Pattern Recognition Letters 33 (2012) 1134–1147

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Background modeling by subspace learning on spatio-temporal patches

Youdong Zhao [a,b,*], Haifeng Gong [b,c,d], Yunde Jia [a], Song-Chun Zhu [b,c]

[a] Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, PR China
[b] Lotus Hill Research Institute, EZhou 436000, PR China
[c] Department of Statistics, UCLA, Los Angeles, CA 90095, Unites States
[d] Google Inc., Mountain View, CA 94043, United States

ARTICLE INFO

Article history:
Received 26 May 2010
Available online 1 February 2012
Communicated by A. Shokoufandeh

Keywords:
Visual surveillance
Background modeling
Spatio-temporal patch
Subspace learning

ABSTRACT

This paper presents a novel background model for video surveillance—Spatio-Temporal Patch based Background Modeling (STPBM). We use spatio-temporal patches, called *bricks*, to characterize both the appearance and motion information. Our method is based on the observation that all the background bricks at a given location under all possible lighting conditions lie in a low dimensional background subspace, while bricks with moving foreground are widely distributed outside. An efficient online subspace learning method is presented to capture the subspace, which is able to model the illumination changes more robustly than traditional pixel-wise or block-wise methods. Experimental results demonstrate that the proposed method is insensitive to drastic illumination changes yet capable of detecting dim foreground objects under low contrast. Moreover, it outperforms the state-of-the-art in various challenging scenes with illumination changes.

1. Introduction

Background modeling is a key component in a video surveillance system with static cameras. In the past decade, various background modeling algorithms (Piccardi, 2004; Elhabian et al., 2008; Bouwmans, 2009) have been proposed and achieved good performance on well-illuminated scenes. However, the challenging scenes with lighting and illumination changes still remain unsolved, such as (1) sudden sunlight changes in daytime, (2) light turned on or off, and (3) car lighting in nighttime outdoor scenes. Especially, in a nighttime outdoor scene, the faint lighting, low signal-noise-ratio (SNR), low contrast, and drastic illumination changes are combined to form a really difficult scenario for background modeling.

Spatial neighborhood information and temporal one are two fundamental elements to understand appearance structure and dynamic motion respectively and *complementary* to each other. For example, in a low contrast environment, the motion of foreground provides most of the visual information; whereas when there are illumination changes, the appearance of foreground contributes the main visual information. However, the traditional **pixel-wise** methods (Wren et al., 1997; Stauffer et al., 2000; Elgammal et al., 2002) model the background as a set of independent pixel processes without considering neighborhood information, the **block-wise** methods use only the spatial correlations between pixels (Seki et al., 2003; Heikkila and Pietikainen, 2006; Lin et al., 2009) or employ the spatial and temporal information separately (Monnet et al., 2003; Wang et al., 2007), and the **motion based** methods (Wixson, 2000) exploit the temporal neighborhood information alone. While it is difficult for these methods to deal with the above scenarios individually, background modeling will benefit from utilizing the spatial and temporal information jointly. Moreover, according to the illumination literature (Belhumeur and Kriegman, 1998; Basri and Jacobs, 2003; Garg et al., 2009), the illumination variations of a static object (e.g., a background patch in a surveillance scene) could be represented by a low-dimensional subspace under the assumption of Lambertian surface.

Motivated by these observations, we propose to build background models on *spatio-temporal patches* (called "bricks"), which characterize both the appearance and motion information in the spatial and temporal neighborhood of a pixel (e.g., $6 \times 6 \times 4$ pixels as shown in Fig. 1). In the proposed method, a brick is the atomic processing unit, which differs from the traditional pixel-wise or block-wise methods. Similar to image patches (or blocks) (Belhumeur and Kriegman, 1998), we observe that under all possible lighting conditions the background bricks extracted from a given location lie in a *low-dimensional* subspace, i.e., *background subspace* or *background model*. Then, we present an efficient online subspace learning method to capture the background model and adapt it to the recent variations in a real scene. The low computation complexity of this

* Corresponding author at: Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, PR China. Tel./fax: +86 10 6891 4849.

E-mail addresses: zhaoyoudong@gmail.com (Y. Zhao), haifeng.gong@gmail.com (H. Gong), jiayunde@bit.edu.cn (Y. Jia), sczhu@stat.ucla.edu (S.-C. Zhu).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2012.01.012

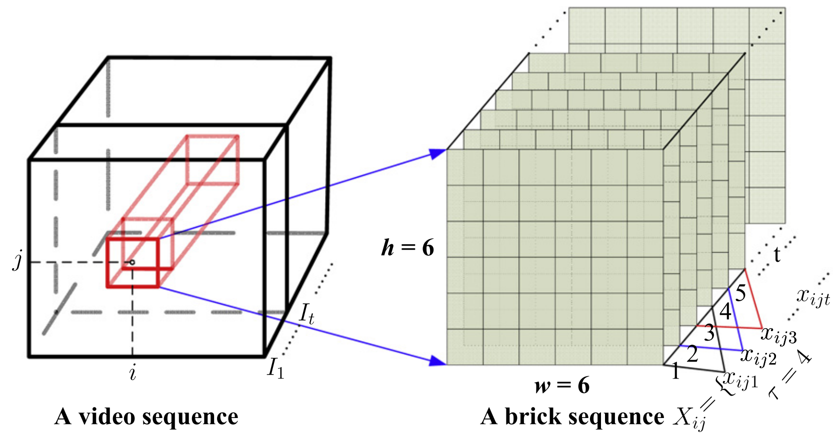

Fig. 1. A brick x_{ijt} is a small spatio-temporal patch with $h \times w \times \tau$ pixels around the point (i,j,t).

method makes it suitable for real-time applications. Here, we capture the background model of bricks at each background location independently.

