BACKGROUND SUBTRACTION USING INCREMENTAL SUBSPACE LEARNING

Lu Wang, Lei Wang, Ming Wen, Qing Zhuo, Wenyuan Wang

Department of Automation, Tsinghua University, Beijing, 100084, China

ABSTRACT

Background modeling and subtraction is a basic component of many computer vision and video analysis applications. It has a critical impact on the performance of object tracking and activity analysis. In this paper, we propose an effective and adaptive background modeling and subtraction approach that can deal with dynamic scenes. The proposed approach uses a subspace learning method to model the background and the subspace is updated on-line with a sequential Karhunen-Loeve algorithm. A linear prediction model is also used to make the detection more robust. Experimental results show that the proposed approach is able to model the background and detect moving objects under various type of background scenarios with close to real-time performance.

Index Terms— Background Subtraction, Object Detection, Subspace Method

1. INTRODUCTION

In many computer vision applications such as real time tracking and video surveillance, one fundamental module is background subtraction - differentiating foreground objects from the static parts of the scene. The information provided by such module can be considered as a valuable low-level cue to perform high-level tasks of motion analysis, like motion estimation, tracking, etc.

A simple yet widely used background modeling algorithm includes two parts: (1) maintaining the background pixel intensity model, and (2) subtracting the new frame from the background scene model and thresholding the difference value to determine the foreground label map. This task looks like fairly simple, but in real world applications, this approach rarely works. Usually background is never static and varies by time due to several reasons such as lighting changes, moving background objects and non-stationary scenes. To overcome these problems, it is crucial to build a stochastic representation of the background and continuously adapt it to the current environmental conditions.

There have been many research projects on this topic trying to build a statistical model of background that allow the video surveillance system to detect the foreground objects. For completely stationary background, the background intensity can be well modeled by a Gaussian function. Pfinder [2] uses a single Gaussian background model per pixel. The pixel intensity is updated recursively by a linear adaptive filter to adapt to slow changes in the scene. In [3], each pixel in the background is modeled by mixture of K Gaussians and the model parameters are updated using an on-line Expectation Maximization (EM) algorithm. Gaussian mixture model is more powerful than single Gaussian because in practice background pixels sometimes have multiple appearance surfaces in different conditions. Although mixture of Gaussian can converge to any arbitrary distribution provided enough number of components, this is not computationally possible for real time applications. Usually three to five components are used per pixel. Another approach to estimate probability distribution of background model is using nonparametric Kernel Density Estimation [4]. This model keeps temporal samples of intensity values per pixel and uses these samples to estimate the density function. Although these pixel-based techniques seem to be reasonable choices for background subtraction, they ignore the correlation of neighboring pixels and are not computationally possible for real-time applications. In [1], a PCA model is used to model the background. This method exploits the correlation of pixels and offers less computation. However, it fails to deal with dynamic background because the PCA model is learned beforehand and fixed during the detection procedure.

In this paper, we extend the work [1] and propose an effective and adaptive background modeling and subtraction approach that can deal with dynamic scenes such as ocean waves, waving trees, rain, moving clouds, etc. We use an online subspace learning method to model the background and updating the model with a sequential Karhunen-Loeve algorithm. A linear prediction model is also used to make the detection more robust.

The remainder of the paper is organized as follows. In section 2, we present the background model and update mechanism. Experimental results are shown in Section 3, followed by Section 4, which concludes the paper.

2. BACKGROUND MODEL

A subspace built by Principle Component Analysis is used to model the background. This subspace can describes the range of appearances such as lighting variations over the day, weather variations, etc. This subspace is updated by an incre-

mental method that updates the subspace of the background model using a variant sequential Karhunen-Loeve algorithm which in turns is based on the classis R-SVD method. Then linear prediction model is employed to make the detection more robust.

