Illumination-Invariant Motion Detection Using Colour Mixture Models[1]

Ming Xu and Tim Ellis
Department of Electrical, Electronic and Information Engineering
City University, London EC1V 0HB
{t.j.ellis, m.xu}@city.ac.uk

Abstract

This paper tackles the problem of robust change detection in image sequences from static cameras. Motion cues are detected using frame differencing with an adaptive background estimation modelled by a mixture of Gaussians. Illumination invariance and elimination or detection of shadows is achieved by using a colour chromaticity representation of the image data. The combination of the colour- and intensity-based models results in some promising applications.

1 Introduction

Frame differencing is a technique widely used for the change detection in dynamic images. It compares each incoming frame with a background image and classifies those pixels of significant variation into foreground. Therefore, the success of frame differencing depends on the robust extraction of the background image. The background can be modeled with a single adaptive Gaussian [8] and learnt during an initialization period when the scene is empty. This method is efficient only in less dynamic scenes but has difficulties with vacillating backgrounds (e.g. swaying trees), background elements moving, and fast illumination changes (flood of sunlight, shadows or lights switched on). A more robust method is to model the background by a mixture of adaptive Gaussians, each distribution of which is updated using the Expectation-Maximization (EM) algorithm [5] or a linear interpolation between the previous estimation and new observation [7]. This method can interpret vacillating backgrounds and background elements that are moving later on. However, it still cannot readily follow fast illumination changes, which cause spurious "foregrounds" and can lose targets in such cases. The reason is that the existing applications use only intensity-based image representations.

To robustly identify a particular object surface under varying illumination has received considerable attention in colour invariance research [1][3][6]. For example, a physics-based method has been proposed for shadow compensation in scenes illuminated by daylight [3]. The daylight is represented as a black body and the colour RGB camera filters are assumed to be of infinitely narrow bandwidth. $(R/B)/(G/B)^A$ is

[1] This work is supported by the EPSRC under grant number GR/M58030.

BMVC 2001 doi:10.5244/C.15.18

found only depending on surface reflection as the illumination changes (A can be pre-calculated from the daylight model and for the specific camera). Under the same assumptions, Finlayson *et al.* [1] found that the log-chromaticity differences (LCDs), $\ln(R/G)$ and $\ln(B/G)$, are independent of light intensity and there even exists a weighted combination of LCDs which is independent of both light intensity and light colour. There also exist adaptive schemes for colour-invariant detection of motion under varying illumination. Wren *et al.* [8] used the normalised components, U/Y and V/Y, of a YUV colour space to remove shadows in a relatively static indoor scene. A single adaptive Gaussian was used to model the background. Raja *et al.* [4] used the hue (H) and saturation (S) of an HSI colour space to obtain a limited level of intensity invariance in an indoor scene. A mixture of Gaussians was used to model a multi-coloured foreground object, in which each Gaussian models one colour in the object.

In this paper, a mixture of adaptive Gaussians is used to model the possibly multiple backgrounds at each pixel. The image representation used is the colour chromaticity, rgb, which is robust to fast illumination changes in an outdoor environment lit by sunlight and shadowed by cloud. A reflection and diffusion model in such a scene is presented in Section 2. As a result, the motion detection is insensitive to large-scale illumination changes. This algorithm differs from the existing applications of mixtures of Gaussians in modelling the intensities of multi-backgrounds [5][7] or the colour hues of a multi-coloured foreground [4].

2 Colour Fundamentals

An intensity image, $\mathbf{F} = (f_R, f_G, f_B)$, taken with a colour camera is composed of sensor responses as:

$$f_K = \int I(\lambda)\rho(\lambda)S_K(\lambda)d\lambda \quad (K = R,G,B) \tag{1}$$

where λ is wavelength, I is the illumination, ρ is the reflectance of an object surface, and S_K is the camera sensitivity. Given a particular colour camera, the image intensity depends only on the reflected light from the object surface:

$$I_{ref}(\lambda) = I(\lambda)\rho(\lambda) \tag{2}$$

Therefore, the appearance of objects in an image is a result of interaction between illumination and reflectance. Either the emergence of an object or illumination variation can cause the image intensity to vary. To be able to identify and track the same object surface (e.g. a background pixel) under varying illumination, it is desirable to separate the variation of the illumination from that of the surface reflection.

