3D Size-Estimation based on the Geodesic Distance Measured by Photogrammetric Scanning Device

Morteza Daneshmand[*1], Andres Traumann[†1], and Gholamreza Anbarjafari[‡1]

[1]iCV Group, Institute of Technology, University of Tartu, Tartu 50411, Estonia

Abstract

This paper introduces an innovative approach for estimating important body measurements according to the depth information extracted by photogrammetric scanning devices, such as Microsoft Kinect II, and verifies its applicability and efficiency through applying it to an experimental use-case concerning classification based on size, to be implemented in virtual fitting rooms. The mathematical framework consists, mainly, in constructing a projection from the screen coordinates associated with the pixels onto the real-world ones, and then integrating the geodesic distance gradients throughout a path connecting the beginning and end points. Besides, in order to reduce the inaccuracy through excluding the possible high-frequency noise, and to obtain more realistic measurements through minimizing the discrepancy between the original path and the one taken into account in the integration module, the path is fitted to a smooth curve in a separate stage before performing the geodesic-distance-calculation iterations. In the end, in order to verify the accuracy and preciseness of the proposed technique, it is applied for obtaining numerous body sizes, where for ensuring that the correct measurements are achieved, imaginary landmarks are placed on the bodies of the subjects. The results show that the accuracy is less than half a centimeter using Microsoft Kinect II.

---

[*]mortezad@ut.ee
[†]andres.traumann@ut.ee
[‡]shb@ut.ee

1   Introduction

The problem of approximating 3D sizes and geodesic distances has been investigated from numerous perspectives by researchers. For instance, some of the strategies from the latter list are developed on the basis of binocular stereo vision [1]. Some others are concerned with devising indices representing anthropomorphic biometric characteristics via home-based imaging approaches [2], which requires taking images of the body with different orientations of the camera, and subsequently, reconstructing the whole body by means of image processing techniques with the consistency between the different segments of the silhouettes as the objective of the optimization algorithm. It is worth noticing that the depth information could be obtained through stereo-vision or structured-lightening techniques [3] even if the existing camera is not able to provide it readily.

In the existing literature, numerous methods and studies have investigated the concept of reliable vision-based 3D measurement estimation, each of which discusses the related technical and theoretical complexities from its own point of view, based on the associated experimental setup, i.e., the approaches proposed heretofore would apply to the specific practical settings considered in the context of the corresponding study, but not necessarily to others.

For devising measurement algorithms functioning properly under the conditions sought from the outset, due attention should be paid to various physical aspects and requirements of the particular devices supposed to be utilized. In what follows, a summary of the details, along with the outstanding advantages and disadvantages, of the most recent studies carried out in the foregoing field is provided, in order to illustrate the current state of the art, as well as to clarify the prospective contributions of the strategy proposed in this paper to the ones that are either directly connected with it or could well benefit from appropriately extending it.

One of the studies that has tried to develop a methodology aiming at solving the range finding problem is [1], which concentrates, particularly, on taking into account the specifications of mobile robots, by means of DSP, through utilizing the so-called "binocular stereo vision". The algorithm is devised such that first paired images of the target objects are captured, which are processed using Gaussian filter and improved Sobel kernels. Afterwards, employing feature-based local stereo matching, the location of the object is determined.

Confidence and left-right consistency filters are further applied to the result of the foregoing procedure, in order to reduce the possibility of the existence of mismatches, which will enhance the quality of the final output. The whole algorithm is implemented using the DSP/BIOS operating system, which will enable real-time control over the position detection process. The point-to-point distance calculation procedure, in ranges up to 120 centimeters, has, reportedly, demonstrated accuracy of 99%, while taking 39 milliseconds in the worst case for performing the task, which could qualify suitable in the associated context. In [4], poor perception of body shape, i.e., invalid thoughts about the proportions, sizes and distances, has been attributed to disorders such as anorexia and bulimia nervosa. Nevertheless, such misconceptions are reported to be diffi-

cult to formulate and represent numerically. In the foregoing study, it has been stated that the previous ones had not provided sufficient analysis of the trend of the changes in the body shapes, for example, in the parts where the level of fat changes considerably. More clearly, the shortcoming lies in the fact that, despite the simplistic nature of most of the approaches proposed so far, the changes of the body shape, normally, do not represent trends that could be resembled by consistently stretching the surface in either of the directions, as proposed by the "Distorting Video Technique (DVT)". In other words, the body shape shows different trends of change in all the parts, which differ in terms of amount, type or complexity, and cannot be simulated using simple linear scaling.

