Self-Adaptive Gaussian Mixture Model for Urban Traffic Monitoring System

Zezhi Chen
Digital Imaging Research Centre
School of Computing and Information Systems
Kingston University London, Kingston Upon Thames, UK
z.chen@kingston.ac.uk

Tim Ellis
Digital Imaging Research Centre
School of Computing and Information Systems
Kingston University London, Kingston Upon Thames, UK
t.ellis@kingston.ac.uk

Abstract

Identifying moving vehicles is a critical task for an urban traffic monitoring system. With static cameras, background subtraction techniques are commonly used to separate foreground moving objects from background at the pixel level. Gaussian mixture model is commonly used for background modelling. Most background modelling techniques use a single leaning rate of adaptation which is inadequate for complex scenes as the background model cannot deal with sudden illumination changes. In this paper, we propose a self-adaptive Gaussian mixture model to address these problems. We introduce an online dynamical learning rate and global illumination of background model adaptation to deal with fast changing scene illumination. Results of experiments using manually-annotated urban traffic video with sudden illumination changes illustrate that our algorithm achieves consistently better performance in terms of ROC curve, detection accuracy, Matthews correction coefficient and Jaccard coefficient compared with other approaches based on the widely-used Gaussian mixture model.

1. Introduction

Identifying moving objects in a video sequence is a fundamental and critical task in video surveillance, traffic monitoring and analysis, human detection and tracking, as well as other visual tracking task, such as gesture recognition in human-machine interface. A static camera observing a scene is a common case of surveillance and monitoring system. Background modelling is often used in different applications to model the background and then detect the moving objects in the scene. The general theory is that the background model is built from the data and objects are detected if they appear significantly different from the background. The foreground pixels are further processed for object detection and tracking. The principal challenges are how to correctly and efficiently model and update the background model and how to deal with shadows. A robust system should be independent of the scene, and robust to lighting effects and changeable weather conditions. It should be capable of dealing with movement through cluttered areas, objects overlapping in the visual field, gradual illumination changes (e.g. time of day, evening and night), sudden illumination changes (e.g. switching a light on or off, clouds moving in front of the sun), camera automatic gain control (e.g. white balance and iris are often applied to optimally map the amount of reflected light to the digitizer dynamic range), moving background (e. g. camera vibration, swaying trees, snowing or raining day), slow-moving objects, cast shadows and geometric deformation of foreground objects.

At the heart of any background subtraction algorithm is the construction of a statistical model that describes the state of each background pixel. Many algorithms have been developed and the most recent surveys can be found in [1, 2, 3, 4]. The simplest form of the background model is a time-averaged background image of the scene without any intruding objects. However, this method suffers from many problems such as requires a large memory and a training period where foreground objects are absent, to ensure that static foreground objects are not considered as a part of the background. This will limit their utility in real time applications. In an urban traffic environment, the 'pure' background isn't available and can always be changed under critical situations like objects being introduced or removed from the scene, slow-moving or stationary objects. To account for these problems of robustness and adaptation, many background modelling methods have been developed. A single Gaussian model was used in real time tracking of the human body [5]. However, individual pixel values can often have a complex distribution and more elaborate models are needed. A Gaussian mixture model (GMM) was proposed for real-time tracking by Stauffer and Grimson [6, 7]. The algorithm relies on assumptions that the background is visible more frequently than foreground and that the model has a relatively narrow variance. The approach has been found to cope reliably with slow lighting changes,

978-1-4673-0063-6/11/$26.00

repetitive motions from clutter, and long-term scene changes. Many adaptive GMM models have been proposed to improve the original background subtraction method. Power and Schoonees [8] used a hysteresis threshold to extend the GMM model. They introduced a faster and more logical application of the fundamental approximation than that used in [7]. The standard GMM update equations were extended to improve the speed and adaptation rate of the model [9][10][11][12]. All these GMMs use a fixed number of components. Zivkovic and Heijden [13] [14] presented an improved GMM model using a recursive computation to constantly update the parameters of a GMM, which adaptively chose the appropriate number of Gaussians to model each pixel on-line, from a Bayesian perspective. We refer to this method as the Zivkovic-Heijden Gaussian mixture model (ZHGMM) later in the text. Greggio et al. [15] proposed self-adaptive Gaussian mixture models for real-time background subtraction recently. The main contribution is that the approach first learnt a description mixture that best described the first video frame, and then use it to initialize the further frames.

However, GMM's have drawbacks. Firstly, they are computationally intensive and the parameters require careful tuning. Second, they are sensitive to sudden changes in global illumination. If a scene remains stationary for a long period of time, the variance of the background components may become very small. A sudden change in global illumination can then turn the entire frame into foreground. For a low learning rate, it produces a very wide and inaccurate model that will have low detection sensitivity. On the other hand, for a high learning rate, the model updates too quickly, and slow moving objects will be absorbed into the background model, resulting in a high false negative rate.

