A Multi-Style License Plate Verification Method Based on Character Classification under Illumination Variances

Zhongyan Liang[1, a *], Jie Hu[1, b], Qiaojin Guo[1, c] and Sanyuan Zhang[2, d]

[1]The 28th Research Institute, China Electronic Technology Group Corporation, Nanjing, P. R. China

[2] College of Computer Science and Technology, Zhejiang University, Hangzhou, P. R. China

[a]leungaptiva@126.com, [b]hujieyc@139.com, [c]guoqiaojin@163.com, [d]syzhang@zju.edu.cn

Keywords: License plate verification; Spurious license plate classification; Spurious character classification; Principal component analysis; Support vector machines

Abstract. License plate verification is essential for locating the license plate region precisely in a vehicle image. Existing verification methods implemented after location are effective for the single-style license plates, but will not do well when distinguishing multi-style ones from complex backgrounds. We propose a method based on segmented character classification, combining Principal Component Analysis and Support Vector Machines, to deal with the verification of license plates with multi-style. Compared with previous methods, the segmented character regions can be analyzed more effectively, and the proposed method gets the high confidence level by calculating the classification result of the segmented characters but not the whole license plates. The experiments were done under complex illumination conditions, and the results show that the proposed method is effective for multi-style license plates for both fixed and moving camera systems.

Introduction

License Plate Verification (LPV) is a technique for judging whether there is a true license plate or not in a license plate candidate region and preventing to report false license plates in a License Plate Recognition (LPR) system. Nowadays, due to the development of the computer vision and camera technologies, the LPR is widely used in many fields, such as stolen vehicle detection, driver navigation support, automated parking attendant, border crossing control, petrol station forecourt surveillance, personalized service via customer identification [1, 2] and so on. In the past few years, many researchers proposed various methods to solve the problem of LPR.

The training and classifying based methods, such as Artificial Neural Networks (ANN) [3] and Convolutional Neural Networks (CNN) [4], used the features extracted from the whole license plates. These approaches achieve good results for license plates with single-style. However, unfortunately, they are unsuitable for multi-style license plates. For example, it is difficult to collect hundreds of license plate styles in the United States. Moreover, the required samples of false license plates may be even more.

The methods proposed in [5-7] used the characteristics of the license plates. These characteristics include height, width, aspect ratio, and the number of character candidates, the number of peaks of the projection, the area, density, and the number of edges in the license plate center, histogram analysis, voting by these features and so on. These characteristics are difficult to describe various license plate styles uniformly, and will fail in dealing with complex scenes, such as grasses, bushes and foliage [5, 8]. The voting method may encounter difficulties in selecting more than one license plate.

It is difficult to identify the license plates in the United States by using the existing approaches because of variant template styles (Fig. 1). To overcome this drawback, we propose a verification method for multi-style license plates based on character classification combining PCA (Principal Component Analysis) and SVM (Support Vector Machines). The experimental results show that the proposed method can work under different illuminations and no matter the images captured by the fixed or moving cameras. Compared with other algorithms, this method does not care about the styles

of the license plates. The aim of this paper is to introduce a verification method applied after character segmentation, but not limited to the position of license plate location.

The rest of this paper is organized as follows. The license plate verification method is proposed in Section II. Experimental results and discussions are shown in Section III and the conclusion in Section IV.

Figure 1. The samples of the license plates in the United States.

The Proposed Method

The peculiarity of the proposed method is to implement the license plate verification after character segmentation. Fig. 2 shows the position of the method within the entire LPR system compared with the traditional one. The preprocessing usually includes the character region refinement, rotation and distortion correction. Fig. 3 shows the overall process of the proposed method.
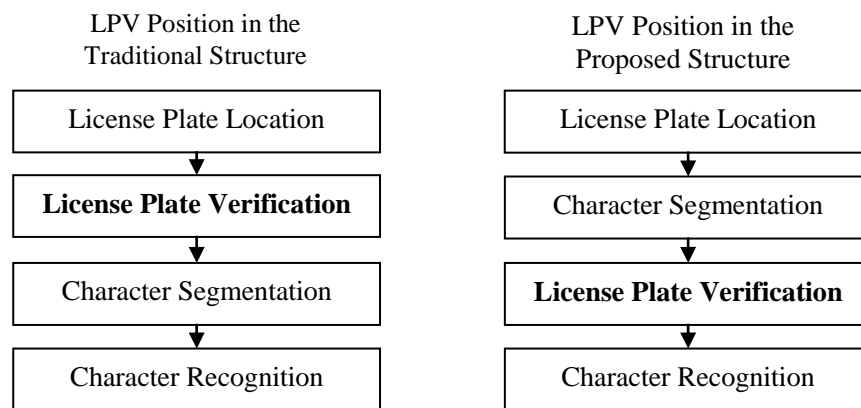

