Machine Vision and Applications (2003) 13: 332–343
Digital Object Identifier (DOI) 10.1007/s00138-002-0091-0

Machine Vision and Applications

Motion detection with nonstationary background

Ying Ren, Chin-Seng Chua, Yeong-Khing Ho

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

Received: 30 July 2001 / Accepted: 20 April 2002

Abstract. This paper proposes a new background subtraction method for detecting moving foreground objects from a nonstationary background. While background subtraction has traditionally worked well for a stationary background, the same cannot be implied for a nonstationary viewing sensor. To a limited extent, motion compensation for the nonstationary background can be applied. However, in practice, it is difficult to realize the motion compensation to sufficient pixel accuracy, and the traditional background subtraction algorithm will fail for a moving scene. The problem is further complicated when the moving target to be detected/tracked is small, since the pixel error in motion that is compensating the background will subsume the small target. A spatial distribution of Gaussians (SDG) model is proposed to deal with moving object detection having motion compensation that is only approximately extracted. The distribution of each background pixel is temporally and spatially modeled. Based on this statistical model, a pixel in the current frame is then 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 nonstationary background. Test cases involving the detection of small moving objects within a highly textured background and with a pan-tilt tracking system are demonstrated successfully.

Key words: Foreground detection – Nonstationary background – SDG model – Background restoration – Background adaptation

1 Introduction

Motion detection and segmentation is a basic problem in computer vision. It is a prerequisite for security surveillance, object tracking, autonomous obstacle avoidance, and image compression. However, it is still in the developmental stage and needs to be robust enough when applied in an unconstrained environment.

Three main approaches have been developed to detect motion, namely, optic flow, temporal difference, and background subtraction [12].

Optic flow [2] can achieve success of motion detection in the presence of camera motion. According to the smoothness constraint, the corresponding points in the two successive frames should not move more than a few pixels. For a moving camera, this means that the camera motion should be relatively small. Moreover, due to the computational cost, this approach is inapplicable for use in real-time without specialized hardware.

Temporal difference [11,19] is a simple method for detecting moving objects in a static environment. As we can expect, the result of the temporal difference is poor because not all feature pixels of the moving regions can be extracted, and the outcome depends heavily on the speeds of the moving objects when the capture interval is constant. The motion of the object is required to be continuous, since the objects cannot be detected when there is no motion between frames. For a nonrigid object, such as the human body, parts of the body may not move within a short duration and, hence, will not be detected in these frames. One of the advantages of this method is the invariance of the visual information under changing illumination.

Background subtraction is a popular technique for finding moving objects in a sequence of images [6,7,9,16–18]. Statistical background modeling makes the foreground detection more robust to illumination changes, shadows, and other artifacts. This approach provides a more complete set of feature data describing the moving targets when compared with other motion detection approaches [12]. Conditionally, the background scene and the viewing sensor are required to be stationary when background subtraction is applied.

Recently, motion detection with a nonstationary viewing sensor has attracted the attention of several research groups [1,4,5,15]. 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 (background motion). Usually a motion model of the back-

Correspondence to: Chin-Seng Chau

ground 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 is 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 3-D 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.

This problem is also encountered when using the temporal difference method [1,4]. Murray et al. [4] utilize morphological operations to eliminate the errors due to the motion compensation and other noises. This is an *all-or-nothing* method, and the result depends on the error due to the background motion compensation, the size of the morphological operator, and the size of the moving object. When the size of the operator is larger than the motion compensation error and smaller than the size of the moving object, this method works well. When the size of the operator is smaller than the motion compensation error, after an erosion-dilation operation, the error remains as before. When the size of the moving object is small compared to the morphological-operator size, the moving object will be eliminated as well. Rowe et al. [15] use a statistical two-class mixture of Gaussians (MOG) to describe the mosaic background on a "visual" image plane. Background subtraction is applied to segment the foreground and reduce the influence of the background on the foreground model when performing the visual tracking. But when the position of the current image is not perfectly aligned with the background mosaic, the error due to the background motion compensation cannot be eliminated. 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. Figure 6a and b show frame 1 and frame 136 of an image sequence involving several moving persons against a stationary background. This image sequence is extracted from a moving handheld video camera (moving sensor). To compensate for sensor movement, an affine motion compensation is applied using a traditional approach [8]. The objective is to detect and track the moving persons, despite the sensor being nonstationary. Referring to Fig. 6c and d, after motion compensation, background subtraction, and a morphological operation, the small targets (moving persons) were submerged by the noise and could not be extracted.

In this paper, we propose a spatial distribution of Gaussians 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. The approach is able to detect and track small moving objects in a highly textured background.

In the remainder of this paper, in Sect. 2, we introduce the SDG model and the technique of background restoration, adaptation, and foreground detection. Section 3 gives experimental results of background restoration, adaptation, and foreground detection with a moving sensor, based on the SDG model. Finally, conclusions are drawn in Sect. 4.

2 Spatial distribution of Gaussians model and foreground detection

The basic idea of this approach is that we can model the intensity value of each background pixel as a Gaussian distribution (or mixture of Gaussians),which can be learned and adapted along the image sequence. For the current frame and the background, the dominant motion is the motion due to the moving sensor. We assume that motion can be approximated by a 2-D parametric transformation, such as affine or projective, in the image plane. The conditions of this assumption are that the scene is far away from the viewing camera and has a small depth, or that the camera undergoes pure rotation and/or zoom. Traditional approaches are used to estimate the transformation parameters, after which the current image is warped to align with the background. Note that, for our approach, only an approximate alignment is assumed. 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, or a mixture of Gaussians. Each Gaussian may correspond to the distribution of background or individual moving objects covering this pixel over time. Note that the distributions are different from pixel to pixel. For each pixel, the distribution is fitted with multiple Gaussians that compose the MOG model. Figure 1 shows a distribution of intensity values $I(\mathbf{x})$ for a given pixel, $\mathbf{x}$, of an image sequence extracted over one hour. We model it as an MOG, and the *probability density function* is given as

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

where B stands for the background, ω_j denotes the intensity classes of the moving objects, and c the number of Gaussians for that pixel. An online learning and adaptive algorithm [13] has been developed to obtain and update the parameters of the MOG. For clarity of discussion, we assume that the intensity distribution for each pixel may be modeled as a two-component MOG, namely, a narrow Gaussian of the background and a flat Gaussian (or a uniform distribution) of the moving objects (refer to Fig. 1). Over a long period of time, different moving objects may cover a certain pixel, and the foreground at this pixel can be of any value within its valid range. The distribution of the foreground is hard to estimate along the image sequence. The widely distributed Gaussian (or uniform distribution) is a reasonable assumption. This assumption is also employed by other researchers [7,14]. Then

Eq. 1 can be modified as

$$p(I) = p(I|B)P(B) + p(I|T)P(T), \tag{2}$$

where T denotes moving targets. 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 to each new incoming frame. A *background image* $\mathcal{I}_\mathrm{b}$ 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 Spatial distribution of Gaussians model

2.2.1 Background motion compensation. For each pixel $I(\mathbf{x}_\mathrm{c})$ in the current image, motion compensation is applied. 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 $\mathcal{I}_\mathrm{c}$ to the background image $\mathcal{I}_\mathrm{b}$. The position after motion compensation is, at best, a predicted position in the background map. Let the predicted position of $\mathbf{x}_\mathrm{c}$ in the background map be $\hat{\mathbf{x}}_\mathrm{b}$. $s\tilde{\hat{\mathbf{x}}}_\mathrm{b} = \mathbf{\Gamma}\tilde{\mathbf{x}}_\mathrm{c}$, where $\tilde{\hat{\mathbf{x}}}_\mathrm{b} = [\hat{x}_\mathrm{b}, \hat{y}_\mathrm{b}, 1]^T$ and $\tilde{\mathbf{x}}_\mathrm{c} = [x_\mathrm{c}, y_\mathrm{c}, 1]^T$ are homogeneous coordinates (hereafter denoted by $\tilde{}$), s is an arbitrary nonzero scalar, and $\mathbf{\Gamma}$ is the transformation matrix for background motion compensation.

