

RESEARCH

Open Access



# A simplified nonlinear regression method for human height estimation in video surveillance

Shengzhe Li<sup>1</sup>, Van Huan Nguyen<sup>2</sup>, Mingjie Ma<sup>2</sup>, Cheng-Bin Jin<sup>1</sup>, Trung Dung Do<sup>1</sup> and Hakil Kim<sup>1\*</sup>

## Abstract

This paper presents a simple camera calibration method for estimating human height in video surveillance. Given that most cameras for video surveillance are installed in high positions at a slightly tilted angle, it is possible to retain only three calibration parameters in the original camera model, namely the focal length, the tilting angle and the camera height. These parameters can be directly estimated using a nonlinear regression model from the observed head and foot points of a walking human instead of estimating the vanishing line and point in the image, which is extremely sensitive to noise in practice. With only three unknown parameters, the nonlinear regression model can fit data efficiently. The experimental results show that the proposed method can predict the human height with a mean absolute error of only about 1.39 cm from ground truth data.

**Keywords:** Camera calibration; Soft biometrics; Human height estimation; Video surveillance

## 1 Introduction

Advances in the image resolution and quality of digital cameras in the last few years have increased the image analysis capability of modern video surveillance systems. Estimating human height is an essential task in video surveillance because it enables many practical applications such as soft biometrics and forensic analyses [1–6]. The key idea behind this technology is a camera calibration system containing a set of parameters for transforming real-world coordinates into image coordinates and vice versa. It is natural to associate a walking or standing human with the camera calibration problem in the context of video surveillance for the following two reasons: a walking or standing person is basically vertical, and his or her height is known. Several camera calibration methods based on walking humans have been proposed. Most such methods rely on estimating vanishing points from walking human. However, as Micusik et al. reported in [7], estimating the vanishing points is usually the bottleneck of these methods because it is extremely sensitive to noise.

Lv et al. [8, 9] proposed a self-calibration method for estimating camera's intrinsic and extrinsic parameters. Their method of computing calibration parameters relies on three vanishing points that can be estimated from a set of automatically extracted head-feet pairs in the video. The initial projection matrix is then refined by minimizing the distance from the original and reprojected head points by using a nonlinear optimization algorithm. Lv's work has inspired many similar methods [7, 10–13].

Krahnstoever et al. [10] introduced a homology-based method. Homology is the transformation from the foot plane to the head plane and contains all geometric information necessary to recover the whole projective matrix in the camera model. The initial projective matrix is updated using a Bayesian framework to obtain estimated parameters. Junejo et al. [11] employed a homology-based method to recover a projective matrix with some modification in the outlier removal stage in which outliers are removed by the Rayleigh quotient.

As reported in Micusik et al. [7], a drawback of the aforementioned method is that it relies on estimating three vanishing points, which is usually the bottleneck of approaches because it is extremely sensitive to noise. Even negligible inaccuracy in a vanishing point can cause huge inaccuracy in the estimate of the focal length, by

\*Correspondence: hikim@inha.ac.kr

<sup>1</sup> Information and Communication Engineering, Inha University, 100 Inha-ro, Nam-gu, 22212 Incheon, Korea

Full list of author information is available at the end of the article

up to 100 %. Therefore, these approaches have limited use in practice. To overcome this problem, Micusik et al. introduced an improved approach based on the quadratic eigenvalue problem without estimating the vanishing points. According to their experiment, this approach outperforms other approaches such as vanishing point-based or the standard eight point-based approaches.

Liu et al. [12, 13] proposed a more automated method for calibrating surveillance cameras based on prior knowledge of the distribution of human heights. The main idea behind this method is based on the observation that objects (pedestrians) in a scene are all roughly the same height. This method is practical in applications that do not require highly precise camera calibration.

Recently, many studies have verified the robustness of camera calibration methods based on vanishing points [14–16]. These methods assume that images are from a “Manhattan” scene with an orthogonal structures and estimate the vanishing points from the scene. Given vanishing points and reference height information, the human height can be computed straightforwardly. The proposed method provides an alternative solution which does not rely on computing vanishing points. This is useful in some cases where vanishing points are difficult to compute.

This study presents a camera calibration method for estimating the human height in video surveillance. In order to provide the best field of view, most surveillance cameras are set at high locations with low tilt angles. Only three camera parameters, namely, the focal length, the tilting angle, and the camera height, are effective with this installation. In the proposed method, these parameters are directly calculated using a nonlinear regression model from the observed head and foot points of a walking human instead of estimating vanishing line and points, which are extremely sensitive to noise in practice. Unlike other methods that estimate all parameters in the camera model, the proposed method estimates only three parameters. With only three unknown parameters, the nonlinear regression model can provide an efficient fit to the data. The experimental results show that the proposed method can predict the human height with a mean absolute error of only about 1.39 cm from ground truth data.

The main advantage of the proposed method is that it provides the simplest solution for the camera calibration problem in video surveillance in comparison to other methods. More specifically, the proposed method 1) does not require vanishing line, which is in generally difficult to estimate and generates many errors in practice, 2) takes only three parameters (the focal length, the tilting angle, and the camera height), and 3) uses no calibration objects, including parallel or perpendicular lines on the ground. These advantages are increasingly valuable

because a growing number of surveillance cameras are being installed and the proposed method can save a lot of time calibrating them.

The rest of this paper is organized as follows: Section 2 describes the proposed method for calibrating cameras and estimating the human height in video surveillance. Section 3 presents the experimental results from the walking human and the ruler-based evaluations. Section 4 analyzes errors, and Section 5 concludes the paper.

