

Integration of multispectral face recognition and multi-PTZ camera automated surveillance for security applications

Research Article

Chung-Hao Chen¹, Yi Yao², Hong Chang³, Andreas Koschan³, Mongi Abidi³

¹ *Electrical and Computer Engineering,
Old Dominion University,
23529 Norfolk VA, USA*

² *GE Global Research Center,
12309 Nikayuna NY, USA*

³ *Electrical Engineering and Computer Science,
The University of Tennessee,
37996 Knoxville TN, USA*

Received 27 December 2012; accepted 28 February 2013

Abstract: Due to increasing security concerns, a complete security system should consist of two major components, a computer-based face-recognition system and a real-time automated video surveillance system. A computer-based face-recognition system can be used in gate access control for identity authentication. In recent studies, multispectral imaging and fusion of multispectral narrow-band images in the visible spectrum have been employed and proven to enhance the recognition performance over conventional broad-band images, especially when the illumination changes. Thus, we present an automated method that specifies the optimal spectral ranges under the given illumination. Experimental results verify the consistent performance of our algorithm *via* the observation that an identical set of spectral band images is selected under all tested conditions. Our discovery can be practically used for a new customized sensor design associated with given illuminations for an improved face recognition performance over conventional broad-band images. In addition, once a person is authorized to enter a restricted area, we still need to continuously monitor his/her activities for the sake of security. Because pan-tilt-zoom (PTZ) cameras are capable of covering a panoramic area and maintaining high resolution imagery for real-time behavior understanding, 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 to Chen and Wang's method [18].

Keywords: PTZ cameras • Surveillance systems • Multispectral images

© Versita sp. z o.o.

1. Introduction

Due to increasing security concerns, a complete security system should consist of two major components, a

computer-based face-recognition system and a real-time

automated video surveillance system. Face recognition has been widely used and has attracted significant research attention because of its wide range of applications in security and surveillance. Appearance variations caused by changes in lighting conditions constitute a major deteriorating factor of the system's recognition rate [1]. Multispectral images have been used to improve face recognition under various illuminations. There are two advantages of multispectral images over conventional images, which we took into consideration as our inspiration of utilizing multispectral images for face recognition. First, it is well known that humans tend to easily spot any color changes in the skin tones. The main obstacle for the universal color use in machine vision applications is that the cameras are not able to distinguish changes of surface color from color shifts caused by varying illumination [28]. Multispectral images in visible domain can provide a new avenue to separate the color of a subject and the illumination. Second, with multispectral images, we have the freedom to emphasize and/or suppress the contribution of images from certain narrowbands. Some of the approaches employed near infrared images that provide more information than the conventional images in the visible spectrum [2]. Pan *et al.* [3] used narrow-band spectral images in near infrared. Our previous work regarding the fusion of narrow-band spectral images [4] in the visible spectrum was the first performance comparison between multispectral images and conventional broad-band images. The fusion of a total or a subset of 25 band images can outperform conventional images for face recognition, especially when the probe and gallery images are acquired under different illuminations. This is due to the freedom to emphasize and/or suppress the contribution of images from certain narrow bands, when using multispectral images. Contrarily, conventional monochromatic and RGB images provide only one- or three-broad-band responses.

In this paper, as an extension of our previous study in spectral range selection for face recognition [5], we investigate the robustness of our algorithm, focusing on two critical steps: probability density function (PDF) estimation and divergence computation. The efficiency of PDF estimation depends on the selection of the kernel function, which may depend on the distribution of the actual input data, in our case the similarity scores of the genuine and imposter sets. The characteristics of the input data may vary according to a large variety of factors, such as the recognition engine and illumination conditions. This raises the question of whether the performance of the band selection algorithm depends on the characteristics of the input data. If the answer is yes, the use of kernel function and distance measure needs to be optimized empirically in

advance according to the specific set of input data, which impedes the application of the proposed selection algorithm in a plug-and-play manner. To maximize its universal applicability, it is desired that the performance of the proposed algorithm is robust to the selection of the aforementioned parameters.

Once a person is authorized to enter a restricted area, we still need to continuously monitor his/her activities for the sake of security. Due to fatigue, the possibility of missing alarms is high, even for well-trained security personnel. These issues lead to the need for a real-time automated surveillance system that automatically detects, tracks, and records security violations. Surveillance systems [10, 11] 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 [12, 13], those works mainly deal with stationary perspective cameras.

Thus, the works of Chen and Wang [14, 18] and Everts *et al.* [19] 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 [15, 16] 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 [14, 18] 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 contributions of this paper are: (1) The robustness and consistency of the proposed algorithm is verified by the observation that identical band ranges are selected *via* various implementations for different input data. Therefore, with the most basic implementation of the Gaussian kernel and Jeffrey divergence, a smaller number of narrow-band images can be selected according to the illumination conditions and fused for an improved recognition performance; (2) Our approach is able to derive the relation of pan, tilt, and zoom values between any pair of PTZ cameras without prior knowledge of their intrinsic

sic parameters and relative positions. In comparison with the reference algorithm [18], 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 [27] without added cost.

The remainder of this paper is organized as follows. Section 2 presents our band selection algorithm and describes various implementations of PDF estimation and divergence computation. Section 3 shows our cooperative mapping method. Experimental results are given in Section 4 and conclusions are drawn in Section 5.

2. Band selection approach

Face recognition starts typically with image preprocessing including segmentation and normalization. Afterward, salient features are extracted based on which similarity scores of a pair of face images, one as the probe and the other as the gallery, are calculated. Let S_{ij}^k denote the similarity score between the gallery image of the i^{th} subject and the probe image of the j^{th} subject collected at the k^{th} band. The similarity scores in each band can be divided into two groups, referred to as the genuine G_k and imposter I_k sets. The genuine and imposter sets are defined as: $G_k : \{S_{ij}^k, i = j\}$ and $I_k : \{S_{ij}^k, i \neq j\}$, respectively. The genuine set contains the similarity scores with probe and gallery images from the same subject while the imposter set consists of similarity scores with the probe and gallery images from different subjects. Without loss of generality, we assume that a higher similarity score indicates a better match. Ideally, the genuine and imposter sets should cluster at the high and low end of the score scale, respectively, without overlap so that an appropriate threshold can be derived to completely separate the genuine matches from the imposter ones. Under such conditions, a perfect 100% recognition rate can be achieved. However, in practical situations, there usually exist overlapped regions between these two sets. An important criterion in evaluating the effectiveness of the recognition system is the separation between the similarity scores of the genuine and imposter sets. Please refer to our previous work [5] for detailed discussions regarding band separation.

