VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum – 590 018

A project report on

"VIRTUAL TRAFFIC POLICE AUTOMATIC HELMET DETECTION AND NUMBER PLATE RECOGNITION"

submitted in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

INFORMATION SCIENCE & ENGINEERING

by

VINOD V.L (1CR16IS123)
VIJETH B.G (1CR16IS121)

Under the guidance of

Mr.Bineet Kumar Jha
Asst.Professor
Dept. of ISE, CMRIT, Bengaluru

CMR INSTITUTE OF TECHNOLOGY
DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING
#132, AECS Layout, IT Park Road, Bengaluru-560037

2019-20

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum – 590 018

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

Certificate

This is to certify that the project entitled, **"VIRTUAL TRAFFIC POILCE AUTOMATIC HELMET DETECTION AND NUMBER PLATE RECOGNITION"**, is a bonafide work carried out by VINOD V.L **(1CR16IS123), VIJETH B.G (1CR16IS121),** in partial fulfillment of the award of the degree of Bachelor of Engineering in Information Science & Engineering of Visvesvaraya Technological University, Belgaum, during the year 2019-20. It is certified that all corrections/suggestions indicated during reviews have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said Degree.

Name & Signature of Guide	Name & Signature of HOD	Signature of Principal
Mr.Bineet Kumar Jha	Dr.Farida Begam	Dr.Sanjay Jain
Asst.Professor	Assc.Professor	Cmrit
Department of ise,cmrit	Department of ise,cmrit	

EXTERNAL VIVA

NAME OF THE EXAMINERS	SIGNATURE
1_____	_____
2_____	_____

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum – 590 018

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

Certificate

We, **VINOD V.L (1CR16IS123), VIJETH B.G(1CR16IS121),** bonafide students of CMR Institute of Technology, Bangalore, hereby declare that the report entitled **"VIRTUAL TRAFFIC POLICE AUTOMATIC HELMET DETECTION AND NUMBER PLATE RECOGNITION"** has been carried out by us under the guidance of **Mr.Bineer Kumar Jha, Asst.professor,** CMRIT Bangalore, in partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering in **Information Science Engineering**, of the Visvesvaraya Technological University, Belgaum during the academic year 2019-2020. The work done in this dissertation report is original and it has not been submitted for any other degree in any university.

VINOD V.L (1CR16IS123)

VIJETH B.G (1CR16IS121)

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Acknowledgment

The satisfaction and euphoria that accompany a successful completion of any task would be incomplete without the mention of people who made it possible, success is the epitome of hard work and perseverance, but steadfast of all is encouraging guidance.

So with gratitude, I acknowledge all those whose guidance and encouragement served as a beacon of light and crowned our effort with success.

I would like to thank **Dr. Farida Begam**, Associate Professor and HOD, Department of Computer Science who shared her opinion and experience through which I received the required information crucial for the seminar.

I consider it a privilege and honor to express my sincere gratitude to my guide **Mr.Bineet Kumar Jha,** Assistant Professor, Department of Computer Science & Engineering, for her valuable guidance throughout the tenure of this review.

I consider it a privilege and honor to express my sincere gratitude to project coordinator Project Coordinators **Dr.Sudhakar K N**, Associate Professor, Department of Information Science & Engineering for her valuable guidance throughout the tenure of this review.

Finally, I would like to thank all my family members and friends whose encouragement and support was invaluable.

Vinod v.l and Vijeth B.g

(1cr16is123 ,1cr16is121)

CHAPTER 1

INTRODUCTION

1.1 GENERAL INTRODUCTION

The social situation in India is fundamentally extraordinary because of issues, for example, neediness, joblessness just as an extensively lower regard for rules. This makes it unfeasible to go for a totally programmed tollbooth. The business requires a programmed vehicle grouping framework in India not to decrease or wipe out human intercession or work, yet to guarantee that human mediation doesn't bring about any budgetary acts of neglect. The business requires a framework that runs out of sight and simply keeps a cross-beware of the manual.

The conventional OCR based methodology for number plate acknowledgment doesn't work for the varieties in painting style of the number plates. In this paper creators have exhibited a picture recovery based strategy to perceive the vehicle number plate caught utilizing an advanced cell to encourage the Car the executives arrangement of a Smart office premise. In the proposed strategy an advanced mobile phone is utilized to catch the pictures and concentrate highlights of the vehicle number plate. These highlights are coordinated against predefined set of same vehicle number plate pictures in the database. The character pictures are coordinated in a proficient way to make it a continuous arrangement.

As effectively expressed, the framework utilizing fiber optics naturally has countless issues separated from the principle worries of significant expense and support. Albeit an IR blind framework decreases the expense essentially, it is still very costly and less expensive options are desired.As practically every one of the tollbooths utilize cameras for security purposes, it was felt that the possibility of a framework utilizing IP cameras ought to be tried.

As for vehicle wellbeing, India meets just two out of the seven vehicle security guidelines by the World Health Organization (WHO). Bikes represent 25% of all out street crash passing. About 75%motorcycle riders engaged with mishaps kept on wearing head protectors, crash records appear. The primary driver of these fatalities is individuals riding bikes affected by liquor results and infringement of traffic rules which later on brings about genuine mishaps.

Nowadays there are some monetarily accessible applications like Google Goggles that perceive pictures and content on the advanced cell, caught by the PDA's camera. Such applications face the accompanying restrictions and difficulties like: (I) necessity of Internet Connection, (ii) prerequisite to extricate singular characters on inserted portable stage which have imperatives as for both memory and processor, (iii) the nearness of specular reflection makes the acknowledgment task increasingly troublesome, (iv) pictures caught are for the most part in the night, so the pictures could be obscured and have low complexity (v) content Localization from the Background of the picture caught is a test. Some example number plates are appeared in Figure 1. So we have attempted to utilize a picture recovery based methodology for perceiving the number plate. One advanced mobile phone based constant picture recovery framework was proposed in [3]. In that work the creators have exhibited a technique to flawlessly interface physical and advanced world utilizing paper records however a similar strategy can't be conveyed in our concern.

1.2 Existing System

The elite fiber optic detectors are used for moving vehicle detection. A run of the mill establishment consists of an interface gadget with transmitter (LED), collector (photographic indicator) and fiber optic detector. As the vehicle ignores the sensors, the sign levels received from the sensors are adjusted. The yield signals from the fiber optic sensors are nourished into a sign handling and information assessment unit involving calculation, which includes hub check figures, pivot separation, vehicle lengths and vehicle classes depending on schedule, separation recipe, and smaller bowing measurement.

In general, an IR window ornament consists of an infrared transmitter and collector. Such window ornaments give the vehicle as it passes through it an unmistakable picture. Nevertheless, the entire vehicle profile can not be obtained by using just one piece of IR drape due to the different vehicle speed that passes the gate. Understanding the vehicle's speed is therefore important. We measure the speed of the vehicle by using the distance between the draperies and the time. With known vehicle speed and heartbeat recurrence realized we can determine the vehicle's right profile.

1.3 Proposed System

This segment introduces the proposed methodology for continuous recognition of bicycle riders without cap which works in two stages. In the primary stage, we identify a bicycle rider in the video outline. In the subsequent stage, we find the leader of the bicycle rider and recognize the number plate and furthermore distinguish whether the rider is utilizing a head protector or not. So as to decrease bogus forecasts, we merge the outcomes from successive casings for conclusive expectation. The square graph shows the different strides of proposed system, for example, foundation subtraction, include extraction, object order utilizing test outlines.

As helmet is important just in the event of moving bicycle riders, so preparing full casing becomes computational overhead which doesn't increase the value of discovery rate. So as to continue further, we apply foundation subtraction on dim scale outlines, with an aim to recognize moving and static items. Next, we present advances associated with foundation displaying.

1.4 Motivation

Two-wheeler is a prevalent method of transportation in pretty much every nation. Be that as it may, there is a high hazard included as a result of less assurance. To lessen the included hazard, it is exceptionally alluring for bicycle riders to utilize head protector. Watching the helpfulness of head protector, Governments have made it a culpable offense to ride a bicycle without cap and have received manual systems to get the violators. In this model edges are separated from the recordings and changed over to grayscale picture, utilizing layout coordinating strategy (cross connection examination) bicycle riders are identified in the casing.

1.5 Objectives of the work

- Use of versatile foundation displaying for the recognition of moving vehicles on occupied streets which handle the difficulties, for example, enlightenment impacts, climate change, and so forth.
- Instead of utilizing hand-made highlights, we have investigated the capacity of convolutional neural system to improve the order execution.

- The proposed approach is assessed on scanty traffic recordings as utilized inas well as on jam-packed traffic recordings gathered from the CCTV Surveillance Network of the Hyderabad City
- This highlight of the calculation referenced above aided in accomplishing quicker character acknowledgment of the tag.
- This character recognition procedure consists of steps such as preparation of the image, Defragmenting, Re dimensioning and Character constraint to be performed on the picture together in order to finish Model Matching.

1.6 Key Features

Tag recognition is one of the procedures used for the proof recognizable purposes of the vehicle. The only goal of this task is to find the most advanced picture (acquired from the camera) approach to interpret the enrollment information. Usually, this procedure involves three stages. The tag restriction is the initial step, paying little consideration for the size and orientation of the tag. The following advance is character division and the last advance is character identification from the mark. Subsequently, during Template Matching, this venture reveals the essential thought of different calculations require to obtain character identification from the tag.The framework can likewise be utilized in profoundly populated territories and exceptionally confined zones to effortlessly distinguish traffic rule disregarded vehicles and proprietor's name, address and other data can be recovered utilizing this framework. This framework can be computerized and it is utilized to perceive vehicles without approval, vehicles that damaged principles at populated zones like shopping centers, colleges, medical clinics and other vehicle parking areas. This can likewise be utilized on account of vehicle use in psychological oppressor activities, carrying, invalid number plates, taken autos and other criminal operations.

