

Improved Gaussian Mixtures for Robust Object Detection by Adaptive Multi-Background Generation

Mahfuzul Haque, Manzur Murshed, and Manoranjan Paul
Gippsland School of Information Technology, Monash University
Churchill Vic 3842, Australia

{mahfuzul.haque, manzur.murshed, manoranjan.paul}@infotech.monash.edu.au

Abstract

Adaptive Gaussian mixtures are widely used to model the dynamic background for real-time object detection. Recently the convergence speed of this approach is improved and a relatively robust statistical framework is proposed by Lee (PAMI, 2005). However, object quality still remains unacceptable due to poor Gaussian mixture quality, susceptibility to background/foreground data proportion, and inability to handle intrinsic background motion. This paper proposes an effective technique to eliminate these drawbacks by modifying the new model induction logic and using intensity difference thresholding to detect objects from one or more believe-to-be backgrounds. Experimental results on two benchmark datasets confirm that the object quality of the proposed technique is superior to that of Lee's technique at any model learning rate.

1. Introduction

The performance of object detection is very crucial in a sound surveillance system for reliable tracking and behavior recognition. Though it can be achieved quite easily within a limited scope using the basic background subtraction (BBS) by storing a reference background image and thresholding the difference in the current scene, but it is often challenged in real-world by the dynamic nature of the background due to sudden and gradual illumination variations, intrinsic repetitive background motions such as movement of tree leaves, and by global motions due to intentional and/or non-intentional camera displacements. To cope with these challenges adaptive background modelling techniques are widely used.

Stauffer and Grimson [7] first introduced the Gaussian mixture model (GMM) technique where each pixel

is modelled using a blend of Gaussian distributions [6] that are continuously learnt by online approximation at a learning rate α . Object detection at the current scene is then performed at pixel-level. First the set of Gaussians that are most likely to represent the background is isolated using the proportion by which a pixel is going to observe the background. Then the membership of the pixel is checked against each Gaussian in that set. Over the past few years, different variants of GMM techniques have been evolved [1, 3, 4] and several multi-stage techniques [2, 10] were also proposed using the pixel-based GMM technique.

Among these works, the statistical framework for the GMM technique developed by Lee [4] improves the convergence speed of the learning mechanism without compromising model stability. This is achieved by adaptively setting the effective learning rate higher than α for new models and introducing a sigmoid function to estimate the probability of a Gaussian to represent the background from its relative weight and variance. While the increased stability has enabled the GMM technique to be used for ad-hoc setups where sensibility analysis of the parameters cannot be performed, the object detection quality of this statistical framework still remains unsatisfactory due to the following reasons: (i) like all other GMM techniques, an established model with sufficiently small variance is unnecessarily duplicated; (ii) the proportion by which a pixel is going to observe the background is assumed constant but in reality it fluctuates constantly depending on the number of objects and their movement patterns; and (iii) the problem intensifies further when intrinsic background motion is involved resulting in a significant number of false positive pixel detection.

In this paper, we propose a new object detection technique using the statistical framework in [4] by systematically addressing the abovementioned problems. The solutions have stemmed from the following observation. With static background, the BBS technique us-

ing a fixed background-foreground intensity difference threshold S can effectively isolate the moving objects from the shadows even in gradually changing lighting condition [5]. BBS fails considerably when the object's intensity does not contrast the background enough or intrinsic background motion is present, including sudden change in lighting condition.

In most of the indoor and traffic scenarios, there would be a single dominating distribution representing the scene background adapted over time, but in the presence of intrinsic repetitive background motions like fluttering tree leaves in the backdrop of clear sky, there would be more than one dominating Gaussians representing sky and waving leaves. If we construct a believe-to-be background for each of these models using the most recently observed value for the model, the object detection decision in the current scene can be made as follows. A pixel is foreground if its intensity differs from all the believe-to-be background values by S . Under this detection principle, two background Gaussians having their means within S will be counter-productive and hence, we safely reject a new model if its mean is within S of any of the existing model.