To estimate the transformation matrix $\mathbf{\Gamma}$, corners are selected as features due to their positional accuracy and low computational cost. Corners are extracted from the background image $\mathcal{I}_\mathrm{b}$ and the current image $\mathcal{I}_\mathrm{c}$ in a coarse-to-fine structure. The coarse-to-fine corner detection and selection promise a homogeneous corner distribution within images [10]. The best l corresponding corner pairs are selected into a set $\{\mathbf{C}_1, \mathbf{C}_2\}$, $\mathbf{C}_1 = \{\mathbf{c}_{1i}, i = 1, 2, \ldots, l\}$, $\mathbf{C}_2 = \{\mathbf{c}_{2i}, i = 1, 2, \ldots, l\}$, and $\mathbf{c}_{1i}$ and $\mathbf{c}_{2i}$ are corner positions in the two images, respectively. The least-square-estimation (LSE) method is used to estimate the transformation matrix $\mathbf{\Gamma}$ according to the assumed transformation model, which is usually the affine or projective transformation. When the influence of the outliers cannot be neglected, the least-median-of-squares (LMedS) method can be applied.

2.2.2 SDG model. For a position $\mathbf{x}_\mathrm{c}$ in the current image, after the compensation for the background motion, its predicted position in the background map is $\hat{\mathbf{x}}_\mathrm{b}$. The corresponding position of $\mathbf{x}_\mathrm{c}$ in the background map is assumed to be Gaussian distributed about $\hat{\mathbf{x}}_\mathrm{b}$ and is expressed as

$$p(\mathbf{x}_\mathrm{b}|\hat{\mathbf{x}}_\mathrm{b}) = \frac{1}{2\pi|\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}_\mathrm{b} - \hat{\mathbf{x}}_\mathrm{b})^T\mathbf{R}^{-1}(\mathbf{x}_\mathrm{b} - \hat{\mathbf{x}}_\mathrm{b})\right), \tag{3}$$

where $\mathbf{R}$ is the covariance matrix of positional errors, which originate from the errors mentioned in Sect. 2.2.1. There is a validation region $\mathcal{A}_{\hat{\mathbf{x}}_\mathrm{b}}$

$$\mathcal{A}_{\hat{\mathbf{x}}_\mathrm{b}} \triangleq \{\mathbf{x}_\mathrm{b} : D_{\mathbf{x}_\mathrm{b},\hat{\mathbf{x}}_\mathrm{b}} \leq \gamma\}, \tag{4}$$

where $D_{\mathbf{x}_\mathrm{b},\hat{\mathbf{x}}_\mathrm{b}} = (\mathbf{x}_\mathrm{b} - \hat{\mathbf{x}}_\mathrm{b})^T\mathbf{R}^{-1}(\mathbf{x}_\mathrm{b} - \hat{\mathbf{x}}_\mathrm{b})$ is the Mahalanobis distance from a random position $\mathbf{x}_\mathrm{b}$ in the background map to the predicted position $\hat{\mathbf{x}}_\mathrm{b}$. $D_{\mathbf{x}_\mathrm{b},\hat{\mathbf{x}}_\mathrm{b}}$ is χ^2 distributed. The positions in the validation region are possible solutions of the real corresponding position of $\mathbf{x}_\mathrm{c}$. The real corresponding position of $\mathbf{x}_\mathrm{c}$ will be found in this region with a certain probability decided by γ.

As described in Sect. 2.1, for a certain position $\mathbf{x}_\mathrm{b}$, the intensity distribution of that pixel is modeled as an MOG, namely, the Gaussians of the background and the targets (Eq. 2). The conditional distribution of the intensity value, given the background at position $\mathbf{x}_\mathrm{b}$, is expressed as

$$p(I|B_{\mathbf{x}_\mathrm{b}}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I - \bar{I}(\mathbf{x}_\mathrm{b}))^2}{2\sigma^2}\right), \tag{5}$$

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

For a certain pixel with the intensity value $I(\mathbf{x}_\mathrm{c})$ in the current frame, if there exists $\mathbf{x}_\mathrm{b}$ in the background map, where $I(\mathbf{x}_\mathrm{c})$ belongs to the background $B_{\mathbf{x}_\mathrm{b}}$ rather than the targets $T_{\mathbf{x}_\mathrm{b}}$ and $\mathbf{x}_\mathrm{b} \in \mathcal{A}_{\hat{\mathbf{x}}_\mathrm{b}}$, we label this pixel as background; otherwise, the pixel is labeled as foreground.

According to the Bayesian decision rule, whether an intensity value belongs to the background or to the targets can be depicted by the likelihood ratio test

$$\frac{p(I|B_{\mathbf{x}_\mathrm{b}})}{p(I|T_{\mathbf{x}_\mathrm{b}})} \geq \frac{P(T)}{P(B)} = \lambda. \tag{6}$$

$P(T)$ and $P(B)$ are the prior knowledge about the probabilities of the background and targets. Here we assume that they are constant with respect to $\mathbf{x}_\mathrm{b}$ and that they are decided by the proportion of the typical time duration a pixel belongs to the background and the foreground, respectively [14]. Replacing $p(I|B_{\mathbf{x}_\mathrm{b}})$ with Eq. 5 yields the likelihood decision form

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I - \bar{I}(\mathbf{x}_\mathrm{b}))^2}{2\sigma^2}\right) \geq \frac{\lambda}{L}. \tag{7}$$

The *logarithm* of Eq. 7 converts the likelihood discriminant problem into a distance discriminant problem,

$$|I - \bar{I}(\mathbf{x}_\mathrm{b})| \leq k\sigma, \tag{8}$$

where

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

Figure 2 shows the changes of k with respect to λ and σ. In this paper, since $P(T)$ and $P(B)$ are constant, λ is also constant.

The problem of foreground detection from a nonstationary background can be regarded as a pixel-wise decision problem based on the SDG model. For a certain pixel $I(\mathbf{x}_\mathrm{c})$ in the

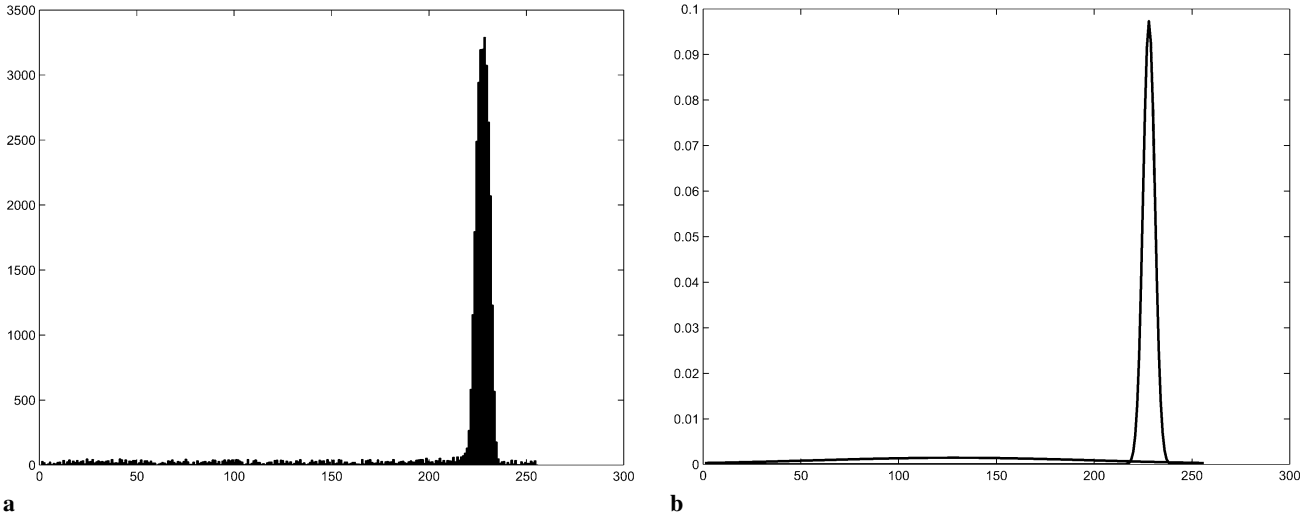


Fig. 1. a Histogram of intensity values for one pixel of an image sequence over one hour. **b** Background distribution can be fitted with a narrow Gaussian and the distribution of the targets can be fitted with a Gaussian with large deviation

