A NOVEL APPROACH TO EXTRACT TEXT FROM LICENSE PLATE OF VEHICLES

R.Radha[1] and C.P.Sumathi[2]

[1]Department of Computer Science, SDNB Vaishnav College for Women, Chromepet, Chennai-44
radhasundar1993@gmail.com
[2]Department of Computer Science, SDNB Vaishnav College for Women, Chromepet, Chennai-44
drcpsumathi@gmail.com

ABSTRACT

In this paper the text found on the vehicle plates is detected from the input image and this requires the localization of number plate area in order to identify the characters present on it. In this work, simple color conversion edge detection and connector measurement technique with the application of median filter as one of the operators is attempted. This paper presents an approach using simple but efficient morphological operations, filtering and finding connected components for localization of Indian number plates. The algorithm has been tested on 100 samples and is found to extract both alphabets and numbers from vehicle license plates images with an accuracy of 93% for both two and four wheeler license plates of Chennai region.

KEYWORDS
Edge Detection, Median Filter, Connected components, Bounding Box, morphological operations

1. INTRODUCTION

Every country uses their own way of designing and allocating number plates to their country vehicles. This license number plate is then used by various government offices for their respective regular administrative task like- traffic police tracking the people who are violating the traffic rules, to identify the theft cars, in toll collection and parking allocation management etc. In India all motorized vehicle are assigned unique numbers. These numbers are assigned to the vehicles by district-level Regional Transport Office (RTO). In India the license plates must be kept in both front and back of the vehicle. These plates in general are easily readable by human due to their high level of intelligence on the contrary; it becomes an extremely difficult task for the computers to do the same. Many attributes like illumination, blur, background color, foreground color etc. will pose a problem. Also, the License Plate Recognition (LPR) in India is difficult because the traffic rules are hardly obeyed and the number plate standards are not strictly practiced. Each one is adopting a different style leading to obtaining variation in parameters like, size of number plate and characters, location of number plate, type of font used, background (white background with black letters for non commercial vehicles and white background with yellow letters for commercial vehicles), different unwanted pictures etc. which makes the task of number plate localization very difficult. The main aim of this article is to implement an efficient method to recognize license plates and extract text from them under Indian conditions.

DOI : 10.5121/sipij.2012.3415

This work is carried over for on car number plates as well as on two wheeler number plates. But this can be applied to any type of motor vehicle. A typical example of an Indian license plate (for car) is shown in the figure 1 with the significance of each character (1. State Code, 2. District Code, 3.Type of Vehicle (car, two wheeler, commercial etc.) 4. Actual Registration Number) [1]. In this we can observe that the license plate is represented with various font size for the state code, district code, type of vehicle and registration number. A sample two wheeler license plate is given in figure 2. Even many two wheeler license plates are arranged in row wise like in figure 1. Thus in India there is no defined alienation followed for writing the license number plates.

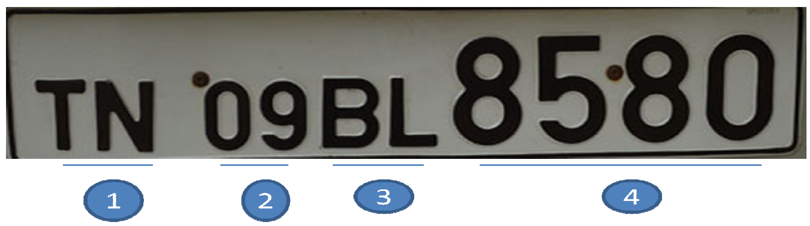


Figure 1 Format of Indian Car License plate

Figure 2 A Two wheeler Valid License Plate format

2. Related Work

In literature we can find many methods for number plate detection and recognition system. The major drawback is that how long it will take to compute and recognize the particular license plates. This is critical and most needed when it is applied to real time applications. However, there is always a trade-off between computational time and performance rate. In order to achieve an accurate result and increase the performance of the system more computational time is required. For number plate detection or localization, techniques based on edge statistic and mathematical morphology gives a very good result as reported in Bai and Liu [2] work that uses vertical edge information to calculate the edge density of the image followed by morphology methods such as dilation to extract the region of interest. This technique works well due to the fact that number plates always have high density of vertical edges. But in this method as unwanted edges in the background are also detected which leads to confusion, it is difficult to apply this method for number plates with complex background.