2.1. Batch PCA

Let $\{I_i\}_{i=1,2,\cdots N}$ be a given set of d dimensional column vector representations of the previous N observations by transforming every image into a 1D vector. We can compute the mean vector μ_b and subtract it from the input image to get zero mean vector $\{X_i\}_{i=1,2,\cdots,N}$ where $X_i = I_i - \mu_b$. Then we can obtain the covariance matrix:

$$C_b = E\{X_i X_i^T\} \approx \frac{1}{N}\mathbf{X}\mathbf{X}^T \tag{1}$$

where $\mathbf{X} = [X_1, X_2, \cdots, X_N]$. This covariance matrix can be diagonalized as:

$$L_b = \Phi_b^T C_b \Phi_b \tag{2}$$

where Φ_b is the eigenvector matrix of C_b and L_b is the diagonal matrix. Φ_b can be calculated through the singular value decomposition (SVD) of $\mathbf{X}$:

$$\mathbf{X} = USV^T \tag{3}$$

The eigenvectors of C_b are the columns of U, while the elements of S are the square root of the corresponding eigenvalues. In order to reduce the dimensionality of the space, only M eigenvectors correspond to the M largest eigenvalues are stored, which leads to a $d \times M$ matrix Φ_M.

2.2. Incremental Subspace Learning

The batch method is computationally inefficient and it might not be possible to execute it at each frame. Therefore, we consider a incremental subspace method based on the sequential Karhunen-Loeve algorithm [5] to update the subspace.

Given a $d \times n$ matrix $\mathcal{I}_p = [I_1, I_2, \cdots, I_n]$ where each column I_i is an observation(a d dimensional image vector), we can compute the singular value decomposition(SVD) of $\mathcal{I}_p = U_p \Sigma_p V_p^T$. When a $d \times m$ matrix of new observations $\mathcal{I}_q = [I_{n+1}, I_{n+2}, \cdots, I_{n+m}]$ is available, the R-SVD algorithm efficiently computes the SVD of the larger matrix $\mathcal{I}_r = [\mathcal{I}_p|\mathcal{I}_q] = U_r \Sigma_r V_r^T$ as follows:

1. Compute the mean of $\mathcal{I}_r$: $\overline{I}_r = \frac{n}{n+m}\overline{I}_p + \frac{m}{n+m}\overline{I}_q$, where $\overline{I}_p, \overline{I}_q$ and $\overline{I}_r$ denote the means of $\mathcal{I}_p, \mathcal{I}_q$ and $\mathcal{I}_r$ respectively.

2. Compute $E = [\mathcal{I}_q - \overline{I}_r \mathbf{1}_{(\mathbf{1}\times\mathbf{m})} | \sqrt{\frac{nm}{n+m}}(\overline{I}_p - \overline{I}_q)]$ where $\mathbf{1}_{(\mathbf{1}\times\mathbf{m})}$ is an m dimensional unit vector.

3. Using $U_p \Sigma_p V_p^T$ and E to obtain $U_r \Sigma_r V_r^T$:

   (a) Use an orthonormalization process (e.g., Gram-Schmidt algorithm) on $[U_p|E]$ to obtain an orthonormal matrix $U^{'} = [U_p|\tilde{E}]$.

   (b) Let the matrix $V^{'} = \begin{bmatrix} V_p & 0 \\ 0 & I_{(m+1)} \end{bmatrix}$ where $I_{(m+1)}$ is a $m+1$ dimensional identity matrix. Then

$$\begin{aligned} \Sigma^{'} &= U^{'T}[\mathcal{I}_p|E]V^{'} \\ &= \begin{bmatrix} U_p^T \\ \tilde{E}^T \end{bmatrix} [\mathcal{I}_p|E] \begin{bmatrix} V_p & 0 \\ 0 & I_{(m+1)} \end{bmatrix} \\ &= \begin{bmatrix} U_p^T \mathcal{I}_p V_p & U_p^T E \\ \tilde{E}^T \mathcal{I}_p V_p & \tilde{E}^T E \end{bmatrix} = \begin{bmatrix} \Sigma_p & U_p^T E \\ 0 & \tilde{E}^T E \end{bmatrix}. \end{aligned}$$

