Dynamic Object Detection, Tracking and Counting in Video Streams for Multimedia Mining

Article *in* IAENG International Journal of Computer Science · January 2008

Source: DOAJ

L. Vibha
B. N. M. Institute of Technology

Chetana Hegde
RNS Institute of Technology

P. Deepa Shenoy
University Visvesvaraya College of Engineering

Venugopal K R
University Visvesvaraya College of Engineering

Dynamic Object Detection, Tracking and Counting in Video Streams for Multimedia Mining

Vibha L, Chetana Hegde, P Deepa Shenoy, Venugopal K R, L M Patnaik *

Abstract—**Video Segmentation is one of the most challenging areas in Multimedia Mining. It deals with identifying an object of interest. It has wide application in the fields like Traffic surveillance, Security, Criminology etc. This paper initially proposes a technique for identifying a moving object in a video clip of stationary background for real time content based multimedia communication systems and discusses one application like traffic surveillance. We present a framework for detecting some important but unknown knowledge like vehicle identification and traffic flow count. The objective is to monitor activities at traffic intersections for detecting congestions, and then predict the traffic flow which assists in regulating traffic. The algorithm for vision-based detection and counting of vehicles in monocular image sequences for traffic scenes are recorded by a stationary camera. Dynamic objects are identified using both background elimination and background registration techniques. Post processing techniques are applied to reduce the noise. The background elimination method uses concept of least squares to compare the accuracies of the current algorithm with the already existing algorithms. The background registration method uses background subtraction which improves the adaptive background mixture model and makes the system learn faster and more accurately, as well as adapt effectively to changing environments.**

Keywords—Background elimination, Frame difference, Object identification, Background registration, Camera calibration, Vehicle tracking.

1 Introduction

Video mining can be defined as the unsupervised discovery of patterns in audio visual content. The motivation for such discovery comes from the success of data mining techniques in discovering non-obvious patterns. In video mining we can discover interesting events in the video even without any prior knowledge about the events. The objective of video mining is to extract significant objects, characters and scenes in a video by determining their frequency of re-occurrence. Some of the basic requirements needed for extracting information in a video mining technique are i) It should be as unsupervised as possible. ii) It should have as few assumptions about the data as possible. iii) It should be computationally simple. iv) It should discover interesting events.

Segmentation can be an extremely easy task if one has access to the production process that has created the discontinuities. For example, the generation of a synthetic image or of a synthetic video implies the modelling of the 3-D world and of its temporal evolution. However, if the segmentation intents to estimate what has been done during the production process, its task is extremely difficult and one has to recognize that the state of the art has still to be improved to lead to robust segmentation algorithms that are able to deal with generic images and video sequences.

Automatic detecting and tracking vehicles in video surveillance data is a very challenging problem in computer vision with important practical applications, such as traffic analysis and security. Video cameras are a relatively inexpensive surveillance tool. Manually reviewing the large amount of data they generate is often impractical. Thus, algorithms for analysing video which require little or no human input is a good solution. Video surveillance systems are focussed on background modelling, moving object classification and tracking. The increasing availability of video sensors and high performance video processing hardware opens up exciting possibilities for tackling many video understanding problems, among which vehicle tracking and target classification are very important. Most occurrences of moving objects in our data are pedestrians and vehicles. Traffic management and information systems depend mainly on sensors for estimating the traffic parameters. In addition to vehicle counts, a much larger set of traffic parameters like vehicle classifications, lane changes, etc.,

*Manuscript received July 19, 2008. Vibha L is with Bangalore Institute of Technology, Bangalore, India (Research scholar of MGR University Chennai) (email: vibhal1@rediffmail.com). Chetana Hegde is with RNS Institute of Technology, Bangalore, India. P Deepa Shenoy and Venugopal K R are with University Visvesvaraya College of Engineering, Bangalore University, Bangalore, INDIA - 560001. L M Patnaik is with Defence Institute of Advanced Technology, Pune, India.

can be computed. Our system uses a single camera mounted usually on a pole or other tall structure, looking down on the traffic scene. The system requires only the camera calibration parameters and direction of traffic for initialization.

The cameras allow operators to monitor traffic conditions visually. As the number of cameras increase, monitoring each of them by operators becomes a difficult task hence videos are recorded and such the videos are usually only monitored after an event of interest (e.g. an accident) has been known to occur within a particular cameras field of view. With suitable processing and analysis it is possible to extract a lot of useful information on traffic from the videos, e.g., the number, type, and speed of vehicles using the road. To perform this task segmenting the video into foreground objects of interest (the vehicles) and the background (road, trees) is required. Advantage of segmenting the video into foreground and background reduces the data rate transmission time of live videos as it is redundant to transmit the background as frequently as the foreground vehicles.

