International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

REAL TIME OBJECT IDENTIFICATION FOR AUTOMATED VIDEO SURVEILLANCE SYSTEM

L.Menaka[1], Dr. M.Kalaiselvi Geetha[2], M.Chitra[3]

[1]*M.E Final CSE, Annamalai University, Chidambaram, India.*
[2]*Associate.Professor/CSE, Annamalai University, Chidambaram, India.*
[3]*M.E Final CSE, Annamalai University, Chidambaram, India.*

Email: l.menaka.it@gmail.com, geesiv@gmail.com, chitrasuja88@gmail.com

Abstract

Intelligent video surveillance system has emerged as a very important topic of research in the field of computer vision in the recent years. It is well suited for a broad range of applications such as to monitor activities at traffic intersections for detecting congestions, and then predict the traffic flow which assists in regulating traffic. Manually reviewing the large amount of data they generate is often impractical. Moving object classification in the field of video surveillance is a key component of smart surveillance software. In this paper, we have proposed robust methodology and algorithms adopted for people and object classification in automated surveillance systems. Object motion can be detected using background subtraction model. The background subtraction and image segmentation based on morphological transformation for tracking and object classification on highways is proposed. This algorithm uses erosion followed by dilation on various frames. Proposed algorithm segments the image by preserving important edges which improves the adaptive background mixture model and makes the system learn faster and more accurately, as well as adapt effectively to changing environments. A probabilistic algorithm for object identification and visual tracking using twin comparison method that incorporates height width based classification method and robust SVM classifier with histogram oriented gradients is utilized for identifying human and vehicles. The experimental results demonstrate the effectiveness of the proposed approach in classifying human and other objects.

Keywords **- HOG, SVM, contour, morphological operator**

I. INTRODUCTION

Traffic monitoring is becoming more and more important, to control the increasing traffic-flow on highways and to meet safety and security standards. Traffic monitoring can be done with automatic event detection systems. In these systems object identification and tracking are major issues. Automatic event detection system requires increase of detection rates and the decrease of false alarm rates. The ability to reliably detect pedestrians from video data has very important applications in many fields like, intelligent transportation, automated surveillance and security, robotics, assistive technology for visually impaired, advanced human machine interfaces, automated driver assistance systems in vehicles etc.

Pedestrian are more vulnerable to accidents and collisions involving pedestrians often produce severe injuries.

Accurately detecting pedestrians from a video is one of the most challenging tasks for object detection which attracts most of the researchers working in this field. This paper proposes an approach for conditions both pedestrian and vehicle identification in real time.

II. RELATED WORK

Vision based vehicle detection is an area of research in the intelligent transportation systems community. In the literature, many studies have been performed on the static images.

The binary classification scheme is an efficient tool which can be used for object detection and matching which is described in [7].In this method, Scale Invariant Feature Transform(SIFT) points were extracted which is invariant to image scaling and rotation and partially invariant to change in illumination and 3D rotation.

A statistical approach has been used in [4], performing vehicle detection using principle component analysis (PCA) and independent component analysis(ICA) to do classification on a statistical model and increased its speed by modeling the PCA and ICA vectors with the weighted Gaussian mixture model. A support vector machine(SVM) approach was used in [1],built multiple detectors using Gabor filters Haar wavelets, PCA, truncated wavelet features using neural networks and SVM classifiers. Template matching is one of the methods used for vehicle detection and tracking. A review of recent template matching methods for detection and tracking of vehicle is presented in [3].

The "irregular blobs" can be detected and then cluster the pieces according to the common motion constraint of the extracted features. It reduces the computational costs by limiting the feature analysis only to "irregular blobs" is presented in [9]. Using 2D and 3D models for detection has also been explored by several authors. A 3D model-based detection approach with background subtraction is presented in [11]. The 2D templates has been created from the 3D vehicle models which is used to generate multiple hypotheses for a given foreground mask. This approach use only the template contours, so like the feature based approaches, its performance on noisy, low resolution and crowded scenes is uncertain.

