IMPROVED BACKGROUND MODELING THROUGH COLOR DE-CORRELATION

Jong Geun Park and Chulhee Lee

Dept. Electrical and Electronic Eng., Yonsei University University
134 Shinchon-Dong, Seodaemoon-Gu, Seoul 120-749, South Korea
phone: + (+82) 2-2123-2779, fax: + (82) 2-312-4584, email: chulhee@yonsei.ac.kr
http://hdsp.yonsei.ac.kr

ABSTRACT

Background modelling and foreground detection, which significantly affect the performance of intelligent visual surveillance systems, are challenging works due to dynamic background, illumination changes, image artefacts, etc. This paper describes an improved algorithm for background modelling. A pixel-wise non-parametric statistical model of the HSV colour components and gradients is used for background modelling. Since the non-parametric statistical model using the kernel density estimation is computationally complex, the probability density functions are estimated as a product of several one-dimensional histograms. Then, foreground regions are detected by using the Bayesian decision rule. The experimental results showed that the proposed algorithm produced more accurate and stable results than existing background modeling methods and the colour de-correlation procedure produced improvements.

1. INTRODUCTION

Foreground detection, which extracts regions of interest (objects) from images, is an important step in most intelligent visual surveillance systems. There are mainly three categories in these foreground detection methods: optical flow, temporal differencing, and background subtraction [1-2]. Optical flow methods are computationally complex and sensitive to noise, and temporal differencing methods frequently fail to extract the complete contours of foreground regions. Background subtraction is a popular foreground detection method since it overcomes the aforementioned problems. In background subtraction methods, background modelling is used. Since background modelling significantly affects the performance of foreground detection, it is important to use a good background model, which is robust against dynamic backgrounds, illumination changes, and shadows.

Stauffer et al. proposed a parametric background modelling method using a mixture of Gaussians (MOG) to represent RGB colour distributions at each pixel [3]. The method has been the basis model in many background modelling methods because the method is simple and fast. The MOG-based methods use predefined parameters to decide whether a pixel is from the background or the foreground region. Therefore, one drawback of the methods is that the performance may be heavily affected by the parameter selection. In order to overcome this problem, Lee et al. used the Bayesian decision rule instead of the predefined standard deviation [4]. Nevertheless, there are some limitations of the MOG method itself. The MOG methods use a K-means algorithm for parameter initialization. When the background involves many modes, a small number of Gaussian distributions may not efficiently model the background region.

Alternatively, nonparametric background modelling methods without any assumptions about the underlying distribution could be used. In [5], W^4 models background regions with maximum and minimum intensity values, and maximum temporal variations in colours. The surveillance system proposed by Li et al. represented the background region as a collection of the most significant and frequent features of colours, gradients, and colour-occurrences in the RGB colour space [6]. Elgammal et al. proposed a background modelling method using kernel density estimation [7]. Based on training samples, this method estimates the probabilities that a pixel is the background region. However, the method requires a long processing time. Park et al. addressed the problem representing the probability density functions for the background regions as a product of several one-dimensional histograms with the assumption that the correlation between channels of the feature vector is zero [8]. However, this assumption may not be always true. Based on this observation, we proposed an improved version of [8]. In the proposed algorithm, we first de-correlated channels of feature vectors using principle component analysis (PCA). Experimental results and comparisons with existing methods on various environments are provided.

2. THE BAYESIAN RULE FOR CLASSIFYING FOREGROUND AND BACKGROUND REGIONS [8]

First, the foreground/background classification based on the Bayesian rule of [8] is briefly described. Let $\mathbf{I}_n(x,y) = \left[I_n^R(x,y), I_n^G(x,y), I_n^B(x,y)\right]$ be an RGB colour image at time n and let $\mathbf{v}$ be a d-dimensional feature vector extracted from $\mathbf{I}_n(x,y)$. Using HSV colour and gradient information, a 4-dimensional feature vector $\mathbf{v}$ at $\mathbf{I}_n(x,y)$ was constructed as follows:

$$\mathbf{v} = \left[v_1, \cdots, v_4\right]^T = \left[h, s, v, e\right]^T, \quad (1)$$

