Object Detection in Video Streams

Sandhya S Deore*

**Assistant Professor Dept. of Computer Engg., SRES COE Kopargaon*

*sandhya.deore@gmail.com

ABSTRACT

Object Detection is the most challenging area in video surveillance that is one part of Multimedia mining. Its task is to detect the object of interest in video. The wide application of this is in field like security, criminology etc. In this paper Background elimination technique is used to identify the object. Background eliminination technique uses the frame difference and background elimination method for object detection. Post processing is used to reduce the noise.

Keywords- *Background elimination, Background object, Foreground object, Frame difference ,Post processing.*

1. INTRODUCTION

Multimedia mining is the large collection of multimedia objects. Multimedia mining can be classified into three major subcategories as Image Mining, Audio Mining, and Video mining [1]. Among all these types, video mining is the most challenging one, as it combines all other media information into a single data stream. Recently, due to the decreasing cost of storage devices, higher transmission rates and improved compression techniques, digital video becomes available at an ever increasing rate. Videos contain a great deal of latest and useful knowledge, so people urgent demands, a way that can fast mining latest and useful patterns from videos.

Video mining is a process of discovering interesting events in the video even without any prior knowledge about the events. The objective of this is to extract significant objects, characters and scenes in a video by determining their frequency of re-occurrence. Segmentation of moving objects in video processing plays an important role in image sequence processing and analysis. The ideal goal of segmentation is to identify the semantically meaningful components of an image and grouping the pixels belonging to such components. Segmenting the video into foreground and background can also reduce the data rate required to transmit live video as it may not be necessary to transmit the background as frequently as the foreground vehicles [9]. Usually video stream contains the moving objects, so the motion segmentation is the most common method for detecting regions corresponding to moving (dynamic) objects such as vehicles, humans etc. Once the moving objects are detected or extracted out, they can serve for varieties of purposes.

The goal of automated surveillance system is to assist the human operator in scene analysis and event classification by automatically detecting the objects and analyzing their behavior using computer vision, pattern recognition and signal processing techniques. An automated surveillance system attempts to detect, recognize and track objects of interest from video obtained by cameras along with information from other sensors installed in the monitored area. The aim of an automated visual surveillance system is to obtain the description of what is happening in a monitored area and to automatically take appropriate action like alerting a human supervisor, based on the perceived description [6]. Manually reviewing the large amount of data they generate is often impractical. Thus, algorithm for analyzing video which require little or no human input is a good solution.

Our system uses a video that mostly contains the humans with constant background. We consider video segmentation with initial frame difference, object detection using background elimination for moving objects in video database.

Organization: remaining paper is organized as section 2 describes the system architecture, algorithms and parameters for implementation, section 3 describes conclusion.

2. RELATED WORK

There are many methods for detecting the moving objects. Their uses depends on different assumptions for e.g. statistical models of the background, minimization of Gaussian differences, minimum and maximum values, adaptively or a combination of frame differences and statistical background models.

Yilmaz A. et.al, in paper [2] describes the various representations of objects. They also discussed the various categories for object detection and the methods included in that categories. Lipton et al. [3] detect moving targets in real video streams using temporal differencing. Meyer et al. [4] compute the displacement vector field to initialize a contour based tracking algorithm, called active rays.

Optical-flow-based methods can be used to detect independently moving objects even in the presence of camera motion. However, most flow computation methods are computationally complex and very sensitive to noise, and cannot be applied to video streams in real time without specialized hardware. Most works however rely on statistical models of the background, assuming that each pixel is a random variable with a probability distribution estimated from the video stream. For example, the Pfinder system ("Person Finder") uses a Gaussian model to describe each pixel of the background image [5]. Pravin Kumar et.al, in paper [6] used the background subtraction method for object detection

3. SYSTEM ARCHITECTURE

Now a day, there are many real time applications like news channels, shopping malls where the camera is at fixed location for surveillance. Vibha L et al. introduced a new algorithm [7] in which the stationary (fixed) background videos are considered. Fig.1 shows the architectural flow of proposed algorithm.

Step performed in background elimination is as follows: Video is read and converted them into number of frames. Then these frames are converted into Gray frames. The second stage is frame difference between frames at regular interval. In third stage the pixels having same values are eliminated. In fourth stage the post processing is applied on the previous results and the final stage for object detection.

2.1 Frame Difference

Frame differencing is a segmentation technique in a video stream where it checks difference between consecutive frames pixel by pixel and if the pixel values are changed, then it is considered that there is something changing in the frame (i.e. something moving).

