International Core Journal of Engineering Volume 6 Issue 5, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

Human Height Measurement in Surveillance Video Based on Vision Technology

Fei Yuan[1, a], Junji He[1, b]

[1]Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China.

[a] yuanfei1995@126.com, [b] jjhe@shmtu.edu.cn

Abstract

Traditional human height measurement technology in surveillance video usually requires a three-dimensional scene reconstructed which combined with the camera calibration to get an accurate measurement. Or sometimes, the approximate height of the human body is determined based on the principle of projective geometry without calibration. In this paper, a new visual measurement method is proposed which can accurately measure the height of a human body in a surveillance video without reconstructing a three-dimensional scene. In the case of camera calibration, this method uses the principle of pinhole imaging, and deduces the homography matrix of the human body plane through the distance between the human body plane and the reference wall, and then establishes the human height model. Moreover, in the implementation process, the corner information in the calibration target is fully used to construct the horizontal vanishing point and the vertical vanishing point in space. Then, the overhead points and the perpendicular points needed in the human height measurement model are extracted effectively. The experimental results show that the error of the measurement result can be less than 1.28%, which can meet the needs of human height measurement.

Keywords

Camera calibration; Surveillance video; Vanishing point; Homography matrix; Height measurement.

1. Introduction

Currently, video surveillance system is widely used in safety assurance and accident investigation, such as fire safety management, fire accident investigation, and hydropower station maintenance monitoring. Retrieving surveillance video and extracting its useful information is also an important means in the process of criminal investigation. The appearance, height, figure and walking posture are usually used in the detention of criminal suspects. In general, it is difficult to identify facial features through surveillance video because the criminals intentionally cover up their facial features by occlusion when they commit crimes, or because the human face occupies few pixels in the video image. Based on this, the measurement of human height, body shape and other characteristics is particularly important for the identification of criminal suspects. In view of the fact that the current video surveillance system is composed of a single camera, and the previous criminal investigation process is usually not accurate in estimating the suspect's height, so this paper proposes a method for measuring human height based on monocular vision to solve this problem.

Vision measurement technology can be divided into measurement with calibration and measurement without calibration according to whether the camera needs to be calibrated. For measurements with calibration, the camera must be calibrated in advance to obtain its internal and external parameters and distortion coefficients. The calibration process is tedious, but high accuracy can be obtained. For

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

measurements without calibration, the projective geometry (e.g., cross ratio, collinear, vanishing points and vanishing lines) is used to achieve the purpose of measurement. The implementation process is more flexible, but its accuracy is usually not as good as the results obtained under calibration.

For measurements without calibration, Criminisi [1] et al. proposed for the first time to use the invariance theory of intersection ratio in projective geometry to measure the height of human body. This method requires higher parallel and vertical information in the scene. Later, Zhijie Gan [2] et al. proposed the OTSU adaptive threshold segmentation method and the fast connected area labeling algorithm to obtain the vertical coordinate of the top projection of the head and the coordinates of the landmark points to calculate the height of the human body. This method requires accurate placement of feature points in advance to ensure that the straight line formed by the feature points is strictly vertical. But the operation process is not easy. Caixia Zhang [3] et al. improved on the basis of Criminisi's theory and constructed virtual horizontal and vertical information to replace the horizontal and vertical information in the real scene through the algorithm. Yu Jiang [4] et al. used the parallel relationship of floor tiles in indoor scenes to construct vanishing points, which further improved the measurement accuracy. Although the operation process is simple, the measurement accuracy is far less effective than that obtained under calibration.

For measurements with calibration, Qiulei Dong [5] et al. used a mixed Gaussian model to extract the feature points of human height for real-time human height measurement. Wei Wei [6] et al. proposed the use of key frame information in the video to measure and analyze human height. Caixia Zhang [7] et al. improved the method of selecting feature points at the top of the head and vertical points based on the research by Qiulei Dong et al.. This method uses the distance from each point on the contour of the human body to the vertical vanishing point in the image to select the feature points at the top of the head and the feature points at the foot, and then combines the calibration parameters to obtain the height of the human body.

