Upper Limb Anthropometric Parameter Estimation through Convolutional Neural Network Systems and Image Processing

Andres Guatibonza[a], Leonardo Solaque[b], Alexandra Velasco[c] and Lina Peñuela[d]
Militar Nueva Granada University, Bogotá, Colombia

Keywords: Estimation, Convolutional Neural Networks, Anthropometry, Upper Limb, Image Processing.

Abstract: Anthropometry is a versatile tool for evaluating the human body proportions. This tool allows the orientation of public health policies and clinical decisions. But in order to optimize the obtaining of anthropometric measurements, different methods have been developed to determine anthropometry automatically using artificial intelligence. In this work, we apply a convolutional neural network to estimate the upper limb's anthropometric parameters. With this aim, we use the OpenPose estimator system and image processing for segmentation with U-NET from a complete uncalibrated body image. The parameter estimation system is performed with total body images from 4 different volunteers. The system accuracy is evaluated through a global average percentage of **71%** from the comparison between measured values and estimated values. A fine-tuning of algorithm hyper-parameters will be used in future works to improve the estimation.

1 INTRODUCTION

Anthropometry is a technique that allows to analyze the physical body features of each person and how affects their performance. It is usually applied in different areas such as nutrition, sports, clothing, ergonomics, and architecture. It is also commonly used to study diseases and to assess the nutritional status of the person (Tovée, 2012; Eaton-Evans, 2013). The anthropometric study makes it possible to calculate a series of measurements such sectional lengths, thicknesses, proportions, weights and sizes in order to obtain information about the individual's physical and nutritional status, which allows treating, as the case may be, certain deficiencies or physical aptitudes (Caballero, 2013; Gallagher et al., 2013).

Anthropometry undoubtedly varies between each person (Tovée, 2012), and the measurement effectiveness of traditional instruments is unquestionable, but when measuring, it usually takes some time, this allows the optimization of certain processes like this to be improved and automated. This fact has led to exploring forms of parametric estimation of anthropometry and how to develop automatic estimation systems, such as computational models based on neural networks.

In this context, one of the most popular applications for human parametric estimation is the pose estimation, and therefore, the estimation of anthropometry (Agha and Alnahhal, 2012; Ayma et al., 2016; Chang and Wang, 2015; Damaševičius et al., 2018). Works related to development of techniques for pose estimation have been widely tackled in literature. For example, in (Batchuluun et al., 2018) convolutional neural networks are used for human identification based on body movements. Moreover, (Brau and Jiang, 2016) proposes a deep convolutional neural network for estimating 3D human pose from monocular images with 2D annotations. In (Cao et al., 2017), a efficient detection of pose in 2D of several people in an image is carried out using non-parametric representation that allows to associate parts of the body with individuals in the image. A human body posture recognition algorithm based on neural networks is developed in (Hu et al., 2016), using the concept of wireless body area networks (WBAN). In (Li et al., 2014), a simultaneous heterogeneous system of a human pose regressor and detectors of articulation points and sliding window body parts is proposed in a deep network architecture for the estimation of human pose from monocular images. In (Tang et al., 2019), a new method is proposed for estimat-

[a] https://orcid.org/0000-0001-6102-563X
[b] https://orcid.org/0000-0002-2773-1028
[c] https://orcid.org/0000-0001-7786-880X
[d] https://orcid.org/0000-0002-1925-9296

Guatibonza, A., Solaque, L., Velasco, A. and Peñuela, L.
Upper Limb Anthropometric Parameter Estimation through Convolutional Neural Network Systems and Image Processing.
DOI: 10.5220/0010566505830591
In *Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2021)*, pages 583-591
ISBN: 978-989-758-522-7

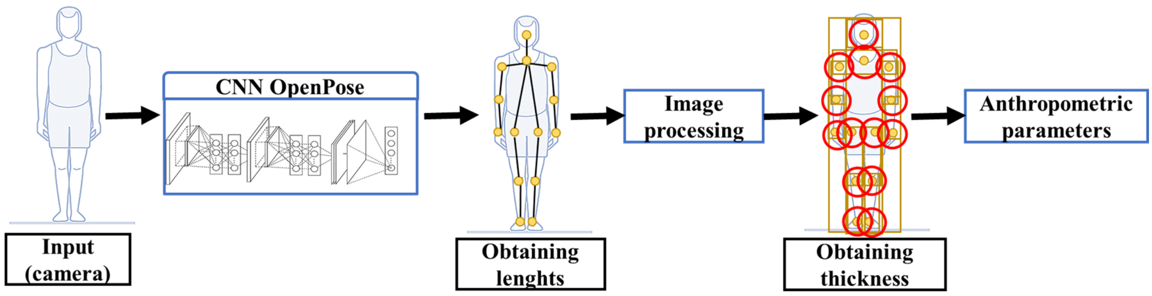

Figure 1: Simplified diagram of anthropometric parameter estimation.

