Short Papers

Robust Environmental Change Detection Using PTZ Camera via Spatial-Temporal Probabilistic Modeling

Jwu-Sheng Hu, and Tzung-Min Su

***Abstract*—This paper proposes a novel procedure for detecting environmental changes by using a pan–tilt–zoom (PTZ) camera. Conventional approaches based on pixel space and stationary cameras need time-consuming image registration to yield pixel statistics. This work proposes an alternative approach to describe each scene with a Gaussian mixture model (GMM) via a spatial–temporal statistical method. Although details of the environment covered by the camera are lost, this model is efficient and robust in recognizing scene and detecting scene changes in the environment. Moreover, the threshold selection for separating different environmental changes is convenient by using the proposed framework. The effectiveness of the proposed method is demonstrated experimentally in an office environment.**

***Index Terms*—Gaussian distributions, machine vision, pattern recognition, surveillance.**

I. Introduction

Detecting environmental changes is crucial in applications such as video surveillance, monitoring, and robot navigation [1], [2]. Static cameras generally cannot be used to capture wide areas of background owing to the limited view angle. The pan–tilt–zoom (PTZ) camera and the omnidirectional camera are two tools used in wide-area surveillance. Despite having a 360° view angle, the omnidirectional camera suffers from image distortion and unevenly distributed resolution. Meanwhile, the PTZ camera suffers from changes in image coordinates when the camera is in motion. Therefore, a precision control mechanism is commonly required to ensure accurate matching of the image coordinates. Otherwise, a time-consuming image recovery or registration procedure must be performed before image processing or recognition, for example, to detect environmental changes in surveillance applications.

A similarity measure between the background model and the test image defines change in the detection of environmental changes. Background modeling includes deterministic methods and statistical methods. However, deterministic methods such as time averaging [3] have been found to have limited effectiveness. Moreover, statistical approaches, including the Gaussian mixture model [4]–[6] and kernel density estimation [7], [8] have been applied for background modeling, considering the effect of lighting and variations in background objects. However, most of these approaches are based on pixel space and stationary cameras. While a wide scene is captured using a PTZ camera or multiple cameras, the space and time-consuming registration of an image [9], [10] from each camera view must be performed to yield pixel statistics. However, matching between the captured image and the background model is difficult because errors accumulate from the PTZ mechanism or image registration. Furthermore, a Gaussian model [11] of spatial distribution is presented to detect moving objects using nonstationary cameras, even with approximately accurate motion compensation, noise, or environmental change. However, it still bases on pixel-statistics, and an approximate alignment is necessary. Moreover, Eric *et al.* [12] assume that the calibration parameters are known reasonably accurately, and then, extend Stauffer and Grimson's technique [5] to a pan/tilt head by incorporating a model for solving motion blur, mixed pixels, and small camera translations, etc.

Instead of using pixel-based statistics, this work proposes an alternative description of the background model using Gaussian mixture model (GMM), called the blob model, which robustly detects changes in the scene without image registration, using low-precision PTZ mechanisms. The proposed method transforms the background image from pixel space to feature space in order to reduce the storage requirement, and transforms the scene into a conceptual description without details. Although the details of the scene are lost in the blob model, the concepts of the scene remains apparent. Furthermore, the measured similarity between two blob models is adopted to calculate the differences between them. The differences between blob models describe the extent of change in the scene, which is important in some surveillance systems, such as manual video monitoring systems. When the scene changes exceed a predefined threshold, an alarm is sounded automatically to notify the security guard for paying attention to the video monitoring system. Additionally, video clips that show potentially dangerous situations can be extracted from the video sequences using the proposed method by assigning levels to the scene changes. Fig. 1 presents the basic workflow of the proposed novel framework, where $T1$ denotes the number of Gaussian distributions of GMM that suffices to select a suitable background model, as described in Section III.

The remainder of this paper is organized as follows. Section II describes the features and the statistical learning method. Section III introduces the spatial and temporal model. Section IV presents the experimental results to demonstrate the performance of the proposed method. Conclusions are finally drawn in Section V.

