Cooperative Mapping of Multiple PTZ Cameras in Automated Surveillance Systems

Chung-Chen Chen, Yi Yao, Anis Drira, Andreas Koschan, and Mongi Abidi
Imaging, Robotics, and Intelligent Systems Laboratory
Department of Electrical Engineering & Computer Science
The University of Tennessee
Knoxville, TN 37996 USA

Abstract

Due to the capacity of pan-tilt-zoom (PTZ) cameras to simultaneously cover a panoramic area and maintain high resolution imagery, researches in automated surveillance systems with multiple PTZ cameras have become increasingly important. Most existing algorithms require the prior knowledge of intrinsic parameters of the PTZ camera to infer the relative positioning and orientation among multiple PTZ cameras. To overcome this limitation, we propose a novel mapping algorithm that derives the relative positioning and orientation between two PTZ cameras based on a unified polynomial model. This reduces the dependence on the knowledge of intrinsic parameters of PTZ camera and relative positions. Experimental results demonstrate that our proposed algorithm presents substantially reduced computational complexity and improved flexibility at the cost of slightly decreased pixel accuracy, as compared with the work of Chen and Wang. This slightly decreased pixel accuracy can be compensated by consistent labeling approaches without added cost for the application of automated surveillance systems along with changing configurations and a larger number of PTZ cameras.

1. Introduction

Surveillance systems [1, 2] with multiple PTZ cameras became popular in the past decade, because of their capacity to simultaneously cover wide area and maintain high resolution imagery. Due to the time-varying relations among PTZ cameras, how to coordinate multiple PTZ cameras by means of changing their poses to achieve a better observation of the object of interest remains challenging. Even though there is a vast amount of literature on automatically calibrating larger camera networks [3, 4], those works mainly deal with stationary perspective cameras.

Thus, the works of Chen and Wang [5, 9] and Everts et al. [10] proposed to use known intrinsic parameters of PTZ cameras to direct their poses, namely pan, tilt, and zoom values, whenever a change is needed. In other words, we have to individually calibrate each PTZ camera [6, 7] to obtain their intrinsic parameters beforehand. This impedes their direct application to automated surveillance systems with changing configurations and a larger number of PTZ cameras. In particular, due to errors in the estimation of intrinsic parameters of PTZ camera, the works of Chen and Wang [5, 9] need one more optimization process, sensitivity analysis, to obtain the pose relation between PTZ cameras. This increases the system's computational complexity in the calibration process. To overcome their limitations, we propose a novel mapping approach that directly derives a unified polynomial model between the pan, tilt, and zoom values of PTZ cameras with unknown intrinsic parameters and setups in the scene.

In summary, the contribution of this paper is to derive the relation of pan, tilt, and zoom values between any pair of PTZ cameras without prior knowledge of their intrinsic parameters and relative positions. In comparison with the reference algorithm [9], our proposed approach not only reduces the dependence on the knowledge of intrinsic parameters of PTZ camera, but improves the degree of autonomy and reduces the system's computational complexity at the cost of slightly decreased pixel accuracy. In general, this slightly decreased pixel accuracy does not affect the overall performance for the application of automated surveillance systems, as long as the desired object can be seen within the field of view and can be compensated by consistent labeling approaches [19] without added cost.

The remainder of this paper is organized as follows. Section 2 addresses the problem. Section 3 describes our cooperative mapping method. Section 4 illustrates our experimental results and Section 5 concludes the paper.

2. Problem statement

The setup of a pair of PTZ cameras is shown in Figure 1. We choose the coordinate of the zero position of a selected camera as the reference world coordinate, where pan and tilt angles are both set to 0. A point $P_i = (X_i, Y_i, Z_i)^T$ in the reference world coordinate is projected onto the j^{th} PTZ camera's image coordinate $(x_{ij}, y_{ij}, \lambda_{ij})$ by

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

$$\begin{bmatrix} x_{ij} \\ y_{ij} \\ \lambda_{ij} \end{bmatrix} = \begin{bmatrix} f_{zoom,j} & s_{zoom,j} & x_{zoom,j} \\ 0 & \alpha_{zoom,j} f_{zoom,j} & y_{zoom,j} \\ 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} \cos\theta_{T,j} & 0 & -\sin\theta_{T,j} \\ 0 & 1 & 0 \\ \sin\theta_{T,j} & 0 & \cos\theta_{T,j} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{P,j} & \sin\theta_{P,j} \\ 0 & -\sin\theta_{P,j} & \cos\theta_{P,j} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}, \quad (1)$$

where $\theta_{P,j}$ and $\theta_{T,j}$ represent the pan and tilt angles of the j^{th} PTZ camera, respectively. $(x_{zoom,j}, y_{zoom,j})$ represents the principal point in the j^{th} PTZ camera. $f_{zoom,j}$ denotes the focal length of the j^{th}. $\alpha_{zoom,j}$ and $s_{zoom,j}$ respectively represent the aspect ratio and skew of the j^{th} PTZ camera.