It can be seen that the above methods either improve the selection of human body feature points, or use each method for comparison between uncalibrated and calibrated cases. The above methods cannot be well solved in the process of human body walking arbitrarily under the surveillance video. Therefore, based on the theory of the pinhole imaging model, a parallel plane homography matrix derivation method is proposed to measure the height of the human body with calibration. The advantage of this method is that the planar homography matrix of the human body at any position in the three-dimensional space can be derived from the distance between the plane of the human body and the reference wall. The method is simple and easy to operate. It has high flexibility and high accuracy.

2. Principle of height measurement based on visual model

2.1 Pinhole imaging model and homography matrix

The core of the video surveillance system is a camera, and its imaging model can be approximated as a pinhole imaging model, as shown in Figure 1. Suppose that any point $P=[X_w, Y_w, Z_w]^{\mathrm{T}}$ in the world coordinate system can be expressed as a point $p=[u,v]^{\mathrm{T}}$ in the image coordinate system after camera imaging. The relationship between the two points is

$$\lambda p = A[R \quad t]P \tag{1}$$

For the formulas above, P and p are the homogeneous coordinates of P and p respectively, λ is the non-zero scale factor, $A=\begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$ is the camera's internal parameter matrix, R is the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

rotation matrix, t is the translation vector. R and t are collectively called the camera's external parameter matrix.

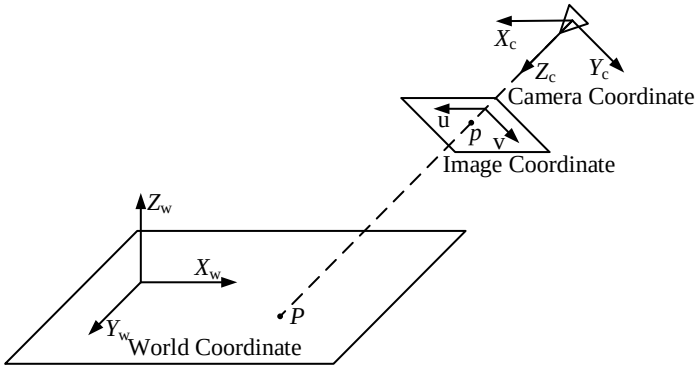


Figure 1. Pinhole imaging model

In order to better describe the position mapping relationship of objects in the world coordinate system and image coordinate system, the concept of homography transformation is introduced. The corresponding transformation matrix is called a homography matrix and is defined as:

$$H = sA\begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \tag{2}$$

where s is a non-zero scale factor, and r_1 and r_2 are the column vectors of the first and second columns of the rotation matrix R.

2.2 Derivation of homography in parallel plane

The position relationship of the human body in the scene is shown in Figure 2.

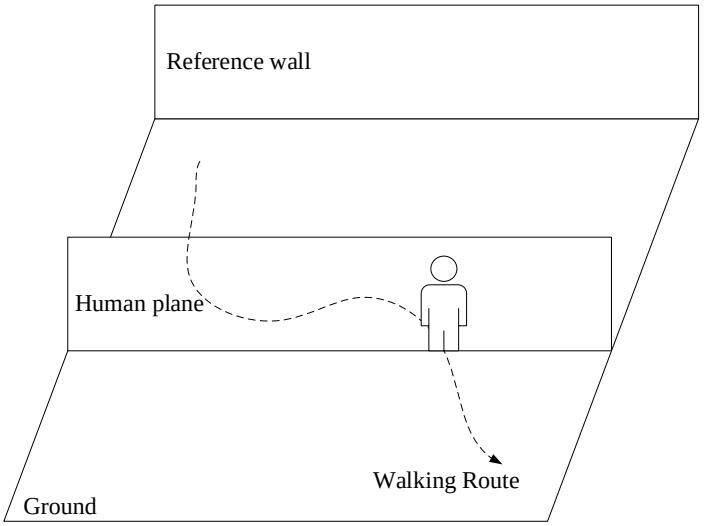


Figure 2. Schematic diagram of the human body in the scene

Among them, the ground is perpendicular to the reference wall, the plane of the human body is parallel to the reference wall, and the dashed line indicates any walking route of the human body in the scene.

