University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2010

On the analysis of background subtraction techniques using Gaussian mixture models

Abdesselam Bouzerdoum
University of Wollongong, bouzer@uow.edu.au

Azeddine Beghdadi

Son Lam Phung
University of Wollongong, phung@uow.edu.au

Philippe L. Bouttefroy
philippe@uow.edu.au

Publication Details

Bouttefroy, P. L., Bouzerdoum, A., Beghdadi, A. & Phung, S. (2010). On the analysis of background subtraction techniques using Gaussian mixture models. 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (pp. 4042-4045). USA: IEEE.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

On the analysis of background subtraction techniques using Gaussian mixture models

Abstract

In this paper, we conduct an investigation into background subtraction techniques using Gaussian Mixture Models (GMM) in the presence of large illumination changes and background variations. We show that the techniques used to date suffer from the trade-off imposed by the use of a common learning rate to update both the mean and variance of the component densities, which leads to a degeneracy of the variance and creates "saturated pixels". To address this problem, we propose a simple yet effective technique that differentiates between the two learning rates, and imposes a constraint on the variance so as to avoid the degeneracy problem. Experimental results are presented which show that, compared to existing techniques, the proposed algorithm provides more robust segmentation in the presence of illumination variations and abrupt changes in background distribution.

Keywords

techniques, analysis, models, mixture, gaussian, background, subtraction

Disciplines

Physical Sciences and Mathematics

Publication Details

Bouttefroy, P. L., Bouzerdoum, A., Beghdadi, A. & Phung, S. (2010). On the analysis of background subtraction techniques using Gaussian mixture models. 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (pp. 4042-4045). USA: IEEE.

This conference paper is available at Research Online: http://ro.uow.edu.au/infopapers/809

ON THE ANALYSIS OF BACKGROUND SUBTRACTION TECHNIQUES USING GAUSSIAN MIXTURE MODELS

P. L. M. Bouttefroy, A. Bouzerdoum, S. L. Phung

School of Electrical, Computer & Telecom. Engineering
University of Wollongong
Northfields Avenue
Wollongong, NSW 2522, Australia

A. Beghdadi

L2TI, Institut Galilée
Université Paris 13
Avenue J. B. Clément
93430 Villetaneuse, France

ABSTRACT

In this paper, we conduct an investigation into background subtraction techniques using Gaussian Mixture Models (GMM) in the presence of large illumination changes and background variations. We show that the techniques used to date suffer from the trade-off imposed by the use of a common learning rate to update both the mean and variance of the component densities, which leads to a degeneracy of the variance and creates "*saturated pixels*". To address this problem, we propose a simple yet effective technique that differentiates between the two learning rates, and imposes a constraint on the variance so as to avoid the degeneracy problem. Experimental results are presented which show that, compared to existing techniques, the proposed algorithm provides more robust segmentation in the presence of illumination variations and abrupt changes in background distribution.

***Index Terms*—** Motion analysis, Video signal processing, Image segmentation, Object detection.

1. INTRODUCTION

Background modeling of indoor and outdoor scenes is a challenging task due to the uncontrolled nature of the environment (*e.g.* changes in illumination). A number of techniques have been developed to extract relevant motion from videos since the frame differencing method proposed by Jain and Nagel in 1979 [1]. For instance, Elgammal *et al.* proposed to model the background through kernel density estimation [2] and, Oliver *et al.* developed an algorithm that generates the eigenbackground of a scene, capturing the stable components or, equivalently, the background [3]. Nevertheless, the computation cost prohibits the use of these techniques for real-time applications. Parametric representations, which limit the processing load to the calculation of a few parameters, are preferred. In particular, Stauffer and Grimson have developed an algorithm for foreground segmentation using the Gaussian mixture model [4].

However, despite the plethora of techniques proposed to model background, there is a lack of rigorous analysis of the background model itself and, in particular, the influence of the update rates. Lee proposed a variation of Stauffer and Grimson's algorithm which adapts quickly to fast variations in background distribution imputable, for example, to changing illumination [5]; unfortunately, this approach leads to pixel saturation due to degeneracy of the variance as we will show later. Other proposed techniques to handle illumination changes rely on chrominance correspondence between the incoming pixel and the reference background model (for more details see [6]). These techniques are therefore inefficient when the reference model has changed.

