A Calibration Algorithm for Multi-camera Visual Surveillance Systems Based on Single-View Metrology

J.C. Neves[1(✉)], J.C. Moreno[1], S. Barra[2], and H. Proença[1]

[1] IT - Instituto de Telecomunicações, Department of Computer Science, University of Beira Interior, 6201-001 Covilhã, Portugal
jcneves@penhas.di.ubi.pt

[2] Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 40, 09124 Cagliari, Italy

Abstract. The growing concerns about persons security and the increasing popularity of pan-tilt-zoom (PTZ) cameras, have been raising the interest on automated master-slave surveillance systems. Such systems are typically composed by (1) a fixed wide-angle camera that covers a large area, detects and tracks moving objects in the scene; and (2) a PTZ camera, that provides a close-up view of an object of interest. Previously published approaches attempted to establish 2D correspondences between the video streams of both cameras, which is a ill-posed formulation due to the absence of depth information. On the other side, 3D-based approaches are more accurate but require more than one fixed camera to estimate depth information. In this paper, we describe a novel method for easy and precise calibration of a master-slave surveillance system, composed by a single fixed wide-angle camera. Our method exploits single view metrology to infer 3D data of the tracked humans and to self-perform the transformation between camera views. Experimental results in both simulated and realistic scenes point for the effectiveness of the proposed model in comparison with the state-of-the-art.

1 Introduction

Video surveillance has been mainly used for traffic monitoring and security systems in both public and private places. Many current surveillance systems need a human operator to record the video and store the data for later analysis. Also, the number of cameras and the area under surveillance are limited by the personal availability. In order to overcome the limitations of traditional video surveillance systems, the development of new methods is a current effort in the computer vision and artificial intelligence community.

A satisfactory surveillance system must be capable of monitoring larger areas as well as presenting high resolution images of an event of interest [1,2]. Systems based on wide-angle cameras are able to monitor an area that covers 180° but have limited scale of range due to the relatively reduced resolution. PTZ cameras are currently used in several surveillance environments due to their capabilities of monitoring large areas of interest providing details of moving objects by

R. Paredes et al. (Eds.): IbPRIA 2015, LNCS 9117, pp. 552–559, 2015.
DOI: 10.1007/978-3-319-19390-8_62

(a) Wide-angle camera

(b) PTZ camera

(c) The surveillance system monitors a large area and zooms-up over a object of interest, i.e. a subject head.

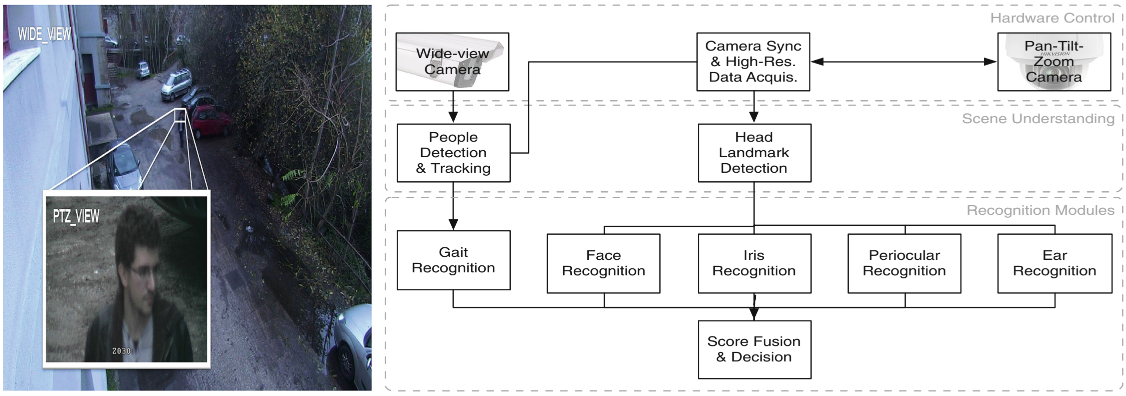


Fig. 1. Dual surveillance system. (a) Wide-angle camera. (b) PTZ camera. (c) The wide-angle camera gives a panoramic view of the scene, while the PTZ camera takes a snapshot of the full body of the tracked subject,in order to provide both physical and behavioural features for an a posteriori biometric recognition.

decreasing or increasing the focal length. A dual camera system can be defined by combining a wide-angle camera and a PTZ camera. The wide-angle camera provides a large resolution coverage of the scene as well as a global tracking of a human activity [3,4], while the PTZ camera is responsible of acquiring close-up views at high resolution [5,6]. The cooperation of both relies on a calibration method where an accurate interaction between the cameras is needed. Further challenges can be found in controlling and scheduling the desired pan and tilt angles of the PTZ camera, in order to better satisfy the competing request. The precision of the pan and tilt estimation mainly depends on the accuracy of the spatial calibrations, determined by mapping onto a common space the pixels of both the cameras [7].