Motivation: Vehicle detection and counting is important in computing traffic congestion and to keep track of vehicles that use state-aid streets and highways. Even in large metropolitan areas, there is a need for data about vehicles that use a particular street. A system like the one proposed here can provide important data for a particular design scenario. Magnetic loop detectors are currently used to count vehicles which pass over them, but vision-based video monitoring systems offer many more advantages. Surveillance and video analysis provide quick practical information resulting in increased safety and traffic flow. For example, objects are defined as vehicles moving on roads. Cars and buses can be differentiated and the different traffic components can be counted and observed for violations, such as lane crossing, vehicles parked in no parking zones and even stranded vehicles that are blocking the roads. Moreover cameras are much less disruptive to install than loop detectors.These were the main factors that motivated us to design the current automated system.

Contribution: A system has been developed to track and count dynamic objects efficiently. Intelligent visual surveillance for road vehicles is a key component for developing autonomous intelligent transportation systems. The algorithm does not require any prior knowledge of road feature extraction on static images. We present a system for detecting and tracking vehicles in surveillance video which uses a simple motion model to determine salient regions in a sequence of video frames. Similar regions are associated between frames and grouped to form the background. The entire process is automatic and uses computation time that scales according to the size of the input Video sequence. We consider image/video segmentation with initial background subtraction, object tracking, and vehicle counting, in the domain of traffic monitoring over an intersection.

Organization: The remainder of the paper is organised as follows Section 2 gives the overview of the related work. Section 3 describes the architecture and modelling for background elimination and background registration. In section 4 the algorithms for identifying background, detection and counting of vehicles is presented. Parameters for implementation and performance are analysed in section 5. Section 6 contains the conclusions.

2 Related Work

A brief survey of the related work in the area of video segmentation and traffic surveillance is presented in this section. Video segmentation helps in the extraction of information about the shape of moving object in the video sequences. Sikora T. [1] used this concept for intelligent signal processing and content-based video coding. Here an image scene consists of video objects and the attempt is to encode the sequence that allows separate decoding and construction of objects. Nack et al., [2] and Salembier et al., [3] have discussed Multimedia content description related to the generation of region based representation with respect to MPEG-4 and MPEG-7.

Video segmentation algorithms can be broadly classified into two types based on their primary criteria for segmentation. Wang D. in [4] proposes a technique for unsupervised video segmentation that consists of two phases i.e. initial segmentation and temporal tracking. Y. Yokahama et al. in [5] discusses concept of initial segmentation as applied to the first frame of the video sequence, which performs spatial segmentation, and then partitions the first frame into homogeneous regions based on intensity. Motion estimation is then computed for determining the motion parameters for each region, and finally motion-based region merging is performed by grouping the regions to obtain the moving objects. L, Wu et al., [6] explains how temporal tracking is performed in detail after initial segmentation.

P. Salembier [7] found better results using spatial homogeneity as the primary criteria, which incorporates luminance and motion information simultaneously. The procedure includes the steps like joint marker extraction [8], [9], boundary decision and motion-based region fusion. Spatio-temporal boundaries are then decided by the watershed algorithm. Choi et al., [10] discusses Joint similarity method for the same purpose and finally, motion-based region fusion is used for eliminating the redundant regions. Initially filters are used to simplify the image and then Watershed algorithm is applied for boundary detection [11]. Later the motion vector is computed using motion estimation and regions with similar motion are merged together to constitute the final object region. As watershed algorithm is being used they generate object boundaries which are more efficient and precise than any other methods. Aach T. et al., [12], discusses the change detection method which is used as the primary segmentation criteria in many applications. The major issue here is to guarantee robust detection in the results, in presence of noise. Many shortcomings are overcome by using Markov random field based on refining method. The position and shape of the moving object is determined using the frame difference concept, followed by a boundary fine-tuning process based on temporal information. Algorithms that deal with spatial domain processing first, without knowing much regarding the motion information will waste much of the computing power in segmenting the background.

Neri et al., [13] describes a solution to eliminate the uncovered background region by applying motion estimation on regions with significant frame difference. The object in the foreground is then identified when a good match is found between two frame differences. The remaining region is then discarded as unwanted areas. Stauder et al., [14] considers the effect of shadow of an object in the background region which affects the output in change detection based approach.

