Acquiring High-resolution Face Images in Outdoor Environments: A Master-slave Calibration Algorithm

João C. Neves[1], Juan C. Moreno[1], Silvio Barra[2] and Hugo Proença[1]
[1]IT - Instituto de Telecomunicações, University of Beira Interior
[2]Department of Mathematics and Computer Science, University of Cagliari
jcneves@ubi.pt, jcmb@ubi.pt, silvio.barra@unica.it, hugomcp@di.ubi.pt

Abstract

Facial recognition at-a-distance in surveillance scenarios remains an open problem, particularly due to the small number of pixels representing the facial region. The use of pan-tilt-zoom (PTZ) cameras has been advocated to solve this problem, however, the existing approaches either rely on rough approximations or additional constraints to estimate the mapping between image coordinates and pan-tilt parameters. In this paper, we aim at extending PTZ-assisted facial recognition to surveillance scenarios by proposing a master-slave calibration algorithm capable of accurately estimating pan-tilt parameters without depending on additional constraints. Our approach exploits geometric cues to automatically estimate subjects height and thus determine their 3D position. Experimental results show that the presented algorithm is able to acquire high-resolution face images at a distance ranging from 5 to 40 meters with high success rate. Additionally, we certify the applicability of the aforementioned algorithm to biometric recognition through a face recognition test, comprising 20 probe subjects and 13,020 gallery subjects.

1. Introduction

The co-existence of humans and video surveillance cameras in outdoor environments is becoming commonplace in modern societies. This new paradigm has raised the interest in automated surveillance systems capable of acquiring biometric data for human identification purposes. Considering that these systems are aimed at covering large areas, the collected biometric data is poorly represented by a small amount of pixels, which greatly degrades recognition performance. To address this issue, several authors have defended the use of PTZ cameras [10, 18, 5, 15], which are capable of acquiring high resolution imagery on arbitrary scene locations.

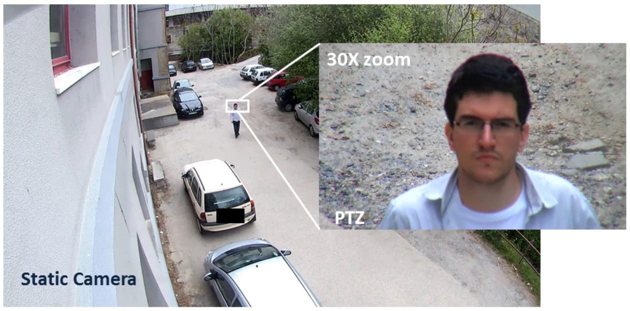


Figure 1. Face image automatically captured using the proposed master-slave calibration algorithm. Our method is accurate enough to use the maximum zoom magnification of the PTZ camera, allowing to acquire high-resolution face images (interpupillary distance greater than 60 pixels) up to 40 m.

In PTZ-based systems, a master-slave configuration is usually adopted, i.e., a static camera is responsible both for detecting and tracking subjects in the scene so that it can instruct the PTZ camera to point to subject faces. While several advantages can be outlined, inter-camera calibration is the major bottleneck of this configuration, since determining the mapping function from static image coordinates to pan-tilt parameters requires depth information.

To address this problem, most master-slave systems use 2D-based approximations, but, in turn, they are compelled to rely on different assumptions (e.g., similar points-of-view [18], intermediate zoom states [2, 14]) to alleviate pan-tilt inaccuracies. The use of multiple optical devices has been pointed as a solution to infer depth information through triangulation. Choi *et* al. [5] and Park *et* al. [15] were the first to exploit this alternative without using stereographic reconstruction, which is not feasible in real-time applications. Instead, they disposed the cameras in a coaxial configuration to ease triangulation. In addition, the authors ascertained the feasibility of facial recognition at-a-distance using the proposed calibration method. However, the highly stringent disposal of the cameras restrains its use in outdoor environments as well as its operational range (up to 15m).

Table 1. Comparative analysis between the existing master-slave systems and the proposed method.

Master-slave system	Pan-Tilt Estimation		Camera disposal		Intermediate Zoom States		Multiple Devices		Calibration Marks	
	Approx.	Exact	Specific	Arbitrary	Yes	No	Yes	No	Yes	No
Zhou *et* al. [22]	×			✓		✓		✓	×	
Liao and Cho [11]	×			✓		✓		✓	×	
Marchesotti *et* al. [14]	×			✓	×			✓	×	
Bodor *et* al. [3]	×		×			✓		✓	×	
Liu *et* al. [12]	×		×			✓		✓		✓
Wheeler *et* al. [18]	×		×			✓		✓	×	
Del Bimbo *et* al. [2]	×			✓	×			✓		✓
Hampapur *et* al. [9]		✓		✓		✓	×		×	
Choi *et* al. [5]		✓	×			✓	×			✓
Park *et* al. [15]		✓	×			✓	×			✓
Senior *et* al. [17]		✓		✓		✓		✓	×	
Fiore *et* al. [8]		✓		✓		✓		✓	×	
Our approach		✓		✓		✓		✓		✓

A comparative analysis between the most relevant master-slave systems is presented in Table 1.