Non-parametric density estimations also lead to flexible models [16][17][18]. These model the pixel as a random variable in feature space with an associated probability density function. Unlike GMMs, non-parametric models do not require the selection of a number of Gaussians to be fitted, but the model adaptation is trivial. The major drawback is its computation cost.

Rapid illumination changes constitute one of the main difficulties when applying background subtraction in a real life setting. Our work is motivated by the need for robust vehicle detection and classification algorithm that can be used in a traffic monitoring system to deal with illumination sudden changes. Martel-Brisson et al. [19] proposed a novel statistical model based on a GMM. This model can deal with scenes with complex and time-varying illumination, including regions that are highly saturated, whilst suppressing false detection in regions where shadows cannot be detected. Vosters et al. [20] combined the eigenbackground algorithm with a statistical illumination model and proposed a background subtraction method to cope with sudden illumination changes. Self-organisation through artificial neural networks can be used to handle scenes containing moving backgrounds, gradual illumination variations and camouflage, can model shadows cast by moving objects [21][22]. Reddy et al. proposed an adaptive patch-based background modelling for improved foreground object segmentation and tracking [23]. Liao et al. [24] developed a scale invariant local ternary pattern operator and pattern kernel density estimation technique to tackle illumination variations and dynamic backgrounds. Pilet et al. [25] presented a fast background algorithm by relaying on a statistical model, not of the pixel intensities, but of the illumination effects. Zhao et al. [26] extended the background GMM to the spatial relations, the joint colors of each pixel-pair are modeled a GMM to suppress the effects of illumination changes. Izadi et al. [27] and Javed et al. [28] employed color and gradient information to build up the background GMM to reduce the foreground detection false alarm rate in pixel level.

Unfortunately, none of the existing background models can achieve desirable performance to sudden changes in global illumination. Moreover, the parameters of a GMM may vary for different scenarios. Existing work either openly admits to setting blending and thresholding parameters by hand, or more commonly, does not mention how they are set.

The algorithm proposed in this paper combines a pixel-level and frame-level process. Systematic statistic analysis shows how to select some significant parameters to obtaining an optimal background model for a given scene. There are four main contributions in the paper: 1) An online dynamical learning-rate adaptation. 2) Global illumination of background model changing adaptation to deal with background illumination fast change. 3) Improving the shadow removal algorithm to deal with the sudden illumination changes. 4) Use of statistical measurements to select some significant parameters to obtain optimal results. Extensive experimental results illustrate that our approach consistently provides better performance than that of ZHGMM.

The remainder of the paper organised as follows. The new algorithm description is given in Section 2. Pixel level performance metrics are introduced in Section 3. Extensive quantitative evaluation is given in Section 4. Finally Section 5 concludes the paper.

2. Algorithm description

2.1. Global illumination changing model

Illumination-invariant change detection model (ICDM) is a process of identifying illumination variation versus time. Changing image illumination causes problems for many computer vision applications operating in unconstrained environments, and especially affect background subtraction methods. The most trivial approach for ICDM is the subtraction of intensities of two sequential video frames. The main disadvantage of such a simple method is its sensitivity to noise. Beiderman et al. proposed an illumination insensitive reconstruction and pattern recognition using spectral manipulation and K-factor spatial transforming [29]. Xie et al. developed a change detection algorithm that deals with sudden illumination changes using order consistency [30]. Various intensity estimation methods are compared in the paper [31]. Usability is evaluated with background classification. They give an accurate non-iterative estimate of the apparent gain factor by experimentally comparing eight algorithms. According to simulation results, many algorithms seem to perform very well. The Median of Quotient (MofQ) performed best, both with and without outlier removal.

For all pixels *s* in set *S*, the MofQ global illumination changing factor *g* between current image i_c and reference image i_r defined as

$$g = median_{s\in S}\left(\frac{i_{c,s}}{i_{r,s}}\right) \quad (1)$$

2.2. Self-adaptive Gaussian mixture model

We have developed a self-adaptive Gaussian mixture model (SAGMM) based on the ZHGMM algorithm. The pixel level ZHGMM is effective in modelling a multiple-modal distributed background. In order to adapt to possible changes, the training set should be updated. We choose a reasonable time adaption period T (for instance 100 frames). Let $x^{(t)}$ be a pixel value at time *t*, we have $X_T = \{x^{(t)}, x^{(t-1)}, \cdots, x^{(t-T)}\}$. For each new sample we update the training data set X_T and re-estimate the density. These samples contain values that belong to both the background (BG) and foreground (FG) object. Therefore, we should denote the estimated density as $\hat{p}(x^{(t)} \mid X_T, BG+FG)$ [13]. We use a GMM with K components (normally between 3-7, we set it as 4).

