Information Sciences 269 (2014) 106–127

Contents lists available at ScienceDirect

Information Sciences

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

Real-time background modeling based on a multi-level texture description

Chia-Hung Yeh [a], Chih-Yang Lin [b,*], Kahlil Muchtar [a], Li-Wei Kang [c]

[a] Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan
[b] Department of Computer Science & Information Engineering, Asia University, Taichung, Taiwan
[c] Graduate School of Engineering Science and Technology-Doctoral Program, Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

ARTICLE INFO

Article history:
Received 13 May 2012
Received in revised form 25 April 2013
Accepted 9 August 2013
Available online 19 August 2013

Keywords:
Background modeling
Motion detection
Gaussian mixture modeling

ABSTRACT

Background construction is the base of object detection and tracking of machine vision systems. Traditional background modeling methods often require complicated computations and are sensitive to illumination changes. This paper proposes a novel block-based background modeling method based on a hierarchical coarse-to-fine texture description, which fully utilizes the texture characteristics of each incoming frame. The proposed method is efficient and can resist both illumination changes and shadow disturbance. The experimental results show that this method is suitable for real-world scenes and real-time applications.

1. Introduction

Due to social security and malicious attacks, the demands for video surveillance camera and its software architecture have increased in the prevention of accidents and the alleviation of risks. In 2007, the Security Industry Association reported that approximately 30 million security cameras have been installed in United States, and the sale for security cameras reached an estimated $3.2 billion, which is one-third of the overall national security market. These cameras, nowadays, have made big inroads into public places, and played a crucial role in detecting terrorist attacks, criminals, or other anti-social behavior on a daily basis. Because of the potential market and the demands of social security, video surveillance techniques have been developed over the past decades and numerous surveillance techniques have been proposed. One of the most important surveillance techniques is background modeling, which is used to detect moving objects—the underlying task of a security system. In other words, background modeling distinguishes foregrounds and backgrounds in order to identify moving objects, which are essential to security concerns.

A number of methods for detecting moving objects, especially different features employed for background modeling, have been proposed in past literature. Once the background modeling is satisfied, more advanced techniques can be applied straightforwardly such as tracking object [12] and face recognition [4]. The most frequently used feature in background modeling is based on color information. For example, a color statistical approach [21] uses average information to accomplish background subtraction and reduce shadow effect. In order to maintain the background flexibly, Wren et al. [22] proposed a one-Gaussian adaptive modeling method instead of using the average for moving object detection. However, one-Gaussian modeling cannot cope with dynamic background changes, such as ripples and screen blinking. Therefore, the Gaussian mixture modeling (GMM) approach [20] proposed by Stauffer and Grimson was developed by using more than one Gaussian for

* Corresponding author. Tel.: +886 4 23323456x20102; fax: +886 4 23305737.
E-mail address: andrewlin@asia.edu.tw (C.-Y. Lin).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.08.014

each pixel [19,20,24]. Pixels that do not fit the GMM would be recognized as foreground areas. An example using GMM has been developed (three Gaussians) for traffic monitoring [8]. Details of other implementations for GMM can be found in [5,18].

Motion-based and edge-based methods are other approaches for background modeling. The motion-based method [21] utilizes optical flow to detect salient motion over frames. This approach often suffers from complicated computations. The edge-based method [15] considers only the edge information of the frames and constructs edge histograms as a feature description for background modeling. The histogram-matching process determines the performance of these kinds of methods.

Recently, Heikkilä and Pietikäinen [9,10] proposed a texture-based background construction method using local binary patterns (LBPs). Although LBPs are tolerant of illumination changes, this method is not robust. When the central pixel value used in LBP is affected by noise, the corresponding LBP histogram will not be stable. This will increase the possibility of false positive or false negative cases. Furthermore, the overlapping block strategy and histogram-matching process proposed in [9,10] make the method of Heikkilä and Pietikäinen inefficient.

An interesting method proposed by McHugh et al.'s [16] is one that uses the hypothesis test and the Markov modeling approach. In this method, the authors use a non-parametric kernel density approach for modeling the background. They assume that the adjacent pixels of a foreground area should have higher probability to be classified as foreground. The modeling process is controlled by the Markov random field (MRF). This method works well in foreground detection, but needs relatively high computational complexity.

In this paper we propose a novel background modeling method based on a hierarchical coarse-to-fine texture description, which is an improved version of our previous work [17]. In [17], the color descriptor is used as the main device for background modeling. This technique, however, suffers from illumination changes and noise interference, hence opening the space for a more robust descriptor. Instead of applying LBPs, a new texture descriptor which employs block truncation coding (BTC) [6] to enhance the tolerance to illumination changes and shadow interference is proposed. Furthermore, the texture descriptor is also extended to a hierarchical coarse-to-fine approach to model the background more accurately and reduce both the false positive and false negative rates. In addition, a method that calculates the number of transitions is used to determine the complexity of each block, and at the same time, to improve the efficiency significantly. Because of the computational simplicity of the proposed method, the proposed background modeling has high efficiency and is suitable for real-time applications.

The remainder of this paper is organized as follows. Section 2 presents a brief review of the mixture of Gaussians and the model by Heikkilä and Pietikäinen's method. Then, Section 3 presents the proposed new background modeling scheme based on texture description. An extended explanation of the proposed method is provided in Section 4. Empirical results and discussions are offered in Section 5, and the conclusion is presented in Section 6.

2. Related work

2.1. Mixture of Gaussians model

A mixture of Gaussians model was first proposed by Grimson and Stauffer [19,20]. The authors proposed this new method to model every background pixel into a K-Gaussians mixture model (GMM). Typically K is a small number ranging between 3 and 5. The weight associated with each Gaussian represents the portion of the data accounted for that Gaussian.

Formally, in the GMM model, each pixel in the scene is modeled by a mixture of K Gaussian distributions. The probability of observing the current pixel value is:

$$P(X_t) = \sum_{j=1}^{K} \omega_{j,t} * \eta\left(X_t, \mu_{j,t}, \sum_{j,t}\right), \tag{1}$$

where X_t is the current pixel value at time t, K is the number of Gaussian distributions, $\omega_{j,t}$ is the weight estimation of the jth Gaussian in the mixture at time t, $\mu_{j,t}$ and $\sum_{j,t}$ are the mean value and covariance matrix of the jth Gaussian in the mixture at time t, respectively, and η is a Gaussian *pdf* (probability density function). The definition of η is presented in the following equation

$$\eta\left(X_t, \mu_{j,t}, \sum_{j,t}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\sum_{j,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu_{j,t})^T \sum_{j,t}^{-1} (X_t-\mu_{j,t})}, \tag{2}$$

where n is the dimension of X_t and $(X_t - \mu_{j,t})^T$ represents the transpose of the vector $(X_t - \mu_{j,t})$. For computational efficiency, the covariance matrix $\sum_{j,t}$ is defined as $\sigma^2_{j,t}\,\mathbf{I}$ for the jth model component, where $\mathbf{I}$ represents the identity matrix. This change of $\sum_{j,t}$ indicates that the components of X_t (i.e., red, green, and blue) are regarded as independent and have the same variances. The simplified $\sum_{j,t}$ improves the efficiency at the cost of some accuracy.