In an outdoor environment, fast illumination changes tend to occur at the regions where shadows emerge or disappear. These shadows may be either large-scale (e.g. those arising from a moving cloud) or small-scale (e.g. those arising from the objects themselves). Here a shadow model derived from that in [2] has been used and is shown in Fig. 1. There is only one illuminant in the scene. Some of the light does not reach the object because of a blocking object, thus creating a shadow region and a lit region on the observed object. The shadow region is not entirely dark but is illuminated by the reflection from each ambient object j:

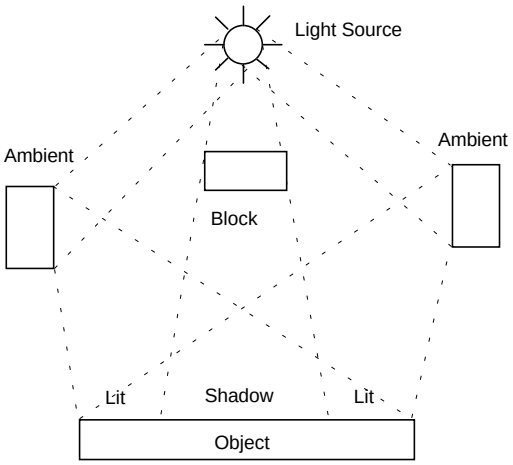


Figure 1: A reflection and shadow model.

$$I_{ambient,j}(\lambda) = I_{incident}(\lambda)\rho_{ambient,j}(\lambda) \tag{3}$$

(1) When the blocking object is opaque

For the lit region the reflected light from the object surface is:

$$\begin{aligned} I_{ref,lit}(\lambda) &= \left[I_{incident}(\lambda) + \sum_j I_{ambient,j}(\lambda) \right] \rho(\lambda) \\ &= I_{incident}(\lambda)\rho(\lambda)\left[1 + \sum_j \rho_{ambient,j}(\lambda) \right] \end{aligned} \tag{4}$$

For the shadow region this becomes:

$$I_{ref,shadow}(\lambda) = I_{incident}(\lambda)\rho(\lambda)\sum_j \rho_{ambient,j}(\lambda) \tag{5}$$

The condition that makes the reflected lights from the lit and shadow regions have the same spectral distribution is that:

Assumption 1: the chromatic average of the ambient objects in a scene is nearly grey, i.e. it is relatively balanced in all visible wavelengths and:

$$\sum_j \rho_{ambient,j}(\lambda) = \alpha \tag{6}$$

where α is independent of λ and may varies over space.

This assumption is realistic for the fast-moving cloud case, in which the only illuminant is the sunlight and both the blocking and ambient objects are grey (or white) clouds. Under such an assumption, the reflected light, $I_{ref}(\lambda)$, from shadow and lit regions will stay in proportion for a given object surface.

$$I_{ref,shadow} \big/ I_{ref,lit} = \alpha/(1+\alpha) \tag{7}$$

(2) When the blocking object is semi-opaque

The reflected light from the lit object surface is the same as that in Eq. (4). The shadow region is not only illuminated by the reflection from the ambient objects but also by the transmitted light from the blocking object:

$$I_{tra}(\lambda) = I_{incident}(\lambda)\tau(\lambda) \tag{8}$$

where τ is the transmittance of the blocking object.

