ROBUST BACKGROUND MODELING VIA STANDARD VARIANCE FEATURE

Bineng Zhong, Hongxun Yao, Shaohui Liu

Department of Computer Science and Engineering, Harbin Institute of Technology
{bnzhong, yhx, shaohl}@vilab.hit.edu.cn

ABSTRACT

In this paper, a novel standard variance feature is proposed for background modeling in dynamic scenes involving waving trees and ripples in water. The standard variance feature is the standard variance of a set of pixels' feature values, which captures mainly co-occurrence statistics of neighboring pixels in an image patch. The background modeling method based on standard variance feature includes two main components. First, we divide image into patches and represent each image patch as a standard variance feature. Then, assuming that standard variance feature fits a mixture of Gaussians distribution, we use mixture of Gaussians models to model it. Experimental results on several challenging video sequences demonstrate the effectiveness of our method.

Index Terms**—** Background Modeling, Standard Variance Feature, Pattern Representation

1. INTRODUCTION

Designing robust background modeling methods is still an open issue, especially considering various complicated variations that may occur in dynamic scenes, e.g., trees waving, water rippling, illumination changes, camera jitters, etc.

To achieve the goal of robust background modeling, a good scene or pattern representation is one of the key issues. Representation issues include: what level (e.g. pixel, patch or frame level) representation and feature are desirable for the description of a pattern, and how to effectively extract the feature from the original input signal.

Intensity or color feature is widely used in the background modeling literature to separate foreground from background. In the famous method [1], every pixel's color feature in a scene is modeled as a Mixture of Gaussian. Elgammal et al. [2] utilize a general nonparametric kernel density estimation technique for building a statistical representation of the scene background using normalized RGB color. A non-statistical clustering technique to construct a background model is presented in [3]. The background is encoded on a pixel-by-pixel basis and samples at each pixel are clustered into the set of codewords. Since most of the dynamic scenes exhibit persistent motion characteristics, a natural approach to model their behavior is via motion information (i.e. optical flow). Sheikh and Shah [4] use both temporal and spatial feature to improve background modeling results. In [5], a non-parametric model of color and optical flow is built for every pixel in a scene. Parag and Elgammal [6] use a boosting method (RealBoost) to choose the best feature to distinguish the foreground for each of the areas in the scene. Wixson [7] presents an algorithm that detects salient motion by integrating frame-to-frame optical flow over time.

Most of the above background modeling methods share the same basic assumption that the time series of observations is independent on each pixel and use the properties of a single pixel as image features (e.g., intensities, color, edges, gradients, or optical flow) directly. This assumption, however, may be too restrictive, especially under difficult conditions such as dynamic scenes. To effectively build background models for dynamic scenes, the correlation between pixels in the spatial vicinity has attracted more and more attentions from researchers. In [8], a 3-stage algorithm is presented, which operates respectively at pixel, region and frame level. Heikkila and Pietikainen [9] propose an approach based on the discriminative Local Binary Pattern (LBP) histogram. However, the method is not so efficiently on uniform regions. In [10], scene is coarsely represented as the union of pixel layers and foreground objects are detected by propagating these layers using a maximum-likelihood assignment. However, the limitations of the method are high-computational complexity and the requirement of an extra offline training step.

Some background modeling methods divide an image into patches and calculate patch-specific features. Change detection is achieved via patch matching. The algorithm presented in [11] uses an edge histogram calculated over the patch area as a feature vector to describe the patch. Matsuyama et al. [12] measure the patch correlation via the NVD (Normalized Vector Distance) measure. Please refer to [13] for a more complete background subtraction methods review.

Inspired by the idea that image variations at neighboring pixels have strong correlation, we propose a novel standard variance feature for background modeling. We divide image into patches and represent each image patch as a standard variance feature. The background model of each image patch's standard variance feature is constructed as a mixture of Gaussians models. The standard variance feature is the standard variance of a set of pixels' feature values, which captures mainly co-occurrence statistics of neighboring pixels in an image patch. Experimental results on several challenging video sequences demonstrate the effectiveness of the proposed method.

The rest of this paper is organized as follows. Section 2 introduces the standard variance feature. The background modeling method based on standard variance feature is described in Section 3. The experiments are given in Section 4. We conclude the paper in Section 5.

