ROBUST VEHICLE NUMBER PLATE TEXT RECOGNITION AND DATA ANALYSIS USING TESSERACT OCR

Dr.G.Padmaja[1], Swetha Pesaru[2],Mr.Desidi Narsimha Reddy[3] , Dr. D. Anitha Kumari[4], Shiva Prasad Maram [5]

[1]Professor & HOD, CSE department, Sreenidhi University, Ghatkesar, Hyderabad
[2]Assistant Professor, Information Technology, Vignana Bharathi Institute of Technology, Hyderabad
.
[3]Data consultant (Data governance ,Data analytics,EPM Enterprise performance management, AI&ML), SONIKS Consulting LLC,USA
[4]Professor, CSE(AI&ML),TKR college of Engineering and Technology
[5]Technical Trainer , Sreenidhi University, Ghatkesar, Hyderabad

Abstract. To detect the vehicle number plate the system should understand the character and integers determined on vehicles. The proposed methodology holds three phases: pre-process, extraction of features and recognition of text. These phases include some operations like grey scale, adaptive threshold, morphological for extraction of characters and numbers from different quality of the images in pre-processing stage. By transforming the images to grey scale, this can remove the extraneous colours and extracts the appropriate values. In morphological removes the borders and removes the background noises. Adaptive thresholding deals with color around the number plate which can increase the contrast. The outcome was transmitted to feature extraction which helps to identify single characters and numbers which may differentiate among many similar letters and numbers.Text recognition last uses OCR techniques to convert the acquired characteristics into readable alphabetic letters. These methods depend on advanced tools like the Tesseract OCR & OpenCV libraries. Tesseract's OCR features make sure accurate character recognition, and OpenCV improves image processing and computer vision features.

Keywords: Computer vision, preprocessing, adaptive thresholding, tesseract, optical character recognition (OCR)

1 Introduction

The system aims to develop a vehicle number plate detection system that can operate under various circumstances. With the rapid increase in the number of vehicles we need

more advanced systems to counter problems like varying speeds, different lightning conditions, non- standard angles.

The system focuses on detecting and recognizing the alphanumerical characters of the vehicle number plates which includes image preprocessing and gray conversion under various conditions like different angles and speeds. This can be achieved by using open cv, tesseract NumPy libraries.

Open cv is used to manipulate the images. This includes gray conversion, adaptive thresholding and contour identification. Py tesseract from Google tesseract is used to extract the text and present them in a correct order. NumPy is used to provide mathematical functions to work on multidimensional arrays. The System also involves the usage of matplotlib if we wanted to visualize the image along with the output. This can be helpful in reducing the time required for identifying the vehicles and is also used to identify the state that it belongs to.

2 LITERATURE SURVEY

There are various systems that are used to detect and recognize the characters of vehicle number plates. In this, we use a simple three stage strategy.

Vanshika Rai and Deepali Kamthania [1] made an attempt to develop a system which includes Guassian smoothing and Guassian thresholding and morphological transformation techniques has been used in preprocessing stage. The proposed system uses K-nearest neighbour algorithm for character recognition. Contours are filtered based on character recognizing and other localizations for segmentation.

Shashidhar R, A S Manjunath, Santhosh Kumar R, Roopa M and Puneeth S B [2] proposed a system which uses You Only Look Once (YOLO) V3 model is used for region of interest (ROI) and Convolution Neural Networks (CNN) is used for optical character recognition. Accuracy of 95.1 percent is obtained for dataset they created containing 6539 images of different Indian number plates with different alpha numerical characters.

Aman Jain, Jatin Gupta, Somya Khandelwal, Surinder Kaur [3] proposed a vehicle license plate detection which uses You Only Look Once (YOLO)- Py Torch deep learning architecture. In this several image processing techniques are used to convert digital information to usable information. This is mainly done in three stages which involves License Plate detection, character segmentation and character recognition. It mainly focuses on detection phase in order to avoid the failures in subsequent phases.

Ravi Kiran Varma P, Srikanth Ganta, Hari Krishna B, Praveen Svsrk [4] explained the limitations of existing automatic number plate detection systems like they are based on the UK number plates which may not be suitable for Indian number plates and they detect the plates which are in straight angle with good clarity only. They proposed a system which can be used to detect the number plates at different angles and different light illuminations. The system uses KNN algorithm for character recognition and also uses contours are filtered based on the character dimensions.