For the shadow region the reflected light from the object surface is:

$$\begin{aligned} I_{ref,\,shadow} &= \left[I_{tra}(\lambda) + \sum_j I_{ambient,\,j}(\lambda) \right] \rho(\lambda) \\ &= I_{incident}(\lambda)\rho(\lambda) \left[\tau(\lambda) + \sum_j \rho_{ambient,\,j}(\lambda) \right] \end{aligned} \tag{9}$$

On the basis of assumption 1, the condition that makes the reflected lights from the lit and shadow regions have the same spectral distribution is that:

Assumption 2: the transmittance of the blocking object is relatively balanced in all visible wavelengths, i.e.:

$$\tau(\lambda) = \beta \tag{10}$$

where β is independent of λ ($\beta \leq 1$) and may vary over space. This assumption is realistic for clouds that do not favour any specific visible wavelength. Under the assumptions 1 and 2, the reflected light, $I_{ref}(\lambda)$, from shadow and lit regions will stay in proportion for a given object surface. We have:

$$I_{ref,\,shadow} \big/ I_{ref,\,lit} = (\beta + \alpha)/(1 + \alpha) \tag{11}$$

Suppose that the camera sensitivity, $S_K(\lambda)$ in Eq. (1), is properly characteristised so that no component of the colour image, **F**, is saturated. The relations shown in Eqs. (7) and (11) lead to the image intensities, f_K, at all colour channels being in proportion, no matter whether the object surface is directly lit or in shadow. The proportionality between RGB colour channels can be better represented using the normalised colour, $\mathbf{f} = (f_r, f_g, f_b)$, each component of which is:

$$f_k = f_K \Big/ \sqrt{f_R^2 + f_G^2 + f_B^2} \tag{12}$$

and will keep constant for a given object surface under varying illumination. Therefore, it is appropriate to model each **f** component using a Gaussian.

3 A Mixture of Colour Distributions

A mixture of *N* Gaussians has been used to model the potentially multiple backgrounds at each pixel. The probability of observing a value, $\mathbf{f}_t$, at a pixel is:

$$P(\mathbf{f}_t) = \sum_{i=1}^{N} \omega_{i,t} G(\mathbf{f}_t, \boldsymbol{\mu}_{i,t}, \boldsymbol{\Sigma}_{i,t}) \tag{13}$$

where G is the Gaussian probability density function, $\boldsymbol{\mu}_{i,t}$ and $\boldsymbol{\Sigma}_{i,t}$ are the (temporal) mean vector and covariance matrix of the i-th distribution, respectively; $\omega_{i,t}$ is an estimate of the weight, which reflects the likelihood that the distribution accounts for the data. To simplify the computation, the rgb values are assumed to be independent and then $\boldsymbol{\Sigma}_{i,t}$ can be approximately represented using the sum of its diagonal elements, $\sigma_{i,t}^2$.

We approximate the initial values of the temporal statistics using the spatial statistics over a local region (n pixels) of the start frame:

$$\begin{aligned}\boldsymbol{\mu}_{0,0}(\mathbf{x}) &= \frac{1}{n}\sum_{\Delta\mathbf{x}} \mathbf{f}_0(\mathbf{x}+\Delta\mathbf{x}) \\ \sigma_{0,0}^2(\mathbf{x}) &= \frac{1}{n-1}\sum_{\Delta\mathbf{x}} \left\|\mathbf{f}_0(\mathbf{x}+\Delta\mathbf{x}) - \boldsymbol{\mu}_{0,0}(\mathbf{x})\right\|^2\end{aligned} \tag{14}$$

For the following frames, every new observation, $\mathbf{f}_t$, is checked against the N Gaussian distributions. A match is defined if $\left\|\mathbf{f}_t - \boldsymbol{\mu}_{i,t-1}\right\| < c\sigma_{i,t-1}$ ($c \approx 3$). The parameters of the matched distribution are updated as:

$$\begin{aligned}\boldsymbol{\mu}_{i,t}(\mathbf{x}) &= (1-\varphi)\boldsymbol{\mu}_{i,t-1}(\mathbf{x}) + \varphi\mathbf{f}_t(\mathbf{x}) \\ \sigma_{i,t}^2(\mathbf{x}) &= (1-\varphi)\sigma_{i,t-1}^2(\mathbf{x}) + \varphi\left\|\mathbf{f}_t - \boldsymbol{\mu}_{i,t}(\mathbf{x})\right\|^2\end{aligned} \tag{15}$$

where φ controls the updating rate, and the weight $\omega_{i,t}(\mathbf{x})$ is increased. For the unmatched j-th distribution ($j \neq i$), $\boldsymbol{\mu}_{j,t}$ and $\sigma_{j,t}$ remain the same, but $\omega_{j,t}(\mathbf{x})$ is decreased.