The above study tries to tackle the aforementioned problems, and achieve a more realistic way of representation of the body shapes, through considering the available biometric information, and performing separate analyses on each body part, which enables to adjust the shape and size of each segment as desired. The results of the foregoing study, which include Perimeter-Area Ratio (PAR), and subsequently, Body Mass Index (BMI), will provide the opportunity to verify the authenticity of the body shape representation perceived by the user by means of comparison. As an outstanding outcome of this study, it has been concluded that the majority of the users tend to overestimate their sizes, where the actual value of BMI is around 15 in case of anorexics, which is the boundary between underweight and emaciation, but is deemed 20 by the instances.

Anthropometric body measurements have been investigated from another perspective in [2], along with the associated ratios, which influence the perception of the body shape and fitness delivered by the user. The main goal of the foregoing study is to devise a home-based imaging system capable of extracting anthropometric body measurements, which consists in capturing multiple images of the same object in front of the camera, or *vice versa*. Subsequently, the entire body can be reconstructed through applying image processing techniques on various silhouette segments. Finally, another processing step is implemented, which improves the consistency between the different parts, taking an energy function into account as the criterion.

Another solution to the problem of estimating 3D distances is proposed in [3], which makes use of stereoscopy in order to create visual presentation on the basis of two images taken from slightly different positions. The latter can be realized through either employing two separate cameras, each of which possessing a single lens, or utilizing a stereo camera with a pair of juxtaposed lenses. The main application of the foregoing method would be to obtain the approximate distance between the camera and the object of interest. The latter might involve numerous physical parameters and constants, such as the focal length of the camera or the distance between the two cameras, in case of using two separate cameras. The procedure is such that first the distance of the object from one of the cameras is calculated, and then the value associated with the other one is estimated through minimizing the dissimilarities that inevitably exist between the two aforementioned values.

As another example of the studies addressing the problem of 3D size estimation, [5] can be mentioned, which aims at developing a strategy for automatic,

computer-based extraction of body sizes and measurements, again according to the anthropometric data. The resulting utility is supposed to be applicable to many industrial contexts, such as security, medicine, sport and clothing. It has been, reportedly, meant not only to minimize the annoying characteristics of the measurement system to the person being measured, but also to maximize the access-ability, ease of use and portability of the service to the operator, which is being tested and verified, in terms of efficiency and reliability, for the specific case-study of garment industry.

It should be noted that the applications of the problem of 3D size estimation are not limited to extracting human body sizes and shapes, i.e. there are numerous fields of industry and research that benefit from the resulting utilities. For example, structural analysis of masonry arches, which are ubiquitous in the sites including architectural heritage, can be mentioned. Such structures could be still analyzed using 3D laser scanners, which are rather cumbersome and expensive. One of the important issues that has been addressed in the foregoing study is that real geometry, rather than idealized one, has to be used, in order to enable the proper use of 3D surveying techniques, which will lead to more realistic results. The proposed methodology is based on TLS, which, as aforementioned, tries to come up with realistic and reliable 3D measurement estimation.

Another field of study and technology that could significantly benefit from the development of 3D measurement techniques is underwater quantification and mensuration. For example, the methodology proposed in [6] can be mentioned, which deals with a laser/video system, consisting of a frame, a roll/pitch motion reference sensor, a video camera and three micro-lasers. In the context of the foregoing study, the orientations of the micro-lasers are taken into account for the purpose of triangulation, which results in the calculation of the position of specific objects within the field of view of the camera. Besides, on the basis of the foregoing information, as well as the one reported by the motion sensors, the magnification is determined throughout the whole field of view. Subsequently, the necessary parameters, such as perspective overlays, the distance from a particular point or location, the scale in each part of the image and measurements of area, can be estimated through employing a prescribed bunch of image processing techniques.