II. Probabilistic Model in the Spatial Domain

Assume that $\boldsymbol{x}_m = [R_{i,j}, G_{i,j}, B_{i,j}, i, j]$ is defined as a five-dimensional (5-D) feature vector associated with the m_{th} pixel that has color information $(R_{i,j}, G_{i,j}, B_{i,j})$ at position (i, j). GMM is then applied to model the background information using a training feature set $\boldsymbol{X} = \{\boldsymbol{x}_m, 1 \leq m \leq M{=}h \times w\}$, where M is the number of pixels, h is the image height, and w is the image width. The procedure from a test image to a feature plane is listed in Fig. 2. Suppose the GMM contains N Gaussian distributions and is estimated using (1), where λ is the GMM parameter set, $\boldsymbol{w}_k = (w_R, w_G, w_B, w_i, w_j)$ is the weight vector, $\boldsymbol{\mu}_k = (\boldsymbol{\mu}_R, \boldsymbol{\mu}_G, \boldsymbol{\mu}_B, \boldsymbol{\mu}_i, \boldsymbol{\mu}_j)$ is the mean vector, and $\sum_{\mathrm{k}} = \mathrm{diag}(\sigma_R^2, \sigma_G^2, \sigma_B^2, \sigma_i^2, \sigma_j^2)$ is the covariance matrix

$$f(\boldsymbol{x}|\lambda){=}\sum_{k=1}^{N}\boldsymbol{w}_k \frac{1}{\sqrt{(2\pi)^d|\sum_k|}} \exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \sum_{k}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)) \tag{1}$$

$$\text{where } \lambda = \{\boldsymbol{w}_{\mathrm{k}}, \boldsymbol{\mu}_{\mathrm{k}}, \sum_{\mathrm{k}}\}, k = 1, 2, \ldots, N and \sum_{k=1}^{\mathrm{N}} ||\boldsymbol{w}_k|| = 1.$$

Manuscript received March 1, 2006; revised December 15, 2006. Recommended by Guest Editors H.-P. Huang and F.-T. Cheng. This work was supported in part by the National Science Council of the Republic of China under Grant NSC94-2218-E009064 and in part by the DOIT TDPA Program under Project 95-EC-17-A-04-S1-054.

The authors are with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: jshu@cn.nctu.edu.tw; tzungminsu@ieee.org).

Digital Object Identifier 10.1109/TMECH.2007.897280

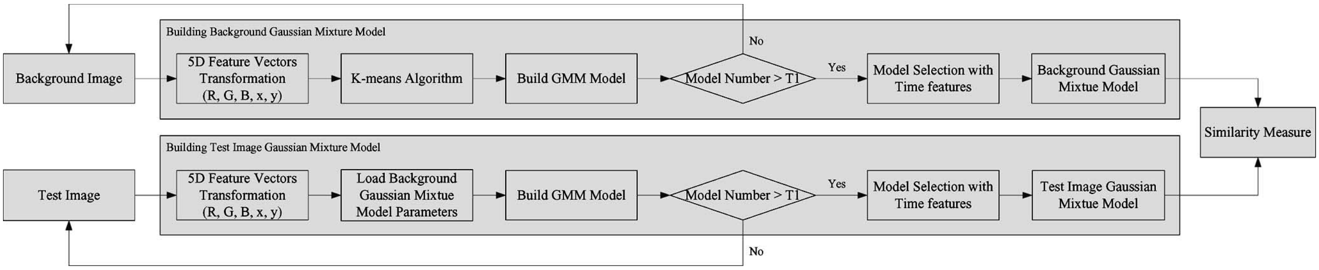

Building Background Gaussian Mixture Model
Background Image
5D Feature Vectors Transformation (R, G, B, x, y)
K-means Algorithm
Build GMM Model
Model Number > TI
No
Yes
Model Selection with Time features
Background Gaussian Mixtue Model
Similarity Measure
Building Test Image Gaussian Mixture Model
Test Image
5D Feature Vectors Transformation (R, G, B, x, y)
Load Background Gaussian Mixtue Model Parameters
Build GMM Model
Model Number > TI
Yes
Model Selection with Time features
Test Image Gaussian Mixtue Model
No

Fig. 1. Basic workflow of proposed novel framework.