Color based techniques are proposed by Paolo *et. al*. [3] and Dai *et. al*. [4]. The draw back with their method is that it performs well when the lighting condition is constant but when there is various illumination condition its performance reduces. But in real-time application normally the images can be obtained with various lighting illumination. Furthermore, the proposed technique is country specific because each country will have different color code for vehicle number plate. In the work of Oz and Ercal [5], Connected Component Analysis (CCA) method is used to detect the number plate region. CCA is useful for simplifying the detection task since it labels binary image into several components based on their connectivity. Based on the problem one can decide on the selection of finding the connected components using 4-adjacency or 8-adjacency of pixels connectivity. Spatial measurement is a measure of spatial characteristics of a connected component such as area, orientation, aspect ratio etc. and filtering is done to eliminate unrelated or unwanted components. When Connected Component Analysis is combined with spatial measurement and filtering produces better result in number plate detection.

A simple horizontal scanning of the image looking for the most repeating brightness changes is the method applied by Kong *et. al*. [6]. A number plate always has significant number of brightness changes due to the transition from the character to background and vice versa. Several image transformation techniques have been used in number plate detection. Among the techniques, Hough transform implemented by Duan *et. al*. [7] yields a satisfactory result. But their method has high computational power and so it is not suitable for real time applications. Vladimir Shaprio *et al*, [8] in their approach deals with stages of preprocessing which involves vertical edge detection and rank filtering. With this preprocessing they obtained a vertical projection of the number plates and detected the horizontal strip loosely locked on the plate and clipped them from the image. The skewed portion are detected and are deskewed. The characters are then segmented and recognized.

Gabor filter is implemented by Kahraman et. al. [9]. As Gabor filter posses the optimal localization properties in both spatial and frequency domain they are well suited for analyzing texture segmentation problem. But it is computationally a time consuming method. Wavelets transform which is implemented by Hsieh et. al. [10] locate multiple plates with different orientations in one image. Their method excel with differently illuminated and oriented license plates with an accuracy of 92.4% . Apart from the above mentioned methods Artificial Neural Network (ANN) , Genetic Algorithm (GA), Principal Component Analysis, Fuzzy systems, Hidden Morkov models etc can also be employed. ANN is experimented by Parisi et. al. [11] and Chang et. al.[12]. Cui and Huang [13], applied GA to locate the number plate from video sequence. AdaBoost algorithm which select a small number of weak classifiers from a very large set of weak classifiers and construct a strong classifier to classify number plate region in the input image under various illumination conditions was designed by Dlagnekov [14] and has obtained 93.5% of detection rate. Additionally, Haar-like features were used in AdaBoost training in Sun et. al. [15]. And they proved that the performance of Gentle Adaboost is better in detection and false positive rate than the discrete AdaBoost and real AdaBoost. (Choo Kar Soon et al, [16]) presents a number plate detector based on combination of AdaBoost and connected component analysis (CCA) algorithms. They used KNN classifier to recognize the characters extracted from the above said method. (Kaushik Deb et al, [17]) adopted a method which segment the images, then applied recursive connected component labeling and filtering for candidate region detection. They used HIS color model for color verification of candidate region and decompose the candidate region using histogram to detect vehicle license plate region.

Anish Lazrus et al [18] presented a robust method of license plate location using segmentation and reorganization of the characters present in the located plate. Wiener2 filter was used to remove noise and Sobel filter for finding and smoothing edges and then the connected components were found. Sandeep Singh et al [19] used several heuristics methods that successfully allow the detection and extraction of alphanumeric regions from correctly identified license plate followed by recognition of characters using Optical Character Recognition (OCR). Pablo Negri et al [20] addressed the license plate detection and recognition (LPR) task on still images of trucks. They hybridized different segmentation algorithms to improve the license plate detection. P.Anishiya et al [21] proposed an algorithm based on a combination of morphological operation, segmentation and Canny edge detector. F.Martin et al [22] adopted an algorithm in which plate location is based on mathematical morphology and character recognition is implemented using Hausdorff distance.

3. RELATED DEFINITIONS

A detection algorithm that employs mathematical morphology, structuring element, median filtering, edge detection and connected components to detect the license plate is detailed below.

3.1 Mathematical Morphology

Mathematical Morphology is set-theoretic method for analyzing the image and extracting image components that are useful in the shape representation and extraction of geometrical structure. They are used to detect boundaries of objects, their skeletons, and their convex hulls. These are the basic operations that has to be carried over for any image pre- and post-processing techniques, that include edge thinning, thickening, region filling ,pruning etc.,. The following operations form the basis of mathematical morphology.

