Towards Fast and Robust Body Measurements Extraction

Mihai Petre
University Politehnica of Bucharest
Splaiul Independenței 313, București 060042
mihai@esenca.app

Cosmin Ciocîrlan
University Politehnica of Bucharest
Splaiul Independenței 313, București 060042
cosmin@esenca.app

Eduard Cojocea
University Politehnica of Bucharest
Splaiul Independenței 313, București 060042
eduard@esenca.app

Traian Rebedea
University Politehnica of Bucharest
Splaiul Independenței 313, București 060042
traian.rebedea@cs.pub.ro

ABSTRACT

The Computer Vision task of extracting body measurements from images has many applications in online shopping, medical, sport & fitness and other fields which require knowing or monitoring body measurements. Ideally, this should be achieved without complex hardware, fast and with high availability. In this paper, we describe a solution which extracts body measurements using any smartphone with a camera and internet access. Using as input two photos – one frontal and one lateral – and the user's height, weight, age and gender, our solution is able to extract more than 100 body measurements with a measurement error less than 5mm, and an inference time of around 5 seconds.

Author Keywords

Body Dimensions Extraction; Body Shape Analysis; Deep Learning; Computer Vision; Body Key-Points Estimation.

ACM Classification Keywords

I.2.10. Artificial Intelligence: Vision and Scene Understanding
I.4.6. Image Processing and Computer Vision: Segmentation
I.4.7. Image Processing and Computer Vision: Feature Measurement

DOI: 10.37789/ROCHI.2022.1.1.16

INTRODUCTION

Finding the perfect size when ordering clothes from online fashion shops is known to be a difficult task. Most shops only publish their products' sizing charts – tables showing correspondence between products' sizes and body measurements in inches or centimeters. Thus, the most challenging part of finding out what size to order is knowing your own measurements. Therefore, many orders end up being wrong, and both the client and the shop can assess that only after the product is received and tried. This causes high return rates, which, in turn, cause profits loss.

The healthcare industry also benefits from remote body measurements extraction. Specialists can consult patients virtually for prosthetics, orthopedic braces or kinesitherapy instruments. They can also monitor their patients body parameters, such as weight, body circumferences, body fat percentage, metrics used by cardiologists and nutritionists. Moreover, MRI and RX scanners can be optimized, using the measurements in the calibration stage.

Accurate human body measurements extraction is a complex problem, tackled in Computer Vision research since late 20th century. The goal in this problem is to extract measurements with a maximum of 5 millimeters (mms) error – the widely-accepted tolerance in tailoring. In the following sections we present an Ensemble Solution, which makes use of traditional Computer Vision techniques, as well as Deep Learning and Statistical Models.

RELATED WORK

Our solution uses techniques such as pose-estimation, semantic segmentation, and depth-estimation. In the following section, the state-of-the-art in these areas will be described.

EfficientNet [1] is a convolutional neural network architecture and scaling method. It can scale depth, width, and resolution uniformly by using a compound coefficient. The coefficients are different from the conventional scales, as they are fixed. The logic behind the compound part is that, if the image is bigger, then the network will need more layers and channels to capture patterns.

The compound scaling method can be generalized to an existing CNN architecture. Choosing a good baseline network is a priority, as the method only enhances the predictive capacity of the base network. The EfficientNet-B0 is based on the inverted bottleneck residual blocks of MobileNetV2 [2].

PyTorch3D [3] is a framework from Facebook Research that handles working with meshes and it is designed to integrate with deep learning to predict and manipulate 3D data. This framework will be used to develop a version of the measurement prediction model, where it will receive the weight, height and the 2 images to generate the mesh of the person, from which more than 200 measurements can be extracted.

In the next few paragraphs, some relevant history of the evolution of anthropometric features extraction will be presented.

As early as 1934, scientists have demonstrated that traditional ways of measuring anthropometric features are prone to errors [4], either because of their human operators, or because of their intrinsic quality. Ever since, people have been trying to grasp control over accuracy and precision of these measurements, which have so long been at the hands of the measurers.

