Motion Detection from Time-varied Background

Ying REN, Chin-Seng CHUA, Yeong-Khing HO, Ying CHEN
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore, 639798
E-mail: {p144775376,ecschua,eykho,eychen}@ntu.edu.sg

Abstract

This paper proposes a new background subtraction method for detecting moving objects from a time-varied background. While background subtraction has traditionally worked well for stationary backgrounds; for a non-stationary viewing sensor, motion compensation can be applied but is difficult to realize to sufficient pixel accuracy in practice. The problem is further compounded when the moving objects are small, since the pixel error in motion compensating the background will subsume the small targets. A Spatial Distribution of Gaussians (SDG) model is proposed to deal with moving object detection having motion compensation approximated. The distribution of each background pixel is temporally and spatially modeled. Based on this statistical model, a pixel in the current frame is classified as belonging to the foreground or background. For this system to perform under lighting and environmental changes over an extended period of time, the background distribution must be updated with each incoming frame. A new background restoration and adaptation algorithm is developed for the time-varied background. Test cases involving the detection of small moving objects within a highly textured background and a pan-tilt tracking system based on 2D background mosaic are demonstrated successfully.

1 Introduction

Motion detection and segmentation is a basic problem in computer vision. It is a prerequisite for security surveillance, object tracking, motion analysis and image compression. Background subtraction is a traditional technique for finding moving objects in a sequence of images [12, 5, 11, 4]. This approach provides a more complete set of feature data describing the moving targets when compared with other motion detection approaches [6]. Conditionally, the background scene and the viewing sensor are required to be stationary when background subtraction is applied.

Recently, motion detection with a non-stationary viewing sensor has attracted the attention of several research groups [3, 7, 1, 10]. The applications include vehicle-borne or airborne video surveillance, object detection and tracking with a pan-tilt camera and others. In these cases, background subtraction cannot be applied directly. Motion compensation is required first to compensate for the motion due to the moving sensor. Usually a motion model of the background is assumed and motion parameters are estimated. Then the background is registered ideally and the foreground can be detected pixel by pixel. The underlying assumptions are that the motion model must be sufficiently accurate, the parameters of the motion model are accurately estimated and the sensing lenses are, more or less, distortion-free. In practice, these assumptions are difficult to realize. Usually, distortion correction, registration refinement and accurate 3D registration are required. These are time-consuming and not suitable for real-time applications. With the approximate estimation of the motion model, the background image and the current image cannot warp and register well. Accurate inter-frame registration of images from a moving sensor is not trivial and false detection cannot be avoided. The problem is further compounded when the moving target to be detected/tracked is small and within a textured background. The target will be subsumed by the pixel error in motion compensation.

In this paper, we propose a Spatial Distribution of Gaussians (SDG) model, which is a temporal and spatial description of the background. The foreground detection is carried out based on the SDG model. The proposed approach is robust even with approximate motion compensation, noise or environmental changes and is able to detect small moving objects in a highly textured background. In Section 2, we introduce the SDG model and the technique of background restoration, adaptation and object detection. Section 3 presents experimental results of background restoration, adaptation and foreground detection with moving sensor based on the SDG model. Finally, conclusions are drawn in Section 4.

2 Spatial Distribution of Gaussians (SDG) model and foreground detection

The basic idea of this approach is that we can model the distribution of the intensity value of each background pixel as a Gaussian (or Mixture of Gaussians (MOG)). Assume that the dominant motion between the current frame and the background is the motion due to the moving sensor and can be approximated by a 2D parametric transformation in the image plane. Traditional approaches are used to estimate the transformation parameters after which the current image is warped to align with the background. For a pixel in the current frame, after compensating for the sensor motion, it should belong to one of the background Gaussians in its local spatial region if it is indeed the background; otherwise, it is regarded as the foreground. Not only the temporal visual information, but also the local spatial information of the pixel, is taken into consideration when deciding the foreground.