4. Compute SVD of $\Sigma^{'} = \tilde{U}\tilde{S}\tilde{V}^T$ then the SVD of $\mathcal{I}_r$ is

$$U_r \Sigma_r V_r = U^{'}(\tilde{U}\tilde{S}\tilde{V}^T)V^{'T} = (U^{'}\tilde{U})\tilde{S}(\tilde{V}^T V^{'T}).$$

Based on the R-SVD method, the sequential Karhunen-Loeve algorithm is able to perform the SVD computation of larger matrix $\mathcal{I}_r = [\mathcal{I}_p|\mathcal{I}_q]$ efficiently using the smaller matrices $U^{'}$, $V^{'}$ and the SVD of smaller matrix $\Sigma^{'}$. Note that this algorithm enables us to store the background model for a number of previous frames and perform a batch update instead of updating the background model every frame.

2.3. Detection

We use linear prediction [6] to detect foreground. Let the current frame be I_t, the previous P frames from the current frame be $I_{t-1}, I_{t-2}, \cdots, I_{t-P}$. The projections of the P frames onto the subspace, $I_{t-1}^{'}, I_{t-2}^{'}, \cdots, I_{t-P}^{'}$ can be computed as:

$$I_{t-i}^{'} = \Phi_M^T (I_{t-i} - \mu_b), i = 1, 2, \cdots, P. \tag{4}$$

Each element of the projection of current frame $I_t^{'}$ can be predicted as:

$$I_t^{'pred}(s) = \sum_{i=1}^{P} a_i I_{t-i}^{'}(s), s = 1, 2, \cdots, M \tag{5}$$

The best fitting values of the coefficients of the linear estimator, $a_1, a_2, \cdots, a_P$ can be computed as the solution to the linear system defined as follows, here s is omitted for convenience :

$$\begin{bmatrix} I_1^{'} & I_2^{'} & \cdots & I_P^{'} \\ I_2^{'} & I_3^{'} & \cdots & I_{P+1}^{'} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & I_{n-1}^{'} \end{bmatrix} \begin{bmatrix} a_P \\ \cdots \\ a_2 \\ a_1 \end{bmatrix} = \begin{bmatrix} I_{P+1}^{'} \\ I_{P+2}^{'} \\ \cdots \\ I_n^{'} \end{bmatrix} \tag{6}$$

The pseudo-inverse solution for the above least squares estimation problem has a $P \times P$ and a $1 \times P$ matrix with components of the form:

$$\sum_i I_i^{'}(s) I_{i+k}^{'}(s), k = 0, 1, \cdots, P+1 \tag{7}$$

For on-line updating the linear prediction model, it is only necessary to maintain the projection vectors of prior P frames, $I'_{t-1}, I'_{t-2}, \cdots, I'_{t-P}$ and update all the $P^2 + P$ components.

Then the current frame I_t can be predicted as:

$$I_t^{pred} = (\Phi_M \Phi_M^T)^{-1} \Phi_M I_t^{'pred} + \mu_b. \quad (8)$$

Finally, difference between the predicted frame and the current frame are computed and thresholded, the foreground points are detected at the location (x,y): $|I_t(x,y) - I_t^{pred}(x,y)| > T$, where T is a given threshold.

2.4. The Proposed Method

Put the initialization, detection and subspace update modules together, we obtain the adaptive background modeling algorithm as follows:

1.**Construct an initial subspace:** From a set of N training images of background $\{I_i\}_{t=1,2,\cdots N}$, the average image μ_b is computed and mean-subtracted images $\mathbf{X}$ are obtained, then the SVD of $\mathbf{X}$ is performed and the best M eigenvectors are stored in an eigenvector matrix Φ_M.

2.**Detection:** For an incoming image I_t, the predicted projection $I_t^{'pred}$ is first computed then it is reconstructed as I_t^{pred}, foreground points are detected at location (x,y) where $|I_t(x,y) - I_t^{pred}(x,y)| > T$.

3.**Update the subspace:** Store the background model for a number of previous frames and perform a batch update of the subspace using sequential Karhunen-Loeve algorithm.

