A

DISSERTATION REPORT

On

Moving Object Detection and Classification for Visual Surveillance System

By

MAHESH KUMAR

2015PEV5121

Under the supervision of

Mr. RAKESH BAIRATHI

Associate professor, ECE Department
MNIT, Jaipur

Submitted in partial fulfillment of requirement of degree of

MASTER OF TECHNOLOGY

to the

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

JULY – 2017

Department of Electronics and Communication Engineering
Malaviya National Institute of Technology, Jaipur

Certificate

This is to certify that this Dissertation report entitled ***"Moving Object Detection and Classification for Visual Surveillance System"*** by **Mahesh Kumar** (2015PEV5121), is the work completed under my supervision and guidance, hence approved for submission in partial fulfillment for the award of degree of ***Master Of Technology*** in ***VLSI Design*** to the Department of Electronics and Communication Engineering, Malaviya National Institute of Technology, Jaipur in the academic session 2016-2017 for full time post-graduation program of 2016-2017. The contents of this dissertation work, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Mr. RAKESH BAIRATHI
Associate professor, ECE Department
MNIT, Jaipur

Declaration

I, hereby declare that the work which is being presented in this project entitled "Moving Object Detection and Classification for Video Surveillance System" in partial fulfilment of degree of Master of Technology in VLSI is an authentic record of my own work carried out under the supervision and guidance of Mr. Rakesh Bairathi in Department of Electronics and Communication, Malaviya National Institute of Technology, Jaipur.

I am fully responsible for the matter embodied in this project in case of any discrepancy found in the project and the project has not been submitted for the award of any other degree. I also confirm that I have consulted the published work of others, the source is clearly attributed and I have acknowledged all main sources of help.

(MAHESH KUMAR)

Acknowledgement

I am grateful to my supervisor Mr. Rakesh Bairathi for his constant guidance and encouragement and support to carry out this work. His excellent cooperation and suggestion provided me with an impetus to work and made the completion of work possible. He has been great source of inspiration to me, all through. I am very grateful to him for guiding me how to conduct research and how to clearly & effectively present the work done.

I would like to express my deepest sense of gratitude and humble regards to our Head of Department Prof. K.K Sharma for providing necessary facility in the Department. I am very thankful to Mr. Bharat Kashyap, Lab Assistant and all other faculty members of ECE, MNIT for their valuable assistance and advice.

I would also like to thank my friends for their support in discussions which proved valuable for me. I am indebted to my parents and family for their constant support, love, encouragement, and sacrifices made by them so that I could grow up in a learning environment.

Finally, I express my sincere thanks to all those who helped me directly or indirectly to complete this work successfully.

(MAHESH KUMAR)

Abstract

Detection of moving the object and classifying the object as a particular class like human, car, motorbikes, etc., under the visual surveillance system is a very challenging task in the computer vision. The surveillance system help in monitoring the activities of a human at different places and vehicles in traffic. In this thesis, we do the detection of moving object and classification using supervised learning classifier. The detection of moving object is done by subtracting the current frame with the background frame which needed background modeling before extracting the object to obtain the background frame. The foreground image of the moving object is obtained using the global thresholding. For the classification of the detected object Support Vector Machine (SVM), a supervised machine learning algorithm is used to classify the detected object. To do this, first set of images of different class in database is made then features are extracted from those images. For extracting the key points, we have used Bag of Words (BOW) to image database as well as to the detected images. The Bag of Words also known as Bag of Features is based on Speeded up Robust Features (SURF) algorithm for features extraction from the images. Now the features obtained using the Bag of Words are labeled with respect to their class and then used to train SVM classifier to create model which gives the accuracy of probability of occurrence of 89.6%. Then image of the detected moving object is tested which gives as result the trained class.

Contents

List of Figure

Glossary

ANN: Artificial Neural Network

BOW: Bag of Words

HOG: Histogram Oriented Gradients

kNN: k – Nearest Neighbour

MSER: Maximally Stable External Region

SIFT: Scale–Invariant Feature Transform

SURF: Speeded Up Robust Features

SVM: Support Vector Machine

Chapter – 1

Introduction

1.1 Introduction

The incorporation of effective image capture, improvement, and processing techniques with computer vision and machine learning technologies have in recent times developed intelligent systems that can impersonate human activities, instinct, and decision making power. Such incorporated technologies have found their use to many application ranging from video surveillance, smart phones, electronic devices, electronic toys, automated vehicles, security systems, entertainment and games industry, etc. In general, an application that can be pleased by the mutual power and capability of the human brain and eyes can be obliged with such technologies.

For executing the above technologies, many important areas of computer vision, image processing, and machine learning have shown rapid growth. As a result, the captured images are of perfect quality, noise free, super resolution and of perfect clarity, for example. Once in the imaging system, perfect quality image is captured, it further needed to be processed to improve the quality and to remove any piece that might have created by the captured devices. The Algorithm in image processing such as for color correction/balancing, distortion removal, white balancing, sharpening, de-blurring, filtering and morphological algorithms are capable of improving the quality of images further, removing any objects that might have presented in the image due to any restriction of the image capture system. When the images are processed, one needs as next stage algorithm of computer vision such as motion detection, foreground object extraction, object detection, anti-shake algorithms, etc., to detect important objects in the frame. After the detection of an object, machine learning algorithm is required to classify their type and for analyzing the behavior. This thesis gives importance in general on moving object detection and object classification technologies.

1.2 Motivation

Visual surveillance is the most dynamic research field in computer vision for humans, animals, and vehicles. Understanding activities of objects moving in a scene by the use of video is both a challenging scientific problem and a very fertile domain with many promising applications. Our aim is to make a visual surveillance system that detect not only a moving object but also classify its class to which it belongs just like human brain does after seeing them so that it can be used in like self-driving car as future applications.

1.3 Overview of the Dissertation

The thesis is organized as follows:

Chapter 2 discusses available different algorithms or previous work which was done previously by researchers on motion detection and classification. Here, we analyzed a several different papers and see their methods to be followed.

Chapter 3 presents various concepts of image processing and different algorithm used for detection of moving object and it also give the concepts on thresholding for segmentation.

Chapter 4 is dedicated to understanding the basic concepts on different classification types and it also discuss the concept of feature extraction using Bag of Words.

Chapter 5 implementation of algorithm on moving object is done first with showing flow chart next then feature extraction is done. After this classification part is done.

Chapter 6 Shows simulation and result and also discussed the conclusion and future work.

Chapter – 2

Literature Survey

2.1 Introduction

Some important terms commonly used in image processing related research is defined in this section. Mainly, the terms like computer vision, object detection, object classification and machine learning are defined.

Definition [16]:

Computer vision is a field that concerned with how computers can be made for acquisition of a high-level understanding of images or videos. It includes methods for acquiring, processing, analyzing, and understanding each image and in general, and in general it deals with automatic extraction, analysis, and understanding of meaningful information from single image frame or sequence of the image frame.

In terms of computer vision, Object detection means detecting occurrences of semantic objects of a definite class like human, birds, motorbikes or cars in image or video.

In the field of computer vision, terminology, Object classification (also known as object recognition) refers to Classification of an apparent object into the definite class of any recognizable human type, like for example motorbikes, car, bird, human, etc. Object classification needs more detail analysis of features for classifying the object into a particular class.

2.2 Literature review:

Shanyi Liu, et al. [1] uses spatial-temporal gradient to detect the moving object. Due to the changes and complexity of the background constraints of refinement mechanism has been adjusted. In this spatial gradient mean has been calculated to improve the detection accuracy by regulating the detection radius. The changes in the background have been detected based on the temporal gradients of the sequential frames.

Hong-Son Vu et al. [2] has proposed a method by merging temporal domain with spatial domain methods to obtain an optimized result of human detection and recognition in real time by using the static camera. First, they used background subtraction to obtain the moving object instead of detecting whole scene image by sliding window. Then for classification purpose, they use the AdaBoost algorithm.

Fergus et al. [3] have given unsupervised learning methods to recognize the object. In this recognition of an object is done in the image without labelling and segmenting the object in the scene, i.e., in an invariant scale manner. To extract the feature entropy- based feature detector, shape, appearance, occlusion, and relative is scale is used. Then its model is used to classify an object in base a manner.

Osama Masoud et al. [4] has proposed an algorithm to detect the vehicle in traffic and classify it in captured videos. They modelled the vehicles by rectangular patches based on the behavior of the vehicles.

Omar et al. [5] used to detect the moving object for videos surveillance system. They used the background subtraction approaches to detect the moving object. In this they decomposed each gray level image obtaining from video into two structure and texture component and then used structures component to surveillance the moving object under fixed camera.

Mori et al. [6] has utilized laser technology to recognize the object based on grid trajectories. On the grid map by voting the scan points, the trajectories of grid are obtained which not only provide scan segmentation but also shows the dimension and speed of the moving object. On the grid trajectory the moving object is classified as a person, a motorbike, a car.

Chapter - 3

Concepts of Image Processing and Detection

3.1 Introduction

In image science technology, image processing is processing of images using mathematical procedures by using any form of signal processing for which the input is an image, a series of images, or a video, such as a snapshot or video frame. Image processing basically refers to digital images processing which uses Mathematical methods and computer algorithms to accomplish image processing on images. As a result, it allow to apply many algorithms on the input image to remove the noise during processing. It is the extraction of meaningful information from digital images for analysis of the features of the image. It can involve simple applications such as reading bar coded tags or more complex application such as face detection. Image processing does not only consist of computer algorithm but also it contains mathematical techniques.

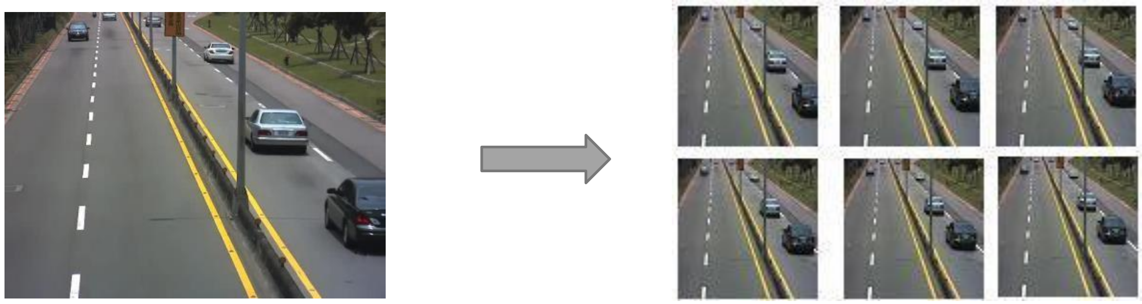
The applications of image processing is continuously expanding in all areas of science such as- Defense, Robotics, machine vision, filtering, optical character recognition, Security, medicine, face detection, thump impression detection in offices, etc. Some of techniques uses in Image Processing are as follows:

3.2 Image Processing

3.2.1 Video to Frame Conversion

Video is basically nothing but a sequence of many frame at the standard rate of 24 frame per second. Since work of image processing is done mainly on the each frame we convert the video into sequence of frames.

We extract the frame from video step by step each.

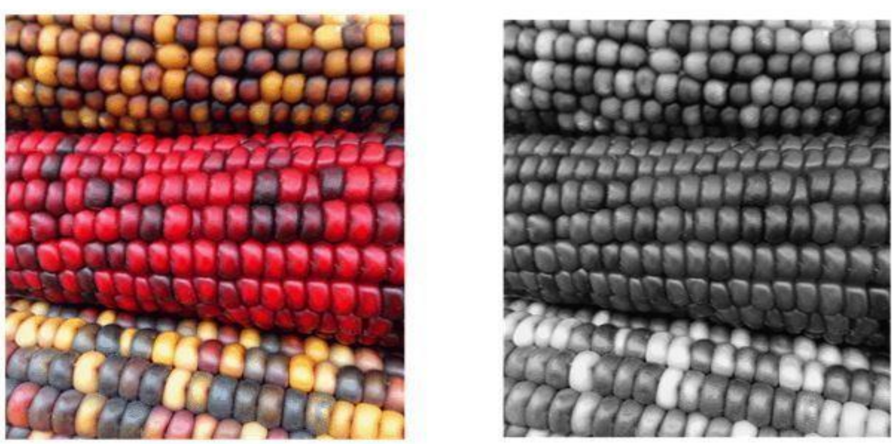


Video Sequence of Frames

Figure 3-1: Video to frame conversion

3.2.2 RGB to GRAY

It is technique which is used to convert RGB image to Gray image which contains each pixel in single sample i.e every pixel of the image contains only intensity information. Image of this short is also known as Black and White images, which consist of shades of gray from very low value of black to high intensity of white. The transformation of RGB to Grayscale image is done by removing the saturation and hue data while retaining the luminance of the image [32].