2.1 Pixel-wise background model

In a sequence of images with size $M \times N$, each pixel $\{\mathbf{x}_i, i = 1, 2, \ldots, M \times N\}$ is modeled as an independent statistical process. The distribution of the intensity value is fitted with multiple Gaussians which compose the MOG model. Each Gaussian may correspond to the distribution of background or individual moving objects covering this pixel over time. The pdf is given as

$$p(I) = p(I|B)P(B) + \sum_{j=1}^{c-1} p(I|\omega_j)P(\omega_j) \tag{1}$$

where B stands for the background, ω_j denotes an intensity class of the moving objects and c, the number of Gaussians for that pixel. For clarity of discussion, we assume that the intensity distribution for each pixel may be modeled as a two-component MOG, that is, the Gaussians of background B and moving objects T. The background distribution should be a narrow Gaussian and the distribution of moving objects, a flat Gaussian. Then Eq. (1) can be modified as $p(I) = p(I|B)P(B) + p(I|T)P(T)$. After the learning stage, the parameters of the distributions of $p(I|B)$, $p(I|T)$ can be obtained.

The background Gaussian distributions of every pixel compose a Background Map, which is adapted frame by frame with each new incoming frame. A Background Image $\mathcal{I}_b$ is extracted by calculating the mean of the background distribution in the Background Map and utilized as the reference frame when performing the motion compensation.

2.2 Background motion compensation

Due to the errors of feature localization, motion model assumption, motion parameter estimation, lens distortion and others, the motion compensation cannot be accurate enough to make a dense registration from the current image $\mathcal{I}_c$ to the Background Image $\mathcal{I}_b$. The position after motion compensation is, at best, a predicted position in the Background Map. For a pixel $\mathbf{x}_c$ in the current image, let the predicted position in the Background Map be $\hat{\mathbf{x}}_b$. $s\tilde{\hat{\mathbf{x}}}_b = \Gamma\tilde{\mathbf{x}}_c$, where $\tilde{\hat{\mathbf{x}}}_b = [\hat{x}_b, \hat{y}_b, 1]^T$ and $\tilde{\mathbf{x}}_c = [x_c, y_c, 1]^T$ are homogeneous coordinates, and Γ is the transformation matrix for background motion compensation.

To estimate the transformation matrix Γ, corners are selected as features due to their positional accuracy and low computational cost. Corners are extracted from $\mathcal{I}_b$ and $\mathcal{I}_c$ in a coarse-to-fine structure. The best l corresponding corner pairs are selected into a set $\{\mathcal{C}_1, \mathcal{C}_2\}$, $\mathcal{C}_1 = \{\mathbf{c}_{1i}, i = 1, 2, \ldots, l\}$, $\mathcal{C}_2 = \{\mathbf{c}_{2i}, i = 1, 2, \ldots, l\}$ and $\mathbf{c}_{1i}$ and $\mathbf{c}_{2i}$ are corner positions in the two images respectively. Least Square Estimation (LSE) Method is used to estimate the transformation matrix Γ according to the assumed transformation model, which is usually affine or projective.

2.3 SDG model

For a position $\mathbf{x}_c$ in the current image, after the compensation for the background motion, its predicted position in the Background Map is $\hat{\mathbf{x}}_b$. The corresponding position of $\mathbf{x}_c$ in the Background Map is assumed to be Gaussian distributed about $\hat{\mathbf{x}}_b$ and is expressed as

$$p(\mathbf{x}_b|\hat{\mathbf{x}}_b) = \frac{1}{2\pi|\mathbf{R}|^{\frac{1}{2}}}\exp(-\frac{1}{2}(\mathbf{x}_b - \hat{\mathbf{x}}_b)^T\mathbf{R}^{-1}(\mathbf{x}_b - \hat{\mathbf{x}}_b)) \tag{2}$$

where $\mathbf{R}$ is the covariance matrix of positional errors. There is a validation region $\mathcal{A}_{\hat{\mathbf{x}}_b}$

$$\mathcal{A}_{\hat{\mathbf{x}}_b} \triangleq \{\mathbf{x}_b : D_{\mathbf{x}_b,\hat{\mathbf{x}}_b} \leq \gamma\} \tag{3}$$

where $D_{\mathbf{x}_b,\hat{\mathbf{x}}_b} = (\mathbf{x}_b - \hat{\mathbf{x}}_b)^T\mathbf{R}^{-1}(\mathbf{x}_b - \hat{\mathbf{x}}_b)$ is the Mahalanobis distance from a random position $\mathbf{x}_b$ in the Background Map to the predicted position $\hat{\mathbf{x}}_b$. $D_{\mathbf{x}_b,\hat{\mathbf{x}}_b}$ is χ^2 distributed. The real corresponding position of $\mathbf{x}_c$ will be found in this region with a certain probability decided by γ.

