BACKGROUND SUBTRACTION BASED ON PHASE AND DISTANCE TRANSFORM UNDER SUDDEN ILLUMINATION CHANGE

Gengjian Xue, Jun Sun, Li Song

Institute of Image Communication and Information Processing,
Shanghai Jiao Tong University, Shanghai, China, 200240
Shanghai Key Lab of Digital Media Processing and Transmission, Shanghai, China, 200240
xgjsword@sjtu.edu.cn, sunjun@cdtv.org.cn, song_li@sjtu.edu.cn

ABSTRACT

Effective foreground detection under sudden illumination change is an active research topic. However, most existing background subtraction approaches, which are intensity based, fail to handle this situation. In this paper, we propose a novel background modeling method that overcomes this limitation by relying on statistical models which use pixel phase instead of intensities. We first extract the phase feature of the pixel using Gabor filters. Then, a phase based background subtraction approach is proposed. In this approach, each phase feature is modeled independently by a mixture of Gaussian models and updated with a novel scheme. Since foreground pixels are scattered in the preliminary detection result, distance transform is implemented on the binary image which transforms the image into a distance map. We segment the distance image with a threshold and get the final result. Experiments on two challenging sequences demonstrate the effectiveness and robustness of our method.

Index Terms— Background subtraction, sudden illumination change, phase, distance transform

1. INTRODUCTION

Moving object detection is an active research subject in computer vision, with several applications such as industrial automation, security and surveillance. Its output is as an input to a high-level process, making it a critical part of the system. A popular method is background subtraction, which consists of obtaining a mathematical model of the static background and comparing it with every new frame from the video sequence. Although many methods have been presented, it is still a challenging task when illumination changes suddenly, such as turning on or off the light.

Background modeling using pixel intensities is a common approach. One of the most widely used methods is the Gaussian Mixture Model(GMM) [1]. In this technique, each pixel is modeled independently using a mixture of Gaussian models and updated by an online approximation. Elgammal et al. [2] proposed another way to represent the background by a nonparametric kernel density estimation (KDE). Recently, Heikkilä and Pietikäinen [3] modeled the background using histograms of local binary patterns (LBP). However, these approaches have proved to be effective only at handling gradual illumination changes and repetitive dynamic backgrounds.

Some authors dealt with this problem in other ways. In [4], the Wallflower system maintains several background models representing different illumination conditions and switches between the learned background when a large number of pixels are detected as foreground. But this approach is very pragmatic. The algorithm is able to handle sudden illumination change only if the model describing the scene after the illumination changes is known a priori. In [5], Binglong Xie et al. assumed the order of pixel values is preserved in local neighborhoods when illumination changes and the probability of order consistency for each pixel is classified to provide the output image. Mittal and Ramesh [6] proposed to combine intensity and rank information to handle illumination changes. But these methods can not handle complex nonlinear brightness changes. Pilet et al. [7] proposed to generate a statistical model combing two texture-based features and the ratios of the current image to the background image in each color channel. However, this method requires priori knowledge. It involves manual extraction of foreground areas during training for modeling the background and foreground distributions.

Phase information, which is as an important feature of the pixel, has the advantage of being insensitive to illumination changes, making it suitable for background modeling. In this paper, we propose a phase based background subtraction approach under sudden illumination change. We first extract phase feature of the pixel using Gabor filters. Then we propose our background modeling method, where each phase feature of the pixel is modeled by a mixture of Gaussian models and updated by a novel scheme. Next, because foreground pixels are scattered in the preliminary result, we transform

This work was supported in part by NSFC (60702044, 60625103, 60632040), MIIT of China (2010ZX03004-003), 973 program (2010CB731401, 2010CB731406) and 863 program (2008AA01A318).

the initial binary image into a distance map using distance transform. After that, we segment the distance image and get the final result through morphology operations. Experimental results indicate that our approach is more robust to sudden illumination changes than two representative methods.

The rest of the paper is organized as follows: In Section 2, we describe the phase extraction, the phase based background modeling method and the blob aggregating technique based on distance transform. Experiments and analysis are the given in Section 3. Finally, conclusions are drawn in Section 4.

2. PROPOSED APPROACH

In this section, we describe our approach to background subtraction. Our algorithm can be divided into three stages. Phase extraction and phase based background modeling are introduced in Section 2.1 and 2.2. In Section 2.3, blob aggregating based on distance transform is described.