(a) (b)

Figure 3-2: shows RGB colour image of corn converted to Grayscale image [32]: (a) RGB colour image and (b) Grayscale image

3.3 Motion Detection

In today's world, video surveillance is used at each place whether it is maintained by governments, private or public organizations to fight against crime and terrorism, public safety at airports, bus stand, railway station, hospitals, and malls. Video surveillance is also used in traffic surveillance for effective management of transport movement and road safety. A video surveillance system includes many processes like motion detection, object classification, tracking, person identification, and object recognition. Out of these process, moving object detection is the first and most important stage and fruitful separation of moving foreground object from the background which further help to result effectively in object classification, tracking, person identification, and behavior analysis.

Motion detection has been categorized by W. Hu et al. [7] into three methods of classes as follows:

1. Temporal Differencing
2. Optical Flow
3. Background Subtraction

3.3.1 Temporal Differencing

In the temporal differencing method, the differences of pixels are used for two or more consecutive frames in the video sequences to detect moving regions. Temporal differencing is very adaptive to active environments, but usually, doesn't work to detect all relevant pixels of moving object when the object has uniform texture or move slowly but when the object stop temporal differencing methods fails to detect the changes in the two or more consecutive frames. In real time video Lipton et al. [8] detect moving object using temporal differencing. After obtaining complete difference between a previous frame and the current frame, a threshold value is used to find the change in the result. Then by using connected component analysis, the extracted moving regions are assembled into moving regions. Instead of two frame difference, three frames are used to improve.

Let $G_n(t)$ be gray level intensity value at t pixel position in n frames of the video image sequence G and $G_{n-1}(t)$ be the gray level intensity value at t pixel position and T be the threshold value whose value is initialized with pre-determined value. Then a pixel can be moving in moving regions if it satisfy the following condition.

$$|(t) - G_{n-1}(t)| > T$$

3.3.2 Optical Flow

This algorithm of motion segmentation detects moving regions in an image sequence by using the flow vector characteristics of moving objects over time. Meyer *et al.* [9] for extracting the artificial object calculated the displacement field vector to reset a silhouette based tracking algorithm, called active rays. The results obtained are then used for gait analysis. In the presence of moving camera this method can be used to detect independently moving objects.

However, most optical flow calculation methods are computationally complex and very delicate to noise, and so can't be used without a particular hardware in the real time.

3.3.3 Background Subtraction

Background subtraction based motion segmentation is a very popular method, mainly under those circumstances with a comparatively stationary background. In this method, the main and first step is to build a background (static) scene. Then to detect the moving regions called foreground in an image comparison is done pixel by pixel by taking the

Difference between each image sequence with background model first then after that the updated version of the background model.

This algorithm has conventionally following approaches using background subtraction for detecting moving objects in videos from static omnidirectional camera [10] [11] [12].

3.3.3.1 Using Frame Differencing

This method of background subtraction is the simplest way to segment the moving object from the background image. In this, an image is taken as background image denoted by B and then current frame C(t) obtained at the time t, is then taken to compare it with background image where corresponding pixels value in both C(t) and B at the same location is subtracted. This difference of image obtained shows some intensity in the pixel locations that has changed in the two frames. After this a threshold value T is used to filtered the pixels intensities difference in the image between C(t) and B.

In mathematical form, it is written as:

P(F(t)) = P[C(T)-P[B]

$$|P[F(t)] = P[C(t)] - P[B]| > \quad T$$

3.3.3.2 Approximate Median

This is second method of background subtraction to segment the moving object from the background image. The principle of the method is the simple subtraction between median frame and next frame. In this there is absent of the background modeling. Accuracy in this method is moderate and take computational time low to moderate.

3.3.3.3 Running Gaussian Average

This approach is based on the principle of Gaussian Probability density of pixels on the most current number of frames It is considered more on suitable for real time applications. In this the statistical computations consume more time.

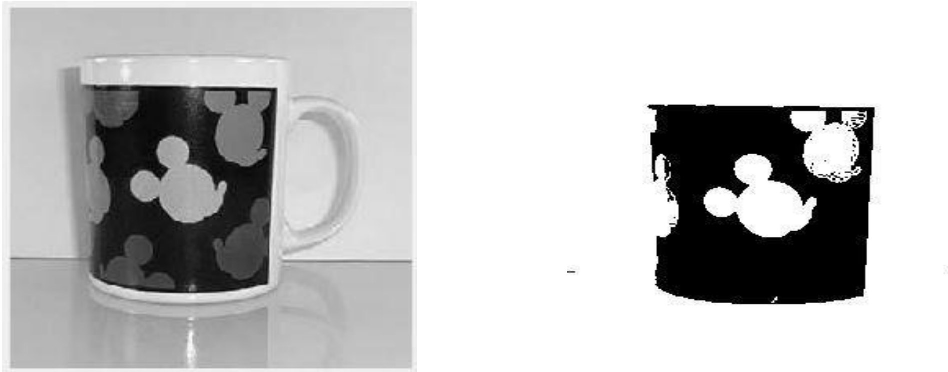
3.3.3.4 Background Mixture of Gaussian

This method of background subtraction is based on approaches by forming each pixel as mixture of Gaussians. For updating the model, it uses online approximation. In this process, assumption is made that intensity of every pixels in the video can be modeled using Gaussian mixture. Here intensity probable to background is determine experimentally and then pixels not matching to it is referred as foreground pixels.

3.4 Threshold

After the result of motion detection algorithm, threshold play an important role in segmenting the image. Whether a pixel represent the moving object or it belongs to a background in the image is defined using a threshold value. Since the object in a gray image describe gray levels many methods of thresholding extract moving object from their background based on the statistics of one-dimensional (1D) histogram of gray levels and two-dimensional (2D) histogram of gray [13].

The commonly used method for thresholding is Otsu method [14], which chooses the global peak threshold by maximizing variance difference between classes based on least square method. This algorithm assumes that the gray- level image consist of foreground and background pixels following bi-modal histogram. Then computes optimum threshold by separating the two classes so that their inter class variance is greatest or their variance of intra class is minimal. Using the Otsu Thresholding, a gray scale image can be shown as in figure 3.3.



(a) (b)

Figure 3-3: shows the image after thresholding. (a) Gray image. (b) Image after thresholding

The mathematical form of Otsu Thresholding is as below:

Suppose that an image in L gray levels is represented [0,1,…L-1]. p_j is the number of pixels at level j and N is the total number of pixels given by N = $p_1 + p_2 + p_3 + \cdots +$. The probability of gray level j is given by

$$P_j = \frac{p_j}{N}, \; P_j \geq 0, \; \sum_0^{j-1} P_j = 1 \qquad (1)$$

With threshold value t, the pixels of an image in the bi-level thresholding method are divided into two classes C_1 *and* C_2 with gray levels for C_1 as [0, 1,…,t] and with gray levels for C_2 as [t+1,….L-1]. For the two classes, the probability distribution of gray level is given as

$$W_1 = \Pr(C_1) = \sum_{j=0}^{1} P_j \qquad (2)$$

$$W_2 = \Pr(C_2) = \sum_{j=t+1}^{L-1} P_j \qquad (3)$$

the class C_1 and C_2 *mean levels are given as*

$$u_1 = \sum_{j=0}^{t} ip_j / W_1 \qquad (4)$$

$$u_2 = \sum_{j=t+1}^{L-1} ip_j / W_2 \qquad (5)$$

u_T Denotes the total mean of gray levels in the image and is given by

$$u_T = w_1 u_1 + w_2 u_2 \quad (6)$$

The variances of the class are given as

$$\sigma_1^2 = \sum_{j=0}^{t} (j - u_1)^2 \quad {}^{P_j}/_{w_1} \quad (7)$$

$$\sigma_2^2 = \sum_{j=t+1}^{L-1} (j - u_2)^2 \quad {}^{P_j}/_{w_2} \quad (8)$$

The within – class variance is given as

$$\sigma_w^2 = \sum_{k=1}^{M} w_k \sigma_k^2 \quad (9)$$

The variance between class is

$$\sigma_B^2 = w_1 (u_1 - u_T)^2 + w_1 (u_2 - u_T)^2 \quad (10)$$

The complete variance of gray levels is given as

$$\sigma_T^2 = \sigma_w^2 + \sigma_B^2 \quad (11)$$

Since the variance among within-class and between-class is constant for diverse partitions, Otsu method select the optimal threshold at time t by maximizing the variance between-class variance, and this equals to reduced variance of the within-class.

Chapter -4

Feature Extraction and Classification

4.1 Introduction

For the goal of smart visual surveillance system, detection of moving object is the essential and significant phase that needed classification to classify the moving detected object from the background in the scene. The extracted moving object can be car, motorbikes, humans, animals, birds, airplanes, etc. The work of the classification in surveillance system, is to classify the detected moving object into classes which are predefined like animal, human, or vehicle, etc. This system is mostly used in examining humans and vehicles. When the detected object is specified that it belongs to the particular class, then it can be further used as application like in identification of person, tracking of the object activity.

The classification of object is an example of pattern recognition. Basically there are two approaches based on which classification of the detected moving objects can be done [7].

1. **Classification based on the Shape of object**:

In this classification, characteristics of the moving detected object region like boxes, silhouettes, blobs and points are used in classification. Kuno et al. [17] use parameters of silhouette patterns of human to classify them from other moving object. Lipton et al. [23] used area and image blob dispersedness of the shape as features to classify detected moving object as human and vehicles. They used time-based steadiness constraints to make the classification more precisely.

2. **Classification based on the Motion of Object:**

The body of object like human is non-rigid and articulated. There is periodicity in the movement and so for the classification of moving object from the other objects in the video frame this property can be used Cutler et al. [24] used attentive tracked object and calculated self-similarity over period. The calculated self-similarity is periodic in nature for the periodically object movement. Then to detect and classify the object time -frequency analysis is done.

For the recognition of moving object Features Extraction and Classification play very important role.

4.2 Bag of Words

Feature extraction of the object defines the significant shape information of object. The aim of feature extraction is to obtain the most applicable features from the object and represent that features in a lesser dimension space. When data like images used to input to machine learning algorithm is very large then input data is need to transformed into less features vector.

For features extraction from the image bag of features is used to create a vocabulary of extracted local point features. Bag of features also known as Bag of Words (BOW) was created by G. Csurka et al. [25]. They used following process to create this bag of features and it shown in figure 4.1

- Image patches from a set of labelled images are first detected and then descripted.
- The descripted patches are assigned to a vocabulary which is pre-determined clusters set with help of a vector quantization.
- Then bag of key points is constructed using clustering which acquire the number of patches allocated to each cluster.