In this paper, we aim at improving the existing master-slave systems, in particular the work of Choi *et* al. [5] and Park *et* al. [15], by extending PTZ-assisted facial recognition to surveillance scenarios. We introduce a calibration algorithm capable of accurately estimating pan-tilt parameters without resorting to intermediate zoom states, multiple optical devices or highly stringent configurations. Our approach exploits geometric cues, i.e., the vanishing points available in the scene, to automatically estimate subjects height and thus determine their 3D position. Furthermore, we have built on the work of Lv *et* al. [13] to ensure robustness against human shape variability during walking. Considering that the proposed calibration algorithm is intended to be integrated in an automated surveillance system, we have also assessed the performance of the proposed algorithm using two challenging scenarios: 1) automatic estimation of head and feet locations using a tracking algorithm; and 2) incorrect vanishing point estimation.

Contributions: 1) A master-slave calibration algorithm is introduced for capturing high-resolution face images (interpupillary distance greater than 60 pixels) at-a-distance (up to 40m). Our approach is capable of accurately estimating the 3D position of the subjects head using geometric cues, without resorting to calibration patterns or specific configurations for the cameras. 2) An experimental evaluation shows that inferring height and 3D position is feasible in surveillance scenarios (89% success rate in acquiring facial images at maximum zoom), even if the algorithm is provided with data automatically obtained from a tracking algorithm (87% success rate). 3) The proposed method has been integrated in an automated surveillance system to acquire high-resolution facial images. A face recognition test with 20 probe subjects and 13,020 gallery subjects evidences the feasibility of the proposed approach to perform biometric recognition at-a-distance (76.5% rank-1 identification accuracy).

Organization: The remainder of this paper is organized as follows. Section 2 summarizes the most relevant PTZ-based approaches according to the proposed taxonomy. Section 3 describes the proposed method. The experimental evaluation of the proposed algorithm is presented and discussed in section 4, whereas its applicability to face recognition is experimentally evidenced in section 5. Finally, section 6 outlines the major conclusions of this work.

2. Related Work

Existing PTZ-based systems can be broadly divided into two principal groups: master-slave configuration and single-camera configuration. In the former, scene monitoring is performed independently in the master camera, while the PTZ camera is treated as a foveal sensor. In contrast, single-camera strategies assign both tasks to the active camera.

Single-camera configuration: In this architecture, the PTZ camera has to perform high range zoom transitions to acquire high resolution biometric data [4, 7]. Taking into consideration that zoom transitions are the most time-consuming actions in these devices, such strategy is highly prone to miss targets. To alleviate this problem some approaches track the target while zooming [1], but, in turn, this strategy requires a greater amount of time observing each subject.

Master-slave configuration: In spite of the multiple advantages of this strategy, its feasibility is greatly dependent on the accurate inter-camera calibration. The lack of depth information poses the mapping between both devices as an ill-posed problem. To that end, several approximations have been proposed to minimize the inter-camera mapping error.

Zhou *et* al. [22] relied on manually constructed look-up tables and linear interpolation to map pixel locations of the static camera to pan-tilt values. In a similar fashion, Liao and Cho [11] approximated the target position as its projection in the reference plane, to which a pixel to pan-tilt mapping had been previously constructed. To alleviate the burden of manual mapping, Liu *et* al. [12] presented an automatic calibration approach by estimating an approximate relation between camera images using feature point matching. Marchesotti *et* al. [14] relied on a 2D-based inexact mapping to provide the active camera with a rough estimate of the target position. Accurate localization was attained by iteratively following the subject while zooming-in. Del Bimbo *et* al. [2] relied on feature point matching to automatically estimate a homography (H), relating the master and slave views with respect to the reference plane. H is used to perform an online mapping between the feet locations in the master to the slave camera and also determine the reference plane vanishing line from the one manually marked on the static view. Despite being capable of determining head location, this strategy has to set the active camera in an intermediate zoom level to cope with the uncertainties of vanishing line location. Xu and Song [20] relied on multiple consecutive frames to approximate target depth. However, this strategy is time-consuming, and consequently, increases the delay between issuing the order and directing the PTZ. You *et* al. [21] estimated the relationship between the static and the active camera using a homography for each image of the mosaic derived from the slave camera.

