Cognitive Number Plate Recognition using Machine Learning and Data Visualization Techniques

Rahul Agrawal[1] Manas Agarwal[2] Rajalakshmi Krishnamurthi[3]
[1,2,3,] *Department of Computer Science & Engineering*
Jaypee Institute of Information Technology
Noida-62, Uttar Pradesh, India
[1]rahul17a1997@gmail.com [2]manasagarwal.m@gmail.com [3]k.rajalakshmi@jiit.ac.in

Abstract— The conventional Automatic Number Plate Recognition (ANPR) is based on image processing mechanism for automatic vehicle authentication using number plate of a vehicle. The applications of number plate recognition include identification and prevention of vehicular crime, traffic management and handling road challan process. However, APNR exhibits incompetent in case of poor plate localization, improper plate sizing and plate disorientation. Moreover, the ANPR algorithms such Optical character recognition, geometric analysis and character segmentation mechanism less efficient in adverse conditions such as illumination flare on the number plate. Hence, this work presents Cognitive Number Plate Recognition (CNPR) system based on Machine Learning techniques and Data Visualization methods. The proposed system performs knowledge generation through data clustering mechanism. Further, the hidden information pattern within the database of detected number plates is used to provide insight into the vehicle information towards decision making and analysis. The experimentation results of the proposed system show 85.3% of from plate recognition, 90.5% for rear plate, 83.2% of localization and 80.5 % of character segmentation and 73.4 % of character recognition.

Keywords — ***Number Plate Recognition, Machine Learning, Segmentation, Localization, Data Visualization***

I. INTRODUCTION

With the ever increasing vehicles on road, implementing vehicles laws manually is not easy task anymore. This is applicable to toll-booths that are situated in highways and parking lots in mall etc., where the vehicle has to stop to pay the toll or the parking fees. Also, in the systems that are used to penalize the vehicles travelling above the allowed speed. Thus, there is wide scope of enhancement available in this aspect. Therefore, to ease the processing of the same and make them more practical to use, a system is needed to acknowledge a vehicle, and then search a query by using vehicle number plate.

Vehicles in every country have a novel identification number allotted by the RTO/RTA, written on its registration number plate. That provides a uniqueness which differentiates one vehicle from the other that is beneficial particularly when comparison is between same build and model. To overcome this, a system that can execute automatic categorization of the unique number plate of each vehicle is essentials. This system will extract each character within the specific area of registered number plate. The registration plate number might be accustomed to retrieve a lot of information regarding the vehicle and its owner, which may be used for additional processes.

Automatic Number Plate Recognition systems are currently in use in many countries like Australia, Korea, USA to name a few. This is possible due to strict implementation of number plate laws in these countries. These countries have a set of defined rules for standard features of the number plates such as: dimensions of plate, border for the plate, color and font of characters, etc., which are in turn used as parameters to localize the number plate and locate the number plate in image, enabling them to recognize the number of the vehicle much easily [1].

Various number plate detection algorithms have been developed in past years. Median filter is used for reducing noise in the image, though it is able remove noise, but this is very time expensive and complex to compute [2]. Also, Canny Edge Detection has been used to locate number plate, but it has to be given an initial point to look around, and is also time consuming [3]. This project streamlines the strategy within which number plate is perceived using confidence related predictions. Additionally, methods are used to limit the search region to bounded areas in a picture. Thus, to overcome, this paper, proposes the use of binarization and elimination of extra regions from a picture. Moreover, simple and effective mechanism such as K-means clustering and data visualization are used for NPR application enhancement.

Key contribution of Cognitive Number Plate Recognition (CNPR) system

- In this approach, initial image process and binarization of an image is meted out.
- Confidence based predication is used to perform distinction between characters and background in registration number plate.
- Binarizing of image is involved, to segregate into the black and white regions.
- Image regionalization is used for elimination stage, which give only the connected regions, which have best chances of enclosing a number plate within it.

In this paper, Section II discusses the various related work that exists in literature. Section III describes the proposed methodology of CNPR system. Section IV addresses the model outcomes, Section V involves results discussion and Section VI concludes the work.

