

Dynamic Control of Adaptive Mixture-of-Gaussians Background Model

Shimada, Atsushi
Department of Intelligent Systems, Kyushu University

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

<https://hdl.handle.net/2324/4036>

出版情報 : CD-ROM Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance 2006, 2006-11

バージョン :

権利関係 :

Dynamic Control of Adaptive Mixture-of-Gaussians Background Model

Atsushi Shimada

Department of Intelligent Systems
Kyushu University, Fukuoka, Japan
atsushi@limu.is.kyushu-u.ac.jp

Daisaku Arita

Department of Intelligent Systems
Kyushu University, Fukuoka, Japan
arita@limu.is.kyushu-u.ac.jp

Rin-ichiro Taniguchi

Department of Intelligent Systems
Kyushu University, Fukuoka, Japan
rin@limu.is.kyushu-u.ac.jp

Abstract

We propose a method for create a background model in non-stationary scenes. Each pixel has a dynamic Gaussian mixture model. Our approach can automatically change the number of Gaussians in each pixel. The number of Gaussians increases when pixel values often change because of Illumination change, object moving and so on. On the other hand, when pixel values are constant in a while, some Gaussians are eliminated or integrated. This process helps reduce computational time. We conducted experiments to investigate the effectiveness of our approach.

1. Introduction

Background subtraction technique was traditionally applied to detection of objects. We can get object areas only by doing subtraction between the observing image and the background image without requiring the prior information about the objects. However, when the background subtraction technique is applied for a surveillance system which captures outdoor scene, it detects not only objects but also a lot of noise since it shows great sensitivity to small changes such as illumination changes.

There are many approaches to handle these background changes[2, 7, 1, 9, 3]. Han et al. proposed a background estimation method. In their method, the mixture-of-Gaussians is used to approximate the background model, and the number of Gaussians is variable in each pixel. Their method can handle variations in lighting since a Gaussian is inserted or deleted according to the illuminant condition. However, it takes a long time to estimate the background model. On the other hand, there are some approaches to estimate background model in less time[6, 4, 5, 8]. Stauffer et al.

proposed a fast estimation method to avoid a costly matrix inversion by ignoring covariance components[8]. However, the number of Gaussians is constant in their background model. When recently observed pixel values frequently change, a constant number of Gaussians is not always enough to estimate the background model, and it is very difficult to determine the appropriate number of Gaussians.

In this paper, we propose a new background estimation method, which can increase and decrease the number of distributions to handle the variations. Section 2 presents the algorithm to estimate the background model proposed by Stauffer et al.[8]. In section 3, we propose an algorithm of our approach. Experimental results are presented in section 4.

2. Background Estimation

2.1. Algorithm

We consider the values of a particular pixel $\{x, y\}$ over time as a “pixel process”, which is a time series of pixel values, e.g. scalars for gray values or vectors for color images. We can represent the recent history of each pixel $\{X_1, \dots, X_t\}$ by a mixture of K Gaussian distributions, where X_t is the pixel value of $\{x, y\}$ at time t . The probability of observing the current pixel value is

$$P(X_t) = \sum_{i=1}^K w_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

where K is the number of distributions, $w_{i,t}$ is an estimate of weight of the i^{th} Gaussian in the mixture at time t , $\mu_{i,t}$ is the mean value of the i^{th} Gaussian in the mixture at time t , $\Sigma_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the

mixture at time t , and where η is a Gaussian probability density function

$$\begin{aligned} \eta(X_t, \mu_t, \Sigma_t) \\ = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)} \end{aligned} \quad (2)$$

K is determined by the available memory and computational power. Also, for computational reasons, the covariance matrix is assumed to be of the form:

$$\Sigma_{k,t} = \sigma_{k,t}^2 \mathbf{I} \quad (3)$$

This assumes that the red, green, and blue pixel values are independent and have the same variances. While this is certainly not the case, the assumption allows us to avoid a costly matrix inversion at the expense of some accuracy.

Thus, the distribution of recently observed values of each pixel in the scene is characterized by a mixture of Gaussians. A new pixel value will be represented by one of the major components of the mixture model and used to update the model. We will describe the background model estimation process in 7 steps.

Step1 Every new pixel value X_t is examined against the existing K Gaussian distributions, until a match is found. A match is defined as a pixel value within 2.5 standard deviations of distribution.

Step2 When a match is found for the new pixel value in Step1, it is regarded as the background if the matched distribution is one of the background models (described in Step7). Otherwise, the pixel value is the foreground.

Step3 The prior weights of the K distributions at time t , $w_{k,t}$, are adjusted as follows

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha(M_{k,t}) \quad (4)$$

where α is the learning rate and $M_{k,t}$ is 1 for the model which matched and 0 for the remaining models. After this approximation, the weights are renormalized.

