

# Moving Object Segmentation using Enhanced Laplacian Thresholding Method for Video Surveillance

P.Vijayakumar,  
Assistant Professor and Head,  
Department of Computer Applications,  
Sri Jayendra Saraswathy Maha Vidyalyaya  
College of Arts & Science,  
Coimbatore – 641005

A.V.Senthil Kumar, Ph.D  
Director, MCA  
Hindustan College of Arts and Science  
Nava India,  
Coimbatore

## ABSTRACT

Identifying moving objects from a video surveillance is a fundamental and critical task in many computer vision applications. Video segmentation is one such application which has been studied for several decades and still remains a difficult problem for the system to automatically and accurately segment moving objects from video sequence with various backgrounds and global motions in real time. This paper proposes a new approach namely ELT (Enhanced Laplacian Thresholding) for real time video segmentation. The aim of this video enhancement method is to improve the visual appearance of the video and future-automated video processing like analysis, detection, segmentation, recognition, and surveillance for traffic, criminal justice systems and other areas that include analysis on large scale.

**Keywords**— Video Segmentation, Preprocessing, Laplacian Pyramid, Otsu's Method.

## 1. INTRODUCTION

Automated video surveillance systems are required to facilitate a wide range of applications in computer vision including human activity understanding, traffic monitoring and analysis, conservation of endangered species and so on, which is gaining substantial interests in the research community in recent years. The advantage of video surveillance automation over traditional closed circuit TV (CCTV)-based system lies in the fact that it is a self-contained system capable of automatic information extraction, detection of moving objects and tracking. Motion detection, which is the segmentation of moving objects in videos streams, is the relevant step in video surveillance system.

Segmentation techniques are sub classified into different approaches as: Amplitude thresholding or window slicing, Component labeling segmentation, Region based segmentation, Template matching, Texture segmentation, Boundary based segmentation.

Moving object detection is the main step in applications where the human detection is applied as in the video surveillance [1]. The goal of Object Detection is to establish similarity of objects and object parts between consecutive frames of video. The moving object detection process classifies the pixels into two categories namely as foreground pixels and background pixels. The foreground pixels are considered as the moving object and the background pixels are considered to be static background.

The detection of moving objects can involve a lot of problems that are associated with image processing like noise, sudden illumination changes, shadows, and few more to add. There are quite a large number of approaches that are proposed for the successful detection of the moving objects in a video sequence. The segmentation of moving objects can be classified into several techniques. In the region based approach, the image is partitioned into connected regions by grouping neighboring pixels of with same intensity levels. Neighboring regions are then combined based on some criterion involving either homogeneity or sharpness of region boundaries. The region based approach includes the traditional background subtraction and optical flow methods. The papers [1]-[3] are using the region based techniques for the detection of the moving objects. The various methods used in the region based approach are background subtraction [1], Markov Random Field [4], Change Detection Mask (CDM) [5], clustering [7] and modified statistical methods [3]. Video segmentation is used to identify the regions in a frame of video that are homogeneous with respect to any given parameter. Different features of homogeneity criteria leads to different segmentation of the same data. For example, segmentation with respect to color, texture, motion etc.

## 2. REGION BASED SEGMENTATION

Region based matching approaches define the velocity vector of an object moving across successive frames as the vector of displacements  $(i, j)$  that produces the best fit between image regions at different times. The best fit means that a distance measure between a region in a frame,  $I_t$ , and its possible location in the next frame,  $I_{t+1}$ , is minimum for the vector  $(i, j)$ , or a similarity measure such as cross-correlation is maximum. These methods are intuitively simple and relatively easy to implement, and the computational load is low since it has to be computed only for the active regions [8].