$$\hat{p}(x^{(t)} \mid X_T, BG+FG) = \sum_{m=1}^{K} w_m \cdot \eta(x^{(t)}; \mu_m, \Sigma_m) \quad (2)$$

$$\eta(x^{(t)}; \mu_m, \Sigma_m) = \sum_{i=1}^{K} \frac{w_{i,t}}{(2\pi)^{d/2}\left|\Sigma_{i,t}\right|^{1/2}} e^{-\frac{1}{2}(x^{(t)}-\mu_{i,t})^T \Sigma_{i,t}^{-1}(x^{(t)}-\mu_{i,t})} \quad (3)$$

Where μ_m is the estimated mean value, Σ_m is the estimated covariance matrix, and w_m are non-negative estimated mixing weights and normalized (add up to 1) for the *m*th GMM at time *t*. For computational reasons (easily invertible), an assumption is usually made that each channel of the colour space is independent from others, so Σ_m is a diagonal matrix. A further assumption is that the red, green and blue pixel values are independent (though may have the same variance [6]). The covariance matrix is assumed to be of the form $\Sigma_m = \sigma_m I$, where *I* is a 3×3 identity matrix. While this is certainly not the case, the assumption allows us to avoid a costly matrix inversion at the expense of some accuracy. Note that a single σ_m may be a reasonable approximation in such a linear colour space, but it may be an over- simplification in non-linear colour spaces.

We developed a new algorithm by incorporating a modified adaptive schedule into the recursive ZHGMM learning procedure. The global illumination change MofQ factor g between the learnt background and the current input image, and a counter *c* for each Gaussian component in the mixture model are introduced. The factor *g* keeps track of how the global illumination changes and the counter *c* keeps track of how many data points have been contributed to the parameter estimation of that Gaussian. Each time the parameters are updated, a learning rate β is calculated based on the basic learning rate α and current accumulative counter of *c*. Given a new data sample $x^{(t)}$ at time *t* the recursive update equations are

$$w_m = (1-\alpha)w_m + \alpha(o_m^{(t)} + c_T) \quad (4)$$

$$\beta_m = \alpha(1+c_m)/c_m \quad (5)$$

$$\mu_m = \mu_m + o_m^{(t)}(\beta_m/w_m)\delta_m \quad (6)$$

$$\sigma_m^2 = \sigma_m^2 + o_m^{(t)}(\beta_m/w_m)(\delta_m^T\delta_m - \sigma_m^2) \quad (7)$$

$$c_m = c_m + 1 \quad (8)$$

where $x^{(t)} = [x_1, x_2, x_3]^T$, $\mu_m = [\mu_1, \mu_2, \mu_3]^T$, $\delta_m = x^{(t)} - \mu_m$ for a 3-channel colour image. Instead of the time interval *T* that was mentioned above, here the constant $\alpha = 1/T$ defines an exponentially decaying envelope that is used to limit the influence of the old data. c_T is a negative complexity prior evidence weight [14], means that we will accept that the class exists only if there is enough evidence from the data for the existence of the class. It will suppress the components that are not supported by the data and we discard the components with negative weights. This also ensures that the mixture weights are non-negative. c_m is the counter. It is increased when parameters of a Gaussian updated. When the Gaussian is re-assigned, it is reset to 1 since the old Gaussian has perished and a new one is initiated with a single value [10]. For a new sample the ownership $o_m^{(t)}$ is set to 1 for the "close" component with largest w_m and the others are set to zero. We define that a sample is "close" to a component if the Mahalanobis distance (MD) from the component is, for example, less than d_c. The squared distance from the m*th* component is calculated as:

$$D_m^2(x^{(t)}) = \hat{\delta}_m^T \Sigma_m^{-1} \hat{\delta}_m \quad (9)$$

where $\hat{\delta}_m = g \cdot x^{(t)} - \mu_m$. If there are no "close" components a new component is generated with

$w_{m+1} = \alpha$; $\mu_{m+1} = x^{(t)}$; $\sigma_{m+1} = \sigma_0$; $c_{m+1} = 1$.

where σ_0 is some appropriate initial variance. If the maximum number of components K is reached we discard the component with the smallest w_m. After each weight update of (4), we need to renormalize such that $\sum_{\forall k} w_k = 1$.

