Fusing Complementary Operators to Enhance Foreground/Background Segmentation

Ahmad H. Al-Mazeed, Mark S. Nixon and Steve R. Gunn
Department of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ, UK
{aha01r,msn,srg}@ecs.soton.ac.uk

Abstract

Foreground/background segmentation is an active research area for moving object analysis. We combine two probabilistic approaches one of which estimates foreground/background probabilistic density and the other uses prior knowledge to decompose the colour space. The observed performance advantages are associated with the fusion of operators with completely different basis. Tests on outdoor and indoor sequences confirm the efficacy of this approach. The new algorithms can successfully identify and remove shadows and highlights with improved moving-object segmentation. A particular advantage of our evaluation is that it is the first approach that compares foreground/background labelling with results obtained from labelling by broadcast techniques.

1   Introduction

The detection of moving objects is an essential part of information extraction in many computer vision applications including: surveillance [9]; traffic control [8, 5] and video coding [10]. Background differencing is a well established basis for moving object extraction. An implementation of background differencing involves producing a reference image (a background model), which is subtracted from each new frame to give an image containing only the moving objects.

In more refined approaches, statistical methods were used to form the background model. Horprasert et. al. [3] introduced a new computational colour model which separates the brightness from the chromaticity component. The algorithm can detect moving objects and can distinguish shadows from a background. Pfinder [11, 4] uses a multiscale statistical model of colour and shape with single Gaussian per pixel to model the background. It succeeded in finding a 2-D representation of head, hands and feet locations of a moving human subject. In contrast, Friedman and Russell [2] took a simpler approach to modelling the statistical nature of the image by using a single distribution to model the whole of the background and two other distributions to model the variability in shadows and moving objects.

In real situations the background is typically multi-modal. A single Gaussian would suffice to approximate the background if each pixel resulted from a single surface under

fixed lighting. Often multiple surfaces appear on a particular background pixel and the lighting conditions change. Therefore, to robustly model a multi-modal background, a multiple adaptive Gaussians can be used. In addition, a mixture of Gaussians model is a very appealing approach to data fitting as it scales favourably with dimensionality of the data, has good analytic properties and many data sets form clusters which are approximately Gaussian in nature [6]. Stauffer and Grimson [9] presented an online algorithm based on a statistical method using a mixture of Gaussians. The persistence and the variance of each of the Gaussians is used to identify background distributions. The approach was designed to deal robustly with bimodal backgrounds, lighting changes, repetitive motions of scene elements. The method lacks the capability to remove shadows and highlights.

Elgammal et al. [1] used a Gaussian density estimator as a kernel in the process of background modelling. The final background model is updated by combing a short and a long term model of the background. A colour representation was used to suppress shadows.

Such techniques form a good base for building a better approach. Skillful fusion of such methods by holding to the strong points and removing the weaknesses can eventually result in a better technique. In the following sections we describe two standard motion extraction approaches based on mixture of Gaussians [9] and another based on statistical properties of the colour model [3]. These are combined in section (3) and we show how moving object segmentation can be improved in section (4). Further comparison of outdoor vs. indoor in section (4) confirms the efficacy for this approach, prior to suggestions for future avenues for research.

2 Motion Extraction

2.1 Mixture of Gaussians Algorithm

This approach models the background with independent distributions that are updated by an efficient on-line method. The recent history of each pixel is modelled as a mixture of K Gaussian distributions. The probability of a pixel intensity, $\mathbf{x} = (x_R, x_G, x_B)$,

$$P(\mathbf{x}) = \sum_{j=1}^{K} w_j \eta(\mathbf{x}, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \qquad (1)$$

where K is the number of distributions (from 3 to 5), w_j is the weight estimate for the Kth distribution, $\boldsymbol{\mu}_j$ is the mean value for the Kth distribution, and $\boldsymbol{\Sigma}_j$ is the covariance matrix for the Kth distribution. η is a Gaussian probability density function formed from the multivariate Gaussian

$$\eta(\mathbf{x}, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_j|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j)\right) \qquad (2)$$

where d is the input dimension which is 3 for the (RGB) colour model and $\boldsymbol{\Sigma}_j$ is assumed to be of the form $\boldsymbol{\Sigma}_j = \sigma_j^2 \mathbf{I}$. Every new pixel value, $\mathbf{x}$, is compared to the existing K Gaussian distributions. The pixel is classified to be in a particular distribution if the pixel is within 2.5 times the standard deviation of the distribution. The pixel is checked against the background distributions first and then to the foreground distributions.