Related work in the area of traffic surveillance is discussed here. Koller et al., [15] , [16], has described algorithms that uses an offline camera calibration step to aid the recovery of the 3D images, and it is also passed through Kalman Filter to update estimates like location and position of the object. Puzicha J. et al., in [17] uses concept of Bayesian technique for image segmentation based on feature distribution. Here a statistical mixture model for probabilistic grouping of distributed data is adopted. It is mainly used for unsupervised segmentation of textured images based on local distributions of Gabor coefficients. Chen et al.,[18],[19] have addressed the issues regarding unsupervised image segmentation and object modelling with multimedia inputs, to capture spatial and temporal behaviour of objects for traffic monitoring. D. Beymer et al.,[20] proposes a real time system for measuring traffic parameters that uses a feature-based method along with occlusion reasoning for tracking vehicles in congested traffic areas. Here instead of tracking the entire vehicle, only sub features are tracked. This approach is however computationally expensive. In [21] tracking and counting pedestrians using a single camera is proposed. Here the image sequences are segmented using background subtraction and the resulting regions are connected then grouped and together as pedestrians and tracked. A. J. Lipton et al., [22] describes vehicle tracking and classification system where one identifies moving objects as vehicles or humans, but however it does not classify vehicles into different classes. Gupta S. et al., in [23] describes algorithms for vision-based detection and classification of vehicles in monocular image sequences of traffic scenes are recorded by a stationary camera. Processing is done at three levels: raw images, region level, and vehicle level. Vehicles are modelled as rectangular patterns with certain dynamic behaviour.

Dailey et al., [24] presents the background subtraction and modelling technique that estimates the traffic speed using a sequence of images from an uncalibrated camera. The combination of moving cameras and lack of calibration makes the concept of speed estimation a challenging job. In [25] Grimson et al., analyses a vision based system that monitors activities in a site, over a period of time using sensor networks.

3 Architecture and Modelling

In many real-time applications like video conferencing, the camera is fixed. Some techniques proposed in paper [12] use global motion estimation and comparison to compensate the change in background due to camera motion. In the present algorithm, we assume that the background is stationary for the video clips considered. The architecture and modelling of the proposed algorithm is shown in Figure 1.

The flow of the algorithm for background elimination is as follows: A video clip is read and it is converted to frames. In the first stage difference between frames are computed i.e. Fi and Fi+k. In the next stage these differences are compared, and in the third stage pixels having the same values in the frame difference are eliminated. The fourth phase is the post processing stage executed on the image obtained in third stage and the final phase is the object detection.

3.1 Frame Difference

Frame differences are computed by finding the difference between consecutive frames but this will introduce computational complexity in case the video clips having slow-moving objects. Moreover this algorithm assumes a stationary background. Hence the difference between the frames at regular intervals (say, some integer k) is considered. If there are n frames, then we will get (n/k) frame differences (FD). The frame difference

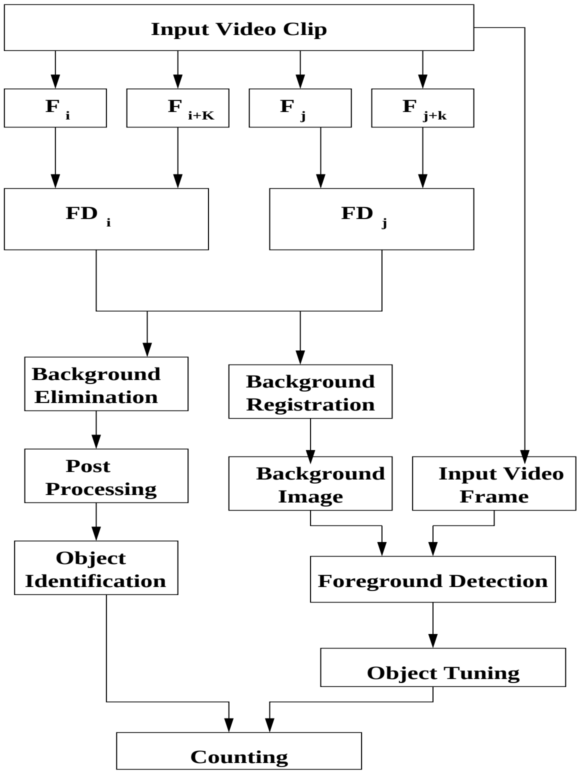


Figure 1: Architecture for Object Identification

follows Gaussian distribution as indicated in equation (1)

$$p(FD) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(FD-\mu)^2}{2\sigma^2}) \qquad (1)$$

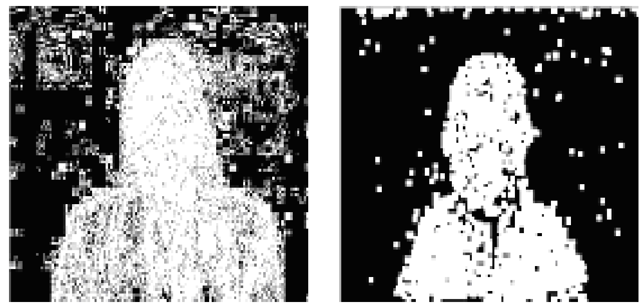
Here, is the mean of FD and is the standard deviation of FD. The frame differences of some test sequences are as shown in Figure 2.

(a) Difference between 30th and 35th frames of Akhiyo (b) Difference between 12th and 15th frames of Claire

Figure 2: Frame Difference

3.2 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 Figure 3.



