MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Measurement of human stature from surveillance camera based on projective geometry

Takuya Inoue
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi,
Kanagawa-ken, 223-8522, Japan
inoue@hvrl.ics.keio.ac.jp

Hideo Saito
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi,
Kanagawa-ken, 223-8522, Japan
saito@hvrl.ics.keio.ac.jp

Makoto Kimura
Samsung Yokohama Research Institute Co.,Ltd.
2-7 Sugasawa-cho, Tsurumi-ku, Yokohama-shi,
Kanagawa-ken, 230-0027, Japan
m.kimura@samsung.com

Abstract

In this paper, we propose a novel measurement method of human stature by single view surveillance camera with minimum and sufficient effort of camera calibration. In this method, we calibrate the camera only towards the vertical direction to the ground by using multiple poles with known length which are placed at the arbitrary positions. Even though the horizontal positions of the poles are unknown, we can obtain the geometrical information which is sufficient to estimate the vertical length of the human according to the length of the poles. In the estimation of the human stature, we compute the homography between the plane along with the tips of the poles (pole-top plane) and the ground plane. Using this homography, we can determine any image point that has the same height from the ground, then we can determine the point of the pole length height in the human region. By comparing with the point and the head top point of the human, we can compute the stature with real scale. We have experimentally confirmed that proposed method can be more accurate and easier measurement than the measurement by the traditional camera calibration method using a chess board.

1 Introduction

In recent years, cameras have generally been used for detecting human features. For instance, Li et al. propose the vision-based gender recognition system from the gait for commercial strategy[1]. Among these researches, image-based human feature detection is greatly investigated like face detection or recognition[2].

The information acquired from images has been limited in qualitative features: cloth color, hair style, physique, and so on. Additional quantitative features of human is useful for sports practice supporting, or identifying oneself[3]. Especially, human stature information is quite useful to identify because it is one of the most basic biological features.

Recently, fixed surveillance cameras are set up in many streets or buildings. The human stature information estimated from those cameras is convenient for detecting suspicious character.

In general, images are shot in perspective, hence real scale of human stature is not estimated directly from images. If a surveillance camera is already calibrated, human stature can be calculated with the assumption that human stands vertically on the ground. However, surveillance camera is rarely calibrated in practice. This means calibration operation is required as pre-treatment or post-treatment.

For camera calibration, some calibrating objects are generally used. Jeges et al. proposed human stature estimation method[4] based on Zhang calibration method[5], which uses the chessboard pattern on the plane object(Figure 1).

In such calibration methods, the size of the calibrating object must be close to the size of the processing target in the scene. In the purpose of measurement of the human stature, the chessboard must be similar size of human, fully rigid, and highly accurate. In addition, human moves in places in the shooting scene, so chessboard have to be shot in many times with changing board's direction and positions in order to keep the high accuracy. Thus, camera calibration in human scale is a hard work. Consequently, high accuracy has not been expected in estimation of human stature.

In this paper, we propose a novel method which estimates human stature from a single view image using projective geometry, which is calculated based on multiple rod-like objects(Figure 2). These rod-like objects size is supposed to be similar length of human

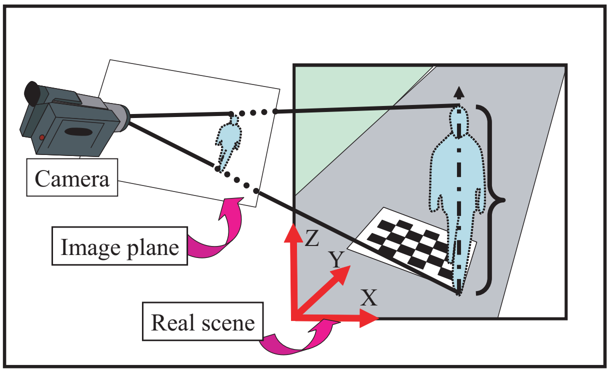


Figure 1. Conception of chessboard calibration method.

stature. In comparison with chessboard, a pole of human scale length is easily available, fully rigid, accurate, and handy to transport. Generally, poles can be easily set up in the real environment, so that we can apply ideal calibration in anywhere in the vast scene. Therefore, we can expect the proposed method is more accurate and more stable than chessboard calibration methods.