In essence, $(x_{zoom,j}, y_{zoom,j})$, $f_{zoom,j}$, $\alpha_{zoom,j}$, and $s_{zoom,j}$ are subject to the changes of zoom value Z_j of the j^{th} camera. The same point is projected onto $\mathbf{p}_{ih} = (x_c, y_c, 1)^T$, the center of the image coordinates of the h^{th} PTZ, by proper pan, tilt, and zoom values:

$$\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} f_{zoom,h} & s_{zoom,h} & x_{zoom,h} \\ 0 & \alpha_{zoom,h} f_{zoom,h} & y_{zoom,h} \\ 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} \cos\theta_{T,h} & 0 & -\sin\theta_{T,h} \\ 0 & 1 & 0 \\ \sin\theta_{T,h} & 0 & \cos\theta_{T,h} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{P,h} & \sin\theta_{P,h} \\ 0 & -\sin\theta_{P,h} & \cos\theta_{P,h} \end{bmatrix} \left(\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + t_{hj} \right), \quad (2)$$

where t_{hj} denotes the translation vector between the optical center of the h^{th} and j^{th} PTZ cameras.

Based on the point correspondences, two equations can be derived from Equation (3) so as to solve for $\hat{\theta}_{P,h}$, $\hat{\theta}_{T,h}$, and $\hat{Z}_h$. In essence, to avoid the needed knowledge of internal and external parameters of each PTZ camera in the scene, we propose to use a set of polynomials to directly relate $(x_{ih}, y_{ih}, \theta_{P,h}, \theta_{T,h}, Z_h)$ and $(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j)$ from a training set. The training set is collected from tracking the same object in two PTZ cameras where the centroid of the object stays at the image center of the h^{th} camera, but can be anywhere in the image of the j^{th} camera. This object in both images maintains a constant-sized pixel resolution for the future applications such as behavior understanding, face recognition, and so forth. As a result, once Equation (4),

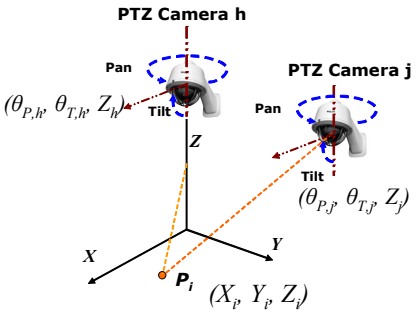


Figure 1: Typical setup of a pair of PTZ cameras.

$$\begin{cases} \hat{\theta}_{P,h} = f_P(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j) \\ \hat{\theta}_{T,h} = f_T(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j), \\ \hat{Z}_h = f_Z(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j) \end{cases} \quad (4)$$

is derived, we can direct the h^{th} PTZ camera to the position where the i^{th} object is supposed to be placed at its image center with a desired pixel size, which is based on the pan, tilt, zoom values and the image coordinates of the i^{th} object in the j^{th} PTZ camera.

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{P,h} & \sin\theta_{P,h} \\ 0 & -\sin\theta_{P,h} & \cos\theta_{P,h} \end{bmatrix} \begin{bmatrix} \cos\theta_{T,h} & 0 & -\sin\theta_{T,h} \\ 0 & 1 & 0 \\ \sin\theta_{T,h} & 0 & \cos\theta_{T,h} \end{bmatrix} \begin{bmatrix} \frac{1}{f_{zoom,h}} & \frac{-s_{zoom,h}}{f_{zoom,h}} & \frac{-x_{zoom,h}}{f_{zoom,h}} \\ 0 & \frac{1}{\alpha_{zoom,h} f_{zoom,h}} & \frac{y_{zoom,h}}{\alpha_{zoom,h} f_{zoom,h}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} - t_{hj}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{P,j} & -\sin\theta_{P,j} \\ 0 & \sin\theta_{P,j} & \cos\theta_{P,j} \end{bmatrix} \begin{bmatrix} \cos\theta_{T,j} & 0 & \sin\theta_{T,j} \\ 0 & 1 & 0 \\ \sin\theta_{T,j} & 0 & \cos\theta_{T,j} \end{bmatrix} \begin{bmatrix} \frac{1}{f_{zoom,j}} & \frac{-s_{zoom,j}}{f_{zoom,j}} & \frac{-x_{zoom,j}}{f_{zoom,j}} \\ 0 & \frac{1}{\alpha_{zoom,j} f_{zoom,j}} & \frac{y_{zoom,j}}{\alpha_{zoom,j} f_{zoom,j}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ij} \\ y_{ij} \\ \lambda_{ij} \end{bmatrix} \quad (3)$$

