A Non-learning-oriented Recognition Algorithm of Vehicle License Plate Characters

Fengkun LI, Benqing DONG, Long LIU
School of Intelligence & Electronic Engineering, Dalian Neusoft University of Information, Dalian 116023, China
lifengkun@neusoft.edu.cn

Abstract. Based on the fact that characters of vehicle license plate are standardized, a recognition algorithm of Vehicle License Plate Characters based on non-learning-oriented method is proposed. This method uses some concepts such as Shape Feature Vector (SFV, for short) to describe the shape of the to be recognized characters. Each character will be evaluated with a vector and the recognition is realized by calculating the similarity of vectors. This method achieves character recognition by vector arithmetic avoiding the learning/training process often used in some algorithms, such as artificial neural network, SVM and so on. The recognition is simplified tremendously. The feasibility of using SFV to recognize vehicle license plate characters is proven theoretically, and then, the character recognition algorithm of vehicle license plate is evaluated by simulation. The result of the simulation shows that SFV can be used as license plate character recognition. And the license plate recognition algorithm based on SFV has an accuracy rate of 97.31%. This method is a helpful attempt of character recognition based on non-learning-oriented method.

1. Introduction
Vehicle License Plate Recognition (VLPR) refers to the automatic extraction of vehicle license plate information and information recognition from the license plate image graphic data. And vehicle license plate information includes Chinese characters, capital English letters, Arabic numerals and license plate color, etc. VLPR is an important part of many intelligent transportation systems, and it has broad application prospect, such as, traffic violation detection, intelligent parking, etc.

With the rapid development of the Internet of things, pervasive computing and other concepts in the application, the application of mobile devices and micro devices is more and more common, such as the smart cameras that can recognize license plates. In this kind of application, the real time performance of such cameras is required to be high. In order to reduce network delay, intelligent processing is best performed at the network "edge", so, the concept of edge computing emerges. Moving "intelligent functions" from the cloud computing center to the network "edge" requires a basic requirement, which is to meet the resource requirements (computing resources, storage resources, etc.) on the edge of the network. Therefore, it has practical significance to explore the method of license plate character recognition with low resource requirements. License plate recognition system mainly includes three important parts: license plate area detection, license plate character segmentation and license plate character recognition [1–2]. Among them, license plate character recognition is a key step to convert the characters of graphic state into characters with character meaning or computable state. At present, the main character recognition can be divided into three categories [3], which is based on template matching, statistical classifier and feature point matching. When the graph size...
changes, the method of template matching is difficult to deal with. Literature [4] analyzes the characteristics of similar characters that are easy to be mistaken, and it improves the recognition rate by combining the idea of grouping and classification. Based on statistical classifier, hidden markov random model, SVM and neural network are adopted. On the basis of SVM scheme, literature [5] introduces structural features and grayscale features to improve the recognition rate of similar characters. In order to reduce the impact of noise, a self-coding neural network recognition scheme which based on stack noise reduction was proposed in literature [6]. In general, graphic images have more pixels, and the input to the statistical classifier is also high dimensional. These factors lead to a complex training or learning process. So, the overall scheme is complex. And it is also difficult to use on embedded devices. In this paper, it presents a lightweight character recognition algorithm. This algorithm uses shape feature vector to reflect the shape difference of license plate characters, and uses shape feature of characters to distinguish different characters. The scheme does not use pattern recognition method and also does not need to use a large number of samples for training. So, the scheme is simple to implement.

2. License plate character recognition scheme

2.1. definition

Definition 1 Character Frame (CF) In a binary bitmap of a license plate, parallelogram that exactly surrounds a character and satisfies the following conditions:

(1) The top and bottom edges are parallel to the top and bottom edges of the image.
(2) The slant angle of left and right edges is the character slant angle.

Definition 2 (Character Frame Subarea) refers to the two opposite sides of the word box midpoint connection line into the word box into four small partitions, which called A, B, C and D partition. A and B are called the left partition, A and D are called the upper partition, C and D are called the right partition, B and C are called the lower partition. c1, c2, c3 and c4 are called angular partition.