If none of the existing distributions matches the current pixel value, we have to either create a new distribution, given less than N existing distributions, or replace the least probable distribution with a new distribution. The distribution(s) with greatest weight is (are) considered as the background model(s):

$$i(\mathbf{x},t) = \underset{j}{argmin}\left\{\omega_{j,t}(\mathbf{x})\right\} \tag{16}$$

One advantage of the Gaussian mixture model is that when something is allowed to become the background, the existing model of the previous background is still maintained. Therefore, if an object is stationary just long enough to become part of the background (e.g. a parked car) and then it moves, the distribution describing the previous background can quickly explain the new object-free background.

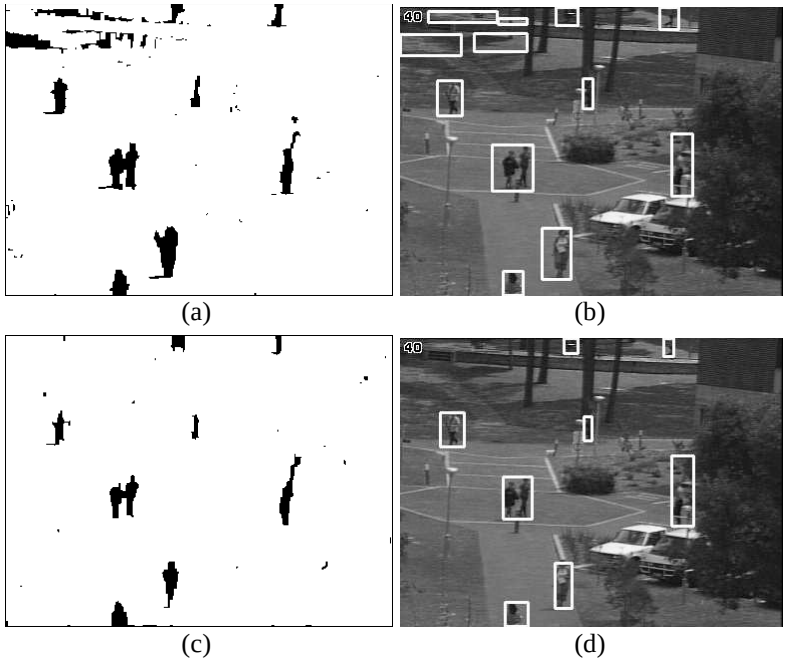
4 Experimental Results

To assess the significance of the colour-invariant motion detection, we evaluated it at both pixel and frame levels using a set of image sequences [9]. The image sequence shown here was captured at a frame rate of 2Hz. Each frame was lossily compressed in JPEG format and has a frame size of 384×288 pixels. This sequence adequately represents the abundant contexts of a daylit outdoor environment, with fast illumination changes, waving trees, shading of the tree canopies, highlights of specular reflection, as well as pedestrians (refer to Fig. 2(b)).

Figs. 2 and 3 show the results of the motion detection in two frames of the image sequence. The foreground pixels in the rgb results are those that go beyond [μ-3.5σ, μ+3.5σ] of the most probable Gaussians. The foreground pixels in the RGB results arise from a global threshold on the difference between the observation and the mean of the most probable Gaussian. The thresholding level is selected so as to produce "blobs" of similar sizes to those in the corresponding rgb results. In order to rule out isolated "foreground" pixels and fill gaps and holes in "foreground" regions, a 3×7 closing (dilation-erosion) operation has been applied to the binary image of detected "foreground" pixels. The grey-level intensity images here were obtained using $I = \sqrt{f_R^2 + f_G^2 + f_B^2} \big/ \sqrt{3}$.

Fig. 2 (at frame 40) is an example comparing the RGB and rgb results under a minor illumination change. The foreground "blobs" extracted using rgb space are as coherent as those using RGB space. Because of the different emphasis of image contexts for both the colour spaces, the corresponding blobs in Figs. 2(a) and (c) may appear as different shapes.

