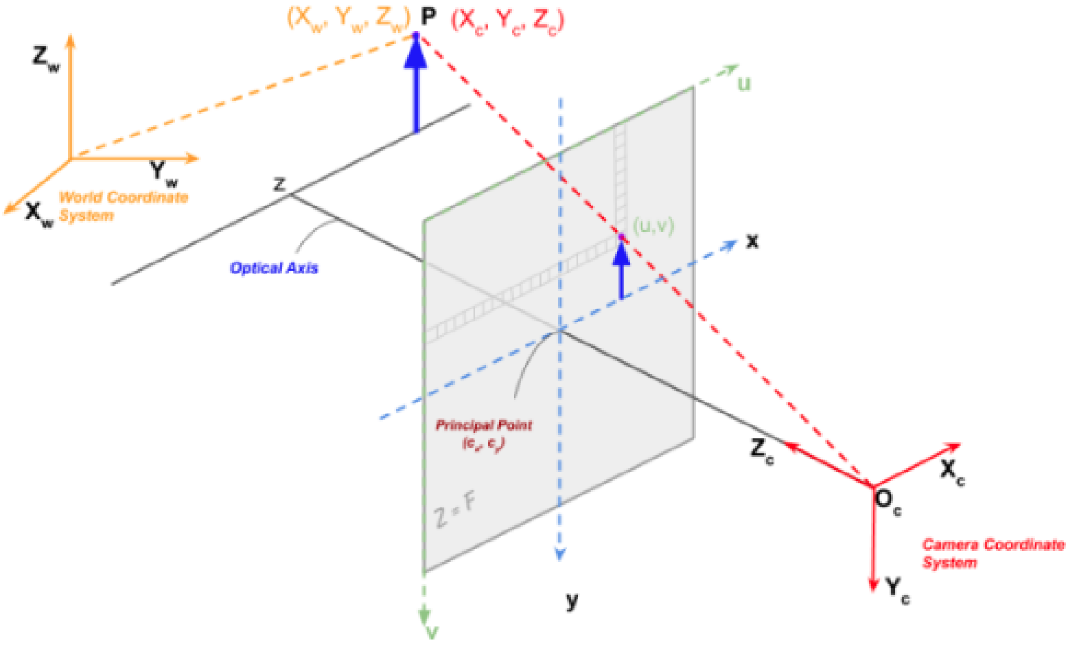


Fig. E.3 illustrates a more realistic scenario when the origin of the image pixel coordinate is in the upper left corner. The camera's intrinsic matrix must take into account the location of the main point, the skew of the axes, and different focal lengths along with different axes. [148].

Equation E.13 requires the x and y pixel coordinates relative to the image's centre. When working with photos, however, the origin is in the upper left corner. Let's define image coordinates as (u, v).

$$\begin{bmatrix} u^{'} \\ v^{'} \\ w^{'} \end{bmatrix} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{E.9}$$

where,

$$u = \frac{u^{'}}{w^{'}\text{(E.10)}}$$

$$v = \frac{v^{'}}{w^{'}\text{(E.11)}}$$

E.1.6.2 Camera Calibration

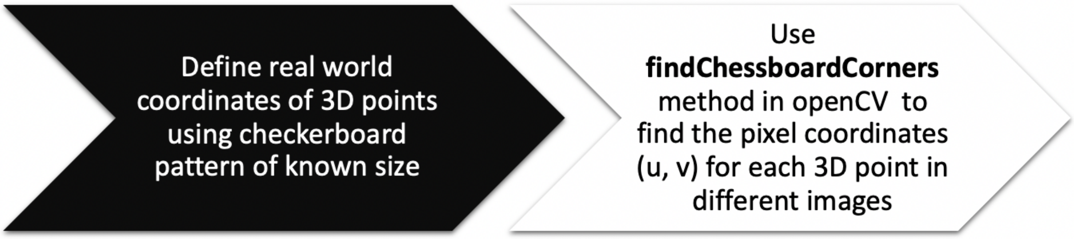
Camera calibration is the process of estimating the parameters of a camera. This means we have all the information (parameters or coefficients) about the camera required to determine an accurate relationships between a 3d point in the real world and its corresponding 2D projection (pixel) in the image captured by that calibrated camera [149]. Usually, this means recovering two kinds of parameters;

1. **Internal** parameters of the camera/lens system (such as; focal length, optical center and radial distortion coefficients of the lens.)

2. **External** parameters which refers to the orientation (rotation and translation) of the camera with respect to some world coordinate system.

The calibration process is needed to find camera parameters, when these are left unknown or deemed unreliable. With calibration we determine the classic: 3x3 matrix intrinsic matrix *K*, 3x3 rotation matrix R, and 3x1 translation vector *t* using a set of known 3D points (X_w, Y_w, Z_w) and their corresponding image coordinates (u, v). By having both values of intrinsic and extrinsic ([R | t]) parameters, the camera is then said to be calibrated.

The following flowchart illustrates the main two steps of camera calibration.


Define real world coordinates of 3D points using checkerboard pattern of known size
Use **findChessboardCorners** method in openCV to find the pixel coordinates (u, v) for each 3D point in different images

E.1.6.2.1 Step 1: Define real-world coordinates with checkerboard pattern Our world coordinates system are fixed by the checkerboard pattern that is given to the participants to face it in front of the camera. the corners in the checkerboards are our 3D points. The origin of the world coordinate systems is any corner of the following checkerboard which can be seen in Figure E.4. The X_w and Y_w are along to the participant's hand which they are face it in front of the camera as well as the Z_w axis is perpendicular. All points on the checkerboard are therefore on the XY plane (e.g. $Z_w = 0$).

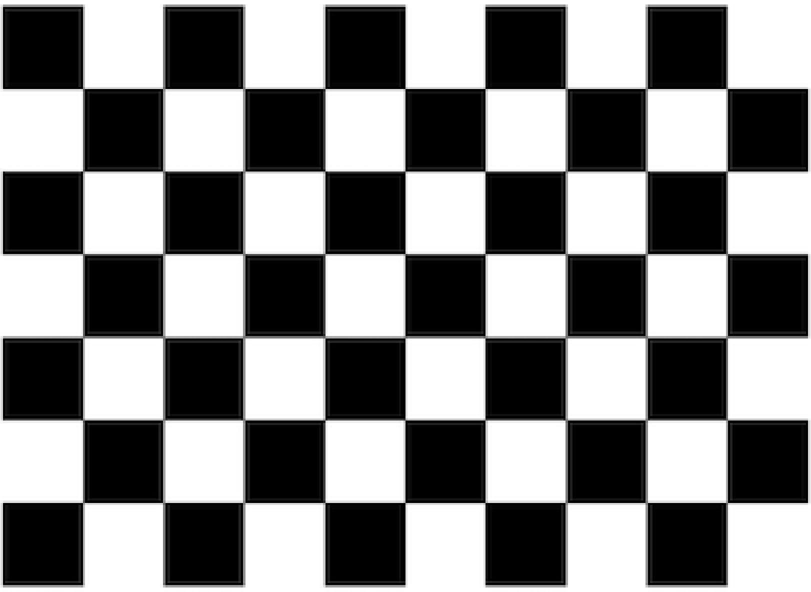


Fig. E.4 Checkerboard

We calculate the camera parameters by a set of know 3D points (X_w, Y_w, Z_w) and their corresponding image coordinates (u, v) in the process of calibrating.

We photograph a checkerboard pattern with a known dimension at many different orientations for the 3D points. Since all the corner points lie on a plane and the world coordinate is attached to the checkerboards, we can arbitrarily choose Z_w for every point to be 0. Since points are equally spaced in the checkerboard, the (X_w, Y_w) coordinates of each 3D point are easily defined by taking one point as reference (0, 0) and defining remaining concerning that reference point.

Why it is important to use checkerboard pattern in calibration?

Since checkerboard pattern is easy to detect in an image, not only that, the corners of squares on the checkerboard are ideal for localizing them, this is because they have sharp gradients in two directions. These corners are also related to the fact that they are at the intersection of checkerboard lines. All these facts are used to robustly locate the corners of the squares in a checkerboard pattern.

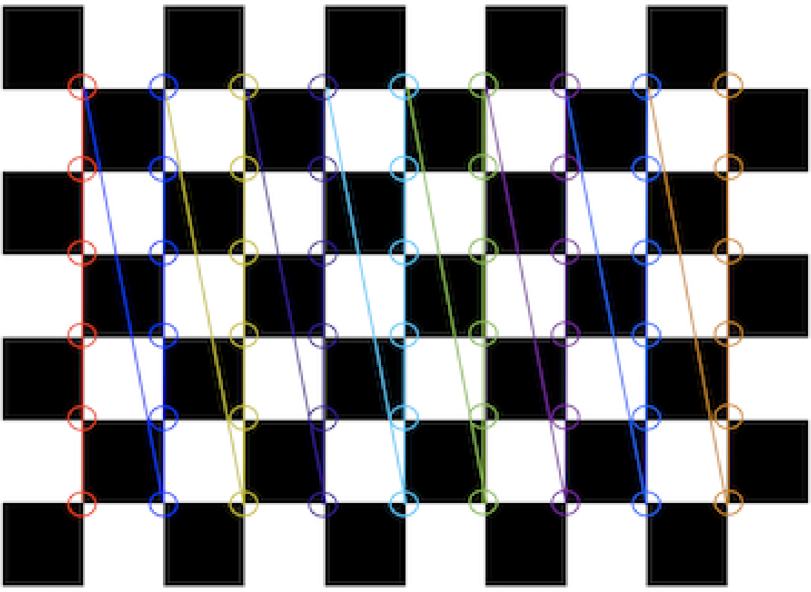


Fig. E.5 The following dummy checkerboard illustrates the result after drawing detected checker board corners

E.1.6.2.2 Find 2D coordinates of checkerboard After defining real-world coordinates, what we need are the locations of the 2D pixels of checkerboards corners in the image which contains two steps. First to call the function that provided by OpenCV **findChessboardCorners** that looks for a checkerboard and return coordinates of the corners. Second, to get good results it is important to collect the position of corners with a sub-pixel level of accuracy which is an OpenCV's function **cornerSubPix** that takes in the original image, the location of corners and looks for the best corner location inside a small neighbourhood of the original location. The algorithm is iterative and therefore, we need to specify the termination criteria (e.g. number of iterations and/or accuracy)

E.2 Method

The first step in camera calibration is to identify the camera's intrinsic parameters, starting with the focal length and principal point, which can be determined by measuring the size of the image in pixels and the size of the object in the image. The next step is to determine the camera's extrinsic parameters, including the position of the camera in 3D physical space relative to the object. This is typically done using a calibration pattern or calibration object.

