

Measuring the Human Body from a Single Camera, with Applications to the Clothing and Fashion Industry

Mohammad MONTAZERIAN, Frederic FOL LEYMARIE
Goldsmiths, University of London, Computing Dept., London, UK

<https://doi.org/10.15221/22.11>

Abstract

Using a single RGB camera to obtain accurate body dimensions rather than measuring these manually or via sophisticated multi-camera or laser-based sensors, has a high application potential for the apparel (fashion) industry.

We present a system that estimates upper human body measurements using a set of computer vision and machine learning technologies. In a nutshell, the main steps involve: (1) using a portable camera (such as with a smartphone); (2) improving the image quality; (3) performing a calibration step; (4) extracting features of the body from the image; (5) indicating markers on the image semi-automatically; (6) producing refined final results.

We experimented with the system on a sample of participants. The results for the upper human body measurements in comparison to the main manual method of tape measurements show $\pm 1\text{cm}$ average differences, which is a good enough result for a number of applications.

1. Introduction

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 digitise measurement capture so that made-to-measure businesses can easily operate online.

Nowadays, especially the younger generations are preferring online shopping. A large and growing amount of clothes is bought via the internet. A majority of orders are returned which proves a real problem for online retailers [7] [12] [14]. Customers will often order the same piece in different sizes to choose the right fit and send back the others. This causes significant costs for the retailer and also has a significant negative impact on the environment.

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, such as 3D scanners based on lasers or multiple fixed camera domes. Such hardware is typically not widely available, not portable, and very expensive.

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].

Several studies have investigated the possibility of automatically estimating human body measurements within seconds from a small number of images. On one hand, we find approaches that require specialized devices (e.g., in-depth camera) to capture 3D images of the human body [9] [19], with hardware still not widely available or current in the public. On the other hand, the use of more common portable camera platforms, such as integrated in smartphones, do not yet provide sufficiently accurate measurements.

Motivated by this current situation, we will propose a simple, yet effective, technique providing a state-of-the-art lightweight measurement system (e.g. based on a smartphone to capture data) that is capable of extracting sufficiently accurate anthropometric measurements to capture the relevant data of the upper human body with applications in apparel/fashion sector.

2. Related Work

Lots of apparel/ fashion companies are pursuing to sell their products online since E-Commerce industry is growing and getting more popular in the last few decades. In their way of doing so, they have faced several issues and one of the main issue of online shopping as mentioned before is fitting issue [16]. To overcome the problem, many research projects have been conducted and one of the best solutions to counter the misfit issue is by estimating the human body measurements from 2D images. However, the prediction is typically for the almost naked body (with tight clothes or underwear). Another group of studies use more complex depth-sensing systems to obtain 3D images that then are used to estimate the body measurements. Below, we present a brief summary of recent research efforts that focus on estimating the human body outer surface measurements.

Xiaohua *et al.* [17] developed an automatic way to extract features points and measure the garment on 3D human bodies. The approach requires a depth-sensing platform, not yet readily available for most online shoppers. It also requires multiple poses to be captured. Another research by Chang *et al.* [9] proposed a dynamic fitting room to allow individuals to visualize a real-time photo of themselves while trying on different digital pieces of clothing. The method is based on the Microsoft Kinect sensing platform, as well as augmented reality technologies. According to head/foot joints and the depth data using two kinect cameras one for taking front and the other one for the side view, the system estimates the user's body height. The system provides sufficiently accurate results (less than a centimeter error).

Chandra *et al.* [8] proposed an approach that estimates human body measurements by utilizing portable cameras commonly found on smartphones. Once the system detects a human body in its field of view, it tracks body outlines. The system then seeks to identify the face, neck and shoulders, as upper body markers. The system has some limitations: users need to set specific environmental conditions (including light, background and type and contrast of clothes), and they must then repeat their measurement capture at least five times.

Francesca *et al.* [10] developed a cloud API, using Amazon Web Services (AWS), that can be easily integrated into other platform to extract up to 24 body measurements by using two photos (front/side view). They use existing open-source algorithm that could be adapted and expanded in online shopping experience including: Tensorflow and PyTorch in collaboration with post-processing function OpenCV and Python. The results that they have provided in their study shows the median difference is below 2.6 cm (or one inch), which is insufficient for commercial applications.