Step4 The μ and σ parameters for unmatched distributions remain the same. The parameters of the distribution which matches the new observation are updates as follows

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (5)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \quad (6)$$

where the second learning rate ρ is

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k) \quad (7)$$

Step5 If none of the K distribution match the current pixel value in Step1, the least probable distribution is replaced with a distribution with the current value as its mean value, an initially high variance, and low prior weight.

Step6 The Gaussians are ordered by the value of w/σ^2 . This value increases both as the distribution gains more evidence and as the variance decreases.

Step7 The first B distributions are chosen as the background model, where

$$B = \arg \min_b \left(\sum_{k=1}^b w_k > T \right) \quad (8)$$

where T is a measure of the minimum portion of the data that should be accounted for by the background. If a small value for T is chosen, the background model is usually unimodal. If T is higher, a multi-modal distribution caused by a repetitive background motion (e.g. leaves on a tree, a flag in the wind, a construction flasher, etc.) could result in more than one color being included in the background model. This results in a transparency effect which allows the background to accept two or more separate colors.

3. Change of the Number of Gaussians

The number of Gaussians is constant in all of the pixels in the background estimation method described in 2.1. When recently observed pixel values are roughly constant, all of the distributions approximate the same values. In such a case, only one distribution should exist and the other distributions are not necessary at all. On the other hand, when recently observed pixel values frequently changes, the pre-defined number of Gaussians is not always enough to estimate the background model. Therefore, we propose a new background estimation method, which can increase and decrease the number of distributions to handle the variations of each pixel.

3.1. Increment of Distribution

Step5 in 2.1 is replaced by following Step5'.

Step5' If none of the K distribution match the current pixel value in Step1, a new Gaussian distribution is made as follow.

$$w_{k+1,t} = W \quad (9)$$

$$\mu_{k+1,t} = X_t \quad (10)$$

$$\sigma_{k+1,t} = \sigma_{k,t} \quad (11)$$

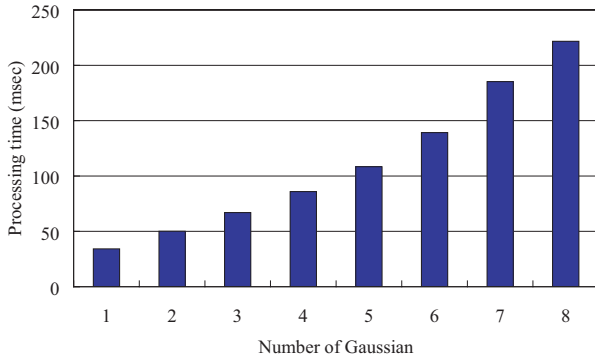


Figure 1. Comparison of computational time when the number of Gaussians is constant

where W is the initial weight value for the new Gaussian. If W is higher, the distribution is chosen as the background model for a long time. After this process, the weights are renormalized.

3.2. Decrement of Distribution

We propose two methods to decrease one of the Gaussians in the mixture. These processes are inserted between Step5 and Step6 in 2.1.

Step5 – 1 When the weight of the least probable distribution is smaller than a threshold, the distribution is deleted, and the remaining weights are renormalized.

Step5 – 2 When the difference between means of two Gaussians (the one is η_a and the other is η_b) is smaller than a threshold, these distributions are integrated into one Gaussian. The integrated Gaussian η_c is calculated as follow

$$w_{c,t} = w_{a,t} + w_{b,t} \quad (12)$$

$$\mu_{c,t} = \frac{w_{a,t} * \mu_{a,t} + w_{b,t} * \mu_{b,t}}{w_{a,t} + w_{b,t}} \quad (13)$$

$$\sigma_{c,t} = \frac{w_{a,t} * \sigma_{a,t} + w_{b,t} * \sigma_{b,t}}{w_{a,t} + w_{b,t}} \quad (14)$$

4. Experimental Result

We made experiments to investigate our approach. We took outdoor scene for 30 minutes (people coming and going in front of the building, moving cars, moving clouds, trees in the wind, and so on). The size of image was 360×240 and each pixel had a 24-bit RGB value. In our experiments, we used a computer which has PentiumD 3.0GHz CPU and 1GB memory.

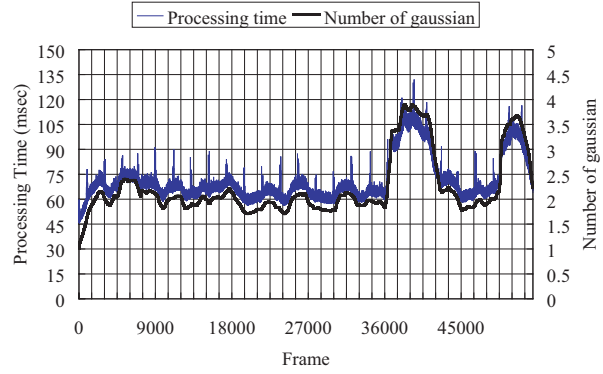


Figure 2. Comparison of computational time when the number of Gaussians is variable.

