Automatic Number Plate Recognition System Using OpenCV And Machine Learning

Basil Baby[1], Jikku Jose[2], Lithin Thomas[3], ShellyShiju George[4]

[1, 2, 3] P.G Scholar of Amal Jyothi College of Engineering, Kanjirappally, Kerala
[4] Asst. Professor of Amal Jyothi College of Engineering Kanjirappally, Kerala

Abstract- Automatic Number Plate Recognition System is a useful application to identify the vehicles from complex security systems, parking sectors and Traffic controls. ANPR is an active research topic in the field of Image Processing using different Algorithms, Methods and Techniques. This paper presents the implementation of ANPR system with the help of Free – Dynamic software Python in support with Open Computer Vision Libraries. The extraction of characters from Number plate is done with Machine Learning, which is a branch of Artificial Intelligence - where Building, Training, Classifying and Predicting data using different Classifiers.

***Keywords*- ANPR; OpenCV; Image Processing; Machine learning; Support Vector Classifier; Character Recognition.**

I. INTRODUCTION

The scientific world is deploying research in intelligent transportation areas which have a significant impact on people's lives. Nowadays, vehicles are increasing rapidly, which makes it complicate to manage all these vehicles. The identification of vehicles are essential in several cases like traffic, gate entrance, etc., [3] for ensuring the security. Every vehicle is uniquely identified on the basis of their number plates. The manual entering of vehicle licence numbers will be a tedious process. This paper discuss the effective method that can be used to reduce the complexity of manual entering and make the process automatic.

The ANPR [4] is a system designed to help in recognition of number plates of vehicles. This system is based on image processing [2]. It helps in the methods like detection of number plates of the vehicles, processing them and using processed data for further functions like storing vehicle details, allowing vehicles to pass or to reject vehicles. ANPR[1] is a mass surveillance method that uses Machine Learning on images to read the licence plates on vehicles. ANPR can be used to store the images that are captured by the cameras as well as the text from the number plate. This systems commonly use infrared (IR) lighting to allow the camera to capture the image at any time of the day. It is an embedded system that has numerous applications and challenges.

Number Plate Recognition Technology is being used all around the world for surveillance applications. This technology is being used to identify and record license plate numbers and offers advantages in law enforcement, security and vehicle access. License Plate Technology is being used for secure and gated entrances, for traffic enforcement, for law enforcement, and at toll gates.

II. LITERATURE REVIEW

There have been several studies on extraction of text from number plate's problem by using various statistical approaches. Through literature review, we only found a few studies which were related to ANPR. But all the contributions that are made in this field are getting more advanced day by day. Some of the information we obtained and discussed are given below:

Miyata et.al [1] proposes the license plate detection process which detects only the edge vertical components, and the sample license plates are narrowed down using the contours obtained by dilation and erosion processing and region fill processing. A Support Vector Machine (SVM) based on negative and positive examples is used to determine whether a sample area is a license plate or not, and finally the position of the license plate is identified. This paper examines how the license plate detection process results in license plate and non-license plate images were affected by differences in aspect ratios, differences in brightness between the vehicle body and license plate, and the number of positive and negative examples used for learning. The success of this

method was confirmed to yield a license plate detection rate of approximately 90%.

Kulkarni et.al [2] proposes the task of recognizing number plate for Indian conditions, in which number plate standards are rarely followed. The system consists of a combination of algorithms that are specifically designed for Indian number plates. The algorithms used are 'Feature-based number plate Localization' for locating the number plate, 'Image Scissoring' for character segmentation and 'statistical feature extraction' for character recognition. The system is capable of recognizing single and double line number plates under widely varying lighting conditions. The system have a success rate of about 82%.

Khare et.al [3] introduces a new concept called partial character reconstruction for segmenting characters of license plates. This increases the performance of license plate recognition systems. Partial character reconstruction is proposed on the basis of stroke width in the Laplacian and gradient domain in a novel way. This results in character components with incomplete shapes. The angular information of character components determined by PCA and the major axis are then studied by considering regular spacing between characters and aspect ratios of character components in a new way for segmenting characters. Then, the same stroke width properties are used for reconstructing the complete shape of each character in the gray domain rather than in the gradient domain, which helps in improving the recognition rate.

Qadri et.al [4] proposes a system which is computationally inexpensive compared to other ANPR systems. The aim of this paper is to model an efficient automatic vehicle identification system by using the vehicle number plates. The system is implemented using Matlab, and its performance is tested on real image. The system have three steps, the first step is the detection and capturing of a vehicle image, the second step is the detection and extraction of number plate from an image, and the third step make use of image segmentation technique to segment the individual character OCR to recognize the individual character with the help of database stored for each and every alphanumeric character. The OCR methods used for the recognition is sensitive to misalignment and different sizes. To improve the OCR recognition from different size and angles, the affine transformation can be used. It is observed that the developed system successfully detects and recognize the vehicle number plate on real images from several experiments.

Merzban et.al [5] proposes a system that use multi-level thresholding of a gray image. This is one of the basic operations in computer vision, with applications in image enhancement and segmentation. One of the various criteria for the selection of threshold level values is the Otsu criterion that uses maximization of between-class variance approach. The computation of the threshold levels with Otsu criterion is a computationally expensive process. The system revisit a dynamic programming algorithm that provides exact and efficient solution to the problem and compare it with modern meta-heuristic algorithms. It provide a rigorous proof for the correctness of the algorithm. The computational cost of algorithm is linear in the number of threshold levels. It compare the algorithm with state of the art algorithms and verify its superior performance. The experiments show that it could gain speedup up to 2.45 x.

Conci et.al [6] introduces with the main objective to show a system that solves the practical problem of car identification for real scenes. All methods, from image scene acquisition to optical character recognition are considered to achieve an automatic identification of number plates. This system can be used for all type of country rules or plates design and adapted to each situation. The system is computationally very efficient and it is suitable for other related image recognition applications.

Selmi et.al [7] proposes an automatic number plate detection and recognition system based on deep learning approach. This system is divided into three parts; detection of number plate, segmentation of characters, and character recognition. To detect a number plate, many pre-treatment steps are made before applying the first CNN model for the classification of plates or non-plates. Then, a few pre-processing steps were applied to segment the number plate and finally to recognize all the characters in upper case format (A-Z) and digits (0-9), using a second CNN model with 37 classes. The performance of the suggested system is tested on two datasets, which contain images under various conditions, such as poor picture quality, image perspective distortion, bright day, night and complex environment. The accuracy of the suggested system shows the great percentage of the results.

III. IMPLEMENTATION

The flow chart of ANPR[1] is as shown in Fig 1. There are various step taken part to identify the Number plate Location and Extraction of characters from that Identified Plates.

A. Creation of TrainingDataset

Machine Learning is a branch of Artificial Intelligence, which deals with various data items, that process and from a pattern that can be used for future prediction. Machine Learning can be categories as Supervised Learning and Unsupervised Learning. Supervised learning make use of known dataset – called Training Dataset – to make the perdition of characters. In this paper, it

will be taking the path of Supervised Learning, since we already have an idea of how As, Bs, Cs and all other characters and letters look like. Supervised Learning can be further divided into 2 categories; Classification and Regression. In this Paper we use Classification category of Supervised Learning because, Character recognition belongs to Classification category.

Creating model for prediction we need to get a Training Dataset, Choose a Supervised Learning Classifier, Train a model and test the model for its accuracy. For Training our model consider the Dataset of 20px by 20px. Note – We need to resize the characters into that size in character Segmentation which will be explained in coming step. Also each letter in the dataset has 10 different Images. For classification there are several classifiers we can use with and each of them having its own advantages and disadvantages. In this paper we use Support Vector classifiers (SVC) as Classifiers for better Performance. However, this does not necessarily mean that SVC is the best classifier.

B. Capturing of Image.

The Vehicle Image is captured using a high resolution Camera. For performing the reaming steps, the characters in the captured image should be in a readable form. **Using a better camera withmore definition and resolution will increase the success ratioof the system.**

C. Pre-process.

Pre-processing of an image involves reducing the cost of computing the image information. The captured I***mages will be either in some multimedia standards or in raw format. Normally, these images will bein RGB mode, with three channels (viz. red, green and blue).Number of channels defines the amount colour informationavailable on the image. The image has to be converted tograyscale.***

D. Localize.

The captured image certainly contains other parts of the vehicle and the Environment, which system doesn't need. The area in the image that interests us is the license plate and needs to be localized from the noise. We use an image processing technique called Thresholding for Localization. The pixels of the image are truncated to two values depending upon the value ofthreshold. For identifying the suitablethreshold value, pre-image analysis is required. Adaptive thresholding technique determines a local optimalthreshold value for each image pixel so as to avoid the problem originating fromnon-uniform illumination.OpenCV provides Otsu method [5] which is a complex and efficient adaptive thresholding algorithm.

E. Connected Component Analysis.

A connectedcomponent algorithm is first applied to the binarized platecandidate, in order to eliminate undesired image areas. Other approaches like edge detection and morphological processing can also be explored. First map all theconnected regions in the binary image and label them. Then draw a rectangle over all the mapped regions. After identifying the Connected Components, the connected components with Rectangular shape and predefined width and height are extracted.

F. Segmentation.

The important phase to facilitate the recognition process is character segmentation [5]. In fact, it consists of extracting the numbers from the Number plate. Several factors make this stage complex, such as the numbering system, the colours, the style (background), the low blur resolution, the noise and the plate rotation. The Connected Component Analysis concept is used here.

G. Character Recognition.

The last stage is Recognising Character from the above Segmented Image.There are several methods used for detection of Characters, one of them is OCR [4]. To improve the accuracy, we use machine learning with SVC model.

We already have a trained model,we can attempt to predict the characters that we earlier segmented.

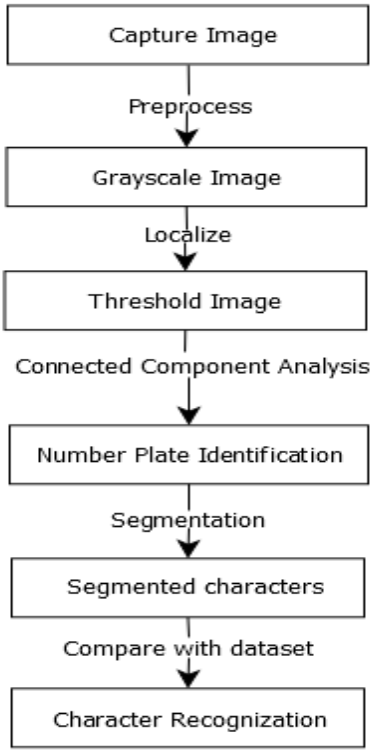

Fig: 1

IV. METHOD OF IMPLEMENTATION

Python is a high-level, powerful dynamic, object-orientedprogramming language that focuses on code readability. While compared to Java or C++, the syntax in Python helps the programmers to do coding in fewer steps. Python supports a strong Integration with other languages and tools. It has a large standard library that has automatic memory management and dynamic features. The main drawback found with the Python is about its speed. It uses Interpreter instead of Complier.

A. Training Dataset.

Learning some properties of a data set and then testing those properties against another data set is known as Machine learning [1]. A common technique used in machine learning is to evaluate an algorithm by splitting a data set into two. We call one of those sets the training set, on which we learn some properties; we call the other set the testing set, on which we test the learned properties. To use SVC we need to install *scikit-learn* package. SVC method will come under *sklearn.svm.* At first load a Dataset, then train and fit the Dataset for predict the classes to which the samples belong to.

```
svc_model.fit(image_data, target_data)
```

B. Capturing of Image

The image can be captured from camera either using Python or other language. Since we can integrate python with other Languages. The following code snippet explainshow to capture a frame from webcam and save it locally for further purpose.

```
camera = cv2.VideoCapture(0)
return_value, image = camera.read()
cv2.imwrite('opencv'+str(i)+'.png', image)
```

C. Pre-process.

Pre-process is done inorder to resize the image and convert the image to grayscale from, so that the image processing with OpenCV can be done quickly.

```
imgBlurred = cv2.GaussianBlur(img, (5,5), 0)
gray =
cv2.cvtColor(imgBlurred,cv2.COLOR_BGR2GRAY
)
```

The above code converts an image to blur and then coverts tograyscale from [7].

D. Localize

Localization is basically a process of binarizing the image. Then, the image is converted to black and white. We use Localization for highlighting characters and Suppressing backgroundprevious researches have concluded thatOtsu's [5] thresholding algorithm is the efficient way ofbinarizing the image [7].

```
ret2,threshold_img =
cv2.threshold(sobelx,127,255,cv2.THRESH_BINAR
Y+
cv2.THRESH_OTSU)
```

E. Connected Component Analysis.

We use findContours to find the rectangle area to detect the number plate position from the image [7].

The below line performs that operation [1].

```
contours, hierarchy=cv2.findContours
(morph_img_threshold,mode=cv2.RETR_EXTER
NAL, method=cv2.CHAIN_APPROX_NONE)
```

F. Character Segmentation.

In this stage we use connected component Analysis to map out all characters from the identified Number plates.

rect_border = patches.Rectangle((x0, y0), x1 - x0, y1 - y0, edgecolor="red",linewidth=2, fill=False)

The above line draws a rectangle line over the characters [4].

resized_char = resize(roi, (20, 20))

We need to resize the characters into 20px by 20px to check with the trained Dataset.

G. Character Recognition.

Character Recognition is the final stage of ANPR. We had already set a trained model, in this stage we attempt to predict the characters that we earlier segmented.

The below code will help to predict the characters from trained Dataset.

Foreach_characterinsegmentation.characters:
each_character=each_character.reshape(1, -1);
result =model.predict(each_character)
classification_result.append(result)

The final output of this stage is the text representing the vehicle number

V. RESULT

A. Accuracy Analysis

Operation	**Sample**	**Success**	**Fail**	**Success Ratio**
License plate localization	100	92	8	92%
Character Separation	92	88	4	95.7%
Character Recognition	88	83	5	94.3%

B. Performance Analysis

The System took more than twenty seconds for recognizing the license plate and extracting the number. One of the key factor that determined performance was the size of input image and its clarity.

VI. FUTURE WORK

Several studies have been conducted onnumber plate detection and recognition. In fact, many researchers have developed various methods and techniques for the solution of this problem. However, all the techniques have their own advantages and disadvantages.

Developing the better image capturing technique will produce a high quality image which will reduce the noise in image. If the image contains noises, develop an efficient method to clean out the noises from the image. Also a convenient method to reduce the image processing time.

The most relevant future work is with characters training and prediction. Efficient training algorithm will produce a efficient model for prediction. Such as using best classifier. Many research is developing in the Machine Learning field and creation of better classification or prediction algorithm will reduce the time latency and gives more accurate result.

VII. CONCLUSION

This paper talks about the techniques how the vehicle Number Plate is identified and recognized using free open source software's. Python and OpenCV is used to perform various image processing techniques. For extracting numbers from the number plate there are various methods such as OCR – *Tesseract*. For improving accuracy rate this paper introduces Machine Learning techniques. But this doesn't mean that it will produces 100 % accurate result. There are several researchers in progress training, Classifying and predicting Data models.

REFERENCES

[1] Shigeharu Miyata, Kenji Oka, "Automated Licence plate Detection Using a Support Vector Machine", 14th International Conference on Control, Automation, Robotics & Vision, 2016.

[2] Prathamesh Kulkarni, Ashish Khatri, Prateek Banga, Kushal Shah, "Automatic Number Plate Recognition (ANPR) System for Indian Conditions", IEEE, 2009

[3]VijetaKhare, PalaiahnakoteShivakumara, Chee Seng Chan, Tong Lu, Liang Kim Meng, Hon Hock Woon, Michael Blumenstein, "A Novel Character Segmentation - Reconstruction Approach for Licence Plate Recognition", Expert Systems with Applications 131, 2019

[4] Muhammed Tahir Qadri, Muhammad Asif, "Automatic Number plate Recognition System for Vehicle Identification Using Optical Character

Recognition", International Conference on Educational Technology & Computer, 2009

[5] Mohamed H. Merzban, Mahmoud Elbayoumi, "Efficient Solution of OTSU Multilevel Image Thresholding: A Comparative Study", Expert Systems with Applications 116, 2019

[6] A. Conci, J. E. R. de Carvalho, T. W. Rauber, "A Complete System for Vehicle Plate Localization, Segmentation and Recognition in Real Life Scene", IEEE Latin America Transactions, Vol. 7, No. 5, pp. 497 – 506, 2009

[7] ZiedSelmi, Mohamed Ben Halima, Adel M. Alimi, "Deep Learning System for Automatic Licence Plate Detection and Recognition", 14th IAPR International Conference on Document Analysis and Recognition, 2017