Research on the Algorithm of Locating and Cutting in License Plate Character Extraction

Bingjie Yuan\textsuperscript{a,1st}, Yuqing Yang\textsuperscript{b,2nd}

University of South China, Hunan Province, China
\textsuperscript{a}2284958574@qq.com, \textsuperscript{b}1143622724@qq.com

Abstract—To improve the real-time and accuracy of intelligent vehicle character recognition algorithm, this paper mainly studies a new method for license plate location and character cutting: license plate location is carried out by combining HSV color space recognition with Sobel edge detection. The outer rectangle of the outermost contour of the character is used to judge the character boundary and segment the character. First, the original picture is converted to the HSV space, then the color information of the license plate is combined with the edge information to locate the license plate roughly, and finally the specific position of the license plate is determined according to the aspect ratio of the license plate. The location of the license plate is extracted for character segmentation preprocessing. Then, by traversing, the contour which meets the aspect ratio condition is found and the corresponding circumscribed rectangle is taken. The coordinate information of the rectangle is extracted and stored as a matrix, which is the basis of character segmentation. It is proved that this method has better accuracy and reliability than the traditional methods.

1. Introduction
Automatic license plate recognition technology plays a very important role in modern intelligent traffic management system. It has a broad application prospect\textsuperscript{[1]} and can be used in various fields such as intersection electronic police, highway automatic toll collection, parking lot safety management. License plate recognition mainly includes three parts: license plate location, character cutting and character recognition. This paper mainly improves two parts: the license plate location and character cutting, using HSV color recognition combined with Sobel edge detection to locate the license plate and a segmentation algorithm based on contour detection to cut characters. The experimental results show that with the proposed algorithm using, the accuracy and real-time of license plate location and character cutting are greatly improved, which plays an important role in the correct recognition of subsequent license plate characters.

2. License plate location

2.1. Image preprocessing
The original image is changeable and the quality is different, so it is necessary to preprocess the image before extracting the features of the image. The main purpose of image preprocessing is to improve the quality of license plate image, so that to improve the accuracy of recognition, to restore useful real information\textsuperscript{[2]}, to enhance the testability of relevant information and to simplify the data we need to to the maximum extent.
2.2. Sobel Edge Detection Algorithm

It is a comparative mature method to extract license plate by the help of edge contour information of license plate. When selecting the edge detection algorithm, both the accuracy of edge positioning and the noise resistance should be considered. When both aspects are considered, appropriate compare [3] should be made. Among many edge detection algorithms, Sobel operator is used in this paper. Because the Sobel operator adopts the local average mode, which has a certain inhibitory effect on the noise.

The edge detection algorithm based on Sobel operator first grayscale the preprocessed image, then uses the Sobel operator to find the edge (as shown in Fig 1(a)). At this time, many other interference edges will be found in addition to the edge of the license plate. Converting the searched gray image to a binary image to reduce the count and using closed operation to bridge the small crack (as shown in Fig 1(b)), then the license plate characters will be connected into a connected domain so as to obtain a rectangular contour, and do a geometric condition on the contour to gain the possible license plate candidate region.

![Figure 1. Sobel edge detection operate](image)

2.3. HSV Space color recognition

The model that combines three color components of hue, saturation and brightness into color space is called HSV color space [4]. Since the RGB color space does not match people's subjective judgment on color similarity, the perceived difference between two colors cannot be represented by the distance between two color points [5-6]. HSV color space is closer to people's subjective understanding of color than RGB color space [7]. Therefore, this paper divides color images based on HSV color space.

![Figure 2. HSV space color recognition](image)

First, it converts the preprocessed image into the HSV color space and getting the value of each component of the H、S、V of the picture. Setting a reasonable threshold range according to the color of the license plate, and traversing each pixel of the converted image, and then judging whether the value of the three components of the H、S、V is to be within the pre-set threshold range. If matched, set to 255, white; Otherwise, set to 0, black. Through the above operation, the binary image of license plate coarse location can be obtained, as shown in Fig 2, where (a) is the image converted to HSV space, and (b) the binary image after color recognition.
2.4. Color recognition combined with edge
Using a single method to identify the license plate cannot be accurate in many cases. It is difficult to locate the license plate accurately under complex background by edge detection, but only color recognition cannot locate the license plate of blue vehicle and the method is easy to be affected by illumination. By combining the advantages of the two methods, the purpose of accurate license plate location in various scenarios is achieved in this paper. The image of the closed operation after Sobel detection and color recognition are then binarized, as shown in Fig 3 where (a) is the image after color combined with edge detection ;(b) the image after closed operation ;(c) the license plate location image ;(d) the license plate image after accurate positioning.

