License Plate Segmentation Based on Connected Component Analysis

V. Karthikeyan\textsuperscript{1}, V. J. Vijayalakshmi\textsuperscript{2}, P. Jeyakumar\textsuperscript{3}

\textsuperscript{1}Assistant Professor, Dept. of ECE, SVS College of Engineering, Coimbatore, India, Karthick77keyan@gmail.com

\textsuperscript{2}Assistant Professor, Department of EEE, SKCET, Coimbatore, India, vijik810@gmail.com

\textsuperscript{3}Student, Department of ECE, Karpagam University, Coimbatore, India, jeyak522@gmail.com

Abstract. License plate recognition (LPR) plays a major role in this busy world, as the number of vehicles increases day by day, theft of vehicles, breaking traffic rules, entering restricted area are also increases linearly, so to block this act license plate recognition system is designed. License Plate Recognition (LPR) systems basically consist of 3 main processing steps such as: Detection of number plate, Segmentation of plate characters and Recognition of each character. Among this, character segmentation is a most challenging task, as the accuracy of the character recognition relies on the accuracy of the character segmentation. Problems of different lighting condition, adhesion, fracture, rivet, rotation degrades the accuracy of the character segmentation. So in order to overcome these problems and uplift the accuracy of character segmentation various algorithms are developed for this work. Purpose of this paper is to categorize and brief them.

Keywords: License plate, License plate recognition (LPR), preprocessing, Character Segmentation.

Introduction

LPR (License Plate Recognition) is an image-processing technology used to identify vehicles by their license plates. The Vehicle Identification Number (VIN) is a unique identification number for every car manufactured. Unlike registration number, license plate number does not change throughout the lifetime. Moreover, the registration number is not unique to a car since it can be carried forward to another car. Since the 21st century, with social development and improvement of living standards, the number of vehicles is continuously increased, the traffic conditions is worsening, which brought huge pressures to the society and environment. License plate recognition system can solve the various road problems generated by the traffic congestion, thus receiving more and more attention. The vehicle license plate recognition system focuses on the key technologies, which include the license plate region location, license plate character extraction and license character segmentation, license plate character recognition. For different countries the type of license plate, characters used in the plate, climatic conditions vary so to cope with this we should establish a good license plate recognition system with higher performance and accuracy rate. But the recognition step mainly relies on the accuracy rate of the character segmentation. Thus we should strengthen the character segmentation step with more powerful and efficient algorithm. One algorithm could work very well for a given country’s plate
but very poorly for another. So based upon our requirement we should select the algorithms. This paper covers various character segmentation algorithms; also the process, advantages, and disadvantages of those algorithms are discussed in this paper. Various algorithms discussed in this paper are character segmentation based on horizontal projection, vertical projection, prior knowledge of the license plate, inherent characteristics of the character, rule based segment analysis, vector quantization, connected component analysis (CCA), blob extraction, morphological operations, filtering, smearing algorithm, line scanning method, image scissoring algorithm, template matching and Hough transform.

All these methods are discussed in this paper in the following section. And the comparison of the accuracy/performance rates of these algorithms also presented in this paper.

Various methods for character segmentation

1. Blob Extraction based Character Segmentation.

In [12], lack of algorithm description, estimation and result analysis of vertical projection and connected component analysis (CCA) overcome by this algorithm. Also the previous methods are not sufficient to handle the low quality images but this proposed algorithm adopting with the previous methods like CCA, vertical projection and binarization methods. This character segmentation procedure consists of three sub parts. Adjusting the input image and generating a binary image is the first part. In this part Ostu’s method not working well and Niblack’s method shows better result and performance than Sauvola’s method so Niblack’s method is used as the binerization algorithm. Analyzing blobs which is a connected set of pixels in the binary image, and removing the noisy blobs and also merging and splitting the blobs is the second part. In this step it starts with CCA, the non-character blobs, Plate boundaries, small dirties, and unrelated marks or bars are excluded. Correcting the rotated plate images and selecting final seven character blobs is the last part. In this module seven blobs with higher matching scores are selected. When analyzing the failure cases closely this robust character segmentation module achieves 97.2% of success rate. But the drawback of this method is that naturally blob analysis failed in some cases. But here a single parameter for Niblack’s binarization method is present. But the parameter is fine for the general images, but it is not the best one for each image.

Figure 1 Results of the character segmentation steps.
2. **Character segmentation using connected component analysis (CCA).**

In [6], based on pixel connectivity the CCA scans and labels the pixels of a binarized image into components. Every pixel is labeled with a value depending on the component to which it was assigned. The connected components are then analyzed to filter out long and wide components and only left the components based on the defined values. But the drawback [], is that in this method the segmented result using may not contain the exact license plate regions.