One of the important properties of the above study is that any characteristic of the whole system, including the "laser wavelength and power, camera sensitivity and resolution, dynamic range and mounting geometry", could be adjusted as desired, in order to accommodate the specifications and particular requirements of the application at hand, as well as the level of accuracy sought from the outset. The system has been claimed to be flexible enough to be used in any sub-sea application, device or platform, since it performs at comparatively low cost, in an automatic manner, for the general purpose of acquiring quantitative data from the standard underwater video streams. The associated software has also the virtue that the video information could be further post-processed in real-time.

The first specific application of the utility resulted from the above study is es-

timating the number of ground-fish species within a "transect area", which are either commercially or recreationally used. The benefit brought about by employing such technique is that the density of particular fish species throughout a "high-relief area"can be estimated directly, but non-intrusively, which is not affordable by typical existing techniques.

The rest of the paper is organized as follows. In the next section, the notion of the method proposed in this paper is described, both literally and mathematically. Afterward, the results of the implementation of the foregoing methodology are visually presented in various scenarios, along with discussion on the reliability of the method, and the authenticity and accuracy of the resulting numerical values. Finally, the paper concludes through summarizing the main points discussed and presented throughout the paper, and providing hints as to how the approach proposed in the paper can be extended and followed up in the ongoing studies.

2 Proposed Size Estimation Algorithm

As the first stage of the 3D measurement method proposed in this paper, the coordinates should be mapped from the screen to the real-world ones. For the latter purpose, a variety of approaches are available in the literature. For example, depth optical cameras can be used, most of which are capable of performing the task automatically. Besides, 3D reconstruction may be considered, as mentioned in Section 1, through stereo-vision or structured lighting.

Assuming that the above goal is accomplished, a function is present in the form $f : \mathbb{N}^2 \to \mathbb{R}^3$, which handles the mapping module, as follows:

$$f(x, y) = (X, Y, Z), \tag{1}$$

in which $x, y \in \mathbb{N}$ and $X, Y, Z \in \mathbb{R}$ stand for the screen and real-world coordinates, respectively.

Every measurement requires a set of starting and end points as P_s and P_e, respectively, which are specified based on the user input. A naive solution for calculating the distance between the foregoing two is the making use of the Euclidean one, which takes the real-world coordinates corresponding to them into account, as follows:

$$d_E(P_s, P_e) = \|f(P_s) - f(P_e)\|. \tag{2}$$

The above attitude, however, does not lead to results correlating with the real geometry, in the most cases, i.e. the value reported by such a formula is correct only where a straight line lies on the surface that connects the two points. To overcome this shortcoming, a more realistic path could be drawn between the points, being obtained while considering the nature of the surface, including the ripples, convexities and concavities, etc., which can be formulated in the following way:

$$d(P_s, P_e) = \int_C h(X, Y) ds, \tag{3}$$

through integrating numerous portions of the whole path, which should be as small as possible, and connect the beginning and end points step by step. In the above relationship, C denotes a path from $f(P_s)$ to $f(P_e)$, resulted from projecting the straight line $P_s P_e$ on the screen onto the real-world coordinates standing for the surface points. Moreover, $h : \mathbb{R}^2 \to \mathbb{R}$ represents a function responsible for outputting the Z coordinate of each projected surface point, which takes the pair (X, Y) as the input.

It is worth noticing that, at least in the context of this paper, the smallest possible components of the path being integrated are determined based on the differences between the real-world coordinates of the points represented by juxtaposing pixes on the screen, which is performed by the aforementioned map.