2.1. Phase extraction

Since pixel intensities vary drastically during sudden illumination change while phase is insensitive to it, we adopt the pixel phase as feature to model the background. We extract the phase feature of each pixel using Gabor filers, for they have the property of optimal joint localization in both the spatial and the spatial-frequency domains. A two-dimensional Gabor function consists of a sinusoidal plane wave of some frequency and orientation, modulated by a two-dimensional Gaussian envelope. The Gabor waveletes can be defined as follows [8]:

$$\psi_{\mu,v}(z) = \frac{||k_{\mu,v}||^2}{\sigma^2} e^{(-||k_{\mu,v}||^2||z||^2/2\sigma^2)} \left[e^{ik_{\mu,v}z} - e^{-\sigma^2/2}\right] \tag{1}$$

where $\overrightarrow{k_{\mu,v}} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v cos\phi_\mu \\ k_v sin\phi_\mu \end{pmatrix}, k_v = f_{max}/2^{v/2}$, $\phi_\mu = \mu(\pi/\mu_{max}), v = 0, \ldots, v_{max}-1, \mu = 0, \ldots, \mu_{max}-1$, v is the frequency and μ is the orientation. The first term in the square brackets in (1) determines the oscillatory part of the kernel and the second term compensates for the DC value. σ determines the ratio of the Gaussian window width to wavelength. The Gabor transformation of a given image is defined as its convolution with the Gabor functions:

$$G_{\mu,v}(z) = I(z) * \Psi_{\mu,v}(z) \tag{2}$$

where the symbol "$*$" represents the convolution operator, $z = (x, y)$ denotes the image position and $G_{\mu,v}(z)$ is the convolution result corresponding to the Gabor kernel at the scale v and orientation μ. The Gabor wavelet coefficient $G_{\mu,v}(z)$ is a complex, which can be rewritten as:

$$G_{\mu,v}(z) = A_{\mu,v}(z) \cdot exp(i\theta_{\mu,v}(z)) \tag{3}$$

with one amplitude item $A_{\mu,v}(z)$ and one phase item $\theta_{\mu,v}(z) \in [0, 2\pi)$. The image is filtered with a set of Gabor filters with different preferred orientations and spatial frequencies that cover appropriately the spatial frequency domain, and each pixel has $v_{max} \times \mu_{max}$ wavelet coefficients after convolutions. The larger amplitude of the coefficient is, the more efficient the coefficient represents the pixel information and the more discriminative of the corresponding phase is. Because the phase value has the rotation property and the range of single phase value is small, we adopt these phase information which correspond to the top L largest amplitudes as the phase feature of the pixel. The phase feature is defined as follows:

$$p(z) = \sum_{i=1}^{L} \theta_i(z) \tag{4}$$

where z denotes the image position, $p(z)$ denotes the phase feature of the pixel, $\theta_i(z)$ means the Gabor phase corresponding to the i_{th} largest Gabor amplitude of the pixel and L denotes the number of the Gabor phase to be extracted.

In order to get more accurate information, we divide the image into $N \times N$ subregions and execute the Gabor convolution in each subregion identically. In our experiment, we set values $N = 4$, $L = 3$ and define the Gabor filters with $v_{max} = 4$ and $\mu_{max} = 6$. Thus, the feature $p(z) \in [0, 6\pi)$.

2.2. Phase based background modeling

Based on the feature $p(z)$, we propose a phase based background subtraction method under sudden illumination change. In the proposed technique, each pixel is modeled independently by a mixture of K Gaussian distributions where each Gaussian represents the phase distribution. Let the k_{th} Gaussian distribution in the mixture be represented by its mean μ_k and variance σ_k^2, and its weight in the mixture be denoted by ω_k such that $\sum_1^K \omega_i = 1$.

When a new phase feature X_t is coming, it is first matched against the existing models to find whether there is a model (say the k_{th}) corresponding to the following conditions. There exists singular regions near 0 and 6π due to the phase rotation property, so we define the matching conditions as follows:

$$\begin{cases} |6\pi - \mu_k + X_t| \le 2.5\sigma_k & \text{if } X_t < \varepsilon \text{ and } 6\pi - \mu_k < \epsilon \\ |6\pi + \mu_k - X_t| \le 2.5\sigma_k & \text{if } \mu_k < \varepsilon \text{ and } 6\pi - X_t < \epsilon \\ |X_t - \mu_t| \le 2.5\sigma_k & \text{others} \end{cases} \tag{5}$$

where ϵ denotes the limit value of the singular region. If such a model exists, its associated parameters are updated in three different ways:

if $\mu_k < \varepsilon$ and $6\pi - X_t < \epsilon$

$$\begin{cases} \mu_k = \mu_k + \alpha(X_t - \mu_k - 6\pi) \\ \sigma_k^2 = (1-\alpha)\sigma_k^2 + \alpha(X_t - \mu_k - 6\pi)^2 \end{cases} \tag{6}$$

if $X_t < \varepsilon$ and $6\pi - \mu_k < \epsilon$

$$\begin{cases} \mu_k = \mu_k + \alpha(X_t - \mu_k + 6\pi) \\ \sigma_k^2 = (1-\alpha)\sigma_k^2 + \alpha(X_t - \mu_k + 6\pi)^2 \end{cases} \tag{7}$$

others:

$$\begin{cases} \mu_k = \mu_k + \alpha(X_t - \mu_k) \\ \sigma_k^2 = (1-\alpha)\sigma_k^2 + \alpha(X_t - \mu_k)^2 \end{cases} \tag{8}$$

the weights are updated as:

$$\omega_k = (1-\alpha)\omega_k + \alpha M_k \tag{9}$$

where α is the learning rate, M_k is 1 for the matched Gaussian and 0 for the others. When the value of μ_k is beyond the scope $[0, 6\pi)$, we have:

$$\begin{cases} \mu_k = \mu_k - 6\pi & \text{if } \mu_k \geq 6\pi \\ \mu_k = \mu_k + 6\pi & \text{if } \mu_k < 0 \end{cases} \tag{10}$$

If there is no matching model, the Gaussian with the lowest ω is replaced by a new Gaussian with $u = X_t$, $\sigma = \sigma_{init}$ and low ω_{init}.

In the foreground detection stage, the Gaussians are sorted in descending order according to the values ω/σ. The first B distributions are chosen as background distributions:

$$B = argmin(\sum_{k=1}^{b} \omega_k > T) \tag{11}$$

where T is a threshold that represents the minimum prior probability that the background is in the scene. If the phase feature matches any one of the B distributions, the corresponding pixel is classified as background, otherwise the pixel is marked as foreground.

2.3. Distance transform based blob aggregating

Because foreground pixels are scattered in the preliminary result, we change the binary image into a distance map using distance transform. The distance transform, which was first introduced by Rosenfeld and Pfaltz [1966], is a general operator forming the basis of many methods in computer vision and geometry, with great potential for practical applications. It converts an binary image of black and white pixels into a representation where each pixel has a value indicating its distance to the nearest white pixel. When the pixel value is 1, the distance of the pixel is 0. We represent a $M \times M$ binary image by $A = \{(i,j) : a(i,j) = 0 \text{ or } 1\}$, for $i, j = 1, \ldots, M$, $W = \{(x,y) : w(x,y) = 1\}$ represent the coordinates of the white pixels of the image. The Euclidean distance transform of pixel $a(i,j)$ is defined as

$$d(i,j) = \min_{(x,y)\in W} \sqrt{(i-x)^2 + (j-y)^2} \tag{12}$$

We then segment the distance image with a threshold to get a more integrated foreground. After noise removing using morphology operations, we obtain the final result. An example of the results before and after blob aggregation is show in Fig. 1.

Fig. 1. Results before and after blob aggregation

3. EXPERIMENTS AND ANALYSIS

In this section, we show results on individual frames of video sequences that feature sudden illumination change. Both visual and numerical methods are used to evaluate our method. Two widely used methods, the GMM [1] and LBP [3] are employed to compare with the proposed approach.

The first sequence is from [4], which is considered as the light switch benchmark. The room is dark at first, then a person enters the room and turns on the light. As shown in Fig.2, GMM and LBP methods can not accurately detect moving objects under sudden illumination change. They detect a large number of background pixels as foreground and classify a amount of foreground pixels as background. Compared with them, our method gives accurate detection and produces good result. The second sequence is captured by ourself. The light is on, then a person goes over and turns off the light. Fig.3 shows the corresponding results produced by the three methods. The GMM and LBP methods can not adapt to sudden illumination change and present incorrect results, whereas our method is more robust and the result produced by ours is much better. The reason for these is that our method uses phase feature for background modeling, which is intensive to illumination change. Table.1 and Table.2 show the performance evaluation of the three methods, which demonstrate that our method outperforms the other two. Here, we use precision and recall as the metrics: (TP: True Positive, FP: False Positive, FN: False Negative)

$$Precision(\%) = \frac{TP}{TP + FP} \times 100 \tag{13}$$

$$Recall(\%) = \frac{TP}{TP + FN} \times 100 \tag{14}$$

Table 1. Precision and recall on the sequence from [4]