## 2 Proposed method

The original pin-hole camera model consists of five intrinsic and six extrinsic parameters such that the intrinsic parameters describe inner properties of the camera, such as the focal length and skewness, and the extrinsic  $\mathbf{I}$  describe the translation and rotation of the camera center from the world coordinate system to the camera  $n$  system [17]. The original camera model is given by

$$\mathbf{P} = \mathbf{K} \cdot \mathbf{R} \cdot [\mathbf{I}|\mathbf{t}], \quad (1)$$

where  $\mathbf{R}$  is the rotation matrix and  $\mathbf{t}$  is the translation vector.

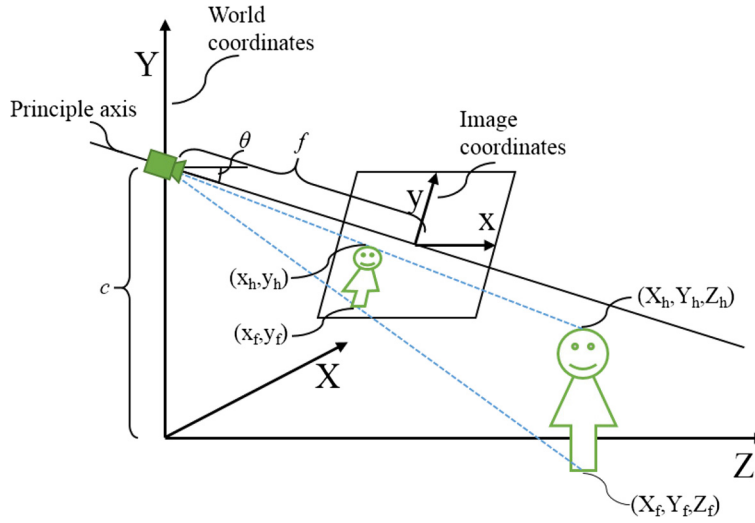
Most cameras for video surveillance are installed at high positions with a slightly tilted angle to ensure the best field of view. Figure 1 shows this type of camera installation and the coordinate system. In this installation, rotation angles along the  $Y$ - and  $Z$ -axis can be assumed as 0 (which are also known as pan and roll), and translations along  $X$ - and  $Z$ -axis can also be assumed as 0. Therefore, the original camera model  $\mathbf{P}$  can be simplified as

$$\mathbf{P} = \mathbf{K} \cdot \mathbf{R}_X \cdot [\mathbf{I}|\mathbf{c}_Y], \quad (2)$$

where  $\mathbf{R}_X$  is the rotation matrix of the camera tilt and  $\mathbf{c}_Y$  is the translation vector along the  $Y$  direction. To further reduce the number of calibration parameters in  $\mathbf{K}$ , zero skew, unit aspect ratio, and known principle points  $[0, 0]^T$  are assumed. Then the camera matrix  $\mathbf{P}$  can be written as

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -c \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} f & 0 & 0 \\ 0 & f \cos \theta & -f \sin \theta \\ 0 & f \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & -c \sin \theta \end{bmatrix}, \end{aligned} \quad (3)$$

where  $f$  is the focal length,  $\theta$  is the tilt angle, and  $c$  is the height of the camera. These three parameters can



**Fig. 1** Camera coordinate system. A typical camera installation and the coordinate system in video surveillance

determine the mapping from world coordinates  $[X, Y, Z]^T$  to image coordinates  $[x, y, w]^T$  as follows:

$$\begin{aligned} \begin{bmatrix} x \\ y \\ w \end{bmatrix} &= \mathbf{P} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f \cos \theta & -f \sin \theta & -fc \cos \theta \\ 0 & \sin \theta & \cos \theta & -c \sin \theta \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4) \\ &= \begin{bmatrix} fX \\ f \cos \theta \cdot Y - f \sin \theta \cdot Z - fc \cos \theta \\ \sin \theta \cdot Y + \cos \theta \cdot Z - c \sin \theta \end{bmatrix}, \end{aligned}$$

which can be represented in Cartesian coordinates as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} fX/(\sin \theta \cdot Y + \cos \theta \cdot Z - c \sin \theta) \\ (f \cos \theta \cdot Y - f \sin \theta \cdot Z - fc \cos \theta)/(\sin \theta \cdot Y + \cos \theta \cdot Z - c \sin \theta) \end{bmatrix}. \quad (5)$$

The walking human in the camera view provides a set of head and foot points in the image plane and a physical height. The walking human is vertical to the ground. Therefore, the  $y$ -coordinate offers more information than the  $x$ -coordinate and can be associated with the physical height of the human. In this regard, the bottom equation in Eq. (5) gives a basic relationship between world coordinates  $Y$  and  $Z$  and the image coordinate  $y$ :

$$y = \frac{f \cos \theta \cdot Y - f \sin \theta \cdot Z - fc \cos \theta}{\sin \theta \cdot Y + \cos \theta \cdot Z - c \sin \theta}, \quad (6)$$

if  $\cos \theta \neq 0$ ,

$$y = \frac{fY - f \tan \theta \cdot Z - fc}{\tan \theta \cdot Y + Z - c \tan \theta}. \quad (7)$$