In the first experiment, we investigated a computational time when the number of Gaussians is constant. Figure 1 shows the result. The horizontal axis shows the number of Gaussians in each pixel and the vertical axis shows the average computational time for per-frame. Considering the online processing, it is preferable that the number of Gaussians is smaller than 4.

In the second experiment, we investigated our approach. Figure 2 shows the relation between the number of distributions and the computational time. The horizontal axis shows the frame number. The left vertical axis shows the computational time and the right one shows the average number of Gaussians per pixel. The number of distributions is closely related to the computational time. The number of distribution is changing from 1.5 to 2.5 in most frames. It turns out that the background model consists of the pixels which have unimodal distribution and the pixels which have multimodal distributions. On the other hand, the computational time is changing from 60 to 70msec. This is proper for the computational time when the number of Gaussians is constant 3, and available online processing. The number of Gaussians increased at a faster rate from 36000 frame. This is caused by the illumination change since the sun was obscured by clouds. Our approach handles this illumination change by increasing the Gaussians. After a few moments, the number of Gaussians gradually decreased since the illumination change had become stable. Figure 3 shows the scene and the background model changing over this period. Figure 3(a) shows the observing images and Figure 3(b) shows the mean value of the most probable distribution. We visualized the number of Gaussians at each pixel in Figure 3(c). The higher pixel value shows that the pixel has a large number of distributions. The background subtraction results are shown in Figure 3(d).

