ISSN No:-2456-2165

An Intelligent System for Object Classification in Video Surveillance

Krishna priya P V
M.Tech CSE, Department of CSE
Sree Narayana Gurukulam College of Engineering ,
Kadayiruppu, Kerala, India

Dr. Smitha Suresh
Professor, CSE Dept.
SNGCE, Kadayiruppu, Kerala

Abstract:- Video surveillance in dynamic environment for detecting human and vehicles, is one of the current challenging research area in computer vision. It is a key technology for efficient management of traffic, public safety, military security. CCTV surveillance systems are more common nowadays because of its easy deployment. But the dependency upon human operator to monitor and classify the objects in the video sequence is very costly and inefficient. Thus an intelligent surveillance system which could detect and classify the moving objects are proposed using this system. Detection of moving objects in video stream is the first relevant step in classification. The proposed method uses Gaussian Mixture Model algorithm (GMM) for object detection and a shadow removal method is also used. The work mainly focused on object classification. Color, texture and gradient methods along with the kNN classifier used for object classification.

***Keywords**:- Visual surveillance, GMM, Shadow removal, Shadow detection, Object classification, LBP, HOG, kNN.*

I. INTRODUCTION

Nowadays , using video cameras for monitoring and surveillance is common in sensitive areas such as banks, highways, crowded places like airport, railway stations, bus stations etc. and thus increased the need and interest in video analysis. But the problem with these type of video surveillance is that they need a human operator to monitor them constantly. Thus the efficiency of these systems are mainly determined by the vigilance of the person monitoring them , not on the technological capabilities. To overcome the limitations of these surveillance methods, a major effort is under way in the computer vision to develop an automated or smart visual surveillance system for real time monitoring of people, vehicle and other objects so that they can monitor both immediate unauthorized behavior and suspicious behavior, and can alert the human operator immediately. So visual surveillance becomes a fast growing research area and topic in computer vision that tries to detect, recognize and to classify objects over a sequence of images. so, these monitoring systems can be improved using vision based techniques, able to extract and classify objects in the scene. The proposed method presenting a smart visual surveillance system for object classification, where there is less human intervention and the entire processing is done by the smart system.

The proposed smart surveillance system is smart enough to detect, classify and identify the real time moving objects in to their respective categories such as group of people, an individual, vehicles etc.. The automatic classification of objects in the video sequence helps to reduce the human effort and thus they can do the actions immediately if any critical situations arises.

The processing of the presented visual surveillance system includes various stages. The different stages of this system is clearly explained in Fig. 1 .

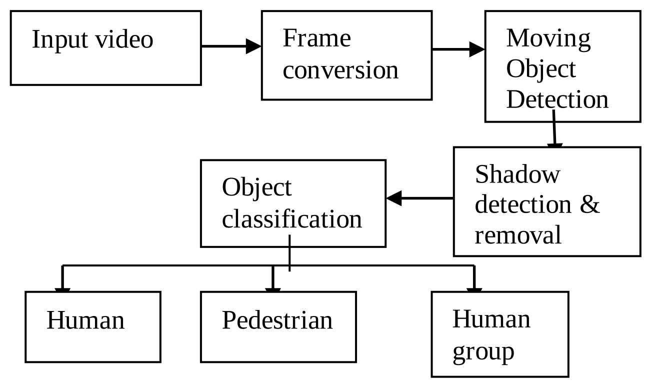


Fig 1:- Proposed System Design

II. LITERATURE SURVEY

Many researchers have introduced various techniques for moving object classification. These classification results have many applications in the real world applications such as the video surveillance, medical imaging, military communication, traffic monitoring and so on. In order to classify these objects in to their respective categories some pre- processing are needed such as object detection, shadow detection and removal. So there have been a number of surveys about moving object detection, shadow detection and removal , classification in the literature.

Moving object detection is the first step in the video surveillance application. Many methods are available for moving object detection, like background subtraction, optical flow, temporal differencing. Among these background subtraction is common. Background subtraction [1,2] or foreground extraction is a widely used method for moving object segmentation in static scenarios. GMM [1,2] is a type of commonly used background modeling and this method is based on pixel level. This method is more robust to the illumination changes and speed of the moving object than the other background subtraction methods like frame differencing, median filtering, etc.