Because each head-foot pair of the  $y$ -coordinate, denoted as  $y_h$  and  $y_f$ , can be measured from the image. Applying Eq. (7) provides a set of equations with three unknowns:

$$\begin{cases} y_f = \frac{fY_f - f \tan \theta \cdot Z_f - fc}{\tan \theta \cdot Y_f + Z_f - c \tan \theta}, \\ y_h = \frac{fY_h - f \tan \theta \cdot Z_h - fc}{\tan \theta \cdot Y_h + Z_h - c \tan \theta}, \end{cases} \quad (8)$$

where  $Y_f = 0$  and  $Y_h$  is  $Y$  coordinate of the head, which is the known physical height of the human, and  $Z_f$  and  $Z_h$  are  $Z$  coordinates of the head and the foot. In practice, measuring  $Z$  requires additional grids or objects on the ground and is more difficult than measuring  $Y$  which is the known height of the human. Therefore, the variable  $Z$  in Eq. (8) is eliminated by substituting  $Z_h$  in the bottom equation with  $Z_f$  in the above equation. This yields an equation containing only  $y_f$  and  $y_h$ :

$$y_h = \frac{f(-c \tan^2 \theta + Y_h - c) y_f + f^2 \tan \theta \cdot Y_h}{\tan \theta \cdot Y_h y_f + f(\tan^2 \theta \cdot Y_h - c \tan^2 \theta - c)}. \quad (9)$$

For real data, the right-hand side of Eq. (9) actually gives an estimated value of  $y_h$  which is denoted as the estimation function  $\hat{y}_h$ . This function takes two arguments  $y_f$  and  $Y_h$ :

$$\hat{y}_h(y_f, Y_h) = \frac{f(-c \tan^2 \theta + Y_h - c) y_f + f^2 \tan \theta \cdot Y_h}{\tan \theta \cdot Y_h y_f + f(\tan^2 \theta \cdot Y_h - c \tan^2 \theta - c)}. \quad (10)$$

Given that real data always come with noise, the observed value of  $y_h$  can be rewritten as

$$y_h = \hat{y}_h(y_f, Y_h) + \epsilon \quad (11)$$

where  $\epsilon$  is the error produced by calibration parameters. Minimizing  $\epsilon$  gives the optimal parameters.

Note that Eq. (10) has a nonlinear form and its parameters can be found by a nonlinear regression:

$$\begin{bmatrix} \hat{f} \\ \hat{\theta} \\ \hat{c} \end{bmatrix} = \arg \min_{f, \theta, c} \sum_{i=1}^N (\hat{y}_{hi} - y_{hi})^2. \quad (12)$$

There are many algorithms for solving this type of problem, including the Levenberg-Marquardt algorithm. Initial parameters  $\theta_0$  and  $c_0$  can be easily approximated through visual measurement, and  $f_0$  can be set as 0.5–1.5 times the image height if the real-world length unit is in centimeter.

Once calibration parameters  $\hat{f}, \hat{\theta}, \hat{c}$  of a camera are obtained, the physical height of a person can be estimated

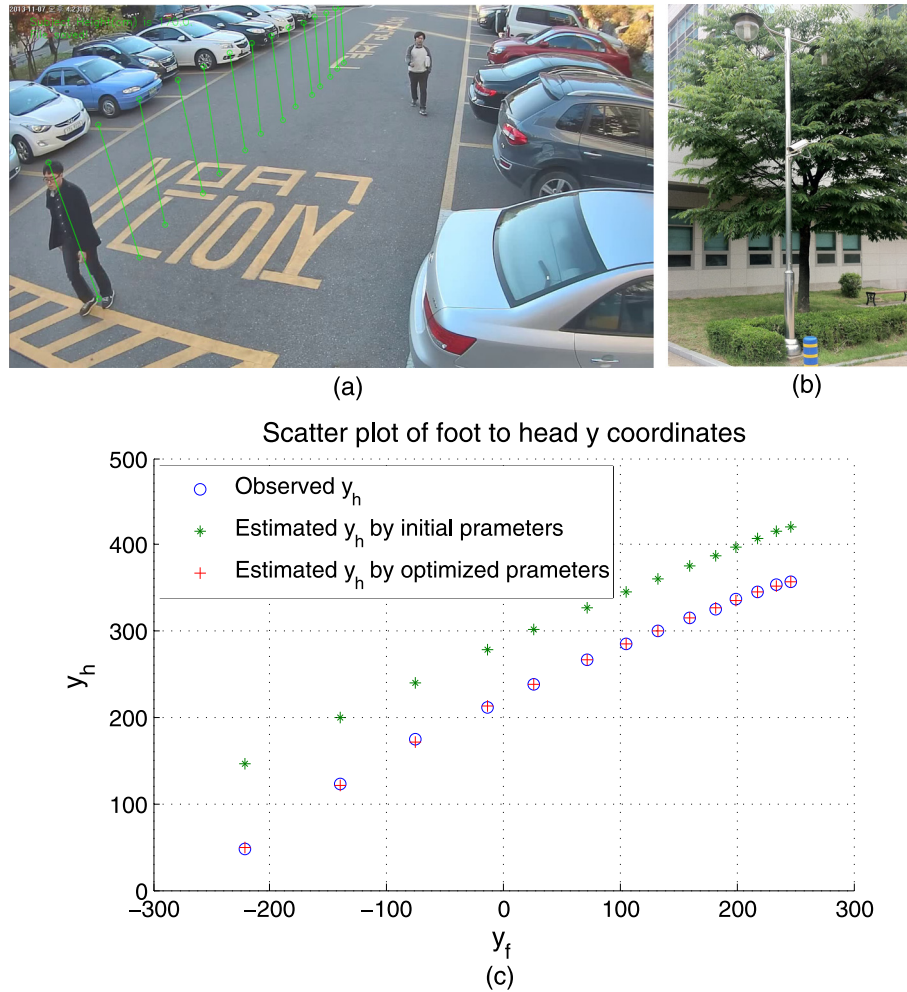
from a pair of head and foot points observed from the image. As in the case of Eq. (10), the estimated physical height  $\hat{Y}$  can be written as a function of  $y_f$  and  $y_h$ :