2 Proposed method

We propose human stature estimation from a single view image. For such purpose, the camera model must conform to the perspective projection, in which the scale in the real world is not in proportion to the scale in the 2D image. Therefore it is difficult to estimate the human stature by only comparison between known object size and human size in the 2D image. The projective geometry is one of the way to handle the camera projection model, but we need to get the real scale value for estimating the human stature.

In this paper, some natural prior conditions in the scene are assumed: the ground is flat, and human is standing upright on the ground. In the target area in the scene, we set up four or more poles whose lengths are common, similar to human stature, and already known. We can also use one pole that placed at different positions instead of using multiple poles. The poles can be set up arbitrary places, but they must be set up vertically on the ground.

Figure 3 shows an example input image for estimating human stature. From the input image, we can get the top position and grounding position of each poles. By estimating the projective geometry from the positions and the grounding positions of the poles, we can obtain the relationship between poles and the human stature that is expressed by homographies of multiple planes.

In this method, input data is each of the image coordinates which are manually entered: top position and grounding position of each poles and head and foot position of human. Then, stature value is estimated as a processing result.

Figure 2. Sample of poles(tripods).

Figure 3. Condition of shooting with four tripods.

2.1 Homography estimation between ground and pole-top plane

As described above, the height of the poles h is already given as the precondition. The homography $\mathbf{H}$ between the ground plane and the pole-top plane can be computed from top points($\boldsymbol{p}_i$) and grounding points($\boldsymbol{q}_i$) correspondence in the image coordinates $(1 \le i \le n)$. n is the number of poles. In this computation, we do not need to know the horizontal position of each pole in the scene.

As shown in Figure 4, the human feet point $\boldsymbol{m}_f$ can be transformed to the point $\boldsymbol{m}_m$ on pole-top plane by the computed homography $\mathbf{H}$. $\boldsymbol{m}_m$ is represented as $\mathbf{H}\boldsymbol{m}_f$. The point $\boldsymbol{m}_m$ must be the height of h in the real world. In this paper, we manually detect the pole's top and grounding points and human feet point in the image.

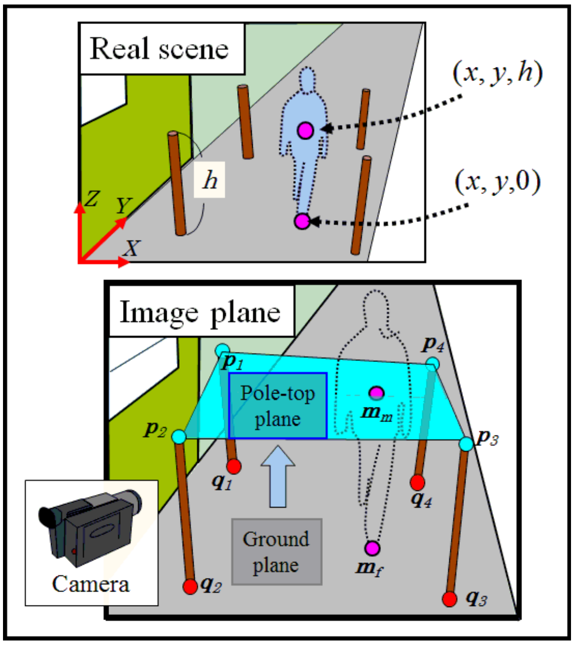


Figure 4. Homography calcuration between ground and pole-top plane.

