Moving Objects Tracking Using Statistical Models

Sara Sharifzadeh
Department of Microelectronics and Electronic Systems
Universitat Autònoma de Barcelona (UAB)
08193, Bellaterra, Spain
E-mail: sarasharifzade@yahoo.com

Abstract

Object detection plays an important role in successfulness of a wide range of applications that involve images as input data. In this paper we have presented a new approach for background modeling by nonconsecutive frames differencing. Direction and velocity of moving objects have been extracted in order to get an appropriate sequence of frames to perform frame subtraction. Stationary parts of background are extracted from differenced frames and joined as patches to complete the background model. There is also a special stage to handle changing regions of dynamic scenes. During the detection phase, the modeled background is updated for every new frame. Since it's not necessary to estimate each pixel gray value like the most common statistical methods, modeling process is not time-consuming. Different experiments show successful results even for challenging phenomena like environmental changes.

Keywords: *Object tracking, Background modeling, Frame subtraction.*

1. Introduction

Background subtraction is one of the major methods for detection of moving objects. This method is based on comparison of observed frame with background model which allows the detection of unmatched pixels between two frames. Then, foreground objects can be extracted by thresholding the difference image. There are cases that the initial background is not available explicitly and it should be modeled. This model is usually updated to overcome variations of the scene due to environmental changes such as moving regions, changes in illumination and camera noise.

Most common methods in background modeling are based on estimation the gray level of each pixel separately. Using adaptive filters like Kalman and Wiener to predict intensity of pixels [1, 2], Parametric mixtured Gaussian models for each pixel [3, 9] are two significant methods. A similar Bayesian approach in this field has been explained in [6]. Another approach is nonparametric modeling of background [4, 8]. Hidden Markov models are also used for background modeling [5, 10]. In some recent works, on-line Adaboost (on-line version of boosting) have been used [7]. Since in our work we join different patches of images and we do not need to estimate every pixel statistically، the method isn't time consuming. These patches are extracted from several appropriate frames. These frames are selected among input stream based on a special arrangement.

First the original sequence is divided in to subsets. Direction and velocity of moving object(s) are calculated in each subset which is used to form main sets. In the resulting sets, subtraction of non-consecutive frames with a special order is performed. This will lead to discover uncovered parts of background in both initial frames. These patches are extracted and joined with each other in the background model. There is a special solution to find dynamic patches like moving leaves in a windy scene that can never be found as static parts in the differenced frames. The modeled background is updated during detection phase for every new frame.

The paper is organized as follows: in section 2 the framework including finding the direction and velocity of moving objects, using movement information in frame subtraction and finding dynamic patches will be explained. The detection and updating phase are described in section 3. Experimental results will be demonstrated in section 4.

2. Background Modeling

When two frames of a scene are differenced from each other, similar parts of background in both frames will be black in the resulting frame. In a particular view of a scene captured from one camera, this feature can be used to find uncovered static parts of the scene corresponding to black regions in the differenced frame. If it is possible to extract all parts of background in this way, the whole background can be modeled by joining these parts.

An error occurs when the color of foreground and background is approximately the same in some regions.

An example is shown in Fig.1(a). It can be seen especially in the legs of the person. This will cause that some of these parts become black in the differenced image while they are covered by foreground in one of initial frames as shown in Fig.1(b). To overcome this problem, first we have removed small scattered points of the differenced frame and then expand all remaining areas with a special rate of dilation. This is illustrated in Fig.1(c). The rate value is defined automatically as a coefficient of the threshold value of each differenced frame. Another problem arises when there is an overlap in the position of foreground objects in two frames. As shown in Fig.2, this overlapping area will be black in the differenced frame, while the corresponding regions in both initial frames are occluded by foreground objects. Although these error regions are covered by extension partly, in order to minimize the probability of remaining any error regions, a quite number of appropriate frames are chosen with out any overlap between the two positions of foreground objects. Consider to have such a sequence of frames as follows:

$$F = \{ f_1, f_1, \ldots f_N \} \tag{1}$$

Where N is the length of the sequence. An example is shown in Fig.3(a) for a few numbers of frames in a scene. If the K^{th} frame be subtracted from $(N/2+K)^{th}$ one, a sequence of differenced frames is achieved as follows:

$$I = \{ I_{d1}, I_{d2}, \ldots I_{dN/2} \} \tag{2}$$

Fig.3(b) illustrates these differenced frames. There isn't any overlapping error in them. To be able to extract the whole background patches, we need to have quite number of frames with similar feature. In fact, the desired condition does not exist in any sequence of input stream especially when several objects are moving in different directions that raises the probability of overlapping.