Closer to our own work, Sahar *et al.* [6] presented an approach to estimate human body measurements with smartphone cameras. They use Haar-based detectors that provide markers focused on the upper, lower and full body areas. Difference between markers are then used to estimate body measurements. Their method can only provide very coarse ranges — such as the commonly used in retail classification in "small, medium, large, XL, XXL". The accuracy achieved is not sufficient for practical use.

3. Method

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 calibration to determine the depth of field (distance from the camera focal point to the human body), (iv) extract body features from the image, (v) semi-automatically set a small number of markers, and (vi) by computing difference between markers, we estimate human upper body measurements.

The overall pipeline is illustrated in (figure 1). 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 A-pose for the frontal view capture, while there are no restrictions for the side views (examples to follow). The 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.

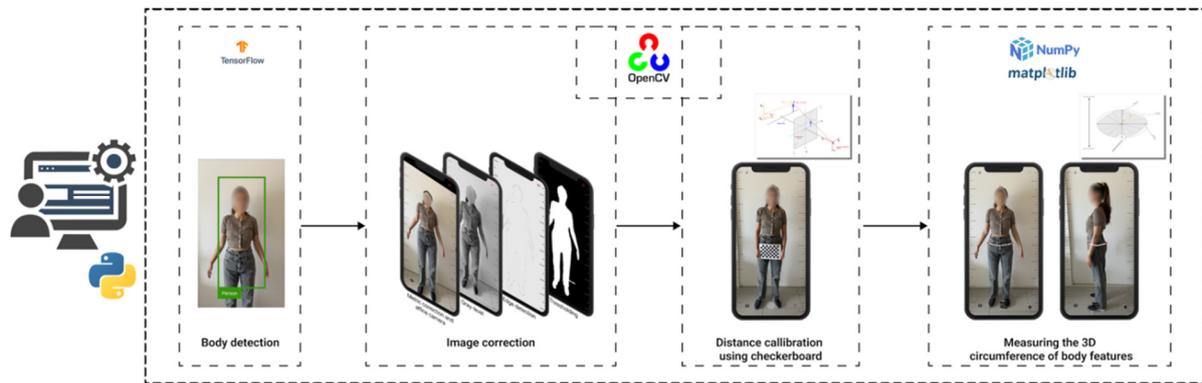


Figure 1: Pipeline - The following flowchart displays every stage to calculate the human upper body measurements

3.1. Body detection

We use a method similar to Sahar *et al.*, which is based on machine learning, to initialise our method, this to accurately locate a human body in an image [6]. This is based on the OpenCV library, using the MobileNet SSD [11] together with a deep neural network (DNN) module.

MobileNet is based on a streamlined architecture that uses depth-wise separable convolutions to build light weight DNNs. MobileNet SSD was chosen over other object detection algorithms as it was designed for mobile and embedded version applications; furthermore, once deployed, it is fast and accurate for single view image processing.

We use a pre-trained MobileNet SSD to differentiate the human body from the surrounding environment (or background) in a single image. Still, the resulting segmentation may include some of the background information near the edges of the body. Also, difficult backgrounds (especially when textured) may create larger bleeding effects at body edges. In our experiments, we achieve good body segmentation where on average 94% or more of the pixels correspond to body locations (when compared to ground truth (hand segmentation)). For comparison, when using other classic methods, such as those based on Haar Cascade detectors [6], we observed difficulties when applying the technique to profile views (side views). Furthermore, such methods give much poorer results than MobileNet when dealing with difficult lighting conditions or textured backgrounds.

In order to collect more accurate pixel data from the upper body area, we need to reduce any potential errors that could creep in on particular views, which can be due to factors like reflections, shadows, low contrast. We consider next how to improve the segmentation results obtained from MobileNet.

3.2. Image corrections

We consider two general categories of problems in this section.

- (i) Image distortions
- (ii) Low-level features enhancement

With a step of affine and metric correction we can improve the visibility and geometry of the various regions into which an image can be partitioned, as well as improve the clarity of image features inside these regions.

