

# Robust Region-Based Background Subtraction and Shadow Removing Using Color and Gradient Information

Mohammad Izadi and Parvaneh Saeedi  
Laboratory of Robotic Vision, Simon Fraser University  
mia4@sfu.ca, psaeedi@sfu.ca

## Abstract

*In this paper, a novel algorithm for foreground detection and shadow removal is presented. The proposed method employs a region-based approach by processing two foregrounds resulted from gradient- and color-based background subtraction methods. The performance of the system is compared against conventional approaches for five indoor and outdoor video sequences. Experimental results confirm that the detection rate exceeds 90%, and the robustness is greatly improved.*

## 1. Introduction

Motion detection in video streams is a primary step for extracting information in many computer vision applications, including video surveillance, tracking, traffic control/monitoring, and semantic annotation. Conventionally, when fixed cameras are used with static backgrounds (e.g. stationary surveillance cameras), background subtraction is utilized to obtain an initial estimate of moving objects. The detection of cast shadows as foreground objects is a common problem that could lead to undesirable consequences. For example, shadows could connect different people walking in a group, generating a single object (blob) as the output of background subtraction. In such cases, it is more difficult to isolate and track individuals.

### 1.1. Previous Works

The pixel-level Mixture of Gaussians (MOG) [10] background model has become very popular due to its efficiency in modeling multi-modal distributions (e. g., waving trees, ocean waves, light reflection, etc), and its adaptation ability to changes in background (e. g., gradual light change) in a real-time implementation. Friedman and Russell [9] modeled the intensity values of a pixel by using a mixture of three Normal distributions for traffic surveillance applications. Stauffer and Grimson [10] presented a method that modeled the pixel intensity by a mixture of K Gaussian

distributions. Zivkovic and van der Heijden [11] proposed an improved GMM algorithm. They incorporated a model selection criterion to choose the proper number of components for each pixel on-line and in this way automatic full adaptation to the scene was obtained.

Shadow detection has been an active area of research. There are many techniques for shadow detection in video sequences [1-6] with a majority of them based on color video sequences. Shadow detection is conventionally based on invariant color features that are not significantly affected by illumination conditions. McKenna et al. [8] employed both the pixel and the edge information at each channel of the normalized RGB color space to detect shadowed pixels. Elgammal et al. [3] also utilized the normalized RGB color space, but included a lightness measure to detect cast shadows. Cucchiara et al. [2] processed the HSV color space to classify pixels having the same hue and saturation values but lower luminosity compared to the background, as shadows. Kim et al. [13] presented a codebook based algorithm for foreground-background. Their method could handle scenes containing moving backgrounds or illumination variations, while properly removing shadows.

In general, the penumbra of the shadow is detected with the assumption that edge intensity within the penumbra is much smaller than the edge intensity of actual moving objects. Clearly, such hypothesis is not held for video sequences containing low-contrast foreground objects, especially for outdoors applications.

### 1.2. Objectives

The objective of this paper is to propose a region-based method for foreground detection and shadow removal in video sequences. The proposed method is implemented to track people in indoor and outdoor scenes. Followings assumptions are made in this work:

- The system includes a background model initialization in which the current scene is viewed over 200 frames in its static state.
- The scene is not over-crowded meaning that

moving objects cover up to 70% of the scene.

## 2. The proposed algorithm

To remove the shadow appropriately, multiple cues (gradient, colors) are combined based on regional processing and consistency with the human mind. The method presented in [8] uses a single-Gaussian model of background gradients and chromaticity values. This method cannot detect the regions of foreground with low chromatic content and low texture as mentioned in [8]. Therefore, our method is developed bearing in mind the mentioned problem.

### 2.1. Background scene modeling

The Gaussians mixture method (GMM) presented by Zivkovic and van der Heijden [11] is adopted here to perform background subtraction in the color domain (modified version originally proposed in [10]). In this method, a mixture of  $M$  Gaussian distributions adaptively models each pixel's color (RGB). First, the GMM function is estimated for each pixel  $x$  using a training set  $\{x^1, \dots, x^{t-T}\}$  where  $T$  represents a reasonable adaptation time period. Here the estimated density for each pixel is computed by:

$$p(x_{i,j}) = \sum_{m=1}^M \pi_m N(x_{i,j}; \mu_m, \sigma_m^2 I) \quad (1)$$

where  $\mu_m$  is the estimated mean, and  $\sigma_m^2$  is the estimated variance for the Gaussian component. The covariance matrices are kept isotropic for computational efficiency.  $I$  represents the identity matrix. The estimated mixing weights, denoted by  $\pi_m$ , are non-negative and they add up to one.