2.1. Getting Appropriate Sequence of Frames

As discussed before, the major requirement in our algorithm is a sequence including enough number of frames so that major moving objects cover their width and leave their initial place through a particular direction. This needs some information about their velocity and direction.

2.1.1 Finding the direction and velocity of Moving Objects

In order to discover the direction of moving objects, variations in row or column numbers of their position in image matrix have been considered. In horizontal movement the column number changes and in vertical movement the row number changes. Providing the difference of consecutive frames and thresholding them, the moving object location would be detected through a sequence of frames. Then scattered unwanted points are removed using appropriate filters and a set of differenced frames is obtained as follows**:**

$$F_d = \{ f_{d1}, f_{d2}, \ldots f_{d(N-1)} \tag{3}$$

Where F_d is the set of consecutive differenced frames. Using connected component labeling, the most prominent changing region in each frame is discovered which deals with the most effective moving object of the scene. The mean values of rows and columns of these regions are calculated in the following sets:

$$Row = \{R_{M1}, R_{M2}, \ldots R_{M(N-1)}\} \tag{4}$$

$$Column = \{C_{M1}, C_{M2}, \ldots C_{M(N-1)}\} \tag{5}$$

The features of Gradient function have been used to detect the direction of moving objects. Any ascending set of numbers have positive gradient values and descending one, have negative gradient values.

$$\nabla Row = \frac{\partial Row}{\partial r} = \{Gr_{M_1}, Gr_{M_2}, \ldots Gr_{M(N-1)}\} \tag{6}$$

$$\nabla Column = \frac{\partial Column}{\partial c} = \{Gc_{M_1}, Gc_{M_2}, \ldots Gc_{M(N-1)}\} \tag{7}$$

Where f is the frame variable which shows the number of rows or columns passed per frame. At this stage positive and negative values in above sets are arranged into four different sets separately to find the direction of moving object:

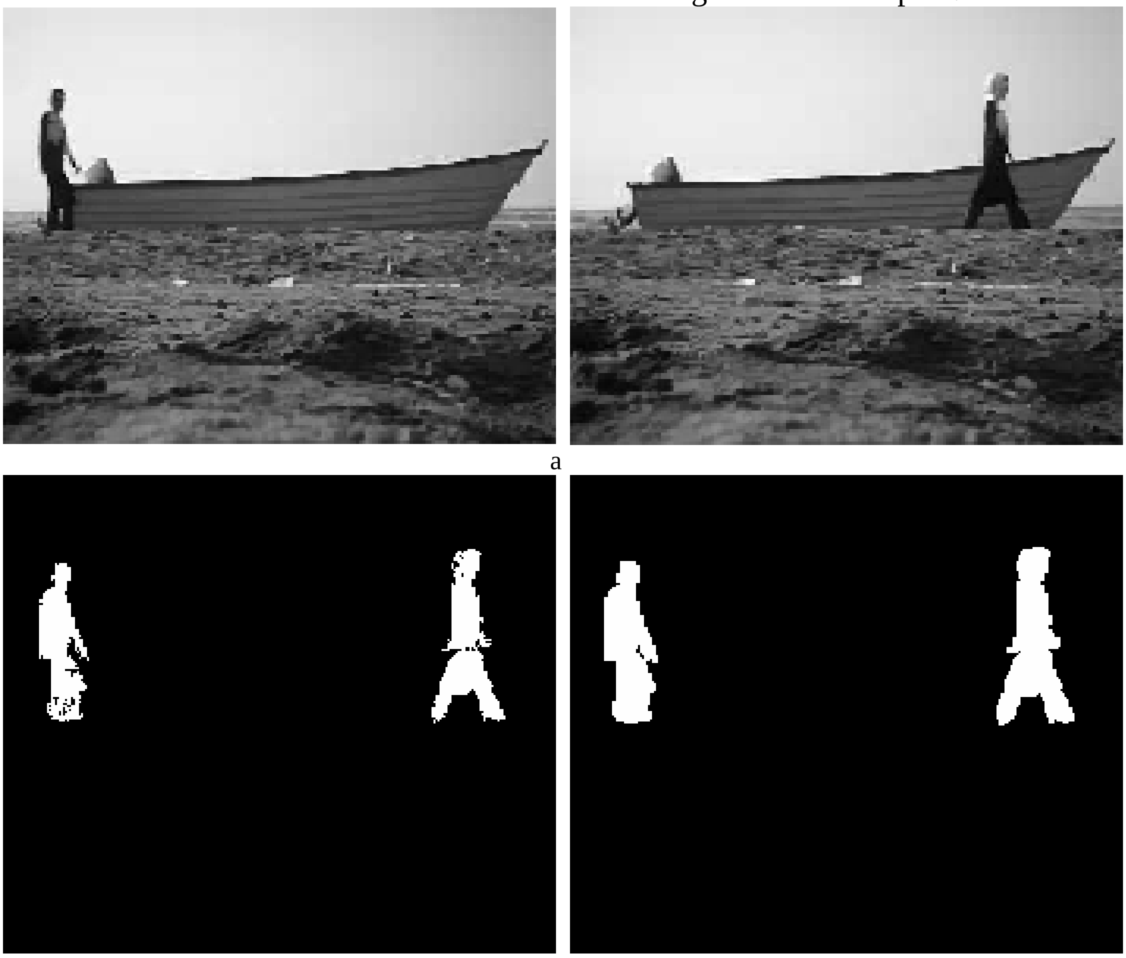
$$\nabla Row_{(p)} = \left\{ Gr_{Mi} \;\middle|\; Gr_{Mi} > 0 \right\} \tag{8}$$

$$\nabla Row_{(n)} = \left\{ Gr_{Mi} \;\middle|\; Gr_{Mi} < 0 \right\} \tag{9}$$

$$\nabla Column_{(p)} = \left\{ Gc_{Mi} \;\middle|\; Gc_{Mi} > 0 \right\} \tag{10}$$

$$\nabla Column_{(n)} = \left\{ Gc_{Mi} \;\middle|\; Gc_{Mi} < 0 \right\} \tag{11}$$

Combining these results with corresponding direction of ascending or descending sets of rows or columns, the direction would be determined. For example if the maximum number belongs to $\nabla Row_{(n)}$ it means that the moving object moves vertically from down to up side as the row numbers have been decreased. Diagonal movements can be classified in to one of these sets based on the angle value of the path.



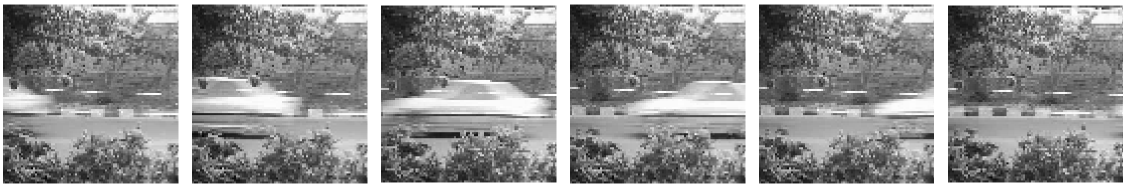
a

(b) (c)

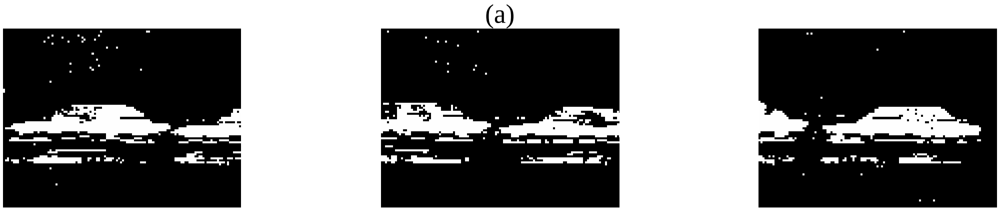
Figure 1. *(a) Frames with similarity between colors of background and foreground in some regions, (b) The differenced frame (c) expanded detected regions.*