![Figure 3. Color Recognition Combined with Edge Detection](image)

3. CHARACTER CUTTING
License plate image character cutting is the premise of next character recognition, because the quality of character cutting will directly affect the recognition rate of next character. At present, the main methods of this research are as follows :1 uses the fixed aspect ratio of characters and the spacing between characters on the license plate as prior knowledge to cut characters [8].

Based on previous studies, this paper proposes a character cutting algorithm based on contour detection.

The domestic license plate consists of seven characters and the size of each character is fixed, so we can cut the characters by judging the geometric characteristics of the outermost contour rectangle of the characters. Since Chinese characters are not necessarily fully connected, it is very likely that there will be cut wrongly of the outermost contours of Chinese characters like "沪" and "川". The character segmentation algorithm based on contour detection proposed in this paper will combine the actual situation of each cutting and the geometric characteristics of the contour to limit the cutting conditions and reduce errors. The specific process is as follows:

1) First, take the outermost outline of each character in the license plate image (the red line in Fig 4(a)), and take the circumscribed rectangle. Our license plate has a standard geometric structure. The length, width, and spacing of characters are fixed. 440 mm in length and 140 mm in width [9]. The width and altitude of each character is 45 mm, 90 mm, respectively and aspect ratio is 1 : 2. It is possible to
determine whether the length, width and aspect ratio of each rectangle fall within the set threshold so that to determine whether the region is where the characters are located. Extracting the eligible rectangle after traversing all outermost contours Shape.

2) Save the coordinate information of four points in each rectangle to matrix A, which is defined as:

\[
A[m][n] = \begin{bmatrix}
A_{11} & A_{12} & A_{13} & A_{14} \\
A_{21} & \ddots & A_{23} & A_{24} \\
A_{(m-1)1} & \ddots & \ddots & \ddots \\
A_{m1} & A_{m2} & A_{m3} & A_{m4}
\end{bmatrix}
\]

where, the \( m \) represents the number of eligible rectangles; the \( n \) represents the four coordinate information of the rectangle 1,2,3,4 correspond to the left coordinate value, the right coordinate value, the upper coordinate value, and the lower coordinate value, respectively. Hence, \( A_{m1} - A_{m4} \) denote respectively the left coordinate, the right coordinate, the upper coordinate and the lower coordinate of the outer rectangle of the \( m \)th outermost contour detected.

3) Since the order of contour detection is not in the order from left to right of the image, each row of the matrix A needs to be sorted from the largest to the smallest from the first column which is the left coordinate value of each rectangle to obtain the matrix B. At this time, each row vector of the B matrix corresponds to the detected contour information from left to right. The expression of B matrix is:

\[
B[m][n] = \begin{bmatrix}
B_{11} & B_{12} & B_{13} & B_{14} \\
B_{21} & \ddots & B_{23} & B_{24} \\
B_{(m-1)1} & \ddots & \ddots & \ddots \\
B_{m1} & B_{m2} & B_{m3} & B_{m4}
\end{bmatrix}
\]

When the number of rows B of the matrix is 7, it means that 7 characters are detected (Fig 4(b) Green box), and then can be cut according to the four coordinate information of each character; if the number of rows B the matrix is not equal to 7, the most possible reason is that the Chinese characters are not connected. Thus, the geometric features of the outermost rectangle can not conform to the set threshold, which can be divided into the following two cases:

a) The number of rows of the matrix is 6 (as shown in Fig 4(c)). The row vectors B of the matrix correspond to the coordinate information of the second character to the seventh character in the license plate respectively. In this case, the upper and lower boundaries of the Chinese character region are set to the second character region of the license plate. And the left boundary \( X_1 \) of the Chinese character region is set to 0, the right boundary of Chinese characters can be obtained according to the aspect ratio of characters \( X_2 \), which should be satisfied:

\[
X_2 = X_1 + (B_{14} - B_{13}) \times \text{rate}
\]

Where, the \( B_{13} - B_{14} \) represents the upper and lower boundary of the second character region of the license plate respectively. The rate represents the ratio of character width to height. The width of each character of our standard license plate is 45 mm, the height is 90 mm, that is, the rate value is 0.5.