![Figure 2 4-neighbourhood and 8-neighbourhood](image)

Figure 2 4-neighbourhood and 8-neighbourhood

![Figure 3 Car images segmented license plates on right](image)

Figure 3 Car images segmented license plates on right

3. **Character segmentation based on the combination of projection and inherent characteristics of the character.**

In [11], preprocessing is strengthened before segmentation stage to improve the accuracy of the character segmentation. Projection based method to segment the characters according to the prior knowledge of the license plate. It consists of two steps: coarse segmentation and fine segmentation. In coarse segmentation vertical projection is passed to the preprocessed plate, several blocks with projection nonzero were obtained. They may be plate characters or noises, they are candidate characters i.e., plate character or noises. Record horizontal start and end locations. Then horizontal projection is applied to each candidate character and record vertical start and end locations. At the end, every candidate character can get a common bounding box. Fine segmentation gives solutions to 3 major problems. Problems and solutions are (i)Coarse segmentation is not ideal so boundaries need adjustment. (ii)For broken characters character merging and splitting must done (iii)The vertical frames may connect with first and last character so it is necessary to separate the plate character and noises. After locating plate, character segmentation will directly affect the quality of character recognition accuracy so this method solves this problem and provides better segmentation. By the combined action of projection and character’s inherent characteristics for tilted characters in horizontal direction there is no need for tilt correction but we can achieve better segmentation. This method solve character bond and break problem and are not sensitive to inclination. So by using this method we can achieve 97% accuracy. But the drawback is that error usually occurs in fuzzy image which is very hard to separate the character from the background.
4. Character segmentation using Hough transformation and the prior knowledge in horizontal and vertical segmentation

In [13], presents a fresh algorithm for character segmentation. This new algorithm uses Hough transformation and the prior knowledge in horizontal and vertical segmentation to overcome the difficulties like image noise, problems in plate frame, rivet, space mark, plate rotation and lighting inconsistency. This method uses information of intensity and avoids the absorption so it overcomes the drawbacks of binarization. The algorithm has three steps: preprocessing, horizontal segmentation and vertical segmentation. For better performance preprocessing is must. Preprocessing step consist of size normalization (plate images are normalized to 160*40 in pixel), determination of plate kind and object enhancement (to improve the quality of images). Then in Horizontal segmentation Hough transformation is used to detect lines. It is hard to obtain the horizontal segment lines for large segmentation by horizontal projection analysis. So Hough transform is preferred. This Hough transformation is used on the midpoints of all subsection lines to remove the wrong subsection lines and combine the correct subsection lines into a whole line. Also the horizontal segment lines of the block with bolt are frequently wrong and can be removed by Hough transformation. This Hough transformation can solve the problem of rivet, rotation, and lighting inconsistency. And then projection based vertical segmentation algorithm is applied using prior knowledge. Thus the use of prior knowledge leads to accurate segmentation and hold back the power of plate frame and space mark. Also segmentation is more accurate and strong than single projection method because of Hough transformation and the prior knowledge. But these methods can't work with some other kind of license plate, like two-row license plate.

5. Character segmentation using line scanning method

In [1], effective character segmentation is performed using line scanning technique. Usually scanning done from left to right of the license plate. The process involves several steps, first the gray scale image is converted into binary image. In order to partition the text on the license plate into lines ‘Lines’ function which uses “clip” function is used. Usually this “Clip” function crops black letter with white background. And then resizing is done. At the end same procedure is repeated on the cropped image till all the characters are segmented.
6. Character segmentation using rule based segment analysis engine

In [8], a segment analysis engine is used to find out the corresponding license plate regions from the edge map images. This work is passed out more than two metro cities in India to produce 2500 ground-truth images. The motivation behind this work is to uplift the accuracy and the efficiency of license plate localization, to overcome the degradation and failure of result due to polluted environment and huge variation in the features of the license plate like different size, color and font of the license plate and to generate a better method to segment the characters of the license plate. For localizing purpose 8-connected edge components are segmented and labeled exclusively using a connected component labeling (CCL) algorithm. But the segmented result using CCL algorithm may not contain the exact license plate regions. So, a rule based segment analysis engine is developed for successful collection of potential license plate regions from the segmented results, using the features: area of the segment (to identify the valid segments), aspect ratio of the segment (to discard the noise segment), vertical edge gradient of the segment (segments under this step are considered as the potential license plate region). This proposed work seems to have a better accuracy in its output and efficiently segment the characters from the license plate. But the drawback of this technique is however well-known vertical edge components visible in some non-plate regions of the images.

