ADAPTIVE FOREGROUND OBJECT EXTRACTION FOR REAL-TIME VIDEO SURVEILLANCE WITH LIGHTING VARIATIONS

Hui-Chi Zeng and Shang-Hong Lai

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
{mr944327,lai}@cs.nthu.edu.tw

ABSTRACT

In this paper we present an adaptive foreground object extraction algorithm for real-time video surveillance. The proposed algorithm improves the previous Gaussian mixture background models (GMMs) by applying a two-stage foreground/background classification procedure to remove the undesirable subtraction results due to shadow, automatic white balance, and sudden illumination change. The traditional background subtraction technique usually cannot work well for situations with lighting variations in the scene. In the proposed two-stage classification, an adaptive classifier is applied to the foreground pixels in a pixel-wise manner based on the normalized color and brightness gain information. Secondly, the remaining foreground candidate pixels are grouped into regions and the corresponding background regions are compared to check if they are foreground regions. Experimental results on some real surveillance video are shown to demonstrate the robustness of the proposed adaptive foreground extraction algorithm under a variety of different environments with lighting variations.

***Index Terms*—** Real-time, surveillance, background subtraction, foreground extraction, lighting variation

1. INTRODUCTION

The main goal of video surveillance is to detect the foreground objects, and background subtraction is the most fundamental and common approach to achieve this goal. In recent years, several different background subtraction techniques are presented. Tuzel, et al. [1] used a Bayesian approach to background modeling. They defined each pixel as a mixture of multivariate Gaussian distributions and estimated the means and covariances of all Gaussian functions from a period of background video frames. Elgammal et al. [2] proposed a non-parametric model for background subtraction. The recent samples of intensity values for each pixel are used to compute the non-parametric probability density function. The drawback of this method is that it requires a considerable amount of memory to store the probability density functions. Stauffer and Grimson [3] proposed to use a mixture of Gaussian functions to model the intensity distribution of each background pixel, and the background model can be gradually adapted to the temporal intensity changes.

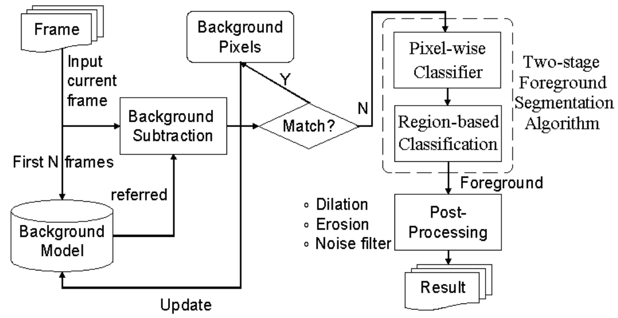


Fig. 1. The system flow chart.

After background subtraction, the subtracted non-background pixels include the foreground objects and background pixels with intensity changes caused by lighting variations or auto white balance. Some shadow detection methods have been proposed in the past. In Porikli and Thornton's work [4], they apply a shadow weak classifier as a pre-filter first, then model the selected shadow pixels using multivariate Gaussians. Huang et al. [5] first segmented each frame into regions based on motion similarity. The intensities of the shadow regions are assumed to be similar to those of the corresponding background regions by a scale. They estimate the scale to determine if a region belongs to a shadow region. Elgammal et al. [2] used the chromaticity coordinates r,g and the ratio of the lighting descent information for shadow detection. Tian et al. [6] presented a normalized cross-correlation algorithm for shadow removal, but it is time-consuming and it can not work well with homogeneous regions.

In many cases, the lighting changes or the auto white balance function is the video camera makes the background modeling very difficult, thus leading to unsatisfactory background subtraction results. In the proposed foreground extraction algorithm, as shown in Figure 1, we employ the mixture of Gaussians approach [3] to model the background, followed by a proposed two-stage procedure for classifying foreground and background pixels under lighting variations. The first step of our algorithm involves using a classifier to pixel-wisely classifying pixels to background or foreground based on the normalized color and intensity gain information. In the second step, we group the remaining pixels into regions based on their gain values and compare their regional color features with those computed from the corresponding background model to decide if the region is background or foreground.

The rest of this paper is organized as follows: In section 2, we describe the background modeling method based on the Gaussian mixture models [3]. The 2-stage foreground/background classification algorithm is discussed in section 3. Some experimental results are shown to demonstrate the robustness of the proposed algorithm under different lighting variation environments in section 4. Finally, we conclude this paper in section 5.