## 3. ELT METHOD

Surveillance and monitoring systems require the segmentation of all moving objects in a video sequence. Indeed, segmentation is a key step since it influences the performance of the other modules. It aims to detect objects of interest in the video stream by using their visual and motion properties. It plays a key role since it reduces the amount of information to be processed by higher processing levels, and object tracking, classification or recognition and locating the position of the targets serve as examples. Our proposed segmentation process uses the

following steps:1) Preprocessing 2) Laplacian pyramid 3) Otsu's method

### 3.1 Preprocessing

Basically several frames in video will not be of good quality, perfect size or with good brightness and contrast. Such frames when taken for further processing may result unexpectedly. So every frame should be enhanced in terms of quality. Similarly, all frames are to be preprocessed initially to make them perfect which includes the following.

#### 3.1.1 Scaling

In video processing, frame scaling is the process of resizing a frame. Scaling is a non-trivial process that involves an exchange between smoothness, efficiency and sharpness. the size of a frame is reduced or enlarged with bitmap graphics, the pixels that form the frame become increasingly visible making the frame appear "soft" if pixels are averaged, or jagged if not. With vector graphics, the trade-off may be in processing power for re-rendering the frame, which may be noticeable as slow re-rendering with still graphics, or slower frame rate and frame skipping in computer animation. One of the simpler ways of doubling its size is nearest-neighbor interpolation, replacing every pixel with four pixels of the same color. The resulting image is larger than the original, and preserves all the actual detail, but has jaggedness that is undesirable.

#### 3.1.2 Denoising

The idea behind a trimmed filter is to reject the noisy pixel from the selected 3 by 3 window. This paper suggests an Unsymmetric Trimmed Median Filter (UTMF) to remove the noises from frames in video. In this UTMF, the selected 3X3 window elements are arranged in either increasing or decreasing order. Then the pixel values 0's and 255's in the frame (i.e., the pixel values responsible for the salt and pepper noise) are removed from the frame. Then the median value of the remaining pixels is considered which is then used to replace the noisy pixel and that is called trimmed median filter as the pixel values 0's and 255's are removed from the selected window.

#### 3.1.3 Contrast Enhancement

Video enhancement techniques involve processing a frame to make it comfortable to look for the human viewers. Normally it is used for post-processing by modifying contrast or dynamic range or both in a frame. The aim of contrast enhancement process is to adjust the local contrast in different regions of the frame so that the details in dark or bright regions are brought out and revealed to the human viewers. Contrast enhancement is usually applied to input frames to obtain a superior visual representation of the frame by transforming original pixel values using a transform function of the form  $g(x; y) = T[r(x; y)]$  (1) where  $g(x; y)$  and  $r(x; y)$  are the output and input pixel values at frame position. Normally it is pleasing to enforce few restrictions on the transformation function  $T$  [20] for correct enhancement. Histogram equalization method is used in this paper for contrast enhancement. An attempt is made to alter the spatial histogram of a frame to match a uniform distribution closely. The core aim of this method is to attain a uniform distributed histogram by using the cumulative density function of the input frame. Using HE

has the following advantages (i) it suffers from the problem of ill-suited for retaining local detail due to its global treatment of the frame. (ii) Small-scale details that are often associated with the small bins of the histogram are removed. The disadvantage is that it is not a suitable property in some applications such as consumer electronic products, where brightness preservation is necessary to avoid annoying artifacts forms the disadvantage of this method. Usually the equalization result is an undesired loss of visual data, quality and intensity scale [6].

In ELT method, background frame is set as the reference frame and the difference from each frame is taken. This process helps to identify the discontinuity of the frames which is done with the help of the laplacian operator.