![Fig. 1 The example of character partitioning](image)

Definition 3 (Shape Ratio) In a binary bitmap of a license plate, take all pixels contained in the word box or a word box partition as the whole, the percentage of a character’s pixels that occupies the total is called the shape ratio. In this paper, take the whole word box as the overall proportion, called the overall shape proportion. The ratio of a given partition to the overall proportion, called the partition ratio. For example, A proportion ratio, left proportion ratio, upper partition ratio and so on, such as shown in Table 1.

| Word box partition | partition ratio name       | partition ratio symbol |
|--------------------|---------------------------|-----------------------|
| word box           | total proportion          | $SR_t$                |
| A partition        | A partition ratio         | $SR_A$                |
| A/B partition      | left partition ratio      | $SR_{AB}$             |
| A/D partition      | upper partition ratio     | $SR_{AD}$             |
| c1 partition       | angular partition 1       | $SR_{c1}$             |
Definition 4 (Shape Feature Vector, SFV) refers to the vector composed of various shape proportions to describe shape features, namely formula (1).

\[
\text{SFV}(c) = [SR_t(c), SR_d(c), SR_p(c), SR_u(c), SR_c(c), SR_{cd}(c), SR_{ca}(c), SR_{cb}(c), SR_{ca}(c), SR_{cd}(c), SR_{ca}(c), SR_{cd}(c), SR_{ca}(c), SR_{cd}(c)]
\]  

(1)

2.2. Feasibility analysis of license plate character recognition based on shape feature vector character

Whether the shape feature vector can be used as the mathematical model of license plate character recognition depends on two aspects:

1. Whether the shape feature vectors of license plate characters are different from each other;
2. Whether the shape feature vector of each license plate character remains unchanged.

This section examines whether the shape feature vector of characters on a license plate are distinguishable. That is, whether a character can be uniquely identified by shape feature vectors.

Chinese license plate characters include 31 Chinese characters, 10 digits and 24 capital English letters. Thereinto, I and O are easily confused with the Numbers 1 and 0, which are no longer used in actual license plate issuance. Then, the following character set and style definitions are available.

Definition 5: License Plate Characters and Style (LPCS) is the People's Republic of China public safety industry standard ga36-2014 specified character set and style. As shown in formula (2).

\[
\text{LPCS} = \{(\text{formatted character } c)|\text{GA36 2014 Specified character set and style}\}
\]  

(2)

Careful observation reveals that the characters in the license plate are limited and neat, this provides a basis for obtaining the shape feature vector of characters. In addition, the characters in the license plate are different in shape. As long as the designed shape Feature Vector is reasonable, the license plate characters can be distinguished by the SFV. The following is to prove that the SFV of these characters are distinguishable. The following is to prove that the SFV of these characters are distinguishable. In the proof process, the geometric composition and proportion of each license plate character are determined by practical measurement and approximate calculation.

Theorem 1 The SFV defined in equation (1) is distinguishable on LPCS, that is formula (3).

\[
\forall c_1 \forall c_2 ((c_1 \in \text{LPCS}) \land (c_2 \in \text{LPCS}) \rightarrow \text{SFV}(c_1) \neq \text{SFV}(c_2))
\]  

(3)

The proof is as follows:

Let the height of the license plate character be h, the width l, and the stroke width w, the SFV of all LPCS characters will be obtained by means of measurement and estimation. Through comparative analysis, formula 3 is established. The SFV of some characters are shown in table 2.

Table 2. The SFVs of character in LPCS set

| character | SR_t | SR_d | SR_p | SR_c | SR_{cd} | SR_{ca} | SR_{cb} |
|-----------|------|------|------|------|---------|---------|---------|
| 0         | \(\frac{5w}{l}+\frac{5w}{h-w}\) \(\frac{w^2}{h}\) | | | | | | |
| 1         | \(\frac{1}{h}\) \(\frac{w}{w^2}\) | | | | | | |
| 8         | \(3\pi(hl-(h-w)(l-w)) \frac{hl}{4hl}\) \(\frac{h}{hl}\) \(\frac{h}{hl}\) \(\frac{h}{hl}\) | | | | | | |
| C         | \(3\pi(hl-(h-w)(l-w)) \frac{hl}{2hl}\) \(\frac{hl}{2hl}\) \(\frac{hl}{2hl}\) \(\frac{hl}{2hl}\) | | | | | | |

2.3. Stability of SFV

Theorem 2 assume s is a character of LPCS, i is the positive image taken by s. Then, SFV(s)=SFV(i) is established.