The distributions are ordered according to the ratio of the weight over the standard deviation of each distribution, w_j/σ_j. This process will rank the most probable (those with high weight and low variance) to the least probable background distributions (those with low weight and high variance). The background model is formed from a number of background distributions

$$B = \arg\min_b \left(\sum_{j=1}^b w_j > T \right) \tag{3}$$

where $T \in [0,1]$ controls the number of modes of variations in the background. If a pixel does not match any of the K distributions, the pixel will be considered as a new distribution replacing the distribution with the smallest w_j/σ_j. The new distribution mean, $\boldsymbol{\mu}_{j,t}$, will be the pixel value. The prior weight of the new distribution will be set to a low weight and the variance to a high variance. After evaluating a new pixel, the K distributions prior weights are updated at time t

$$w_{j,t} = (1-\alpha)w_{j,t-1} + \alpha(M_{j,t}) \tag{4}$$

where α is the learning rate. $M_{j,t}$ is 1 for the matching distribution, and 0 for the remaining distributions. The weights are normalised after this process. The value of $\boldsymbol{\mu}_{j,t}$ and $\sigma_{j,t}^2$ are updated only for the matching distribution

$$\boldsymbol{\mu}_{j,t} = (1-\rho)\boldsymbol{\mu}_{j,t-1} + \rho \mathbf{x}_t \tag{5}$$

$$\sigma_{j,t}^2 = (1-\rho)\sigma_{j,t-1}^2 + \rho(\mathbf{x}_t - \boldsymbol{\mu}_{j,t})^T(\mathbf{x}_t - \boldsymbol{\mu}_{j,t}) \tag{6}$$

where

$$\rho = \alpha\eta(\mathbf{x}_t, \boldsymbol{\mu}_{j,t-1}, \boldsymbol{\Sigma}_{j,t-1}) \tag{7}$$

In this algorithm if a non-background pixel (part of a moving object) does not move over a period of time, its distribution weight over time will increase and its variance will decrease until this pixel becomes part of the background model.

2.2   Statistical Background Disturbance Technique

This algorithm decomposes the colour space using prior knowledge established on a statistical computational model to separate the chromaticity from the brightness component. The algorithm initially uses N frames to form the background model. From these frames, the mean and the variance is computed for each colour band (RGB) in each pixel. The chrominance distortion CD and the brightness distortion β between a background pixel and a new pixel $\mathbf{x}$ are computed as

$$CD = \sqrt{\sum_{c \in \{R,G,B\}} \left(\frac{x_c - \beta\mu_c}{\sigma_c} \right)^2} \tag{8}$$

and

$$\beta = \frac{\sum_{c \in \{R,G,B\}} \left(\frac{x_c \mu_c}{\sigma_c^2} \right)}{\sum_{c \in \{R,G,B\}} \left(\frac{\mu_c}{\sigma_c} \right)^2} \tag{9}$$

where μ_c and σ_c now respectively represent the mean and the standard deviation for each background pixel colour band. The normalised chrominance distortion $\widehat{CD}$ and the brightness distortion $\widehat{\beta}$ are used to classify the new pixel

$$M(i) = \begin{cases} Foreground & : & \widehat{CD}_i > \tau_{CD} & or & \widehat{\beta}_i < \tau_{\beta lo}, & else \\ Background & : & \widehat{\beta}_i < \tau_{\beta 1} & and & \widehat{\beta}_i > \tau_{\beta 2}, & else \\ Shadow & : & \widehat{\beta}_i < 0 & else \\ Highlight & : & otherwise \end{cases} \tag{10}$$

