

# Classifiers Combination for Improved Motion Segmentation

Ahmad Al-Mazeed, Mark Nixon, and Steve Gunn

University of Southampton, Southampton, SO17 1BJ, UK  
{aha01r,msn,srg}@ecs.soton.ac.uk

**Abstract.** Multiple classifiers have shown capability to improve performance in pattern recognition. This process can improve the overall accuracy of the system by using an optimal decision criteria. In this paper we propose an approach using a weighted benevolent fusion strategy to combine two state of the art pixel based motion classifiers. Tests on outdoor and indoor sequences confirm the efficacy of this approach. The new algorithm can successfully identify and remove shadows and highlights with improved moving-object segmentation. A process to optimise shadow removal is introduced to remove shadows and distinguish them from motion pixels. A particular advantage of our evaluation is that it is the first approach that compares foreground/background labelling with results obtained from ground truth labelling.

## 1 Introduction

The objective of achieving the best performing pattern recognition classifiers leads to different designs of high performance algorithms. Classifiers differ in their classification decision suggesting that different classifiers designs potentially offer complementary information about the patterns to be classified which can be harnessed to improve performance of the selected classifier [1].

In this paper two motion classifiers are combined using Bayes theorem while considering the confidence of each classifier to optimise the motion classification process.

## 2 Motion Detection

The detection of moving objects is an essential part of information extraction in many computer vision applications including: surveillance and video coding. Background differencing is a well established basis for moving object extraction. In more refined approaches, statistical methods were used to form the background model. Horprasert et. al. [2] 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 [3] uses a multiscale statistical model of colour and shape with a 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 [4] 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. Elgammal et al. [5] used a Gaussian density estimator as a kernel in the process of background modelling. The final background model is updated by combining a short and a long term model of the background.

Often multiple surfaces appear on a particular background pixel and the lighting conditions change. Therefore, to robustly model a multi-modal background, the 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 [7] 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 multimodal backgrounds, lighting changes, repetitive motions of scene elements. The method lacks the capability to remove shadows and highlights. This method was further extended using an EM algorithm in [8] to track motion and in [9] to track faces. The method was also used with image mosaicing techniques to build panoramic representations of the scene background [10]. Magee [11] used a projective ground-plane transform within the foreground model to strengthen object size and velocity consistency assumptions with the mixture of Gaussians background modelling method.

Such techniques form a good base for building a better approach. Skillful combination 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 pixel-based motion extraction approaches based on mixture of Gaussians [7] and another based on statistical properties of the colour model [2]. These are combined in Sect.(4). The segmentation analysis is given in Sect.(5). Further comparison of outdoor vs. indoor extraction in Sect.(5) confirms the efficacy for this approach, prior to suggestions for future avenues for research.

### 3 Motion Extraction

#### 3.1 Mixture of Gaussians Algorithm (MOG)

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

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

where  $K$  is the number of distributions,  $w_j$  is the weight estimate for the  $K^{th}$  distribution,  $\boldsymbol{\mu}_j$  is the mean value for the  $K^{th}$  distribution, and  $\boldsymbol{\Sigma}_j$  is the co-

variance matrix for the  $K^{th}$  distribution.  $\eta$  is a Gaussian probability density function formed from the multivariate Gaussian

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

where  $d$  is the input dimension which is 3 for the (RGB) colour model and  $\boldsymbol{\Sigma}_j$  is approximated by  $\boldsymbol{\Sigma}_j = \sigma_j^2 \mathbf{I}$ . 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,  $w_j/\sigma_j$ . 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

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

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  $w_j/\sigma_j$ . 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  $t$

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

where  $\alpha \in [0, 1]$  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

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

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

where

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

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.