II. RELATED WORK

In ANPR systems, researchers till now, have integrated several algorithms associated with different stages of the system such as locating number via feature-based localization, extraction of segmented characters and

978-1-7281-5493-0/20/$31.00 ©2020 IEEE

feature pertaining to number plate fonts for Indian vehicles [1] ,as no solid rules of number plate for standard features like fonts are obeyed.

At first presence of vehicle is checked followed by triggering the camera to capture the image. There are plenty of localization algorithms but none is suitable upon Indian license plate. Therefore, feature based number plate localization method was designed and implemented, and for pre-processing the grayscale image is being used to find the corresponding binary image by using Otsu's method [2].From the practical outcomes, the execution of Otsu's method is found to be limited by the small object size, the large variances of the object and the background intensities, the noise, and some more.

Step by step localization is done further as initially white pixel were located then image-size calculation, pixel density now height matching moved to width ratio each step proceeds with elimination of location not above threshold then finally number plate is located and noises are removed [3],otherwise detection becomes a tough task to be performed by Otsu's method. Now moving forward with character segmentation for that square of size 2X2 from the binary image are extracted. Various methods are tested but the best suitable is Image scissoring algorithms designed to vertically scan and cut the area. Works for both single and double line license plate The same process is followed horizontally, and taking width as the threshold parameter. This results in individual characters being segmented. Moving ahead to create binary image of segmented image for helping in recognition process works well. Statistical feature extraction is used here individual character is divided and features are extracted which is compared with the stored features and the max value provides the required character.

Authors in [4] have proposed method in which the first step for image pre-processing is that the input image was captured by a camera, and in case of video, it converted into frames and appropriate frame is selected. Then the image in converted from RGB to grayscale. Next step is localizing the number in the image. Sobel edge filter is used to locate the pixels where there is a sudden change in gradient intensity as compared to previous adjacent pixel. Any sudden change in intensity helps locate edges. Sobel edge filtering thus helps find boundaries for connected objects. Then bounding box method is used to separate each character in the detected number plate into a separate single image. Then each of the individual images are compared with template of images, and each character is matched pixel by pixel for a match.

In other paper [5] the authors have used the Stroke Width Transformation as a local image operator. Stroke width for each pixel of the picture is calculated by this operator. In this way, the resulting image will include the widths of the strokes associated with corresponding pixels. After the calculation of the SWT of the input image based on the Canny edges map and gradient directions, the letter candidates are extracted. Analysis of filtered letter candidates results in detection of words regions. This returns a list of the number plate candidates. The it determines the true plate image among the extracted list. It uses PCA for classification of characters, which gives better result as compared to another classifier.

In the paper [10] the authors have discussed about ways of retrieval of image on the basis of method that has been used to recognize the number in number plate, with the help mobile application. The framework of recognizing car registration number plate helps simplify the vehicle management system for an office parking lot.

In smart buildings like office where employees work late night, calls cab to go home, have guard note down the cab number, along with the employee details. Smart card which all employee has, can be used on the Smartphone with one swipe. So that's why the need for collecting the registration number with details need to be digitalized for convenience and time saving. This has been implemented using ORC based method, but this creates problem on its own, as there is no standard font being used for number plate in India and across the world. So the security in-charge is equipped with the ask to capture the image of each vehicle leaving, and correspondingly the swiping the cards of the employee inside to create a log entry. This method is not very viable option as time consumption is high which leads to jamming the exit.

III. PROPOSED METHODOLOGY

In this work, the Cognitive Number Plate Recognition (CNPR) system is proposed for recognizing the number plate and consists of four main functions, as listed below.

- Preprocessing
- License plate localization
- Character segmentation
- Character recognition

This is followed by creating a web app for visualizing the dataset consisting of data entry log of cars in particular parking area for instance car park, society.

For the dataset thus obtained, k-means clustering, a famous technique for cluster analysis in data mining, has been applied to it. The purpose of the method is to create k clusters in order to divide 'n' observations such that each of the 'n' observation goes to one of the cluster with the value closest to mean. As a result, several cells are obtained out of the data. Let $m^t_a \ldots m^t_k$ initial set of k means. Let, any y_P goes to any C_i cluster based on two parameters namely positioning and updating.

