Pattern Recognition Letters 29 (2008) 2145–2155

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Motion detection using Fourier image reconstruction

Du-Ming Tsai *, Wei-Yao Chiu

Department of Industrial Engineering and Management, Yuan-Ze University, 135 Yuan-Tung Road, Nei-Li, Tao-Yuan, Taiwan, ROC

ARTICLE INFO

Article history:
Received 15 August 2007
Received in revised form 18 June 2008
Available online 17 August 2008

Communicated by S. Dickinson

Keywords:
Motion detection
Surveillance
Foreground segmentation
Fourier transforms

ABSTRACT

In video surveillance, detection of moving objects from an image sequence is very important for object tracking, activity recognition and behavior understanding. The conventional background subtraction suffers from slow updating of environmental changes, and temporal difference cannot accurately extract the moving object boundaries. In this paper, a Fourier reconstruction scheme for motion detection is proposed. A series of consecutive 2D spatial images along the time axis are first reorganized as a series of 2D spatial–temporal images along a spatial axis. In each of the 2D spatial–temporal images, a static background region forms a vertical line pattern, and a moving object creates an irregular, non-vertical structure in the image. Fourier transforms are applied to remove the vertical line pattern (i.e. the background) and retain only the foreground in the reconstructed image. The proposed method is a global approach that identifies the moving objects based on structural variations in the whole patterned image. It is therefore very robust to accommodate noise and local gray-level variations. It can well extract the shapes of foreground objects at various moving speeds, and is very responsive to dynamic environments. High computational cost is the major drawback of the proposed method. However, it can still achieve 11 frames per second for small images of size 150×200.

1. Introduction

Detection of moving objects in image sequences is very important for the success of object tracking, incident detection, activity recognition and behavior understanding in video surveillance. Motion detection aims at segmenting foreground pixels corresponding to moving objects from the background in a scene image. There are two main requirements for an effective motion detection algorithm: detection accuracy in shape, and responsiveness to environmental changes. The shape of a moving object should be extracted as complete as possible so that the subsequent high-level recognition applications can be reliably executed. The motion detection should also be highly responsive to dynamic environments such as gradual changes in outdoor light, sudden on/off switching of indoor lights, door opening/closing and parking cars on the sidewalk.

Temporal difference and background subtraction are two commonly-used techniques to segment moving objects in image sequences from a static camera. Temporal difference (Lipton et al., 1998; Wang and Brandstein, 1998) calculates the difference of pixel features between two consecutive scene frames in an image sequence. It is very computationally efficient, and well accommodates environmental changes, but generally can only extract partial shapes of moving objects. Background subtraction detects moving objects in an image by evaluating the difference of pixel features between the current scene image and the reference background image. This approach is very computationally fast, but is very sensitive to environmental changes without adaptively updating the reference background. In this study, we propose a Fourier reconstruction scheme for moving object detection in image sequences. It is as responsive as the temporal difference method to environmental changes in both indoor and outdoor scenes, and yet gives good shape boundaries of moving objects.

In order to make background subtraction adaptive to environmental changes, many background updating strategies were proposed. Piccardi (2004) presented a review of background subtraction techniques with emphasis on background modeling algorithms for detecting moving objects from a static camera. Background model updating methods are generally based on the analysis of the gray-level (or color) histogram taken by each individual pixel over a limited number of recent frames. Wren et al. (1997) modeled each pixel of the background over time with a single Gaussian distribution. Single Gaussian updating models were also adopted by many researchers (Olson and Brill, 1997; Eveland et al., 1998; Kanade et al., 1998; Cavallaro and Ebrahimi, 2001) for background subtraction.

A more robust background modeling is to represent each pixel of the background image over time by a mixture of Gaussians, which was originally proposed by Stauffer and Grimson (1999, 2000). The background model can deal with multimodal distributions caused by shadows, swaying branches, etc. It can handle slow

* Corresponding author. Fax: +886 (03) 463 8907.
E-mail addresses: iedmtsai@saturn.yzu.edu.tw, s968902@mail.yzu.edu.tw (D.-M. Tsai).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.08.005

lighting changes by slowly adapting the parameter values of the Gaussians. Since the estimation of Gaussian parameter values for each pixel in the image using standard algorithms such as Expectation Maximization is computationally prohibited, recursive updating using a simple linear adaptive filter is applied for real-time implementation. The background model is generally considered as a linear combination of the current background and the current scene image with a specific learning rate. A slow updating of the background model cannot promptly respond the background changes, whereas a fast updating may absorb the slow moving object as the background. It may fail to detect multiple moving objects with different moving velocities in a scene, e.g. slow walking pedestrians and fast moving vehicles on streets.

KaewTrakulPong and Bowden (2001) proposed updating formulations of the Gaussian mixture to improve the slow learning at the beginning of the updating process. Lee (2005) presented an adaptive learning rate of the background model to improve the updating convergence without compromising model stability. McFarlane and Schofield (1995) and Manzanera and Richefeu (2007) used a recursive approximation of the temporal median for background estimation. The background is updated with a constant increment/decrement of 1 based on the sign of the gray-level difference between the current frame and the background. The method requires only a small cost in memory consumption and computational complexity. Manzanera and Richefeu showed that if the increment/decrement depends linearly on the gray-level difference, the updating model is equivalent to the classical exponential filter (i.e. moving average). The variance of each pixel over time used to detect a foreground is also increased/decreased by 1 at each frame. The statistics of the background must be constantly updated with a pre-determined frequency. This approach may not be fast enough to update radical changes of the background. A moving object may not be promptly identified during the updating. In a background updating model, the representation of the true current background may be degraded after a very long-run due to the accumulated errors of background estimation in each update operation.