Proof

As known from the camera imaging principle, the camera works by using the law that a convex lens will form an inverted, reduced real image between the first and second focal lengths of an object at two focal lengths. The real image produces a physical (digital camera) or a chemical (film camera) reaction on the photosensitive material, so the plane image can be got. That is, the different images of characters can be approximately simulated by the character s with the same proportional size change.
The proof is as follows: It is well known that the area of an irregular plane graph can be solved by dividing it into regular graphs.

\[ S = \sum a \cdot w \cdot h, \]

thereinto, \( a \) is coefficient, if it is a triangle, \( a = \frac{1}{3} \), if it is a rectangle, \( a = 1 \), if it is a circular, \( a = \pi \).

The total area of the word box:

\[ S(s) = \sum a(s)w(s)h(s). \]

Similarly, the partition areas of \( S_{\mathcal{A}}, S_{\mathcal{B}}, S_{\mathcal{C}} \) and \( S_{\mathcal{D}} \) have the same form.

Similarly, the total area of pixels occupied by the character \( s \), \( S_{\mathcal{R}}(s) = \sum a(s)w(s)h(s) \), and the pixel area of character \( c \) in each partition \( S_{\mathcal{R}}(c), S_{\mathcal{R}}(s) \) have the same expression.

Similarly, \( S_{\mathcal{R}}(i) \) can be got. thereinto, \( S_{\mathcal{R}}(i) = \sum a(s)w(s)h(s) \).

That is:

\[ S_{\mathcal{R}}(i) = S_{\mathcal{R}}(s). \]

Similarly, \( S_{\mathcal{R}}(i) \) have the expression for that.

As known from theorem1, characters in LPCS character set can be distinguished by SFV. As known form theorem2, All the positive images of LPCS characters have the same SFV. Combining the two theorems, it can be concluded that SFV can be used for LPCS character recognition.

3. Character recognition algorithm

Theorems 1 and 2 prove the feasibility of SFV for recognition of LPCS. This section introduces an algorithm for recognizing license plate characters based on shape feature vectors. The algorithm consists of the following five steps.
(1) For all collected license plate character samples, according to formula (1), the SFV of each character is calculated, denoted as: SFV(s). And then SFV of all license plate characters’ library is established. That is denoted as SFV\(_{\text{LPCS}}\).

(2) Get the characters to be recognized for denoising and shape correction.

(3) As same as formula (1), to calculate its SFV, denoted as SFV'(x).

(4) Calculate the Euclidean metric of SFV'(x) and every character of SFV\(_{\text{LPCS}}\) library, that denoted as: \(d_{\text{ks}}\).

(5) Sort \(d_{\text{ks}}\) from small to large, the result of recognition is the smallest distance from \(SFV'(x)\).

4. Experimental verification

The 652 LPCS character samples were analyzed by Matlab which is used as the simulation experimental tool. The sample covers the entire LPCS character set. First, the SPF library should be calculated. Some characters’ data is shown in table 3.

Observation shows that the measured values in table 3 conform to the laws reflected in the theoretical values in table 2.

4.1. Experimental study on special characters

In the LPCS character set, some characters and Numbers have similar shapes. So Character recognition algorithms often have difficulty in recognizing them. These characters include the letters D and 0, B and 8. The recognition schemes proposed in this paper is analyzed below.

| characters | \(SR_{i}\) | \(SR_{A}\) | \(SR_{B}\) | \(SR_{C}\) | \(SR_{A_{i}}\) | \(SR_{B_{i}}\) | \(SR_{C_{i}}\) | \(SR_{A_{c}}\) | \(SR_{B_{c}}\) | \(SR_{C_{c}}\) |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 0          | 0.49       | 0.47       | 0.50       | 0.48       | 0.53       | 0.50       | 0.49       | 0.48       | 0.50       | 0.16       |
| D          | 0.53       | 0.55       | 0.57       | 0.49       | 0.53       | 0.54       | 0.53       | 0.56       | 0.51       | 1.03       |
| 1          | 0.93       | 0.92       | 0.96       | 0.95       | 0.94       | 0.93       | 0.95       | 0.94       | 0.94       | 0.61       |
| 8          | 0.56       | 0.57       | 0.56       | 0.56       | 0.57       | 0.56       | 0.57       | 0.56       | 0.56       | 0.04       |
| B          | 0.59       | 0.68       | 0.57       | 0.52       | 0.61       | 0.65       | 0.55       | 0.63       | 0.57       | 0.95       |
| 6          | 0.44       | 0.38       | 0.58       | 0.54       | 0.26       | 0.32       | 0.56       | 0.48       | 0.40       | 0.00       |

Fig. 2 the characters "B" and "8"

The characters "B" and "8" are difficult to distinguish in various recognition algorithms. It can be seen from definition 2, table 3 and figure 2, the characters "B" and "8" have little difference in A, B, C, D, upper (AD), lower (BC), left (AB), right (CD), C3 and C4. These characters can’t be distinguished in these partitions.