3. Cooperative mapping method

Our cooperative mapping methodology is inspired by the work of Chen et al [16]. They pointed out that. Existing algorithms [17, 18] in the area of spatial mapping between the omnidirectional and PTZ cameras need to have prior knowledge of project models of cameras, namely internal and external parameters, and the environment geometry. This impedes their direct application to surveillance systems with changing configurations. This is similar to surveillance systems with multiple PTZ cameras. Thus, our proposed cooperative method can be divided into two phases, the data acquisition phase and the data fitting phase. Figure 2 illustrates the flow chart of these two phases. The purpose of data acquisition phase is to collect desired information to relate directly $(x_{ih}, y_{ih}, \theta_{P,h}, \theta_{T,h}, Z_h)$ and $(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j)$. The purpose of data fitting phase is to derive Equation (4) by the collected data set from data acquisition phase.

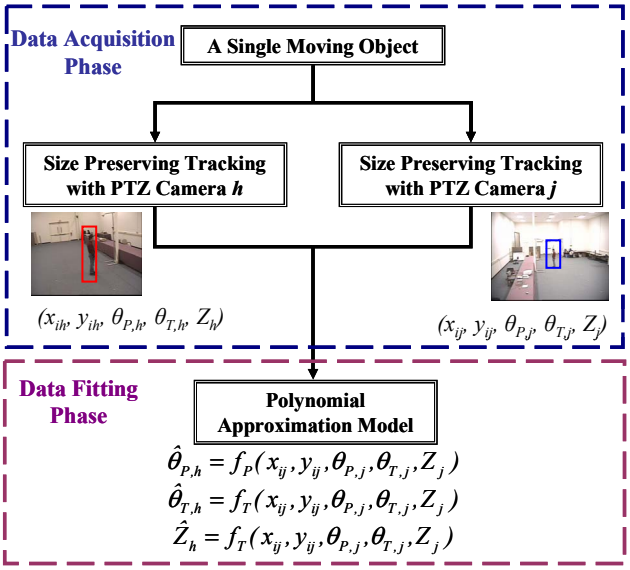


Figure 2: Illustration of our proposed cooperative mapping method.

3.1. Data acquisition phase

At first, a single object moves around randomly in the overlapped field of views (FOVs) of the j^{th} and h^{th} PTZ cameras to collect its motion trajectory including $(x_{ih}, y_{ih}, \theta_{P,h}, \theta_{T,h}, Z_h)$ and $(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j)$. The centroid of the object stays at the image center of the h^{th} camera but can be anywhere in the image of the j^{th} camera. This object in both images maintains a constant-sized pixel resolution for the future applications such as behavior understanding, face recognition, and so forth. Since the focus of this paper is not developing a size preserving tracking approach, we utilize the algorithm proposed by Fayman at el. [13] in here. Once $(x_{ih}, y_{ih}, \theta_{P,h}, \theta_{T,h}, Z_h)$ and $(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j)$ are collected, we enter to data fitting phase to obtain Equation (4).

3.2. Data fitting phase

Since the derivations for pan, tilt, and zoom functions are similar, in the following discussion, we will take the pan angle, $\hat{\theta}_{P,h} = f_P(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j)$, as an example to save space. In general, we first fit a model with all possible predictor variables [16] such as $\theta_{P,j}, \ldots, \theta^n_{P,j}, \theta_{T,j}, \ldots, \theta^n_{P,j}, \ldots, Z_j, \ldots, Z^n_j, \ldots, x_{ij}, y_{ij}, x^2_{ij}, Z_j x_{ij} y_{ij}, y^2_{ij}, \ldots, x^n_{ij}, x_{ij} y^{n-1}_{ij}, \ldots$, and $\theta^n_{T,j} y^n_{ij}$. Let w_i, with $i = 1, \ldots k$, represent these k predictor variables. The pan angle in a complete model can then be expressed as:

$$\hat{\theta}_{P,h(C)} = \gamma_0 + \gamma_1 w_1 + \gamma_2 w_2 + \ldots + \gamma_k w_k + \varepsilon_C \quad (5)$$

Where γ_i denotes the model fitting parameter and ε_C is a random error term with $E\{\varepsilon_C\} = 0$.