Several methods have been proposed in the past, in order to perform a dual camera calibration system: Hampapur et al. [8] controlled a PTZ camera by triangulating 3D positions computed from two calibrated wide-angle cameras. Chen et al. [9] connected multiple 3D position camera models by fitting the best polynomial, to better describe the camera projection model and the spatial relations. Senior et al. [10] mapped wide-angle camera coordinates to PTZ camera coordinates by estimating the homography between cameras. Zhou et al. [11] triangulated the target position between cameras at pixel level (2D geometry) using different pixel locations of the image given by the wide-angle camera and then mapping the pixels by linear interpolation over the pan-tilt angles. You et al. [12] estimated the relationship between wide-angle and PTZ cameras using a mosaic of images created by snapshots derived from the slave camera, in order to define a linear interpolation for describing the calibration of the dual surveillance system. Liu et al. [13] presented a calibration approach that interpolates two camera models by matching feature points. Even if many of these methods are accurate enough, they heavily depend on the precision of the spatial mappings based either on 3D or 2D geometry. In the former, target position is based on stereo vision, while in the latter, because of the lack of informations about depth, the more the distance between camera and target grows, the more the system becomes inaccurate. Many of these methods are based on scene features which, by defining a fundamental matrix, allow to move from a wide-angle to PTZ camera [14].

In this work, we propose an automatically geometry-based calibration system (no particular system setup or human intervention is needed) that estimates the 3D world coordinates of a given pixel in the panoramic view by allowing the computation of both the pan and tilt angles of the PTZ camera. The proposed method is based on solving a system of equations that directly relates image pixel coordinates with its corresponding position in the space, via a projection matrix. Then, the intrinsic and extrinsic parameters of each camera need to be determined. For computing the 3D world-coordinates of a chosen pixel on a human head, we first need to fix a Euclidean structure on the image scene, such that the vanishing lines, as well as the related vanishing points can be identified from a single view image. Once the height of the detected human is computed, a 3×3 system of equation is solved to determine the 3D position of the human head. Figure 1(c) shows the final iteration of the calibration module which is followed by the recognition module.

The remainder of the paper is organized as follows. Section 2 introduces the calibration method between wide-angle and PTZ cameras. In Sect. 3, experiments are implemented and the experimental results are shown. Finally, Sect. 4 concludes the paper.

2 Dual Surveillance System Description

The proposed dual surveillance system is based on a wide-angle camera, responsible of pointing to a target on a large panoramic scene, and a PTZ camera focusing at the same position and taking a close-up snapshot of the head of the tracked person. The image coordinate system is the usual xy affine image frame. A point $\mathbf{X}$ with homogeneous coordinates in space is projected to the image point $\mathbf{x}$ using the following 3×4 projection matrix $\mathbf{P}_1$ [15]:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \underbrace{\mathbf{K}_1[\mathbf{R}_1\ \mathbf{T}_1]}_{:=\mathbf{P}_1} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{1}$$

where $\mathbf{T}_1$ is a 3×1 vector that represents the camera location, $\mathbf{R}_1$ is a 3×3 rotation matrix that represents the orientation of the camera, with respect to an absolute coordinate frame, and $\mathbf{K}_1$ is a 3×3 upper triangular matrix encompassing the parameters of the camera.

Our aim is to compute the height of an object relative to a reference. To this aim we use minimal calibration information as well as a perspective approach for mensuration. The minimal calibration condition is based on the vanishing line of a ground plane, the vertical point and a reference height [16,17]. The vanishing line (or line at infinite) is the line through two or more vanishing points of the plane. A vanishing point is the intersection point of a set of parallel lines in the image plane [18]. If $\mathbf{v}$ is the vanishing point for the vertical direction, $\mathbf{l}$ is the vanishing line of the ground plane, $\mathbf{t}_r$ and $\mathbf{b}_r$ are the top and base point of a reference with height equal to Z_r and $\mathbf{t}_x$ and $\mathbf{b}_x$ are the top and the base point of the tracked human, we compute the height using the following equations:

$$Z_x = -\frac{\|\mathbf{b}_x \times \mathbf{t}_x\|}{\alpha(\mathbf{l}.\mathbf{b}_x)\|\mathbf{v} \times \mathbf{t}_x\|}, \tag{2}$$

with

$$\alpha = -\frac{\|\mathbf{b}_r \times \mathbf{t}_r\|}{Z_r(\mathbf{l}.\mathbf{b}_r)\|\mathbf{v} \times \mathbf{t}_r\|}. \tag{3}$$

It can be noticed that the success of the height computation of the tracked human depends on the accurate estimation of the vertical vanishing point $\mathbf{v}$ and the vanishing line $\mathbf{l}$ of the reference plane.

