

Background modeling using special type of Markov Chain

Thi Thi Zin^{1a)}, Pyke Tin¹, Takashi Toriu¹, and Hiromitsu Hama²

¹ Graduate School of Engineering, Osaka City University

² Research Center for Industry Innovation, Osaka City University

3–3–138 Sugimoto, Sumiyoshi-ku, Osaka 558–8585, Japan

a) thithi@info.eng.osaka-cu.ac.jp

Abstract: Background modeling is important in video surveillance for extracting foreground regions from a complex environment. In this paper, we present a novel background modeling technique based on a special type of Markov Chain. The method is a substantial extension to the existing background subtraction techniques. First, a background pixel is statistically modeled by a linear regressive Gamma Markov distribution. Then, these statistical estimates are used as important parameters in background update schemes. The experimental results show that the proposed model is less sensitive to movements of the texture background and more robust for real time segmenting the foreground object accurately.

Keywords: background modeling, special type of Markov Chain, Gamma Markov, video surveillance

Classification: Sensing hardware

References

- [1] S. Fazli, H. Bouzari, and H. Moradi, “A Novel GMM-Based Motion Segmentation Method for Complex Background,” *Proc. 5th IEEE Gulf Cooperation Council (GCC) Conf.*, Kuwait, March 2009.
- [2] H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama, and K. Kogure, “Background Subtraction using Generalized Gaussian Family Model,” *IET Electron. Lett.*, vol. 44, no. 3, pp. 189–190, Jan. 2008.
- [3] Y. B. Chin, L. W. Soong, L. H. Siong, and W. W. Kit, “Extended fuzzy background modeling for moving vehicle detection using infrared vision,” *IEICE Electron. Express*, vol. 8, no. 6, pp. 340–345, March 2011.
- [4] Q. Wan and Y. Wang, “Background subtraction based on adaptive non-parametric model,” *Proc. 7th World Congress on Intell. Control Automation*, Chongqing, China, pp. 5960–5965, June 2008.
- [5] A. Mittal and N. Paragios, “Motion-Based Background Subtraction Using Adaptive Kernel Density Estimation,” *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, Washington, D.C, USA, vol. 2, pp. 302–309, July 2004.
- [6] T. Rodríguez and N. García, “An adaptive, real-time, traffic monitoring system,” *Machine Vision Applications*, vol. 21, no. 4, pp. 555–576, 2010.

- [7] R. M. Platarfod, “A linearly regressive Gamma Markov process,” *Stochastic Hydrol. Hydraul.*, vol. 1, pp. 155–160, 1987.

1 Introduction

Many video analysis applications require the separation of foreground of interest from the background, before any detailed analysis can be carried out on the foreground. The problems of image segmentation and grouping remain a great challenge in Computer Vision. The primary purpose of background model is to detect and extract the foreground object from the background scene in a video sequence. Because background changes due to lighting conditions, scene movements, weather conditions and etc., developing a robust background modeling algorithm is very important for many video surveillance applications. The principle is to subtract the long time-average background image from the current image frame. The pixels belonging to the foreground objects is expected to be different in intensity and chromaticity values, when compared with the background model pixel. These pixels can be differentiated from the other pixels by a heuristic threshold process.

There have been many works that tried to obtain accurate foregrounds with suppression of false alarms. Most of existing methods used a single Gaussian model to model the distribution of the background. However, it is insufficient to represent the background intensity by using only one Gaussian distribution when the background is complex. In [1, 2], they proposed the frameworks using mixture of Gaussians to deal with complex backgrounds. These approaches can have multiple hypotheses for the background so that it can be adapted for complex scenes such as waving trees, streaming waters and refreshing monitors. However, they cannot deal with the uncertainties in the correct manner. In [3, 4], a non-parametric approach to deal with the uncertainties in an accurate manner has been developed. Mittal and Paragios combined the non-parametric approach with motion information as optical flow to deal with persistent dynamic behavior in time [5].

Apart from these, the speed of the methods is also considered as a crucial for practical application systems [6]. In this context, a number of ways, such as updating the background partially and simplified model of background, were put forward to improve the speedup performance of monitoring systems. Unfortunately, the accuracy cannot be guaranteed as the extraction speed rises up because of the limitation of system resources.

In this paper, a linear regressive Markov chain based background model, accompanies with an updating method is presented. The main goal of this method is to real-timely classify the way pixel changes when different motion occurs and update the background model accordingly. The remainder of this paper is organized as follows. In Section 2, an overview of the proposed method is presented. Section 3 describes the background model and its updating process. We present the experimental results in Section 4 and

the conclusions are given in Section 5.

2 Overview of proposed method

We observe that the background modeling can be viewed as a labeling process of background and foreground from the frames of a video sequence. Most of existing research assumed the foreground/background label is a random variable of individual pixel processes governed by independent probability distributions such as Gaussians, mixture of Gaussians, family of Gaussians and so on. To take dependency nature of pixels we consider all the labels in the modeling process are formed a sequence of Markov random variables. Specifically, we assume that the state probability of each pixel only depends temporally on the state of its neighborhood pixels.