$$\hat{Y}(y_f, y_h) = \frac{-\hat{f}\hat{c}(\tan^2 \hat{\theta} + 1) \cdot (y_f - y_h)}{\tan \hat{\theta} \cdot y_f y_h - \hat{f} y_f + \hat{f} \tan^2 \hat{\theta} y_h - \hat{f}^2 \tan \hat{\theta}}. \quad (13)$$

### 3 Experimental results

#### 3.1 Experiment setup

Two types of experiments were conducted to evaluate the accuracy and robustness of the proposed method: 1) an evaluation based on the walking human and 2) an evaluation based on the ruler. A dataset was collected from a video surveillance site in use. Cameras at the site were installed at entrances and corridors of a building as well



**Fig. 2** An example of the proposed camera calibration system. **a** A sample video was taken, and some head and foot points were manually marked. **b** The camera used for collecting the dataset. **c** A scatter plot of the y-coordinate of observed and estimated head points with respect to the observed foot points. The initial parameters  $f = 720$ ,  $\theta = -30$ , and  $c = 300$  are approximated by a visual estimation of the installed camera in **(b)**, and optimal parameters are found as  $f = 547.7$ ,  $\theta = -38.6$ , and  $c = 270.2$  based on the nonlinear regression method



as at an outside parking lot. The video resolution for this dataset was  $1280 \times 720$ . For each camera, 15 pairs of points were collected in the ruler-based evaluation, and 5–30 pairs of points were collected in the walking human-based evaluation. These points were collected in a broad range of camera view, and they covered near and far positions.

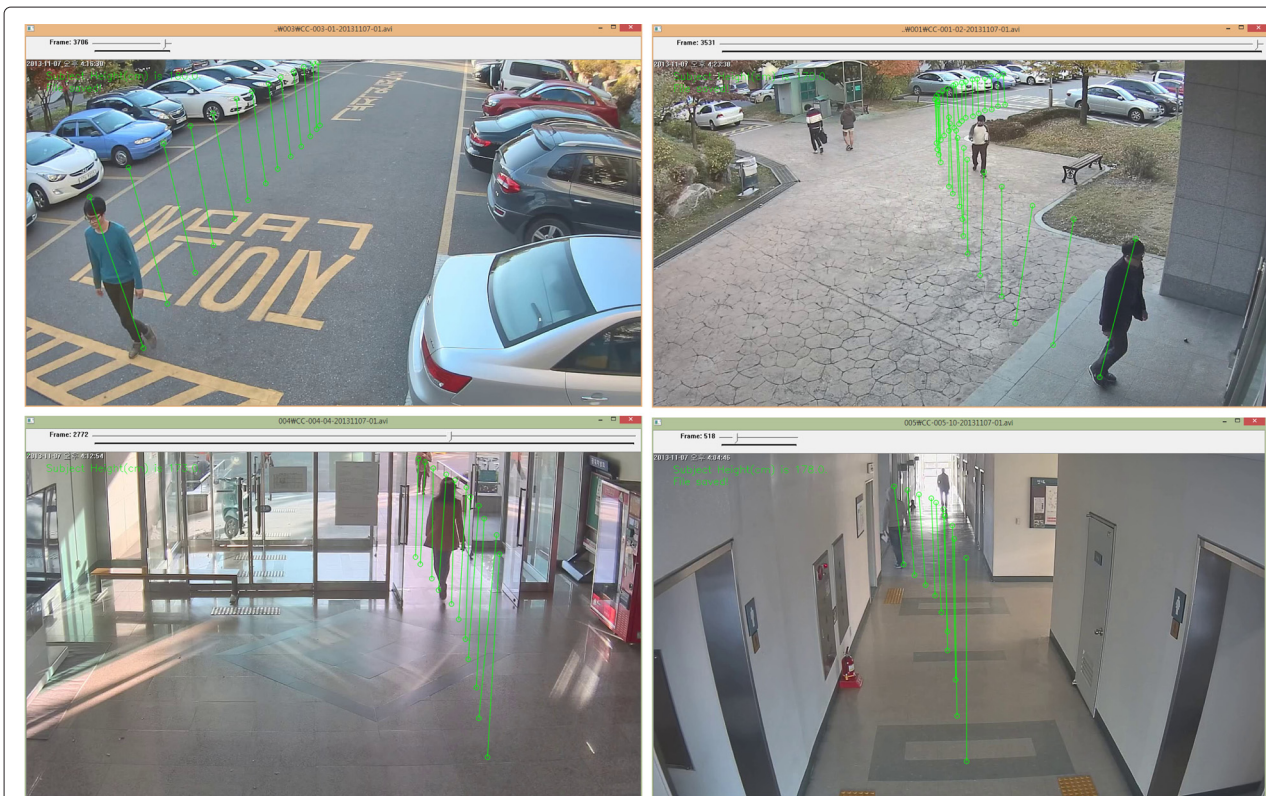
Figure 2 shows the camera calibration process using the proposed method. First, head and foot points were marked and saved, as shown in Fig. 2a. Initial values of the focal length, the tilt angle, and the camera height are set to 720,  $-30$ , and 300 by default or to values obtained by the visual measurement of the camera location. Estimated parameters were found by the nonlinear regression method described in Section 2. More specifically, this experiment adopted the Levenberg-Marquardt algorithm. Optimized parameters  $\hat{f}$ ,  $\hat{\theta}$ , and  $\hat{c}$  for this camera are 547.7,  $-38.6$ , and 270.2, respectively.