### 3.2 Statistical Background Disturbance Technique (SBD)

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,  $\beta$ , between the background model and a new pixel,  $\mathbf{x}$ , are computed as

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

where  $\mu_c$  and  $\sigma_c$  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

$$M(i) = \begin{cases} FG : \widehat{CD}_i > \tau_{CD} \text{ or } \widehat{\beta}_i < \tau_{\beta lo}, \text{ else} \\ BG : \widehat{\beta}_i < \tau_{\beta 1} \quad \text{and } \widehat{\beta}_i > \tau_{\beta 2}, \text{ else} \\ S : \widehat{\beta}_i < 0 \quad \text{else} \\ H : \text{otherwise} \end{cases} \quad (9)$$

where  $FG$ ,  $BG$ ,  $S$  and  $H$  resembles: foreground, background, shadow and highlights respectively.  $\tau_{CD}$  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 [2]. 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.

## 4 Combining Motion Classifiers

The combination of the two classification algorithms evolved due to the performance of both algorithms in classification and the shadow extraction feature provided by the SBD algorithm. The fact that both algorithms operate using pixel wise operations facilitated the process of combination. The two classification algorithms are combined using Bayes theorem,

$$P(w|\mathbf{x}) = \frac{p(\mathbf{x}|w)P(w)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|w)P(w)}{\sum p(\mathbf{x}|w)P(w)}. \quad (10)$$

Whenever the classifiers agree on a certain decision (whether a pixel is a foreground pixel or a background pixel), the decision will be set to such decision. On the other hand, if the classifiers disagree then the conditional probability for

the chosen class by each classifier is calculated. The conditional probability for the Statistical Background Disturbance technique for a pixel  $\mathbf{x}$  being part of the background class is calculated as follows

$$p(\mathbf{x}|w_{BG}) = 1 - \frac{D}{Var} \quad (11)$$

where  $D$  is the distance between the tested pixel and the mean of the background distribution, and  $Var$  is the background variance. The Mixture of Gaussians algorithm provides the conditional probability for the background. The foreground conditional probability for the MOG algorithm Eqn. (12) is calculated from  $p(\mathbf{x}|w_{BG})$  for the closest background distribution for the pixel  $\mathbf{x}$ .

$$p(\mathbf{x}|w_{FG}) = 1 - p(\mathbf{x}|w_{BG}) \quad (12)$$

For the SBD algorithm the conditional probability for the foreground,  $p(\mathbf{x}|w_{BG})$ , is approximated by

$$p(\mathbf{x}|w_{FG}) = h\left(\frac{D}{Var} - 1\right) \quad \text{where} \quad h(y) = \begin{cases} y & \text{when } y \leq 1 \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

The decision is then made according to the following equation

$$\arg \max_{i \in \{1,2\}} p(\mathbf{x}|w_{CLSF_i})P(w_{CLSF_i}) \quad (14)$$

where  $w$  is a class of either a background (BG) or a foreground (FG) for the classifier  $CLSF_i$ . In Eqn. (14) the maximum conditional probability for each classifier is used with the classifier's confidence measure  $P(w_{CLSF_i})$  to find the final decision for the algorithm.  $P(w_{CLSF_i})$  satisfy a sum to unity condition

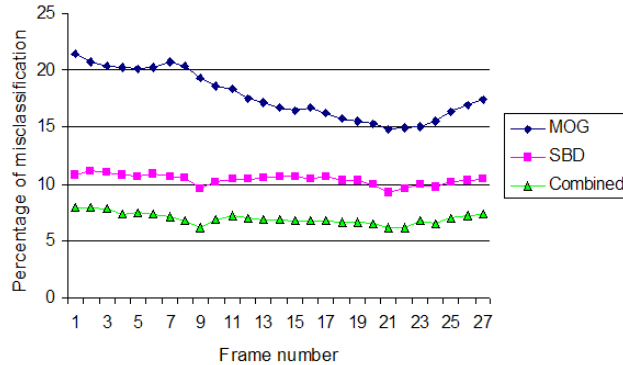
$$P(w_{FG,CLSF1}) + P(w_{BG,CLSF2}) = 1 \quad (15)$$

The priors  $P(w_{FG,CLSF1})$  and  $P(w_{BG,CLSF2})$  are calculated using a training set of  $N$  frames. In the training process, an exhaustive search method is performed by changing the weights incrementally between zero and one until an optimal value is reached giving the minimum classification errors.

