

# The Measurement of Human Height Based on Coordinate Transformation

Xiao Zhou<sup>1</sup>, Peilin Jiang<sup>2,3</sup>, Xuetao Zhang<sup>1,4</sup>,  
Bin Zhang<sup>2</sup>, and Fei Wang<sup>1</sup>(✉)

<sup>1</sup> The Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University,  
No.28 Xianning West Road, Xi'an 710048, China  
{xuetaozh,wfx}@mail.xjtu.edu.cn

<sup>2</sup> The School of Software Engineering, Xi'an Jiaotong University,  
No.28 Xianning West Road, Xi'an 710048, China  
{pljiang,bzhang82}@mail.xjtu.edu.cn

<sup>3</sup> National Engineering Laboratory for Visual Information Processing  
and Application, Xi'an Jiaotong University,  
No.28 Xianning West Road, Xi'an 710048, China

<sup>4</sup> Shaanxi Digital Technology and Intelligent System Key Laboratory,  
Xi'an Jiaotong University, No.28 Xianning West Road, Xi'an 710048, China

**Abstract.** This paper presents a new method for measuring human height from video frames. Based on coordinate transformation, the positional relationship between the coordinates of a person's feature points on the image can be transformed into the person's height with the knowledge of intrinsic parameters and rotation angle of the camera. In our method, the distance between the camera and the target is not a necessary parameter, which, in contrast, can be estimated by our algorithm. From experiments, we conclude that our method can be simply implemented to estimate a person's height from video frames in a controllable error scale.

**Keywords:** Visual measuring · Coordinate transformation · Camera calibration · Video

## 1 Introduction

Various human-recognition technology based on image processing have been presented in recent years, such as face-recognition, gait analysis, morphological analysis of human and so forth, which are mature enough to be utilized in many fields. However, the research on vision measurement plays an important role in surveillance video analysis. Specifically, the information of human height obtained from a normal surveillance camera system makes up the important characters of a target [1, 2].

This paper presents a coordinate transformation based method using a calibrated Pan Tilt Zoom (PTZ) camera to evaluate the walking object's height. Our method can work very well as long as the object is walking. Moreover, any translation or rotation of the surveillance camera will not invalidate the measurement system or reduce its precision. Certainly, to guarantee the accuracy, several parameters are indispensable for

our method. Firstly, the intrinsic parameters of the camera are used to transform the image coordinate system to the camera coordinate system. The rotation angle between the camera's optical axis and the horizontal plane can help modify the camera coordinate system.

## 2 Related Research

Extensive research has been done in human height measurement using images. There are about two different approaches: camera-only geometry based method and multi-device based method. One of the multi-device based methods [3], considers a camera as the main hardware and a fixed laser beam is used for signal emission. Because of good identifiability of the laser beam in image, it is easy to extract distance between the laser beam projection image and the center of image, which is used to estimate the human height. This method is simple, but with a high cost due to the laser generator. Sonia Das proposed the Direct Linear Transformation method [4], to get human height variation. Specifically, the intrinsic and extrinsic parameters of the camera are needed to compute the Z (the vertical direction) coordinates of the person's sole and head using DLT method, and then the difference between them is the person's height. The method proposed by Richard Hartley [5] is one of the camera-only geometry based methods. It is necessary to extract the information told by the reference lines in a real vertical reference plane. In addition, from the whole measurement procedure, the camera cannot be moved or rotated at all. Once the camera's position is changed, the information in the reference image should be re-extracted, which in a large degree restrains the camera's vision. While our coordinate transformation based method can solve the problems mentioned above perfectly, which is one of the camera-only geometry based methods, and neither need a real reference plane nor get invalid on account of the camera's translation or rotation.

### 2.1 The Coordinate Transformation Based Algorithm

Our method is based on the coordinate transformation. In short, the intrinsic parameters are used to transform the image coordinate system to the camera reference frame. The angle between the camera's optical center and the ground can help modify the camera coordinate system. At last, the height of the camera is regarded as a reference to compute the height of the walking object.

Generally speaking, a person's height can be judged by the distance between the head of the person and his/her sole when the person stands upright on the floor. In other words, we can entirely establish a coordinate system matching the corner of the wall, as a result, a person's height can be estimated as the absolute value of the difference between the ordinates of the person's head and the sole. Therefore any coordinate systems without rotation compared with the corner of wall are available for us to measure a person's height. Well, as the link between the world coordinate system and the image reference frame, the camera reference frame is the best choice.

In a standard camera coordinate system (the optical axis of the camera parallels the ground and the imaging plane is perpendicular to the ground),  $TB$  denotes the person in the vision of the camera,  $H$  is the height of the person,  $tb$  is the image of  $TB$  on the virtual imaging plane,  $D$  shows the distance between the camera and the person; The front view of the geometric model is shown in Fig. 1.