The mapping could be thought of as a weighted graph, such that each pixel, being considered a node, is connected to its four neighbours, which are all mapped from the screen coordinates to that of the real world. As a major difference between the methodology proposed in this paper and the approaches suggested in the previous studies, in order to achieve physically-sound results, while calculating the weights, the contribution of the Z-coordinate value is discarded, i.e. only the X- and Y-coordinates are taken into account, which is also the main difference between the foregoing definition of distance and euclidean distance.

For the purpose of determining the shortest possible path connecting the beginning and end points, the Dijkstra Algorithm is employed, followed by piecewise calculation of the distance based on the foregoing path. Assuming that p_n stands for the n-th point visited while iterating through the path, the overall distance, D, could be reliably estimated as follows:

$$D = \sum_{n=2}^{m} \|p_{n-1} - p_n\|, \tag{4}$$

where m is the total number of the points visited.

The underlying fact is that the aforementioned path is determined in such a way that it would resemble the shortest real 3D line lying on the surface, despite other studies reported in the literature that mostly consider the Euclidean distance between the points, which, due to the convexities and concavities that normally appear on the faces of almost all objects, does not necessarily lead to the actual measurement supposed to be calculated.

As a relevant point, since the body surface does not usually present sudden change in its depth, being denoted by the Z-coordinate, if appeared, they are considered wrong information, i.e. noise, which might have arisen from various grounds, including electrical or mechanical fault. To be more clear, since extreme fluctuations of the body surface are not sought from the outset, in case of emergence, they should be properly denoised, which would, otherwise, lead to undesired overestimation of the sizes.

In order to reduce the chances of the above negative impact, averaging of the path is considered, which maintains the general trend of the changes in the body surface, but ignores ripples smaller than a certain threshold, which are, almost certainly, caused by noise, malfunction or miscalculation. Fig. 1 shows a
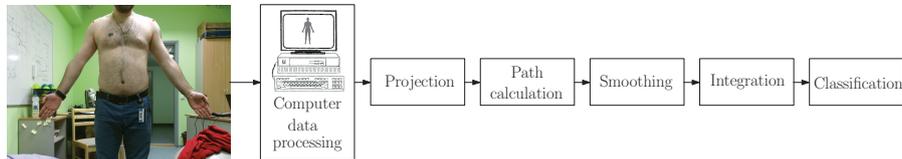
Figure 1: A sample measurement scenario, with the red marks representing the beginning and end points, accompanied with a schematic diagram summarizing the main modules of the proposed method.

sample scenario, along with a schematic diagram summarizing the fundamental elements of the proposed 3D measurement procedure. The accuracy accomplished, on the basis of the exhaustive experiments conducted, is lower than *0.5* centimeter, while using the Microsoft Kinect as the optical depth camera.

The main contribution of the proposed method is making it feasible to grasp a rough approximation of the essential body sizes and 3D geodesic distances, for example, bust, waist and hip girths, via optical cameras, rather than 3D scanners, which typically, impose considerably more restrictive physical limitations in practical and commercial contexts.

3 Experiments, Results and Discussion

The experimental results required taking RGB-D images using the Kinect camera, analysing the visual data, extracting the measurements, and classifying the body shapes based on the size, according to the known gender. The images were taken through a simple setup with white background, from the people who were tried to be as diverse as possible, in terms of the foregoing criteria, so that they could suffice for representing all the possible classes. It is worth mentioning that the images should be taken under circumstances where the depth information is not vanished, which would otherwise mean that the parts of the image undergone depth loss would appear as totally black.

One of the advantages of the approach suggested in this paper is that it significantly improves the ease of use and convenience, as well as reduces the costs and hardware requirements, in that the images are obtained using a single camera, while taking a single shot is enough. Most of the similar methods proposed in the literature so far demand making use of either a 3D scanner or multiple cameras, and in the most cases, multiple shots are required for gathering and processing the information needed for extracting the 3D distances and sizes.