In contrast to the previous approaches, the use of multiple static cameras has also been introduced to solve the lack of depth information in master-slave systems. However, these systems either rely on stereographic reconstruction [9], which is computationally expensive, or dispose the cameras in a specific configuration to ease object triangulation [5, 15], which is not practical for real-world scenarios.

3. Proposed Method

We start by introducing the notation used in our description:

- (X, Y, Z) : the 3D world coordinates;
- (X_s, Y_s, Z_s) : the 3D coordinates in the static camera world referential;
- (X_p, Y_p, Z_p) : the 3D coordinates in the PTZ camera world referential;
- (x_s, y_s) : the 2D coordinates in the static camera image referential;
- (x_t, y_t) : the 2D coordinates of a head in the static camera image referential;
- (x_p, y_p) : the 2D coordinates in the PTZ camera image referential;
- $(\theta_p, \theta_t, \theta_z)$: the pan, tilt and zoom parameters of the PTZ camera.

In the pin-hole camera model, the projective transformation of 3D scene points onto the 2D image plane is governed by:

$$\lambda \begin{pmatrix} x_t \\ y_t \\ 1 \end{pmatrix} = \underbrace{\mathbf{K}\,[\,\mathbf{R}\,|\,\mathbf{T}\,]}_{:=\,\mathbf{P}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{1}$$

where λ is a scalar factor, $\mathbf{K}$ and $[\,\mathbf{R}\,|\,\mathbf{T}\,]$ represent the intrinsic and extrinsic camera matrices, which define the projection matrix $\mathbf{P}$.

Let $\mathbf{p}_t = (x_t, y_t)$. Solving equation (1) for (X, Y, Z) yields an under-determined system, i.e., infinite possible 3D locations for the face. As such, we propose to solve equation 1 by determining one of the 3D components previously.

By assuming a world coordinate system (WCS) where the XY plane corresponds to the reference ground plane of the scene, the Z component of a subject's head corresponds to its height (h). The use of height information reduces the equation (1) to:

$$\lambda \begin{pmatrix} \mathbf{p}_t \\ 1 \end{pmatrix} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad h\mathbf{p}_3 + \mathbf{p}_4] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}, \tag{2}$$

where $\mathbf{p}_i$ is the set of column vectors of the projection matrix $\mathbf{P}$. As such, our algorithm works on the static camera to extract (x_t, y_t) and infer the subject position in the WCS using its height.

3.1. Height Estimation

To perform height estimation, we rely on the insight that surveillance scenarios are typically urban environments with useful geometric information that can be exploited, such as vanishing points and vanishing lines.

As in [6], three vanishing points $(\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z)$ are used for the X, Y and Z axis, in order to infer the height of a subject, which is vertical to a planar surface. $\mathbf{v}_x$ and $\mathbf{v}_y$ are determined from parallel lines contained in the reference plane, so that the line $\mathbf{l}$ defined by these points represents the plane vanishing line. The point $\mathbf{v}_z$ corresponds to the intersection of two lines perpendicular to the reference plane.

Given $\mathbf{l}$, $\mathbf{v}_z$, the head $(\mathbf{p}_t)$ and feet $(\mathbf{p}_b)$ points in an image, the height of a person can be obtained by:

$$h = -\frac{\|\mathbf{p}_b \times \mathbf{p}_t\|}{\alpha(\mathbf{l}.\mathbf{p}_b)\|\mathbf{v}_z \times \mathbf{p}_t\|}, \tag{3}$$