ing the human pose in 3D from color images and in-depth images through an RGBD camera using convolutional neural networks. In (Zhu et al., 2017), an estimation of the human pose is performed in fixed images, using a multiple resolution convolutional neural network (MR-CNN) to train and learn the characteristics of multiple scales of each part of the body. In (Zhang et al., 2019), the study of pose estimation is mainly focused on the dynamics of the human limbs for the recognition of actions based on skeletons by means of a graph edge convolutional neural network (CNN). For some specific applications, (Farulla et al., 2016) presents the estimation based on artificial vision for robot-guided hand tele-rehabilitation, or the work proposed by (Chen et al., 2018), where a surveillance human posture dataset (SHPD) based on more specialized human pose reference points is presented for surveillance tasks, and (Liu et al., 2018) where a comprehensive pose network (IntePoseNet) is proposed to assess the level of customer interest in a business' merchandise.

These studies lead to the integration of new methods or techniques associated with the estimation of anthropometric indices. These indices are usually obtained through body metrics or using anthropometric measurement ratios to estimate human proportions or complementary data that can be obtained from the estimation of human pose or other methods with neural networks as in (Agha and Alnahhal, 2012), where a neural network and multiple linear regressions are developed to predict the dimensions of school children for an ergonomic design of school furniture. Moreover, in (Ayma et al., 2016) an approach to carry out nutritional evaluations to children under five years of age, through a system focused on the estimation of anthropometric indices, using image processing techniques and neural network is presented. In (Chang and Wang, 2015) a learning-based algorithm is proposed for estimating body shape using multi-view clothing images.

In the same way, other types of parameters can be obtained; for instance, in the work proposed by (Damaševičius et al., 2018) an automatic system is developed by using 3D anthropometric measurements of human subjects for gender detection with k-nearest neighbors (KNN), Support Vector Machines (SVM) and neural networks classifying parameters. Authors in (Rativa et al., 2018) present a complete study of the application of learning models to estimate height and weight from anthropometric measurements using support vector regression, Gaussian process and artificial neural networks, as well as in (Sarafianos et al., 2016) a regression-based method is proposed to use privileged information (LUPI) to estimate height using human metrology. Anthropometric-based true height estimation using multilayer perceptron ANN architecture in surveillance areas is tackled in (Sriharsha and Alphonse, 2019). All these proposed techniques generate new alternatives to improve or propose estimation strategies with several applications.

In this paper, a study of the application of convolutional neural networks and image processing to estimate anthropometric proportions is presented. We have developed a method based on neural networks and pre-calculated systems, such as OpenPose and U-NET, that allows optimizing anthropometric measurement processes in time and without using measurement instruments. The process is carried out from image analysis. This study aims to estimate anthropometric parameters from a single uncalibrated image, using convolutional neural networks for pose estimation and image processing. In this way, we obtain the lengths and widths of the upper limb. We apply neural networks and the Euclidean minimization Random Search technique (EMRS) to estimate the real anthropometric parameters of a person's upper limb. This work is limited to the anthropometric estimation of the upper limb, but it can be applied to the entire human body. This kind of neural networks models creates new opportunities for applications related to human anthropometry in safety and health.

This document is organized as follows: section II presents the methodology followed to obtain the estimation of human pose, anthropometric lengths, thicknesses and perimeters, and a description of the complete system that includes the networks used, and the techniques of measurement for the estimation. In section III we present the experimental results obtained by the application of this algorithm to the images of 4 different healthy volunteers. We compare the results with the actual anthropometric measurements of the participant. Finally we present, in Section IV, the conclusions of the work carried out and some future actions.

2 METHODOLOGY

A neural network system is developed for the acquisition of people's anthropometric data. The system estimates anthropometric parameters from an uncalibrated image. The first step is resizing and calibrating the image. For this, an algorithm based on convolutional neural networks for pose estimation finds the points of the joints. Then, the limb lengths are calculated through the Euclidean distance between the points that make up the body structure. Finally, the original image is processed to find the contours and segmentation of the person to isolate it from the environment. Additionally, other measurements such as perimeters and thicknesses are calculated using the EMRS technique. In summary, the system is composed of 3 main phases:

1. Estimation of the human pose using convolutional networks from an input image.
2. Estimation of the anthropometric lengths from the keypoints obtained in the convolutional network.
3. Estimation of thickness of the upper limb joints using image processing to obtain segmental contours of the person.

