

A New License Plate Character Segmentation Algorithm Based on Priori Knowledge Constraints

Liu Yongjun, Yan Jiao, Xie Conghua, Zhang Mingxin

¹Changshu Institute of Technology, Jiangsu Changshu 215500, China
yongjun1981@126.com

Abstract

The license plate character segmentation is one of the three key technologies of the automatic license plate recognition system and the character segmentation is the foundation of the character recognition. The large amount of calculations and a long time for processing are the drawbacks of traditional character segmentation algorithm based on the connected domain character segmentation algorithm, so this paper presents an improved algorithm, which makes full use of priori knowledge for character initial segmentation, and then realizes complete character segmentation based on the connected domain, mean while, this paper improves the traditional iteration binarization threshold algorithm by the priori knowledge, reducing iteration times. The experimental results show that the character segmentation algorithm provided by this paper greatly cuts down the processing time and meets real time requirement under the premise of accurate extraction of license plate character.

Keywords: character segmentation; binarization; connected domain; priori knowledge

1. Introduction

License Plate Recognition (LRP) is very important in the study of Intelligent Transportation System (ITS), which involves Pattern Recognition 、 Computational Vision 、 Digital Image Processing and other theoretical knowledge and can be widely used in highway Electronic Toll Collection, a large stop Yard management, the Intelligent Community vehicle access control and many other fields. Generally speaking, a complete License Plate Recognition System Includes four major components: Image Capture, license plate location, Character-delimited, Character

Recognition, mainly from the license plate recognition technology starting core, focusing on the license plate character segmentation algorithm.

Under the precondition for accurate positioning plates, character segmentation's results have a direct follow-up to determine the accuracy of character recognition. The main task of Vehicle Plate Characteristics Segment is to locate the license plate characters within split up one by one in order to prepare for the subsequent character recognition step. Since most of the license plate image captures in open natural environment, influenced by light, plus the license plate itself dusts contamination and other reasons, it is difficult to find a common segmentation method, common main Vertical Projection method ^[1], Template matching method ^[2] and Interconnected Domain method ^[3, 4]. That is the number of pixels in the Vertical Projection method to calculate the horizontal direction each column belongs to the license plate characters, so the character will get a local minimum gap characters in horizontal projection, so the correct position of the character should be divided in the vicinity of the local minimum. Vertical Projection method program logic design is simple, short, easy to implement, but the characters on the license plate adhesion problems handling is not very satisfactory, it is lack of some presence when address some of the characters are not in communication. Template Matching method makes full use of the width of characters in the license plate arrangement rule features, design plate template string, the character position is determined by template matching, although this method has a good character segmentation, but this method is very complex, processing time long real time inferior, in addition to how to design a matching template is a difficult. Interconnected Domain method is determined by the method of scanning a number of adjacent pixels of the target pixel according to a certain relationship between the target pixel and the criteria for the communication between the adjacent pixels and

marking, thereby extracting the connected area of each separation. The Interconnected Domain method to noise, image tilt is not sensitive for identifying a large number of connected domain image, which has been widely used. However, the traditional method of top-down connected domain, from left to right to start scanning the license plate image after binarization, when scanning to the first white pixel, the pixel seeds as a regional growth 8 direction or 4 direction chain code tracking, tracking the results as an array to store, connected domain extraction method needs to traverse all the images pixel, the computation time is longer, it is difficult to meet the needs of real-time systems.

In response to these problems, in this full use of China's current laws and universal license plate characters arranged in geometric features, combined with the known prior knowledge, we propose a character segmentation method based on improved connectivity domain. Based on prior knowledge and background of the license plate character analysis, improvement of Iterative Solutions image optimal segmentation threshold algorithm, making the license plate image after binarizing more clarity and precision; appropriate use of a priori knowledge of the traditional connected domain France adaptive improvement, making the characters more accurate segmentation of various quality license plate images and adaptable.

2. Algorithm principle

2.1 Pretreatment

After a license plate after plate positioning the extracted image are often due to the shooting angle image acquisition devices, lighting and other factors, it can't be directly character segmentation, the need for Pretreatment operations necessary before division, pretreatment for license plate characters segmentation and character recognition are very important and a good preprocessing algorithm can effectively improve the accuracy [5]. Preprocessing operations primarily relates the size of the normalized image, color image binarization, and remove the license plate frame and rivets and other processing steps.

