MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

13—27

Foreground Detection with Non-stationary Background

Ying REN, Chin-Seng CHUA*, Yeong-Khing HO
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

Abstract

Background subtraction is a traditional technique for finding moving objects (foreground). With a non-stationary viewing sensor, this approach usually assumes that the motion compensation for the background must be sufficiently accurate. In practice, it is difficult to realized this assumption and the background subtraction algorithm will fail for a moving scene. The problem is further compounded when the moving target to be detected/tracked is small, since the pixel error in motion compensating the background will subsume the small target. This paper proposes a Spatial Distribution of Gaussians (SDG) model to deal with the moving object detection with motion compensation which are only approximately accurate. Unlike the traditional approach, this algorithm integrates not only the visual information at each pixel but also its local spatial information and classifies each pixel based on a statistical model. This approach extends the application of the background subtraction to the case of a moving sensor and is robust even with less accurate motion compensation, noise, or environmental changes. Test cases involving the detection of small moving objects with a highly textured background are demonstrated successfully.

1 Introduction

Motion detection and segmentation is a basic problem in security surveillance, object tracking, autonomous obstacle avoidance and image compression. Background subtraction is a traditional technique for finding moving objects in a sequence of images. It requires that the background scene and the viewing sensor are stationary. With a non-stationary viewing sensor, motion compensation is required to compensate for the motion due to the moving sensor. First, a motion model is assumed and then motion parameters are estimated. The background is registered ideally and the foreground can be detected pixel by pixel. This approach assumes that the motion model is approximated well enough and the parameters of the motion model are accurately estimated [1, 2]. In practice, it is difficult to realize this assumption due to the approximated motion model and computational errors of the parameter estimation. The background image and the current frame cannot warp and register well. It is not trivial for most of the methods of motion detection with mobile viewing sensor. This problem is further compounded when the moving target to be detected/tracked is small,being subsumed by the pixel error in motion compensation. Hence, the background subtraction algorithm will fail to detect moving objects in most practical implementations involving moving sensors. Figure 2 (*a*) and (*b*) show Frame 1 and 25 of an image sequence involving a moving person against a stationary background. This image sequence is extracted from a moving hand-held video camera (moving sensor). To compensate for sensor movement, an affine motion compensation is applied (traditional approach). The objective is to detect and track the moving person, despite the sensor being non-stationary. Referring to Figure 2(*c*) and (*d*), after motion compensation, background subtraction and morphological operation, the target (moving person) disappears.

In this paper, we propose a Spatial Distribution of Gaussians (SDG) model, which is established in the spatial and visual domain. The SDG model is a simplified form of the spatial-visual statistical optimization. The foreground detection based on the SDG model is more robust even with less accurate motion compensation, noise or environmental changes and is also able to detect and track small moving objects in a highly textured background.

The remainder of this paper is arranged as follows. Section 2 introduces the Spatial Distribution of Gaussians (SDG) model and the technique of object detection with a moving viewing sensor. Section 3 describes two applications of the SDG model. One application demonstrates active human detection and tracking with a moving pan-tilt camera in an indoor surveillance system while the other illustrates the mechanism of background restoration,

*Address: Block S1, Nanyang Avenue, Singapore 639798. E-mail: ecschua@ntu.edu.sg

adaptation and moving object detection.

2 Spatial Distribution of Gaussians (S-DG) model

The basic idea of this paper is that we can model the intensity value of each background pixel as a Gaussian distribution, 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 2D parametric transformation, such as affine or projective, in the image plane. Traditional approaches are used to estimated 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 Gaussian distributions in its local spatial region if it is indeed the background; otherwise, it is regarded as foreground.