The partial occlusion problem is addressed by a feature based tracking algorithm in [14]. The detection is done based on the "corner" features and then grouped them according to a "common motion constraint". However, both algorithms depend purely on the accuracy of feature detection and matching, which makes them error prone in noisy, low resolution videos. Common motion constraint is not applicable in very crowded scenes, where the vehicles are forced to move at similar speeds.

There are many technologies that are currently being used for pedestrian and vehicle detection such as ultrasonic sensors, Doppler radar sensors piezo-metric sensors etc which is presented in [8]. These sensors while being very effective have various drawbacks ranging from cost effectiveness to durability. Video based detection emerged as an important aspect of research, as proliferation high performance cameras and faster inexpensive computing systems became assessable.

This paper introduces twin comparison method at classification stage that classifies the object features into pedestrian features and vehicle features. A pixel level processing method for classifying the objects using height and width. Again the classification is done by binary Support Vector Machine (SVM) classifier. The method combines appearance based feature training using a Histogram oriented Gradient based descriptor and SVM. The training of the SVM classifier is carried out offline.

The proposed classification procedure is useful not only for increasing the speed and reliability of feature matching but also for reducing the computational load associated with the robust estimation of the registration parameters.

III. OVERVIEW

In this work, initially foreground objects are segmented from the background. Next, the foreground object motion between the current and previous frames is obtained.

After background subtraction, initially height, width and number of pixels are measured, based on which classification is performed. Secondly, HOG based SVM is used to obtain robust results. A discussion on SVM is given in section 5.2. The Fig.1 shows the overview of the proposed system.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

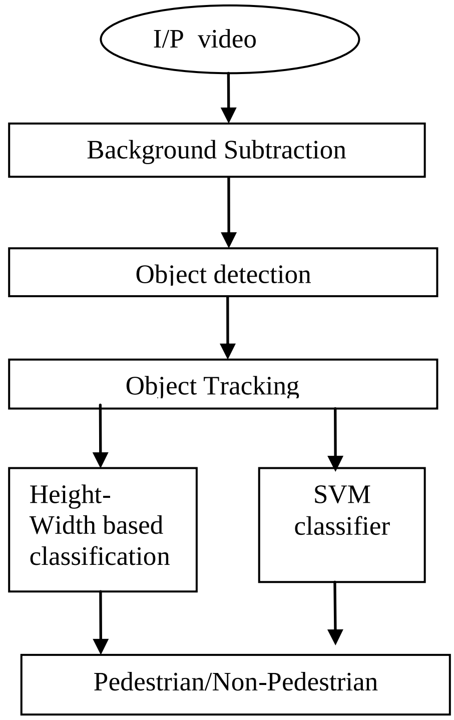


Fig 1.Overview of the proposed System

3.1. Background Subtraction

The main reason of using background subtraction is that it works well if the background is static for long time. In traffic surveillance system, camera often remains static. This approach uses mixture of Gaussian for background subtraction.

3.1.1. Mixture Of Gaussians

Mixture of Gaussians method maintains a density function for every pixel and is capable of handling multi model backgrounds.

The Equation for Mixture of Gaussian is

$$p(x|\pi, \mu, \Sigma) = \sum_{k=1}^{K} \pi_k \; N(x|\mu_k \Sigma_k)$$

Where, *x* random variable

$\pi_k$mixing coefficient of the k^{th}Gaussian
μ_k mean of the k^{th}Gaussian
$\Sigma_k$covariance of the k^{th}Gaussian

Mixture of Gaussians robustly deal with lighting changes, repetitive motions, clutter, introducing or removing objects from the scene and slowly moving objects. It can be updated without having to store large number of frames in buffer hence reducing memory costs. The learning rate is passed into MOG model. The learning rate is the rate at which the model adapts to changes in the video image. Low values correspond to a slowly adapting model. High values make the model adapt quickly to scene changes. The output of Background subtraction is shown in Fig.2(a) and Fig 2(b).