Another problem occurs while classifying the moving objects is the presence of shadows. Shadows appear in frames when the light coming from the light resource is completely or partially occluded by elevated objects [3]. Y.I Shedlovska [3] proposed a new shadow removal algorithm based on boundary processing. In the existing work a threshold based method is used for shadow detection and then the shadow removal and boundary processing methods are applied to obtain a shadow free image. Aqel *et al* [4] presented a shadow detection and elimination model for traffic surveillance. Main focus is for the visual traffic surveillance system on highway through the elimination of moving cast shadows.

Dale Joshua[5] *et al* presented another shadow detection and removal method and used ViBe as background subtraction algorithm. The algorithm make use of the combination of HSV color space information, texture information and some post processing. In order to remove the candidate shadow pixels , the internal edge mask pixels are subtracted from the candidate shadow pixel mask. In another work done by Chankiang Wang *et al* a new technique [6] for effective detection and elimination of shadows on road was proposed and make use of the SVM based on color saliency space and gradient field.

Anoopa *et al* [7] proposed another shadow detection and removal method using a Tri- Class Based Thresholding model and shadow matting technique respectively. Another shadow detection and removal method available is based on an 'automatically learning the features in a supervised manner' using ConvNets[8]. A Bayesian formulation is used to extract the shadow matte accurately, using the detected shadow masks and to remove the shadows later. Shadows can be detected and removed [9] using texture based information.

Image classification can also be done using texture information and SVM[6,10] classifier along with HOG feature[10]. Color feature is one of the commonly used visual feature for object classification. Another existing work for object classification is based on color feature[11]. Hong-Son Vu *et al* proposed [12] an effective method for real time moving object detection and classification. Objects classification based on their intensity[13] value are also available in literature. Lun Zhang *et al* [14] proposed a method based on Multi-block Local Binary Pattern (MB-LBP), which can overcome the limitations of other classification methods. In [15] L´opez *et al* proposed a hybrid classification method which uses the static KNN clasifier and dynamic classifiers for classification.

III. PROPOSED METHODOLOGY

In the proposed system, or in any surveillance application, the foremost step in processing is to convert the captured video into successive frames. These frames are used for further processing. Here background subtraction is used as a moving object detection method. So that foreground and background images get separated and can obtain the target foreground pixel. But while processing with the foreground image, there may be chances to contain shadow in the foreground pixel. If so, the shadow region also will interpret as a part of foreground image. This may lead to false interpretation of the objects in the object classification stage. Therefore it is necessary to detect the shadow region from the foreground pixels and to remove this part from the foreground target image. This shadow detection and shadow removal stage can be considered as a pre- processing step in the video object classification applications. After the shadow removal stage proposed system will generate shadow free foreground image sequences. These image sequences can be used for classification task. The classification process need a classification algorithm and features need to be extracted. Using this features and classifier proposed method can categorize the detected objects into different classes such as pedestrian /human, vehicles, human group.

A. Moving Object Detection

First stage in the video analysis is the moving object detection. A number of methods exists for moving object detection, such as background subtraction, statistical methods, temporal differencing and optical flow.

To estimate the background is the first and the most important step in all the video surveillance applications. Background subtraction [2] or foreground extraction is a widely used method for moving object segmentation. In simple background subtraction is the process to separate foreground objects from the background objects. There are many methods in background subtraction. One of the effective method is Gaussian Mixture Model(GMM) [1] which is used in this work. The advantage of using gaussian mixture model is that this recursive algorithm estimates the parameters of the mixture and simultaneously selects the number of components for each pixel. Also this method is more robust to the illumination changes and can handle fast moving objects .

GMM is a type of background modeling and this method is based on pixel level in which every pixel in the background is is modeled using a Mixture of K-Gaussian distributions rather than using any other type of distributions. Normally K value ranges from 3 to 5. K value used in the proposed method is 4. The K value determines the changes in the background pixels. In order to use Gaussian mixture model or for learning the background model, need to define three gaussian parameters [1] namely weight (W), mean(μ), and variance(σ^2). These gaussian parameters are used to obtain the representations of the background.