Given a new data sample at time  $t$ , GMM is updated using the following recursive equations:

$$\pi_m \leftarrow \pi_m + \alpha(o_m^t - \pi_m) - \alpha c_T \quad (2)$$

$$\mu_m \leftarrow \mu_m + o_m^t (\alpha / \pi_m) \delta_m \quad (3)$$

$$\sigma_m^2 \leftarrow \sigma_m^2 + o_m^t (\alpha / \pi_m) (\delta_m^T \delta_m - \sigma_m^2) \quad (4)$$

where  $\delta_m = x^t - \mu_m$ , and  $\alpha = 1/T$ . Here  $c_T$  is a constant value ( $c_T$  is set to 0.01 [11]). For a new sample, the ownership  $o_m^t$  is set to 1 if there is a ‘‘close’’ component with the largest  $\pi_m$ . A sample is ‘‘close’’ to a component if the Mahalanobis distance between the two is less than three. If there is no ‘‘close’’ component, a new component is generated with  $\pi_{M+1} = \alpha$ ,  $\mu_{M+1} = x^t$ , and  $\sigma_{M+1}^2 = \sigma_0$ , where  $\sigma_0$  is an initial variance. If the maximum number of

components is reached, the component with the smallest  $\pi_m$  is discarded. At the end of each update cycle, components with negative weights are removed from the rest of the process.

The weights are sorted in decreasing order and the first  $B$  distributions are selected as the background [25].

$$B = \arg \min_b \left( \sum_{m=1}^b \pi_m > 1 - c_f \right) \quad (5)$$

here  $c_f$  is a measure of the maximum portion of the data that could belong to foreground objects without influencing the background model. The Expectation Maximization (EM) algorithm is implemented for finding the best solution. A pixel belongs to the background, if the background model is greater than a threshold,  $c_{th}$ . Using the above background model an image is generated at time  $t$  with the following attribute:

$$I'(x, y) = \begin{cases} 1 & \text{foreground pixel} \\ 0 & \text{background pixel} \end{cases} \quad (6)$$

### 2.2. Gradient-based subtraction

In the gradient-based background subtraction, a single Gaussian distribution [7] is assumed to model the gradient value of each pixel. Therefore, the gray-scale image at each frame is filtered using Sobel kernel. The gradient magnitude for each pixel,  $\Delta_{i,j}$ , is used to update the Gaussian model. The training set,  $\{\Delta^1, \dots, \Delta^{t-T}\}$ , is employed to initialize the Gaussian model. The estimated density for a pixel at position  $(x, y)$  is denoted by:

$$p(\Delta_{x,y}) = N(\Delta_{x,y}; \mu_{x,y}, \sigma_{x,y}^2) \quad (7)$$

here  $\mu_{x,y}$  and  $\sigma_{x,y}^2$  are the estimated mean and variance. For each data sample at time  $t$ , the model is updated by:

$$\mu_{x,y} \leftarrow \mu_{x,y} + \alpha (\Delta_{x,y}^t - \mu_{x,y}) \quad (8)$$

$$\sigma_{x,y}^2 \leftarrow \sigma_{x,y}^2 + \alpha ((\Delta_{x,y}^t - \mu_{x,y})^T (\Delta_{x,y}^t - \mu_{x,y}) - \sigma_{x,y}^2) \quad (9)$$

where  $\alpha = 1/T$ . Now, a binary map  $I'_\Delta(x, y)$  representing the pixel segmentation at position  $(x, y)$  for time  $t$  in foreground and background is defined by:

$$I'_\Delta(x, y) = \begin{cases} 1 & \left| \Delta_{x,y}^t - \mu_{x,y} \right| \geq \beta \sigma_{x,y} \\ 0 & \left| \Delta_{x,y}^t - \mu_{x,y} \right| < \beta \sigma_{x,y} \end{cases} \quad (10)$$