Fig 2(a) Input Frame

Fig 2(b)Background subtracted frame

Background Model:

Input: **i Frames,** where **i** representing current frame
Output: Background model

(1) For i Frames**,** get the median intensity
(2) Remove noises (thresholding)
(3) Remove small regions (which is usually false foreground object during background subtraction)
(4) Update median intensity measurement
(5) Repeat Step 1

IV. HEIGHT WIDTH BASED CLASSIFICATION

Objects must be initially classified as pedestrians or vehicles. The proposed approach uses height-width based classification method as seen in Fig.3.

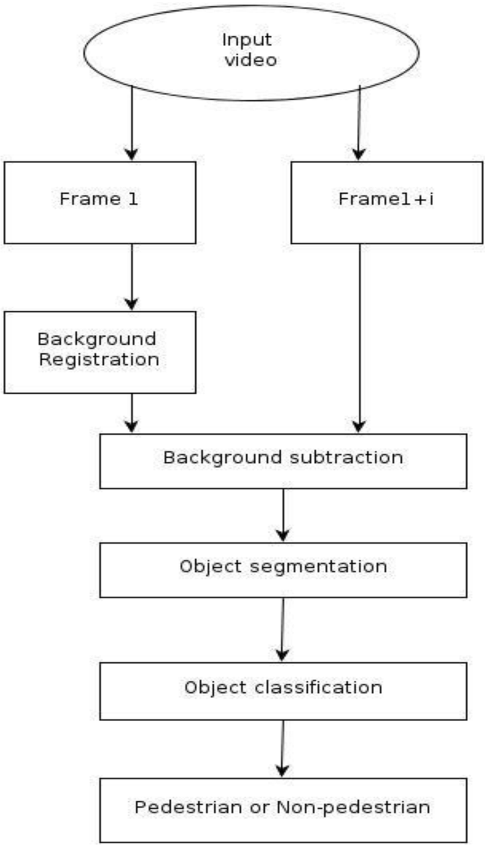


Fig 3. A model for height-width based classification method

4.1. Object Detection

In the proposed algorithm, the first frame is considered as background. The mixture of Gaussians method is used in background subtraction. At the segmentation stage, morphological operators are used. This approach uses dilation and erosion as morphological operators. Mathematical morphology is used for analyzing object shape characteristics (such as size and connectivity). Basic operation of a morphology-based approach is the translation of a structuring element over the image.

After background subtraction successfully extract out the foreground, it reproduces them into a binary image as seen in Fig 2(b). To make the objects more recognizable and informative, these objects should be marked and recorded.

Contouring is the process of marking out the edge of the object, making it more recognizable and informative.

Once background subtraction is done, moving objects are marked using contours. An example of object detection is shown in Fig 4.

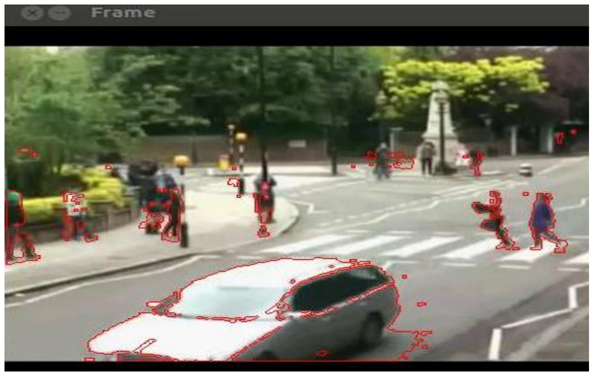


Fig4: Object Detection

4.2. Object Identification

In computer vision field, object's edge is usually identified as the area of pixels that contrasts with its neighbour pixels or area of pixels motion. After getting the edge of the object , the height of the object can be measured by subtracting the top left pixel position of the rectangle and bottom left pixel position of the rectangle . the width of the object can be measured by subtracting the top left pixel position of the rectangle and top right pixel position of the rectangle. The threshold value set to 1.126 or below for pedestrian and from1.126 to 2.72 for vehicle from the ground truth results. Based on threshold value for the height and width, we can classify the object whether it is a pedestrian or vehicle. An example is shown in Fig.5.