2. MIXTURE OF GAUSSIANS MODEL FOR BACKGROUND SUBTRACTION

Stauffer and Grimson [3] propose a mixture of K Gussian distributions (K is a small number from 3 to 5) to model the intensity distribution for each pixel. Assume the history of a particular pixel, $\{x_0, y_0\}$, at any time t be given by

$$\{X_1, ..., X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\}, \tag{1}$$

where I is the image sequence. $\{X_1, ..., X_t\}$ is modeled by a mixture of K Gaussian distributions, and the probability of the observed pixel with value X at time t is estimated as:

$$P(X) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X, \mu_{i,t}, \Sigma_{i,t}), \tag{2}$$

where $\mu_{i,t}$ and $\Sigma_{i,t}$ are the mean value and the covariance matrix of the i^{th} Gaussian in the mixture model at time t, and

$$\eta(X, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_t)^T \Sigma^{-1} (X-\mu_t)}, \tag{3}$$

and

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

is the estimated weight of the i^{th} Gaussian in the mixture models at time t, where α is the learning rate, and $M_{i,t}$ is 1 for the matched Gaussian and 0 for the remaining Gaussians. The update equations of μ_t and ${\sigma_t}^2$ are as follows:

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X, \tag{5}$$

$${\sigma_t}^2 = (1-\rho){\sigma_{t-1}}^2 + \rho(X-\mu_t)^T(X-\mu_t), \tag{6}$$

where $\rho = \alpha\eta(X|\mu_i, \sigma_i)$. For computational reasons, the red, green, and blue pixel values are assumed to be independent and have the same variances, so the form of the covariance matrix is $\Sigma_{i,t} = {\sigma_i}^2 I$. The K Gaussian distributions are ordered by the value of ω/σ, and the first B distribution are used as the background model, where

$$B = \arg\min_b(\sum_{i=1}^{b} \omega_i > T). \tag{7}$$

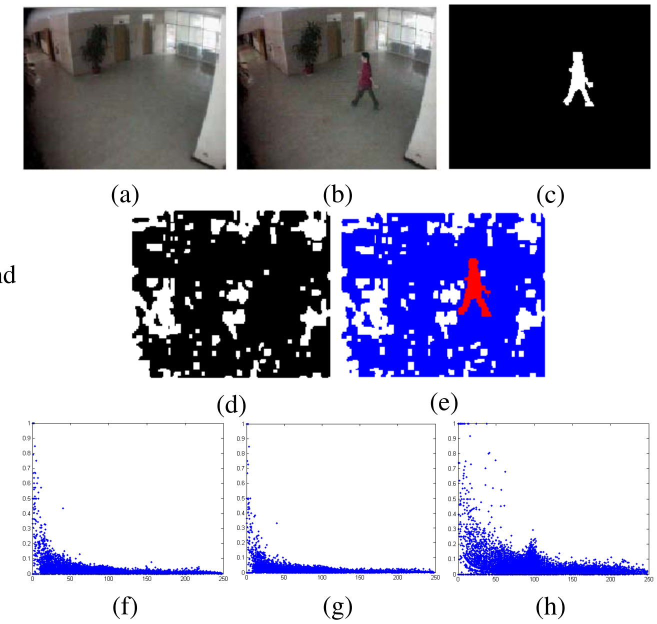


Fig. 2. The light change information retrieval. (a) The estimated background. (b) The current frame. (c) The ground truth. (d) Black pixels are results after background subtraction. (e) Red pixels are real foreground objects, and blue pixels are light changing pixels. (f)(g)(h) The distributions of the absolute r, g difference between current pixel and corresponding background pixel and $|gain|$ of the blue pixels in (e), and the x-axis is the background intensity$(0 \sim 255)$.

The threshold T is the minimum portion of the total weight given to background model. Foreground pixels are the pixels which are more than 2.5 standard deviations away from any of the B distributions. For more details, we refer to [3]. In our implementation, the parameters K and T are set to 3 and 0.4, respectively.

