AN AUTOMATIC EXTRACTOR AND RECOGNIZER FOR INDIAN VEHICLE LICENSE PLATES

INDU.V, P.RAVI KUMAR.

Abstract: Automatic License Plate Recognition (ALPR) deals with extracting vehicle license plate information from an image or a sequence of images. Though standards are specified for the vehicle number plates by law, they are not followed strictly in India; and, the High Security Registration Plates (HSRP) are yet to be implemented. The dissimilarities in the license plates character extraction and character recognition difficult in the case of vehicle number plates This paper proposes an elegant and robust method -for extracting registration numbers from localized license plates and recognizing them by computer techniques. After binarization of the image and noise removal a clustering algorithm is implemented along with connected component analysis to locate the characters. Then the characters are re extracted and binarised for better results. The extracted characters are recognized using a feed forward back propagation neural network. The proposed work can be used for applications like access control, traffic law enforcement etc. which need to process blurred images obtained by ALPR cameras.

Keywords: Automatic License Plate Recognition (ALPR), back propagation neural network, character extraction, character recognition, clustering algorithm, Niblack binarization.

1. Introduction: Vehicles in India have unique license numbers which is written on their license plates. Automatic License Plate Recognition (ALPR) is emerging as an actively researched area spanning wide disciplines [1] like image processing, pattern recognition, neural networks etc. License plate recognition has numerous real life applications in law enforcement, intelligent transportation systems and parking and access control.Difficulties in license plate recognition are mainly due to the quality of the image captured at outdoors. These are affected by different illumination, reflections, skew, non-standard and fancy writing etc.ALPR consists of the following four steps.Image acquisition consists of acquiring vehicle images for processing. License plate localization deals with extraction of license plate from the vehicle image. Character segmentation extracts characters from the localized image and character recognition identifies the characters.Lot of algorithms has been proposed for license plate recognition. Connected component analysis [2] is one of the widely used methods in character segmentation. Other methods include taking horizontal and vertical projections [3], [4], segmentation based on previous knowledge about number of characters [5], and the methods using combined features [6]. Methods used in character recognition mainly consist of recognition using templates and recognition using extracted features [7], [8].

During our study, it was found that studies conducted on license plate recognition of Indian license plates is very less compared to the vast amount of research already conducted in this area. Studies have mostly been conducted on manually acquired images of stationary vehicles [9], [10] (mostly close zoom and clear visibility) Vehicle images captured from a road-side surveillance camera automatically throughout day and night has also been studied [11]. Satadal Saha et al., proposed[11] template matching based on matching factor for real field images. D.Renuka Devi and D .Kanagapushpavalli[9] proposed template matching using normalized correlation for manually acquired images. Kumar Parasuraman and P.Vasantha Kumar[10] performed character recognition using chain code concept for manually acquired images. In this paper, we propose a different approach based on neural networks for character recognition for images captured by ALPR cameras installed at traffic junctions.Of the four steps in ALPR, character extraction and character recognition are dealt with in this paper. The localized license plate image is binarised and connected components are extracted based on size. Then a clustering algorithm is employed to group similar components and exclude non characters. After re extraction and binarization, features of the candidate characters are extracted and a back propagation neural network is used to classify

the characters. The license plate number can be reconstructed from these characters.

The rest of the paper is organized as follows. The proposed system is explained in Section 2.Preprocessing steps are given in Section 3.Section 4 details character segmentation. Character recognition based on neural networks is given in Section 5. Experimental results are presented and discussed in Section 6. Conclusions are given in Section 7.

2. Proposed system: The proposed system takes localized gray scale license plate images as input. The input to the system is obtained from RLVDS (Red Light Violation Detection System) [12] done by CDAC (Centre for Development of Advanced Computing) Trivandrum.Block diagram of the proposed method is shown in Fig.1.The first step is preprocessing, during which the gray scale image is binarised using Niblack binarization[13] method and noises are removed after connected component analysis. Preprocessing is explained in detail in Section 3.

After preprocessing, the binarised image is segmented by examining the connected components and selecting and grouping the characters using a clustering algorithm. Then these characters are re extracted from the localized gray scale license plate and binarised using another local thresholding technique so that better results are obtained during recognition. The algorithm along with re extraction and binarization is explained in Section 4. To recognize these characters so that the number plate can be reconstructed, the features of the characters are extracted and they are classified using the feed-forward

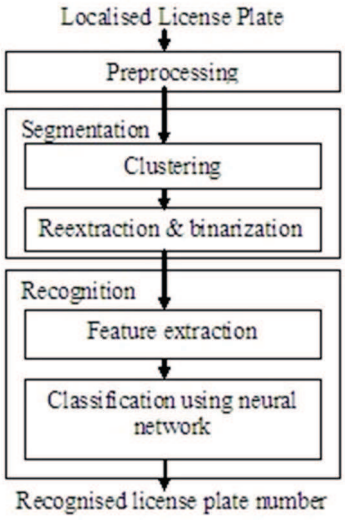


Fig 1.Proposed method

back propagation network which was trained earlier using files [14] which contains 650 samples of images of English characters with 25 images each character. The details of these are given in Section 5.

