Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

Research on Blob Analysis Algorithm-based Segmentation of Characters in License Plate

Chunhui Han, Zhao Chen *

School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China
* *chenzhao970316@163.com*

ABSTRACT. *A license plate character segmentation method based on Blob analysis algorithm is proposed to improve the precision of license plate character segmentation. Firstly, the Blob analysis algorithm is combined with the prior knowledge of license plate characters to perform rough segmentation of characters. Secondly, using vertical projection method to achieve the fine segmentation of the characters. Experimental results show that the algorithm can effectively and practicability solve the problem of segmentation and characters cohesion of disconnected Chinese characters.*

KEYWORDS: *Image processing; License plate character segmentation; Blob analysis algorithm; Connected domain; Vertical projection method.*

1. Introduction

Licesnse plate recognition (LPR) technology [1-2] is the core part of intelligent traffic system (ITS). However, license plate character segmentation (LPCS) is a vital part of vehicle license plate recognition (LPR) system, which directly affects the accuracy of subsequent license plate recognition (LPR) system. Common license plate character segmentation methods include connected domain-based segmentation method [3], vertical projection-based segmentation method [4], template matching-based segmentation method [5] and clustering-based segmentation method [6]. The segmentation method based on connected domain marks connected numbers and letters in binary images, so as to achieve accurate segmentation, but the characters of the disconnected Chinese characters, glued and broken characters cannot be processed well. The segmentation method based on vertical projection is to segment characters by analyzing the peaks and troughs of the vertical projection of the license plate gray scale, the algorithm is simple and the running time is short, but the segmentation results for various situations such as license plate tilt, license plate blur, character sticking and character break are not at

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

the desired level. The segmentation method based on template matching is segmented by the template characteristics of the license plate itself, which is less affected by noise, but the segmentation of the left and right borders needs to be improved [7]. The advantage of clustering-based segmentation method is that the algorithm has good flexibility, and the disadvantage is that the efficiency of the clustering algorithm needs to be improved [8].

According to the above, this paper proposes a license plate character segmentation method based on Blob analysis algorithm. Firstly, the license plate image is preprocessed (including the grayscale, binarization, denoising and frame removal of the license plate image) . Secondly, the grayscale mutation region of the binary image connected domain is analyzed and processed. In the process of processing, the line of the imaged is operated, not the single pixel is analyzed one by one, obtain the geometric attributes of the license plate characters (including the minimum circumscribed rectangle, aspect ratio, etc.) to roughly segment the characters. Finally, the vertical projection method is used to segment the conglutination characters, which can effectively solve the problem of conglutination and ensure the accuracy of the entire character segmentation algorithm.

2. Image Preprocessing

2.1 Gray Processing

The color image is generated into a black-and-white image with 256 gray levels. As shown in Figure 1, the color of each pixel of the color image is composed of three components of R, G and B, which are red, green and blue. Among the three colors, the human eye is the most sensitive to green and the lowest sensitive to blue. Therefore,the weighted average method was used to gray the color image and obtain the best grayscale image effect. The weight formula is shown in (1), and the grayscale image is shown in Figure 2.

$$f(i,j) = 0.11B(i,j) + 0.3R(i,j) + 0.59G(i,j) \quad (1)$$

Figure 1 Color image of license plate

Figure 2 Gray image

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

2.2 License plate binarizatio

The binarization of the license plate is to separate the characters and the background in the license plate [9], so as to make the character outline more clear and facilitate the segmentation of characters. The Blob analysis algorithm analyzes the image after the binarization process. Therefore, the appropriate threshold T is required to binarize the grayscale image. If the threshold T is too large,the break of the character is prone to occur, if the threshold T is too small, it will cause adhesion of the characters. In view of this problem, this paper uses the Ostu's to binarize grayscale images.

The principle is as follows:

Step 1: Firstly, an initial threshold T is used to segment the image, according to the threshold value T, threshold T is used to divide the gray value of pixels in the image into foreground and background, which are represented by C_1 and C_2 respectively. The threshold of C_0 is smaller than T, and the threshold of C_1 is greater than T.

Step 2: Set the total number of pixels of class C_0 be W_1, the total number of pixels of class C_1 be W_2, the average grayscale of their pixels is represented by μ_1 and μ_2 respectively. The total average gray level of the image is $\mu_T = W_1 \times \mu_1 + W_2 \times \mu_2$.