Once the background model is learnt and adapted, we could perform "foreground detection" or "background subtraction" by thresholding the residual errors of incoming bricks on it. The residual errors of background bricks are usually distinct from those of foreground bricks and an adaptive threshold is proposed for reliable detection. Extensive experiments demonstrate the robustness of the proposed method in overcoming various illumination changes and low contrast in real-world video surveillance. The proposed brick based method advances the state-of-the-art in three aspects:

- *It is robust to both sudden and gradual illumination changes and achieves superior performance to the state-of-the-art.*
- *It is sensitive to dim moving objects under low contrast.*
- *It is simple yet effective for almost all the real challenging cases including indoor and outdoor, and daytime and nighttime scenes.*

1.1. Related work

While most of the **pixel-wise methods** (Wren et al., 1997; Stauffer et al., 2000; Elgammal et al., 2002; Kaewtrakulpong and Bowden, 2001; Lee, 2005) could handle gradual illumination changes by adapting their models, they often have difficulty dealing with sudden changes and are vulnerable to noises. Some methods explicitly alleviate the illumination effects by an extra illumination estimation (Messelodi et al., 2005) or a color model (Kim et al., 2005; Patwardhan et al., 2008). Recently, instead of modeling the intensities of pixels, Pilet et al. (2008) model the ratio of intensities between a stored background image and an input image by Gaussian Mixture Models to deal with sudden illumination changes. Their method successes in coping with sudden illumination changes, such as light switch in indoor scenes.

Block-wise methods use spatial correlations between pixels to improve robustness to noises and illumination changes. Seki et al (2003) exploit the cooccurrence of adjacent blocks for background subtraction. Heikkila and Pietikainen (2006) present a texture-based method (TBMOD), which employs the Local Binary Pattern (LBP) operator and can tolerate considerable illumination variations. In (Yao and Odobez, 2007), a multi-layer method is proposed, which combines the LBP feature and a color feature. In (Grabner and Bischof, 2006; Lin et al., 2009), classification based methods are proposed using image blocks. Edge (or gradient-based) features are used to model the background for its robustness to illumination changes in (Yang and Levine, 1992). A fusion of color and edge information is used in (Jabri et al., 2000). Noriega and Bernier (2006) combine local kernel histograms and contour-based features for background subtraction.

Motion based methods exploit temporal neighborhood information for foreground detection. Wixson (2000) defines a salient motion that tends to move in a consistent direction over time and detects the salient motion by integrating frame-to-frame optical flow. The information of successive frames enhances the saliency of moving objects despite of the similarity of appearances to the background.

A subspace learning based method for background modeling is first introduced by Oliver et al. (2000). This method establishes a global subspace over the whole frames, i.e., the eigen-background model, which can handle the global illumination changes to a certain degree. But they cannot deal with local illumination changes, and fail to distinguish slow moving foreground objects. While some improvements (Li, 2004; Skočaj et al., 2007; Skočaj and Leonardis, 2008) are made to deal with the slow moving objects, the local illumination change still remains unsolved. Some other methods (Monnet et al., 2003; Wang et al., 2007) operate on image blocks and exploit an additional prediction model (e.g., on-line auto-regression model) to predict future frames to capture the dynamic changes in temporal domain. While their methods capture the spatial and temporal information by two separate models respectively and can deal with the local illumination changes, they still miss detections in low contrast cases like other block-wise methods, which could be addressed by our brick-based method that models the spatio-temporal variations jointly in the brick space.

Some researchers also use the **spatio-temporal information** for background modeling. Pless (2005) builds background models based on the responses of spatiotemporal derivative filters at each pixel. Wang et al. (2006) integrate spatial and temporal dependencies for foreground segmentation and shadow removal via a dynamic probabilistic framework based on the conditional random field.

Stereo information is also employed to construct the background model invariant to illumination changes (Ivanov et al., 2000). Their method requires an off-line construction of disparity fields mapping the primary background images via using two or more cameras and suffers from both missing and false detections due to certain geometric considerations. Lim et al. (2005) introduce an improvement to Ivanov's method (Ivanov et al., 2000) to alleviate the false detections and some other issues.

1.2. Paper organization

The rest of the paper is arranged as follows: Section 2 presents two conjectures about the distribution of a video brick sequence, which is the motivation of the novel proposed algorithm.