where h, s and v represent the hue, saturation and value components of the HSV colour space. The last element (e) represents a gradient obtained from the RGB colour space:

$$e = \tan^{-1}\left(\frac{\sum_c \left(E_V * I_n^c(x,y) \right)}{\sum_c \left(E_H * I_n^c(x,y) \right)} \right) \qquad c \in \{R, G, B\}. \tag{2}$$

where the symbol $*$ denotes convolution. E_V and E_H represent the vertical and horizontal masks of an edge detector. Then the posterior probability that the feature vector $\mathbf{v}$ at a pixel comes from the background (foreground) region is given by

$$P^{x,y}(\omega \mid \mathbf{v}) = \frac{p^{x,y}(\mathbf{v} \mid \omega) P^{x,y}(\omega)}{p^{x,y}(\mathbf{v})} \qquad \omega \in \{\omega_b, \omega_f\}, \tag{3}$$

where ω_b and ω_f represent the background and foreground regions. $P^{x,y}(\omega)$ represents the prior probability of the background or foreground regions at (x, y) and $p^{x,y}(\mathbf{v})$ represents the probability density function of the feature vector. Using the Bayesian decision rule, pixel (x, y) is classified as background if

$$P^{x,y}(\omega_b \mid \mathbf{v}) > P^{x,y}(\omega_f \mid \mathbf{v}). \tag{4}$$

Substituting (3) into (4), the pixel is classified as a background region if the following condition was met:

$$p^{x,y}(\mathbf{v} \mid \omega_b) P^{x,y}(\omega_b) > p^{x,y}(\mathbf{v} \mid \omega_f) P^{x,y}(\omega_f). \tag{5}$$

3. THE MODELING OF THE STATISTICS BASED ON KERNEL DENSITY ESTIMATION

To use Eq. (5), the density functions and prior probabilities need to be estimated in advance. One can use nonparametric techniques that directly estimate the density function from the data such as kernel density estimation. Assuming that feature vectors $S = \{\mathbf{v}_i \mid i = 1,2,\cdots,n\}$ extracted from the first n frames, $\hat{p}^{x,y}(\mathbf{v} \mid \omega_b)$ can be estimated as follows:

$$\hat{p}^{x,y}(\mathbf{v} \mid \omega_b) = \frac{1}{n} \sum_{i=1}^{n} \varphi_{\mathbf{\Sigma}}(\mathbf{v} - \mathbf{v}_i), \tag{6}$$

where $\varphi_{\mathbf{\Sigma}}$ is a kernel function with a symmetric positive definite $d \times d$ bandwidth matrix $\mathbf{\Sigma}$ such that $\varphi_{\mathbf{\Sigma}}(\mathbf{v}) = |\mathbf{\Sigma}|^{-1/2} \varphi(\mathbf{\Sigma}^{-1/2}\mathbf{v})$. Typically, the normal distribution $N(0, 1)$ is used for the kernel function.

To avoid the long processing time of non-parametric kernel density estimation, we used multi-dimensional histograms to approximate density functions. However, a problem is that such multi-dimensional histograms require a large amount of memory. For a d-dimensional feature vector, if each element is quantized to L levels, the histogram contains L^d bins. Also, a limited number of samples are available for histogram estimation so that the number of training samples is inadequately small compared to the number of bins. In order to address the problems, the multi-dimensional histogram was approximated as a product of one-dimensional histograms assuming that the channels of the feature vector are independent [8]. However, this assumption may be unreasonable and potentially risk because the correlation between colour channels (elements of the feature vector) is relatively high.

In this paper, we make the elements of the feature vector uncorrelated by applying PCA. Let $\mathbf{\Phi}^{x,y} = \left[\boldsymbol{\varphi}_1^{x,y}, \cdots, \boldsymbol{\varphi}_d^{x,y}\right]$ be a matrix consisting of eigenvectors, $\boldsymbol{\varphi}_j^{x,y}$ ($j = 1, \cdots, d$), acquired from S at the location of (x,y). Then the uncorrelated vector, $\boldsymbol{\upsilon}_i$, can be obtained as follows:

$$\boldsymbol{\upsilon}_i = \mathbf{\Phi}^{x,y}(\mathbf{v}_i - \boldsymbol{\mu}_S), \tag{7}$$