As described in Section 2.1, for a certain position $\mathbf{x}_b$, the intensity distribution of that pixel is modeled as a MOG, namely, the Gaussians of the background $B_{\mathbf{x}_b}$ and the targets $T_{\mathbf{x}_b}$. The conditional distribution of the intensity value given the background at position $\mathbf{x}_b$ is expressed as

$$p(I|B_{\mathbf{x}_b}) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(I - \bar{I}(\mathbf{x}_b))^2}{2\sigma^2}) \tag{4}$$

where $\bar{I}(\mathbf{x}_b)$ and σ are the mean and standard deviation of the background distribution at $\mathbf{x}_b$. In the process of foreground detection, it is probable that the foreground intensity value appears in any position within its valid range. If $\mathbf{x}_b$ is the corresponding position of $\mathbf{x}_c$, we are concerned with whether the intensity value belongs to the background $B_{\mathbf{x}_b}$ or the targets $T_{\mathbf{x}_b}$, instead of which target Gaussian it belongs to. For the consideration of the above reasons and to minimize the computational cost, we regard that the foreground is uniformly distributed with $p(I|T_{\mathbf{x}_b}) = 1/L$, L is decided by the valid range of the intensity value I. According to the Bayesian decision rule, whether an intensity value belongs to the background or the targets can be depicted by a likelihood ratio test.

$$\frac{p(I|B_{\mathbf{x}_b})}{p(I|T_{\mathbf{x}_b})} \geq \frac{P(T)}{P(B)} = \lambda \tag{5}$$

$P(T)$ and $P(B)$ are assumed constant with respect to $\mathbf{x}_b$ and decided by the proportion of the typical time duration a pixel belongs to the background and the foreground respectively [9]. Replacing $p(I|B_{\mathbf{x}_b})$ with Eq. (4) yields the $logarithm$ of the likelihood decision form

$$|I - \bar{I}(\mathbf{x}_b)| \leq k\sigma \tag{6}$$

where $k = \sqrt{-2\ln(\sqrt{2\pi}\sigma\lambda/L)}$.

For a certain pixel with the intensity value $I(\mathbf{x}_c)$ in the current frame, there is a corresponding SDG model in the Background Map. This SDG model is composed of the local background Gaussians centered at $\hat{\mathbf{x}}_b$. The size of SDG is decided according to Eq. (3). The problem of foreground detection from a non-stationary background can be regarded as the pixel-wise decision problem based on a SDG model. If there exists $\mathbf{x}_b$ in the Background Map, where $I(\mathbf{x}_c)$ belongs to the $B_{\mathbf{x}_b}$ rather than $T_{\mathbf{x}_b}$ and $\mathbf{x}_b \in \mathcal{A}_{\hat{\mathbf{x}}_b}$, that is, $I(\mathbf{x}_c)$ belongs to one of the background Gaussians of its SDG model, the pixel is labeled as background; otherwise, it is labeled as foreground.

2.4 R and the size of the SDG model

The covariance matrix $\mathbf{R}$ of the positional errors in Eq. (2) is important when deciding the size of the SDG model. In this paper, we assume that $\mathbf{R}$ is constant and approximately estimated as

$$\mathbf{R} = \alpha\hat{\mathbf{E}} \tag{7}$$

where $\hat{\mathbf{E}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{c}_{1i} - \mathbf{c}_{1i}^{'})(\mathbf{c}_{1i} - \mathbf{c}_{1i}^{'})^T$ and $s\tilde{\mathbf{c}}_{1i}^{'} = \Gamma\tilde{\mathbf{c}}_{2i}$, s is a nonzero scaler and $\{(\mathbf{c}_{i1}, \mathbf{c}_{i2}), i = 1, 2, \ldots, n\}$ is the set of the wholly available corresponding corner pairs in the two images. According to Eq. (3), when the confidence probability is given, the size of the SDG model is mainly decided by $\mathbf{R}$. With $\hat{\mathbf{E}}$ being estimated, as α increases, the size of the SDG model increases and different results of the detection will be obtained accordingly. False Alarm Rate (FA) and Miss Detection Rate (MD) [8] are applied to evaluate the efficiency of the foreground detection. The problem is to decide the value of α to ensure a balance between $\mathbf{FA}$ and $\mathbf{MD}$. Figure 3 shows the relationship between $\mathbf{FA}$, $\mathbf{MD}$ and coefficient α (the size of the SDG model). When $\hat{\mathbf{E}}$ has been estimated, the increase of α will result in an increase of the size of the SDG model and hence decrease the FA (Figure 3(a), the method we propose) and slightly increase the MD (Figure 3(b), the method we propose). With the SDG model, we can eliminate much of the false detection due to the registration errors. It is not true that the larger the size of SDG model, the better the results we can obtain. Some true foreground detection will be missed when the foreground pixel happen to belong to the background of a certain position within its SDG model. In applications, the proper α can be set according to the balance of the FA and MD.