Each and every pixel value is matched with current set of models to discover the match. If no match is found[1], the least model that is acquired is rejected and it is substituted by new Gaussian with initialization by the existing pixel value means the pixel value that don't suit into the background are taken to be background. If matching occurs[2];

ISSN No:-2456-2165

$$\frac{||X-\mu j||}{\sigma j} < \tau \, for \; some \; j\varepsilon[1..k] \qquad (1)$$

Where X be value of the given pixel in time and τ is some threshold. Next step is to update the gaussian parameters. Updation equations[1] for these parameters such as mean, weight and variance are given as follows:

$$w_k(t) = (1-\alpha)w_k(t-1) + \alpha M_k(t) \quad (2)$$

$$\mu_k\ (t) = (1-\alpha)\mu_k\ (t-1) + \beta X \qquad (3)$$

$$\sigma_k^2\,(t) = (1-\beta)\sigma_k^2\,(t-1)||X - \mu_k\ (t)\,||^2 \,(4)$$

Where M_k(t) is equal to one for the matching component j and otherwise it is zero. Then sort the gaussian by decreasing weight to standard deviation ratio to represent the background. A threshold is applied to the sum of weights to find the set {1….B} of gaussian modeling the background and can be defined [1,2] as

$$B = argmin_{KB}\ (\textstyle\sum_{k=1}^{Kb\le K} w_k\ > \lambda\,) \qquad (5)$$

The background components are determined by assuming that the background contains B highest probable colors. The probable background colors are the one which stay longer and more static and so they are considered as background.

B. Shadow Detection and Removal

Next most important step in the moving object classification for a video surveillance application is the shadow detection and shadow removal . In the first step of object classification that is moving object detection, the methods like background subtraction cannot resolve the problems caused by shadows.

Proposed work make use of the texture based information. Texture method is found to be effective for this work, because this feature requires only less processing time than the color space information method. Shadow detection algorithm[17] consist of three steps.

- *Edge Detection*

An edge detection method along with bilateral filtering is used for detecting the shadow region from the foreground image pixels. Canny edge detector is used for edge detection, which can detect background and foreground edge. Bilateral filter is used to preserve the edges. Multi threshold edge detection [17] is applied in which multiple thresholds varying from 0.9,0.8,…,0.2 is used.

- *K_H Calculation*

K_H [17] is defined as the spectrum ratio and is a constant independent of wavelength and also is different in H=R,G,B channel. This step uses the dilated regions along with the edge detected regions to compute the K_H value. K_H value can be computed using the following equation.

$$K_H = (\frac{F_H + 14}{f_H + 14})^{2.4}$$

Where F_H and f_H are the gamma corrected pixel values of the non shadow and shadow parts. To overcome noise and obtain fast computing, for each edge, we calculate the mean pixel values on its both sides in each channel to determine the spectrum ratios. After this the detected regions are enlarged by using dilation operation with six pixels on both sides.

- *Edge Verification*

The following two criteria should be satisfied for edge verification [17].

- Criteria 1: K_R>K_G>K_B
- Criteria 2:

$$\begin{cases} K_R - K_G > \varepsilon \wedge K_G - K_B > \varepsilon \;\; If\ K_R > K_R^{80} \\ K_R - K_G > \frac{\in}{2} \wedge K_G - K_B > \frac{\in}{2} \;\; If\ K_R \le K_R^{80} \end{cases}$$

Where ε is the lower bound for the difference between K_R and K_G and based on the difference between $K_R - K_G < K_G - K_B$, ε value can be defined as

$$\varepsilon = \frac{K_R^{80} - K_R^{80}}{2}$$

Once the shadow regions are detected, the next step to do is to perform necessary to remove these detected shadow regions. If the above two criteria satisfy, the edges will be verified and classified as shadows .The shadow regions can be removed by applying a ratio calculation method. In the proposed method the ratio between the shadow and non shadow region is calculated. Then this value is multiplied with the adjacent pixels of the shadow region, so that the pixel value will become same as that of the non shadow region.

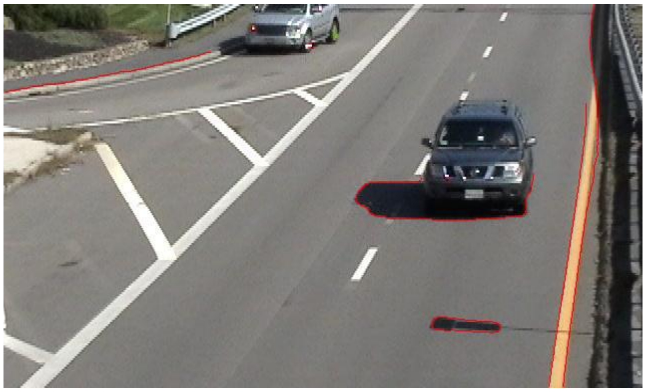


Fig 3.1:- Shadow Detection

Fig 3.1. shows the shadow detected image using the proposed method and Fig. 3.2 shows the shadow removed frame.

