ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

Moving Foreground Object Detection and Background Subtraction Using Adaptive-K GMM: A Survey

P. Rajan[1]
Computer Science Dept
APEC
Chennai – India

Dr. S. Prakash[2]
Electronics Communication and Engineering Dept
JCE
Chennai – India

Abstract: Video segmentation is an important phase in video based traffic surveillance applications. The basic task of traffic video segmentation is to classify pixels in the current frame to road background or in the foreground moving vehicles, and casting shadows should be taken into account if exists. This segmentation used for video based traffic surveillance applications. The numerous approaches are available for background subtraction. This paper discusses moving foreground object detection and background subtraction using Adaptive-K Gaussian Mixture Model (AKGMM). The Gaussian distributions of the adaptive mixture model are then evaluated to determine which are most likely to result from a background process. Each pixel is classified based on whether the Gaussian distribution which represents it most effectively is considered part of the background model. This model gives better results in illumination changes, motion changes and changes in the background geometry. We also discuss the improvements in this model and compare with other methodologies.

Keywords: Adaptive-K Gaussian Mixture Model [AKGMM], Support vector machine [SVM], Relevance vector machine [RVM], Mixture of Gaussians [MoG], Kernel Density Estimation [KDE], Sequential Kernel Density Approximation [SKDA].

I. INTRODUCTION

Background modelling methods can be classified into two categories: 1) Methods that employ *local* (pixel-wise) models of intensity and 2) Methods that have *regional* models of Intensity [10]. Most of the background modelling approaches tends to fall into the first category of pixel-wise models. The pixel wise models again classify into two categories: 1) Parametric model for Background subtraction and 2) Non-Parametric model for Background subtraction. In this paper we discuss pixel-wise parametric model [AKGMM] and also briefly discuss non-parametric model.

Background subtraction is a widely used approach to identify foreground moving objects in videos from static cameras. In most applications, objects are of interest, not the scene such as trees, buildings like that. So normally we identifying of moving objects used for Traffic monitoring, human action recognition, human-computer interaction and computer vision such as digital forensics.

The general requirements for a background removal algorithm are the accuracy in object contour detection (spatial accuracy) and temporal stability of the detection (temporal coherency). Moreover, the ability to detect changes of small magnitude (sensitivity) and providing good accuracy under varying conditions such as illumination changes (robustness).

A reliable and robust background subtraction algorithm should handle the sudden or gradual illumination changes examples lighting changes, clouds like that, High frequency and repetitive motion in the background examples as tree leaves, flags, sea waves and long-term scene changes (a car is parked for a week or month). In these cases, a single valued background is not an adequate model.

The goal of a moving object detection algorithm is to detect significant changes occurring throughout the video sequence while rejecting unimportant ones. The following describes pre-processing steps used to filter out common types of unimportant changes before making the object detection decision. These steps generally involve geometric and radiometric (i.e. intensity) adjustments [10]. Others involve using image derivatives or depth information as an information source to the moving object detection algorithm. For real-time systems, frame-size and frame-rate detection are commonly used to reduce the data processing rate. Simple temporal and/or spatial smoothing is often used in the early stage of pre-processing to reduce camera noise such as rain and snow captured in outdoor applications.

The Gaussian distributions of the adaptive-K mixture model are then evaluated to determine which are most likely to result from a background process. If we model each pixel as a adaptive-K mixture of Gaussians to determine whether or not a pixel is part of the background, then we will arrive at an effective approach to separate the background and foreground, which can be used for real-time tracking [1].

The rest of the paper is organized as follows: section 2 describes about the multivariate Gaussian distributions. Section 3 presents the background subtraction and moving foreground object detection. Section 4 explains experimental results. Section 5 explains about the improvement of multimodal Gaussians. Section 6 discusses the other background subtraction models. Section 7 compares the multimodal Gaussian with SVM. Section 8 discusses the pixel-based performance metrics for choose correct classifier for application specific. Section 9 Conclusive remarks are addressed at the end of this paper.