2. STANDARD VARIANCE FEATURE

To effectively employ the co-occurrence statistics of neighboring pixels, we extract the standard variance feature in an image patch. In the following section, we will first define the standard variance feature, and then give some typical examples to show why it contains substantial evidence for dynamic background modeling.

Let R be an $N \times N$ image patch. (For simplicity, we assume that the shape of image patch is square.) For a pixel $p(x, y)$, let $I(p)$ denote its intensity. The standard variance feature can be defined as follows:

$$\sigma = \sqrt{\frac{1}{N \times N} \sum_{p \in R} (I(p) - \mu)^2} \quad (1)$$

where $\mu = \frac{1}{N \times N} \sum_{p \in R} I(p)$ is the mean of pixels' intensities.

The advantages of using the standard variance feature as co-occurrence statistics descriptor for dynamic background modeling are as follows. First, it explicitly considers the meaningful correlation between pixels in the spatial vicinity. For example, an image patch's center pixel in current frame would be a neighboring pixel in the next frame due to the small movements of dynamic scenes. The center pixel's intensity will change non-periodically. However, the image patch's standard variance feature is unchanged due to the spatial co-occurrence correlations between the center pixel and its neighboring pixels are modeled by Eq. (1). Second, the image noises are largely filtered out with the average filter during the computation of standard variance feature. Third, the standard variance feature is invariant to mean changes such as identical shifting of intensities. This is very valuable when scenes are under changing illumination conditions. Finally, the standard variance feature results in a low dimensional scalar as representation of each image patch. This avoids expensive computation in background modeling phase, which is an important property for practical applications.

To illustrate why the standard variance feature contains substantial evidence for dynamic background modeling, we check how the standard variance features of some image patches' from an outdoor scene [6] change over a short period of time (650 frames-30 seconds). The scene and the sample image patches are shown in Fig.1

Fig.1. Outdoor scene with three red rectanges showing the locations of three sample image patches respectively, static patch A, dynamic patch B and foreground patch C.

Fig.2 plots the evolving curves of the standard variance feature over time, for the three patches separately. It is obviously to see that the standard variance feature distribution of Patch A is fairly stable near zero, in consistent with the fact that Patch A is a flat region in the sky. For the dynamic Patch B, the fluctuation of the corresponding standard variance feature is relatively small. When there is no foreground object occupies the Patch C, the distribution of the corresponding standard variance feature is also relatively stable. However, when foreground objects frequently occupy the Patch C, the standard variance feature abruptly changes. This indicates that current extracted standard variance feature deviates the underlying background model. Thus, we can achieve robust foreground objects detection based on these phenomena.

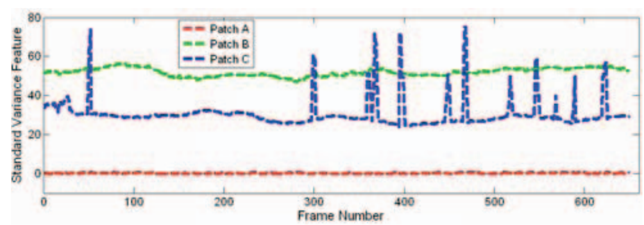


Fig.2. Evolving curves of the standard variance feature over time, for the three sample image patches in Fig.1.

3. BACKGROUND MODELING BASED ON STANDARD VARIANCE FEATURE

In this section, we introduce background modeling mechanism based on the standard variance feature described above. The goal is to construct and maintain a statistical representation of the scene that the camera sees. We partition each new video frame into a set of non-overlapping patches with $N \times N$ pixels and represent each patch by a standard variance feature. In the following, we explain the background modeling procedure for each image patch.

Inspired by the Stauffer and Grimson' work [1], we consider the standard variance feature of a particular patch over time as a patch process, and model the background model for this patch as a mixture of Gaussians models. As shown in Fig.3, it's obvious to see that the standard variance features can fit mixture of Gaussians distribution well.

For a patch X at time t, the probability of its standard variance feature value can be written as:

$$P_{patch}(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \qquad (2)$$

where K is the number of Gaussian components, η is a Gaussian probability density function $\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1}(X_t - \mu_t)\right)$, and $\omega_{i,t}$, $\mu_{i,t}$ and $\Sigma_{i,t}$ are the time adaptive mixture coefficients, mean and variance, respectively, of the ith Gaussian of the mixture associated with X_t. At each time instant, the Gaussian components are evaluated in descending order with respect to ω/Σ to find the first matching with X_t (a match occurs if the value falls within 2.5Σ of the mean of the component). The first B components are chosen as the background model, where

$$B = \operatorname{argmin}_b\left(\sum_{k=1}^{b} \omega_k > T\right) \qquad (3)$$

where T is the minimum portion of the background model. The weight $\omega_{i,t}$ is adjusted as follows:

$$\omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha(M_{i,t}) \qquad (4)$$

where α is the learning rate and $M_{i,t}$ is 1 for the model which is matched and 0 for others. In implementation, two significant parameters T and α are needed to be set. For further details, please see [1].