Usually not all predictor variables are equally significant. A subset of these variables can be found forming a reduced model:

$$\hat{\theta}_{P,h(R)} = \gamma_0 + \gamma_1 w_1 + \gamma_2 w_2 + \ldots + \gamma_g w_g + \varepsilon_R, \quad (6)$$

where $g < k$ and ε_R is a random error term with $E\{\varepsilon_R\} = 0$. Let SSE_C and SSE_R denote the sum of squared error of the complete and reduced models:

$$SSE_C = \Theta^T_{P,C}\Theta_{P,C} - \Theta^T_{P,C} W_{P,C} (W^T_{P,C} W_{P,C})^{-1} W^T_{P,C} \Theta_{P,C},$$
$$SSE_R = \Theta^T_{P,R}\Theta_{P,R} - \Theta^T_{P,R} W_{P,R} (W^T_{P,R} W_{P,R})^{-1} W^T_{P,R} \Theta_{P,R} \quad (7)$$

where $\Theta_{P,C}$ / $\Theta_{P,R}$ is the vector of all response variables in a complete/reduced model and $W_{P,C} / W_{P,R}$ is the vector of all predictor variables w_k / w_g in a complete/reduced model.

Intuitively, if w_1, w_2, …, and w_k are important information contributing variables, the complete model should have a smaller prediction error than the reduced model: $SSE_C \le SSE_R$. The greater the difference $(SSE_R - SSE_C)$ is, the stronger is the evidence to support the complete model that w_1, w_2, …, w_k are significant information contributing terms and to reject the reduced model: $H_0 : \beta_{g+1} = \beta_{g+2} = \ldots = \beta_k = 0$. Conversely, the

$$\hat{\theta}_{P,h} = 155.376 - 16.612\theta_{P,j} + 37.412\theta_{T,j} - 10.290x_j + 2.977y_i + 5.469\theta_{P,j}^2 - 23.364\theta_{T,j}^2 + 2.067x_j^2 - 0.804\theta_{P,j}\theta_{T,j}$$
$$+ 6.764\theta_{P,j}x_j - 1.940\theta_{T,j}y_j - 0.658x_jy_j$$
$$\hat{\theta}_{T,h} = -7.964 - 29.955\theta_{T,j} - 6.465y_j - 0.900\theta_{P,j}^2 + 24.060\theta_{T,j}^2 - 0.558y_j^2 - 1.386\theta_{P,j}x_j + 0.7291\theta_{P,j}y_j - 1.940\theta_{T,j}y_j \quad (10)$$
$$\hat{Z}_h = -0.439 + 0.7324\theta_{P,j} - 0.6218Z_j + 0.1221y_j + 0.0817\theta_{P,j}^2 + 0.086\theta_{T,j}^2 + 0.5934Z_j^2 + 0.0218x_j^2 - 0.0141y_j^2$$
$$+ 0.0153\theta_{P,j}\theta_{T,j} + 0.0723\theta_{T,j}y_j + 0.0596Z_jx_j + 0.0596Z_jy_j + 0.0125x_jy_j$$

acceptance of the reduced model suggests that the additional predictors in the complete model, $w_{g+1}, w_{g+2}, \ldots, w_k$, introduce no improvement to fitting accuracy. The predictors, w_1 , $w_2, \ldots, w_g$ in the reduced model are sufficient and more significant information contributing terms than predictors, $w_{g+1}, w_{g+2}, \ldots, w_k$. In other words, this becomes a model selection problem. Thus, we use the recently proposed extension to Akaike's information criterion called information complexity (ICOMP) [14] as our fitness function, which has been proved more efficient than existing fitness functions such as F test used in [16]. Other than its efficiency, another rationale for ICOMP as our fitness function is that it combines a badness-of-fit term with a measure of complexity of a model by taking into account the interdependencies of the parameter estimates, as well as the dependencies of the model residuals. This can increase the accuracy of estimation [16].