Due to this wearing cap is obligatory according to traffic rules, infringement of which draw in robust fines. Inspite, countless motorcyclists don't comply with the standard. By and by, every single significant city previously conveyed enormous video observation system to keep a vigil on a wide assortment of dangers. In this manner utilizing such previously existing framework will be a cost proficient arrangement, anyway these frameworks include an enormous number of people whose exhibition isn't maintainable for extensive stretches of time. Ongoing examinations have

demonstrated that human reconnaissance demonstrates ineffectual, as the length of observing of recordings expands, the mistakes made by people likewise increments.

CHAPTER 2

LITERATURE SURVEY

Authors: AmirgaliyevBeibut, KairanbayMagzhan, Kenshimov Chingiz

Title: Effective Algorithms and Methods for Automatic Number Plate Recognition

Published in: IEEE 2018

An automatic number plate recognition (ANPR) system is a key aspect in traffic congestion. This will help to minimize the different kind of violations on the road. Advanced systems for tracking and identifying stolen, unauthorized vehicles are based on automated number plate recognition technology. This paper's main objectives are to review other methods and propose our own algorithm. A short review is performed on the various methods of number plate recognition algorithms. Further explanations of the proposed algorithm are illustrated in graphical forms to show how the algorithm works. This paper concluded with tests and evaluation results.

Merits – The car number plate pictures were taken from different sides and in different climate conditions and the accuracy of proposed algorithm is 90%.

Demerits - The distance from camera to the vehicle and the weather conditions decreases the performance of the system. Segmentation part as well as optical character recognition can be improved using other popular algorithms like Artificial Neural Network.

Authors: Yuan Jing, BaharYoussefi, MitraMirhassani, RobertoMuscedere

Title: An Efficient FPGA Implementation of Optical Character Recognition for License Plate Recognition

Published in: 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)

Optical Character Recognition system (OCR) can be used in intelligent transportation systems for license plate detection. However, most times the systems are unable to work with noisy and imperfect images. In this work, a robust FPGAbasedOCR system has been designed and tested with imperfect and noisy license plate images. The OCR system is based on a feedforward neural networks, which uses an efficient and precise neuron. The neuron transfer function is based on an approximation of the Hyperbolic Tangent Activation Function.

The neuron is utilized in a 189 160 36 feed forward neuralnetwork configuration. The network parameters were optimized and then tested with noisy images of license plates numbers. The network was able to maintain a 98:2% accuracy in recognizing the characters despite the image imperfections.

Merits: A robust FPGA based OCR system has been designed and tested with imperfect and noisy license plate images. It was able to maintain a 98:2% accuracy in recognizing the characters despite the image imperfections.

Demerits: It requires a proper system with selection of network size, and the optimized and efficient setup of the neuron activation function.

Authors: FaridBounini, Denis Gingras, Vincent Lapointe, HervePollart Title:
Autonomous Vehicle And Real Time Road Lanes Detection And Tracking
Published in: 2015 IEEE

Advanced Driving Assistant Systems, intelligent and autonomous vehicles are promising solutions to enhance road safety, traffic issues and passengers' comfort. Such applications require advanced computer vision algorithms that demandpowerful computers with high-speed processing capabilities. Keeping intelligent vehicles on the road until its destination, in some cases, remains a great challenge, particularly when driving at high speeds. The first principle task is robust navigation, which is often based on system vision to acquire RGB images of the road for more advanced processing. The second task is the vehicle's dynamic controller according to its position, speed and direction. This paper presents an accurate and efficient road boundaries and painted lines' detection algorithm for intelligent and autonomous vehicle. It combines Hough Transform to initialize the algorithm at each time needed, and Canny edges' detector, leastsquare method and Kalman filter to minimize the adaptive region of interest, redict the future road boundaries' location and lines parameters. The scenarios are simulated on the Pro-SiVIC simulator provided by Civitec, which is a realistic simulator of vehicles' dynamics, road infrastructures, and sensors behaviors, and OPAL-RT product dedicated for real time processing and parallel computing.

Merits - Robust against exogenous perturbations and different constraints, but good enough to control the vehicle with a simple couple of fuzzy logic laws.

Demerits – The fuzzy controller handles the vehicle's steering, which has a limitation for a maximum speed of 70 km/h in sharp turns.

Authors: Mahesh Babu K,M V Raghunadh

Title: Vehicle Number Plate Detection and Recognition using Bounding Box Method

Published in: 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)

The use of vehicles in our life is increasing exponentially day by day and as increasing vehicles are violating the traffic rules, theft of vehicles, entering in restricted areas, high number of accidents lead to increase in the crime rates linearly.

This paper has four major steps as follows: Preprocessing of captured image, Extracting license number plate region, Segmentation and Character Recognition oflicense plate. In pre-processing the desired vehicle imageis taken through the digital camera, brightness of image is adjusted, noise removal using filters and image is converted to gray scale. Exactions of license plate region consist of finding the edges in the image where exact location of licenses plate is located and crop it into rectangular frame. Segmentation plays a vital role in vehicle licenses plate recognition; the legibility of character recognition completely relies on the segmentation done. The approach which we have used is simple but appropriate. First we segmented all characters in the image (LP) using Bounding box method. Finally, recognition of each character is done. The template matching method is used for recognition each character inthe vehicle license plate. Merits – After noise removal, character segmentation and recognition the algorithm gives an accuracy of around 91.11%.

Demerits - Blur Images, Broken Number Plate, Similarities between some characters such as O and D; 5 and S; 8 and B, E; O and 0 etc.

Authors: Worawut Yimyam, Mahasak Ketcham

Title: The Automated Parking Fee Calculation Using License Plate Recognition System

Published in: 2017 2nd International Conference on Telecommunication and Networks (TEL-NET 2017)

Design automated parking fee calculation with license plate recognition in order to reduce manual license plate identification that is mostly employed at parking area of other leading malls where the vehicle circulation is high.

Merits - It decrease manual operation including license plate identification and fee calculation, cut off the staff wage, and save time for identifying plate and

figuring out parking fee as well.

Demerits - It requires the license plate images are clear, sharp, no reflection and the background should be white or light color and no pattern.

Image Recognition for Automatic Number Plate

Authors: P.Meghana, S. SagarImambi, P. Sivateja, K. Sairam

Published in: International Journal of Innovative Technology and Exploring Engineering (IJITEE)

ISSN: 2278-3075, Volume-8 Issue-4, February 2019

Automatic number plate recognition is a well known proposal in todays world due to the rapid growth of cars, bikes and other vehicles. This automatic number plate recognition system uses image processing technology for identification of the vehicles. This system can be used in highly populated areas and higly restricted areas to easily identify traffic rule violated vehicles and owners name, address and other information can be retrieved using this system. This system can be automated and it is used to recognize vehicles without authorization ,vehicles that violated rules at populated areas like malls, universities, hospitals and other car parking lots. This can also be used in the case of car usage in terrorist activites, smuggling, invalid number plates, stolen cars and other illegal activities. It can also be used in highway electronic toll collection. Image of the car number plate is captured and detection is done by image processing ,character segmentation which locate the alpha numeric characters on a number plate.Then the segmented characters are translated into text entries using optical character recognition(ocr).ANPR systems are already available but efficiency is not gained thoroughly. These systems are developed using different methodologies butsome factors like vehicle speed, different font styles,font sizes, language of vehicle number and light conditions are required to be explored .These can affect a lot in the overall recognition rate. ANPR systems use (ocr) optical character recognition to scan the vehical number plates, and it can be retrieved whenever required. The other details of the owners of the vehicles like address and mobile number can be manipulated whenever necessary by contacting the system administrative. The purpose of this paper is to recognize a car number plate using ann, image segmentation. We intended to develop a system in mat lab which can perform detection as well as recognition of a car number plate.

Merits - Efficient and first computing technique for identifying vehicle number plate. Computationally very inexpensive as compared with most of the conventional methods.

Demerits - Limited area coverage.

Authors: MuayadAliHamood Bakhtan, Dr. Munaisyah Abdullah, Dr. AedahAb Rahman

Title: A Review on License Plate Recognition System Algorithm

Published in: IEEE 2016

Image enhancement (preprocessing) and number

plate extraction.

This is a review paper hence no merits and demerits are listed.

Authors: Y. Y. Nguwi, W. J. Lim

Title: Number Plate Recognition in Noisy Image

Published in: IEEE 2016

The system is able to tolerate noise level up to 20% with recognition rate of 85%.

Merits - This system is able to tolerate noise level up to 20% with recognition rate of 85%. It utilizes a combination of filters and morphological transformation for segmenting the number plate.

Demerits - It caters only to same font and size number plates

Authors: . Riazul Islam, Kazi Fatima Sharif, Satyen Biswas

Title: Automatic Vehicle Number Plate Recognition Using Structured Elements

Published in: IEEE 2015

An efficient and first computing technique for identifying vehicle number plate. Limited amount of computations are employed.

Merits -. Efficient and first computing technique for identifying vehicle number plate.

Computationally very inexpensive as compared with most of the conventional methods Demerits - Limited area coverage.

Authors: I. Sina, A. Wibisono, A. Nurhadiyatna, B. Hardjono, W. Jatmiko Title:

Vehicle Counting and Speed Measurement Using Headlight Detection.

Published in: 2018 IEEE

We can use a few methods to detect and estimate vehicle speed at night by using CCTV Camera. Normalized cross-correlation has given us better detection accuracy.

Merits - Better detection accuracy than the area-centroid-difference method. Pin-hole model has given us a better accuracy.

Demerits - The miscalculation of the vehicle counting happens because of the high density of the read.

Authors: Wei Wang

Title: License Plate Recognition System Based on the Hardware Acceleration Technology on the ZYNQ

Published in: IEEE 2017

For the disadvantage of high cost and poor practicability of traditional license plate recognition technology based on PC, apply this technology on the ZYNQ to implement the hardware acceleration of the license plate recognition algorithm. The platform consists of programmable logic (PL) and a processing system (PS). The hardware acceleration of the algorithm of license plate location is completed through the part PL; On the PS runs the Linux system. In terms of license plate recognition algorithm, we use the edge detection algorithm, vertical projection method and template matching method respectively for the location of license plate image, character segmentation and recognition.

and an OCR scanner should be able to threshold images. In other words, it should replace each pixel in an image with a black or a white pixel. It is a method of image segmentation.