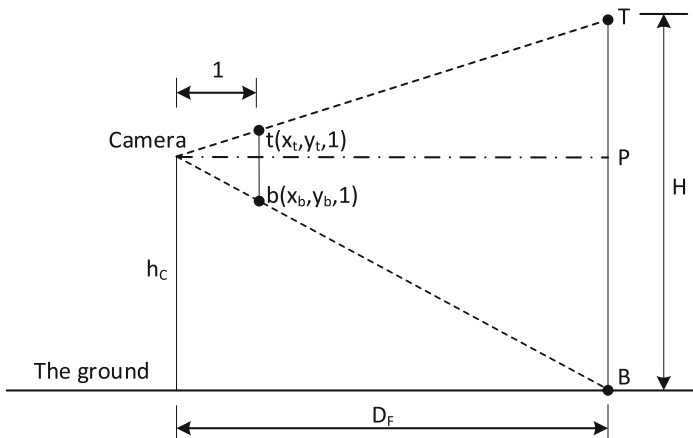


Fig. 1. The front view of the geometric model of the real scene

Supposing that the 3D coordinates, which are normalized along direction of the optical axis of the camera ( $z$ -axis), of the points  $t$  and  $b$  are  $(x_t, y_t, 1)$  and  $(x_b, y_b, 1)$  respectively. Then using similar triangles theorem, the height of the person denoted by  $H$  can be easily computed as follows:

$$H = h_c \left(1 - \frac{y_t}{y_b}\right) \tag{1}$$

Where is the height of the camera. Similarly, when the optical axis of the camera crosses through the person's body, the perpendicular distance between the lens and the person which is denoted by  $D_F$  can be computed as follows:

$$D_F = \frac{h_c}{y_b} \tag{2}$$

Nevertheless, when the optical axis of the camera doesn't crosses through the person's body,  $D_F$  is just the projection of the perpendicular distance between the lens and the person, so the real distance denoted by  $D$  in such circumstances should be modified by the following equation:

$$D = \frac{h_c}{y_b} \sqrt{1 + \frac{(x_t + x_b)^2}{4}} \tag{3}$$

Strictly speaking,  $x_t$  and  $x_b$  are identical when the target is a vertical segment with no width. However, the width of the walking object cannot be omitted, so we use the mean of  $x_t$  and  $x_b$  to represent the abscissa of the walking object.

So far, we have managed to estimate the height and the distance on the condition of the knowledge of the coordinates of the points  $t$  and  $b$ . However, the three-dimensional coordinates of  $t$  and  $b$  in the camera reference frame cannot be gained directly from the image. Therefore we have to use coordinate transformation to calculate these coordinates as follows.

**Coordinate Transformation.** There are two types of camera parameters including the intrinsic and the extrinsic. Intrinsic or internal camera parameters describe the projection of objects onto the camera image [6]. They establish the relationship between the points in the camera reference frame and the pixel coordinates of the points on the images captured from the camera.

Assuming that the pixel coordinates of  $t$  in the image reference frame is denoted by  $(u_t, v_t)$ , then its coordinates denoted by  $(x_1, y_1, 1)$  in the camera reference frame can be estimated as follows:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = A^{-1} \begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} \tag{4}$$

$A$  is the intrinsic parameters matrix and given by

$$A = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where  $f_x, f_y$  are the equivalent focal length in  $x$  and  $y$  direction.  $u_0, v_0$  are the principal point in  $x$  and  $y$  direction. Likewise, the coordinates of  $b$  denoted by  $(x_2, y_2, 1)$  in the camera reference frame can be computed from the Eq. (4).

**Modifying the Camera Reference Frame.** The camera reference frame might not be a standard camera coordinate system because of the optical axis might not parallel the ground, in other words, there may exist a nonzero angle denoted by  $\beta$  between the optical axis and the ground. Thus we have to rotate the current camera reference frame to make it a standard camera coordinate system. The rotation angle of the transformation is the angle between the optical axis and the ground which can be acquired directly from a PTZ camera, and the rotation direction is the direction that can lessen the angle. The coordinates of  $t$  after the rotation transformation denoted by  $(x_t, y_t, 1)$  is shown in the Eq. (6).

$$\lambda_1 \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = R \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (6)$$

R is the rotation matrix and given by:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & \sin \beta \\ 0 & -\sin \beta & \cos \beta \end{bmatrix} \quad (7)$$

Where  $(x_1, y_1, 1)$  is obtained by the Eq. (4). Apparently, the coordinates gained from the Eq. (6) is normalized along the z-axis. Likewise, the coordinates of b after the rotation transformation denoted by  $(x_b, y_b, 1)$  can be estimated by the Eq. (6) with the change of the right side into  $(x_2, y_2, 1)$ .