(a) Image of Moving Hands (b) Image of Claire

Figure 3: After Background Elimination

3.3 Background Registration

A general tracking approach is to extract salient regions from the given video clip using a learned background modelling technique. This involves subtracting every image from the background scene and thresholding the resultant difference image to determine the foreground image. Stationary pixels are identified and processed to construct the initial background registered image.

Here we go by the fact that vehicle is a group of pixels that move in a coherent manner, either as a lighter region over a darker background or vice versa. Often the vehicle may be of the same colour as the background, or may be some portion of it may be camouflaged with the background, due to which tracking the object becomes difficult. This leads to an erroneous vehicle count.

3.4 Foreground Detection (Object Tracking)

Most vision based traffic monitoring system must be capable of tracking vehicles through the video sequence. Tracking helps in eliminating multiple counts in vehicle counting applications and it also helps in deriving useful information while computing vehicle velocities. Tracking information can be used to refine the vehicle type and also to correct errors caused due to occlusions. After registering the static objects the background image is subtracted from the video frames to obtain the foreground dynamic objects. Post processing is performed on the foreground dynamic objects to reduce the noise interference.

3.5 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. Initially, order-statistics filters are used, which are the spatial filters and whose response is based on ordering (ranking) the pixels contained in the image area encompassed by the filter. The response of the filter at any point is then determined by the ranking result. The current algorithm uses Median filter which is the best-known order-statistics filter. This filter replaces the value of a pixel by the median of the gray levels in the neighbourhood of that pixel. The formula used is

$$\hat{f}(x, y) = median\{g(s, t)\} \tag{2}$$

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

$$A \circ B = (A \Theta B) \oplus B \tag{3}$$

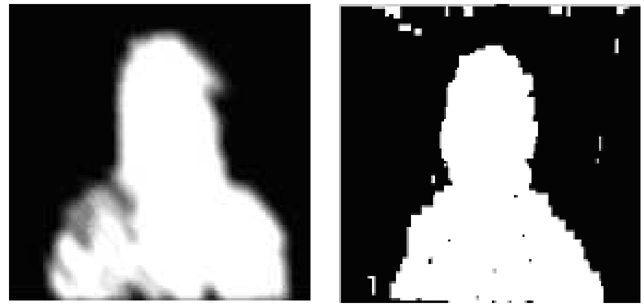
Here, A is the image and B is a structuring element. After applying the above explained pre-processing techniques, the new image obtained is as shown in the Figure 4.

3.6 Object tuning

This is a post processing technique applied to the traffic surveillance system application. In the current algorithm we use a median filter for noise elimination in both the i.e. object and background. As the object boundaries are not very smooth, a post processing technique is required on the foreground image. The final output of the object tuning phase is a binary image of the objects detected termed mask1.

3.7 Object Identification

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 a certain threshold, then, those pixels are replaced by the pixels of the original frame. This process identifies the moving object as shown in Figure 5.



(a) Image of Moving Hands (b) Image of Claire

Figure 4: After Post-Processing

(a) Moving Hands (b) Claire

Figure 5: Identification of Objects

3.8 Object counting

The tracked binary image mask1 forms the input image for counting. This image is scanned from top to bottom for detecting the presence of an object. Two variables are maintained i.e., count that keeps track of the number of vehicles and countregister countreg, which contains the information of the registered object. When a new object is encountered, it is first checked to see whether it is already registered in the buffer, if the object is not registered then it is assumed to be a new object and count is incremented, else it is treated as a part of an already existing object and the presence of the object is neglected. This concept is applied for the entire image and the final count of objects is present in variable count. A fairly good accuracy of count is achieved. Sometimes due to occlusions two objects are merged together and treated as a single entity.

4 Algorithm

4.1 Problem Definition

This consists of a video clip which is a sequence of traffic images in AVI format, the objectives are:

1. Given a video clip in the format of QCIF (176 x 144) for object identification and in AVI format for traffic surveillance, 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 and compare it with the already existing algorithms.

2. The video clip consists of a sequence of traffic images in AVI format, the objectives are:

   (a) To develop a vision based surveillance system capable of identifying vehicles in the scene.

   (b) To track the vehicles as they progress along the image sequence.

   (c) To count the number of vehicles in the image.

Assumptions: The background of a video sequence is stationary.

4.2 Algorithm

Three major functions are involved in the proposed technique. The first function is to read a given video clip and convert it into frames as shown in Table 1. The second function is to implement the major procedures like finding frame differences, eliminating the background, post-processing and then identifying the moving object as described in Table 2. The last function used is to implement the Least Square Method (LSM) on the outputs obtained for performance comparison shown in Table 3. It was tested for QCIF (176 x 144) video sequences. The algorithms/pseudo codes for various steps involved are as shown below.