Figure 2c plots the relationship between the observed value of  $y_h$  and the estimated value of  $\hat{y}_h$  with respect to the observed value of  $y_f$ . Note that the slope of the initially estimated value of  $\hat{y}_h$  was very close to that of the observed  $y_h$  but that the scale diverged. This is because the visually approximated height and tilt can be relatively accurate, whereas the focal length cannot.

### 3.2 Walking human-based evaluation

In the evaluation based on the walking human, the video dataset consisted of 11 subjects and 9 cameras. The height of each subject was measured with shoes before recording the video. Head and foot points were manually marked in the videos, although there are many algorithms for automatic human detection. Some studies have suggested that the height of the human is more accurate in the phase in which two feet cross each other [9]. The manual marking in this experiment was done according to this clue. Some videos and marking results are shown in Fig. 3. Each camera was calibrated by measuring subject 1, and the error was evaluated by remaining subjects.

In the experiment, the mean and standard deviation of heights were computed from observations from each camera. Figure 4 shows the height estimation error in the evaluation based on the walking human, including the distribution and the limits of agreement (LOA; also known as the Bland-Altman plot). The purpose of the LOA is to investigate the difference between the true height (measured based on the ruler) and the estimated height. The mean absolute error is 1.39 cm, and the standard deviation is 1.91 cm. The 95 % limits of agreement are 3.32 and  $-4.71$  cm, respectively, which are computed by the  $\pm 1.96$  standard deviation. Table 1 provides the detailed results.



**Fig. 3** The walking human dataset. The evaluation dataset collected from a video surveillance site in use. Eleven subjects were requested to walk past nine cameras. The head and foot points were manually marked in the video to evaluate camera calibration and height estimation accuracy

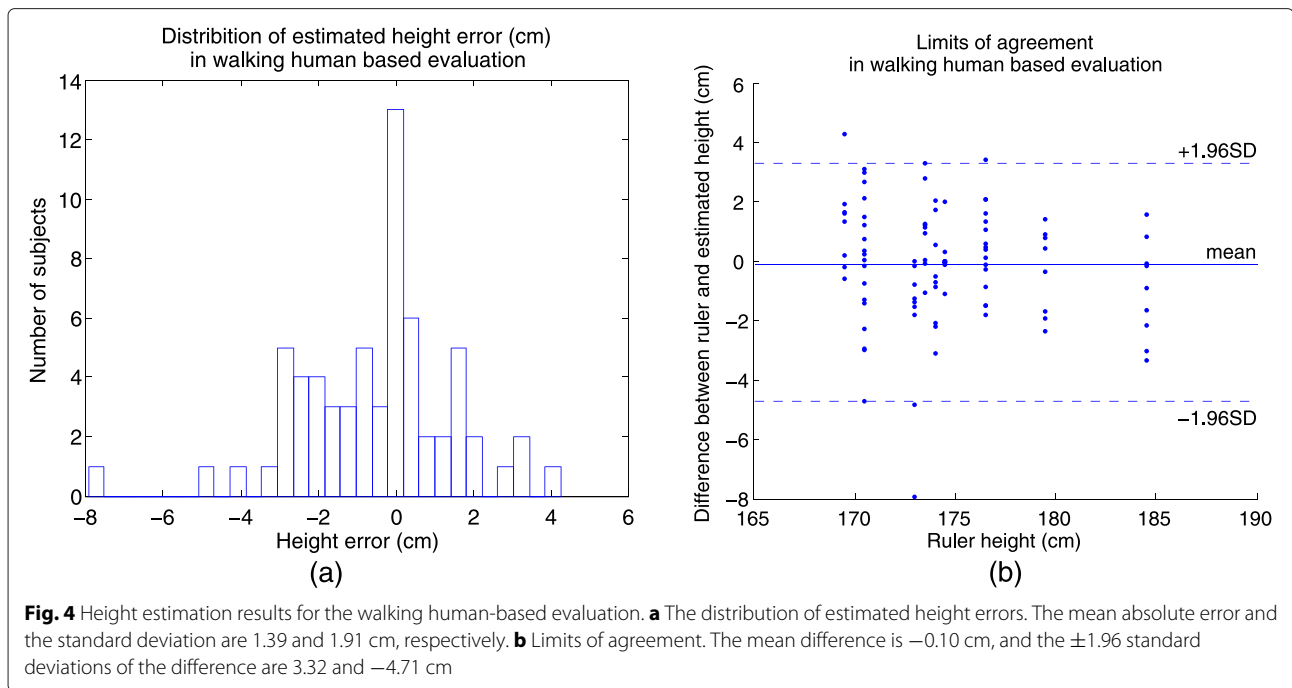


Figure 5 demonstrates no correlation is found between the height estimation error (cm) and x- and y-coordinates, which indicates that the height estimation error does not depend on the position of the human.