7. Image scissoring algorithm based character segmentation:

In [7], image scissoring algorithm is used to identify the license plate characters, license plate is divided into several images that each contain one isolated character. Several steps involves in this segmentation process, first image contrast is done. The Ostu method is used to perform thresholding of plate image then connected components are searched and labeled and the as per the standard height and width all the characters are resized. Since license plates do not have a fixed number of characters for all plates, vertical projection and connected-component method does not segment the license plate, so image scissoring algorithm is used in this paper. Using this method the performance of character segmentation reaches 95%.
8. Morphological and partition based character segmentation.

In [10], stated character segmentation job is very difficult due to several factors like image noise, plate frame, rivet, and rotation and lighting variance. So in order to get a good performance of character segmentation preprocessing stage is significant. Initially, image is filtered and noises are removed. During the threshold processing several small objects that directly affect the segmentation process may grow on the threshold image due to the problems of various lighting conditions, low quality camera and motion effect. So a morphological process which looks for the whole image for small connected elements and remove it. Then to separate the characters that are close with each other, dilation operator is applied to the image. Following this stage, in order to extract the character from the plate partition scanning is projected. In this partition method based on partition value character and background are separated. After conducting several experiments, a Partition value larger than 0.7-0.8 is measured as background, if not it is measured as character. At last the plate is divided into two blocks with digits in first block and letter in the second block. The input color image with the size of 640x480. And tested the images under a variety of lighting conditions and distance. As per this algorithm performs well on different types of vehicles including Iranian car and motorcycle plates as well as miscellaneous conditions. By this method character segmentation achieves accuracy of 94%. But hardly any images of tremendously poor quality attempted more than three probable strips.


In [5], for character segmentation prior knowledge like size of the license plate, size of a character, size of the interval between the characters and the number of characters within a license plate is used. As we have the prior knowledge of the license plate, if there is only one type of license plate this method will be very effective. But projection based segmentation is the
traditional method that is used in majority of the system. Horizontal and vertical projections are used to perform horizontal and vertical segmentation respectively. Horizontal segmentation is used to remove the unwanted top and bottom parts of the license plate image. And vertical segmentation is used to separate the characters that are in the same license plate image. But when projection method is used alone it will leads to problems like rivet, rotation and illumination variance. When these problems appear it will leads to several errors in the character segmentation. So prior knowledge about the license plate is used along with projection method in order to eliminate the problems when projection method is used alone and helps in deleting the wrong horizontal segmentation and combining the correct lines. So character segmentation using prior knowledge of license plate simple and fast.

10. Projection Based character segmentation

In [5], a projection based method and binarization is used for character segmentation. In order to improve the segmentation rate this method considered the relationship among characters. In normal case, for binarization fixed threshold is used so it is difficult to separate the characters during random lighting conditions and a dirt license plate. Thus for block-based thresholding, an extracted LP of size \( P \times Q \) into non-overlapped blocks of size \( M \times N \), where \( M = P/4 \) and \( N = Q/4 \). But even some specific characters like “F” are not binarized it is still sufficient for the extraction of top and bottom boundaries of characters for further processing. After binarization, the top and bottom boundaries of the characters are confirmed, and the horizontal projection is carried out. But even some other least projection points which may hinder the detection accuracy. In order to manage with this, the searching strategy is given with, the imaginary height of the License plate and vertical coordinate of License plate. Initially the bottom boundaries of License plate should locate within the vertical region and then the top boundaries of License plate should locate within the vertical region. At the end the boundary of every character can be resolute by the horizontal and vertical projections. Based on the extracted boundaries, all the characters are cropped from the original grayscale basis and then normalized to 10x20. Then again the cropped grayscale basis will be binarized with its mean. Then from the binarized output the object with utmost area size will pass through connected component processing. This projection based method achieves good accuracy. That the accurate rates of the character segmentation by this method can achieve 95.2

![Figure 8 License plate binarization with various situations](image)

<table>
<thead>
<tr>
<th>Binarized results</th>
<th>Original grayscale source (d)</th>
<th>( B ) (e)</th>
<th>( B' ) (f)</th>
<th>( B'' ) (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal condition  (a)</td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
</tr>
<tr>
<td>Dirty condition (b)</td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
</tr>
<tr>
<td>Lighting condition (c)</td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
<td><img src="image" alt="image" /></td>
</tr>
</tbody>
</table>

Figure 8 License plate binarization with various situations
11. Smearing algorithms, filtering and morphological algorithms based character segmentation

In [9], smearing algorithms, filtering and some morphological algorithms are used for character segmentation. License plate is segmented into its ingredient parts and the characters are obtained individually in the segmentation process. This segmentation process consists of four steps. Filtering the image for enhancement and removing noises and unrelated parts are the first step. Then separating the characters that are close with each other by dilation operation is the second step. Finding the character regions by horizontal and vertical smearing is the third step. Cutting the plate characters is the final step of this process. This algorithm shows greater performance. Character segmentation method achieves accuracy of about 96%.