### 3.2 Laplacian Pyramid

The Laplacian Pyramid is a decomposition of the original image into a hierarchy of images such that each level corresponds to a different band of image frequencies. This is done by taking the difference of levels in the Gaussian pyramid. For image  $I$  the Laplacian pyramid  $L(I)$  is:

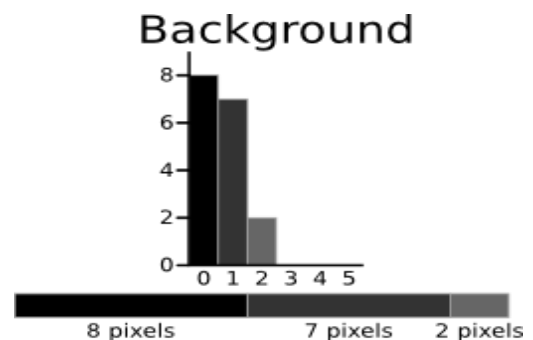
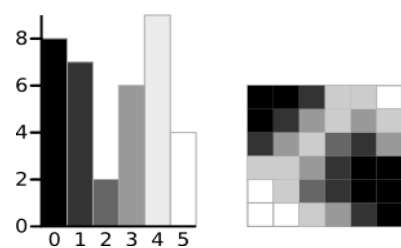
$$L_i = G_i - \text{EXPAND}(G_{i+1})$$

The original image can always be reconstructed from its Laplacian pyramid by reversing this process. The particular frame that belongs to background frame or foreground frame based on threshold value for that process is found next for which Otsu's method is used.

### 3.3 Otsu's Method

Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold, i.e., the pixels which may either fall in the foreground or in the background of the frame Finding the threshold value where the sum of foreground and background spreads is at its minimum forms the aim.

A 6-level grayscale image and its histogram



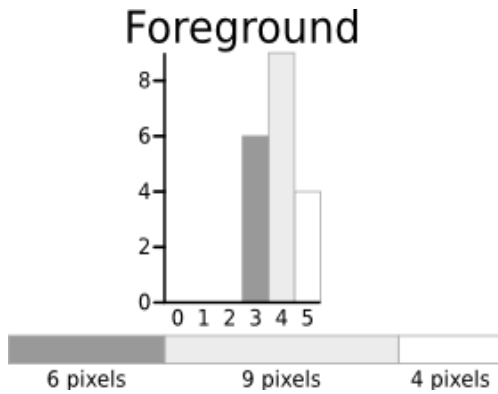
$$\text{Weight } W_6 = \frac{8+7+2}{36} = 0.4722$$

$$\text{Mean } \mu_6 = \frac{(0 \times 8) + (1 \times 7) + (2 \times 2)}{17} = 0.6471$$

$$\text{Variance } \sigma_6^2 = \frac{((0 - 0.6471)^2 \times 8) + ((1 - 0.6471)^2 \times 7) + ((2 - 0.6471)^2 \times 2)}{17}$$

$$= \frac{(0.4187 \times 8) + (0.1246 \times 7) + (1.8304 \times 2)}{17}$$

$$= 0.4637$$



$$\text{Weight } W_7 = \frac{6+9+4}{36} = 0.5278$$

$$\text{Mean } \mu_7 = \frac{(3 \times 6) + (4 \times 9) + (5 \times 4)}{19} = 3.8947$$

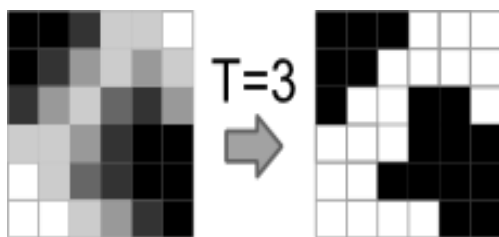
$$\text{Variance } \sigma_7^2 = \frac{((3 - 3.8947)^2 \times 6) + ((4 - 3.8947)^2 \times 9) + ((5 - 3.8947)^2 \times 4)}{19}$$

$$= \frac{(4.8033 \times 6) + (0.0097 \times 9) + (4.8864 \times 4)}{19}$$

$$= 0.5152$$

$$\text{Within Class Variance } \sigma_w^2 = W_6 \sigma_6^2 + W_7 \sigma_7^2 = 0.4722 \times 0.4637 + 0.5278 \times 0.5152$$