Step 3: Traversing T from the minimum grayscale value to the maximum grayscale value, when T makes the variance $\sigma^2 = W_1 \times (\mu_1 - \mu_T)^2 + W_2 \times (\mu_2 - \mu_T)^2$ maximum, T is the optimal threshold for the segmentation. The binarization graph based on Otsu method is shown in Figure 3.

Figure 3 Binary image

In order to avoid the influence of shadows and external noise on the segmentation effect, the binarized license plate image is processed by morphological operation. For a given image I , the structural element S , the morphological opening and closing operations are shown in equation (2). If the structural element S is selected properly, the opening operation can eliminate the burrs on the contour and play a role in smoothing the outline of the image. The closed operation can eliminate the small holes in the license plate area, and fills in small gaps and contour gaps. The binarized license plate is shown in Figure 4 and Figure 5 after being processed by the open operation and the closed operation.

$$\begin{cases} I_{open} = \left(I(x,y) \Theta S\right) \oplus S \\ I_{close} = \left(I(x,y) \oplus S\right) \Theta S \end{cases} \tag{2}$$

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

Figure 4 Open operation

Figure 5 Closed operation

2.3 Removal of Borders

The existence of license plate borders and rivets can affect the accurate segmentation of license plate characters. The removal of the border of this paper mainly adopts the row scanning algorithm. Through testing, when *L*=Width/7 (the Width is the width of the license plate) and the number of jumps 13 is the threshold value, the progressive scan effect of the license plate image is best. After the line scan of the license plate image, if the length of the obtained line segment is greater than *L* and the number of jumps obtained is less than 13, the scanned area is considered to be the upper and lower borders and the rivets, Remove the upper and lower borders and the rivets as shown in Figure 6.

Figure 6 Remove the upper and lower borders and rivets

3. Blob Analysis Algorithms

Blob analysis algorithm is a feasibility analysis for the same connected domain pixels [10]. It is a basic method for analyzing and processing closed target shape, this connected domain is called Blob. Blob analysis can provide the location, number, shape and direction of the region in the image, as well as the topological structure between the relevant regions. It is commonly used to extract and classify graphic features from target image, such as Chip pin detection. According to the idea of the Blob analysis algorithm, the Blob analysis is based on the connectivity of the whole license plate. The common license plate standard in China is unified and sonsisting of seven characters, except that the first character is Chinese characters, the rest is composed of numbers and letters. However, the numbers and letters are

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

connected. Therefore, the license plate can be segmented by analyzing the connectivity the connectivity between numbers and letters.

The algorithm steps are as follows:

Step 1: The license plate image after binarization is defined as I, whose rows and columns are denoted by X and Y . Sets m=0; Takes the middle row x_0 as the scanning line, where x_0=Y/2; Starts the cycle from y_0 =1 until $y_0 = Y$; If $I(x_0, y_0) \neq 0$, then go to step 2.

Step 2: Find the coordinate position where the connected domain of (x_0, y_0) is located, so that the pixel value of the corresponding coordinate position is 0 , and return the row and column coordinates (x, y) where the connected area is located.

Step 3: Calculate the basic characteristic values of the region by through the row and column coordinates (x, y) of the connected domain , which are the width w and height h . If $w >$ y/5 or $h < x/4$, then the connected region is not a character region. Then go to step 1 and proceed to the next round, otherwise go to step 4.

Step 4: Let $m = m + 1$. Run the following formula: blob(m).$w = w$; blob(m).$h = h$; blob(m).z = [(min(x) + max(x))/2 , (min(y)+max (y))/2] ; blob(m).P=[x , y] ; Go to step 1 for the next round.

Step 5: Return Blob and m.

4. Character Segmentation

4.1 Segmentation of Connected Characters

The license plate model is A1A2•A3A4A5A6A7, the idea of character segmentation is to use Blob analysis algorithm to obtain the location, area and parameters of the external rectangle of Blob(A2)~Blob(A7) in the license plate. Then, according to the location of the Blob and the parameters of the external rectangle, combined with the basic height of the license plate characters and the proportion of the middle distance between the adjacent characters, the second to seventh characters of the license plate are segmented [11].