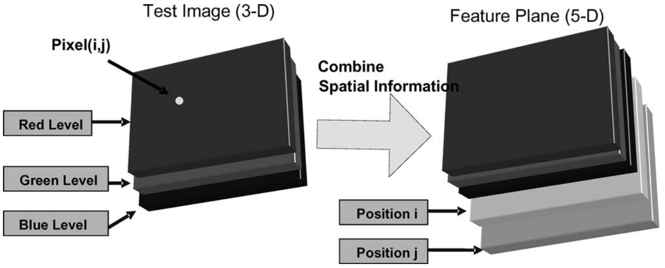
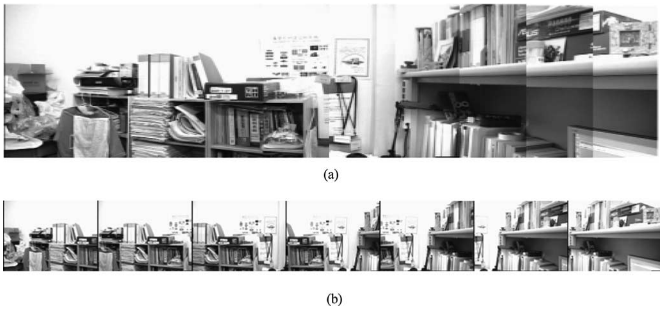


Fig. 2. Procedure from a test image to a feature plane, where the feature plane is composed of the color information $(R_{i,j}, G_{i,j}, B_{i,j})$ and the position information (i, j).


(a)

(b)

Fig. 3. Panoramic view of wide scene produced manually from seven single views. (a) Panoramic view. (b) Seven partitions.

The expectation maximization (EM) algorithm [13] is an extensively adopted method for estimating λ using the expectation step and the maximum step iteratively, which are described by (2) and (3).

1) *Expectation step:*

$$\boldsymbol{\beta}_{m,n} = \frac{\boldsymbol{w}_n f(\boldsymbol{x}_m|\boldsymbol{\mu}_n, \sum_n)}{\sum_{k=1}^{N} \boldsymbol{w}_k f(\boldsymbol{x}_m|\boldsymbol{\mu}_k, \sum_k)}, \quad 1 \leq \mathrm{n} \leq N, 1 \leq m \leq M \tag{2}$$

where $\beta_{m,n}$ is the *a posteriori* probability that the feature vector $\boldsymbol{x}_m$ belongs to the n_{th} Gaussian distribution and M is the number of feature vectors.

2) *Maximum step:*

$$\hat{\boldsymbol{w}}_n = \frac{1}{N} \sum_{m=1}^{M} \boldsymbol{\beta}_{mn}$$

$$\hat{\boldsymbol{\mu}}_n = \sum_{m=1}^{M} \boldsymbol{\beta}_{mn} \boldsymbol{x}_m \Big/ \sum_{m=1}^{M} \boldsymbol{\beta}_{mn}$$

$$\hat{\sum_n} = \sum_{m=1}^{M} \boldsymbol{\beta}_{mn} (\boldsymbol{x}_m - \hat{\boldsymbol{\mu}}_n)(\boldsymbol{x}_m - \hat{\boldsymbol{\mu}}_n)^T \Big/ \sum_{m=1}^{M} \boldsymbol{\beta}_{mn}. \tag{3}$$

The termination criteria of the EM algorithm are listed next, where T_2 and T_3 are two predefined values.

1) The increment between the new and the last log-likelihood value is below a minimum increment threshold T_2.
2) The number of iterations exceeds a maximum iterative count threshold T_3.

For the acceleration of the EM convergence, GMM parameters are initialized by applying the K-means algorithm [14] to the training feature set $\boldsymbol{X}$.

III. Modeling and Matching in the Spatial-Temporal Domain

Although the K-means algorithm is a simple clustering method, and can improve the convergence speed of the EM algorithm, the intuitive selection of initial parameters in the K-means algorithm, such as cluster numbers and the initial center of each cluster, is responsible for the uncertainty of background model representation. To cope with this, a method that combines Jeffrey divergence [15] and temporal feature (time) is proposed in this work.