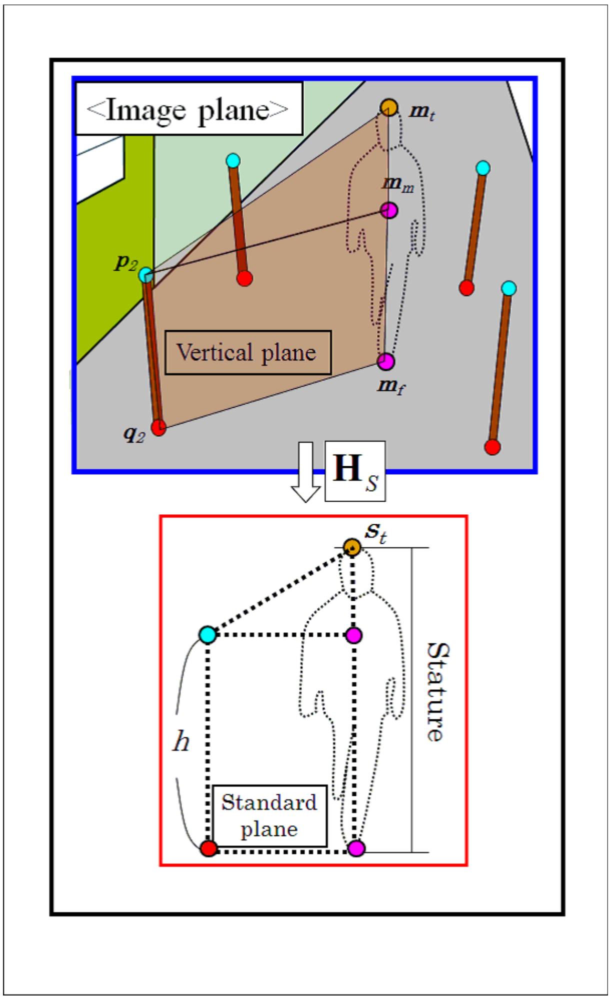


Figure 5. How to get human stature.

2.2 Homography estimation between human and a pole

As shown in Figure 5, we can consider a 2D quadrangle composed by $\boldsymbol{m}_m$, $\boldsymbol{m}_f$, and any pole's $\boldsymbol{q}_i$, $\boldsymbol{p}_i$. In this paper, we adopt a pole which makes maximum quadrangle with the points $\boldsymbol{m}_m$ and $\boldsymbol{m}_f$. In the real world, the 2D quadrangle corresponds to a 3D vertical rectangle of height h, because both of the segment $\boldsymbol{m}_m$-$\boldsymbol{m}_f$ and the pole are vertical to the ground.

Here, we assume a virtual standard plane with a rectangle of height h and arbitrary width, so that we can calculate the homography $\mathbf{H}_S$ as the translating matrix from vertical plane to standard plane from four point's ($\boldsymbol{m}_m$, $\boldsymbol{m}_f$, $\boldsymbol{q}_i$, and $\boldsymbol{p}_i$) and vertexes of standard plane point's correspondence.

Next, we detect the head position $\boldsymbol{m}_t$ on the image manually, so that the coordinate can be transformed into $\boldsymbol{s}_t$ in the standard plane by $\mathbf{H}_S$. $\boldsymbol{s}_t$ is represented as $\mathbf{H}_S\boldsymbol{m}_t$. On the standard plane, height scale is corresponding to the real world, because we used the pole's known height h. Therefore, the human stature in the real scale can be calculated by $\boldsymbol{s}_t$ on the standard plane.

3 Experiments

In this section, we demonstrate three experiments using real images. At first, the proposed method is applied in outdoor field in order to check the accuracy. Next, the proposed method is applied to the indoor camera set up like surveillance camera. Finally, the accuracy of the proposed method is compared with the existing calibration method using chessboard pattern. In all experiments, input resolution is VGA (640×480).

3.1 Outdoor field

In order to evaluate the accuracy of the proposed method, human stature was estimated in the outdoor scene (Figure 6(a)). For the camera calibration, we used four tripods shown in Figure 2. The true value of target human stature was 1.69m, and the length of the tripods were 0.97m. The size of area where there was a human was about 1.5m×3.0m.

In ten times trial of this condition, the error was always less than 30mm (about 2%). Therefore, human stature was estimated through the easy camera calibration.

3.2 Indoor scene

For simulating the indoor surveillance camera, the human stature was estimated in the scene like Figure 6(b). The calibration objects and the target person were same as section 3.1. The size of area where there was a human was about 2.0m×5.0m.

In this case, the position of the human was changed every time in the trial. In ten times trial of this condition, error rate was always less than 30mm again. Therefore, human stature was estimated as well as 3.1.

3.3 Comparison with strong calibration

In order to compare the proposed method with Zhang method for strong calibration using a chessboard[5], the human stature was estimated based on the strong calibration. For strong calibration, A0 size chessboard pattern was used as shown in Figure 6(c). The position of the human was changed every time in ten times trial. The size of area where there was a human was about 2.0m×5.0m. At camera calibration, chessboard was shot in twenty times with different posture and location.