In this paper, we conduct an investigation into the effects of the Gaussian Mixture parameters, in particular the influence of the variance, on the performance of Stauffer and Grimson's and Lee's algorithms. Our study shows that the variance can degenerate creating the so-called *saturated pixels*, which appear due to a change in illumination. Furthermore, we propose a method that can cope with rapid changes in the background model. Contrary to existing techniques, the proposed method handles abrupt variations during the mixture update and not in a pre- or post-processing stage. The rest of the paper is organized as follows. Section 2 presents the general formulation of the mixture of Gaussians for background modeling. Section 3 describes the problem of variance degeneracy and defines saturated pixels. Section 4 introduces the adaptive mixture of Gaussians algorithm. Section 5 presents the experimental results, followed by concluding remarks in Section 6.

2. BACKGROUND MODELING USING GAUSSIAN MIXTURES

Background modeling by Gaussian mixtures is a pixel based process. Let $\mathbf{x}$ be a random process representing the value of a given pixel in time. A convenient framework to model the probability density function of $\mathbf{x}$ is the parametric Gaussian mixture model where the density is composed of a sum of Gaussians. Let $p(\mathbf{x})$ denote the probability density function of a Gaussian mixture comprising K component densities:

Fig. 1. Original (*first row*) and foreground segmentation (*second row*) of an object passing through a saturated zone (aggregation of saturated pixels) in the *HighwayII* sequence. The saturated region is delimited with a rectangle.

$$p(\mathbf{x}) = \sum_{k=1}^{K} w_k \, \mathcal{N}(\mathbf{x}; \mu_k, \sigma_k), \tag{1}$$

where w_k are the *weights*, and $\mathcal{N}(\mathbf{x}; \mu_k, \sigma_k)$ is the normal density of mean $\mu_\mathbf{k}$ and covariance matrix $\Sigma_k = \sigma_k I$, (I denotes the identity matrix). The mixture of Gaussians algorithm, proposed by Stauffer and Grimson [4], estimates these parameters over time to obtain a robust representation of the background. First, the parameters are initialized with $w_k = w_0$, $\mu_k = \mu_0$ and $\sigma_k = \sigma_0$. If there is a *match*, i.e.

$$\|\mathbf{x} - \mu_j\|/\sigma_j < \tau \text{ for some } j \in [1..K], \tag{2}$$

where $\tau(> 0)$ is some threshold value, then the parameters of the mixture are updated as follows:

$$w_k(t) = (1-\alpha)\, w_k(t-1) + \alpha\, M_k(t), \tag{3}$$

$$\mu_k(t) = (1-\beta)\mu_k(t-1) + \beta\, \mathbf{x}, \tag{4}$$

$$\sigma_k^2(t) = (1-\beta)\, \sigma_k^2(t-1) + \beta\, ||(\mathbf{x} - \mu_k(t))||^2, \tag{5}$$

where $M_k(t)$ is equal to 1 for the matching component j and 0 otherwise. If there is no match, the component with the lowest weight w_k is re-initialized with $w_k = w_0$, $\mu_k = \mathbf{x}$ and $\sigma_k = \sigma_0$. The learning rate α is constant and β is defined as:

$$\beta = \alpha\, \mathcal{N}(\mathbf{x}; \mu_k, \sigma_k)\,. \tag{6}$$

Finally, the weights w_k are normalized at each iteration to add up to 1. Stauffer and Grimson proposed to sort the Gaussians by decreasing weight-to-standard-deviation ratio, w_k/σ_k, to represent the background. A threshold λ is applied to the cumulative sum of weights to find the set $\{1...B\}$ of Gaussians modeling the background, defined as:

$$B = \text{argmin}_{K_B} \left(\sum_{k=1}^{K_B \leq K} w_k > \lambda \right). \tag{7}$$

Intuitively, Gaussians with the highest probability of occurrence, w_k, and lowest variability in the distribution, measured by σ_k, indicating a representative mode, are the most likely to model the background.