Table 1 is the result of Blob analysis on Figure 7(a), the column values and row values in the table represent the position information of the Blob, indicating the coordinates of the starting point, width and height are Blob's width and height respectively. From Table 1, it can be seen that nine connected regions can be obtained after Blob analysis of Figure 7(a) and expressed in red rectangular boxes, the results are shown in Figure 7(a). In this paper, in the process of connected character segmentation, the connected domain is judged by the prior knowledge of license plate characters. Therefore, we can judge that Blob9 is a non-character region, which can be discarded in the process of segmentation.

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

Table 1 Blob analysis data example

Blob	Column value	Line value	width	high
Blob1	1.5	2.5	44	40
Blob2	5.5	48.5	36	40
Blob3	57.5	1.5	44	88
Blob4	136.5	1.5	46	88
Blob5	192.5	0.5	48	89
Blob6	250.5	1.5	47	89
Blob7	307.5	1.5	47	89
Blob8	366.5	1.5	46	80
Blob9	420.5	0.5	8	91

The segmentation result of the connected characters is shown in Figure 7(a) is a labeled connected domain, it can be seen that there is a vertical frame noise region. Figure 7(b) is the result of segmentation by the traditional connected domain segmentation method, it can be seen that the traditional connected domain segmentation can not solve this situation very well. The Figure 7(c) is the segmentation method of this paper, which can effectively solve the existence of vertical border noise and make the character segmentation more accurate.

(a) Connect domain tags

(b) Traditional segmentation results

(c) The segmentation results of this paper

Figure 7 Concatenation of characters

4.2 Segmentation of Disconnected Characters

The first character of the common standard license plate in China is Chinese character. Because of the disconnection of Chinese character, when marking the

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

connected area, they will be marked as multiple areas, such as“吉, 苏,鄂”(Chinese characters) .etc. For such problems, the segmentation of disconnected Chinese characters in this paper is mainly based on the prior knowledge of license plate characters. Through Blob analysis, we can know the coordinates, length and width of the upper left corner of the second character of the license plate, mark the starting position *L* of the second character. According to the average width Wave of the divided characters, we can determine the starting point *L1=L*-Wave, and the static boundary method is used to segment Chinese characters, above-mentioned the Wave can be expressed as (3). The segmentation of the non-connected Chinese characters is shown in Figure 8, and can be seen from the figure that the method of this paper can effectively solve the segmentation of non-connected Chinese characters and make the segmentation result more accurate.

$$Wave = \frac{1}{6}\sum_{i=2}^{7} blob(X_i).w \tag{3}$$

(a) Connect domain tags

(b) The segmentation results of this paper

Figure 8 Disconnected characters

4.3 Segmentation of Sticky Characters

Due to the presence of stains between the characters on the license plate, or the inclination of the shooting angle leads to the deformation of the obtained license plate image, which may lead to the character adhesion. When connecting domain marks, the sticky characters will be marked together, which leads to the inaccurate segmentation of the characters. In order to solve this kind of problems, the marked binary image is projected vertically. Because the projection value of the adhered part is smaller than that of the other parts, therefore, the Sticky characters can be segmented by calculating the projection value of each projection. Through experiments, the algorithm in this paper is compared with the Traditional connected domain marking method. Through comparison Figure 9 we can see that the method in this paper can accurately segment the sticky characters.

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

(a) Character sticky license plate

(b)Traditional connected domain segmentation method

(c) The segmentation results of this paper

Figure 9 Comparison of different segmentation methods

5. Experimental Results

The test environment of this paper is Intel(R) Pentium(R) 1.8 GHz CPU, 2G memory, the program running platform is Matlab R2012a. Using 200 tilt correction successful license plate images for the segmentation experiment of this paper, the number of license plates correctly segmented is 196, and the segmentation accuracy is 98%. The main reason for the failure of segmentation are the serious pollution of the license plate handwriting and the illegible handwriting. Due to space reasons, the partial segmentation results given in this paper are shown in Figure 10.