One of the earliest uses of Computer Vision in measuring anthropometric features was attempted by Meunier, P., & Yin, S. in 2000 [5], as an alternative to traditional and three-dimensional methods of measuring the human body. By taking 2 pictures of the subject (frontal and lateral) at the same time on a background populated with landmarks and then applying body segmentation and landmark detection algorithms, they were able to estimate the following anthropometric variables with up to ± 5mm error: stature, neck circumference, chest circumference, waist circumference, hip circumference, sleeve length. These measurements were done on a pool of 349 male and female subjects and compared to measurements extracted by traditional means. This proposal, particularly extracting the features by using only 2 pictures, seems to have inspired the majority of the subsequential approaches. However, when taking into consideration larger pool of subjects and more body measurements, errors tend to rise. To date, there is no known robust body measurement solution, that extracts any body measurement of any person – of any nationality, ethnicity, gender, and age.

When researching this subject, one would be tempted to start with Body Shape Analysis.

Body Shape Analysis means determining the shape, figure, or type of the human body, fully or partially, by taking into consideration pictures, 3D models, body dimensions, geographical, biological, hereditary, or other type of information. This type of analysis is crucial for anthropometry, because, even though there are clues that each of the human body dimensions can be a continuum between some intervals, it is known that the limits of those intervals cannot vary too much, as the human body has its limits, to preserve its shape.

Body Shapes and Somatotypes are two different concepts. The somatotype concept is a type of classification of the human body shapes, popularized in the 1940s by psychologist W. H. Sheldon. The somatotype taxonomy is a popular, yet highly disputed way to categorize the human body into classes like ectomorph, mesomorph, and endomorph. This taxonomy, W. H. Sheldon claims, can differentiate between different psychological and personality traits, shared only by people whose bodies belong to a specific category. In what follows, I will describe each category without detailing the psychological traits.

Even though W. H. Sheldon's claims regarding the psychological traits connected to those body types have been since dismissed, this classification remains the most popular in our society's culture.

Below, we are going to detail some other attempts to body shape analysis and classification, which are more scientifically reliable.

Anthropometric measurements in combination with dimensionality reduction and clustering techniques show promise for partitioning individuals into distinct groups [6].

Most research in this direction falls into three categories:

- body shape clustering using tabular data
- body shape clustering using photos and/or 3D data
- partial body shape clustering – clustering the lower body, the upper body, the head shape etc.

Research in these directions – particularly in partial body shape clustering and whole-body clustering using tabular data – usually build up on the somatotypes, do not explicitly describe their processes, deliver inconsistent data, or check all three problems.

One research paper that stands out is [6], that focuses on clustering body dimensions data of New Zealand Defense Force soldiers. The dataset used in this paper is made of 84 anthropometric measurements of 1003 participants – 212 females and 791 males –, data also present in NZDFAS dataset. In this paper, the authors applied PCA to find the most important variables for clustering. They concluded that, for both male and female upper body, the most important variables for clustering are the body height and the waist circumference, whereas for the male lower body, the most important variables for clustering are inseam length and the waist circumference, while for females, the variables are inseam length and maximum hip circumference. Finally, they used a combination of two-step and k-means clustering to derive cluster characteristics.

Two-step cluster analysis is a technique developed by Punj and Steward [7] and firstly published in 1983, and then further developed by Chui et al. [8] in 2001. This technique is an exploratory tool designed to reveal clusters within a dataset that would otherwise not be apparent. It uses a likelihood distance measure which assumes that variables in the cluster model are independent. Furthermore, each continuous variable is assumed to have a normal distribution and each categorical variable is assumed to have a multinomial distribution. Empirical internal testing indicates that the procedure is fairly robust to violations of both the assumption of independence and the distributional assumptions, but you should try to be aware of how well these assumptions are met.