Let I_t denote the image of the video sequence at time t . Each pixel m in the sequence is associated with intensity value $p_t(m)$ and the two-state random variable of label $l_t(m)$ takes one of the two values: F and B standing for foreground and background respectively. In terms of Bayesian data analysis, $p_t(m)$ is the observation and $l_t(m)$ is the hidden state. In this paper, a special type of Gamma Markov Chain [7] is applied to represent the temporal connections of pixel labels between label images at different times. In one image sequence, we consider a temporal connection C in the Markov random sequence defined as follows:

C : Connection between the corresponding pixels at two adjacent image frames.

The connection C is designed to model the relationship between two adjacent image frames in the sequence and represents the probability of pixel state conditioned on the state of its corresponding pixel at next or previous image. It also represents the state transition probability of the pixel label $l_t(m)$.

We now assume the sequence random variables $\{p_t(m)\}$ for $t = 0, 1, 2, \dots$ form a Gamma Markov Chain with the conditional probability density of $p_{t+1}(m) = x$ given $p_t(m) = y$ is given by

$$f(x|y) = e^{\frac{-(x+\rho y)}{\alpha(1-\rho)}} \sum_{j=0}^{\infty} \frac{(\rho x)^j x^{j+r-1}}{j! \Gamma(r+j) \{\alpha(1-\rho)\}^{2j+r}}, \quad (1)$$

where $\Gamma(\bullet)$ is a gamma function, ρ is the serial correlation coefficient of lag 1 and $0 < x, y < \infty, r > 0$.

An extensive investigation of the processes has been carried out by Phatarfod [7] who obtained them as output processes of a counter system. They show that the stationary distributions of the two processes are given by, the Gamma density:

$$f(x) = \frac{1}{\alpha^r \Gamma(r)} x^{r-1} e^{-x/\alpha}, x > 0. \quad (2)$$

From Eq. (1) and (2) we obtain the conditional mean and variance of current pixels depending on the previous pixels as follows:

$$E[p_{t+1}(m)|p_t(m)] = (1-\rho)r\alpha + \rho p_t(m), \quad (3)$$

$$\sigma^2 [p_{t+1}(m)|p_t(m)] = (1 - \rho)^2 r\alpha^2 + 2(1 - \rho)r\alpha p_t(m), \quad (4)$$

Denoting $\frac{\rho p_t(m)}{\alpha(1 - \rho)}$ by λ , we have the conditional Laplace Transform (L.T) of $p_{t+1}(m)$ given $p_t(m)$ as:

$$L(\theta, p_{t+1}(m)|p_t(m)) = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j \{1 + \theta\alpha(1 - \rho)\}^{-(r+j)}}{j!}. \quad (5)$$

Eq. (5) shows that given $p_t(m)$, $p_{t+1}(m)$ is a Poisson-Gamma mixture, i.e. $p_{t+1}(m)$ has a Gamma ($U+r$, $\alpha(1 - \rho)$) distribution, where U is a Poisson variable with mean λ . The results allow us to simulate the sequence $p_0(m)$, $p_1(m)$, $p_2(m)$, ... as follows. First we generate a value of a Poisson random variable with mean $\lambda = \frac{\rho p_0(m)}{\alpha(1 - \rho)}$. Then generate a value of gamma random variable with parameters $u+r$ and $\alpha(1 - \rho)$. This procedure is repeated n times to give the background image sequence $p_0(m)$, $p_1(m)$, ..., $p_n(m)$. Substituting these values in Eq. (3) and (4) we can estimate the conditional mean and variance of background images so that we put the background pixel $B(x, y)$ as:

$$B(m) = E[p_n(m)], \quad (6)$$

with corresponding variance as $\sigma = \sqrt{\sigma^2(p_n(m))}$.

3 Background subtraction and model update scheme

For the non-background pixel, we calculate the difference between this pixel in current image and in background model. Only the pixel with the difference over the threshold $\beta\sigma$ is labeled as foreground pixel. The higher β values, the more background pixels are covered. In this stage, the output foreground mask is given by

$$M_t(m) = \begin{cases} 1 & \text{if Mahalanobis distance } |p_t(m) - B(m)| > \beta\sigma, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

It is intuitive that the change caused by a moving object can be large while the change caused by noise and varies only around the mean value of the corresponding pixel in the background frames. However, a generic background point will have a small variance, while a point in moving object will have a higher variance value. Hence, the background variance can be used as a threshold to decide whether the pixel belongs to the background or occluding region.

The change detection mask can be obtained by thresholding on the normalized statistics of the difference between $p_t(m)$ and $p_{t-k}(m)$ that is

$$D_t(m) = \begin{cases} 1 & \text{if } |p_t(m) - p_{t-k}(m) - \mu(d)| > \beta\sigma(d), \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

where $\mu(d)$ and $\sigma(d)$ are the mean and the standard deviation of $p_t(m) - p_{t-k}(m)$.