The algorithm provides an on-line clustering algorithm. Usually, intruding foreground objects will be represented by some additional clusters with small weights w_m. Therefore, we can approximate the background model by the first B largest clusters:

$$\hat{p}\left(x^{(t)} \mid X_T, BG\right) \sim \sum_{m=1}^{B} w_m \cdot \eta\left(x^{(t)}; \mu_m, \Sigma_m\right) \tag{10}$$

If the components are sorted to have descending weights w_m, we have

$$B = \arg\min_b \left(\sum_{i=1}^{b} w_i > \left(1 - c_f\right) \right) \tag{11}$$

where c_f is a measure of the maximum portion of the data that should be accounted for by the foreground objects without influencing the background model.

This results in three significant advantages for this approach: i) When an object is allowed to become a part of the background, it doesn't destroy the existing model of the background. The original background values remain in the GMM if the object remains static for long enough, and its weight becomes larger than c_f. If the object then moves, the distribution describing the previous background still exists with the same estimated mean and variance. ii) From the dynamic learning rate update equation (5), it can be seen that if background changes quickly, the value of c_m will become smaller and the new learning rate β_m will increase. So the background model will update quickly. The model will quickly achieve good estimation of mean and variance. Maintaining a dynamic learning rate for each Gaussian component will improve convergence and approximation of a smaller data cluster. Otherwise, if the background is stable, as more data samples are included in its parameter estimation, β_m will approach the basic learning rate α, while still maintaining the same temporal adaptability, because the weights update equation (4) still uses the basic learning rate. iii) From formula (9), we can see that the MD calculation will compensate for the global illumination change. This makes the MD insensitive to sudden illumination change.

2.3. Shadow removal

The GMM is susceptible to both global and local illumination changes, such as shadows and highlight reflections. These can cause the failure of consequent processes, e.g. tracking, classification, etc. One of the main challenges in these applications is identifying shadows which objects cast and which move along with them in the scene. It is desirable to discriminate between targets and their shadows. It is also important to recognize the type of features utilized for shadow detection. Basically, these features are extracted from three domains: spectral, spatial and temporal. Some approaches exploit different spectral features, i.e. using gray level or colour information [32]. Others improve results by using spatial information working at the region level or at the frame level instead of the pixel level [16]. The shadow removal algorithm used in this paper is based on the algorithm proposed by Horprasert [33]. Extensive experiments for shadow removal in 5 colour spaces have been done in the paper [34]. The experimental results show that the method proposed by Horprasert is very stable in RGB colour space. We have updated the shadow removal algorithm to deal with sudden global illumination change by introducing the MofQ factor *g*.

The distortion measurement in RGB space can be decomposed into two components, brightness distortion and chromaticity distortion. The *j*th pixel's brightness distortion B_j is a scalar value that brings the observed colour close to the expected chromaticity line.

For the *j*th pixel value of $I_j = \left[I_{Rj}, I_{Gj}, I_{Bj}\right]^T$ in RGB space, the estimated mean is $E_j = \left[\mu_{Rj}, \mu_{Gj}, \mu_{Bj}\right]^T$. Considering balancing colour bands by rescaling the colour values by the pixel standard deviation σ_i, the brightness and chromaticity distortion become

$$B_j = \frac{gI_j E_j}{E_j^2} \qquad CD_j = \frac{\sqrt{gI_j - B_j E_j}}{\sigma_j} \tag{12}$$

Then a pixel in the foreground that is obtained by the GMM above is

$$\begin{cases} shadow & CD_j < \gamma_1 \;\; and \;\; B_j < 1 \\ Highlight & CD_j < \gamma_1 \;\; and \;\; B_j > \gamma_2 \end{cases} \tag{13}$$

γ_1 is a selected threshold value used to determine the similarities of the chromaticity between the learnt background by GMM and the current observed image. A dark pixel from a moving object can be misclassified as a shadow because the value of the dark pixel is close to the origin in RGB space. Since all chromaticity lines in RGB space meet at the origin, a colour point is considered to be close or similar to any chromaticity line. To avoid this problem we introduce a threshold γ_2 for the normalized brightness distortion. $\gamma_2 = 1/\left(1 - \varepsilon\right)$, where ε is a lower band for the normalized brightness distortion. An automatic threshold selection method was provided in [33].

3. Pixel based performance evaluation matrix

Once a ground truth has been established, there are several standard methods for comparing the ground truth to a candidate binary foreground map. Generally, it is useful to identify the following four important quality measures for pixel-based object detection.

True positive (TP): the pixels belong to the intruding objects are correctly assigned to the foreground.

False positive (FP): the background pixels are incorrectly detected as the foreground.

False negative (FN): the pixels belong to the intruding objects are incorrectly classified as the background.

True negative (TN): the background pixels are correctly detected as the background.