So, the four coordinates of the first character rectangular region can be expressed as:

\[
\text{char}[1][4] = [a_{11} - \Delta \text{gap} \quad a_{11} - \Delta \text{gap} \quad a_{13} \quad a_{14}]
\]

The line matrix is inserted into the first line of the matrix B and the character is cut according to the character coordinate information stored in the matrix (see Fig 4(d)).

b) The number of rows of the matrix is greater than 7 (as shown in Fig 4(e), the m-5 to m lines of the matrix correspond to the coordinate information of the second character to the seventh character in the license plate, Lines 1 to m-6 are wrongly detected rectangular coordinate information due to the disconnection of Chinese characters. In this case, the method used in this paper is to set the upper and
lower boundaries of the first character region to $B_{13}$, $B_{14}$. If the left boundary of the first character area is set to $X_1$, $X_2$ the right boundary shall be satisfied:

$$X_2 = B_{[m-6][2]}$$

(5)

So, the four coordinates of the first character rectangular region can be expressed as:

$$char[1][4] = [B_{11} B_{[m-6][2]} B_{13} B_{14}]$$

(6)

It removes lines 1 to m-6 of the matrix B, and inserts matrix char to the first line, and then cuts characters based on the information stored in the Rect1 matrix (see Fig 4 (f)).

![Character segmentation images](image)

Figure 4. character segmentation

4. CONCLUSION

This paper mainly discusses the main problems of license plate location and character segmentation. A method of HSV color recognition combined with Sobel edge detection is proposed for location problem. This method enhances the adaptive ability and robustness of the system, and greatly improves the accuracy of license plate location, which has strong applicability. A character cutting algorithm based on contour detection is proposed for character segmentation, which reduces the error segmentation of Chinese characters compared with other traditional methods.

REFERENCES

[1] Zhuge bin, Zhou Heqin. Algorithm for Car License Plate Location Based on Fractal Dimension of Waveforms[J]. Pattern Recognition and Artificial Intelligence, 2006,19(06):818-824.

[2] Zhang Lu, Lin Kaisi. Research on Skew Distortion License Plate Correction Based on Radon Transform [J]. Jining Normal University, 2020,42(05):32-35.

[3] Zhang Xuelan, Li Jun, Xin Peihong, et al. Method for Paper Defect Image Detection Based on Prewitt Operator and Morphology [J]. Paper and Paper Making, 2012,31(08):25-28.

[4] SHAN B M. License plate character segmentation and recognition based on RBF neural network[C]. Second International Workshop on Education Technology and Computer Science (ETCS), 2010.

[5] Xu Ying. Research on Seawater Pearl Detection and Recognition Technology Based on Information Entropy and Feature Extraction [D]. Hainan University, 2013.

[6] Huang Xiaoming, Gao Chenqiang, Tian Yangyang. Text Detection in Natural Scene Images [J]. Journal of Chongqing University of Posts and Telecommunications, 2015,27(05): 700-705.
[7] Yuan Weikang, Murat Hamit*, et al. Research of Xinjiang Uygur Herbal Medicine Image Based on Color Moment Feature[J]. Bulletin of Science and Technology, 2015, 31(03): 29-33.
[8] YAO Wenfeng, ZHEN Tong, et al. Research on technology of segmentation and recognition of license plate character[J]. Modern Electronics Technique, 2020, 43(19): 65-69.
[9] Wang Meiqin, Yi Min, Guan Caizhong, Wang Mingyi. Design of Automatic License Plate Recognition System[J]. Electronic Instrumentation Customers, 2020, 27(10): 5