We propose using the band separation between the genuine and imposter sets to select the optimal spectral range of face images for given illumination conditions. Figure 1 illustrates the pipeline of the face recognition algorithm with our automated band selection mechanism. Based on the features extracted from probe and gallery images, similarity scores for all pairs of probe and gallery images are computed. Then, band selection is performed as follows. (1) The distribution of the similarity scores of the genuine and imposter sets are estimated using kernel functions. (2) Divergence is calculated to quantitatively describe the separation between these two distributions. (3) The optimal m bands can be chosen by sorting the divergence values in a descending order. The m bands corresponding to the first m divergence values in the sorted sequence are selected. Finally, the images from the selected bands are fused and fed into a classification engine that outputs the recognition rate.

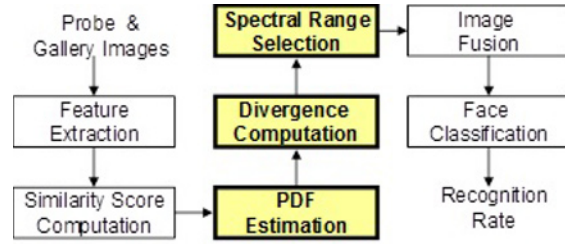


Figure 1. Illustration of the algorithm pipeline. The proposed band selection algorithm is highlighted in bold

To achieve automated selection of optimal multispectral bands, we need an accurate estimation of the PDFs of the genuine $\hat{p}_{G,k}(x)$ and imposter $\hat{p}_{I,k}(x)$ sets for the k^{th} band and a quantified measure D to evaluate the separation between them. In this paper, we investigate the performance of the proposed algorithm with various implementations of PDF estimation and divergence computation. Our motivation is to show that our algorithm is sufficiently robust so that its performance is independent of the implementation of the PDF estimation and divergence computation. This is an attractive attribute and is important for practical implementation.

From the similarity scores of various subjects, the distributions of the genuine and imposter sets, $\hat{p}_{G,k}(x)$ and $\hat{p}_{I,k}(x)$, are estimated by using kernel density estimation (KDE) [6]:

$$\hat{p}_{G,k}(x) = \frac{1}{Nh_{G,k}} \sum_{i=1}^N K\left(\frac{x - S_{ii}^k}{h_{G,k}}\right), \quad (1)$$

$$\hat{p}_{I,k}(x) = \frac{1}{N(N-1)h_{I,k}} \sum_{i=1}^N \sum_{j=1, j \neq i}^N K\left(\frac{x - S_{ij}^k}{h_{I,k}}\right), \quad (2)$$

where $K()$ denotes the kernel function with the property of $\int K(t) dt = 1$, $h_{G,k}/h_{l,k}$ is the smoothing parameter, and N is the total number of subjects.

The quality of a kernel estimate depends on both the shape of the kernel and the value of its smoothing parameter. The following kernel functions are commonly used: triangle $(1 - |t|)$, Gaussian $\frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$, Epanechnikov $\frac{3}{4}(1-t^2)$, biweight/quartic $\frac{15}{16}(1-t^2)^2$, triweight $\frac{35}{32}(1-t^2)^3$, and cosine $\frac{\pi}{4} \cos(\frac{\pi}{2}t)$. The Gaussian function is defined in $(-\infty, \infty)$ while others are defined in $[-1, 1]$.

As its name suggests, the smoothing parameter controls the smoothness of the density estimate. A smaller smoothing parameter leads to spiky estimates. The bias in the density estimate is small but the variance is large. In contrast, a larger smoothing parameter results in over-smoothing with a smaller variance but a larger estimation bias. Minimizing the asymptotic mean integrated square error (AMISE) [7] is the most commonly used method of choosing the smoothing parameter, which is normally denoted as h_{AMISE} :

$$h_{AMISE} = \left[\frac{\rho(K)}{N\mu(K)^2\sigma(p')} \right]^{1/3}, \quad (3)$$

where $\rho(K) = 2 \int_{-\infty}^{\infty} xK(x)K_l(x)dx$, $\mu(K) = \int_{-\infty}^{\infty} x^2K(x)dx$, and $\sigma(p') = \int_{-\infty}^{\infty} p'(x)^2dx$ with $K_l(x) = \int_{-\infty}^x K(x)dx$. A more advanced approach of estimating the smoothing parameter explores a more complicated criterion that considers the trade-off between the estimation bias and variance. The optimal parameter h_{ICOMP} is obtained by minimizing the information complexity (ICOMP) defined as follows [8]:

$$\begin{aligned} ICOMP(K, h) = & 2n \ln(n-1) + 2n \ln(h) + \\ & - 2 \sum_{i=1}^n \ln \left[\sum_{j \neq i}^n K \left(\frac{x_i - x_j}{h} \right) \right] + \\ & + 2C_1 \left(\text{Cov}(\hat{\theta}) \right), \end{aligned} \quad (4)$$

where the covariance matrix is given by $\text{Cov}(\hat{\theta}) = \hat{F}_i^{-1} \hat{R} \hat{F}_i^{-1}$. $\hat{F}_i^{-1}$ is the Inverse Fisher Information Matrix (IFIM) and $\hat{R}$ is the estimated outer-product form of the Fisher information. n represents representative principal components. The $C_1(\bullet)$ information complexity is defined by:

$$C_1 \left(\text{Cov}(\hat{\theta}) \right) = \frac{s}{2} \ln \left[\frac{\text{trace} \left(\text{Cov}(\hat{\theta}) \right)}{\text{rank} \left(\text{Cov}(\hat{\theta}) \right)} \right] - \frac{1}{2} \ln \left| \text{Cov}(\hat{\theta}) \right|, \quad (5)$$

where trace refers to the trace of the matrix. Equation (4) measures the lack of fit of the model, and Equation (5) 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 [22]. In this paper, we only use generic algorithm (GA) as 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 [8].