Assuming that the human body is walking upright and perpendicular to the ground, the homography matrix H_0 between the reference plane and the camera imaging plane is established on the premise that the reference wall, the human plane and the camera imaging plane are parallel. According to the constraints of the parallelism of the three planes, the homography matrix H_n between the human plane and the camera imaging plane at any position can be derived.

The positional relationship of three parallel planes in three-dimensional space is shown in Figure 3, where W is the projection center, q is the center of the camera imaging plane, and $f = qW$ is the focal length of the camera. Assume that the reference wall is located on the world coordinate system

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

$X_{w0}-Y_{w0}-Z_w$, the human body plane is located on the world coordinate system $X_{w1}-Y_{w1}-Z_w$, and the two planes are parallel in the three-dimensional space coordinate system. Let point P be the vertex of the head in the world coordinate system, then PM is the height of the human body, so $P_0M_0 = P_1M_1$.

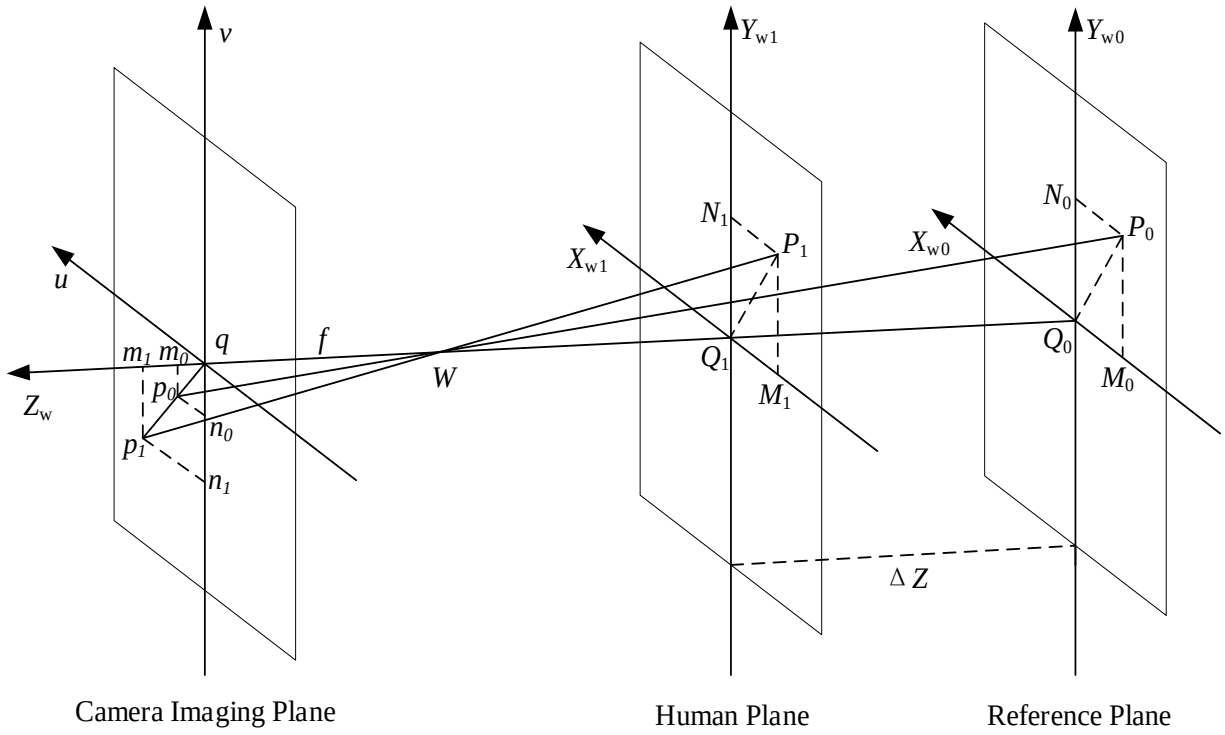


Figure 3. Positional relationship between camera imaging plane, human plane and reference plane