2.1 Application of Convolutional Networks for Human Pose Estimation

The pose estimation is formulated as a CNN-based regression problem at the body joints. It consists on detecting the coordinates of a 2D pose (x, y) for each joint from a RGB image. This pose estimation has been heavily reshaped by CNN since the introduction of "DeepPose" by (Toshev and Szegedy, 2014). As part of the development of this work, the method proposed by OpenPose for the estimation of human posture is used with a non-parametric representation called Part Affinity Fields (PAF), to "connect" each of the body joints found in an image. In essence, the model consists of pieces based on heat maps for an approximate location, and a module to sample and

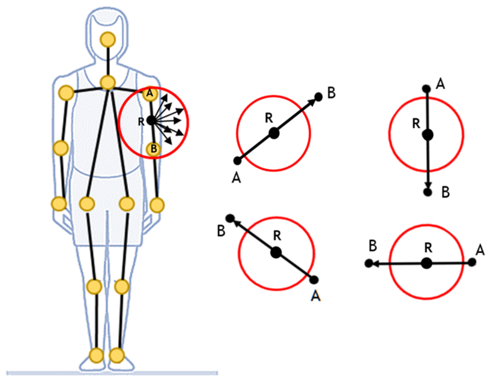

Figure 2: Concept of random search of distance minimization.

clip the convolution characteristics at a specified point (x,y) for each joint as shown in Fig. 1, as well as an additional convolutional model for fine adjustment. Specifically, this system is a pose machine (Wei et al., 2016a) that takes as input a color image, and produces the 2D locations of anatomical keypoints. This system consists of a feed-forward network that predicts a set of 2D confidence maps of body part locations and a set of 2D vector fields of part affinity fields (PAFs), which encode the degree of association between parts. The architecture iteratively predicts the affinity fields that encode the association between parts, and the detection confidence maps. The iterative prediction architecture refines the predictions in successive stages with intermediate supervision at each stage. If the reader wants more information about the operation of Openpose, please refer to (Cao et al., 2019; Wei et al., 2016a; Simon et al., 2017; Cao et al., 2017; Wei et al., 2016b)

In our case, the estimation of the human posture was limited to the detection of a single person in the image due to the focus of this work. We use the COCO model from Microsoft (Lin et al., 2014), which includes an extensive default dataset of 118.000 images for training and 5.000 images for validation that can be used for object detection, segmentation, and captioning dataset, among other features.

For the COCO model, 14 points are produced, associated as follows: p_0 = forehead, p_1 = Chest, p_2 = Right shoulder, p_3 = Right elbow, p_4 = right wrist, p_5 = Left shoulder, p_6 = Left elbow, p_7 = Left wrist, p_8 = Right thigh, p_9 = Right knee, p_{10} = Right ankle, p_{11} = Left thigh , p_{12} = Left knee, p_{13} = Left ankle. Once the image is passed to the model, the network makes a prediction of the coordinates where the previously defined points are located. The model produces confidence maps and affinity maps of the pieces that would be concatenated. Then it verifies whether each keypoint belongs to the image. The keypoint location is obtained by finding its maximum value on the confidence map, and also, includes an algorithm to find false detections, for more information please refer to (Cao et al., 2019).

2.2 Estimation of Anthropometric Lengths and Thicknesses of the Upper Limb from Estimation of Human Pose and Image Processing (Regionprops)

Once the coordinates of the body's joints are obtained with the pose estimation, the extraction of the characteristics related to the upper limb is carried out. We consider the distance from shoulder to elbow SEd, elbow to wrist EWd, and the biacromial width Bw. These characteristics can be obtained by calculating the Euclidean distance between two points in pixels expressed by

$$d_i = \sqrt{(p_{x(i+1)} - p_{x(i)})^2 + (p_{y(i+1)} - p_{y(i)})^2}, \quad (1)$$

Where d_i is the Euclidean distance between two points for $i = 1,..,8$ upper limb joints; $(P_x i, P, yi)$ are the cartesian position coordinates of the current keypoint, and $(P_x(i+1), P, y(i+1))$ are the cartesian position coordinates of the next adjacent keypoints.

Regarding the characteristics associated with the thicknesses and the perimeters, such as the wrist width Ww, the arm perimeter Ap, the forearm perimeter Fp and the elbow width Ew, we obtain them through image processing. A common way to perform segmentation is through a Fully Convolutional Network (FCN), better known as U-NET. In the same way, this network has pre-trained models with an extensive image dataset of people with pointers or identifiers to perform the segmentation. The result obtained at the output of the network is a mask of the person excluding the background. This approach to image processing will allow to segment it into regions. Similarly, if the reader wants more information about the operation of U-NET, please refer to (Ronneberger et al., 2015).