Once the intrinsic and extrinsic parameters have been determined, the next step is to measure the distance between the camera and the object. This is done by measuring the size of the object in pixels and then converting the pixel units to centimeters. This can also be achieved using a calibration pattern (and the calculated camera parameters). We can then convert pixel units to centimeters using a conversion factor determined by dividing the distance between the camera and the object by the size of the object in pixels: Distance in centimeters = (Pixel distance in pixels) * (Focal length in centimeters) / (Image width in pixels)

In practice, our current (classic) camera calibration process involves capturing images of a calibration object — in our case, a checkerboard pattern — from multiple angles, and using the captured images to estimate and refine the camera parameters. The first step in the calibration process is to create a set of object points that represent the 3D coordinates of the calibration object corners. These points are defined in the coordinate system of the calibration object, and their values remain constant throughout the calibration process. We detect the checkerboard corners in the captured images using the OpenCV cv2.findChessboardCorners() function, which returns the pixel coordinates of the detected corners.

Once the object and image points have been obtained for each image, the function cv2.calibrateCamera() is called to estimate the camera matrix, distortion coefficients, rotation and translation vectors. The camera matrix contains the intrinsic parameters of the camera, including the focal length, principal point, and skew. The distortion coefficients describe the lens distortion caused by the camera optics, and the rotation and translation vectors describe the camera's position and orientation in the 3D world.

We note that the cv2.calibrateCamera() function uses a mathematical optimization algorithm to minimize the reprojection error between the observed image points and the projected 3D object points. The reprojection error is the difference between the observed image points and the corresponding points reprojected onto the image plane using the estimated camera parameters. The optimization algorithm minimizes the sum of the squared reprojection errors across all images, which results in the most accurate estimate of the camera parameters.

To summary the following are the steps involved in our camera calibration algorithm:

1. Take a set of images of a checkerboard pattern from different viewpoints.
2. For each image, find the corners of the checkerboard pattern.
3. Use the corners of the checkerboard pattern to estimate the intrinsic and extrinsic parameters of the camera.
4. Save the intrinsic and extrinsic parameters of the camera.

The following are the mathematical equations used in the camera calibration algorithm:

The intrinsic parameters of the camera are estimated using the following equation:

$$\mathbf{K} = \frac{1}{f_x}\begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_y \\ 0 & 0 & 1 \end{bmatrix} \tag{E.12}$$

where f_x is the focal length in the x-direction, f_y is the focal length in the y-direction, u_0 is the principal point in the x-direction, and v_0 is the principal point in the y-direction.

The extrinsic parameters of the camera are estimated using the following equation:

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & r_{22} & R_{23} \\ R_{31} & R_{32} & r_{33} \end{bmatrix} \tag{E.13}$$

where R_{ij} is the element of the rotation matrix in the ith row and jth column.

The translation of the camera is estimated using the following equation:

$$\mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{E.14}$$

The following are the metrics used in the camera calibration algorithm:

The reprojection error is the error between the actual and reprojected points. The reprojection error is calculated using the following equation:

$$e = \sqrt{\sum_{n=1}^{N} (x_i - x_i^i)^2 + (y_i - y_i^i)^2} \tag{E.15}$$

where x_i and y_i are the actual coordinates of the ith point, and $x_i^{'}$ and $y_i^{'}$ are the reprojected coordinates of the ith point.

The number of inliers is the number of points that are within a certain threshold of the reprojection error. The number of inliers is calculated using the following equation:

$$N_i = \sum_{n=1}^{N} \mathbb{I}(e_i < \in) \tag{E.16}$$

where $\mathbb{I}(\cdot)$ is the indicator function, and $\in$ is the threshold.

The reprojection error is used to evaluate the performance of the camera calibration algorithm. The lower the reprojection error, the better the performance of the algorithm.

The number of inliers is used to determine the number of points that are needed to calibrate the camera. The more inliers, the more accurate the calibration will be.

The following are the mathematical algorithms used in the camera calibration algorithm:

The Levenberg-Marquardt algorithm is used to estimate the intrinsic and extrinsic parameters of the camera. The Levenberg-Marquardt algorithm is a nonlinear least-squares algorithm that is used to solve problems with multiple variables.

The RANSAC algorithm is used to estimate the number of inliers. The RANSAC algorithm is a robust estimator that is used to estimate parameters in the presence of outliers.

The following are the limitations of the camera calibration algorithm in our software:

- The algorithm assumes that the checkerboard pattern is perfectly planar. This is not always the case in practice, as the checkerboard pattern may be slightly warped or distorted. This can lead to errors in the calibration results.
- The algorithm assumes that the camera is perfectly still during the calibration process. This is not always the case in practice, as the camera may be slightly moved or shaken during the calibration process. This can also lead to errors in the calibration results.
- The algorithm assumes that the lighting conditions are constant during the calibration process. This is not always the case in practice, as the lighting conditions may change during the calibration process. This can also lead to errors in the calibration results.
- The algorithm is sensitive to noise in the image data. This can lead to errors in the calibration results.

- The algorithm is not robust to outliers. This means that if there are a few points in the image data that are not from the checkerboard pattern, these points can cause errors in the calibration results.

To avoid these limitations, we ran the input image through advanced image processing and segmentation to minimise the error that can be caused by the camera calibrations which we did mention them Chapter 6. This can help to ensure that the checkerboard pattern is as planar as possible, that the camera is as still as possible, and that the lighting conditions are as constant as possible. It can also help remove any noise or outliers from the image data.

E.3 Results

E.3.1 Statistical Analysis and Comparison of Chest Circumference Measurements: Distance between 0.5 to 3 meters vs. Distance less than 0.5 m or more than 3 meters.

The T-test was conducted to determine whether there is a statistically significant difference between the chest circumferences measured at a distance between 0.5 to 3 meters and those measured at a distance less than 0.5 m or more than 3 meters. It was hypothesized that the stress level of the latter group (M = 1.02, SD = 0.58, n = 78) would be greater than the stress level of the former group (M = 0.48, SD = 0.35, n = 78) based on the data received from the participants and by considering the distance from the camera. The results indicated a statistically significant difference between the two groups (t(127) = -7.14, $p < 0.001$, one-tail). This suggests that the two data sets are likely different, with the stress level being higher for measurements taken at a distance less than 0.5 m or more than 3 meters.

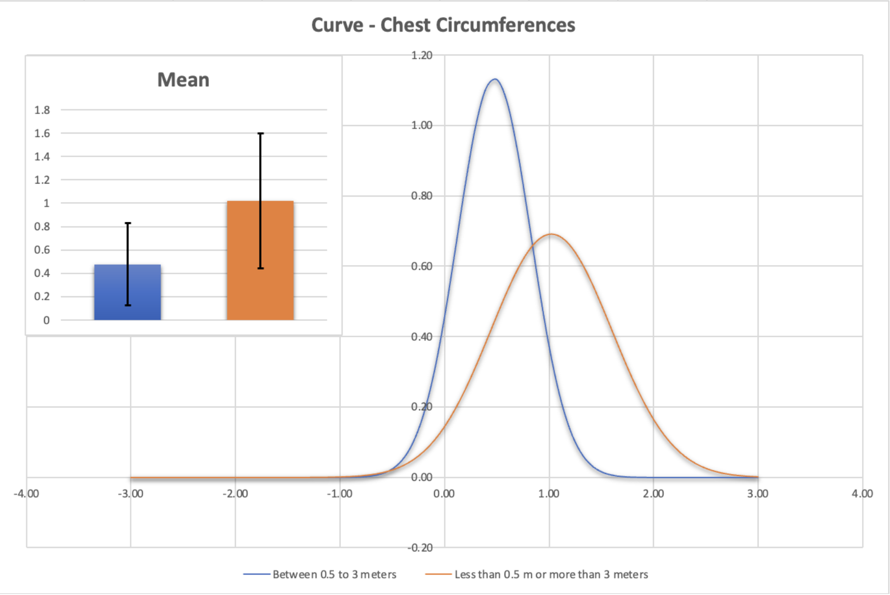


Fig. E.6 : demonstrate the each background's stress level on the bell curve.

The descriptive statistics for the chest circumference measurements at a distance between 0.5 to 3 meters showed a mean of 0.48 (SE = 0.04), with a standard deviation of 0.35. The median was 0.48, and the mode was 0.56. The measurements ranged from 0.01 to 1.73, with a sample size of 78.

For the chest circumference measurements taken at a distance less than 0.5 m or more than 3 meters, the mean was 1.02 (SE = 0.07), with a standard deviation of 0.58. The median was 0.99, and the mode was 1.02. The measurements ranged from 0.02 to 3.13, with a sample size of 78.

The analysis of error differences revealed that, for the distance between 0.5 to 3 meters group, 53.85% of the measurements had an error difference of less than 0.5 cm, 39.74% had an error difference between 0.5 and 1 cm, 5.13% had an error difference between 1 and 1.5 cm, and 1.28% had an error difference between 1.5 and 2 cm. In contrast, for the distance less than 0.5 m or more than 3 meters group, the percentages were as follows: 16.67% had an error difference of less than 0.5 cm, 37.18% had an error difference between 0.5 and 1 cm, 33.33% had an error difference between 1 and 1.5

cm, 7.69% had an error difference between 1.5 and 2 cm, 1.28% had an error difference between 2 and 2.5 cm, 2.56% had an error difference between 2.5 and 3 cm, and 1.28% had an error difference between 3 and 3.5 cm.

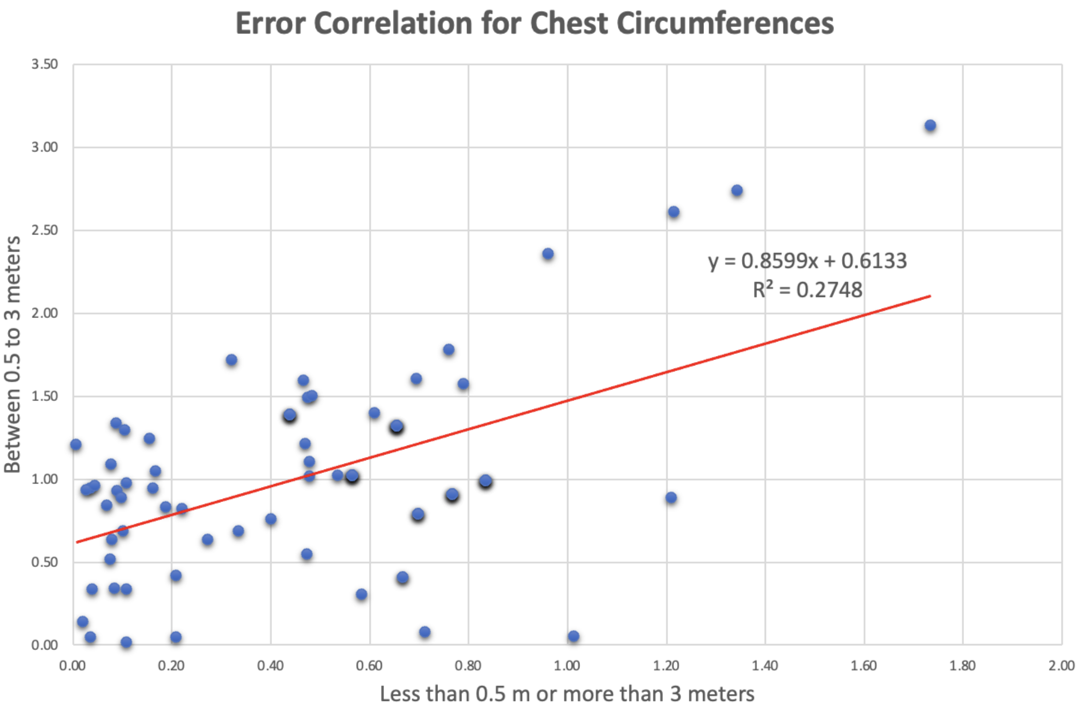


Fig. E.7 : displays the error correlation for the chest circumferences measurements

The error correlation between the two sets of measurements was found to be 0.52, indicating a moderate positive correlation between the errors in the two groups.