First assume that:

$$\begin{cases} qm_0 = u_0 \\ qm_1 = u \\ Q_0M_0 = Q_1M_1 = X_w \end{cases} \tag{3}$$

From $\Delta P_0Q_0M_0 \sim \Delta p_0qm_0$ and $\Delta P_1Q_1M_1 \sim \Delta p_1qm_1$, it is not difficult to get:

$$\begin{cases} \dfrac{P_0Q_0}{Q_0M_0} = \dfrac{p_0q}{qm_0} \\ \dfrac{P_1Q_1}{Q_1M_1} = \dfrac{p_1q}{qm_1} \end{cases} \tag{4}$$

The following relationship can be obtained from the pinhole imaging model:

$$\begin{cases} \dfrac{P_0Q_0}{p_0q} = \dfrac{Q_0W}{qW} = \dfrac{Z_0}{f} \\ \dfrac{P_1Q_1}{p_1q} = \dfrac{Q_1W}{qW} = \dfrac{Z_0-\Delta Z}{f} \end{cases} \tag{5}$$

where Z_0 is the distance between the camera imaging plane and the reference wall, and ΔZ is the distance between the human plane and the reference wall.

According to equations (4) and (5),

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

$$\begin{cases} \dfrac{Q_0M_0}{qm_0} = \dfrac{Q_0P_0}{qp_0} = \dfrac{Z_0}{f} \\ \dfrac{Q_1M_1}{qm_1} = \dfrac{Q_1P_1}{qp_1} = \dfrac{Z_0 - \Delta Z}{f} \end{cases} \tag{6}$$

That is

$$\begin{cases} \dfrac{X_w}{u_0} = \dfrac{Z_0}{f} \\ \dfrac{X_w}{u} = \dfrac{Z_0 - \Delta Z}{f} \end{cases} \tag{7}$$

Therefore,

$$u = \frac{Z_0}{Z_0 - \Delta Z} u_0 \tag{8}$$

Similarly,

$$v = \frac{Z_0}{Z_0 - \Delta Z} v_0 \tag{9}$$

Equations (8) and (9) are the positional relationships of points on the human body in parallel planes in the image coordinate system. When equations (8) and (9) are written in a matrix form, the following forms can be obtained:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{Z_0}{Z_0 - \Delta Z} & 0 & 0 \\ 0 & \dfrac{Z_0}{Z_0 - \Delta Z} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ 1 \end{bmatrix} \tag{10}$$

Under the premise that the reference wall is calibrated to obtain the homography matrix between the reference wall and the camera imaging plane, the following relationship can be obtained:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{Z_0}{Z_0 - \Delta Z} & 0 & 0 \\ 0 & \dfrac{Z_0}{Z_0 - \Delta Z} & 0 \\ 0 & 0 & 1 \end{bmatrix} H_0 \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \tag{11}$$

Similarly, the homography matrix between the human body plane parallel to the reference wall and the camera imaging plane has the following relationship:

$$H_n = \begin{bmatrix} \dfrac{Z_0}{Z_0 - \Delta Z} & 0 & 0 \\ 0 & \dfrac{Z_0}{Z_0 - \Delta Z} & 0 \\ 0 & 0 & 1 \end{bmatrix} H_0 \tag{12}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

2.3 Distance ΔZ between human plane and reference wall

During the derivation of the homography matrix between any human plane parallel to the reference wall and the camera imaging plane, the distance between the human plane and the reference wall needs to be solved. The horizontal vanishing point is solved using the geometric position relationship in space, as shown in Figure 4.

In the figure, the human foot drop point is recorded as the point P and the intersection of the wall and the ground is the straight line where the AC is located. In the camera imaging plane, the line AB and CD on the ground are perpendicular to the wall, that is:

$$\begin{cases} AB \perp AC \\ CD \perp AC \end{cases} \tag{13}$$

And the line AB and CD intersect at the horizontal vanishing point O. It is not difficult to infer from the principle of projective geometry that the line OP between the horizontal vanishing point O and the human foot point P is also perpendicular to the wall (line AC). The vertical point is denoted as the point M. Then the distance in the world coordinate system of the line MP is the distance ΔZ between the human plane and the reference wall.