To obtain the widths and perimeters, we start from the original image $M_{original}$ which is entered into pre-trained U-NET model where the matrix of the identified person mask is obtained at the output M_{mask}. Additionally, noise filtering is performed to soften the edges of the mask and eliminate unwanted dispersion. Filtering is done with a morphological transformation since the image can be described as a binary. The morphological operators used for the decision are erosion and threshold. A kernel array $H(u,v)$ slides through the image (as in 2D convolution) where 1 is assigned if all pixels below the kernel are greater than the threshold; otherwise it erodes (reduces to zero). The matrix resulting from filtering M_{filter} can be obtained by

$$M_{filter(i,j)} = M_{mask(i,j)} * H(u,v)$$
$$\sum_{i=0}^{n}\sum_{j=0}^{n} M_{filter(i,j)} \begin{cases} 1 & \text{if } M_{filter(i,j)} > threshold \\ 0 & \text{if } M_{filter(i,j)} \leq threshold \end{cases}, \quad (2)$$

where $M_{filter(i,j)}$ is the resulting matrix after the filtering; $H(u,v)$ is the kernel matrix that performs the convolution across the image and evaluates with respect to the threshold. The selection of the threshold is free, typically, a value 0.5 is used.

Then, the background is eliminated by subtracting the matrix belonging to the original image $M_{original}$ from the matrix resulting from the filtered mask M_{filter}. After that, a blur filter is applied to the binary image to eliminate the residual noise and preserve the smoothness of the edges. To do this, we apply the image convolution. Then, we take the average of all the pixels below the kernel area, to replace the central element with this average. We use cv2.blur () function of OpenCV with this purpose. For more information of this library the reader is encouraged to review (Mordvintsev and K,).

From the binarized image, the contours of the person can be easily obtained by evaluating the matrix of pixels when a change from 0 to 1 is detected, and viceversa. Each individual contour is a matrix of coordinates (x,y) of the object border points. These contours are the limits of a shape with the same intensity. We use OpenCV function cv2.findContours () with this purpose. This function generates contours eliminating redundant points and keeping those that offer a great amount of information to save processing memory. The stored points that belong to the contour are included in the original image.

Subsequently, the Canny edge detection algorithm (Canny, 1986) is applied. This algorithm first removes noise in the image with a 5x5 Gaussian filter, since edge detection is susceptible to noise in the image. Then, the intensity of the gradient is found for horizontal, vertical and diagonal detection at the edges of the filtered image. Sobel's edge detection operator (Patnaik and Yang, 2012) for example, returns a value for the first derivative in the horizontal direction (Gx) and the vertical direction (Gy). Then, the edge gradient and direction are determined by

$$EdgeGradient(G) = \sqrt{G_x^2 + G_y^2}$$
$$Angle(\theta) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{3}$$

The pixels of the edges are kept or eliminated by the hysteresis threshold in the magnitude of the gradient. In our case, we used a Gaussian width $\sigma = 1$. The function skimage.feature.canny () of Scikit-image library performs this detection (Canny, 2020).

After calculating the edges, an exact Euclidean distance transform is used to evaluate the distances between the points obtained from the edges and their correlation using a heat map. For this, we used the function distance_transform_edt() of SciPy (Scipy, 2020).

Using the resulting distance map, we calculate the coordinates of the local maxima found in the image. If the peaks are flat, that is, several adjacent pixels have identical intensities, the coordinates of all those pixels are returned as local maxima. The calculation of local maxima is be defined by a region of $2 * mindistance + 1$ where the selection of *mindistance* will define the amount of points that the algorithm manages to find. We use the function feature.peak_local_max() of scikit-image library (OpenCV, 2020b).

Then, with the definition of the local maxima, edge segmentation is used where the limits between the highlighted labeled regions are applied to the original image. We use the function segmentation.mark_boundaries() of scikit-image library for this purpose (OpenCV, 2020a). Subsequently, the markers obtained by the segmentation will serve to fill in the areas enclosed by the segmentation, evaluating the color values of the pixels in each area including the background, and then calculating the average values to normalize the segmented area. The background of the image is normalized by assigning a user-defined color that represents a differentiation with the person to avoid confusing the colors between the person and the background.

Then, the thickness and the perimeters are estimated on the resulting image using a random search technique to minimize the euclidean distances. This technique consists in selecting the joint and midpoint between the joints and evaluating the euclidean distance between a series of random points located outside the person's space and the chosen reference point. The workspace radius that restricts this evaluation is defined as shown in Fig. 2.