The formula for Positioning is given by the equation (1) below,

$$C_i = \{y_P : |y_P - m^t_a| \leq |y_P - m^t_b| \ \forall\ b, 1 \leq b \leq k\} \quad \text{---} \quad (1)$$

Similarly, the formula for Updating is given by equation (2) below,

$$M_a^{(t+1)} = \frac{1}{abs|C_i|} \sum_{Y_p \in C_i} Y_n \quad \text{------} \quad (2)$$

The proposed methodology for Cognitive Number Plate Recognition (CNPR) involves seven steps namely preprocessing, license plate localization, Covering the number plate bounding rectangle is positioned, character localization, character segmentation, dataset generation and cluster.

1. Preprocessing: There are quite a few challenges faced in implementing the framework of automatic registration plate recognition. Hence, this phase is most crucial to improve the input image for the following steps. In preprocessing the initial steps done is to apply minimum filter to the image so as to improve the dark values in the image by expanding their region. Basically, this is done to negate any impact of light strips and highlighting the characters and the plate edges. To build this partition between colors, the saturation of image is increased and to distinct the background from highlights, the image contrast is enhanced. At this point the image is changed over to grayscale.

2. License Plate Localization: Here, from the preprocessed image, the position of the registration numberplate is known and a sub-image containing only the number plate is obtained as the output of this phase.

3. **Covering the number plate bounding rectangle is positioned:** During this phase a rectangle that contains the registration number plate is placed (may contain some additional part from the sides), and for further processing this rectangle is the input to approaching phase.

4. In this stage, **characters are localized** in the bounding rectangle, using predefined constraints for size of characters (numbers) in the number plate. Now the segmented characters are further processed for recognition.

5. **The segmented characters** are now parsed for character recognition, using machine learning techniques. Then characters are checked for similar looking characters already stored in an array, this is done to double check any characters which may have wrongly recognized. Confusion matrix allows for better character recognition.

6**. The dataset** is then used for web app, which has been implemented using JavaScript libraries like D3.js (data driven documents) which is a JavaScript library used to create interactive visualizations in the browser; DC.js (dimensional charting) which is a charting library for exploring large multi-dimensional datasets.

7. **k-means clustering**, which is easy to apply to large data sets, is then applied. The application of k-mean clustering involves areas like computer vision, marked segmentation and astronomy among many other domains. It then creates cluster according to 'State' and 'Company'. The clusters formed are often used as first step, for finding starting configuration.

Initially the data was collected, and various preprocessing methods to the data were carriedout, and the result was stored accordingly. The flow diagrams for various steps applied are depicted as in figure 1 below.

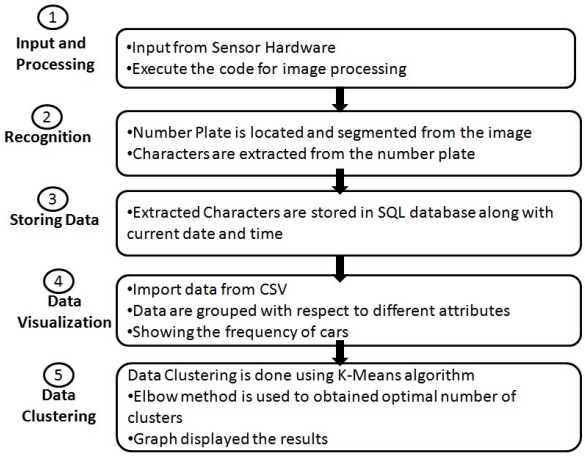

Figure 1. Flow diagram of the Enhanced ANPR

IV. MODEL OUTCOMES

The proposed CNPR system aims toward recognition of different number plates of India. The picture is to be collected from the user, and the pre-processing steps are applied which responds with where the number plate is located in the image by firstly getting the grey scale image and then corresponding adaptive binary image, followed by localization of image which marks out regions for further classification and features addition for next steps.