ICOMP is computed using Equations 8 and 9.

$$ICOMP = -2\log(L(\hat{w}_i)) + 2C_1(F^{-1}(\hat{w}_i)), \quad (8)$$

and

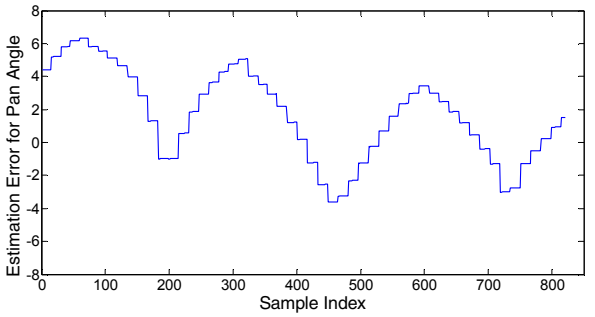
$$C_1(F^{-1}) = \frac{s}{2}\log\left[\frac{tr(F^{-1})}{s}\right] - \frac{1}{2}\log\left|F^{-1}\right|, \quad (9)$$

where $\hat{w}_i$ is the maximum likelihood estimator of w_i, L represents the likelihood function, F^{-1} is the inverse Fisher information matrix (IFIM), C_1 denotes the maximal information complexity of F^{-1}, s is the rank of F^{-1}, $|\ .\ |$ refers to the determinant,

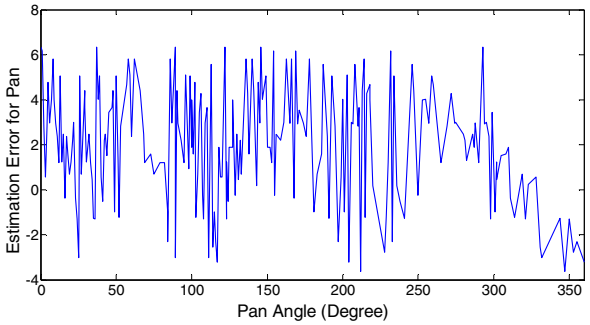
tr refers to the trace of the matrix. Equation (8) measures the lack of fit of the model, and Equation (9) measures the complexity of the estimated IFIM, which gives a scalar measure of the celebrated Cramér-Rao lower bound matrix. This takes into account the accuracy of the estimated parameters. The minimum value of ICOMP reveals the feature variable-subset is optimal in dimensionality and information content. More details behind the derivation of this formulation are available in [14]. In this paper, we use generic algorithm (GA) as a searching method along with the use of ICOMP criteria as the fitness function. How to use a GA-based procedure with informational complexity as the fitness function employed in this work is detailed in Bearse and Bozdogan [15].

4. Experimental results

In our experiments, we compare our proposed cooperative mapping approach with the reference algorithm [9] in an indoor surveillance system including two Pelco PTZ cameras(Spectra III SE dome with 640×480 pixels, $0^o \sim 360^o$ pang angle, $0^o \sim 90^o$ tilt angle, and 1 ~ 184 zoom position). To compare the accuracy between our and the reference algorithms [9], we conduct the following experiment. In our cooperative mapping approach, a total of 825 samples uniformly distributed in the scene are collected by a single moving person as the training set for the correspondence functions, which are shown in Equation (10). Figure 3 shows the estimation error in pan values, where Figure 3(a) and (b) indicate the estimation error in comparison with the original sample set (825 sample) and relative pan angles ($0^o \sim 360^o$), respectively. Figure 4 shows the estimation error in tilt values, where Figure 4(a) and (b) indicate the estimation error in comparison with the original sample set (825 sample) and relative tilt angles ($0^o \sim 90^o$), respectively. Figure 5 shows the estimation error in zoom values, where Figure 5(a) and (b) indicate the estimation error in comparison with the original sample set (825 sample) and relative zoom positions (1 ~ 184). In average, the estimation error in pan angle is less than ± 6.3°, which depicts the averaged relative estimation error with respect to complete pan angle 360^o is 1.7% (it is calculated by $\frac{6.3}{360}$). The estimation error in tilt angle is less than ± 8.5°, which depicts the averaged relative estimation error with respect to complete tilt angle 90^o is 9.4% (it is calculated by $\frac{8.5}{90}$). The estimation error in zoom value is less than ± 19.5°, which depicts the averaged relative estimation error with respect to complete zoom value 184 is 10.5% (it is calculated by $\frac{19.5}{184}$). For the reference algorithm, we manually calibrate two PTZ cameras to learn their intrinsic parameters fist. This manual intervention impedes their direct application to surveillance systems with changing setups and larger number of PTZ cameras in the scene.