3. VARIANCE DEGENERACY AND PIXEL SATURATION

The Gaussian mixture algorithm used for background subtraction by Stauffer and Grimson updates the mean μ_k and the variance σ_k with the same learning rate β. Lee proposed a modified approach where the learning rate β is increased in the initial learning phase of the algorithm, hence providing a quicker adaptation when a new surface appears [5]. In this section, we focus on the approach proposed by Stauffer and Grimson [4] and Lee [5] to analyze the learning rates[1]. It is convenient herein to introduce the term "*saturated pixel*" that defines a pixel which responds neither to a foreground nor to a background object. In other words, a saturated pixel is classified either as foreground or as background, regardless of the surface present. The phenomenon is illustrated in Fig. 1. In this example, the saturated pixels are classified as background even though the zone inside the square alternates between foreground and background, due to flow of cars.

The increase of the learning rate β and large variations of the pixel value distribution (e.g., recurrent changes in illumination) are accountable for the saturation of pixels. Indeed, the update of the variance (Eq. 5) with the square of the difference between the mean and the pixel value is prone to result in large, overestimated, values for the variance. The variance increases until the component saturates the mixture by spanning over the entire pixel color range. As a consequence, pixels are constantly classified as either foreground or background, depending on the weight of the Gaussian component; indeed, Eq. (2) always holds if the variance is large enough.

To the best of our knowledge, this phenomenon has never been investigated and was considered as an adaptation of the background to a changing distribution. However, it can be shown from Eq. (3) that this assumption does not hold because the minimum time t_{min} for a new distribution to be part of the background model is[2]

$$t_{min} \geq \frac{1}{\ln(1-\alpha)} \ln \left[\frac{K + \lambda - 3}{(K-2)(\alpha - 1)} \right]. \tag{8}$$

[1] Techniques which implement an expectation maximization algorithm are not considered due to their computational cost.

[2] The derivation of the equation is omitted due to lack of space.

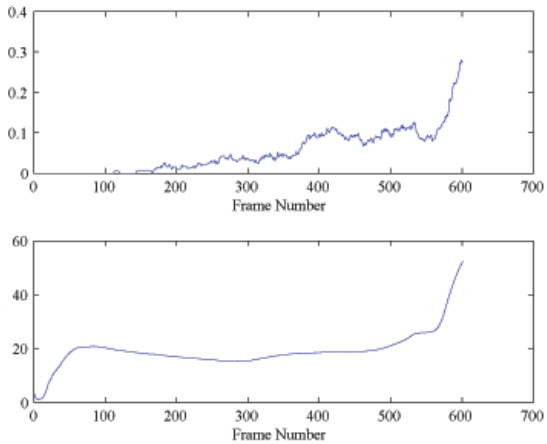


Fig. 2. Percentage of saturated pixels: (a) Stauffer and Grimson's method (*top*), (b) Lee's method (*bottom*).

There is equality when a new object presents a surface with a constant color value. Consider $\lambda = 0.7$ and $\alpha = 0.005$, as in [5], a component is integrated to the background after 72 frames for $K = 3$. Such a large number discards the assumption of fast background adaptation to explain the phenomenon presented in Fig. 1. Figure 2 displays the percentage of saturated pixels for the Stauffer and Grimson's method and Lee's method. The learning speed-up in Lee's method leads to a significant percentage of the image being saturated.

4. PROPOSED APPROACH

As seen in Section 3, the method proposed by Stauffer and Grimson lacks adaptation to fast changes due to the use of a constant learning rate. On the other hand, the method proposed by Lee suffers from pixel saturation because of the fast adaptation of the variance. Here, we propose an algorithm that handles the degeneracy of the variance via adequate update of the parameters. To this end, the learning rates for the mean and the variance terms are decoupled. The first one is an adaptive learning rate γ_k, updating the parameter μ_k that includes clues of the relative probability $q_k = \mathcal{N}(\mathbf{x}; \mu_k, \sigma_k)$ of a pixel belonging to the kth Gaussian:

$$\gamma_k(t) = \gamma_k(t-1) + \frac{K+1}{K} q_k - \frac{1}{K} \sum_{i=1}^{K} q_i \,. \qquad (9)$$