In the next step, the difference between the frames at certain intervals is calculated. This is achieved by comparing any two frame differences, say, FDi and FDj. The matching pixels in FDi and FDj are considered to be a part of background and they are set to zero. All other pixels are unaltered. The image obtained by this procedure must undergo some post-processing techniques to remove the possible noise. After post-processing, the image is compared with the one of the original frames (usually, the first frame). If the pixels are less than certain threshold, then they are ignored. Otherwise, they are replaced by the pixels of original image. This resulting image will be consisting of the moving object ignoring the background and hence satisfying our requirement.

Table 1: Algorithm to Read Video

1. Initialise an array M_Array[ ] to an empty array.
2. for i = 1 to No_of_Frames in steps of 1 with Interval 4 frames
3. Convert movie structures stored in M_Array[ ] into images.
4. Convert the images obtained in Step 3 from RGB to Gray format.
5. Store all these gray images in an array viz.Im_Array[ ].

Table 2: Algorithm for Object Identification

```
ALGORITHM Idovs(Im_Array[],Rows, Cols, NumFrames)
//Input: An array of frames that are converted into images in gray colour format viz. Im_Array[], Rows and Cols indicating size of image and NumFrames indicating total number of images in Im_Array[].
//Output: An image showing the moving object.

p:=1; k:=5;// any pre-defined value

// finding the frame differences

for i:=1 to Rows in steps of 1
    for j:=1 to Cols in steps of 1
        for m:=1 to NumFrames in steps of k
            FD[i,j,p]:=Im_Array[i,j,m+k] Im_Array[i,j,m];
            p:= p+1;
        end for
    end for
end for

//Background Elimination
p:=2; q:=4;//any two pre-defined values

for i:=1 to Rows in steps of 1
    for j:=1 to Cols in steps of 1
        if (FD[i,j,p]==FD[i,j,q]) then
            BackElim[i, j]:= 0;
        else
            BackElim[i, j] := 255;
        end if
    end for
end for

//Post processing
K:= MedianFilter(BackElim);
G:= MorphologicalOpening(K);

//Object Identification
for i:=1 to Rows in steps of 1
    for j:=1 to Cols in steps of 1
        //TH is some observed threshold
        if G(i, j) ¿= TH then
            Object[i,j] = Im_Array[i, j, 1];
        end if
    end for
end for
```

5 Implementation and Performance Analysis

Here two algorithms are proposed background elimination and background registration method which are implemented using Matlab 7. The performance analysis is done through the method of Least Squares. The least square method is normally used to find the best-fit, given two sets of data. According to the method of least squares, the best-fit must satisfy the rule given by equation (4).

Table 3: Algorithm for Backgorund Registration

ALGORITHM *BGRegist()*
//Input: M_Array
//Output: An Image with Registered Background in bg array
//Initialize array [b] to zeros

1. for i:=1 to m
   for j:=1 to n
   for k=1 to l-1
   if abs(double(T(i,j,l-k))-double(T(i,j,k)))¡10
   b(i,j)=T(i,j,k)
   end if
   end for
   end for
   end for
2. Convert b array values to unsigned integers and store it into array called *background*.
3. Fill the hole regions in image *background* and store it in *bg* array
4. Show the output images *background*, *bg*.
5. Declare two global variables m and n which stores the row and column values of video frames respectively.

Table 4: Least Square Method

ALGORITHM *LS(O[], BE[], Rows, Cols)*
//Algorithm for finding least square value comparing
//original and identified images
//Input: Original Image O and identified image BE
//with Rows and Cols indicating size of images.
//Output: An integer value showing the least square
//value

1. LSValue:=0;
2. for i:=1 to Rows in steps of 1
   for j:=1 to Cols in steps of 1
   LSValue:= LSValue + O(i, j) BE(i, j)2
   end for
   end for

$$\Pi = d_1^2 + d_2^2 + ... + d_n^2 = \sum_{i=1}^{n} d_i^2 = min \tag{4}$$

Table 5: Algorithm for Counting

ALGORITHM *Count()*
//Input: d is specific video frame
//Output: An image with Foreground Objects is stored in c
//Initialize count=0 and count register buffer
//countveg=0

1. Traverse the mask1 image to detect an object
2. If object encountered then check for registration in *countveg*
3. If the object is not registered then increment count and register the object in *countveg*, labelled with the new count.
4. repeat steps 2-4 untill traversing not completed.