where $\mathbf{v}_i$ is a feature vector and $\boldsymbol{\mu}_S$ is the mean of feature vectors. Since the transformed vector $\boldsymbol{\upsilon}_i$ is now uncorrelated, the multi-dimensional histogram of $\boldsymbol{\upsilon}_i$ can be approximated as a product of one-dimensional histograms as follows:

$$\hat{p}^n_{\boldsymbol{\upsilon}|\omega_b}(\mathbf{k}) = \prod_{j=1}^{d} \hat{p}^n_{\langle \boldsymbol{\upsilon} \rangle_j | \omega_b}(k_j), \tag{8}$$

where

$$\hat{p}^n_{\langle \boldsymbol{\upsilon} \rangle_j | \omega_b}(k_j) = \frac{1}{n} \sum_{i=1}^{n} \varphi_{\sigma_j} \left(\frac{\langle \mathbf{c}_\mathbf{k} - \boldsymbol{\upsilon}_i \rangle_j}{\sigma_j} \right). \tag{9}$$

where $\mathbf{c}_\mathbf{k}$ is the centre of the hypercube represented by vector $\mathbf{k} = [k_1, \cdots, k_j, \cdots, k_d]^T$ in the transformed feature space. In other words, we divided the transformed feature space into a number of hyper-cubes and estimated the probability density at the centre of each cube. σ_j is a bandwidth for each dimension and $\langle \mathbf{x} \rangle_j$ is defined as the j-th element of vector $\mathbf{x}$.

We also construct the histograms for the pdf of the foreground region: $p^{x,y}(\mathbf{v} \mid \omega_f)$. Since the foreground region may exhibit arbitrary colours and gradients, we assume that

the underlying densities of $p^{x,y}(\mathbf{v} \mid \omega_f)$ have a uniform distribution:

$$\hat{p}^n_{\mathbf{v}|\omega_f}(\mathbf{k}) = \prod_{j=1}^{d}\left(\frac{1}{K_j}\right). \quad (10)$$

where K_j represents the number of bins of each dimension. Initially, the prior probabilities, $P^{x,y,n}(\omega)$ ($\omega \in \{\omega_b, \omega_f\}$), are set to 0.5. Finally, we define a reference background image $\mathbf{B}_n(x,y) = \left[B_n^R(x,y), B_n^G(x,y), B_n^B(x,y)\right]$ as follows:

$$\mathbf{B}_n(x,y) = \underset{i\in\{1,2,\cdots,n\}}{\mathbf{med}} \mathbf{I}_i(x,y), \quad (11)$$

where **med** represents a pixel-wise median filtering operation, which is used to exclude the pixels of moving foreground objects in the training images.

4. BACKGROUND SUBTRACTION

After computing the histograms for the background regions from *n* training images, foreground detection for a new entering image is performed using the Bayesian decision rule [8]. Then the statistics (the histograms and the prior probabilities) are updated to adapt to gradually changing backgrounds.

4.1. Foreground detection

We first compute the difference image between a newly entering image and the reference background image and then apply a threshold operation:

$$F_n(x,y) = \begin{cases} 1, & if \ \underset{c\in\{R,G,B\}}{\vee}\left(\left| I_n^c(x,y) - B_n^c(x,y)\right| > T_c\right) \\ 0, & otherwise \end{cases}, \quad (12)$$

where $F_n(x,y)$ is a foreground candidate binary image, T_c is an adaptive threshold value for each colour channel generated using [9] and the symbol $\vee$ represents a logical OR operation. For each foreground candidate pixel (i.e., $F_n(x,y) = 1$), we extract a feature vector $\mathbf{v}$ and transform it as follows:

$$\boldsymbol{\upsilon} = \boldsymbol{\Phi}^{x,y}\left(\mathbf{v} - \boldsymbol{\mu}_S\right), \quad (13)$$

Then the pixel is changed to a background region if the following rule is met:

$$\hat{P}^n(\omega_b)\cdot\prod_{j=1}^{4}\hat{p}^n_{\langle\boldsymbol{\upsilon}\rangle_j|\omega_b}(\kappa_j) > \hat{P}^n(\omega_f)\cdot\prod_{j=1}^{4}\left(\frac{1}{K_j}\right) \quad (14)$$

where κ_j is the bin index that $\langle\boldsymbol{\upsilon}\rangle_j$ falls into. After classifying all the pixels, there may have been several small holes inside the foreground regions and noises in the background regions. Since most of the noise regions are small in size, these erroneous regions can be easily eliminated by applying morphological operations.