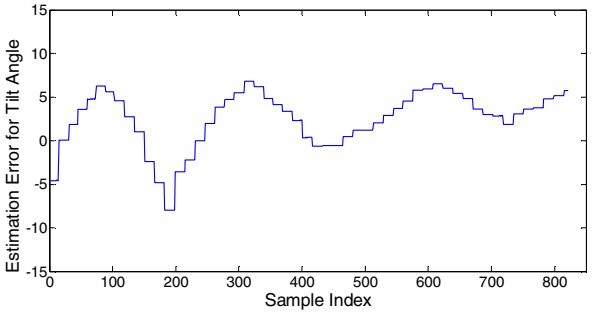


(a)

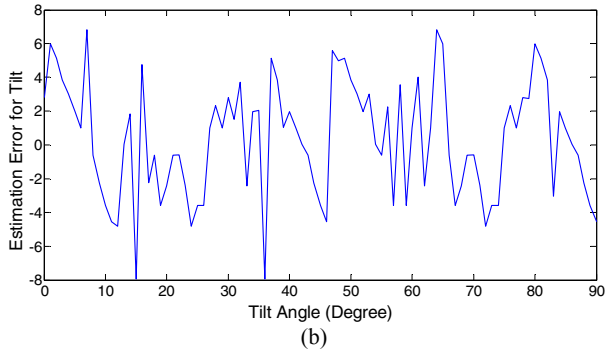


(b)

Figure 3: Estimation errors in pan values: (a) comparison to the original sample set (825 samples), (b) relative pan angle ($0^o \sim 360^o$).

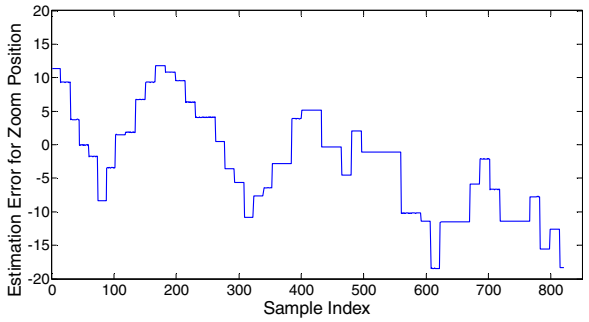


(a)

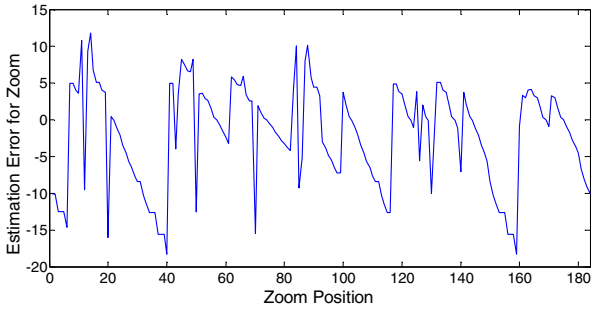


(b)

Figure 4: Estimation errors in tilt values: (a) comparison to the original sample set (825 samples), (b) relative tilt angle ($0^o \sim 90^o$).



(a)



(b)

Figure 5: Estimation errors in zoom values: (a) comparison to the original sample set (825 samples), (b) relative zoom position (1 ~ 184).

Table 1: Comparison between our and reference algorithms

	Averaged Pixel Distance Deviation	***Averaged Pixel Size Deviation***
Our Method	11.1%	16.7%
Chen and Wang [9]	9.2%	15.2%

Then we have 20 points forming a rectangular pattern in a 1 meter high table to estimate pose relationship based on back projections. Afterwards, we compare their accuracy to infer pixel correspondences between two PTZ cameras, where a single moving person is tested in the scene. Table 1 illustrates the comparison between our and reference algorithms. In Table 1, the averaged pixel distance deviation indicates the distance between the centroid of the object in the image and image center (320×240), when normalized with respect to the half of image width (320). The averaged pixel size deviation indicates the difference between the derived pixel size of the object and the desired pixel size (50×170=7500 pixels), when normalized with respect to the desired pixel size (7500). We can see that our proposed approach reduces the dependence on the knowledge of intrinsic parameters of the PTZ camera and improves the degree of autonomy at the cost of slightly decreased pixel accuracy, as compared to Chen and Wang' method.

Figure 6 shows real-time video sequences for our proposed, and Chen and Wang's approaches. In Figure 6,

the j^{th} PTZ camera uses Equation (10) to obtain $\hat{\theta}_{P,h}$, $\hat{\theta}_{T,h}$, and $\hat{Z}_h$ to direct the h^{th} PTZ camera to place the object in the center of the image with desired pixel size (7500) ideally. Figure 6 (a) shows the example where the single object is far away (18 meters) from the h^{th} PTZ camera (The tilt angle of the h^{th} PTZ camera is about 17^o). Figure 6 (b) shows the example where the single object is close to (3 meters) the h^{th} PTZ camera (The tilt angle of the h^{th} PTZ camera is about 75^o). In both Figure 6(a) and (b), the first row shows five different locations in images of the j^{th} PTZ camera, the second row shows their respective pixel locations and sizes, derived by our approach, in images of the h^{th} PTZ camera, and the third row shows their respective pixel locations and sizes, derived by Chen and Wang's approach, in images of the h^{th} PTZ camera. In both examples, the averaged pixel distance deviations are 12.6% and 10.3% for our proposed, and Chen and Wang's methods, respectively. The averaged pixel size deviations are 14.6% and 12.7% for our proposed, and Chen and Wang's methods, respectively.