Once the PDFs of the similarity scores from the genuine and imposter sets are estimated, the remaining question is how to quantitatively evaluate the distance between the two PDFs. Probabilistic distance measures are exploited. To simplify the notations, we use $p_1(x)$ and $p_2(x)$ to represent the density functions of two sets, which in our case are the genuine and imposter sets. Table 1 defines a list of probabilistic distance measures often found in literature [9]. These distances have the following relations. (1) The Bhattacharyya distance is a special case of the Chernoff distance with $\alpha_1 = \alpha_2 = \frac{1}{2}$. (2) The Matsumoto distance is related to the Bhattacharyya distance by $D_M = \sqrt{2[1 - \exp(-D_B)]}$. (3) The relation between the Kullback-Leibler and Jeffrey divergence (a symmetric version of the Kullback-Leibler divergence) is given by $D_J(p_1, p_2) = D_{KL}(p_1||p_2) + D_{KL}(p_2||p_1)$. (4) The Kolmogorov distance is a special case of the Lissack-Fu distance with $\alpha_1 = 1$.

3. Cooperative mapping approach

The setup of a pair of PTZ cameras is shown in Figure 2. 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}

Table 1. List of probabilistic distances and their definitions, where $0 < \alpha_1, \alpha_2 < 1$, $\alpha_1 + \alpha_2 = 1$, and π_1 and π_2 are prior probabilities of classes 1 and 2, respectively

Distance	Definition
Bhattacharyya	$D_B(p_1, p_2) = -\log \left\{ \int_X \sqrt{p_1(x) p_2(x)} dx \right\}$
Chernoff	$D_C(p_1, p_2) = -\log \left\{ \int_X p_1^{\alpha_1}(x) p_2^{\alpha_2}(x) dx \right\}$
Kullback-Leibler	$D_{KL}(p_1 p_2) = \int_X \hat{p}_1(x) \log \frac{\hat{p}_1(x)}{\hat{p}_2(x)} dx$
Jeffrey	$D_J = \int [p_1(x) - p_2(x)] \log \frac{p_1(x)}{p_2(x)} dx$
Matusita	$D_M(p_1, p_2) = \sqrt{\int_X [\sqrt{p_1(x)} - \sqrt{p_2(x)}]^2 dx}$
Patrick-Fisher	$D_{PF}(p_1, p_2) = \sqrt{\int_X [p_1(x) \pi_1 - p_2(x) \pi_2]^2 dx}$
Lissack-Fu	$D_{LF} = \int_X p_1(x) \pi_1 - p_2(x) \pi_2 ^{\alpha_1} [p_1(x) \pi_1 + p_2(x) \pi_2]^{\alpha_2} dx$
Kolmogorov	$D_K = \int_X p_1(x) \pi_1 - p_2(x) \pi_2 dx$

PTZ camera's image coordinate $(x_{ij}, y_{ij}, \lambda_{ij})$ by

pan, tilt, and zoom values:

$$\begin{bmatrix} x_{ij} \\ y_{ij} \\ \lambda_{ij} \end{bmatrix} = \begin{bmatrix} f_{zoom,j} & s_{zoom,j} & x_{zoom,j} \\ 0 & \alpha_{zoom,j} f_{zoom,j} & y_{zoom,j} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos \theta_{T,j} & 0 & -\sin \theta_{T,j} \\ 0 & 1 & 0 \\ \sin \theta_{T,j} & 0 & \cos \theta_{T,j} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{P,j} & \sin \theta_{P,j} \\ 0 & -\sin \theta_{P,j} & \cos \theta_{P,j} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}, \quad (6)$$

$$\begin{bmatrix} x_C \\ y_C \\ 1 \end{bmatrix} = \begin{bmatrix} f_{zoom,h} & s_{zoom,h} & x_{zoom,h} \\ 0 & \alpha_{zoom,h} f_{zoom,h} & y_{zoom,h} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos \theta_{T,h} & 0 & -\sin \theta_{T,h} \\ 0 & 1 & 0 \\ \sin \theta_{T,h} & 0 & \cos \theta_{T,h} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{P,h} & \sin \theta_{P,h} \\ 0 & -\sin \theta_{P,h} & \cos \theta_{P,h} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + t_{hj}, \quad (7)$$

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 $p_{ih} = (x_c, y_c, 1)^T$, the center of the image coordinates of the h^{th} PTZ, by proper