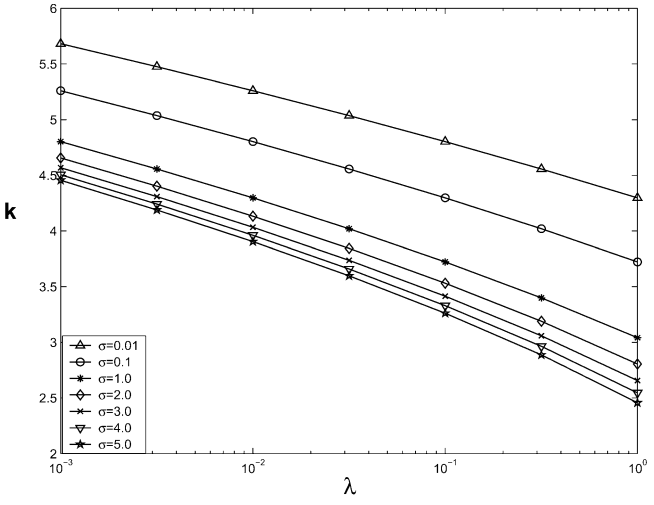


Fig. 2. Relationship of k, λ, and σ

current image, there is a corresponding SDG model in the background map. This SDG model is composed of the local background Gaussians centered at $\hat{\mathbf{x}}_{\mathrm{b}}$, which is the position after motion compensation. The size of SDG is decided according to Eq. 4. If $I(\mathbf{x}_{\mathrm{c}})$ belongs to any of the background Gaussians of its SDG model, it is labeled as background. If no corresponding background distribution can be found in its SDG model, the pixel $I(\mathbf{x}_{\mathrm{c}})$ is regarded as the foreground.

2.2.3 **R** *and the size of the SDG model.* The covariance matrix **R** of the positional errors in Eq. 3 is important when deciding the size of the SDG model. As mentioned in Sect. 2.2.1, after motion compensation, the sources of spatial errors include motion model error, motion parameter estimation error, feature localization error, distortion of lens, and residual error. In fact, **R** may be different from pixel to pixel and is difficult to calculate directly. In this paper, we assume that **R** is constant and approximately estimated as

$$\mathbf{R} = \alpha\hat{\mathbf{E}} \tag{9}$$

and

$$\hat{\mathbf{E}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{c}_{1i} - \mathbf{c}_{1i}^{'})(\mathbf{c}_{1i} - \mathbf{c}_{1i}^{'})^{T} \tag{10}$$

where $s\tilde{\mathbf{c}}_{1i}^{'} = \mathbf{\Gamma}\tilde{\mathbf{c}}_{2i}$, s is a nonzero scaler and $\{(\mathbf{c}_{1i}, \mathbf{c}_{2i}), i = 1, 2, \ldots, n\}$ is the set of the wholly available corresponding corner pairs in the two images. According to Eq. 4, the locus of the boundary of the SDG model is an ellipse. When the confidence probability is given, the size of the SDG model is mainly decided by **R**. With $\hat{\mathbf{E}}$ being estimated, as α increases, the size of the SDG model increases and different results of the detection are obtained accordingly.

To evaluate the efficiency of the foreground detection, two kinds of error rate are defined as follows:

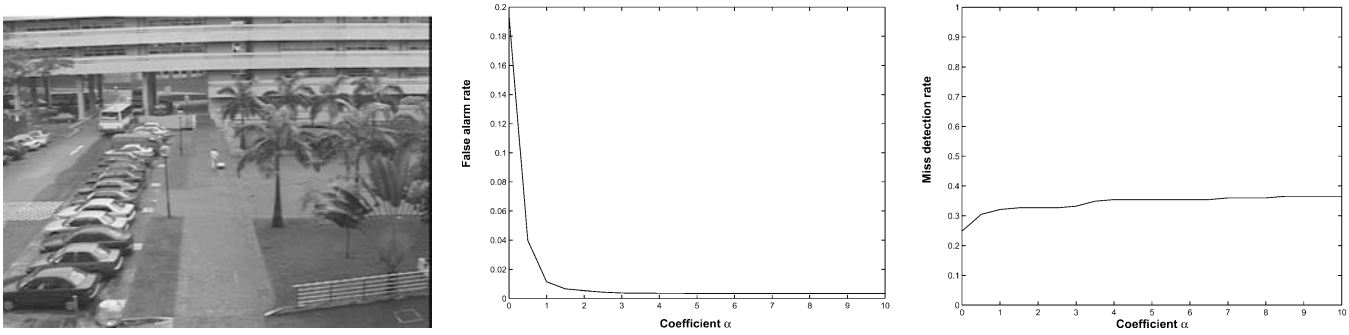
False-alarm rate (FA)

$$\mathbf{FA} = \frac{\text{falsely detected area (size)}}{\text{whole image area (size)}} \tag{11}$$

Miss-detection rate (MD)

$$\mathbf{MD} = \frac{\text{undetected foreground area (size)}}{\text{target area (size)}} \tag{12}$$

The problem is to decide the value of α to ensure a balance between the FA and the MD. Figure 3 shows the relationship between the FA, the MD, and coefficient α. The analyzed images of a small moving object and highly textured background are extracted from a moving camera (Fig. 3a). When $\hat{\mathbf{E}}$ has been estimated according to (Eq. 10), the increase of α results in an increase in the size of the SDG model and, hence, decrease the FA (Fig. 3b) and slightly increase the MD (Fig. 3c). With the SDG model, we can eliminate much of the false detection due to the registration errors. It is not true



a **b** **c**

Fig. 3a–c. Error analysis with the spatial covariance setting. **a** A typical analyzed image from an image sequence. The foreground is a person and a trolley in the center of the image. **b** False-alarm rate vs. coefficient α. **c** Miss-detection rate vs. coefficient α

that the larger the size of the SDG model, the better the results we can obtain. Some true foreground detection will be missed when the foreground pixels happen to belong to the background of a certain position within the SDG model. In our application, a proper α can be set according to the balance of the FA and MD.

2.2.4 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, it is not precisely known which Gaussian within the SDG model this current pixel should be used to update. 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, in terms of deciding which Gaussian corresponds to the background and, hence, to the SDG model at this position.

As described in Sect. 2.1, each pixel along an image sequence is considered as a statistical process. The distribution of the intensity value can be modeled as an MOG, namely, the Gaussians of background and targets. Learning the parameters of the MOG for a static background and, hence, the detection of the foreground, is a broadly studied topic [6,7,16]. For a moving background, the problem is to decide the correspondence of the pixels in the previous and current frames during the learning stage. The problem is further complicated when there is an occlusion and/or uncovered background.

The background coordinate system is defined with respect to the background of the first frame. Since the pixel process is considered to be an independent process along the image sequence, a random pixel $\mathbf{x}_\mathrm{b}$ is used to illustrate the procedure for background restoration and adaptation. As described in Section 2.1, the distribution of the intensity value at $\mathbf{x}_\mathrm{b}$ is modeled as a c mixture of Gaussians (knowing c in advance is not required). We initialize the first Gaussian with the intensity value at $\mathbf{x}_\mathrm{b}$ 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 where an occlusion and/or uncovered background appear; and with this, a new Gaussian distribution should be created.