M. A. Jawale, P. William, A. B. Pawar, Nikhil Marriwala [5] explained the implementation of automatic number plate detection for vehicle registration using IOT and recognition using CNN. The proposed system contains four main steps which involves License Plate Extraction, Image preprocessing, Character segmentation, Character recognition. Many unique methods have been proposed for the first three

stages but for character recognition they use four different methods like Convolution Neural Network (CNN), Mobile Net, Inception V3, Res Net 50.

Nur A Aalam Munna, Mominul Ahsan, Md Abdul Based, Julfikar Haider [6] proposed a vehicle detecting system using Convolution Neural Networks (CNN). It mainly involves two stages. They are number plate detection and number plate recognition. Segmentation of each character using bounding box method and CNN method is used for feature extraction and classification.

Anugu Hansika, Thaviti Sony, Sohaib Ibrahim [7] explained the need of reliable and robust vehicle number plate detections in real world scenarios. They proposed a system which can detect the vehicle number in different light conditions, variable speed and different angles. They aimed to prepare a system which can be adaptable to different environment conditions and other factors as well.

3 EXISTING SYSTEM

There are many existing technologies which can be used to detect vehicle number plate. But with the increase in technology in the present these existing systems are failing to detect the number plate accurately because of various conditions like speed, varying light illuminations. The performance and accuracy of the existing system is gradually going down because of its failure in detecting the number plates in various conditions. It can only work when the number plate is given in full bright light. The conditions in which the existing system is failing to detect the number plate are

Lightning variations: When it is required to detect the number plates in dim lights or highly illuminated environments, the existing systems doesn't provide accurate detections or they miss few characters.

Variation in angles: Existing system cannot detect the number plates accurately in different angles when it is required the most. They cannot detect when the number plate is in non-standard angles.

Speed: Precision can be affected when vehicles move at high speed. System can not detect accurately which proves its inefficiency at varying speeds.

Adaptability: A system should be developed which can adapt itself to different environmental conditions like different light conditions, variation in speed, able to detect in non-standard angles.

4 PROPOSED SYSTEM

The main aim is to propose a system which deals with all situations like varying light conditions, different angles like cross angles, non-standard angles, variation in speeds and its adaptability to different environmental conditions. Proposed system mainly performs 3 operations. They are Image preprocessing, feature extraction and text extraction.

Preprocessing: Raw images taken from multiple environmental conditions including low illumination, motion blur, & weather interference are pre-processed to boost plate & character recognition performance Preprocessing takes image as an input. The main aim of preprocessing is to reduce the complexity of the image and at the end of this stage all the essential features are retained and will be efficient to detect the characters. Dynamic lighting, shadows, and headlights may all have an influence on image clarity,

therefore brightness and contrast are modified to optimise plate visibility. This guarantees that the characters on the plate are unique from the backdrop. In this stage there are other operations like gray conversion, adaptive thresholding and noise removal techniques. By the end of this stage the image will be efficient to detect the characters appropriately. This stage was essential for identification of number and letters with help of scalling the images in different quality or backgrounds.

Feature Extraction: The data provided from pre-process has transmitted to feature extraction for the unique identification. By avoiding background in the image it's necessary to extract numbers and letters with unique patterns. As pre-processing and edge recognition phases has retrieved shapes, edges or pixel patterns. Extraction of features includes some operations like morphological and edge detection to retrieve meaningful number plate. Where morphological is essential for extracting image color textures. Edge detection holds the frame edges and boundaries with approproiate pixel level. Now OCR is utilized for reading and understand the accurate pattern. Finally the purposes of this phase will return the approporiate number plate. Therefore, cuts down on noise and improves the accuracy of the system. Effective feature extraction is ultimately the foundation of dependable car number identification since it enables ANPR systems to differentiate between characters that seem same, even in difficult situations.

Feature extraction is crucial step in computer vision as it is used to convert complex data into meaningful transformations which can further be used for recognition of patterns and visualize the information.

Text Extraction: Text Extraction is the identification and separation of regions in an image most likely containing text, like the alphabetic letters on a number plate. Text detection systems find and highlight the area including the license plate after pre-processing & edge detection. Commonly utilised techniques are CCA and MSER as they can efficiently find groups of pixels that most certainly reflect characters depending on shape & intensity. The technology segments the characters once it finds the license plate area. Since license plates vary in size, text, and spacing and images may be impacted by angle, motion blur, or illumination changes, this stage is very crucial. This reduces processing time and improves character identification accuracy by precisely recognising the text area and passing only relevant image components to the future steps. Text extraction from number plates is done by the PyTesseract, a python library that integrates with the Tesseract OCR machine. PyTesseract is used to extract the alphanumerical characters and other appropriate information required for identification. The integration of PyTesseract with Tesseract OCR proved to be reliable in number plate detection.

4.1 Read Image: This step(Fig 1) is used to examine the images and perform operations like object detection which can further be used for computer vision. This can be done with the help of OpenCV which is efficient enough to visualize the data and reading images.

4.2 Gray Scaling: Gray scaling(from fig 1) is used to convert colored image to gray image. Gray scale is used to reduce the complexity of images making it more efficient for

identification and visualizing the data. Gray scaling helps in representing the images in a standardized manner which can be used gain more accurate and better results. Gray scaling can solve the problem of inconsistent and varying lightning conditions. In a greyscale image, each pixel is a level of brightness between black and white. In an 8-bit image, this level of brightness goes from 0 (black) to 255 (white). Ideal for uses including object recognition, edge detection, & image analysis, this reduction lowers computing complexity by expressing every pixel with one value instead of three. Grayscaling is a typical pre-processing step in vehicle number identification systems as it emphasises number plate character intensity contrasts with background. This increase in contrast makes it easier for later processing steps, such as thresholding or edge recognition, to work better. Grayscaling serves to simplify the identification process by eliminating colour distractions and concentrating only on intensity, therefore enabling algorithms to more precisely and effectively identify features.

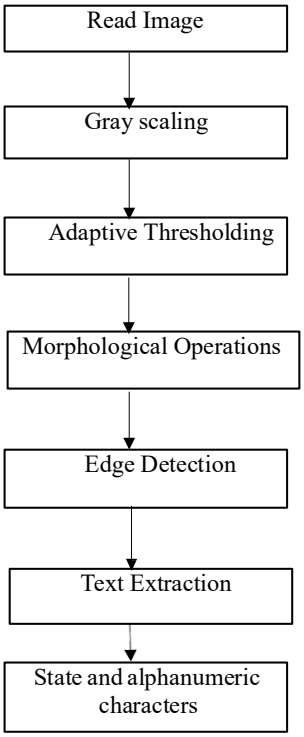

Fig 1- Flowchart that depicts the methodology.

4.3 Adaptive Thresholding: This method(from fig 1) is used to separate objects from background. Converting in to binary format will help in avoiding background objects and also reaching the region of interest. Region of interest will help in ignoring the noise, background objects and all other irrelevant information. Adaptive thresholding computes distinct threshold values for various areas, in contrast to global thresholding, which applies a single threshold value across the image. Light adjustment and shadows common in images which has to be adjusted with dynamic thresholds. For this adaptive threshold was efficient which divide the image nto small tiny parts. To calculate the threshold, it depends on the weight or mean of Gaussian which can be derived by pixel rates. By this the similar pixels are matched and computed as a single threshold. To the similar values or mid point the above the margin are considred as white and below as black. Moreover, this value retrieves the essential feature of the image even in different light situations. Adaptive thresholding is very useful in vehicle number identification in segmenting the number plate text from the backdrop so that letters are discernible even with shadows or uneven illumination. It focuses on the relevant information which in turn improve the performance and accuracy of all the next further steps. Adaptive Thresholding a vital role computer vision application.

4.4 Morphological Operations: Morphological operations are a class of image processing algorithms that examine and alter the shape & structure of the objects in an image. These operations(from fig 1) are used to manipulate the images. These are often employed as pre-processing stages for computer vision applications such as OCR, ANPR, & barcode recognition. Here are some morphological operations: Opening, Top-hat Transformation, Skeletonization, and Thinning. They remove the noise and improve the quality of the number plates. These operations are done either by expanding the boundaries or by shrinking the boundaries depending on the number plate to be detected.