Experiments showed that an initial value of $\gamma_k(0) = 0.05$ provides good results for the tested datasets. The above equation shows that the learning rate γ_k increases if the value of the kth Gaussian component is higher than the average value of other components. In other words, γ_k is increased if the probability of the pixel belonging to the kth component is greater than the average probability of belonging to another component; otherwise, it is decreased. An adaptive coefficient allows a quicker update of the Gaussian mixture mean as in [5], and copes with fast changes in illumination, thus addressing the problem encountered with Stauffer and Grimson method. The learning rate γ_k replaces β in Eq (4). The variance is updated independently from the mean so as to avoid the saturation phenomenon; a fast update would lead to the degeneracy problem highlighted in Section 3. Here we propose to use a semi-parametric model for the variance, which enables a quasi-linear adaptation in case of small deviations from the mean and a flattened response for large deviations. The sigmoid function is used to this aim:

$$f_{a,b}(\mathbf{x}, \mu_k) = a + \frac{b-a}{1 + e^{-s\varepsilon(\mathbf{x},\mu_k)}} \,, \qquad (10)$$

where $\varepsilon(\mathbf{x}, \mu_k) = (\mathbf{x} - \mu_k)^T (\mathbf{x} - \mu_k)$. The parameter s controls the slope of the sigmoid and is equal to 0.005 in the experiments . The update of the variance in Eq. (5) is thus

$$\sigma_k^2(t) = (1-\eta)\, \sigma_k^2(t-1) + \eta f_{a,b}(\mathbf{x}, \mu_k(t-1)) \,, \qquad (11)$$

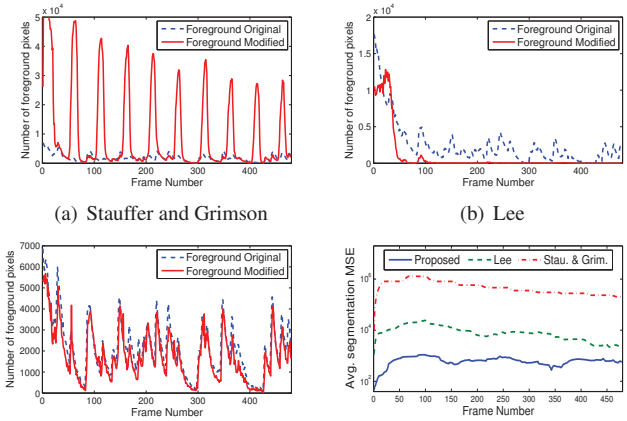
where $\eta = 0.6$ is a constant learning rate. The function $f_{a,b}(\mathbf{x}, \mu)|_{\Re^+}$ bounds the variance value to the domain $\mathcal{D} \in [\frac{a+b}{2}, b]$. The values of the parameters a and b are chosen so that the domain $\mathcal{D}$ spans over one Kth of the pixel range value.

5. EXPERIMENTAL RESULTS

The proposed algorithm is compared with the methods developed by Stauffer and Grimson [4] and by Lee [5] on a database containing several hours of video sequences from indoor and outdoor environments. The database is composed of video-surveillance footage, pedestrian and car sequences. We first test the three algorithms by adding illumination changes, generated artificially, to a video sequence of real scenes. Evaluation of the performance on natural illumination changes is then presented.

Changes in illumination are fast but smooth due to the gradual transition from penumbra to umbra as described in [7]. We therefore model the variation with a time varying sine term $\sin(0.02\pi t)$, whose value is added to every pixel value in each frame of the *HighwayII* sequence, resulting in a modified sequence. To test the adaptation performance to illumination changes, the foreground from the original and the modified sequences are extracted by the three algorithms. Then, the pixels of the extracted foreground are counted for each frame. Here, the segmentation of the original sequence is used as reference to compare with, or pseudo-ground truth. Figures 3(a)–(c) display the count of foreground pixels through time and Fig. 3(d) shows the average (over 50 frames) of the mean squared error (MSE). The difference in count between the original and the modified sequence represents the error of segmentation attributable to the illumination changes. The number of pixels extracted from the original sequence, the reference, is consistent across the

three algorithms (dashed line). However, with illumination changes, Stauffer and Grimson's method is too slow updating the model: some background is detected as foreground and portions of sine are clearly visible. Lee's method shows only residual foreground detection after 120 frames due to the degeneracy of the variance for most pixels. Only the proposed algorithm is able to accurately segment the foreground in the presence of illumination changes.