3.Preprocessing : The steps in preprocessing are shown in Fig.2.

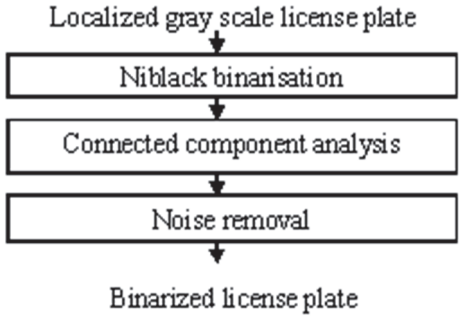
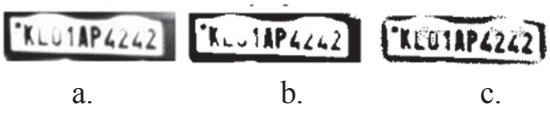


Fig 2.Preprocessing

A. Binarisation

Image binarization [15] converts a gray scale image to a binary image.

In global thresholding methods, as in Otsu's method [6], a single threshold is chosen for the entire image as opposed to local thresholding methods. Niblack method of binarization is a local thresholding method where value of threshold at any point (x,y) in an image depends on properties of neighborhood of (x,y). The reason for choosing local thresholding method is that it gives consistent quality results when the image has non uniform illumination, which is true in the case of our license plate images. Field images may have different lighting at different portions, may be blurred etc. Images of license plate binarised using Otsu and Niblack are shown in Fig.3.



a. b. c.

Fig 3. License plate binarization
a. Gray scale image b. Image binarised using Otsu c. Image binarised using Niblack

In Niblack binarization, threshold t at each pixel is calculated as a function of mean m and standard deviation s of all pixels in the N*N window as given by Eq.1 $t = m + k * s$ (1)

The window height (N) chosen gives better results when N approximates character height, (experimentally found to be 15) for the images chosen for this study. k is a constant which is usually taken

to be -0.2[13].

B. Finding connected components

The procedure [16] for finding connected components is as follows.

1. In the image, search for the next unlabeled pixel p.
2. Use a flood fill algorithm (8 connectivity) to label all the pixels connected to p.
3. Repeat steps 2 and 3 till all pixels are labeled.

C. Noise Removal

Even though binarization suppresses the back ground features, unwanted elements, i.e. non-characters will still be present in the binary image. These should be eliminated .In this step all components having area less than a specified area, (i.e., size of the smallest character) is removed. For this purpose, the filled area (area of foreground pixels alone) is calculated. This is done as distorted shapes or noises may pass off as characters when their bounding area exceeds the size, even though their filled area is less. Also all components comparable to the height and width of license plate are removed. This is done to remove border areas of license plate as well as the separating lines in multiline license plates. The resulting binarised image is passed to the next stage for segmentation.

4. Character Segmentation: In this stage, the properties of the bounding boxes [17] of the connected components in the binarised image are extracted. Then a clustering algorithm is applied to extract characters alone and separate multi lines. The characters are re extracted and binarised so that better results are obtained during recognition.

A. Clustering Algorithm

The input to this algorithm is the properties of the bounding boxes and the outputs are the candidate clusters. The x and y values of the top left corner of the bounding box as well as the height and width of the box form the properties. The clustering algorithm is as follows. 1. Sort the bounding boxes according to their y values. (This is done so that clusters will be formed in the same order as the license plate number)

2. Initialize the first bounding box as the first cluster.
3. For each bounding box encountered, perform the following steps.

a. Find the cluster c, which showed the least difference d in standard deviation of height among the existing bounding boxes. The cluster should also satisfy the property that the difference in mean of the y value should fall within a threshold (experimentally set to 3.)

b. If this least difference d lies within a limit (experimentally set to 4), add the bounding box to the same cluster. Else create a new cluster.

4. Identified clusters are output to the next stage.

B. Reextracting characters and binarization

The image obtained during first binarization contains characters which are recognizable to human eye. But for machine recognition, we need characters which are of better quality. Also some characters will be merged in Niblack binarization. To overcome these difficulties, the clusters which were found earlier are re extracted from the original gray scale image of the localized license plate and binarised again.

Here, we use a different form of local thresholding .For each pixel, the threshold is found by calculating the mean of pixels in the neighborhood. The neighborhood size is chosen as 2*2 for the images we have studied with. To find the threshold of pixels along the border, the image is added along the border with the contents of the border pixels.

After this binarization, images of characters gave better results during recognition. Also, if two or more characnt ters are merged together they are separated. This is shown in Fig. 4.

a. b.

Fig 4.Merged characters separated during re extraction and binarization
a.Niblack binarised image b.Re extracted and binarised characters.