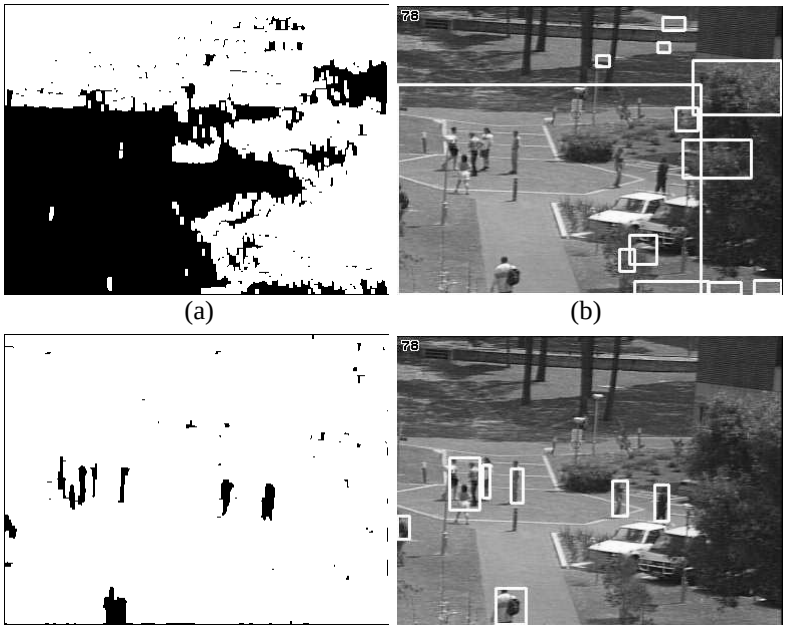
Fig. 3 (at frame 78) shows the RGB and rgb results under a major illumination change (refer to Fig. 5). In the RGB result, Fig. 3(a), a large area of the background is detected as a huge foreground object, in which the ground-truth targets (pedestrians) are submerged and lost. On the other hand, in the rgb result, Fig. 3(c), fast illumination changes give no additional "foreground" blob and the "ground truth" targets are clearly visible. Note the poor detection of some foreground blobs on the left of the frame is caused by the stationary pedestrians that are being absorbed into the estimated "background" by the adaptive Gaussian model.



(a) (b)

(c) (d)

Figure 2: Motion detection at frame 40 with little illumination change: the detected blobs (left) and corresponding bounding boxes overlaid on the frame (right) using the RGB (top) and rgb (bottom) spaces.



(a) (b)

(c) (d)

Figure 3: Motion detection at frame 78 with a major illumination change: the detected blobs (left) and corresponding bounding boxes overlaid on the frame (right) using the RGB (top) and rgb (bottom) spaces.

The foreground pixels above are clustered into foreground "blobs" using a connected component analysis. A minimum number of foreground pixels is set for each blob to rule out small disturbances. Due to the varying sizes of possible foreground targets, e.g. from a pedestrian to two intersecting trucks, the selection of the corresponding maximum number is not so trivial as that of the minimum number and has not been used here in order to differentiate the results of the intensity- and colour-based models. Each detected "foreground" blob is labelled by a rectangular bounding box, as shown in Figs. 2(b)(d) and 3(b)(d).

Table 1 shows the number of the detection errors in the same image sequence, from frame 16 (skipping the learning period) to frame 100. Multiple objects are considered as a single ground-truth object if they are grouped. Most of the undetected positives occur when ground-truth objects are lost in large-scale illumination changes. Most of the false positives occur when a piece of background under illumination changes is determined as a "foreground" object or occasionally the trees are waving. The colour-based model is much more successful in dealing with illumination changes.

Models	Intensity-based	Colour-based
No. of ground-truth objects	514	
No. of undetected positives	159	2
No. of false positives	418	9

Table 1: The detection errors in an image sequence with fast illumination changes.