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Overview on Machine Learning

Machine learning is an application of artifificial intelligence (AI) that gives systems the ability to automatically learn and evolve from experience without being specially programmed by the programmer.The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The main aim of machine learning is to allow computers to learn automatically and adjust their actions to improve the accuracy and usefulness of the program, without any human intervention or assistance.

Traditional writing of programs for a computer can be defined as automating the procedures to be performed on input data in order to create output artifacts. Almost always, they are linear, procedural and logical. A traditional program is written in a programming language to some specification, and it has properties like:

• We know or can control the inputs to the program.

• We can specify how the program will achieve its goal.

• We can map out what decisions the program will make and under what conditions it makes them.

• Since we know the inputs as well as the expected outputs, we can be confident that the program will achieve its goal.

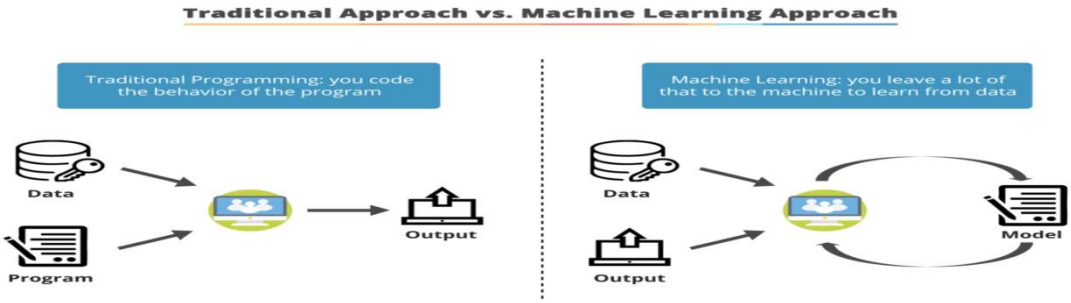


Figure 3.1: Traditional Programming vs Machine Learning

Traditional programming works on the premise that, as long as we can define what a program needs to do, we are confident we can define how a program can achieve that goal.

This is not always the case as sometimes, however, there are problems that you can represent in a

computer that you cannot write a traditional program to solve. Such problems resist a procedural and logical solution. They have properties such as: • The scope of all possible inputs is not known beforehand.

• You cannot specify how to achieve the goal of the program, only what that goal is.

• You cannot map out all the decisions the program will need to make to achieve its goal.

• You can collect only sample input data but not all possible input data for the program.

Problems like this resist traditional programmed solutions because manually specifying a solution would require a disproportionate amount of resources. Furthermore, when new inputs arise, the rules may change, thereby necessitating changes to the program. In such cases as these, machine learning might be the optimum approach to use in deriving a solution to the way the problem is represented on the computer.

Machine learning techniques can be broadly categorized into the following types: Supervised learning takes a set of feature/label pairs, called the training set. From this training set the system creates a generalized model of the relationship between the set of descriptive features and the target features in the form of a program that contains a set of rules. The objective is to use the output program produced to predict the label for a previously unseen, unlabeled input set of features, i.e. to predict the outcome for some new data. Data with known labels, which have not been included in the training set, are classified by the generated model and the results are compared to the known labels. This dataset is called the test set.

The accuracy of the predictive model can then be calculated as the proportion of the correct predictions the model labeled out of the total number of instances in the test set. Unsupervised learning takes a dataset of descriptive features without labels as a training set. In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data.

The goal now is to create a model that finds some hidden structure in the dataset, such as natural clusters or associations. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system does not figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data. Unsupervised learning can be used for clustering, which is used to discover any inherent grouping that are already present in the data. It can also be used for

association problems, by creating rules based on the data and finding relationships or associations between them. Semi-supervised machine learning falls somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy.

Usually, semisupervised learning is chosen when the acquired labeled data requires skilled and

relevant resources in order to train it / learn from it. Otherwise, acquiring labeled data generally does not require additional resources. Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance.

Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal. Machine learning algorithms are tools to automatically make decisions from data in order to achieve some over-arching goal or requirement. The promise of machine learning is that it can solve complex problems automatically, faster and more accurately than a manually specified solution, and at a larger scale. Over the past few decades, many machine learning algorithms have been developed by researchers, and new ones continue to emerge and old ones modified. In this project, we have focused on only supervised learning methods since our dataset contains labels. Since the problem requires us to classify the synthetic dataset and creating them, we will be using exclusively on classification algorithms to classify synthetic based on whether they are Hazy or Non-Hazy

3.2 Machine Learning Tools

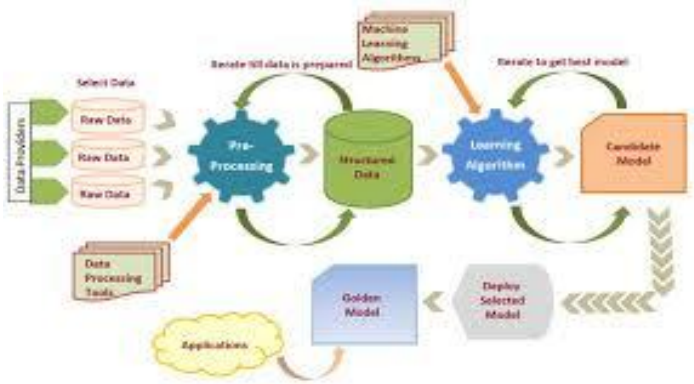


Figure 3.2 machine learning steps

There are many different software tools available to build machine learning models and to apply these models to new, unseen data.

There are also a large number of well defined machine learning algorithms available. These tools typically contain libraries implementing some of the most popular machine learning algorithms. They can be categorized as follows :

• Pre-built application-based solutions.

• Programming languages which have specialized libraries for machine learning. Using programming languages to develop and implement models is more flexible and gave us better control of the parameters to the algorithms. It also allows us to have a better understanding of the output models
produced. Some of the popular programming languages used in the field of machine learning are:

• **Python**: Python is an extremely popular choice in the field of machine learning and AI development. Its short and simple syntax make it extremely easy to learn and use. It supports both object oriented and functional oriented styles of programming as well. In addition to this, Python has a number of libraries which have been developed for machine learning, such as Numpy, Pandas and SciKit-learn.

• **MATLAB**: MATLAB is a programming language developed by MathWorks. Created primarily for numerical computing, MATLAB is also extremely popular among machine learning programmers. It is heavily used in statistical analysis and complex systems. MATLAB excels at handling matrices, making it especially useful in image recognition. It is an extremely

versatile language which allows matrix manipulations, implementation of algorithms and creation of user interfaces. It is also able to interface with other programming languages like C, Fortran and Java.

3.3 **SciKit-learn:** SciKit learn is an open source machine learning library built for python. Since its release in 2007, Scikit-learn has become one of the most popular open source machine learning libraries. Scikit-learn (also called sklearn) provides algorithms for many machine learning tasks including classification, regression, dimensionality reduction and clustering. It also provides utilities for extracting features, processing data and evaluating models.

It provides in-built code for many of the popular machine learning algorithms. The documentation for scikit-learn is comprehensive, popular and well maintained. Sklearn is built on mature Python Libraries such as NumPy, SciPy, and matplotlib. It has a very active development community with regular update releases of the library. While languages such as R and MATLAB are extremely popular and useful for machine learning, we decided to choose Python along with its SciKit-learn libraries as our programming language of choice. The reasons for this are:

• We already have some familiarity and exposure to Python, and thus have a smaller learning curve.

• Both Python and Scikit-learn have excellent documentation and tutorials available online.

• The number of classic machine learning algorithms that come with Scikit-learn, and the consistent patterns for using the different models i.e., each model can be used with the same basic commands for setting up the data, training the model and using the model for prediction. This makes it easier to try a range of machine learning algorithms on the same data.

• The machine learning algorithms included with ski-learn have modifiable parameters known as hyper-parameters that effect the performance of the model. These usually have sensible default

values, so that we can run them without needing a detailed knowledge or understanding of their semantics.

• The IPython notebook, which is an interactive computational environment for Python, in which a user can combine code execution, rich text, mathematics and plots in a web page

This functionality allows us to provide the notebooks we used to run our experiments almost as an audit and in a presentable and easily understood way that allows for reproducible research.

3.4 Dataset

FRIDA and FRIDA2 are databases of numerical images easily usable to evaluate in a systematic way the performance of visibility and contrast restoration algorithms. FRIDA comprises 90 synthetic images of 18 urban road scenes. FRIDA2 comprises 330 synthetic images of 66 diverse road scenes. The view point is closed to the one of the vehicle's driver. To each image without fog is associated 4 foggy images and a depthmap. Different kind of fog are added on each of the 4 associated images: uniform fog, heterogeneous fog, cloudy fog, and cloudy heterogeneous fog. These scenes can be used to test visibility and contrast restoration algorithms intensively and in an objective way, as well as "shape from fog" algorithms. The calibration parameters of the camera are given.

3.5 Data Preprocessing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a method used to resolving such issues. Real world data is generally incomplete: lacking attribute values, lacking certain attributes of interest, or containing only aggregate data. It can also contain noise such as errors or outliers. The steps involved in data preprocessing are:

1. Data Cleaning: The data can have many irrelevant and missing parts. To handle this part, data cleaning is done. It involves handling of missing data and noise.

• Missing Data:

a) Remove tuples with missing values: We simply remove those tuples which contain missing values. This is only feasible if we have a large dataset which will not be affffected by losing rows.

b) Imputation: It involves filling the missing values manually. It can be done by taking the mean, median or most probable value using the values in the same column. Creation of Synthetic Dataset for Haze Removal and Haze Classification

• Noisy Data:

a) Binning:This method works on sorted data in order to smooth it. The whole data is divided into segments of equal size and then various methods are performed to complete the task. Each segmented is handled separately. One can replace all data in a segment by its mean or boundary values can be used to complete the task.

b) Regression: Here data can be made smooth by fitting it to a regression function.The regression used may be linear or multiple .

c) Clustering: This approach groups the similar data in a cluster. The outliers may be undetected or it will fall outside the clusters.