A. *Measuring Similarity Using Jeffrey Divergence*

Jeffrey divergence is a modification of the Kullback–Leibler distance that is numerically stable, symmetric, and robust against noise and histogram bin size, and is regarded as a similarity measure of two statistical models from an information-theoretic perspective. Suppose f_0 and f_1 are defined as two GMMs; the Jeffrey divergence between f_0 and f_1 is defined as

$$D(f_1||f_0) = \sum_{k=1}^{M} (f_1(\boldsymbol{x}_k) \log(f_1(\boldsymbol{x}_k)/a) + f_0(\boldsymbol{x}_k) \log(f_0(\boldsymbol{x}_k)/a)) \tag{4}$$

where $a = (f_0(\boldsymbol{x}_k) + f_1(\boldsymbol{x}_k))/2$ and $\log(\cdot)$ refers to the two-based logarithm.

B. *Model Selection with Temporal Features*

Suppose $\boldsymbol{I} = \{I_t, 1 \leq t \leq T_1\}$ is defined as a set of captured frames of a period T_1, and I_t is the frame captured at time t. Moreover, the corresponding GMM set of $\boldsymbol{I}$ is defined as $\boldsymbol{F} = \{f^t, 1 \leq t \leq T_1\}$, where f^t denotes the GMM of the captured frame I_t. The selection of T_1 depends on the robustness of the model representation and the computing time. If T_1 is defined as a small value, the robustness of the model representation decreases and the computing time also decreases.

TABLE I
ROBUSTNESS TEST OF BACKGROUND MODELS

Partitions	Recognition Rate			
	$\mathbf{E}_1^{\mathrm{p}}$ (%)	$\mathbf{E}_2^{\mathrm{p}}$ (%)	$\mathbf{E}_3^{\mathrm{p}}$ (%)	Avg. (%)
-30^o	94	100	94	96
-20 ^{o}	97	93	90	93.33
-10^o	100	100	100	100
0^o	100	100	100	100
10^o	100	100	99	99.67
20^o	96	91	98	95
30^o	97	100	100	99
Avg.(%)	97.71	97.71	97.29	---

TABLE II
COMPARISON OF THREE METHODS TO THE PROPOSED METHOD (GMM)

Method	C_1^{p}	C_2^{p}	C_3^{p}	C_4^{p}	C_5^{p}	C_6^{p}	C_7^{p}	C_8^{p}
IM_DIF	97.8	97.7	96.3	95.9	93.5	74.1	---	---
PCA	97.3	95.3	93.3	89.3	84.2	61.3	---	---
EX_HI	97.7	97.7	97.7	96	95	79.1	90.5	87.1
GMM	97.7	97.7	97	96.7	96	83.2	97.3	96.7

In the selection of the suitable GMM $f^{\hat{t}}$ as the background model, a correlation matrix $\boldsymbol{S}$ is then, defined by (4) using $\boldsymbol{F}$, and is described as

$$\boldsymbol{S} = D(f^{t_1}||f^{t_2}), \qquad 1 \leq t_1 \leq T_1, 1 \leq t_2 \leq T_1. \tag{5}$$

After $\boldsymbol{S}$ is determined, the sum of the rows in $\boldsymbol{S}$ is defined as $\boldsymbol{S}_r$ which is given by (6), and then, the frame $\hat{t}$ that has the minimum $S_r(\hat{t})$ is selected to build up the suitable background model. The suitable GMM $f^{\hat{t}}$ for detecting the environmental change is estimated using (7), where $\hat{t}$ represents the index of GMM in $\boldsymbol{F}$

$$\boldsymbol{S}_r = \sum_{t_2=1}^{T_1} D(f^{t_1}||f^{t_2}), \qquad 1 \leq t_1 \leq T_1 \tag{6}$$