The orientation of the points near A and B to the reference point R determine the arms orientation in the photo (raised or relaxed) and whether it is the right arm or the left arm. Then we define the radius of the workspace on which the search is carried out by using $r = 0.5 * \sqrt{(A_x^2 - B_x^2) + (A_y^2 - B_y^2)}$

Subsequently, a series of random points is generated on the radial space, and each distance towards the reference point is evaluated iteratively until obtaining the smallest possible distance that occurs when the edge of the person is reached. In this way, we obtain the estimates of the thicknesses and the perimeters of the upper limb. Then it is necessary to make a conversion of the estimates calculated from pixels to centimeters as

$$ExtractedLength_{cm} = \alpha * ExtractedLength_{px} \tag{4}$$

where α is a proportionality constant associated with a known horizontal distance between the camera and the person.

3 EXPERIMENTAL RESULTS

For the validation of the estimation system, a set of images with known anthropometric values are used. Four volunteers (2 men and 2 women) participated in the experiment. Their anthropomorphic measurements were previously obtained with measurement instruments. We used rulers to measure distances and calipers to measure thicknesses. Due to current sanitary restrictions, we worked with a limited number of volunteers, however, in future work we will extend the participation to give a greater validity to the estimation method.

The images are inserted into the subsystem to obtain the pose estimation, calculating the coordinates (x,y) associated with the joints and other defined points. The images were taken in uncontrolled environments to validate the effectiveness of the algorithm. For this reason, in this paper we present the cases of four volunteers in a non-controlled environment, with raised arms, which is a particular example where the person is doing any pose gesture and the background is not one color. In Fig.3 we show the resulting estimation of the defined points and the association of these points to each joint. The input image has a resolution of 500x500 pixels. The algorithm returns the parts that belong to the background and those that belong to the person's body to place the points of the joints. Some points of the lower limbs are not fully centered in the human body joints due to the similarity of colors with the background. Then, the lengths in pixels of SEd, EWd and Bw are calculated using (1). The proportionality of distances calculated in the image in Fig. 3 keeps a certain consistency if it is compared with the proportionality of the

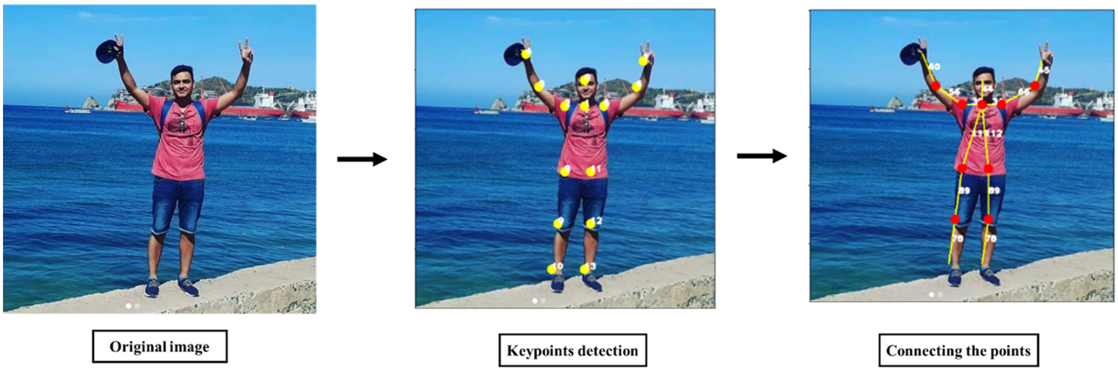

Figure 3: Sample image pose estimation.