2. Data Transformation: In this step we transform the data into a form suitable for the data mining process.

a) Normalization: It is a method used to scale all the data values to a specified range. This helps standardize all the values.

b) Feature Selection: It is the process of reducing the size of the input by selecting only a subset of
features from all the originals. It helps reduce the degree of cardinality.

c) Discretization: It is used to change continuous values into discrete intervals.

d) Concept Hierarchy: Here we create a hierarchy that allows attributes to get converted into an attribute which is higher in the hierarchy.

3. Data Reduction: While working with huge volume of data, analysis became harder in such cases. In order to get rid of this, we uses data reduction technique. It aims to increase the storage efficiency and reduce data storage and analysis costs. The various steps to data reduction are:

a) Data Cube Aggregation: Aggregation operation is applied to data for the construction of the data cube. Raw data is gathered and expressed in a summary form for statistical analysis.

b) Attribute Subset Selection: The highly relevant attributes should be used, rest all can be discarded. For performing attribute selection, one can use level of significance and p- value of the attribute.the attribute having pvalue greater than significance level can be discarded. Creation of Synthetic Dataset for Haze Removal and Haze Classification

c) Dimensionality Reduction: This reduce the size of data by encoding mechanisms.It can be lossy or lossless. If after reconstruction from compressed data, original data can be retrieved, such reduction are called lossless reduction else it is called lossy reduction. The two effective methods of dimensionality reduction are:Wavelet transforms and PCA (Principal Component Analysis).

3.6 **Machine Learning Algorithms**

As mentioned before, we looked towards only supervised learning methods for this project. This is due to the fact that our dataset has labels already given. Since the aim of this project is to Create synthetic dataset and Classify them to hazy or non hazy, we will be looking at classification algorithms which will be used to classify the images as Hazy or Non-Hazy. We used the following algorithms for our project

YOLO Algorithm

YOLO You Only Look Once (YOLO) was developed to create a one step process involving detection and classification. Bounding box and class predictions are made after one evaluation of

the input image.

Confidence here refers to the probability an object exists in each bounding box and is defined as: C = P r (Object) ∗ IOUtruth pred (1) where IOU, intersection over union, represents a fraction between 0 and 1.

OCR Algorithm

a. Image Acquisition

The first step is to acquire images of paper documents with the help of optical scanners. This way, an original image can be captured and stored. Most of the paper documents are black and white and an OCR scanner should be able to threshold images. In other words, it should replace each pixel in an image with a black or a white pixel. It is a method of image segmentation.

CHAPTER 4

SYSTEM REQUIREMENT SPECIFICATION

System configuration is the process of characterizing the technology, parts, modules, interfaces and information for the system in order to fulfill the defined requirements. Configuration of the frameworks could be seen as the use of the frameworks hypothesis for the advancement of the item. Item located research and strategies are becoming the most widely used techniques for the structure of PC frameworks. The design of frameworks is therefore the way to characterize and build frameworks to satisfy the client's specific needs. The UML has become the standard language in object arranged investigation and plan.

4.1 System Analysis

The primary objective of the analysis is to remove traffic data from video, for example, vehicle counter and vehicle's speed estimation. This procedure involvesseveral steps, which are recovering picture arrangement from video, picture twofold change, mass discovery, fog light matching, following, lastly vehicle checking and speed estimation.

4.2 Functional Requirement

- System should catch the number plate of the vehicle
- System should refresh the violation report to framework.
- System ought to naturally identify the number plate of vehicle
- System should naturally sending the update message to the proprietor of the vehicle of the infringement

4.3 Non Functional Requirement

Usability:

- Easy Interface for catch of picture and Updated in the cloud Dependability:
- Capture the low light effectively.
- Violation Filing.

Execution:

- Should not take unnecessary time in handling the caught picture and preparing it.

Supportability:

- Contain straightforward code with arrangements for future improvement.

Tools and Technologies

Hardware Requirements:

- Camera
- Laptop

Software Requirements:

- Open CV
- Python 3
- Noobsos

Python language is being used worldwide as a wide extent of use headway and system improvement programming language. Colossal brands and web searcher are using python programming to make their task more straightforward. It is flexible, strong and far reaching.

Chapter 5

SYSTEM ANALYSIS

Analysis is the process of finding the best solution to the problem. System analysis is the process by which we learn about the existing problems, define objects and requirements and evaluates the solutions. It is the way of thinking about the organization and the problem it involves, a set of technologies that helps in solving these problems. Feasibility study plays an important role in system analysis which gives the target for design and development.

5.1 **Feasibility Study:** All systems are feasible when provided with unlimited resource and infinite time. It is a formally documented output that summarizes results of the analysis and evaluations conducted to review the proposed solution and investigate project alternatives for the purpose of identifying if the project is really feasible, cost-effective and profitable. It describes and supports the most feasible solution applicable to the project. So it is both necessary and prudent to evaluate the feasibility of the system at the earliest possible time. If project risk is great, the feasibility of producing quality software is reduced. In this case there are three key considerations involved in the feasibility analysis.

5.1.1 **Economical Feasibility**: As we all know hospitals are expensive and going to it cannot be ignored if you are sick. Going to a hospital and getting tested regularly is very costly. This code helps us to early identify the disease within first few visits. The money required for multiple testing is saved.

5.1.2 Technical Feasibility: The algorithm takes data from multiple sources and brings it down to a few specific categories. This gives accurate information of the condition of the patient and the accurate output if the image is Hazy or not.

5.1.3 Social Feasibility: As the machine does most of the analyzing part of the data generated. The doctors and staff is free to handle other more necessary tasks in hand and it helps in betterment of the health care society

5.2 Analysis

5.2.1 Technical Analysis The performance of the system can be increased if the technical analysis is done well. The systems hardware requirements must be taken into consideration. The software must go hand in hand with the hard ware else the efficiency of the system deteriorates.

1. Changes to bring in the system: All changes should be in positive direction, there will be increased level of efficiency and better customer service.

2. Required skills: Platforms tools used this project are widely used.

3. Acceptability: The structure of the system is kept feasible enough so that there should not be any problem from the users point of view

CHAPTER 6

SYSTEM DESIGN

System configuration is the procedure for characterizing a framework's architecture, parts, components, interfaces, and information to fulfill the specified preconditions. Configuration of frameworks could be interpreted as using the theory of frameworks to advance objects. The study and techniques situated in Article develop into the most commonly used methods for the design of PC frameworks. In this way, the configuration of frameworks is the way to characterize and build frameworks to meet the client's defined needs. The UML has become the standard language in object situated investigation and structure.

6.1 System Architecture

Structure design is an applicable model that characterizes the framework's structure and behavior. This includes the frame pieces and the relationship explaining how they work together to modify the overall structure.

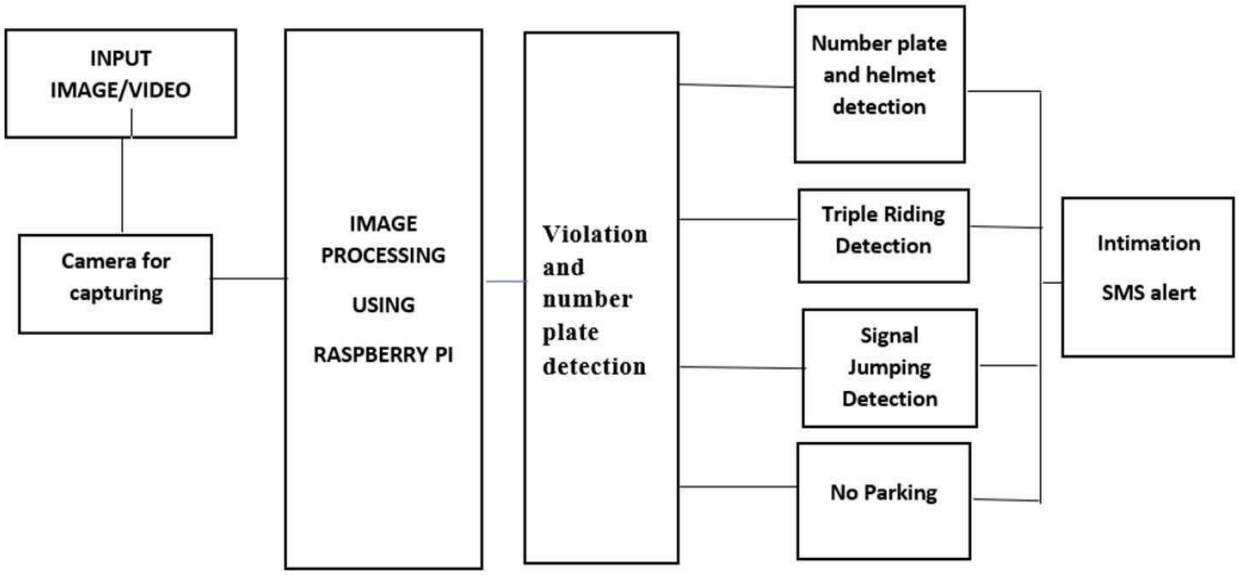


Fig No.6.1.1: Architectural Design

The fig.6.1.1 depicts the basic system architecture which includes the input image as a video through camera and this image is processed and if any violations then helmet, triple riding, signal

jumping and no parking are all detected

6.2 Input / Output Design

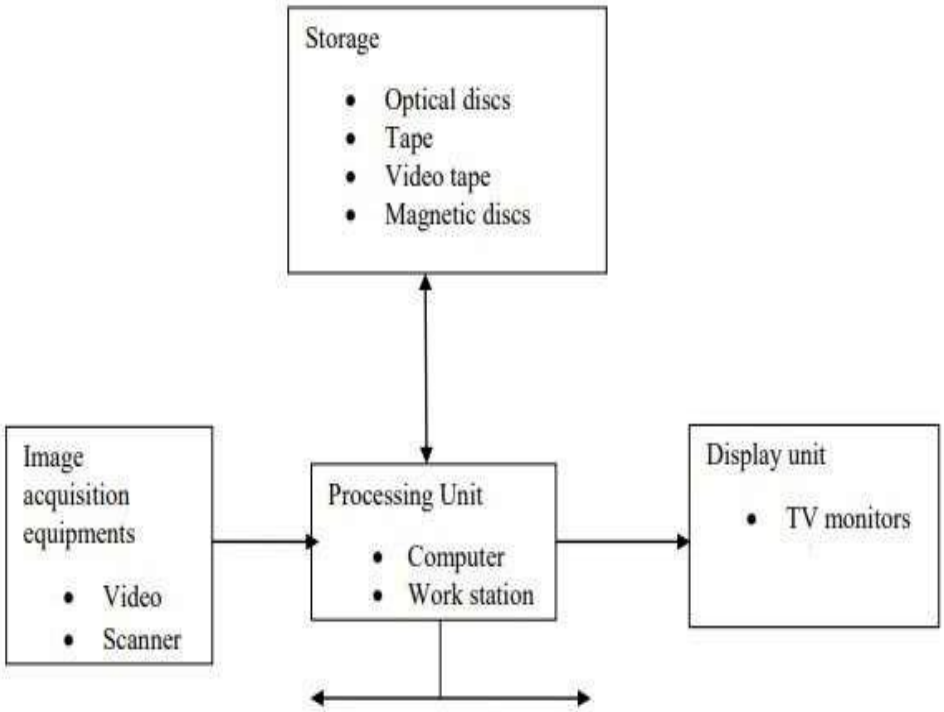


Fig No.6.2.1: Input/output design

The Fig 6.2.1 shows the input and output diagram in input the image acquisition equipments are video, scanner this is further processed through raspberry pi and it is displayed in the computer monitor and which are stored in optical discs, video tape and magnetic discs.