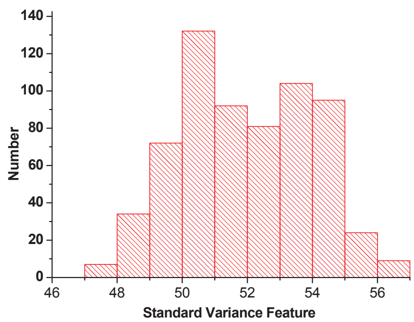
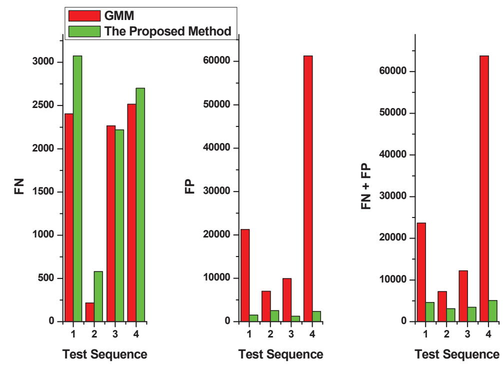


Fig.3. Histogram of the standard variance feature of the sample image patch B in Fig.1. It can fit mixture of Gaussians distribution well.

4. EXPERIMENTS

The performance of the standard variance feature based method for modeling the background is evaluated in this section. The algorithm is implemented using C++, on a computer with Intel-Core 2 1.86 GHz processor. It achieves the processing speed of 20 fps at the resolution of 320×240 pixels. We compare the performance of our method to the widely used methods of [1]. We refer to this method as GMM in the rest of our experiments. Both qualitative and quantitative comparisons are used to evaluate our approach. The quantitative comparison is done in terms of the number of false negatives (the number of foreground pixels that are missed) and false positives (the number of background pixels that are marked as foreground).

(a)



(b)

Fig.4. Comparision results of GMM and the proposed method. a) is the original test sequences and some detection results of the GMM and the proposed method. b) is the test results. FN and FP stand for false negatives and false positives, respectively.

In fig. 4(a), we show some results of the proposed method using four test sequences. The sequences used in the experiment include dynamic scenes, i.e., swaying trees, water ripples, and camera jitters. The frames on the first two columns contain heavily swaying trees, in which the challenge is due to the vigorous motion of the trees and bushes. The frames on the first two columns are from [8] and [6] respectively. The frames on the third column contain a moving bottle in foreground, with dynamic background composed of ripples in the water. The last frames on the fourth column are from [4], which contain average camera jitter of about 14.66 pixels. The challenges contained in above dynamic scenes cause classical GMM method that rely only on color information to fail. The proposed method gives good results because the standard variance feature exploits co-occurrence statistics of neighboring pixels in an image patch.

In order to provide a quantitative perspective about the quality of foreground detection with the proposed method, we manually mark the foreground regions in five frames from each sequence to generate ground truth data, and make comparison with the widely used GMM method. The numbers of error classifications are achieved by summing the errors from the frames corresponding to the ground truth frames. The corresponding quantitative comparison is reported in Fig. 4(b). For all test sequences, the proposed method achieves best performance in terms of false positives, and false negatives are acceptable. Since the standard variance feature is obtained by capturing mainly co-occurrence statistics of neighboring pixels in an image patch, it is robust against dynamic background. It should be noticed that, for the proposed method, most of the false negatives occur on the contour areas of the foreground objects (see Fig. 4(a)). This is because standard variance feature is a patch description feature. According to the overall results, the proposed method outperforms the GMM method for the used test sequences. The values for our method parameters are given in Table 1. Identical parameters are used in the four sequences.

Table 1. The parameter values of our method for the results in Fig.4(a).