Then the characters are segmented from the number plate, and are send for the recognition, which then gives an output that's the ASCII of the characters obtained. Further the number detected will be stored in the database for accessing later, and other usage.

Phase 1 - Number Plate Detection:

Phase 1 is the most significant preliminary process toward number plate recognition. It is at this stage that the position of the number plate is determined. The pseudo code for number plate detection is given below. Also, Figure 2 depicts the grey scale image and figure 3 depicts the binary image.

Pseudo Code# 1 : **Number Plate Localization**

1: input the image,
IF necessary: resize
End if
2: Convert input image into grayscale
3: Find Otsu's threshold
4: Convert into adaptive binary, and get connected regions
5: Define expected plate dimension (min, max: height and width)
6: for each region:
If region area less than threshold: continue
If region height & width satisfy plate dimension:
Region maybe number plate
End for
7: returns potential regions

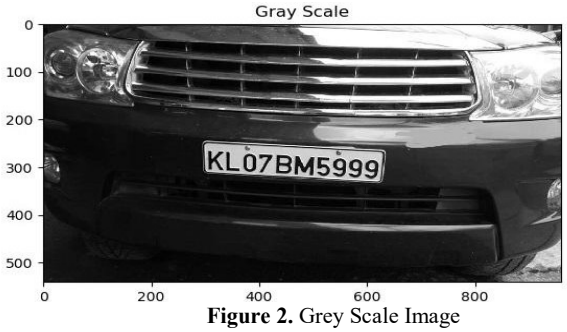

Figure 2. Grey Scale Image

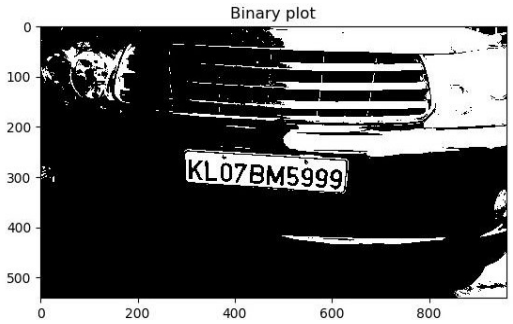

Figure 3. Binary Image

First, the input image was converted into the gray scale image then the grey scale image is adaptively converted into binary image using the Otsu's method. Then further steps involve reducing of pixels setting vertical and horizontal axis most suitable for our number plate localization, removing noise is also one of the main step in determination of number plate.

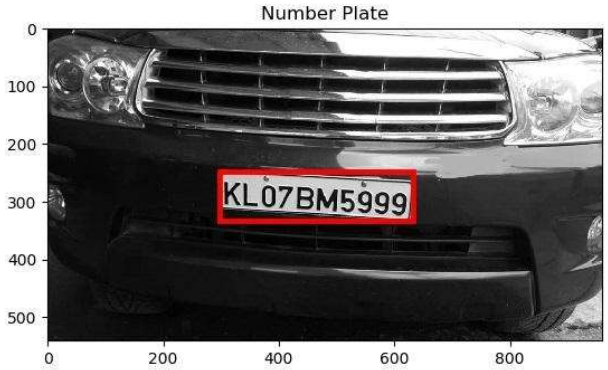

Figure 4. Number Plate localized

Phase 2 - Character Segmentation:

In this phase, the characters on the number plate are mapped out and segmented into individual images. The pseudo code 2 below depicts the various steps involved in segmentation of character from the number plate. Also, the Figure 5 show the result of character segmentation.

Pseudo Code# 2 : **Character Segmentation**

1: for each connected regions
2: define expected character dimension (min, max: height and width)
3: for each sub region in plate:
If region height & width satisfy plate dimension:
    Add box around character
    Add region to character list
End for
4: returns characters and their coordinates

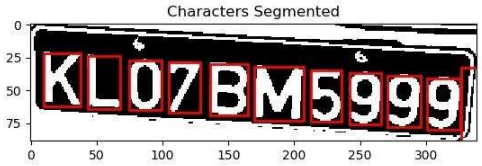

Figure 5. Characters Segmented

Phase 3 - Character Recognition:

The characters earlier segmented are now identified using machine learning techniques, by which the program recognizes characters and stores them in database, for various uses.
The supervised machine learning technique is used to recognize the characters on the number plates. This paper uses 20 images for each of number and alphabet in total 36 characters. First step for character recognition was to convert the images into a one-dimensional array consisting of value of each pixel of an image. Then each of these array is being used as a feature for training our model.