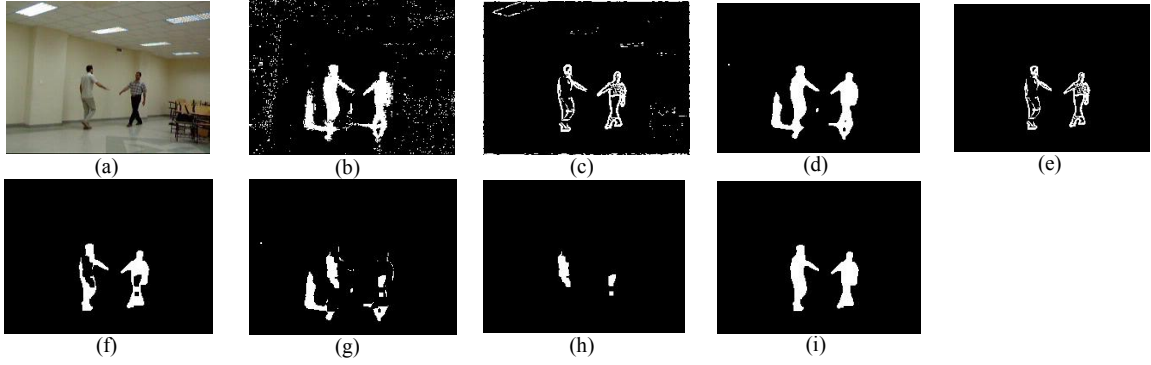


Figure 1: (a) A frame of a sequence, (b) the binary map  $I'(x, y)$ , (c) the binary map  $I'_\Delta(x, y)$ , (d) the filtered  $I'(x, y)$ , (e) the noise-free  $I'_\Delta(x, y)$ , (f) morphological close filtered  $I'_\Delta(x, y)$ , (g) subtraction of  $I'_\Delta(x, y)$  from  $I'(x, y)$ , (h) non-shadow regions, (i) resulting image.

here  $\beta$  is a constant value that for this work it is set to 3.

### 2.3. Shadow removal

At this point of the process (time= $t$ ) there are two binary maps extracted using described algorithms in Sections 2.1 and 2.2. The binary map  $I'(x, y)$  (Figure 1(b)), extracts foreground pixels well, but it could contain shadows of moving objects that must be removed.

The binary map  $I'_\Delta(x, y)$  includes only some parts of the moving objects as foreground (Figure 1(c)). This is under the condition that the shadows of moving objects are appropriately removed from  $I'_\Delta(x, y)$ .

The proposed idea in this paper is that any foreground region that corresponds to an actual object and does not exist in  $I'_\Delta(x, y)$  could be recovered from  $I'(x, y)$ .

Following steps describe the proposed algorithm:

1. A median filter, for the noise removal purpose, is applied to the binary image  $I'(x, y)$  (Figure 1(d)).
2. To remove noise from  $I'_\Delta(x, y)$ , any pixel of  $I'_\Delta(x, y)$  whose value is equal to one and its corresponding pixel value in  $I'(x, y)$  is zero, is set to zero (Figure 1(e)).
3. A morphological close filtering is performed on the resulting image using a circular structuring element of 3-pixel diameter to fill the gaps and smooth outer edges (Figure 1(f)).

4. A binary image  $I'_S(x, y)$  is generated by subtracting  $I'_\Delta(x, y)$  from  $I'(x, y)$ , Figure 1(g).
5. A connected component algorithm [14] is applied to  $I'_S(x, y)$ .
6. For each region  $R$ , its outer boundary  $\nabla R$  is extracted by subtracting the dilated region from the original region.
7. A region  $R$  is declared as a shadow region if it satisfies the following inequality:

$$\frac{\sum (\nabla R(i, j) I'_S(x, y))}{\sum \nabla R(i, j)} \leq P \quad (11)$$

This equation calculates the percentage of a region boundary in common with the moving object's boundary. In this equation  $P$  is a constant ( $0 \leq P \leq 1$ ).

When  $P$  is equal to zero, no region is considered as a shadow region. The larger the  $P$  the higher the probability of finding shadow regions. All non-shadow regions (Figure 1(h)) are added to the enhanced image  $I'_\Delta(x, y)$  (Figure 1(f)) to generate the final image as shown in Figure 1(i).

