# BACKGROUND MODELING BASED ON SUBPIXEL EDGES

*Vishal Jain   Benjamin B. Kimia   Joseph L. Mundy*

Brown University, Division of Engineering
Providence, RI, USA

## ABSTRACT

We propose an approach to model the background of images in a video sequence based on subpixel edge map. This work is motivated by the observation that intensity based background models are sensitive to changes in illumination and camera parameters, *e.g.*, gain control. In addition, the false positive rate is higher due to accidental alignment of figure intensities with the background model. Background models of edge maps, however, are more localized and thus reduce the likelihood of accidental alignment. We argue that the discretization error in pixel-level background models is also responsible for some of the false positives and develop a method based on subpixel edges whose background is thus highly selective. This method models the edge position and orientation using a Mixture of Gaussians model. This approach has been tested on a wide range of videos and the resulting background models are a much more selective figure-ground segregation.

***Index Terms***— Subpixel Edges, Mixture of Gaussians, Segmentation, Tracking, Figure-ground Segregation

## 1. INTRODUCTION

Methods for the analysis of moving objects in video sequences obtained from stationary cameras, *e.g.*, for surveillance and monitoring, typically model the stationary background and detect moving objects as those pixels which do not fit this model. Averaging frames over time is a simple method of constructing a background model which is effective if objects move continuously over the scene and lighting does not change rapidly. Background modeling using multiple distributions is used to handle images with slowly moving objects, slight lighting variations, and repetitive object movements [1, 2, 3, 4, 5]. The most popular schemes use the Mixture of Gaussian (MoG) model for each pixel. The intensity at each pixel is modeled using a fixed number of Gaussians which are updated on every observation. Any pixel which is unlikely to come from the MoG is classified as foreground.

Methods for modeling background intensity typically suffer from two limitations. First, they are susceptible to sudden changes in illumination, either global changes, *e.g.*, due to the sun coming out of clouds, or local changes, *e.g.*, due to partial reflection from a brightly colored objects passing nearby, Figure 1. Handling different illuminations requires either a broader distribution model or adding a new distribution to the mixture, both of which reduce sensitivity to figure segmentation. Second, these models are susceptible to changes in the camera model. For example, automatic gain control can change the overall intensity distribution as a bright object enters the field of view as illustrated in Figure 2. Another drawback of the intensity-based methods is that numerous observation frames are required, especially (i) when the illumination is changing and (ii) the scene is constantly occupied with moving objects or when objects are moving slowly.

An alternative to modeling background intensities is to model the background *intensity gradient*. Jabri *et al*. [6] augments the traditional intensity background model with models of the intensity gradient as captured by the Sobel edge responses. Large changes in either intensity or in edges are fused. However, the involvement of the intensity model retains the sensitivity to sudden changes in illumination. Javed *et al*. [7] in contrast, require significant changes in both the intensity and intensity gradient. The use of a gradient model removes many false alarms due to small illumination changes. However, intensity gradients arising from large illumination changes can still signal a figure when none exists, Figure 1(b).

A key limitation of intensity and intensity gradient background models is that background models do not take spatial interactions into account. Alternatively, edge maps tag those background pixels which maximize local gradient in a neighborhood of pixels. This tagging increases selectivity which in turn reduces both the number of pixels which would have been discarded from the background model and the pixels would have been erroneously labeled as foreground. Yang and Levine [8] modeled background edge-maps using robust statistics where edges diagnosed as outliers correspond to the foreground edges. Zhang *et.al* [9] identify an edge as a figure if its intensity is sufficiently different from an intensity background model. As a result, the output becomes insensitive to changes in focus and illumination. Kim and Hwang [10] detect the edges of current frame as well as of the difference image of consecutive frames. They compare the edge-locations of both maps with a background edge-map and detect foreground edges if the distance is within threshold.

Discretization errors in pixel-based edge maps lead to unnecessary broad background models: a background edge halfway

