DYNAMIC BACKGROUND MODELING AND SUBTRACTION USING SPATIO-TEMPORAL LOCAL BINARY PATTERNS

Shengping Zhang, Hongxun Yao, Shaohui Liu

School of Computer Science and Engineering, Harbin Institute of Technology, Harbin, 150001, China
{spzhang, yhx, shaohl}@vilab.hit.edu.cn

ABSTRACT

Traditional background modeling and subtraction methods have a strong assumption that the scenes are of static structures with limited perturbation. These methods will perform poorly in dynamic scenes. In this paper, we present a solution to this problem. We first extend the local binary patterns from spatial domain to spatio-temporal domain, and present a new online dynamic texture extraction operator, named spatio-temporal local binary patterns (STLBP). Then we present a novel and effective method for dynamic background modeling and subtraction using STLBP. In the proposed method, each pixel is modeled as a group of STLBP dynamic texture histograms which combine spatial texture and temporal motion information together. Compared with traditional methods, experimental results show that the proposed method adapts quickly to the changes of the dynamic background. It achieves accurate detection of moving objects and suppresses most of the false detections for dynamic changes of nature scenes.

Index Terms— Background modeling, object detection, spatio-temporal features, local binary patterns

1. INTRODUCTION

In many video analysis applications, detecting moving objects from a video sequence captured by a static camera is one of the basic tasks. A common approach for this task is background modeling and subtraction, which first builds an adaptive statistical background model, and then new pixels that are unlikely to be generated by this model are labeled as foreground. It is always desirable to achieve very high accuracy in the detection of moving objects with the lowest possible false detection rate. The performance of background subtraction depends mainly on the background modeling technique used.

The simplest background model assumes that intensity values of a pixel can be modeled by a single Gaussian distribution [1]. In [2], the mixture of Gaussians approach was used to model the background. However, these models based on Gaussian distributions do not work well in dynamic scenes since they are based on the assumption that the scenes to be modeled are of static structures with limited perturbation. When the assumption is not valid, for example, in dynamic natural scenes which include repetitive motions like swaying vegetation, waving trees, rippling water, etc, they can not accurately model the background with just a few Gaussians distributions and they may fail to provide accurate detection [3].

Pixel-based methods mentioned above assume that the time series of observations is independent on each pixel which is an unreasonable assumption and restrict their use in dynamic background. In contrast, region based methods have been proposed in [4]. The textured-based method was used in [5], and it modeled the background with a group of histograms based on local binary patterns. This method, to a certain extent, can avoid labeling some moving background pixels as foreground since it extracts region texture features. However, its detection performance will sharply decline when scenes have strong changes.

Description of dynamic textures has attracted growing attention. A volume local binary patterns (VLBP) based dynamic texture recognition method was proposed in [6]. VLBP is an extension of the ordinary local binary patterns. However it has two significant shortcomings: 1) it can only be used offline because it demands the whole video sequence is available before computing dynamic texture, 2) it considers too many neighboring pixels, hence, it is time-consuming. Both of these shortcomings limit its usefulness in real-time applications.

In this paper, we first extend ordinary local binary patterns from spatial domain to spatio-temporal domain, and propose a new online dynamic texture extraction operator, named spatio-temporal local binary patterns (STLBP). We then present a new method of dynamic background modeling and subtraction based on STLBP histograms which combine spatial texture and temporal motion information together. Experimental results indicate that the proposed method can adapt quickly to changes in the dynamic background. Compared to the work proposed in [2] [7], it achieves more accurate detection of moving objects and suppresses most of the false detections for dynamic changes of nature scenes.

This work is supported by the National Natural Science Foundation of China (Grant No 60775024) and New Century Excellent Talents in University (NCET-05-03-34).

This paper is organized as follows. Section 2 describes the ordinary local binary patterns and the novel spatio-temporal local binary patterns. STLBP based dynamic background modeling and subtraction is also described in this Section. Experiments and analysis are then given in Section 3. Finally, conclusions are drawn in Section 4.