Fig 3.2:- Shadow Removal

C. Moving Object Classification

The proposed method mainly concentrate on the moving object classification. Due to some uncontrollable environment conditions, shadows, or illuminations changes makes outdoor object classification challenging. Ultimate aim of most of the video surveillance application is object classification. The main target of interest in the surveillance applications are generally human, vehicles or human group. Categorizing the detected moving or non moving objects is the crucial step of any visual surveillance applications. Here proposed a new classification approach which combines the texture feature, gradient feature and color feature information. Local Binary Pattern (LBP) is used for texture feature extraction , Histogram Oriented Gradient (HOG) feature is used as the structure feature extraction means and for getting the color information histogram count is used. The first step of every classification system is to train a dataset based on the features mentioned. After that the features are extracted for the detected image and then it is fed into the classifier and the classifier compares this with the dataset and classify the data based on this. The proposed work make use of K-Nearest Neighbor (KNN)classifier. Since we use the combination of these three feature extractions, the accuracy of the object classification will be high. The use of the classifier is to train the dataset provided. The functional block diagram of classification process is shown in Fig. 3.3.

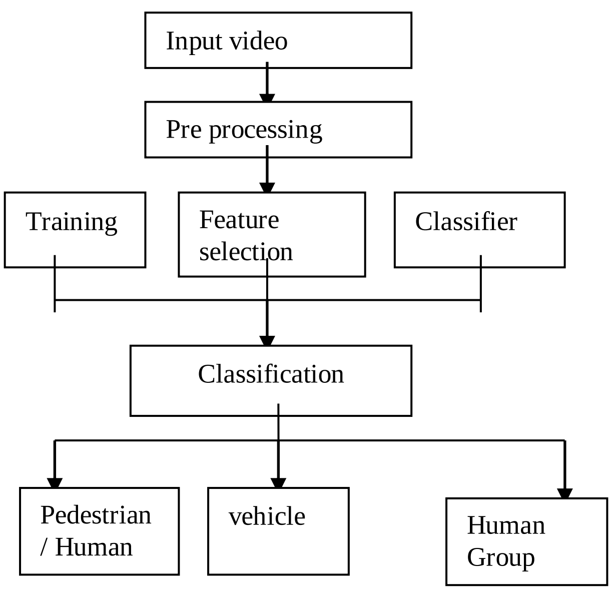


Fig 3.3:- Functional Block Diagram of Classification process

A training dataset is created based on the features selected (Histogram, HOG, LBP) for object classification. In the classification step, after detecting the moving object, extract the respective features like HOG, LBP, Histogram of the image.

- *LBP*

Local Binary Pattern or LBP [16] is the most commonly used texture classification feature. LBP gray-scale invariant texture measure computed from the analysis of a 3x3 local neighborhood over a central gray pixel. This code is built by thresholding a local neighborhood by the gray value of its center. The neighbors are labeled using a binary code {0, 1} obtained by taking difference of values to the central pixel value. If the tested gray value is below the gray value of the central pixel, then it is labeled 0, otherwise it is assigned the value 1. LBP code at location (x, y) is obtained using the following equation [16]

$$LBP_{x,y} = \sum_{b=0}^{B-1} S(P_b - P_c)2^b \qquad (6)$$

$$S(z) = \begin{cases} 1 & z > 0 \\ 0 & z < 0 \end{cases} \qquad (7)$$

where *Pc* is the pixel value at *(x, y)* or the central pixel, *Pb* is the pixel value estimated using bilinear interpolation from neighboring pixels in the *b*-th location on the circle of radius *R* around *Pc* and *B* is the total number of neighboring pixels. The obtained value is then multiplied by weights given to the corresponding pixels. The weight is given by the value 2^b. Summing the obtained values gives the measure of the LBP.