For the pixel $\mathbf{x}_\mathrm{b}$, assume that we already have m Gaussian distributions $p(I|\omega_j)$ $(j = 1,\ldots,m;\ 1 \le m \le c)$, which correspond to the background and individual targets

$$p(I|\omega_j) = \mathcal{N}(I; \bar{I}_{\omega_j}(\mathbf{x}_\mathrm{b}), \sigma^2_{\omega_j}(\mathbf{x}_\mathrm{b})), \tag{13}$$

where $\bar{I}_{\omega_j}(\mathbf{x}_\mathrm{b})$ and $\sigma_{\omega_j}(\mathbf{x}_\mathrm{b})$ are the mean and standard deviation of the jth Gaussian distribution at $\mathbf{x}_\mathrm{b}$. With a new frame, the motion parameters are estimated, and the background is transformed and warped to the current frame. The predicted position of $\mathbf{x}_\mathrm{b}$ in the current frame is $\hat{\mathbf{x}}_\mathrm{c}$, where $\tilde{\hat{\mathbf{x}}}_\mathrm{c} = s\mathbf{\Gamma}^{-1}\tilde{\mathbf{x}}_\mathrm{b}$. As described in Sect. 2.2.2, if there is no occlusion and/or uncovered background (boundary can be regarded as the cases of an occlusion and/or uncovered background), there should be a corresponding pixel of $\mathbf{x}_\mathrm{b}$ in the current frame. The position of the corresponding pixel is assumed to be Gaussian distributed about $\hat{\mathbf{x}}_\mathrm{c}$ and is expressed as

$$p(\mathbf{x}_\mathrm{c}|\hat{\mathbf{x}}_\mathrm{c}) = \mathcal{N}(\mathbf{x}_\mathrm{c}; \hat{\mathbf{x}}_\mathrm{c}, \mathbf{R}). \tag{14}$$

Accordingly, there is a validation region $\mathcal{A}_{\hat{\mathbf{x}}_\mathrm{c}}$ about $\hat{\mathbf{x}}_\mathrm{c}$ in the current frame. The pixels in this validation region are the feasible corresponding pixel of $\mathbf{x}_\mathrm{b}$. The positions that are not the corresponding position of $\mathbf{x}_\mathrm{b}$ 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 $\mathbf{x}_\mathrm{c} = \mathbf{x}_{\mathrm{c}i}$ in the current frame and the predicted position $\hat{\mathbf{x}}_\mathrm{c}$, where $\mathbf{x}_{\mathrm{c}i} \in \mathcal{A}_{\hat{\mathbf{x}}_\mathrm{c}}, i = 1,\ldots,n$.

$\theta_i \triangleq \{\mathbf{x}_{\mathrm{c}i}$ is the corresponding position of $\mathbf{x}_\mathrm{b}.\}$
$\theta_0 \triangleq$ none of the positions is the corresponding position of $\mathbf{x}_\mathrm{b}\}$.

$P(\theta_i)$ is the probability of event θ_i that $\mathbf{x}_{\mathrm{c}i}$ is the corresponding position of $\mathbf{x}_\mathrm{b}$ in the current frame; and $P(\theta_0)$ is the probability of the event that there is no corresponding position of $\mathbf{x}_\mathrm{b}$ in the current frame. $\sum_{i=0}^{n} P(\theta_i) = 1$. In [3], BarShalom gives the details of the derivation of each $P(\theta_i)$,

$$P(\theta_i) = \frac{e_i}{b + \sum_{j=1}^{n} e_j}, i = 1,\ldots,n;$$
$$P(\theta_0) = \frac{b}{b + \sum_{j=1}^{n} e_j}, \tag{15}$$

where $e_i \triangleq \exp\{-\frac{1}{2}(\mathbf{x}_{ci} - \hat{\mathbf{x}}_c)^T\mathbf{R}^{-1}(\mathbf{x}_{ci} - \hat{\mathbf{x}}_c)\}$, and $b \triangleq (2n\pi^2/\gamma)(1 - P_D P_G)/P_D$, P_G is the probability that the corresponding position of $\mathbf{x}_b$ will fall within the validation region $\mathcal{A}_{\hat{\mathbf{x}}_c}$, and P_D is the detection probability. When parameter γ is given, $P(\theta_i)$ is mainly decided by the Mahalanobis distance $D_{\mathbf{x}_{ci},\hat{\mathbf{x}}_c} = (\mathbf{x}_{ci} - \hat{\mathbf{x}}_c)^T\mathbf{R}^{-1}(\mathbf{x}_{ci} - \hat{\mathbf{x}}_c)$ from a certain position $\mathbf{x}_{ci}$ to the predicted position $\hat{\mathbf{x}}_c$. 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 that have already been learned. 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 the position corresponds to $\mathbf{x}_b$ in the current frame, find the plausible ω_{j_i} it belongs to:

$$\begin{aligned}\omega_{j_i} &= \arg\max_{0\le j\le m} P(\omega_j|I(\mathbf{x}_{ci}),\theta_i)\\ &= \arg\max_{0\le j\le m} p(I(\mathbf{x}_{ci})|\omega_j,\theta_i)P(\omega_j|\theta_i)P(\theta_i)\\ &= \arg\max_{0\le j\le m} p(I(\mathbf{x}_{ci})|\omega_j)P(\omega_j),\end{aligned} \tag{16}$$

where $p(I(\mathbf{x}_{ci})|\omega_j,\theta_i) = p(I(\mathbf{x}_{ci})|\omega_j)$ and $P(\omega_j|\theta_i) = P(\omega_j)$. $P(\omega_j)$ can be approximated by the proportion of the time duration of the pixel belonging to the individual Gaussian. When we do not assure the prior probability, we usually assume that $P(\omega_j)$ is equal and satisfies

$$\kappa + \sum_{j=1}^{m} P(\omega_j) = 1. \tag{17}$$

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

$$\begin{aligned}(\mathbf{x}_c^*,\omega_j^*) &= \arg\max_{\mathbf{x}_{ci}\in\mathcal{A}_{\hat{\mathbf{x}}_c},1\le j_i\le m} P(\omega_{j_i}|I(\mathbf{x}_{ci}),\theta_i)\\ &= \arg\max_{\mathbf{x}_{ci}\in\mathcal{A}_{\hat{\mathbf{x}}_c},1\le j_i\le m} \frac{p(I(\mathbf{x}_{ci})|\omega_{j_i},\theta_i)P(\omega_{j_i},\theta_i)}{p(I(\mathbf{x}_{ci}),\theta_i)}\\ &= \arg\max_{\mathbf{x}_{ci}\in\mathcal{A}_{\hat{\mathbf{x}}_c},1\le j_i\le m} \frac{p(I(\mathbf{x}_{ci})|\omega_{j_i})P(\omega_{j_i})P(\theta_i)}{p(I(\mathbf{x}_{ci}))}\\ &= \arg\max_{\mathbf{x}_{ci}\in\mathcal{A}_{\hat{\mathbf{x}}_c},1\le j_i\le m} \frac{p(I(\mathbf{x}_{ci})|\omega_{j_i})P(\theta_i)}{p(I(\mathbf{x}_{ci}))}\end{aligned} \tag{18}$$

under the constraint

$$p(I(\mathbf{x}_c^*)|\omega_j^*)P(\omega_j^*)P(\theta_{i^*}) = \frac{1}{\sqrt{2\pi}\sigma_j^*}\exp\left(-\frac{(I(\mathbf{x}_c^*) - \bar{I}(\omega_j^*))^2}{2\sigma_j^{*2}}\right)\cdot\frac{1-\kappa}{m}\cdot P(\theta_{i^*}) \ge \frac{1}{L}\cdot\kappa\cdot P(\theta_0).$$

Table 1. The background restoration and adaptation algorithm

Begin Algorithm (for a certain pixel $\mathbf{x}_b$)
- Initialization with the first frame. NumberOfGaussian ← 1; Gaussian[1].Mean ← PixelValue of frame 1; Gaussian[1].Variance = σ^2.
- FOR Frame = 2 TO N
  - Motion compensation and obtaining $\hat{\mathbf{x}}_c$.
  - Find $(\mathbf{x}_c^*,\omega_j^*)$ (Eq. 18), where $\mathbf{x}_{ci} \in \mathcal{A}_{\hat{\mathbf{x}}_c}; 1 \le j \le$ NumbrOfGaussian
  - GaussianNumber = ω_j^*; Value = $I(\mathbf{x}_c^*)$
  - IF $D_i > D$(Eq. 19) THEN
  - NumberOfGaussian++ ; Gaussian[NumberOfGaussian]. Mean← $I(\hat{\mathbf{x}}_c)$; Gaussian[NumberOfGaussian].Variance $= \sigma^2$.
  - ELSE
  - Gaussian[GaussianNumber].Count++;Update the parameters of Gaussian[GaussianNumber] with Value.
  - END IF
  - Find Gaussian[j]: $\max_{1\le j\le NumberOfGaussian}$ Gaussian[j]. Count/Gaussian[j].Variance
  - Update the background map
- END FOR

End Algorithm

The corresponding distance constraint is

$$|I(\mathbf{x}_c^*) - \bar{I}(\omega_j^*)| \le \sqrt{-2\sigma_j^{*2}\left(\ln\left(\frac{\sqrt{2\pi}\kappa m}{1-\kappa}\right) + \ln(\sigma_j^*) + \ln P(\theta_0) - \ln P(\theta_{i^*})\right)} = D, \tag{19}$$

where D is a dissimilarity threshold. 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 the constraint cannot be satisfied, it means that no corresponding pixel of $\mathbf{x}_b$ is 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 a fatal error when using the SDG model to detect the foreground.