Frame differencing makes use of the pixel-wise differences between two or three consecutive frames in an image sequence to extract moving regions. Frame differencing is very adaptive to dynamic environments. This may become computationally complex when video having slow moving objects. Hence, the difference between the frames is taken at regular intervals. Hence if there are n frames, then we will get (n/k) frame differences (FD), otherwise videos having fast moving objects then we can decrease the interval between the frames. Frames are segmented by the following general equation:

$|\text{frame } I_c - \text{frame } I_b| > T$

Where,

frame I_c – Current image.
frame I_b – Background image.
T – Threshold value

When the pixel difference between the two frames is above the threshold value T, then that pixel is considered as foreground pixel.

Algorithm:

1. Initialize an empty array.
2. Take two variables such as p and k.
3. Initialize p=1 and k=2
4. For i=1 to no_frames in step of p with interval 2 frames do the following steps
   i. Convert movie structure stored in array into frames.
   ii. Convert frames obtained in previous step in double array .
   iii. Convert frames from RGB to Gray format.
   iv. Compute difference between frames p and k at regular interval.

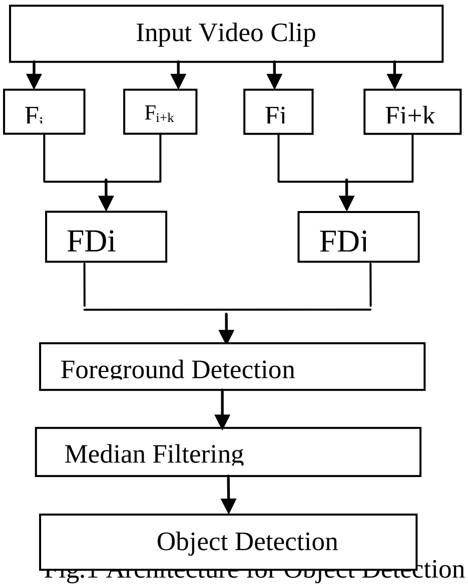


Fig.1 Architecture for Object Detection

Fig 2. Shows the result of frame difference on some test sequence.

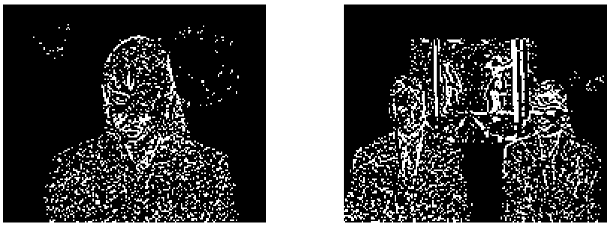
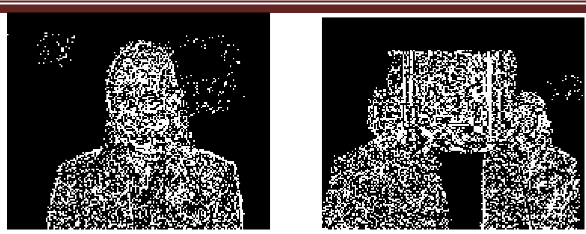


Fig 2. (a) Frame difference between 54th and 56th (b) Frame difference between 2nd and 4th

2.2 Background Elimination

Once the frame differences are computed the pixels that belong to the background region will have a value almost equal to zero, as the background is assumed stationary. Many a times because of camera noise, some of the pixels belonging to the background region may not tend to zero. These values are set to zero by comparing any two frame differences, say, FDi and FDj. Thus, the background region is eliminated and only the moving object region will contain non-zero pixel values. The images obtained after background elimination is as shown in the Fig.3.



(a) Akiyo (b) News

Fig. 3 Background Elimination

Algorithm:

- After computing frame difference, pixels having same value considered as background.
- So set those pixels value as zero.
- Therefore background is eliminated.
- Moving object will contain the non-zero value.

2.3 Post Processing

Many a times due to camera noise and irregular object motion, there always exists some noise regions both in the object and background region. Moreover the object boundaries are also not very smooth; hence a post processing technique is required. Most of the post processing techniques are applied on the image obtained after background elimination. The current algorithm uses Median filter [8] which is the best-known order-statistics filter.