Furthermore, we need to: clean the image from various type of noises, enhancing the contrast among adjacent regions and features, simplify the image via selective smoothing and the elimination of features at small scales while retaining other features at desirable scales [13].

In our current pipeline, we proceed as follows: (i) metric correction of an affine camera model, (ii) grey level mapping of an RGB input, together with removal of small image regions and image smoothing [15] [5], (iii) detecting edges [13], (iv) mask operations on matrices and thresholding [5]. Figure 2 illustrates some of these steps of image improvement in the proposed system.



Figure 2: demonstrates the different stage of the image corrections

3.3. Distance calibrating using checkerboard

The calibration process is needed to find camera parameters. 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.

We first define real world coordinates of 3D points using a checkerboard pattern of known size and second find the pixel coordinates (u, v) for each 3D point in different images. For this purpose, we make use of the OpenCV method `findChessboardCorners` (see figure 3).

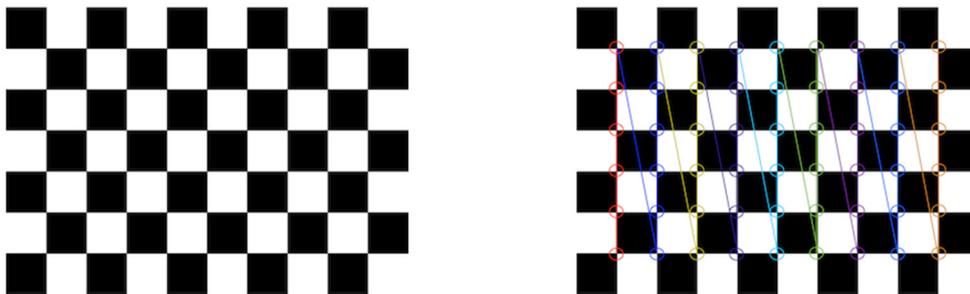


Figure 3: The following checkerboards display the result before and after detecting checkerboard corners.

3.4. Measuring the 3D circumference of body features

The primary goal of our feature extraction stage is to identify particular interest points — we call *markers* — from the detected body. These markers are then used to estimate the measurements of specific horizontal body sections: waistline, chest, hip, shoulders.

We identify markers by pairs, along horizontal body "slices". To do so, we first find a best vertical body line used as an approximate mirror-like splitting axis. Then, we look for right and left pairs of body extremities, i.e. pixel locations at the edge of the body.

We first identify a top central head point — head tip — which we use to center our vertical splitting line. From there we walk along the body contour on one side (left or right) and monitor the slope of a line segment between the head tip and the current body edge point. When this slope goes through a large change in value, we determine a potential marker location. In our experiments, this works well for markers in the neck and shoulder areas. Note that we use the A-pose to determine shoulder markers as, then, the shoulder line (between the pair of shoulder markers), corresponds to the recommended way by ISO (the International organization for standardization) to measure such segments.

Currently, in our system, other markers for the chest and waist measurements require a semi-automatic method where a user moves an horizontal virtual line along the medial vertical axis to identify the useful additional pairs of markers. The automatic detection of such markers should be achieved in the future, e.g. based on machine learning (and this requires additional training data, which we have yet to produce).

Once our set of marker pairs are obtained, we can approximate upper human body (3D) circumferences for each slice, by fitting an ellipse to the data points (using both the frontal and side views). Using an ellipse as an approximate geometric model this way has been validated before [18].

Then, by simply calculating the semi-axes of the ellipse from the two images (front and side), we get an estimate of the circumference of a human body "slice" with respect to the selected marker pair (e.g., waist, circumference). Figure 4 illustrates an example of semi-auto markers around the shoulders-bust-waist-hips line.



Figure 4: Indicating markers (semi-auto): Our advanced computer vision algorithms detect the dressed human body on 2D photos taken with any smartphone on any background and based on the detected body will indicate semi-auto markers. Users are also able to indicate their own markers.