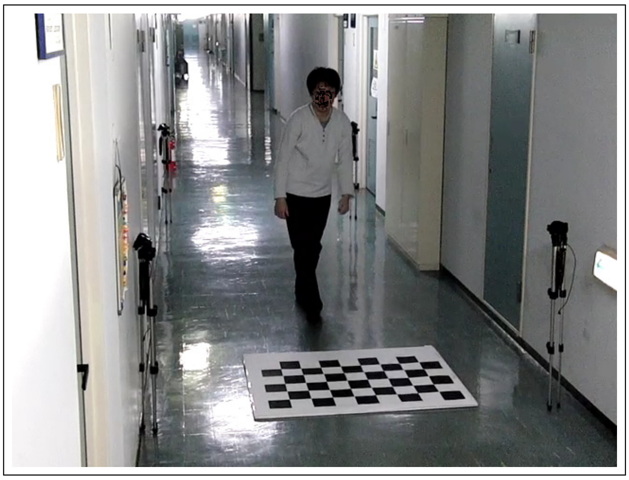
In the method using a chessboard, the maximum error rate was about 3%. On the other hand, in our proposed method, the maximum error rate was less than 2%. In this experiment, the accuracy of proposed method was comparable to the strong calibration method using chessboard or more. In the proposed method, multiple poles can be easily set up in vast environment in real space, so that accuracy of estimation is stable in vast scene compared with the method using chessboard.

4 Conclusion

In this paper, we proposed a novel human stature measurement method from single view image based on

(a) Experimental scene in outdoor field

(b) Experimental scene in indoor environment



(c) Experimental scene with chessboard

Figure 6. Condition of shooting environment

projective geometry which is calibrated by using multiple poles. In the proposed method, human stature is estimated easily and accurately in vast scene.

In consideration of the human size, strong calibration is difficult and inaccurate to operate in real situation because it requires human size object (e.g. chessboard pattern). In addition, it is difficult to calibrate the camera using chessboard because the target scene should be vast environment.

In contrast, the proposed method requires just putting poles in the scene. Though the camera calibration limited in vertical direction using poles, proposed method is specialized for estimating human stature and can be applied in vast environment. In the experiments, four poles were used. However, one pole can be used with changing position instead of using multiple poles. Thus, the proposed method is easy to operate and easy to prepare.

In the future work, the preconditions about human pose – standing upright on the ground – should be eased. The human stature in natural motion (e.g. walking) can not be measured in the current method. The major difficulty of the walking human's stature measurement is the vertical body motion. In such a case, human stature may be estimated from the measured values in multiple frames.

Currently, all image coordinates (the pole's top, the pole's grounding, the human foot and the human head) are given in manual. In the future work, they should be detected automatically[6][7].

Acknowledgment

This research was supported by National Institute of Information and Communications Technology, Japan.

References

[1] X. Li, S. J. Maybank, S. Yan, D. Tao, D. Xu,: "Gait Components and Their Application to Gender Recognition," IEEE Transactions on Systems, Man, and Cybernetics.Part C: Applications and Reviews, vol. 38, no. 2, pp.145–155, 2008.

[2] J. Wright, A.Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma,: "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, no.2, pp.210–227, 2008.

[3] Y. Iwashita, R. Kurazume: "Person identification from human walking sequences using affine moment invariants," IEEE International Conference on Robotics and Automation, 2009.

[4] E. Jeges, I. Kispal, Z. Hornak: "Measuring human height using calibrated cameras," Conference on Human System Interactions, pp.755–760, 2008.

[5] Z. Zhang: "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.11, pp.1330–1334, 2000.

[6] H. Ando, H. Fujiyoshi: "Human-Area Segmentation by Selecting Similar Silhouette Images Based on Weak-Classifier Response," International Conference on Pattern Recognition, 2010.

[7] Z. Zhang, P. L. Venetianer, A. J. Lipton: "A Robust Human Detection and Tracking System Using a Human-Model-Based Camera Calibration," The Eighth International Workshop on Visual Surveillance, 2008.