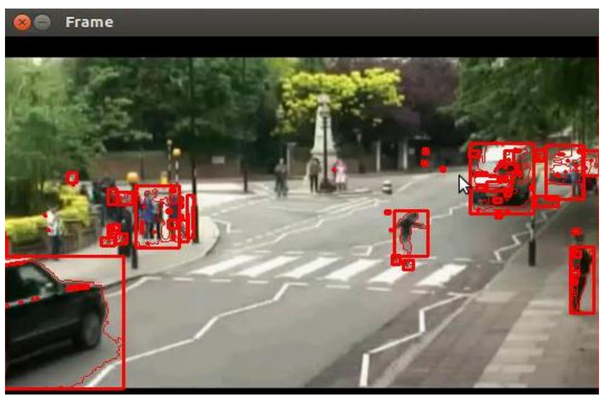


Fig 5. object identification

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

V. SVM BASED CLASSIFICATION

A feature is a descriptive aspect extracted from a video frame. Visual data exhibit numerous types of features that could be used to recognize or represent the information it reveals. Identification of objects in a video frame relies on competent use of these features that provide discriminative information useful for high level analysis. The following subsections present the description of these features used in this work.

5.1. HOG

Color and shape feature of the object are not discriminative enough to separate the positive examples from the negative ones. On the other hand, positive and negative examples have significantly different line distributions and orientations. Because of these reason, a histogram of oriented gradients (HOG) based method, similar to [9] and [12], is used to extract features from the object.

HOGs evaluate local histograms of image gradient orientations in a dense grid. The idea is that the local appearance and shape of the objects can often be well characterized by the distribution of the local edge directions, even if the corresponding edge positions are not accurately known. This idea is implemented by dividing the image into small regions called cells. Then, for each cell, a histogram of the gradient orientations over the pixels is extracted.

The original HOG technique was proposed by Dalal and Triggs [9], they presents two different kinds of configurations, called Rectangular HOG (R-HOG) and circular HOG (C-HOG), depending on the geometry of the cells used. Specifically, the former involves a grid of rectangular spatial cells and the latter uses cells partitioned in a log-polar fashion. We follow the configuration described in [10] to divide the image sub region into cells and blocks to facilitate the HOG feature extraction. The detailed steps are listed below:

1. Scale the image to 32x32 pixels and smooth it with a Gaussian filter.
2. Divide the resulting image into 16 cells, with each cell 8x8 pixels.
3. Each group of 2x2 neighbouring cells forms a block (A total of 9 overlapping blocks are generated).
4. Calculate the horizontal and vertical gradients (dx and dy respectively) for each pixel, $I(x, y)$ in the block using:
   $dx = I(x + 1, y) - I(x - 1, y)$
   $dy = I(x, y + 1) - I(x, y - 1)$
5. Calculate the gradient orientation, \$ for each pixel:
   $\$(x, y) = tan-1\ (dy/dx)$
6. Accumulate the orientations into 8 histogram bins

The eight histogram values from each block are concatenated to form a 72-dimensional HOG feature vector. This feature vector is used for the classifier training.

This method has two advantages: 1) Since the training examples are automatically generated from the background subtraction, they have various sizes. This method does not require any resizing or alignment, which is very important since these operations would not be very accurate in this uncontrolled setting. 2) By using a 2 × 2 grid, some spatial information is also captured.

5.2. SVM

SVM is based on the principle of structural risk minimization. For linearly separable data, SVM finds the separating hyper plane which separates the data within the largest margin. For, linearly inseparable data, it maps the data in the input space into high dimension space $x \in R^I \mapsto \Phi(x) \in R^H$ with kernel function $\Phi(x)$, to find the separating hyper plane. SVM was originally developed for two class classification problems. The N class classification problem can be solved using N SVMs. Each SVM separates a single class from all the remaining classes (One-vs.-rest approach).

Given a set of frames corresponding to N classes for training, N SVMs are trained. Each SVM is trained to distinguish a class and other classes in the training set. During testing, the class label y of a class x can be determined using:

$$y=\begin{cases} n, \text{if } d_n(x) + t > 0 \\ 0, \text{if } d_n(x) + t \le 0 \end{cases}$$

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Where, $d_n(x) = \max\{d_i(x)\}^N_{i=1}$, and $d_i(x)$ is the distance from x to the SVM hyper plane corresponding to frame i, the classification threshold is t, and the class label y=0 stands for unknown. The model for SVM based classification is given in Fig.6.