The median filter is a spatial filter, which replaces the center value in the window with the median of all the pixel values in the window. Median filter is a non-linear signal enhancement technique for smoothing of signals, suppression of impulse noise, and preservation of edges. In the one-dimensional case it consists of a sliding window of an odd number of elements and replaces the center sample by the median of the samples in the window.

$$Y(n) = \text{med}\,(y_{n-k}, y_{n-k}, \ldots, y_{n-1}, x_n, \ldots x_{n+k}) \qquad (3.3)$$

This filter replaces the value of a pixel by the median of the gray levels in the neighbourhood of that pixel. The general equation for median filter is given in equation 3.4

$$\hat{f}(x, y) = median\{g(s,t)\}$$

This filter replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel.

(a) Akiyo (b) News

Fig. 4 Post Processing

$$f(x, y) = \text{median}\{g(s, t)\} \quad (1)$$

After applying the median filter, the resulting image is converted into a binary image. The morphological Opening technique [8] is applied on this binary image. The Opening of A by B is simply erosion of A by followed by dilation of the result by B. This can be given as

$$A \circ B = (A \ominus B) \oplus B \quad (2)$$

Here, A is the image and B is a structuring element. After applying the above explained pre-processing techniques, the new image obtained is as shown in the Fig.4.

2.4 Object detection

The image obtained after the pre-processing step has relatively less noise, so, the background area is completely eliminated. Now, if the pixel values of this image are greater than zero, then, those pixels are replaced by the pixels of the original frame. This process identifies the moving object as shown in Fig.5.

(a) Akiyo (b)

Fig 5. Object Detection

Given a video clip in the format of QCIF (176 x 144) for object identification. The objectives are:

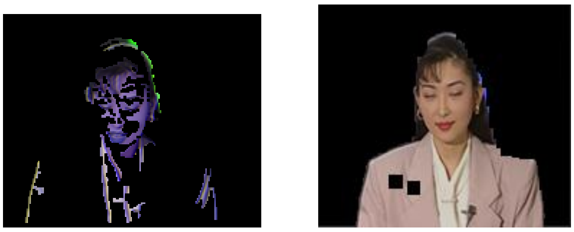
(a) To detect a moving object using the concept of background elimination technique.

(b) To improve the clarity of the moving object.

4. PERFORMANCE ANALYSIS

It is observed that the clarity of the image obtained using our proposed algorithm is much clearer than the existing algorithm. Simulation was carried out on standard QCIF sequences.

The results obtained from proposed algorithm are compared with those of background registration method. The outputs obtained through two different techniques on the QCIF sequence video are as shown in Fig.6.



(a) (b)

Fig.6 a) result from background registration b) result from Elimination

5. CONCLUSION

In this paper, we propose an efficient algorithm for detecting a moving object using background elimination technique. Initially we compute the frame differences (FD) between frames F_i and F_{i+k}. The frame differences obtained are then compared with one another which help in identifying the stationary background image. The moving object is then isolated from the background. In the post processing step, the noise and shadow regions present in the moving object are eliminated using a morphological gradient operation that uses median filter without disturbing the object shape. This could be used in real time applications involving multimedia communication systems. The experimental results obtained indicate that the clarity of the image obtained using background elimination technique is much better.

REFERENCES

[1] Manjunath T.N, Ravindra S Hegadi, Ravikumar G K," A Survey on Multimedia Data Mining and Its Relevance Today", IJCSNS International Journal of Computer Science and Network Security, vol. 10, no. 11, November 2010.

[2] Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking:A survey. ACM Comput. Surv. 38, 4, Article 13 (Dec. 2006), 45 pages.

[3] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proc. IEEE Workshop Applications of Computer Vision*, 1998, pp. 8–14.

[4] D. Meyer, J. Denzler, and H. Niemann, "Model based extraction of articulated objects in image sequences for gait analysis," in *Proc. IEEE Int. Conf. Image Processing*, 1998, pp. 78–81

[5] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland,"Pfinder: Real-Time Tracking of the Human Body", IEEE Transactions on pattern analysis and machine intelligence, vol. 19, no. 7, july 1997

[6] Praveen Kumar, Ankush Mittal and Padam Kumar , " Study of Robust and Intelligent Surveillance in Visible and Multimodal Framework", Department of Electronics and Computer Engineering, Indian Institute of Technology, Roorkee, India 247667

[7] Vibha L., Chetana Hegde et. al., "Dynamic Object Detection, Tracking and Counting in Video Streams for Multimedia Mining", *IAENG International Journal of Computer Science*, pp.35, August 2008.

[8] Damien LEFLOCH, report on "Real-Time People counting system using Video Camera"

[9] Tun-Yu Chiang, Wilson Lau, "Segmentation of Vehicles in Traffic Video",EE392j Project Report", Stanford University