$$\hat{t} = \arg \min_{\forall \text{all}\, t_1} \boldsymbol{S}_r. \tag{7}$$

IV. EXPERIMENTAL RESULTS

This section describes various experiments that demonstrate the effectiveness of the proposed method using real image sequences acquired using a PTZ camera. For demonstration, only the results concerning the pan motion of the camera are presented. Moreover, an 800×240 wide scene is divided into seven 320×240 partitions by rotating the camera from $-30°$ to $30°$ in $10°$ increments to explain the proposed method. Fig. 3 shows the panoramic view of the wide scene and the seven partitions. Additionally, suppose the background image of each partition is modeled using five Gaussian distributions ($N = 5$), and an appropriate background model is selected with nine image frames ($T_1 = 9$). The threshold values in the termination criteria of the EM algorithm are defined as $T_2 = 10^{-6}$ and $T_3 = 50$. However, the proposed method can be applied to combine more elaborate motion of the PTZ camera and the other selections of N and T_1. The robustness of the background models is demonstrated, and the models are then adopted to detect environmental changes. The experiments have three stages, as described next.

A. Robustness of the Proposed Background Model

Suppose $\boldsymbol{E}_{\mathrm{c}}^{\mathrm{p}} = \{I_{c,p}^t, 1 \leq c \leq 3, 1 \leq p \leq 7, 1 \leq t \leq 100\}$ are defined as three sets that contain 100 images in each partition. In $\boldsymbol{E}_1^{\mathrm{p}}$, test images are captured from the same direction as training images of each partition. In $\boldsymbol{E}_2^{\mathrm{p}}$, test images are captured from different direction, which is $2°$ away from the direction of training images of each partition. In $\boldsymbol{E}_3^{\mathrm{p}}$, test images that differ by 5% from the original scene in the image content are captured by placing objects into the scene. However, some errors exist in the directions defined in $\boldsymbol{E}_{\mathrm{c}}^{\mathrm{p}}$ according to the errors of the control mechanism of the PTZ camera. The similarity between a test image and each partition is calculated using each background model and the GMM of the test image via Jeffrey divergence. The partition that has the smallest Jeffrey divergence is regarded as the partition to which the test image belongs. The recognition results of the $\boldsymbol{E}_1^{\mathrm{p}}, \boldsymbol{E}_2^{\mathrm{p}}$, and $\boldsymbol{E}_3^{\mathrm{p}}$ are listed in Table I. The scene can be recognized correctly via GMM with high recognition rates. However, there exists an issue about the uncertainty of the K-means algorithm [5] resulting from the intuitive selection of the cluster center. In order to solve the problem, a method for selecting a suitable model using the Jeffrey divergence and temporal features is proposed in this work, and the results are listed in the next experiments.

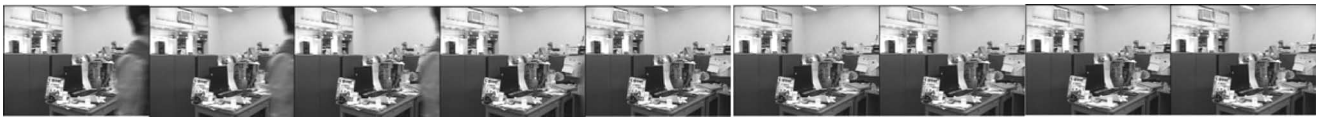
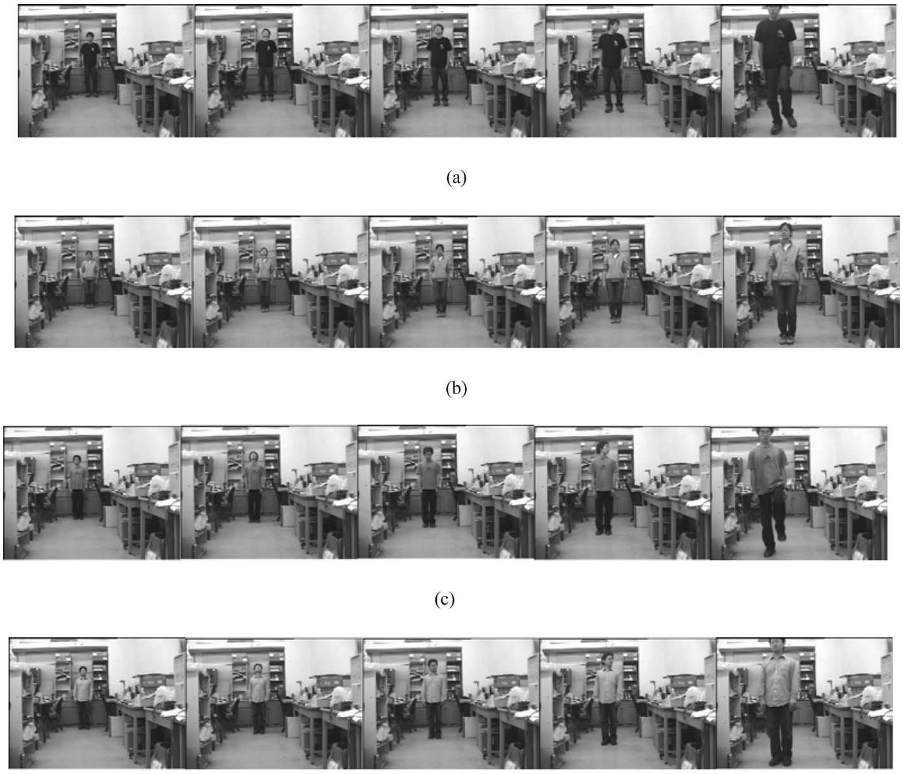