Regardless of our proposed or Chen and Wang's methods, a consistent labeling approach is needed to identify the object of interest in both PTZ cameras after the occurrence of changing pose. Since this object of interest is maintained within the field of view of the h^{th} PTZ camera by both methods and maximal estimation errors for pan and tilt angles are 6.3^o and 8.5^o for our proposed method. Consistent labeling approaches can be carried out without added cost in here, because existing consistent labeling approaches such as scale-invariant feature transform (SIFT) [19] had been proved efficient when viewing angle is less than 50 degree. In other words, this slightly decreased pixel accuracy in our proposed approach has comparable result for the application of automated surveillance systems, as compared with Cheng and Wang's method. However, we reduce the dependence on the knowledge of intrinsic parameters of PTZ camera, thus holding the direct application to automated surveillance systems with changing configurations and a larger number of PTZ cameras.

5. Conclusion

In this paper, we proposed a novel approach that directly derives a unified polynomial model between the pan and tilt values of PTZ cameras with unknown intrinsic parameters and system setups in the scene. This reduces the dependence on the knowledge of intrinsic parameters of the PTZ camera, which most existing algorithms find challenging. Experimental results showed that our proposed method improves the feasibility and autonomy of the spatial mapping between PTZ cameras and reduces system's computational complexity at the cost of slightly decreased pixel accuracy, as compared with the work of Chen and Wang. This slightly decreased pixel accuracy can be compensated by consistent labeling approaches without added cost for the application of automated surveillance systems along with changing configurations and a larger number of PTZ cameras.

Acknowledgements

This work was supported in part by the University Research Program in Robotics under grant DOE-DE-FG52-2004NA25589.

References

[1] A. D. Bagdanov, A. D. Bimbo, and W. Nunziati, "Improving Evidential Quality of Surveillance Imagery Through Active Face Tracking", 18th Int'l Conf. on Pattern Recognition, 2006.

[2] F. Angella, L. Reithler, and F. Gallesio, "Optimal Deployment of Cameras for Video Surveillance Systems", IEEE Conf. on Advanced Video and Signal Based Surveillance, Sept. 2007.

[3] T. Svoboda, D. Martinec, and T. Pajdla, "A Convenient Multi-Camera Self-Calibration for Virtual Environments", PRESENCE: Teleoperators and Virtual Environments, vol. 14, pp. 407-422, 2005.

[4] Y. Sugimure and J. Sato, "Camera Calibration and Reconstruction from the Chain Connection of Mutual Camera Projections", 17th Int'l Conf. on Pattern Recognition, 2004.

[5] I.-H. Chen and S. J. Wang, "Efficient vision-based calibration for visual surveillance systems with multiple PTZ cameras", IEEE Int'l Conf. on Computer Vision Systems, 2006.

[6] H. Li and C. Shen, "An LMI Approach for Reliable PTZ Camera Self-Calibration", IEEE Conf. on Advanced Video and Signal Based Surveillance, Nov. 2007.

[7] L. Agapito, E. Hayman, and I. Reid, "Self-Calibration of Rotating and Zooming Cameras", Int'l Journal of Computer Vision, vol. 45, no. 2, pp. 107-127, 2001.

[8] M. H. Kutner, C. J. Nachtsheim, and J. Neter, Applied Linear Regression Models, McGraw Hill, 2004.

[9] I.-H. Chen and S.-J. Wang, "An Efficient Approach for the Calibration of Multiple PTZ Cameras", IEEE Trans. on Automation Science and Engineering, vol. 4, no. 2, 2007.

[10] I. Everts, N. Sebe, and G. A. Jones, "Cooperative Object Tracking with Multiple PTZ Cameras", the 14th Int'l Conf. on Image Analysis and Processing, Sept. 2007.

[11] D. Johnson, Applied Multivariate Methods for Data Analysis, Duxbury, 1998.