2.1.1 Size normalization

Due to the distance when shooting the image acquisition equipment and vehicles between the angle and the vehicle itself is different volume size when shooting fast-moving vehicles plus other reasons, so that the license plate

captured image size of the final dimensions are different, which gives the license plate character segmentation brought a greater adverse effects. Based on this, the first ,we should make image of the license plate size normalized to 160×40 pixels, the experiments show that the license plate on this scale character image segmentation is best [6].

2.1.2 Binarization

The binarization purpose is to make the license plate image characters and the background with white and black to represent pixels respectively. Although the license plate recognition technology make color image converted to black and white binary image color information will be lost partial license plate, but almost no effect on the license plate character information, when the amount of information the binary image processing less than color images, without affect the accuracy of the premise, greatly improve the processing speed.

Character image binarization result that depends on the clarity of the characters selected threshold, when the selected threshold value is too small, will cause the character adhesions, but when threshold is too large, will cause the character breaks. Obtaining image binarization threshold are many commonly used are histogram method, Otsu method, co-occurrence matrix method and other. Due to the license plate inevitably affected by noise (pollution, defect, uneven exposure and other factors) interference, a single fixed threshold difficult to meet the requirements of the universal system, so using an iterative method to strike a dynamic threshold. The principle of iterative method as follows:

Step1: Obtaining image's the minimum gray value and the maximum gray value, denoted as V_l and V_h , the first test threshold Th_0 :

$$Th_k = \frac{AV_f + AV_b}{2} \quad (1)$$

Step2: According threshold Th_0 make image divided into foreground and background images were obtained both the average gray value AV_f and AV_b :

$$AV_f = \frac{\sum_{V(i,j) < Th_0} V(i,j)}{N_f}, \quad AV_b = \frac{\sum_{V(i,j) > Th_0} V(i,j)}{N_b} \quad (2)$$

In the formula (1), $V(i,j)$ is the pixel value of the image (i,j) points, N_f and N_b are the pixels of the foreground image and the total number of the background image.

Step3: Calculating a new threshold Th_o :

$$X_a = 160 \times \frac{45+12}{309} \quad (3)$$

If the formula $Th_k = Th_o$ is correct, then Th_k is the final threshold, the end of the iteration, otherwise go to Step2 continue iteration.

The traditional iterative threshold method is based on distribution of image gray value image itself, multiple iterations, making the final threshold is optimal, with strong stability and adaptability. However, the final binarization threshold generally go through several iterations, computing capacity, long processing time, if they can be bound by certain conditions, to reduce the number of iterations will increase binarization processing speed. Based on this, according to the knowledge previously been validated, the license plate characters share pixel ratio of about 20% [6, 7], the first gray value of all pixels in descending order, Th_o value of gradation values of row gray value at around 20%, thus avoiding the traditional iterative methods initial threshold value Th_o blindness, thus reducing unnecessary iterations, shortening the time binarization processing experiment has achieved good results, As shown in Figure 1.



Figure 1 Plate grayscale and binary diagram

2.1.3 Removing rivets and borders

Through the license plate image binary mathematical morphology after opening and closing operations [8], remove the rivets and dust on the plate and other smaller disturbances, through horizontal and vertical projection operations [2], to remove the border.

The results shown in Figure 2.



Figure 2 Renderings after the removal of border and rivets

2.2 License plate character segmentation

For any standard license plate, under ideal conditions, the license plate characters except the first one, the other six characters (numbers or letters) pixels in the binary image of the license plate (in this case white pixels) constitutes a separate connected domain, based on the analysis of this prior knowledge, just to get the ranks of the starting and ending positions of each connected domain, and thus constitute a minimum bounding rectangle area, you can quickly achieve character segmentation. For the first character, as some characters are not connectivity, such as "皖", "川", "吉", etc., will form two or more connected domain, which deal with characters, later do specific analysis.

Traditional segmentation algorithm based on the extraction of the license plate to determine the connected domain connected domain ranks starting position requires a top-down vertical scanning lines traverse the plate image pixel binarization after the characters find their first pixel, as a connectivity domain starting point, according to the region growing algorithm to complete the extraction of character region. Although such algorithm can accurately segment the license plate characters and then traverse the entire image, calculate the amount is too large, it is difficult to meet real-time requirements, the subsequent researchers such algorithms have been improved, reducing the amount of computation, shortens the processing time [3]. However, the improvement is limited after all, the principle of the pixel information of the image itself, to reduce the conventional traverse eight neighborhood pixels traversing the neighborhood of three pixels, the processing time is shortened is very limited. In fact, the characters are arranged on the law license information is known, if we can make full use of the character arrangement information, will be able to help quickly locate characters, thus achieving rapid segmentation is based on this expanded character segmentation study.