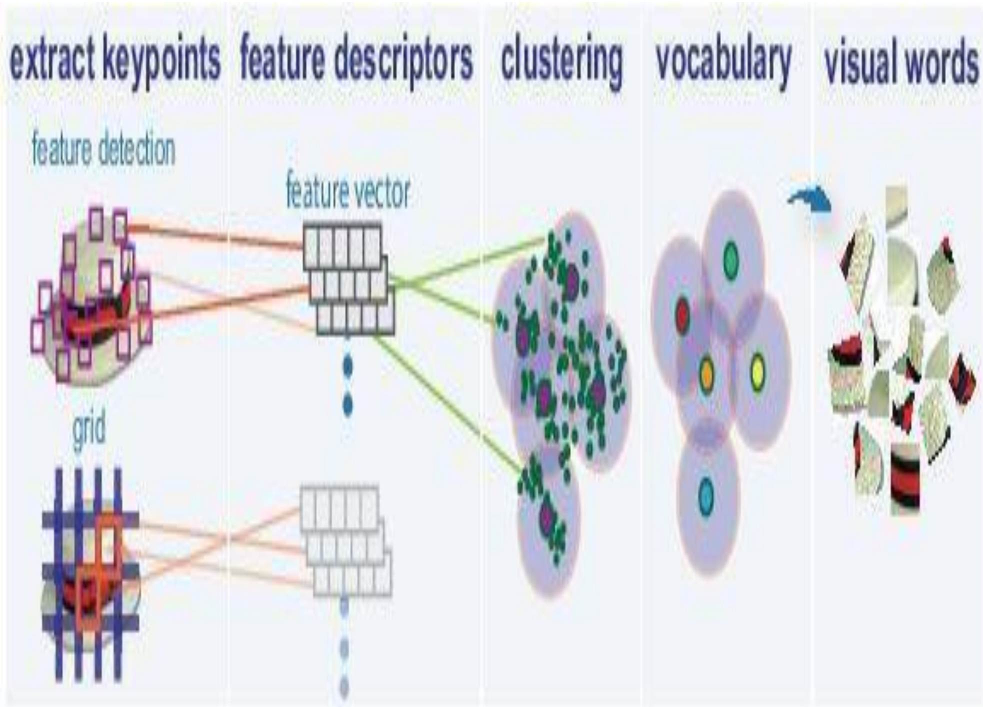


Figure 4-1: Show the process involved in Bag of features

4.2.1 Detection and Description

The patches of labeled images can be detected and descripted using Scale–Invariant Feature Transform (SIFT) [18], Maximally Stable External Region (MSER) [22], Speeded Up Robust Features (SURF) [19], Histogram Oriented Gradient (HOG) [20], etc. Csurka used SIFT for the detect and description of features but now we can use SURF Features for detecting and extracting as the extraction of blob-like features and descriptors from images instead of SIFT. The reason to use this is that size used for descriptor in standard SIFT is of 128 floating point which is reduced in SURF to a value of 64 floating point. A descriptor vector with length of 64 floating point value is created around each key point in nearby using a histogram of gradient orientations. SURF descriptor vector can be created as shown in Figure 4.2

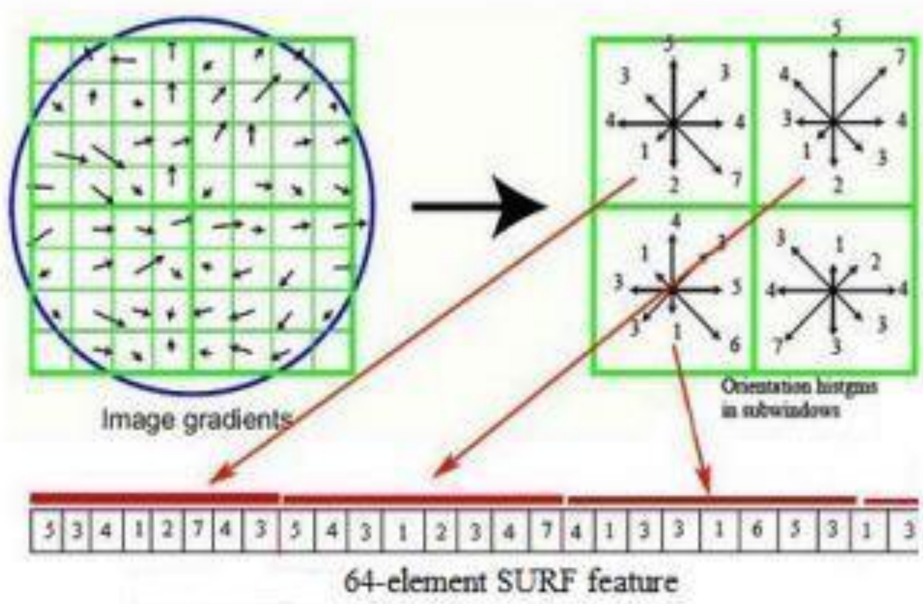
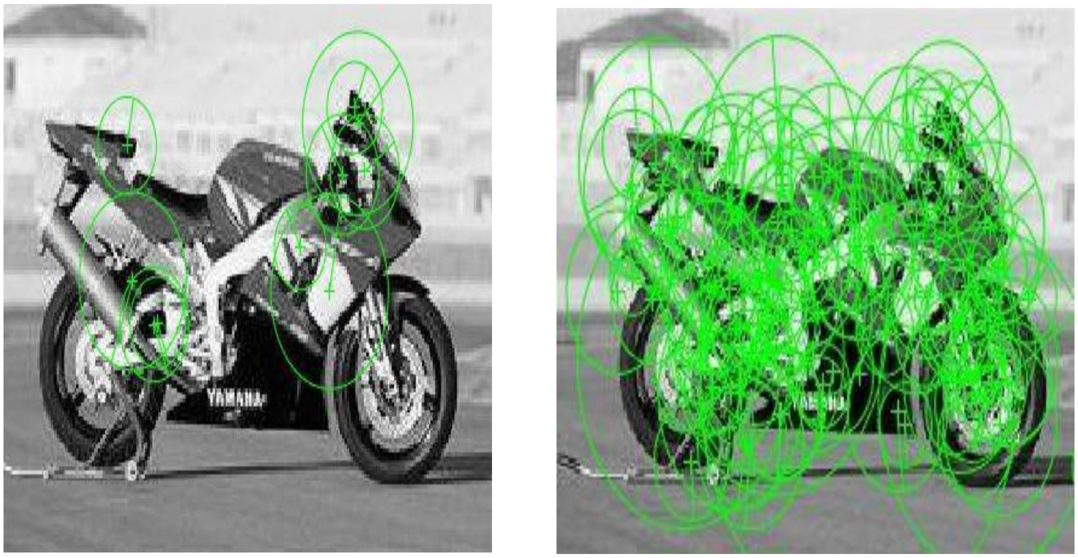


Figure 4-2: Showing SURF feature vector that is built from gradient of images.

The principle of SURF algorithm is same as SIFT algorithm but detail in each step of SURF is different. The algorithm of SURF is consist of detection of interest points, description of local neighboring and matching which is done by equating the descriptors acquired from two different images. Figure 4.3 shows the detected interest point using SURF in the Yamaha motorbikes.



(a) (b)

Figure 4-3: Shows the detected interest point in the image of Yamaha motorbike: (a) selected interest points is 10. (b) Selected interest points is 1000.

4.3 Classification

After doing the features extraction process supervised machine learning classifier are used for the classification of classes. A summary of object classification is shown in figure 4.4 and some of the classifier of machine learning are discussed here.

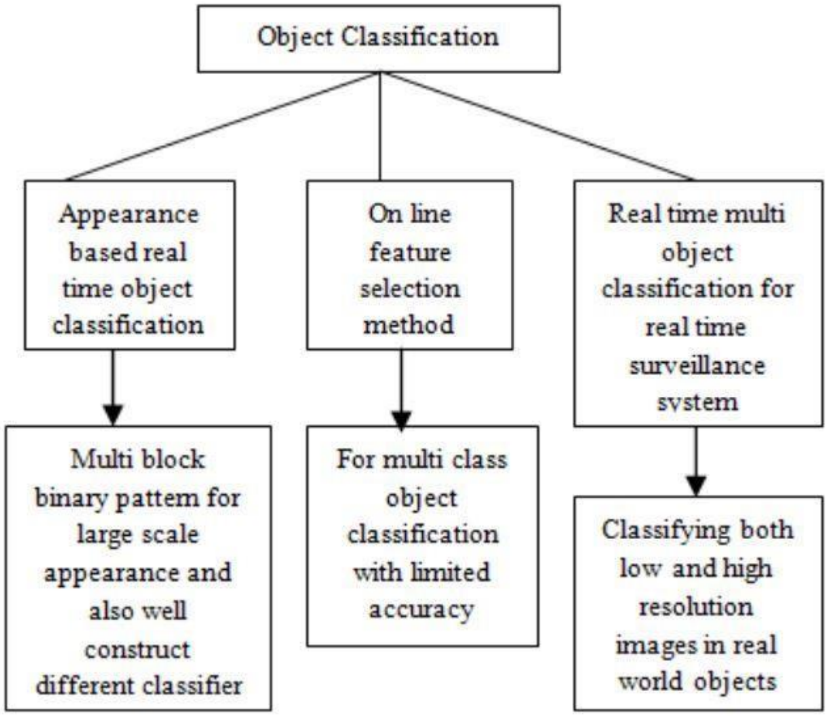


Figure 4-4: Summary of object classification approach [27]

4.3.1 Decision Tree

Decision tree learning developed by Quinlan [30] is one the commonly used classifier in case for spontaneous learning. It uses decision tree as a predictive model to go from observations represented in the branches to conclusion about the item's target value represented in leaves. The models of the decision tree where variable target take discrete target set are called classification trees. In this tree structures, a labelled class are represented by leaf nodes and concurrences of features are resembled by branches which direct to those class labels. Decision trees are known as regression trees when the target object variable take continuous values.

Any data set can be classified following the nodes and branches accordingly until a leaf node is reached. The Decision Tree process target to achieve the classification correctly from all data which are being trained. One benefit of Decision Trees is that the generated classification structures are simple, so they are easy to handle and interpret.

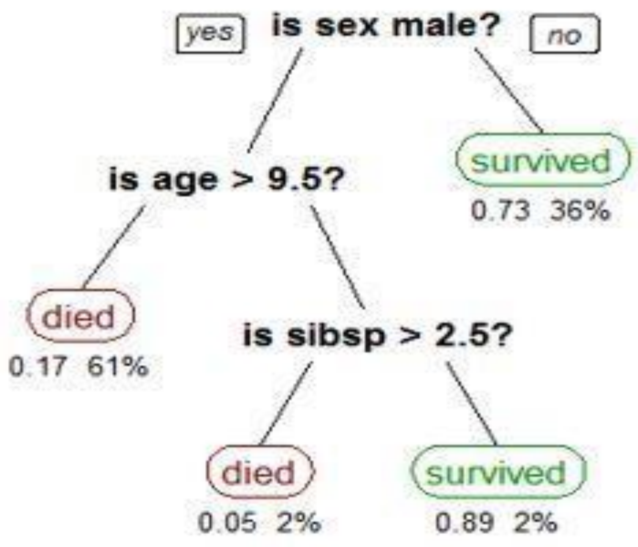


Figure 4-5: an example of Classification using Decision Tree [28]

In figure 4.4 the classification of passengers on Titanic is shown where probability of survival is shown under leaves and observations percentage under leaf. Here, first feature as gender, age, and sibsp (the number of spouses or siblings aboard). If gender feature of person is not male then they are survived by 36% else classification is done on age basis, where if age is more than 9.5 they died else we check sibsp feature. In this case, if sibsp is more than 2.5 than 2% is died else classified as 2% survived.

4.3.2 k-Nearest Neighbour

The k-Nearest Neighbour (kNN) algorithm used for classification is the simplest non-parametric method in the pattern recognition. In the industry, it is widely used in classification problems. All data to be trained in kNN are stored and queries data are classified according to k nearest neighbors of the majority class in the dataset given. kNN uses Euclidean distance for calculating the distance between two data items. The accuracy of classification in kNN can be corrected considerably with the help of algorithms like Large Margin Nearest Neighbor if the space metric is erudite. All training data are re-entered, in spite of this, kNN have costly prediction time.