Suppose F(x,y) is a grayscale image and B(x,y) is the structuring element, The morphological operations which can be applied on this image is given by

Dilation :

Dilation will cause objects to grow in size as it will exchange every pixel value with the maximum value within a 3x3 window size around the pixel. i.e it adds pixels to the boundaries of objects in an image. The process may be repeated to create larger effects. The size and shape of the structuring element decides the number of elements to be added to the image under processing. It can be represented mathematically as:

$$(F \oplus B)(x, y) = \max F(x - s, y - t) + B(s, t) \tag{1}$$

Erosion :

Erosion works the same way except that it will cause objects to decrease because each pixel value is exchanged with the minimum value within a 3x3 window size around the pixel. i.e it removes pixels from the boundaries of objects in an image. The size and shape of the structuring element decides the number of elements to be removed from the image under processing.. It can be represented mathematically as:

$$(F \ominus B)(x, y) = \min F(x + s, y + t) - B(s, t) \tag{2}$$

Opening :

Opening is an important morphological operator. It is defined as erosion, followed by dilation. Erosion tries to eliminate some of the foreground (bright) pixels from the edges of regions of foreground pixels. The disadvantage is that it will remove all regions of foreground pixels indiscriminately. Opening gets around this by performing both erosion and dilation on the image.

$$F \circ B = (F \ominus B) \oplus B \tag{3}$$

Closing :

Closing is similar in some ways to dilation in that it tends to enlarge the boundaries of foreground (bright) regions in an image (and shrink background color holes in such regions), but is less destructive of the original boundary shape. Closing is defined as dilation, followed by erosion. The effect of the operator is to preserve *background* regions that have a similar shape to this structuring element, or that can completely contain the structuring element, while eliminating all other regions of background pixels.

$$F \cdot B = (F \oplus B) \ominus B \qquad (4)$$

3.2 Structuring Elements

In order to carry over the dilation and erosion operations on images the structuring element are used. A structuring element is a matrix with m X n size. The values in this matrix are of binary value i.e either a 1 or 0. The pixels with values of 1 next to each other are called the neighborhood pixels.

In a morphological operation, the origin of the structuring element is compared with every pixel along with its neighbors in the input image and translated to each pixel position in the corresponding output image. The outcome of this comparison depends upon the type of morphological operator and size of structural element used.

Sample structuring elements are given in figure 3.

1	1	1	1	1
1	1	1	1	1
1	1	**1**	1	1
1	1	1	1	1
1	1	1	1	1

square 5 x 5 element

0	0	1	0	0
0	1	1	1	0
1	1	**1**	1	1
0	1	1	1	0
0	0	1	0	0

Diamond-shaped 5 x 5 element

0	0	1	0	0
0	0	1	0	0
1	1	**1**	1	1
0	0	1	0	0
0	0	1	0	0

Cross-shaped 5x5 element

1	1	1	1	1
1	1	**1**	1	1

2x5 structuring element

Figure 3. Examples of simple structuring elements

The red color in each matrix represents the origin. This paper uses a (2,4) structuring element.

3.3 Median Filter

The median filter is a non-linear filtering technique used to remove noise from image under consideration. While it helps in removing the impulse noise it preserves the edges. As the impulse noise spikes are much brighter than their neighboring pixels, they are generally placed in the extreme top or bottom end of the brightness ranking while analyzing the neighborhood of input pixels. As a consequence, these extremes values with noise which lie far away from the median value are removed by the filter which leads to dramatic reduction of noises from the image. Repeated application of median filter make the image with uniform regions that are very effective when classified for segmentation. Because of its nonlinearity it is unsuitable for common optimization techniques. A list of pixel is generated in the filter kernel and by sorting them gives median which is situated in the middle of the list is found. In the general case, this algorithm's per pixel complexity is O(r2 log r), where r is the kernel radius. As a matter of fact, median filter is a statistical non-linear filter that is often described in the spatial domain. A median filter also smoothens the image by utilizing the median of neighborhood. In this experiment, a 2×4 median filter is used mainly because, this filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. It behaves like low-pass filtering in smoothening and reducing the noise in the image while preserving discontinuities and smooths the pixels whose values differ significantly from their surroundings without affecting the other pixels which is lacking in low-pass filters.