5 Applications

Colour and intensity reflect the two distinct characteristics of an image. Motion detection based on only one aspect may fail in some specific situations. We have combined the motion detection results using the intensity, $I = \sqrt{f_R^2 + f_G^2 + f_B^2} / \sqrt{3}$, with those using the rgb colour space. Such an (r,g,b,I) colour space is a complete representation of the image information in that it can be invertibly transformed to and from the RGB colour space. However, this provides some promising applications that are not readily obtained from RGB space only.

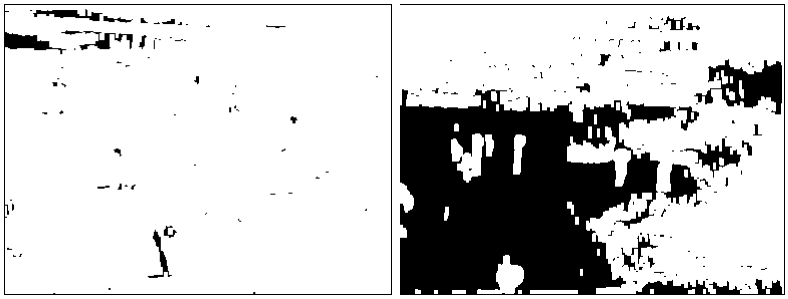
(1) Illumination change detection

An intensity-based model is sensitive to both foreground targets and illumination changes. A colour-based model responds only to targets. Therefore, a region can be determined as being shadowed or re-lit if the rgb components are stable but the I component has a significant change. Suppose S_I and S_C are the binary sets of motion detection using intensity- and colour-based models, respectively, and a value of 1 represents the detected foreground (0 for background). The regions where illumination varies include a set of pixels, **x**, which satisfy:

$$S_1(\mathbf{x}) = S_I(\mathbf{x}) \cdot \overline{(S_C \oplus B)(\mathbf{x})} \tag{17}$$

where $\oplus$ denotes the morphological dilation and B is the structuring element. The dilation operation gives some tolerance of the different foreground profiles detected using intensity- and colour-based models. The results of illumination change detection are shown in Fig. 4, where B is 3×7 sized.

Self-shadow detection can guide positioning and orientating of light sources in a scene. In an environment with large-scale illumination changes, the detection result can guide when to use an intensity- or colour-based model.



(a) (b)

Figure 4: The regions with illumination changes detected at frame 40 (a) and 78 (b), respectively.

(2) Switching between models

A colour-based model is vulnerable to failure when detecting targets in dim or saturated regions, where the colour chromaticity is unreliable. In contrast, an intensity-based model has a rather consistent performance, provided no drastic illumination change oc-

curs. Therefore, the ratio of illumination-varying regions to entire image, as well as the average intensity (illumination level), can be used as the indicator of when to use an intensity- or colour-based model. For example, at sunset when the average intensity in a scene is very low, the colour-based model is switched off and only the intensity-based model keeps working. Only when the average intensity becomes higher than some threshold, both models are applied to the image sequence simultaneously. The final detection result is switched to that of the colour-based model if the level of illumination change is higher than some threshold; Otherwise it is switched to either the result of the intensity-based model, that of the colour-based model, or the combination of both (see data fusion in (3)).

Fig. 5 shows the normalised average intensity and the ratio of illumination-varying regions to entire image, through the previous image sequence. The normalised average intensity is bounded between 0 (black) and 1 (white). For the ratio of illumination-varying regions, the peak at frame 0 arises from the learning errors of the initial model parameters, and each of the other smaller peaks (with a magnitude of 0.05-0.15) corresponds to a local illumination change due to the flooding of sunlight; The highest peak arises from a global illumination change.