Referring to Fig. 1, the background distribution is a narrow Gaussian with a high prior probability $P(B)$. We regard ω_j as having a higher frequency and lower standard deviation corresponding to the background.

3 Applications and experimental results

Two applications have been developed based on the approach we propose. The performances of the background map restoration and adaptation and foreground detection are evaluated with a sequence of images that are extracted from a moving handheld video camera. First, with several initial frames, the background map is restored, and the parameters are learned,

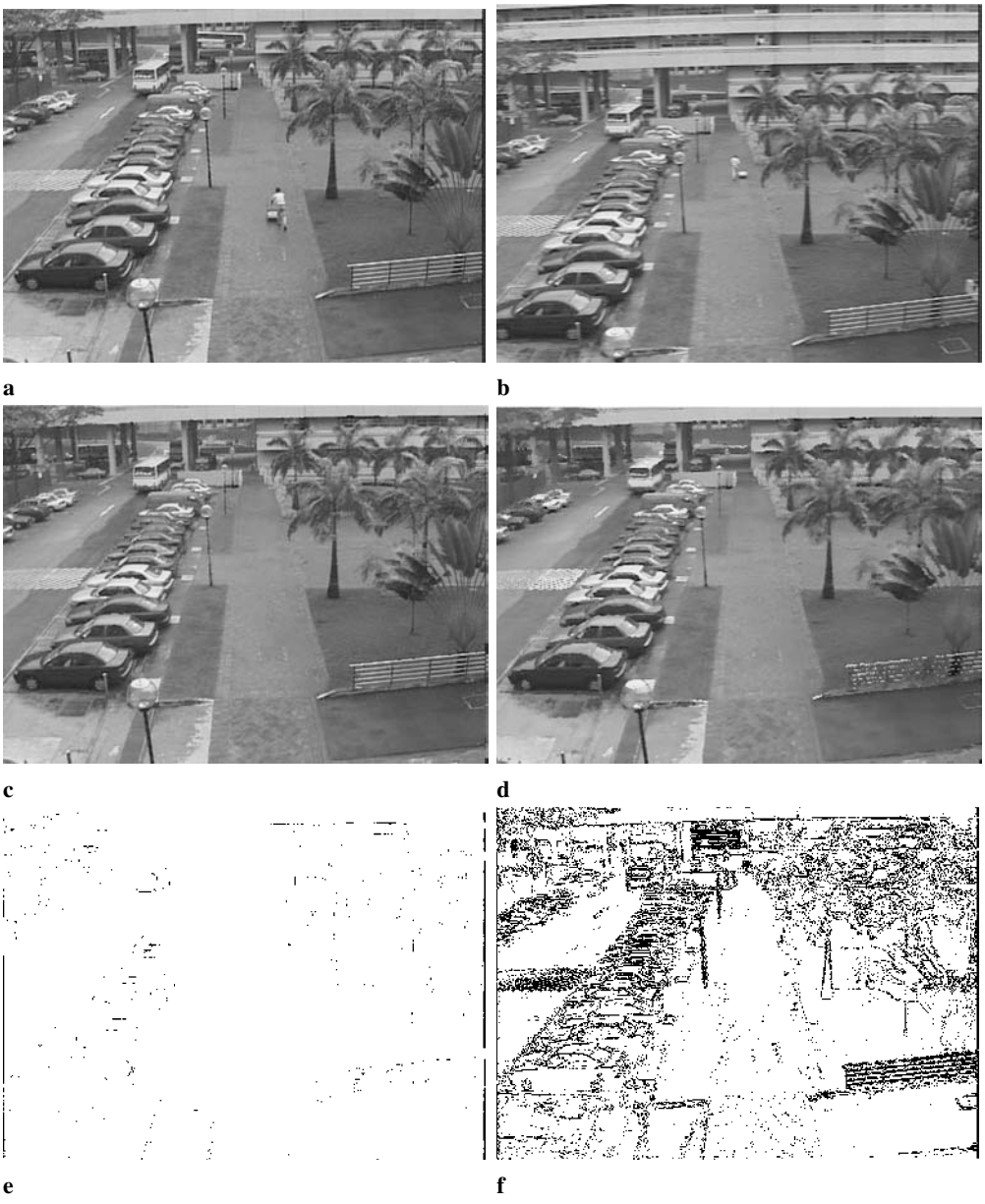


Fig. 4a–f. Background restoration and adaptation. Frame 1 **a** and frame 25 **b** of an image sequence with image resolution of 384×288, extracted from a moving handheld video camera. **c** Background restoration and adaptation based on our approach at frame 25. **d** Background restoration and adaptation after motion compensation directly at frame 25. The error maps between the restored background images **e** and an ideal background image **f**

even though no background map is available in advance. In the frames that follow, the foreground is detected and the background map is adapted at the same time. The experimental results of these two parts will be introduced in Sects. 3.1 and 3.2, respectively. In Sect. 3.3, we will describe the application of the SDG model for the moving object detection with a pan-tilt camera.

3.1 Background map restoration and adaptation

Figure 4 shows the experimental results of the restored and adapted background. We make a comparison of two methods, one is the method we proposed above, another is the one that uses background restoration and adaptation directly after motion compensation without a local search of the corresponding pixel in the current frame. Figure 4a and b are frame 1 and frame 25 (frame rate is 5 frames/s) of the sequence. Figure 4c shows the result of the restored background image $\mathcal{I}_{\mathrm{b}}$ at frame 25 using our method, where $\sigma = 3.0$. Figure 4d shows the restored background image using the direct method; the restored background image is more blurred than the one using our approach. Figure 4e and f are the error maps between the restored background images and an ideal background image for both methods. The results show that there is an accumulation of motion compensation errors, and the efficiency of foreground detection will be impacted when using this direct

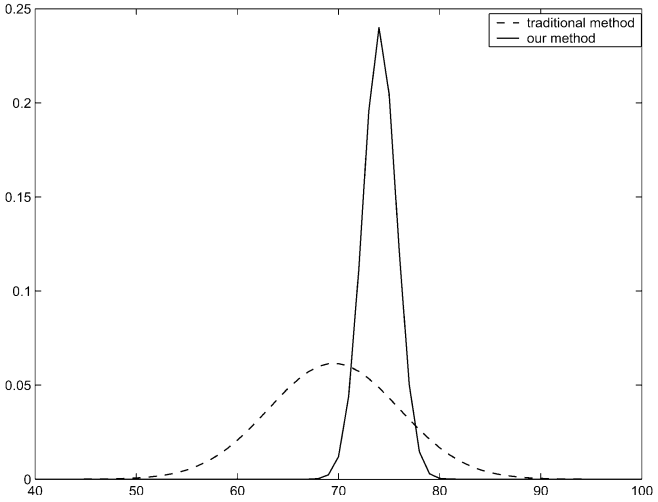


Fig. 5. The fitted background Gaussians according to the approach we proposed and the direct method after motion compensation

method. Figure 5 shows the learned background Gaussians of pixel (165,120) in Fig. 4c and d using the methods mentioned above. The background Gaussian learned directly after motion compensation is more flat than the Gaussian learned by our approach. This is due to the motion compensation errors. Using the flat background Gaussian to detect the foreground will decrease the efficiency of foreground detection.