4.2. Updating of the statistics and background image

The prior probabilities and histograms must be updated to adapt to gradually changing backgrounds as new images are entered. We use a simple IIR filter to update the prior probabilities and histograms for background. If $\mathbf{I}_n(x,y)$ is classified as a background region, the statistics are updated as follows:

$$\begin{aligned} &\hat{P}^{n+1}(\omega_b) = (1-\alpha)\hat{P}^n(\omega_b) + \alpha\ , \\ &\hat{p}^{n+1}_{\langle\upsilon\rangle_j|\omega_b}(\kappa_j) = (1-\alpha)\hat{p}^n_{\langle\upsilon\rangle_j|\omega_b}(\kappa_j) + \alpha \quad j \in \{1,\cdots,d\}. \end{aligned} \quad (15)$$

where α represents a learning rate which determines the adaptation speed. A large value of α allowed the background model to adapt quickly to a new image sequence. After updating, the prior probabilities and histograms were normalized. The background image was also updated as follows [8]:

If $\mathbf{I}_n(x,y)$ is classified as a background region,
then $\mathbf{B}_{n+1}(x,y) = (1-\alpha)\cdot\mathbf{B}_n(x,y) + \alpha\cdot\mathbf{I}_n(x,y)$. (16)
Otherwise, $\mathbf{B}_{n+1}(x,y) = \mathbf{B}_n(x,y)$.

5. EXPERIMENTAL RESULTS

5.1. Data sets

The proposed algorithm was tested with the image sequences in [6]. The data set[1] was acquired using a fixed camera in indoor and outdoor environments and presented potentially problematic scenarios (dynamic background, sudden illumination changes, etc) for background modelling. The data set includes nine image sequences (buffet restaurant: BR, campus: CAM, meeting room: MR, subway station: SS, fountain: FT, airport: AP, lobby: LB, shopping centre: SC, and water surface: WS). The data set also provides the ground truth data of randomly selected 20 images for each image sequence (a total of 180 ground truth images). The image resolutions range from 160×120 to 320×256 pixels.

5.2. Quantitative and qualitative evaluations

The proposed method requires several parameters ($K_{j=1,\cdots d}$: the number of bins, $\sigma_{j=1,\cdots,d}$: the bandwidth of the kernel functions, and α : learning rate). In our tests, the parameters K_j, σ_j, and α were fixed to 200, 3.5, and 0.02, respectively. The Jaccard similarity is used for quantitative performance assessment and defined as follows:

[1] Available at a website http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html.

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}, \quad (17)$$

where S_1 represents the foreground pixel set obtained by an algorithm and S_2 represents the pixel set of the ground truth data. Table 1 shows the Jaccard similarities of the proposed method and other methods. The overall performance of the proposed method was better than that of other methods.

Figs. 1-4 show experimental results for the SC, CAM, SS, and LB image sequences: a test image (top, left), the ground truth data of the test image (top, right), the results of the proposed method (bottom, left), and the results of Li's method (bottom, right). In indoor environments, there are significant specular reflections and shadows on the floor. These cause illumination changes. The three sequences (B, AP, and SC) were captured in such environments. The proposed algorithm produced more satisfactory results and was more robust against shadows than the Li's method since the proposed method used the HSV colour space [2, 8].

The proposed method also provided good results for the sequences with dynamic background regions (CAM, MR, FT, and WS). These sequences included swaying tree branches, moving curtains due to winds, water streams, and rippling water surfaces. The experimental results show that the proposed method effectively handled such difficult scenes and more accurately detects foreground regions than the Li's method. On the other hand, the Li's method showed slightly better performance than the proposed method in the MR image sequence since the proposed method lost a part of the person whose clothes' colour was similar to the curtain colour. Also, the proposed method incorrectly classified a part of dynamic curtains as foreground regions.

