

Accurate 3D measurement using optical depth information

A. Traumann, M. Daneshmand, S. Escalera and G. Anbarjafari[✉]

A novel three-dimensional measurement technique is proposed. The methodology consists in mapping from the screen coordinates reported by the optical camera to the real world, and integrating distance gradients from the beginning to the end point, while also minimising the error through fitting pixel locations to a smooth curve. The results demonstrate accuracy of less than half a centimetre using Microsoft Kinect II.

Introduction: The problem of vision-based measurement calculation has been dealt with through different perspectives and methodologies in the literature. In what follows, a selection of the studies conducted in the foregoing field is briefly reviewed so as to provide a background to the approach proposed and implemented in this Letter. In [1], the focus is specifically on finding the range for mobile robots through DSP (digital signal processing) with the so-called ‘binocular stereo vision’. The proposed methodology consists in capturing paired images of the target, along with Gaussian filter and improved Sobel kernels, and then determining its location through implementing feature-based local stereo matching. As the final stage, for alleviating mismatching possibilities, which would lead to more reliable performance, confidence, and left–right consistency filters are applied. Moreover, so as to achieve real-time control over the location detection algorithm, it is developed on the basis of a DSP/BIOS (basic input/output system) operating system. All in all, their suggested algorithm has claimed to be able to operate with more than 99% accuracy for point-to-point distance range of 120 cm within 39 ms in the worst case, which could be considered efficient and reliable. The perception of body shape has been investigated from another perspective in [2], which deals with the disorders, such as anorexia and bulimia nervosa, caused, supposedly, by invalid perception of the body shape and sizes, being considered difficult to represent numerically. The foregoing study deems previous ones reported in the literature inadequate in the sense of properly illustrating the trend of the changes of the body shape while the level of fat is either increasing or decreasing. More precisely, the body shape does not vary in such a way that could be represented by simply stretching it in either of the directions consistently, as the one entitled ‘distorting video technique (DVT)’; it rather demonstrates different amounts and types of changes in each part, i.e. the body change follows a specific pattern which is much more complex than simple linear scaling. The foregoing study tries to come up with a more realistic body-shape representation through taking actual biometric information into account, and investigating the changes in the body parts separately, which leads to flexibility in providing the opportunity to manipulate them in the desired manner. One of the virtues of the latter study is that one could calculate the perimeter–area ratio, and subsequently, body mass index (BMI), which helps to evaluate the validity of the user’s perception through comparison.

Similarly, Li *et al.* [3] deal with specific anthropometric body measurements and ratios affecting the perception of the users of body fitness. The main idea in the latter study is developing a home-based imaging system for automatic extraction of anthropometric body measurements. The work reported in [4] focuses on stereoscopy, promoting a strategy helping to produce a visual illustration based on two separate images taken at not exactly the same locations, which can be realised through either making use of two cameras with a single lens each or employing a stereo camera containing two juxtaposed lenses. The main functionality of such an approach is to measure the distance between the camera and the object under consideration. Obviously, while calculating the latter values, physical parameters, including the focal lens and the distance between the two cameras have to be taken into account.

In [5], the depth information obtained by the sensors is analysed for extracting biometric soft indicators, such as lengths and girths. Supervised training is considered while making use of multi-part pose clusters. Afterwards, an iterative process is applied to match the three-dimensional (3D) body shape descriptor model. The process obtains accurate measurements and segmentation, which outperforms the random forest detection [6] in the sense of demanding a smaller amount of training data.

From another perspective, Konovalov *et al.* [7] deal with the problem of detecting hands based on RGB (red–green–blue)–depth information of the upper part of the body for human–computer interaction purposes,

and involves overcoming variability of the possible appearance of the hand due to the high flexibility of the hand, especially at the wrist part. The underlying assumption, which has been verified through conducting several experiments, is that the hand landmarks are always located in certain positions with constant geodesic distances from an anatomical reference point, which is also automatically detected. The methodology consists in segmenting the human body on the basis of the depth information and afterwards obtaining a graph representation including the geodesic paths originating from the aforementioned reference point, $G^t = (V^t, E^t)$, where $V^t = \mathbf{B}^t$ denotes the vertices, and $E^t \subseteq V^t \times V^t$ stands for the edges, such that

$$E^t = \left\{ (\mathbf{p}_{ijk}, \mathbf{p}_{i'j'k'}) \in V^t \times V^t : \left\| (i, j, k)^T - (i', j', k')^T \right\|_\infty < 1 \right\} \quad (1)$$