To determine which number of clusters is "best", each of these cluster solutions is compared using Schwarz's Bayesian Criterion (BIC) or the Akaike Information Criterion (AIC) as the clustering criterion.

In this research paper, the two most important variables, as defined by the PCA step, were used in the Two-step cluster analysis part. By minimizing the AIC and BIC, the optimal number of clusters for females was discovered to be 6, and for men was discovered to be 10. Finally, the authors used k-means to cluster the individuals.

This research paper cites [9], which shows, on a rather small dataset of 382 men and 391 women, that two-step cluster analysis can detect clusters within an anthropometric dataset, based on different body types.

PROPOSED SOLUTION

We present a novel approach of the human body measurement problem, using an Ensemble Model containing Computer Vision (CV) techniques, Deep Learning (DL) models, classic Machine Learning (ML) and Statistical Methods, that extracts body measurements with an error of maximum 5 mms, by only using some basic user information – weight, height, age, gender – and by taking 2 images – frontal and lateral.

The Ensemble Model consist of two principal solutions: one based on Machine Learning and Statistical Methods, and the other based on Deep Learning and Computer Vision techniques.

The Machine Learning and Statistical Methods solution consists of a double-input Neural Network, that is fed raw data and a Multivariate Gaussian's predictions on the raw data and that outputs an array of more than 50 predicted body dimensions. This solution does not require pictures and the predictions are mean to serve as anomaly detectors, correctors, and failsafe for the Computer Vision solution. Moreover, this solution has been deployed as a standalone fit predictor – sizing recommender based on predicted body measurements.

The Computer Vision solution consists of a complex combination of different models, one for pose estimation, one for segmentation, one for depth estimation and another one for extracting the actual dimensions based on the key-points, body binary mask, weight, and height. The pose estimation model is based on PoseNet [10], developed by Google and published as an Open-Source Solution. The semantic segmentation model has U2-Net [11] as a backbone and was trained specifically on a dataset containing only images with humans. The depth-estimation model was trained on a dataset containing images with rooms, thus predictions on a close picture with a person would have a relatively high accuracy. The final model receives the 2 original images as input, the 2 resulted masks, the 2 depth-estimated images, the key-points for both images, height, and weight. The output is an array of over 100 body dimensions.

By extracting these 100 body measurements, we can then create the user's 3D Avatar, and then further extract any dimension, regardless of its specifications.

To ensure that the users are correctly positioned in the frame to take the photos, we have built an in-house Positioning Model, that runs inside the UI in the browser. We describe it below.

POSITIONING MODEL

This model is based on [10] and extracts 17 body key-points, that the model further uses to verify whether the user is fully visible in the frame, whether they have correctly positioned their phone, at the height of their hips and to further guide them to raise their hands, turn around and overall act as a failsafe that ensures correctly taken photos. The guidance is done both visually, and by playing audio snippets.

There are 2 main solutions for creating a real-time experience during the positioning phase, first one would be to create a live stream between the user's device and a server, where on the server would run the same positioning code, thus not relying on the user's device for processing power. But this process would be dependent on the internet speed, which happens to be relatively bad in most of the places around the globe and it would also need a lot of processing power on the server side (or multiple servers) in the case of having simultaneous connections. This approach currently runs at a maximum of 15 fps, given the best-case scenario.

A second solution would be a very light-weight model that could be compatible with most of the devices from nowadays and that would work in real-time. In this approach, the model is loaded on the device's processor, which could take up to 15 seconds, depending on the type of the processor, and how new the device is. After the model is loaded, though, the key-point prediction is done in less than 40ms, thus resulting in a consistent 20fps performance, which can be considered real time.

Having both video and audio guide for the positioning solution, creating a real-time version affected the audio guidance, so that it would play the same instruction on and on, or it would play more than one instruction at a time, which would certainly confuse the user. In this situation, the audio system had to be changed too, by creating a custom data structure that would hold and play only one audio file at a time, depending on the priority. For example, the guidance will stop any instruction if the user is in the correct position, and it will wait for the image to be taken before continuing with the other instructions. The positioning process is mainly built as a complex state-machine.

