# A PROJECT REPORT ON

"AUTOMATIC NUMBER PLATE RECOGNIZATION"

**Submitted**

**In partial fulfillment of the requirements**

**for the Degree of**

**MASTERS OF COMPUTER APPLICATIONS**

**By**

**SAURABH BHATT**

**(Enrollment Number: PV-D1921159)**

**Under the Supervision of**

**AMIT JUYAL**

**( Assistant Professor, GEHU)**


**SCHOOL OF COMPUTING**

**GRAPHIC ERA HILL UNIVERSITY,DEHRADUN (2022-2023)**

# CERTIFICATE

I hereby declare that the work which is being presented in the project entitled, “HEART DISEASEPREDICTION” has been carried out by **SAURABH BHATT** for the partial fulfilment of the requirements for the award of the MCA, submitted in the School of Computing, **GRAPHIC ERA HILL UNIVERSITY** is an authentic record of our own **AMIT JUYAL**. I further declare that the matter embodied in this project has not been submitted by us for the award of any other degree.

**PROJECT INCHARGE**

**AMIT JUYAL**

**(ASSISTANT PROFESSOR)**

# ACKNOWLEDGEMENT

This is a great opportunity to acknowledge and to thanks all those persons without whose support and help this project would have been impossible. I would like to add a few heartfelt words for the people who were part of this project in numerous ways.

I am obliged and thankful to my project SUPERVISOR, **AMIT JUYAL**, for her continuous encouragement, motivation and professional guidance during the work of this project which has proven to be an integral part of it. Without her valuable support and guidance, this project could not elevate up this level of development from our point of view. I would like to thank all the Faculty members, School Of Computing, Graphic Era Hill University for their valuable time spent in requirements analysis and evaluation of the project work.

I would like to express my sincere and cordial gratitude to the people those who have supported me directly, purveyed mental encouragement, evaluated and criticized my work in several phases during the development ofthis project and for preparing this dissertation indirectly.

# ABSTRACT

This final project develops an algorithm for automatic number plate recognition (ANPR). ANPR has gained much interest during the last decade along with the improvement of digital cameras and the gain in computational capacity.
The text is divided in four chapters. The first, introduces the origins of digital image processing, also a little resume about the following algorithms that are needed for develop the system ANPR.

The second chapter presents the objectives to be achieved in this project, as well as, the program used for his development. The following chapter explains the different algorithms that compound the system, which is built in five sections; the first is the initial detection of a possible number plate using edge and intensity detection to extract information from the image.

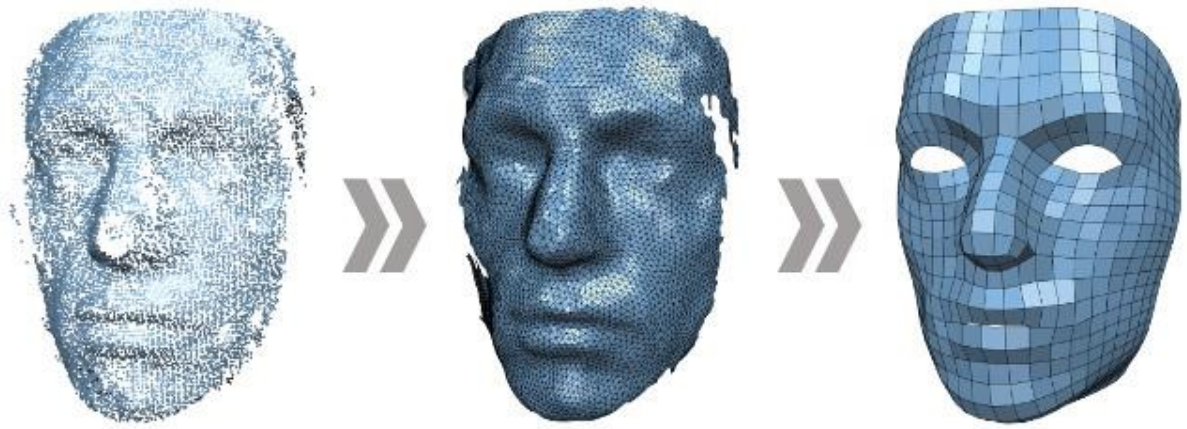
The second and the third step, thresholding and normalization, are necessary to use the images in the following stages; the text of the plate is found and normalized. With the segmentation, each character of the plate is isolated for subsequent recognition. The last step reads the characters by correlation template matching, which is a simple but robust way of recognizing structured text with a small set of characters. It is evaluated the system's speed and his error rate. Finally, the conclusions and future works are shown in the chapter four.

# Table of Contents

# CHAPTER-1
# INTRODUCTION

## DIGITAL IMAGE PROCESSING:

It refers to process real world images digitally by a computer. It is a broad topic, which includes studies in physics, mathematics, electrical engineering, computer science. It studies the conceptual foundations of the acquisition and deployment of images and in detail the theoretical and algorithmic processing as such. It also aims to improve the appearance of the images and make them more evident in certain details that you want to note. This chapter doesn't intend to provide a detailed explanation of digital image processing, but yes an overview of those concepts and methods more important for the realization of this project.



## IMPORTANCE OF THE IMAGES:

The human uses the senses to iterate with the world they live. The senses allow you to know reality. This way we grasp information about the world around us. We can feel objects, identify smells, hear sounds, detect flavors and most importantly we can see the space in which we live. Of all the senses the most developed is in sight. It is the means by which we receive information. It allows us to perceive and understand the world around us and accounts for nearly seventy percent of the information we receive. Among this type of information include the identification of faces, reading, images, etc...

# AUTOMATIC NUMBER PLATE RECOGNITION:

Due to the mass integration of information technology in all aspects of modern life, there is a demand for information systems for data processing in respect of vehicles.These systems require data to be archived or by a human or by a special team which is able to recognize vehicles by their license plates in real-time environment and reflect the facts of reality in the information system. Therefore, several techniques have been developed recognition and recognition systems are license plates used today in many applications.

In most cases, vehicles are identified by their license plate numbers, which are easily readable by humans but not machines. For machines, a registration number plate is just a dark spot that is within a region of an image with a certain intensity and luminosity. Because of this, it is necessary to design a robust mathematical system able to perceive and extract what we want from the captured image.

These functions are implemented or mathematical patterns in what is called "ANPR Systems" (Automatic Numbers Plate Recognition) and mean a transformation between the real environment is perceived and information systems need to store and manage all that information.

The design of these systems is one of the areas of research in areas such as Artificial Intelligence, Computer Vision, Pattern Recognition and Neural Networks. Systems of automatic recognition of license plates are sets of hardware and software to process a signal that is converted into a graphical representation such as static images or sequences of them and recognize the characters in the plate.

The basic hardware of these systems is a camera, an image processor, an event logger memory and a storage unit and communication. In our project we have relied on images of cars in which we can see their license plate. The license plate recognition systems have two main points: The quality of license plate recognition software with recognition algorithms used and the quality of imaging technology, including camera and lighting.

Elements to consider: maximum recognition accuracy, achieve faster processing speed, handling as many types of plates, manage the broadest range of image qualities and achieve maximum distortion tolerance of input data.

Ideally, for extreme conditions and with serious problems of normal visibility, would have special cameras ready for such an activity, such as infrared cameras that are much better to address these goals and achieve better results.

This is because the infrared illumination causes reflection of light on the license plate is made of special material which causes a different light in that area of the image relative to the rest of it, causing it to be easier to detect.

There are five main algorithms that the software needed to identify a license plate:

1. Location license plate, responsible for finding and isolating the plate in the image. It should be located and extracted from the image for further processing.

2. After the number plate is located and extracted from the image, it can be transformed into into a standard format for brightness and contrast

3. Orientation and plate size, offset angles that make the plate look "crooked" and adjust the Size.

4. Segmentation of individual characters is present in plate.

5. Optical Character Recognition (OCR) for each image we segmented individual character. The output of the recognition of each character is processed as ASCII code associated with the image of the character. By recognizing all successive images of the characters are completely read the license plate.

# Optical Character Recognition:

Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo).

Widely used as a form of data entry from printed paper data records – whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation – it is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

## Pre-processing :

OCR software often "pre-processes" images to improve the chances of successful recognition. Techniques include:

- De-skew – If the document was not aligned properly when scanned, it may need to be tilted a few degrees clockwise or counterclockwise in order to make lines of text perfectly horizontal or vertical.
- Despeckle – remove positive and negative spots, smoothing edges
- Line removal – Cleans up non-glyph boxes and lines
- Layout analysis or "zoning" – Identifies columns, paragraphs, captions, etc. as distinct blocks. Especially important in multi-column layouts and tables

## Text Recognition:

Matrix matching involves comparing an image to a stored glyph on a pixel-by-pixel basis; it is also known as "pattern matching", "pattern recognition", or "image correlation". This relies on the input glyph being correctly isolated from the rest of the image, and on the stored glyph being in a similar font and at the same scale. This technique works best with typewritten text and does not work well when new fonts are encountered. This is the technique the early physical photocell-based OCR implemented, rather directly.

## Post-processing :

OCR accuracy can be increased if the output is constrained by a lexicon – a list of words that are allowed to occur in a document.[15] This might be, for example, all the words in the English language, or a more technical lexicon for a specific field. This technique can be problematic if the document contains words not in the lexicon, like proper nouns. Tesseract uses its dictionary to influence the character segmentation step, for improved accuracy. The output stream may be a plain text stream or file of characters, but more sophisticated OCR systems can preserve the original layout of the page and produce

## Uses of OCR:

- Data entry for business documents, e.g. check, passport, invoice, bank statement and receipt
- Automatic number plate recognition
- In airports, for passport recognition and information extraction
- Automatic insurance documents key information extraction
- Traffic sign recognition
- Extracting business card information into a contact list
- More quickly make textual versions of printed documents, e.g. book scanning for Project Gutenberg
- Make electronic images of printed documents searchable, e.g. Google Books
- Converting handwriting in real time to control a computer (pen computing)
- Defeating CAPTCHA anti-bot systems, though these are specifically designed to prevent OCR. The purpose can also be to test the robustness of CAPTCHA anti-bot systems.
- Assistive technology for blind and visually impaired users
- Writing the instructions for vehicles by identifying CAD images in a database that are appropriate to the vehicle design as it changes in real time.
- Making scanned documents searchable by converting them to searchable PDFs

## OpenCV :

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.

Along with well-established companies like Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota that employ the library, there are many startups such as Applied Minds, VideoSurf, and Zeitera, that make extensive use of OpenCV. OpenCV's deployed uses span the range from stitching streetview images together, detecting intrusions in surveillance video in Israel, monitoring mine equipment in China, helping robots navigate and pick up objects at Willow Garage, detection of swimming pool drowning accidents in Europe, running interactive art in Spain and New York, checking runways for debris in Turkey, inspecting labels on products in factories around the world on to rapid face detection in Japan.

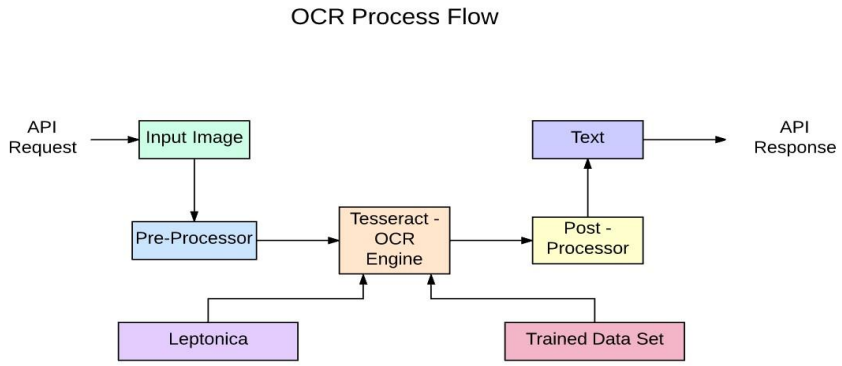
It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being

actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

## Tesseract Engine:

An open-source OCR engine that has gained popularity among OCR developers. Even though it can be painful to implement and modify sometimes, there weren't too many free and powerful OCR alternatives on the market for the longest time. Tesseract began as a Ph.D. research project in HP Labs, Bristol. It gained popularity and was developed by HP between 1984 and 1994. In 2005 HP released Tesseract as an open-source software. Since 2006 it is developed by Google.

Tesseract is an open source text recognition (OCR) Engine, available under the Apache 2.0 license. It can be used directly, or (for programmers) using an API to extract printed text from images. It supports a wide variety of languages. Tesseract doesn't have a built-in GUI, but there are several available from the 3rdParty page. Tesseract is compatible with many programming languages and frameworks through wrappers that can be found here. It can be used with the existing layout analysis to recognize text within a large document, or it can be used in conjunction with an external text detector to recognize text from an image of a single text line.




Tesseract Architecture

# SYSTEM REQUIREMENT SPECIFICATION :

It specifies the hardware and software requirements that are required in order to run the application properly. The Software Requirement Specification (SRS) is explained in detail, which includes overview of dissertation as well as the functional and non-functional requirement of this dissertation.

A SRS document describes all data, functional and behavioral requirements of the software under production or development. SRS is a fundamental document, which forms the foundation of the software development process. It is the complete description of the behavior of a system to be developed. Requirement Analysis discusses the conditions to be met for a new or altered product. Requirement Analysis is critical to the success to a development project. Requirement must be documented, measurable, testable, related to in identified business needs or opportunities, and defined to a level of detail sufficient for system design.

## FUNCTIONAL REQUIREMENTS

Functional Requirement defines in detail how the system must respond to the various kinds of input that is given to the system. It also talks about the expected system behavior under certain conditions. In this system following are the functional requirements:

- **1. Mobility:** The device identifying the License plate should be movable to capture the LP on the go
- **2. Convenience:** The system will make it efficient to access for vehicles and prevent congestion at entry and exit points
- **3. User-Interface:** The system shall provide an easy-to-use user-interface.
- **4. Transparency:** Users should be able to possess a general knowledge and understanding of the ALPR process.
- **5. Flexibility:** The system shall be flexible in that it allows a variety of formats to ingrate the scanning of characters to

- **6. Support for Disabled:** The system shall cater to the needs of physically challenged voters (e.g. blind voters).
- **7. Accuracy:** The system shall accurately convert the image to characters
- **8. Uniqueness:** The system is trained to handle unique and varying types of number plates
- **9. Documentation and Assurance:** The design, implementation, and testing procedures must be well documented so that the confidence is ensured.
- **10. Cost-effectiveness:** Should be affordable and efficient

## NON-FUNCTIONAL REQUIREMENTS:

Non-functional requirements are the requirements which are not directly concerned with the specific function delivered by the system. They specify the criteria that can be used to judge the operation of a system rather than specific behaviours. They may relate to emergent system properties. **Non-functional requirements** are requirements that are not specifically concerned with the functionality of a system. They normally place restrictions on the product being developed and the development process. Non-functional requirements may be regarded as parameters of functionality in that they determine how quickly, how accurately, how reliably, how securely, etc., functions must operate. Some of the ALPR non-functional requirements are as follows:

- The system may issue a receipt to remove any papers printed and make it a green initiative
- The system must be working at 100% peak efficiency
- When checking the database for errors, a 100% scan of the data is required, rather than selecting a sample set
- A process must be devised to support normal precinct business hours
- The system should provide documentation to inform users of system functionality and any change to the system
- The system should provide friendly graphical Interface to ensure ease of use when end users utilize system functionality

## HARDWARE REQUIREMENTS:

· Processor : Intel I5 2.1 Ghz.
· Storage : 100 GB.
· RAM : 4 GB

## SOFTWARE REQUIREMENTS:

- **Platform:** Windows/Linux/macOS
- **Language used:** Python
- **Technologies used:** PyTesseract, OpenCV, Selenium, Chrome Driver, MySQL connector

**Pytesseract :** Library to use the Tesseract-OCR. Tesseract is an optical character recognition engine for various operating systems. Tesseract is considered to be one of the most accurate opensource OCR engines available.

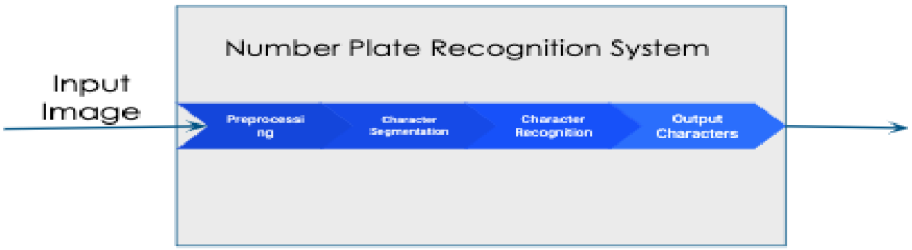
**OpenCV :** OpenCV stands for *Open Source Computer Vision*. It is an open source computer vision and machine learning software library. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc.

**Selenium :** It is a free (open-source) automated testing framework used to validate web applications across different browsers and platforms. You can use multiple programming languages like Java, C#, Python etc to create Selenium Test Scripts. Testing done using the Selenium tool is usually referred to as Selenium Testing. Selenium Software is not just a single tool but a suite of software, each piece catering to different testing needs of an organization.

# CHAPTER-2

# PROJECT-DESIGN

This section elaborately describes the architecture of a Number plate recognition system and an information extractor . The proposed system detects the vehicle number plate and extracts the vehicle number plate in characters. In this module of vehicle number plate recognition, the input image is preprocessed and the characters in the image are segmented. The characters that are segmented are cropped and recognized. The recognized characters are then returned as a string.




ARCHITECTURE

The basic architecture of the system is demonstrated in Fig 5.1. The system has the following set of modules:

- **Number Plate Recognition System**: The input image is captured by the system through a camera feed or an image file. The image is pre-processe**d** by converting the image to a grayscale image and then bilateral filter is applied to smoothen the image further. The edges in the image are detected using the canny edge detection algorithm. Once the canny edges are detected. Contours with 4 edges are ranked . The contour with the highest rank is considered as the number plate region and it is cropped. Pytesseract is an OCR which performs character recognition on this region of the image.

# CHAPTER-3

# FUNCTIONALITY

## IMPLEMENTATION :

## ALGORTIHMS:

**Algorithm to Recognize the Number Plates**

The sequence of processes associated with Number plate recognition is given below. The file upload process is initiated by the data owner entity.
**Input**: Uploading the image file from camera
**Output**: Vehicle number plate in characters

1) Read the original image or Capture the image

2) Resize the image

3) Convert it to grayscale.

4) Apply Bilateral Filter. *What is a bilateral filter* ? A bilateral filter is a non-linear, edge preserving, and noise-reducing smoothing filter for images. It replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels.

5) Identify and store the Canny edges. *What are Canny edges* ? The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images.

6) Find the contours in from the edges detected and sort the top 30 contours.

7) Get the perimeter of each contour and select those with 4 corners.
8) Mask all other parts of the image and show the final image.

9) Read the text using Tesseract OCR

10) Standardize the text to Indian vehicle number plate format

11. Stop

On upload of the image file to the system, the number plate recognition system performs its functions to provide the output.

# CODE:

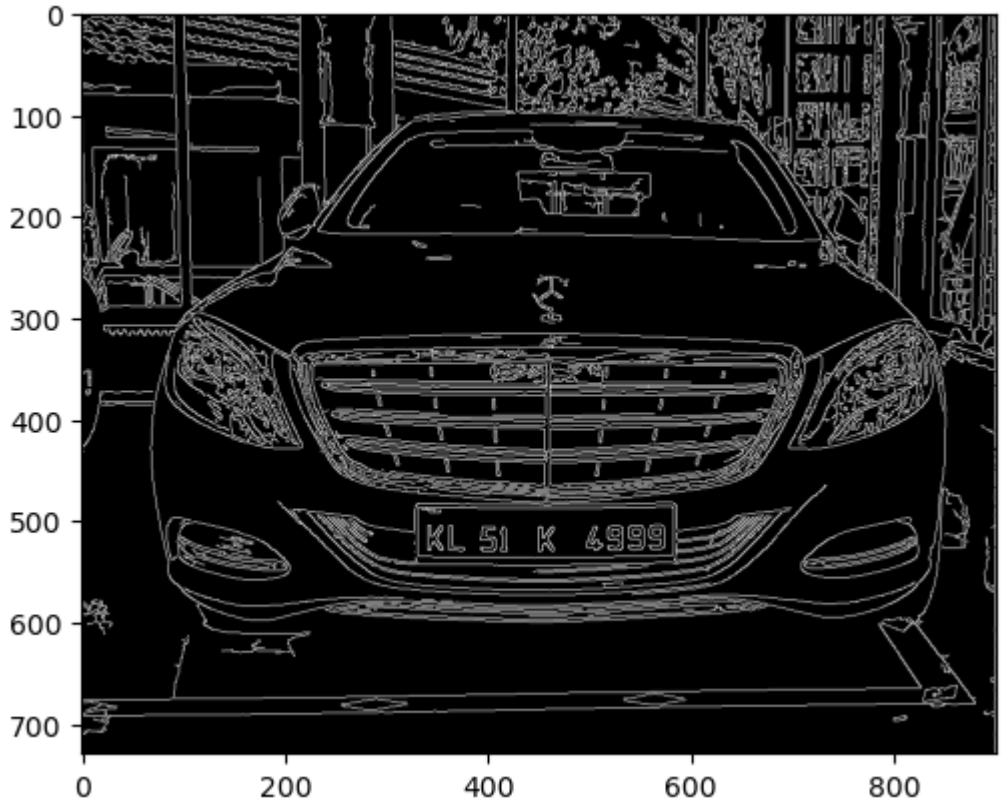
### Code to capture/read the image of the vehicle number plate

```
img=cv2.imread(r'C:\Users\hp\Desktop\ANPR\car6.jpg')
```

### Code to recognize the vehicle number plate

```
grayimg=cv2.cvtColor(img,cv2.COLOR_BGR2RGB)
plt.imshow(cv2.cvtColor(grayimg,cv2.COLOR_BGR2RGB))
```

```
bfilter=cv2.bilateralFilter(grayimg,11,17,17)
edged=cv2.Canny(bfilter,30,200)
plt.imshow(cv2.cvtColor(edged,cv2.COLOR_BGR2RGB))
```

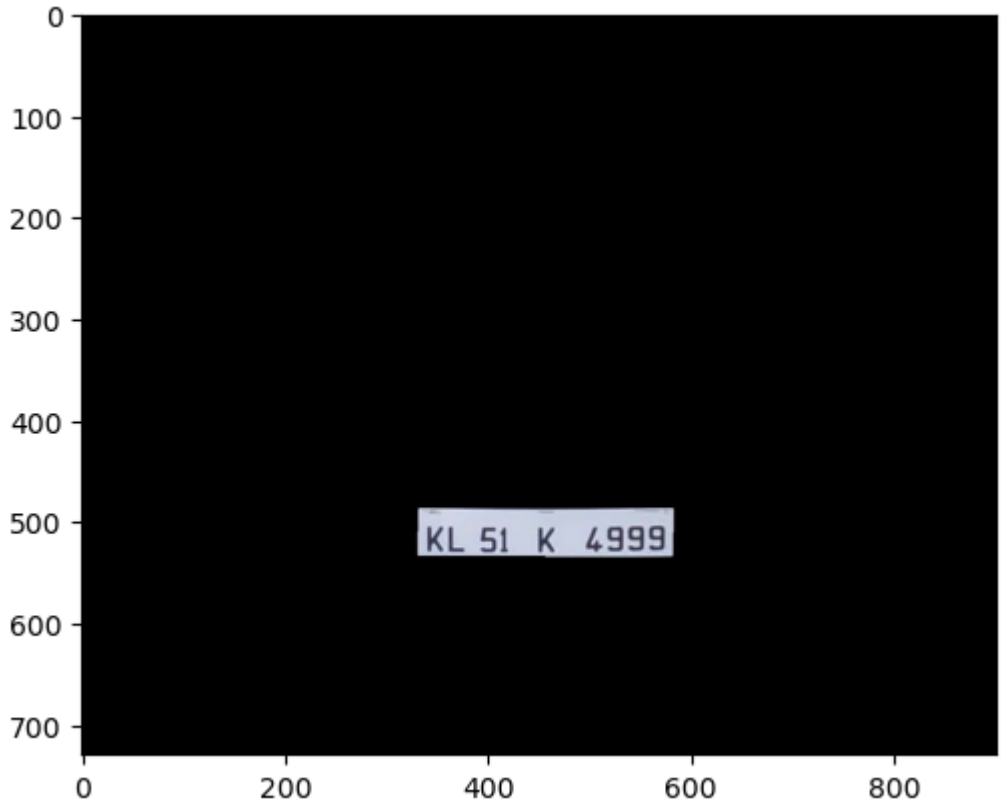



```
keypoints
=cv2.findContours(edged.copy(),cv2.RETR_TREE,cv2.CHAIN_APPROX_SIMPLE)
contours=imutils.grab_contours(keypoints)
contours=sorted(contours,key=cv2.contourArea,reverse=True)[:10]
```

```
location=None
for contour in contours:
  approx=cv2.approxPolyDP(contour,10,True)
  if len(approx)==4:
      location=approx
      break
location
```

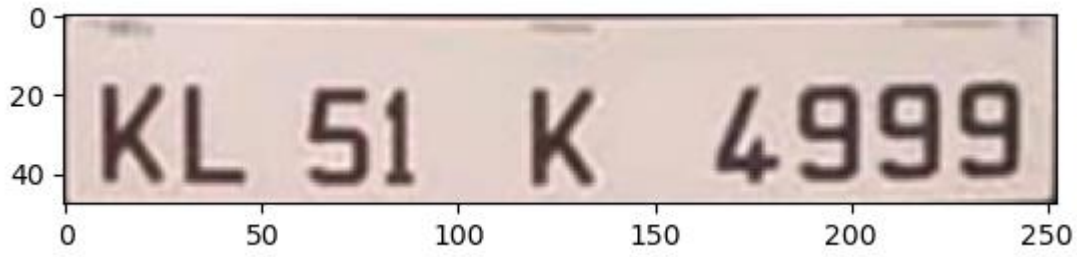
```
mask=np.zeros(grayimg.shape[:2],np.uint8)
new_image=cv2.drawContours(mask,[location],0,255,-1)
new_image=cv2.bitwise_and(img, img, mask=mask)
```

```
plt.imshow(cv2.cvtColor(new_image,cv2.COLOR_BGR2RGB))
```




```
(x,y)=np.where(mask==255)
(x1,y1)=(np.min(x),np.min(y))
(x2,y2)=(np.max(x),np.max(y))
croped_img=grayimg[x1:x2+1,y1:y2+1]
```

```
plt.imshow(cv2.cvtColor(croped_img, cv2.COLOR_BGR2RGB))
result=pytesseract.image_to_string(croped_img)
print(result)
```



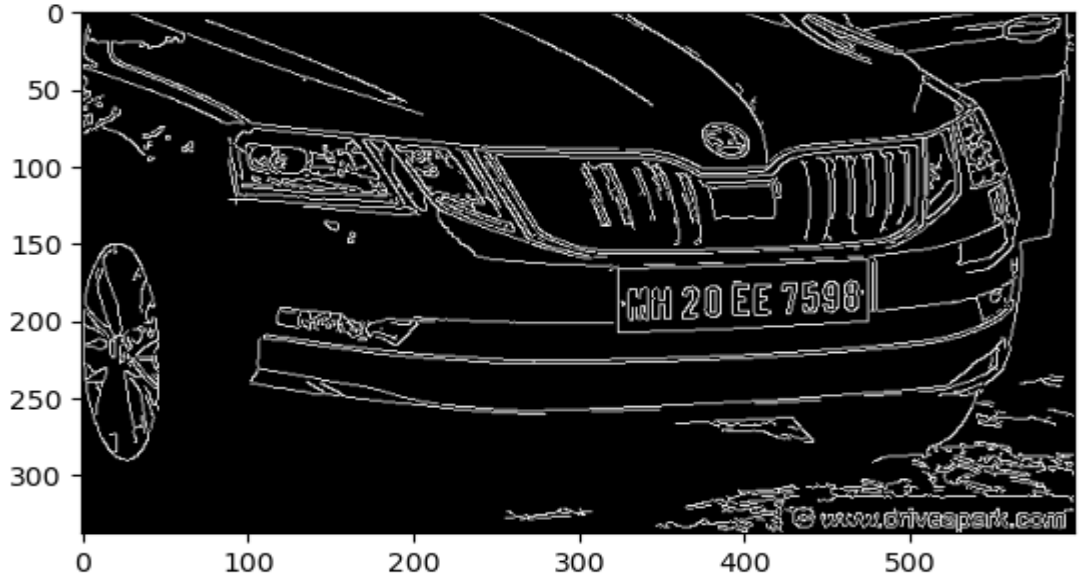

```
text = result[:-1]
font = cv2.FONT_HERSHEY_SIMPLEX
res =  cv2.putText(img, text=text, org=(00,190), fontFace=font, fontScale=1,
color=(0,255,0),thickness=4)
res =  cv2.rectangle(img,tuple(approx[0][0]), tuple(approx[2][0]),(
0,255,0),3)
```