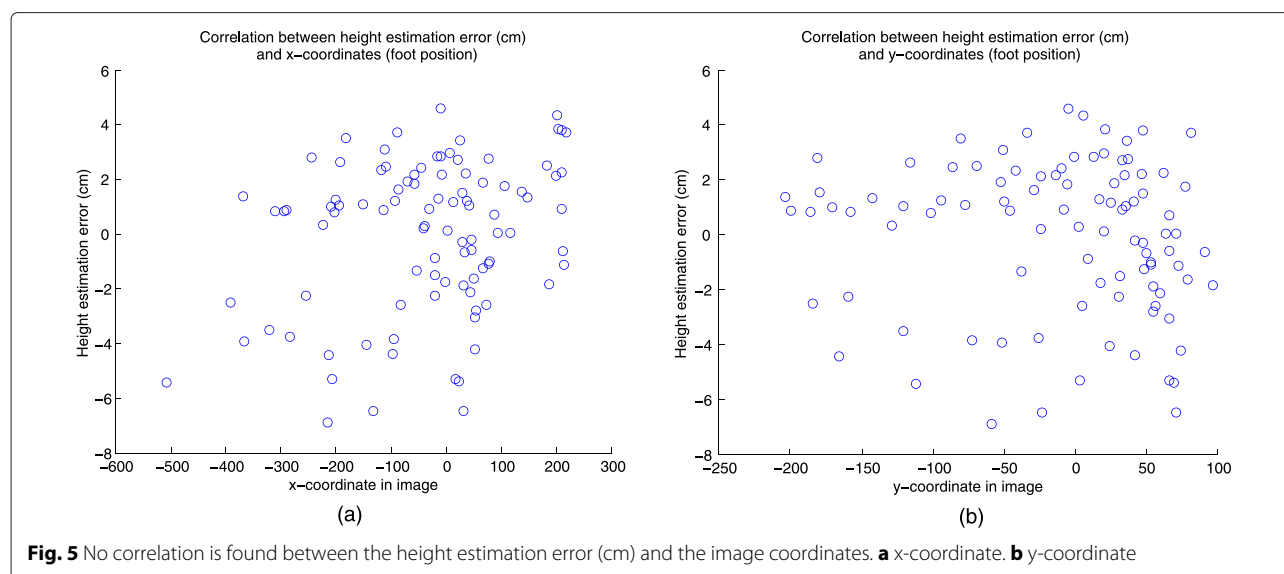
Table 2 compares the results for the proposed method with the existing methods. The empty fields indicate no available result (N/A) in these studies. Mean absolute error, standard deviation, and maximum error were chosen as measures for comparing accuracy. As shown in the table, the proposed method provided more accurate height estimates while requiring only a walking human as the calibration object.

### 3.3 Ruler-based evaluation

In the ruler-based evaluation, a vertical ruler was used instead of a walking human in order to isolate the error caused by the wrong annotation of head and foot points. Figure 6 shows the devices and the data collection procedure. The experiment was conducted in an indoor environment to clearly identify ruler labels. Calibration parameters were estimated by measuring 200 cm and the measurement error was evaluated by measuring 160 to 210 cm with 10-cm increasing intervals. Figure 7 shows the height estimation error in the ruler-based evaluation, including the distribution and the LOA. The mean

**Table 1** Height estimation results. Each camera was calibrated by measuring subject 1, and the error was evaluated by remaining subjects

ID	Ruler	Cam01	Cam02	Cam03	Cam04	Cam05	Cam06	Cam07	Cam08	Cam09
1	174.5	174.4(2.8)	174.4(2.2)	174.4(1.8)	174.4(1.5)	174.4(4.3)	174.4(2.1)	174.4(0.8)	172.4(2.0)	174.4(0.8)
2	176.5	177.5(2.5)	178.5(1.5)	179.9(1.4)	176.5(1.5)	176.9(2.2)	177.0(2.0)	176.4(0.6)	173.5(0.7)	176.6(0.8)
3	169.5	169.6(1.5)	169.3(2.3)	171.4(1.8)	173.7(2.0)	171.1(1.8)	170.7(1.7)	168.9(1.7)	170.2(2.0)	171.1(0.6)
4	184.5	184.4(1.8)	185.3(2.3)	186.0(1.5)	183.6(1.3)	184.3(5.9)	182.8(1.1)	182.3(2.5)	180.5(1.2)	181.1(1.1)
5	170.5	165.8(2.1)	170.7(1.1)	169.7(2.6)	169.0(1.9)	168.2(1.5)	167.5(1.2)	167.5(2.2)	168.5(1.8)	170.3(0.6)
6	179.5	179.0(1.8)	180.9(1.8)	180.3(1.5)	180.3(2.4)	179.1(2.9)	179.9(1.3)	177.1(3.1)	176.6(1.3)	177.8(0.3)
7	170.5	173.1(1.1)	173.5(1.0)	172.6(1.9)	173.5(3.1)	171.7(2.0)	170.8(1.1)	170.5(1.7)	170.6(2.5)	171.9(1.3)
8	173.5	174.6(1.0)	174.4(1.1)	174.7(1.3)	176.8(1.7)	176.2(4.0)	174.7(1.0)	172.4(1.9)	172.5(1.5)	173.4(2.3)
9	176.5	178.5(1.7)	176.2(1.8)	177.8(1.2)	178.1(1.5)	174.8(3.5)	176.9(1.8)	175.6(0.7)	174.1(1.7)	175.0(0.6)
10	174	170.9(3.0)	173.1(1.5)	174.5(2.1)	176.0(2.3)	175.7(2.3)	173.2(2.5)	171.9(2.3)	171.0(2.6)	173.5(1.5)
11	173	171.2(1.6)	171.7(4.4)	172.8(1.8)	172.9(2.8)	168.1(3.8)	171.6(1.3)	172.2(1.1)	170.5(0.7)	165.0(1.8)



absolute error is 0.42 cm, and the standard deviation is 0.54 cm. The 95 % limits of agreement are 1.25 and  $-0.98$  cm which are computed by the  $\pm 1.96$  standard deviation. The mean difference is  $-0.10$  cm, which is close to 0, indicating that the measurement error is not biased.