2.1 Pixel-wise Background Model

In a sequence of images, each pixel is modeled as an independent statistical process, a mixture of Gaussian. Each Gaussian corresponds to the distribution of background and different 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 which compose the Gaussian mixture model. Figure 1 shows a distribution of intensity values $\mathbf{I}(\mathbf{x})$ for a given pixel $\mathbf{x}$ of an image sequence extracted over one hour. We model it as a mixture of Gaussians:

$$p(\mathbf{I}) = p(\mathbf{I}|\mathbf{B})P(\mathbf{B}) + \sum_{i=1}^{c-1} p(\mathbf{I}|\omega_i)P(\omega_i) \qquad (1)$$

where B denotes the background distribution, ω_i denotes the distributions of different moving objects and c denotes the number of classes for that pixel. An on-line learning and adaptive algorithm [3] is developed to obtain the parameters, especially the one pertaining to the background distribution. We established a background map which consist of the entire background Gaussians distributions of every pixel. This background map is adapted frame by frame with each new incoming frame.

2.2 Spatial Distribution of Gaussians (S-DG) model

For each pixel $\mathbf{I}(\mathbf{x}_c)$ in the current image, motion compensation and distortion correction is applied.

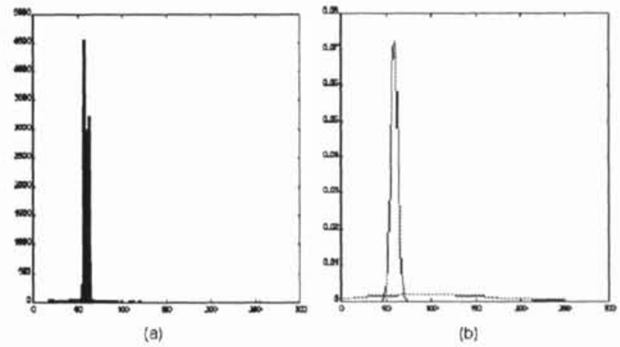

Figure 1: (a) Histogram of intensity values for one pixel of an image sequence over one hour. (b) Background distribution can be fitted with a narrow gaussian and others can be fitted with a Gaussian with large deviation.

Let the predicted position in the background map be $\hat{\mathbf{x}}_b$, where $\hat{\mathbf{x}}_b^T = \mathbf{P}\mathbf{x}_c^T$, and $\mathbf{P}$ is the transformation matrix for motion compensation. We assume that the real position is $\bar{\mathbf{x}}_b$, and $\bar{\mathbf{x}}_b \in \{\mathbf{x}_b\}$, where $\mathbf{x}_b$ is Gaussian distributed and centered on the prediction position $\hat{\mathbf{x}}_b$. This is expressed as

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

where $\mathbf{R}$ is the covariance of positional errors, which include errors due to motion modeling, distortion and parameter estimation. As described in the Section 2.1, for each position $\mathbf{x}_b$ around $\hat{\mathbf{x}}_b$, there is an intensity value distribution (Eq. (1)). The background distribution at position $\mathbf{x}_b$ is:

$$p(\mathbf{I}|\bar{\mathbf{I}}(\mathbf{x}_b)) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp(-\frac{(\mathbf{I} - \bar{\mathbf{I}}(\mathbf{x}_b))^2}{2\sigma^2}) \qquad (3)$$

where $\bar{\mathbf{I}}(\mathbf{x}_b)$ and σ are the mean and standard deviation of the background distribution at $\mathbf{x}_b$. These background distributions consist of a background map. For a pixel $\mathbf{I}(\mathbf{x})$ in the current frame, the corresponding distribution in the background map should satisfy:

$$\mathbf{x}_b^* = \arg\max_{\mathbf{x}_b} p(\mathbf{x}_b|\hat{\mathbf{x}}_b)^{\omega} p(\mathbf{I}(\mathbf{x}_c)|\bar{\mathbf{I}}(\mathbf{x}_b))^{(1-\omega)} \qquad (4)$$

and

$$p(\mathbf{x}_b|\hat{\mathbf{x}}_b) \geq \mathbf{T_1} \qquad (5)$$

$$p(\mathbf{I}(\mathbf{x}_c)|\bar{\mathbf{I}}(\mathbf{x}_b)) \geq \mathbf{T_2} \qquad (6)$$

where $\mathbf{T_1}$ and $\mathbf{T_2}$ are decided by the learnt *prior* probability of visual and spatial background and the conditional probability of the non-background distribution according to the MAP decision rule. ω is the weight to decide the importance of the two components.