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

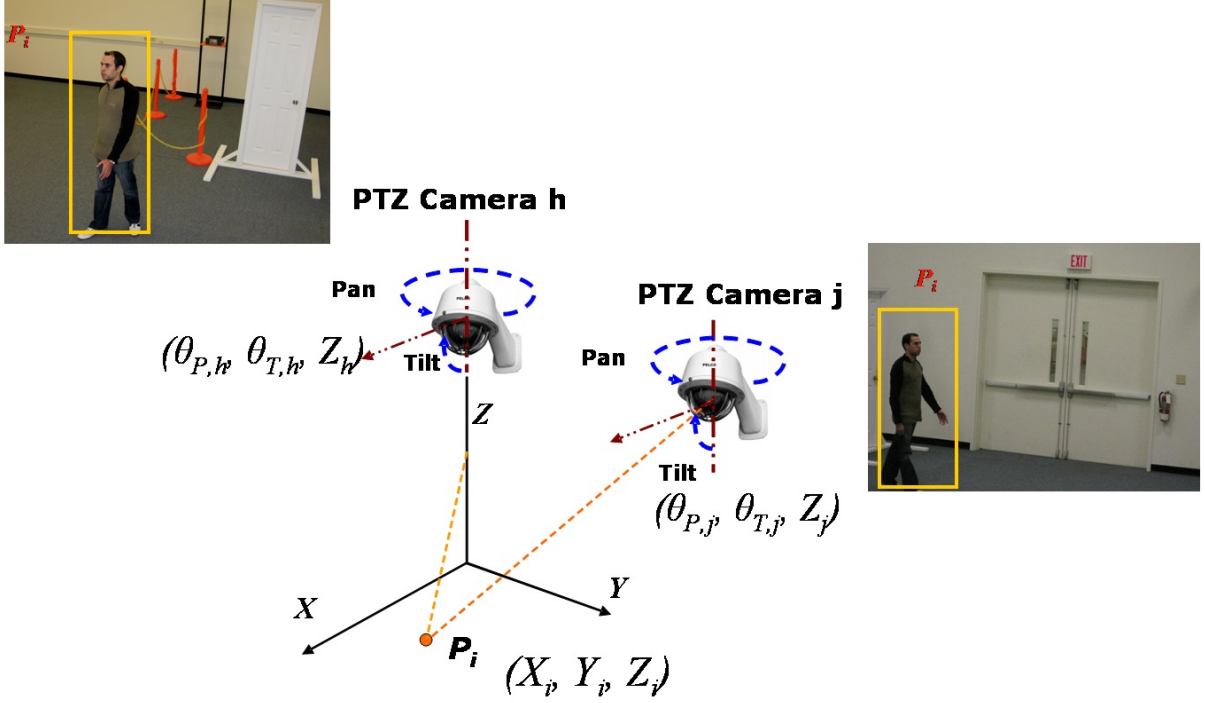


Figure 2. Typical setup of a pair of PTZ cameras

Based on the point correspondences, two equations can be derived from,

$$\begin{aligned}
 \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{P,h} & \sin \theta_{P,h} \\ 0 & -\sin \theta_{P,h} & \cos \theta_{P,h} \end{bmatrix} \cdot \begin{bmatrix} \cos \theta_{T,h} & 0 & -\sin \theta_{T,h} \\ 0 & 1 & 0 \\ \sin \theta_{T,h} & 0 & \cos \theta_{T,h} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{f_{zoom,h}} & \frac{-s_{zoom,h}}{f_{zoom,h}} & \frac{-x_{zoom,h}}{f_{zoom,h}} \\ 0 & \frac{1}{\alpha_{zoom,h} f_{zoom,h}} & \frac{y_{zoom,h}}{\alpha_{zoom,h} f_{zoom,h}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ y_c \\ 1 \end{bmatrix} - t_{hj} \\
 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{P,j} & -\sin \theta_{P,j} \\ 0 & \sin \theta_{P,j} & \cos \theta_{P,j} \end{bmatrix} \begin{bmatrix} \cos \theta_{T,j} & 0 & \sin \theta_{T,j} \\ 0 & 1 & 0 \\ \sin \theta_{T,j} & 0 & \cos \theta_{T,j} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{f_{zoom,j}} & \frac{-s_{zoom,j}}{f_{zoom,j}} & \frac{-x_{zoom,j}}{f_{zoom,j}} \\ 0 & \frac{1}{\alpha_{zoom,j} f_{zoom,j}} & \frac{y_{zoom,j}}{\alpha_{zoom,j} f_{zoom,j}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{ij} \\ y_{ij} \\ \lambda_{ij} \end{bmatrix}, \quad (8)
 \end{aligned}$$

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 (9),

$$\begin{cases} \hat{\theta}_{P,h} = f_P(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j) \\ \hat{\theta}_{T,h} = f_T(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j) \\ \hat{Z}_h = f_Z(x_{ij}, y_{ij}, \theta_{P,j}, \theta_{T,j}, Z_j) \end{cases} \quad (9)$$

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.

Our cooperative mapping methodology is inspired by the work of Chen *et al.* [24]. They pointed out that existing algorithms [21, 25, 26] 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 3 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 (9) by the collected data set from data acquisition phase.

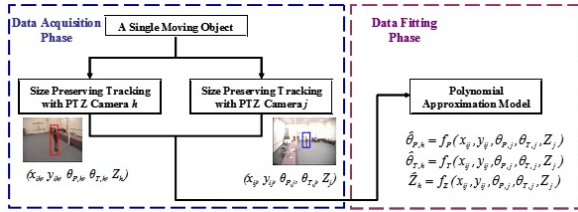


Figure 3. Illustration of our proposed cooperative mapping method

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 *et al.* [22] 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 (9).

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 [17, 23] with different n^{th} -order terms such as $\theta_{P,j}, \dots, \theta_{P,j}^n, \theta_{T,j}, \dots, \theta_{P,j}^n, \dots, Z_j, \dots, Z_j^n, \dots, x_{ij}, y_{ij}, x_{ij}^2, Z_j x_{ij} y_{ij}, y_{ij}^2, \dots, x_{ij}^n, x_{ij} y_{ij}^{n-1}, \dots$, and $\theta_{T,j}^n y_{ij}^n$. Let w_i , with $i = 1, \dots, k$, represent these k predictor variables. The pan

angle in a complete model can then be expressed as:

$$\hat{\theta}_{P,h(C)} = \gamma_0 + \gamma_1 w_1 + \gamma_2 w_2 + \dots + \gamma_k w_k + \varepsilon_C, \quad (10)$$

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:

$$\hat{\theta}_{P,h(R)} = \gamma_0 + \gamma_1 w_1 + \gamma_2 w_2 + \dots + \gamma_g w_g + \varepsilon_R, \quad (11)$$

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:

$$\begin{aligned} SSE_C &= \Theta_{P,C}^T \Theta_{P,C} + \\ &\quad - \Theta_{P,C}^T W_{P,C} (W_{P,C}^T W_{P,C})^{-1} W_{P,C}^T \Theta_{P,C}, \\ SSE_R &= \Theta_{P,R}^T \Theta_{P,R} + \\ &\quad - \Theta_{P,R}^T W_{P,R} (W_{P,R}^T W_{P,R})^{-1} W_{P,R}^T \Theta_{P,R}, \end{aligned} \quad (12)$$

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 $w_1, w_2, \dots$, and w_k are important information contributing variables, the complete model should have a smaller prediction error than the reduced model: $SSE_C \leq SSE_R$. The greater the difference ($SSE_R - SSE_C$) is, the stronger is the evidence to support the complete model that $w_1, w_2, \dots, w_k$ are significant information contributing terms and to reject the reduced model: $H_0: \gamma_{g+1} = \gamma_{g+2} = \dots = \gamma_k = 0$. Conversely, the acceptance of the reduced model suggests that the additional predictors in the complete model, $w_{g+1}, w_{g+2}, \dots, w_k$, introduce no improvement to fitting accuracy. The predictors, $w_1, w_2, \dots, w_g$ in the reduced model are sufficient and more significant information contributing terms than predictors, $w_{g+1}, w_{g+2}, \dots, w_k$. 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) [8] as our fitness function, which is briefed in Section 2. ICOMP has been proved more efficient than existing fitness functions such as F -test used in [20, 23]. 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 [23].