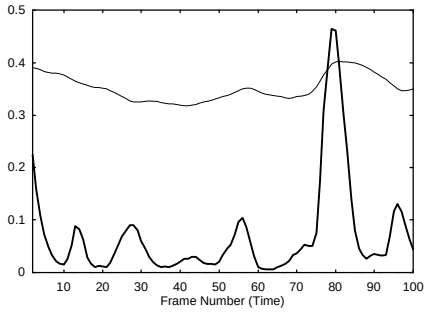


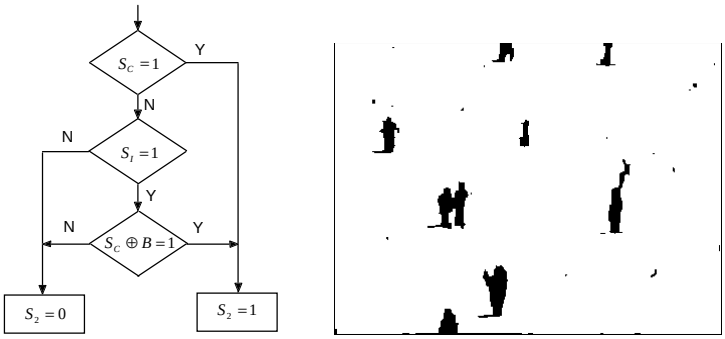
Figure 5: The normalised average intensity (thin line) and ratio of illumination-varying regions to entire image (thick line) through the previous image sequence.

(3) Data fusion between models

The intensity-based model fails to detect targets with a similar intensity to background, and the colour-based model misses targets with a similar colour chromaticity to the background. Therefore, both sets of the results can be combined to give better detection. One combination scheme favouring the colour-based model is to add some points of S_I, which is spatially close to S_C, into S_C. This can compensate for the loss of the colour-based model due to the similarity of the colour chromaticity between targets and background. This combination scheme can be configured into a decision tree shown as in Fig. 6(a). The set of the fused foreground pixels, S_2, can be computed as:

$$S_2(\mathbf{x}) = S_C(\mathbf{x}) + \overline{S_C(\mathbf{x})} \cdot S_I(\mathbf{x}) \cdot (S_C \oplus B)(\mathbf{x}) \tag{18}$$

Fig. 6(b) shows the result of applying such a combination scheme to frame 40. It is noted that the blob at the bottom is dilated to the same size as that for the intensity-based result (Fig. 2(a)). The combination is favourable to the colour-based result in that the regions with illumination change (in the upper-left corner) are still excluded and the right-most blob is as complete as that in the colour-based result (Fig. 2(c)).



(a) (b)

Figure 6: (a) a decision tree for data fusion using intensity- and colour-based models. (b) the fused result at frame 40 (compare with Figs. 2(a) and (c)).

6 Conclusions

A Gaussian mixture model based on the rgb colour space has been presented for maintaining a background image for motion detection. This scheme is especially successful when applied to outdoor scenes illuminated by daylight and is robust to fast illumination changes arising from moving cloud. The success results from a realistic reflection model in which shadows are present. We are currently working on the matching and tracking of the pedestrians in an outdoor environment, in which the principal colour chromaticity of each target plays a central role.

References

[1] G. D. Finlayson and S. D. Hordley. Colour invariance at a pixel, In *Proc. BMVC*, pp. 13-22, 2000.

[2] R. Gershon, A. D. Jepson and J. K. Tsotsos. Ambient illumination and the determination of material changes, *J. Opt. Soc. of Am.*, 3(10):1700-1707, 1986.

[3] J. A. Marchant and C. M. Onyango. Shadow invariant classification for scenes illuminated by daylight, to appear in *J. Opt. Soc. of Am.*, 2000.

[4] Y. Raja, S. J. McKenna and S. Gong. Segmentation and tracking using colour mixture models, In *Proc. Asian Conf. on Computer Vision*, 1998.

[5] S. Rowe and A. Blake. Statistical background modelling for tracking with a virtual camera, *Proc. BMVC*, pp. 423-432, 1995.

[6] J. M. Rubin and W. A. Richards. Color vision: representing material changes, AI Memo 764, MIT Artificial Intelligence Lab., 1984.

[7] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking, In *Proc. IEEE CVPR Conf.*, 1999.

[8] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland. Pfinder: real-time tracking of the human body, *IEEE Trans. PAMI*, 19(7):780-785, 1997.

[9] M. Xu and T. Ellis, Colour-invariant motion detection under fast illumination changes, *European Workshop on Advanced Video-based Surveillance Systems*.