4.3.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) learning algorithm is one of the trending classifier based on the structure, processing method and learning ability of brain given by Haykin et al.[26]

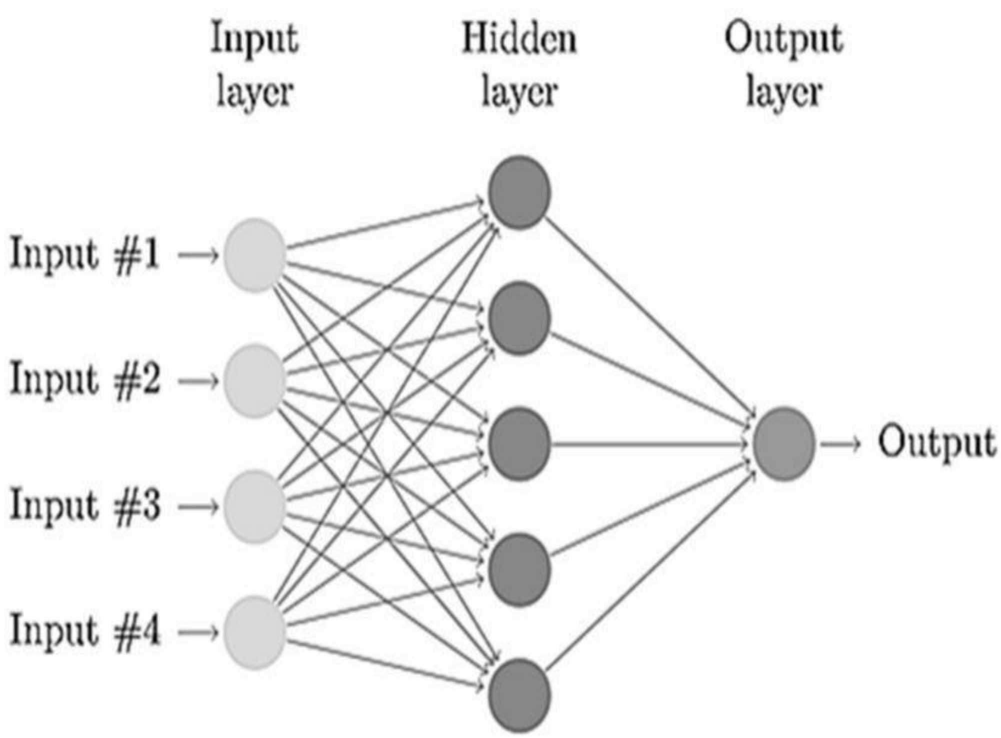


Figure 4-6: shows the structure of neural network [20]

Its structure is inspired by biological neurons networks. Their artificial neurons also termed as nodes, are liable in even more than one layers. Each node of the structure is linked to one or more than one nodes through weighted connections.

It is used in many application as classifier to classify the classes such as in vehicles identification, face recognition, text recognition and moving object classification. Figure 4.4 shows the basic structure model of neural network that composed three layer as input layer, hidden layer, and output layer. Neuron present in the first input layer provide the information about features. The hidden second layer implements calculation required for the third layer to provide the output.

The representation and knowledge about data is acquired and stored in an ANN by adjusting connections' weights. There are several algorithms for ANN training. They usually try to adjust the ANN weights to approximate the outputs of the ANN to the desired outputs known for training data. The algorithm named back-propagation is based on this error correcting concept.

In general, among the advantages of ANNs are their robustness to noisy data and their ability to represent linear and non-linear functions of various forms and complexities. Disadvantages include the need of parameter tuning and the difficulty in interpreting the concepts learned by the ANN, which are codified in the weights.

4.3.4 Support Vector Machines

Support Vector Machines (SVMs) [29] is most popular supervised binary classification algorithm of the machine learning algorithm. This was introduced based on the concepts from the Statistical Learning Theory by Coertes and Vapnik[ ] . From a dataset D composed of n pairs (xi, yi), in which xi ∈ Rm and yi ∈ {−1, +1}, SVMs search for a hyper plane $w \cdot \Phi(x) + b = 0$ that with minimum error able to isolate the data in D with maximizing the margin of separation between the classes vectors. In this equation, a mapping function is represented by Φ which maps the data in D to a space of higher dimension, so that the different classes become separable linearly.

In the training and predictions process of SVM, the mapping function is in the form of dot products as $\Phi(xi) \cdot \Phi(xj)$, which are efficiently calculated with the help of Kernel functions,. Some kernel function used in Support Vector Machine are the Linear, Radial-Basis Function, and Gaussian functions.

Chapter – 5

Implementation

In the previous two chapter, we discussed the concepts of some image processing and machine learning algorithm for the case of motion detection and classification respectively. In this chapter 5, we are going to implement them for moving object detection and classification.

The mechanism to detect and classify moving object involves three important stage that are moving object detection, feature extraction and classification also followed by pre-processing steps. For this video is taken on which detection part is done and for training the SVM classifier image database from Caltech_101 [21][31] and LabelMe[33] is taken. The entire process is operated on MATLAB R2015a. In first part detection process is done in video which includes three major stages background modeling, background subtraction and moving object segmentation. then features are extracted using bag of features of the detected object in the images. In second part features are extracted using the same bag of features from the image set in database and then trained the classifier using the Classifier learner of Matlab. These steps are revealed below in brief.

5.1 Algorithm and Flow Chart

5.1.1 Detection of Moving Object

Step1: Image pre-processing

- A video is taken and then successive frames are extracted from it
- Read image
- RGB color image is converted to grayscale image

Step2: Detection of moving object using Background Subtraction

- Background Modelling is done. During a video sequence pixels remain unchanged. So a background modeling

  Specify that pixels is a background pixels or foreground pixels in the frame. Than threshold value of previous frame and current frame is calculated using Otsu threshold.
- After the background frame generation the background subtraction is done using frame difference.
- For any change in frame, current background pixel value is updated by the current frame pixel value if it is lower than a average threshold value of current frame and background frame.
- Then moving foreground object is segmented using Otsu thresholding to obtain the binary image.

Step3: Morphology Filtering and holes filling

- The binary image obtain after thresholding is filtered through closing operation of morphological process and then holes are filled in the foreground object.

5.1.2 Flow Chart of the Moving Object Detection

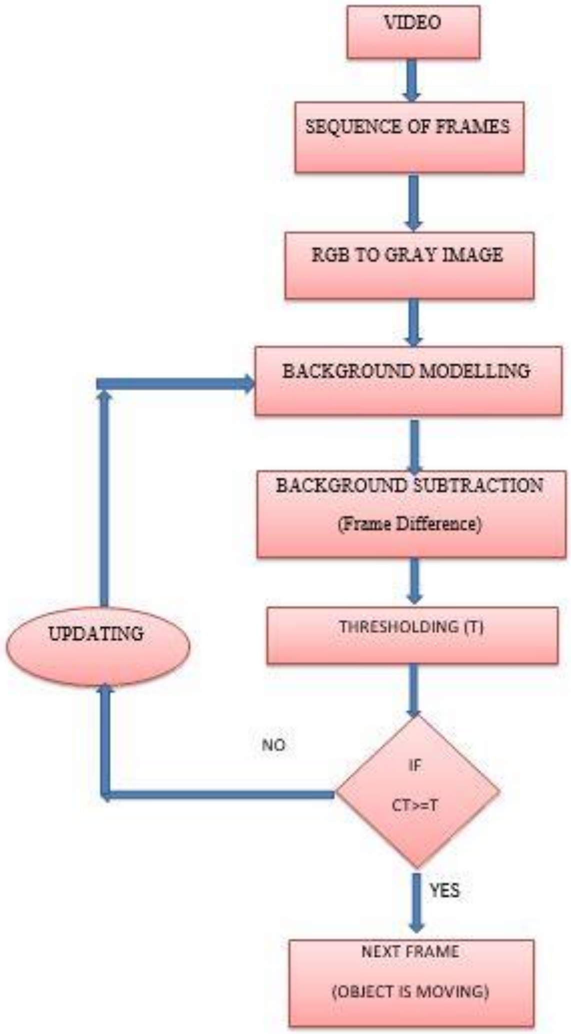


Figure 5-1: Flow chart of Moving Object Detection

5.1.3 Training Image

In this stage images from database are used to train the classifier to make a classification model. For this step used are as below:

Step1: First total 280 images are used to create image database in four groups with equal number of images in each group.

Step2: Features are extracted using Bag of Features from those image set as same as it is extracted from detected images. These features are clustered into 100 vocabulary size.

Step3: Once features are clustered then each cluster are encoded to form visual words and histogram is obtained from that. The encoded features are arranged in a table and then label to it with respect to the classes.

Step4: Now with classification learner, the classification model of SVM classifier is made which gives the confusion matrix as result. This confusion matrix provide the accuracy of probability of occurrence of images.

5.1.4 Flow Chart of the Training process

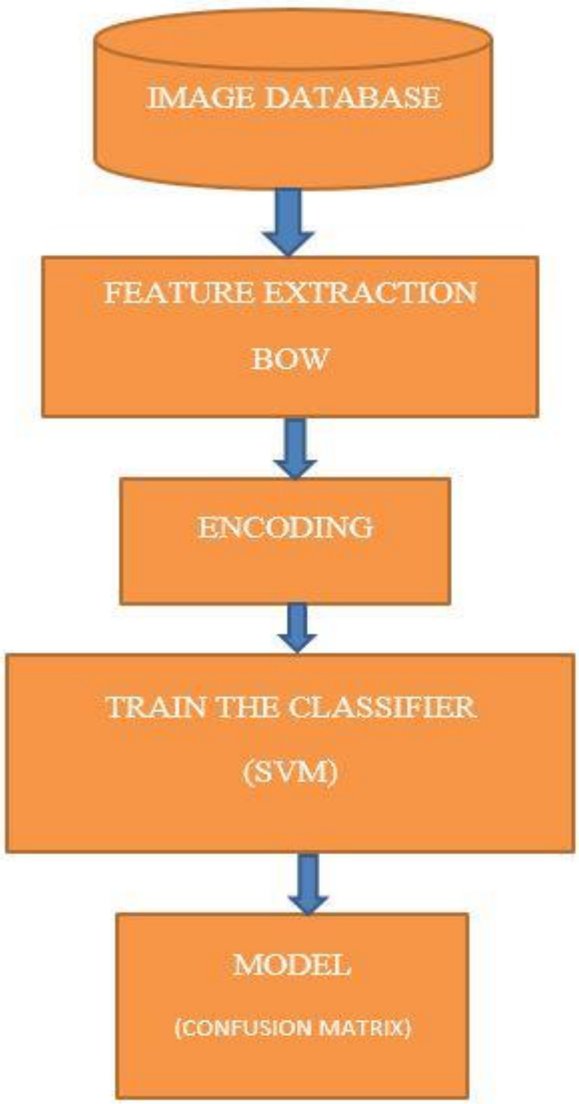


Figure 5-2: Flow chart of the Training process

5.1.5 Testing of the Image

The image obtained after the detection of the moving object in section 5.1.1 is used to test with model of SVM classifier obtained in section 5.1.3. The steps involve in this is discussed below:

Step1: We used the images obtained from detected moving object to test.

Step2: The features are extracted using the same bag of features from the test images. Here also the extracted features are clustered into 100 vocabulary size.

Step3: Once features are clustered then each cluster are encoded to form visual words and histogram is obtained from that. Then we arrange these features into a table.

Step4: Now the features arranged in table and the model created of SVM classifier are used to predict the class of the tested images to which it belongs.