II. ADAPTIVE –K MIXTURE OF MODEL

The values of a particular pixel are modelled as a mixture of adaptive Gaussians. Mixture required for multiple surfaces appear in a pixel. Adaptive required for lighting conditions change. At each iteration Gaussians are evaluated using a simple heuristic to determine which ones are mostly likely to correspond to the background. Pixels that do not match with the “background Gaussians" are classified as foreground. Foreground pixels are grouped using two dimensional connected component analysis [2].

Normal or Gaussian distribution (N):

$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

Mixture of Gaussian distribution (N):

$$\mathcal{N}(\mathrm{X} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)} \qquad (2)$$

At any time t, what is known about a particular pixel, (x0; y0), is its history:

$$\{X_{1,\ldots,}X_t\} = \{\mathrm{I}(\mathrm{x}_0, y_0, i) : 1 \le i \le t\} \qquad (3)$$

This history is modelled by a mixture of K Gaussian distributions:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \mathcal{N}(\mathrm{X}_t \mid \mu_{i,t}, \Sigma_{i,t}) \qquad (4)$$

Where

$$\mathcal{N}(\mathrm{X}_t \mid \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_{i,t}|^{1/2}} e^{-\frac{1}{2}(X_t-\mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t-\mu_{i,t})} \qquad (5)$$

Here K is the number of Gaussian distributions, $\omega_{i,t}$ is an estimate of the weight of the i^{th} Gaussian in the mixture at time t, is the mean value of the i^{th} Gaussian in the mixture at time t, is the co-variance matrix of the $\Sigma_{i,t}$ i^{th} Gaussian in the mixture at time t, is the Gaussian probability density function and D represents no of dimension taken. If we assume gray scale images and set K = 5, history of a pixel will be something like this Fig 1:

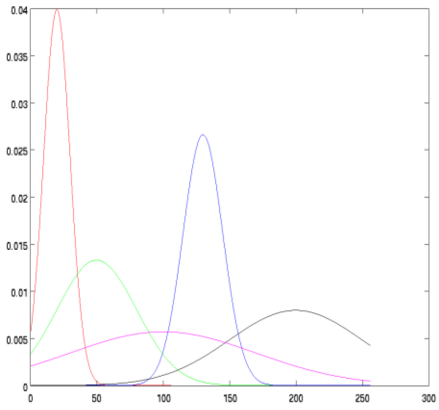


Fig 1. Adaptive-K Gaussian Mixture Model

An adaptive K-means approximation is used to update the Gaussians. If a new pixel value, X_{t+1}, can be matched to one of the existing Gaussians (within 2.5σ), that Gaussian's $\mu_{i,t+1}$ and $\sigma^2_{i,t+1}$ are updated as follows:

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1} \qquad (6)$$

$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho(X_{t+1} - \mu_{i,t+1})^2 \qquad (7)$$

Where

$$\rho = \alpha N(X_{t+1} \mid \mu_{i,t}, \sigma^2_{i,t}) \qquad (8)$$

and α is a learning rate.

Prior weights of all Gaussians are adjusted as follows:

$$\omega_{i,t+1} = (1-\alpha)\omega_{i,t} + \alpha(M_{i,t+1}) \qquad (9)$$

Where $M_{i,t+1} = 1$ for the matching Gaussian and $M_{i,t+1} = 0$ for all the others. If X_{t+1} do not match to any of the K existing Gaussians, the least probably distribution is replaced with a new one. "Least probably" means ω/σ. New distribution has $\mu_{t+1} = X_{t+1}$, a high variance and a low prior weight.

III. BACKGROUND MODEL ESTIMATION

Heuristic: the Gaussians with the most supporting evidence and least variance should correspond to the background. The Gaussians are ordered by the value of ω/σ (high support and less variance will give a high value). Then simply the first B distributions are chosen as the background model:

$$B = \arg\min_b \left(\sum_{i=1}^{b} \omega_i > T\right) \qquad (10)$$

Where T is minimum portion of the image which is expected to be background in Fig 2.