2. The proposed technique

In the proposed technique, each pixel of a scene is modelled independently by a mixture of at most K Gaussian distributions where each Gaussian represents the intensity distribution of one of the different environment components e.g., moving objects, shadow, illumination changes, sky, tree leaves, and static background, observed by the pixel over time. Let the k th Gaussian in the mixture be denoted as η_k with mean μ_k , variance σ_k^2 , the most recently observed pixel value m_k , the number of observed pixel values c_k , and weight ω_k such that $\sum_{\forall k} \omega_k = 1$. Let $\eta_k(x)$ denote the probability pixel intensity x in Gaussian η_k .

2.1 Model learning

The system starts with no model in the mixture of a pixel and then for every new observation x_t of the pixel at time t , it is first matched against each of the existing models where x_t is no further than 3 standard deviations or S from the mean. As setting S low has shown guaranteed high quality object detection for a wide range of surveillance test sequences in [5], we propose to use $S \approx 10\%$ of the maximum possible value of a pixel for all operating environments. Of all the matched models, the one (say η_i) with the maximum weight times the probability of x_t in the model is selected as follows:

$$i = \arg \max_{\forall k: |x_t - \mu_k| \leq \max(3\sigma_k, S)} \{\omega_k \eta_k(x_t)\} \quad (1)$$

and its associated parameters are updated as follows:

$$m_i \leftarrow x_t; c_i \leftarrow c_i + 1; \beta_i \leftarrow (1 - \alpha)/c_i + \alpha; \quad (2)$$

$$\sigma_i^2 \leftarrow (1 - \beta_i)\sigma_i^2 + \beta_i(x_t - \mu_i)^2; \quad (3)$$

$$\mu_i \leftarrow (1 - \beta_i)\mu_i + \beta_i x_t; \quad (4)$$

$$\omega_i \leftarrow (1 - \alpha)\omega_i + \alpha. \quad (5)$$

If no match is found, a new Gaussian (say η_i) is introduced with $m_i = \mu_i = x_t$, $\sigma_i = 30$, $c_i = 1$, and $\omega_i = \alpha$. The weights of the remaining Gaussians are updated as

$$\forall k \neq i : \omega_k \leftarrow (1 - \alpha)\omega_k \quad (6)$$

in both the cases. Finally weights of all the models are normalized such that $\sum_{\forall k} \omega_k = 1$.

2.2 Object detection

For each pixel, all the existing models in the mixture are sorted in descending order of their background probabilities $P(\eta_k)$'s such that after sorting $P(\eta_1) \geq P(\eta_2) \geq \dots \geq P(\eta_K)$ where the probability is defined using the following sigmoid function:

$$P(\eta_k) = 1/(1 + e^{-a\omega_k/\sigma_k + b}); \quad (7)$$

where the constants $a = 96$ and $b = 3$ are suggested in [4] after sensitivity analysis.

Now, η_1 always represents the most dominating background. However, to cater for intrinsic repetitive background motion, we also test the remaining models in sorting order whether it should be included in the set of the most dominating Gaussians. Two different statistics are utilized for this test. One is obviously the weight of the Gaussian, as the existence of similarly weighted models corresponds to the existence of a repetitive multimodal background. The other is the observation stability. Standard deviation σ_k of model η_k is a good measure of its stability as low σ_k corresponds to stable observation and high σ_k indicates varying intensities. However, this is not true when the static background is revealed after a long observation of varying intensities due to moving foregrounds as it may introduce a new model η_i with very high σ_i . To avoid this situation, we use an alternative measure $d_k = |m_k - \mu_k|$ for each Gaussian η_k . This measure is always closer to zero and shorter in Gaussians with stable observations than those representing fluctuating observations. The test for Gaussian η_k , $k = 2, 3, \dots$ is carried out as follows:

$$|\omega_1/d_1 - \omega_k/d_k|d_1/\omega_1 < f; \quad (8)$$

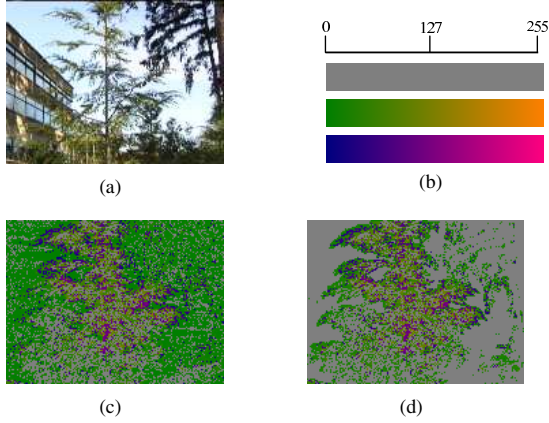


Figure 1. Visualisation of model quality and number: (a) frame 200 of the *Waving Trees* sequence; (b) distance colour mapping for single (gray), double (green-orange), and many (blue-pink) models; (c) Lee's; and (d) the proposed techniques.

where the constant $f = 0.05$ is determined from a sensitivity analysis.

If B models are identified in the set of most dominating background models, a pixel with value x_t at time t is considered background if

$$\exists i_{i=1, \dots, B} : |m_i - x_t| \leq S/i. \quad (9)$$

Note that the test threshold S/i decreases linearly with i to make sure that enough intensity band is left to represent the foreground even for large B . This measure is found to reduce false negative error where a foreground pixel is undetected.

3 Experiments

The proposed technique is evaluated by quantitative analysis and visual comparisons against Lee's [4] technique on 14 test sequences from the *PETS* [9] and *Wallflower* [8] datasets, including both indoor and outdoor surveillance scenarios from different camera angles. No post-processing was applied to evaluate the unaided strengths of each technique.

The proposed preventive measure to avoid redundant models was found significant in reducing the overall number of models from Lee's technique while concomitantly improving the object detection quality. The proposed and Lee's techniques used on average 1.745 and 2.373 models per pixel respectively. This 26.5% reduction in the number of models contributes a proportionate computational complexity improvement by

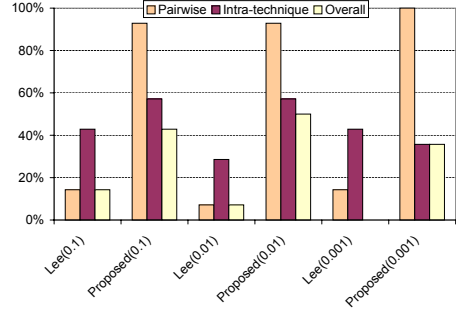


Figure 2. For three different learning rates, $\alpha = 0.1, 0.01,$ and 0.001 , pairwise (inter-technique for the same α), intra-technique, and the overall best detection rates.

the proposed technique. The object detection quality improvement was due to higher model quality, which can be measured by the average of pairwise difference of model means. Figure 1 visualises model numbers and quality combinedly for both the techniques at frame 200 of the *Waving Trees* [8] sequence with an intuitive colour mapping.

Error rate of a detection is the ratio of false positive (FP), incorrect object, and false negative (FN), incorrect background, pixels to the number of pixels in the test frame. Given a group of setups, best detection rate for each setup is the fraction of 14 test sequences where the setup's error rate is within a small constant (say 0.1%) from the best error rate among all setups. In Figure 2, best detection rates are presented in three different comparison setups. The superiority of the proposed technique is evident in the pairwise comparison at all learning rates as well as the overall comparison among the six setups. The intra-technique comparison reveals that the detection rate of neither technique was highly sensitive to the learning rate.

Table 1 presents the error rates of test sequences at medium learning rate and the standard deviation of error rates at different learning rates for both the techniques (pairwise best in bold). The proposed technique was more robust as its error rates were far less sensitive to learning rates. The superiority of the proposed technique in producing negligible noise, shadow and trailing effects can be visually verified in Figure 3.