2.5 Dynamic restoration and adaptation of the Background Map

The previous sections assumed that the Background Map is available. This map needs to be updated under two conditions: (1) when the camera pans and tilts to image new, uncovered background scenes, or (2) when the background scene changes due to lighting or environmental changes. The objective of the previous sections was to classify a given pixel in the current image as a foreground pixel or a background pixel using the SDG model within the Background Map. However which Gaussian within the SDG model should this current pixel be used to update is not precisely known. This section deals with the problem of using the appropriate pixel in the current frame to update a given Gaussian of the MOG at a certain position, and decides which Gaussian is corresponding to the background and hence the SDG model at this position.

As described in Section 2.1, each pixel along an image sequence is considered as a statistical process. The distribution of the intensity value can be modeled as a MOG. Learning the parameters of the MOG and hence the detection of the foreground is a broadly studied topic for a static background [5, 11, 4]. For a moving background, the problem is to decide the correspondence of the pixels in the previous and current frame during the learning stage. The problem is further complicated when there is occlusion and/or uncovered background.

The background coordinate system is defined with respect to that of the first frame. Since the pixel process is considered as an independent process along the image sequence, a random pixel $\mathbf{x}_b$ is used to illustrate the procedure of background restoration and adaptation. Assume that the distribution of the intensity value at $\mathbf{x}_b$ is modeled as a c mixture of Gaussians (c is not required to be known in advance). We initialize the first Gaussian with the intensity value at $\mathbf{x}_b$ in the first frame as the mean and a predefined variance σ^2. The first Gaussian is not guaranteed to be the background Gaussian. ω_0 is defined to describe the cases that occlusion and/or uncovered background appear; with this a new Gaussian distribution should be created.

With a new frame, the motion parameters are estimated and the background is transformed and warped to the current frame. The predicted position of $\mathbf{x}_b$ in the current frame is $\hat{\mathbf{x}}_c$, where $\tilde{\hat{\mathbf{x}}}_c = s\Gamma^{-1}\tilde{\mathbf{x}}_b$. Assume that we already have m Gaussians $p(I|\omega_j)$ ($j = 1, ..., m;\ 1 \leq m \leq c$), which are corresponding to the background and individual targets. If there is no occlusion and/or uncovered background, there should be a corresponding pixel of $\mathbf{x}_b$ in the current frame. The position of the corresponding pixel is assumed to be Gaussian distributed about $\hat{\mathbf{x}}_c$ and is expressed as $p(\mathbf{x}_c|\hat{\mathbf{x}}_c) = \mathcal{N}(\mathbf{x}_c; \hat{\mathbf{x}}_c, \mathbf{R})$. Accordingly, there is a validation region $\mathcal{A}_{\hat{\mathbf{x}}_c}$ about $\hat{\mathbf{x}}_c$ in the current frame. The pixels in this validation region are feasible corresponding pixels of $\mathbf{x}_b$. The positions which are not the corresponding position of $\mathbf{x}_b$ are modeled as independent identically distributed (IID) random variables with uniform spatial distribution. Events θ_i and θ_0 are defined to describe the relationship of the position $\mathbf{x}_c = \mathbf{x}_{ci}$ in the current frame and the predicted position $\hat{\mathbf{x}}_c$, where, $\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}$, $i = 1, ..., n$. $\theta_i \triangleq \{\ \mathbf{x}_{ci}\ is\ the\ corresponding\ position\ of\ \mathbf{x}_b\}$, $\theta_0 \triangleq \{\ none\ of\ the\ positions\ is\ the\ corresponding\ position\ of\ \mathbf{x}_b\}$, and $\sum_{i=0}^{n} P(\theta_i) = 1$. BarShalom gives the details of the derivation of each $P(\theta_i)$ in [2]. $P(\theta_i)$ is mainly decided by the Mahalanobis distance $D_{\mathbf{x}_{ci}, \hat{\mathbf{x}}_c} = (\mathbf{x}_{ci} - \hat{\mathbf{x}}_c)^T \mathbf{R}^{-1} (\mathbf{x}_{ci} - \hat{\mathbf{x}}_c)$. Since $\mathbf{R}$ is assumed to be constant globally, $P(\theta_i)$ can be calculated in advance and a lookup table can be used when performing background restoration and adaptation.