GATHERING TRAINING DATA

Both solutions need some anthropometric datasets for training, validation, and testing. Since the beginning of this project, we have analyzed 4 datasets that we have used throughout our work: ANSUR & ANSUR II, CAESAR,

and a proprietary dataset, gathered by us in Costinești, Romania.

The **Costinești dataset** was gathered by our team on a 3-day trip to the seaside, in the summer of 2021, organized for data collection. The location was chosen so that the people posing would be comfortable in wearing only their swimsuit, the situation improving the measuring process (the accuracy of the measurements should be higher if the images do not contain noise - such as baggy clothes). The purpose was to build a dataset of as many people as possible, from which we would measure 12 dimensions: weight, height, neck circumference, chest circumference, waist circumference, bottom circumference, both biceps circumference, both forearms circumference, both wrists circumference, both thighs circumference, both calves circumference and both ankles circumference.

The steps of the measurement process are the following:

1. The person is weighted, and their height is measured.
2. One team member is measuring 12 body parts of the person, using a tape measure, while another team member is writing them down.
3. The person is photographed using a specifically built app (that would automatically take the pictures when the person positions themselves correctly).

The problem encountered during this data collection trip was that the shorts worn by the male participants would interfere with the thigh measurements (the shorts being too loose). Besides this, it was hard to find interested participants and it would take a relatively long time to get through the entire measuring process, because it had to be done thoroughly.

ANSUR [12] and **ANSUR II** [13] are some of the most comprehensive anthropometry datasets publicly available. Created by the US Army, they contain over 100 measurements of approximately 10 000 reserve soldiers. ANSUR was published in 1988, while ANSUR II was published in 2014. The main objective of these datasets was to extract body-size information to have a guide of design and sizing of clothing and protective equipment, while providing diversity in data when it comes to race, gender, and age. Moreover, both datasets come with reports attached, which provide information about data acquisition and, most importantly, the exact body position where the measurement was extracted from and indications on how the subject should position themselves for the measurements. On top of these comprehensive descriptions, the report also contains photos with each measurement pointed out on a sample human body.

CAESAR - Civilian American and European Surface Anthropometry Resource - was designed to provide researchers with the most current measurements for today's body. This dataset was developed because of a comprehensive research project that brought together representatives from numerous industries including apparel, aerospace, and automotive. CAESAR began as a partnership between government and industry to collect and organize the most extensive sampling of consumer body measurements for comparison. The project collected and organized data on 2,400 U.S. & Canadian and 1,900 European civilians and a database was developed. [14] The CAESAR database contains anthropometric variability of men and women, ages 18-65. Representatives were solicited to ensure samples for various weights, ethnic groups, gender, geographic regions, and socio-economic status. The study was conducted from April 1998 to early 2000 and includes three scans per person in a standing pose, full coverage poses and relaxed seating pose. Data collection methods were standardized and documented so that the database can be consistently expanded and updated.

BODY MEASUREMENTS EXTRACTION MODEL

Both multi-input models were created using the Keras Functional API. The multi-input model was invented to conquer the need of having more information formed of mixed data, because there are some cases where it is needed to have more than just image or numeric data. Alternatively, there are cases where it is needed to have multiple inputs from the same type of data - such as more than one image in case of super-resolution, used for example in extracting more information from MRI scans, used for an accurate assessment of cardiovascular physiology.

First Version

The first generation of the model is constructed as a multi-input classification model. The 2 types of input data will be a simple image (the frontal picture received from the widget) and the weight and the height of the person in the picture. The output will be one of the following 10 size classes: [“xxs”, ”xs”, “s", “m", “l", “xl", “xxl", “xxxl", “xxxxl", “xxxxxl”]. The dataset used for this generation and for the next is formed of the 115 measured people described in the earlier section.