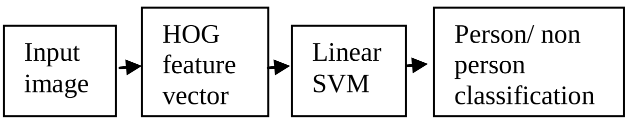


Fig 6.Work flow of the SVM approach

VI. Experimental Results

Exhaustive experiments are conducted to evaluate the performance of the proposed approach. Experiments were conducted on Intel Pentium IV processor with 2.67 GHz speed and the object identification was done in C++ with opencv library in [6].

6.1. Experimental Datasets

To examine the performance of the proposed method, data set was collected from a traffic surveillance video of six hour long, single day, and single view video and it is sampled at a resolution of 320 X 240 and at a rate of 20 frames per second. The moving sequences of objects were automatically extracted using background subtraction method.

6.2. Performance Metrics

6.2.1. ROC curve

The Receiver Operating Characteristic (ROC) curve is a common technique for assessing the performance of a classifier.

It is a plot of operating points showing the possible trade-off between a classifier's true positive rate (TPR) and false positive rate (FPR). The operating points are obtained by varying some underlying parameters of the classifier and then measuring its TPR and FPR.

The area under the ROC curve can also be used as an overall measure for the classifier's performance.

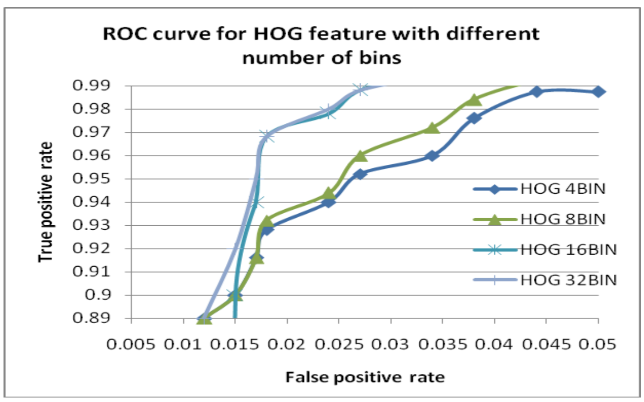
The analysis of the ROC curve can also help in the selection of a suitable operating point for the classifier implemented in the final object identification system.

6.2.2. Performance of the HOG Feature using Different Number of Histogram Bins

In the HOG feature extraction, the orientation angles are equally divided into several histogram bins. The number of histogram bins determines the fineness of the orientation's details captured by the features. Having too many bins may capture a lot of irrelevant orientations originated from the image's noise or the background clutter. It will also slow down the classification process. On the other hand, using too few histogram bins may miss out some important object's features and thus degrade the classification performance. The ROC curves for HOG feature with different no. of bins is depicted in Graph 1. Based on the graph, the area under the ROC curve is calculated and it is shown in Table 1 with the detection rate.

Detection Rate

From the ground truth results the detection rate is evaluated. It was calculated by counting the total number of objects detected over the total number of actual objects that appeared in the video frames.



Graph 1.ROC curves for HOG feature with different number of histogram bins

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Table 1.
Performance of HOG feature with different number of histogram bins

Number of histogram bins	Detection Rate(%)	Area under ROC curve	Processing Time(ms)
4	86.16	0.9874	13.32
8	90.81	0.9920	22.61
16	96.72	0.9968	61.02
32	97.18	0.9973	276.23

From the results, it can be seen that features with higher numbers of orientation bins perform better. When doubling the number of bins from 4 to8 and from 8 to 16, there are significant improvements in performance (better by 4.7% and 5.9% respectively). However, only a small improvement (<1%) is achieved when doubling the histogram bins from 16 to 32. These results show that some vehicle's features are represented in the finer orientation's details. However, at 16 bins, most of the salient orientation's features have been extracted. Therefore, increasing the histogram bins beyond this number will only gain a little improvement. Another important finding from this experiment is that the processing time for the classification increases drastically for every increase in the number of features. This is because the complexity of the classifier increases exponentially with the number of inputs.

