Moving Vehicle Identification using Background Registration Technique for Traffic Surveillance

Vibha L*, Venkatesha M*, Prasanth G Rao*, Suhas N*, P Deepa Shenoy*, Venugopal K R*, L M Patnaik**

Abstract- **Real-time segmentation of moving regions in image sequences is a fundamental step in many vision systems including automated visual surveillance and human-machine interface. In this paper 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 present algorithm for vision-based detection and counting of vehicles in monocular image sequences for traffic scenes are recorded by a stationary camera. The method is based on the establishment of correspondences between regions and vehicles, as the vehicles move through the image sequence. Background subtraction is used which improves the adaptive background mixture model and makes the system learn faster and more accurately, as well as adapt effectively to changing environments. The resulting system robustly identifies vehicles at intersection, rejecting background and tracks vehicles over a specific period of time. Real-life traffic video sequences are used to illustrate the effectiveness of the proposed algorithm.**

Index Terms- Camera calibration, Motion segmentation, Vehicle tracking, Multimedia data mining, Background Registration.

I. INTRODUCTION

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. A vehicle tracking and classification system is described as one that can categorize moving objects as vehicles and further classifies the vehicles into various classes. 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., 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. Two common themes associated with tracking traffic movement and recognizing accident information from real time video sequences are first, the video information must be segmented and turned into objects, second, the behaviour of these objects are monitored (they are tracked) for immediate decision making purposes.

One such application is closed-circuit television cameras which are becoming increasingly common on freeways and are used for traffic management; 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 camera's 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

*Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, INDIA -560001. (e-mail: vibhal1@rediffmail.com, shenoypd@yahoo.com)

** Microprocessor Applications Laboratory, Indian Institute of Science, Bangalore, INDIA-560012

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.

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. In section 4 the algorithm for detection and counting of vehicles is presented. Parameters for implementation and performance are analysed in section 5. Section 6 contains the conclusions.

II. Related Work

A brief survey of the related work in the area of traffic surveillance is presented here. Koller et al., [1], [2] 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. The authors in [3] use 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., [4], [5] have addressed the issues regarding unsupervised image segmentation and object modelling with multimedia inputs to capture the spatial and temporal behaviour of the object for traffic monitoring. D. Beymer et al., [6] 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 however is very computationally expensive. In [7] tracking and counting pedestrians using a single camera is proposed. Here the image sequences are segmented using background subtraction and the resulting connected regions are then grouped together into pedestrians and tracked. A. J. Lipton et al., [8] 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. In [9] 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.

Daniel et al., [10] 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 [11] Grimson et al., analysis a vision based system that monitors activities in a site, over a period of time using sensor networks. The idea here is to classify detected objects, learn common patterns of activities for different kinds of objects and to identify unusual patterns. Cheng and Kamath [12] compare the performance of a large set of different background models on urban traffic video. They experimented with sequences filmed in weather conditions such as snow and fog, for which a robust background model is required.

Kanhere et al., [13] applies a feature tracking approach to traffic viewed from a low-angle off-axis camera. Vehicle occlusions and perspective effects pose a more significant challenge for a camera placed low to the ground. Deva et al., [14] proposes a concept to automatically track the articulations of people from video sequences. This is a challenging task but contains a rich body of relevant literature. It can identify and track individuals and count distinct people. Karmann and Brandt [15] discuss the segmentation approach using adaptive background subtraction that uses Kalman filtering to predict the background. Segmentation requires vehicles to be accurately separated from the background with minimal amount of initialization.

Toufiq P. et al., in [16] describes background subtraction as the widely used paradigm for detection of moving objects in videos taken from static camera which has a very wide range of applications. The main idea behind this concept is to automatically generate and maintain a representation of the background, which can be later used to classify any new observation as background or foreground. In [17] background subtraction also involves computing a reference image and subtracting each new frame from this image and thresholding the result. This method is an improved version of adaptive background mixture model, it is faster and adapts effectively to changing environments.

III. ARCHITECTURE AND MODELING

In many real-time applications like video conferencing, the camera is fixed. Some techniques use global motion estimation and comparison to compensate the change in background due to camera motion. In this algorithm, we assume a stationary background for all video sequences.

The architecture and modelling of the present paper is presented in Figure 1. Initially, a video clip is read and decomposed into a number of frames. Next, using these frames as inputs, the stationary background image is registered. The next phase is identifying the foreground dynamic objects, which is obtained by subtracting background image from the given input video frame. The following phase is the post processing phase where interference of noise is being minimised. Then, the frame consisting of only dynamic objects is obtained which is then converted into a binary image, where presence of an object is indicated as a white patch while the rest of the area is made to appear black. This is achieved using morphological processing techniques (dilation), which is applied to the binary image to group the different segments of a single object into one logical object. Structuring elements for dilation are chosen based on the video sequences. A counting algorithm is then applied to the resulting image to assist in counting the number of objects.

Background Modelling: 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 use the fact that the 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.