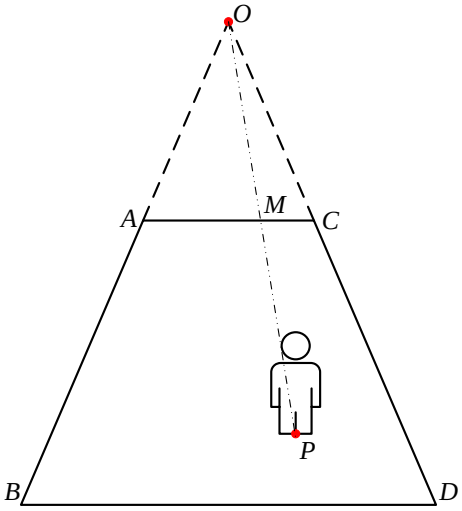


Figure 4. Distance between human plane and reference wall

2.4 Soulution of human height

Assume that the human plane is in the world coordinate system X_wOY_w. According to equations (1) and (2), there is

$$\begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = sH^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{14}$$

Among them, s is a non-zero scale factor, and H is a human plane homography matrix.

Set the homogeneous coordinate of the overhead point $point_h$ at a frame in the surveillance video in the image coordinate system as $[u_1 \quad v_1 \quad 1]^{\mathrm{T}}$, and the homogeneous coordinate of the foot point $point_f$ in the image coordinate system as $[u_2 \quad v_2 \quad 1]^{\mathrm{T}}$. It is not difficult to obtain two points in the world coordinate system. The corresponding coordinate values are recorded as $[X_{w_1} \quad Y_{w_1} \quad 1]^{\mathrm{T}}$ and $[X_{w_2} \quad Y_{w_2} \quad 1]^{\mathrm{T}}$, respectively, then the human height value at this frame is

$$height = \sqrt{(X_{w_1} - X_{w_2})^2 + (Y_{w_1} - Y_{w_2})^2} \tag{15}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

3. Selection of human feature points

3.1 Extraction of human foreground

Because the experiments in this paper are all performed indoors, the background is roughly fixed except for the effects of light. The background difference method is used to extract the foreground of the human body. This algorithm is simple to implement. The average value of consecutive frame images in the video is usually used as the background model

$$f_{bg_*}(x,y)=\frac{\sum_{i=1}^{N} f_*(x,y)}{N} \tag{16}$$

Among them, $*$ represents R, G, B channels.

Using the grayscale image of the current frame and the grayscale image of the background model for the difference operation, the foreground area of the current frame can be obtained by the following formula

$$f_{fg}(x,y)=\begin{cases}1 & \left|f(x,y)-f_{bg}(x,y)\right|>thr\\ 0 & other\end{cases} \tag{17}$$

For the formulas above, $f_{fg}(x,y)$ represents the extracted foreground area of the current frame, $f(x,y)$ represents the grayscale image of the current frame, and $f_{bg}(x,y)$ represents the grayscale image of the background model; thr is a given threshold for binarizing the foreground image.

Finally, the maximum connected domain extraction is performed on the binarized image. The effect shown in Figure 5 can be obtained.

(a) Background modele (b) Current frame (c) Foreground image

Figure 5. Human foreground extraction

3.2 Feature point extraction

When the human body is standing or walking normally, the height of the human body is the vertical distance from the vertex of the human head to the ground, i.e., the distance between the vertex of the human head and the foot point.

The human body contour extracted from the foreground is used to determine the position of the vertex of the human head by the distance from the vertical vanishing point to the human body contour. If the vertical vanishing point is above the contour of the human body, the point with the minimum distance from the vertical vanishing point to the contour of the human body can be regarded as the head of the human body. Similarly, if the vertical vanishing point is below the contour of the human body, the maximum distance from the vertical vanishing point to the contour of the human body can be regarded as the head of the human body.