These findings suggest that there is a significant difference in chest circumference measurements depending on the distance from the camera. Measurements taken at a distance less than 0.5 m or more than 3 meters tend to have higher stress levels compared to those taken at a distance between 0.5 to 3 meters. However, it is important to note that the overall accuracy of the software remains acceptable in both cases.

E.3.2 Statistical Analysis and Comparison of Bust Circumference Measurements: Distance between 0.5 to 3 meters vs. Distance less than 0.5 m or more than 3 meters.

The T-test was conducted to determine whether there is a statistically significant difference between the bust circumferences measured at a distance between 0.5 to 3 meters and those measured at a distance less than 0.5 m or more than 3 meters. It was hypothesized that the stress level of the latter group (M = 1.02, SD = 0.53, n = 78) would be greater than the stress level of the former group (M = 0.50, SD = 0.30, n = 78) based on the data received from the participants and by considering the distance from the camera. The results indicated a statistically significant difference between the two groups (t(154) = -7.42, p < 0.001, one-tail). This suggests that the two data sets are likely different, with the stress level being higher for measurements taken at a distance less than 0.5 m or more than 3 meters.

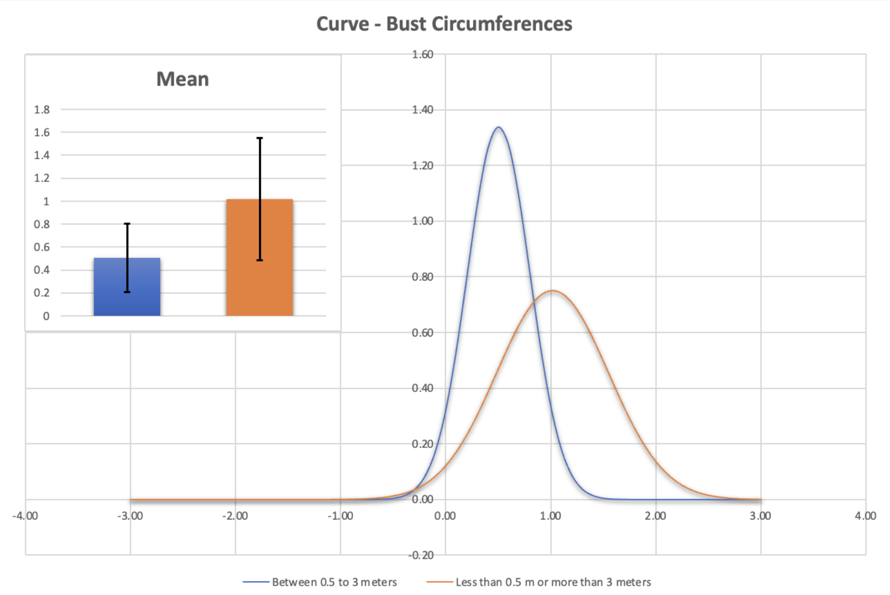


Fig. E.8 : demonstrate the each background's stress level on the bell curve.

The descriptive statistics for the bust circumference measurements at a distance between 0.5 to 3 meters showed a mean of 0.50 (SE = 0.03), with a standard deviation of 0.30. The median was 0.56, and the mode was 0.70. The measurements ranged from 0.00 to 1.25, with a sample size of 78.

For the bust circumference measurements taken at a distance less than 0.5 m or more than 3 meters, the mean was 1.02 (SE = 0.06), with a standard deviation of 0.53. The median was 1.01, and the mode was 1.02. The measurements ranged from 0.02 to 2.64, with a sample size of 78.

The analysis of error differences revealed that, for the distance between 0.5 to 3 meters group, 42.31% of the measurements had an error difference of less than 0.5 cm, 53.85% had an error difference between 0.5 and 1 cm, and 3.85% had an error difference between 1 and 1.5 cm. In contrast, for the distance less than 0.5 m or more than 3 meters group, the percentages were as follows: 15.38% had an error difference of less than 0.5 cm, 30.77% had an error difference between 0.5 and 1 cm, 39.74% had an error difference between 1 and 1.5 cm, 8.97% had an error difference between 1.5 and 2 cm, 3.85% had an error difference between 2 and 2.5 cm, and 1.28% had an error difference between 2.5 and 3 cm.

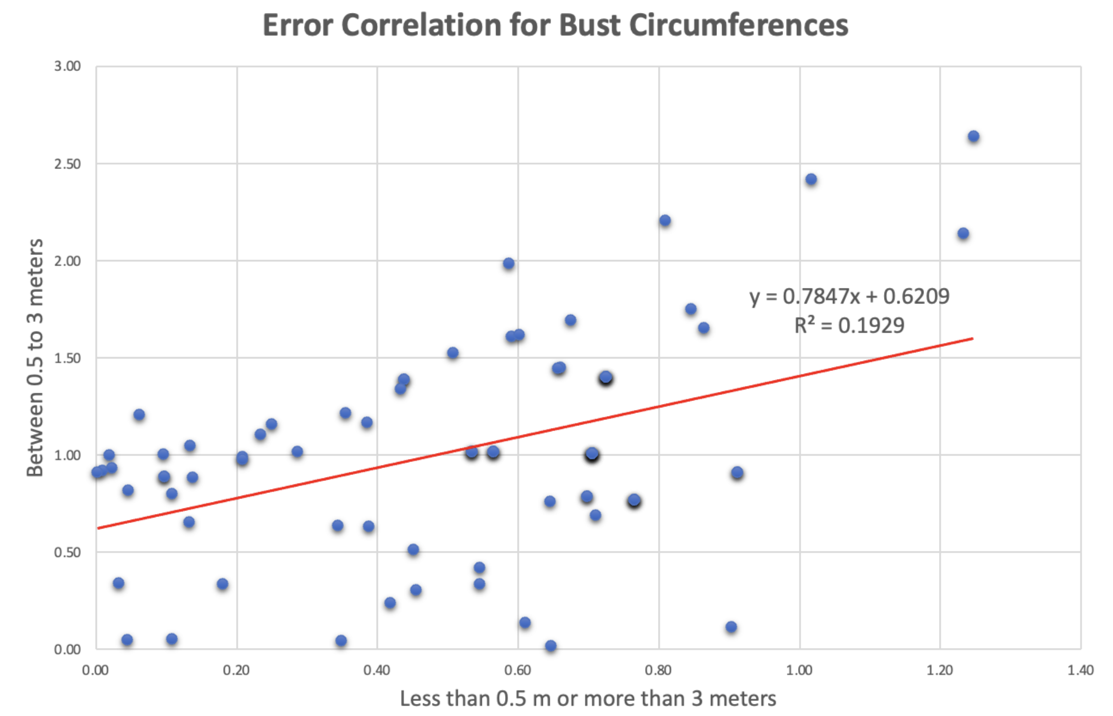


Fig. E.9 : displays the error correlation for the bust circumferences measurements

The error correlation between the two sets of measurements was found to be 0.43, indicating a moderate positive correlation between the errors in the two groups.

These findings suggest that there is a significant difference in bust circumference measurements depending on the distance from the camera. Measurements taken at a distance less than 0.5 m or more than 3 meters tend to have higher stress levels compared to those taken at a distance between 0.5 to 3 meters. However, the overall accuracy of the software remains acceptable in both cases.

E.3.3 Statistical Analysis and Comparison of Waist Circumference Measurements: Distance between 0.5 to 3 meters vs. Distance less than 0.5 m or more than 3 meters.

The T-test was conducted to determine whether there is a statistically significant difference between the waist circumferences measured at a distance between 0.5 to 3 meters and those measured at a distance less than 0.5 m or more than 3 meters. It was hypothesized that the stress level of the latter group (M = 1.29, SD = 0.62, n = 78) would be greater than the stress level of the former group (M = 0.56, SD = 0.35, n = 78) based on the data received from the participants and by best matching the human body shape. The results indicated a statistically significant difference between the two groups (t(122) = -9.13, p < 0.001, one-tail). This suggests that the two data sets are likely different, with the stress level being higher for measurements taken at a distance less than 0.5 m or more than 3 meters.

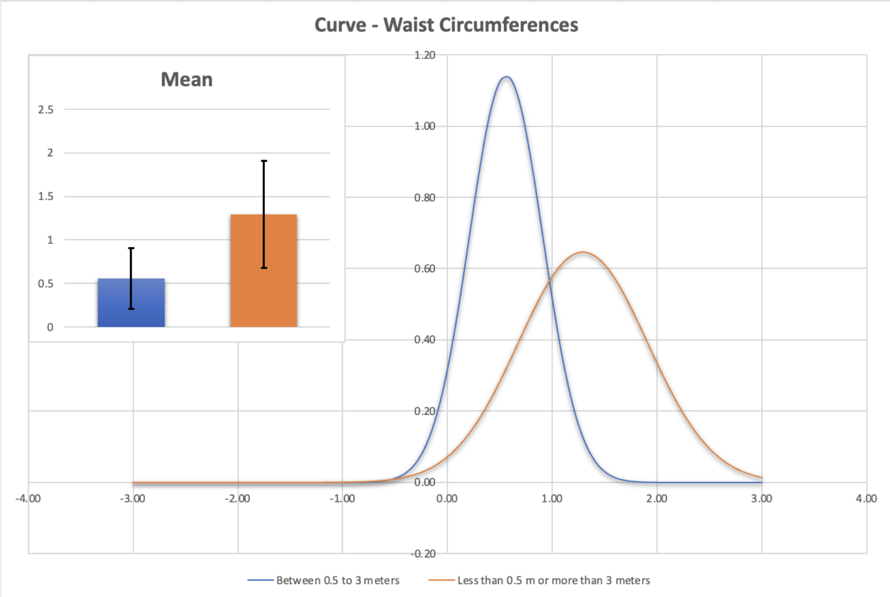


Fig. E.10 : demonstrate the each background's stress level on the bell curve.

The descriptive statistics for the waist circumference measurements at a distance between 0.5 to 3 meters showed a mean of 0.56 (SE = 0.04), with a standard deviation of 0.35. The median was 0.56, and the mode was also 0.56. The measurements ranged from 0.02 to 1.75, with a sample size of 78.