Five performance metrics are used to compare the two algorithms:

$$DR = \frac{TP}{TP + FN} \tag{14}$$

$$FAR = \frac{FP}{FP + TN} \tag{15}$$

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{16}$$

$$JC = \frac{TP}{TP + FP + FN} \tag{17}$$

$$MCC = \frac{TP \times TN - FP \times FN}{(TP + FP)(TP + FN)(TN + FP)(TN + FN)} \tag{18}$$

Detection rate (DR), false alarm rate (FAR) and accuracy (ACC) are used to qualify how well each algorithm matches the ground truth. However, ACC tends to give misleading estimates when the amount of change is small compared to the overall image. The Jaccard coefficient (JC) overcomes this problem to some degree by minimizing or eliminating the effect of the expected large volume of true negatives [35]. While there is no perfect way of describing the confusion matrix of true and false, positives and negatives by a single number, the Matthews correlation coefficient (MCC) is generally regarded as being one of the best such measures. The MCC is in essence a correlation coefficient between the observed and predicted binary classifications; it returns a value between −1 and +1. A coefficient of +1 represents a perfect prediction, 0 an average random prediction and −1 an inverse prediction.

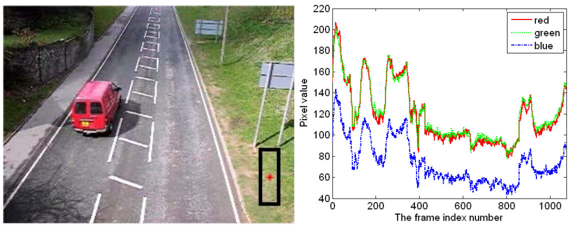
4. Quantitative evaluation

Extensive experiments have been performed to compare the results of SAGMM with ZHGMM. For the first experiment, the input data is a traffic video captured under intermittent cloudy conditions, where the clouds move in front of the sun. A total of 1079 frames are analysed from this video. A sample RGB frame is shown in Figure 1 (a). The variation of red, green and blue values of a particular pixel (red asterisk in the centre of black rectangle) in an area where no intruding object appears over time is shown in Figure 1 (b). It can be seen that the range of intensity variation is very large. The standard variation of R, G and B is 28.3, 27.4 and 21.2, respectively, over a period of 42.9 seconds. The ground-truth data was created by Viper [36]. The receiver operating characteristic (ROC) is used to analyse the model's performance. For a GMM, any input distribution can be converted to ROC curves. One may combine multiple ROC plots for different values of some of the fixed parameters. In our case, the two parameters of most significant interest are the threshold MD and the learning rate α. For a typical fixed α=0.001, if $c_f = 0.1$, we know that the moving object should be static for at most $\log(1-c_f)/\log(1-\alpha) \approx 105$ frames before it will merge into the background. As the MD increases, DR and FAR decrease accordingly. However, we are interested in a higher DR and lower FAR. MCC can be used to explicitly select the probability tradeoff. The variation of MCC corresponding to a different threshold for MD is given in Figure 2 (a). It shows that the SAGMM is much better than ZHGMM for all thresholds MD ∈ [1, 40], (the increment of MD is 2). The maximum MCC of SAGMM is 0.42, but the maximum MCC is 0.38. Each point of MCC here is the mean of each threshold across the entire video (of which only 484 frames include the moving vehicle). Obviously, the best cutting point threshold is MD=7. Under the optimal threshold of MD, the ROC curve is shown in Figure 2(b). We are interested in the lower rate of FAR, for instance, FAR<0.021, illustrated by the green dashed vertical line in figure 2(b), the DR of SAGMM is much higher than that of ZHGMM. At the point of FAR = 0.021, the DR of SAGMM is 0.7301, but the DR of ZHGMM is only 0.5683. The comparison of ACC and JC values in Figure 3 also shows the advantages of our algorithm. In order to get stable and reasonable performance metrics, the results during the learning stage weren't included in the performance metric calculations above.

To get a sense of how well our background model, estimated by SAGMM, matches the original input image, we compare the similarity between the predicted background and the original input frame. We use the mean of the Euclidean distance as the similarity measure. A test area is selected from a rectangular area that is placed in a part of the scene where the background is always visible, i.e. no intruding object appears (the black rectangular area in figure 1(a)). The similarity is defined as:

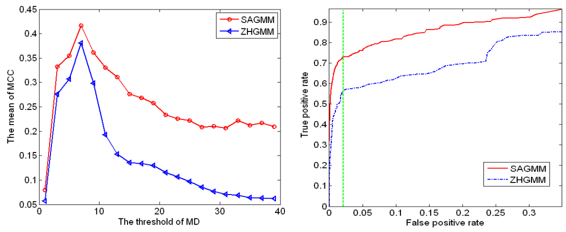
$$S = \frac{\sum_{j \in A} \left(\sum_{p \in [R,G,B]} \left(I_{j,p}^{cr} - I_{j,p}^{bg} \right)^2 / 3 \right)^{1/2}}{A_\Delta} \quad (19)$$

Where, S is the average colour components difference over the rectangular area A; A_Δ is the area of A in pixels. $I_{j,p}^{cr}$ and $I_{j,p}^{bg}$ represent the jth pixel value in R, G or B component in the area A in the current input image and current modeled background, respectively. The box plot figure of the similarity distance is given in Figure 4. The red solid line is the median of the results. The mean is 12.25 pixels and the std (standard variation) is 5.22 pixels. The blue dash line is the result of ZHGMM. The mean is 18.52, and the std is 8.13. It shows that the proposed method has a good background prediction.