The VGG19 part of the model receives as input the frontal picture taken with the widget, where the user is forming an A pose. This layer from the model is meant to extract all the relevant visual information from the image and might replace segmentation, depth estimation and the pose-estimation models, thus basically recreating the human contour and estimating the body shape that would fit the correct size. The AutoEncoder will receive the height and the weight of the user. The 2 models are then concatenated and on top of them are added 3 more Dense layers with the last of them giving an output size of 10 (number of classes used in the classification). Inside the VGG, the activation functions used are ReLU, and in the AutoEncoder we used both ReLU and sigmoid activations.

Second Version

The second generation of the model would receive 2 images as input in the VGG layer, the frontal, and the side pictures, but keeping the same architecture (only the input layer is changed to handle the updated input size). The images are concatenated one on top of the other forming a 6-channel picture. The AutoEncoder will maintain the same input form and architecture, still receiving the height and the weight of the subject.

Third Version

The third version is being tested on the same dataset with 3 sets of images: the original images, the binary masks from the segmentation and the depth-estimation outputs, along with the weight and the height.

Its architecture is more complex than the other versions, having 3 VGGs in composition instead of 1. Besides this, the key-points coming from the positioning model are added along with the height and the weight. Having more information about the environment, the model should behave better under those circumstances. The only inconvenience about this model would be that the pose-estimation, semantic segmentation, and depth-estimation will be used as they "are", meaning that they will not be improved in the process, but will only output data for the model. They will basically be a "black-box".

Current Version

The main difference between the current version and the earlier ones is that this one is trained on the 3D CAESAR dataset. Having more data to train the models on is a huge improvement, even for the earlier generations of models, which can also be improved by training on these datasets.

The VGGs from the third version were replaced by a single EffiecientNetB0 model and along with an AutoEncoder, that receives the height and the weight of the user, will create the multi-input model. The results of this iteration tend to be better by having an accuracy of 90% (in the classification scenario).

The double-input model that uses tabular data and no pictures was constructed with the same idea: classifying subjects into the same 10 output sizes. The first input of this model is the raw tabular data containing the subject's information and their body measurements, plus their body type – a categorical feature calculated by us. The second input is made up of the Multivariate Gaussian's predictions on the subject's raw data. The input would be concatenated in a third dimension to fit into a 3-channel input. Next, there are 2 fully connected layers, one with 32 neurons and one with 64 neurons, each followed by a Dropout layer with 0.5 dropout rate. The number and size of the layers is subject to change following ongoing experiments.

Preliminary results of this iteration are promising: the double-input neural network seems to correct the predictions of the Multivariate Gaussian, having an accuracy of approximately 70% in the classification scenario.

Research on a Different Approach

Having a large dataset of 3D meshes of people with 72 of their respective measurements, it tends to open a discussion about creating a different model, such that would receive the same input as the initial model (weight, height and the 2 images - front and side) but will output a mesh of the person. From the resulted mesh, any body measurement can be extracted, using a method of computing the distances between vertices. This type of model would be similar to a GAN architecture, where it would generate the body mesh based on the input.

The idea would be to split the dataset into the 2 genders, male and female, and train 2 GAN models for each one of them, thus increasing the accuracy. This would be great in the current situation where there is plenty of data (more than 2000 meshes from each gender). The approach will start from the idea used in generating point-clouds from images [14], most of them representing simple objects, such as chairs, planes, or guns. In the current situation, the model will receive as input 2 images, which would increase the information received, thus creating a more accurate 3D model.

After the mesh will be generated, all the measurements will be extracted using the vertices and ignoring the missing ones, we can still extract more than 200 important dimensions of the body, like circumferences - chest, waist, neck, bottom - or linear dimensions, such as leg length, arm length, neck height, crotch height and many others. There are endless possibilities in computing those measurements, as we could calculate the distance from any vertex A to any vertex B.

There could also be the case where some meshes would not correspond to the images and we must deform them to match the person's body. The method used is called "Free-form deformation".