Figure 2. The position (bold font) of the proposed method compared with traditional structures within the LPR system.
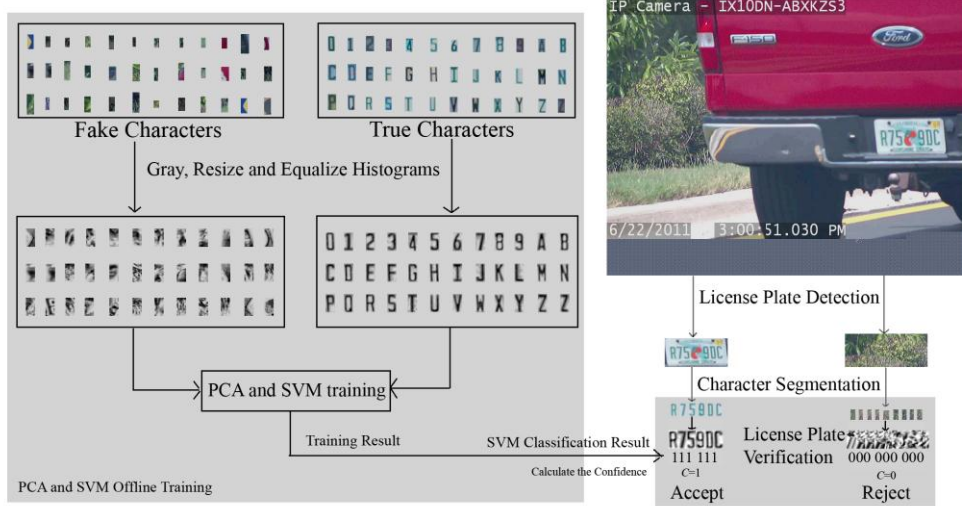
Figure 3. The example of the proposed method in the entire LPR system. The proposed method is shown in the shaded area.

Extract Features and Classification. Because the character segmentation is not a part of the proposed method, we use the vertical projection [9] and bounding boxes [10] to do this job.

Firstly, after gotten the character candidates from character segmentation, we should label them as two data sets: false characters and true characters (Fig. 3), respectively.

Secondly, for each character image, we use the following steps to normalize the PCA [11] features. PCA is a dimensionality reduction technique and expresses the data in a way of identifying patterns as to highlight the similarities and differences.

1) To make the algorithm work on both color and gray images, we need to convert color images to gray ones.

2) Once the gray image has been obtained, an image resizing is performed to change the image to a standard size, such as 20 x 40.

3) Histogram equalization is applied to balance the illumination intensity of the resized images.

4) We can use the equalized images to get PCA features, such as 16-dimension.

5) In order to make the SVM more stable, we should normalize the features by scaling in the interval [-1 1].

As a last step, we use the SVM [12] algorithm to classify the previous normalized PCA features. The SVM is effective for two-class problems. In this paper, we use Radial Basis Function (RBF) as the kernel.

In the training phase, we need to solve the following problem:

$$\min_{\mathbf{w},\xi,b} \max_{\alpha,\mu} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i - \sum_i \alpha_i [y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i] - \sum_i \mu_i \xi_i \right\} \tag{1}$$

$$\text{s.t.} \alpha_i, \mu_i \geq 0$$

Where μ_i are the Lagrange multipliers introduced to enforce positivity of the slack variables ξ_i [12].