For $\mathbf{x}_b$, if there is a corresponding pixel $\mathbf{x}_c^*$ in the current frame, the intensity value $I(\mathbf{x}_c^*)$ should belong to a certain Gaussian of the $m-$MOG which have been already learnt. If there is no corresponding pixel in the current frame (the presence of the foreground and/or uncovered background), no pixel in the current frame belongs to any Gaussian of the $m-$MOG. ω_0 is active and a new Gaussian should be created. $P(\omega_0)$ is given by κ and $p(I|\omega_0) = 1/L$. We decide the corresponding pixel $\mathbf{x}_c^*$ of the position $\mathbf{x}_b$ in the following steps:

1. For each pixel $\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}$ with the same $P(\theta_i)$ and $I = I(\mathbf{x}_{ci})$, if it is the corresponding position of $\mathbf{x}_b$ in the current frame, find the plausible ω_{j_i} it belongs to:

$$\begin{aligned} \omega_{j_i} &= \arg \max_{0 \leq j \leq m} P(\omega_j | I(\mathbf{x}_{ci})) \\ &= \arg \max_{0 \leq j \leq m} p(I(\mathbf{x}_{ci})|\omega_j) P(\omega_j) \end{aligned} \tag{8}$$

If no better argument is available about the prior probability $P(\omega_j)$, we usually assume that $P(\omega_j)$ is equal and satisfies $\kappa + \sum_{j=1}^{m} P(\omega_j) = 1$.

2. For all $\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}$ with plausible ω_{j_i}, the corresponding pixel $\mathbf{x}_c^*$ and the updated Gaussian are decided by

$$\begin{aligned} (\mathbf{x}_c^*, \omega_j^*) &= \arg \max_{\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}, 0 \leq j_i \leq m} P(\omega_{j_i} | I(\mathbf{x}_{ci})) P(\theta_i) \\ &= \arg \max_{\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}, 0 \leq j_i \leq m} \frac{p(I(\mathbf{x}_{ci})|\omega_{j_i}) P(\omega_{j_i}) P(\theta_i)}{\sum_{n_i=0}^{m} p(I(\mathbf{x}_{ci})|\omega_{n_i}) P(\omega_{n_i})} \end{aligned} \tag{9}$$

If $\omega_j^* \neq 0$, the pixel $\mathbf{x}_c^*$ is regarded as the corresponding pixel of $\mathbf{x}_b$ and the parameters of Gaussian ω_j^* are updated with $I(\mathbf{x}_c^*)$. If $\omega_j^* = 0$, or for all $\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}$, $\omega_{j_i} = \omega_0$, that means no corresponding pixel of $\mathbf{x}_b$ in this frame and $I(\hat{\mathbf{x}}_c)$ is used to initialize a new Gaussian with variance σ^2. This may cause a small disturbance of position in the restored background. But this position disturbance will not cause fatal error when using the SDG model to detect the foreground.

As mentioned in Section 2.1, the background distribution is a narrow Gaussian with a high prior probability $P(B)$. We regard ω_j which has a higher frequency and lower standard deviation corresponding to the background.

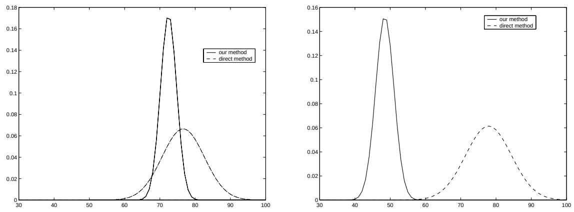
3 Applications and experimental results

Two applications of the SDG model based motion detection are developed, they are foreground detection of an outdoor scene from a hand-held, moving camera and a human tracking system with a pan-tilt camera.

3.1 Background Map restoration, adaptation and foreground detection of an outdoor scene

The performances of the Background Map restoration, adaptation and foreground detection are evaluated with a sequence of images which are extracted from a moving handheld video camera. With the first several frames, the Background Map is restored, while the parameters are learnt even though no Background Map is available in advance. In the following frames, foreground is detected and the Background Map is adapted based on the SDG model.