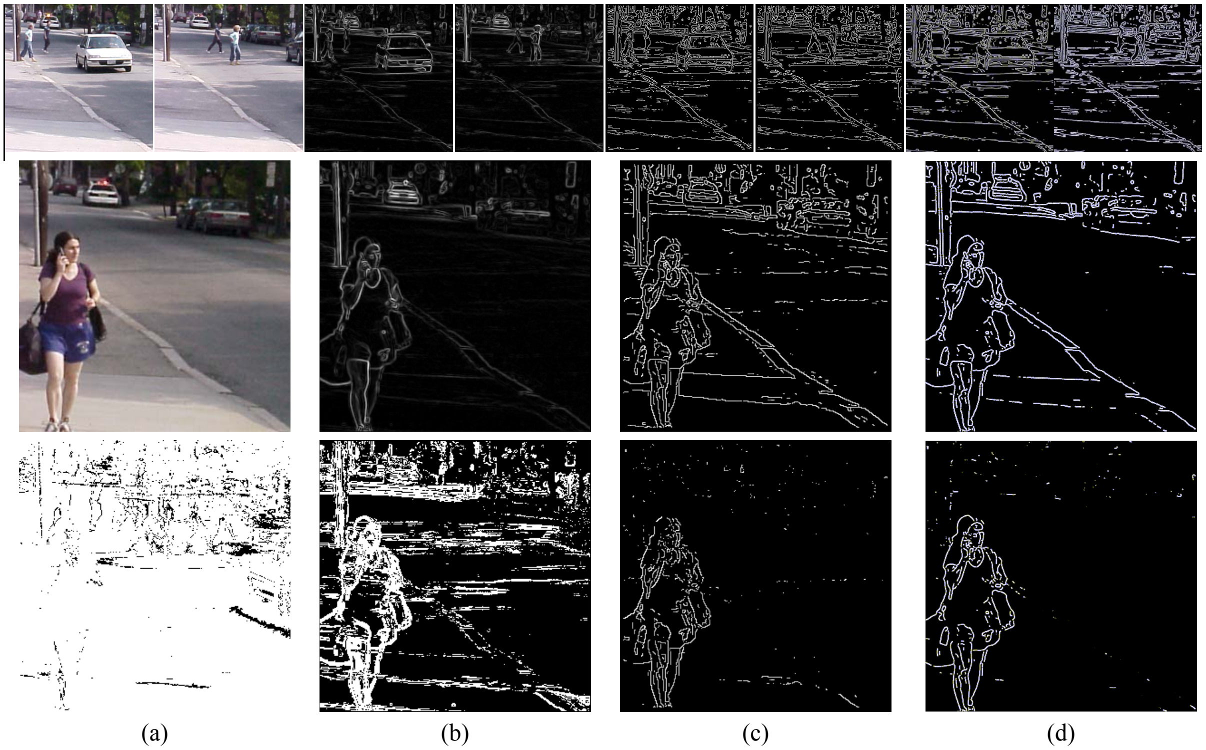


**Fig. 1**. The effect of sudden illumination change on different background modeling schemes is illustrated. Top Row: a pair of typical background images, and their (b) gradient maps, (c) edges and (d) subpixel edges. Middle row: (a) a frame when the illumination has changed and its (b) intensity gradient, (c) edge map and (d) subpixel edgemap. Bottom row: Foreground detection using (a) intensity, (b) intensity gradient, (c) edge map and (d) subpixel edgemap.