where τ_{CD} and $\tau_{\beta lo}$ are thresholds used to specify the borders of the foreground. $\tau_{\beta 1}$ and $\tau_{\beta 2}$ are thresholds used to identify the borders of the background. These thresholds are determined automatically through a statistical learning procedure [3]. Through the background building process a histogram is constructed for $\widehat{CD}$ and $\widehat{\beta}$. The thresholds are then computed after fixing a detection rate which fixes the expected proportions of the image contents.

3 Fusing Complementary Techniques

Here we show not only how one of the techniques can be extended, but also how the two techniques can be fused in two different ways.

3.1 Extending the Mixture of Gaussians to Remove Shadows and Highlights

Our first improvement is to extend the mixture of Gaussians approach to remove highlights and shadows. As such, the highlight and shadow selection parameters of the statistical background disturbance technique were deployed to ensure that these factors were removed. This led to some improvement.

3.2 The Enhanced Foreground/Background Selector

The extended mixture of Gaussians algorithm is joined with the statistical background disturbance technique by ANDing the results of both operators. A pixel is labelled as foreground only if both techniques labelled it as such independently, otherwise it is labelled as background. The process gave better results than either of the original algorithms on the data sets considered.

3.3 Dynamic Gaussian Background Distributions

The mixture of Gaussians has more attractive properties than the statistical background disturbance technique. The mixture of Gaussians updates the background model during

the extraction process repeatedly using the online input sequence. In contrast, the statistical background disturbance technique uses a fixed background model. This led us to combine the two techniques to predispose processing to favour the more robust technique. Hence, the mixture of Gaussians forms the basis of the model, and the statistical background disturbance technique is used to force the mixture of Gaussians to expand the background distributions only when a pixel is labelled as a motion pixel and the background disturbance technique disagrees.

The distribution size is an essential parameter for the quality of extraction in the mixture of Gaussians technique. Using a large size for the background distributions will accommodate more background pixels but at the same time will affect the accuracy of identifying the moving object pixels, especially if both (the background pixels and the moving object pixels) are similar in colour space. In contrast, using small background distribution sizes will maintain the accuracy of the extracted moving object but will not have the ability to contain all the background pixels. In this algorithm we present a better way to satisfy the feature of accommodating the background pixels as well as preserving the accuracy of moving pixel extraction quality.

The algorithm first specifies two distribution sizes as a new compromise model: one fits most of the background pixels without affecting the detection of a moving object. The other is larger, to approximate the remaining background pixels. Then the extended mixture of Gaussians algorithm is used with the small distribution size as an initial distribution variance for each new distribution. The pixels identified as moving objects are tested using the statistical background disturbance technique. The testing algorithm will determine whether it agrees with the labelled motion pixels, if not the background distributions will be enlarged (using the larger background distribution size) trying to fit the pixels.

To find the proper alternating sizes for a background distribution we propose the following:

- **The small distribution size:**
  The smaller distribution size is determined by constructing a histogram for all the image pixels of the variance within each pixel. The histogram is built from the combined data through a long background sequence. The small distribution size is then obtained from the histogram by thresholding it at a chosen percentage. The threshold is chosen while considering the quality of motion extraction and the background noise (the threshold was set between $85\% - 95\%$ in our test sequences).