The final 3D position $\mathbf{X} = (X, Y, Z_x)$ of the main object in the scene is then computed by solving the wide-angle camera system (1) at the pixel point $\mathbf{t}_x$. By including the computed height, Z_x the equation systems (1) is reduced to the 3×3 equation systems

$$\begin{pmatrix} \mathbf{t}_x \\ w \end{pmatrix} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad Z_x\mathbf{p}_3 + \mathbf{p}_4] \begin{pmatrix} X \\ Y \\ W \end{pmatrix}, \tag{4}$$

with w, W representing the homogenous coordinate of the extended points $\mathbf{t}_x$, (X, Y) in the projective space, and $\mathbf{p}_i$ (i=1,2,3,4) are the column vectors of the projection matrix $\mathbf{P}_1$. From (4) an unique solution is computed. Therefore the final 3D position of the main object in the scene is $\mathbf{X} = (X/W, Y/W, Z_x)$.

Without loss of generality, we start the PTZ camera with pan angle $\theta = 0$ and some tilt angle ϕ. By panning with respect to the $Y-$axis, the point $\mathbf{X}$ changes with respect to the PTZ camera coordinate system by first translating the opposite of the center of projection C and the rotation matrix R:

$$C = \begin{pmatrix} 0 \\ \rho \sin\theta \\ \rho \cos\theta \end{pmatrix}, \quad R = \begin{pmatrix} \cos\phi & 0 & -\sin\phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin\phi & 0 & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

where ρ is the focal length of the PTZ camera. According to (5), the point $\mathbf{X}$, with respect to the PTZ camera, is given by

$$\begin{pmatrix} X_{PTZcam} \\ Y_{PTZcam} \\ Z_{PTZcam} \end{pmatrix} = \begin{pmatrix} X\cos\phi + Z\sin\phi \\ Y + \rho\cos\theta \\ -X\sin\phi + Z\cos\phi - \rho\sin\theta \end{pmatrix}. \tag{6}$$

The final pan and tilt angles for the PTZ camera are

$$\theta = \arccos\left(\frac{Z_{PTZcam}}{\sqrt{X_{PTZcam}^2 + Y_{PTZcam}^2 + Z_{PTZcam}^2}}\right), \quad \phi = \arctan\left(\frac{Y_{PTZcam}}{X_{PTZcam}}\right). \quad (7)$$

3 Experimental Results

In order to validate our dual calibration system, we designed the following experiment: 1) A pixel (x_1, y_1) from the panoramic image using the wide-angle camera is selected; then we compute its corresponding position in the space $\mathbf{X}_1 = (X_1, Y_1, Z_1)$ using the Eqs. (2) and (4). 2) With respect to the PTZ camera we select the same pixel used in the wide-angle view (x_2, y_2) in order to compute the ray $\mathbf{r}$ of projection of (x_2, y_2). $\mathbf{r}$ is determined by the system of equations

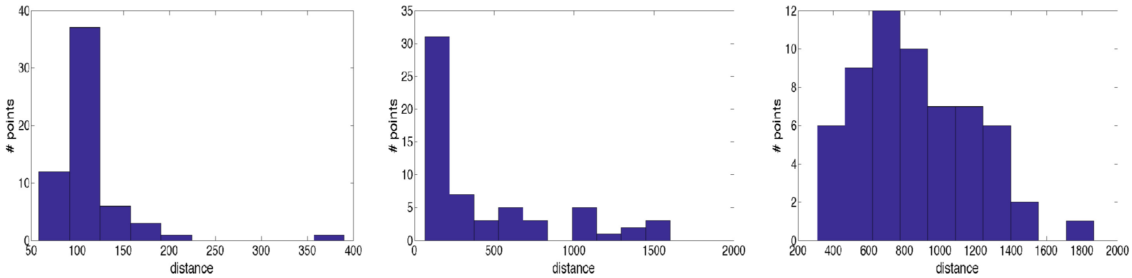
$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \mathbf{P}_2 \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{pmatrix}, \quad (8)$$

with $\mathbf{P}_2$ the projection matrix associated to the PTZ camera. Then, we calculate the distance $d(\mathbf{r}, \mathbf{X}_1)$ between the line $\mathbf{r}$ and the point $\mathbf{X}_1$.