4. Experimental results

First, the performance as well as robustness of the band selection method is investigated *via* a variety of choices of kernels, smoothing parameters, and distance measures. Two experiments are conducted with gallery and probes images collected from different illuminations. Next, we compare our proposed cooperative mapping approach with the reference algorithm [18] in an indoor surveillance system including two Pelco PTZ cameras (Spectra III SE dome with 640×480 pixels, $0^\circ \sim 360^\circ$ pang angle, $0^\circ \sim 90^\circ$ tilt angle, and $1 \sim 184$ zoom position).

4.1. Fluorescent gallery and halogen probe

In this experiment, the spectral bands of multispectral face images under halogen light are selected *via* the proposed algorithm while gallery images are under a different indoor lighting, fluorescent light. There are 25 sets of probe images, sub-spectral narrow-band images between wavelength 480 nm and 720 nm with an increment of 10 nm. We investigate the ranking results *via* various distance measures of these 25 bands. The PDFs are estimated using different kernel functions with the smoothing parameter optimized by the AMISE and ICOMP criteria.

Table 2 lists the top three bands with the highest separation between the genuine and imposter sets. It is obvious that regardless of the different combinations of kernel functions, smoothing parameters, and distance measures, the same band range, 610 nm–640 nm, is identified. We could conclude that the ranking results of bands are robust to the selection of parameters. The normalized distances with respect to the band wavelength are shown in Figure 4. To save space, only the results based on the Gaussian and cosine kernels with h_{AIMSE} are shown. Similar observations apply to other combinations. Even though the distances show various values at certain wavelength, the trends and ranking results from the largest distance values to the smallest distance values are clearly similar. For example, the top band is 610 or 620 nm for all the tested kernels and distance measures. The above experiment verifies the robustness of the proposed algorithm. We now study the recognition performance of the images obtained by the fusion of multispectral narrow-band images of the chosen bands. Figures 5 and 6 demonstrates the rank-one recognition rate of various probes, including the single subspectral band, conventional broadband, and fused images from two and three bands. As expected, the fused images from the selected narrowbands yield a higher recognition rate, indicated by an increase of 20% relative improvement in the rank-one rate in comparison with the conventional broadband image set.

The rank-one recognition rate for (610 nm, 630 nm, and 640 nm) and (610 nm, 620 nm, and 640 nm) are the same (97.14%), as shown in Figure 5, which outperforms the conventional broad-band images by approximately 8.58%. This demonstrates the effectiveness of our band selection algorithm.

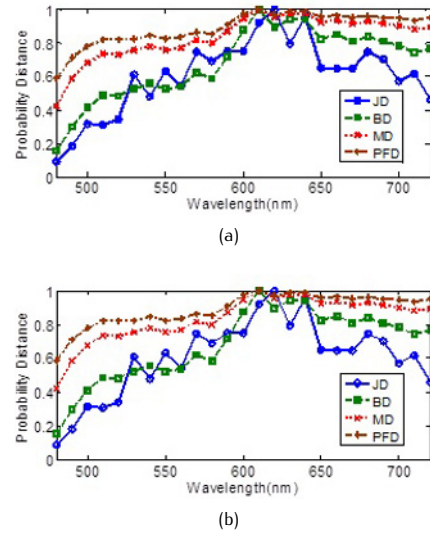


Figure 4. Normalized probability distances along the visible spectrum based on the Jeffery divergence (JD), Bhattacharyya distance (BD), Matusita distance (MD), and Patrick-Fisher distance (PFD). (a) Gaussian kernel and (b) cosine kernel. The smoothing parameter is obtained by AMISE. The distance values are normalized to [0, 1] for comparison purpose

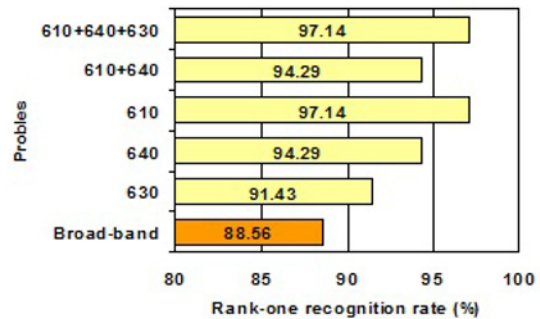


Figure 5. Rank-one recognition rate of different probe sets, including conventional broad-band images, single sub-spectral images, and fused images from selected spectral range in the experiment of fluorescent gallery and halogen probes

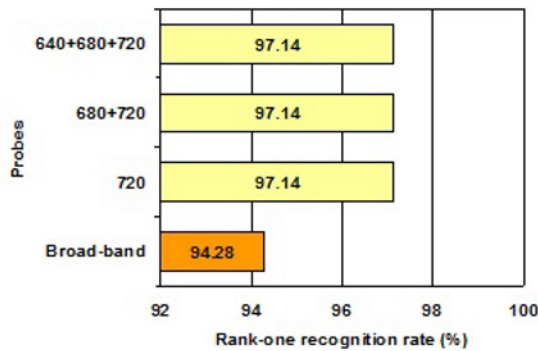
4.2. Fluorescent gallery and daylight probe

In this experiment, a more challenging lighting condition, daylight, is used for probe sets. To simulate practical