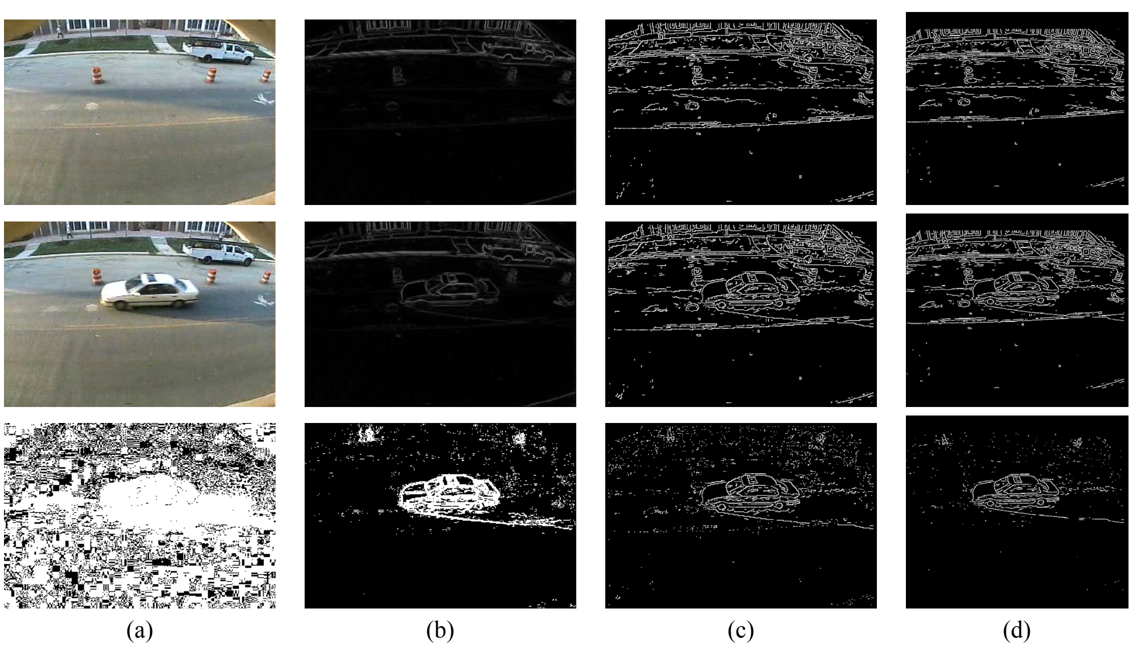


**Fig. 2**. The effect of change in the gain of the camera is depicted for different background modeling schemes. Top Row: a background input image, and its (b) gradient map, (c) edgemap and (d) subpixel edgemap. Middle row: (a) a new input image with a change in the gain of camera, and (b) intensity gradient, (c) edge map and (d) subpixel edgemap. Bottom row: Foreground detection results based on (a) intensity, (b) intensity gradient, (c) edge map and (d) subpixel edgemap. Observe that the extent of the spurious responses reduces from left to right.

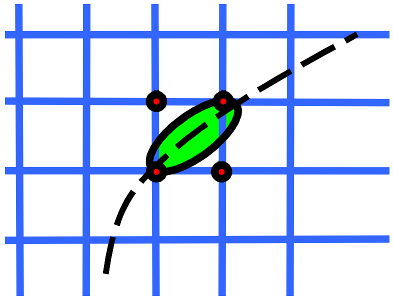

**Fig. 3**. Covariance $\Sigma_{x,y}$ (green ellipse) for an edge distribution and four red dots shown represents the sites for the edge under consideration. Any edge which lies in the distribution is considered to be background.

between the pixels will require both pixels modeled as background, thus unnecessarily "blurring" the background model, which in turn reduces sensitivity to detecting figures. Instead, we propose a background model based on *subpixel* edge-maps of the images. In our approach, we model position $(x, y)$ and orientation $\theta$ of subpixel edges which disambiguates between edges of the same orientation but at different positions and vice versa. Subpixel edge-maps attains high precision and accuracy in addition to being invariant to illumination changes and accommodates small translations easily. Another advantage is that the algorithm requires fewer frames to build the background model even in case of slow moving objects and busy scene. The advantage of modeling sub-pixel edges becomes evident in scenes with cluttered backgrounds where edges from a figure can share the same pixel as well as the same orientation as shown in Figure 4.

## 2. APPROACH

Subpixel edge-maps are obtained on a video sequence using a modified Canny edge detector as represented by a set $(x, y, \theta)$ for each edge, where $x, y \in \Re$ and $\theta \in [0, 2\pi)$. Since the edges are subpixel, we associate each edge to its corresponding *sites* (neighboring pixels) as illustrated by red dots in Figure 3. Note "sites" only facilitates indexing of distribution for the edges and it still allows for subpixel accuracy for modeling the distribution of edges across frames. Each of the sites has a mixture of 3D Gaussian distributions in which each Gaussian component represents the history of observations for an edge *w.r.t* its orientation and location.

Let $\chi_{x,y,\theta}$ be a random variable representing observations for each edge across frames. The Gaussian distribution of random variable $\chi$ is

$$\eta(\chi) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma_{x,y,\theta}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\chi-\mu(x,y,\theta))^T \Sigma_{x,y,\theta}^{-1}(\chi-\mu(x,y,\theta))} \quad (1)$$

At each site, we store mixture of such Gaussian distributions along with their weights $\omega$ which essentially represent the *frequency* of the same edge in the observation history. The probability of the current edge $e(x, y, \theta)$ being observed in the past is given by

