Computer Science Department of The University of Auckland
CITR at Tamaki Campus (http://www.citr.auckland.ac.nz/)

CITR-TR-114 Aug 2002

Evaluation of an Adaptive Composite Gaussian Model in Video Surveillance

Qi Zang and Reinhard Klette[1]

Abstract

Video surveillance systems seek to automatically identify events of interest in a variety of situations. Extracting a moving object from background is the most important step of the whole system. There are many approaches to track moving objects in a video surveillance system. These can be classified into three main groups: feature-based tracking, background subtraction, and optical flow techniques. Background subtraction is a region-based approach where the objective is to identify parts of the image plane that are significantly different to the background. In order to avoid the most common problems introduced by gradual illumination changes, waving trees, shadows, etc., the background scene requires a composite model. A mixture of Gaussian distributions is most popular.

In this paper, we classify and discuss several recently proposed composite models. We have chosen one of these for implementation and evaluate its performance. We also analyzed its benefits and drawbacks, and designed an improved version of this model based on our experimental evaluation. One stationary camera has been used.

Keywords: video sequence analysis, surveillance systems, Gaussian mixture models.

---

1 Centre for Image Technology and Robotics, Tamaki Campus, Building 731, The University of Auckland, Morrin Road, Glen Innes, Auckland 1005, New Zealand.

Evaluation of an Adaptive Composite Gaussian Model in Video Surveillance

Qi Zang and Reinhard Klette
CITR, Computer Science Department
The University of Auckland
Tamaki Campus, Auckland, New Zealand

Abstract

Video surveillance systems seek to automatically identify events of interest in a variety of situations. Extracting a moving object from background is the most important step of the whole system. There are many approaches to track moving objects in a video surveillance system. These can be classified into three main groups: feature-based tracking, background subtraction, and optical flow techniques. Background subtraction is a region-based approach where the objective is to identify parts of the image plane that are significantly different to the background. In order to avoid the most common problems introduced by gradual illumination changes, waving trees, shadows, etc., the background scene requires a composite model. A mixture of Gaussian distributions is most popular.

In this paper, we classify and discuss several recently proposed composite models. We have chosen one of these for implementation and evaluate its performance. We also analyzed its benefits and drawbacks, and designed an improved version of this model based on our experimental evaluation. One stationary camera has been used.

Keywords: *video sequence analysis, surveillance systems, Gaussian mixture models*

1 Introduction

Video surveillance is a well-studied subject area with both existing application systems and new approaches still being developed. Research subjects are background modelling, moving object detection and tracking. Normally, video surveillance systems have three separate processing phases:

1. *low-level processes* are based on pixel models, they detect signal changes and update the background model,
2. *middle-level processes* are based on region models, they allow region splitting or merging,
3. *high-level processes* deal with final tasks such as object recognition or tracking.

Obviously, the low-level phase is most fundamental for the whole system: the correct extraction of all pixels defining a moving object and a background is the key step for the next two phases. A variety of methods has been developed and used in video surveillance applications, like *W4* [1] which analyzes pixel changes between frames, records pixel minimum and maximum values that are used in subtracting moving objects from the background; *wallflower* [2] which models and maintains the background in three levels: a pixel level, a region level and a frame level; *P-finder* [3] which models the background using a single Gaussian distribution and uses a multi-class statistical model for the tracked object; *Stauffer* [4] which uses a mixture of Gaussians to model the background and is considered to be robust against changes in outdoor scenes.

The primary goal of this paper is to critically discuss the use of mixtures of Gaussians to model a background. A second goal is to inform about an implementation of a previously already published method suggesting a mixture of Gaussians, by reporting about its performance and proposing an improvement.

The paper is structured as follows: in Section 2, we discuss recent approaches for modeling a background using Gauss distributions, following [16]. Section 3 presents performance results of the chosen model for implementation. Section 4 introduces and discusses our improvements. Section 5 finally informs about the obtained analysis and gives our conclusion.