The SS sequence shows dynamic background regions (moving escalators). The proposed method successfully adapted to these dynamic background regions, but occasionally failed to detect colours similar to the escalator colour. The Li's method also failed in this case. Furthermore, there are sudden background illumination changes caused by the auto gain controlling of video recording devices.

The main difficulty in the LB sequence is that the background region suddenly changed due to light switching. The LB images in Fig. 4 show the situation when some lights were switched off. In this case, the proposed method incorrectly classified some background regions as foreground regions due to these sudden changes while the Li's method correctly handled this problem.

Compared to the Park's method [8], the proposed method shows some improvement due to the proposed de-correlation procedure.

6. Conclusions

In this paper, we proposed a background modelling method using nonparametric kernel density estimation based on the Bayesian theory were proposed. We further used the de-correlation procedure to improve performance. Experimental results using the data sets obtained in indoor and outdoor environments showed that the proposed method is robust for complex, dynamic, multi-modally background and produced more accurate and stable results than other background modelling method. Furthermore, the de-correlation procedure produced some improvement compared to [8].

References

[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., 34(3), pp. 334-352, 2004.

[2] Y. Shan and R. Wang, "Improved algorithms for motion detection and tracking," Opt. Eng. 45(6), 067201, 2006.

[3] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. Pattern Anal. Mach. Intell. 22(8), pp. 747-757, 2000.

[4] D. S. Lee, J. J. Hull, and B. Erol, "A Bayesian framework for Gaussian mixture background modelling," Proc. IEEE Intl. Conf. on Image Process., pp. 973-976 2003.

[5] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-time surveillance of people and their activities," IEEE Trans. Pattern Anal. Mach. Intell. 22(8), pp. 809-830, 2000.

[6] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Statistical modelling of complex backgrounds for foreground object detection," IEEE Trans. Image Process. 13(11), pp. 1459–1472, 2004.

[7] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modelling using nonparametric kernel density estimation for visual surveillance," Proc. IEEE 90, pp. 1151-1163, 2001.

[8] J. G. Park, and C. Lee, "Bayesian rule-based complex background modelling and foreground detection," SPIE Optical engineering, 49(2), 027006 (February 2010).

[9] P. L. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection," Proc. British Machine Vision Conf., pp. 347-356, 1995.

Table 1. Performance comparison of the proposed method and other methods.

Image sequence	Proposed	Park [8]	Li [6]	MOG [3]
BR	**0.686**	0.648	0.564	0.358
CAM	0.779	**0.794**	0.683	0.48
MR	0.900	0.899	**0.911**	0.445
SS	0.641	**0.649**	0.534	0.277
FT	0.791	**0.792**	0.674	0.663
AP	**0.692**	0.677	0.508	0.28
LB	0.693	0.671	**0.706**	0.421
SC	**0.710**	0.683	0.645	0.423
WS	0.892	**0.900**	0.851	0.536
Average	**0.754**	0.746	0.675	0.431

Fig. 1. Experimental results for the SC image sequence.

Fig. 2. Experimental results for the CAM image sequence.

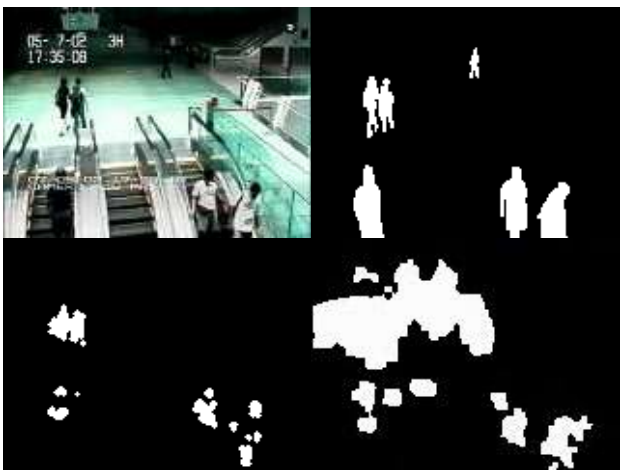

Fig. 3. Experimental results for the SS image sequence.

Fig. 4. Experimental results for the LB image sequence.