3.2 Foreground detection of an outdoor scene from a handheld, moving camera

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 6a and b are frames 1 and 136 (frame rate is 5 frames/s) of the sequence. Figure 6c is the result of foreground extraction using background subtraction after affine motion compensation at frame 136, and Fig. 6d is the result of Fig. 6c after a 3×3 morphological operation [4]. We can see that, with this traditional approach, the moving object is submerged by noise. The restored background image, using the technique described above, is illustrated in Fig. 6e. Figure 6f is the result of segmenting the moving objects of frame 136, based on the SDG model.

Figure 7 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, the MD increases dramatically using the traditional method, while our approach exhibits minimal increase in the 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 subsequently fails. 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.

3.3 Indoor active-human tracking with pan-tilt camera

Another application of the foreground detection based on the SDG model is an indoor active-human tracking system with a pan-tilt camera. The coordinate system of the pan-tilt unit is illustrated in Fig. 8. Three coordinate systems are defined in 3-D space: the camera coordinate system, whose origin $\mathbf{O}_{\mathrm{c}}$ coincides with the optical center of the camera; the reference coordinate system, whose origin $\mathbf{O}_{\mathrm{r}}$ is defined as the rotation center of the pan-tilt unit; and the world coordinate system. We assume that the world coordinate system and the camera coordinate system, which is the original position of the camera with no pan or tilt, are aligned.

A calibrated camera with intrinsic parameters $\mathbf{C}$ projects a 3-D point $\mathbf{P}_{\mathrm{c}} = [X, Y, Z]^T$ in the camera coordinate system into a 2-D image frame pixel $\mathbf{p}_{\mathrm{p}} = [u, v]^T$ by

$$s\tilde{\mathbf{p}}_{\mathrm{p}} = \mathbf{\Gamma}_{\mathrm{pro}}\tilde{\mathbf{P}}_{\mathrm{c}}, \tag{20}$$

where

$$\mathbf{\Gamma}_{pro} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} fk_x & fk_x \cos\theta & u_0 & 0 \\ 0 & fk_y/\sin\theta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{21}$$

Assuming that there is no shearing in the u- and v-axes, Eq. 21 can be rewritten as

$$\mathbf{\Gamma}_{\mathrm{pro}} = \begin{bmatrix} A & 0 & u_0 & 0 \\ 0 & B & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{22}$$

Due to the noncoincidence, the optical center is displaced from the rotation center by ρ_Y (in the Y-direction) and ρ_Z (in the Z-direction). Hence, there is a translation $\mathbf{\Gamma}_{\mathrm{tra}}$ from the camera coordinates to the reference coordinates. $\mathbf{P}_{\mathrm{r}}$ in the reference coordinate system is satisfied by

$$\tilde{\mathbf{P}}_r = \mathbf{\Gamma}_{tra}\tilde{\mathbf{P}}_{\mathrm{c}} \tag{23}$$

Let the camera position before any pan(α) or tilt(β) be the original position. For a certain point $\mathbf{P}_w$ in the world coordinate system, its original coordinates p_{p1} and current coordinates (after pan or tilt) p_{p2} in the image frame can be described as

$$s_1\tilde{\mathbf{p}}_{\mathrm{p1}} = \mathbf{\Gamma}_{\mathrm{pro}}\tilde{\mathbf{P}}_w \tag{24}$$

and

$$s_2\tilde{\mathbf{p}}_{\mathrm{p2}} = \mathbf{\Gamma}_{\mathrm{pro}}\mathbf{\Gamma}_{\mathrm{tra}}^{-1}\mathbf{\Gamma}_{\mathrm{tilt}}\mathbf{\Gamma}_{\mathrm{pan}}\mathbf{\Gamma}_{\mathrm{tra}}\tilde{\mathbf{P}}_w. \tag{25}$$

Getting rid of $\tilde{\mathbf{P}}_w$ from Eq. 24 and Eq. 25, we obtain the 2-D transformation equations where

$$\Delta C = \frac{A(-\rho_Y \sin\beta + \rho_Z \cos\alpha \cos\beta - \rho_Z)}{Z}; \tag{28}$$

$$\Delta D = \frac{A^2 \rho_Z \sin\alpha}{Z}; \tag{29}$$

$$\Delta E = \frac{AB(\rho_Y \cos\beta + \rho_Z \cos\alpha \sin\beta - \rho_Y)}{Z}. \tag{30}$$