The resulting image contains moving objects without their shadows. By employing the above method, any sudden luminance change, i.e. turning on a flash light in the scene, will not cause spurious foreground regions.

### 3. Experimental results

In this section experimental results and the quantitative comparisons of the three methods—the proposed algorithm, the method in [8] and codebook-based model in [13]—are presented. To test the

algorithm, first a set of sequences was chosen to form a complete and nontrivial benchmark suite. Five scenarios including three outdoors and two indoors under different lighting conditions and perspectives are employed here (*campus I*, *campus II*, *shop*, *laboratory*, and *classroom* sequences). The ground truth was prepared manually over twenty frames for each video sequence representative of different situations including dark/light objects, multiple objects or single object, occlusion and non-occlusion cases.

To evaluate the moving object detection algorithms quantitatively, three metrics are employed: the *Detection Rate* (DR), the *Specificity* (Spec), and the *False Alarm Rate* (FAR).

$$\begin{cases} DR = \frac{N_{TP}}{N_{TP} + N_{FN}} \\ Spec = \frac{N_{TN}}{N_{TN} + N_{FP}} \\ FAR = \frac{N_{FP}}{N_{FP} + N_{TP}} \end{cases} \quad (12)$$

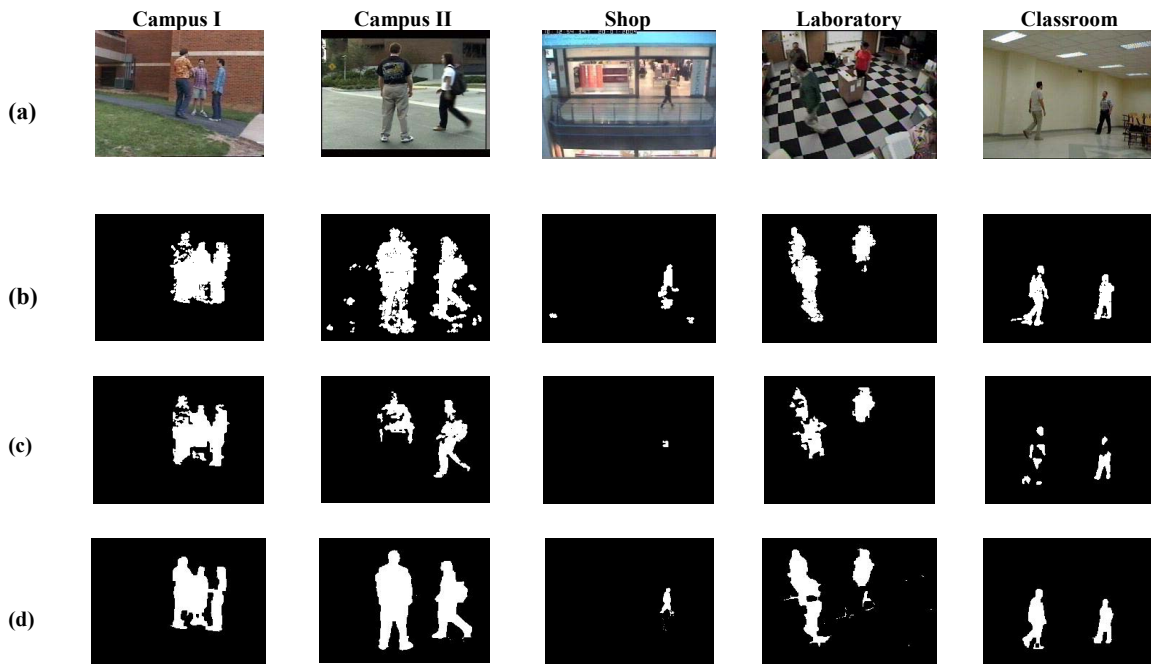
For the purpose of comparison, the proposed algorithm is evaluated using the following parameter setting:  $M = 4$ ,  $\alpha = 0.001$  and  $c_T = 0.01$ .

Parameters  $\beta$  and  $P$  are determined using Receiver Operating Characteristic (ROC) analysis (to minimize FAR) for each video sequence as well as parameters in the codebook-based method.