(a) (b) (c)

Figure 2. *(a, b) Frames with overlap in the position of foreground objects (c) Differenced frame.*



(a)



(b) (C) (d)

Figure 3. *(a) Frames in which the car moves along the initial place completely (b) The difference of the first and the 4^{th} frame (c) The difference of the second and the 5^{th} frame (d) The difference of the third and the 6^{th} frame.*

The algorithm should work in any real condition without any information about the number of moving objects or their size, direction and velocity. But in order to minimize the probability of presence of several moving components, it is necessary to limit the number of frames that should be analyzed for the direction. On the other hand, as mentioned before, it is important that the object moves along its initial place in a particular direction through these frames. Size and velocity of moving objects play a significant role in this case. Obviously more frames are required for slow moving objects as well as bigger ones and for faster or smaller ones fewer numbers of frames is needed. We have limited the number of frames to a certain value. It is obtained due to this fact that our experiments are done in real world and as an example the biggest and slowest object can be considered a truck in about 4 meter width and at the very slow speed of $3_{Km/Sec}$ .According to these considerations, it takes **5** seconds that such a slow big object moves along its covered area. As a result, assuming to have **5** frames per Second, we have directed each **25** frame separately.

Surely, these frames are also sufficient for objects at higher speed or smaller size to move along their covered area.

In order to calculate the velocity of a moving object in a particular direction, absolute values of corresponding gradients of row or column changes are used. This is done for the set that shows the direction of moving object among the four sets in Eq. (8)–(11). While the absolute value of each member in this set shows the instantaneous velocity of moving object per frame, the average of these velocities shows the mean velocity. For example if the set $\nabla Group$ shows the direction, we have:

$$|V| = \{ \ |Gr_1|, |Gr_2|, ..., |Gr_X| \ \} \tag{12}$$

$$V_m = \frac{1}{K} \sum_{i=1}^{k} V_i = \frac{1}{k} \sum_{i=1}^{k} |Gr_i| \tag{13}$$

Where $|V|$ is the velocity set, V_m is the average velocity and K is the length of this set.

2.1.2 Forming the main set and nonconsecutive frames differencing

As mentioned before, certain number of subsets with **25** frames is directed and the mean velocity is calculated for each one. An analysis is performed to determine the number of these subsets which will be explained in the following paragraph. Then the average of these mean velocities will be calculated to estimate the velocity of moving object(s) in the whole subsets. This is an appropriate criterion to determine the length of main sets which are used in final step of this background modeling algorithm. As described in section **2**, subtraction of nonconsecutive frames in the main sets, will lead to extraction of different background patches. Assuming to have frames with **320** pixel width and **240** pixel height, if the estimated average velocity is V_0*col/sec*, at least $320/V_0$ frames are needed so that moving objects moves along the view of the camera, as well as $240/V_0$ frames for V_0*row/*sec average velocity. To ensure that all background regions are extracted via this set, tree main sets are prepared. Then the direction of moving objects in each one is calculated again and unidirectional ones will be merged together. This will help us in situations that some parts of background are occluded with temporary standing objects. Considering maximum required length for slowest objects ($L_{\max}$), the number of initial subsets with **25** frames can be estimated as follows:

$$3 \times L_{\max} = n \times 25 \tag{14}$$

At this stage, subtraction of nonconsecutive frames can be accomplished in main sets. This will result in a complete model of background as described in section 2 expect in some special situations that there are regions with dynamic nature such as moving leaves in windy situation. Since these regions are continuously changing, they will never appear as black stationary parts in the differenced frames. They remain as black points in the final background model.