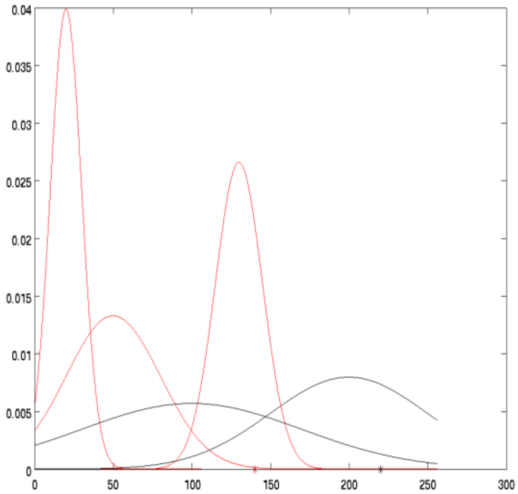


Fig 2.After background model estimation red distributions become the background model and black distributions are considered to be foreground.

The above formula T represents threshold value. Thresholding is a fundamental method to convert a gray scale image into a binary mask, so that the objects of interest are separated from the background [15]. In the difference image, the gray levels of pixels belonging to the foreground object should be different from the pixels belonging to the background. Thus, finding an appropriate threshold will solve the localization of the moving object problem. The output of the thresholding operation will be a binary image whose gray level of 0 (black) will indicate the pixel belonging to the background and a gray level of 1 (white) will indicate the object.

Many shadow detection algorithms are proposed, but most of them are too complex. In our method, we adopt a simple shadow detection algorithm called Normalized Cross-Correlation algorithm (NCC-algorithm) proposed by [3] to refine the segmentation if dynamic casting shadow exists. NCC explores the relationship between casting shadow and background, that is, the intensity of shadowed pixel is linear to the corresponding background, so the background image provided by Adaptive-K Gaussian Mixture Model is used to detect shadows.

IV. Experimental Results

In our program execution resulting Fig 3. Given the original image, foreground image and also detecting the moving object such as car. The Gaussian mixture model speed is intermediate and memory requirements also intermediate [4] and [5].

The Gaussian method having following merits:

- A different “threshold" is selected for each pixel.
- These pixel-wise ‘thresholds" are adapting by time.
- Objects are allowed to become part of the background without destroying the existing background model.
- Provides fast recovery.

Fig 3.(a)original image (b) foreground image (c)moving object detection such as car.

The Gaussian method having following demerits:

- Cannot deal with sudden, drastic lighting changes.
- Initializing the Gaussians is important (median filtering).
- There are relatively many parameters, and they should be selected intelligently.

The goal of a moving object detection algorithm is to detect significant changes occurring throughout the video sequence while rejecting unimportant ones. The following describes pre-processing steps used to filter out common types of unimportant changes before making the object detection decision. These steps generally involve geometric and radiometric (i.e. intensity) adjustments [10]. Others involve using image derivatives or depth information as an information source to the moving object detection algorithm. For real-time systems, frame-size and frame-rate detection are commonly used to reduce the data processing rate. Simple temporal and/or spatial smoothing is often used in the early stage of pre-processing to reduce camera noise such as rain and snow captured in outdoor applications. These pre-processing steps are necessary because in the result Fig 3-b foreground image contains lots of noise Such as waving trees, buildings like that.

V. IMPROVING MULTIMODAL GAUSSIANS

Mixture of Gaussian method has been used on greyscale, RGB, HSV and local linear filter responses. But this method should be capable of modelling any streamed input source in which our assumptions and heuristics are generally valid. We are investigating use of this method with frame-rate stereo, Infra Red cameras, and including depth as a fourth channel(R, G, B, D). Depth is an example where multi-modal Gaussian distributions are useful, because while disparity estimates are noisy due to false correspondences, those noisy values are often relatively predictable when they result from false correspondences in the background [1].