2 Previous Work

An important property of Gaussian distributions is that they still remain to be Gaussian distributions after any linear transformation. They are widely used in adaptive systems. Especially in video surveillance applications, normally a Gaussian distribution is assumed in order to make the system adaptive to uncontrolled changes like in illumination,

outdoor weather, etc. The Gaussian is defined as

$$p(x) = N(x; \mu, \sigma^2)$$

also expressed by notation

$$x \sim N(\mu, \sigma^2)$$

which states that x is normally distributed with the corresponding mean and variance [13]. Approaches using the Gaussian can be classified into there categories:

1. *Single Gaussian*: the background distribution is modelled using a single Gaussian in HSV space, see [5];

2. *Combined Gaussians*: use of one Gaussian distribution to model a person's face and another Gaussian to model the body (shirt); this is actually a color-based tracking approach, see [6];

3. *Gaussian Mixture*: model the background by using a mixture model in order to capture changes in illumination, waving trees, etc., see [4].

The Gaussian mixture model belongs to a class of density models which have several functions as additive components. It can be stated as

$$P(\mathbf{X}_t) = \sum_{i=1}^{K} \omega_{i,t} \eta(\mathbf{X}_t; \mu_{i,t}, \Sigma_{i,t}) \ . \tag{1}$$

These functions are combined together to provide a multimodal density function, which can be employed to model colors of a dynamic scene or object. Conditional probabilities can be computed for each color pixel while a model is constructed.

The papers [7, 8, 9] are all based on using the Gaussian mixture model. In [9], a number of Gaussian functions are taken as an approximation of a multimodal distribution in color space and conditional probabilities are computed for all color pixels, probability densities are estimated from the background colors and peoples' clothing, heads, hands etc. Two assumptions are made, one is that a person of interest in an image will form a spatially contiguous region in the image plane. Another is that the set of colors for either the person or the background are relatively distinct, the pixels belonging to the person may be treated as a statistical distribution in the image plane.

MIT

An adaptive technique based on the Gaussian mixture model is discussed in [4] for the tracker modul of a video surveillance system. This technique is to model each background pixel as a mixture of Gaussians. The Gaussians are evaluated using a simple heuristics to hypothesize which are most likely to be part of the "background process". Each pixel is modeled by a mixture of K Gaussians:

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

where K is the number of distributions: normally K is between 3 to 5 in practice. $\omega_{i,t}$ is an estimate of the weight of the ith Gaussian in the mixture at time t, $\mu_{i,t}$ is the mean value of the ith Gaussian in the mixture at time t. $\Sigma_{i,t}$ is the covariance matrix of the ith Gaussian in the mixture at time t. Every new pixel value $\mathbf{X}_t$ is checked against the existing K Gaussian distributions until a match is found. Based on the matching results, the background is updated as follows:

$\mathbf{X}_t$ matches component i, that is $\mathbf{X}_t$ decreases by 2.5 standard deviations of the distribution, then the parameters of the ith component are updated as follows:

$$\omega_{i,t} = \omega_{i,t-1} \tag{3}$$
$$\mu_{i,t} = (1-\rho)\mu_{i,t-1} + \rho \mathbf{I}_t \tag{4}$$
$$\sigma^2_{i,t} = (1-\rho)\sigma^2_{i,t-1} + \rho(\mathbf{I}_t - \mu_{i,t})^{\mathsf{T}}(\mathbf{I}_t - \mu_{i,t}) \tag{5}$$

where $\rho = \alpha \Pr(\mathbf{I}_t | \mu_{i,t-1}, \Sigma_{i,t-1})$.

The parameters for unmatched distributions remain unchanged, i.e., to be precise:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} \tag{6}$$
$$\mu_{i,t} = \mu_{i,t-1} \quad \text{and} \tag{7}$$
$$\sigma^2_{i,t} = \sigma^2_{i,t-1} \ . \tag{8}$$

If $\mathbf{X}_t$ matches none of the K distributions, then the least probable distribution is replaced by a distribution where the current value acts as its mean value, the variance is chose to be high and the a-priori weight is low [4].