2. PROPOSED APPROACH

2.1. Ordinary Local Binary Patterns

Local Binary pattern (LBP) is a gray-scale invariant texture description. The operator [8] labels the pixels of an image by thresholding the eight neighborhood of each pixel with the center value and considering the result as an eight bit binary number (LBP code):

$$LBP(x_c,\, y_c) = \sum_{p=0}^{7} s(g_p - g_c)2^p \tag{1}$$

where g_c corresponds to the grey value of the central pixel $(x_c,\, y_c)$ and g_p to the grey values of the eight neighboring pixels. The function $s(x)$ is defined as follows:

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2}$$

LBP features are robust to monotonic gray-scale changes and very fast to compute which are very important for background subtraction in real-time application.

2.2. Spatio-temporal Local Binary Patterns

In order to model the dynamic scenes using both spatial texture and temporal motion information together, we extend ordinary local binary patterns from spatial domain to spatio-temporal domain, and propose a new online dynamic texture extraction operator, named spatio-temporal local binary patterns (STLBP).

Let f_t be the current frame at time t and f_{t-1} be the previous frame at time $t-1$. In the frame f_t, the central pixel is $(x_{t,c},\, y_{t,c})$ with grey value $g_{t,c}$. P equally spaced neighboring pixels $(x_{t,0},\, y_{t,0}), \ldots, (x_{t,P-1}\, y_{t,P-1})$ with grey values $g_{t,0}, \ldots, g_{t,P-1}$ on a circle of radius R_{LBP} in f_t are defined to be the spatial neighboring pixels of $(x_{t,c},\, y_{t,c})$. In the f_{t-1}, the corresponding position pixels of P spatial neighboring pixels is $(x_{t-1,0},\, y_{t-1,0}), \ldots, (x_{t-1,P-1},\, y_{t-1,P-1})$ which are defined to be the P temporal neighboring pixels of $(x_{t,c},\, y_{t,c})$ with grey values $g_{t-1,0}, \ldots, g_{t-1,P-1}$. The central pixel and its spatial and temporal neighboring pixels are shown in Fig. 1. Using these neighboring grey values, we can compute two P-bit LBP codes for central pixel $(x_{t,c},\, y_{t,c})$ as follows:

$$LBP^{t}_{P,R}(x_{t,c},\, y_{t,c}) = \sum_{p=0}^{P-1} s(g_{t,p} - g_{t,c})2^p \tag{3}$$

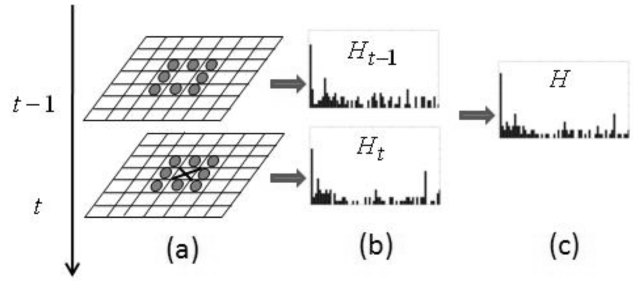

Fig. 1. The computing procedure of STLBP histogram on center pixel (marked with X). (a) The spatial neighborhood and temporal neighborhood (marked with circles) of the center pixel. (b) Histograms computed in current and previous frames. (c) The STLBP histogram.

$$LBP^{t-1}_{P,R}(x_{t,c},\, y_{t,c}) = \sum_{p=0}^{P-1} s(g_{t-1,p} - g_{t,c})2^p \tag{4}$$

$LBP^{t}_{P,R}(x_{t,c},\, y_{t,c})$ and $LBP^{t-1}_{P,R}(x_{t,c},\, y_{t,c})$ are called spatial and temporal local binary patterns of pixel $(x_{t,c},\, y_{t,c})$, respectively. The former extracts the spatial texture features and the latter extracts the motion information of neighboring two frames. Let R be a circular region of radius R_{rigion} centered on the pixel $(x_{t,c},\, y_{t,c})$ in the frame f_t, we can compute two histograms H_t and H_{t-1} over this region as follows:

$$H_{t,i} = \sum_{(x,\, y) \in R} I\{LBP^{t}_{P,R}(x,\, y) = i\}, i = 0, \ldots, 2^p - 1 \tag{5}$$

$$H_{t-1,i} = \sum_{(x,\, y) \in R} I\{LBP^{t-1}_{P,R}(x,\, y) = i\}, i = 0, \ldots, 2^p - 1 \tag{6}$$

where $H_{t,i}$ and $H_{t-1,i}$ are the histogram values at i^{th} bin of H_t and H_{t-1}, respectively, and $I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{else} \end{cases}$

Then these two histograms can be sum up to form a spatio-temporal local binary pattern (STLBP) histogram H as follows:

$$H_t = \omega H_{t-1,i} + (1-\omega) H_{t,i}, \qquad i = 0, \ldots, 2^p - 1 \tag{7}$$

where H_i is the histogram value at i^{th} bin of H. Parameter ω is the spatio-temporal rate. Fig. 1shown the whole computing procedure of STLBP.