For the waist circumference measurements taken at a distance less than 0.5 m or more than 3 meters, the mean was 1.29 (SE = 0.07), with a standard deviation of 0.62. The median was 1.10, and the mode was 1.02. The measurements ranged from 0.13 to 3.16, with a sample size of 78.

The analysis of error differences revealed that, for the distance between 0.5 to 3 meters group, 38.46% of the measurements had an error difference of less than 0.5 cm, 53.85% had an error difference between 0.5 and 1 cm, 5.13% had an error difference between 1 and 1.5 cm, and 2.56% had an error difference between 1.5 and 2 cm. In contrast, for the distance less than 0.5 m or more than 3 meters group, the percentages were as follows: 10.26% had an error difference of less than 0.5 cm, 20.51% had an error difference between 0.5 and 1 cm, 33.33% had an error difference between 1 and 1.5 cm, 24.36% had an error difference between 1.5 and 2 cm, 8.97% had an error difference between 2 and 2.5 cm, 1.28% had an error difference between 2.5 and 3 cm, and 1.28% had an error difference between 3 and 3.5 cm.

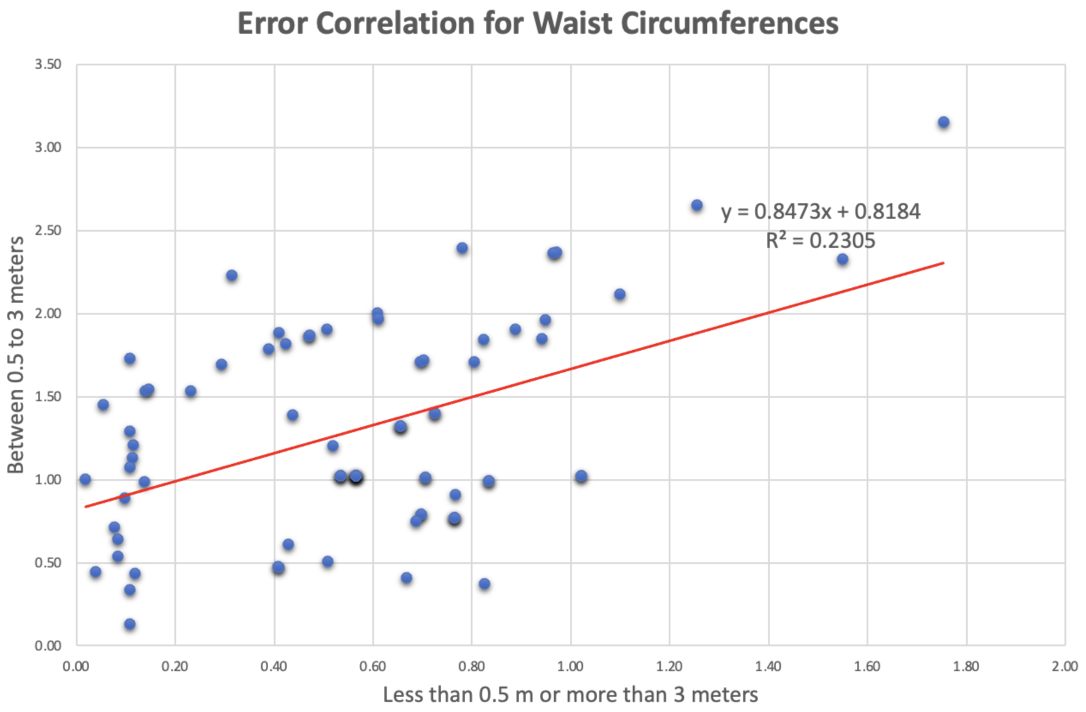


Fig. E.11 : displays the error correlation for the waist circumferences measurements

The error correlation between the two sets of measurements was found to be 0.48, indicating a moderate positive correlation between the errors in the two groups.

These findings suggest that there is a significant difference in waist circumference measurements depending on the distance from the camera. Measurements taken at a distance less than 0.5 m or more than 3 meters tend to have higher stress levels compared to those taken at a distance between 0.5 to 3 meters. However, the overall accuracy of the software remains acceptable in both cases.

E.3.4 Statistical Analysis and Comparison of Hips Circumference Measurements: Distance between 0.5 to 3 meters vs. Distance less than 0.5 m or more than 3 meters.

A T-test was conducted to examine whether there is a statistically significant difference between the hips circumferences measured at a distance between 0.5 to 3 meters and those measured at a distance less than 0.5 m or more than 3 meters. It was hypothesized that the stress level of the latter group (M = 0.99, SD = 0.60, n = 78) would be greater than the stress level of the former group (M = 0.52, SD = 0.42, n = 78) based on the data received from the participants and by best matching the human body shape. The results revealed a statistically significant difference between the two groups (t(138) = -5.56, p < 0.001, one-tail), indicating that the two data sets are likely different, with the stress level being higher for measurements taken at a distance less than 0.5 m or more than 3 meters.

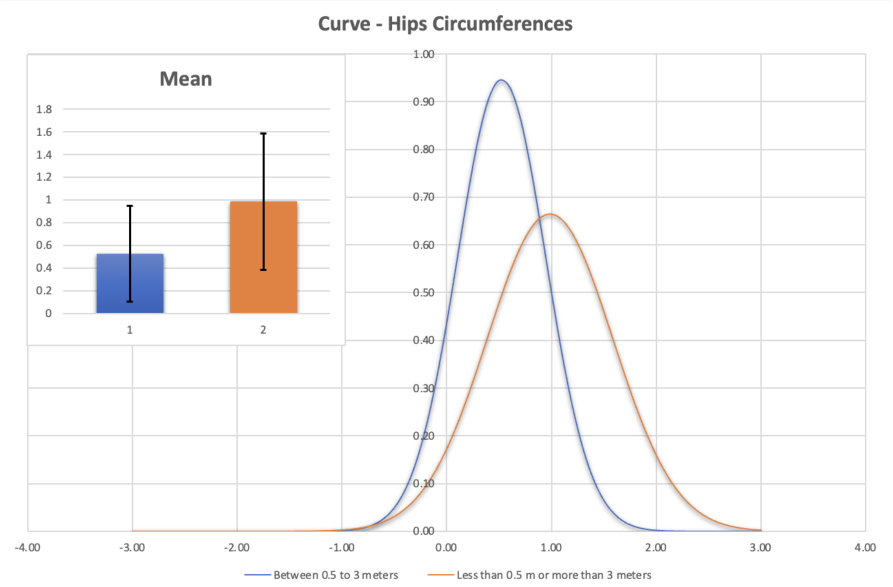


Fig. E.12 : demonstrate the each background's stress level on the bell curve.

The descriptive statistics for the hips circumference measurements at a distance between 0.5 to 3 meters showed a mean of 0.52 (SE = 0.05), with a standard deviation of 0.42. The median and mode were both 0.53. The measurements ranged from 0.02 to 3.02, with a sample size of 78.

For the hips circumference measurements taken at a distance less than 0.5 m or more than 3 meters, the mean was 0.99 (SE = 0.07), with a standard deviation of 0.60. The median was 0.98, and the mode was 1.02. The measurements ranged from 0.02 to 3.81, with a sample size of 78.

The analysis of error differences revealed that, for the distance between 0.5 to 3 meters group, 46.15% of the measurements had an error difference of less than 0.5 cm, 47.44% had an error difference between 0.5 and 1 cm, 5.13% had an error difference between 1 and 1.5 cm, and 1.28% had an error difference between 3 and 3.5 cm. In contrast, for the distance less than 0.5 m or more than 3 meters group, the percentages were as follows: 20.51% had an error difference of less than 0.5 cm, 32.05% had an error difference between 0.5 and 1 cm, 32.05% had an error difference between 1 and 1.5 cm, 11.54% had an error difference between 1.5 and 2 cm, 2.56% had an error difference between 2 and 2.5 cm, and 1.28% had an error difference between 3.5 and 4 cm.

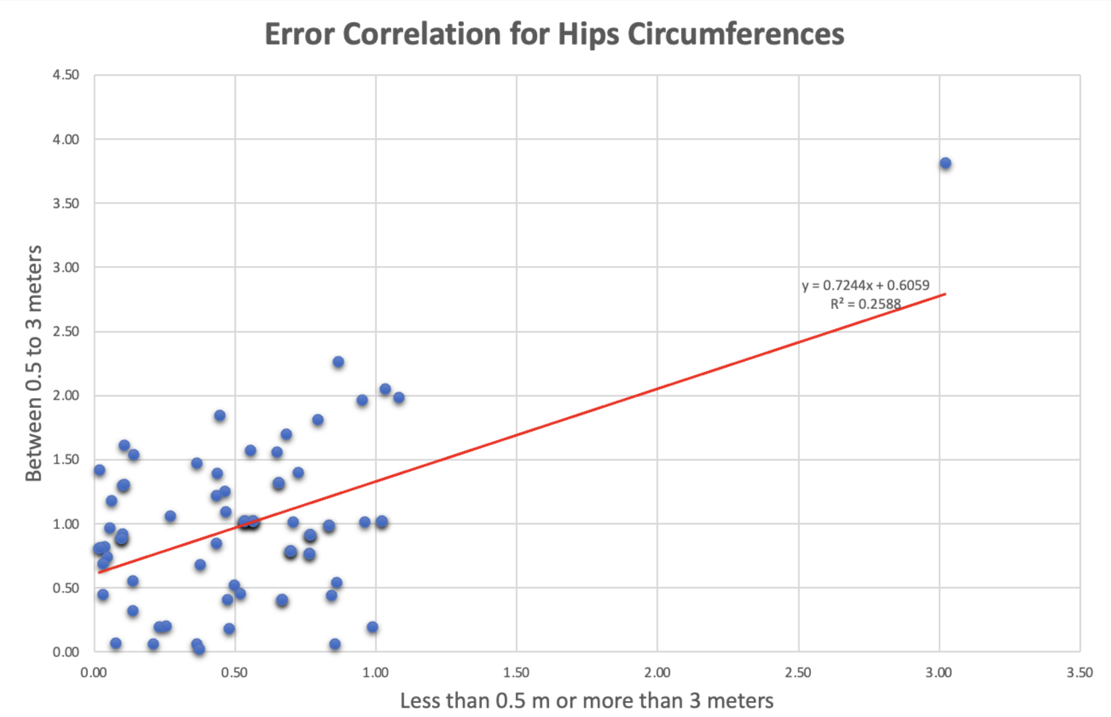


Fig. E.13 : displays the error correlation for the hips circumferences measurements

The error correlation between the two sets of measurements was found to be 0.51, indicating a moderate positive correlation between the errors in the two groups.

In conclusion, the findings suggest a significant difference in hips circumference measurements depending on the distance from the camera. Measurements taken at a distance less than 0.5 m or more than 3 meters tend to exhibit higher stress levels compared to those taken at a distance between 0.5 to 3 meters. However, the overall accuracy of the software remains satisfactory in both scenarios.