(a) (b)

Figure 1: A sample image (a) and the variation of red, green, blue of a pixel (red asterisk in the center of black rectangle) value over time (b). The black rectangle is used to measure the similarity between the modeled background and the original input image.



(a) (b)

Figure 2: The variation of MCC corresponds to different threshold of MD (a), ROC curve (b).

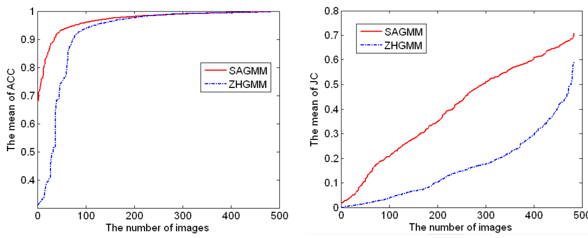


Figure 3: ACC (left) and JC (right).

Two samples of background subtraction image results are given in Figure 5. The first column shows the original images. The second column shows the results from ZHGMM, and the third column are the results from SAGMM. Because of sudden global illumination change, it can be seen from the first row that the ZHGMM performs poorly, but SAGMM can detects most of foreground.

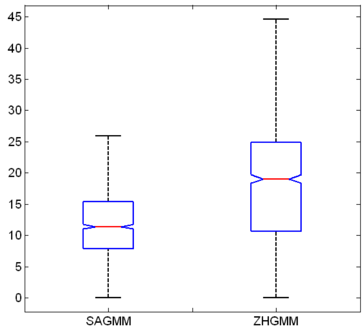


Figure 4: The comparison of similarity measurements between modeled background and original input image.

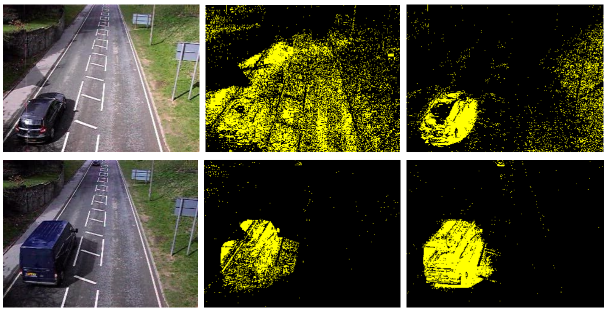
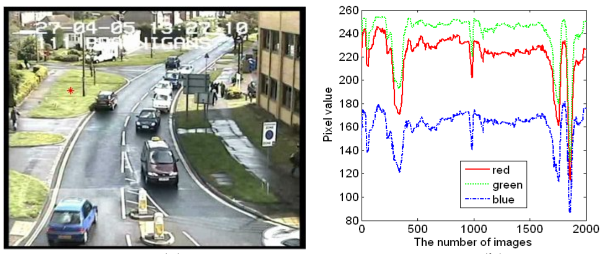


Figure 5: Example frames (#151 and #189). The first column is original images. The second column is the results from ZHGMM. The third column is the results from SAGMM.



(a) (b)

Figure 6: A sample image of i-LIDS data (a) and the variation of red, green and blue of a pixel (red asterisk) value over time (b).

The second experiment was performed using part of the video from i-LIDS (image library of intelligent detection systems) dataset which is provided by the Home Office of the United Kingdom. A sample image and the variation of a pixel (red asterisk) values are shown in Figure 6. This shows that the variation of intensity is smaller than in the previous sample video, but it is still large. There are 2000 frames included in the video. 494 randomly selected frames were annotated using Viper to create ground-truth data. Parameters α =0.001, MD=5 were used for this experiment. The ROC and JC curve are shown in Figure 7. A sample of image results is given in Figure 8.