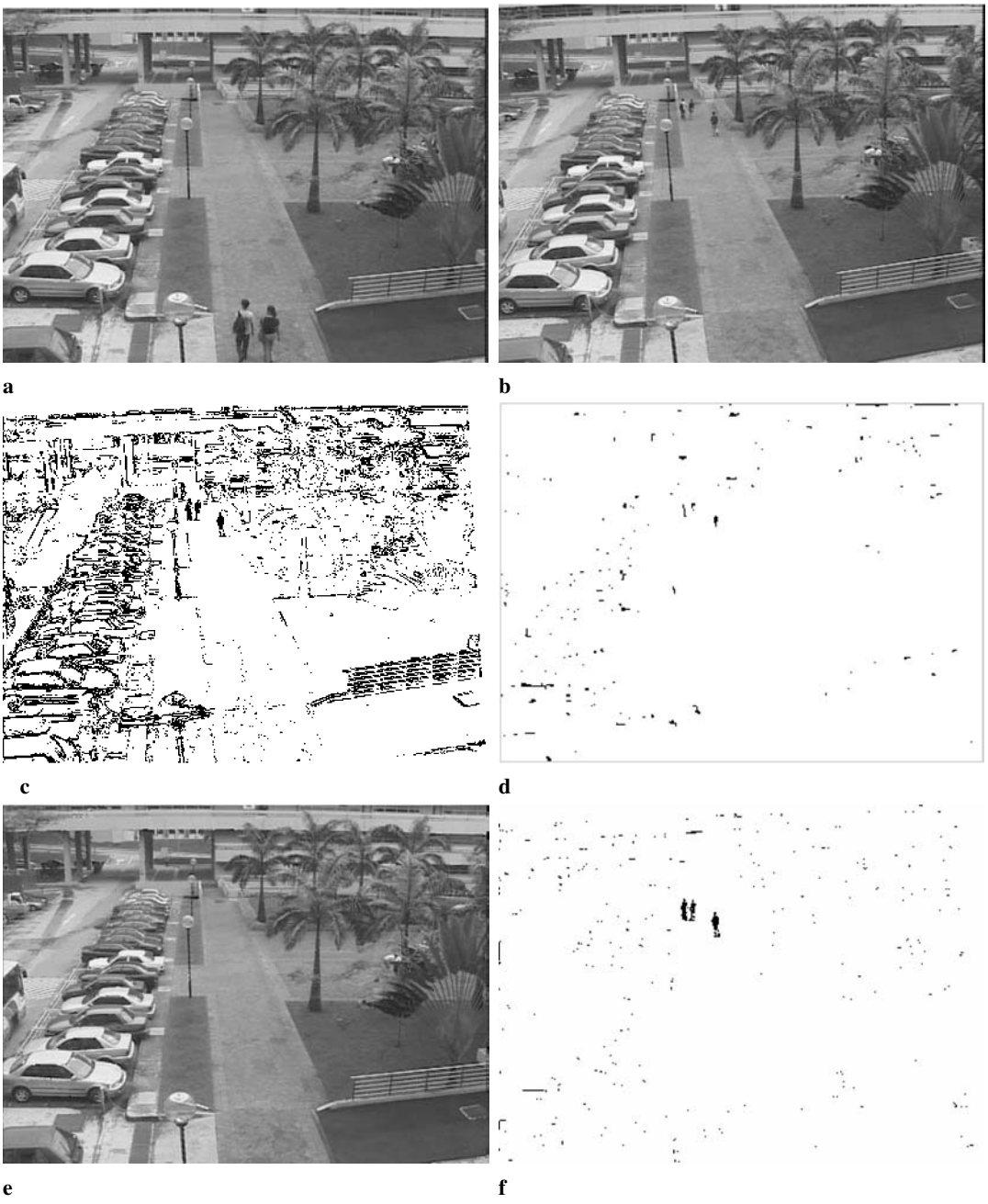


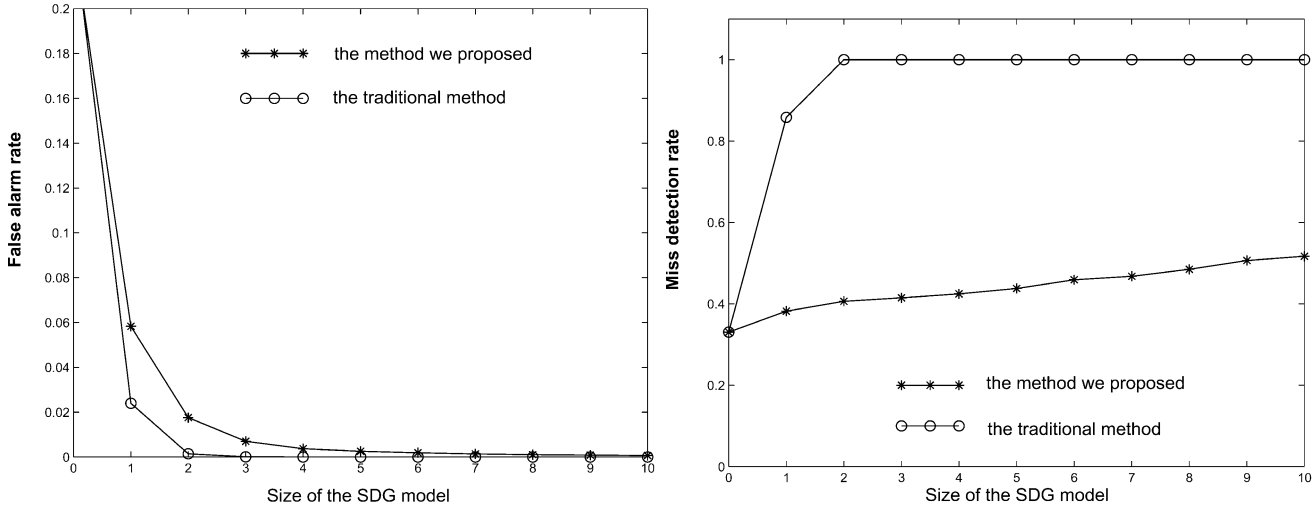
Fig. 6a–f. Background restoration, adaptation, and foreground detection. Frame 1 **a** and frame 136 **b** of an image sequence, extracted from a moving handheld video camera. **c** Foreground detection using background subtraction after affine motion compensation. **d** Result of **c** after morphological operation. **e** Restored background at frame 136. **f** Moving targets correctly extracted (at frame 136) based on our approach

In some papers, the pan and tilt are assumed to rotate about the optical center (that is, $\rho_Y = \rho_Z = 0$). This may be applicable if the distance from the camera to the scene is much larger than the depth of the scene itself (that is, $Z \gg \Delta Z$). If such is the case, then Eq. 26 and Eq. 27 can be simplified to

$$x_1 = \frac{\cos\alpha x_0 + A\sin\alpha}{-\frac{1}{A}\sin\alpha\cos\beta x_0 - \frac{1}{B}\sin\beta y_0 + \cos\alpha\cos\beta} \quad (31)$$

$$y_1 = \frac{-\frac{B}{A}\sin\alpha\sin\beta x_0 + \cos\beta y_0 + B\cos\alpha\sin\beta}{-\frac{1}{A}\sin\alpha\cos\beta x_0 - \frac{1}{B}\sin\beta y_0 + \cos\alpha\cos\beta}. \quad (32)$$

However, for our indoor scene, the above assumption is not applicable. ΔC, ΔD, and ΔE will contribute to registration errors. As an illustration of the potential errors that exist, an experiment was conducted involving nine 500-frame sequences of clean background. These sequences were extracted from nine pan/tilt positions at pan angles of $-25.7°, 0°$, and $25.7°$ and tilt angles of $-15.42°, 0°$, and $15.42°$. Each sequence of background was learned as one Gaussian. The nine sequences were warped to the same coordinate system according to Eqs. 31 and 32. The Gaussians covering the same pixel consisted of an MOG, and all these MOGs formed the background



a **b**

Fig. 7a,b. Error analysis with the varying size of the SDG model. **a** False-alarm rate with the varying size of the SDG model. **b** Miss-detection rate with the varying size of the SDG model

$$x_2 = u_2 - u_0 = \frac{\cos\alpha x_1 + A\sin\alpha + \frac{A\rho_Z \sin\alpha}{Z}}{-\frac{1}{A}\sin\alpha\cos\beta x_1 - \frac{1}{B}\sin\beta y_1 + \cos\alpha\cos\beta + \frac{-\rho_Y \sin\beta + \rho_Z \cos\alpha\cos\beta - \rho_Z}{Z}} = \frac{D + \Delta D}{C + \Delta C} \tag{26}$$

$$y_2 = v_2 - v_0 = \frac{-\frac{B}{A}\sin\alpha\sin\beta x_1 + \cos\beta y_1 + B\cos\alpha\sin\beta + \frac{B(\rho_Y\cos\beta + \rho_Z\cos\alpha\sin\beta - \rho_Y)}{Z}}{-\frac{1}{A}\sin\alpha\cos\beta x_1 - \frac{1}{B}\sin\beta y_1 + \cos\alpha\cos\beta + \frac{-\rho_Y\sin\beta + \rho_Z\cos\alpha\cos\beta - \rho_Z}{Z}} = \frac{E + \Delta E}{C + \Delta C} \tag{27}$$

map mosaic. That is, after the transformation, the distribution of the intensity value of each pixel in the background map mosaic was modeled as an MOG. All this work could be done offline. Figure 10a shows the result of the background map mosaic.

In the tracking stage, for each pixel in the current frame, the motion compensation with an arbitrary pan/tilt angle can be calculated according to Eqs. 31 and 32. 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. Figure 9 shows an example of the error analysis. The background map mosaic is constructed from two sequences at $\mathrm{pan} = 0^\circ$ and $\mathrm{pan} = 25.7^\circ$. Figure 9a–f show the errors in the x-direction when the camera positions are at $\mathrm{pan} = 0^\circ, 5.14^\circ, 10.28^\circ, 15.42^\circ, 20.56^\circ$, and 25.7° (tilt is unchanged). The vertical coordinate describes the error between the approximate real position and the calculated position in the x-direction. The error analysis is of pixel-level accuracy. The approximate real position is localized based on the correlation operator. The calculated position is computed according to Eqs. 31 and 32. As Fig. 9 shows, the error is larger when the pan position departs from the two original positions ($\mathrm{pan} = 0^\circ$ and $\mathrm{pan} = 25.7^\circ$). In this test case (PTU-46-17.5 of Directed Perception Inc.), with a resolution step of 3.086 arc minute; $\rho_Y \doteq 0.05m$, $\rho_Z \doteq 0.07m$; $f = 752$), the maximum error is between -4 and 4. Even with the MOG background map mosaic, motion compensation error cannot be eliminated. Using the SDG model, referring to Fig. 10d, moving objects (human subjects) are detected and tracked accurately. Compared with the traditional method (Fig. 10c), the detection based on our approach is able to extract the desired target without significant noise clutter.