Appendix F

Chapter-7*

F.1 Evaluation and Comparison of Elliptical Mathematical Models for Human Body Measurements

F.1.1 Evaluate elliptical mathematical models

Before proceeding, let's first try to understand what perimeter means. The perimeter is the distance around any shape's contour or edge. It is also known as the circumference of the ellipse.

A practical example of measuring the perimeter of an ellipse would be the distance walked along the perimeter of an elliptical field. Or the length of fence necessary to surround it. In the following paragraphs, we shall determine the ellipse's perimeter.

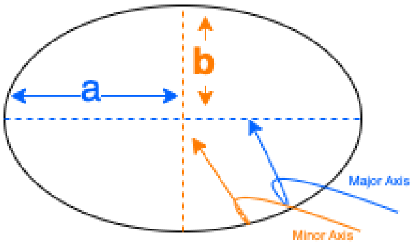


Fig. F.1 Ellipse perimeter

Perimeters of ellipses and other figures of the conic section cannot be calculated exactly (or accurately) using a standard formula, unlike most other shapes. However, there are numerous approximation formulas for calculating the perimeter's approximate value. Six equations have been chosen in order to calculate the ellipse's perimeter, which is represented by ***P***. The selection of equations with a low complexity level was chosen because it was thought to be the most effective way for calculating the length of the circumference of an ellipse, such as:

(A) Approximation I

This equation is used to determine the circumference of a circle or an ellipse. ""dist$_a$[1]"" and ""dist$_b$[2]"" are variables that represent the lengths of the semi-major and semi-minor axes of the ellipse or circle, respectively.

Because the perimeter of an ellipse is equal to the circumference of a circle with the same semi-major and semi-minor axes lengths, the equation can be applied to both circles and ellipses. The equation approximates the true perimeter of the circle or ellipse since it assumes that the ellipse is a perfect circle and that all locations along the ellipse's perimeter are equally distant from its centre.

$$P = \Pi(dist_a + dist_b) \tag{F.1}$$

where,

- $dist_a$ = $dist_b$
- the length the major axis is 2($dist_a$)
- the coordinates of vertices are ($\pm$dist$_a$, 0)
- the length the major axis is 2($dist_b$)
- the coordinates of vertices are ($\pm$dist$_b$, 0)

(B) Approximation II

[1] "$dist_a$" is semi major axis

[2] "$dist_b$" is semi minor axis

This is the second approximation of the ellipse equations, when the shape resembles a circle but has two axes of varying lengths, known as the major and minor axes. The major axis of an ellipse has the longest diameter, while the minor axis has the shortest diameter.

This equation is derived from the formula for the circumference of an ellipse, which is based on the fact that the circumference of any ellipse is equal to the product of the lengths of its major and minor axes multiplied by the constant Π [20], [33].

$$P = \Pi\sqrt{2dist_a^2 + dist_b^2} \tag{F.2}$$

Where,

- $dist_a \geq$dist$_b$ (where major axis is not three times bigger than minor axis)
- the length the major axis is 2$(dist_a)$
- the coordinates of vertices are ($\pm$dist$_a$, 0)
- the length the major axis is 2$(dist_b)$
- the coordinates of vertices are ($\pm$dist$_b$, 0)

(C) Approximation III

The Equation 7.2 is used when one of the values of "dist$_a$" is slightly bigger than the "dist$_b$". where "dist$_a$" is a major and "dist$_b$" minor axes. It is important that in this formula "dist$_a$" is not three times bigger than "dist$_b$" since it will effect on the average error (i.e., the ellipse is not overly "compressed").

$$P \approx 2 \times \Pi \times \sqrt{\frac{(dist_a^2) + (dist_b^2)}{2}} \tag{F.3}$$

Where,

- $dist_a \geq$dist$_b$ (where major axis is not three times bigger than minor axis)
- the length the major axis is 2($dist_a$)
- the coordinates of vertices are ($\pm$dist$_a$, 0)
- the length the major axis is 2($dist_b$)
- the coordinates of vertices are ($\pm$dist$_b$, 0)

(D) Ramanujan Formulas

An ellipse has equation (x/dist$_b$)^2 + (y/dist$_b$)^2 = 1. If the ellipse is close to the circle, the following approximation is very good.

$$P \approx \pi\left[3(dist_a + dist_b) - \sqrt{(3dist_a + dist_b)(dist_a + 3dist_b)}\right] \tag{F.4}$$

Where,

- $dist_a$ > $dist_b$ (where major axis is three times bigger than minor axis)
- the length the major axis is 2($dist_a$)
- the coordinates of vertices are ($\pm$dist$_a$, 0)
- the length the major axis is 2($dist_b$)
- the coordinates of vertices are ($\pm$dist$_b$, 0)

The advantage of Ramanujan's equation is that it more accurately describes the shape of an ellipse. In particular, it is able to better approximate the shape of an ellipse when it is highly elongated or has a large eccentricity. In addition, it can be used to accurately describe the shape of an ellipse even when the coefficients of the equation are not known.

(E) Ramanujan Formulas II

Other equations were also developed by Ramanujan.

$$P \approx \Pi((dist_a)+(dist_b))(1+\frac{3h}{10+\sqrt{(4-3h)}}) \tag{F.5a}$$

Where h is:

$$h = \frac{(dist_a - dist_b)^2}{(dist_a + dist_b)^2} \tag{F.5b}$$

Where,

- $dist_a$ > $dist_b$ (where major axis is three times bigger than minor axis)
- the length the major axis is 2($dist_a$)
- the coordinates of vertices are (±dist$_a$, 0)
- the length the major axis is 2($dist_b$)
- the coordinates of vertices are (±dist$_b$, 0)

(F) The perimeter and elliptic integrals

The parametric equations of the ellipse are; x = dist$_a$ cos θ and y = dist$_b$ sin θ.

Using the arc length formula of parametric equations, we have the arc length of a function (x(θ), y(θ)) over the interval [dist$_a$, dist$_b$] is given by $\int_{dist_a}^{dist_b}$ $(x'\ (\theta))^2$ + $(y'\ (\theta))^2$ dθ. Applying this formula for the ellipse over the interval [0, π/2], we get the ellipse circumferences perimeter for only in the first quadrant. Thus, we can estimate the total perimeter, by multiplying the resultant integral by four.

$$P = 4\int_0^{\pi/2} \sqrt{dist_a^2 cos^2\theta + dist_b^2 sin^2\theta d\theta} \tag{F.6}$$

Where,

- $dist_a$ > $dist_b$ (where major axis is three times bigger than minor axis)

- the length the major axis is 2($dist_a$)
- the coordinates of vertices are ($\pm$dist$_a$, 0)
- the length the major axis is 2($dist_b$)
- the coordinates of vertices are ($\pm$dist$_b$, 0)

Equation F.6 was proposed by Xun Wang et al. [10] who developed an approach that can estimate anthropometric dimensions based on 2D images by extracting landmarks through a convolutional neural network and by building a general multi-ellipse model in which body shape information is added to obtain more accurate results. This equation approaches the perfect integral formula of the ellipse equations pretty closely.

F.1.1.1 Findings

Now, the question is which of these equations is the best option for our software. The Ramanujan Formula I is the most popular of the existing technologies, according to our analysis. This is due to the accuracy of the Ramanujan Approximation. It calculates the perimeter of an ellipse to within 0.3% of its actual perimeter. This precision is impressive, as it is significantly superior to the most used estimate for determining the perimeter of an ellipse, the Archimedes-Euler approximation. The Archimedes-Euler approximation approximates the perimeter to within 2.6% of its true value.

The Ramanujan Approximation is also quite efficient. Using the Archimedes-Euler approximation to calculate the perimeter of an ellipse involves two integrals, which can be computationally demanding. The Ramanujan Approximation uses only a single integral, which is considerably quicker than the Archimedes-Euler approximation. In addition, the Ramanujan Approximation can be calculated with a small number of terms, which increases its efficiency.

Lastly, the Ramanujan Approximation is straightforward and simple to understand. The Ramanujan Approximation formula is easy and may be implemented simply in any programming language. As a result, it is an excellent option for computing the perimeter of an ellipse.

F.1.2 Results

F.1.2.1 Analysis of Variance (ANOVA) Results and Post-hoc Analysis for Chest Circumferences

This report presents the results of an Analysis of Variance (ANOVA) test performed on data collected from three different groups - Equation 7.2, Equation 7.3, and Equation F.6 for Chest circumferences. The context for this analysis is that these mean values represent error differences in centimeters, hence the group with the lower mean signifies higher accuracy.

In the summary statistics, the ANOVA test is used to see if there is a statistically significant difference between the ellipse formulas by testing for differences in the mean. Based on our collected data from the participants, and by best fitting the human body shape (horizontal slices) by ellipses, we observed that for elliptic profiles such that their major axis were not more than 3 times longer than their minor axis, Equation 7.2 has the lowest mean value of 0.345. Equation 7.3 has a mean value of 0.777, and Equation F.6 has the highest mean value of 0.779. The variances of the three groups are quite close, with Equation 7.2 at 0.105, Equation 7.3 at 0.101, and Equation F.6 at 0.101.

The ANOVA test, which evaluates if there are significant differences between the three groups, indicates a Between Groups sum of squares (SS) of 3.260 and a Mean Square (MS) value of 1.630. The F-statistic is 15.88, and the critical value of F is 3.118.

The p-value obtained from the test is 1.77752E-06, which is smaller than the commonly used significance level of 0.05. This means that we reject the null hypothesis of equal means and conclude that there is a significant difference between at least two of the groups.

The Post-hoc analysis reveals that Equation 7.2 is significantly different from Equation 7.3 and Equation F.6. This finding supports the ANOVA results. However, when comparing Equation 7.3 and Equation F.6, no significant difference is found, suggesting that their means are statistically equivalent.

In conclusion, the ANOVA test provides significant evidence that the means of the three groups are not all equal. Equation 7.2 is notably different from Equation 7.3 and Equation F.6. This may suggest that the factors influencing Equation 7.2 are different from those impacting Equation 7.3 and Equation F.6, or that Equation 7.2 is exposed to a different set of conditions.