FREE-FORM DEFORMATION

Free form deformation [15] represents a method that a mesh could be deformed in any way by using control points and scale or translate them. This method can be used to deform a baseline model to reach a form like the user's body. This can be done by modifying a humanoid model and scale it to reach the predicted body dimensions (received from the measurement model).

The process of free-form deformation begins with loading a mesh similar to the user's body. For the base there are 3 categories of human models: short, medium, and tall, and for each category there are 3 body types: slim, regular, and fat. In total, there are 9 models from where the algorithm can choose. Based on the user's dimension it will pick the most similar one and that one will be used as a baseline for

the free-from deformation process. After that, different body parts are scaled to match the measurements, starting from the torso, and then the upper body, legs, and hands.

First Version

The first version of the algorithm would select all the vertices in an interval and deform them using the Bernstein method. The steps would be to determine the min and max values of the vertices on the 3 axes, after that we would generate the control points based on the body part we would like to deform, and a final step would be to iterate through all the vertices that are in the selected interval and alter them using the control points and the Bernstein equation. We could also do a smoothing method, that would save the mean between the new deformed value and the old value of the vertices, so that there would not appear huge changes in the body.

The problem with this first method was that it would not have smooth edges between the body parts, so that they would appear deformed unnaturally, like in the left side of Figure 2.

Second Version

The smoothing problem presented in the first version was fixed by creating an algorithm that would split the body part into many smaller pieces and would deform them according to the position in the selected interval. For example, if we would have to deform the waist, we would select all the vertices, starting from the lower-waist level and ending below the chest level and we would gradually deform each level until it reaches the expected deformation value needed. The right side of Figure 2 represents an example of the difference between the two versions.

Also, for deforming the arms and the legs, we would have to rotate and translate the body for the body part we want to deform to be in the origin, so that the algorithm will perform correctly.

Current Version – Full Body Deformation

The current version of the algorithm basically uses the second version with a slight of performance improvement. It represents an iterative deformation of each body part at a time. In the left side of Figure 1, one can observe the belly deformation of the mesh, while in the right side of Figure 1, the deformation of the entire human body can be observed.

VIRTUAL TRY-ON

Having a 3D model similar to a user's body, there is a small step until we can represent clothes on the resulted mesh. As discussed in the beginning of the free-form deformation method, having 9 body-shapes should be enough to represent any body type, but we could have a single model of a T-Shirt and we could mold it to fit the resulted body mesh of the user, thus creating a virtual try-on experience, as seen in Figure 2.

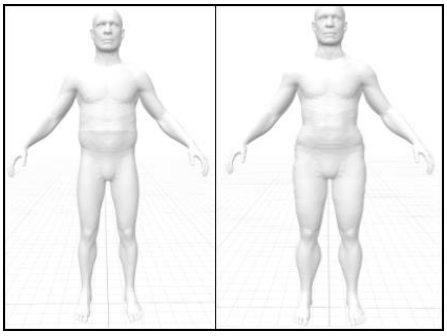

Figure 1 - Full Body FFD. Left: Belly deformation; Right: Multiple body dimensions deformation

Figure 2 - Virtual Try-On demo, using the deformed mesh.

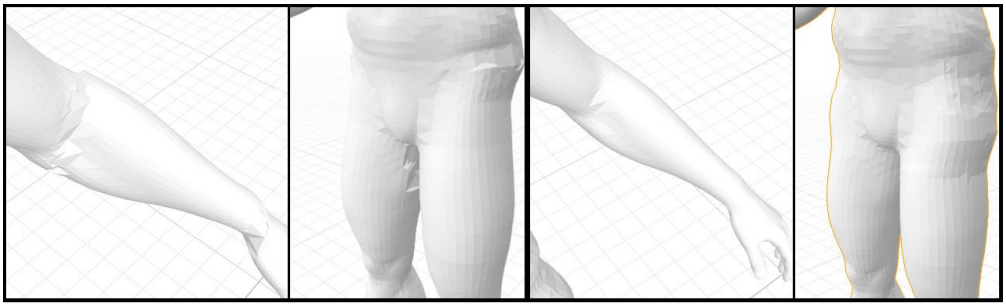

Figure 3: Comparison between different FFD implementations. From left to right: FFD on the forearm and on the thighs, one can observe that the edges between body parts are not smoothed out; 2nd FFD iteration, major improvements in smoothing out edges