## 4 Discussion

### 4.1 Lens distortion correction

Lens distortion causes substantial error in edges of the recorded area, particularly in some wide-angle cameras. To solve this problem, a commonly used radial distortion correction method [18] was applied. The image coordinates were converted into distortion-free coordinates before the calibration.

### 4.2 Ground surface

Some cameras are placed lower than the height of adult subjects, such that their main function is to recognize the face. The proposed method can be applied without any modification. The condition for using the proposed method is that the camera pan and roll are both equal to 0. If this condition is satisfied and the human's head/foot points are observable, then calibration parameters can be

estimated in the same way as general cases. In such cases, the camera height  $c$  is lower than that of adult subjects.

Another case may be the ground surface not being in the same level or the floor not being flat. In such case, substantial error of height estimation will be caused. The solution might be to consider the different level as a new floor and perform the calibration separately.

### 4.3 Pose of the walking human

Some subjects habitually walk with a bowed head. Figure 8 demonstrates that subject 11 walked with a forward-leaning pose in cameras 3, 5, and 9. Estimated errors are  $-0.14$ ,  $-4.84$ , and  $-7.93$  cm. The results show that walking with a leaning pose leads to an underestimation of the subject height.

## 5 Conclusions

This paper proposes a simple camera calibration method for estimating the human height in video surveillance. The proposed method requires neither any special calibration object nor a special pattern on the ground, such as parallel or perpendicular lines. Only three parameters are retained in the camera model, making the estimation of parameters

**Table 2** Comparison of the proposed method with the existing methods

	Calibration object	Mean absolute error	Standard deviation	Maximum error
Krahnstoever [10]	Walking human	5.80 %	N/A	N/A
Lee [19]	Cubix box or line	N/A	N/A	5.50 %
Gallagher [20]	Grid pattern	N/A	2.67 cm	3.28 cm
Jeges [21]	Grid pattern	2.03 cm	4.17 cm	N/A
Proposed	Walking human	1.39 cm (0.80 %)	1.91 cm (1.1 %)	7.93 cm (4.5 %)



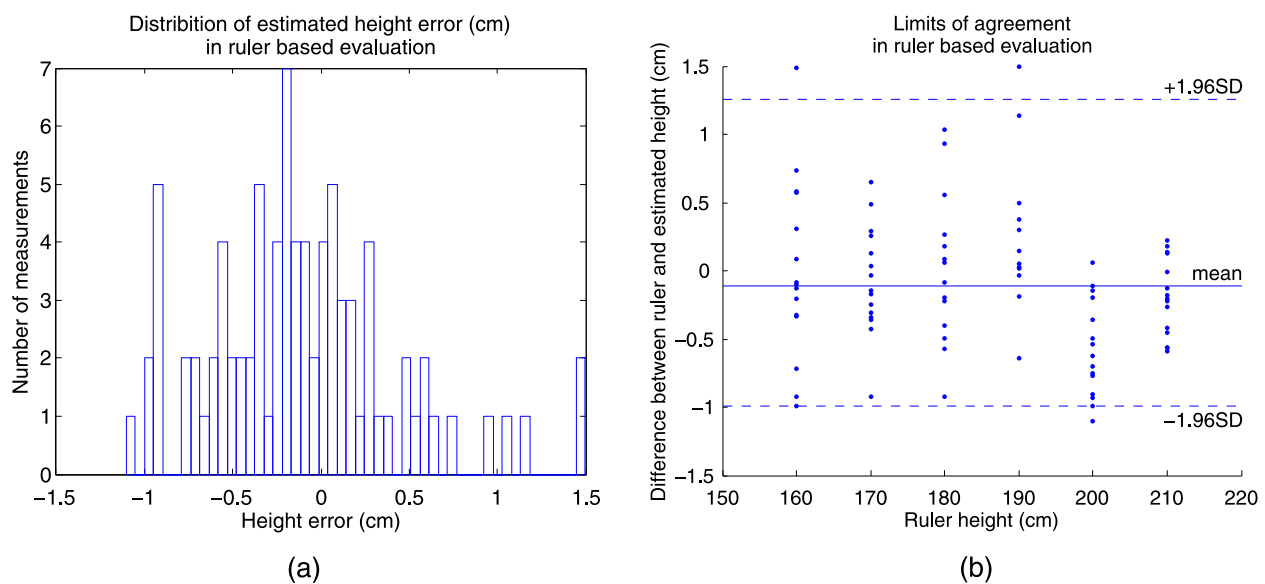
**Fig. 6** The ruler-based evaluation. **a** An aluminum ruler was used as the reference height. **b** A spirit level was attached to the ruler to maintain it at a vertical position. **c** The process of height measurement (the experimenter raised the hand to indicate the ruler is in a vertical position)

more efficient. In addition, the proposed method does not rely on computing vanishing points, which is difficult to estimate in practice.