4. Go to step 2.

3. EXPERIMENTS

In order to confirm the effectiveness of the proposed method, we conduct experiments using two different image sequences. The first is the scene of the ocean front which involves waving water surface, blowing grass, illumination changes, etc. The second is the scenario of the fountain which involves long term changes due to fountaining water, illumination changes and waving tree leaves. In order to reduce complexity, the images are divided into equal size blocks and each block is updated and detected individually in our experiments.

Two widely-used method, Mixture of Gaussians[3] and Kernel Density Estimation [4] are employed to compare with the proposed method. Simple spatial and temporal filtering was used for all algorithms. Examples of detection results are shown in Figure 1 and Figure 2. From the comparisons, we can see that the proposed method outperformed the Mixture of Gaussians and Kernel Density Estimation.

Figure 3 shows ROC curves formed by gradually changing the threshold value used for processing. The horizontal axis shows the rate of incorrect detections in the background, and the vertical axis shows the rate of correct detections of foreground, where the true data of the person region was created manually. The ROC curves showed that the proposed method is more effective than the others in dynamic scenes involving the background image variations.

Our current implementation of the proposed method in MATLAB runs about 5 frames per seconds on a Pentium IV 3GHz processor and can certainly be improved to operate in real-time.

4. CONCLUSION

In this paper, we proposed an effective and adaptive background subtraction approach that (1) updates the subspace online using the sequential Karhunen-Loeve algorithm; (2) employs linear prediction model for foreground detection. The advantage of the proposed approach is its ability to model complex background. We claim that the proposed method is able to model the background and detect moving objects under various type of background scenarios and with close to real-time performance.

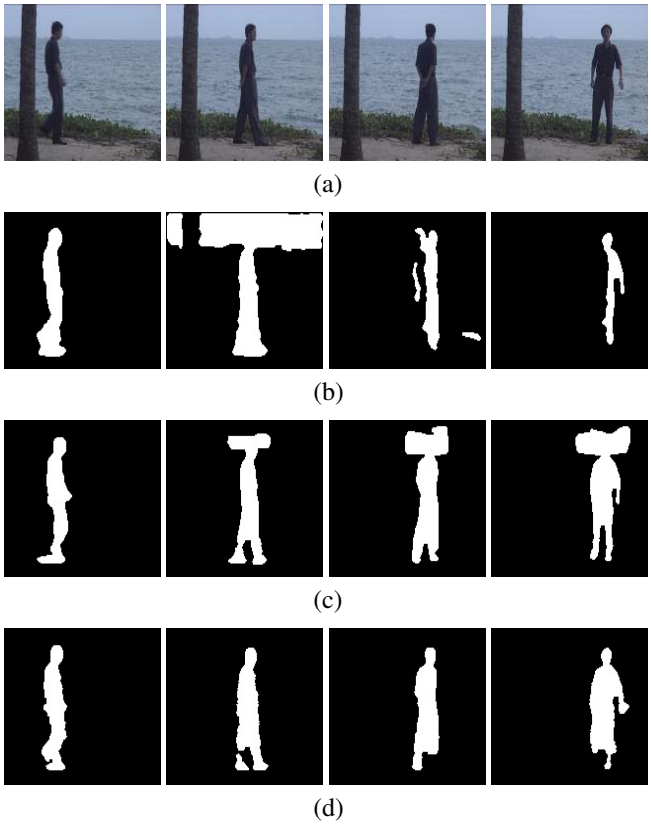


Fig. 1. (a) Original Images. Detection results using (b) Mixture of Gaussians model, (c) Kernel Density Estimation, (d) Proposed method.