- **The large distribution size:**
  In the mixture of Gaussians algorithm, the decision of labelling a pixel as a background or a foreground distribution depends on whether the pixel is within 2.5 times the standard deviation of the distribution [9]. To find the large distribution size an incremental process is performed on the variance of each pixel distribution to obtain a minimum error for the extracted subject. A constant, γ, is incremented in the fitting criteria 2.5 ($\gamma \times \sigma_{BG,i}$), where γ is larger than or equals 1. $\sigma_{BG,i}$ is the standard deviation for background distribution i. Whenever the constant is incremented it is tested on M input frames and the mean of the total error for the tested sequence is calculated. The incrementing procedure will continue (and the error will decrease) until the error level reaches a steady state. Figure (1) shows the effect of increasing the distribution size on the extracted subjects with less erroneous background

(a) Moving Human Subject
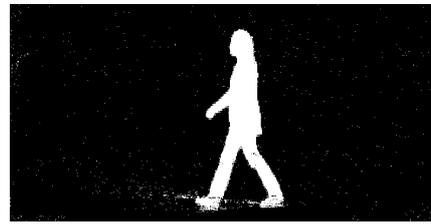
(b) $\gamma = 1.0$
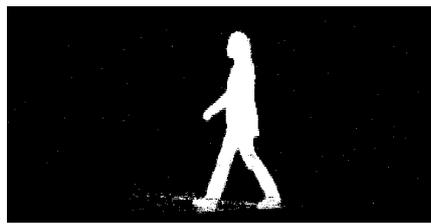
(c) $\gamma = 2.4$
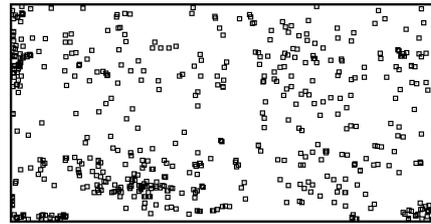
(d) The difference image with the removed pixels highlighted

Figure 1: The effect of increasing the distribution size on the extracted subjects in the dynamic Gaussian background distributions algorithm

points for $\gamma = 2.4$ with differences highlighted for visibility. Figure (2) shows the error rate for the same tested sequence, where the error decreases as γ is increased.
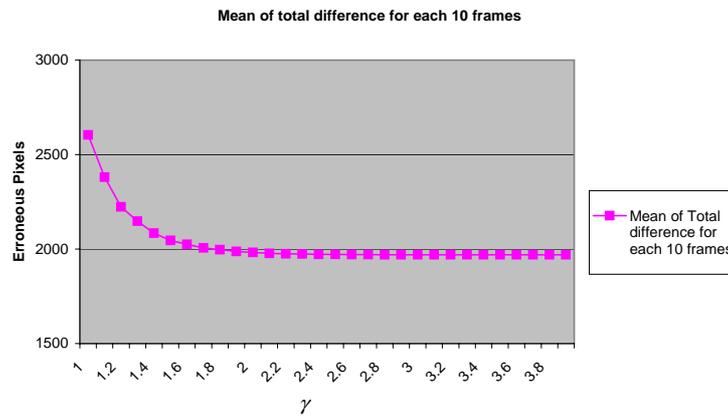

Figure 2: The mean error for a tested sequence

4   Experimental Results

The presented algorithms were tested on indoor and outdoor sequences of walking human subjects. The quality of the resulting extracted moving objects is far better than the mixture of Gaussians and the statistical background disturbance techniques. In testing the algorithms we used outdoor sequences of size 220×190 pixels and indoor sequences of size 720×367 pixels . The resulting extracted motion frames for the indoor sequences are compared with the silhouettes provided by the University of Southampton [7]. The silhouettes were generated by chroma-key extraction of the green background. The total difference between the algorithm's extractions and the silhouettes is calculated for each image as a count of the mismatching pixels. This facilitates the comparison of the extraction procedure with a form of ground truth (derived by a more conventional broadcast technology).