5.1.6 Flow Chart of the Testing process

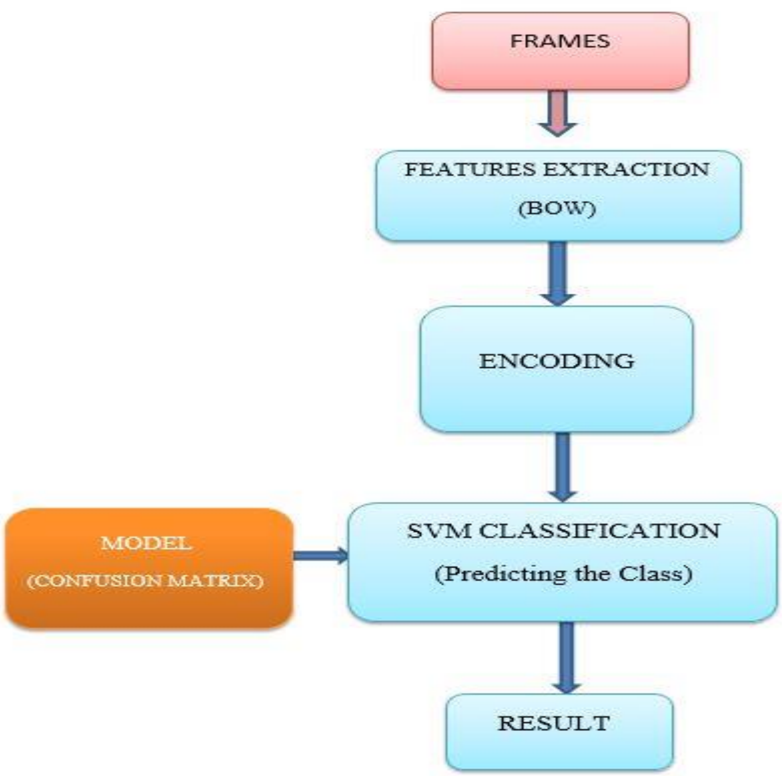


Figure 5-3: Flow chart of the Testing process

Chapter - 6

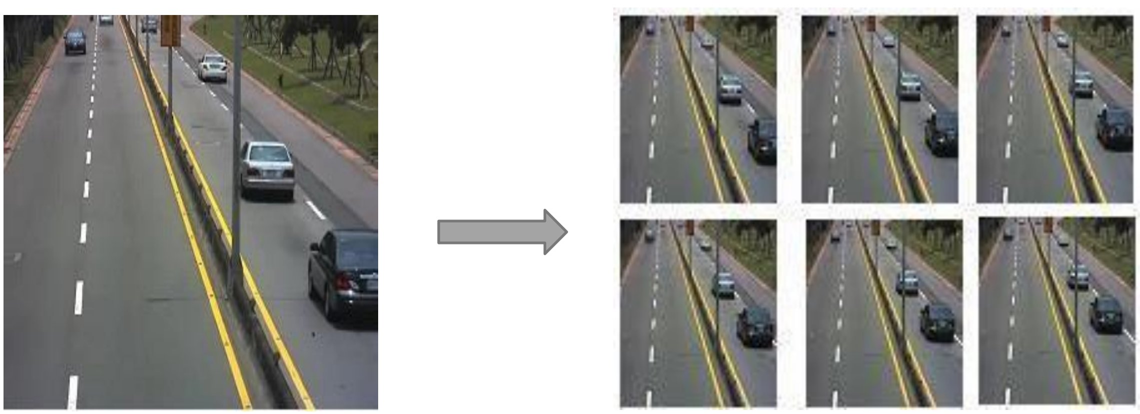
Result and Conclusion

In previous chapter we have discussed the algorithm of moving object detection and classification. Now in this chapter we are going to show the simulation results of each stage of the algorithm implementation .The different stages of algorithm implementation are image pre-processing, feature extraction followed by classification

6.1 Result of Moving Object Detection

6.1.1 Video to Frame Conversion

From a recorded video frames are extracted as shown in the figure 6.1



Video *Sequence of Frames*

Figure 6-1: Shows a video converted into sequence of frame

6.1.2 Frame Differencing

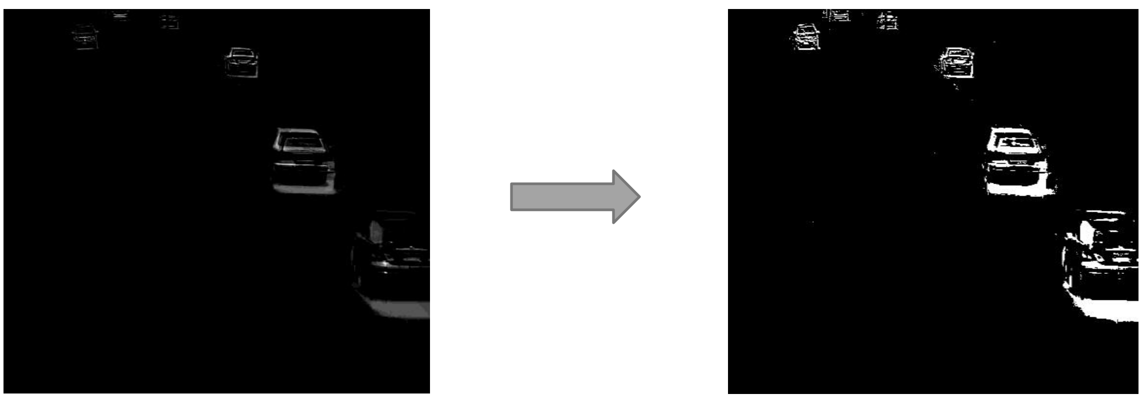
Difference of the image from previous frame (frame-1) and next frame (frame-2) is obtained as shown in figure 6.2

Previous Frame Next Frame Difference Image

Figure 6-2: Shows the Difference of the two image

6.1.3 Image after thresholding and binarization

After doing the background subtraction the moving object in gray difference image is obtained as it is shown in the figure 6.3, then the object is segmented using thresholding and binarized.



Difference Image Binary image

Figure 6-3: Showing binary image after thresholding

6.1.4 Removing noise in binary image

Binary image obtained contains some noise. These are removed by removing all connected components that have fewer than 15 pixels from binary image and then performs morphologically closing on the binary image. If there is any holes left then they are filled. These all are shown in figure 6.4

The object obtained shows that it is moving and so we can say that moving object has been detected.

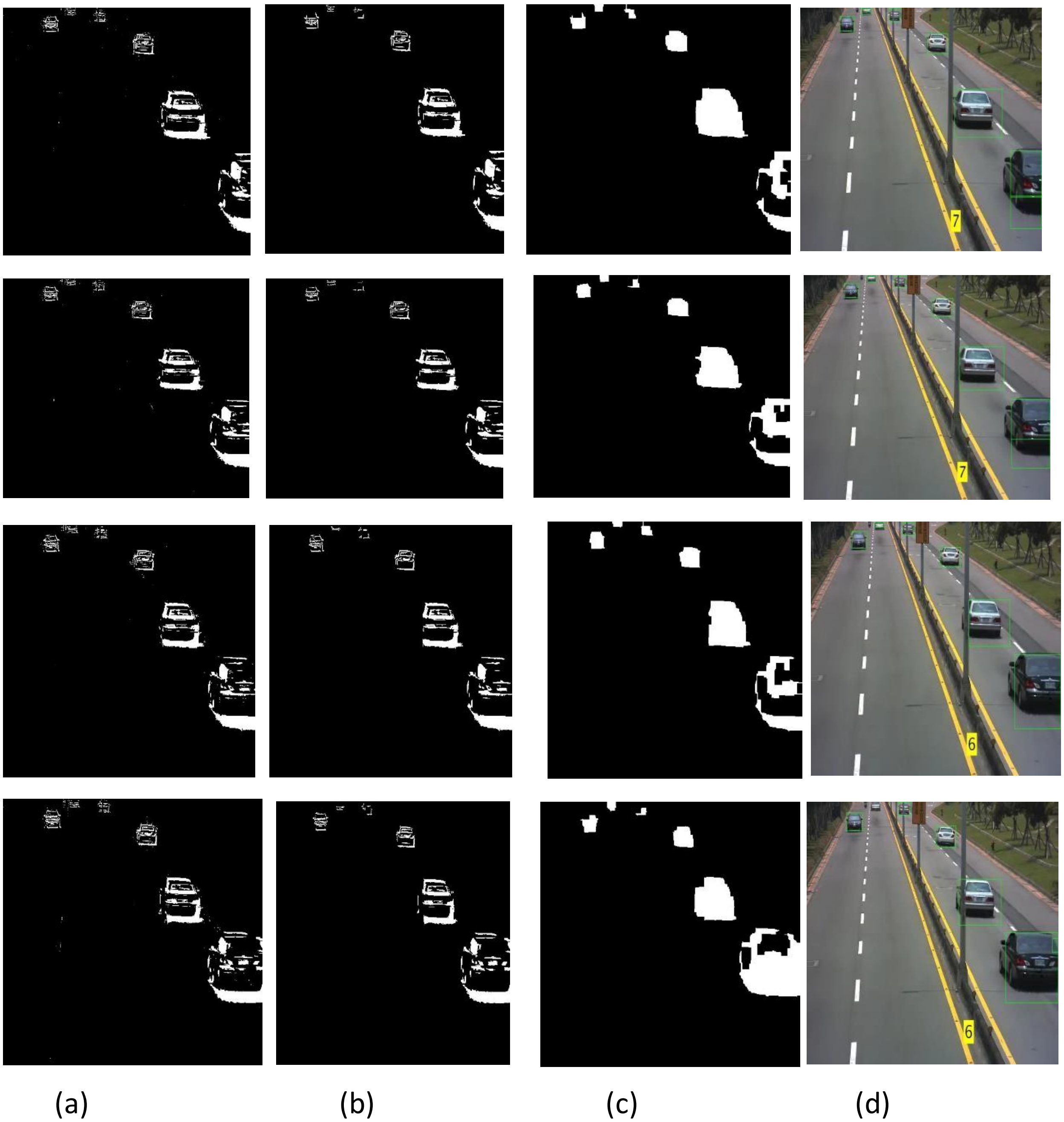


Figure 6-4: Showing binary image after removing noise: (a) Binary image with noise (b)Binary image without noise (c)Morphological closing (d) Showing detected image in original frame.

6.2 Result of Training Classifier

Classes used to train classifier for creating a model are human, car, motorbikes, and bicycle. Images of these taken from Caltech_101 and LabelMe as sample are shown here.

6.2.1 Images taken for the database

6.2.1.1 Human

Images of human used for training as sample is shown in the figure 6.5

Figure 6-5: Images of human

6.2.1.2 Car

Images of car used for training as sample is shown in the figure 6.6

Figure 6-6: Images of car

6.2.1.3 Bicycle

Images of Bicycle used for training as sample is shown in the figure 6.7

Figure 6-7: Images of Bicycle

6.2.1.4 Motorbikes

Images of Motorbikes used for training as sample is shown in the figure 6.8

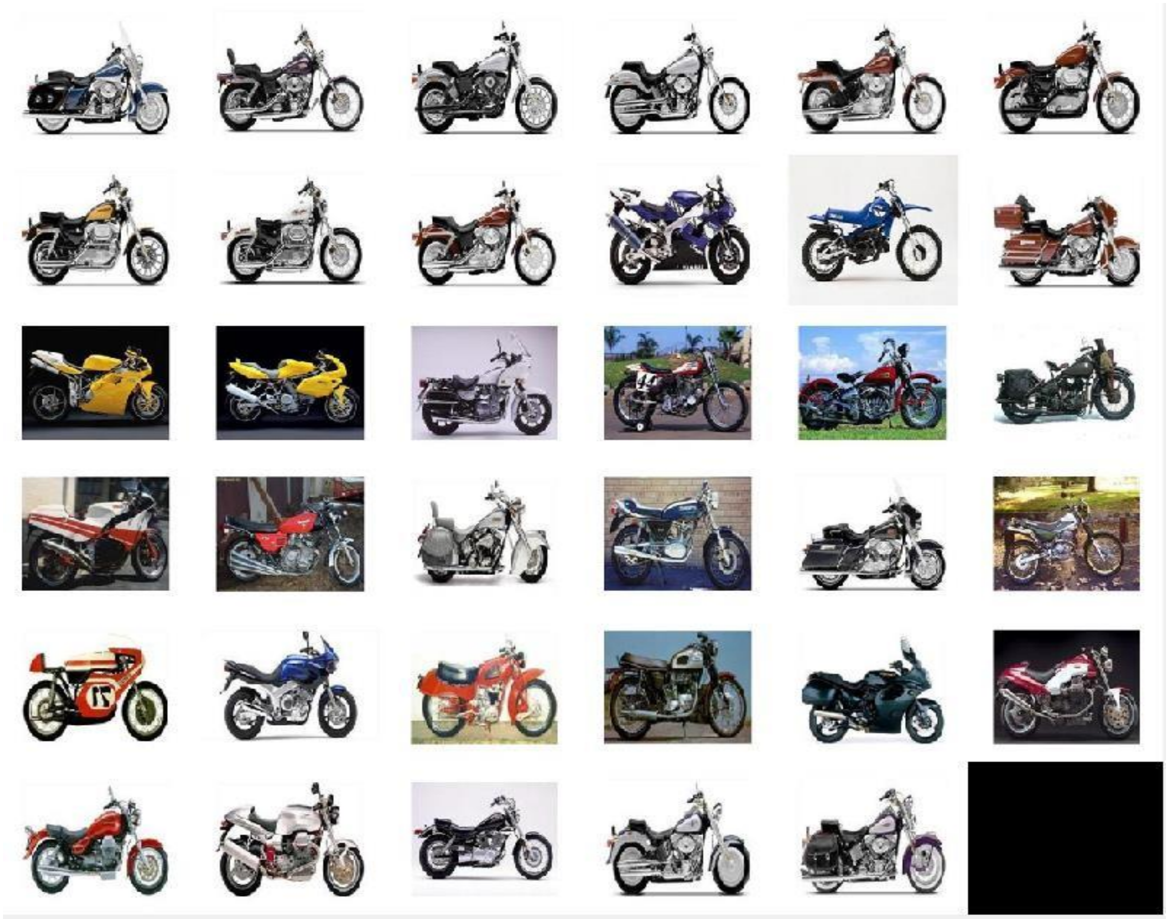


Figure 6-8: Images of Motorbikes

6.2.2 Histogram representation of Visual word obtained from features extraction

6.2.2.1 Histogram representation of bicycle

The histogram representation of visual words of respective bicycle as a sample is shown in figure 6.9.