The shadows are removed using the detection criteria in the Statistical Background Disturbance algorithm only, where there is no such feature in the Mixture of Gaussians algorithm. To optimise this process a threshold distance between the background mean and a virtual border for the shadow class is determined (using the same process used for the priors). Any shadow pixel with a distance exceeding the shadow border will be considered as a motion pixel.

## 5 Experimental Results

The presented algorithms were tested on indoor and outdoor sequences of walking human subjects. In testing the algorithms we used outdoor sequences of size  $220 \times 220$  pixels and indoor sequences of size  $720 \times 367$  pixels, with 77-81 images



**Fig. 1.** Comparing the tested algorithms.

per sequence. The resulting extracted motion frames for the indoor sequences are compared with the silhouettes provided by the University of Southampton database [12]. 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.

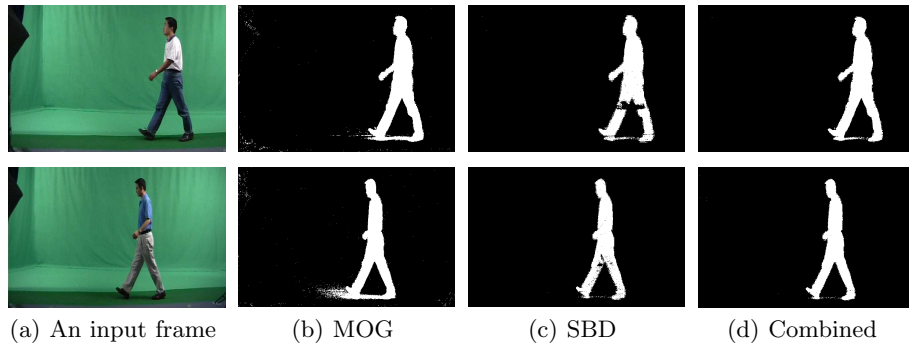
The Statistical Background Disturbance algorithm and the Mixture of Gaussians algorithm were trained initially with a background sequence of 50 frames. The MOG was used with 5 distributions per pixel. Each new distribution created was set to an initial weight of 0.05 and an initial variance equal to the largest variance of all the background pixels for the indoor sequences (double the background variance for the outdoor). The frames were tested with a background threshold ( $T$ ) of 0.4 (0.6 for outdoor) and a learning rate of 0.05.

The combined algorithm was tested on ten indoor sequences. All the combined algorithm results were better than both the Mixture of Gaussians (with 5 distributions) and the Statistical Background Disturbance techniques. The results of the mean error for the extraction of 10 indoor sequences is shown below in Fig.(1). The chart values are produced by finding the percentage of all misclassified pixels (comparing the current extraction with the silhouette).

The performance measure used to evaluate each method is

$$\begin{aligned}
 \text{Performance} &= 100 - \text{Misclassified Percentage} \\
 &= 100 - \left( \frac{\text{Misclassified Pixels}}{\text{White Silhouette Pixels}} \times 100 \right) \quad (16)
 \end{aligned}$$

By this performance measure, the combined algorithm shows an improvement of more than 10% over the MOG algorithm and more than 3% over the SBD



**Fig. 2.** Two examples of indoor images extracted with the tested algorithms.

algorithm. Table (1) shows the overall performance of the tested algorithm on all the indoor sequences.