12. Template Matching and projection based character segmentation.

In [2], for license plate character segmentation a new method has been proposed. The originality lies in this case in its treatment which not only combines the projection and template match, but also improves the techniques. In order to overcome the problems of different light strength in different time, partial light shadow in the image and adhesive characters new method is used. Projection method is done in the horizontal direction to detect the edge of the single license plate in the vertical direction. Refined character segmentation is done using template matching as templates designed in the literature of other works are difficult to meet the images that are shot random. These templates are designed on the basis of prior knowledge about the license plate, and take similarity measures to match the image segmented by projection. So by this method accuracy requirements are satisfied along with proper segmentation. Image normalization is necessary due to the usage of template matching method. Then to improve the accuracy of image segmentation minimum distance classifier is used. As single method is very difficult to perform
the task so a combination of algorithms Thorny floating-point operations are restricted by the template matching method. It is only the use of arithmetical operations. The overall algorithm reduces the complexity, also has great robust. The algorithm is simple and effective. As per this method, overall the segmentation rate is 98.8% and time for performing the segmentation is about 10ms.

13. Vector quantization based character segmentation

In [3], presents a fresh algorithm for license plate character segmentation problems by using local vector quantization. To extract the correct boundary and segment the plate region the binary split tree is used for vector quantization which becomes ready for the optical character recognition. The output of the system is the bounding box for the license plates detected and their segmented characters. Gabor filter performs well at detection, but its segmentation performance is poor. So nonlinear vector quantization is applied to remove the false alarms and to segment the license plate characters to its accurate boundary, Vector quantization is a process of allocating pixel values in one of a finite number of vectors. These vectors are determined in such a way using binary split tree method so such that quantization error is minimized. And at the end, connected component analysis is applied to the quantized image to gain the character segments. This method is effective segmenting the plate characters. In this method the performance rate of this License plate segmentation is 94.2%. But the drawback of this method is computationally very expensive.

![Figure 13 segmentation results of various sizes and forms of license plate](image)

**Performance/Accuracy Rate Analysis of Different Vehicle License Plate Character Segmentation Techniques**
Table 1

<table>
<thead>
<tr>
<th>S.No</th>
<th>Algorithms used for character segmentation</th>
<th>Percentage of performance/ accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blob Extraction [12]</td>
<td>97.2</td>
</tr>
<tr>
<td>2</td>
<td>Combination of projection and inherent characteristics of the character [11]</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>Image scissoring algorithm[7]</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>Morphological and partition based method[10]</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>Projection based method[5]</td>
<td>95.2</td>
</tr>
<tr>
<td>6</td>
<td>Smearing algorithm, filtering and morphological algorithm[9]</td>
<td>96</td>
</tr>
<tr>
<td>7</td>
<td>Template matching and projection based method[2]</td>
<td>98.8</td>
</tr>
<tr>
<td>8</td>
<td>Vector quantization[3]</td>
<td>94.2</td>
</tr>
</tbody>
</table>

Conclusion

This paper presents various algorithms for license plate character segmentation. License Plate Recognition (LPR) system plays a vital role in Intelligent Transportation System (ITS). It controls various problems like traffic congestion, and used in various applications like border crossing, law enforcement, parking, automatic toll gates etc., this LPR system comes across with various problems like lighting condition, rivet, various size, form, break in the plate, so in order to produce an effective result the recognition rate must be good. So, as we know recognition rate mainly rely on the accuracy rate of the character segmentation. If segmentation part fails or the output of character segmentation is not fair then it will have some effect on the recognition output. Thus to generate a valuable output we presented various algorithms for character segmentation. Among all the above methods projection based method is the traditional method used in various systems, but this method is not much effective alone. So if the characteristic of the license plate and prior knowledge about the license plate is used along with projection method we can get a better accuracy and effective result. Character segmentation accuracy can also improved by strengthening the preprocessing before segmentation process. Algorithms presented in this paper have both positive and negative sides. Certain algorithm works well in one country but not for other countries, so based upon our requirement we should select the algorithm for our system.

References


[14] Adebayo daramola.s1, e. adetiba1, a. u. adoghe1, j. a. badejo1, i. a samuell1 and t. fagorus