3.4 Edge Detection Method

We may use edges to measure the size of the objects in an image to isolate particular objects from their background, and to recognize or classify objects. There are many algorithms like Sobel, Roberts, Canny, Laplacian, Hoor classify etc. In this paper, we have adopted Laplacian Of Gaussian (LOG) edge detection method.

4 METHODOLOGY ADOPTED

Generally the text in number plates are written with contrast background and foreground like black letters on white background and black letters on yellow background and based on this property of text a localization technique has been proposed in this paper. The work is divided into three major parts namely preprocessing, text localization and extraction and text/non-text classification.

4.1 Preprocessing

In the preprocessing step, the colored input image is converted to gray scale image. The image is then binarized. A binarized image is a must for doing all morphological operations like opening, closing, thinning, skeletonization , region filling etc. To this binarized image median filtering is applied to remove any noise if presents. LOG edge detection algorithm is applied on the resultant image to extract the edges. LOG edge detector detects the edge points through searching for spots which two-order differential coefficient is zero so that it finds the correct places of edges and tests a wider area around the pixel.

4.2 Text localization and Extraction

In this phase, the morphological dilation operation is performed on the edge image obtained from the previous step. Since texts are normally aligned in the horizontal direction a 2x4 rectangular structuring element is used. All Connected Components are then extracted.

4.3 Text/non-text classification

The extracted components from the above step contains both text and non-text components. They are separated and eliminated by a two way process. First, the initial bounding boxes are drawn for all objects (figure 4). The required texts in the connected components are extracted and placed in a jpeg file.

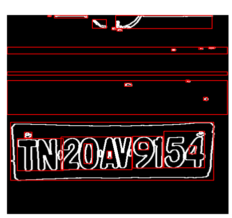


Figure 4. Connected Components

The flowchart explaining the algorithm is given in figure 5.

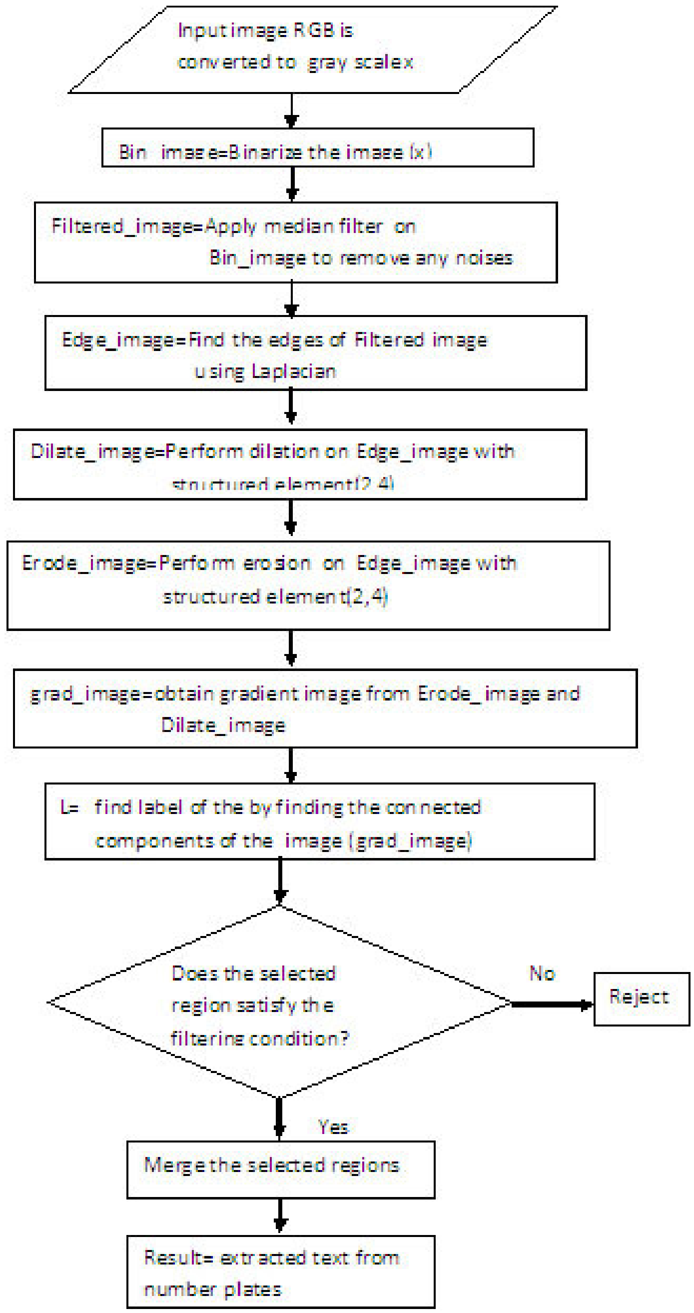


Figure 5: The proposed algorithm for extraction of number plates

The figure 6(a,b,c,d) shows the various stages of the above mentioned steps with a sample of 2 cars and 2 two wheeler. The figure clearly shows how the unwanted regions are eliminated by the algorithm and only the license plate region gets extracted.