Finally, we have investigated the precision ratio and the

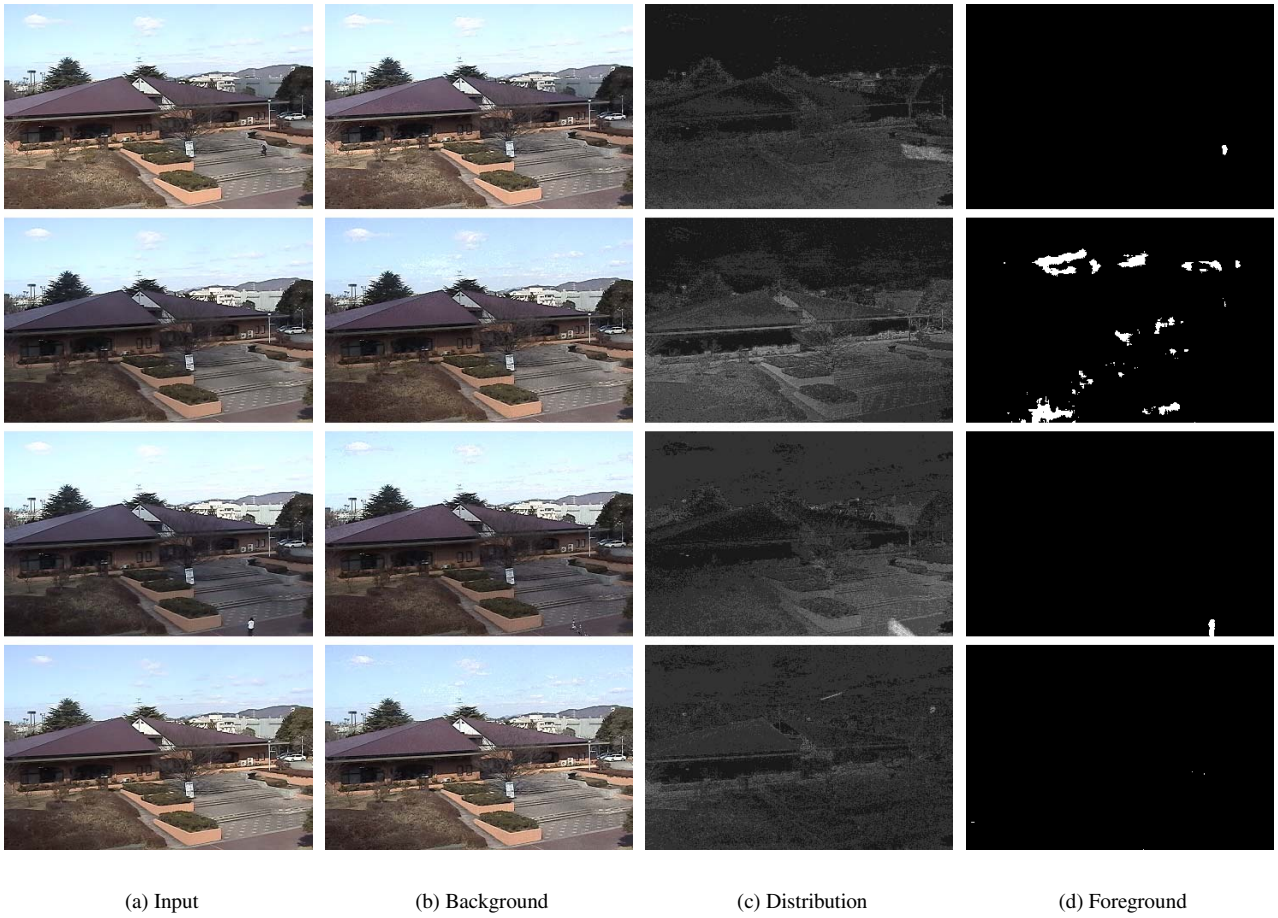


Figure 3. Adaptive background model and the number of Gaussians when observing outdoor scene.

recall ratio of the traditional approach and our approach. Precision is the ratio of the number of correct detections in the detected regions to the number of detected regions. Recall is the ratio of the number of correct detections in the detected regions to the ground truth data, i.e., the number of true foreground regions. Here, the true foreground regions are extracted manually. Figure 4 shows the result. The horizontal axis shows the number of Gaussians in each pixel and the vertical axis shows the ratio. With the increase of the number of Gaussians, the precision ratio tended to decrease, and the recall ratio tended to increase. The dotted lines plotted in the figure show the precision ratio and the recall ratio of our approach. The precision ratio was 84.40% and the recall ratio was 92.93%. Our approach can handle the variation of background since it can change the number of Gaussians dynamically without setting the number in advance, and, therefore, our approach is superior to the traditional approach.

5. Conclusion

In this paper, we proposed a new approach to estimate the background model with the mixture-of-Gaussians. Our approach can increase and decrease the number of Gaussians in each pixel. When recently observed pixel values are roughly constant, one of the Gaussians is deleted or integrated into the other distribution. When recently observed pixel values frequently changes, a new Gaussian is inserted into the background model. In our experiment, we got a good result that the computational time was 60 ~ 70msec (about 15fps) and the precision ratio and the recall ratio were superior to the traditional approach. We are now researching for immediate handling to rapid changing of the background, and for reducing computational time.

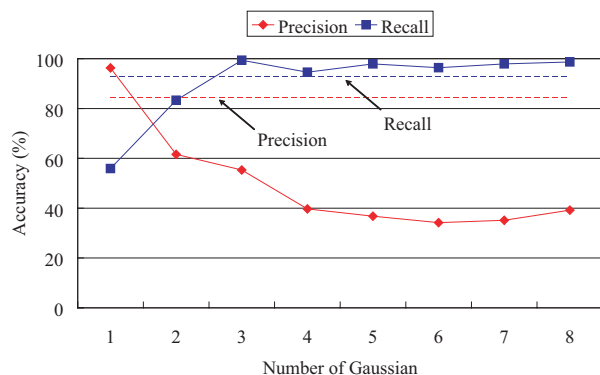


Figure 4. Precision and recall for the number of distributions.

References

- [1] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis. "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance". *Proceedings of the IEEE*, 90(7):1151–1163, 2002.
- [2] B. Han, D. Comaniciu, and L. Davis. "Sequential Kernel Density Approximation through Mode Propagation: Applications to Background Modeling". *Asian Conference on Computer Vision 2004*, pages 818–823, 2004.
- [3] M. Harville. "A Framework for High-Level Feedback to Adaptive, Per-Pixel, Mixture-of-Gaussian Background Models". *the 7th European Conference on Computer Vision*, (III):543–560, 2002.
- [4] D.-S. Lee. "Online Adaptive Gaussian Mixture Learning for Video Applications". *ECCV Workshop on Statistical Methods for Video Processing*, pages 105–116, 2004.
- [5] D.-S. Lee. "Effective Gaussian Mixture Learning for Video Background Subtraction". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):827–832, 2005.
- [6] D.-S. Lee, J. J. Hull, and B. Erol. "A Bayesian framework for Gaussian mixture background modeling". *IEEE International Conference on Image Processing*, pages 14–17, 2003.
- [7] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. "Illumination Normalization with Time-Dependent Intrinsic Images for Video Surveillance". *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26:1336–1347, 2004.
- [8] C. Stauffer and W. Grimson. "Adaptive background mixture models for real-time tracking". *Computer Vision and Pattern Recognition*, 2:246–252, 1999.
- [9] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. "Wallflower: Principle and Practice of Background Maintenance". *International Conference on Computer Vision*, pages 255–261, 1999.