Although the methodology introduced in this paper is capable of calculating any arbitrarily-specified 3D measurement, the experiments showed that the cross shoulder and torso width at waist values were sufficient for the aim and scope of this study. More clearly, the main goal of this paper is to validate the 3D measurement approach in the sense of providing reliable information for classification of human bodies into the classes of small, medium and large, though it

has been devised as a general -purpose method, which may have various applications of numerous sorts, some of which were discussed in Section 1.

Many researchers have been engaged in trying to fulfill the above task, which is vital for the aim of utilities such as virtual fitting rooms, and have proposed solutions approaching the problem from different perspectives. In [7], for example, it has been concluded that the most important measurements that could help make a mannequin robot resemble a body shape realistically are bust, waist and hip circumferences.

For the aim of this paper, on the other hand, the problematic constraint was that the method was supposed to lead to the required classification information just through a single shot, which means that it is not readily possible to calculate, or even reliably estimate, the circumferences, since there is no access to the rear side of the body, i.e. there is only a frontal image. As a consequence, the choices of the sizes were limited to the ones that lied completely in the frontal part of the body, meaning that only widths and heights could be taken into consideration.

As shown in Fig. 2, the way the initial set of the widths and heights were selected was such that the ones that, according to our previous experiments and studies, had proven helpful were tried, most of which were located on the parts of body that separate two or more juxtaposing segments - they stand for the most extreme changes in the shape and size of the cross-section contours, which are crucial for deciding whether the cloth, at least, would not be simply too big or small when it comes to satisfying the minimum requirements of fitting the body.

 Further exploration of the usefulness and informativeness of the aforementioned measurements led to the conclusion that, among the ones shown in Fig. 2, the cross shoulder and torso width at waist band were not only the most important, but also enough for deciding on the class the body size belongs to, i.e. the choice of the foregoing values as the criteria for classification was achieved through consultations with the experts from Fits.me [8], who have had substantial experiences in this area, and constant involvement of them with this research project while supervising the results of the implementation of the 3D measurement and classification algorithms, intended at ensuring that the results correspond to the real commercial standards and user expectations.

The above process included, mainly, trying out as many combinations of the conditions on the sizes, along with different thresholds, and checking whether the results of classification on such basis leads to acceptable results. Heretofore, this has been implemented rather manually, through considering picking conditions randomly, and adjusting the thresholds via trial-and error after starting with an initial guess decided intuitively. However, a more systematic way of dealing with this problem would be to find the best possible solution through an optimization procedure or, if computationally possible, searching for it exhaustively, which is beyond the scope of this paper.

The proposed 3D measurement algorithm was applied to 156 instances, and 7 important body sizes were calculated for all of them, where in 98% of the cases, the classification result corresponded to the one sought from the outset, meaning
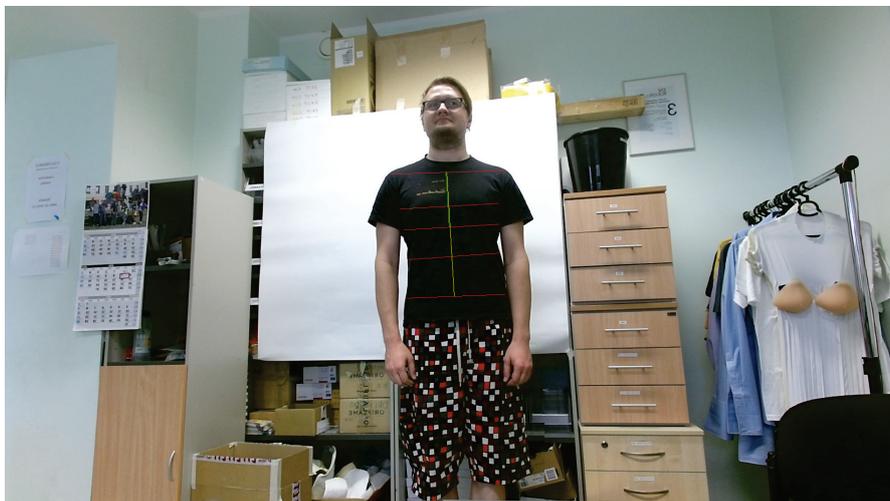
Figure 2: A sample image showing the placement of the imaginary markers for calculating the seven important body measurements on one of the instances, which include, among others, shoulder and waist widths.

that both geodesic 3D measurement and classification procedures are reliable and almost accurate. As aforementioned, only shoulder and waist widths were taken into account for the classification purpose, which resulted in a reasonable performance. For the sake of illustration, the seven important measurements of a couple of sample body sizes, along with the associated size classes decided by the classification algorithm are provided in Table 1.

