Engineering Applications of Artificial Intelligence 25 (2012) 1338–1348

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

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

Motion detection with pyramid structure of background model for intelligent surveillance systems

Shih-Chia Huang*, Fan-Chieh Cheng

Department of Electronic Engineering, National Taipei University of Technology, Taipei, 106, Taiwan

ARTICLE INFO

Available online 27 February 2012

Keywords:
Video surveillance
Motion detection
Background model
Pyramidal structure

ABSTRACT

This paper proposes a pyramidal background matching structure for motion detection. The proposed method utilizes spectral, spatial, and temporal features to generate a pyramidal structure of the background model. After performing the background subtraction based on the proposed background model, the moving targets can be accurately detected at each frame of the video sequence. In order to produce high accuracy for the motion detection, the proposed method also further includes a noise filter based on Bezier curve to smooth noise pixels, after which the binary motion mask can be computed by the proposed threshold function. Experimental results demonstrate that the proposed method substantially outperforms existing methods by perceptional evaluation.

1. Introduction

Video surveillance systems are very essential for safety and security. Actually, the automatic surveillance system has tremendously progressed due to the high applicability in public institutions, private firms, and houses. In fact, smart surveillance is a hot issue of extensive research, such as human actions recognition (Park and Aggarwal, 2004), traffic flow visualization (Shastry and Schowengerdt, 2005), association analysis of home videos (Pan and Ngo, 2007), home activity recognition (Naeem and Ham, 2009), explorative visualization and analysis (Buter et al., 2011), telecommunication applications (Rahman and Pathan, 2011), human–machine interaction (Halim et al., 2011), and many others. In the developing video surveillance systems, several significant functionalities must be taken into consideration, but not limited to, motion detection (Wren et al., 1997; Manzanera and Richefeu, 2004, 2007; Shoushtarian and Ghasem-aghaee, 2003; Wang et al., 2008), tracking (McFarlane and Schofield, 1995; Stauffer and Grimson, 2000), identification (Wang et al., 2003; Hayfron-acquah et al., 2003; Chong and Tanaka, 2010), data hiding (Wang et al., 2010), and feature recognition (Panganiban et al., 2011). This paper focuses on the design of motion detection since the first function significantly affects the performance of the surveillance system.

In general, motion detection can be performed by three categorized methods, including temporal difference, optical flow, and background subtraction methods (Hu et al., 2004). Temporal difference method can effectively accommodate environmental changes, but the shapes of moving objects are often not complete. On the other hand, optical flow method generally shows the projected motion on the image plane with good approximation. However, one common limitation of optical flow is that the computational complexity is often too high, making it difficult to implement. Background subtraction method consists of detecting moving objects that deviate from a maintained up-to-date background model. This is the most popular method for motion detection because it requires less computational complexity and provides high quality motion information. In other words, the background subtraction method is the most effective way to solve motion detection problems. The existing method of background subtraction computes the absolute difference between each pixel of the incoming video frame and background model. The threshold is then applied to get the binary motion detection mask (Pai et al., 2010). Although the existing background subtraction method can be easily implemented, threshold selection is still a critical operation for the noise immunity.

In order to distinguish background and foreground components, spectral, spatial, and temporal features are usually extracted from video sequences. Spectral features represent the gray-scale of an image frame. Spatial features are associated with regional variations of the local structure in the same frame. Temporal features represent changes in pixels within video streams. The organization of existing background models are broadly classified into pixel-based methods and block-based methods. Pixel-based methods use gray-level changes for each pixel in the video sequence to represent the spectral and temporal features. Block-based methods utilize variations within spaces between different frames for the representation of spatial and temporal features.

* Corresponding author.
E-mail address: schuang@ntut.edu.tw (S.-C. Huang).

0952-1976/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2012.02.002

This paper presents a new background model that incorporates both pixel-based and block-based methods to represent the spectral, spatial, and temporal features in a video stream. The organization of the proposed method is as follows:

(1) An intelligent matching process that uses both pixel-based and block-based techniques for producing an updated background model.
(2) Using Bezier curve smoothing method to suppress the possible noise pixels.
(3) Achieving complete motion detection through the proposed automatic selection of a threshold using Probability Mass Function and Cumulative Distribution Function after deriving a good quality background model.

Compared with other state-of-the-art methods, our method is the most effective, as indicated by the qualitative and quantitative results of the performance study, which presented a wide range of natural video sequences. The rest of this paper is organized as follows: Section 2 gives a fairly compact overview of some of the compared approaches for background subtraction. Section 3 describes the proposed method in detail. Section 4 reports the experimental results and discusses this paper. Section 5 contains our concluding remarks.

2. Related work

Typically, reliable background models can offer significantly improved motion detection masks that are able to detect moving objects completely. Some state-of-the-art methods of background subtraction include Simple Background Subtraction (SBS), Running Average (RA) (Wren et al., 1997), Σ-Δ Estimation (SDE) (Manzanera and Richefeu, 2004), Multiple Σ-Δ Estimation (MSDE) (Manzanera and Richefeu, 2007), Simple Statistical Difference (SSD), Temporal Median Filter (TMF) (Shoushtarian and Ghasemaghaee, 2003), and Running Average with DCT domain (RADCT) methods (Wang et al., 2008).