$$= 0.4909$$



Result of Otsu's Method

The algorithm will be demonstrated using the simple 6x6 image shown below. The histogram for the image is shown next to it. To simplify the explanation, only 6 grayscale levels are used. The calculations for finding the foreground

and background variances (the measure of spread) for a single threshold are now shown where the threshold value is 3 in our case. The next step is to calculate the 'Within-Class Variance'. This step simply sums the two variances multiplied by their associated weights. The final value is calculated as the 'sum of weighted variances' for the threshold value 3 and the same calculation requires it to be performed for all the possible threshold values 0 to 5. It can be found that for the threshold equal to 3, as well as the one used in the example, has the lowest sum of weighted variances. Hence, this forms the final selected threshold. All pixels with a level below 3 are background, and those with a level equal to or greater than 3 are considered as foreground [9].

### ALGORITHM

**STEP 1:** Get the input video which is to be processed as X.

**STEP 2:** After the video is fetched, retrieve each frame present in the video separately as images (A).

**STEP 3:** For the purpose of scaling, every frame is subjected to Bicubic Interpolation using the formula,

$$B(r, c) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} r^i c^j$$

Where  $a_{ij}$  is the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column pixel of the 3 x 3 window considered. The process of bicubic interpolation is done to every frame by considering each frame into several 3 x 3 windows.

**STEP 4:** The video fetched may contain some noises in it so after the process of Bicubic interpolation the frames are filtered using an unsymmetrical trimmed median filter.

**STEP 5:** In order to enhance the frames and its quality its contrast are improved using histogram equalization. It is performed using the formula given below,

$$H(p = i) = \frac{N_i}{N}$$

Where,  $H$  is the probability that the pixel  $p$  of a frame is equal to the level (gray level)  $i$ .  $N_i$  is the number of occurrence of the selected gray level and  $N$  is the total number of pixel.

**STEP 6:** Consider the first frame of the video as the reference frame and calculate the difference between every frame and the reference frame.

**STEP 7:** Discontinuities present in the difference image  $D_1, D_2, D_3, D_4, \dots$ , is removed by using the laplacian pyramid. Each difference image is subjected to laplacian pyramid using the formula,

$$E(j) = D_j - \text{Expand}(D_{j+1})$$

Where  $j$  is the frame considered,  $E$  is the discontinuity removed output of  $j^{\text{th}}$  frame,  $L_j$  and  $L_{j+1}$  are the various laplacian levels considered for the  $j^{\text{th}}$  frame.