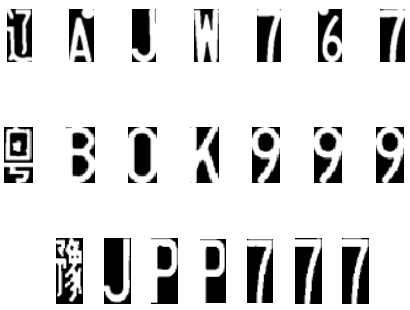

Figure 10 Split the results

In order to illustrate the effectiveness of this algorithm, experiments are carried out to compare the Traditional connected domain marking method, Vertical projection method and the time and accuracy of character segmentation in this algorithm. It can be seen from Table 2 that test time of the algorithm is obviously

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

lower than that of Traditional connected area marking method and Vertical projection method, and moreover the accuracy rate is nearly 3 percentage points higher than that of Traditional connected area marking method and 4 percentage points higher than that of Vertical projection method. This shows that the algorithm of this paper can effectively improve the accuracy of character segmentation and shortens the time of character segmentation.

Table2 Character segmentation comparison test

Segmentation method	Test the picture/Amplitude	Time consuming/s	Accuracy/%
Traditional connected domain labeling method	200	9.88	95.6
Vertical projection method	200	7.79	94
Algorithm of this paper	200	6.09	98

6. Conclusion

The effective segmentation of license plate characters is a key part of the license plate recognition system. In this paper, based on the Blob analysis algorithm, the license plate character segmentation is firstly rough segmentation of the license plate characters. In the rough segmentation stage, the Blob analysis is used to effectively mark the connected domain, and the minimum circumscribed rectangle and the aspect ratio of the connected domain are obtained, combined with the license plate characters. The prior knowledge of the Chinese characters is accurately segmented, and the interference of vertical border noise is removed. Secondly, the vertical projection method is used to finely segment the license plate characters, which effectively solves the problem of character adhesion. Compared with the traditional connected domain segmentation algorithm and vertical projection algorithm, the segmentation accuracy of this algorithm is significantly improved.

References

[1] Geng Qintian and Zhao Hongwei (2013). License plate recognition based on fractal and hidden Markov feature. Optics and Precision Engineering, vol.21, no.12, p.3198-3203.

[2] Guo Runhua, Su Tingting and Ma Xiaowei (2013). License plate recognition system using a BP neural network. Journal of Tsinghua University (Science and Technology) , vol.53, no.9, p.1221-1226.

[3] Shi Longzhao and Wang Kai (2016). Complex license plate character segmentation based on connected component. Computer Engineering and Design, vol.37, no.8, p.2138-2142.

ISSN 2616-5775 Vol. 2, Issue 2: 39-48, DOI: 10.25236/AJCIS.010037

[4] Ran Lingfeng (2012). An Algorithm of Character Segmentation based on Vertical Projection for License Plate. Communication Technology, vol.04, no.45, p.89-91.
[5] Pei Mingtao, Wang Yongjie and Jia Yunde (2014). License plate character segmentation based on multiple scale templates matching and Par-based model. Transactions of Beijing Institute of Technology, vol.34, no.9, p.962-965.
[6] Li Xu, Su Shuchang, You Yucai and Zang Sanyuan (2012). Segmentation method for personalized American car plate based on clustering analysis. Journal of Zhejiang University(Engineering Science), vol.46, no.12, p.2155-2159.
[7] Zheng Chengyong and Li Hong (2010). Vehicle license plate characters segmentation method using character's entirety and blob analysis. J. Huazhong Univ. of Sci.& Tech. (Natural Science Edition), vol.38, no.3, p.88-91.
[8] Wu Zhipan, Zhao Yuelong, Luo Zhongliang and Du Huaying (2017). License plate recognition technology based on PSO-BP neural network . Acta Scientiarum Naturalium Universitatis Sunyatseni, vol.56, no.1, p.46-51.
[9] Lu Jingyi, Xue Zheng, Shao Keyong and Li Miao (2015). License plate character segmentation based on improved connected component and vertical projection. Techniques of Automation and Applications, vol.34, no.12, p.93-97.
[10] Xu Min, Tang Wanyou, Ma Qianli , et al. (2011) . Research of printing defect on-line detection based on Blob algorithm. Packaging Engineering, vol. 32, no. 9, p.20-23.
[11] Zhang Yuhui, Wu Jiangmei , Sun Ying and Liu Haimeng (2013). Locating of human hand based on Gaussian mixture model and blob analysis. Science Technology and Engineering, vol.13, no.31, p.9371-9374.