3. TWO-STAGE FOREGROUND SEGMENTATION ALGORITHM

The light change includes brightening and darkening, which may be due to illumination changes, shadowing or white balance. For bath cases, the influence is not on some individual independent pixels, but on a semi-transparent and gradually growing region. In the chromaticity coordinates, a pixel caused by the light change is considered unaffected. Let the red, green, and blue values of a pixel be R, G, and B. The chromaticity coordinates of the pixel, r, g, and b, are

$$r = \frac{R}{R+G+B},\ g = \frac{G}{R+G+B},\ b = \frac{B}{R+G+B}, \tag{8}$$

and $r+g+b = 1$. Thus, let the chromaticity coordinates of the background model of a pixel be r_b, g_b, and b_b and of the observed pixel value be r_o, g_o, and b_o. These three pairs should be very similar. That is,

$$r_b \sim r_o,\ g_b \sim g_o,\ b_b \sim b_o. \tag{9}$$

However, the chromaticity coordinates of the pixels in the dark area can vary a lot even the light changing slightly in the RGB space since the value $R+G+B$ in equation (8) is small. Thus, we provide a 2-stage algorithm to alleviate the problem due to light changes. Firstly, a pre-learned classifier is used to pixel-wisely remove the pixels with slight light changes, which is done by the background subtraction, especially the pixels in the dark area. Secondly, the remaining pixels are segmented into regions according to the gain, that is the ratio between the light change and the corresponding background value, given by

$$gain = \frac{I_o - I_b}{I_b}, \tag{10}$$

where I_b is the background model intensity and I_o is the observed pixel intensity. The pixels with similar gain are grouped into a region. Then, the average values $\overline{r}_o$ and $\overline{g}_o$ of each region are compared with their corresponding background average values $\overline{r}_b$ and $\overline{g}_b$ to determine if the region is foreground or background with light changes.

4. PIXEL-WISE CLASSIFIER

Figure 2(b)(c) are the 14-th frame and its ground truth image in the test video 1 released from the IPPR contest[1]. We collect a lot of images that are manually labeled with foreground and background regions as depicted in Figure 2(c). The blue area in Figure 2(e) shows the pixels passing the background subtraction due to white balance and shadow of the foreground object. This information is taken to learn the classifier for pixel-wisely removing the background pixels with light changes. We generate three distributions to find the relation between the light change and the background model intensity. The first two distributions are related with these two equations

$$d_r = |r_o - r_b| \tag{11}$$

and

$$d_g = |g_o - g_b|. \tag{12}$$

Figure 2(f) and (g) are the distributions of d_r and d_g to the background model intensity, respectively. However, the disadvantage of chromaticity coordinates is the lack of lightness information. So we add an extra component, $gain$, into the decision. The distribution of the absolute gain, $|gain|$, is shown in Figure 2(h). We can see the distribution of the background pixels under different lighting changes bounded by a decreasing envelop function of the following form

$$y_k = a_k + b_k \times e^{-c_k I_b}, \tag{13}$$

[1]IPPR contest: http://archer.ee.nctu.edu.tw/contest/

Table 1. The accuracy of the proposed algorithm on 3 IPPR contest test video sequences.

	Data1	Data2	Data3	Average
Total error pixels	47927	43085	22970	37994
Error pixels per frame	319	287	153	253
Accuracy rate (%)	99.6%	99.6%	99.8%	99.7%

where a_k, b_k and c_k are the parameters to be determined from the distribution, k = 'r', 'g' or 'gain', I_b is the corresponding background pixel intensity, and the function y provides the boundary for foreground/backgroundclassification. The current pixel is classified as a background pixel if it satisfies all of the following conditions:

$$d_r < y_r,\ d_g < y_g,\ |gain| < y_{gain}. \tag{14}$$

5. REGION-BASED CLASSIFICATION

After the pixel-wise classification, the remaining pixels are the foreground objects and the pixels suffering strong lighting changes. Based on their gain values, these pixels are grouped into regions based on a region growing technique. The pixels belonging to the same light change should have similar $gain$ values even if they are of different color. Thus, they are grouped into a region. We compare the average values $\overline{r}_o$ and $\overline{g}_o$ of each region with their corresponding background average values $\overline{r}_b$ and $\overline{g}_b$. In addition, the average $gain$ of the region, $\overline{gain}$, is also taken into the classification. The variation of the intensity in the background region due to light changes should not be too large. Thus, a background region should satisfy the following conditions:

$$\overline{r}_o \sim \overline{r}_b,\ \overline{g}_o \sim \overline{g}_b,\ \overline{gain} < T_{gain}, \tag{15}$$