Then the segmented characters are recognized using the trained the model. This initially gives us an imprecise result, after the first test. In the next step, the output is passed through a template matching function. We then take all the characters from output, and find the characters appearing in an array, containing a list of characters which have similar looking characters on typical Indian plate font, for example 2 has '5', 'S', 'Z'. Thus, possible array for each character is formed.

Then for each of these characters, testing is performed to identify whether target character was falsely identified or not. This helps to reduce the count of wrongly identified characters. The output obtained is the final result, the character recognized from the plate, which is then displayed to the user and also stored in the database, along with current timestamp.

Here is a look at theinterface created for the system and output for an image on which E- ANPR was run.

Figure 6. Predicted text is shown in user interface

The predicted number (output/ result) can been above, along with database entry being made below.

```
mysql> select * from anpr where moment like '2018-12-13%';
+----+------------+----------------------------+
| id | plate_text | moment                     |
+----+------------+----------------------------+
| 32 | KL07BH5999 | 2018-12-13 09:51:27.921409 |
+----+------------+----------------------------+
1 row in set (0.00 sec)
```

Figure 7. Predicted number stored in database

Phase 4–Web Application for Visualization

The dataset thus created will be used for data analysis, and k-means clustering. Next, a web app has been created on this dataset for an interactive interface. This has been implemented using JavaScript libraries like D3.js (data driven documents) which is a JavaScript library used to create interactive visualizations in the browser; DC.js (dimensional charting) which is a charting library for exploring large multi-dimensional datasets.

For creating the web app, we have first imported the data from .csv file into the json object 'record'. Then the crossfilter instance for the data is applied. Crossfilter is a JS library, which allows multi-dimensional dataset having large number of entries, to be used for visualization in browser.
It supports extremely fast interaction with datasets containing a million or more records. Thedata is then grouped for each attribute like Month, Day, Brand etc. Then a new dimension is created using the crossfilter. This is then to define the dimension of each individual attribute. Then charts are defined, and respectively placed in html file.

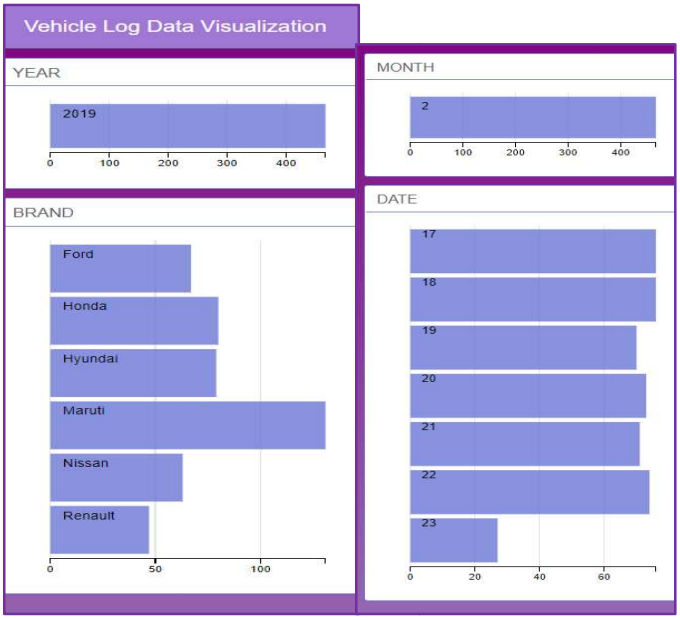

Figure 8. Web Application for Data Analysis

Phase 5–k-means clustering for Visualization

Now k-means clustering has been performed on the dataset. After performing clustering on the dataset, it is observed that the clustering algorithm demonstrates an overall positive correlation between dependent variable and independent variables. While this is a more simplistic example and could be modelled through linear regression analysis, there are many instances where relationships between data will not be linear and k-means can serve as a valuable tool in understanding the data through clustering methods.