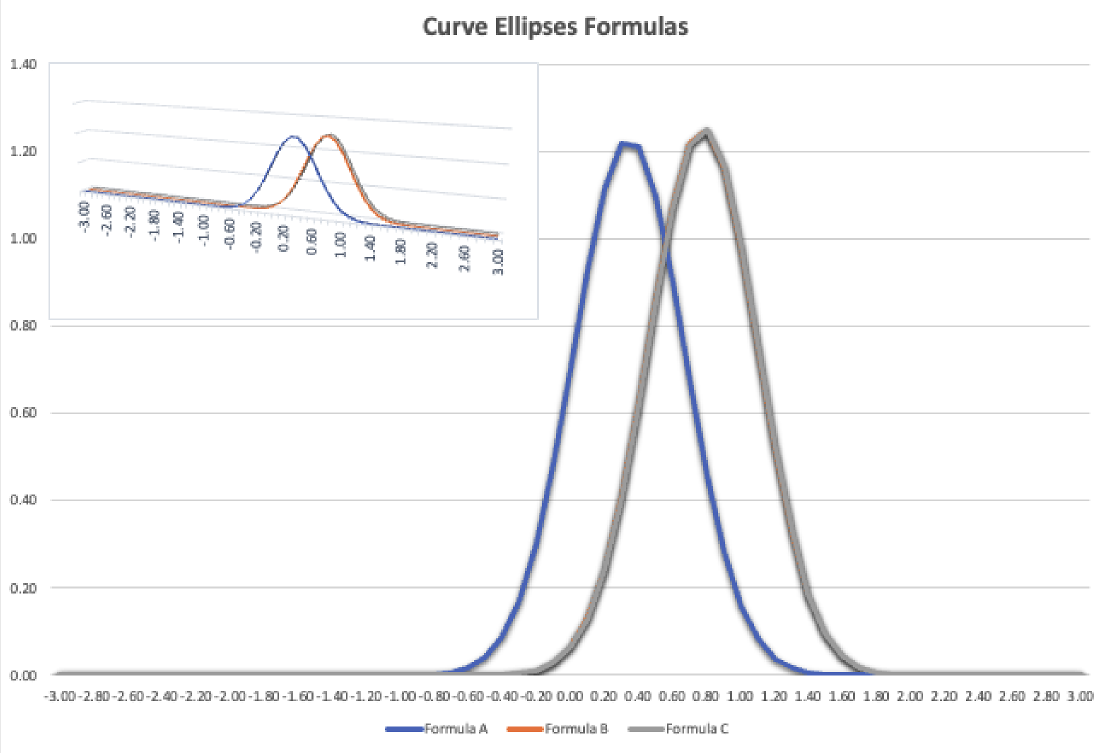


Fig. F.2 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

On the other hand, it was hypothesized that when the major axis of the body shape is three (or more) times longer than the minor one, it is observed that Equation 7.2 has the highest mean value of 0.918, followed by Equation 7.3 with a mean value of 0.542, and Equation F.6 with the least mean value of 0.537. The variances of the three groups are relatively similar, with Equation 7.2 having the highest value of 0.159, and Equation F.6 having the lowest value of 0.123.

The ANOVA test, which seeks to ascertain whether there are significant differences between these three groups, shows a Between Groups sum of squares (SS) of 4.971 and a Mean Square (MS) value of 2.486. The F-statistic is 18.5, and the critical value of F is 3.055.

The p-value obtained from the test is 6.36507E-08 (practically zero when considering typical significance levels), which is less than the commonly used significance level of 0.05. This indicates that the null hypothesis of equal means is rejected, and we can infer that there is a significant difference between at least two of the groups.

In our Post-hoc analysis, it is discovered that group Equation 7.2 is significantly different from the other two groups (Equation 7.3 and Equation F.6). This supports the ANOVA results. However, when comparing Equation 7.3 and Equation F.6, no significant difference is found, indicating that the means of Equation 7.3 and Equation F.6 are not statistically different from each other.

In conclusion, the ANOVA test provides significant evidence that the means of the three groups are not all equal, with Equation 7.2 being significantly different from both Equation 7.3 and Equation F.6. This could imply that the factors affecting Equation 7.2 are different from those influencing Equation 7.3 and Equation F.6, or that Equation 7.2 is subjected to a different set of conditions.

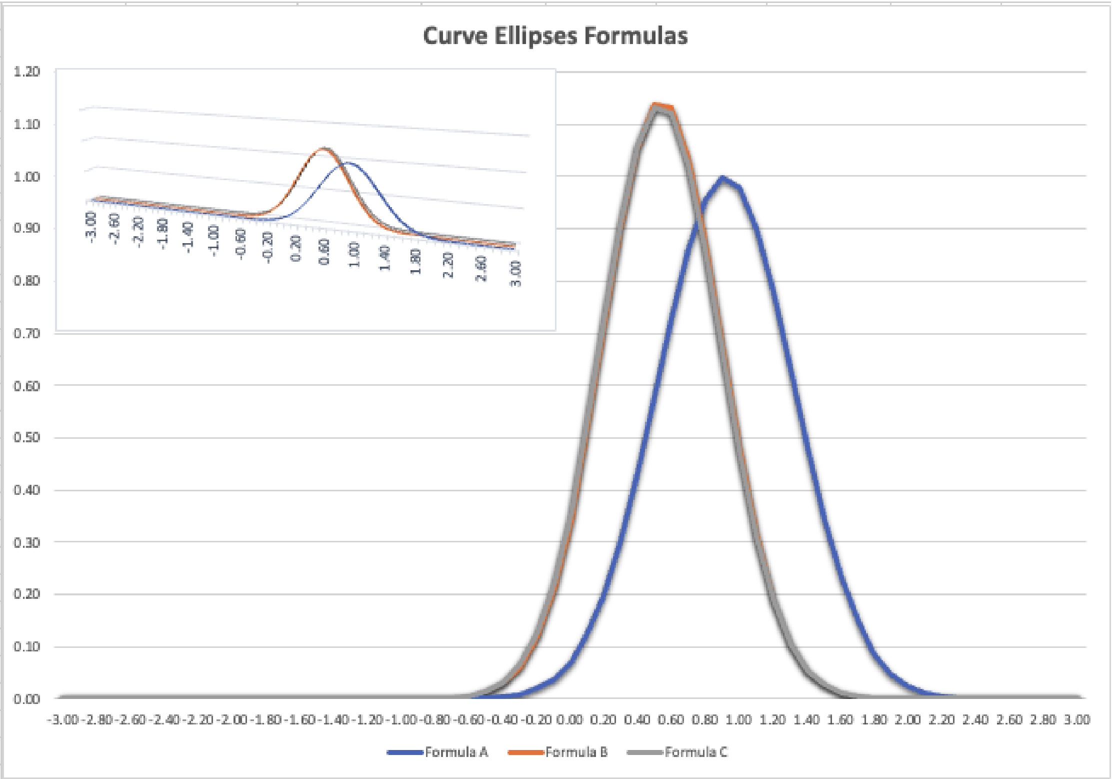


Fig. F.3 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

F.1.2.2 Analysis of Variance (ANOVA) Results and Post-hoc Analysis for Bust Circumferences

This report presents the results of an Analysis of Variance (ANOVA) test performed on data collected from three different groups - Equation 7.2, Equation 7.3, and Equation F.6 for bust circumferences. The context for this analysis is that these mean values represent error differences in centimeters, hence the group with the lower mean signifies higher accuracy.

Based on our collected data from the participants, and by best fitting the human body shape (horizontal slices) by ellipses, we observed that for elliptic profiles such that their major axis were not more than 3 times longer than their minor axis, Equation 7.2

has the lowest mean value of 0.509. Equation 7.3 has a mean value of 0.920, and Equation F.6 has a similar mean value of 0.918. The variances of the three groups are more disparate, with Equation 7.2 having the smallest variance of 0.086, and Equation 7.3 and Equation F.6 having larger, similar variances of 0.188 each.

The ANOVA test, which assesses whether there are significant differences between the three groups, indicates a Between Groups sum of squares (SS) of 4.253 and a Mean Square (MS) value of 2.127. The F-statistic is 13.816, and the critical value of F is 3.078.

The p-value obtained from the test is 4.38645E-06, which is smaller than the commonly used significance level of 0.05. This means that we reject the null hypothesis of equal means and conclude that there is a significant difference between at least two of the groups.

The Post-hoc analysis shows that Equation 7.2 is significantly different from Equation 7.3 and Equation F.6. This finding corroborates the ANOVA results. However, when comparing Equation 7.3 and Equation F.6, no significant difference is found, suggesting that their means are statistically similar.

In conclusion, the ANOVA test provides significant evidence that the means of the three groups are not all equal. Equation 7.2 is notably different from Equation 7.3 and Equation F.6. This may suggest that the factors influencing Equation 7.2 are different from those impacting Equation 7.3 and Equation F.6, or that Equation 7.2 is exposed to a different set of conditions.

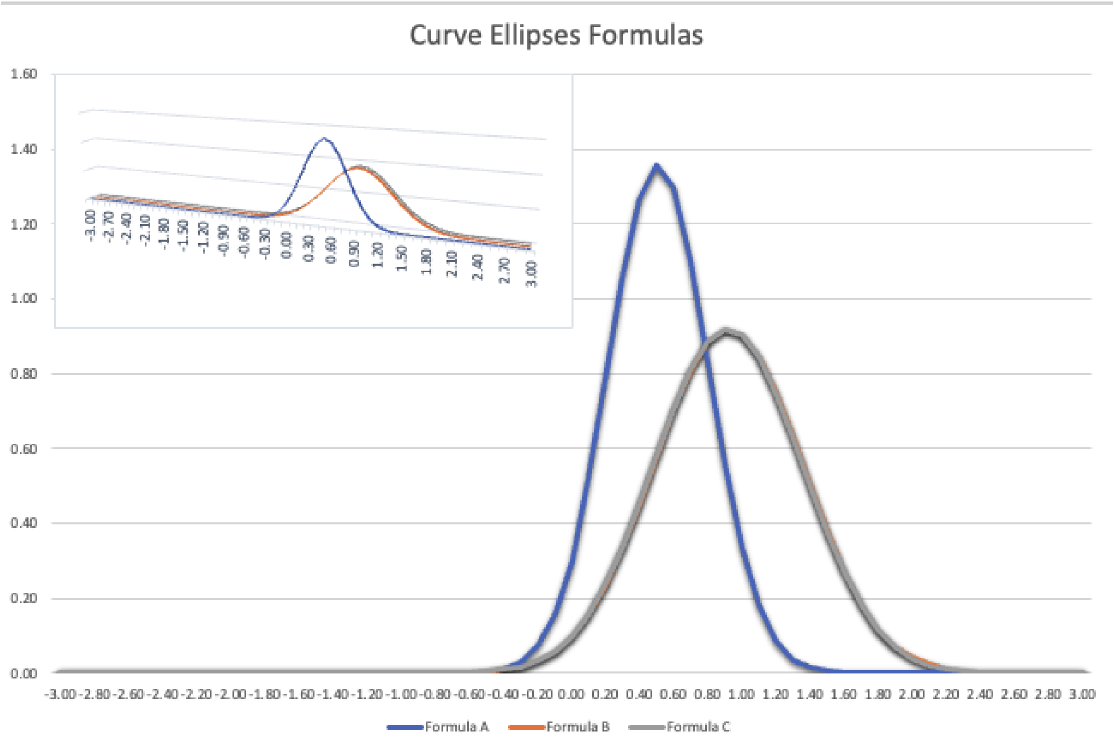


Fig. F.4 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

On the other hand, it was hypothesized that when the major axis of the body shape is three (or more) times longer than the minor one, Equation 7.2 has the highest mean value of 0.882, followed by Equation 7.3 with a mean of 0.507 and Equation F.6 with a mean of 0.506. The variances for these three groups are similar, with Equation 7.2 having a slightly higher value of 0.163, and Equation F.6 having the lowest value of 0.103.