Fig. 4. Image sequence used for testing the proposed method for selecting a suitable background model with temporal features.

TABLE III
MODEL SELECTION WITH TEMPORAL FEATURES

Index	The results of each frame								
I_i	1	2	3	4	5	6	7	8	9
$S_r^{'}$	13.2	17.6	9.44	5.04	5.00	5.04	5.01	4.92	4.97
$I_{\hat{i}}$	8	9	7	6	3	5	4	1	2



(a)

(b)

(c)

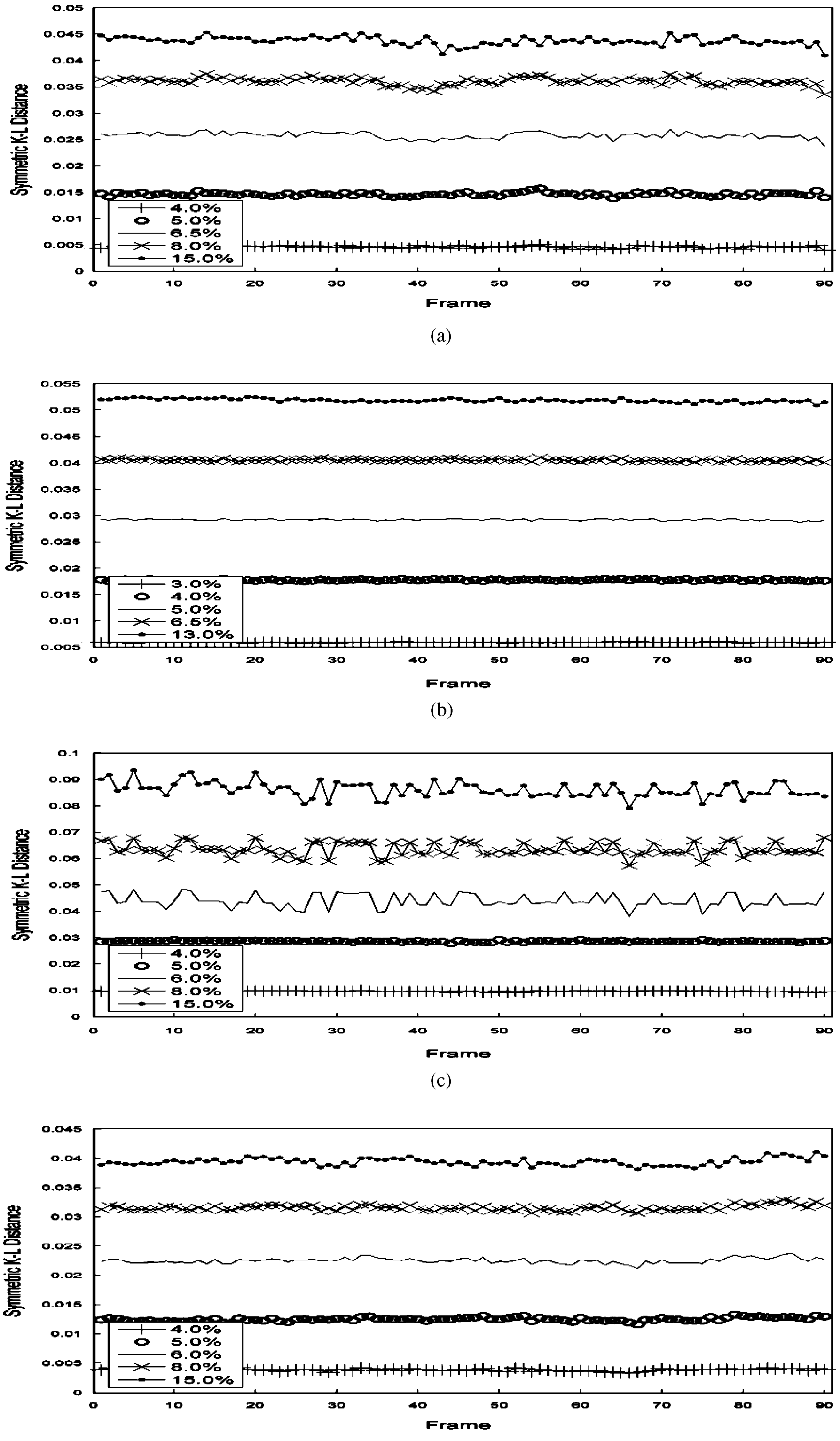
(d)