- *HOG*

HOG[10,18] feature representation is especially appropriate for some objects which have strong edges and corners such as vehicles and people. HOG computation is shown in Fig. 3.4

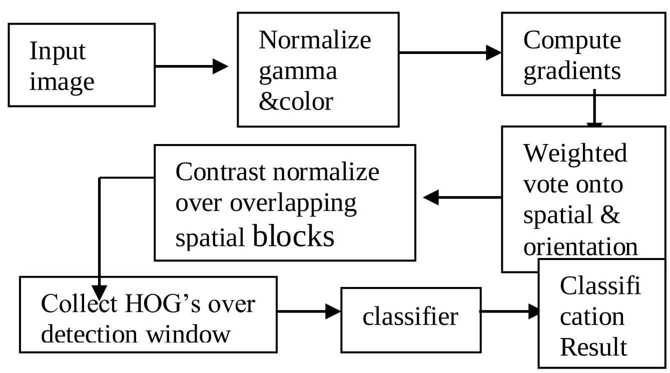


Fig 3.4:- HOG Computation(Source:[18])

- *Histogram Count*

Histogram count is one of the common color visual feature for feature selection process. In the proposed system, histogram of the RGB color space is taken and it is represented as

$$FeatVector=[Hist_{red}, Hist_{green}, Hist_{blue}]$$

ISSN No:-2456-2165

This Feat Vector can be used along with the other two features namely LBP and HOG.

- *kNN Classifier*

k-Nearest Neighbor is the simplest among the machine learning or image classification algorithm. In case of image classification, labels are associated with each image, therefore it is easy to predict and return the actual category of the particular image. This algorithm simply relies on the distance between feature vectors. Usually Euclidean Distance is used for obtaining the K value. Euclidean distance finds the distance between two pixels. Euclidean distance between two points can be find using equation

$$ED = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (8)$$

After extracting the features the method compares the extracted feature with the data in the dataset. If there is a match occurs with any of the items in the dataset, then the KNN classifier will classify the data item into its resultant class. Here the classification is said to be stronger because a double comparison is performed before the final classification result, which means the result obtained after the KNN classifier is again checked with the bounding box ratio. A different ratio range is predefined for every class, human/pedestrian, vehicle or human group .Since this method uses a double classification, the accuracy of the classification result is comparatively high with the other classification methods.

IV. RESULT AND ANALYSIS

The proposed method has been tested on different videos under different conditions. The results of technique presented in this work have been analyzed under qualitative and quantitative point of view. The analysis result proven that the proposed method works well in different scenarios of video surveillance. Fig. 4.1 shows some the original frames of video1. Video 1 is a standard traffic surveilance video taken during a sunny day.

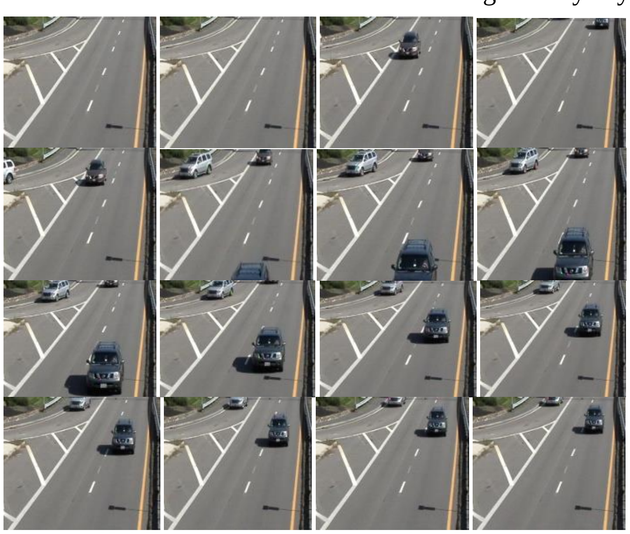


Fig 4.1:- Video 1 original Frames

After applying the shadow detection and removal method presented above, the shadow regions are removed and obtained shadow free image. Fig 4.2 shows the result and analysis of shadow free image with histogram.