For the initial testing, we validate the interaction between the cameras in our dual surveillance system. For this purpose, we manually selected 60 points on the floor, as shown in Fig. 2(a) (*green* points). The Euclidean structure on the fixed scene is determined by $Z = 0$ millimetres (mm) in the space. The performance of the distance $d(\mathbf{r}, \mathbf{X}_1)$ over all the points in the proposed experiment are measured in millimeters with results visible in terms of the histogram displayed in Fig. 2(b). From the histogram graph, we can appreciate that the biggest and the lowest distance measure 390.1 mm and 58.3 mm, respectively. Comparisons with previous methods based on homography mappings also validate the improvements of our results. Figure 2(c) and (d) display the results according to the works due to Hartley et al. [14] and Liu et a.l [13] respectively. By implementing our model we have 37 points such that the distance measure is between 91.5 and 125.3 mm, while in the Hartley et al. [14] approach there are 31 points with distance measure between 66.2 and 219.7 mm; moreover the biggest distance measure is equal to 1520.4 mm; Liu et al. [13] method presents 12 points with distance measure between 662.3 and 775.7 mm and highest and lowest distance measure equal to 1865.4 mm and 308.5 mm, respectively.

For the next experiment, we validate the complete dual surveillance system by applying the single view metrology techniques in order to compute height of people in the scene and use the information to focus the PTZ camera in the selected pixel over the human head. In this case we have $Z \neq 0$, then we need to compute the heights by implementing Eqs. (2) and (3) using the scene display in Fig. 3(a). The reference height Z_r is taken from the door with pixels $\mathbf{b}_r$ and $\mathbf{t}_r$

(a)



(b) Our approach (c) Hartley et al. approach [14] (d) Liu et al. approach [13]

Fig. 2. Validation of the proposed dual surveillance system calibration. (a) Panoramic view given by the wide angle camera. (b), (c) and (d) show the histograms validating our approach and the models [14] and [13] . In particular, they display the number of points (y axis) lying at several distances (in mm, x axis), according to the manually selected points (*green* dots in (a)) (Color figure online).

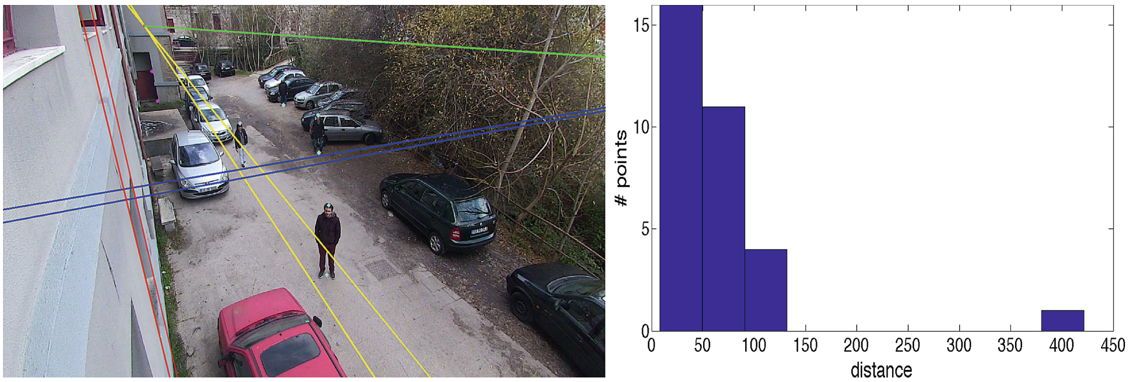


Fig. 3. Validation of the complete dual surveillance system calibration. The image on the left shows a panoramic view together with the lines defining the vanishing line (*green* line) and vanishing point (*red* lines intersection) as well as the reference height given by the distance between the two points in *magenta* color. On the right, the histogram shows the performance of the proposed method, when the distance $d(\mathbf{r}, \mathbf{X}_1)$ is calculated with respect to the points selected on the heads of the people in the scene(Color figure online).

indicated with *magenta* color. The height of the people in the scene is computed using the head and foots marked with *cyan* color. For the final validation we made the people fill seven different positions. Computation of the distance $d(\mathbf{r}, \mathbf{X}_1)$ is implemented over the people head and the results are displayed in the histogram (2), where it is appreciable that the biggest distance measure is equal to $401.2\,mm$ and the lowest distance measure is equal to $8.2\,mm$. In the 32 panoramic view points, we have 16 points with distance measure between 8.2 and 49.5 mm.