RESULTS

The evaluation of the measurement will only be related to the accuracy of the resulted dimensions of the users in environments where background, lighting and clothing can vary. The pose will not vary as the poisoning will always be the same, no maJer the situation. Plus, evaluating the positioning side of the widget will cover two aspects: prediction time and prediction accuracy.

Firstly, we are going to evaluate the positioning server.

Testing was performed on two different versions of ResNet50 PoseNet models: first one uses an output stride of 32, 1 quant byte and 200x200 input resolution; second one uses the same output stride, 4 quant bytes and 480x640 input resolution. The devices used were a Samsung Galaxy S21 Ultra and an iPhone X, both with relatively bad internet connection (10 ping, 8Mbps download speed, 5Mbps upload speed). Table 1 shows PoseNet transmission and processing speed on the S21 Ultra, while Table 2 shows the same on the iPhone X.

#	1st version	2nd version	Current version
1	353.856 ms	502.08 ms	46.445ms
2	382.997 ms	482.21 ms	52.027ms
3	376.940 ms	458.10 ms	51.901ms
4	391.269 ms	491.40 ms	48.341ms
5	369.679 ms	535.43 ms	47.072ms

Table 1 - performance of PoseNet on S21 Ultra

An observation taken during testing was that the positioning algorithm performs better on iPhones. The prediction time would be the same as the server does not change, but the total time is smaller, which can indicate the fact that the compression of binary images is better, or the transmission uses a more efficient algorithm.

#	1st version	2nd version	Current version
1	327.38 ms	509.31 ms	41.40 ms
2	300.445 ms	381.38 ms	42.12 ms
3	307.641 ms	328.11 ms	41.88 ms
4	337.525 ms	339.25 ms	41.76 ms
5	380.376 ms	468.88 ms	41.29 ms

Table 2 - performance of PoseNet on iPhone X

As it can be observed from both Table 1 and Table 2, the current version, running an in-house modification of PoseNet, directly on the device, runs considerably faster. Thus, it can be concluded that this version runs in real time.

The measuring algorithm will be evaluated by comparing real measurements with predicted measurements of the same subject. The prediction time will be compared from the different versions of the algorithm.

The whole purpose of the algorithm is to output accurate dimensions to various body parts of the human, eventually used for creating custom pieces. In Table 3, two subjects will be measured with the tape and the results will be compared to the predicted ones. The prediction and the actual measurement were done in a matter of minutes, thus short-term body changes do not apply here.

Body Dimension (girths)	Subject 1		Subject 2	
	Actual	Predicted	Actual	Predicted
Chest	103	**102.67**	106.50	107.65
Neck	38.5	**39.00**	40.50	**40.59**
Up. Waist	89	**88.97**	96.00	**96.31**
Waist	89.5	90.12	97.00	**96.80**
Lw. Waist	94.5	**95.33**	98.00	95.56
Bottom	107	105.40	105.00	**105.47**
L. Biceps	35.5	34.0	33.00	31.58
R. Biceps	35.5	34.6	34.00	33.38
L. Forearm	28.5	**29.05**	28.00	27.00
R. Forearm	28.5	**28.92**	29.00	**28.85**
L. Wrist	18.00	**18.25**	18.00	16.46
R. Wrist	18.00	**17.63**	18.00	**17.59**
L. Thigh	56.00	59.09	58.50	59.19
R. Thigh	56.00	59.19	59.5	**60.01**
L. Calf	39.00	34.31	37.00	30.69
R. Calf	39.00	34.34	39.00	32.93
L. Ankle	26.00	27.42	25.00	22.03
R. Ankle	26.00	**25.73**	24.00	**23.79**

Table 3 - Actual vs. Predicted results on real humans

The measured subjects are 183 cm tall, weighting 85 kgs, and 184 cm tall, weighting 87 kgs, respectively.