Fig.	Sequences	K	T	α	$N \times N$
4(a)	1-4	3	0.7	0.01	10×10

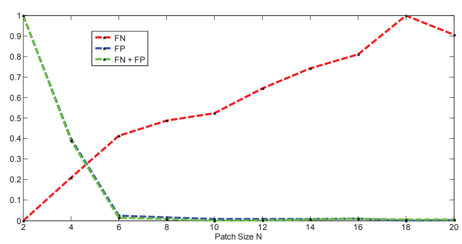


Fig.5. Number of false negatives (FN) and false positives (FP) for different parameter values for the second sequence of Fig.4. While the parameter value of patch size is varied, other parameters are kept fixed at the values given in Table 1. The results are normalized between zero and one: $x = (x - \min(x))/\max(x - \min(x))$.

Since our method is a patch-based method, there naturally arise the following questions: 1) How sensitive the proposed method is to small changes of the values of patch size? 2) How easy or difficult is it to obtain a good set of the values of patch size? To answer these questions, we calculate the error classifications (e.g. FN, FP and FN + FP) for different parameter configuration. Because of a huge amount of different combinations, only the parameter of patch size is varied. The measurements are made for several image sequences. The results for the second sequence of Fig.4 are plotted in Fig.5. Obviously, for the parameter of patch size, a good value can be chosen across a wide range of values. The same observation is identical for all the test sequences. This property significantly eases the selection of the parameter values of patch size.

5. CONCLUSION AND FUTURE WORK

This paper proposes a novel background modeling method based on standard variance feature, in which co-occurrence statistics of neighboring pixels in the spatial vicinity are captured. The main contributions of the proposed method are: (1) The standard variance feature is completely new for background modeling; (2) Different from the features (e.g., intensities, color) widely used in background modeling literature, the standard variance feature can offer relatively stable information for background modeling task; (3) We have validated the proposed method by conducting the experiments on four challenging sequences involving trees waving, water rippling, illumination changes, camera jitters, etc. Although the proposed method has not achieved pixel-level accuracy for the moving objects, it outperforms the GMM method for the used test sequences. Our future work will focus on how to improve contour accuracy for the moving objects.

Acknowledgments: This work is supported by National Basic Research Program of China (2009CB320906), National Natural Science Foundation of China (60775024).

6. REFERENCES

[1] C. Stauffer and W.E.L. Grimson, Learning Patterns of Activity Using Real-Time Tracking, TPAMI, vol. 22, no. 8, pp. 747-757, August 2000.

[2] A. Elgammal, D. Harwood, and L. Davis, Non-parametric Model for Background Subtraction, ECCV, vol.2, pp.751-767, June 2000.

[3] K. Kim, T.H. Chalidabhongse, D. Harwood and L. Davis, Real-time Foreground-Background Segmentation using Codebook Model, Real-Time Imaging, vol.11, issue 3, pp.167-256, June 2005.

[4] Y. Sheikh and M. Shah, Bayesian Modeling of Dynamic Scenes for Object Detection, TPAMI, vol. 27, no. 11, pp. 1778-1792, November 2005.

[5] A. Mittal and N. Paragios, Motion-based Background Subtraction using Adaptive Kernel Density Estimation, CVPR, vol.2, pp.302-309, July 2004.

[6] T. Parag, A. Elgammal and A. Mittal, A Framework for Feature Selection for Background Subtraction, CVPR, vol.2, pp.1916-1923, 2006.

[7] L. Wixson, Detecting Salient Motion by Accumulating Directionally Consistent Flow, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 774-780, Aug. 2000.

[8] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, Wallflower: Principles and Practice of Background Maintenance, ICCV, vol.1, pp.255-261, 1999.

[9] M. Heikkila and M. Pietikainen, A Texture-Based Method for Modeling the Background and Detecting Moving Objects, TPAMI, vol. 28, no. 4, pp. 657-662, April 2006.

[10] K.A. Patwardhan, G. Sapiro and V. Morellas, Robust Foreground Detection in Video Using Pixel Layers, TPAMI, vol. 30, no. 4, pp. 746-751, April 2008.

[11] M. Mason and Z. Duric, Using Histograms to Detect and Track Objects in Color Video, Proc. Applied Imagery Pattern Recognition Workshop, pp. 154-159, 2001.

[12] T. Matsuyama, T. Ohya, and H. Habe, Background Subtraction for Non Stationary Scenes, ACCV, pp. 622-667, 2000.

[13] M.Piccardi, Background Subtraction Techniques: a Review, SMC (4), pp. 3099-3104, 2004.