This paper uses the least square method for comparing the outputs. Let Oij be any frame of the input video clip, BEij be the result obtained through Background Elimination technique and BRij be the result obtained through Background Registration technique. Here, i=1, 2,m and j=1,2,n. where m and n indicate rows and columns (i.e. size) of the image. The values are calculated using the formulae

$$V1 = \sum_{i=1}^{m}\sum_{j=1}^{n}(O_{ij} - BE_{ij})^2 \tag{5}$$

and

$$V2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(O_{ij} - BR_{ij})^2 \tag{6}$$

It is observed through simulation that, $V1 < V2$ for various test sequences. The actual values obtained for test sequences are given in Table VI. The outputs obtained through two different techniques are as shown in Figure 6. It is also 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 and on sequences captured in our laboratory. The results obtained from proposed algorithm are compared with those of background registration method. The Graph Showing Error Rates computed through Least Square Method is shown in Figure 6.

5.1 Simulation Software

Simulation is performed using Matlab Software. This is an interactive system whose basic data element is an array that does not require dimensioning. It is a tool used for formulating solutions to many technical computing problems, especially those involving matrix representation. This tool emphasises a lot of importance on comprehensive prototyping environment in the solution of digital image processing. Vision is most advanced of our senses, hence images play an important role in humans perception, and Matlab is a very efficient tool for image processing.

5.2 Performance Analysis

This system was implemented on an Intel Core 2 Duo 4.0 GHz PC. We have tested the system on image sequences on different scenarios like traffic junction intersection, highways etc. The entire processing requires approximately about 60 frames. Real life traffic video sequence are used to demonstrate the knowledge discovery process i.e., vehicle tracking from traffic video sequences using the proposed framework. All the videos chosen for vehicle tracking have same light intensity and have been taken during day time. We convert the colour video frames to gray scale images. Multimedia data mining techniques are used to count the number of vehicles passing through the road intersection in a given time duration.

This video segmentation method was applied on three different video sequences two of which are depicted below. For the first video sequence Figure 7(a) depicts the original image, Figure 7(b) shows the background registered image, Figure 3 (c) the foreground detected objects obtained after background subtraction, and finally Figure 7(d) shows the count of the detected objects. The same is repeated for the next video sequence and is indicated in Figure 8. The system is able to track and count most vehicles successfully. Although the accuracy of vehicle detection was 100%., the average accuracy of counting vehicles was 94%.This is due to noise which causes detected objects to become too large or too small to be considered as a vehicle. However, two vehicles will persist to exist as a single vehicle if relative motion between them is small and in such cases the count of vehicles becomes incorrect.

An added advantage of this algorithm is, the segmentation logic is not intensity based, and hence vehicles whose intensities are similar to the road surface are not missed out. The results were successfully carried out on three videos; the accuracy of detecting the objects was 100%. as shown in Table VII. The detected objects are then counted and the accuracy of counting is shown in Table VIII.

6 Conclusions

In this paper, we propose an efficient algorithm for detecting a moving object using background elimination technique. Initially we compute the frame differences

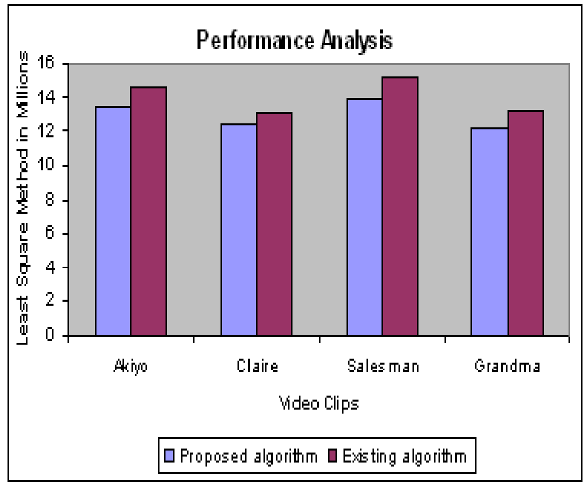


Figure 6: Performance Analysis os Proposed and Existing Algorithm

Table 6: Comparision of Error Rates

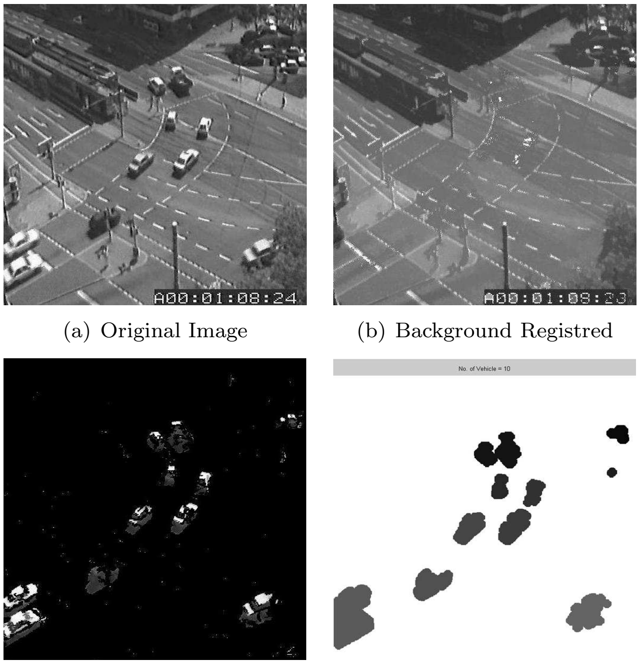
Video Sequence Name (i)	Proposed Algorithm (Background Elimination Technique) (ii)	Existing Algorithm (Background Registration Technique) (iii)	Difference in Values (iv)=(ii)-(iii)
Akiyo	13400352	14648832	-1248480
Claire	12442454	13141224	-698770
Salesman	13964820	15104228	-1139408
Grandma	12137846	13267234	-1129388