Computational complexity increases directly with the parameter  $M$ . However, the suggested value in here satisfies the requirement. Moreover it was noticed that a larger  $M$  had no significant effect on the final results. These parameters have identical values in all test cases. Also, to evaluate method in [8], the parameters setting was chosen according to the values presented in this paper except for parameter  $\alpha$  which is set to 0.001. The provided results are not very sensitive to the parameter  $P$  in the suggested algorithm. Moreover the presented method is robust with respect to parameter  $\beta$  within the range of [1.5,3].

To establish a fair comparison, all algorithms do not implement any background updating process. Instead, the reference image and other parameters are computed from the first  $N$  frames (with  $N$  varying for each sequence).

The visual representation of the segmentation results are shown in Figure 2. Besides the visual comparison, the results are evaluated quantitatively using the three mentioned metrics in comparison with the “ground truth” images. These results, summarized in Table 1, confirm the superior performance of the proposed method for all cases. As shown in Figure 2, the resultant images of the proposed method (Figures 2(d)s) are more precise than those of others. From these results, the method presented in [8] removes shadows appropriately. However it also removes some regions of the foreground with low texture and chromatic values. The codebook-based model has properly removed shadows and its specificity rate is high. Unfortunately, it has also removed some of the



	Campus I			Campus II			Shop			Laboratory			Classroom		
	DR (%)	Spec (%)	FAR (%)	DR (%)	Spec (%)	FAR (%)	DR (%)	Spec (%)	FAR (%)	DR (%)	Spec (%)	FAR (%)	DR (%)	Spec (%)	FAR (%)
The codebook method	89.94	97.71	21.95	88.44	96.92	17.29	90.31	97.12	35.84	73.88	97.38	24.52	79.45	99.52	9.44
The method in [8]	87.10	98.72	14.13	34.83	99.29	10.31	36.87	99.36	20.06	45.35	98.49	12.43	46.67	99.55	6.24
The proposed algorithm	91.37	98.96	11.09	93.95	99.40	4.51	93.76	99.47	13.99	87.40	98.68	11.75	90.15	99.67	5.93

Table 1: Quantitative evaluation of different methods.

foreground regions and therefore its detection performance falls lower than the other methods for most sequences, especially for the noisy sequences.

These results verify that the proposed method performs better in removing shadows; they also suggest a more robust performance in detecting moving objects. In general, the proposed method achieves the best performance, and offers highest robustness with respect to noisy sequences.

#### 4. Conclusion

In this paper, a novel approach for foreground segmentation and shadow removing in video sequences is presented. The improved GMM-based background subtraction and edge information are employed for object detection and shadow removal process. The proposed method uses region-based processing results to remove the object's shadows. Its performance is evaluated for five indoor and outdoor video sequences and it is compared against two other methods' performances. Experimental results verify that the proposed method performs significantly better for situations including non-stationary background, camouflage and shadows in color video sequences than the conventional approaches.

#### 5. References

[1] S. Y. Chien, S. Y. Ma, and L.G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. On CSVT*, 12(7):577–586, 2002.

[2] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. PAMI*, 25(10):1337–1342, 2003.

[3] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. of the IEEE*, 90(7):1151–1163, 2002.

[4] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features,"

*Computer Vision and Image Understanding*, 95(2):238–259, 2004.

[5] B. Shoushtarian and H. E. Bez, "A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking," *Pattern Recognition Letters*, 26(1):91–99, 2005.

[6] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Trans. on Multimedia*, 1(1):65–76, 1999.

[7] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. on PAMI*, 19(7): 780-785, 1997.

[8] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, 80(1):42–56, 2000.

[9] N. Friedman, and S. Russell, "Image segmentation in video sequences: a probabilistic approach," *Proc. of the 13th Conf. on Uncertainty in AI*, pp. 175-181, 1997.

[10] C. Stauffer and W. Grimson, "Adaptive background mixture model for real-time tracking," *Proc. IEEE Comp. Soc. Conf. CVPR*, vol. 2, pp. 246–252, 1999.

[11] Z. Zivkovic, and F. van der Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Trans. on PAMI*, 26 (5):651–656, 2004.

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

[13] K. Kim, T. H. Chalidabhongse, D. Harwood and L. Davis, "Real-time Foreground-Background Segmentation using Codebook Model," *Real-Time Imaging*, 11(3):172-185, 2005.

[14] M. R. Haralick, and G. S. Linda, "Computer and Robot Vision," vol. I, Addison-Wesley, pp. 28-48, 1992.