From the prior knowledge, after the perspective transformation of the camera, the overhead point $point_h$, the foot point $point_f$ of the human body in the image and the vertical vanishing point $point_v$ in the scene should be on the same line. With the constraint condition that the human foot point $point_f$ is in line with the left and right foot points, the position of the human foot point can be obtained, and the extracted effect is shown in Figure 6.

The specific implementation steps of the algorithm are as follows:

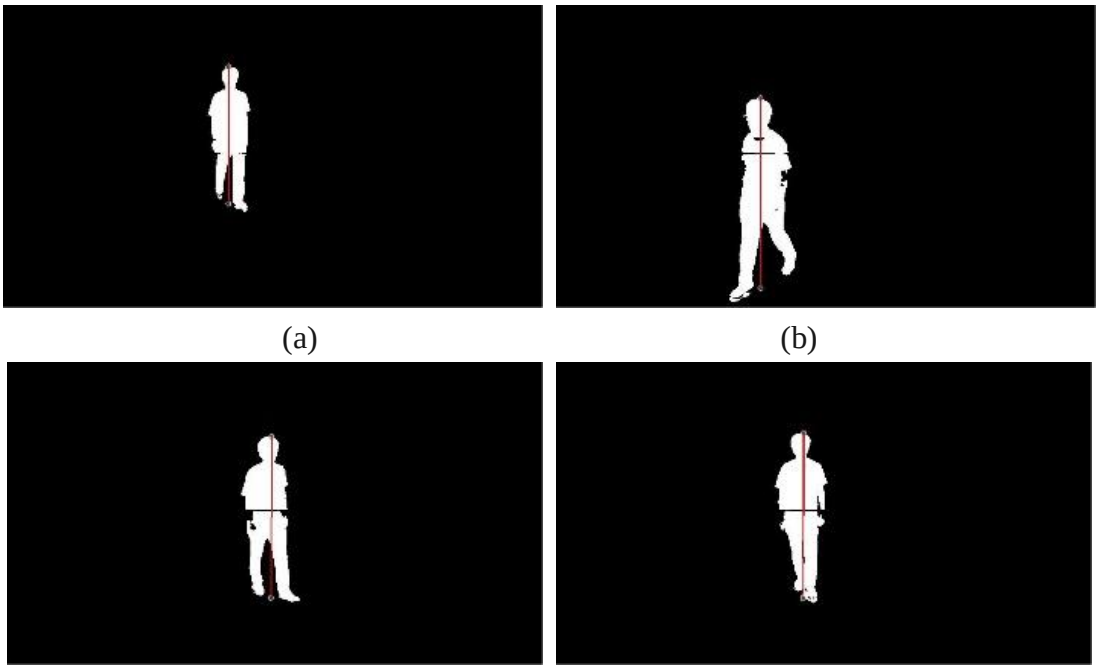
ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

Step1: Use the positional relationship of the corner points in the checkerboard to obtain the vertical vanishing point $point_v$.

Step2: If the vertical vanishing point $point_v$ is above the contour of the human body, take the point with the smallest distance as the overhead point $point_h$, and the point with the largest distance as a foot point. In contrast, if the vertical vanishing point $point_v$ is below the human contour, take the distance with the largest point as the overhead point $point_h$, and the point with the smallest distance as a foot point.

Step3: Divide the contour of the human body into two parts by connecting the vertical vanishing point with the apex of the human head, and then use the method in Step2 to find another foot point.

Step4: Take the line connecting the two foot points as constraint condition 1, the line connecting the vertical vanishing point and the vertices of the overhead point as constraint condition 2, and set up the equations of constraint conditions 1 and 2. The last point of intersection is the foot of the human body.



(a) (b)

(c) (d)

Figure 6. Selection of human feature points

4. Experimental results and analysis

In this paper, Zhengyou Zhang's plane target calibration algorithm is used in the calibration process [8]. The parameters in the camera after calibration are

$$\begin{bmatrix} 1028.24 & 0 & 625.61 \\ 0 & 1060.05 & 317.35 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Experiment 1: Benchmark height measurement

Because the human body is not a rigid body, the measured height will fluctuate during walking. In order to verify the reliability of the algorithm, this paper first measure the benchmarks at different positions. Due to limited paper length, Figure 7 shows the position of the benchmarks of 1.25m, 1.55m and 1.85m.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

1.25m 1.55m 1.85m

(a) Position 1

1.25m 1.55m 1.85m

(b) Position 2

1.25m 1.55m 1.85m

(c) Position 3

Figure 7. Benchmark placement

The measured data are shown in Table 1.