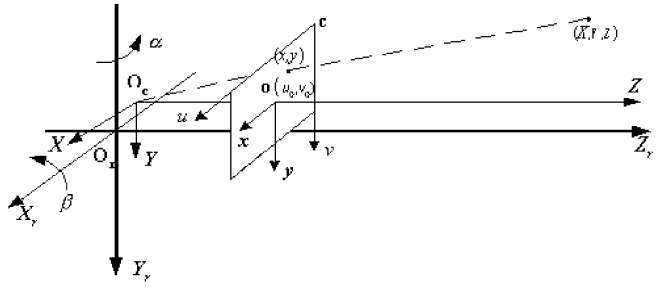


Fig. 8. Image, camera, and reference coordinate system

4 Conclusions

This paper proposes an SDG model that is used to detect the foreground from a nonstationary background. The detection based on the SDG model can maintain the shape of the detected object and shows good results, even when the detection is applied to small moving objects in a highly textured background. With a nonstationary background, an algorithm is proposed for the background restoration and adaptation. The SDG model and the statistically modeled background extend the application of the background subtraction to the moving sensor and make the detection robust, even with approximate motion compensation, noise, or environmental changes. These algorithms are in pixel-wise case; and no iterative computations

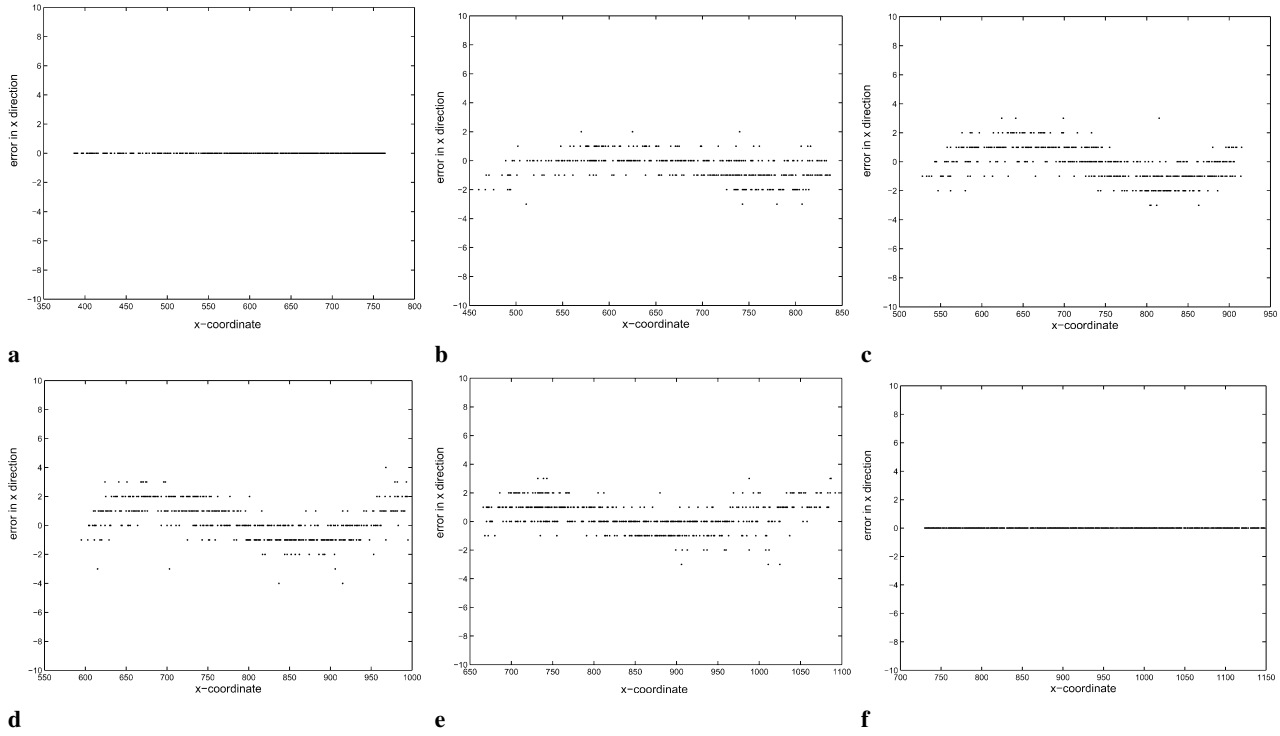


Fig. 9a–f. Position error analysis. **a** Errors when pan$= 0^{\circ}$; **b** errors when pan$= 5.14^{\circ}$; **c** errors when pan$= 10.28^{\circ}$; **d** errors when pan=15.42°; **e** errors when pan=20.56°; and **f** errors when $pan = 25.7^{\circ}$

Fig. 10a–d. Human detection and tracking with pan-tilt camera. **a** Background map mosaic; **b** a sequence of images obtained from a pan-tilt camera; **c** extracted foreground using background subtraction after motion compensation; and **d** extracted foreground (moving human) using background subtraction based on SDG model

are required. As such, they are suitable for parallel implementations for real-time considerations.

References

1. Araki S, Matsuoka T, Takemura H, Yokoya N. (1998) Real-time tracking of multiple moving objects in moving camera image sequences using robust statistics. In: Proceedings of the 14th IEEE International Conference on Pattern Recognition, pp 1433–1435
2. Barron J, Fleet D, Beauchemin S (1994) Performance of optical flow techniques. Int J Comput Vision 12(1):42–77
3. BarShalom Y, Fortmann TE (1989) Tracking and data association. Academic Press, New York
4. Murray D, Basu A (1994) Motion tracking with an active camera. IEEE Trans Patt Anal Mach Intell 16(5):449–459
5. Eledath J, McDowell L, Hansen M, Wixson L, Pope A, Gendel G (1998) Real-time fixation, mosaic construction and moving object detection from a moving camera. In: Proceedings of the IEEE International Conference on Application of Computer Vision, pp 284–285
6. Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. In: Proceedings of the 6th European Conference on Computer Vision, pp 751–767
7. Friedman N, Russell S (1997) Image segmentation in video sequences: A probabilistic approach. In: Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence, pp 175–181
8. Hansen M, Anandan P, Dana K, van der Wal G, Burt P. (1994) Real-time scene stabilization and mosaic construction. In: Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision, pp 54–62
9. Haritaoglu I, Davis LS, Harwood D (1998) W4: who? when? where? what? a real time system for detecting and tracking people. In: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp 222–227
10. Zoghlami I, Faugeras O, Deriche R (1997) Using geometric corners to build a 2D mosaic from a set of images. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp 420–425
11. Jain R, Martin WN, Aggarwal JK (1979) Segmentation through the detection of changes due to motion. Comput Graph Image Process 11:13–34
12. Kanade T, Collins RT, Lipton AJ (1998) Advances in cooperative multi-sensor video surveillance. In: Proceedings of DARPA Image Understanding Workshop (IUW), pp 3–24
13. Ren Y, Chua CS, Ho YK (2000) Multiple-model based human tracking. In: Proceedings of the IEEE International Conference on Visual Interface, pp 280–285
14. Rittscher J, Joga J, Blake A (2000) A probabilistic background model for tracking. In: Proceedings of the European Conference on Computer Vision, pp 336–351
15. Rowe S, Blake A (1996) Statistical mosaic for tracking. Image Vision Comput 14:549–564
16. Stauffer C, Grimson W (1999) Adaptive background mixture models for real-time tracking. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp 246–252
17. Toyama K, Krumm J, Brumitt B,Meyers B (1999) Wallflower: Principles and practice of background maintenance. In: Proceedings of IEEE International Conference on Computer Vision, pp 255–261
18. Wren CR, Azarbayejani A, Darrell T, Pentland A (1997) Pfinder: Real-time tracking of the human body. IEEE Trans Patt Anal Mach Intell 19(7):780–785
19. Yalamanchili S, Martin WN, Aggarwal JK (1982) Extraction of moving object description via differencing. Comput Graph Image Process 18:188–201

Ying Ren received her Bachelor and Masters degrees, both in electrical engineering, from Tianjin University, Tianjin, P. R. China in 1991 and 1994, respectively. Currently, she is a PhD student in the school of electrical and electronic engineering, Nanyang Technological University, Singapore. Her research interests include motion segmentation, visual tracking, multiple cues integration, pattern recognition, and visual surveillance.

Chin-Seng Chua received a BEng degree from Nanyang Technological University, Singapore, in 1991, and PhD degree in computer vision from Monash University, Australia, in 1995. From 1995 to 1997, he was with the Defence Science Organisation and is currently a faculty member of Nanyang Technological University. His research interests include computer vision, face recognition, and surveillance.

Yeong-Khing Ho received a BSc (Honours) from Strathclyde University, UK, and MSc (London) from the Imperial College of Science, Technology and Medicine, UK. He also received a PhD from the Nanyang Technological University, Singapore, and is currently a faculty member of the University. His research interests include robotics, automation, and computer vision.