The extended algorithm and the enhanced background foreground selector were tested on eight indoor sequences. All the extended mixture of Gaussian results were better than both the mixture of Gaussians and the statistical background disturbance techniques. The enhanced background foreground selector was tested on the same input sequences. The results of the mean error for all the tested sequences is shown below in Figure (3). Samples of extracted indoor sequences is shown in Figure (4). Similar results were obtained when the algorithms were tested on outdoor sequences.
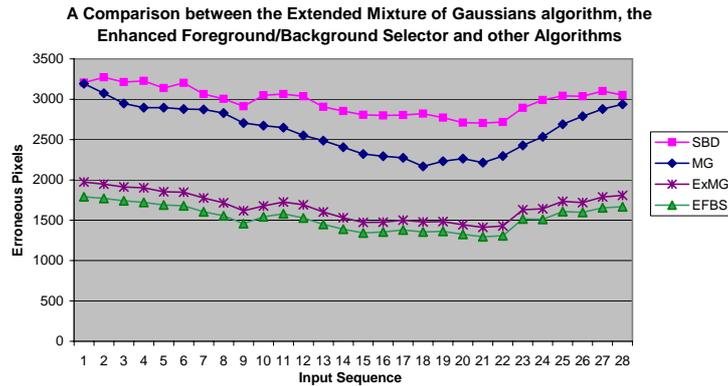

Figure 3: A comparison between the mixture of Gaussians (MG), the statistical background disturbance technique (SBD), the extended mixture of Gaussians (ExMG) and the enhanced foreground/background selector (EFBS).

The dynamic background distributions algorithm was tested on indoor and outdoor sequences as well. The performance of this algorithm appears better than the mixture of Gaussians algorithm and the background disturbance algorithm. Figure (5) shows a comparison. A moving extracted subject is shown in Figure (6) for an outdoor frame to compare the used algorithms with the resulting extraction. Isolated noise points are removed from the images using connected components algorithm. The extracted frames emphasise the advantage of our method where the use of relatively small size of distribu-

(a) The input frame

(b) Mixture of Gaussians

(c) Statistical Background Disturbance

(d) Extended Mixture of Gaussians

(e) EFBS

Figure 4: Indoor images extracted using the extended mixture of Gaussians, the enhanced foreground/background selector (EFBS) compared with other different algorithms.

tions (6 b) did not accommodate all the background pixels. While larger size distributions (6 c) resulted in affecting some of the motion pixels and were labelled as background. The dynamic background distribution technique in (6 e), had successfully accommodated most of the background pixels and at the same time preserved the quality of motion pixel extraction.

5   Conclusions

This paper presented two new motion extraction algorithms, the dynamic background distribution algorithm and the enhanced foreground/background selector. These are achieved by logical combination of two popular foreground/background motion extraction algorithms and by an adaptive combination of the same two approaches. A comparison between new algorithms and their original versions was prepared using controlled laboratory data, and outside data. The dynamic background distribution algorithm and the enhanced foreground/background selector gave the best performance among the other tested algo-

A Comparison between the Dynamic Distributions algorithm and three other extraction algorithms
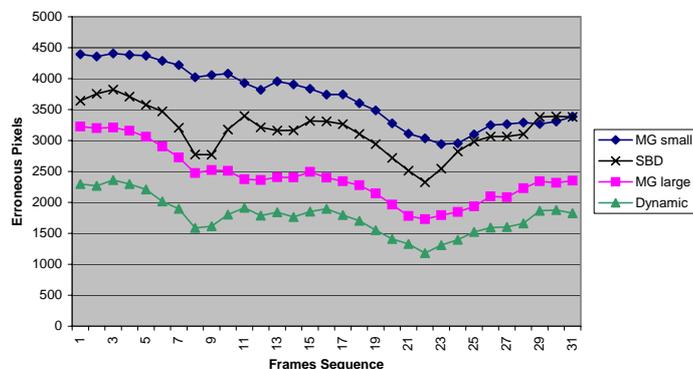
Figure 5: A comparison between the dynamic background distribution, mixture of Gaussians with two different distribution sizes (MG small and MG large) and the statistical background disturbance technique (SBD).

rithms and had improved the capability to select correct pixels as shown by comparison with ground truth labelling.

Acknowledgment

References

[1] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis. Background and foreground modeling using non-parametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163, 2002.