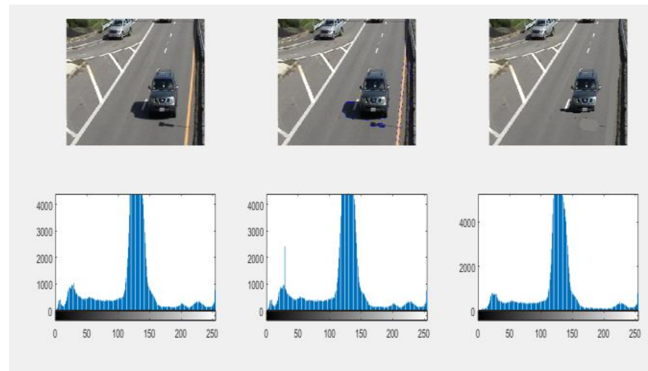
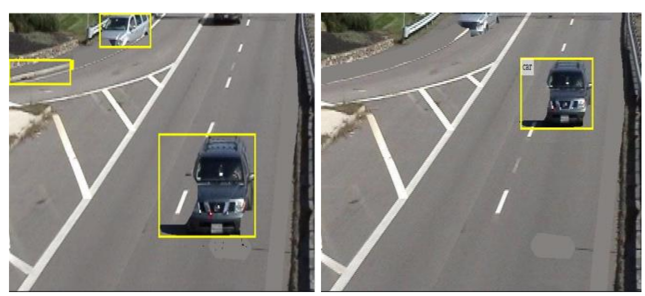


Fig 4.2:- Shadow detection , removal and its corresponding histogram

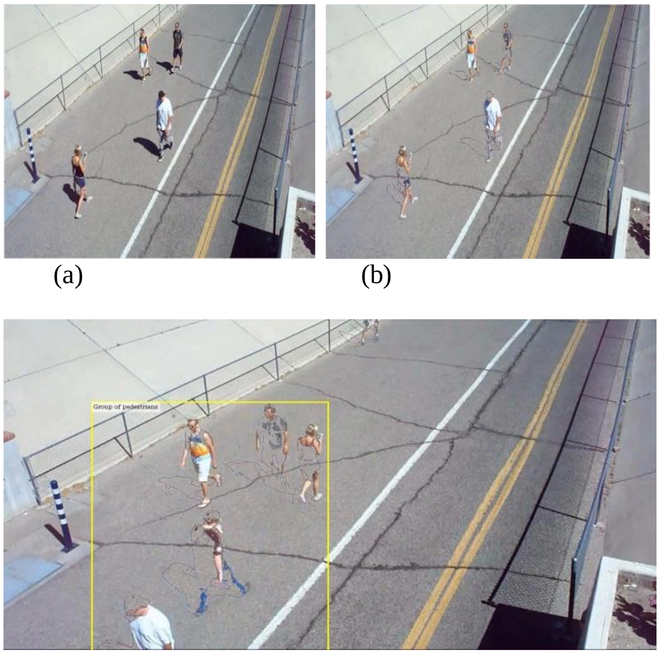
In Fig. 4.2 , image in the first column is the original frame and its histogram , the second column shows the shadow detected frame and the third column represents the shadow free image with its histogram. Next stage is the classification stage. The output obtained after the shadow removal stage is fed into classification stage. In this stage the classifier classify the item by extracting the features and comparing it with the training set. In this work we are using two classes, they are ‘car/Vehicle’ and ‘Pedestrian’. The result of classification stage is shown in Fig. 4.3. Fig 4.3 (a) shows the detected object for classification and its classification output is shown in Fig. 4.3 (b). Using the proposed method the classifier classify the detected object as ‘car’.



(a) (b)

Fig 4.3:- (a) Detected Object (b) Classification

The proposed method is tested for different videos and all those videos shows a satisfactory result. Fig 4.4 shows the results a Video 2.



(a) (b) (c)

Fig 4.4:- (a) Original Frame (b) Shadow Free Image (c) Object Classification

Fig. 4.4 (a) shows the original frame of Video 2 which is a standard street video dataset. Fig. 4.4 (b) shows the result after applying the shadow detection and removal method. Fig. 4.4(c) shows the classification result of video 2, which classify the shadow free detected objects as 'Group of pedestrian'. Detection accuracy and classification accuracy can be calculated using the following equations.

$$detection\ accuracy = \frac{No\ of\ detected\ objects}{total\ number\ of\ objects} * \ 100$$

$$class.accu = \frac{total\ no\!:of\ correct\ classification}{Total\ number\ of\ detected\ objects} * 100$$

Table 4.1 shows the detection and classification accuracy of different videos

Video	Detection Accuracy(%)	Classification Accuracy(%)
Video 1	100	100
Video 2	80	60
Video 3	100	100
Video 4	90	80
Video 5	100	80

Table 4.1 Accuracy of Videos

V. CONCLUSION

In the proposed work , an advanced method for moving object classification is presented. To make an efficient moving object classification, in to their respective categories some pre-processing is done which include object detection and shadow removal. To classify objects we extract a combination of texture, gradient and color based information and fed into kNN classifier. Combination of these three features makes the classification result more accurate. The results of the techniques presented in this work have been analyzed under qualitative and quantitative point of view. The analysis results proved the efficiency and accuracy of this classification techniques with low processing requirements. In future, this work can be extended to automatic capturing of the identity information of the detected objects such as a pedestrian.