Table 2. The top three bands selected by different distance measures with four different kernels for the experiment of fluorescent gallery and halogen probe

	h_{AMISE}			
	Gaussian	Triangle	Epanechnikov	Cosine
Jeffrey	610 620 640	610 620 640	610 620 640	610 620 640
Bhattacharyya	610 630 640	610 630 640	610 630 640	610 630 640
Matusita	610 630 640	610 630 640	610 630 640	610 630 640
Patrick-Fisher	610 630 640	610 630 640	610 630 640	610 630 640
	h_{ICOMP}			
	Gaussian	Triangle	Epanechnikov	Cosine
Jeffrey	610 620 640	610 620 640	610 620 640	610 620 640
Bhattacharyya	610 630 640	610 630 640	610 630 640	610 630 640
Matusita	620 630 640	610 630 640	610 630 640	610 630 640
Patrick-Fisher	610 620 640	610 630 640	610 630 640	610 630 640

face recognition, stable indoor fluorescent light is used for gallery images while all the probes are acquired under varying daylight. The spectral range is selected among 13 sets of narrow-band spectral images from wavelength 480 nm to 720 nm with an increment of 20 nm. Identical bands (640 nm, 680 nm, and 720 nm) are selected from various implementations of the proposed algorithm. The fused images from these selected bands produce a 97.14% rank-one recognition rate, 2.86% higher than that of the broad-band images, as shown in Figure 6.

**Figure 6.** Rank-one recognition rate of different probe sets, including band 680 nm, 700 nm, 720 nm, broad-band image, and fused images from selected spectral range in the experiment of fluorescent gallery and daylight probe

4.3. Comparisons for mapping approaches

To compare the accuracy between our and the reference algorithms [18], 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 (13) based on Equation (9). Figure 7 shows the estimation error in pan values, where Figure 7(a) and 7(b) indicate the estimation error in comparison with the original sample set (825 sample) and relative pan angles ($0^\circ \sim 360^\circ$), respectively. Figure 8 shows the estimation error in tilt values, where Figure 8(a) and 8(b) indicate the estimation error in comparison with the original sample set (825 sample) and relative tilt angles ($0^\circ \sim 90^\circ$), respectively. Figure 9 shows the estimation error in zoom values, where Figure 9(a) and 9(b) indicate the estimation error in comparison with the original sample set (825 sample) and relative zoom positions ($1 \sim 184$). The estimation error is based on how many degrees the system is supposed to pan, tilt, or zoom to keep the object in the center of image. In average, the estimation error in pan angle is less than ± 6.3 . The estimation error in tilt angle is less than ± 8.5 . The estimation error in zoom value is less than ± 19.5 . For the reference algorithm, we manually calibrate two PTZ cameras to learn their intrinsic parameters first. This manual intervention impedes their direct application to surveillance systems with changing setups and larger

number of PTZ cameras in the scene:

$$\begin{aligned}
 \hat{\Theta}_{P,h} &= 155.376 - 16.612\Theta_{P,j} + 37.412\Theta_{T,j} + \\
 &\quad - 10.290x_j + 2.977y_j + 5.469\Theta_{P,j}^2 + \\
 &\quad - 23.364\Theta_{T,j}^2 + 2.067x_j^2 - 0.804\Theta_{P,j}\Theta_{T,j} + \\
 &\quad + 6.764\Theta_{P,j}x_j - 1.940\Theta_{T,j}y_j - 0.658x_jy_j, \\
 \hat{\Theta}_{T,h} &= -7.964 - 29.955\Theta_{T,j} - 6.465y_j - 0.900\Theta_{T,j}^2 + \\
 &\quad + 24.060\Theta_{T,j}^2 - 0.558y_j^2 - 1.386\Theta_{P,j}x_j + \\
 &\quad + 0.7291\Theta_{P,j}y_j - 1.940\Theta_{T,j}y_j, \\
 \hat{Z}_h &= -0.439 + 0.7324\Theta_{P,j} - 0.6218Z_j + 0.1221y_j + \\
 &\quad + 0.0817\Theta_{P,j}^2 + 0.086\Theta_{T,j}^2 + 0.5934Z_j^2 + \\
 &\quad + 0.0218x_j^2 - 0.0141y_j^2 + 0.0153\Theta_{P,j}\Theta_{T,j} + \\
 &\quad + 0.0723\Theta_{T,j}y_j + 0.0596Z_jx_j + 0.0596Z_jy_j + \\
 &\quad + 0.0125x_jy_j.
 \end{aligned} \tag{13}$$

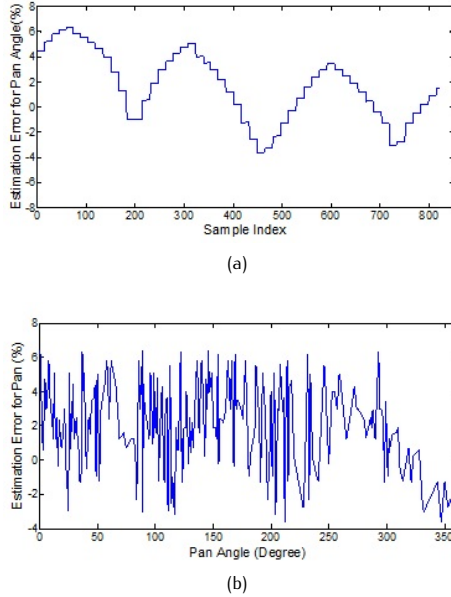


Figure 7. Estimation errors in pan values: (a) comparison to the original sample set (825 samples), (b) relative pan angle ($0^\circ \sim 360^\circ$)

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 3 illustrates the comparison between our and reference algorithms. In Table 3, 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).