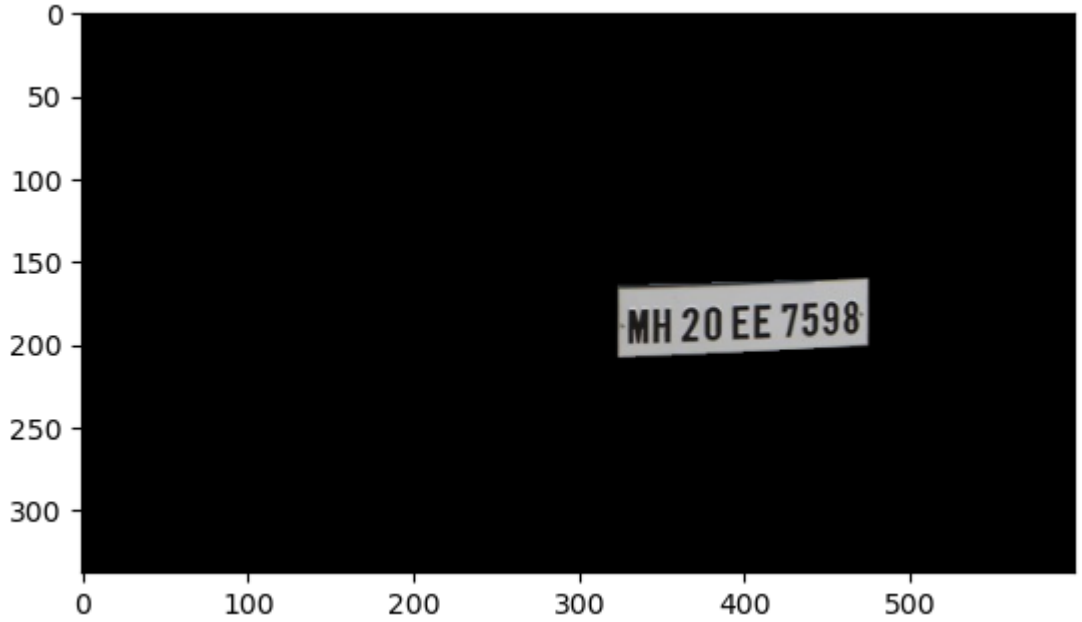
```
plt.imshow(res)
```

# CHAPTER-4
# RESULT AND TESTING

Testing is the process of checking whether the actual results obtained meet the expected results and to ensure that the software system is free from any kind of defects. It involves execution of a software component or system component to evaluate one or more properties of interest. Software testing also helps to identify errors, gaps or missing requirements in contrary to the actual requirements. It can be either done manually or using automated tools.

Testing is an important phase in the development life cycle of the product. Testing performs a very critical role for quality assurance and ensuring the reliability of the software. Each test has a different purpose, all work to verify that all the system elements have been properly integrated and perform allocated functions. The testing process is actually carried out to make sure that the product exactly does the same thing that it is supposed to do. Testing is the final verification and validation activity within the development environment.


MH 20 EE 7598
© www.driveapark.com


MH 20 EE 7598


MH 20 EE 7598


MH 20 EE 7598
MH 20 EE 7598
© www.drivespark.com
0
50
100
150
200
250
300
0
100
200
300
400
500

# CHAPTER-5
# CONCLUSION AND FUTURE WORK

Through this project it is possible to recognise Vehicle registration numbers through digital image processing. From this system we have effortlessly obtained the various results such as

- Whether the vehicle which is registered is blacklisted or not.
- This also enables one single user to effectively monitor the traffic, and can easily locate the traffic violated vehicle.
- The data can be easily stored and transferred which makes the system more efficient.

The system has been designed using a modular approach which allows easy upgrading and/or substituting of various sub-modules thus making it potentially suitable for a large range of vision applications. The performances of the system makes it a valid choice among its competitors especially in those situations when the cost of the application has to be maintained at reasonable levels. Furthermore, the modular architecture makes it extremely flexible and versatile.

The earlier methodologies which have been implemented have not been as accurate and efficient as the designed Recognition system , this is because of the implementation of digital Image Processing which gives an accuracy of 90% under normal conditions

This Project is based on automatic vehicle license plate recognition, in which it is observed that the existing techniques don't pay much attention towards improving the system's efficiency in terms of its power consumption. As the objective in our proposed design is to reduce power consumption of the system, with the successful implementation of the same it will play a very important role in traffic management and security systems such as automobile theft prevention, parking lot management etc. implementations of the software algorithm have shown promising results.
The system can be made more robust if high precision cameras can be used to increase overall accuracy if this system is implemented in real time applications. Also a sensor can be designed to allow the camera to capture the image only when required to save power.

# FUTURE SCOPS

The future scope is that the automatic vehicle recognition system plays a major role in detecting threats to defense Also it can improve the security related to the women’s as they can easily detect the number plate before using cab or other services. The system robustness can be increase if bright and sharp camera is used. Government should take some interest in developing this system as this system is money-saving and eco- friendly, if applied effectively in various areas...

# REFERENCES

https://en.wikipedia.org/wiki/Automatic_number-plate_recognition

Machine Learning for Absolute Beginners: A Plain English Introduction (Third Edition)