REFERENCES

[1]. Addesselam Bouzerdoum, azeddine beghadi, Son lam Phung, Philippe Bouttefroy, "On the analysis of background subtraction techniques using Gaussian mixture models", IEEE International Conference on Acoustics, Speech, and signal processing, USA, 2010.

[2]. C. Stauffer and W.E.L Grimson. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, August 2000.

[3]. Y.I Shedlovska, V.V Hnatushenko "Shadow Removal Algorithm with Shadow area Border Processing", International Young Scientists Forum on Applied Physics and Engineering, 2016.

[4]. Siham Aqel, Abdellah Aarab, My Abdelouahed Sabri "Traffic video surveillance: background modeling and shadow elimination" IEEE 2016.

[5]. Dale Joshua R. Del Carmen and Rhandley D. Cajote "Moving Shadow Detection and Removal for Video-based Traffic Monitoring" APSIPA, 2016.

[6]. Chunxiang Wang, Liuyuan Deng, Zhiyu Zhou, Ming Yang, Bing Wang, "Shadow Detection and Removal for Illumination Consistency on the Road", International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2017.

[7]. Anoopa S, Dhanya V, Dr. Jubilant J Kizhakkethottam, "Shadow Detection and Removal Using Tri-Class Based Thresholding and Shadow Matting Technique", International Conference on Emerging Trends in Engineering, Science and Technology, 2015.

[8]. S. H. Khan, M. Bennamoun, F. Sohel and R. Togneri, "Automatic Shadow Detection and Removal from a Single Image", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

[9]. Shiping Zhu, Zhichao Guo and Li Ma, "Shadow removal with background difference method based on shadow position and edges attributes", EURASIP Journal on Image and Video Processing , 2012.

[10]. Seung-Hyun Lee, MinSuk Bang, Kyeong-Hoon Jung and Kang Yi, "An Efficient Selection of HOG Feature for SVM Classification of Vehicle", IEEE International Symposium on Consumer Electronics (ISCE), 2015 .
[11]. T. Balaji, "An Automatic Color Feature Vector Classification Based on Clustering Method", International Journal of Research and Scientific Innovation (IJRSI) Volume IV, Issue IV, April 2017.
[12]. Hong-Son Vu, Jia-Xian Guo, Kuan-Hung Chen, Shu-Jui Hsieh, and De-Sheng Chen, " A Real-Time Moving Objects Detection and Classification Approach for Static Cameras", International Conference on Consumer Electronics-Taiwan, 2016.
[13]. Adesh Hardas, Dattatray Bade , Vibha wali "Moving Object Detection using Background Subtraction Shadow Removal and Post Processing.", International Conference on Computer Technology (ICCT) 2015.
[14]. Lun Zhang, Stan Z. Li, Xiaotong Yuan and Shiming Xiang, "Real-time Object Classification in Video Surveillance Based on Appearance Learning", Center for Biometrics and Security Research & National Laboratory of Pattern Recognition, IEEE 2015.
[15]. Miriam M. L´opez , Lucio Marcenaro and Carlo S. Regazzoni, "Advantages Of Dynamic Analysis In HOG-PCA Feature Space For Video Moving Object Classification", ICASSP 2015.
[16]. Li Liu , Lingjun Zhao , Yunli Long , Gangyao Kuang , Paul Fieguth, "Extended local binary patterns for texture classification", Image and Vision Computing, 2012.
[17]. Jiandong Tian, Xiaojun Qi, Liangqiong Qu, YandongTang "New spectrum ratio properties and features for shadow detection", journal of pattern recognition , 2015.
[18]. Dalal, N., Triggs, B., "Histograms Of oriented gradients for human detection" In: Proc. CVPR 2005, vol. 1.