Table 1. Benchmark measurement results

Position	Actual value (mm)	Measured value（mm）	Relative error
Position 1	1250	1240.5	0.76%
	1550	1558.0	0.52%
	1850	1899.6	2.68%
Position 2	1250	1227.2	1.82%
	1550	1562.6	0.81%
	1850	1883.8	1.83%
Position 3	1250	1222.7	2.18%
	1550	1538.5	0.74%
	1850	1872.3	1.21%

It can be seen from the above measurement results that the error of the benchmark between 1250 and 1850mm is within 3.0%. In most cases, the relative error of the benchmark height can be controlled within 2.0%. Some points have larger errors, which may be due to the larger base area of the benchmark and the lowest point of the benchmark was not well extracted. The actual height of human body occupies a large proportion within 1600 ~ 1850mm, so the algorithm proposed in this paper is feasible.

Experiment 2: Different human height measurements

Based on Experiment 1, three people with different heights (1598mm, 1736mm, and 1832mm) were selected for human height measurement experiments. Figure 8 shows the height curves of three people

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

of different heights during walking in 1 second, i.e., the height measurement values corresponding to each frame of images of people with different heights during walking for 25 consecutive frames.

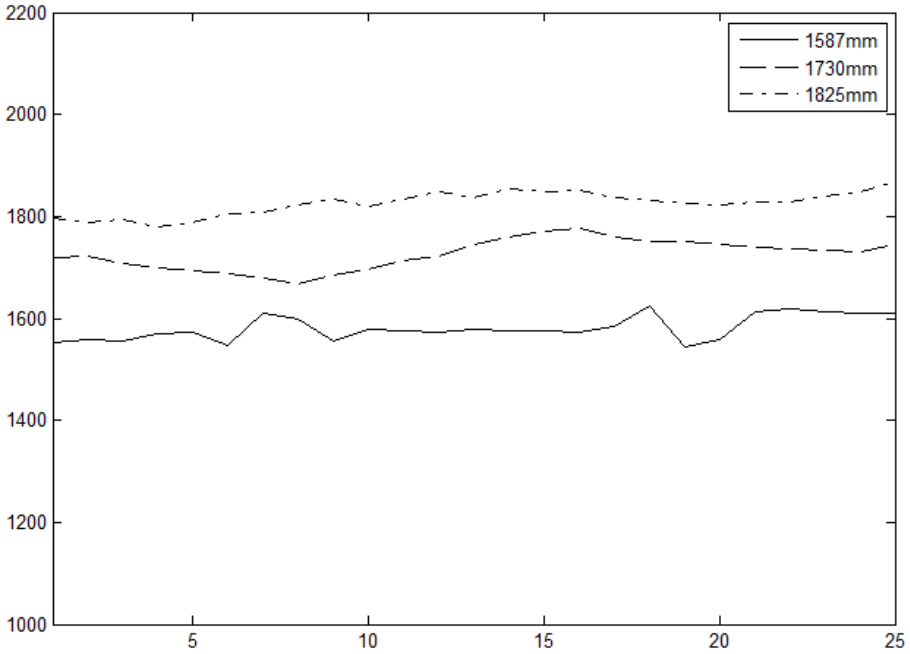


Figure 8. Measurement curves of different heights

The average of the measured values of the human body height in the 25 consecutive images is taken as the final result of the measurement. The measured values and relative errors are shown in Table 2.

Table 2. Measurement results of humans of different heights

Actual height (mm)	Measured height (mm)	Relative error
1598.00	1580.82	1.08%
1736.00	1725.17	0.62%
1832.00	1825.44	0.36%

The above measurement results can be found that although the height of the human body at different positions may be larger than the actual height of the human body or smaller than the actual height of the human body, the average value of the height measurement value can reasonably eliminate part of the measurement deviation. So that the final measurement result can better meet the actual measurement needs.