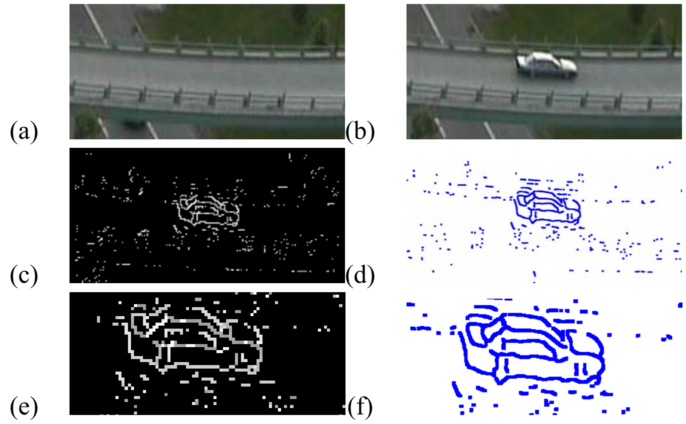

**Fig. 4**. This figure compares the accuracy of background modeling using pixel and subpixel edge maps. (a) One of the background images, (b) Foreground image, (c) Pixel-edge foreground detection, (d) Subpixel-edge foreground detection, (e) zoomed image of (c), (f) zoomed image of (d). Note how the roof of the car disappears in the case of pixel-edge modeling as the railing behind the car has the same orientation as of the roof edges.

$$P(e) = \max_{s\in S} \sum_{j=1}^{N} \omega_{sj} * \eta_{sj}\left(\chi, \mu(x, y, \theta), \Sigma_{x,y,\theta}\right) \quad (2)$$

where $S$ are the sites for $e$, $N$ is the number of components in the mixture, $\omega_{sj}$ and $\eta_{sj}$ are the weight and Gaussian distribution respectively, for $j^{th}$ component of the mixture at site $s$. The $N$ components of the mixture are ordered by the ratio $w_j/|\Sigma|$. The first $N_b$ components having sum of weights greater than a threshold $\tau$ are defined as background components. We use $\tau = 0.4$ to $0.6$.

**Foreground Edge Detection:** The foreground edges are detected if an observed edge $e_k(x_k, y_k, \theta_k)$ does not lie within 2.5 standard deviation away from the mean of any of the background components of the distribution at all its sites, *i.e.*,

$$(\chi_k - \mu_j(x, y, \theta))\Sigma_j^{-1}(\chi_k - \mu_j(x, y, \theta))^T < 2.5^2 \quad \forall j \in N_b \quad (3)$$

**Updating the Background Model:** The Gaussian components $j$ at corresponding sites $s$ which match the observation value $\chi_{k+1}$ are updated by the following equations [2],

$$\omega_{sj}^{k+1} = \omega_{sj}^{k} + \frac{1}{k+1}(\eta_{sj}(\chi_{k+1}) - \omega_{sj}^{k})$$

$$\mu_{sj}^{k+1} = \mu_{sj}^{k} + \frac{\eta_{sj}(\chi_{k+1})}{\sum_{p=1}^{k+1}\eta_{sj}(\chi_p)}(\chi_{k+1} - \mu_{sj}^{k})$$

$$\Sigma_{sj}^{k+1} = \Sigma_{sj}^{k} + \frac{\eta_{sj}(\chi_{k+1})}{\sum_{p=1}^{k+1}\eta(\chi_p)}(\chi_{k+1} - \mu_{sj}^{k})^T(\chi_{k+1} - \mu_{sj}^{k})$$

Most edges except noise edges are samples of a curve in the image, Figure 3. Since an edge can slide along the curve, one would expect a large variance in the tangential direction and small variation along normal as shown in Figure 3. Empirically we also get larger variation along the curve.

