An Approach to the Estimation of Dimensions from Static Images

František Jakab[1,*], *Ondrej* Kainz[1,**], *Miroslav* Michalko[1,***], and *Rastislav* Petija[1,****]

[1]*Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, Košice, Slovak Republic*

Abstract. The broad goal of this study is the estimation of dimensions of real-world multidimensional objects. The focus is on the human being and its dimensional parameters, primarily the height. The approach to the estimation of the height in a single camera environment is presented and is experimentally verified. The proposed method presumes that the distance from the optical device to the human body and its Vertical Field of View (VFOV) are known. The VFOV has to be calculated or acquired only once for a specific device. The accuracy of the proposed method is validated by a series of test scenarios including, e.g., statistical height measurements.

1 Related work

Momeni-K et al. [1] proposed the estimation of the human height from a single calibrated image. These authors used the information on the camera height and pitch angle with respect to a reference horizontal plane and also a vanishing point. The overall results yield the largest deviation of 11.4 mm in the case of larger pitch angles. Gallagher et al. [2] also presented the utilization of a calibrated image with the goal to extract the height of a person, along with the demographic classification of an individual. This classification included height, age, and gender. The authors claim the proposed method to have a largest deviation of 27 mm. A study using the intrinsic camera parameters, the height and tilt angle of the camera as calibration parameters was presented by Kispál and Jeges [3]. The authors use in this case a surveillance camera. The maximal and minimal points of a person in the image are selected as points of interest for the further calculation. The standard deviation of the measurements ranges from 38 to 43 mm (see Tab.1 of [3]). The results of [4] reported a largest deviation of only 10 mm. The height of the figure was determined using a number of three calibrated devices that were placed at a height of 2.5 meters at precisely defined positions. The height was determined from the video, and the calculations were done on individual frames.

2 Proposal to the estimation of the height

The main goal of this paper is to propose a solution that would enable the extraction of the human height in a single camera environment. For the calculation of the height, the estimation of angles from

---
[*]e-mail: frantisek.jakab@tuke.sk
[**]e-mail: ondrej.kainz@tuke.sk
[***]e-mail: miroslav.michalko@tuke.sk
[****]e-mail: rastislav.petija@tuke.sk

the still scene is required. As the reference angle, we consider the extracted value of the Vertical Field of View (VFOV) of the camera and its further utilization in the calculation of the object angle towards the camera. The distance of the object from the camera is presumed to be accurately known.

The extraction process involves the following specific steps given in the flowchart in the Fig. 1. Most of the input data do not require user input or can be fully automatized, except for the distance and the calculated VFOV. At this phase, no human body detection algorithm is implemented, however background subtraction technique was utilized in both the testing scenarios and in the developed experimental software solution. The calculation of angles is based on the procedure described below.
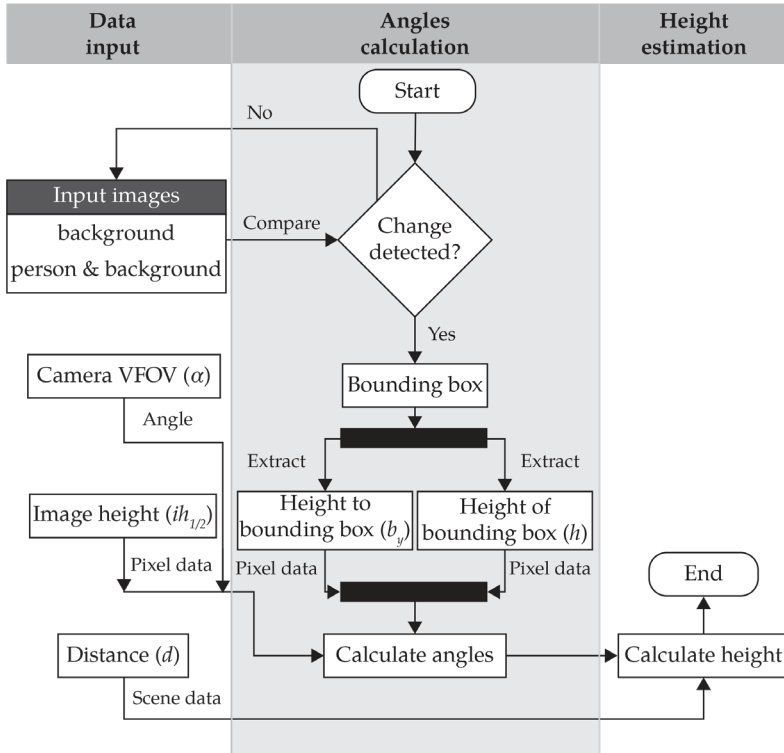

Figure 1. Main steps of the approach to the height estimation using angles

The shots taken by a single camera are considered as the input images. The distance d from the person to the camera is another input parameter. For the extraction of the dimensions, the distance in pixels (the height h, in Fig. 2), between the maximum and minimum points of the person is fixed. The number of pixels from the bottom of the image to the minimum point of the bounding box determines the horizontal location of the bounding box, in Fig. 2 (right) denoted as b_y. The value for location of upper number of pixels is denoted as t_y and is calculated from b_y, h and the height of the image. The reference value $ih_{1/2}$ (half of the image height in pixels), being half of the overall image vertical size, is used to acquire ht_y and hb_y which are used for the calculation.