Experiment 3: Measurements on different walking routes

In order to verify that the algorithm in this paper is suitable for walking measurement under various routes, and the derivation of the homography matrix is correct, this paper conducts height measurement experiments on human walking under different walking routes. The height measurement of the human body is performed, and the schematic diagram of the route walking is shown in Figure 9.

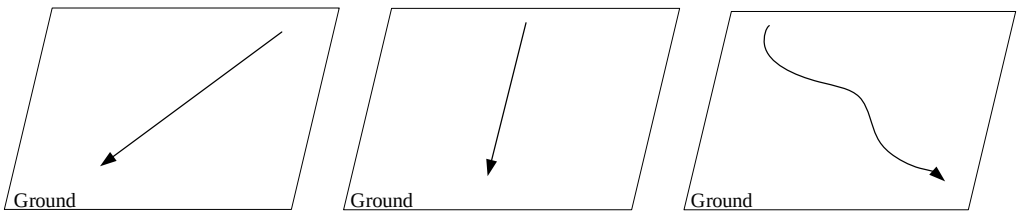


Figure 9. Experimental route diagram

ISSN: 2414-1895 DOI: 10.6919/ICJE.202005_6(5).0030

As in Experiment 2, this experiment performed a height measurement on the human body with the height of 1736mm under the walking route shown above. The measurement result is the average of 25 consecutive frame measurements. The results and relative errors are shown in Table 3.

Table 3. Measurement results under different walking routes

Actual height (mm)	Measured height (mm)	Relative error
1736.00	1722.59	0.77%
	1758.18	1.28%
	1727.37	0.48%

The measurement results show that the measurement results of the algorithm in this paper are stable, and good measurement results can be obtained for the measurement of human height during any walking process. The relative error can be controlled within 1.28%.

5. Conclusion

This paper proposes a new method for measuring human height in a video surveillance system. This method uses the distance from the human plane to the reference wall to derive the homography matrix of the human plane, and then obtains the human height. This method does not require the reconstruction of a three-dimensional scene, and is easy to operate. At the same time, the measurement accuracy is higher than that of the uncalibrated case. The experimental results show that the method proposed in this paper has relatively accurate measurement results of human height measurement under any walking route, and the relative error of the measurement can be guaranteed to be within 1.30%, which can basically meet the needs of human height measurement. The research results of this paper have important reference value for further development of human height measurement in indoor and outdoor scenes.

References

[1] Criminisi A. Accurate Visual Metrology from Single and Multiple Uncalibrated Images[M]. Springer Press,2002.

[2] Zhijie Gan, Yun Liu. Real-time Stature Measurement Algorithm Based on Monocular Vision Technique [J]. Journal of Qingdao University of Science and Technology (Natural Science Edition), 2008, 029(004):366-369.

[3] Caixia Zhang, Huanli Fu. Visual Metrology for Height of Pedestrian from Uncalibrated Video[J]. Computer Engineering and Applications, 2017, 53(21):162-166.

[4] Yu Jiang. Human Height Measurement Based on Monocular Vision[J]. Computer Knowledge and Technology, 2017, 13(13):164-165+170.

[5] Qiulei Dong, Yihong Wu, Zhanyi Hu. Video-based Real-time Automatic Human Height Measurement[J]. Acta Automatica Sinica, 2009, 35(02):137-144.

[6] Wei Wei, Yan Lv, Kehong Shi. Body Height Measurement Algorithm in Key Frame from Surveillance Video[J]. Computer Systems Applications, 2012, 21(6):195-198

[7] Caixia, Yingdong Gu. Dynamic Human Height Measurement under Video[J]. Journal of North China University of Technology, 2014, 26(01):10-15.

[8] Zhang Zhengyou. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11):1330-1334.

[9] Jingjing Wei, Junji He. Calibration Method for Height Detection of Yard Container Based on Monocular Vision[J]. Industrial Control Computer, 2015, 28(10):95-96+99.