Fig. 5. Four intruders with five intrusion covering ratios on the same background image in an indoor environment.

Image differences [IM_DIF, pixel-based image representation (IR)], extended histograms [16] (EX_HI, histogram-based IR), and principal component analysis [17] (PCA, feature-based IR) are compared with the proposed method (GMM, feature-based IR) to determine the efficiency of the proposed method. Moreover, the similarity measure is calculated using Jeffrey divergence. Suppose $C_c^{\mathrm{p}} = \{\hat{I}_{c,p}^t, 1 \leq c \leq 8, 1 \leq p \leq 7, 1 \leq t \leq 100\}$ are defined as eight image sets that contain 100 images in each partition. C_1^{p} is the set of test images that have no error in any partition, $C_2^{\mathrm{p}}, C_3^{\mathrm{p}}, C_4^{\mathrm{p}}, C_5^{\mathrm{p}}$, and C_6^{p} are the sets of test images with one-degree, two-degree, three-degree, four-degree and five-degree position errors from each partition, where the position error means the difference between the new direction for capturing test images and the original direction for capturing training images. C_7^{p} and C_8^{p} are the sets of test images that differ by 5% and 10% from those of the original scene. According to the results listed in Table II, the expanded histogram and GMM are both robust against position errors. However, GMM is more robust than the expanded histogram against noise variations according to C_7^{p} and C_8^{p} in Table II.

B. *Efficiency of Model Selection*

A method for selecting a suitable model using the Jeffrey divergence and temporal features is proposed to deal with the uncertainty of the K-means algorithm and variations in the indoor environment. In Fig. 4, the image sequences used to establish the background model contain a man in the first four frames. The representative background model can



(a)

(b)

(c)

(d)

Fig. 6. Distributions of Jeffrey divergence with five intrusion covering ratios.

be selected using the proposed correlation matrix $\boldsymbol{S}$ described by (5)–(7), and the results are listed in Table III, where I_i is the image index of the image sequence; $\boldsymbol{S}'_r$ is defined as $S'_r = 100 \times S_r$ for the display, and $I_{\hat{t}}$ is the priority order after sorting the frames by $\boldsymbol{S}'_r$. According to $I_{\hat{t}}$ listed in Table III, the first four images have a lower priority than the last four images for being the frame to build the suitable GMM, and the eighth image is selected as the frame for training the suitable background model.