Table 7: Detection of Moving Objects

Input Video	Format	Actual Moving Objects	Detected Moving Objects	Accuracy %
Video 1	Grayscale	11	11	100
Video 2	RGB	3	3	100
Video 3	Grayscale	11	10	90

Table 8: Accuracy of Counting

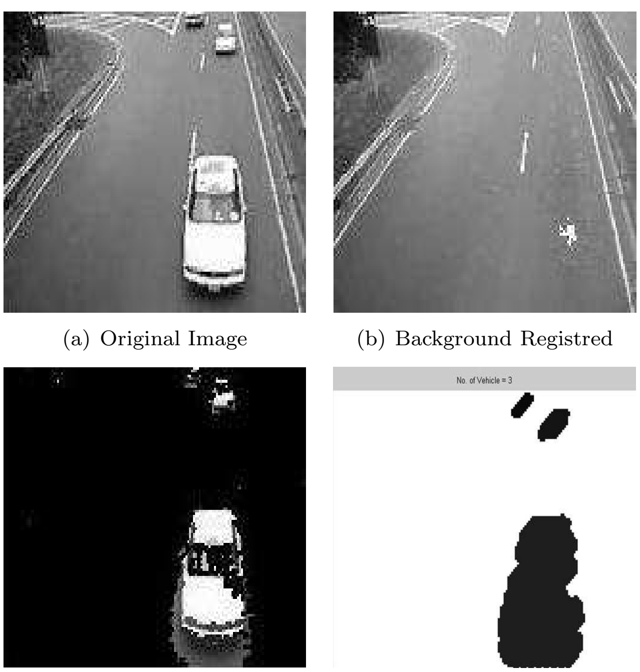
Input Video	Size	Actual Number of Vehicles	Detected Moving of Vehicles	Accuracy %
Video 1	512*512	11	9	82
Video 2	160*120	3	3	100
Video 3	768*576	7	7	100



(a) Original Image (b) Background Registred

(c) Foreground Detection (d) Objects Tracked

Figure 7: Video 1



(a) Original Image (b) Background Registred

(c) Foreground Detection (d) Objects Tracked

Figure 8: Video 2

(FD) between frames Fi and Fi+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 than using background registration technique.

Good segmentation quality is achieved efficiently. This paper also discusses an application system of traffic surveillance. Here we develop an algorithm to track and count dynamic objects efficiently. The tracking system is based on a combination of a temporal difference and correlation matching. The system effectively combines simple domain knowledge about object classes with time domain statistical measures to identify target objects in the presence of partial occlusions and ambiguous poses in which the vehicles are moving. The background clutter is effectively rejected. The experimental results show that the accuracy of counting vehicles reached 94%, although the vehicle detection was computational complexity of our algorithm is linear to the size of a video clip and the number of vehicles tracked. As a future work a combination of higher dimensional features with some additional constraints may be tried so that adverse effects of some features can be compensated by contribution of others.

References

[1] Sikora, T., The MPEG-4 Video Standard Verification Model, *IEEE Transactions, Circuits Systems, Video Technology,* vol. 7, pp. 19-31, Feb. 1997.

[2] Nack F. and Lindsay A. T., Everything you Wanted to Know about MPEG-7: Part 2, *IEEE Multimedia,* vol. 6, pp. 64-73, Dec. 1999.

[3] Salembier P. and Marques F., Region-based Representations of Image and Video: Segmentation Tools for Multimedia Services, *IEEE Transactions, Circuits Systems, Video Technology,* vol. 9, pp. 1147-1169, Dec. 1999.

[4] Wang D., Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking, *IEEE Transactions, Circuits Systems, Video Technology,* vol. 8, pp. 539-546, Sept. 1998.

[5] Y. Yokoyama, Y. Miyamoto and M. Ohta, Very Low Bit Rate Video Coding Using Arbitrarily Shaped Region-Based Motion Compensation, *IEEE Trans-*

actions, Circuits System. Video Technology, vol. 5, pp. 500-507, Dec. 1995.