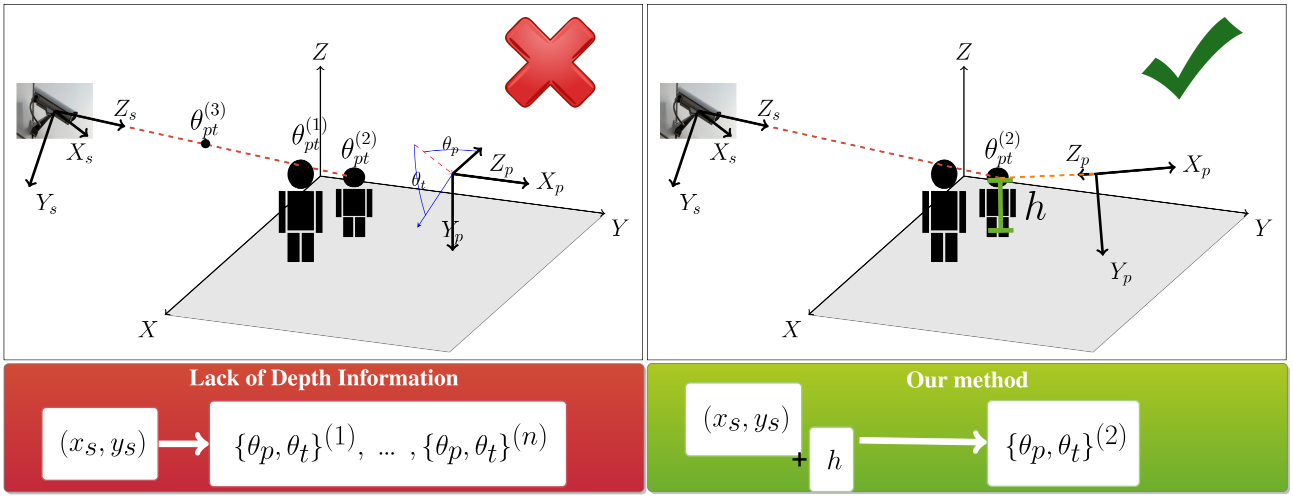


Figure 2. Illustration of the principal bottleneck of master-slave systems and the proposed strategy to address this problem. The same image pixel (x_s, y_s) corresponds to different 3D positions and consequently to different pan-tilt $\{\theta_p, \theta_t\}$ values. Our work is based on the premise that human height can be exploited to infer depth information and avoid that ambiguity.

where $\alpha = -\|\mathbf{p}_{rb} \times \mathbf{p}_{rt}\| / (h_r(\mathbf{l}.\mathbf{p}_{rb})\|\mathbf{v}_z \times \mathbf{p}_{rt}\|)$, whereas $\mathbf{p}_{rt}$ and $\mathbf{p}_{rb}$ are the top and base points of a reference object in the image with height equal to h_r.

3.2. Pan-Tilt Angle Estimation

Considering the referential depicted in Figure 2, the center of rotation of the PTZ camera is given by $C = (0, \rho \sin\theta_t, -\rho\cos\theta_t)$, being ρ the displacement between the mechanical rotation axis and the image plane (which can be approximated by the camera focal distance f).

Given the 3D coordinates (X, Y, Z) of an interest point in the WCS, the location of that point with respect to the PTZ referential is obtained by:

$$\begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix} = [\,\mathbf{R}\,|\,\mathbf{T}\,] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{4}$$

and the corrected coordinates are given by:

$$\begin{pmatrix} X'_p \\ Y'_p \\ Z'_p \end{pmatrix} = \begin{pmatrix} X_p \\ Y_p - \rho\sin\theta_t \\ Z_p + \cos\theta_t \end{pmatrix}. \tag{5}$$

The corresponding pan and tilt angles are given by:

$$\theta_p = \arctan\left(\frac{X'_p}{Z'_p}\right), \tag{6}$$

and

$$\theta_t = \arcsin\left(\frac{Y'_p}{\sqrt{(X'_p)^2 + (Y'_p)^2 + (Z'_p)^2}}\right). \tag{7}$$

4. Experimental Results

To validate the proposed approach, the following procedure was adopted: given (x_s, y_s) and its corresponding (x_p, y_p) point, the algorithm error $(\Delta\theta)$ was determined by the angular distance between the estimated (X_p, Y_p, Z_p) and the 3D ray associated with (x_p, y_p). As compared to the typical reprojection error, this strategy was advantageous in the sense that it allowed a direct comparison with the PTZ field of view (FOV). Additionally, the height estimation performance in surveillance scenarios was also assessed by determining the deviation (Δh) from true subjects height.

The performance of our approach was assessed by carrying out three distinct evaluations: 1) height estimation performance; 2) independent performance analysis; 3) integration in an automated surveillance system.

In all evaluations, we used videos of ten different persons - comprising more than 1,000 frames - acquired both by the static and the PTZ camera while walking throughout an outdoor parking lot. Each pair of corresponding frames was annotated to mark the pixel location of the head and feet, in order to determine the performance of the proposed method with respect to Δh and $\Delta\theta$, which, in this case, corresponds to the angular distance between the estimated face location and its real position.

Besides, it is worth noting that in all evaluations a comparative analysis between inferring intrinsic and extrinsic camera parameters from calibration patterns (CB) and vanishing points (VP) was performed.

Furthermore, the inherent difficulties in accurately estimating vanishing points locations were taken into account. To assess the impact of incorrect vanishing point estimation, the previous experiments were replicated and the vanishing points location corrupted by a zero mean, Gaussian noise