The ANOVA test, used to determine if there are significant differences between these three groups, shows a Between Groups sum of squares (SS) of 3.757 and a Mean Square (MS) value of 1.879. The F-statistic is 15.142, and the critical value of F is 3.074.

The p-value obtained from the test is 1.418E-06, which is far below the standard significance level of 0.05. This indicates that we reject the null hypothesis of equal means, and conclude that there is a significant difference between at least two of the groups.

In our Post-hoc analysis, it is found that Equation 7.2 is significantly different from the other two groups, Equation 7.3 and Equation F.6. This finding supports the ANOVA results. However, when comparing Equation 7.3 and Equation F.6, no significant difference is found, indicating that their means are statistically similar.

In conclusion, the ANOVA test provides significant evidence that the means of the three groups are not all equal, with Equation 7.2 being significantly different from both Equation 7.3 and Equation F.6. This could imply that the factors affecting Equation 7.2 are different from those influencing Equation 7.3 and Equation F.6, or that Equation 7.2 is subjected to a different set of conditions.

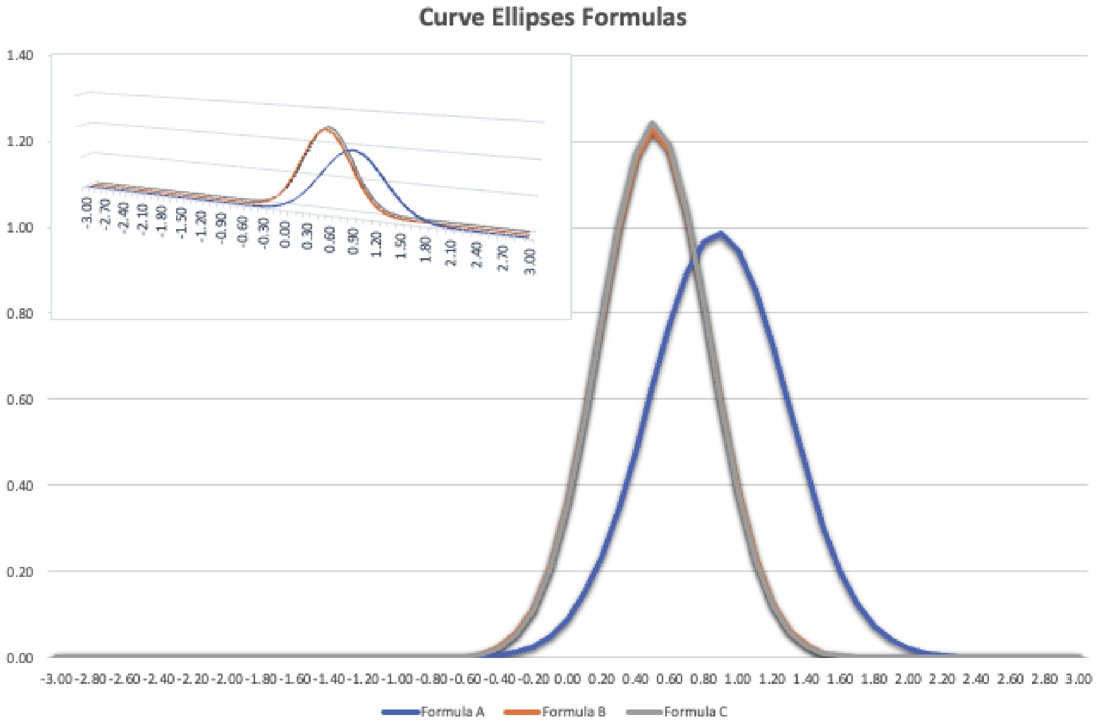


Fig. F.5 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

F.1.2.3 Analysis of Variance (ANOVA) Results and Post-hoc Analysis for Waist Circumferences

The following report provides an analysis of the Analysis of Variance (ANOVA) test performed on three different groups - Equation 7.2, Equation 7.3, and Equation F.6 for waist circumferences. The context for this analysis is that these mean values represent error differences in centimeters, hence the group with the lower mean signifies higher accuracy.

Based on our collected data from the participants, and by best fitting the human body shape (horizontal slices) by ellipses, we observed that for elliptic profiles such that their major axis were not more than 3 times longer than their minor axis, Equation 7.2 has a mean of 0.532, while Equation 7.3 and Equation F.6 have similar mean values of 0.755 and 0.756 respectively. Variance is also comparable across all three groups, ranging from 0.073 in Equation 7.2 to approximately 0.084 in Equation 7.3 and Equation F.6.

The ANOVA test was carried out to determine whether there are significant differences between the means of these three groups. The Between Groups sum of squares (SS) is calculated to be 0.637, with a corresponding Mean Square (MS) value of 0.318. The computed F-statistic is 3.971, and the critical value of F is 3.168.

The p-value of the test is 0.0246, which is lower than the common significance level of 0.05. This suggests that we can reject the null hypothesis of equal means, indicating that there are significant differences between at least two of the groups.

Upon performing the post-hoc analysis, it was found that Equation 7.2 is significantly different from Equation 7.3 and Equation F.6, aligning with the results from the ANOVA. However, there is no significant difference observed between Equation 7.3 and Equation F.6, implying that their means are statistically similar.

In conclusion, the ANOVA test provides evidence that there are significant differences in the means between Equation 7.2 and the other two groups. Equation 7.2 appears to be distinct from Equation 7.3 and Equation F.6, suggesting a different set of influencing factors or conditions for Equation 7.2.

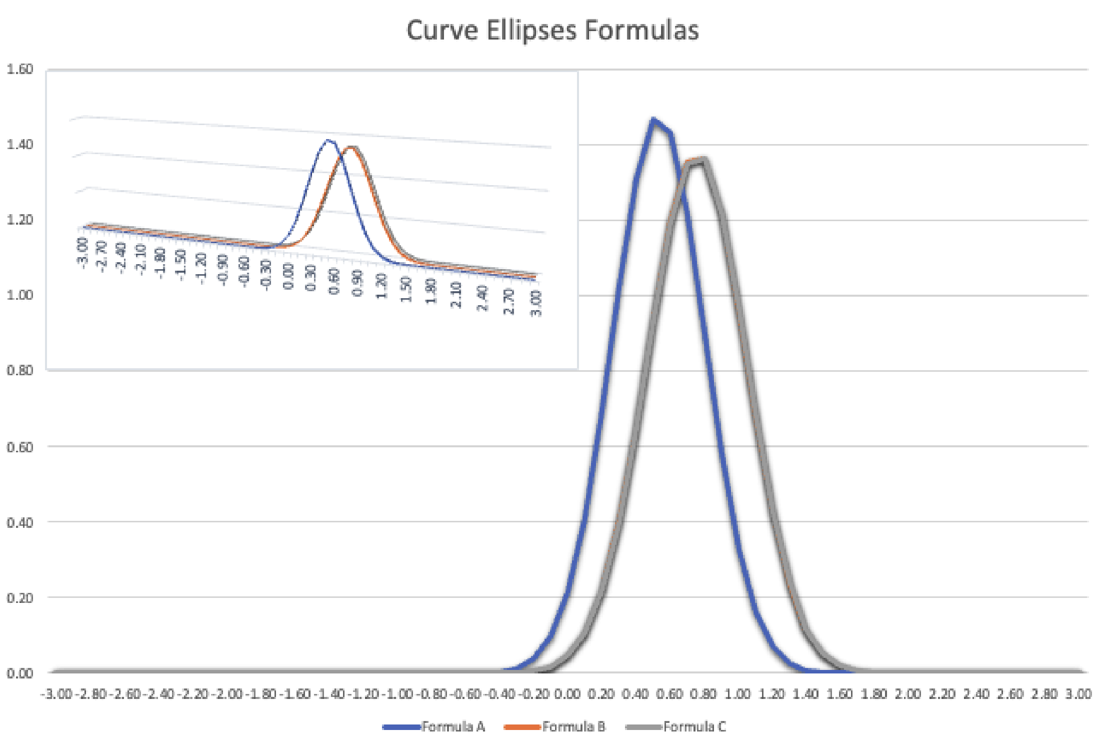


Fig. F.6 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

On the other hand, it was hypothesized that when the major axis of the body shape is three (or more) times longer than the minor one, Equation 7.2 has the highest mean value of 0.952, while Equation 7.3 and Equation F.6 have similar mean values of 0.567 and 0.565 respectively. The variance within each group is also similar, with Equation 7.2 having the highest variance of 0.232 and Equation 7.3 and Equation F.6 having variances of 0.140 each.

The ANOVA test was conducted to determine if there are significant differences between the mean values of the three groups. The results show a Between Groups sum of squares (SS) of 5.864 and a Mean Square (MS) value of 2.932. The F-statistic is 17.222, with a critical value of F being 3.048.

The p-value obtained from the test is 1.4993E-07, which is significantly smaller than the typically used significance level of 0.05. This implies that the null hypothesis of equal means is rejected, indicating that there are significant differences between at least two of the groups.

Post-hoc analysis reveals that Equation 7.2 is significantly different from Equation 7.3 and Equation F.6, which is in agreement with the results from the ANOVA. However, when comparing Equation 7.3 and Equation F.6, there is no significant difference found, indicating that their means are statistically similar.

In conclusion, the ANOVA test provides strong evidence that the means of the three groups are not all the same. Equation 7.2 is significantly different from both Equation 7.3 and Equation F.6. This suggests that the factors influencing Equation 7.2 are distinct from those affecting Equation 7.3 and Equation F.6, or that Equation 7.2 is exposed to a different set of conditions.