[6] L. Wu, J. Benoise-Pineau, P. Delagnes and D. Barba, Spatio-temporal Segmentation of Image Sequences for Object-Oriented Low Bit-Rate Image Coding, *Signal Processing: Image Communication.,* vol. 8, pp. 513-543, 1996.

[7] P. Salembier, Morphological Multiscale Segmentation for Image Coding, *Signal Processing,* vol. 38, pp. 359-386, 1994.

[8] N. T. Watsuji, H. Katata and T. Aono, Morphological Segmentation with Motion Based Feature Extraction, *In International Workshop on Coding Techniqus for Very Low Bit-Rate Video,* Tokyo, Nov. 8-10, 1995.

[9] W. H. Hong, N. C. Kim and S. M. Lee, Video Segmentation Using Spatial Proximity, Color and Motion Information for Region-Based Coding, *In Proceedings SPIE Visual Communications and Image Processing,* vol. 2308, pp. 1627-1633, 1994.

[10] Choi J. C., Lee, S .W. and Kim, S. D., Spatio-Temporal Video Segmentation Using a Joint Similarity Measure, *IEEE Transactions, Circuits Systems, Video Technology,* vol. 7, pp. 279-289, Apr. 1997.

[11] Vibha L., Venugopal K. R. and L. M. Patnaik, A Study of Breast Cancer Using Data Mining Techniques, *Technical Report, University Visvesvaraya College of Engineering, Bangalore University,* August 2003.

[12] Aach T., Kaup A. and Mester R., Statistical Model-Based Change Detection in Moving Video, *Signal Processing,* vol. 31, pp. 165-180,Mar. 1993.

[13] Neri A., Colonnese S., Russo G. and Talone P., Automatic Moving Object and Background Separation, *Signal Processing,* vol. 66, pp. 219-232, Apr. 1998.

[14] Stauder J., Mech R. and Ostermann J., Detection of Moving Cast Shadows for Object Segmentation, *IEEE Transaction Multimedia,* vol. 1, pp. 65-76, Mar. 1999.

[15] D. Koller, J. Weber, T. Haung and J. Malik, Moving Object Recognition and Classification based on Recursive Shape Parameter Estimation, *In Proceedings of the 12th Israeli Conference on Artificial Intelligence, Computer Vision and Neural Networks,* pp. 359-368, Tel-Aviv, Israel, December 1993.

[16] D. Koller, J. Weber, T. Haung, J. Malik, G. Ogasawara, B. Rao and S. Russel, Towards Robust Automatic Traffic Scene Analysis in Real-Time, *In Proceedings of the 12th International Conference on Pattern Recognition (ICPR-94),* pp. 126-131, Jerusalem, Israel, October 9-13, 1994.

[17] Puzicha J., Hofmann T. and Buhmann J. M., Histogram Clustering for Unsupervised Image Segmentation, *In IEEE Computer Society Conference Computer Vision and Pattern Recognition,* Fort Collins, CO, pp. 602-608, June 1999.

[18] Chen S. C., Shyu M. L. and Zhang C., An Unsupervised Segmentation Framework for Texture Image Queries, *In 25th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC),* Chicago, Illinois, USA, Oct. 2000.

[19] Chen S. C., Shyu M. L. and Zhang C., An Intelligent Framework for Spatio-Temporal Vehicle Tracking, *4th International IEEE Conference on Intelligent Transportation Systems,* Oakland, California, USA, Aug. 2001.

[20] D. Beymer, P. McLauchlan, B. Coifman and J. Malik, A Real-time Computer Vision System for Measuring Traffic Parameters, *In Proceedings of IEEE Conference Computer Vision and Pattern Recognition,* Puerto Rico, pp. 496501,June, 1997,

[21] O. Masoud and N. P. Papanikolopoulos, Robust Pedestrian Tracking Using a Model-based Approach, *In Proceedings of IEEE Conference on Intelligent Transportation Systems,* pp. 338343, Nov. 1997.

[22] A. J. Lipton, H. Fujiyoshi and R. S. Patil, Moving Target Classification and Tracking from Real-time Video, *In Proceedings of IEEE Workshop Applications of Computer Vision,* pp. 814, 1998.

[23] Gupte S., Masoud O., Martin R. F. K. and Papanikolopoulos N. P., Detection and Classification of Vehicles, *In IEEE Transactions on Intelligent Transportation Systems,* vol. 3, no.1, pp. 37 47, March 2002.

[24] Dailey D. J., Cathey F. and Pumrin S., An Algorithm to Estimate Mean Traffic Speed Using Uncalibrated Cameras, *In IEEE Transactions on Intelligent Ttransportations Systems,* vol. 1, No. 2, pp. 98-107, June, 2000.

[25] Grimson W. E. L., Stauffer C., Romano R. and Lee L., Using Adaptive Tracking to Classify and Monitor Activities in a Site, *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceeding,* pp. 22-31, 1998.