6.3 Object Oriented Design

During the organized stage, the point of view on the application made during the raised level arrangement is isolated into modules and tasks. Justification design is cultivated for each program and a while later recorded as program points of interest. For each program, a unit test plan is made.

6.4 Class Diagram:

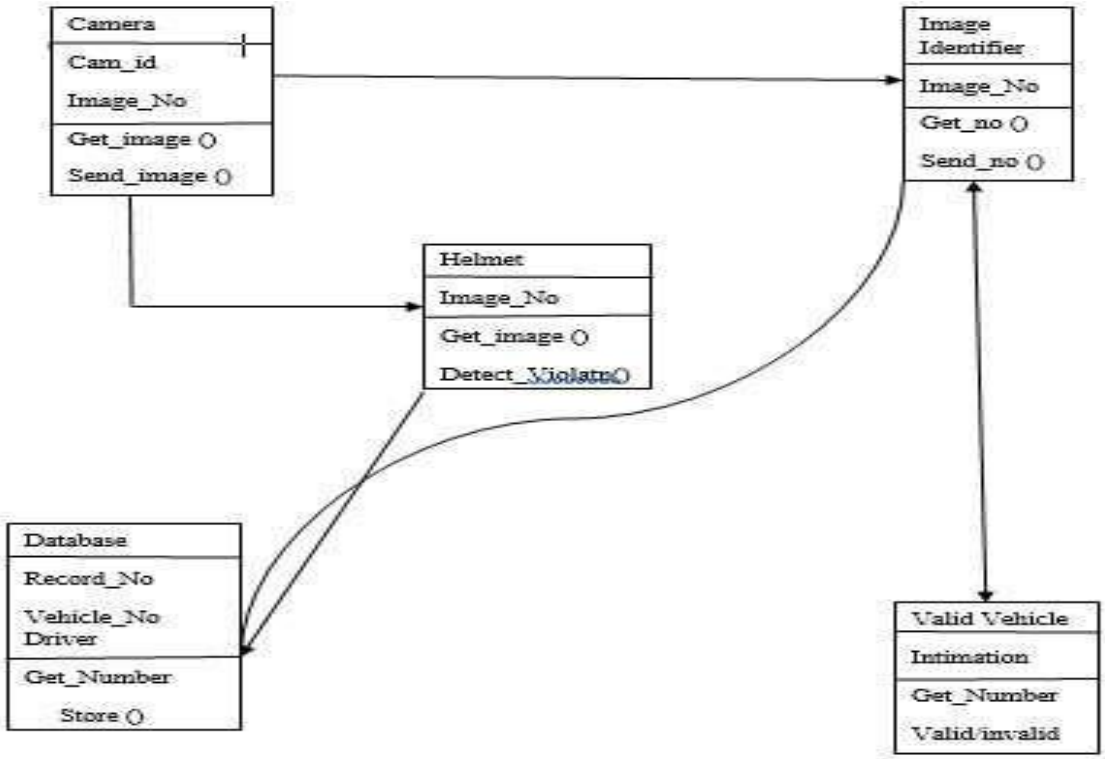


Fig No.4.3.1: Class diagram

The Fig 4.3.1 shows the class diagram in which their are five classes they are camera,image iidentifier,helmet,database,valid vehicle

6.5 USE CASE DIAGRAM

The use case definition at its simplest is an illustration of the client's relationship with the application that demonstrates the client's relation to the distinctive use cases that include the client. A case graph of use identifies the various types of customers of a process and the specific cases of use and is often followed by different types of graphs as well. While an utilization case itself may bore into a great deal of insight concerning each probability, an utilization case graph can help give a more significant level perspective on the framework. It has been said before that "Utilization case graphs are the plans for your framework". They give the improved and graphical portrayal of what the framework should really do.

Use case 1:

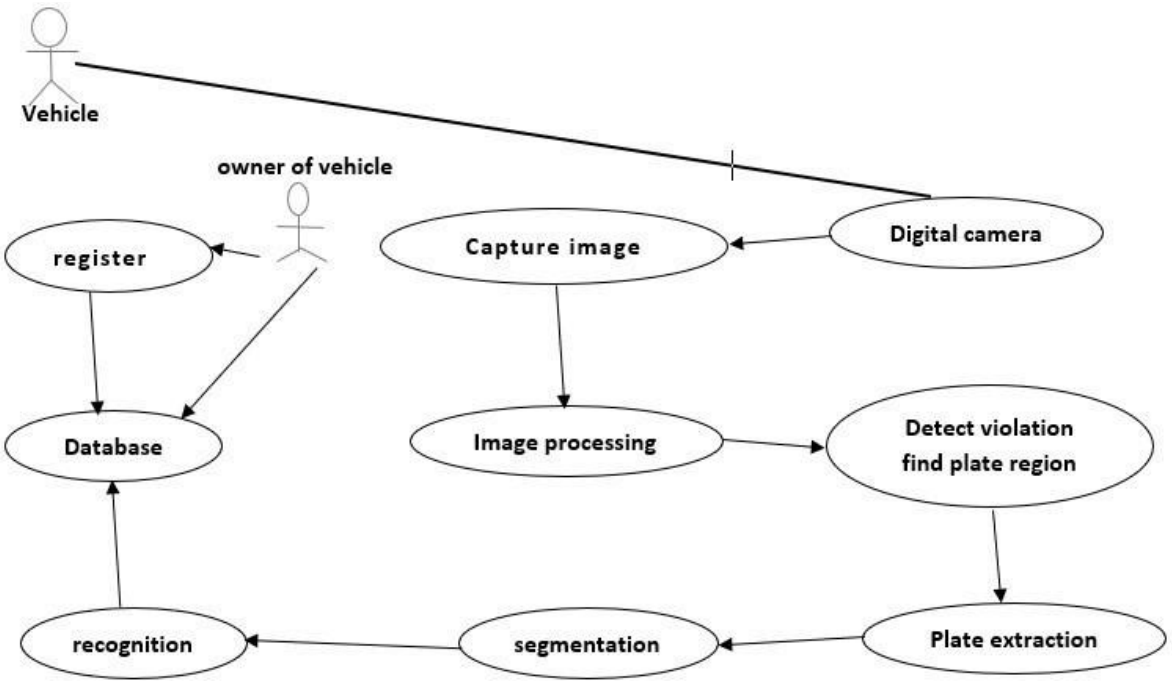


Fig No.6.5.1: Use case1 diagram

The Fig 6.5.1 shows the use case diagram in which camera will capture the image and image processing step take place in this diagram.

Use case2:

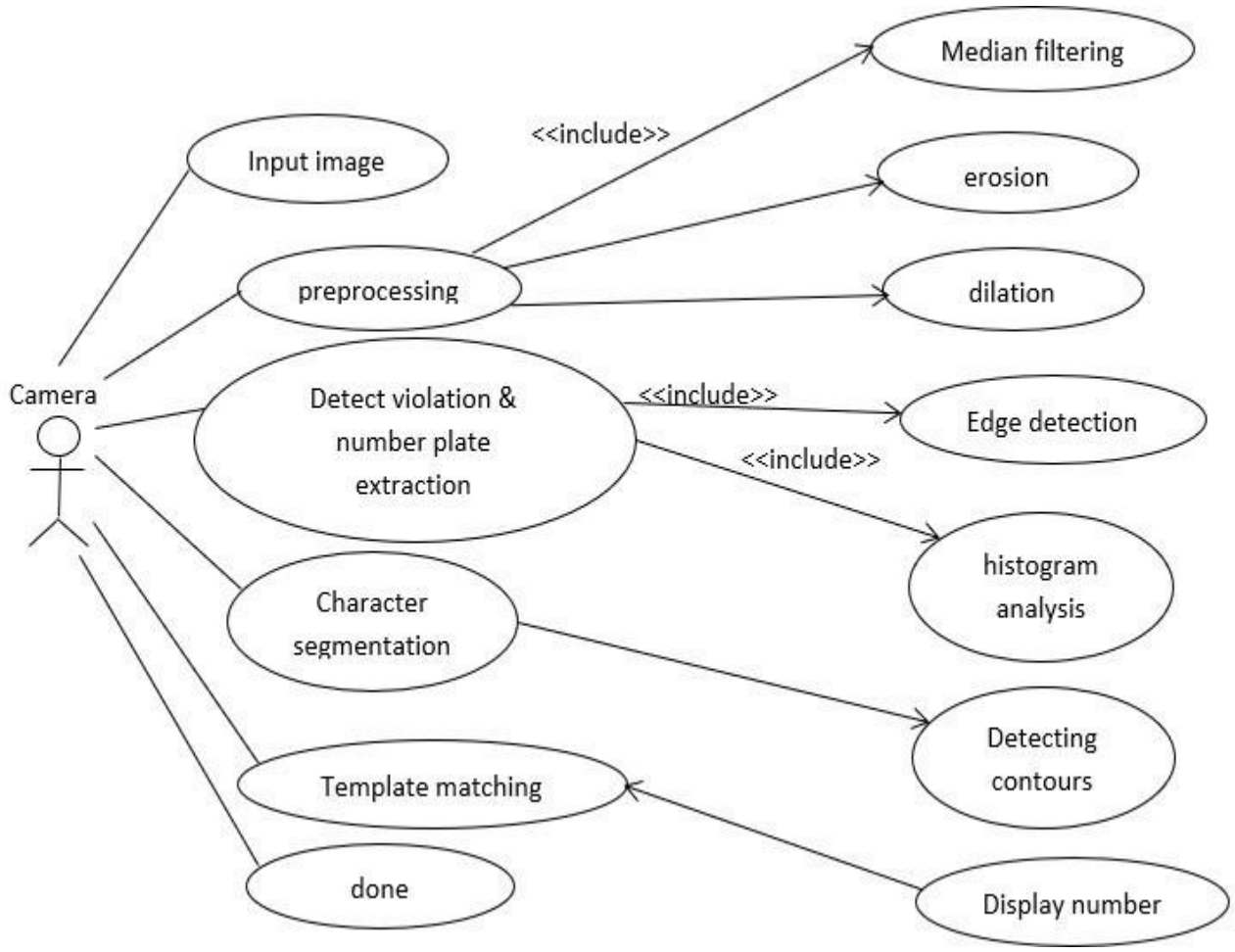


Fig No.6.5.22: Usecase2 diagram

The Fig 6.5.2shows the use case diagram in which camera will capture the image and then image processing steps take place .

6.6 Sequence Diagram:

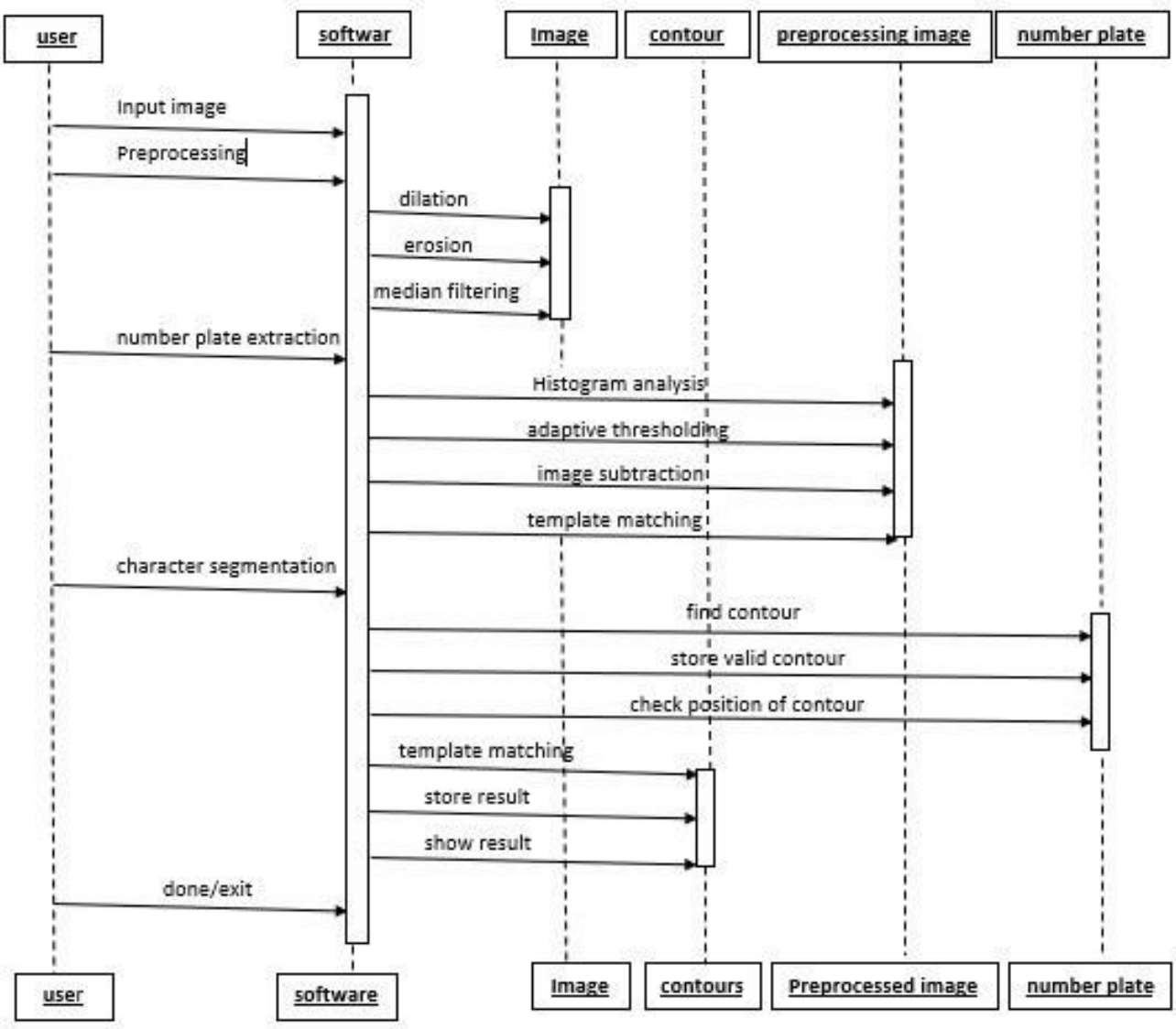


Fig No.6.6: Sequence Diagram

The Fig 6.6 shows the sequence of procedure take place in the modules.

6.7 Activity Diagram:

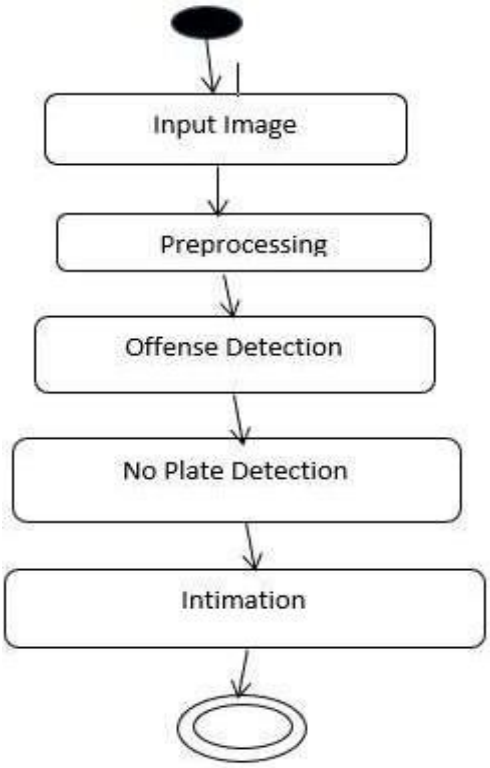


Fig No.6.7: Activity diagram The

Fig 6.7 shows the activity take place in project.

6.8 DataFlow Diagram

Level 0:

The information flow diagram is a schematic diagram of the "path" of the data system, showing the angles of the process. A DFD is used regularly as a first step to make a framework outline without broad explanation, which can be explained later. DFDs can also be used to reflect the handling of data. The DFD demonstrates what kind of data will contribute to and produce from the system, how information will advance through the framework, and where information will be discarded.

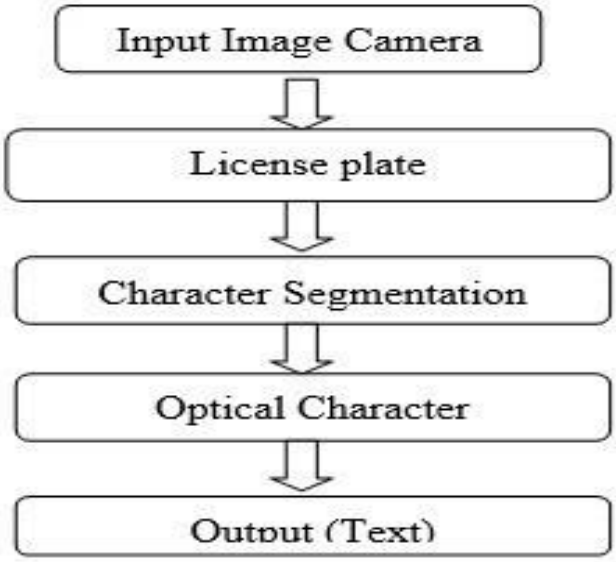


Fig No.6.8: Dataflow Level0 outline

This Fig 6.8 Dataflow Level0 outline shows the reflect of handling of data. Level1:

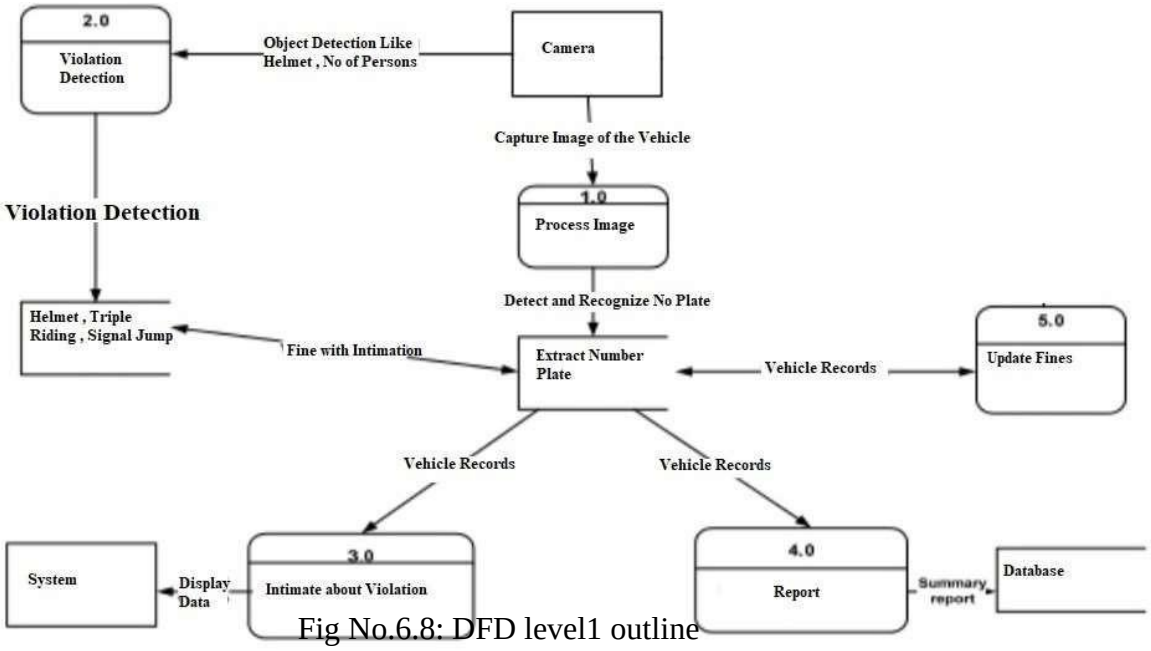


Fig No.6.8: DFD level1 outline

This Fig 6.8 Dataflow Level1 outline shows the storing of data in the database.