Compare than mixture of Gaussian method no of improved model available recently for foreground detection. They are Kernel Density Estimation [KDE], Sequential Kernel Density Approximation [SKDA] and Support Vector Machine [SVM] to give better results [4,9]

VI. KERNEL DENSITY ESTIMATION (KDE) AND SEQUENTIAL KERNEL DENSITY APPROXIMATION (SKAD)

A robust, non-parametric background model and background subtraction mechanism that works with colour imagery. The model can handle situations where the background of the scene is not completely static but contains small motions such as tree branch motion. The model is based on estimating the intensity density directly from sample history values. The main feature of the model is that it represents a very recent model of the scene and adapts to change quickly. A second stage of the background subtraction was presented was presented to suppress false detection that are due to small motions in the scene background based on spatial properties.

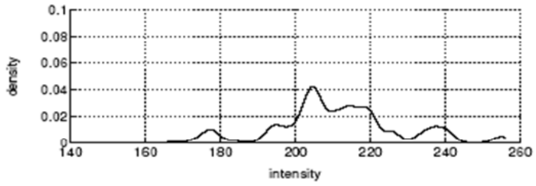
In the non-parametric background model KDE Fig 4-b shows the estimated background probability where brighter pixels represent lower background probability pixels [6]. We seen that Fig easily identify one human, we cannot identify easily another human but Fig 4-b shows two humans visibly. One human cycling and another human walking around bushes and trees in the left side.

(a) (b)

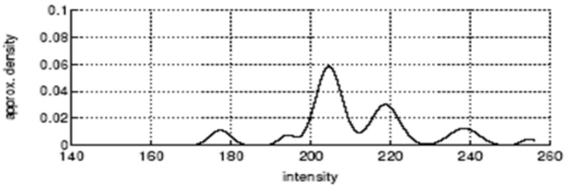
Fig 4.Background Subtraction (a) original image (b) Estimated probability image

The comparison between GMM and KDE models. We seen that if both models are given the same amount of memory, and the parameters of the two models are adjusted to achieve the same false positive rates, then the non-parametric model has much higher sensitivity in detection than the mixture of K Gaussians [6].

Sequential Kernel Density Approximation [SKDA] is approximates mean and variances of KDE. So, SKDA memory utilization is less compare than KDE [7]. Fig 5 shows both of the algorithms result.



(a) KDE



(b) SKDA

Fig 5. Kernel Density Estimation [KDE] versus Sequential Kernel Density Approximation [SKDA].

VII. MIXTURE OF GAUSSIANS (MOG) VS SUPPORT VECTOR MACHINE (SVM)

In recent trends so many of background subtraction techniques are available such SVM, Relevance Vector Machine [RVM] like that. The Principal Component Analysis [PCA] technique is used to identify moving objects. Its works faster than Gaussian mixture model [4] and [8]. In this paper MoG compare with SVM [9].

Background subtraction results of three algorithms SVM, KDE, and GMM in the subway, fountain and caviar sequence images are presented in Fig 6.

(a) *subway* sequence (b) *fountain* sequence (c) *caviar* sequence

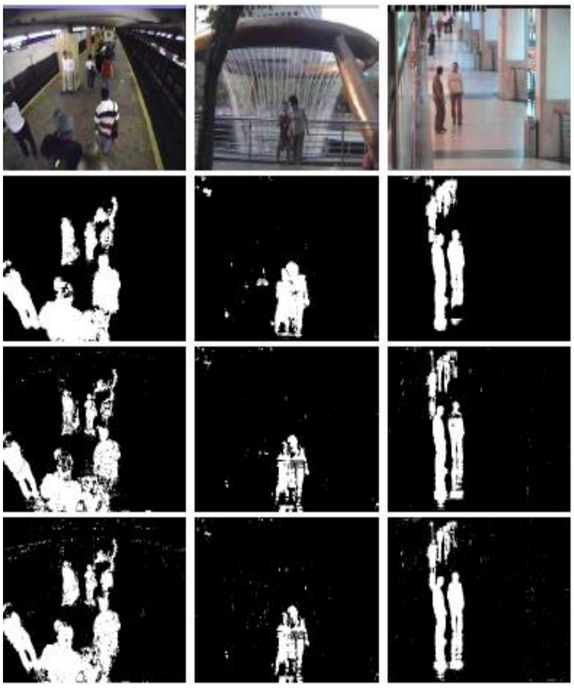


Fig 6. From top to bottom, the original images and the results of SVM, KDE, and GMM are presented

Note that there is a moving car on the left side of the child in the fountain sequence, which is clearly visible but not detected by KDE or GMM [9].