4 Conclusion

In this paper, we proposed a novel framework for a video surveillance system that automatically obtains the pan and tilt angles required for a PTZ camera to take a snapshot of a human head. Our method starts by detecting the 2D position of a human head in an image, from where the corresponding 3D position in the scene is estimated, using the wide-angle camera view and the matrix projection that relates the image coordinate system and the space. The proposed algorithm is validated empirically in comparison to two previously published dual calibrated systems [13,14], enabling us to conclude about consistent improvements in performance. As further work, results might even be improved by using adaptive prediction models for tracking human trajectories in the scene. Such a models should learn specific image trajectories with respect to positions in the scene, improving the calibration dual system and therefore the accurate interaction between surveillance cameras.

Acknowledgements. This work is supported by '*FCT - Fundação para a Ciência e Tecnologia*' (Portugal) through the research grant 'SFRH/BD/92520/2013', and the funding from 'FEDER - QREN - Type 4.1 - *Formação Avançada*', co-founded by the European Social Fund and by national funds through Portuguese 'MEC - *Ministério da Educação e Ciência*'. It is also supported by the *IT - Instituto de Telecomunicações* through 'PEst-OE/EEI/LA0008/2013'.

References

1. Valera, M., Valestin, S.A.: Intelligence distributed surveillance systems: a review. IEE Proc. Vis. Image Sig. Proc. **152**(2), 192–204 (2005)
2. Xu, Y., Song, D.: Systems and algorithms for autonomous and scalable crowd surveillance using robotic ptz cameras assisted by a wide-angle camera. Auton. Robots **29**(1), 53–66 (2010)
3. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional wide-angle, and fish-eye lenses. IEEE Trans. Pattern Anal. Mach. Intell. **28**(8), 1335–1340 (2006)
4. Wang, J., Shi, F., Zhang, J., Liu, Y.: A new calibration model of camera lens distortion. Pattern Recogn. **41**(2), 607–615 (2008)
5. Sinha, S.N., Pollefeys, M.: Pan-tilt-zoom camera calibration and high-resolution mosaic generation. Comput. Vis. Image Underst. **103**(3), 170–183 (2006)

6. Wu, Z., Radke, R.J.: Keeping a pan-tilt-zoom camera calibrated. Image Vis. Comput. **35**(8), 1994–2007 (2013)
7. Del Bimbo, A., Dini, F., Lisanti, G., Percini, F.: Exploiting distinctive visual landmark maps in pan-tilt-zoom camera networks. Comput. Vis. Image Underst. **114**(6), 611–623 (2010)
8. Hampapur, A., Pankanti, S., Senior, A., Tian, Y.-L., Brown, L., Bolle, R.: Face cataloger: multi-scale imaging for relating identity to location. In: Proceedings of the IEEE Conference on Advance Video and Signal Based Surveillance, pp. 13–20, Miami, USA (2003)
9. Chen, C.-H., Yao, Y., Page, D., Abidi, B., Koschan, A., Abidi, M.: Heterogeneous fusion of omnidirectional and ptz cameras for multiple object tracking. IEEE Trans. Circuits Syst. Video Technol. **18**(8), 1052–1063 (2008)
10. Senior, A.W., Hampapur, A., Lu, M.: Acquiring multiscale images by pan-titl-zoom control and automatic multicamera calibration. In: Proceedings of the 7th IEEE Workshop on Application ofComputer Vision, vol.1, pp. 433–438, Breckenridge, USA (2005)
11. Zhou, X., Collins, R., Kanade, T., Metes, P.: A master-slave system to acquire biometric imagery of human at distance. In: Proceedings of the 1st ACM International Workshop on Video Surveillance, pp. 113–120, Berkeley, USA (2003)
12. You, L., Li, S., Jia, W.: Automatic weak calibration of master-slave surveillance system based on mosaic images. In: Proceedings of the 20th International Conference on Pattern Recognition, pp. 1824–1827, Istanbul, Turkey (2010)
13. Liu, Y., Lai, S., Zuo, S., Shi, H., Zhang, M.: A master-slave surveillance system to acquire panoramic and multiscale videos. Sci. World J. **2014**, 11 (2014). Article ID 491549
14. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York, USA (2004)
15. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
16. Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. Int. J. Comput. Vision **40**(2), 123–148 (2000)
17. Wang, G., Hu, Z., Wu, F., Tsui, H.-T.: Single view metrology from scene constraints. Image Vis. Comput. **23**(9), 831–840 (2005)
18. Schaffalitzky, F., Zisserman, A.: Planar grouping for automatic detection of vanishing lines and points. Image Vis. Comput. **18**(9), 647–658 (2000)