The foreground detection algorithm according to the upper criterion is computationally intense. Here, we use its simplified form, Spatial Distribution of Gaussian (SDG) model. The SDG model is established based on the background Gaussian distribution for each pixel - background map. For pixel $\mathbf{I}(\mathbf{x}_c)$ in the current frame, the center of its SDG is $\hat{\mathbf{x}}_b$. According to Eq. (2), the size of the SDG model is decided by the spatial distance d_s, where

$$d_s = (\mathbf{x}_b - \hat{\mathbf{x}}_b)^T \mathbf{R}^{-1} (\mathbf{x}_b - \hat{\mathbf{x}}_b). \quad (7)$$

Using Eq. (3),the visual distance d_v, is defined as

$$d_v = \frac{(\mathbf{I}(\mathbf{x}_c) - \bar{\mathbf{I}}(\mathbf{x}_b))^2}{\sigma^2}. \quad (8)$$

The corresponding distribution in the background map is decided by the following three conditions:

$$d_v \leq \gamma_1 \quad (9)$$

$$d_s \leq \gamma_2 \quad (10)$$

and

$$\mathbf{x}_b^* = \arg\min_{\mathbf{x}_b} d_s \quad (11)$$

where γ_1 and γ_2 can be obtained from the χ^2 distribution table. If no corresponding background distribution can be found in its SDG model, the pixel $\mathbf{I}(\mathbf{x}_c)$ is regarded as foreground. Unlike the traditional approach, this algorithm integrates not only the visual information but also the local spatial information and classifies the pixel based on optimal decision rules. With this approach, the detection of small moving objects in a highly textured background works well. Another application of this approach is in background restoration and adaptation. Referring to Figure 2 (*e*), the background (Frame 25) is restored and adapted without the prior knowledge of a "clean" background. With the temporal background adaptation, this approach is more robust to noise and illumination changes.

3 Applications and experiments

3.1 Background restoration, adaptation and moving object detection

Another application of SDG model is the background restoration, adaptation and moving object detection with a moving camera in an outdoor surveillance system. Referring to Figure 2, the camera motion cannot be neglected and the moving object (human) is small. An affine transformation model and least square technique are applied for the estimation of the motion parameters and motion compensation. Figure 2(*c*) is the result of foreground extraction using background subtraction after affine motion compensation at Frame 25, and (*d*) is the result of (c) after a 5×5 morphological operation. We can see that with this traditional approach the moving object is submerged by noise. To overcome the above problem, each pixel is modeled as a mixture of Gaussians (as described in Section 2.1). The parameters of the mixture of Gaussians are learnt on-line and the background is restored and adapted with each incoming image based on the SDG model. Figure 2(*e*) is the restored and adapted background at Frame 25. Figure 2(*f*) is the result of moving object segmentation of Frame 25 based on the SDG.

3.2 Indoor active human tracking with a moving pan-tilt camera

The foreground detection based on the SDG model is applied to an indoor active human tracking system with pan-tilt camera and wide angle lens. Projective transformation model is applied in the motion compensation. The lens distortion and the depth of the scene [4] are taken into consideration for each pixel's SDG model. Referring to Figure 3, moving objects (human subjects) are detected and tracked accurately. Compared with the traditional method (the second row of Figure 3), the detected objects based on our technique (the bottom row of Figure 3) significantly extract the desired target without noise clutter.

4 Conclusions

This paper proposes a SDG model which is based on the theory of optimal decision. It extends the application of the background subtraction to the moving sensor and is robust even with less accurate motion compensation, noise, or environmental changes. The detection based on SDG model can keep the shape of the detected object perfectly, and show good results even when the detection is applied to small moving objects in a highly textured background. The algorithm is a pixel-wise case; no iterative computations are required. As such, it is suitable for parallel implementations for real-time consideration.