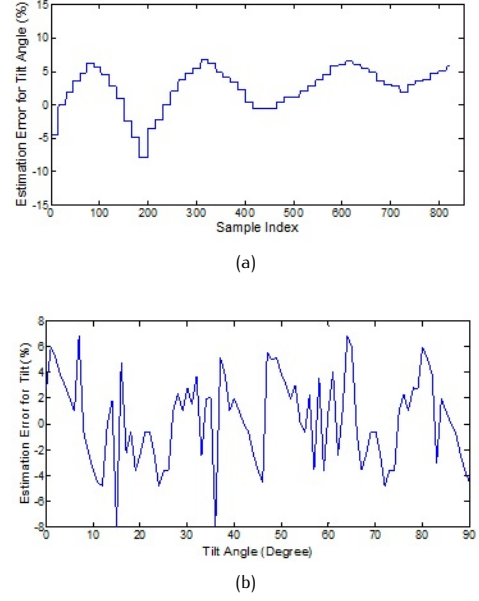


Figure 8. Estimation errors in tilt values: (a) comparison to the original sample set (825 samples), (b) relative tilt angle ($0^\circ \sim 90^\circ$)

The averaged pixel size deviation indicates the difference between the derived pixel size of the object and the desired pixel size ($50 \times 170 = 7500$ 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 10 and 11 show real-time video sequences for our proposed, and Chen and Wang's approaches. In Figures 10 and 11, the j^{th} PTZ camera uses Equation (13) 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 10 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°). Figure 11 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°). In both Figures 10 and 11, 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 pro-

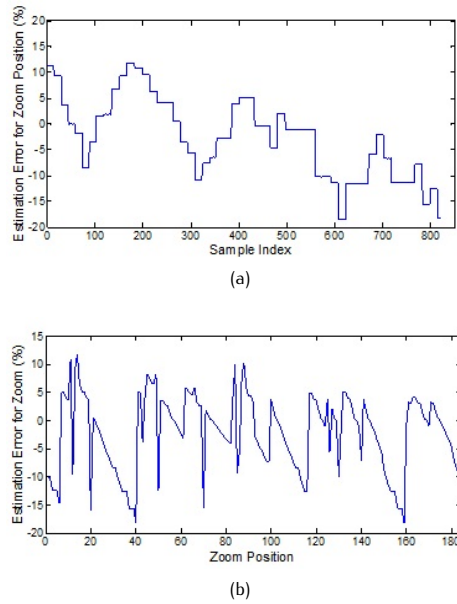


Figure 9. Estimation errors in zoom values: (a) comparison to the original sample set (825 samples), (b) relative zoom position (1 ~ 184)

posed, 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. Figure 12 illustrates how we calculate their pixel distance deviation and pixel size deviation.

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° and 8.5° 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) [26] 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.

Table 3. Comparison between our and reference algorithms

	Averaged Pixel Distance Deviation	Averaged Pixel Size Deviation
Our Method	11.1%	16.7%
Chen and Wang [18]	9.2%	15.2%

5. Conclusion

In this work, we investigated two studies: 1) using narrow-band spectral images instead of conventional broad-band images to improve recognition performance; 2) directly deriving a unified polynomial model between the pan and tilt values of PTZ cameras with unknown intrinsic parameters and system setups in the scene. We demonstrated the robustness and consistency of the automated band selection algorithm under various implementations of kernel functions, smoothing parameters, and distance measures. An improved face recognition rate over the conventional broad-band images was achieved under various illumination conditions by the fusion of images from the selected bands. The robustness of the algorithm facilitates the application of the proposed algorithm in a plug-and-play manner that is independent of the characteristics of the input data. The second proposed approach, which 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, has proven to reduce 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.

References

- [1] Phillips P.J., Moon H., Rizvi S. A., Rauss P.J., The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, 2002, 1090-1103
- [2] Kong S.G., Heo J., Boughorbel F., Zheng Y., et al.,



Figure 10. Performance of our proposed and reference methods for a real-time multiple PTZ cameras system. 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. In this experiment, 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°)

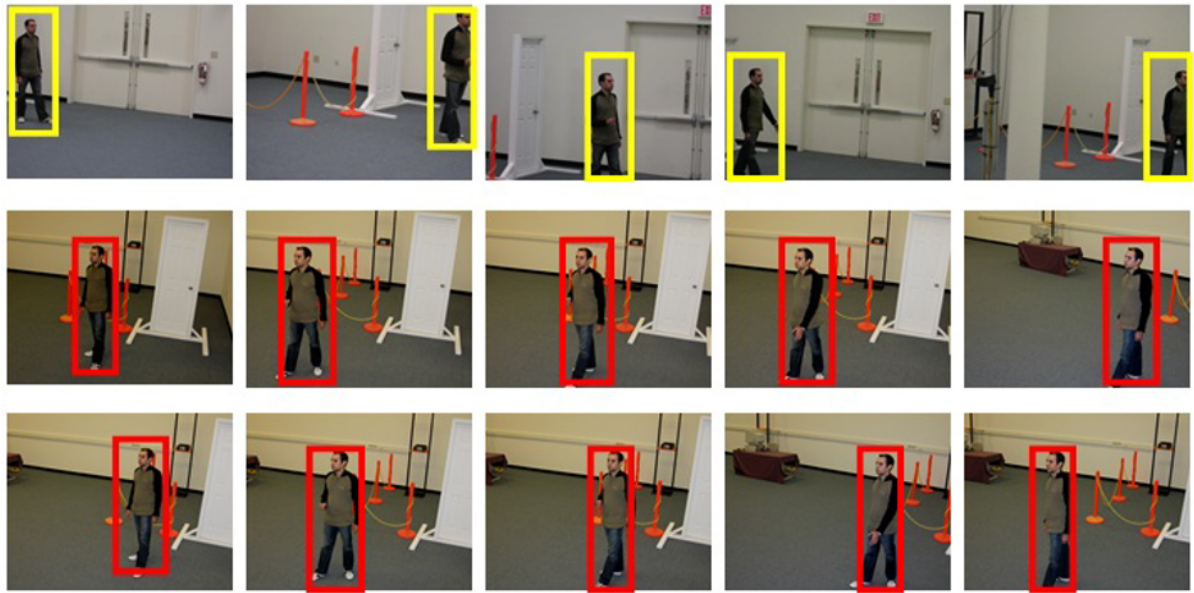


Figure 11. Performance of our proposed and reference methods for a real-time multiple PTZ cameras system. The first row shows five different locations in images of the j^{th} PTZ camera, the second row shows their respective pixel locations and sizes, derived by our approach, in images of the h^{th} 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. In this experiment, 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°)