It should be noted that in the above formulation, $\|\cdot\|_\infty$ returns the infinity norm of its vector component, and $(i, j, k)^T$ and $(i', j', k')^T$ denote the 3D coordinates of the points \mathbf{p}_{ijk} and $\mathbf{p}_{i'j'k'}$, respectively, in \mathbf{B}^t , which, if being neighbours, are connected with the edge $e = (\mathbf{p}, \mathbf{p}') \in E^t$, corresponding to the weight $w(e) = \|\mathbf{p} - \mathbf{p}'\|_2$, on which basis, the geodesic distance between the two points is defined as follows:

$$d_G(\mathbf{p}, \mathbf{p}') = \sum_{e \in EP(\mathbf{p}, \mathbf{p}')} w(e) \quad (2)$$

which includes the edges along the shortest path between the points, determined based on the min-path Dijkstra’s algorithm [8].

Proposed method: The proposed method requires a coordinate mapping from screen coordinates to real-world coordinates. It can be achieved in various ways: it is possible to use a depth camera that provides the mapping, or it can be obtained using some form of 3D reconstruction method based on stereo vision, structured lighting etc. From here on, it is assumed that there exists a function $f: \mathbb{N}^2 \rightarrow \mathbb{R}^3$ that maps screen coordinates to real-world coordinates:

$$f(x, y) = (X, Y, Z) \quad (3)$$

where $x, y \in \mathbb{N}$ are the screen coordinates and $X, Y, Z \in \mathbb{R}$ are the corresponding real-world coordinates.

For each measurement two points of screen coordinates are required: a starting and an ending point, let us denote the starting point by P_s and the ending point by P_e . These two points are obtained from the user input. The most obvious measurement can be performed by finding the Euclidean distance between the real-world points corresponding to these two points:

$$d_E(P_s, P_e) = \|f(P_s) - f(P_e)\| \quad (4)$$

This distance is rarely correct; however, as the real distance depends on the nature of the surface between these two points, the Euclidean distance would be correct only when the path from start to end along the surface is a completely straight line. In reality, the distance should be given by the following line integral:

$$d(P_s, P_e) = \int_C h(X, Y) ds \quad (5)$$

where C is the geodesic path from $f(P_s)$ to $f(P_e)$ that is obtained by projecting the line segment $P_s P_e$ on the real-world coordinate surface, and $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function that gives the Z coordinate of the projected path at the surface point (X, Y) .

The proposed algorithm approximates the true distance using the following method. The pixel locations are mapped to their corresponding real-world coordinates. This situation can be represented by a weighted graph where each node is connected to four nodes that were created based on the real-world coordinates obtained by mapping the neighbouring pixel locations to real-world coordinate space.

The main contribution of this Letter arises from the fact that, instead of measuring 3D geodesic paths, as performed in [7], the shortest path is computed by considering only X - and Y -coordinates, which, from the mathematical point of view, means that the graph where the Dijkstra’s algorithm [8] is applied is constructed as $\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$, whereas the Euclidean distance would be $\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}$. More clearly, the underlying notion is that the path obtained via the approach suggested in this Letter is more realistic in the sense that it is the result of projecting the path connecting the beginning and end points onto the surface,

rather than simply considering a straight line drawn between them, which is tantamount to excluding the Z-coordinate values from the calculations and avoiding bypassing the convexities and concavities constructing the actual surface.

After obtaining the path, the algorithm starts to iterate through the path and sums up the total Euclidean distance between all the iterated points. Let p_n denote the n th visited point. The Euclidean distance from p_{n-1} to p_n is added to a total sum of distances D . After iterating all the pixels the total sum D is a close approximation of the real distance. The described approximation can be written as

$$D = \sum_{n=2}^m \|p_{n-1} - p_n\| \quad (6)$$

where m denotes the total number of visited points. The described method works by following the approximate path between the starting and the ending points. The path is then divided into several smaller parts; each of those smaller parts is approximated using the Euclidean distance. The final output of the algorithm is achieved by summing up all the lengths of the smaller path segments.

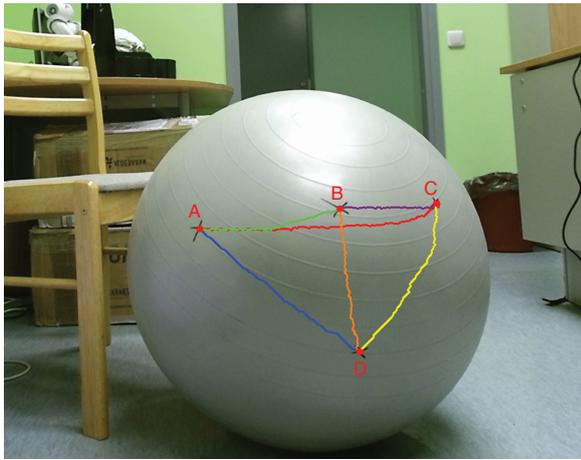


Fig. 1 Ball with markers used for measurements

Note that the accuracy of the whole system depends on correct denoising of the iterated path coordinates. Often the Z-coordinates can fluctuate between two similar values even in the case of a totally flat surface. It might cause a big overestimation of the distance due to the nature of the resulting path. To avoid that a strong averaging was used, so that the shape of the path was retained and all smaller changes were successfully removed.