5. EXPERIMENTAL RESULTS AND COMPARATIVE RESULTS

The tests were conducted on 100 images taken with the help of Sony Cybershot DSC-W530/W550 14.1 mega pixels digital camera and MATLAB 7 software was used for the experiment. About 95% of the number plates were localized correctly and 5% images resulted in the localization of number plates along with unwanted non candidate regions, because of the damage in the number plates. Except for the unwanted regions, the algorithm works robust under different illumination and brightness. An example for recognizing unwanted non-candidate region is given in figure 7.

Table 1 shows the recognition accuracy rates on real scene samples implemented using different algorithms.

Table 1: Recognition Accuracy Rate on Real Scene Sample

S.No	Research papers	Real Time Data	Images correctly detected	Results accuracy
1.	Bai Hongliang , Liu Changping [2]	9825	9786	99.6%
2	Vladimir Shapiro et al [8]	150	135	90%
3	Kaushik Deb et al [17]	40	33	82.5%
4	Anish Lazrus et al [18]	50	46	98%
5	Sandeep Singh et al[19]	250	220	88%
6	Pablo Negri et al [20]	623	589	94.6%
7	P.Anishiya et al [21]	288	276	96%
8	F. Martín et al [22]	105	84	80.39%
9	Proposed Algorithm	100	93	93%

6. CONCLUSION

This article proposes a text localization and extraction technique from vehicle number plates. The suggested method is tested with various types of vehicles like two wheeler, four wheelers etc. and with different background. The number plates with yellow background and black foreground as well as number plates with white background and foreground with black images were extracted with almost 100% accuracy**.** The algorithm identifies the connected components of most of the number plates except for the damaged ones. In figure 7, one can note that the maximum unwanted objects are removed but the unwanted candidate region like border of the number plate is also extracted. This leads to the false positive detection. In this work, main emphasis is in eliminating false positives The future work will involve in recognizing the individual characters from the above extracted text.

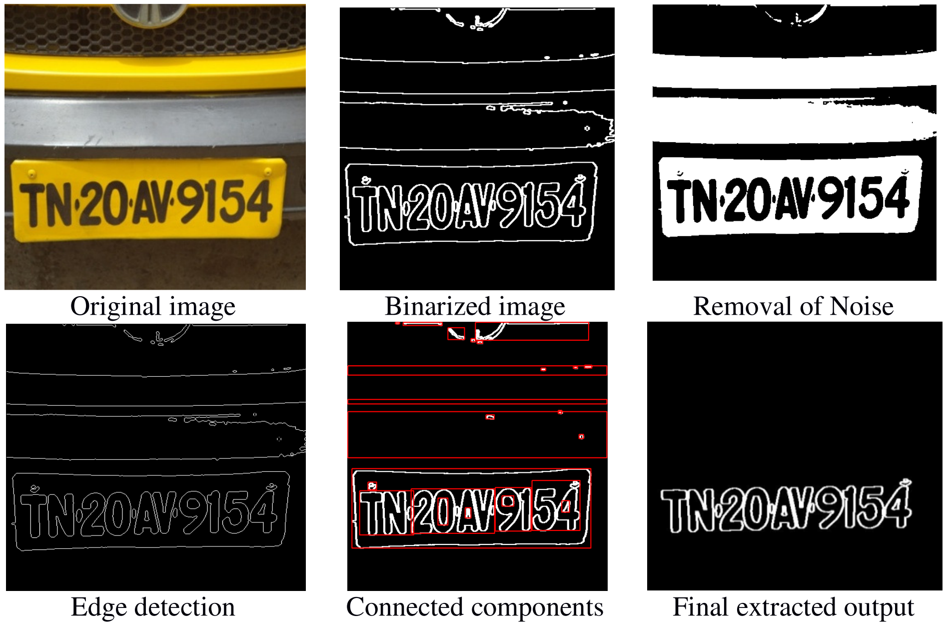


Figure 6a: The stages of the algorithm- applied on a car (Example 1)