6.2.3. Performance Evaluation of Twin Comparison Method

In this method, the sample image is first given into height-width based classification method. The aspect ratio is calculated based on the height and width of the bounding box. Again the same sample is given into the SVM based classification method and aspect ratio is calculated based on the height and width of the bounding box. Once the pedestrian is identified with the height-width based classification method's aspect ratio then it is compared with the SVM based classification method's aspect ratio. If both have similar aspect ratio of pedestrian then the given object is identified as a pedestrian, otherwise it is identified as non-pedestrian which is clearly represented in Fig 7 and Fig 8.

A video segment showing the object identification obtained with height-width based classification method and SVM based classification method are shown in Fig 9 and Fig 10. The drawback of the proposed system is that in some cases, one method identifies the object as pedestrian and another method identified the same object as non-pedestrian.

The aspect ratio is calculated by

$$\text{Aspect ratio}= \frac{\text{Width}}{\text{Height}} \qquad (1)$$

In our work, from the ground truth results, the performance can be evaluated by true positive rate and false positive rate. This quantity measures recall and localization. Based on the equation (2) and (3) the area of the ROC curve is measured.

TPR is defined by

$$\text{TPR}= \frac{\text{detected objects}}{\text{Total no.of objects}} \qquad (2)$$

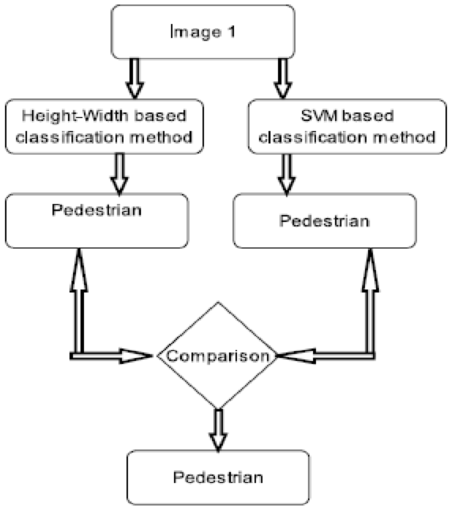


Fig 7.Framework for the twin comparison method in the case of both object are pedestrian

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

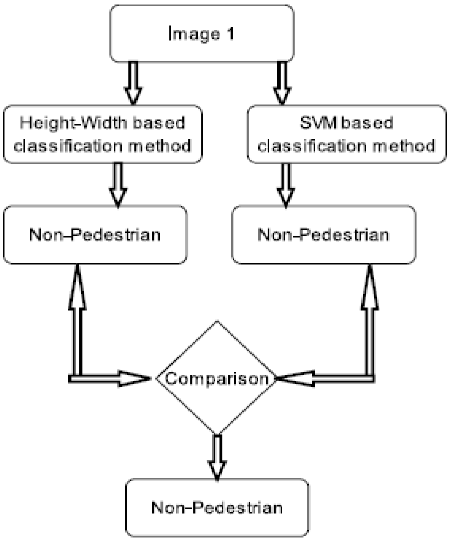
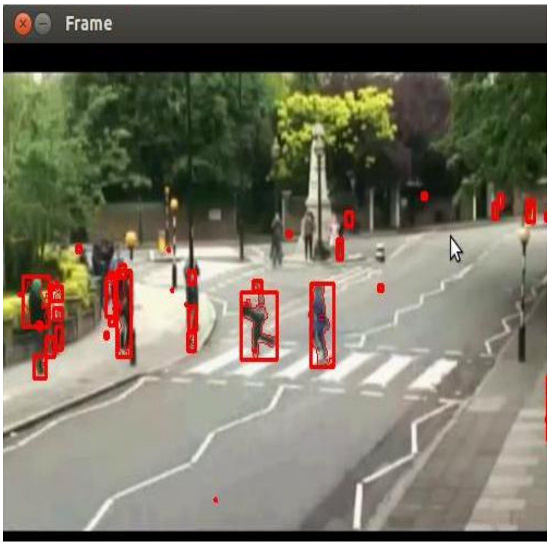
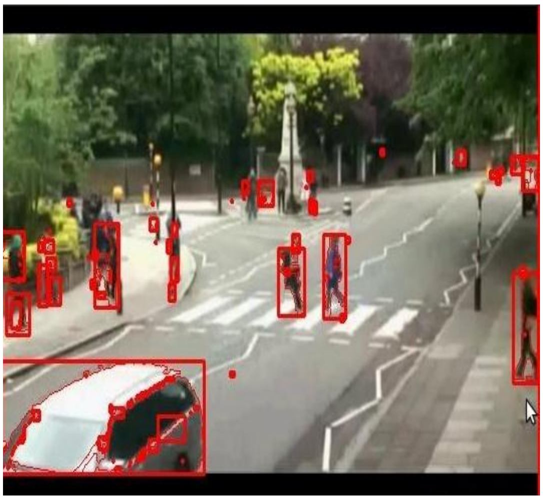