References

[1] D. Murray and A. Basu, "Motion Tracking with an Active Camera," IEEE Transactions on Pat-

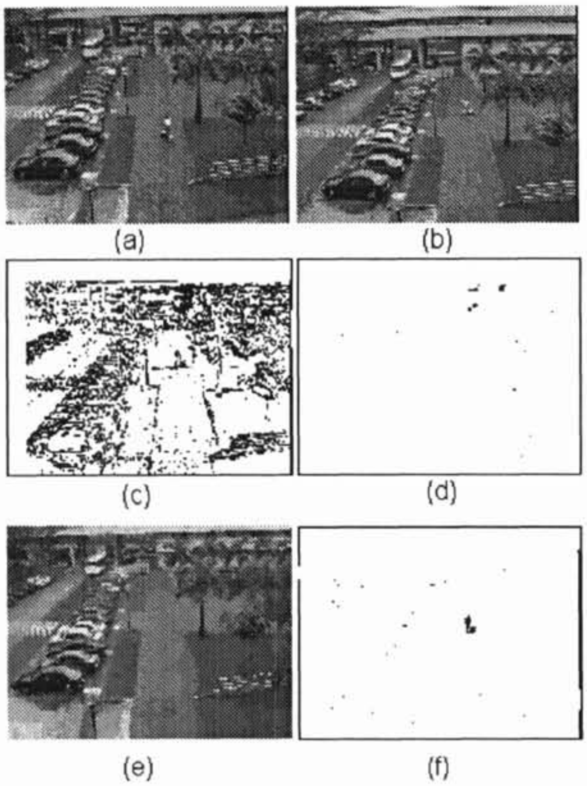

Figure 2: Background restoration, adaptation and moving region segmentation. (a,b) Frame 1 and 25 of an image sequence, extracted from a moving hand-held video camera. (c) Foreground extraction using background subtraction after affine motion compensation (traditional approach)using our approach at Frame 25. (d) Result of (c) after morphological operation. (e) Restored and adapted background at Frame 25. (f) Moving target correctly extracted using background subtraction (at Frame 25) based on our approach.

tern Analysis and Machine Intelligence, 16(5): 449-459, 1994.

[2] S. Araki, T. Matsuoka, H. Takemura and N. Yokoya, "Real-time Tracking of Multiple Moving Objects in Moving Camera Image Sequences Using Robust Statistics," Proceedings of the *14th* International Conference on Pattern Recognition, pages: 1433-1435, 1998.

[3] Y. Ren, C. S. Chua and Y. K. Ho , "Multiple-model Based Human Tracking," Proceedings of the International Conference on Visual Interface, pages: 280-285, 2000.

[4] G. Xu and Z. Y. Zhang, "Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach," Kluwer Academic Publishers, 1996.

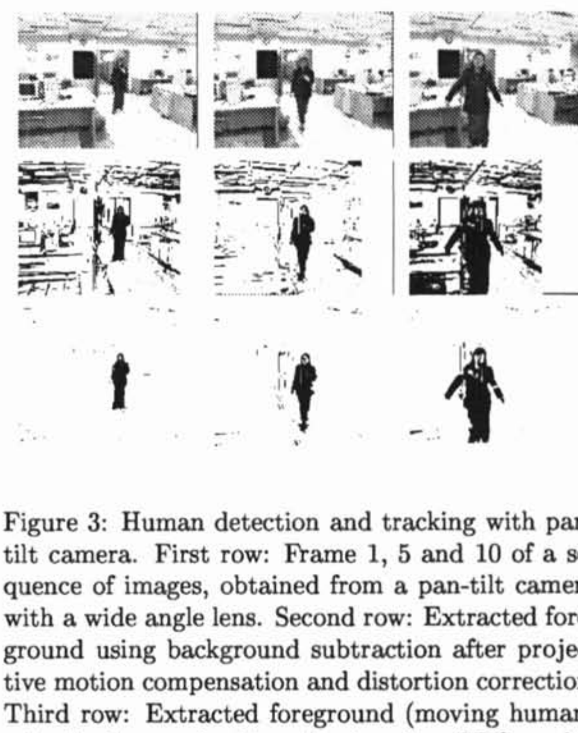

Figure 3: Human detection and tracking with pan-tilt camera. First row: Frame 1, 5 and 10 of a sequence of images, obtained from a pan-tilt camera with a wide angle lens. Second row: Extracted foreground using background subtraction after projective motion compensation and distortion correction. Third row: Extracted foreground (moving human) using background subtraction based on SDG model.