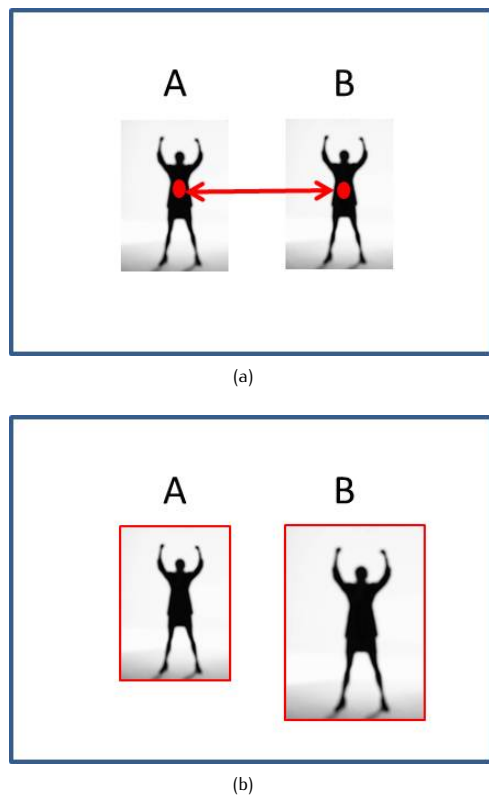


Figure 12. Illustration of deviation calculation. (a) Object A represents the desired pixel position. Object B represents the actual pixel position calculated by our proposed or Chen and Wang's method. The distance deviation is calculated by the pixel distance between their mass points. (b) Object A represents the desired size. Object B represents the actual pixel size calculated by our proposed or Chen and Wang's method. The size deviation is calculated by the pixel size between their pixel sizes

Adaptive fusion of visual and thermal IR images for illumination-invariant face recognition, *Int'l J. of Computer Vision*, vol. 71, no. 2, 2007, 215-233

- [3] Pan Z., Healey G., Prasad M., Tromberg B., Face recognition in hyperspectral images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, 2003, 1552-1560
- [4] Chang H., Koschan A., Abidi B., Abidi M., Physics-based fusion of multispectral data for improved face recognition," *IEEE Proc. Int'l Conf. on Pattern Recognition*, vol. III, 2006, 1083-1086
- [5] Chang H., Yao Y., Koschan A., Abidi B., et al., Spectral range selection for face recognition under various illumination, *IEEE Int'l Conf. on Image Processing*, San Diego, CA, Oct. 2008
- [6] Wasserman L., *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics, 2005
- [7] Mugdadi A.-R., Munthali E., Relative efficiency in kernel estimation of the distribution function, *Journal of Statistical Research*, vol. 15, no. 4, 2003, 579-605
- [8] Bozdogan H., Akaike's information criterion and recent developments in information complexity, *Journal of Mathematical Psychology*, vol. 44, 2000, 62-91
- [9] Zhou S., Chellappa R., From sample similarity to ensemble similarity: probabilistic distance measures in reproducing kernel hilbert space, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, 2006, 917-929
- [10] Bagdanov A.D., Bimbo A.D., Nunziati W., Improving Evidential Quality of Surveillance Imagery Through Active Face Tracking, *18th Int'l Conf. on Pattern Recognition*, 2006
- [11] Angella F., Reithler L., Gallesio F., Optimal Deployment of Cameras for Video Surveillance Systems, *IEEE Conf. on Advanced Video and Signal Based Surveillance*, Sept. 2007
- [12] Svoboda T., Martinec D., Pajdla T., A Convenient Multi-Camera Self-Calibration for Virtual Environments, *PRESENCE: Teleoperators and Virtual Environments*, vol. 14, 2005, 407-422
- [13] Sugimure Y., Sato J., Camera Calibration and Reconstruction from the Chain Connection of Mutual Camera Projections, *17th Int'l Conf. on Pattern Recognition*, 2004
- [14] Chen I.-H., Wang S.-J., Efficient vision-based calibration for visual surveillance systems with multiple PTZ cameras, *IEEE Int'l Conf. on Computer Vision Systems*, 2006
- [15] Li H., Shen C., An LMI Approach for Reliable PTZ Camera Self-Calibration, *IEEE Conf. on Advanced Video and Signal Based Surveillance*, Nov. 2007
- [16] Agapito L., Hayman E., Reid I., Self-Calibration of Rotating and Zooming Cameras, *Int'l Journal of Computer Vision*, vol. 45, no. 2, 2001, 107-127
- [17] Kutner M.H., Nachtsheim C.J., Neter J., *Applied Linear Regression Models*, McGraw Hill, 2004
- [18] Chen I.-H., Wang S.-J., An Efficient Approach for the Calibration of Multiple PTZ Cameras, *IEEE Trans. on Automation Science and Engineering*, vol. 4, no. 2, 2007
- [19] Everts I., Sebe N., Jones G. A., Cooperative Object Tracking with Multiple PTZ Cameras, the *14th Int'l Conf. on Image Analysis and Processing*, Sept. 2007
- [20] Johnson D., *Applied Multivariate Methods for Data Analysis*, Duxbury, 1998
- [21] Kannala J., Brandt S., A generic camera calibration method for fish-eye lenses, the *17th Int'l Conf. on Pattern Recognition*, Cambridge, UK, Aug. 2004
- [22] Fayman J. F., Sudarsky O., Rivlin E., Rudzsky M.,

- Zoom tracking and its applications, *Machine Vision and Applications*, vol. 13, 2001, 25-37
- [23] Bearse P.M., Bozdogan H., Subset selection in vector autoregressive (VAR) models using the genetic algorithm with informational complexity as the fitness function, *Systems Analysis Modeling Simulation*, 1998, 61-91, 1998
- [24] Chen C.-H., Yao Y., Page D., Abidi B.R., et al., Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, No. 8, August 2008, 1052-1063
- [25] Cui Y., Samarasekera S., Huang Q., Greiffenhagen M., Indoor monitoring via the collaboration between a peripheral sensor and a foveal sensor, in *Proc. IEEE Workshop Visual Surveillance*, Mumbai, India, Jan. 1998
- [26] Scotti G., Marcenaro L., Coelho C., Selvaggi F., et al., Dual camera intelligent sensor for high definition 360 degrees surveillance, *IEE Proc.-Vis., Image, Signal Process.*, vol. 152, no. 2, Apr. 2005, 250-257
- [27] Lowe D.G., Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol. 60, no.2, 2004, 91-110
- [28] Li S., Jain A., *Handbook of Face Recognition*. New York: Springer, 2004