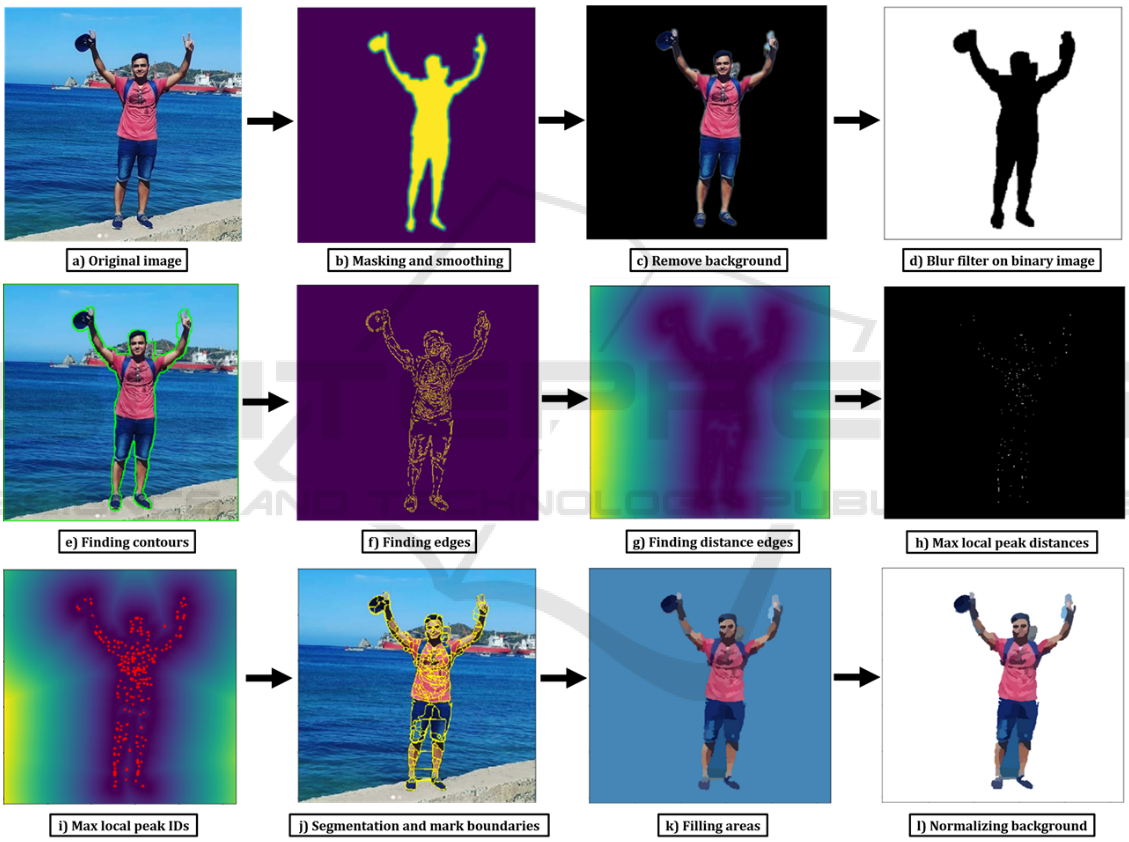

Figure 4: Segmentation process and image processing.

Table 1: Measured vs estimated anthropometric values.

	Volunteer 1		Volunteer 2		Volunteer 3		Volunteer 4	
Parameter	Measured (cm)	Estimated (cm)	Measured (cm)	Estimated (cm)	Measured (cm)	Estimated (cm)	Measured (cm)	Estimated (cm)
SEd	35	36,6	33	38,4	30	35	24	31,3
EWd	26	31,5	27	38	26	35,6	24	27,1
Bw	42	41,4	40	53,4	38	47,8	38	42,8
Ww	5,7	4,32	5,6	5,1	5,4	8,6	5,7	3,7
Ap	30	33,06	35	73	33	34,9	35	21,6
Fp	25	21,18	25	10,5	20	10,7	28	32,1
Ew	8,9	5,64	8,7	10	8,9	7	8,9	3,7

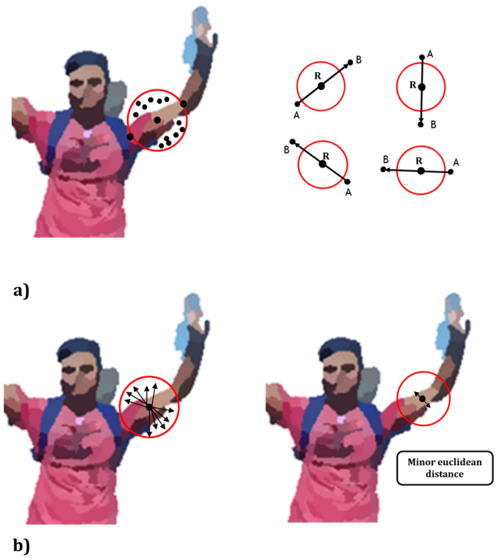

Figure 5: Shortest distance search in processed image.

anthropometric data in (Avila-Chaurand et al., 2007). Proportionality must maintain a fixed value for all the measured parameters. Table. 2 shows that SEd and EWd have a fixed proportionality value of 0.3 ± 0.025 but Bw differs in 0.15 units approximately, this is due to the fact that the estimation of the points that belong to the shoulders' biacromial width are not accurately located.

Table 2: Proportionality estimation values.