**STEP 8:** Find the threshold value of each frame to create the segmented video frames using OTSU thresholding

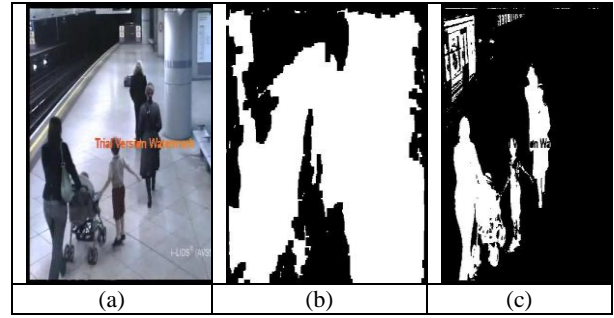
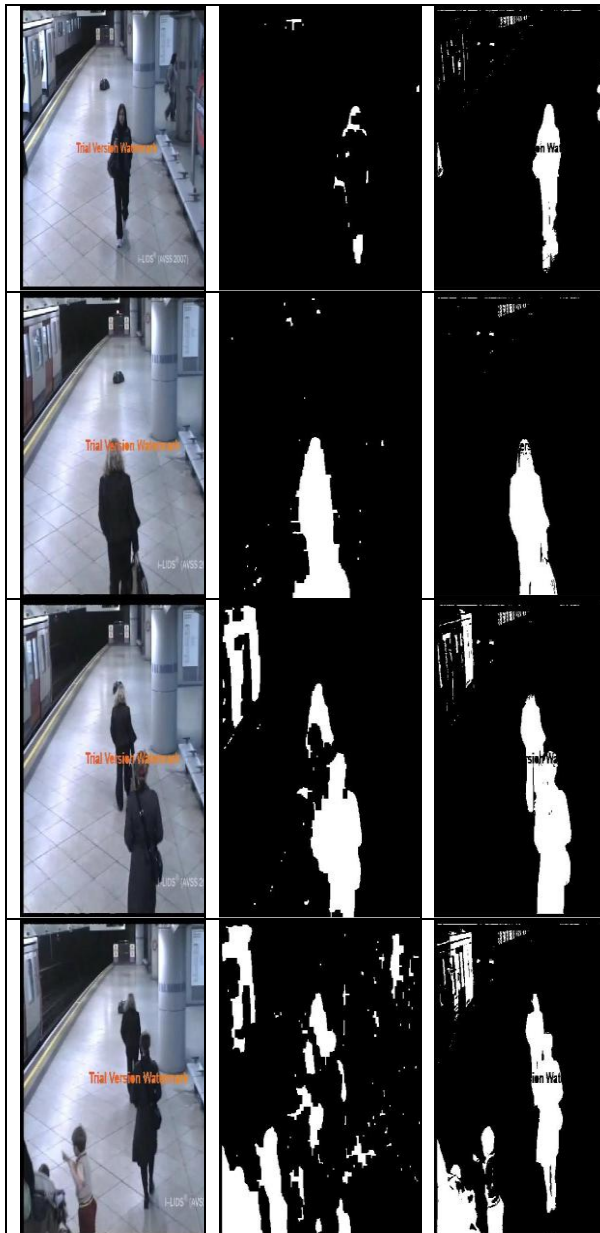
$$\sigma_{\omega}^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t)$$

Where Weights  $\omega_i$  are the probabilities of the two classes separated by a threshold  $t$  and  $\sigma_i^2$  variances of these classes.

**STEP 9:** Repeat the steps from 6 to 8 till the last frame of the video.

**STEP 10:** Finally the motion object is obtained from the segmented video frames.

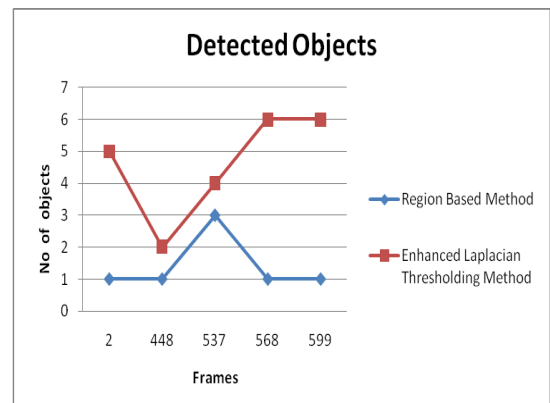
#### 4. EXPERIMENTAL RESULTS



**Fig.1. Comparison on Video sequence using Region based segmentation and proposed method (a)Input Video Frames (b) Results from Region based segmentation Method (c) Results from ELT Method**

This paper proposed effective video segmentation approach named ELT method. ELT method was implemented in Matlab with Intel core2duo processor, 4GB RAM and tested with different video datasets and it delivers good performance in many different situations. From the experimental results, first 600 frames are taken from the video. The results are shown in Fig.1. In Fig.1. first row (a) represents the original frames second row (b) represents region based segmentation results and third row (c) shows the results of proposed segmentation.