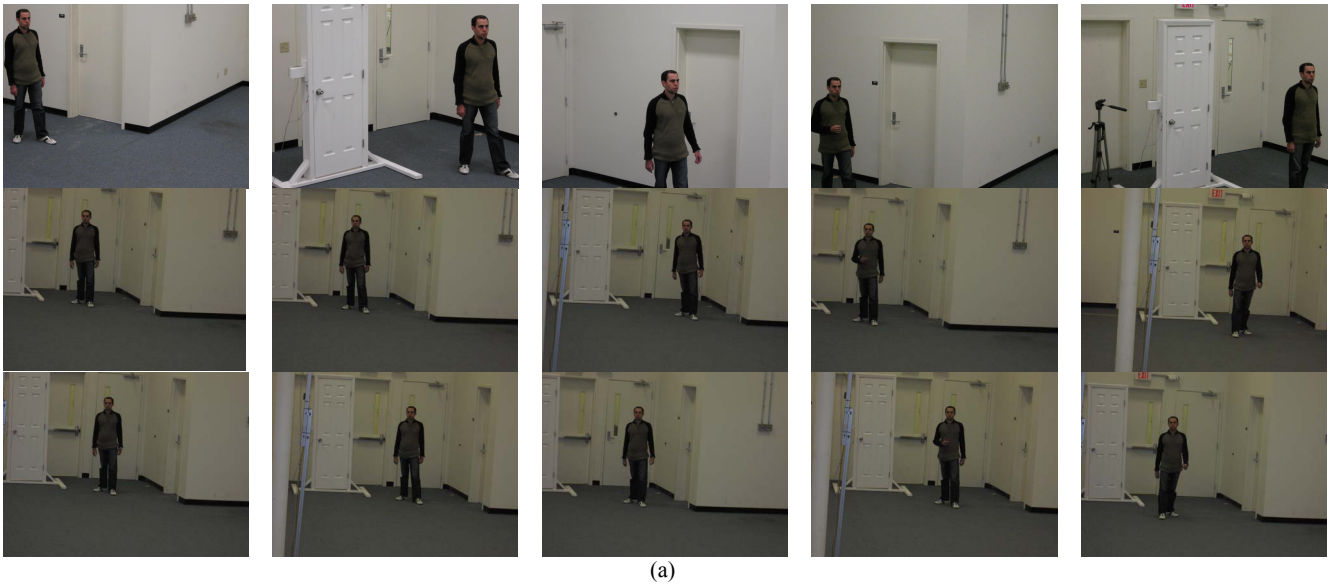
[12] J. Kannala and S. Brandt, "A generic camera calibration method for fish-eye lenses", the 17th Int'l Conf. on Pattern Recognition, Cambridge, UK, Aug. 2004.

[13] J. F. Fayman, O. Sudarsky, E. Rivlin, and M. Rudzsky, "Zoom tracking and its applications," Machine Vision and Applications, vol. 13, pp. 25-37, 2001.

[14] H. Bozdogan, "Akaike's Information Criterion and Recent Developments in Information Complexity", Journal of Mathematical Psychology, Vol. 44, NO. 1, 2000, pp. 62-91.

[15] P. M. Bearse, H. Bozdogan, "Subset selection in vector autoregressive (VAR) models using the genetic algorithm

with informational complexity as the fitness function," Systems Analysis Modeling Simulation, pp.61-91, 1998.
[16] C.-H. Chen, Y. Yao, D. Page, B.R. Abidi, A. Koschan, M. Abidi, "Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 18, No. 8, pp. 1052-1063, August 2008.
[17] Y. Cui, S. Samarasekera, Q. Huang, and M. Greiffenhagen, "Indoor monitoring via the collaboration between a peripheral sensor and a foveal sensor," in Proc. IEEE Workshop Visual Surveillance, Mumbai, India, Jan. 1998.
[18] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, and C. S. Regazzoni, "Dual camera intelligent sensor for high definition 360 degrees surveillance," IEE Proc.-Vis., Image, Signal Process., vol. 152, no. 2, pp. 250–257, Apr. 2005.
[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, Vol. 60, no.2, pp. 91-110, 2004.



(a)

(b)

Figure 6: Performance of our proposed and reference methods for a real-time multiple PTZ cameras system. In both (a) and (b), the first row shows five different locations in images of the j^{th} PTZ camera, the second row shows their respective pixel locations and sizes, derived by our approach, in images of the h^{th} PTZ camera, and the third row shows respective pixel locations and sizes, derived by Chen and Wang's approach, in images of the h^{th} PTZ camera. (a) The single object is far away (18 meters) from the h^{th} PTZ camera (The tilt angle of the h^{th} PTZ camera is about 17^{o}). (b) The single object is close to (3 meters) the h^{th} PTZ camera (The tilt angle of the h^{th} PTZ camera is about 75^{o}).