4   Conclusion

This paper proposed a novel strategy for almost accurate estimation of body measurements and sizes based on the depth information provided by photogrammetric optical scanning devices. The procedure starts with constructing a map relating the screen coordinates to that of the real word, and subsequently, integrating the geodesic distance gradients through iterating the connecting points on the path obtained via projecting the straight line connecting the beginning and end points on the screen onto the real line lying on the physical surface. Moreover, for the aim of overcoming the negative effect of the high-frequency noise present in the path determined, which emerges in form of extreme fluctuations of the surface depth, a smoothing fitting is performed, averaging the path connecting points, and removing the undesired ripples. The methodology is implemented on an industrial application, namely, classification of human bodies according to size, as a case-study, which is utilized in the context of virtual fitting rooms employed for online try-on of clothes. The experimental

Table 1: A set of sample body measurements, as well as the body size classes determined by means of the proposed classification technique. All values are in millimeters.

Sizes⇒ Sample⇓	S	s_1	s_2	s_3	s_4	s_5	s_6	s_7	Class
1	M	462	394	380	406	412	169	480	Medium
2	M	477	373	377	445	410	239	520	Large
3	M	437	361	368	363	392	209	448	Medium
4	M	424	351	353	359	370	174	493	Small
5	W	381	302	292	259	348	189	420	Medium

Legend
S: Sex
M: Man
W: Woman
s_1: Cross shoulder
s_2: Trunk width at bust
s_3: Trunk width under the bust
s_4: Torso width at waist band
s_5: Trunk width at hip
s_6: Chest height
s_7: Upper-body height

results certify the credibility and reliability of the approach introduced, where the accuracy of the distance estimation is bounded to less than half a centimeter using Microsoft Kinect as the optical camera.

5 Acknowledgement

References

[1] X.-b. Lai, H.-s. Wang, and Y.-h. Xu, "A real-time range finding system with binocular stereo vision," *Int J Adv Robotic Sy*, vol. 9, no. 27, 2012.

[2] Z. Li, W. Jia, Z.-H. Mao, J. Li, H.-C. Chen, W. Zuo, K. Wang, and M. Sun, "Anthropometric body measurements based on multi-view stereo image reconstruction," in *35th Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 366–369.

[3] J. Mrovlje and D. Vrancic, "Distance measuring based on stereoscopic pictures," in *9th International PhD Workshop on Systems and Control, Izola, Slovenia*, 2008.

[4] M. J. Tovée, P. J. Benson, J. L. Emery, S. M. Mason, and E. M. Cohen-Tovée, "Measurement of body size and shape perception in eating-disordered and control observers using body-shape software," *British Journal of Psychology*, vol. 94, no. 4, pp. 501–516, 2003.

[5] T. Kohlschütter, "Human body modelling by development of the automatic landmarking algorithm," 2012.

[6] D. M. Kocak, F. M. Caimi, T. H. Jagielo, and J. Kloske, "Laser Projection Photogrammetry and Video System for Quantification and Mensuration," in *OCEANS'02 MTS*, vol. 3. IEEE, 2002, pp. 1569–1574.

[7] M. Daneshmand, A. Aabloo, and G. Anbarjafari, "Size-Dictionary Interpolation for Robots Adjustment," *Frontiers in Bioengineering and Biotechnology*, vol. 3, p. 63, 2015.

[8] "Fits.me, Virtual Fitting Room." [Online]. Available: http://fits.me/