The experimental results show that the proposed method can predict the human height from observed head

and foot points in the video. The experimental results show that the mean absolute error is only about 1.39 cm from ground truth data in a walking human-based evaluation.

The proposed method can be integrated with automated human detection methods to fully perform



**Fig. 7** Height estimation results for the ruler-based evaluation. **a** The distribution of estimated height errors. The mean absolute error and the standard deviation of error are 0.42 and 0.54 cm, respectively. **b** Limits of agreement. The mean difference is  $-0.10$  cm, and the  $\pm 1.96$  standard deviations are 1.25 and  $-0.98$  cm





**Fig. 8** Walking with a bowed head. Subject 11 walked with a bowed head, as shown in cameras 3, 5, and 9. Estimated errors are  $-0.14$ ,  $-4.84$ , and  $-7.93$  cm. **a** Camera 3. **b** Camera 5. **c** Camera 9

autocalibration, and this provides a useful avenue for future research. In addition, future research should introduce lens distortion parameters to a simplified camera model.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-15-1282-00010002, Suspicious pedestrian tracking using multiple fixed cameras).

The source code and the dataset are available for download at <https://github.com/lishengzhe/ccvs>.

#### Author details

<sup>1</sup>Information and Communication Engineering, Inha University, 100 Inha-ro, Nam-gu, 22212 Incheon, Korea. <sup>2</sup>Visionin Inc., Inha University Incubation Center, 100 Inha-ro, Nam-gu, 402751 Incheon, Korea.

Received: 21 October 2014 Accepted: 29 September 2015

Published online: 12 October 2015

#### References

1. A Dantcheva, C Velardo, A D'Angelo, JL Dugelay, Bag of soft biometrics for person identification. *Multimed. Tools. Appl.* **51**(2), 739–777 (2011)
2. B Hoogeboom, I Alberink, M Goos, Body height measurements in images. *J. Forensic. Sci.* **54**(6), 1365–1375 (2009)
3. N Ramstrand, S Ramstrand, P Brolund, K Norell, P Bergstrom, Relative effects of posture and activity on human height estimation from surveillance footage. *Forensic. Sci. Int.* **212**(1–3), 27–31 (2011)
4. D Reid, M Nixon, S Stevenage, Soft biometrics; human identification using comparative descriptions. *IEEE Trans. Pattern. Anal. Mach. Intell.* **36**(6), 1216–1228 (2014)
5. P Tome, J Fierrez, R Vera-Rodriguez, M Nixon, Soft biometrics and their application in person recognition at a distance. *IEEE Trans. Inf. Forensics. Secur.* **9**(3), 464–475 (2014)
6. SX Yang, PK Larsen, T Alkjaer, B Juul-Kristensen, EB Simonsen, N Lynnerup, Height estimations based on eye measurements throughout a gait cycle. *Forensic. Sci. Int.* **236**(0), 170–174 (2014)
7. B Matusik, T Pajdla, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Simultaneous surveillance camera calibration and foot-head homology estimation from human detections (IEEE, 2010), pp. 1562–1569
8. F Lv, T Zhao, R Nevatia, in *International Conference on Pattern Recognition (ICPR)*. Self-calibration of a camera from video of a walking human, vol. 1 (IEEE, 2002), pp. 562–567
9. F Lv, T Zhao, R Nevatia, Camera calibration from video of a walking human. *IEEE Trans. Pattern. Anal. Mach. Intell.* **28**(9), 1513–1518 (2006)
10. N Krahnstoeber, PR Mendonca, in *International Conference on Computer Vision (ICCV)*. Bayesian autocalibration for surveillance (IEEE, 2005)
11. I Junejo, H Foroosh, in *IEEE International Conference on Video and Signal Based Surveillance (AVSS)*. Robust auto-calibration from pedestrians (IEEE, 2006)
12. J Liu, RT Collins, Y Liu, in *British Machine Vision Conference, Dundee*. Surveillance camera autocalibration based on pedestrian height distributions (BMVA, 2011)
13. J Liu, RT Collins, Y Liu, Robust autocalibration for a surveillance camera network. *IEEE Work. Appl. Comp. Vis. (WACV)*, 433–440 (2013)
14. JP Tardif, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Non-iterative approach for fast and accurate vanishing point detection, (2009), pp. 1250–1257
15. E Tretiyak, O Barinova, P Kohli, V Lempitsky, Geometric image parsing in man-made environments. *Int. J. Comput. Vis.* **97**(3), 305–321 (2012)
16. H Wildenauer, A Hanbury, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Robust camera self-calibration from monocular images of Manhattan worlds (IEEE, 2012), pp. 2831–2838
17. R Hartley, A Zisserman, *Multiple View Geometry in Computer Vision*. (Cambridge University Press, New York, 2004)
18. Z Zhengyou, A flexible new technique for camera calibration. *Pattern. Anal. Mach. Intell. IEEE Trans.* **22**(11), 1330–1334 (2000)
19. K-Z Lee, in *IEEE Conference on Computer and Robot Vision (CRV)*. A simple calibration approach to single view height estimation (IEEE, 2012), pp. 161–166
20. AC Gallagher, AC Blose, T Chen, in *International Conference on Computer Vision (ICCV)*. Jointly estimating demographics and height with a calibrated camera (IEEE, 2009), pp. 1187–1194
21. E Jeges, I Kispal, Z Hornak, in *Human System Interactions*. Measuring human height using calibrated cameras (IEEE, 2008), pp. 755–760