VIII. PIXEL BASED PERFORMANCE METRIC

Ground truth (or gold standard) generation can be viewed as the process of establishing the "correct answer" for what *exactly* the algorithm is expected to produce, which is generally application-specific [11]. For example, in video surveillance, it is generally undesirable to detect the "background" revealed as a consequence of camera and object motion as change, whereas in remote sensing, this change might be considered significant (e.g. different terrain is revealed as a forest recedes).

Once a ground truth has been established, there are several standard methods for comparing the ground truth to a candidate binary foreground map. The following quantities are generally involved:

- True positives (TP): the number of foreground pixels correctly detected;
- False positives (FP): the number of background pixels incorrectly detected as foreground (also known as false alarms);
- True negatives (TN): the number of background pixels correctly detected; and
- False negatives (FN): the number of foreground pixels incorrectly detected as background (also known as misses).

Based on the above mentioned quantities, Rosin [12] described three methods for quantifying a classifier's performance:

The Percentage Correct Classification

$$PCC = \frac{TP + TN}{TP + FP + TN + FN} \quad (11)$$

The Jaccard Coefficient

$$JC = \frac{TP}{TP + FP + FN} \quad (12)$$

The Yule Coefficient

$$YC = \left| \frac{TP}{TP + FP} + \frac{TN}{TN + FN} - 1 \right| \quad (13)$$

Combining all four values to form the PCC is the most widespread method in computer vision for assessing a classifier's performance. However, it tends to give misleading estimates when the amount of change is small compared to the overall image [12].

The Yule and Jaccard coefficients overcome this problem to some degree by minimizing or eliminating the effect of the expected large volume of true negatives. Note that the Yule coefficient cannot be applied when the algorithm correctly detected no change in the image (since one denominator becomes zero). Within the sequences there might be frames in which no change occurs which can be analyzed separately to monitor the effects of noise and compression artifacts when no real activity exists in the sequence.

Cheung and Kamath [13] compare the performance of a number of popular background removal techniques using two information retrieval measurements, recall and precision, to quantify how well each algorithm matches the ground-truth [11]. They are defined in their context as; *Recall* is the ratio of the number of foreground pixels correctly identified by the algorithm to the number of foreground pixels in ground truth, while *Precision* is defined as the ratio of the number of foreground pixels correctly identified by the algorithm to the number of foreground pixels detected by the algorithm. Recall and precision values are both within the range of 0 and 1. Typically, there is a trade-off between recall and precision - recall usually increases with the number of foreground pixels detected, which in turn may lead to a decrease in precision. A good background algorithm should attain as high a recall value as possible without sacrificing precision [13].

Given the ground truth, Nascimento and Marques [14] detect several types of errors i) splits of foreground regions, ii) merges of foreground regions, iii) simultaneously split and merge of foreground regions, iv) false alarms (detection of false objects) and v) the detection failures (missing active regions). They then compute statistics for each type of error.

Object-based metrics usually involve both spatial and temporal accuracy metrics. Additionally, some metrics like the criticality can be considered as spatio-temporal, since they simultaneously cover spatial and temporal aspects of the complexity of a sequence. The results obtained with each metric are normalized to the range [0, 1] (see [11] for details).

IX. CONCLUSION

This paper has shown a novel, probabilistic method for background subtraction. It involves modelling each pixel as a separate mixture model. We implemented a real-time approximate method which is stable and robust. The method requires only two parameters, α and T. These two parameters are robust to different cameras and different scenes.

This system has been successfully used to track cars in outdoor environments. This system achieves our goals of real time performance over extended periods of time without human intervention. Recently Kernel Density Estimation, Sequential Kernel Density Approximation, Support Vector Machine used for background subtraction. Those are given better results comparing than Mixture of Gaussians.