Parameter	Estimated (px)	Measured (cm)	Proportion
SEd	56	35	0.375
EWd	40	26	0.35
Bw	35	42	0.2

Then, we calculate the thicknesses and the perimeters using the original image for the segmentation system with U-NET. The result of the estimation is shown in Fig.4.b where the identified matrix of the person mask M_{filter}, with noise filtering and edge smoothing using a threshold of 0.5 is obtained. By subtracting the matrix $M_{original}$ from the M_{filter} we eliminate the background, as shown in Fig.4.c. In addition to matrix M_{filter}, a blur filter is applied to eliminate residual noises that may have remained in the first filtering. The idea is to approximate the contour of the person as closely as possible so the estimation can be improved. The result of the second filtering is shown in Fig.4.d. Subsequently, the outlines of the binary and filtered image are found using regionprops and combined with the original image as shown in Fig.4.e. Then, the characteristic edges and strokes found throughout the body are calculated with Canny's edge detection algorithm (Canny, 1986) using a Gaussian width $\sigma = 1$, as shown in Fig.4.f. The correlation of the edges and distances between the points obtained from the edge are shown in Fig.4.g, where later in Fig.4.h-i, the local max peak distances and IDs are calculated using a region with $mindistance = 5$. Then, the segmentation of the points obtained is carried out and the segmented areas are repainted with the average RGB value as shown in Fig.4.j-k. The background color has on average a bluish color so when repainting the areas, the background is also repainted. Finally, the background is normalized by assigning a RGB value of (255,255,255) corresponding to white color, thus resulting in the image of Fig.4.l as the output image for estimating thickness and perimeters.

Now, we calculate the thicknesses and the perimeters Ww, Ap, Fp and Ew. For this, we identify the orientation of the points A and B with the slope of those points and we evaluate the highest and lowest values of the coordinates in (x, y) in the cartesian plane of the image as shown in Fig.5.a. Then, we define the searching radius (red circle), and we evaluate the minor euclidean distance performing the random search algorithm with 100 points as shown in Fig.5.b.

Then, the obtained anthropometric proportions estimation results are converted from px to cm with $\alpha = 0.6$. This procedure was repeated for the other volunteers. Table 1 shows a comparison of the estimated and measured data. In average, the accuracy is 71% between the measured values and the estimated values. SE_d is the parameter with the highest average accuracy (83%), while A_p is the parameter with the lowest average accuracy (59%). Volunteer 1 has the highest average accuracy (84%), while volunteer 2 has the lowest average accuracy (60%). The accuracy can be increased using an image with a more controlled environment. On the other hand, the resolution of the camera, the image processing indices and the number of points in the EMRS technique can also improve the estimation.

4 CONCLUSIONS

In this paper, we presented a system for the estimation of anthropometric parameters. This system was implemented through the use of convolutional neural networks and image processing. We proposed a system based on 3 phases: estimation of the human pose, estimation of anthropometric lengths and estimation of thickness using image processing.

Human pose estimation is made from an image of total body. Four volunteers participated in this study. A particular example of an uncalibrated image was analyzed. This image was taken in a random environment where the pose estimation algorithm finds adequately the joint points. Then, the anthropometric lengths are calculated from the coordinates obtained from the human pose by means of a euclidean calculation. Then, we obtained the thicknesses and the perimeters of the upper limb through segmentation, image processing and a Euclidean minimization random search (EMRS). The results obtained were compared with measurements previously taken with rulers for the distances and calipers for the thicknesses, getting a global average accuracy of 71% between the measurements in all the subjects. We define empirically the hyperparameters, but other strategies can be proposed in the future to make a more accurate fine-tunning of these hyperparameters in order to get better results. This method promises to optimize the estimation of anthropometric measurements automatically without using other instrumentation, only from an image and the distance between the camera and the person.

Currently we are performing more tests of the estimation system, considering a higher number of participants. In this way we will have different anthropometric proportions in order validate and generalized the algorithm by fine-tuning the hyperparameters of the estimation. In the future we will also generalize the estimation to obtain other data such as muscle strength to use it for example, in rehabilitation applications.

ACKNOWLEDGEMENTS

This work is supported by Universidad Militar Nueva Granada- Vicerrectoria de Investigaciones, under research project IMP-ING-3127, entitled 'Diseño e implementación de un sistema robótico asistencial para apoyo al diagnóstico y rehabilitación de tendinopatías del codo'.

REFERENCES

Agha, S. R. and Alnahhal, M. J. (2012). Neural network and multiple linear regression to predict school children dimensions for ergonomic school furniture design. *Applied Ergonomics*, 43(6):979–984.

Avila-Chaurand, R., Prado-León, L., and González-Muñoz, E. (2007). *Dimensiones antropométricas de la población latinoamericana : México, Cuba, Colombia, Chile / R. Avila Chaurand, L.R. Prado León, E.L. González Muñoz.*

Ayma, V. A., Ayma, V. H., and Torre, L. G. A. (2016). Nutritional assessment of children under five based on anthropometric measurements with image processing techniques. In *2016 IEEE ANDESCON*. IEEE.