Figure 2 shows the experimental results of the restored and adapted background. We make a comparison of two approaches, one is the method we proposed above, another is the direct method which restores and adapts the background directly after motion compensation. Figure 2(a) and



(a) (b)

Figure 1.

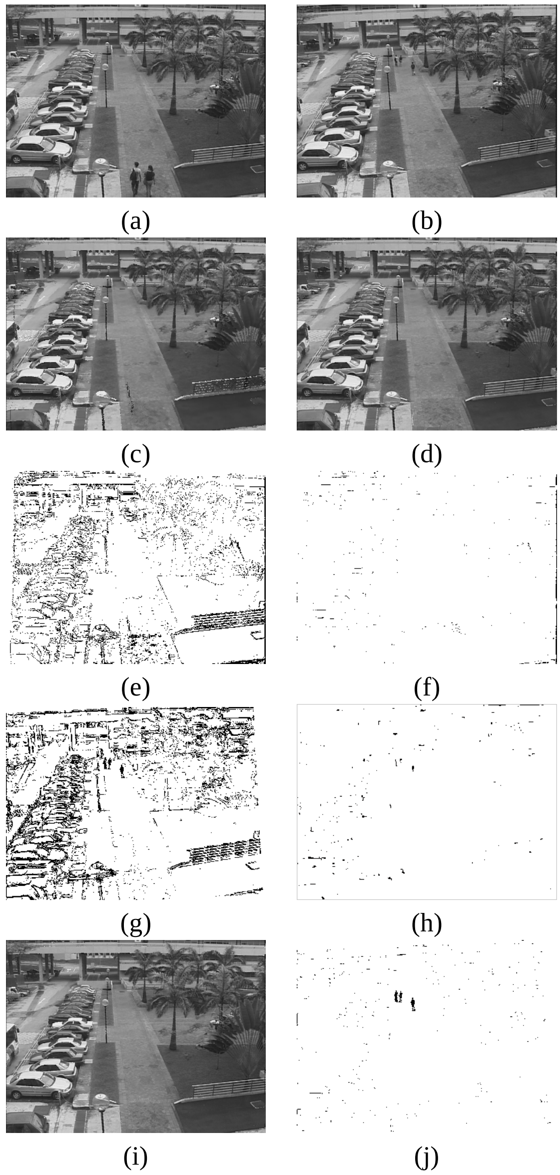


Figure 2. Background restoration, adaptation and foreground detection.

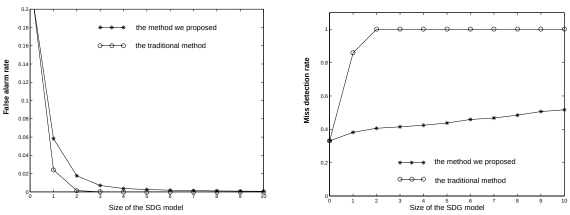
(b) are Frame 1 and 136 (frame rate is 5 frames/sec) of the sequence. Figure 2(c) shows the result of the restored Background Image at frame 50 using the direct method. Figure 2(d) shows the restored Background Image using our method. The restored Background Image using the direct method is more blur than the one using our approach. Figure 2(e) and (f) are the error maps between the restored Background Images and an ideal Background Image for both methods. Due to the inaccurate motion compensation, there is an accumulation of the motion compensation errors. As showing in Figure 1, the background Gaussian learnt directly after motion compensation is more flat than the Gaussian learnt by our approach. Sometimes, the estimated Gaussian is bias (Figure 1(b)). With these background Gaussians, the efficiency of foreground detection will be impacted.

In this test case, the camera motion cannot be neglected and the moving objects (humans) are small. An affine transformation model is applied for the estimation of the motion parameters and motion compensation. Figure 2(g) is the result of foreground extraction using background subtraction after affine motion compensation at Frame 136, and (h) is the result of (c) after a 3×3 morphological operation. With this traditional approach, the moving objects are submerged by noise. The restored Background Image using the technique described above is illustrated in Figure 2(i). Figure 2(j) is the result of segmenting the moving objects of Frame 136 based on the SDG model.

Figure 3 shows the results of the error analysis based on the traditional method and our approach. In the traditional method, let the size of the morphological mask be equal to the size of the SDG model. When increasing the size, the FA decreases for both methods. Significantly, MD increases dramatically using the traditional method while our approach exhibits minimal increase in MD. When the size of the operator is larger than 2 pixels, the MD using the traditional method reaches 100% even though its FA is decreased. Detection fails subsequently. On the contrary, the MD using our approach is not as sensitive to the size of the SDG model. It shows that our approach is insensitive to motion compensation error and is able to detect small objects.