3.4.1. Final measurement step

For each pair of markers, we need to find the actual distance between these points, by converting pixel units to centimeters. We can then fit our measured distances for front (f) and side (s) views in the following equations (adapted from [16]):

$$P \approx 2 \times \Pi \times \frac{\sqrt{(dist_a^2) + (dist_b^2)}}{2} \tag{Equation 1}$$

$$P \approx \Pi [3(dist_a + dist_b) - \sqrt{(3dist_a + dist_b)(dist_a + 3dist_b)}] \tag{Equation 2}$$

As mentioned earlier, the human body horizontal circumferences (slices) are, for a majority of them, approximately elliptic in shape. In particular, to obtain a good measurement of the human body torso, we evaluate it as the perimeter of a best fitted ellipse.

Experimentally, we have determined that to obtain more accurate results we need more than one formula to measure the human body circumferences. For ellipses with semi-major axes which are not more than three times longer than their semi-minor axes Equation 1 provides better accuracy; otherwise, when the ellipse is more squashed, Equation 2 provides better results.

Algorithm 1: Human Upper Body Measurements

Result: Set of Measurements \mathbf{M} , where t^{th} row is the measurement around joint j_t

while *Calibration=failure* **do**

 Call findChessboardsCorners;

if *User Detected* **then**

 Affine correct (convert image to grayscale, smooth it, detect edges);

 Compute principal axes (u, v, w);

 Request users to Draw Circles;

 Get head point;

 Call analyze chessboard;

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 to distance;

 Request user to detect points;

 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 1;

if m_t is available for all T joints **then**

 Save \mathbf{M} ;

 Break;

if *Semi-Major axes is 3 times longer than Semi-Minor axes* **then**

 Use Equation 2;

if m_t is available for all T joints **then**

 Save \mathbf{M} ;

 Break;

Print Perimeter;

4. Result and Discussion

A key to the successful design of combined machine learning models with additional computer vision techniques is a good training set of sufficient scale, with accurate annotation (in our case, pairs of specific markers). In our research, we need a set of data that contains human body images and their associated measurements for each body corresponding to: chest, bust, waist and hip circumferences. There is a lack of good adapted such datasets and we have had to start building our own dataset.

Participants in the dataset are volunteers who self-identified as female and over 18 years old. Thus far, we have collected for training purposes, a relatively small dataset, 55 female body, representing a variety of body shapes and heights. We note that we have been very constrained due to the ongoing pandemic. Nevertheless, our results are promising.

Ground truth references are obtained from each individual by using tape measurements. Then, participants were photographed in different locations with different environmental conditions varying lighting, backgrounds, body posture and distance to camera. Note that each participant performed at least three separate such photometric sessions. The images were taken by using an iPhone 10 (and above) model and a Galaxy S21, as these smartphones are good representatives of the most widely used such devices. We have followed the standard procedure as proposed in ISO 8559-1 [2], ISO 8559-2 [3] and ISO 8559-3 [4] for obtaining our tape measurements (as references), in order to reduce human error.

Table 1: Summary of the differences between our software and tape-measured data for the 55 participants (275 cases in total) included in this study. Every participant performed at least 4 case studies (clothing that has creases, plain/cluttered background, different body posture, variety of lighting conditions and different distance from the camera point of view).

Body Type	Chest	Bust	Waist	Hips
Mean Differences (cm)	0.880	0.927	1.1020	0.862
Median Differences (cm)	0.811	0.811	0.893	0.795
Max Differences (cm)	3.134	3.138	4.432	3.812
Min Differences (cm)	0.001	0.002	0.017	0.015
Standard Deviation (cm)	0.511	0.501	0.561	0.534

The results showing the average differences between measurements obtained via our system and those tape measurements, for all 55 participants are reported in Table 1. Overall, our upper human body measurements show less than ± 1 cm average difference. Also, you can find the error correlation for every individual participants in Figure 5.

It is worth mentioning that the maximum measured differences from ground truth occurred when a participant was wearing clothing that has creases, or when the background was highly cluttered, or when the lighting was not well diffused (creating strong shadows over the body), or when the camera's focal was too close or too far away from the participant (i.e. either less than 0.5m or more than 3m).

The test performed in this study have shown the importance of good quality images being uploaded by the users. Current RGB cameras, generally, produce a high quality resolution for this type of application. Having good instructions, and some examples in a tutorial, help alleviate difficult conditions. In our experience, this can reduce the mean difference for our software up to ± 0.5 cm.