By now, we have obtained the normalized three-dimensional coordinates of t and b in the standard camera reference frame, as long as we have the knowledge of the camera's height, we can estimated the person's height and the perpendicular distance between the lens and the person from the Eqs. (1) and (3).

As we mentioned earlier,  $h_c$  is the height of the camera, which is actually the distance between the optical center of the camera and the horizontal plane that the camera lies on. In fact, we cannot get the accurate height of the camera using a tapeline since we don't know the exact position of the optical center of the camera. But we can estimate it according to the Eq. (1): put a reference object with known height denoted by  $h_R$  in front of the camera and then  $h_c$  can be estimated as follows:

$$h_c = \frac{y_b}{y_b - y_t} h_R \quad (8)$$

### 3 Experiments

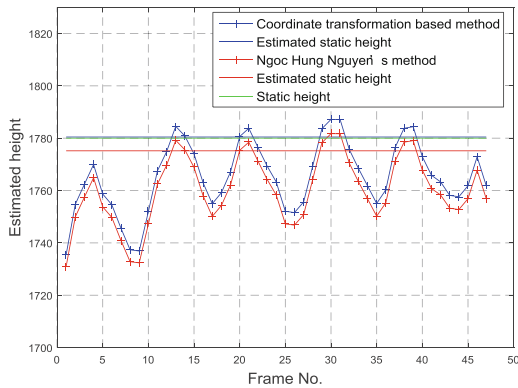
To testify feasibility and accuracy of our method, we designed two groups of experiments. Control group is based on Ngoc Hung Nguyen's method [7]. The first group of experiments is named feasible measurement experiment, which is implemented to estimate the human height from a series of video frames. The second group is used to testify the accuracy of our method, which considers the image of calibration plate as experimental subjects.

#### 3.1 Human Height Measurement

The first step of our measurement system is to distinguish and extract the walking subject from a fixed background, which can be completed by using GMM [8] method. Then the ordinates of the top of the human head and his/her sole on each video frame can be easily acquired from the foreground image exported by GMM algorithm.

Finally, the measurement algorithm will be implemented to compute the height of the walking subject. In our experiments, the volunteer passed through the vision of the camera in a relatively low speed. We changed the angle between the optical axis of the camera and the horizontal plane from  $7^\circ$  to  $13^\circ$ . Correspondingly, we got 7 videos that captured the volunteer’s motion.

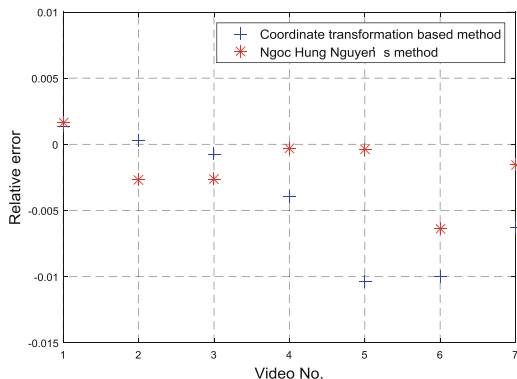
As we mentioned before, the author’s method needs six parameters including three ordinates of the reference lines on the image and their heights in the real world to evaluate the human height. Thus we have to get 7 images capturing the three reference lines on the vertical reference plane. Besides, we have to measure their real heights and extract their ordinates on each image corresponding to every different angle manually. However, our method requires only one particular parameter, which is the angle between the optical axis of the camera and the horizontal plane and can be immediately obtained from the PTZ camera. The experimental results are shown in Fig. 2.



**Fig. 2.** The estimated heights of the volunteer when the angle between the optical axis of the camera and the horizontal plane is  $7^\circ$ : the star shaped points and the cross shaped points correspond to the heights computed by the author’s method and our method respectively, while the static height is shown by the green horizontal straight line. The red and blue straight lines denote the estimates of the static height, which are the average of the maxima of the heights curves. (Color figure online)

Figure 2 shows the height variation of the volunteer. It can be easily observed that the results obtained by the two methods seem to match each other while the human height variation is quite significant. Figure 3 shows the relative error of the heights computed by the author’s method and our method respectively in each video. Apparently, the estimated static heights computed by our method are more accurate than the author’s method in the first three videos, while the others are not. More importantly, we can see the largest relative error of our method is about 1.5 % from Fig. 3, which is absolutely acceptable in measuring a person’s height.

According to Ngoc Hung Nguyen’s method, three reference lines are needed to compute the height of the target. One of the reference planes is arranged as shown in Fig. 4(a), and the three red circles on the calibration plate are chosen to be the reference



**Fig. 3.** The relative error of the estimated static height, which is the average of the maxima of each estimated heights' curves corresponding to every single video. The star shaped points and the cross shaped points denote the relative error pf heights computed by the author's method and our method respectively.