Fig. 2 Illustration of some experimentation scenarios

Results: The proposed method was tested in several scenarios, one of which was a large ball with marks on it, as shown in Fig. 1, and some others are illustrated in Fig. 2. In the former, there are four points marked on the ball and measurements were performed between

all of them. We refer to these four points by A , B , C , and D . The results of the automatic measurements using the proposed algorithm are compared with the measurements obtained by manual measurements using a measuring tape. These results are shown in Table 1.

Table 1: Results of measurements

Points	Proposed method (cm)	Manual measurement (cm)
AB	20.87979	20.7
AC	37.34988	37.2
AD	28.36199	28.0
BC	16.29063	16.2
BD	19.71188	19.3
CD	24.80985	24.6

It can be seen that the measurement results of the proposed method are very accurate and are close to the results that were obtained manually. The maximum difference was only 0.41 cm, while the average difference was 0.23 cm. On average, the accuracy of the proposed automatic measurement for different scenarios is 0.25 cm, as shown in Fig. 1. Hence, it is concluded that the proposed method can be used to measure distances along surfaces with high accuracy.

Conclusion: In this Letter, a novel 3D measurement technique has been proposed. The proposed algorithm maps the screen coordinates to the real-world coordinate and then integrates distance gradients from the beginning to the end point. The error of fitting pixel locations to a smooth curve is also minimised. The conducted experimental results show an accuracy of less than half a centimetre for the proposed technique.

Acknowledgments: The research was supported by the ERDF program Estonian higher education information and communications technology and research and development activities state program 2011–2015 (ICT program), Estonian Research Council Grant (PUT638) and by the Spanish project TIN2013-43478-P.

© The Institution of Engineering and Technology 2015

Submitted: 19 April 2015

doi: 10.1049/el.2015.1345

One or more of the Figures in this Letter are available in colour online.

A. Traumann, M. Daneshmand, S. Escalera and G. Anbarjafari (*iCV Group, Institute of Technology, University of Tartu, Tartu 50411, Estonia*)

✉ E-mail: shb@ut.ee

S. Escalera (*Computer Vision Center, Universitat de Barcelona, Spain*)

References

- Lai, X.-B., Wang, H.-S., and Xu, Y.-H.: 'A real-time range finding system with binocular stereo vision', *Int. J. Adv. Robotic Syst.*, 2012, **9**, (27), pp. 1–9
- Tovée, M.J., Benson, P.J., Emery, J.L., Mason, S.M., and Cohen-Tovée, E.M.: 'Measurement of body size and shape perception in eating-disordered and control observers using body-shape software', *Br. J. Psychol.*, 2003, **94**, (4), pp. 501–516
- Li, Z., Jia, W., Mao, Z.-H., Li, J., Chen, H.-C., Zuo, W., Wang, K., and Sun, M.: 'Anthropometric body measurements based on multi-view stereo image reconstruction'. Thirty Fifth Annual Int. Conf. on Engineering in Medicine and Biology Society (EMBC). IEEE, Osaka, Japan, July 2013, pp. 366–369
- Mrovlje, J., and Vrancic, D.: 'Distance measuring based on stereoscopic pictures'. Ninth Int. PhD Workshop on Systems and Control, Izola, Slovenia, 2008
- Madadi, M., Escalera, S., González, J., Roca, F.X., and Lumberras, F.: 'Multi-part body segmentation based on depth maps for soft biometry analysis', *Pattern Recognit. Lett.*, 2015, **56**, pp. 14–21
- Jiang, R., Tang, W., Wu, X., and Fu, W.: 'A random forest approach to the selection of epistatic interactions in case-control studies', *BMC Bioinf.*, 2009, **10**, (Suppl 1), p. S65
- Konovalov, V., Clapés, A., and Escalera, S.: 'Automatic hand detection in RGB-depth data sequences'. Conf. of the Catalan Association of Artificial Intelligence, Osona, Spain, October 2013, pp. 91–100
- Bertsekas, D.P.: 'Network optimization: continuous and discrete methods', *Athena Sci.*, 1998, **8**, pp. 51–98