[2] N. Friedman and S. Russell. Image segmentation in video sequences: a probabilistic approach. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 175–181, 1997.

[3] T. Horprasert, D. Harwood, and L. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *In Proceedings IEEE ICCV'99*, pages 1–19, 1999.

[4] P. Maes, T. Darrell, B. Blumberg, and A. Pentland. The alive system: wireless, full-body interaction with autonomous agents. *ACM Multimedia Systems*, 5:105–112, 1997.

(a) Outdoor input frame

(b) Mixture of Gaussians - small distributions

(c) Mixture of Gaussians - large distributions

(d) The Statistical Background Disturbance technique

(e) Dynamic Background Distributions technique
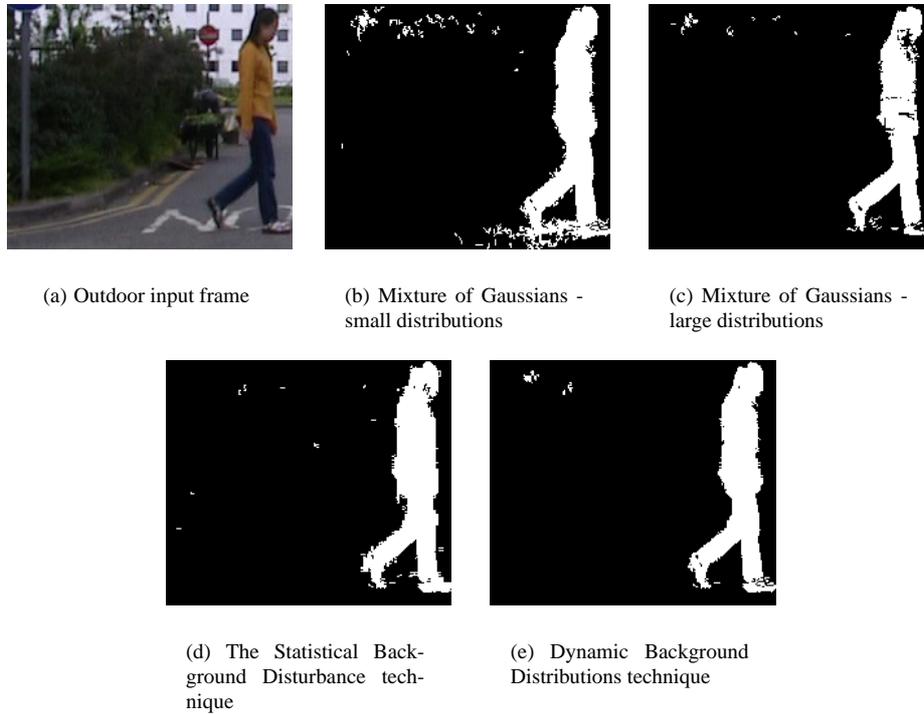
Figure 6: Outdoor images extracted using the dynamic background distributions algorithm compared with other algorithms.

[5] P. Michalopoulos. Vehicle detection video through image processing: the autoscope system. *IEEE Transactions on Vehicular Technology*, 40(1):21–29, 1991.

[6] S. Roberts, D. Husmeier, I. Rezek, and W. Penny. Bayesian approaches to Gaussian mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1133–1142, 1998.

[7] J. Shutler, M. Grant, M. Nixon, and J. Carter. On a large sequence-based human gait database. In *In Proc. of RASC 2002*, pages 66–71, 2002.

[8] C. Smith, C. Richards, S. Brandt, and N. Papanikolopoulos. Visual tracking for intelligent vehicle-highway systems. *IEEE Transactions for Vehicular Technology*, 45(4):744–759, 1996.

[9] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE TPAMI*, 22(8):747–757, 2000.

[10] C. Su, H. Hang, and D. Lin. Global motion parameter extraction and deformable block motion estimation. *IEICE Trans. Inf. and Syst.*, E82-D(8):1210–1218, 1999.

[11] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE TPAMI*, 19(7):780–785, 1997.