References

1. Stauffer and Grimson, “Adaptive background mixture models for real-time tracking” In Computer Vision and Pattern Recognition, volume 2, pages 252–258, 1999.
2. Tan, Hong, Jin and Fang “Traffic Video Segmentation Using Adaptive –K Gaussian Mixture Model” Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China.
3. Julio Cezar, C.R. Jung and S.R.Musse, “Background Subtraction and Shadow Detection in Grayscale Video Sequences”, in Proceedings of SIBGRAPI 2005-Natal-RN-Brazil, 2005.
4. M.Piccardi,”Background Subtraction Techniques: A Review”, in 2004 IEEE International Conference on Systems, Man and Cybernetics, PP.3099-3104.
5. T.Bouwmans, F.El Baf and B.Vachon, “Background Modelling Using Mixture of Gaussians for Foreground Detection – A Survey” Proceedings in Recent Patents on Computer Science 1, 3(2008) 219-237.
6. Elgammal, A., Harwood, D., and Davis, L.S., “Non-parametric Model for Background Subtraction”, Proc. of ICCV '99 FRAME-RATE Workshop, 1999.
7. B. Han, D. Comaniciu, and L. Davis, "Sequential kernel density approximation through mode propagation: applications to background modelling,“ Proc. ACCV -Asian Conf. on Computer Vision, 2004.
8. N. M. Oliver, B. Rosario, and A. P. Pentland, “A Bayesian Computer Vision System for Modelling Human Interactions,” IEEE Trans. on Patt. Anal. and Machine Intell., vol. 22, no. 8, pp. 831-843, 2000.
9. Bohyung Han and S.Davis “Density-Based Multifeature Background Subtraction with Support Vector Machine”, IEEE Trans. on Patt. Anal. and Machine Intell., vol. 34, no. 5, pp. May 2012.
10. Yaser Sheikh and Mubarak Shah “Bayesian Modelling of Dynamic Scenes for Object Detection”, IEEE Trans. on Patt. Anal. and Machine Intell., vol. 27, no. 11, pp. November 2005.
11. Elhabian, Sayed and Ahmed “Moving Object Detection in Spatial Domain using Background Removal Techniques – State of Art”, Bentham Science Publishers Ltd, vol. 1, no. 1, pp. 32-54, 2008.
12. P.Rosin and E.Ioannnidis “Evaluation of global image thresholding for change detection”, Pattern Recognition Letters vol.24, no. 14, pp. 2345-2356, October 2003.
13. S-C Cheung and C.Kamath “Robust techniques for background subtraction in urban traffic video”, SPIE Visual Comm Image Proce vol.53, no. 08, pp. 881-892, October 2003.
14. J.Nascimento and JS Marques “Performance evaluation of object detection algorithms for video surveillance”, IEEE Trans. on Multimedia., 2005.
15. R.Jain, R.Kasturi and GB.Schunk Machine Vision. McGRAWHILL Int. Editions, 1995.

AUTHOR(S) PROFILE

P. Rajan, received the B.E. degree in Computer Science and Engineering from SBCE College, Anna University, Chennai, Tamil Nadu, India and the M.E. degree in Computer Science and Engineering from GKM College of Engineering, Anna University, Chennai, Tamil Nadu, India in 2004 and 2007 respectively.

Presently he is a working as Assistant Professor in Computer Science and Engineering at Adhiparasakthi Engineering College, Melmaruvathur, Tamil Nadu, India.

Dr. S. Prakash, received the Ph.D degree in Instrumentation from IISc Bangalore in 1997. The M.E. degree in Electronics Communication and Engineering from Birla Institute of Technology, Mesra and The B.E. degree in Electronics Communication and Engineering from Madurai Kamaraj University, Tamil Nadu, India in 1992 and 1990 respectively.

Presently he is a working as Professor in Electronics Communication and Engineering at Jerusalem College of Engineering, Chennai, Tamil Nadu, India.