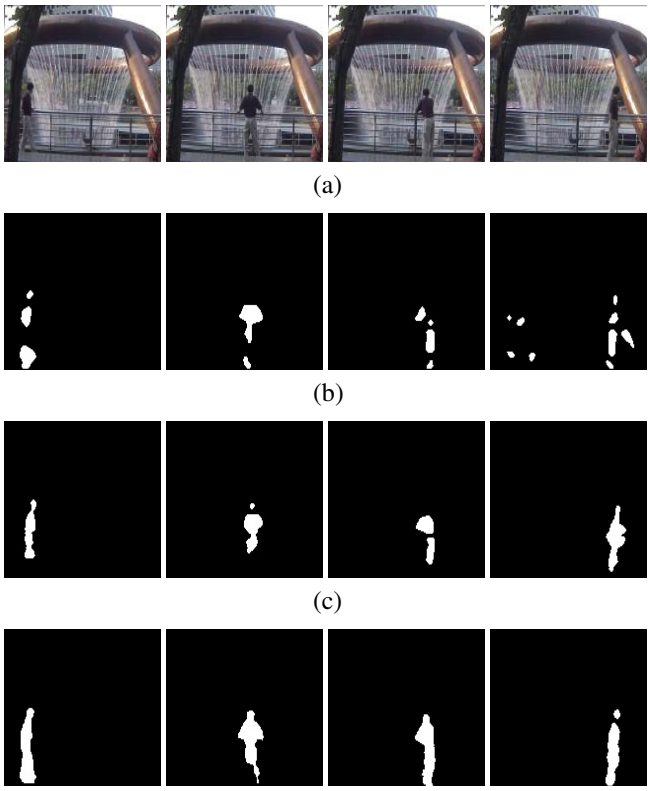
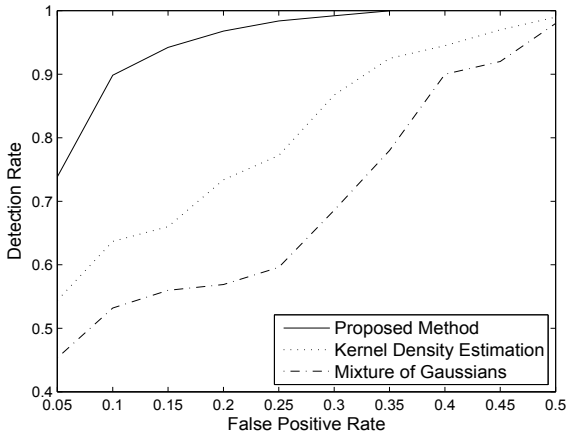
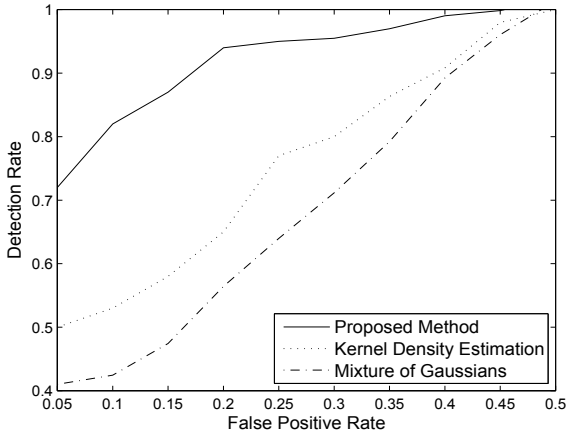


Fig. 2. (a) Original Images. Detection results using (b) Mixture of Gaussians model, (c) Kernel Density Estimation, (d) Proposed method.



(a)



(b)

Fig. 3. Receiver-Operator Characteristic(ROC) curves for (a) the first sequence and (b) the second sequence for (i) Mixture of Gaussians,(ii) Kernel Density Estimation, and (iii) Proposed method.

5. REFERENCES

[1] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, 831-843, 2000.

[2] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, 780-785, 1997.

[3] C. Stauffer, W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 246-252, 1999.

[4] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Volume 2. 302-309, 2004

[5] R. David.,J. Lim, M.H Yang, "Adaptive probabilistic visual tracking with incremental subspace update," Proceedings of the Eighth European Conference on Computer Vision. Volume 2. 470-482 2004

[6] R. Pless , J. Larson , S. Siebers, B. Westover , "Evaluation of local models of dynamic backgrounds", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 73–78, 2003