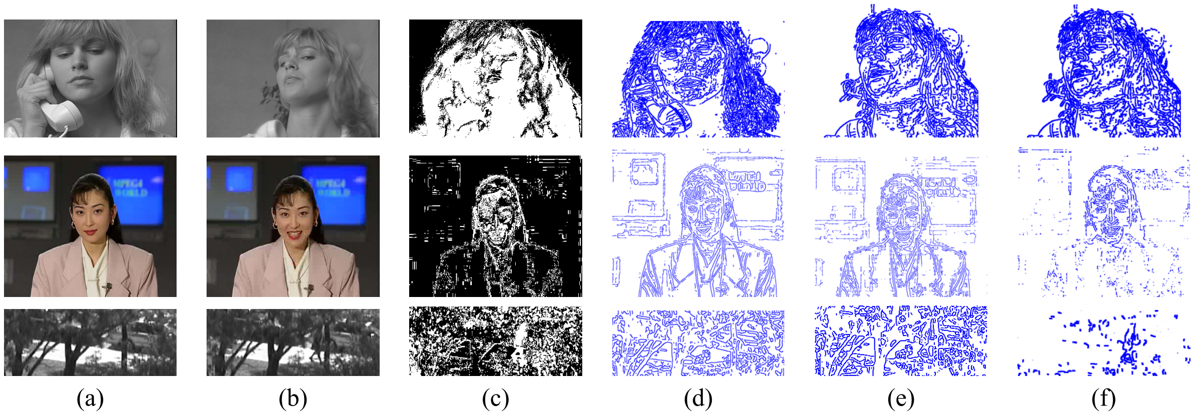


**Fig. 5**. A comparison on standard video sequence “Susie”, “Akiyo” and the “occluded pedestrian” video sequence.(a) a typical frame used for the background model, (b) a new frame for which a figure needs to be segregated, (c) foreground detected using intensity, (d) a typical subpixel edgemap for background modeling, (e) a new subpixel edgemap needed to be segregated, (f) foreground subpixel edgemap. It is clear in all cases the subpixel edge is more selective.

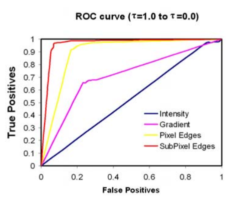


**Fig. 6**. ROC curve obtained for foreground detection using Intensity (Black), Gradient (Pink), Pixel Edges (Yellow) and Subpixel Edges (Red) for sequence Figure 1

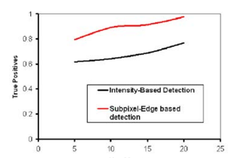


**Fig. 7**. Plot of detection rate vs number of frames to build the model on sequence in Figure 4 for intensity and subpixel edges based method.

## 3. EXPERIMENTS & RESULTS

We compared four background models:*(i)* intensity, *(ii)* gradient, *(iii)* pixel edges, and *(iv)* subpixel edges. First, one can *qualitatively* observe the differences among foreground detected by these models in Figures 1 and 2. We also show results on two widely used video sequences *(i)* Susie and *(ii)* Akiyo in Figure 5. Observe in the first and second sequence that the foreground occupies most of the scene and its very slowly moving, so it becomes really difficult to model the background for the video sequence. Our method is able to detect foreground as compared to the intensity based methods. The last sequence shows how robust our method is in images with trees and bushes. Second, we quantify these differences in the form of ROC curve, Figure 6, computed on the sequence shown in Figure 1. Ground truth was marked for each of the four methods manually for 5 frames and false positives and true positives were recorded by varying the detection threshold $\tau$. Also the plot in Figure 7 shows subpixel edge-based methods require less frames to build the model. Clearly, a background model based on subpixel edgemaps outperforms the others.

## 4. CONCLUSION

We have proposed a novel idea to model subpixel edges which provides superior foreground detection, especially in case of sudden illumination change. It provides detection of figure with high precision and good accuracy.

## 5. REFERENCES

[1] C. Stauffer and W.E.L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE TRANS. PAMI*, vol. 22, no. 8, pp. 747–757, August 2000.

[2] P. KaewTraKulPong and R. Bowden, “An improved adaptive background mixture model for real-time tracking with shadow detection,” in *Proc. on AVBSS*, Sept 2001.

[3] P. W. Power and J.A. Schoonees, “Understanding background mixture models for foreground segmentation,” *Imaging and Vision Computing New Zealand*, 2002.

[4] Dar-Shyang Lee, Jonathan J. Hull, and Berna Erol, “A Bayesian framework for Gaussian mixture background modeling,” in *Proc. ICIP*, 2003.

[5] Nicolas Martel-Brisson and Andre Zaccarin, “Moving cast shadow detection from a gaussian mixture shadow model,” in *CVPR*. 2005, pp. 643–648, IEEE Computer Society.

[6] Sumer Jabri, Zoran Duric, Harry Wechsler, and Azriel Rosenfeld, “Detection and location of people in video images using adaptive fusion of color and edge information.,” in *ICPR*, 2000.

[7] O. Javed, K. Shafique, and M. Shah, “A hierarchical approach to robust background subtraction using color and gradient information,” in *Motion02*, pp. 22–27.

[8] Yee-Hong Yang and Martin D. Levine, “The background primal sketch: An approach for tracking moving objects,” *MVA*, vol. 5, no. 1, pp. 17–34, December 1992.

[9] J. Zhang, L. Zhang, and H.M. Tai, “Efficient video object segmentation using adaptive background registration and edge-based change detection techniques,,” in *ICME04*, 2004.

[10] Changick Kim and Jenq-Neng Hwang, “Video object extraction for object-oriented applications,” *J. VLSI Signal Process. Syst.*, vol. 29, no. 1-2, pp. 7–21, 2001.