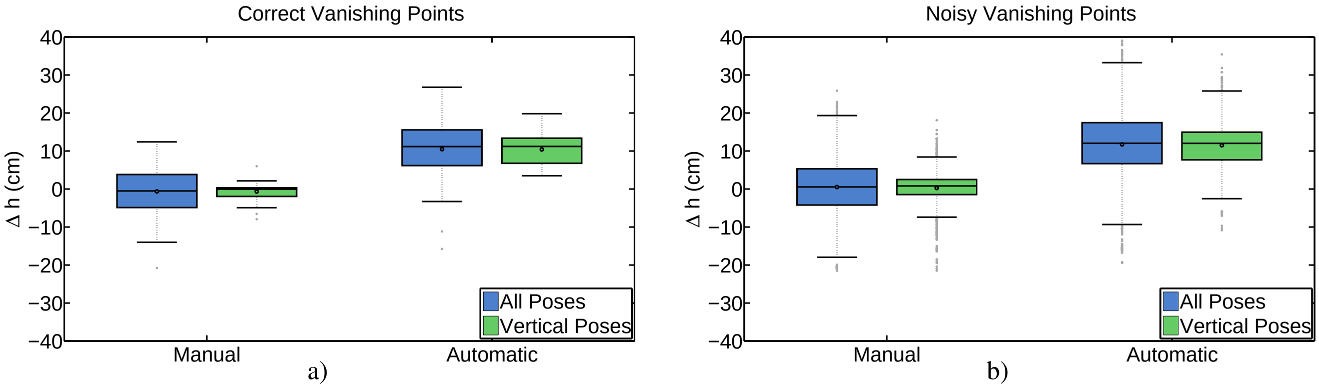


Figure 3. Height estimation performance in surveillance scenarios. Two distinct evaluations were carried out: an independent analysis (using manually annotated data) and the integration in an automated surveillance system (using data automatically obtained from a tracking module). Also, the accuracy of height estimation for vertical poses is presented, as well as the impact of noisy vanishing points.

with standard deviation of 10 pixel.

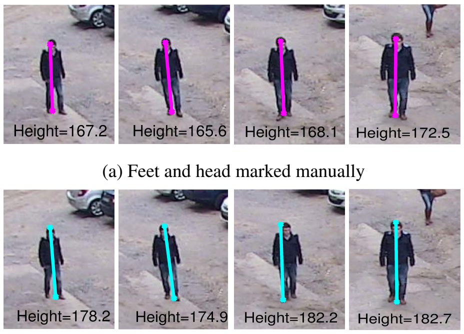
Finally, the feasibility of our approach in surveillance scenarios was determined by confronting $\Delta\theta$ with the PTZ FOV at different zoom magnifications (Figure 6). The percentage of faces successfully acquired summarizes the overall performance of the proposed calibration algorithm. The attained results for the several evaluations described are presented in Table 2 and compared with the work of Senior *et* al. [17].

4.1. Height Estimation Performance

The obtained results for height estimation in surveillance scenarios are presented in Figure 3. With regard to the type of data used, the distribution of Δh evidences that, in average, automatic height estimation is accurate ($\Delta h \approx 0$ and $\sigma_{\Delta h} = 6$ cm) for manually annotated data, while, it tends to overestimate the subjects height when using automatic annotations. We believe that a more robust tracker is likely to provide closer approximations to the manual annotations. Figure 4 illustrates three examples of incorrect height estimation due to the output of the tracking algorithm.

Furthermore, the approximately similar distribution of the second and third quantiles for correct and noisy vanishing points suggest that in the majority of the cases the aggregated noise did not affect significantly the height estimation performance. Only strong deviations in the vanishing points - typically more than 10px (the standard deviation used in our experiments) - affect severely height estimation performance.

Finally, it is worth noting that by building on [13], it is possible to narrow the height estimation error by relying solely on vertical poses. This result constitutes the basis of our future work, since the method accuracy may be improved by correcting height estimation on non-vertical poses with information obtained from the vertical ones.



(a) Feet and head marked manually

(b) Feet and head marked automatically

Figure 4. Examples of height estimation in surveillance scenarios using manually annotated data and automatic annotations obtained from a tracking algorithm. Note that the true height is 168 cm.

4.2. Independent Performance Evaluation

To assess the performance of the proposed calibration algorithm apart from the errors induced by the preceding phases of a surveillance system, the test videos were manually annotated, as described in section 4. The attained results are presented in Figure 5.

Regarding the strategy used for determining the camera projection matrix, it is evident that the use of vanishing points is advantageous in surveillance scenarios. The failure of typical calibration algorithms using planar calibration patterns can be explained by the arduousness in estimating the extrinsic parameters in outdoor scenarios. Ground irregularities and the reduced size of calibration patterns when compared to the extent of surveillance environments are the principal factors of inaccurate estimation of the rotation and translation matrices. On the contrary, in such scenarios, vanishing points are straightforward to determine using