We use STLBP statistics H as the dynamic texture description of the central pixel of the region R which combines spatial texture and temporal motion information together. Parameter ω reflects the importance of temporal motion information in histogram statistics. The computing of STLBP histogram only needs the current fame and previous frame. It is online and can be used in real-time application.

STLBP features have three advantages: 1) it is robust to monotonic gray-scale changes; 2) it is online and very fast to compute; 3) it can extract spatial texture and temporal motion information of a pixel. These three advantages are all very important for modelling the dynamic natural scenes.

2.3. Background Subtraction Using STLBP

Based on STLBP histograms, we proposed a new method of dynamic background modeling and subtraction. In proposed method, the dynamic background model of a pixel is built using a group of STLBP histograms. When a new frame is arriving, a STLBP histogram of the pixel can be computed using the current frame and previous one which was stored in memory at last time. Then labeling the pixel and updating its background model is the same as in [5]. Notice that the previous frame of the video sequence is just stored in memory to compute STLBP histograms for pixels in the current frame.

Parameter ω in Eq. 7 should be set according to changes degree of the dynamic background. A small value is sufficient for scenes which have small changes, whereas a larger value is required in the scenes which have strong changes. The method proposed in [5] only considers spatial texture information of a pixel, which ignores its temporal motion information. In fact, it is the special case of our method when ω is set to 0.

3. EXPERIMENTS AND ANALYSIS

In order to confirm the effectiveness of the proposed method for dynamic scenes, we conduct experiments using two images sequences. The first is the scene from [9] which involves heavily waving trees. The second is the scene of the fountain from [10] which involves three sources of nonstationarity: 1) The tree branches oscillate, 2) the fountains, and 3) the shadow of the tree on the grass below. We use both visual and numerical methods to evaluate proposed method. Numerical evaluation is done in terms of the number of false negatives (the number of foreground pixels that were missed) and false positives (the number of background pixels that were marked as foreground). The ground truth frames are obtained by manually labeling some frames from the image sequences. Two widely-used methods, Mixture of Gaussians (MoG) [2] and Kernel Density Estimation (KDE) [7] are employed to compare with the proposed method (STLBP).

Examples of detection results are shown in Fig. 2 and Fig. 3. In these two figures, the top row are the original frames, the second row are the ground truth frames, and the third row results are obtained by the proposed method. The fourth and fifth rows are the results obtained by KDE and MoG, respectively. Note that morphological operators were not used in the results.

As shown in Fig. 2 and Fig. 3, traditional pixel based MoG and KDE methods can not accurately detect moving objects in dynamic scenes. They label large numbers of moving background pixels as foreground and also output a huge amount of false negatives on the inner areas of the moving object. However, the proposed method can accurately distinguish moving background pixels and true moving objects. It dramatically outperforms other methods. The reason for this is that proposed method integrates spatial texture and temporal motion information together by STLBP, which is very important to accurately label those moving background pixels. Table. 1 and Table. 2 show the corresponding number evaluation of Fig. 2 and Fig. 3. From the comparisons, overall performance of our method was better than other methods. It should be noticed that, for the proposed method, most of the false positives occur on the contour areas of the moving objects. This is be-