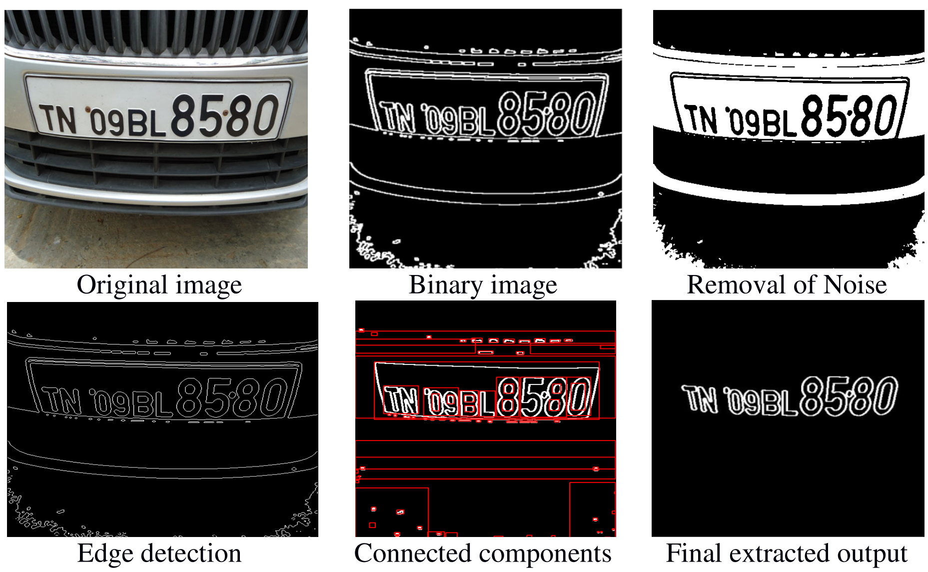


Figure 6b: The stages of the algorithm -applied on a car (Example 2)

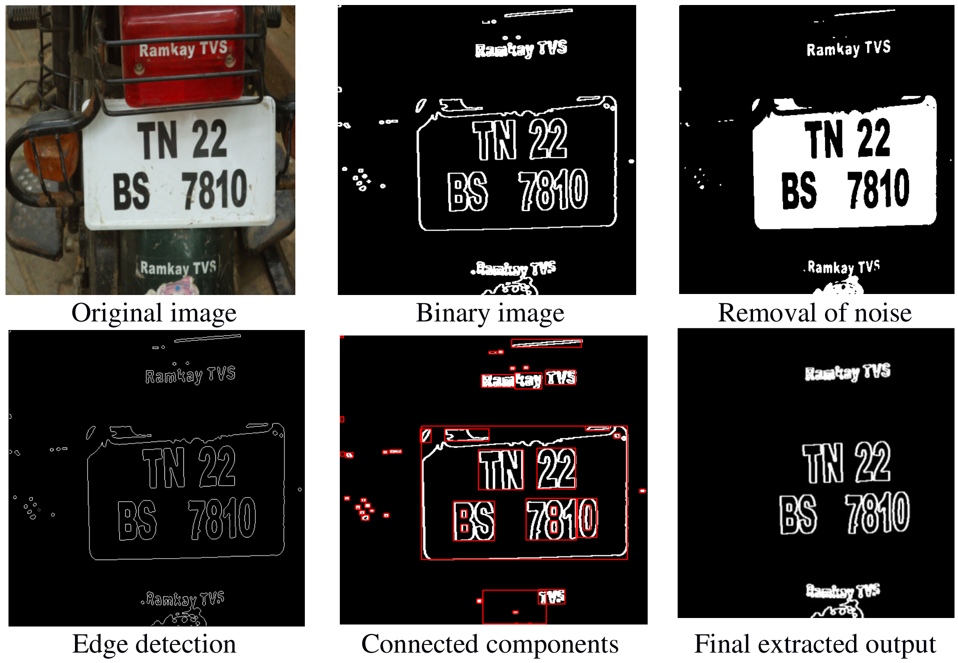


Figure 6c: The stages of the algorithm – applied on a two wheeler(Example 3)