4 Conclusion

By eliminating near-duplicate models and using basic background subtraction from believe-to-be back-

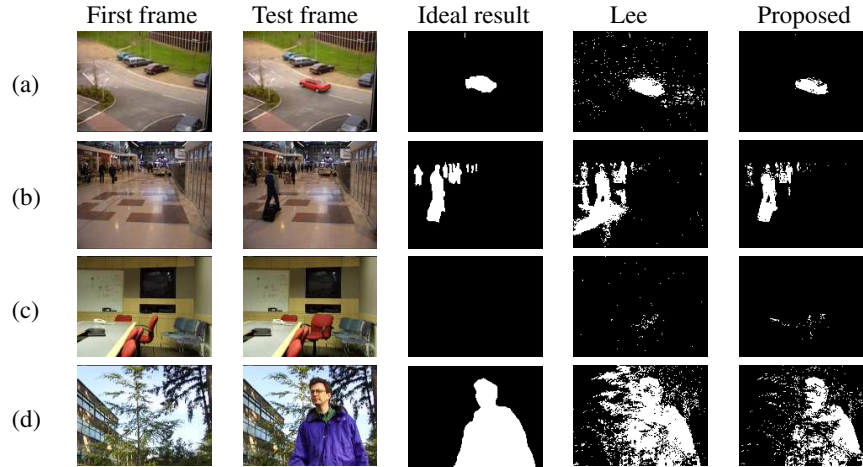


Figure 3. Visual comparison results at medium learning rate, $\alpha = 0.01$ for test sequence (a) PETS2000; (b) PETS2006-B1; (c) Moved Object; and (d) Waving Trees.

Table 1. Error rates at medium learning rate and the standard deviation of the error rates over three learning rates.

Test Sequence	%Error Rate (FP + FN)			
	$\alpha = 0.01$		Stddev	
	Lee	Proposed	Lee	Proposed
PETS2000	4.1	1.6	1.3	0.0
PETS2006-B1	10.3	4.2	1.2	0.5
PETS2006-B2	3.8	2.8	0.3	0.3
PETS2006-B3	5.6	2.5	1.1	0.3
PETS2006-B4	11.3	5.6	1.1	0.9
Bootstrap	13.3	11.8	2.1	1.3
Camouflage	29.8	13.6	9.6	1.5
Fground Aper.	67.2	15.8	7.4	0.1
Light Switch	86.1	84.0	32.9	14.4
Moved Object	0.5	0.4	3.3	3.6
Time Of Day	4.1	5.7	7.0	0.6
Waving Trees	19.2	15.8	0.5	0.6
Football	33.4	21.7	10.8	2.0
Walk	0.5	0.3	0.6	0.1

grounds from Gaussian mixture models, we have developed a robust object detection technique with reduced computational complexity and superior quality demonstrated on benchmark test sequences.

References

- [1] M. S. Allili, N. Bouguila, and D. Ziou. A robust video foreground segmentation by using generalized gaussian mixture modeling. In *Fourth Canadian Conf. on Computer and Robot Vision*, pages 503–509, 2007.
- [2] S. S. Huang, L. C. Fu, and P. Y. Hsiao. Region-level motion-based background modeling and subtraction using mrfs. *IEEE Trans. Image Process.*, 16:1446–1456, 2007.
- [3] P. KaewTraKulPong and R. Bowden. An improved adaptive background mixture model for realtime tracking with shadow detection. In *2nd European Workshop on Advanced Video Based Surveillance Systems*, 2001.
- [4] D. S. Lee. Effective gaussian mixture learning for video background subtraction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27:827–832, 2005.
- [5] J. C. Nascimento and J. S. Marques. Performance evaluation of object detection algorithms for video surveillance. *IEEE Trans. Multimedia*, 8:761–774, 2006.
- [6] P. W. Power and J. A. Schoonees. Understanding background mixture models for foreground segmentation. In *Image and Vision Comp. New Zealand*, pages 267–271, 2002.
- [7] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Comp. Soc. Conf. on Comp. Vision and Patt. Recog.*, volume 2, pages 246–252, 1999.
- [8] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Seventh IEEE Int. Conf. on Computer Vision*, volume 1, pages 255–261, 1999.
- [9] www.cvg.rdg.ac.uk/slides/pets.html. Pets: Performance evaluation of tracking and surveillance, oct 2007.
- [10] H. C. Zeng and S. H. Lai. Adaptive foreground object extraction for real-time video surveillance with lighting variations. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 1, pages 1201–1204, 2007.