In Fig. 2, α denotes the VFOV angle. For the calculation of the other angles, the angle α was used together with the values (in pixels) of b_y and h. The value of d is expected to be known.
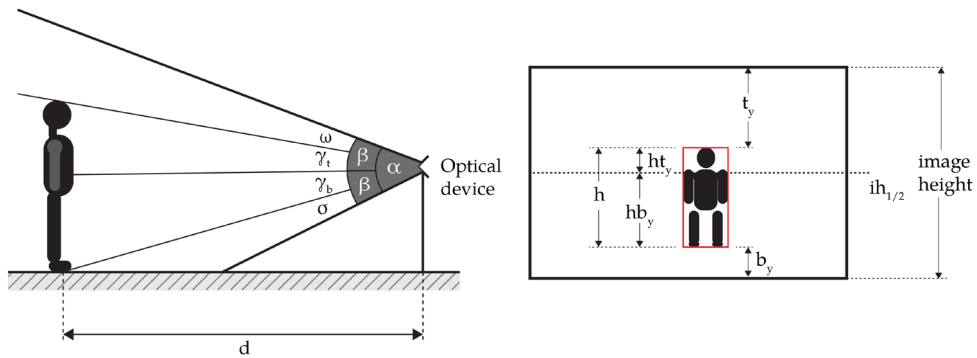
Figure 2. Calculated angles and distances in the image

Two values are extracted from the image:

- b_y, the extracted pixel value from the image, i.e., the minimal value from the image to the lowest point of the bounding box,
- h, the height of the person extracted from the image in pixels.

In order to extract the height of the person, the values of t_y, hb_y, ht_y, σ, ω, γ_b, and γ_t have to be determined. Once acquired, the height of the human figure in the image is obtained from:

$$h_{\text{real}} = d \cdot \tan(\gamma_b) + d \cdot \tan(\gamma_t). \tag{1}$$

This calculation is valid for an arbitrary tilt angle of the optical device, however the optimal value is get at an angle of $0°$.

3 Experimental results

An experimental software solution was developed for testing purposes of the proposed approach. The calculation of the real height is possible through the calculation of partial results having as inputs the VFOV and the distance from the person to the device. The VFOV depends on the type of the camera and has to be entered or estimated only once (for a specific camera).

The experimental verification of the proposed approach proves valid, having an error in the laboratory controlled conditions of up to 1 % (the worst case). In a real environment, the error is ranging from 1 to 2 %. The error in percent considers a reference height of 1700 mm. If a high-end imagining device is used, the expected deviation is up to 1 %. For optical devices with smaller image sensor size, the deviation may increase up to 2 % (the worst case). The utilization of the proposed solution remains valid even for the height extraction of the larger number of persons if necessary (e.g., for the statistical purposes). There are, however, specific limitations, these include:

- As the distance d between the object and camera gets larger, the height accuracy determination may severely decrease. At distances larger than 10 meters, values 1500 ± 500 mm were obtained in the case of 16 MP camera sensors.
- During testing of the proposed solution, only small pitch angles of the camera were considered (up to 10 degrees). Larger pitch angles result in larger deviation. Momeni-K et al. [1], who also utilized the pitch angle, formulated the same conclusion.

- The posture of the human figure has to be taken into account. Under certain circumstances the human body may be considered as lying in a single vertical plane. However, this may often not be the case due to incorrect posture.
- The estimation depends on the type of the optical device, the lower the resolution, the worse the results. In our study a 16 MP camera was utilized and with the current trends in the development of technologies, it is expected that low resolution devices will be deprecated.
- The precise measurement of the distance d from the camera is a precondition for the achievement of an accurate extraction.

4 Conclusion

This study was focused on the extraction of dimensional information. The principal goal was the extraction of the height of a person from a static still image in a single camera environment. The secondary aim was to minimize the need of the user defined input to the system. The method for extraction of angles, using VFOV, was defined. The approach to the extraction of the height was described and experimentally verified.

The proposed method for the extraction of the height using angles, deployed in both laboratory and real conditions, is comparable to the prior studies. The verification of the proposed method, as well as validation of the experimental software solution, were carried out on series of real tests. When testing the overall accuracy, the highest deviation was 18 mm, the lowest only 3 mm.

The input requirement of the distance may be considered as the main drawback of this approach. In future studies we plan to enhance the current method with the estimation of the distance through a calibration procedure. We will presume the camera as static and not moving with respect to the environment. Then, we will calibrate the image with an object of known height using the ratio of the real dimension to the pixel value.

Acknowledgments

References

[1] M. Momeni-K, S.C. Diamantas, F. Ruggiero, B. Siciliano, *Height estimation from a single camera view,* in Proceedings of the *International Conference on Computer Vision Theory and Applications*, Volume **2**: VISAPP (Rome, Italy, February 24–26, 2012), pp. 358–364

[2] A.C. Gallagher, A.C. Blose, T. Chen, *Jointly estimating demographics and height with a calibrated camera,* in *IEEE 12th International Conference on Computer Vision* (Kyoto, Japan, September 29 – October 2, 2009), pp. 1187–1194

[3] I. Kispál, E. Jeges, *Human height estimation using a calibrated camera* (2006) Retrieved from https://pdfs.semanticscholar.org/8cd1/ca2ead8f5dd56c409c52dfca74fa0f2450dc.pdf

[4] E. Jeges, I. Kispál, Z. Hornák, *Measuring human height using calibrated cameras,* in *2008 Conference on Human System Interactions* (Krakow, Poland, May 25–27, 2008), pp. 755–760