The background estimation problem is solved by specifying the Gaussian distributions, which have the most supporting evidence and the least variance. Because the moving object has larger variance than a background pixel, so in order to represents background processes, first the Gaussians are ordered by the value of $\omega_{i,t}/\|\Sigma_{i,t}\|$ in decreasing order. The background distribution stays on top with the lowest variance by applying a threshold T, where

$$B = \text{argmin}_b \left(\frac{\sum_{i=1}^{b} \omega_{i,t}}{\sum_{i=1}^{K} \omega_{i,t}} > T \right) \ . \tag{9}$$

All pixels $\mathbf{X}_t$ which do not match any of these components will be marked as foreground.

[10] suggested an improvement of this technique by using depth estimates: a similar Gaussian mixture model as in [4] is used, except that it is formulated in YUV color space, and it also utilizes depth values instead of disparities. This makes the algorithm applicable in systems that

compute depth not only by window-matching techniques, but also by methods based on active illumination, lidar, or other means. The progress reported in [10] is about controlling of shadows, color camouflages and high-traffic areas.

[11] also reports improvements on shadow detection based on [4]. Actually this paper combines methods proposed in [4] and [12]. Shadows remained to be the main problem in [4], and [12] uses a chromatic color space model to detect and eliminate moving object shadows. It separates chromatic and brightness components by making use of the [4] mixture model, comparing a non-background pixel against the current background components. If the difference in both chromatic and brightness components are within some thresholds, then the pixel is considered to be shadow.

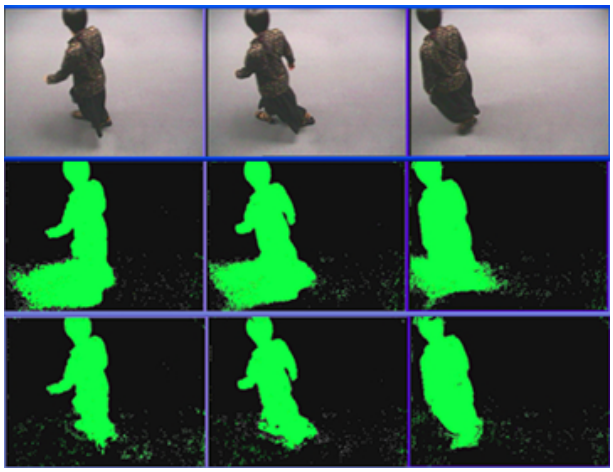

Figure 1. The top row shows an original indoor image sequence. The middle row shows results after background subtraction, still affected by shadows. The bottom row are the results after eliminating shadows.

3 Implementation and Evaluation

We implemented the method reported in [4], 'a statistical adaptive Gaussian mixture model for background subtraction' (Our implementation is on Linux in order to achieve fast processing.) We tested the program both on indoor and outdoor image sequences. In this section we report about our evaluation results.

PLUS: There are only two parameters that need to be defined in advance, and they do not need to be changed during sequence processing. (These two parameters need to be estimated/fixed during an initialization period.) The method is able to cope with many of the common problems that may happen in video surveillance applications, such as gradual illumination changes, waving trees, etc. It is stable and robust. It suits different types of cameras. It works especially very well for fast moving objects in complex environments.

MINUS: Shadows could not properly be detected/removed in [4]. This is the main problem of the method. Another problem is, while an object is moving very slowly, it will be treated as part of the background, or just detected based on differences between the current frame and previous frames, and the overlapping regions of the moving object cannot be detected as foreground. Similar outcomes happened while testing a large moving object, leaving 'holes' at the overlapping regions. This is because a slowly moving object has a small variance, which will match the background model, and, as a result, the slowly moving object was absorbed by the background. A second learning parameter ρ is not necessary to be recalculated here, because it is too small. This causes the background model to be refined too slowly. Another issue to be considered for simple indoor scenes: there are no problems such as waving trees, the background is not affected by bad weather, etc., so there is actually no need to use a Gaussian mixture model. A single Gaussian is sufficient. In another words, the mixture model is not suitable for simple indoor environments.