Background subtraction is a technique which eliminates static components from a video sequence. An important assumption in this application is that camera remains stationary. The whole idea is to create a reference frame which consists of all stationary components in the given image sequence. Frame differences can be computed by finding the difference between consecutive frames but this will introduce additional computational complexity. Hence the difference between the frames at regular intervals is considered. Frame difference at an interval of four frames is computed and after thresholding it, the reference frame is constructed. The frame difference follows Gaussian distribution as

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

Here, μ is the mean of frame differences (FD) and σ is the standard deviation of FD.

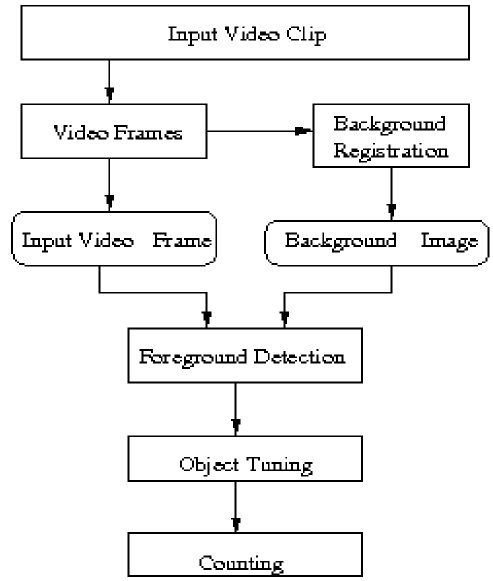

Fig 1. Architecture and Modelling.

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 also be used to refine the vehicle type and also to correct errors which are 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.

Object tuning: Usually due to the camera noise and irregular object motion, there always exist 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 applied on the foreground image. First, order-statistics filters termed spatial filters are used, 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 final output of the object tuning phase is a binary image of the objects detected termed *mask1*.

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.

IV. ALGORITHM

A. *Problem Definition:* This consists of a video clip which is a sequence of traffic images in AVI format, the objectives are:

i) To develop a vision based surveillance system capable of identifying vehicles in the scene.
ii) To track the vehicles as they progress along the image sequence.
iii) To count the number of vehicles in the image.

B. *Algorithm:* Four major functions are involved in the proposed technique. The first function is to read and divide a give video clip into number of frames. This is shown in Table I. Second function is to implement the major procedures like finding frame differences and identifying the background registered image which is depicted in Table II. Next post-processing is performed, and the background is eliminated thus maintaining only the foreground objects as indicated in Table III. The last function assists in counting the detected objects which is described in Table IV. The aim of the algorithm is to design an efficient tracking and counting system. The algorithms/pseudo codes for various steps involved are as shown below. Given a video clip, the initial problem is segregating it into number of frames. Each frame is then considered as an independent image, which is in RGB format and is converted into Gray scale image. Next the difference between the frames at certain intervals is computed. This interval can be decided based on the motion of moving object in a video sequence. If the object is moving quite fast, then the difference between every successive frame is considered.

TABLE I
Algorithm to Read Video

Step 1: Initialise an array *M_Array*[ ] tozeros.
Step 2: Declare two global variables *m* and *n* which stores the row and column values of video frames respectively..
Step 2: for i = 1 to No_of_Frames in steps of 1 with Interval 4 frames
- Read each frame of video clip and store it.
- Store the video frames into *M_Array*[ ].
- Increment variable *k* which stores the total number of frames in *M_Array*.

end for
Step 3: for i = 1 to k
Convert images obtained in step 3 from RGB to gray format and store that in a three- dimensional array $T[m,n,l]$.

But, if the motion is slow, then the difference at the intervals of 3 to 5 frames is sufficient. The background region of the image is then identified.

TABLE II
Algorithm for Background Registration

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

Step 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

Step 2: Convert b array values to unsigned integers and store it
          into array called background..
Step 3: Fill the hole regions in image background and store it
          in bg array
Step 4: Show the output images background, bg.
```

TABLE III
Algorithm for Background Elimination

```
ALGORITHM BGElimin( )
//Input: d is a specific Video Frame
//Output: An image with Foreground Objects is stored
   in'c'
 //Initialize c array to zeros
Step 1: for i = 2 to m-1
            for j = 2 to n-1
                if the difference between the pixel values of d
                   array with bg or b array is less than 20
                         store value 0 for that pixel in array c
                else
                      store pixel value
                end if
              end for
          end for

Step 2: Convert the values in c array and apply median
          filter.
 Step 3: Show the output images c, d
```

TABLE IV
Algorithm for Counting