2.1.3 Repairing Dynamic Regions

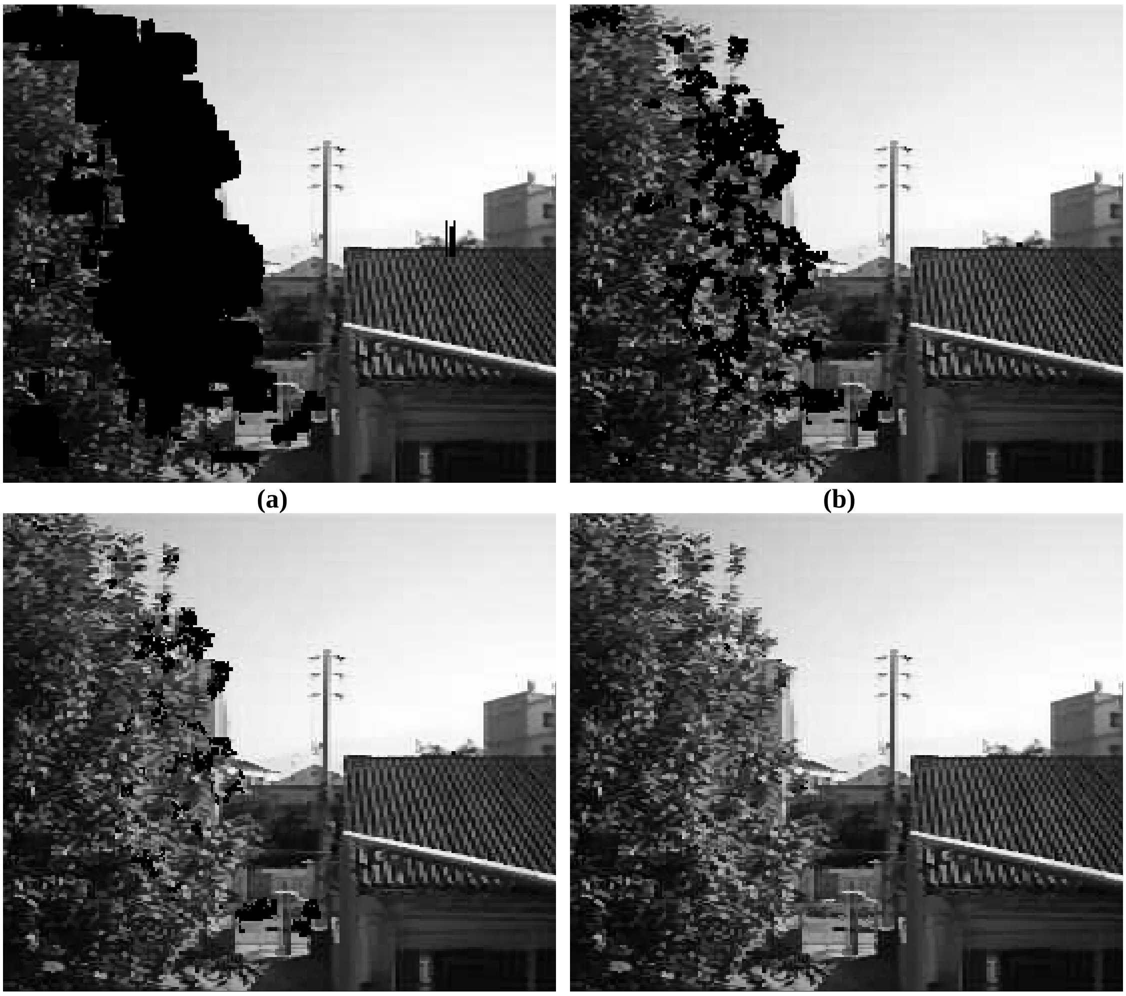
In order to repair constantly changing regions of background, these regions should be replaced with the corresponding gray values in the last main set to handle the most recent changes. Therefore nonconsecutive subtraction is performed in the last main set and extension process is done with a smaller rate of dilation in order to prevent these points appearing as white foreground regions. These frames include the most recent samples of the remaining pixels which are appropriate to repair them.