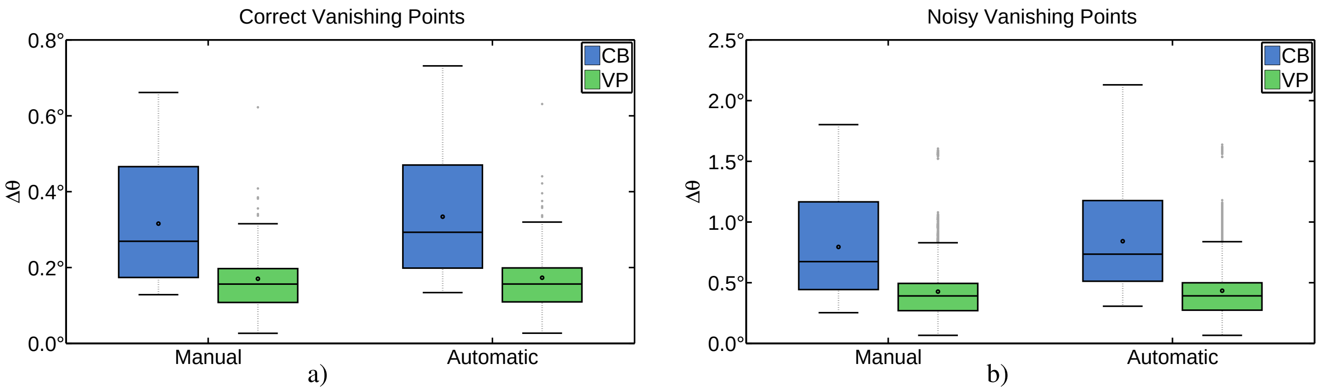


Figure 5. Overall performance of the proposed system. Two distinct evaluations were carried out: an independent analysis (using manually annotated data) and the integration in an automated surveillance system (using data automatically obtained from a tracking module). Additionally, two calibration strategies are compared, as well as the impact of noisy vanishing points.

pairs of parallel lines. Also, small inaccuracies in their estimation do not affect severely the performance of our approach (compare the differences in the average of $\Delta\theta$ when using VP), which provides additional support to the idea that a calibration based on vanishing points is preferred in surveillance scenarios.

Finally, the overall performance of the proposed algorithm has been summarized as the percentage of faces successfully acquired. This analysis was performed by comparing $\Delta\theta$ to the PTZ FOV at a given distance and the attained results are presented in Table 2. A comparative analysis with the results presented in [17], which also used height information to determine the 3D location of subjects, evidences a great improvement in the success rate, when considering the independent performance of the calibration module (manual data).

4.3. Integration in an Automated Surveillance System

Contrary to the previous evaluations, this experiment aims at analysing the impact of inaccuracies yielded by a tracking algorithm. For this purpose, the test videos were provided to an adaptive background subtraction algorithm [23] to automatically obtain head and feet locations through morphological operations.

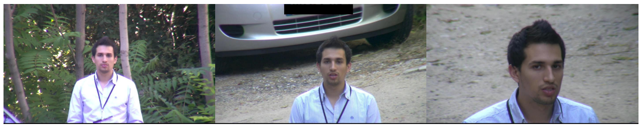
The attained results are presented in Figure 5 and, as in section 4.2, it is clear that the use of vanishing points is advantageous in surveillance scenarios when compared to typical calibration approaches. With regard to the type of data used, it is interesting to note that the integration of the proposed algorithm in an automated surveillance system does not affect severely the accuracy of the method (note the small differences in the average of $\Delta\theta$). Also, the same conclusion holds when comparing the use of correct and noisy vanishing points. These conclusions are also supported by the attained results presented in Table 2. An automated surveillance system using the proposed method attains a 87% success rate in capturing facial images at a distance using the maximum camera zoom, which outperforms the 4.8% success rate attained in [17].

Table 2. Percentage of faces successfully acquired. The performance of our method is compared for different calibration strategies (CB and VP) with [17], when using manual and automatic annotations.

Evaluation	Method	Without Noise (%)	With Noise (%)
Manual	Senior *et* al. [17]	30.3	-
	Our approach (CB)	58.4	57.2
	Our approach (VP)	89.4	89.1
Automatic	Senior *et* al. [17]	4.8	-
	Our approach (CB)	54.0	52.12
	Our approach (VP)	87.6	87.5

In order to provide further insights about the success rate of the proposed approach with respect to the zoom magnification used, we compare the pan ($\Delta\theta_p$) and tilt ($\Delta\theta_t$) displacements with the PTZ FOV for different zoom magnifications in Figure 6. Notice the extremely narrow FOV when using large zoom magnifications, and consequently, the importance of an accurate estimation of pan-tilt parameters.

5. Biometric Recognition



a) 35 m b) 25 m c) 15 m

Figure 7. Examples images captured by the PTZ camera used to randomly build gallery and probe sets.