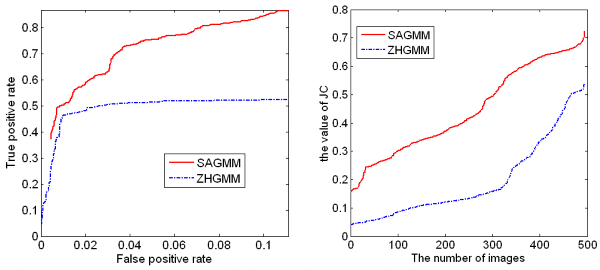
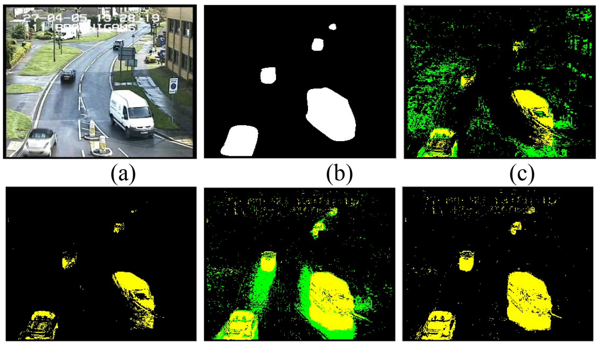


Figure 7: ROC curve (left) and JC variation (right) of i-LIDS data set.



(a) (b) (c)

(d) (e) (f)

Figure 8: Example frame. (a) Original image. (b) Annotated ground truth. (c) and (d) The background subtraction result with/without shadow from ZHGMM. (e) and (f) The background subtraction result with/without shadow from SAGMM. The yellow pixels are foreground and the green pixels are shadow. Black pixels are detected background.

5. Conclusions and future work

We proposed a new online, self-adaptive Gaussian mixture model and improved the shadow removal algorithm to deal with the sudden illumination changes. The algorithm has a dynamically adaptive learning rate and models global illumination changes of the background frame by frame. At the cost of only one additional parameter per Gaussian, this modification dramatically improves the convergence and the accuracy of background subtraction whilst maintaining the same temporal adaptability. This is achieved by incorporating a modified adaptive schedule (a counter keeping tracking the number of pixels that have contributed to a Gaussian component in temporal space) into a recursive filter. Comprehensive analysis of experimental results through statistical evaluation performance metrics obtained using annotated ground truth from traffic videos shows that our method leads to a significantly better performance than others, in situations subject to sudden illumination changes.

We believe that no perfect system exists. Background modeling and subtraction in itself is application oriented. Some of these problems cannot be solved simultaneously because of the differing needs associated with the semantic interpretation of the moving foreground and background. In this paper we have only discussed the basic pixel-level processes and frame-level global illumination changes, excluding any additional pre- or post processing operations, for example, checking for spatial and temporal correlation, ghost removal, etc. Furthermore, we intend to evaluate the performance over a wider range of condition, including sequences captured during the evening and at night, and under heavy rain conditions. We also intend to combine low level and high level processes in order to identify vehicles that stop for long periods, e.g. in front of traffic lights.

6. Acknowledgements

We are grateful for use of the i-LIDS dataset provided by the UK Home Office which were used for evaluation and comparison.

References

[1] M. Piccardi. Background subtraction techniques: A review. In IEEE International Conference on System , Man, and Cybernetics, 3099-3104, 2004.

[2] R.J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. IEEE transaction on Image Processing. 14(3): 294-307, 2005.

[3] T. Bouwmans, F.EI Baf, B. Vachon. Background modeling using mixture of Gaussian for foreground detection: A survey. Recent Patents on Computer Science, 1(3):219-237, 2008.

[4] S.Y. Elhabian, K.M. El-Sayed and A.H. Ahmed. Moving object detection in spatial domain using background removal techniques-state-of-art. Recent patents on computer science, 1(1):32-54, 2008.

[5] C.R., Wren, A. Azarbayejani, T. Darrell, A. Pentland. Pfinder: Real-time tracking of the human body. IEEE Transaction on Pattern Analysis and Machine Intelligence, 19(7): 780-785, July, 1997.

[6] C. Stauffer, W. Grimson. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 246-252, 1999.

[7] C. Stauffer, W. Grimson. Learning patterns of activity using real-time tracking. IEEE Transaction on Pattern Analysis and Machine Intelligence, 22(8): 747-757, 2000.

[8] P.W. Power, J.A. Schoonees. Understanding background mixture models for foreground segmentation. Proceedings of Image and Vision Computing, New Zealand, November, 2002.

[9] P. KaewTraKulPong and R. Bowden. An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection. Proc. of 2nd European workshop on Advanced Video Based Surveillance Systems, chapter 11, 135-144, Sept. 2001.

[10] D-S. Lee. effective Gaussian mixture learning for video background subtraction. IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(5): 827-832, 2005.

[11] Sen-Ching S. Cheung and Chandrika Kamath. Robust techniques for background subtraction in urban traffic video. Visual Communication and Image Processing, Vol. 5308, No. 1. 881-892, 2004.