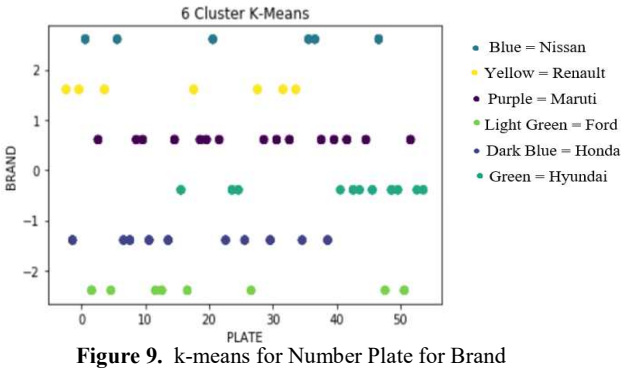

Figure 9. k-means for Number Plate for Brand

Figure 10.k-means for Number Plate for State

V. RESULTS AND DISCUSSION

In the proposed CNPR model, there exists constraint on the size of number plate and character dimension for much improved result. This helped by removing other non-useful connected regions beforehand. Table 1 depicts the results obtained for different plate dimension of region selected with constraints range. Similar, the table 2 depicts the results obtained for character region selection with constraints.

Table 1. Plate Dimension constraints

Input (x,y)	Height Range	Width Range	Selected
(533, 800)	(32,80)	(90, 240)	(72, 115)
(600, 800)	(36, 90)	(90, 240)	(49, 181)
(510, 827)	(29.6, 74)	(90, 240)	(50, 193)
(200, 251)	(16.0, 40.0)	(37.65, 100.4)	(19, 74)
(338, 600)	(27.04, 67.6)	(90, 240)	(54, 94)

Table 2. Character Dimension constraints

Input (x,y)	Height Range	Width Range	Selected
(72, 115)	(15.2, 32.3)	(4.76, 17.85)	(17, 6)
(49, 181)	(19.6, 41.65)	(7.24, 27.15)	(20, 15)
(50, 193)	(20, 42.5)	(7.72, 28.95)	(24, 13)
(19, 74)	(7.6, 16.15)	(2.96, 11.2)	(11, 6)
(54, 94)	(19.2, 40.8)	(6.08, 22.8)	(20, 9)

For experimentation purpose, the database consisting of different sized, color and quality JPEG images. Images with different backgrounds are used, and also with different illumination conditions. The experiment results concur that the algorithm used, for number plate localization, and character segmentation is indeed in good working condition. It also exhibit reduction in noise, illumination variance for the certain images. However, still there exists difficulty in the recognition part of the algorithm, in case of, similar looking characters.

Table 3. Accuracy of our system (in percentage)

Front Plate	Rear Plate	Localization	Character Segmentation	Character recognition
92.3	94.5	97.2	84.5	73.4

These values are taken by calculating average outcomes of our system after running the process. The average results are also taken by the following calculation.

92.3 + 94.5 + 97.2 + 84.5 + 73.4 = 441.9 / 5 = 88.38%.

VI. CONCLUSION

This research paper provided a summary of significant advances made in this exciting area of research with a focus on number plate detection algorithms, although no such system has yet been made oriented for use in India. In this paper proposes an enhanced strategy called Cognitive Number Plate Recognition (CNPR) using Machine Learning and Data Visualization Techniques. The aim is to tackle the issues faced in localization of Indian number plates. The CNPR model allows number plate to be placed in any part of the image. When an input image is given we first extract the region which may be containing the number plate. This region is then processed to segregate the characters in this particular connected region. And finally identification of the segregated characters is done using machine learning technique.