In order to have a preliminary assessment about the face

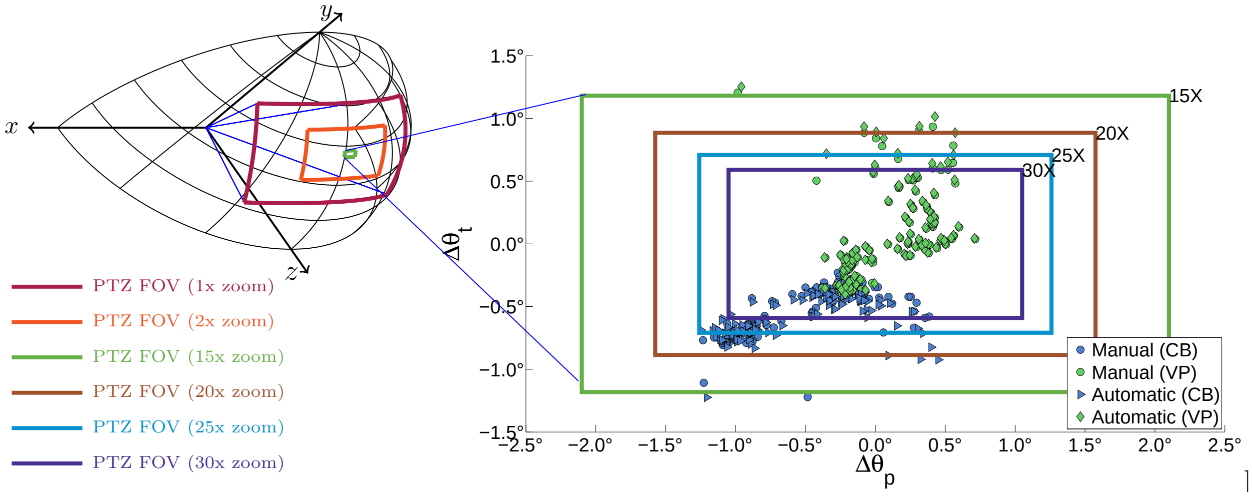


Figure 6. Accuracy of the proposed calibration algorithm. The pan and tilt errors are presented for two distinct calibration strategies (VP and CB). Notice the wide difference between the field of view (FOV) of the minimum (1x) and maximum (30x) zoom magnifications and the importance of accurate pan-tilt estimation (at the maximum zoom the tilt error can not exceed 0.5° to ensure a successful capture). Even so, our method has a success rate of 89%. when provided with manual data.

recognition capability of the proposed system, we have devised the following experiment: 20 subjects were asked to walk through the scene so that they could be imaged at 15m, 25m and 35m, as illustrated in Figure 7. This procedure was repeated five times for each person, yielding 15 images per subject.

To construct the gallery five images of each subject were used, and subsequently added to 13,000 images of 13,000 subjects of the MORPH database [16], in order to increase the gallery size and face recognition complexity. The remaining 10 images per subject were used as probe, comprising 200 probe images.

A face recognition algorithm based on sparse representation [19] was applied to perform recognition on grayscale face images. For comparison purposes, face recognition performance was evaluated as in [5]. The algorithm matching score was used to reject probe images with a score lower than t_r. The rank-1 identification accuracy attained in this experiment for $t_r = 0.18$ (20% rejection) is presented in Table 3 along with the results reported by Choi *et* al. [5] for one PTZ view. Even though the obtained results did not outperform the accuracy attained in [5], this can be explained by the fact that in [5] probe images are captured in an indoor environment while our dataset is acquired in an outdoor scenario. The dynamic illumination experienced in outdoor scenarios affects both face appearance and subjects facial expressions. As such, the attained results evidence that face recognition at-a-distance in outdoor environments is feasible, but further improvements are still required to improve face recognition performance in these scenarios. Dynamic lighting, head pose, motion blur, occlusions and non-neutral expressions are some key degradation factors that need to be taken into account by future face recognition algorithms to address face recognition assisted by PTZ cameras in surveillance scenarios.

Table 3. Face recognition accuracy in images acquired by the proposed master-slave calibration algorithm.

Method	t_r	Rank-1 identification accuracy (%)
Choi *et* al. [5]	0.31	64.5
	0.45	78.4
Our approach	0.2	76.5

6. Conclusion

In this paper, we introduced a master-slave calibration algorithm to accurately estimate the pixel to pan-tilt mapping using 3D information. Our work was based on the premise that inverse projective transform was feasible if one of the 3D components was known. Accordingly, we have shown that subjects height - which can be estimated from geometric cues available in the scene - is a valid information to solve this problem.

The workability of the proposed system was evidenced by the following results: 1) automatic height estimation is feasible in surveillance scenarios, particularly if combined with a vertical pose filter; 2) the typical displacement between the estimated 3D position and the actual face location enables face acquisition in 89% of the cases and in 87% of the cases when integrating the calibration algorithm in automated surveillance system; 3) the system performance is not severely affected by small deviations in vanishing point locations (up to 10 px).