From the designer's point of view, the ability to measure human body circumferences within ± 1 cm average difference from the customer's body shape enables more efficient processing going from purchased to finished products, and it can help a lot to reduce returns and wastes.

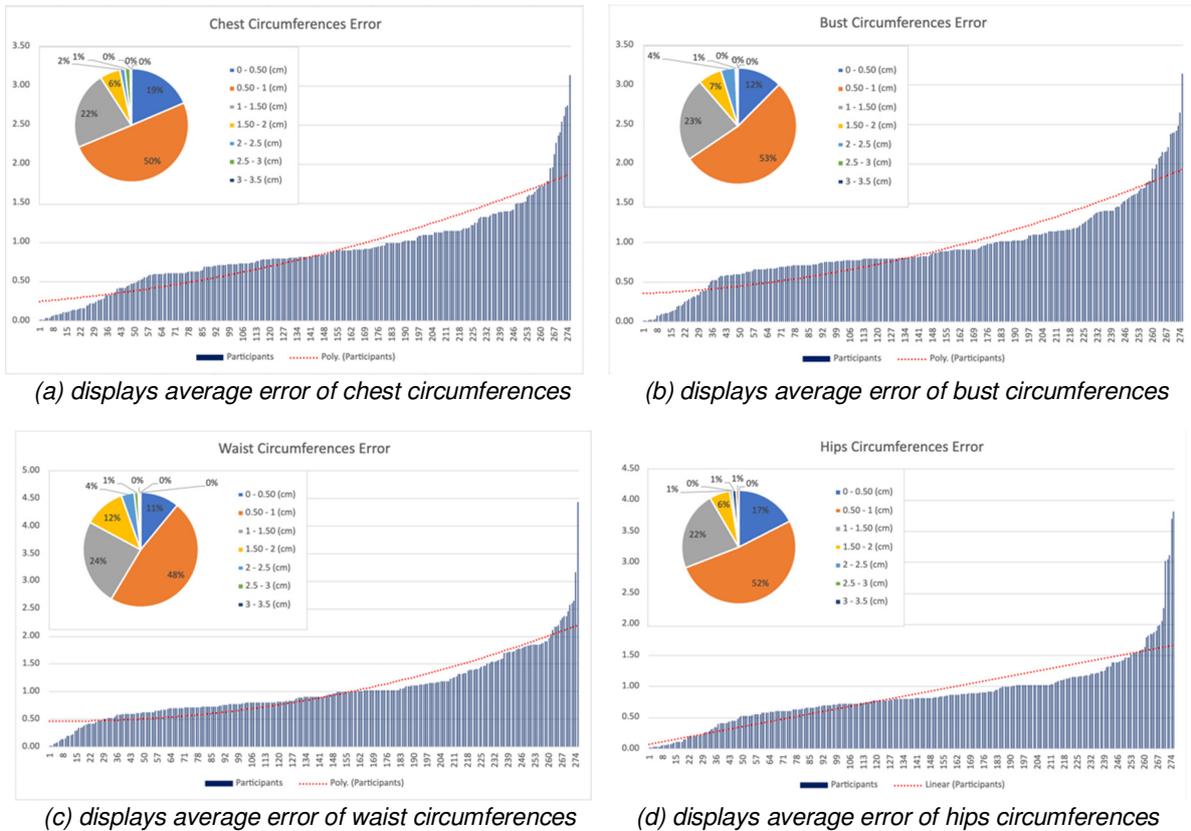


Figure 5: A plot of the error correlation values of the Chest-Bust-Waist-Hips Circumferences in comparison to the tape measurements. Images contain plain/cluttered background, different body posture (Stand-Pose, A-Pose/T-Pose), different lighting conditions, different distance from the camera point of view.

5. Conclusion

In this research, the proposed technique aims to improve and facilitate the experience of E-Commerce apparel/fashion industry with a simple way for everyday people to estimate their characteristic human body measurements from a pair of 2D images taken by smartphones.