[12] M. Haque, M. Murshed and M. Paul. Improved Gaussian mixtures for robust object detection by adaptive multi-background generation. In Proceedings of 19th International Conference on Pattern Recognition, 1-4, 2008.

[13] Z. Zivkovic and F. van der Heijden. Recursive unsupervised learning of finite mixture models. IEEE Transaction on Pattern Analysis and Machine Intelligence, 26(5): 651-656, May, 2004

[14] Z. Zivkovic and F. van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letters, 27(7): 773-780, May 2006.

[15] N. Greggio, A. Bernardino, C. Laschi, P. Dario, J. Santos-Victor. Self-adaptive Gaussian mixture models for real-time video segmentation and background subtraction. 10th International Conference on Intelligent Systems Design and Applications (ISDA), pp.983-989, 2010.

[16] A.M. Elgammal, A. Harwood, L.S. Davis. Non-parametric model for background subtraction. Lecture Notes in Computer Science, vol. 1843, Proceedings of the 6th European Conference on Computer Vision-Part II, 751-767, 2000.

[17] B. Han, D. Comaniciu and L. Davis. Sequential kernel density approximation and its application to real-time visual tracking. IEEE Transaction on Pattern Analysis and Machine Intelligence, 30(7), 1186-1197, 2008.

[18] B. Zhong, S. Liu and H. Yao. Local spatial co-occurrence for background subtraction via adaptive binned kernel estimation. 9th Asian Conference on Computer Vision, vol. III, 152-161, 2009.

[19] N. Martel-Brisson and A. Zaccarin. Learning and removing cast shadows through a multidistribution approach. IEEE Transaction on Pattern Analysis and Machine Intelligence, 29(7):1133-1146, 2007.

[20] L.P.J. Vosters, C. Shan and T. Gritti. Background subtraction under illumination changes. In Proceedings of 7th International Conference on Advanced Video and Signal Based Surveillance, pp. 384-391, 2010.

[21] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. IEEE Transactions on Image processing 17(7):1168-1177, 2008.

[22] Y. Singh, P. Gupta and V.S. Yadav. Implementation of a self-organising approach to background subtraction for visual surveillance approach. International Journal of Computer Science and Network Security, 10(3):136-143, 2010.

[23] V. Reddy, C. Sanderson, A. Sanin, B.C. Lovell. Adaptive patch-based background modelling for improved foreground object segmentation and tracking. In Proceedings of 7th International Conference on Advanced Video and Signal Based Surveillance, pp. 172-179, 2010.

[24] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, S.Z. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scene. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1301-1306, 2010.

[25] J. Pilet, C. Strecha, P. Fua. Making background subtraction robust to sudden illumination changes. European Conference on Computer Vision, 2008.

[26] S.L. Zhao, H.J. Lee. A spatial-extended background model for moving blob extraction in indoor environments. Jounal of Information Science and Engineering, 25, 1819-1837, 2009.

[27] M. Izadi, P. Saeedi. Robust region-based background subtraction and shadow removing using color and gradient information. 19th International Conference on Pattern Recognition (ICPR), 1-5, 2008.

[28] O. Javed, K. Shafique, M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. Workshop on Motion and Video Computing, 22-27, 2002

[29] Y. Beiderman, E. Rivlin, M. Teicher and Z. Zalevsky. illumination insensitive reconstruction and pattern recognition using spectral manipulation and K-factor spatial transforming. Recent Patents on Signal Processing, 2, 22-27, 2010.

[30] B. Xie, V. Ramesh and T. Boult. Sudden illumination change detection using order consistency. Image and Vision Computing, 22: 117-125, 2004.

[31] P.J. Withagen, K. Schutte and F.C.A. Groen. Global intensity correction in dynamic scenes. International Journal of Computer Vision, 86: 33-47, 2010.

[32] T. Joachims. Training Linear SVMs in Linear Time. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06), 217-226, Philadelphia, Pennsylvania, USA, Aug. 2006.

[33] T. Horprasert, D. Harwood, L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. Proceedings of IEEE ICCV'99 Frame rate workshop, 1999.

[34] Z. Chen, N. Pears, Michael Freeman and Jim Austin. Background subtraction in video using recursive mixture models, spatio-temporal filtering and shadow removal. Proceedings of 5th International symposium on Visual Computing (ISVC), Las Vegas, NV, USA, Nov. 30-Dec. 2, 2009. Lecture Notes in Computer Science, Vol. 5876, 2009, pp. 1141-1150.

[35] P. Rosin, E. Ioannidis. Evaluation of global image thresholding for change detection. Pattern Recognition Letters, 24(14): 2345-2356, 2003.

[36] D. Doermann, and D. Mihalcik. Tools and techniques for video performances evaluation. ICPR, pages 167-170, 2000 (http://viper-toolkit.sourceforge.net/).