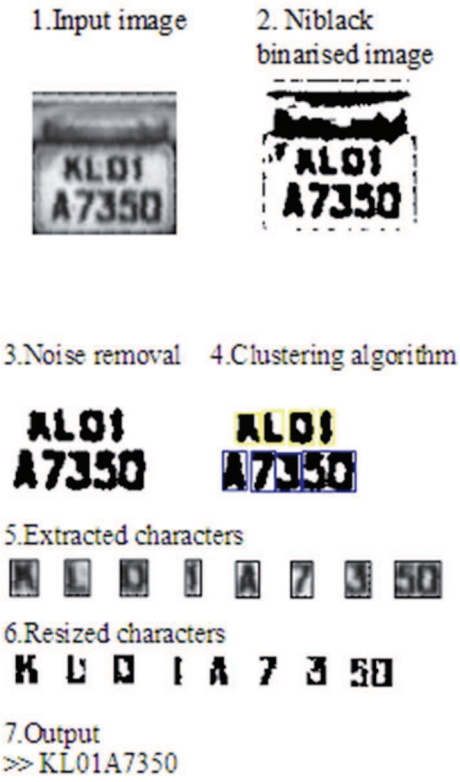
5. er recognition, the characters extracted during the previous stage were resized; their features extracted and were passed to a feed forward back propagation neural network for classification.

The network was trained earlier using features extracted from training data. Training data consists of alphabets obtained from ref [14], and numbers created as bmp files using paint. The character database contains 25 samples of each alphabet (A to Z) in bmp format of size 50*50, of which 10 samples each were taken for training. Files containing numbers from 0 to 9 (10 samples each) were also

created in the same format. Features mentioned in ref [18], were extracted for the same purpose.

6. Results: The results obtained during study are given below.

A. Different stages in implementation



B. Test results

For the 110 images tested, the character segmentation step was successful in 96.36% of the cases. In the successful cases, character recognition was accurate in 90.56 %. Thus the entire system was successful in 87.27% of test images.

Character segmentation failed for cases in which images were of very poor quality, i.e., difficult to be read even by humans and in cases were characters were broken. Performance of character recognition can be improved by improving performance of neural networks, by incorporating more features and increasing the number of sample images used for training.

7. Conclusion: In this paper, a connected component and neural network based approach for character extraction and recognition was presented. A simple and efficient clustering algorithm for character segmentation using connected component analysis is proposed. During preprocessing, Niblack binarization instead of Otsu for achieving better results in binarization. Clustering algorithm identified the candidate characters in correct order and separated multi line characters. Re extraction and binarization using another local thresholding method helped in separating merged characters as well obtaining better results in character recognition using neural network. Overall the system showed a success rate of 87.25%.

References

1. Shan Du et al.," ALPR:A State of the Art Review", Copyright (c) 2011 IEEE, in press
2. A K Ghosh et al., ALPR for Bangladeshi Vehicles, Global Journal of Comp.Science and Tech. Volume 11 Issue 21 Version 1.0 December 2011
3. Xia H., Liao D., "The study of license plate character segmentation algorithm based on vetical projection", CECNet 2011, pp. 4583–4586.1074
4. Fikriye Ozturka, Figen Ozena, "A New License Plate Recognition System Based on Probabilistic Neural Networks", Insode 2011
5. Zhen-Xue Chen et al. ,"Automatic License-Plate Location and Recognition Based on Feature Salience", IEEE Transactions on Vehicular Technology, vol. 58, no. 7, September 2009
6. N.Otsu,"A Threshold Selection Method from Gray-Level Histograms" ,IEEE Transactions on SMC, vol. SMC-9, NO. 1, January 1979
7. Jianhong Xie,Optical Character Recognition Based on LS_SVM, Third Int. Symposium on Intelligent Information Technology ,2009
8. Subhadip Basu, et al. "A hierarchical approach to recognition of handwritten Bangla characters",Pattern Recognition 42 (2009) 1467 – 1484
9. D.Renuka Devi et al., "Automatic License Plate Recognition" 978-1-4673-0132-91111$26.00@2011 IEEE
10. Kumar P., P.V. Kumar P., 2010,"An Efficient Method for Indian Vehicle License Plate Extraction and Character Segmentation ", 2010 IEEE Int. Conf. on Computational Intelligence and Computing Research
11. Satadal Saha et al. , "Automatic Localization and Recognition of License Plate Characters for Indian Vehicles ",Int. J Comp Sci. Emerging Tech Vol-2 No 4 August, 2011
12 .CDAC-ITNS Project Report (RLVD)

13. W.Niblack, An Introduction to Digital Image Processing. Prentice Hall, Engle woodCliffs, (1986).
14. http://mathworks.in/matlabcentral/fileexchange/26158-hand-drawn-sample-images-of-english-characters/all_files
15. R.C.Gonzalez, R.E.Woods, Digital Image Processing ,Third Edition, Pearson Education
16. http://www.mathworks.in/help/images/ref/bwconncomp.html
17. http://www.mathworks.in/help/images/ref/regionprops.html
18. http://arxiv.org/ftp/arxiv/papers/1202/1202.3884.pdf

* * *

CIS., SIST, Thiruvananthapuram-695883,Kerala.induvenugopal@gmail.com
Joint Director, ITNS, CDAC,Thiruvananthapuram-695033,Kerala.ravikumar@cdac.in