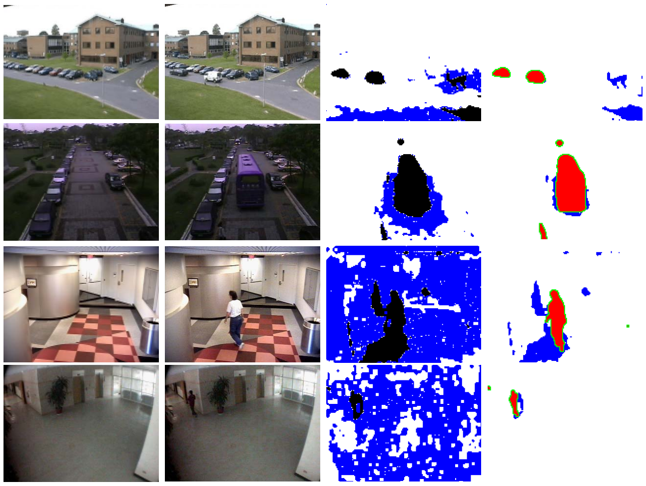
where T_{gain} is a threshold. In our experiment, $T_{\overline{gain}}$ is set to 0.5.

6. EXPERIMENTAL RESULTS

The proposed method is used for real-time video surveillance on a static web camera, and it processes about 24 frames per second for color images at size 320×240 on PC with a 3GHz Pentium IV CPU.

A robust initial background model is constructed in 5 seconds by using GMMs. Then our proposed robust background segmentation works well in a variety of environments with **the same parameter setting**.

The tested video sequences include different environments, such as outdoor scenes, indoor scenes with auto white balance, and light turning on/off. The proposed background segmentation algorithm provides satisfactory results in real-time. Some of the results are depicted in Fig. 3 and Fig. 4.



(a) (b) (c) (d)

Fig. 3. The name of the video from top to bottom : PetsD1TeC1 video of PETS 2001, Test data 3 of IPPR contest, IndoorGTTest2 video of IBM research, Test data 1 of IPPR contest. The upper two rows and lower two ones are the classified results on outdoor and indoor environments , respectively. (a) The background image. (b) The current frame. (c) The result of the pixel-wise classifier. The non-white pixels are the foreground region decided by the background subtraction. The blue pixels are the background with lighting changes classified by the pixel-wise classifier, while the black pixels are the classified foreground region. (d) The result after the region-based classification. The blue pixels are the region similar to the background. The red pixels are the foreground object and the green pixels are its boundary.

Furthermore, we evaluated the accuracy of our proposed method on 3 IPPR contest test video with ground truths. Each video sequences contains 150 frames at size 320×240. The accuracy of our algorithm is shown in Table 1. It is evident that the proposed algorithm can provide very accurate foreground/background segmentation results.

7. CONCLUSIONS

In this paper, we presented a robust foreground object extraction algorithm for real-time video surveillance under lighting variations. The proposed algorithm first employs a mixture-of-Gaussians model for background subtraction, followed by the proposed two-stage foreground/background segmentation algorithm. The first step is a pixel-wise foreground/background classifier, which is based on applying decreasing exponential curves as the separation function for foreground and background pixels based on the normalized color and gain values, respectively. The second step consists of a pixel grouping process and a region classification based on comparing the regional color features of the current and the background

(a) (b) (c) (d)

Fig. 4. The background subtraction results for the lab scene sequences. The top, second, and bottom rows show the cases with indoor diffusion shadow, automatic white balance, and turning light off,respectively. (a), (b), (c), and (d) depict the intermediate processing results, which are the same as those in Figure 3. The small blobs shown in (d) are caused by the movement of humans in the background.

model. Experimental results show the proposed foreground object extraction algorithm is robust against different types of lighting variation under different environments.

8. REFERENCES

[1] O. Tuzel, F. Porikli, and P.Meer, “A bayesian approach to background modeling,” *IEEE Workshop on Machine Vision for Intelligent Vehicles*, June 2005.

[2] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” *European Conference on Computer Vision*, June/July 2000.

[3] C. Stauffer and W.E.L. Grimson, “Adaptive background mixture models for real-time tracking,” *IEEE International Conference on Computer Vision and Pattern Recognition*, June 1999.

[4] F. Porikli and J.Thornton, “Shadow flow: A recursive method to learn moving cast shadows,” *IEEE International Conference on Computer Vision*, October 2005.

[5] S.S. Huang, L.C. Fu, and P.Y. Hsiao, “Region-level motion-based foreground detection with shadow removal using mrfs,” *Asian Conference on Computer Vision*, 2006.

[6] Y.L. Tian, M. Lu, and A. Hampapur, “Robust and efficient foreground analysis for real-time video surveillance,” *IEEE International Conference on Computer Vision and Pattern Recognition*, June 2005.