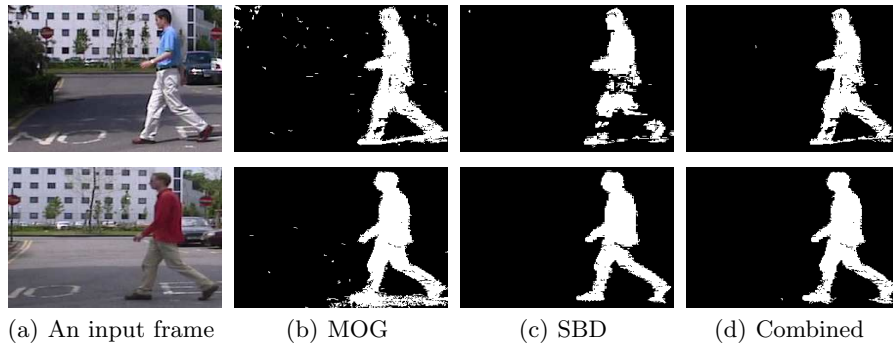
**Table 1.** Performance of the tested algorithms

Classifier	Misclassification (%)
MOG	17.54
SBD	10.44
Combined	6.96

Samples of extracted indoor sequences are shown in Fig.(2). The samples were chosen so as to show the performance advantage of the new algorithm over the MOG and the SBD algorithms. The output image for the Mixture of Gaussian algorithm gave a fine motion extraction but with noise on the background and shadows accompanying the moving object. Some of the output images produced by the SBD algorithm have holes in the moving object (usually holes vary in sizes). The extraction by the SBD (Fig 2.c) has misclassified small parts of the legs though with less shadow and a cleaner background. The best result is given by the output image for the combined algorithm with a clean background and fine extracted moving object (in some of the extracted sequence small parts of the shadow still persist). For outdoor sequences since the environment is more complex, it is possible to have more pixels mistakenly labelled in the combined algorithm. The combined algorithm can improve the outdoor motion extraction as shown in Fig.(3).

## 6 Conclusions

This paper presents a new motion extraction algorithm by combining two motion classifiers. A comparison between the new algorithm and their original



**Fig. 3.** Two examples of outdoor images using the three extraction algorithms.

versions was prepared using controlled laboratory data, and outside data. The combined algorithm shows that combination of pixel-based motion segmentation algorithms can improve segmentation performance. This suggests that applying more advanced ensemble methods could provide further performance improvement.

## References

1. Kittler, J., Hatef, M., Duin, R., Matas, J.: On combining classifiers. *IEEE TPAMI* **20** (1998) 226–239
2. Horprasert, T., Harwood, D., Davis, L.: A statistical approach for real-time robust background subtraction and shadow detection. In: *Proc. ICCV'99*. (1999) 1–19
3. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfunder: Real-time tracking of the human body. *IEEE TPAMI* **19** (1997) 780–785
4. Friedman, N., Russell, S.: Image segmentation in video sequences: a probabilistic approach. In: *Proc. UAI97*. (1997) 175–181
5. Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.S.: Background and foreground modeling using non-parametric kernel density estimation for visual surveillance. *Proceedings of the IEEE* **90** (2002) 1151–1163
6. Roberts, S., Husmeier, D., Rezek, I., Penny, W.: Bayesian approaches to Gaussian mixture modeling. *IEEE TPAMI* **20** (1998) 1133–1142
7. Stauffer, C., Grimson, W.: Learning patterns of activity using real-time tracking. *IEEE TPAMI* **22** (2000) 747–757
8. KaewTraKulPong, P., Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection. In: *Proc. AVBS'01*. (2001)
9. S.J. McKenna, Y.R., Gong, S.: Tracking colour objects using adaptive mixture models. *Image and Vision Computing* **17** (1999) 225–231
10. Mittal, A., Huttenlocher, D.: Scene modeling for wide area surveillance and image synthesis. In: *Proc. CVPR'2000*. Volume 2. (2000) 160–167
11. Magee, D.R.: Tracking multiple vehicles using foreground, background and motion models. *Image and Vision Computing* **22** (2004) 143–155
12. Shutler, J., Grant, M., Nixon, M., Carter, J.: On a large sequence-based human gait database. In: *Proc. of RASC 2002*. (2002) 66–71