Instead of modeling the features of each pixel by Gaussian distributions, Elgammal et al. (2002) and Ianasi et al. (2005) evaluated the probability of a background pixel using kernel density estimation from very recent historical samples in the image sequence. Elgammal et al. (2003) further used Fast Gauss Transform (Greengard and Strain, 1991) to improve the computation of the Gaussian kernel density estimation.

The background subtraction and temporal difference techniques reviewed above basically identify the foreground regions based on gray-level (or color) statistics or difference of each individual pixel over time. Motion detection from the statistical viewpoint on separated pixels causes the background updating models less responsive to radical changes of the environment, and the temporal difference methods less accurate to extract the complete shape of a moving object. In this paper, we propose a Fourier reconstruction scheme from the structural aspect of a global image pattern for motion detection in video images.

By considering the 2D spatial images in the x–y plane over time t as a 3D x–y–t spatial–temporal image, one cross-sectional slice of the 3D image in the spatial-axis of y constructs a 2D spatial-temporal image in the x–t plane. If there are no moving objects over a limited number of consecutive frames, the x–t slice will appear as a vertical line pattern in the image. It can then be considered as a homogeneous structure texture. Conversely, if there are moving objects in the stationary background, the resulting x–t slice will show non-vertical variations in the image, and the homogeneity of the vertically structured texture is violated. The vertical line pattern in the x–t sliced image can be removed by setting the corresponding high-energy frequency components in the Fourier domain image to zero and back-transforming to the spatial domain image. In the reconstructed image, the homogeneous structure texture representing the stationary background in the original x–t image will have an approximately uniform gray-level, whereas the foreground regions associated with moving objects will be distinctly preserved. By scanning all 2D x–t sliced images in the spatial y-axis, the moving objects in all consecutive frames under detection can be well constructed and extracted. The proposed Fourier reconstruction scheme is less computationally efficient, compared with the background subtraction and temporal difference methods. However, it can effectively and responsively extract the shapes of moving objects under dynamic environments. Niyogi and Adelson (1994) presented saptio-temporal patterns in the x–y–t space for analyzing walking persons in video images. Their method was based on the observation that a translating head generates a slanted stripe, and walking legs generate braids in the x–t slices. It mainly focused on the gait recognition of individual walkers, but not the foreground segmentation of moving objects.

This paper is organized as follows: Section 2 presents the Fourier reconstruction process for static background removal. Section 3 describes the experimental results on three sets of video scenarios, and compares the performance of the proposed method with that of temporal difference and background subtraction. Conclusions are given in Section 4.

2. Fourier reconstruction for background removal

As aforementioned, background subtraction techniques cannot promptly respond changes of the environment such as moving chairs, placing newspapers on a table, opening/closing a curtain in a living room, and parking and moving cars in a street. The proposed method is based on the temporal consistency of pixels in the same location over a limited number of most-recent frames.

2.1. Fourier transform of video images

Let $f_t(x,y)$ be the 2D spatial scene image of size $R \times C$ at frame t. Given a series of N consecutive frames, a 3D spatial–temporal image of size $R \times C \times N$ can be constructed as $f(x,y,t) = f_t(x,y)$, for $x = 0,1,2,\ldots, R-1$; $y = 0,1,2,\ldots, C-1$ and $t = T, T-1, \ldots, T-N+1$, where T denotes the current time. One cross-sectional slice of the 3D image $f(x,y,t)$ along the y-axis gives a 2D spatial–temporal image $f_y(x,t)$ in the x–t plane. It can be expected that the gray values for each location x over time frames $t = T, T-1, \ldots, T-N+1$ will be approximately the same if the location x is a part of the static background over the entire N observed frames. Fig. 1 depicts the 3D spatial–temporal image and a 2D spatial–temporal image sliced along the y-axis. Fig. 2a and b further demonstrates the stacking of a series of 2D spatial images without and with a moving object. Fig. 3a and b shows the 2D slices of the 3D spatial–temporal images in Fig. 2a and b, respectively. It can be seen from Fig. 3a that the 2D spatial–temporal image $f_y(x,t)$ appears as a vertical-line structure texture. This is because there are no moving objects and the background remains steady throughout a short period of time for these N consecutive frames. Since each location x has similar gray values over the N time frames, a vertical line appears in the x–t image plane. Observing from Fig. 3b, we see that the static background region also forms a vertical-line pattern, but the foreground region shows an irregular, non-vertical line structure in the x–t image plane. Therefore, detection of moving objects in a series of video images can now be equivalent to removing the vertical line texture corresponding to the static background in the scene and identifying non-vertical line pattern associated with foreground objects in the x–t image plane. By scanning all x–t image planes along the y-axis in the 3D spatial-temporal image $f(x,y,t)$,