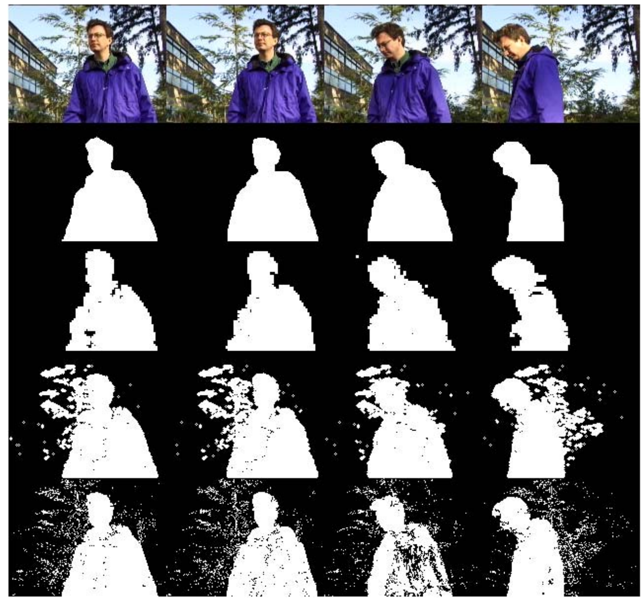

Fig. 2. Comparisons results on waving trees sequence. the top row are the original 247^{th}, 249^{th}, 254^{th} and 256^{th} frames, the second row are the ground truth frames, and the third row results are obtained by the proposed method. The fourth and fifth rows are the results obtained by MOG and KDE, respectively. Note that morphological operators were not used in the results.

Table 1. Comparison results of false positives (FP) and false negatives (FN) on waving trees sequence

Method		247^{th}	249^{th}	254^{th}	256^{th}
STLBP	FP	7	0	13	27
	FN	32	14	21	237
KDE	FP	1375	1686	639	1754
	FN	65	117	118	120
MoG	FP	799	828	440	957
	FN	119	271	1058	207

Fig. 3. Comparisons results on fountain sequence. the top row are the original 594^{th}, 598^{th}, 610^{th} and 616^{th} frames, the second row are the ground truth frames, and the third row results are obtained by the proposed method. The fourth and fifth rows are the results obtained by MOG and KDE, respectively. Note that morphological operators were not used in the results.

Table 2. Comparison results of false positives (FP) and false negatives (FN) on fountain sequence

Method		594^{th}	598^{th}	610^{th}	616^{th}
STLBP	FP	53	314	253	370
	FN	0	17	11	9
KDE	FP	578	759	3335	3733
	FN	823	672	1855	954
MoG	FP	1177	865	1951	1309
	FN	1548	1767	3077	2289

cause the STLBP is computed over a region. We do not take these false positives into account since they are not important for most high level vision application.

4. CONCLUSION

In this paper, we have extended the ordinary local binary patterns from spatial domain to spatio-temporal domain and proposed a new online dynamic texture extraction operator, named spatio-temporal local binary patterns (STLBP). It is online and very fast to compute. The proposed dynamic background modeling and subtraction method based on STLBP histograms is very robust to dynamic movement in natural scenes such as swaying vegetation, waving trees and rippling water. It achieves detection of moving objects with high accuracy and suppresses most of the false detections by traditional methods. The proposed method can be used in real-time visual surveillance applications.

5. REFERENCES

[1] C.R. Wren, A. AZarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780–785, July 1997.

[2] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1999, vol. 2, pp. 246–252.

[3] D.S. Lee, J.J. Hull, and B. Erol, "A bayesian framework for gaussian mixture background modelling," Proc. IEEE Conf. Image Processing, 2003, vol. 3, pp. 973–976.

[4] F. Porikli and O. Tuzel, "Human body tracking by adaptive background models and mean-shift analysis," Proc. IEEE Workshop on Performance Evaluation of Tracking and Surveillance, 2003.

[5] M. Heikkilä, M. Pietikäinen, and J. Heikkilä, "A texture-based method for detecting moving objects," Proc. British Machine Vision Conf., 2004, vol. 1, pp. 187–196.

[6] G. Zhao and M. Pietikäinen, "Dynamic texture recognition uing volume local binary patterns," Proc. Workshop on Dynamical Vision, 2007, pp. 165–177.

[7] A. Elgammal, D. Harwood, and L.S. Davis, "Non-parametric model for background subtraction," European Conf. Computer Vision, 2000, vol. 2, pp. 751–761.

[8] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[9] K. Toyama, B. Brumitt J. Krumm, and B. Meyers, "Wallflower: Principles and practice of background maintenance," Proc. International Conference on Computer Vision, 1999, vol. 1, pp. 255–261.

[10] Y. Sheikh and M. Shah, "Bayesian object detection in dynamic scenes," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005, vol. 1, pp. 74–79.