From Table 3, it can be observed that, although there are improvements to be done, our body measurement model measures the human body consistently and with an error of under 0.5 cm for many dimensions.

CONCLUSIONS AND FURTHER RESEARCH

The proposed solution aims to remove any third-party model from the initial solution (pose, segmentation, depth-

estimation), by embedding them into one larger architecture. The goal of the initial research study regarding the improvement of the positioning side of the widget was reached and the next steps would be to improve the accuracy of the extracted dimensions.

In the next generation of models, there is expected a working GAN model that would generate the body mesh (without any missing vertices), from which the extracted measurements would have a maximum of 5mm error, on which a user could see how the clothes would actually fit their body, thus creating the ultimate real-time measuring experience.

Ensemble, multi-input models seem to perform best on this kind of task, where the diversity of the input data is so large. Thus, when it comes to future work, we are thinking of building a more complex multi-input ensemble model, trained on both tabular data and pictures, plus some more feature-engineered data.

Regarding body shape analysis, there seems to be two main promising avenues of research: the first would be in-depth Principal Component Analysis on the body measurements, extracting a number of features that explain most of the dataset's variance and using them to find clusters in data; the second avenue of research would be to use two-step cluster analysis to determine the number of clusters, and then to use some other clustering technique to actually discover the clusters. Finally, one would build a classification model that would be able to pin-point what cluster should an individual be part of.

REFERENCES

1. Koonce, B. (2021). EfficientNet. In Convolutional neural networks with swim for tensorflow (pp. 109-123). Apress, Berkeley, CA
2. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear boilenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520)
3. https://pytorch3d.org/, last accessed 20 June 2022
4. Davenport, C. B., Steggerda, M., & Drager, W. (1934, February). Critical examination of physical anthropometry on the living. In Proceedings of the American Academy of Arts and Sciences (Vol. 69, No. 6, pp. 265-284). American Academy of Arts & Sciences.
5. Meunier, P., & Yin, S. (2000). Performance of a 2D image-based anthropometric measurement and clothing sizing system. Applied Ergonomics, 31(5), 445-451.
6. Kolose et al., "Cluster Size Prediction for Military Clothing Using 3D Body Scan Data.", Applied Ergonomics, 96th edition, 2021
7. Punj, G., & Stewart, D. W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research, 20(2), 134–148
8. Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001, August). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining (pp. 263-268).
9. Majumder, J., & Sharma, L.K. (2015). Identifying Body Size Group Clusters from Anthropometric Body Composition Indicators. Journal of Ecophysiology and Occupational Health, 15(3/4), 81.
10. PoseNet. Available at https://www.tensorflow.org/. Last accessed on 13th July 2022.
11. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., & Jagersand, M. (2020). U2- Net: Going deeper with nested U-structure for salient object detection. Pattern Recognition, 106, 107404.
12. Gordon, C. C., Churchill, T., Clauser, C. E., Bradtmiller, B., McConville, J. T., Tebbetts, I., & Walker, R. A. (1989). Anthropometric survey of US Army personnel: Summary statistics, interim report for 1988. Anthropology Research Project Inc Yellow Springs OH.
13. Gordon, C. C., Blackwell, C. L., Bradtmiller, B., Parham, J. L., Barrientos, P., Paquette, S. P., ... & Kristensen, S. (2014). 2012 anthropometric survey of us army personnel: Methods and summary statistics. Army Natick Soldier Research Development and Engineering Center MA.
14. CAESAR, Available: http://www.shapeanalysis.com/, last accessed on 13th July 2022.
15. Sederberg, T. W., & Parry, S. R. (1986, August). Free-form deformation of solid geometric models. In Proceedings of the 13th annual conference on Computer graphics and interactive techniques (pp. 151-160).