At present, China's current common license plate standard format: $P_i C_1 X_1 X_2 X_3 X_4 X_5$, where p_i is a

Chinese characters, ranging from the provinces, municipalities and autonomous regions referred; C_i is an English letter, on behalf of the name of the prefecture-level city; X_1, X_2, X_3, X_4, X_5 are English characters or Arabic numerals, and there is a solid dots between C_i and X_1 . The total length of the license plate character is 409 mm, where a single character width is 45 mm, height is 90 mm, Spacing between C_i and X_1 is 34 mm, a small dot in the middle of which is 12 mm, other inter-character spacing is 12mm^[5], the license plate characters arranged in the order shown in Figure 3.

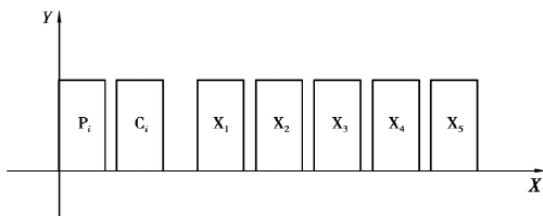


Figure 3 Schematic arrangement of license plate characters

Use of license plate character fixed width, spacing's fixed proportional relationship and other priori knowledge as constraint conditions. Take the following steps to determine the character boundaries, extracting a single character.

Step1: Calculating character C_i starting position. After pretreatment, the rest just left 7 simply characters on the license plate image, after the size of the normalized plate, license plate into a 160×40 pixels in size, so based on a priori knowledge of the starting position can be calculated

C_i abscissa X_a

$$X_a = 160 \times \frac{45 + 12}{309} \quad (4)$$

Bottom-up traversal $X = X_a$ this vertical split line's pixel, find the first white pixel, and thus the white pixels as a starting point accordance with region growing algorithm to find the character a complete connected domain, thereby dividing the character. Considering the license plate positioning and pre-processing errors, in order to accurately locate the starting pixel characters, set the threshold δ , so that the vertical dividing lines $X = X_a + \delta$, δ value of 2.

Step2: Accordance with Step1 sequentially calculated X_1, X_2, X_3, X_4, X_5 is a starting position, and then segmentation the subsequent five characters.

Step3: For the first Chinese characters, considering the connectivity of part Chinese characters. Accordance with Step1 character segmentation can't be completed. However, after a Step1 and Step2, only the first one on the license plate image Chinese character, so the combination of license plate character alignment rule and width of a priori knowledge, largely determine the area where the characters, in the region to merge all the connected domain, complete the kanji characters' extracted.

After the above three steps, the final result of the character segmentation in Figure 4.



Figure 4 Character segmentation results

3. Conclusion

To further shorten the processing time of license plate character segmentation, after analyzing the existing typical character segmentation algorithm, this paper improved binarization traditional iterative algorithm in another way by joining the license plate constructed prior knowledge, on the basis of a full analysis of binary image pixel distribution license plate features. It greatly reducing the number of iterations. The pretreatment time is shortened. While improving the traditional character-based segmentation algorithm connected domain, without having to traverse all the pixels will be able to quickly locate the position of the starting character column, dramatically reducing processing time.

References

- [1] Pan zhongjie, fangzhou Tan. A new license plate character segmentation method of template matching and vertical projection Combination [J]. Automation and Information Engineering, 2007, 28(2): 12-13
- [2] Meng tao. Research and Implementation of License Plate Recognition of key technologies [D]. Wuhan: Huazhong University of Science and Technology, 2006

- [3] Gan ling, xiaojing Lin. Based connected domain extracted license plate character segmentation algorithm [J]. Computer Simulation, 2011, 28 (4): 336-339
- [4] Li fang, bing Wu,Zhang hongying. Based on fast 8-connected component labeling's a new algorithm of extracting video subtitles [J]. TV Technology, 2009, 33 (2): 69-71
- [5] Zhang ying, yunhe Pan,.Vehicle license plate character recognition for preprocessing algorithm [J]. Application Research of Computers, 1999, 16 (7): 85-87
- [6] Zhang yungang, changshui Zhang. Use of Hough transform and a priori knowledge's the license plate character segmentation algorithm [J]. Journal of Computers, 2004, 27 (1): 130-135
- [7] HEGT J, DELAHAYE R, KHAN N. A high performance license plate recognition system [J]. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, San Diego, California, 1998: 4357-4362
- [8] Song chenguang, haijian Ye. Segmentation algorithm based on mathematical morphology license plate characters [J]. Optoelectronic Technology & Information, 2005, 18(6) : 58-62