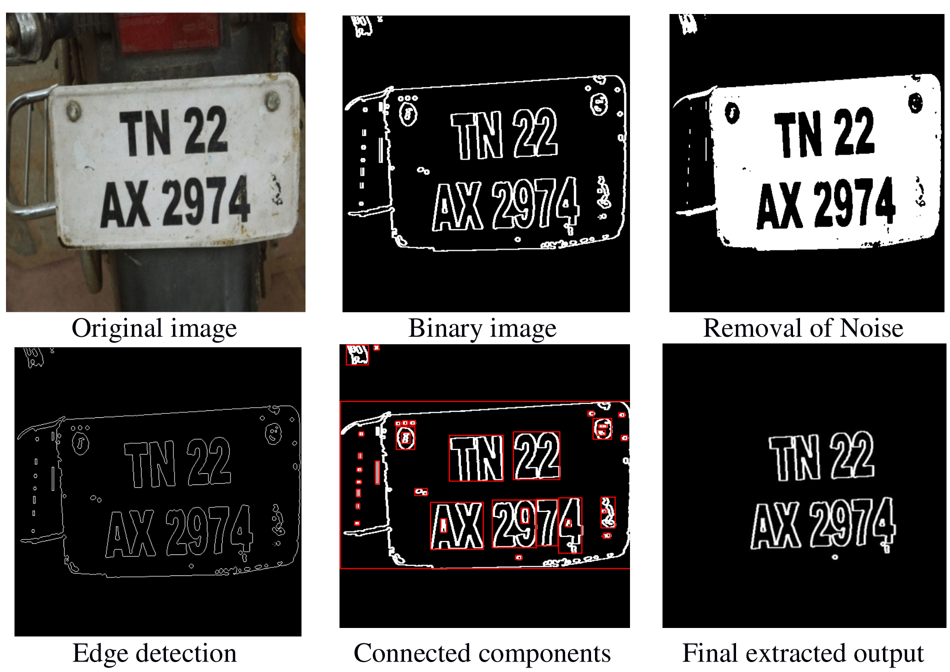
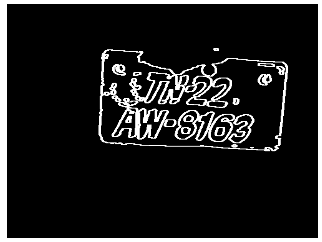
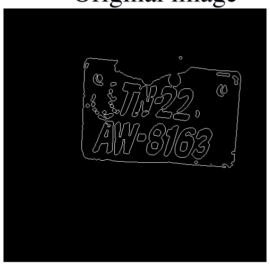
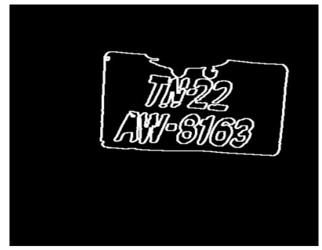


Figure 6d: The stages of the algorithm - applied on a two wheeler (Example 4)



Original image Binary image Noise removed





Edge enhanced Connected components Final output

Figure 7: Recognition of unwanted non-candidate regions

ACKNOWLEDGEMENT

The authors are grateful to the University Grants Commission for sanctioning the minor research grant (Reference Number –No F.MRP-3725/11(MRP/UGC-SERO), Dated 08/09/2011) for this project. They are also thankful to the Principal and the Management of SDNB Vaishnav College for Women, Chennai, for their continued and relentless support for the project.

References

[1] http://en.wikipedia.org/wiki/Vehicle_registration_plates_of_India

[2] Bai Hongliang , Liu Changping ,2004 . A hybrid License Plate Extraction Method Based On Edge Statistics and Morphology. Proceeding ICPR '04 *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)* Vol. 2 831-834.

[3] Paolo Comelli, M.N.G., Paolo Ferragina, F. Stabile, 1995. Optical recognition of motor vehicle License plates,*IEEE Transactions on Vehicular Technology*, 44: 790-799.

[4] Dai Yan, L.J., Ma Hongqing, L. Langang, 2001. A high performance license plate recognition system based on the web technique. *Proceedings of International Conference on Intelligent Transportation Systems*, 325-329.

[5] Oz, C., F. Ercal, 2005. A practical license plate recognition system for real-time environments (3512).*Springer-Verlag*.

[6] Jun Kong, Y.L., Xinyue Liu, X. Zhou, 2005. A novel license plate localization method based on Textural feature analysis. *Proceedings on IEEE International Symposium on Signal Processing and Information Technology*, 275-279.