In this segmentation method, the background frame is considered as an initial frame to compare with other input frames. The experimental result, Lablacian pyramid method is used to detect the discontinuities of the frames and using Otsu's Method to set the threshold values and determine whether the particular frame is related to background or foreground based on the threshold value, and the combination of these two was used to generate proposed method. Unlike other algorithms, here Block matching method is used instead of pixel based comparison of each frames. This reduces the computational time and mismatching of the pixel was also avoided. Thus noises can be reduced and background difference mask is considered instead of change detection mask. This also improves the quality of the moving object detection in video surveillance.



**Fig.2. Detcted Objects**

The Graph shows the performance and accuracy of ELT method in which X-axis represents the number of moving objects and Y-axis represents the particular frame in the video sequence. From the above graph it is seen that in

second frame the region based segmentation method detects 1 moving object and our ELT method detects 5 objects, in the 448th frame the region based segmentation method detects 1 moving object and our ELT method detects 2 objects, in 537th frame region based segmentation method detects 3 moving objects and our ELT method detects 4 objects. In the 568th frame, region based segmentation method detects 1 moving object and our ELT method detects 6 objects, and in 599th frame region based segmentation method detects 1 moving object but our ELT method detects 6 objects .

## 5. CONCLUSION

A new method is proposed to automatically segment multiple moving objects in a video sequence. Different types of videos were collected and the algorithms were applied which segmented the moving objects very accurately to compare that with region based segmentation method. The algorithms proposed here, saved the hardware resources and also reduced the processing time for each frame. This method automatically removes the noise from each frame. It can be concluded that for the moving object segmentation, this proposal achieves comparable results. This suggests that the proposed methodology is reliable for a fully automatic segmentation of multiple moving objects in video surveillance system and also this is encouraging for the extension of this framework to denser environments such as crowd of moving people.

## 6. REFERENCES

- [1] B.S.M. Madhavi and M.V.Ganeswara Rao. —A fast and reliable motion human detection and tracking based on background subtraction, I IOSR Journal of Electronics and Communication Engineering (IOSRJECE), ISBN: 2278-2834 vol 1, Issue 1, PP 29-34, May 2012.
- [2] Z.Zhu and Y.Wang. —A hybrid algorithm for automatic segmentation of slowly moving objects, I Int. J. Electron. Commun. (AEÜ) 66 (2012) pp.249–254, 2012.
- [3] S.Vahora, N. Chauhan and N. Prajapati. —A Robust Method for Moving Object Detection Using Modified Statistical Mean Method, International Journal of Advanced Information Technology (IJAIT) vol. 2, no.1, February 2012.
- [4] W.Zeng and W.Gao, —Accurate moving object segmentation by a hierarchical region labeling approach, Proceedings of IEEE conference on Acoustics, Speech and Signal Processing, May 2004.
- [5] B. N. Subudhi and P. K. Nanda —Detection of Slow Moving Video Objects Using Compound Markov Random Field Model, I TENCON 2008-2008 IEEE region conference, Nov 2008.
- [6] A.K.Jain —Fundamentals of digital image processing, Prentice-Hall PvtLtd. 3rd edition 1997.
- [7] Qing-Zhong Li, Dong-Xiao He and Bing Wang, —Effective Moving Objects Detection Based on Clustering Background Model for Video Surveillance, Proceedings of the 2008 IEEE International Conference on Image and Signal Processing - CISP, pp. 656-660, 27-30 May, 2008.
- [8] Manya V.Afonso, Jacinto C.Nascimento and Jorge S.Marques – Automatic estimation of Multiple Motion Fields From Video Sequences Using a Region Matching Based Approach, IEEE Transactions on Multimedia, Vol.16, NO.1, January 2014
- [9] Vijay Jumb, Mandar Sohani, Avinash Shrivastava —Color Image Segmentation Using K-Means Clustering and Otsu's Adaptive Thresholding, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-3, Issue- 9, February 2014
- [10] N. Indhumadhi, G. Padmavathi—Enhanced Image Fusion Algorithm Using Laplacian Pyramid and Spatial frequency Based Wavelet Algorithm, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-5, November 2011