	Precision	Recall
GMM	22.03	76.40
LBP	18.40	73.05
Our method	78.53	97.48

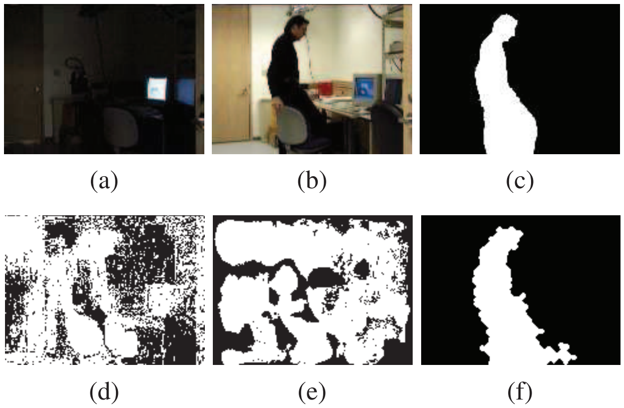

Fig. 2. Detection results for the light switch sequence from [4]. (a) Input image. (b) Test image. (c) Manually segmented ground truth. (d) The output of the GMM method. (e) Result obtained by the LBP method. (f) Result obtained by our method.

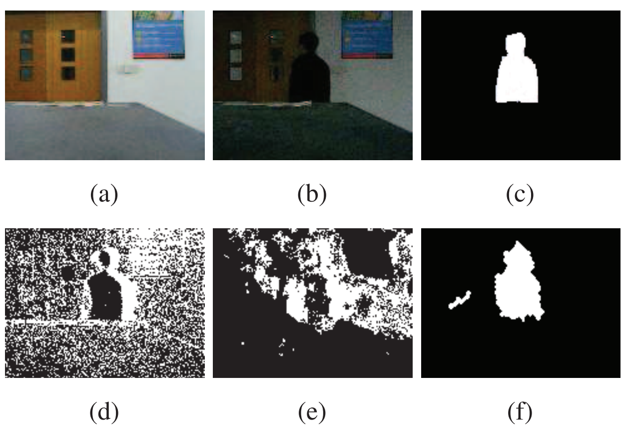

Fig. 3. Detection results for our sequence. (a) Input image. (b) Test image. (c) Manually segmented ground truth. (d) The output of the GMM method. (e) Result obtained by the LBP method. (f) Result obtained by our method.

Table 2. Precision and recall on our sequence

	Precision	Recall
GMM	4.52	26.01
LBP	14.27	46.91
Our method	73.58	92.25

4. CONCLUSIONS

In this paper, we have proposed a background modeling method under sudden illumination change. Unlike traditional methods that use intensity information, our method adopts the phase feature. The proposed method is based on estimating the phase probability density function of the background difference using Gaussian mixture models. Then distance transform is employed on the preliminary detection result, which transforms the binary image into a distance map. Finally, we segment the distance image and get the result. Experiments on two sequences demonstrate the effectiveness and robustness of our method. Our future work will focus on the complexity reduction of the algorithm.

5. REFERENCES

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

[2] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance,” *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[3] M. Heikkilä and M.Pietikäinen, “A texture-based method for modelling the background and detecting moving objects,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 657–662, 2006.

[4] K. Toyama, B. Brumitt, J. Krumm, and B. Meyers, “Wallflower: Principles and practice of background maintenance,” *Proc. IEEE International Conferences on Computer Vision*, vol. 1, pp. 255–261, 1999.

[5] Binglong Xie, Visvanathan Ramesh, and Terrance Boult, “Sudden illumination change detection using order consistency,” *Image Vision Computing*, vol. 22, no. 2, pp. 117–125, 2004.

[6] Anurag Mittal and Visvanathan Ramesh, “An intensity-augmented ordinal measure for visual correspondence,” *Proc. IEEE Computer Vision and Pattern Recognition*, vol. 1, pp. 849–856, June 2006.

[7] Julien Pilet, Christoph Strecha, and Pascal Fua, “Making background subtraction robust to sudden illumination changes,” *Proc. European Conference on Computer Vision*, pp. 567–580, October 2008.

[8] Baochang Zhang, Shiguang Shan, Xilin Chen, and Wen Gao, “Histogram of gabor phase patterns(hgpp): A novel object representation approach for face recongnition,” *IEEE Trans on Image Processing*, , no. 1, pp. 57–68, 2007.