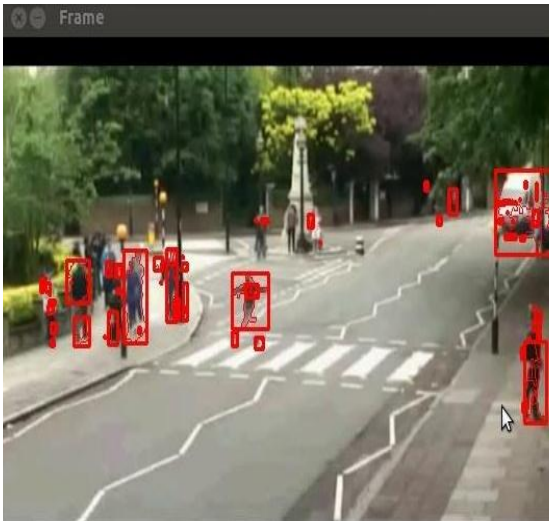
Fig 8. Framework for the twin comparison method in the case of both object are non-pedestrian



Frame (4)



Frame (10)



Frame (20)

Fig 9. Video frames showing the object identification using height-width based classification

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Frame(40)

Frame(46)

Frame(57)

Fig 10. Video frames showing the object identification using SVM based classification

The false positive rate FPR is the proportion of detection that were not true objects. We assess the FPR by dividing the number of false positives by the total number of detections. This is the percentage of erroneous detection. FPR is a measure of precision and localization. It is defined by

$$FPR = \frac{\text{false positives}}{\text{detected objects} + \text{false positives}} \quad (3)$$

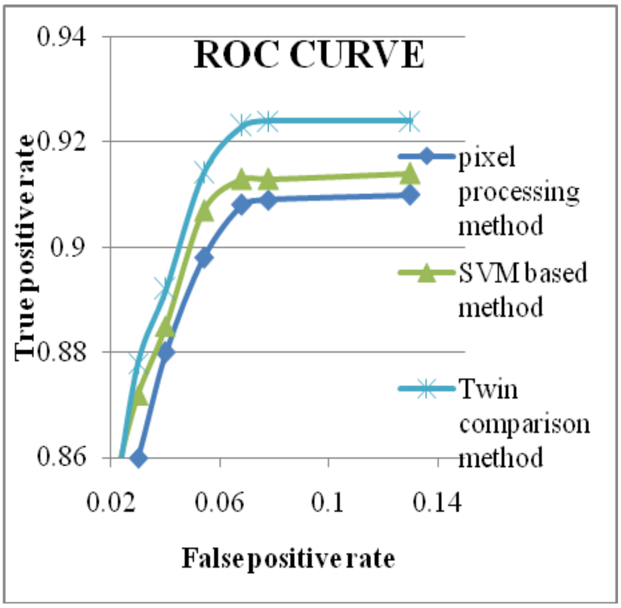
Table 1.
Performance analysis for all the methods

Identification/ Tracking System	TPR (%)	FPR (%)
Height width based classification method	91.3	6.8
SVM based clasiffication method	89.5	15.8
Twin comparison method	92.4	6.2

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

The Table 1 shows the True positive ratio and False positive ratio for both the methods. Based on the TPR and FPR, the ROC curve should be drawn which is showed in graph 2.