(a)

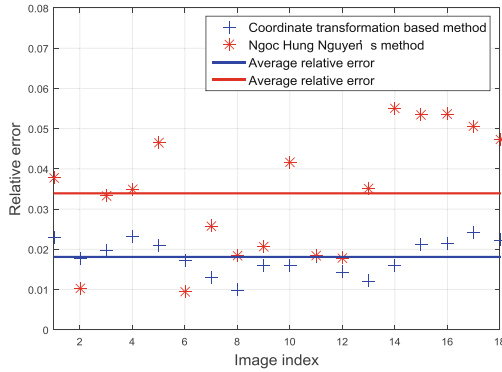


(b)

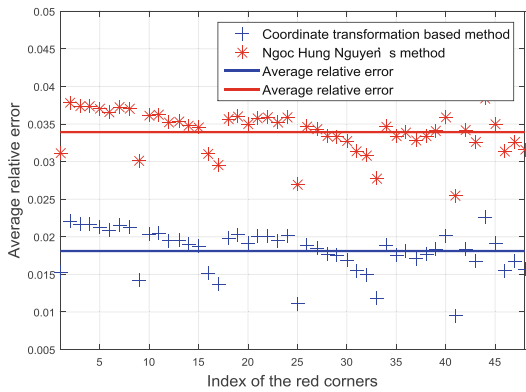
**Fig. 4.** (a) The reference plane with three red circles on the calibration; (b) One of test images (Color figure online)

lines. The real heights of the red corners and the ordinates of them on the image are extracted manually to compose the parameters of the author's method. One of the test images is shown in Fig. 4(b), in which we place the calibration plate about 5 meters away from the camera. We change the angle between the optical axis of the camera and the horizontal plane from  $7^\circ$  to  $24^\circ$  so that we get 18 images of the calibration plate. 48 corners are marked in red circles on each image. We compute the real height of each of the red corners on every image using both Ngoc Hung Nguyen's approach and ours. The experimental results are shown in Figs. 5 and 6.

Obviously, Fig. 5 illustrates that the average relative error of each image obtained by our approach is about 1.81 % while that of the author's method is about 3.39 %, which is nearly twice ours. Figure 6 shows the same result.



**Fig. 5.** The average relative error of the measurement height of 48 red corners on each calibration plate image for both methods: the blue star-like marks indicate the average relative errors of the height computed by our coordinate transformation based method, and the red marks denote Ngoc Hung Nguyen's. (Color figure online)



**Fig. 6.** The average relative error of the estimated height of each red corner on 18 calibration plate images for both methods: the blue star-like marks indicate the average relative errors of the height computed by our coordinate transformation based method, and the red marks denote Ngoc Hung Nguyen's. (Color figure online)

## 4 Conclusion

Our experiments show that human height can be accurately measured by using a calibrated PTZ camera with a measurement system. Compared with other methods, the arrangement of our system is very simple, including calibration of the camera and accurate measurement of the camera's height. Our height estimation algorithm can handle various situations. Especially when the camera moves or rotates instead of being fixed on the wall, our coordinate transformation based method can still work very well as long as the angle between the optical axis and the horizontal plane can be exactly obtained.



**Acknowledgements.** This work was supported by the National Natural Science Foundation of China under Grant Nos: 61273366 and 61231018 and the program of introducing talents of discipline to university under grant no: B13043.

## References

1. Yadav, A., Patil, T.B.M.: Study of imaged based human height measurement & application. *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)* **1**(6), 68–69 (2012)
2. Tsai, H.C., Wang, W.C., Wang, J.C., Wang, J.F.: Long distance person identification using height measurement and face recognition. In: *IEEE Region 10 Conference, TENCN 2009*. IEEE, Singapore (2009)
3. Das, S., Meher, S.: Automatic extraction of height and stride parameters for human recognition. In: *2013 Students Conference on Engineering and Systems (SCES)*. IEEE, Allahabad (2013)
4. Wang, C.M., Chen, W.Y.: The human-height measurement scheme by using image processing techniques. In: *2012 International Conference on Information Security and Intelligence Control (ISIC)*, Yunlin, Taiwan (2012)
5. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge (2003)
6. Zhang Z.: A flexible camera calibration by viewing a plane from unknown orientations. In: *Proceedings of ICCV 1999*, pp.666–673, Corfu, Greece (1999)
7. Nguyen, N.H., Hartley, R.: Height measurement for humans in motion using a camera: a comparison of different methods. In: *International Conference on Digital Image Computing Techniques and Applications*, Fremantle, Western Australia, 3–5 December 2012
8. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: *Proceedings of the 17th International Conference on ICPR 2004*. IEEE, Cambridge (2004)