3. Detection and Updateing

The resulting background model is used to detect moving objects by subtraction of new frames from the model. For each new frame the model is updated in regions that have been repaired in the last stage of modeling process.

4. Experimental Results

The experiments have been done on images with highly textured scenes and environmental changes in which different size objects move with various speeds. We achieved a frame rate of 5 frames per second with the size of 240×320 pixel for our experiments. Fig 4.(a)-(c) shows three steps in background modeling of a scene. Because of windy condition and day light changes, leaves of trees are changing continuously. Fig.4.(d) shows the completed background model. Fig.5 illustrates the results of moving object detection. The average velocity of moving cars is estimated as **17.35** $_{col/\sec}$ and totally **50** frames have been used to model the whole background. The average of falsely detected pixels in our experiments is less than 10 percent.



(a) **(b)** **(c)** **(d)**

Figure 4. *(a-c) Three steps in background modeling (d) Completed background model*

(a) **(b)**

Figure 5. *(a, b) Original frames in the top row and detection results in the second row*

5. Conclusion

In this paper a novel approach for background modeling is introduced in order to detect and track moving objects. General works in this field, estimate each pixel value according to statistical models. In the proposed method, background is modeled by joining different patches. These patches of images are uncovered parts of background which are discovered from the difference of special nonconsecutive frames. These frames are selected according to a special analysis based on velocity and direction of moving objects in the original sequence. There is also a particular stage to model dynamic parts of the scene. The modeled background has been used to detect and track moving objects in highly textured indoor and outdoor scenes successfully. It is capable to handle dynamic changes of the scene because the model is updated for every new frame.

References

[1] Koller D, Weber J. Huang T. Malik J. Ogasawara G. Rao B. Russell S, ***"Towards robust automatic traffic scene analysis in realtime."*** in Proceedings of the 33rd IEEE Conference on Decision and Control (Cat. No.94CH3460-3). IEEE. Part vol.4, pp. 3776 – 3781, 1994.

[2] K.P. Karmann, A. von Brandt, ***"Moving Object Recognition Using an Adaptive Background Memory,"*** Signal Processing V: Theories and Application, Elsevier Science Publishers B.V, pp. 297-307, 1990.

[3] P. KaewTraKulPong and R. Bowden. ***"An Improved Adaptive Background Mixture Model for Real time Tracking with Shadow Detection,"*** In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01. pp. 149-158, Sept 2001.

[4] A. Elgammal, R. Duraiswami, D. Harwood, L.S. Davis, ***"Background and foreground modeling using nonparametric kernel density estimation for visual surveillance"***, In Proc. IEEE Computer Vision Lab, Vol. 90, pp. 1151- 1163, Jul 2002.

[5] B. Stenger, V. Ramesh, N. Paragios, F.Coetzee and J.M. Buhmann***. "Topology Free Hidden Markov Models: Application to Background Modeling,"*** In proc.8th Intl. Conf on Computer Vision, Vancouver, Canada, PP. 294-301, 2001.

[6] O. Tuzel, F. Porikli, P. Meer, ***"A Bayesian Approach to Background Modeling,"*** IEEE Workshop on Machine Vision for Intelligent Vehicles (MVIV), Vol. 3, pp. 58, June 2005.

[7] H. Grabner, P. M. Roth, M. Grabner, and H. Bischof. ***"Autonomous learning of a robust background model for change detection,"*** In Proc. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp 39, 46, 2006.

[8] A. Elgammal, D. Harwood, and L. Davis. ***"Non-parametric model for background subtraction,"*** In ECCV ,Dublin, Ireland, PART II, PP. 751–767, May 2000.

[9] C. Stauffer, W. E. L. Grimson, ***"Adaptive background mixture models for real-time tracking,"*** In Proceedings. 1999 IEEE ComputerSociety Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). IEEE Comput. Soc. Part Vol. 2, PP. 252-256, 1999.

[10] J.Rittscher, Jakarta, S.Joga and A.Black. ***"A Probabilistic background Model for Tracking,"*** In proc. Europe ,on Conf .on Computer Vision and Pattern Recognition, Fort Collins,CO,volume 2,1999.