Batchuluun, G., Naqvi, R. A., Kim, W., and Park, K. R. (2018). Body-movement-based human identification using convolutional neural network. *Expert Systems with Applications*, 101:56–77.

Brau, E. and Jiang, H. (2016). 3d human pose estimation via deep learning from 2d annotations. In *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE.

Caballero, B. (2013). Nutritional assessment: Clinical examination. In *Encyclopedia of Human Nutrition*, pages 233–235. Elsevier.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698.

Canny, J. (2020). Canny edge detector. https://scikit-image.org/docs/dev/auto_examples/edges/plot_canny. [Online; accessed october-2020].

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., and Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*.

Chang, W.-Y. and Wang, Y.-C. F. (2015). Seeing through the appearance: Body shape estimation using multi-view clothing images. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE.

Chen, Q., Zhang, C., Liu, W., and Wang, D. (2018). SHPD: Surveillance human pose dataset and performance evaluation for coarse-grained pose estimation. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE.

Damaševičius, R., Camalan Seda, Sengul Gokhan, Misra Sanjay, and Maskeliūnas Rytis (2018). Gender detection using 3d anthropometric measurements by kinect.

Eaton-Evans, J. (2013). Nutritional assessment: Anthropometry. In *Encyclopedia of Human Nutrition*, pages 227–232. Elsevier.

Farulla, G. A., Pianu, D., Cempini, M., Cortese, M., Russo, L., Indaco, M., Nerino, R., Chimienti, A., Oddo, C., and Vitiello, N. (2016). Vision-based pose estimation for robot-mediated hand telerehabilitation. *Sensors*, 16(2):208.

Gallagher, D., Chung, S., and Akram, M. (2013). Body composition. In *Encyclopedia of Human Nutrition*, pages 191–199. Elsevier.

Hu, F., Wang, L., Wang, S., Liu, X., and He, G. (2016). A human body posture recognition algorithm based on BP neural network for wireless body area networks. *China Communications*, 13(8):198–208.

Li, S., Liu, Z.-Q., and Chan, A. B. (2014). Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network. *International Journal of Computer Vision*, 113(1):19–36.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014*, pages 740–755. Springer International Publishing.

Liu, J., Gu, Y., and Kamijo, S. (2018). Integral customer pose estimation using body orientation and visibility mask. *Multimedia Tools and Applications*, 77(19):26107–26134.

Mordvintsev, A. and K, A. Smoothing images. https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_filtering/py_filtering.html.

OpenCV (2020a). Module: segmentation. https://scikit-image.org/docs/dev/api/skimage.segmentation.html. [Online; accessed october-2020].

OpenCV (2020b). peak local max. https://scikit-image.org/docs/0.7.0/api/skimage.feature.peak.html. [Online; accessed october-2020].

Patnaik, S. and Yang, Y.-M., editors (2012). *Soft Computing Techniques in Vision Science*. Springer Berlin Heidelberg.

Rativa, D., Fernandes, B. J. T., and Roque, A. (2018). Height and weight estimation from anthropometric measurements using machine learning regressions. *IEEE Journal of Translational Engineering in Health and Medicine*, 6:1–9.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.

Sarafianos, N., Nikou, C., and Kakadiaris, I. A. (2016). Predicting privileged information for height estimation. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE.

Scipy (2020). distance transform edt. https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.distance_transform_edt.html. [Online; accessed october-2020].

Simon, T., Joo, H., Matthews, I., and Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*.

Sriharsha, K. V. and Alphonse, P. (2019). Anthropometric based real height estimation using multi layer peceptron ANN architecture in surveillance areas. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE.

Tang, H., Wang, Q., and Chen, H. (2019). Research on 3d human pose estimation using RGBD camera. In *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. IEEE.

Toshev, A. and Szegedy, C. (2014). DeepPose: Human pose estimation via deep neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

Tovée, M. (2012). Anthropometry. In Cash, T., editor, *Encyclopedia of Body Image and Human Appearance*, pages 23 – 29. Academic Press, Oxford.

Wei, S., Ramakrishna, V., Kanade, T., and Sheikh, Y. (2016a). Convolutional pose machines. *CoRR*, abs/1602.00134.

Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y. (2016b). Convolutional pose machines. In *CVPR*.

Zhang, X., Xu, C., Tian, X., and Tao, D. (2019). Graph edge convolutional neural networks for skeleton-based action recognition. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–14.

Zhu, A., Jin, J., Wang, T., and Zhu, Q. (2017). Human pose estimation via multi-resolution convolutional neural network. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE.