Automatic Vehicle Number Plate Recognition Approach Using Color Detection Technique

Muhammad Ayaz¹, Dr. Said Khalid Shah¹, Dr. Muhammad Javed¹, Muhammad Assam², Wasiat Khan³, Fahad Najeeb¹

¹Department of Computer Science University of Science and Technology Bannu, Pakistan.
²College of Computer Science and Technology Hangzhou 310027, China.
³Department of Software Engineering, University of Science and Technology, Bannu, KP Pakistan

*Muhammad Ayaz Email: ayazkhan1ttt@gmail.com

Received | Dec 07, 2021; Revised | Dec 28, 2021 Accepted | Dec 31, 2021; Published | 8 Jan, 2022.

An Automatic Vehicle Number Plate Recognition System (AVNPR) is a key research area in image processing. Various techniques are developed and tested by researchers to improve the detection and recognition rate of AVNPR system but faced problems due to issues such as variation in format, lighting conditions, scales, and colors of number plates in different countries or states or even provinces of a country. Douglas Peucker Algorithm for shape approximation has been used in this research to detect the rectangular contours and the most prominent rectangular contour is extracted as a number plate (NP) and the connected component analysis is used to segment the characters followed by optical character recognition (OCR) to recognize the number plate characters. A custom dataset of 210 vehicle images with different colors at various distances and lighting conditions was used for the proposed method captured on my smart phone Galaxy J7 Model SM-j700F at roads and parking. The dataset contains various types of vehicles (i.e. Trucks, motorcars, mini-buses, tractors, pick-ups etc). The proposed method shows an average result of 95.5%. The novelty used in this method is that it works for different colors simultaneously because in Pakistan, several colors are used for vehicle NPs.

**Keywords:** Number plate recognition, color detection, HSV color, Douglas Peucker Algorithm, OCR

**INTRODUCTION**

The number plate recognition system is a crucial factor of an intelligent transportation system. In this system, vehicles are recognized based on their number plates. It has a variety of applications, including a border management system, automatic car parking system, automatic toll collection on highways and enforcement of traffic laws [1, 2, 3]. It is also used to detect stolen vehicles, fake NPs holders, and non-standard number plates. An intelligent gate management system is also based on this criterion to restrict the entry of unauthorized
vehicles into an organization [4, 5]. It has wide applications to recognize high-valued vehicles (i.e., government officials, police vans, school buses, ambulances, etc.) to provide a safe way to them in bulk of vehicles and can be used to manage one-way roads and restrict heavy trucks to enter cities and towns [2].

In general, the number plate recognition system consists of four steps [6, 7]. In the first step, the input image is preprocessed for noise removal and foreground enhancement. In the second step, the region of interest (ROI) is detected and extracted. In the third step, the characters are segmented, and in the fourth step, recognition is done. Researchers adopt different approaches and each approach has its pros and cons. The majority of systems are limited to some specific formats and territories. There are two main approaches used to recognize the number plates; traditional image processing approach and the deep learning approach. In the traditional image processing approach, simple mathematical functions are used to locate and extract the number plate and character segmentation is used to save the characters as separate images. Template matching is used to recognize the characters and convert them into text form.

In the deep learning approach, a supervised learning strategy is used. An object detection model is developed and trained using a relevant dataset to recognize the desired object in a video or image. Nowadays deep learning-based models are used widely for the number plate detection and recognition with high accuracy. Hundreds and thousands of image based datasets are used with proper annotation to train the model and it is not an easy task that need high power systems with GPUs to work smoothly. A model trained for one format may not be compatible for other formats of NPs. On the other hand, in traditional image processing approach, the localization/detection ratio of NPs does not work well in natural scene images and videos having noise and bad illumination and it is the main step of AVNPR system and the remaining steps are highly dependent on this step [8][9].

This section contains different methods obtained from literature survey and are used for number plate detection and recognition. Pavani and Mohan [1] used simple image processing techniques and connected component analysis (CCA) was used for plate extraction and again CCA was used to segment the characters. Sharma et al. [2] used a Sobel edge detector to detect the foreground edges and thresholding technique was used for image binarization to locate the number plates. Morphological operations were used to enhance the plate area for extraction and minimization of the non-plate locations. Template matching was used to recognize the number plate characters. This system was developed for the specific format of Nepali number plates. Yogheedha et al. [8] used Otsu's thresholding and background subtraction to locate the number plate and Built-in OCR was used to recognize the characters. Out of 14 images, the system recognized 13 images of the car. It was developed as a training project for Perlis University of Malaysia's intelligent gate management system to restrict the entry of private vehicles. Kulkarni et al. [9] developed a model for helmet detection of motorcycle riders and also number plate recognition system was used to find those riders having no helmet. They used thresholding and morphological operations for plate detection and OCR was used to recognize the plate. Wang et al. [10] used the color detection method to detect the red lights and then the middle of the lights was extracted as the region of interest. It worked for the rear NP of a vehicle having red color backlights. Babbar et al. [11] used preprocessing techniques
and binarization to filter out the strong objects. CCA was used to locate NP and aspect ratio to extract the NP successfully and OCR was used for NP recognition.

Rajathilagam et al. [12] used traditional image processing techniques to develop the system. The input image was converted to grayscale and dynamic adaptive thresholding was used for NP localization. Morphological operation dilation was used to highlight the region of interest and erosion was used to eliminate the false NP areas. OCR was used to recognize the plate characters. That method worked only for very clear images which are taken at a very short distance. Indira et al. [13] used Euclidian distance to locate the possible NP region with the predefined threshold value. The stroke width was used with maximally stable extreme regions (MSER) to detect and extract the NP and template matching was used to recognize the plates. Haider and Khuram [14] used simple morphological operations to detect the number plate. After morphological operations and image filtering, CCA was used to extract the number plate and feature matching was used for character recognition. Hendry and Chen [15] used the sliding window technique with a deep learning-based object detection model “You Only Look Once (YOLO)” to detect and recognize NPs of Taiwan motor car. They used AOLP dataset for training the model having 2049 images of Taiwan LPs. Babu et al. [16] used a custom data set of 6500 car NPs images with annotation and trained Darknet framework YOLO v3 for NPs detection [17, 18]. The same model was trained again for character recognition of plates with 91% accuracy. Pustokhina et al. [19] used the improved Bernsen method to enhance the image illumination. Connected Component Analysis was used to detect the number plates. Optimal k-means clustering was used to segment the characters and Convolutional Neural Network was used to recognize the characters. Saboo et al. [20] used preprocessing techniques to enhance the image quality and the prewitt edge detector was used to detect the vertical edges. Thresholding technique was used to segment the images to remove the weak blobs. Morphological operations dilation and erosion were used to locate the vehicle number plate. Aspect ratio was used to extract the rectangular blob as expected number plate region. The extracted image (NP) was enhanced using histogram equalization to improve the shapes of characters and CCA was used to segment the characters. K-nearest neighbor and distance tree classifier were used to recognize the characters. Tenzin et al. [21] used preprocessing for image enhancement and Robert’s edge vertical detection with a fixed value of thresholding (0.0375) to filter the image for NP and the rectangular shape was extracted as NP. The bounding box method was used to segment the characters and template matching was used to recognize the plate characters. Tourani et al. [16] used histogram equalization for image enhancement and trained Darkent framework YOLOv3 for NP detection and again the same model was trained for character recognition. A custom dataset of Iranian vehicles with proper annotation was used for training the model. Pustokhina et al., [19] developed a model for mobile application using Tinny YOLO for NP detection and again convolutional neural network (CNN) was used for character recognition. Saif et al. [20] developed and trained a model for Bangladeshi standard NPs format using YOLOv3 for NP detection using custom dataset and again Darknet framework YOLOv3 was trained with a dataset of characters to recognize the plate characters. Henry et al. [21] used Tiny YOLOv3 for NP detection and direct recognition without segmentation was done using YOLOv3-SPP (Spatial Pyramid Pooling) model. Quadri and Asif [22] used a color-based technique to extract the NPs having a yellow color. After preprocessing, the input image was searched for yellow pixels and set to 1 and other than
yellow pixels were set 0 to extract the NP region easily. OCR was used to recognize the characters. Khan et al. [23] used LAB color space transformation and L-channel was selected and contours were drawn. Otsu’s thresholding was used after preprocessing to segment the ROI. Morphological operation erosion and dilation were used and eroded image was subtracted from dilated image to highlight NP’s characters and 2d convolution was used to extract the NP region. Support Vector Machine was used to recognize the characters.

M Rizwan Asif et al. [24] used traditional image processing techniques to detect and recognize vehicle number plates using blue color properties. Image was converted from RGB to YDbDr color space to detect the blue color and intensity adjustment was used to eliminate low intensity objects. Thresholding was used for image segmentation and CCA was used to detect the candidate regions and aspect ratio was used to detect the expected plate among the list of ROIs. [25] Converted RGB image to HSI format for color segmentation. Morphological operations were used to highlight the ROI. Connected component analysis was used with aspect ratio to filter the rectangular shape (expected LP) among the existing blobs. F. S. Khan et al. [26] combined shape and color properties based on late fusion to detect objects. [27] Used two different approaches and comparatively studied the results. Initially, they used color-based approach to detect the LP of vehicles and then used the shape-based LP detection method.

In this research, a color-based NP detection technique with shape approximation is used. The acquired image is converted to HSV color format to filter the image for NP localization and extraction. In the previous literature, color base techniques were utilized for single color [28]. The novelty used in this method is that it works for different colors simultaneously because in Pakistan several colors are used for vehicle NPs. After detecting the NP, we proceed with the recognition step.

PROPOSED METHOD

In this study, the proposed method was discussed in detail Figure 1. showed the flow diagram of the proposed system.

![Figure 1. General steps of the proposed method](image-url)
In this research, traditional image processing techniques were used to detect and extract the number plate region. Connected component analysis was used to segment the characters and built-in OCR was used to recognize the characters. The input image was converted to HSV color format and the mask was generated. Binarization and morphological operations were used to locate the strong blobs and Ramer Dauglous Peauceler shape approximation algorithm was used with aspect ratio to detect the rectangular shapes and the most prominent rectangular shape was obtained as NP area. The extracted NP image was enhanced using pre-processing operations and connected component analysis was used to segment the plate characters and built-in OCR was used to recognize the plate. Three primary steps i.e. extraction, segmentation and recognition were elaborated in detail as below.

**Number Plate Detection and Extraction**

The input image was resized and pre-processing techniques were used to enhance the quality of the input image. It was then converted from RGB color format to HSV color format to easily filter the desired color intensity using a NumPy array of lower and upper bound values. Separate lower and upper bound values were used for each NP's color detection because each color had a separate range of values. The mask image was generated as shown in Figure 2.

![Figure 2](image)

Figure 2. (a) Input image with yellow plate, (b) Mask image, (c) Input Image with green Plate and (d) Mask image

The image was converted to grayscale and thresholding was used to eliminate the non-plate regions because in many cases, there is a probability of existence of some non-plate regions. As the thresholding operation was used to remove the non-plate regions but it also erases some portion of ROI called outlines of number plate and it was difficult to take it as a single contour region. Morphological operation gradient was used to strengthen the outlines of the region of interest (ROI) that was used to highlight the object to contour the ROI easily. Edge detection method was used to find the edges and contour function was used to find all the contours.

*Shape Approximation*: In most of the cases the largest contour contains vehicle number plate but the problem arises when the input image contains a non-plate blob/area which is
larger than the NP area with color intensities similar to the region of interest [29, 30]. To eliminate this problem Ramer Dauflous Peacuker shape approximation algorithm was used. Before shape approximation method, the internal region of each contour is filled completely to approximate the shape accurately. Without filling the blob, shape approximation function approximates each NP character as separate contour.

**Ramer Dauflous Peacuker algorithm** developed in 1973 by David H. Douglas and Thomas K. Peucker was widely used for geometrical shape approximation [31, 32]. This algorithm converts a curve/contour composed of a linear line segment to a similar curve by reducing number of points. The points on the converted curve are nothing but subset of the points that defines the original curve. The algorithm treats each curve/contour C as an order set of points and the distance dimension is $\varepsilon > 0$.

$$C = (p_1, p_2, \ldots, p_n)$$

First of all the algorithm automatically marks the first and the last points to be kept. The line segment obtained from $p_1$ to $p_n$ by joining the kept points is the first approximation of the curve $C$. The algorithm then searches the point in the inner points that has largest distance from the line segment $p_1$ to $p_n$. If the distance of the largest point is smaller than $\varepsilon$, then simplification is been done and all the inner points can be put away without the simplification curve being worse than $\varepsilon$. If the distance of the largest distance point is greater than $\varepsilon$ then the largest distance point must be kept. The algorithm repeats itself with the first point and largest distance point and with the largest distance point and end point. When the repetition is complete, the output curve represents the set of all kept point called shape approximation [33].

**Character Segmentation**

After successful detection and extraction of the number plate region, the next step is character segmentation. This step is mainly used to prepare the extracted plate image as an input to the recognition step to easily recognize the characters. In this step, the input image was resized and converted to the grayscale image. Thresholding was used for image binarization. Connected Component Analysis was used to declare each area composed of similar and connected pixels as an individual object. Width, height and area properties were used for each connected component to filter out the objects that are too small or too large.

**Character Recognition**

This step was used to read the plate image and convert the characters and digits into readable form called character recognition. There are different approaches used for this purpose. In this system a built-in OCR based on deep learning called Pytesseract was used for character recognition [34, 35].

**Results and Discussion**

The proposed method is based on color intensities with shape approximation. Due to the prominent colors used in our country for NPs it becomes easy to detect the ROI easily using color based segmentation and in more than 80% cases, the plates using their color intensities are easily distinguished shown in Figure 3.
Figure 3. Shows some sample images after finding contours

To increase the accuracy rate we used shape approximation in the second phase to overcome the problems that shown in Figure 4.

Figure 4. Shows largest contour as non-plate area

There are three main colors of NPs in Pakistan used in different territories/provinces for the last decade. The green color is used for government vehicles. For public vehicles, the yellow color is used in Khyber Pakhtunkhwa and Sindh provinces; green-yellow is used in Punjab and Baluchistan provinces. Sample images of Number Plates used in our country are shown in Figure 5.

Figure 5. Sample Images of NP

A custom dataset of 210 vehicle images with different colors at various distances and lighting conditions was used for the proposed method captured on my smart phone Galaxy J7 Model.
SM-j700F at roads and parking. The dataset contains various types of vehicles (i.e. Trucks, motorcars, mini-buses, tractors, pick-ups etc). The proposed method shows an average result of 95.5%. Table 1 shows the result of extraction and recognition of all the four types of NPs.

Table 1. Shows the results of the proposed system

<table>
<thead>
<tr>
<th>Type</th>
<th>Item</th>
<th>Extraction</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow</td>
<td>90</td>
<td>88</td>
<td>83</td>
</tr>
<tr>
<td>Green</td>
<td>50</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>Green_Yellow</td>
<td>70</td>
<td>67</td>
<td>65</td>
</tr>
<tr>
<td>Total</td>
<td>210</td>
<td>201</td>
<td>191</td>
</tr>
</tbody>
</table>

There is no public dataset of Pakistani vehicle NPs as per the knowledge obtained by the survey and efforts were also made for comparative study. The dataset used for comparisons with some state-of-the-art methods are shown in Table 2.

Table 2. Comparisons

<table>
<thead>
<tr>
<th>S#</th>
<th>Methods</th>
<th>Year</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Qadri and Asif [28]</td>
<td>2009</td>
<td>74 %</td>
</tr>
<tr>
<td>2</td>
<td>Haider and Khurshid [10]</td>
<td>2017</td>
<td>85 %</td>
</tr>
<tr>
<td>3</td>
<td>K. Deb et al. [31]</td>
<td>2017</td>
<td>91 %</td>
</tr>
<tr>
<td>4</td>
<td>Babar et al. [7]</td>
<td>2018</td>
<td>87 %</td>
</tr>
<tr>
<td>5</td>
<td>Pustkkhina et al. [14]</td>
<td>2020</td>
<td>93 %</td>
</tr>
<tr>
<td>6</td>
<td>Proposed</td>
<td>2021</td>
<td>95.5 %</td>
</tr>
</tbody>
</table>

CONCLUSION AND FUTURE WORK

There is no public dataset of Pakistani vehicle NPs as per the knowledge obtained by the survey and efforts were also made for comparative study. The number plate recognition system is not used in Pakistan currently due to various factors. Still this research article proposed an efficient method for the development on an automatic number plate recognition system in Pakistan. Despite being trained on a small dataset, it can show better and accurate results comparable to other number plate’s recognition systems. The proposed method may provide a base for the implementation of an automatic number plate recognition system in Pakistan. In the proposed research article, simple image processing techniques are used to first extract the NP, and then OCR is used to recognize the plate. The extraction rate of the plate may be increased by imposing rules on the drivers to keep the number plates safe from dust and scratches.

Acknowledgement

The Author says thanks to Mr. Waheed Ullah Post Graduate Student at department of Mathematics University of Science and Technology Bannu for his valuable suggestions and support regarding the implementation of Ramer Dauglous Peacuker Algorithm.

Conflict of Interest

It is declared by the authors that this manuscript has no conflict of interest.
REFERENCES


Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, pp. 1-5, 2017. Article (CrossRef Link)


Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.