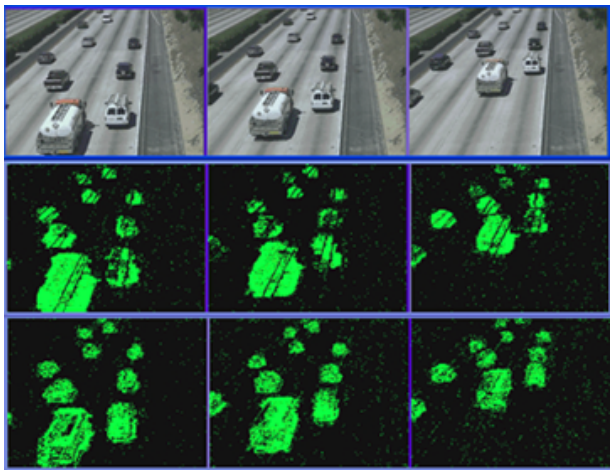

Figure 2. The top row shows an original image sequence of outdoor traffic on a highway. The middle row are results after background subtraction without shadow elimination, and the bottom row are the results after eliminating shadows.

4 Improvement

Based on the above analysis of the Gaussian mixture model, we concluded to improve as follows:

(i) Drop the second learning parameter ρ: This value is normally very small. If it is too small, the model will be refined very slowly, and this causes that the system adapts to background changes very slowly. Instead we use a reasonable value: we set this value according to specific situations in the captured scenes to control the speed of adaptation. Originally parameter ρ is still defined in an initialization phase.

(ii) Detect shadows: As we discussed before, [11] proposes to detect shadows by combining a Gaussian mixture model and [12] to detect shadows directly, however from the analysis results reported in [14] we can see that the shadow detection method of [12] works better for indoor scenes compared to outdoor scenes, and the Gaussian mixture model works better for outdoor scenes compared to indoor scenes. We derived another method which works well for outdoor scenes, and combined with a Gaussian mixture model it represents a robust approach for outdoor image sequences processing.

Our system works in RGB space. First we model the background by using a Gaussian mixture model. The number of contributing components will be set according to different environments. The minimum value is $K = 3$. The system can start at any stage, it does not matter whether the background is static at this moment or or not. The initial mean value was set to a very small value, and standard deviation was set to a high value. As in [4], for computational reasons, we also assume that the red, green and blue values are independent and have the same variances. This simplifies the calculation of a covariance matrix: instead we approximate the value by calculating the Mahalanobis distance between the pixel of a current frame and the background model. Values which are within an interval of 2.5 times standard deviation are matching one of the mixture components. After resorting the background model using weight/variance values, the best match is detected: if the matching is within the background component, the pixel will be marked as background, otherwise, the pixel will be marked as foreground and grouped together for further processing. If none of these components match, a new component will be initialized using the current pixel value as its mean value, and using a high variance and a low weight. The parameters of matched components will be updated.

The next step is to remove shadows. Here we use a method similar to [11]. The detection of brightness and chromaticity changes in the HSV space are more accurate than in RGB space, especially in outdoor scenes, and the HSV color space corresponds closely to human perception of color [15]. At this stage, only foreground pixels need to be converted to hue, saturation and intensity triples. Shadow regions can be detected/eliminated as followings: let E represent the current pixel at time t, and B represents the background pixel at time t. For each foreground pixel, if it satisfies the constraints

$$\begin{aligned} |\mathbf{E}_h - \hat{\mathbf{B}}_h| &< \mathbf{T}_h \ , \\ |\mathbf{E}_s - \hat{\mathbf{B}}_s| &< \mathbf{T}_s \ , \text{ and} \\ \mathbf{T}_{v1} < \mathbf{E}_v / \hat{\mathbf{B}}_v &< \mathbf{T}_{v2} \end{aligned}$$

then this pixel will be removed from the foreground mask. Parameters of shadow pixels will not be updated. Finally, we obtain the moving objects mask, which is applicable for object tracking. The three thresholds used for HSV are obtained from testing data, they are $\mathbf{T}_h$=0.5, $\mathbf{T}_s$=0.1, $\mathbf{T}_{v1}$=0.1,$\mathbf{T}_{v1}$=0.7, respectively. The frame size used was 320 x 240, the real-time processing rate we achieved was 5 to 7 frames per second. The indoor testing data are captured in the CITR vision lab, the background was modelled by using a single multi-dimensional Gaussian distribution. Shadows were eliminated by detecting changes in lighting and chromaticity in RGB space. The outdoor traffic testing data have been downloaded from a website, showing scenes for different weather situations, such as sunny or snowing.