We have reported in this paper our results on the upper part of the human body. While Junfeng *et al.* [18] have previously reported that the ellipse is suitable to approximate human body cross sections including the torso and limbs, our results indicate that this method can be used to measure other parts of the human body sufficiently accurately for practical use.

These techniques helped us to improve the state-of-the-art in terms of accuracy to a maximum of ± 1 cm average difference (in comparison to the tape measurements methods) on 55 participants (who self-identified as female).

In the future we are planning to remove the dependency from a calibrating checkerboard and make the marker identification step fully automatic. We shall also create a larger database and test our system on a more diverse human body population.

References

- [1] 3DLook. 3dlookme. [Accessed: 21/06/2021]. [Online]. Available: <https://3dlook.me>, June 2021.
- [2] ISO 8559-1. Size designation of clothes—part 1: Anthropometric definitions for body measurement, 2017.
- [3] ISO 8559-2. Size designation of clothes — part 2: Primary and secondary dimension indicators, 2017.
- [4] ISO 8559-3. Size designation of clothes — part 3: Methodology for the creation of body measurement tables and intervals, 2018.
- [5] Rosebrock Adrian. 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/>, July 2019.
- [6] Sahar Ashmawi, Maram Alharbi, Ameerah Almaghrabi, and Areej Alhothali. Fitme: Body measurement estimations using machine learning method. *Procedia Computer Science*, 163:209–217, 2019.
- [7] Anthony L Brooks and EvaPettersson Brooks. Towards an inclusive virtual dressing room for wheelchair-bound customers. In *2014 International Conference on Collaboration Technologies and Systems (CTS)*, pages 582–589. IEEE, 2014.
- [8] Richard Nathaniel Chandra, Fajar Febriyan, and Theresia Herlina Rochadiani. 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, 2018.
- [9] Hsien-Tsung Chang, Yu-Wen Li, Huan-Ting Chen, Shih-Yi Feng, and Tsung-Tien Chien. A dynamic fitting room based on microsoft kinect and augmented reality technologies. In *International Conference on Human-Computer Interaction*, pages 177–185. Springer, 2013.
- [10] Coette Johnson Francesca Danadoni, Jayne Mechan. “ps for you”: A contactless, remote, cloud based body measurement technology powered by artificial intelligence. In *Proceedings of 3DBODY.TECH 2021 12th Int. Conference and Exhibition on 3D Body Scanning and Processing Technologies*, October 2021.
- [11] Andrew G. Howard, Menglong Zhu, Bo Chen, D. Kalenichenko, Weijun Wang, Tobias Weyand, M. Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv*, abs/1704.04861, 2017.
- [12] Mains Madeline. Consumers to return half of online clothing purchases this holiday season, Dec 2018.
- [13] Petros Maragos. Morphological filtering for image enhancement and feature detection. *The Image and Video Processing Handbook*, pages 135–156, 2005.
- [14] Halilday Sandra. Online fashion returns soar as shoppers lack size info, Mar 2018.

- [15] Linda Shapiro and George Stockman. Computer vision. Pages 145–205. [Accessed: 21/06/2021]. [Online]. Available: <http://lmvicente.com/ee7716/docs/art4ShapiroComputerVision.pdf>, March 2000.
- [16] Bazli Tomi, Mohd Shahrizal Sunar, Farhan Mohamed, Takeshi Saitoh, Mohd Khalid Bin Mokhtar, and Simplicio Michael Luis. Dynamic body circumference measurement technique for a more realistic virtual fitting room experience. In 2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e), pages 56–60, 2018. doi: 10.1109/IC3e.2018.8632636.
- [17] Tan Xiaohui, Peng Xiaoyu, Liu Liwen, and Xia Qing. Automatic human body feature extraction and personal size measurement. Journal of Visual Languages & Computing, 47:9–18, 2018.
- [18] Junfeng Yao, Hongming Zhang, Hanhui Zhang, and Qingqing Chen. R&d of a parameterized method for 3d virtual human body based on anthropometry. Age, 20(30):30–40, 2008.
- [19] Vladimir Kolmogorov Yuri Boykov. Skeleton tracking and body measurements with kinects. [Accessed: 21/06/2021]. [Online]. Available: https://github.com/oveis/Body_Measurement_with_Kinect, Mar 2012.