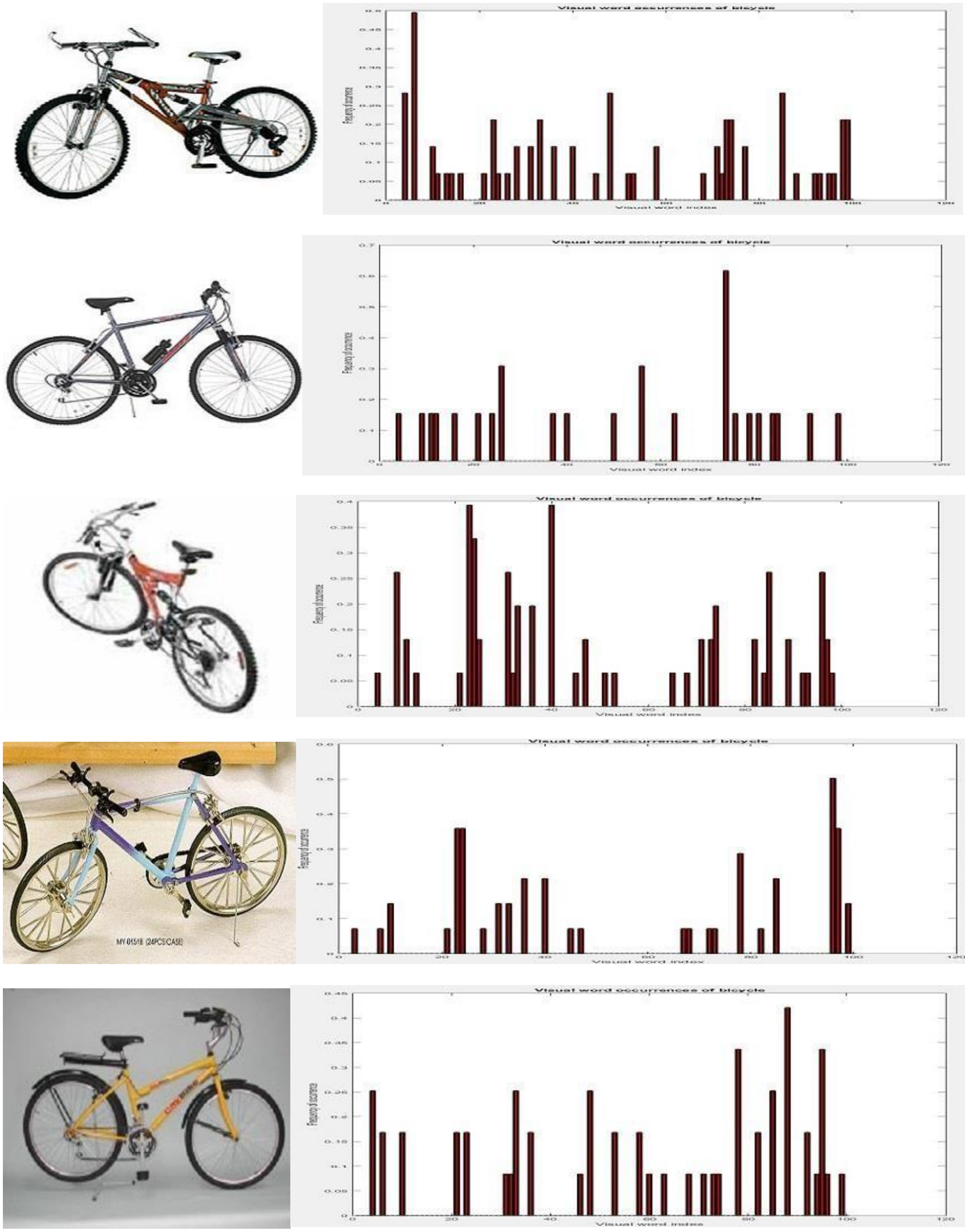


Figure 6-9: Shows image of bicycles in left side with respective histogram in the right side

6.2.2.2 Histogram representation of car

The histogram representation of visual words of respective car as a sample is shown in figure 6.10

Figure 6-10: Shows image of Motorbikes in left side with respective histogram in the right side

6.2.2.3 Histogram representation of human

The histogram representation of visual words of respective human as a sample is shown in figure 6.11

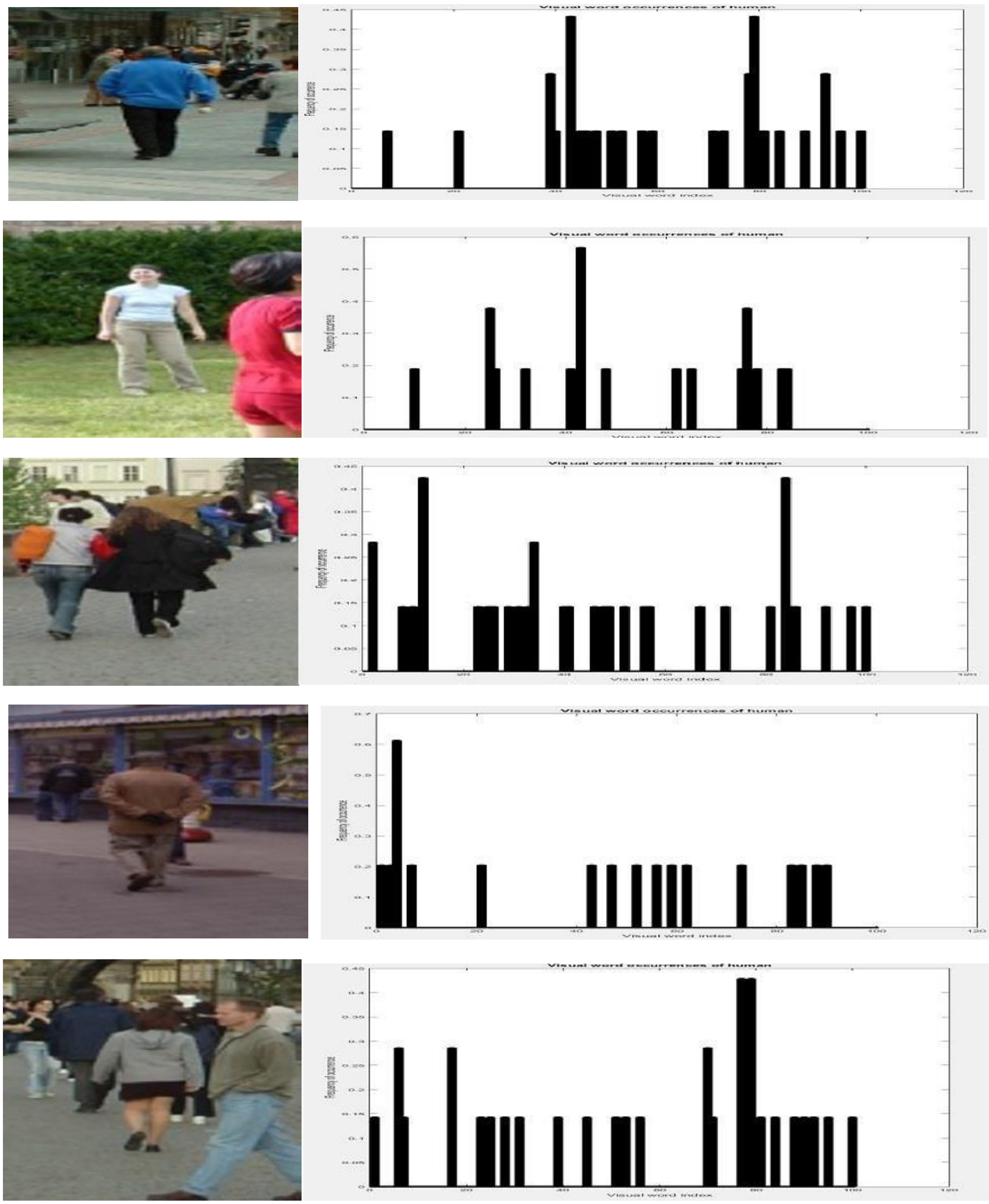


Figure 6-11: Show image of human in left side with respective histogram in right side

6.2.2.4 Histogram representation of motorbikes

The histogram representation of visual words of respective motorbikes as a sample is shown in figure 6.12

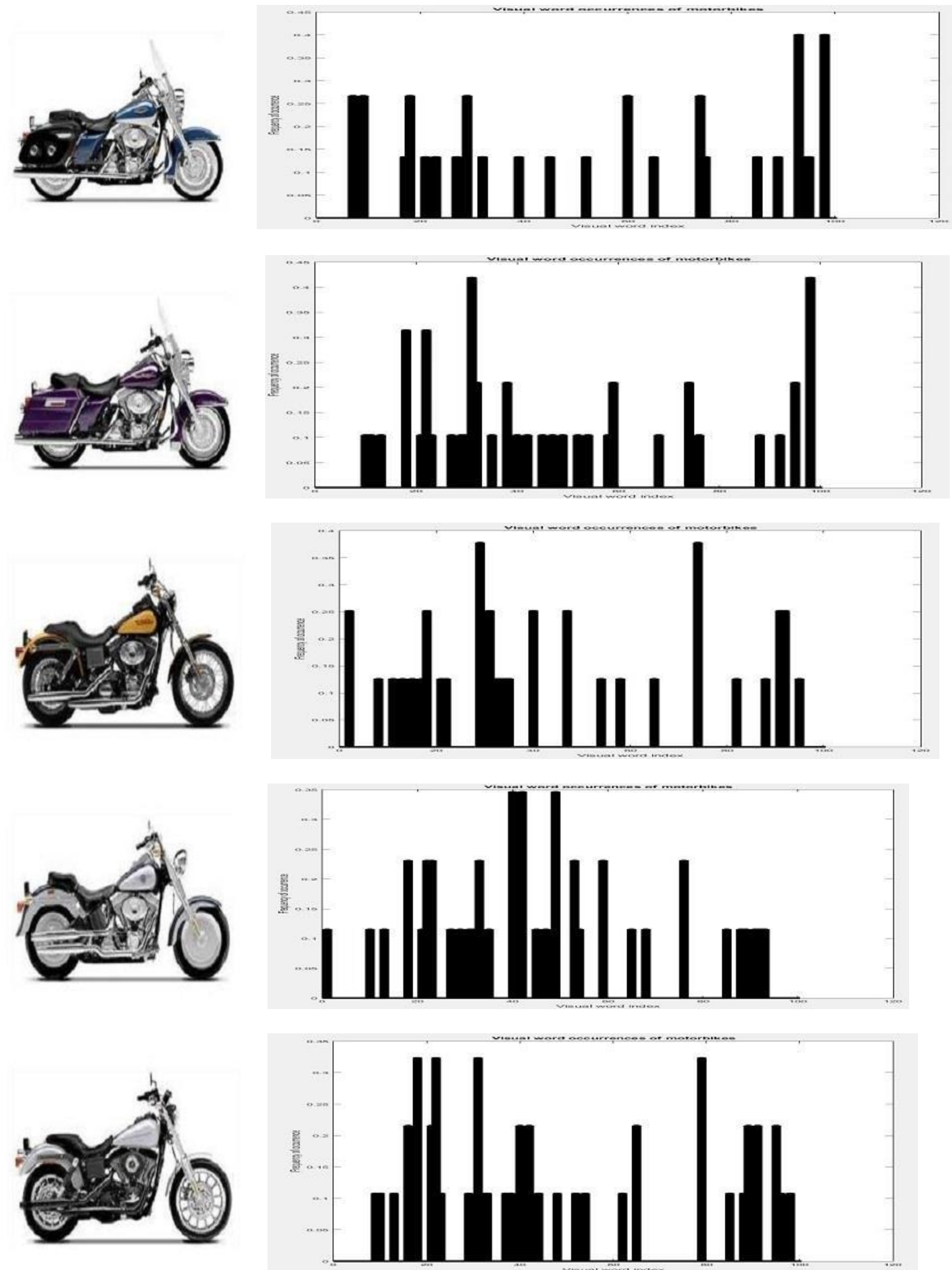


Figure 6-12: Shows image of Motorbikes in left side with respective histogram in the right side.