```
ALGORITHM Count( )
//Initialise count = 0 and counted register buffer
  countveg = 0
      Step 1.   Traverse the mask1 image to detect an
      object
      Step 2.  If object encountered then check for
               registration in countveg
      Step 3.  If the object is not registered then increment
               count and register the object in countveg,
               labelled with the new count.
      Step 4.  repeat  steps 2-4 untill traversing  not
               completed.
```

V. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. *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.

B. *Performance Analysis*:

This system was implemented on an Intel Core 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 3 (a) depicts the original image, Figure 3 (b) shows the background registered image, Figure 3 (c) the foreground detected objects obtained after background subtracted, and finally Figure 3 (d) shows the count of the detected objects. The same is repeated for the next video sequence and is indicated in Figure 4. 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 V. The detected objects are then counted and the accuracy of counting is shown in Table VI.

VI. CONCLUSIONS

In this paper, we present a background registration technique to detect moving objects. A system has been developed to track and count dynamic objects efficiently.

TABLE V
Detection of Moving Object

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 VI
Accuracy of Counting

Input Video	Dimensions	Actual number of vehicles	Counted number of vehicles	Accuracy %
Video 1	512*512	11	9	82
Video 2	160*120	3	3	100
Video 3	768*576	7	7	100

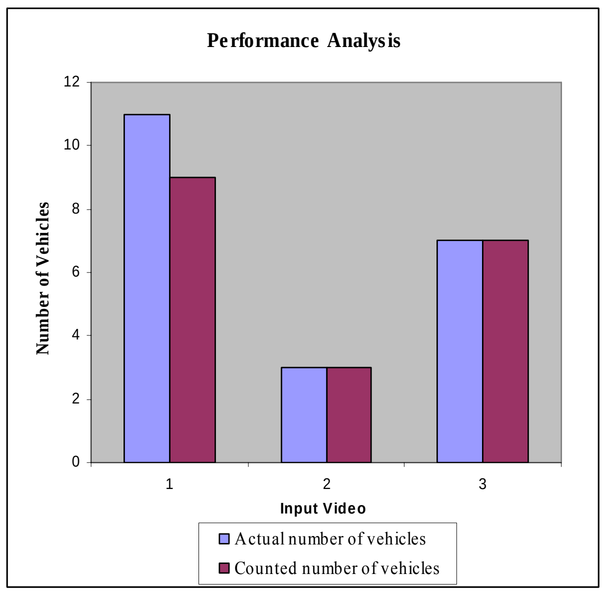

Fig. 2. Graph Depicting Counting Accuracy

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, and the background clutter is effectively rejected. The experimental results show that the accuracy of counting vehicles was 94%, although the vehicle detection was 100% which is attributed towards partial occlusions. The computational complexity of our algorithm is linear in the size of a video frame 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.

ISBN: 978-988-98671-8-8

Fig. 3 a. Original Image Fig. 3 b. Background Registered

Fig. 3 c. Foreground Detection Fig. 3 d. Object Tracked

Fig. 4 a. Original Image Fig. 4 b. Background Registered

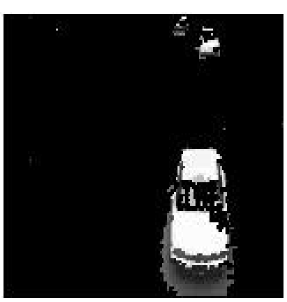

Fig. 4 c. Foreground Detection Fig. 4 d. Object Tracked

REFERENCES

1. 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.
2. 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.
3. 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.
4. Chen S. C., Shyu M. L. and Zhang C., "An Unsupervised Segmentation Framework for Texture Image Queries", *In the 25th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC),* Chicago, Illinois, USA, Oct. 2000.
5. 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.
6. D. Beymer, P. McLauchlan, B. Coifman and J. Malik, "A Real-time Computer Vision System for Measuring Traffic Parameters", *In Proceeding. IEEE Conference. Computer Vision and Pattern Recognition*, Puerto Rico, June, 1997, pp. 496–501.
7. O. Masoud and N. P. Papanikolopoulos, "Robust Pedestrian Tracking using a Model-based Approach", *In Proceedings of IEEE Conference on Intelligent Transportation Systems*, Nov. 1997, pp. 338–343.
8. 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*, 1998, pp. 8–14.
9. 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, March, 2002, pp. 37 – 47.
10. 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.
11. 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.
12. S. Cheung and C. Kamath, "Robust Techniques for Background Subtraction in Urban Traffic Video", *In Video Communications and Image Processing, SPIE Electronic Imaging*, San Jose, January, 2004.
13. N. Kanhere, S. Pundlik and S. Birchfield, "Vehicle Segmentation and Tracking from a Low-Angle Off-Axis Camera", *In IEEE Conference on Computer Vision and Pattern Recognition*", San Diego, June, 2005.
14. Deva R., David A., Forsyth and Andrew Z., " Tracking People by Learning their Appearance", *In IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no1, Jan. 2007.
15. K. P. Karmann and A. Von Brandt, "Moving Object Recognition using an Adaptive Background Memory", *In Proceedings of Time-Varying Image Processing and Moving Object Recognition*, vol. 2, V. Capellini, Ed., 1990.
16. Toufiq P., Ahmed Egammal and Anurag Mittal, "A Framework for Feature Selection for Background Subtraction", *In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06),* 2006.
17. P. Kaewtra Kulpong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection", *In Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems*, Sept. 2001.