SBS method: The most basic method detects moving objects by taking the absolute difference between the reference image $B(x,y)$, which contains static background of the observed scene, and the incoming video frame $I_t(x,y)$, which possibly contains the moving objects and the background. The reference image $B(x,y)$ and the incoming video frame $I_t(x,y)$ are taken from the video sequence. A binary motion detection mask $D(x,y)$ is calculated as follows:

$$D(x,y) = \begin{cases} 1 & \text{if } |I_t(x,y) - B(x,y)| > \tau, \\ 0 & \text{if } |I_t(x,y) - B(x,y)| \le \tau, \end{cases} \tag{1}$$

where τ is an empirically selected threshold to distinguish pixels with either the moving objects or the background in an image frame.

When the absolute difference between a reference image and a current image frame exceeds τ, the pixels of the detection mask are labeled with "1" which means it contains moving objects; otherwise non-active ones are labeled with "0". The primary problems of the SBS method are the noise in the incoming video frame $I_t(x,y)$ and the occurrence of static objects in the reference image $B(x,y)$, which causes a failure to respond accurately in most real video sequences.

RA method: The problem can be modified substantially through an up-to-date background model using the RA method to adapt the temporal changes in the video sequence. Different from the SBS method, the RA background model guarantees the reliability of motion detection because each background image frame $B_t(x,y)$ of the adaptive background model is updated frequently. The current background image is integrated into the new incoming video frame $I_t(x,y)$ and the previous background frame $B_{t-1}(x,y)$. The adaptive background model is achieved by the following first-order recursive filter:

$$B_t(x,y) = (1-\beta)B_{t-1}(x,y) + \beta I_t(x,y), \tag{2}$$

where β is an empirically adjustable parameter. While faster background adaptation can be performed with a greater parameter β, the production of artificial "ghost" trails is also caused behind moving objects in the adaptive background model.

Based on the generated adaptive background model, the binary motion detection mask $D(x,y)$ is defined as follows:

$$D(x,y) = \begin{cases} 1 & \text{if } |I_t(x,y) - B_t(x,y)| > \tau, \\ 0 & \text{if } |I_t(x,y) - B_t(x,y)| \le \tau, \end{cases} \tag{3}$$

where $I_t(x,y)$ is the current incoming video frame, $B_t(x,y)$ is the current background model, and τ is an experimentally predefined threshold by which the binary motion detection mask can be detected.

SDE method: This method computes the temporal statistics for the pixels of the original video sequence based on the pixel-based decision framework. In the first background estimate, the computation uses the sgn function to estimate the background value. The sign function sgn can be expressed as follows:

$$\text{sgn}(a) = \begin{cases} 1 & \text{if } a > 0, \\ 0 & \text{if } a = 0, \\ -1 & \text{if } a < 0, \end{cases} \tag{4}$$

where a is the input real value.

Based on sgn function, the background estimation can be expressed as the following recursive filter:

$$B_t(x,y) = B_{t-1}(x,y) + \text{sgn}\,(I_t(x,y) - B_{t-1}(x,y)), \tag{5}$$

where $B_t(x,y)$ is the current background model, $B_{t-1}(x,y)$ is the previous background model, and $I_t(x,y)$ is the current incoming video frame. The intensity of the background model increases or decreases by a value of one through the evaluation of the sgn function at every frame.

The background model is estimated through the sgn function for a simple increment or decrement of one at every frame. Then the absolute difference $\Delta_t(x,y)$ is computed as the differential estimate between $I_t(x,y)$ and $B_t(x,y)$:

$$\Delta_t(x,y) = |I_t(x,y) - B_t(x,y)|. \tag{6}$$

Similarly, sgn function is also used to compute the time-variance $V_t(x,y)$ representing the measure of motion activity to decide whether the state for each pixel belongs to "background" or "moving object":

$$V_t(x,y) = V_{t-1}(x,y) + \text{sgn}(N \times \Delta_t(x,y) - V_{t-1}(x,y)), \tag{7}$$

where $V_t(x,y)$ is the current time-variance, $V_{t-1}(x,y)$ is the previous time-variance, and N is the predefined parameter which ranges from 1 to 4.

Based on the generated current time-variance $V_t(x,y)$, the binary motion detection mask $D(x,y)$ is detected as follows:

$$D_t(x,y) = \begin{cases} 1 & \text{if } \Delta_t(x,y) > V_t(x,y), \\ 0 & \text{if } \Delta_t(x,y) \le V_t(x,y). \end{cases} \tag{8}$$

MSDE method: Nevertheless, the SDE method is characterized by its updating period with a constant time to build a background model, which induces a limitation in certain complex scenes, typically in some cases of scenes with a lot of moving objects, or with the moving objects slowing down or stopping. Therefore, the MSDE method was developed in order to solve the problem. The adaptive background model of MSDE method can be expressed as