5 Conclusion

The Gaussian mixture models are a type of density models which are composed of a number of components (functions). These functions can be used to model the colors of objects or backgrounds in a scene. This allows to achieve color-based object tracking and background segmentation. When a model is generated, conditional probabilities can be estimated for all color pixels. Adaptive Gaussian distributions are applicable for modelling changes, especially also related to fast moving objects such as cars on a highway. How to use Gaussian distributions has to be based on the application context. It can provide analysis results for long duration scenes. It is also quite suitable for complex scenes or multiply-colored objects. For simple indoor scenes or objects appearing monocolored, a Gaussian mixture model is not necessary if care has to be taken about computation time and system efficiency.

Shadow is a main drawback for all video surveillance applications and affects the accuracy of the system performance. We combined Gaussian mixture models and shadow elimination methods, which resulted into a more robust and efficient system, which may be used in traffic analysis and control systems. Future work will also include IR image data.

References

[1] I. Haritaoglu, D. Harwood, L. S. Davis: W4: Who? When? Where? What? A real-time system for detecting and tracking people. In Proc. *3rd Face and Gesture Recognition Conf.*, pages: 222-227, 1998.

[2] K. Toyama, J. Krumm, B. Brumitt, B. Meyers: Wallflower: Principles and practice of background maintenance. In Proc. *Internat. Conf. Computer Vision*, pages: 255-261, 1999.

[3] C. Wren, A. Azabayejani, T. Darrell, A. Pentland: Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Analysis Machine Int. Volume: 19 Issue: 7* pages: 780-785, 1997.

[4] C. Stauffer, W. E. L. Grimson: Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition, Volume: 2* pages: 246-252, 1999.

[5] A. R. J. Franois, G. G. Medioni: Adaptive color background modeling for real-time segmentation of video streams. In Proc. *Int. Conf. Imaging Science, Systems, and Technology*, pages: 227-232, 1999.

[6] S. Waldherr, S. Thrun, R. Romero: A neural-network based recognition of pose and motion gestures on a mobile robot. *Neural Networks, 1998. Proceedings. Vth Brazilian Symposium on, 1998* , pages: 79-84, 1998.

[7] S. J. Mckenna, Y. Raja, S. Gong: Object tracking using adaptive colour mixture models. In Proc. *ACCV'98*, pages: 615-622, 1998.

[8] Y. Raja, S. J. Mckenna, S. Gong: Tracking colour objects using adaptive mixture models. *Image Vision Computing Volume: 17* pages: 225-231, 1999.

[9] Y. Raja, S. J. Mckenna, S. Gong: Tracking and segmenting people in varying lighting conditions using colour. In Proc *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, 1998* , pages: 228-233, 1998.

[10] M. Harville, G. Gordon, J. Woodfill: Foreground segmentation using adaptive mixture models in color and depth. In Proc. *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on, 2001* , pages: 3-11, 2001.

[11] P. KaewTraKulPong, R. Bowden: An improved adaptive background mixture model for real-time tracking with shadow detection. *2nd European Workshop on Advanced Video Based Surveillance System. AVBS01. Sept 2001. http://www.ee.surrey.ac.uk/Personal/R.Bowden/publications/avbs01/avbs01.pdf*

[12] T. Horparasert, D. Harwood, L. A. Davis: A statistical approach for real-time robust background subtraction and shadow detection. In Proc. *ICCV'99: Frame Rate Workshop*, pages: 1-19, 1999.

[13] Y. Bar-Shalom, X. R. Li: *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, Boston, 1993.

[14] A. Prati, I. Mikic, M. Trivedi, R. Cucchiara: Detecting moving shadows: formulation, algorithms and evaluation. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on . Volume:2* pages: 571-576, 2001.

[15] N. Herodotou, K. N. Plataniotis, A. N. Venetsanopoulos: A color segmentation scheme for object-based video coding. In Proc. *IEEE Symp. Advances in Digital Filtering and Signal Proc.*, pages: 25-29, 1998.

[16] A. McIvor, Q. Zang, R. Klette: The background subtraction problem for video surveillance systems. *International Workshop RobVis 2001*. (2001) Page(s): 176-183.