Graph2. ROC curves for pixel processing method, SVM based method and Twin comparison method.

VII. CONCLUSION

A twin comparison based object identification and tracking for automated video surveillance system has been introduced. Using a height-width based classification method, a object identification and tracking system has been implemented, and a thorough quantitative analysis has been presented. This is very simple and low time consuming method.

Classification is again done with the HOG feature based SVM classifier. The performance of the HOG feature is generally better than the other feature. The HOG feature also requires a shorter processing time since its feature set is more compact. The HOG reduced features set trained on the SVM, is selected to be implemented in the proposed object identification and tracking system. Based on both results the object is identified as a pedestrian or non-pedestrian. This dual comparison method gives a better result.

REFERENCES

1. Journal Papers

[1] Sayanan sivaraman, mohan manubhai trivedi " A General Active Learning Framework for on-road vehicle Recognition and tracking" in Transactions on Intelligent Transportation systems vol no 11, issue 2, pages 267-272,June 2010 IEEE.

[2] P.M.Daigavane and P.R.Bajaj," Real Time Vehicle Detection and Counting Method for Unsupervised Traffic Video on Highways", International Journal of Computer Science and Network Security, VOL.10 No.8, pages112-117August 2010.

[3] Rajiv Kumar NAth,swapan kumar Deb, " On Road Vehicle /Object Detection and Tracking using Template", on Indian Journal of Computer Science and Engineering vol. No .2 pages 98-107,2009.

[4] C. Wang and J.-J. J. Lien, "Automatic vehicle detection using local features—A statistical approach," IEEE Transactions on Intelligent Transportation Systems, vol. 9,Issue no. 1, pp. 83–96, Mar. 2008.

[5] A.Elgammal R.Duraisamy,D.Harwood and Larry S.Devis ," Background & Foreground modeling using non parametric Kernal density estimation for visual surveillance",on proceeding of the IEEE, vol.90,no 7, pages 1151 - 1163 july 2002.

2. Text Book

[6] The OpenCV Reference Manual Release 2.4.2 published on july 2012.

3. Conference Proceedings

[7] F.Dornaika,san Sebastian, Bilbao, P.Chakik,"Efficient object detection and Matching using Feature classification on Pattern Recognition",20th International Conference on Pattern Recognition(ICPR), 2010 pages 3073 – 3076.

[8] S.Schraml, A.N.Belbachir, N.Brandle,"A Real-Time Pedestrian Classification Method for Event-Based Dynamic Stereo Vision" in IEEE 2010.

[9] G. Jun, J. K. Aggarwal, and M. Gokmen, "Tracking and segmentation of highway vehicles in cluttered and crowded scenes", in IEEE Workshops on Applications of Computer Vision, pages 1-6,2008.

[10] Zhenjun Han, Qixiang Ye, and Jianbin Jiao," Online feature evaluation for object tracking using Kalman Filter" in 19th International Conference on Pattern Recognition, 2008. (ICPR 2008), pages 1 –4, 2008.

[11] X. Song and R. Nevatia," A model-based vehicle segmentation method for tracking" IEEE Proceedings of the 10th International Conference. on Computer Vision, pages 1124 - 1131 Vol. 2 2005.

[12] N Dalal, B Triggs,"Histograms of Oriented Gradients for Human Detection" in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005.

[13] C.Stagffer and W.E.L.Grimson, " Adaptive background mixture models for real-time tracking", in IEEE International conference on computer vision and pattern recognition,1999, pp.246-252

[14] B. Coifman and D. Beymer,"A real-time computer vision system for vehicle tracking and traffic surveillance",on Transportation Research Part C: Emerging Technologies,Volume 6, Issue 4, August 1998, Pages 271–288.