However, the difference in C1 and C2 is significant. This is the difference in the location of the light green boxes shown in figure 2(b).

In the same way, figure 4 shows the reasons that SFV distinguish between characters "D" and "0".
4.2. Algorithm comparison and computational analysis

The recognition accuracy comparison with other class methods is shown in Table 4. The method in this paper is not very different from other improved methods on accuracy. However, this method is superior to other class methods in computational and storage complexity, as shown in Table 5. In Table 5, size is the LPCS character set size; \( \mu \) is the average feature vector number of each character; \( \mu \) is generally 1. The "high dimension" in the table refers to the number of picture pixels after image normalization; it is generally above 300. For example, literature [4] is 320 dimensions, literature [6] has 784 dimensions. It can be seen that the computational and storage complexity is very low, since the SFV is only a 9-dimensional vector.

| Recognition methods                  | Recognition accuracy /% |
|--------------------------------------|-------------------------|
| Improved template-based approach [4] | 91.46                   |
| SVM [6]                              | 95.12                   |
| he article's methods                 | 97.31                   |

| Recognition methods                  | Computing unit | Computing times | Storage unit | Storage quantity | If need training |
|--------------------------------------|----------------|-----------------|--------------|------------------|-----------------|
| Template method                      | Higher dimensional Euclidean distance | Size \( \times \mu \) | Higher dimensional feature vector | Size \( \times \mu \) | No |
| SVM                                  | High dimensional polynomial | 1                | High dimensional polynomial coefficient | At least Size \( \times \mu \) | Yes |
| Neural network methods               | High dimensional polynomial | 1                | High dimensional polynomial coefficient | At least Size \( \times \mu \) | Yes |
| he article's methods                 | Nine dimensional Euclidean distance | Size \( \times \mu \) | 9 dimensional feature vector | Size \( \times \mu \) | No |

5. Conclusion

Under the guidance of the idea of edge calculation, in order to reduce the resource requirements of the license plate character recognition method, the concept of SFV is proposed, and a non-learning license plate character recognition method is designed based on the concept. Based on shape feature vector, it has the following advantages. (1) There is no need to normalize the license plate image, and the character information filtered in the normalization process can be well used. (2) The calculation of SFV is simple, after the picture binarization, counting the number of 1. (3) This method does not need complex training and more storage space to store the data after training.
For example, in a neural network approach, various coefficients of a high-dimensional neural network need to be stored. The main problem in the implementation of this method is that the frame solution is not accurate when the license plate character is tilted. The next step is to optimize the solving process of the word box to further improve the character recognition accuracy.

Acknowledgments
This work was financially supported by the doctor launching fund of Liaoning Province (20170520398) and general project of science and technology of education department of Liaoning Province (L2015041).

References
[1] LETS, TRAN V N, Hamamoto K. A robust and flexible license plate detection method[C]// Proceedings of International Conference on Advanced Technologies for Communications (ATC). Hanoi: IEEE Press, 2014: 326-331.
[2] DU S, IBRAHIM M, SHEHATA M. Automatic License Plate Recognition (ALPR): A State-of-the-Art Review [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(2):311-325.
[3] Li Yan, Zhang Hongbo, Shi Lianying. License plate recognition method based on SIFT feature matching. Computer Engineering and Applications, 2016, 52(12): 194-200.
[4] Qu Zhong, Chang Qingli, Wang Yongkun. An improved algorithm for classification and recognition of license plate similar characters [J]. Computer Engineering and Design. 2013, 34(4): 1336-1340.
[5] Luo Huiwu, Tang Yuanyan, Wang Yi. License plate character recognition method based on structure feature and grayscale feature [J]. Computer science, 2011, 38(11): 267-302.
[6] Jia Wenqi, Li Ming, Zhu Meiqiang. License plate character recognition based on stack noise reduction self-coding neural network [J]. Computer Engineering and Design, 2016, 37(3):750-756.