(a) Stauffer and Grimson (b) Lee

(c) Proposed (d) Segmentation MSE

Fig. 3. Number of detected foreground pixels for the original (*dashed line*) and modified (*solid line*) *HighwayII* sequence; (a), (b) and (c) represent Stauffer and Grimson's, Lee's and the proposed method, respectively. (d) is the average segmentation MSE for the three techniques.

Natural changes in illumination are difficult to evaluate since the ground truth is not available. Traditionally, a qualitative evaluation is performed. Figure 4 shows excerpts for foreground segmentation. The displays are raw foreground masks, *i.e.* no post processing (opening, closure, median filtering, etc.) has been applied. The proposed algorithm outperforms Stauffer and Grimson's and Lee's techniques for global illumination changes (Fig. 4, rows 1 to 3) and for shadow edge (Fig. 4, row 4). It is worthwhile noting that saturated pixels here appear as foreground (black) for Lee's method.

6. CONCLUSION

This paper proposed a new technique for background subtraction by mixture of Gaussians that can handle large variations in the background intensity distribution. It was shown that the phenomenon of pixel saturation is due to the degeneracy of the variance of some Gaussian mixture components, a consequence of a large learning rate. To address this problem, the variance is bounded, and the mean and variance of the Gaussian components are updated with different learning rates. The mean is updated with an adaptive rate, providing an accelerated update when sudden illumination changes occur. Experimental results were presented, which show that the

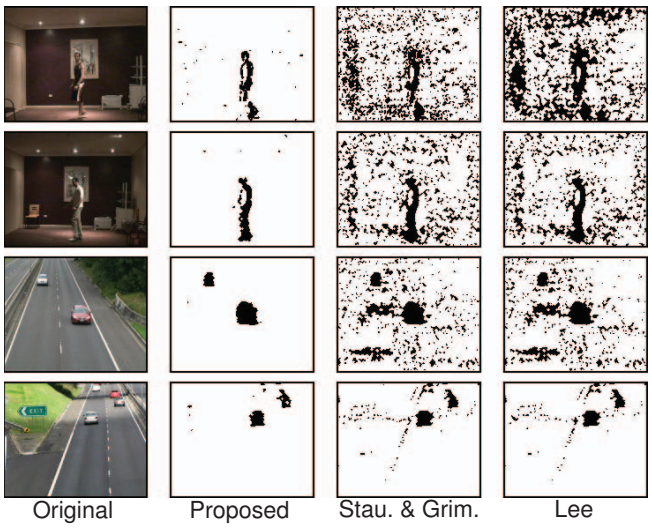


Fig. 4. Examples of foreground mask for indoor and outdoor scenes. *From left to right*: original, proposed algorithm, Stauffer and Grimson, Lee.

proposed method is robust to large variations in pixel intensity values and abrupt changes in background distribution.

7. REFERENCES

[1] R.C. Jain and H.H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 206–213, 1979.

[2] A. Elgammal, D. Harwood, and L. Davis, "Nonparametric model for background subtraction," in *European Conf. on Computer Vision*, Dublin, 2000, vol. 1843 of *Lecture Notes in Computer Science*, pp. 751–767.

[3] N. M. Oliver, B. Rosario, and A. P. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, 2000.

[4] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1999, vol. 2, pp. 246–252.

[5] D.-S. Lee, "Effective gaussian mixture learning for video background subtraction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 827–832, 2005.

[6] N. Martel-Brisson and A. Zaccarin, "Learning and removing cast shadows through a multidistribution approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1133–1146, 2007.

[7] L. Zhou, H. Kaiqi, T. Tieniu, and W. Liangsheng, "Cast shadow removal with gmm for surface reflectance component," in *Proc. of IEEE International Conference on Pattern Recognition*, 2006, vol. 1, pp. 727–730.