4.5 Edge Detection: Edge detection is an important part of car number detection because it helps figure out where the number plate ends and the characters on it begin. Within this process, computers find edges where there are big differences in the intensities of the pixels that are next to each other. Usually, these edges are where the license plate's borders are different from the car or where the characters on the plate stand out from the backdrop. These boundaries are highlighted by the use of popular edge detection methods such as the Sobel, Canny, & Laplacian operators. In car number recognition, edge detection helps separate the number plate's oblong shape and gets the image ready for character segmentation. A popular edge detector is the Canny, which accurately finds strong edges while reducing noise. Edge recognition cuts down on the image to just the most important lines and shapes, making it easier to work with in later steps. This makes it easier to find the license plate, even if the image is taken at an angle, in different lighting, or with other things in the background. This (from fig1) is used to detect the boundaries of image. By analyzing these edges, we can decide shape and location of number plates. Once the useful information is retained using thresholding operations further transformations like Hough can be used to work on the number plate further to detect the number plate more accurately.

4.6 Text Extraction: After performing all the above operations efficiently text extraction is done in which all the alphanumerical characters are extracted.

4.7 Output: After extracting the text we get the state that the vehicle belongs to and the alphanumerical characters in the number plate.

5 RESULTS AND DISCUSSION

5.1. Dataset: The Kaggle dataset named "Car Number Plate Detection" by users elysian0 is intended to train models to identify and recognise car number plates. This sort of information is very valuable for automated number plate recognition (ANPR) systems, that provide a broad range of applications including traffic control, toll collecting, and security. The dataset includes a collection of photos of automobiles with visible number plates. These photographs may comprise a variety of viewpoints, lighting situations, and settings, making the dataset ideal for model training that must generalise effectively.

https://www.kaggle.com/datasets/elysian0 1/car-number-plate-detection.

*5.2. Experimental Results:*The experiments are done using OpenCv to implement the image processing techniques. The accuracy was evaluated on tested data where the images are taken with different angles and lights conditions.

Table 1 – Accuracy of different I/O.

Image	Conditio ns	Output	Accuracy
	Slightly blurred	Delhi “NL3C AM0857	90%
	Straight angled	Maharasht ra MH12DE14 33	100%
	In motion	24 BZ 0768	70%

6 CONCLUSION:

Number plate prediction was essential in different fraud aspects. The proposed methodology includes three phases pre-processing, extraction of features and text detection. Initial stage involves pre-processing with some operations like grey conversion, adaptive thresholds, and morphological. These techniques retrieve's the essential data to feature extraction stage for identification of characters and numbers. This phase retrieve features and form as unique patterns to form a meaningful text. This data was transmitted to text detection phase with several angles and light environment. This was extracted using Tesseract OCR engine & OpenCV libraries. This technique can involve many dependability uses like automated tooling, law, and monitoring traffic. This application can work in real-time.

7 FUTURE SCOPE:

Required advanced methodologies for the detection of number plate and to prediction fast in different light environments.

References

1. Vanshika Rai and Deepali Kamthania Automatic Number Plate Recognition is published on July 3 2021 which can be used to detect the stolen vehicles and e- challan surveillance.
2. Shashidhar R, A S Manjunath, Santhosh Kumar R, Roopa M and Puneeth S B Vehicle Number Plate Detection and Recognition using YOLO-V3 and OCR Method. DOI:10.1109/ICMNWC525 12.2021.9688407. Conference: 2021 IEEEInternational Conference on Mobile Networks and Wireless Communications (ICMNWC).
3. Aman Jain, Jatin Gupta, Somya Khandelwal, Surinder Kaur Vehicle License Plate Recognition June 2021 4(1):15-21.DOI: 10.5281/zenodo.5171216
4. Ravi Kiran Varma P, Srikanth Ganta, Hari Krishna B, Praveen Svsrk prepared a novel on various image processing techniques in vehicle number plate detection systems. It is available at https://doi.org/10.1016/j.procs.2020.03.32 4.
5. M. A. Jawale, P. William, A. B. Pawar, Nikhil Marriwala explained the implementation of number plate detection for vehicle registration using IOT and recognition using CNN published on May 1 2023 available at https://doi.org/10.1016/j.measen.2023.10 0761.
6. Nur A Aalam Munna, Mominul Ahsan, Md Abdul Based, Julfikar Haider proposed a sytem foe vehicle number plate detection and recognition using YOLO-V3 and OCR Method published on December 2021. IEEE International Conference on Mobile Networks and Communications(ICMNWC) at Karnataka. Volume: 978-0-7381-4637- 9/21/$31.00.
7. Anugu Hansika, Thaviti Sony, Sohaib Ibrahim Vehicle number plate detection and recognition under different environmental conditions.