6.9 Algorithms

YOLO Algorithm

YOLO You Only Look Once (YOLO) was developed to create a one step process involving detection and classification. Bounding box and class predictions are made after one evaluation of the input image.

Confidence here refers to the probability an object exists in each bounding box and is defined as: C = P r (Object) ∗ IOUtruth pred (1) where IOU, intersection over union, represents a fraction between 0 and 1.

The class-specific probability for each grid cell is defined as:

$$\begin{aligned} &Pr\left(Class_i|Object\right) * Pr\left(Object\right) * IOU_{pred}^{truth} \\ =&Pr\left(Class_i\right) * IOU_{pred}^{truth}. \end{aligned}$$

YOLO uses the following equation below to calculate loss and ultimately optimize confidence:

$$\begin{aligned} Loss = & \\ & \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{A} \mathbb{1}_{ij}^{obj} [(b_{x_i} - b_{\hat{x}_i})^2 + (b_{y_i} - b_{\hat{y}_i})^2] \\ & + \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{A} \mathbb{1}_{ij}^{obj} [(\sqrt{b_{w_i}} - \sqrt{b_{\hat{w}_i}})^2 + (\sqrt{b_{h_i}} - \sqrt{b_{\hat{h}_i}})^2] \\ & + \sum_{i=0}^{s^2} \sum_{j=0}^{A} \mathbb{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 \\ & + \lambda_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^{A} \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{s^2} \mathbb{1}_{i}^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2. \end{aligned} \tag{3}$$

OCR Algorithm

b. Image Acquisition

The first step is to acquire images of paper documents with the help of optical scanners. This way, an original image can be captured and stored. Most of the paper documents are black and white,

and an OCR scanner should be able to threshold images. In other words, it should replace each pixel in an image with a black or a white pixel. It is a method of image segmentation.

c. Preprocessing: The goal of preprocessing is to make raw data usable by computers. The noise level on an image should be optimized and areas outside the text removed. Preprocessing is especially vital for recognizing handwritten documents that are more sensitive to noise. Preprocessing allows obtaining a clean character image to yield better results of image recognition.

d. Segmentation: The process of segmentation is aimed at grouping characters into meaningful chunks. There can be predefined classes for characters. So, images can be scanned for patterns that match the classes.

e. Feature Extraction

This step means splitting the input data into a set of features, that is, to find essential characteristics that make one or another pattern recognizable. As a result, each character gets classified in a particular class.

f. Training a Neural Network

Once all the features are extracted, they can be fetched to a neural network (NN) to train it to recognize characters. A training dataset and the methods applied to achieve the best output will depend on a problem that requires an OCR-based solution.

g. Post-Processing

This stage is the process of refinement as an OCR model can require some corrections. However, it isn't possible to achieve 100% recognition accuracy. The identification of characters heavily depends on the context. The verification of the output requires a human-in-the-loop approach.

CHAPTER 7

SYSTEM IMPLEMENTATION

Utilization is the affirmation of an application or execution of a course of action, thought, model, plan, specific, standard, estimation, or system. In that capacity, Use is a declaration of a computer, programming or other PC process through programming and programming action.
There may be different executions for a given insurance or standard.

Modules

- Data Collection
- Python OpenCV
- Object Detection
- Tensorflow
- OCR

Modules description

Data Collection

Smart Camera:

The sharp camera contains a twofold processor which enables the camera to work honorably with applications, for example, design coordinating, optical character affirmation and data organize code continuously. The whole plan conveys significant standards diminish scale pictures which are indispensable for the execution of the persistent eye following estimation. Exactly when the sagacious camera is related through FTP or Ethernet hard drive, the data would then have the option to be moved to various devices. In this venture we are utilizing web camera to illustrate.

Python OpenCV

OpenCV is a library of PC vision produced initially by Intel and now maintained by Willow Garage. It is safe to use under the license of the open source BSD. It's a cross platform library. For the most part, it centers around the preparation of the picture. If the library discovers

the system's Integrated Performance Primitives, these business updated schedules will be used

to speed it up. OpenCV isn't a piece of code that pictures are running and processing. You've got to write script.

The Visual Studio Express Edition of Microsoft can be downloaded (for nothing). It's an outstanding IDE. The Visual C++ 2010 Express must be downloaded.

Likewise, OpenCV is not an executable file that you replicate and it will start to work. It is code, library files, and DLL documents that are unadulterated. You "link" to these library records when you write your own code to get to the capabilities of OpenCV.

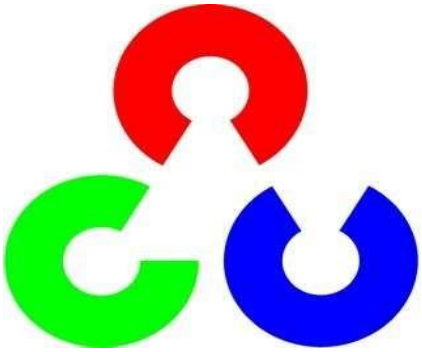


Fig No.7.1: OpenCV

Why OpenCV?

There are two or three for what reason to incline toward OpenCV over Matlab.

Explicit OpenCV was created for the handling of the image. In view of the Image Processing Coder, that capability and information structure has been designed. Matlab is very non-exclusive, then again. As a tool cache, you get almost anything on the planet. From monetary tool kits to DNA tool kits that are exceptionally specific.

Speedy:

Matlab is just too moderate. Matlab is based on Java itself. Moreover, Java is based on C. So when you run a Matlab program, your PC will be caught trying to translate all that code from Matlab. It transforms it into Java at that point and then finally executes the code.

You don't burn through this time on the off chance that you use C / C++. You send the PC machine language code straightforward, and it is executed. And eventually you get to plan more videos, not to translate more.

Sure you take care of speed – a progressively enigmatic language to manage, yet it's very justified, despite all the trouble. You can do much more. You could do some extremely unpredictable arithmetic on pictures with C and still pull off sufficient rates for your application.

Efficient:

Matlab uses only an excessive amount of framework assets. For OpenCV, for a real-time program, you can pull off as little as 10 mb RAM. In any case, the RAM factor is definitely not a big thing to stress about with the current PCs. You need to take care of spills in your brain, but it's not that disturbing. This report on the control of memory can be viewed in OpenCV on the off chance you like.

Requirements for OpenCV:

Operating System:

32- bit MS Windows (95/98), 32-bit MS Window, All POSIX (Linux/BSD/UNIX-like OSes), OS X, Linux, Win2K, WinXP

- It should also be compatible with Windows XP SP3 and newer.
- OpenCV 2.1 is compatible with VC++ 2008 and VC++ 2010.

Object Detection

Article Detection is the way toward discovering true item occasions like vehicle, bicycle, TV, blooms, and people in still pictures or Videos. It takes into account the acknowledgment, confinement, and recognition of numerous items inside a picture which gives us a greatly improved comprehension of a picture everything in all. It is widely used in appliances such as image recovery, safety, observation, and propelled driver help frameworks (ADAS).

Article Detection should be possible by means of various ways:

- Detection of objects based on their characteristics
- Detection of objects by viola Jones
- SVM HOG identification features

- Advanced Object Learning Identification.

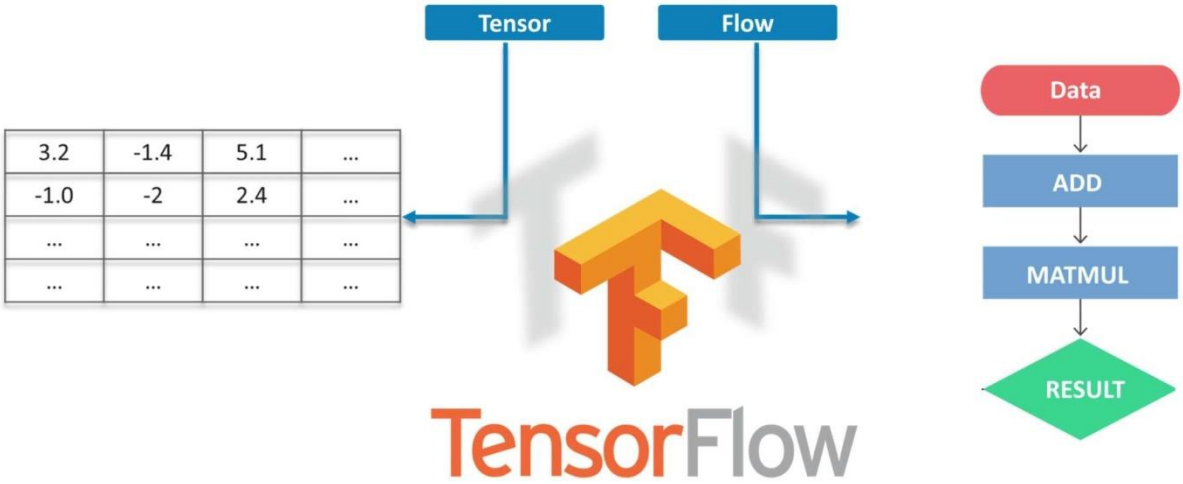
Applications of Detecting the Object

Facial Recognition:

The face acknowledgment framework the "DeepFace" was created by a Facebook gathering of experts who recognize human faces in an advanced picture successfully. Google utilizes its very own facial acknowledgment framework in Google Photos, which consequently isolates all the photographs dependent on the individual in the picture. Facial recognition is correlated with various parts such as the ears, nose, mouth, and eyebrows.