The resultant binary mask with representing occlusion regions can be generated as

$$O_t = M_t \cup D_t, \quad (9)$$

We then define a foreground pixel m in O_t is considered as shaded background if the condition $Th_1 < p_t(m)/B(m) < Th_2$ holds. Th_1 and Th_2 are the thresholds for limiting the shadows. In other words we define the shaded background region

$$S_t(m) = \begin{cases} 1 & \text{if } Th_1 < p_t(m)/B(m) < Th_2, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

This gives the resultant foreground object mask,

$$F_t = S_t \cap O_t \quad (11)$$

The background subtraction approaches are usually very sensitive to variations of the illumination; to elevate this problem the background model must be updated. In the proposed robust update, the background model is dynamically updated with incoming images. The update scheme is different for pixel positions which belong to foreground, as occluded regions and part of the background:

$$B_{t+1}(m) = \begin{cases} B_t(m) & \text{if } m \in F_t, \\ p_t(m) & \text{if } m \in O_t, \\ E[p_t(m)|p_{t-1}(m)] & \text{otherwise.} \end{cases} \quad (12)$$

4 Experimental results

This background modeling algorithm is designed for the application of video surveillance systems. We did our experiment in both indoor and outdoor environment. In our experiment, we tested 20 video sequences under various illumination conditions. Each video is 5–10 min long. Below are some of our experimental results. Fig. 1 shows the experimental results of outdoor scene video sequence. The first column is the image frame from the video. The second column is the background labeling result using our algorithm. Moreover, the effectiveness is also confirmed in complex situation using the video sequence taken at International Airport (see Fig. 2). To evaluate the accuracy of our proposed algorithm more precisely, we randomly selected 10 frames from each video sequence in different scenes and created ground-truth segmentation masks by manual segmentation. Then we calculate the segmentation error of the proposed method by using Eq. (13).

$$\text{Error Rate} = \frac{\text{number of error pixels}}{\text{number of real foreground pixels}} \times 100\%. \quad (13)$$

The average error rates of the proposed algorithm are lower than those of the conventional methods in most scenes. The average error rate is only about 0.4%.

Moreover, experimental results show that the proposed method works well even on low-frame-rate video sequences, sequences with complex lighting conditions, and outdoor video sequences, which cannot be correctly processed by the previous works.

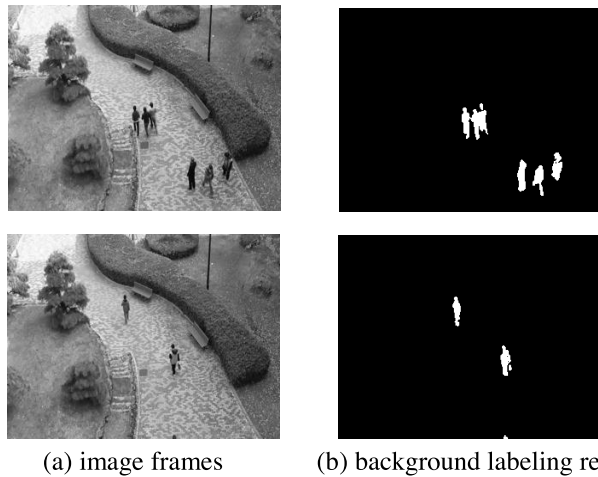


Fig. 1. Example of outdoor scene under illumination changes.

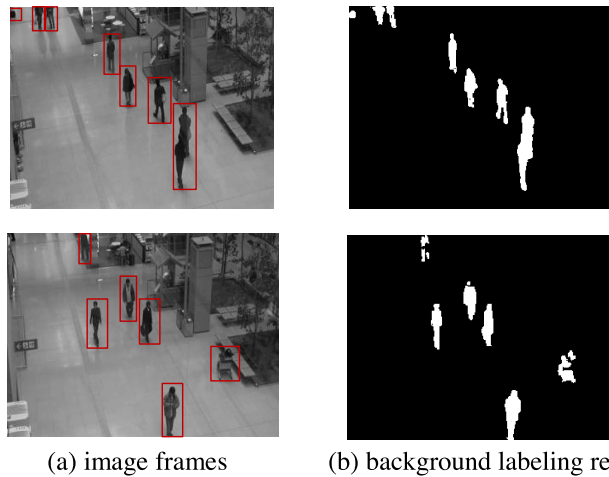


Fig. 2. Example of indoor scene in complex environment.

5 Conclusion

In this paper, a robust background modeling technique is proposed based on a Gamma Markov sequence, which can effectively classify the foreground and background. We also had proposed a new background update approach. Conventional methods used only the difference between an input image and a background image to update the background model. In the contrary, we utilize the difference between consecutive frames for updating backgrounds. Our background update method adapts the change of backgrounds very well. Almost all of the methods require that the training sequence is free of any foreground objects. In practical cases, for example, in a busy road or in a public area, it is hard to control the environments. Such a requirement cannot be always satisfied. However, our proposed method can handle such complex cases without any restrictions on the initial scenes.

Acknowledgments

This work was supported in part by SCOPE: Strategic Information and Communications R&D Promotion Program (10103768).