3.2 Indoor active human tracking with pan-tilt camera

Another application of the SDG model is an indoor active human tracking system with a pan-tilt camera. When the pan and tilt are assumed to rotate about the optical center, or the distance from the camera to the scene is assumed to be much larger than the depth of the scene itself, a simplified 2D transformation can be applied to calculate the pixel transformation. With the calibration parameters and pan-tilt angle α-β, a panorama Background Map is established



(a) (b)

Figure 3. Error analysis with the varying size of the SDG model.

Figure 4. Human detection and tracking with pan-tilt camera.

and foreground can be detected. In practice, the assumption is difficult to realize. The non-coincidence of the optical center and the rotation center makes the accuracy of the 2D transformation affected by the depth and bring errors to the results of the transformation.

In this application, the SDG model is applied to eliminate the errors that originate from the approximation of the 2D transformation as well as other artefacts. 5 five-hundred-frame sequences of clean background are captured from 5 different pan/tilt positions. The intensity value of every pixel in each background sequence is modeled as one Gaussian. The 5 sequences are wrapped to the same coordinate system. The Gaussians covering the same pixel consist a MOG, and all these MOGs form the Background Map mosaic. All these work can be done off-line. Figure 4(a) shows the result of the Background Map mosaic. In the tracking stage, for each pixel in the current frame, the motion compensation with arbitrary pan/tilt angle can be calculated. Due to the random rotation angle of the pan-tilt camera, the position of the camera cannot be the exact position where the sequences of background images are captured. Even with the MOG Background Map mosaic, motion compensation error cannot be eliminated (Figure 4(c)). With the SDG model, referring to Figure 4(d), moving objects are detected and tracked accurately. The detection based on our approach is able to extract the desired target without significant noise clutter.

4 Conclusions

This paper proposes a SDG model which is used to detect the foreground from a non-stationary background. It extends the application of the background subtraction to the moving sensor and is robust even with approximate motion compensation, noise, or environmental changes. The detection based on the SDG model shows good results even when the detection is applied to small moving objects in a highly textured background. With a non-stationary background, an algorithm is proposed for the background restoration and statistically modeling the background, which is required when detecting the foreground. These algorithms are in pixel-wise case; no iterative computations are required. As such, they are suitable for parallel implementations for real-time considerations.

References

[1] S. Araki, T. Matsuoka, H. Takemura, and N. Yokoya. "Real-time tracking of multiple moving objects in moving camera image sequences using robust statistics". In *Proc. of ICPR*, pages 1433–1435, 1998.

[2] Y. BarShalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1989.

[3] D.Murray and A. Basu. "Motion tracking with an active camera". *IEEE Trans. on PAMI*, 16(5):449–459, 1994.

[4] A. Elgammal, D. Harwood, and L. Davis. "Non-parametric model for background subtraction". In *Proc. of ECCV*, pages 751–767, 2000.

[5] I. Haritaoglu, L. S. Davis, and D. Harwood. "W4: who? when? where? what? a real time system for detecting and tracking people". In *Proc. of ICAFGR*, pages 222–227, 1998.

[6] T. Kanade, R. T. Collins, and A. J. Lipton. "Advances in coorperative multi-sensor video surveillance". In *Proc. of DARPA*, 1998.

[7] L.Marcenaro, F.Oberti, and C.Regazzoni. "Change detection methods for automatic scene analysis by using mobile surveillance cameras". In *Proc. of ICIP*, pages 244–247, 2000.

[8] Y. Ren, C. S. Chua, and Y. K. Ho. "Motion detection with non-stationary background". *to appear in Proc. of ICIAP*, 2001.

[9] J. Rittscher, J. Joga, and A. Blake. "A probabilistic background model for tracking". In *Proc. of ECCV*, pages 336–351, 2000.

[10] S.Rowe and A. Blake. "Statistical mosaic for tracking". *IVC*, 14:549–564, 1996.

[11] C. Stauffer and W. Grimson. "Adaptive background mixture models for real-time tracking". In *Proc. of CVPR*, pages 246–252, 1999.

[12] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. "Pfinder: Real-time tracking of the human body". *IEEE Trans. on PAMI*, 19(7):780–785, 1997.