Tensorflow

Tensorflow is Google's Open Source Machine Learning Platform for programming data flow over a count of tasks. Hubs refer to numerical functions, while the edges of the diagram speak to the multi-dimensional points of information (tensors) within them.



Tensors are simply multidimensional clusters, an enlarge to information with a higher measurement of 2-dimensional tables. Tensorflow's numerous highlights make it suitable for profound learning. By burning through whenever, along these lines, how do we interpret how we can perform object detection using Tensorflow.

What is YOLO and for what reason is it Useful?

The R-CNN group of strategies we found in Part 1 fundamentally use districts to confine the articles inside the picture. The system doesn't take a gander at the whole picture, just at the pieces of the pictures which have a higher possibility of containing an article.

The YOLO system (You Only Look Once) then again, manages object recognition in an alternate way. It takes the whole picture in a solitary case and predicts the bouncing box arranges and class probabilities for these containers. The greatest favorable position of utilizing YOLO is its eminent speed – it's fantastically quick and can process 45 casings for each second. YOLO additionally comprehends summed up object portrayal.

OCR

OCR (optical character acknowledgment) is the utilization of innovation to recognize printed or transcribed content characters inside advanced pictures of physical archives, for example, a filtered paper record. The essential procedure of OCR includes analyzing the content of an archive and interpreting the characters into code that can be used to handle information. OCR is now and again additionally alluded to as content acknowledgment.

OCR frameworks are comprised of a blend of equipment and programming that is utilized to change over physical records into machine-discernible content. Equipment, for example, an optical scanner or particular circuit board is utilized to duplicate or understand content while programming commonly handles the propelled preparing. Programming can likewise exploit computerized reasoning (AI) to actualize further developed techniques for insightful character acknowledgment (ICR), like distinguishing dialects or styles of penmanship.

The procedure of OCR is most regularly used to transform printed version legitimate or noteworthy records into PDFs. When put in this delicate duplicate, clients can alter, configuration and search the record as though it was made with a word processor.

How optical character acknowledgment functions

The initial step of OCR is utilizing a scanner to process the physical type of a record. When all pages are duplicated, OCR programming changes over the archive into a two-shading, or high contrast, form. The examined in picture or bitmap is dissected for light and dim territories, where the dull regions are distinguished as characters that should be perceived and light regions are recognized as foundation.

The dull zones are then prepared further to discover alphabetic letters or numeric digits. OCR projects can shift in their methods, yet ordinarily include focusing on one character, word or square of content at once. Characters are then recognized utilizing one of two calculations:

Example acknowledgment OCR programs are encouraged instances of content in different text styles and arrangements which are then used to analyze, and perceive, characters in the filtered report.

Highlight identification OCR programs apply rules with respect to the highlights of a particular letter or number to perceive characters in the examined record. Highlights could incorporate the

quantity of calculated lines, crossed lines or bends in a character for examination. For instance, the capital letter "A" might be put away as two corner to corner lines that meet with a flat line over the center. At the point when a character is distinguished, it is changed over into an ASCII code that can be utilized by PC frameworks to deal with further controls. Clients should address fundamental blunders, edit and ensure complex formats were dealt with appropriately before sparing the record for sometime later.

OCR can be utilized for an assortment of utilizations, including:

- Scanning typed renditions that can be changed by word processors such as Microsoft Word or Google Docs.
- Web crawler print material indexing.
- Automating the passage, extraction and preparation of information.
- Deciphering content documents that can be perused to allow anyone to hear impeded or daze customers from the outside.
- Archiving noteworthy data, for example, papers, magazines or phonebooks, into accessible configurations.
- Electronically storing checks without the requirement for a bank employee.
- Placing significant, marked l

The main advantages of OCR technology are saved time, decreased errors and minimized effort. It also enables actions that are not capable with physical copies such as compressing into ZIP files, highlighting keywords, incorporating into a website and attaching to an email.

CHAPTER 8

SYSTEM TESTING

Testing

Testing is the method of testing a system or its components with the intention of deciding whether it meets the specified specifications or not. Testing involves running a program to detect any holes, bugs or incomplete specifications which contradict the actual requirements.

Testing Principle

A software engineer must grasp the fundamental theory that drives the software testing before implementing methods to design successful test cases. All the results will be traceable according to consumer needs.

Testing Methods

There are various methods which can be used for evaluating applications. They are

1. Black-Box Testing

The research technique is called black-box testing without understanding the implementations of the inner workings. The tester does not know the architecture of the framework, and has no access to the source code. Typically, when running a black-search, a tester may interact with the device's user interface, presenting inputs and evaluating outputs without knowing how and where the inputs are being processed.

2. White-Box Testing

White-box testing is a thorough study of the internal logic and the structure of the code. White-testing is also called glass testing or open-testing. A tester needs to learn the internal workings of the code to do the white-box testing on an application. To assess which unit / chunk of code is behaving inadequately, the tester will look inside the source code.

Manual and Automated Testing

Manual Testing

This is a type of black-box testing, to be evaluated based on the software requirements. The application is reviewed by providing feedback, and it then discusses the results that need to conform to the functionality for which it was designed. Manual software testing is performed on a complete, integrated system to determine the system's compliance with its specified specifications. When testing a function program there are 5 steps involved.

• Determining the features to be exercised by the intended program.

• Production of test data according to application requirements.

• Quality based on the specifications of the test data and the program.

• Drawing up test scenarios and running test cases.

• Review of real and anticipated outcomes, based on the performed test cases.

Automated Testing

Testing an application from its automated attributes is based on this section. Automated testing includes testing software from non-functional specifications, but essential in nature such as performance, protection, user interface, etc. Testing can be performed at various SDLC rates.

Unit Testing

Unit testing is a software development process in which the smallest testable parts of an program, called units, are tested individually for proper functioning. Unit monitoring is often automated but can be performed manually, too. The aim of unit testing is to isolate each part of the system, and to demonstrate that the specifications and functionality of individual components are right. The tables detail test cases and measurements.

Unit testing benefits

• Unit testing improves trust that code is changed / maintained.

• Development is faster.

• The cost of repairing a defect found during device testing is lower than that of higher-level found defeats

• Easy to debug.

• Codes are more reliable.

S1 #Test Case: -	**UTC-1**
Name of Test: -	**Image Capture and Helmet Detection**
Items being tested: -	**Helmet Detection**
Sample Input: -	**Image**
Expected output: -	**Helmet should be detected**
Actual output: -	**Detection of helmet is done successfully**
Remarks: -	**Pass**

Table No.8.1: Helmet Detection

Sl # Test Case : -	**UTC-2**
Name of Test: -	**Triple riding**
Items being tested: -	**No of persons on vehicle**
Sample Input: -	**Image**
Expected output: -	**System should detect no of persons**
Actual output: -	**No of persons detected successfully**
Remarks: -	**Pass**

Table No.8.2: Unit Test Case for Person Detection

Sl #Test Case: -	**UTC-3**
Name of Test: -	**Signal jumping**
Items being tested: -	**Red Signal Detection and Intimation**
Sample Input: -	**Image**
Expected output: -	**Signal Violation Detection**
Actual output: -	**Vehicle detection on signal jumping is achieved**
Remarks: -	**Pass**

Table 8.3: Unit test case for Signal Jumping Table

Sl #Test Case: -	**UTC-4**
Name of Test: -	**No parking**
Items being tested: -	**Parking Area Detection and Intimation**
Sample Input: -	**Image**
Expected output: -	**Parking Area Detection**
Actual output: -	**Vehicle detection on parking area is achieved**
Remarks: -	**Pass**

Table 8.3: Unit test case for Parking Area

Integration Testing

Integration testing is a type of software testing where single units are combined and tested as a group. The goal of this test level is to identify faults in the relationship between integrated units. Test drivers and test stubs are used to help in performance checking. Integration testing is defined as checking the combined parts of an application to see if they function properly. It occurs after device testing and before validation tests. Integration testing can be carried out in two ways: from the bottom up, integration testing and from the top down, integration testing.

Bottom-up Integration

The cycle starts with unit testing followed by testing combinations of progressively higher level units called modules or builds.

Top-down Integration

The highest-level modules are tested in this study first, and then, lower-level modules are tested gradually. Bottom-up testing is typically performed first in a detailed software development environment, followed by top-down testing. The cycle ends with several full-application evaluations, preferably in environments designed to simulate real situations.

Sl # Test Case : -	**ITC-1**
Name of Test: -	**Synchronization Testing**
Items being tested: -	**Helmet, Signal jumpings, Triple Riding with No Plate, No Parking**
Sample Input: -	**Image**
Expected output: -	**System Should Automatically Check Violation and Intimate**
Actual output: -	**All functions work properly**
Remarks: -	**Pass**

Table 8.5: Integration test case

Chapter 7

RESULTS AND DISCUSSION

A program is being developed in this research to identify the motorcyclists who violate the laws of wearing the helmet. The program consists primarily of three parts–motorcycle identification, helmet identification, and license plate recognition of motorcyclists riding without a helmet. The key criterion is to determine whether or not the picture captured is of a motorcycle using HOG, and to test whether or not the motorcyclist is wearing a helmet using CNN. If the motorcyclist is marked without a helmet, then the motorcyclist's license plate is recognized using tesseract OCR. The accuracy of the motorcycle / non-motorcycle classification is 93%, the helmet / no-helmet classification is 85% and the license plate identification is 51%, resulting in an overall accuracy of about 76%. Increasing the training data collection and image quality will increase the accuracy.