Cognitive Number Plate Recognition can be adapted as a surveillance system that captures the image of vehicle and recognizes their number on number plate. Other applications of an automated system are residential society parking support system, highway/car park system, traffic surveillance, journey time monitoring and many more. In systems like these, the task of number plate recognition of the vehicles, is being automated, this proves to be both time saving and cost efficient. The preliminary outcomes achieved from our study give about 80% accuracy, which is significant, given the small amount of training sample used. Though the proposed model accuracy is not very high, but the vast development in other ML method like Tensorflow, the accuracy can be increased much more. Further, the task of real time application of this system can be carried out as future scope.

REFERENCES

[1] P. Kulkarni, A. Khatri, P. Banga and K. Shah, "Automatic Number Plate Recognition (ANPR) system for Indian conditions," 2009 19th International Conference Radioelektronika, Bratislava, 2009, pp. 111-114.

[2] A. Beibut, K. Magzhan and K. Chingiz, "Effective algorithms and methods for automatic number plate recognition," 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), Astana, 2014, pp. 1-4.

[3] R. Islam, K. F. Sharif and S. Biswas, "Automatic vehicle number plate recognition using structured elements," 2015 IEEE Conference on Systems, Process and Control (ICSPC), Bandar Sunway, 2015, pp. 44-48.

[4] K. M. Babu and M. V. Raghunadh, "Vehicle number plate detection and recognition using bounding box method," 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).

[5] I. M. Gorovyi and I. O. Smirnov, "Robust number plate detector based on stroke width transform and neural network," 2015 Signal Processing Symposium (SPSympo), Debe, 2015, pp. 1-4.

[6] M. Mondal, P. Mondal, N. Saha and P. Chattopadhyay, "Automatic number plate recognition using CNN based self-synthesized feature learning," 2017 IEEE Calcutta Conference (CALCON), Kolkata, 2017, pp. 378-381.

[7] A. Choudhury and A. Negi, "A new zone based algorithm for detection of license plate from Indian vehicle," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, 2016, pp. 370-374.

[8] R. Boliwala and M. Pawar, "Automatic number plate detection for varying illumination conditions," 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, 2016, pp. 0658-0661.

[9] D. Renuka devi and D. Kanagapushpavalli, "Automatic license plate recognition," 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011), Chennai, 2011, pp. 75-78.

[10] A. Visvanathan and T. Chattopadhyay, "Car number plate recognition for smart building management," 2014 19th International Conference on Digital Signal Processing, Hong Kong, 2014, pp. 73-77.

[11] Y. Shi and Y. Chen, "License Plate Detection Based on Convolutional Neural Network and Visual Feature," 2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Huhhot, 2018, pp. 514-519.

[12] S. K. Ingole and S. B. Gundre, "Characters feature based Indian vehicle license plate detection and recognition," 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017, pp. 1-5

[13] B. Pechiammal and J. A. Renjith, "An efficient approach for automatic license plate recognition system," 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM), Chennai, 2017, pp. 121-129.

[14] A. Agarwal and S. Goswami, "An Efficient Algorithm for Automatic Car Plate Detection & Recognition," 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, 2016, pp. 644-648.

[15] W. Liu and C. Lin, "A hierarchical license plate recognition system using supervised K-means and Support Vector Machine," 2017 International Conference on Applied System Innovation (ICASI), Sapporo, 2017, pp. 1622-1625.

[16] A. K. Singh and S. Roy, "ANPR Indian system using surveillance cameras," 2015 Eighth International Conference on Contemporary Computing (IC3), Noida, 2015, pp. 291-294.

[17] S. Subhadhira, U. Juithonglang, P. Sakulkoo and P. Horata, "License plate recognition application using extreme learning machines," 2014 Third ICT International Student Project Conference (ICT-ISPC), Nakhon Pathom, 2014, pp. 103-106.

[18] B. Y. Amirgaliyev, C. A. Kenshimov, K. K. Kuatov, M. Z. Kairanbay, Z. Y. Baibatyr and A. K. Jantassov, "License plate verification method for automatic license plate recognition systems," 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO), Almaty, 2015, pp. 1-3.

[19] A. Sasi, S. Sharma and A. N. Cheeran, "Automatic car number plate recognition," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-6.