6.2.2.5 Histogram representation of all the images of classes

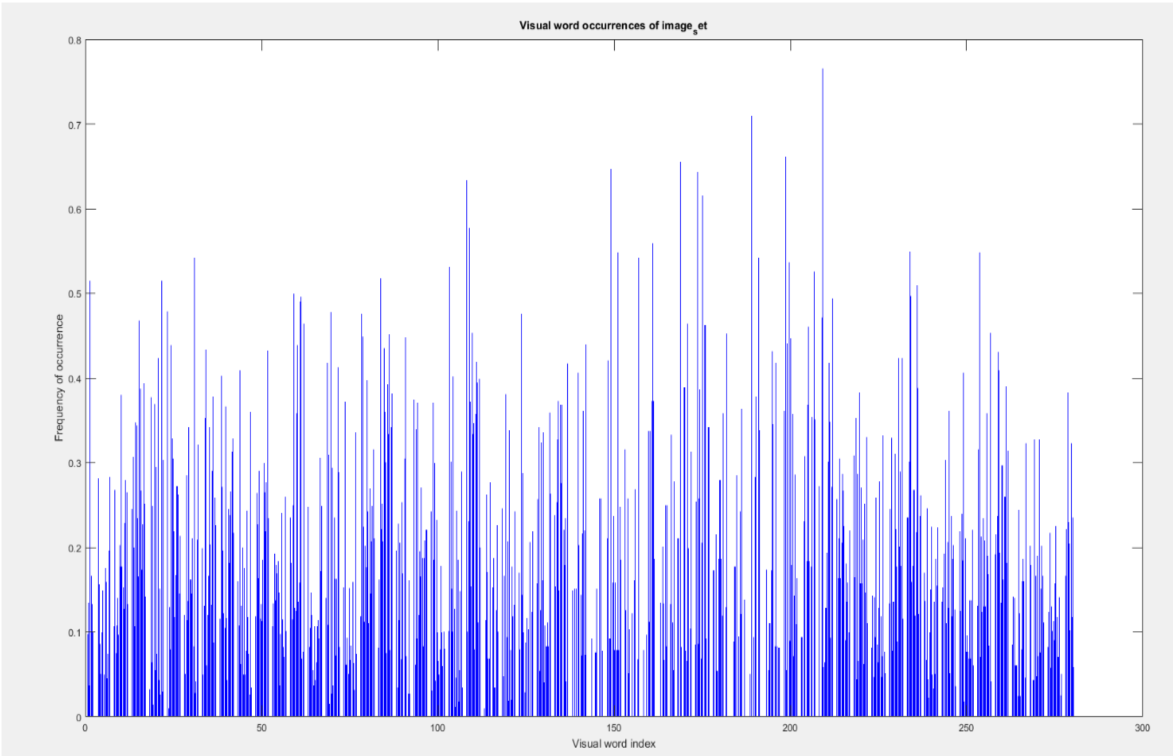


Figure 6-13: Shows the histogram of all class images

The histogram of features obtained from the 280 images of entire training class is shown in figure 6.13

6.2.3 Model of the classifier

The extracted features from the images after training create a model of SVM classifier and this is shown as scatter plotting.

6.2.3.1 Scatter Plot of Train model

The scattering plot of the SVM classifier data and its prediction of the train model is shown in the figure 6.14 and figure 6.15 respectively. Here data of each vector of the class is shown with dot in figure 6.14 and then probability of misclassified with cross in figure 6.15.

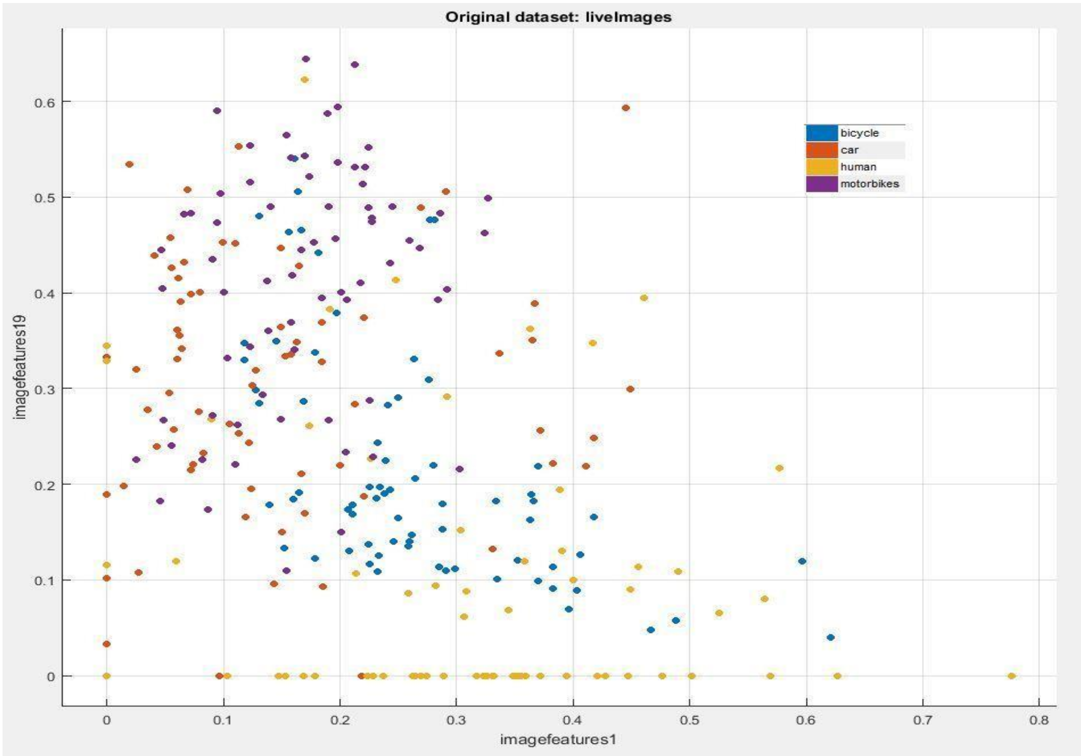


Figure 6-14: Scatter plot of data of 280 images of the classes

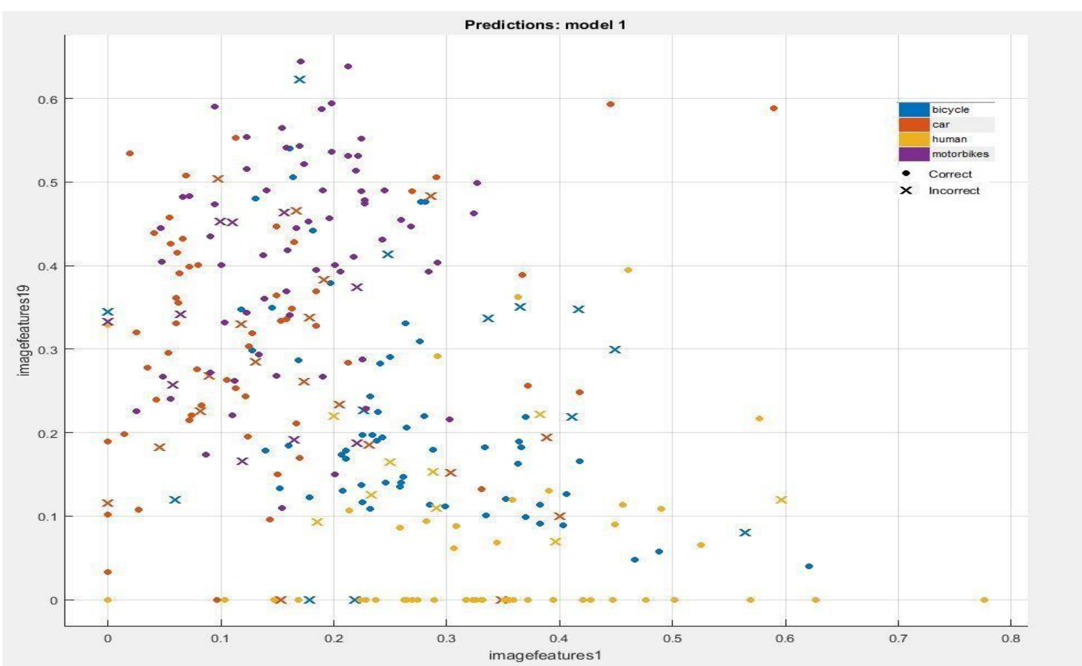


Figure 6-15: Scatter plot of predicted data of 280 images of the classes

6.2.3.2 Confusion matrix obtained of the classes

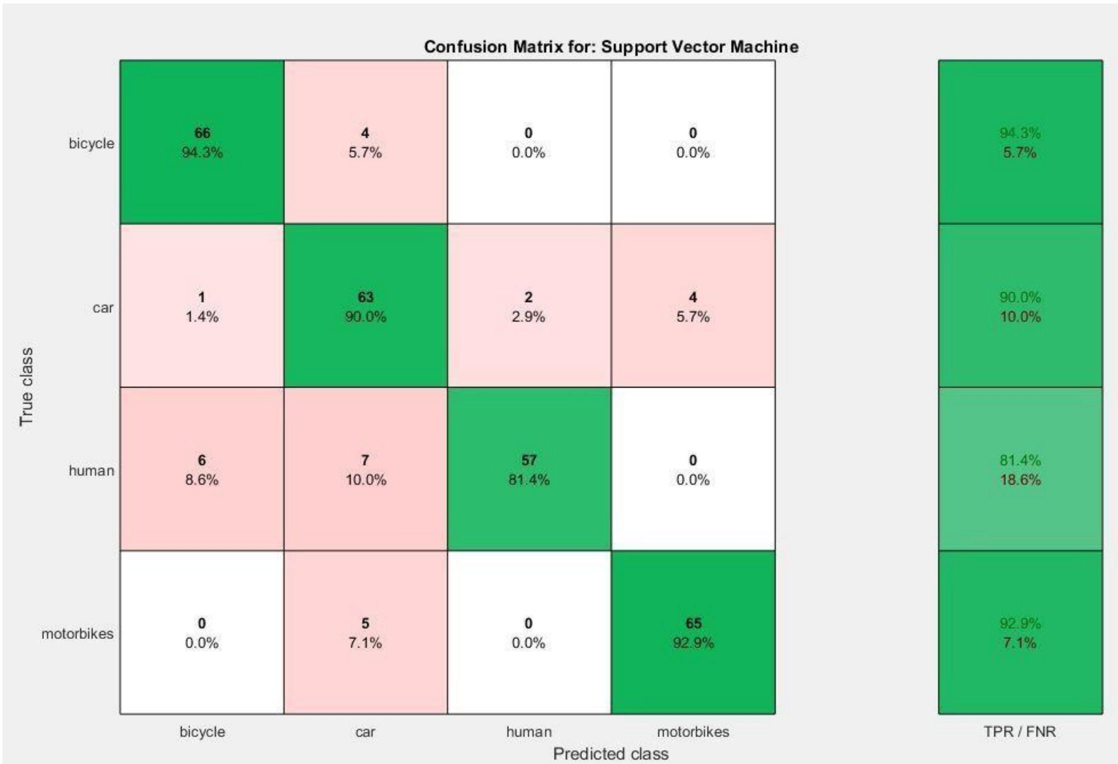


Figure 6-16: confusion matrix of the 280 images of classes

The confusion matrix of the trained classes of total 280 images is shown in the figure 6.16. It shows the probability of occurrence of object with accuracy of 89.6%. From confusion matrix we conclude that classifier SVM can predict bicycle from 70 images as 66 images as bicycle and with error of prediction of image 4 images as other class car. Similarly from 70 images predict 63 images as motorbike, predict 57 images as human, and predict 65 images as car. The box in the diagonal of the matrix give over all correction in a class with percentage and rest boxes in row and column are showing prediction error with percentage. Also the probability of correct prediction of a class is given by column in right.