C. Threshold Selection of Different Covering Ratios

One technical issue with the proposed method is that the background model depends on the image. Consequently, the threshold values vary from case to case. This paper finds that a relationship exists between the model and the complexity of the scene. That is, a mathematical measure of complexity can be applied to access the model parameters such as the number of Gaussian distributions. This work employs the minimum description length [18] to optimize the number of the Gaussian distributions N in GMM. However, it is found that a low value of N, about four to six, is sufficient to represent the scene in each view captured by the camera in the experiments. The problem of developing a systematic method for selecting N for different images remains unsolved. However, N can be determined if the background scene is assigned in advance. The threshold value is directly related to the sensitivity of the surveillance system. The stability of this value with respect to changes in the image and whether different intrusion covering ratios have distinct threshold values are of primary concern. In this experiment, real image sequences with different intrusion covering ratios are captured in the indoor environment and tested using Jeffrey divergence. If the corresponding background model of the present view angle is known, automatic threshold selection can be performed despite the complexity of the background image and the texture of intruders.

Two classes of image sequences are utilized to determine the effectiveness of threshold selection by the proposed method. Fig. 5 shows four intruders with five intrusion covering ratios on the same background image. Although the covering ratios vary only slightly, the threshold values for seperating different covering levels are easily determined with a 99% confidence level (three-standard deviations), and the results are presented in Fig. 6.

V. Conclusion

This paper has presented an abstract model of background images to represent the environment captured by a PTZ camera. Although details of each scene captured by the PTZ camera are lost, the model efficiently detects environmental changes, as in an intrusion alert in a surveillance system. For a PTZ camera, the proposed method eliminates the image registration procedure in pixel-based modeling and the issue of positioning errors in the PTZ mechanism. A threshold for determining the environmental changes can be estimated efficiently to improve the sensitivity of the surveillance systems, e.g., the manual video monitoring system. Additionally, the processing time for calculating the similarity between two GMMs is about 0.5 s, based on P4 2.8G CPU and 512M RAM. The computational requirements of the proposed method remain high. Increasing the efficiency of the proposed method is an issue for future study.

References

[1] B. J. A. Krose, N. Vlassis, R. Bunschoten, and Y. Motomura, "A probabilistic model for appearance-based robot localization," *Image Vis. Comput.*, vol. 19, no. 6, pp. 381–391, Apr. 2001.

[2] G. N. Desouza and A. C. Kak, "Vision for mobile robot navigation: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 237–267, Feb. 2002.

[3] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in *Proc. 13th Conf. Uncertainty Artif. Intell.*, Providence, RI, Aug. 1997, pp. 175–181.

[4] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. CVPR*, Ft. Collins, CO, Jun. 1999, vol. 2, pp. 246–252.

[6] P. K. T. Pong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Proc. 2nd Eur. Workshop Adv. Video Based Surveill. Syst. (AVBS01)*, Kingston, U.K., Sep. 2001, pp. 149–158.

[7] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.

[8] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. 6th Eur. Conf. Comput. Vis.*, Dublin, Ireland, Jun./Jul. 2000, pp. 751–767.

[9] A. Mittal and D. Huttenlocher, "Scene modeling for wide area surveillance and image synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Hilton Head Island, SC, Jun. 2000, vol. 2, pp. 160–167.

[10] J. Kang, I. Cohen, and G. Medioni, "Continuous tracking within and across camera streams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Madison, WI, Jun. 2003, vol. 1, pp. 267–272.

[11] Y. Ren, C. S. Chua, and Y. K. Ho, "Motion detection with non-stationary background," *Mach. Vis. Appl.*, vol. 13, pp. 332–343, 2003.

[12] E. Hayman and J. O. Eklundh, "Statistical background subtraction for a mobile observer," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nice, France, Oct. 2003, vol. 1, pp. 67–74.

[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[14] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, Berkeley, CA, 1967, vol. 1, pp. 281–297.

[15] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.

[16] L. J. Latecki, "Image similarity measures for video analysis," in *Proc. IS&T/SPIE Conf. Internet Imag. IV*, Santa Clara, CA, Jan. 2003, vol. 5018, pp. 219–227.

[17] B. J. A. Krose, N. Vlassis, R. Bunschoten, and Y. Motomura, "A probabilistic model for appearance-based robot localization," *Image Vis. Comput.*, vol. 19, pp. 381–391, Apr. 2001.

[18] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Statist.*, vol. 11, no. 2, pp. 417–431, 1983.