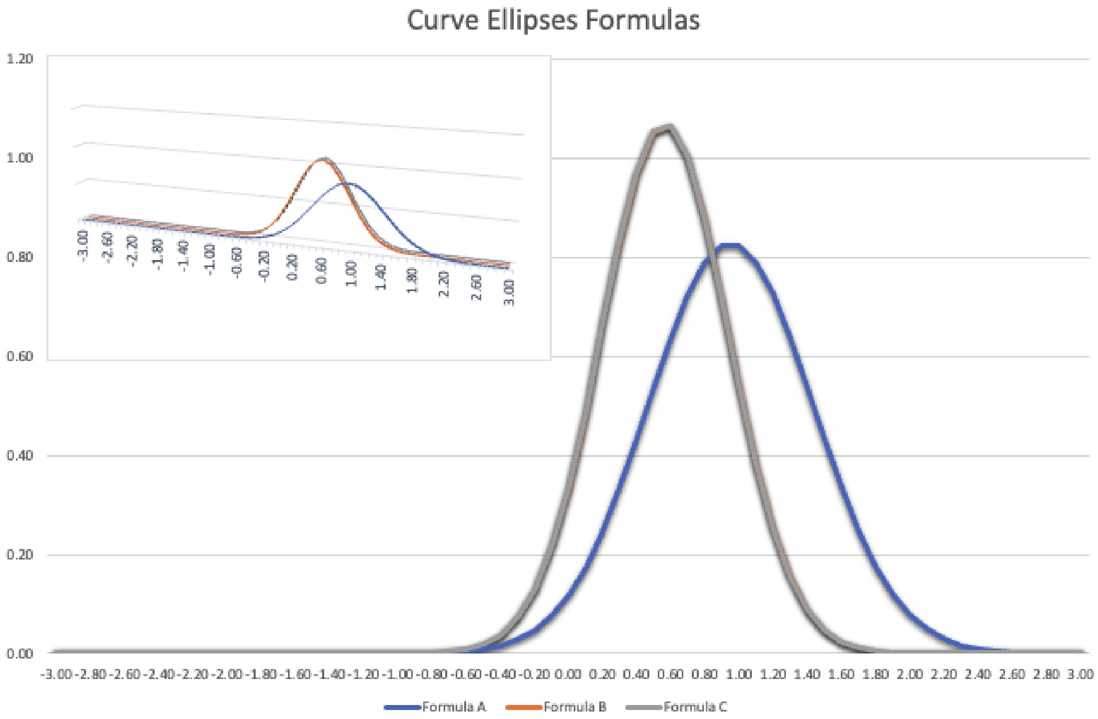


Fig. F.7 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

F.1.2.4 Analysis of Variance (ANOVA) Results and Post-hoc Analysis for Hips Circumferences

This report provides an analysis of the Analysis of Variance (ANOVA) test carried out on three groups - Equation 7.2, Equation 7.3, and Equation F.6 for hips circumferences. The context for this analysis is that these mean values represent error differences in centimeters, hence the group with the lower mean signifies higher accuracy.

Based on our collected data from the participants, and by best fitting the human body shape (horizontal slices) by ellipses, we observed that for elliptic profiles such that their major axis were not more than 3 times longer than their minor axis, Equation 7.2 group has an average of 0.544, while Equation 7.3 and Equation F.6 show higher and quite similar average values of 0.969 and 0.971, respectively. The variance is also relatively comparable across the groups, with a range from 0.282 in Equation 7.2 to 0.549 in Equation F.6.

The ANOVA test was conducted to examine if there are significant differences among the means of the three groups. The Sum of Squares (SS) for Between Groups is 4.243, and the corresponding Mean Square (MS) value is 2.122. The resulting F-statistic for the test is 4.635, and the critical F-value is 3.085.

The p-value of the ANOVA test is 0.0118, which is below the conventional threshold of 0.05 for statistical significance. This indicates that we can reject the null hypothesis of equal means across all groups. There is evidence to suggest that there is a significant difference between at least two of the groups.

Upon further analysis with post-hoc tests, it is established that Equation 7.2 significantly differs from both Equation 7.3 and Equation F.6. This finding aligns with the results of the ANOVA test. However, no significant difference was found between Equation 7.3 and Equation F.6, suggesting that their means are statistically similar.

In conclusion, the ANOVA test results show significant differences in the means of Equation 7.2 and the other two groups. Equation 7.2 appears to be distinct from Equation 7.3 and Equation F.6, implying different influencing factors or conditions in Equation 7.2. Additional investigation is advised to comprehend the significance of these observed differences.

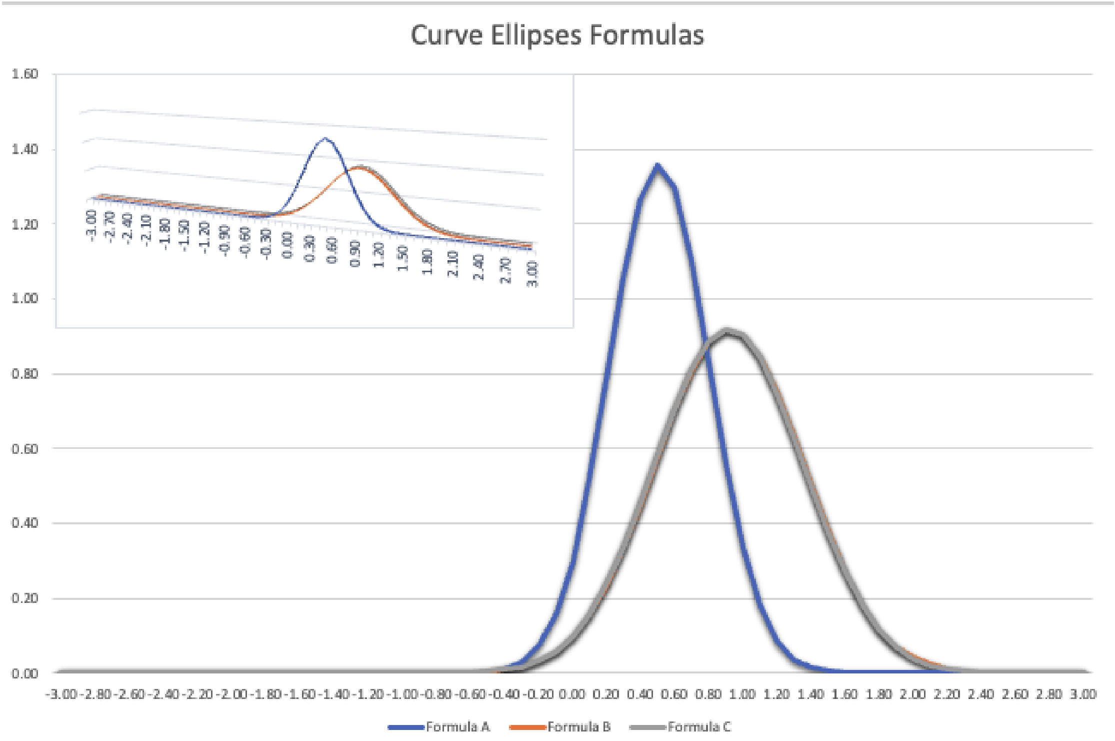


Fig. F.8 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

On the other hand, it was hypothesized that when the major axis of the body shape is three (or more) times longer than the minor one, Equation 7.2 has a mean value of 0.918, whereas Equation 7.3 and Equation F.6 both have similar mean values of 0.506 and 0.507, respectively. The variance across all three groups is also somewhat comparable, ranging from 0.096 in Equation 7.3 and Equation F.6 to 0.177 in Equation 7.2.

The ANOVA test was executed to investigate whether there are significant differences among the means of these three groups. The Between Groups sum of squares (SS) comes out to be 4.849, with a corresponding Mean Square (MS) value of 2.425. The computed F-statistic for the test is 19.689, and the critical F-value is 3.068.

The p-value of the ANOVA test is extremely small (3.63E-08), which is significantly below the conventional significance level of 0.05. This suggests that we can reject the null hypothesis that assumes equal means across the groups. It indicates that there are substantial differences between at least two of the groups.

In the post-hoc analysis, it was discovered that Equation 7.2 significantly differs from Equation 7.3 and Equation F.6. This finding aligns with the ANOVA results. However, there is no statistically significant difference between Equation 7.3 and Equation F.6, implying that their means are statistically similar.

In summary, the ANOVA test reveals significant differences in the means between Equation 7.2 and the other two groups. Equation 7.2 appears to differ from Equation 7.3 and Equation F.6, indicating different underlying factors or conditions impacting Equation 7.2. Further investigation is recommended to understand the implications of these observed differences.

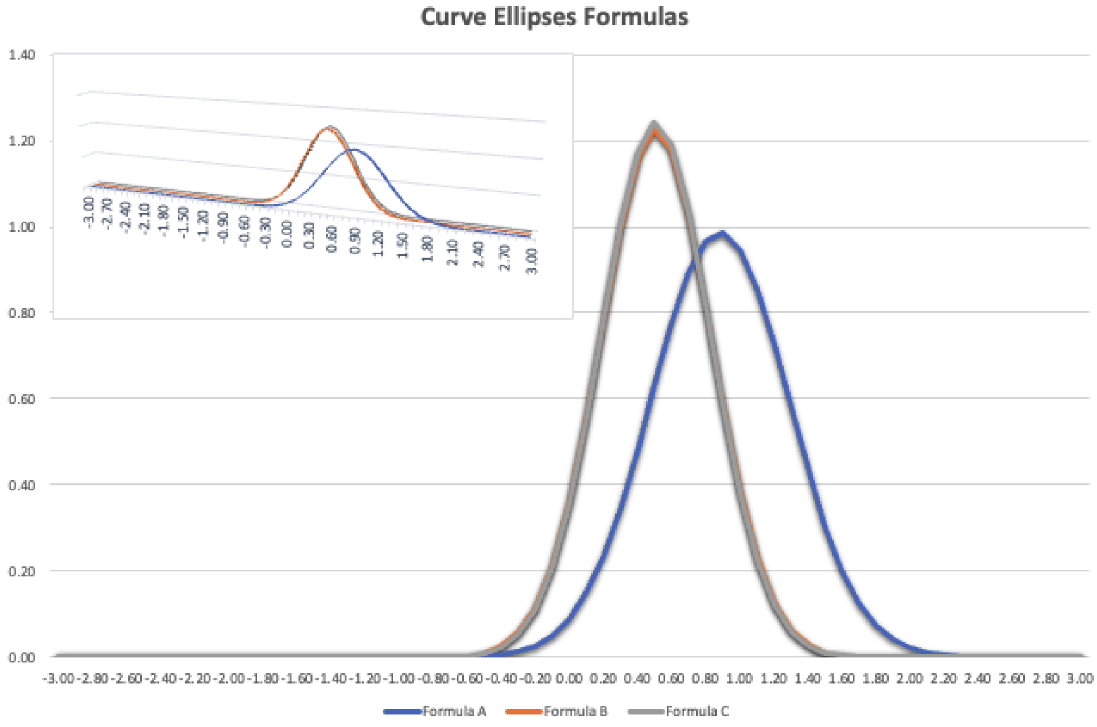


Fig. F.9 displays each formula's stress level on the bell curve Note that Equation 7.3 and Equation F.6 provide almost the same results, which is why the curve only displays a single colour.

Appendix G

Ethical Approval Forms

This thesis, entitled "Designing a Contactless, AI System to Measure the Human Body using a Single Camera for the Clothing and Fashion Industry" has undergone the requisite ethical review and has been formally approved by the Ethical Review Board of the Goldsmiths, University of London, in compliance with the university's ethical guidelines and standards. The approval was granted to facilitate the data collection process integral to the research conducted for this thesis.

For verification purposes and further details, the related documents and approval forms can be located within the supplementary materials of this thesis under the '**Ethical-Approval-Forms**' folder. These documents provide evidence of the thorough and careful consideration given to the ethical implications associated with the research undertaken, ensuring the protection and respect of all participants involved, the integrity of the research process, and the responsible use of the collected data.