6.3 Testing of the Image

Now the image obtained of moving object is tested with the trained model to predict the class to which it belongs. Here after testing, we get the result of Car and this is shown in figure 6.17 and also it gives as error of human when test second time and this is shown in figure 6.18.

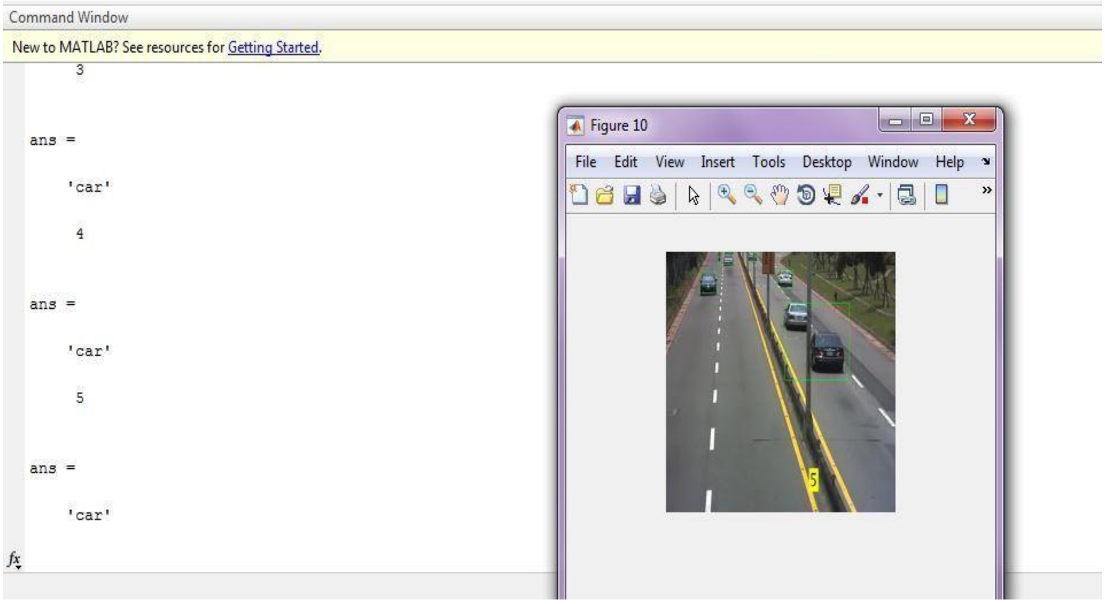


Figure 6-17 Shows tested image of the car as result.

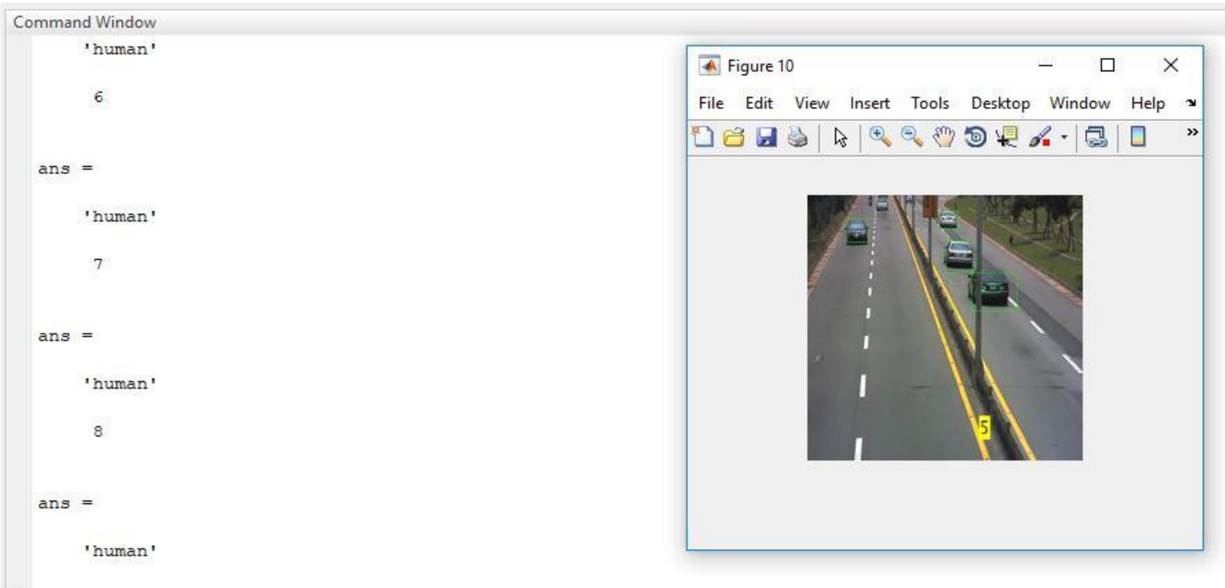


Figure 6-18 Shows tested image of the car as error in result

6.4 Conclusion and Future Work

We have used algorithm to detect the moving object and classify them. The detection of the moving object is done using the background subtraction frame difference method and the classification is done using SVM. Like the most methods in the literature, we have modelled the background frame of the video sequence and then we have computed absolute difference from the next frame to get the moving object. After this we used the supervised machine learning SVM classifier to train the images taken from database Caltech_101 and LabelMe and then classify the moving object. The trained classifier gives the accuracy of 89.6% as the probability of occurrence of the image. Then tested the detected moving object image which result as car.

The work in thesis is given on fixed camera which in future can be implemented with moving camera for the detection as well as classification. The classified object can also be tracked with many tracking algorithm and also their speed and distance can be determined which finally can be utilized in self-auto mobiles in future work.

Bibliography

[1] S. Liu, "An improved ViBe moving object detection algorithm based on spatial-temporal gradient of image," *2016 International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, China, 2016, pp. 192-197. doi: 10.1109/PIC.2016.7949493

[2] H. S. Vu, J. X. Guo, K. H. Chen, S. J. Hsieh and D. S. Chen, "A real-time moving objects detection and classification approach for static cameras," *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, Nantou, 2016, pp. 1-2. doi: 10.1109/ICCE-TW.2016.7521014

[3]R. Fergus, P. Perona and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* 2003, pp. II-264-II-271 vol.2. doi: 10.1109/CVPR.2003.1211479

[4]O. Masoud, S. Gupte, R. F. K. Martin and N. P. Papanikolopoulos, "Detection and classification of vehicles," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 3,no. 1, pp. 37-47, Mar 2002. doi: 10.1109/6979.994794

[5] O. E. L. Harrouss, D. Moujahid, S. E. Elkaitouni and H. Tairi, "Moving objects detection based on thresholding operations for video surveillance systems," *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, Marrakech, 2015, pp. 1-5. doi: 10.1109/AICCSA.2015.7507180

[6] T. Mori, T. Sato, H. Noguchi, M. Shimosaka, R. Fukui and T. Sato, "Moving objects detection and classification based on trajectories of LRF scan data on a grid map," *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, 2010, pp. 2606-2611. doi: 10.1109/IROS.2010.5651516

[7] Weiming Hu, Tieniu Tan, Liang Wang, and S. Maybank. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 34(3):334–352, August 2004.

[8]A. J. Lipton, H. Fujiyoshi and R. S. Patil, "Moving target classification and tracking from real-time video," *Applications of Computer Vision, 1998. WACV '98. Proceedings., Fourth IEEE Workshop on*, Princeton, NJ, 1998, pp. 8-14. doi: 10.1109/ACV.1998.732851

[9] Meyer, Dorthe, Joachim Denzler, and Heinrich Niemann. "Model based extraction of articulated objects in image sequences for gait analysis." *Image Processing, 1997. Proceedings., International Conference on*. Vol. 3. IEEE, 1997.

[10] Haritaoglu, Ismail, David Harwood, and Larry S. Davis. "W/sup 4: real-time surveillance of people and their activities." *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000): 809-830.

[11] McKenna, Stephen J., et al. "Tracking groups of people." *Computer vision and image understanding* 80.1 (2000): 42-56.

[12] Stauffer, Chris, and W. Eric L. Grimson. "Adaptive background mixture models for real-time tracking." *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* Vol. 2. IEEE, 1999.

[13] D. Liu and J. Yu, "Otsu Method and K-means," *2009 Ninth International Conference on Hybrid Intelligent Systems*, Shenyang, 2009, pp. 344-349. doi: 10.1109/HIS.2009.74

[14] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." *IEEE transactions on systems, man, and cybernetics* 9.1 (1979): 62-66.

[15] M. A. Rashidan, Y. M. Mustafah, Z. Z. Abidin, N. A. Zainuddin and N. N. A. Aziz, "Analysis of Artificial Neural Network and Viola-Jones Algorithm Based Moving Object Detection," *2014 International Conference on Computer and Communication Engineering*, Kuala Lumpur, 2014, pp. 251-254. doi: 10.1109/ICCCE.2014.78

[16] Bradski, Gary, and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. "O'Reilly Media, Inc.", 2008.

[17] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, "Automated detection of human for visual surveillance system," in Proc. Int. Conf. Pattern Recognition, 1996, pp. 865–869.

[18] Lowe, David G. "Object recognition from local scale-invariant features." Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, 1999.

[19] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer vision–ECCV 2006 (2006): 404-417.

[20] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005

[21] L. Fei-Fei, R. Fergus and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004 (2016): 76.

[22] Matas, Jiri, et al. "Robust wide-baseline stereo from maximally stable extremal regions." Image and vision computing 22.10 (2004): 761-767.

[23] A. J. Lipton, H. Fujiyoshi and R. S. Patil, "Moving target classification and tracking from real-time video," Applications of Computer Vision, 1998. WACV '98.Proceedings. Fourth IEEE Workshop on, Princeton, NJ, 1998, pp. 8-14. doi: 10.1109/ACV.1998.732851

[24] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 781-796, Aug 2000.doi: 10.1109/34.868681

[25] Csurka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C., 2004, May. Visual categorization with bags of keypoints. In Workshop on statistical learning in computer vision, ECCV (Vol. 1, No. 1-22, pp. 1-2).

[26] Haykin, Simon S., et al. *Neural networks and learning machines*. Vol. 3. Upper Saddle River, NJ, USA: Pearson, 2009.

[27] Mishra, Pawan Kumar, and G. P. Saroha. "A study on classification for static and moving object in video surveillance system." *International Journal of Image, Graphics and Signal Processing* 8.5

[28] https://en.wikipedia.org/wiki/Decision_tree_learning/

[29] Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

[30] Quinlan, J. Ross. "Induction of decision trees." Machine learning 1.1 (1986): 81-106.

[31] http://www.vision.caltech.edu/Image_Datasets/Caltech101/

[32]http://in.mathworks.com/help/matlab/ref/rgb2gray.html/

[33] http://people.csail.mit.edu/brussell/research/LabelMe/intro.html