[7] Duan, T.D., T.L.H. Du, N.V. Hoang, 2005. Building an automatic vehicle license-plate recognition system. *Proceedings of International Conference on Computer Science*, 59-63.

[8] Vladimir Shapiro, Dimo Dimov, Stefan Bonchev, Veselin Velichkov, and Georgi Gluhchev “Adaptive License Plate Image Extraction” , *Proceedings of the International Conference on Computer Systems and Technologies - CompSysTech’2003.*

[9] Kahraman, F., B. Kurt, M. Gokmen, 2003. License plate character segmentation based on the gabor transform and vector quantization, *In Computer and information sciences-iscis* (2869: 381-388). *Springer Berlin / Heidelberg..*

[10] Ching Tang Hsieh, Y.S.J., K.M. Hung, 2005. Multiple license plate detection for complex background. *Proceedings of International Conference on Advanced Information Networking and Applications,* 2:389-392.

[11] Parisi, R., G., E.D. Di Claudio, G. Orlandi, 1998. Car plate recognition by neural networks and imageprocessing. *Proceedings of IEEE International Symposium on Circuits and Systems*, 3: 195-198.

[12] Shyang-Lih Chang, Y.C.C., Li-Shien Chen, S.W. Chen, 2004. Automatic license plate recognition, *IEEE Transactions On Intelligent Transportation Systems*, 5: 42-53.

[13] Cui, Y., Q. Huang, 1997. Character extraction of license plates from video. *Proceedings of IEEEConference on Computer Vision and Pattern Recognition,* 502-507.

[14] Dlagnekov, L., 2004. License plate detection using adaboost. Available from http://cseweb.ucsd.edu/classes/fa04/cse252c /projects/louka.pdf.

[15] Sun, J., D. Cui, D. Gu, H. Cai, G. Liu, 2009. Empirical analysis of adaboost algorithm on license plate detection. *Journal of Combinatorial Theory*, 3497-3502.

[16] Choo Kar Soon, Kueh Chiung Lin, Chung Ying Jeng and Shahrel A. Suandi Malaysian Car Number Plate Detection and Recognition System, *Australian Journal of Basic and Applied Sciences,* 6(3): 49-59, 2012 , ISSN 1991-8178.

[17] Kaushik Deb, Hyun-Uk Chae and Kang-Hyun Jo,2009, Vehicle License Plate Detection Method Based on Sliding Concentric Windows and Histogram, *Journal Of Computers*, Vol. 4, No. 8, pp 771-777, August 2009.

[18] Anish Lazrus, Siddhartha Choubey, Sinha G.R., 2011. An Efficient Method Of Vehicle Number Plate Detection And Recognition, *International Journal Of Machine Intelligence*, Volume 3, Issue 3, Pp-134-13

[19] Sandeep Singh ,Chhabada, Rahul Singh, and Atul Negi , 2011, Heuristics For License Plate Detection And Extraction, *World Journal of Science and Technology* 1(12): 63-67

[20] Pablo Negri, Mariano Tepper, Daniel Acevedo, Julio Jacobo, Marta Mejail, 2010. Multiple clues for license plate detection and recognition, *Proceeding CIARP'10 Proceedings of the 15th Iberoamerican congress conference on Progress in pattern recognition, image analysis, computer vision, and applications* Pp 269-276

[21] P.Anishiya, Prof. S. Mary Joans, 2011 , Number Plate Recognition for Indian Cars Using Morphological Dilation and Erosion with the Aid Of OCRs, *International Conference on Information and Network Technology IPCSIT vol.4 (2011)* , IACSIT Press, Singapore

[22] F. Martín, M. García and J.L. Alba, 2002, New Methods for Automatic Reading of VLP's, *Proceedings of IASTED International conference on Signal Processing, Pattern Recognition, and Applications (SPPRA).*

Authors

Dr.R.Radha, aged 44, is working as an Associate Professor in Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, Chromepet, Chennai, Tamil Nadu, India. She has 20 years of teaching experience. She did her research in *Fuzzy Based Data Mining for Effective Decision Support in Bio-Medical Applications.* Her research interest is in bio informatics. She has published 6 international and 1 national publications. She has presented papers in 1 national and 1 State level conferences.

Dr.C.P.Sumathi, aged 46, is working as Head and Associate Professor in Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, Chromepet, Chennai, Tamil Nadu, India. She has 22 years of teaching experience and 7 years of research experience. Her research interests are Image Processing and Neural Networks.. She has 13 international publications to her credit.