Additionally, the applicability to face recognition was verified by using the proposed algorithm to automatically acquire high-resolution images of subjects at-a-distance. A face recognition test with 20 probe subjects and 13,020 gallery subjects has shown a rank-1 accuracy of 76.5%.

6.1. Acknowledgements

This work is supported by '*FCT - Fundação para a Ciência e Tecnologia*' (Portugal) through the project 'UID/EEA/50008/2013', 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*'.

References

[1] K. Bernardin, F. v. d. Camp, and R. Stiefelhagen. Automatic person detection and tracking using fuzzy controlled active cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, USA, 2007.

[2] A. D. Bimbo, F. Dini., G. Lisanti, and F. Pernici. Exploiting distinctive visual landmark maps in pan-tilt-zoom camera networks. *Computer Vision and Image Understanding*, 114(6):611–623, 2010.

[3] R. Bodor, R. Morlok, and N. Papanikolopoulos. Dual-camera system for multi-level activity recognition. In *IIEEE/RSJ International Conference on Intelligent Robots and Systems.*, volume 1, pages 643–648, 2004.

[4] Y. Cai, G. Medioni, and T. Dinh. Towards a practical PTZ face detection and tracking systems. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 31–38, USA, 2013.

[5] H.-C. Choi, U. Park, and A. Jain. Ptz camera assisted face acquisition, tracking & recognition. In *Proceedings of the Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*, pages 1–6, USA, 2010.

[6] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.

[7] C. Ding, B. Song, A. More, J. A. Farrel, and A. K. Roy-Chowdhury. Collaborative sensing in a distributed ptz camera network. *IEEE Transactions on Image Processing*, 11(7):3282–3295, 2012.

[8] L. Fiore, D. Fehr, R. Bodor, A. Drenner, G. Somasundaram, and N. Papanikolopoulos. Multi-camera human activity monitoring. *Journal of Intelligent and Robotic Systems*, 52(1):5–43, 2008.

[9] A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, and R. Bolle. Face cataloger: multi-scale imaging for relating identity to location. In *Proceedings of the IEEE conference on Advance Video and Signal Based Surveillance*, pages 13–20, USA, 2003.

[10] A. Jain, D. Kopell, K. Kakligian, and Y.-F. Wang. Using stationary-dynamic camera assemblies for wide-area video surveillance and selective attention. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 537–544, 2006.

[11] H. C. Liao and Y. C. Cho. A new calibration method and its application for the cooperation of wide-angle and pan-tilt-zoom cameras. *Information Technology Journal*, 7(8):1096–1105, 2008.

[12] Y. Liu, S. Lai, C. Zuo, H. Shi, and M. Zhang. A master-slave surveillance system to acquire panoramic and multi-scale videos. *The Scientific World Journal*, 2014.

[13] F. Lv, T. Zhao, and R. Nevatia. Self-calibration of a camera from video of a walking human. In *16th International Conference on Pattern Recognition*, volume 1, pages 562–567, 2002.

[14] L. Marchesotti, S. Piva, A. Turolla, D. Minetti, and C. S. Regazzoni. Cooperative multisensor system for real-time face detection and tracking in uncontrolled conditions. In *Proceedings SPIE 5689, Image and Video Communications and Processing*, USA, 2005.

[15] U. Park, H.-C. Choi, A. Jain, and S.-W. Lee. Face tracking and recognition at a distance: A coaxial and concentric PTZ camera system. *IEEE Transactions on Information Forensics and Security*, 8(10):1665–1677, 2013.

[16] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 341–345, 2006.

[17] A. W. Senior, A. Hampapur, and M. Lu. Acquiring multi-scale images by pan-titl-zoom control and automatic multi-camera calibration. In *Proceedings of the $7th$ IEEE workshop on Application of Computer Vision*, volume 1, pages 433–438, USA, 2005.

[18] F. Wheeler, R. Weiss, and P. Tu. Face recognition at a distance system for surveillance applications. In *Proceedings of the Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*, pages 1–8, USA, 2010.

[19] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[20] Y. Xu and D. Song. Systems and algorithms for autonomous and scalable crowd surveillance using robotic ptz cameras assisted by a wide-angle camera. *Autonomus Robots*, 29(1):53–66, 2010.

[21] L. You, S. Li, and W. Jia. Automatic weak calibration of master-slave surveillance system based on mosaic images. In *Proceedings of the $20th$ International Conference on Pattern Recognition*, pages 1824–1827, Turkey, 2010.

[22] X. Zhou, R. Collins, T. Kanade, and P. Metes. A master-slave system to acquire biometric imagery of human at distance. In *Proceedings of the $1st$ ACM International workshop on Video Surveillance*, pages 113–120, USA, 2003.

[23] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 2, pages 28–31, 2004.