To maximize the dual form:

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{2}$$

$$\text{s.t.} 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0$$

RBF kernel: $K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$. $\tag{3}$

In the test phase, calculate the sign of the following formula to classify:

$$f(\mathbf{x}) = \sum_{i=1}^{Ns} \alpha_i y_i K(\mathbf{v}_i, \mathbf{x}) + b \tag{4}$$

Where Ns is the number of support vectors ($\mathbf{v}_i$), α_i is Lagrange multiplier, $\mathbf{x}_i \in \mathbf{R}^d$ is the training data, and $y_i \in \{-1, 1\}$ is the label, positive (1) or negative (-1).

While using the RBF kernel, the SVM training can produce the number of centers (Ns), the centers themselves ($\mathbf{v}_i$), the weights (α_i), and the threshold (b) automatically [12]. Therefore, we need to select the soft margin C and the kernel parameter γ by the grid-search the cross-validation [5].

The License Plate Candidate Judgment. We calculate the confidence value for the whole license plate by the following formula:

$$C = \frac{\sum_{i=1}^{N} T_i}{N} \tag{5}$$

Where C denotes the confidence value, N denotes the number of the segmented character candidates in a license plate. If the SVM reports the character is true, $T_i = 1$, otherwise, $T_i = 0$. (Fig. 3)

If $C > Th$, such as $Th = 0.3$, we think the license plate candidate is true.

Other Characteristics. Furthermore, we use other characteristics (e.g. the number of characters, the aspect ratio and the area of the license plate, etc.) to filter out the false regions.

Experimental Results and Discussions

Experimental Results. To validate the effectiveness of the proposed approach, we use two data sets which we collect during this year. The training set and testing set consist of the images got from a fixed camera and a moving camera during the day, respectively.

Table 1 shows the number of instances in each set.

Table 1. The number of instances in each set

	Positive instances	Negative instances
The Training Set	8619	1883
The Testing Set	3694	807

We use the true identification rate (*TIR*) and false identification rate (*FIR*) [13] to evaluate the approach:

$$TIR = \frac{N_{tp}}{N_p} \tag{6}$$

$$FIR = \frac{N_{fn}}{N_n} \tag{7}$$

Where N_{tp} denotes the number of true positives, N_p denotes the number of positive instances, N_{fn} denotes the number of false negatives, and N_n denotes the number of negative instances.

TIR and *FIR* can evaluate the ability of keeping the true license plates and removing the false license plates, respectively.

To compare with the training based verification method after the license plate location; we implement the SVM and ANN classification algorithm dealing with the PCA features without segmentation, respectively. ANN used the Back-propagation algorithm (using the same configuration as the sample letter_recog.cpp in Open CV).

Table 2 shows the experimental results.

Table 2 Experimental Results

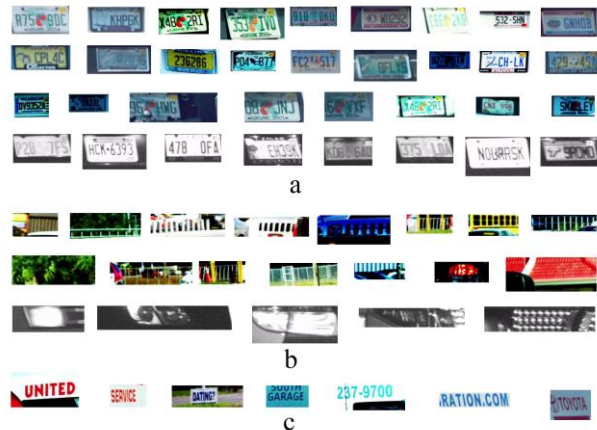| Data set	TIR (%)	FIR (%)	TIR (%)	FIR (%)
	The Training Set		**The Testing Set**	
PCA and SVM without Segmentation	**99.98**	**0.00**	95.57	67.03
PCA and ANN without Segmentation	99.23	**0.00**	83.17	22.25
The proposed	99.04	0.04	**97.74**	**1.65**

Figure 4. Samples of the verification results, (a) shows the true license plates, including color images and infrared black and white images (b) shows the false license plates (true negatives), and (c) shows the false positive results.

Performance. The proposed method is fast enough (about 3.271 ms on average) for real-time LPR system and almost does not affect the performance of the system. We implemented the algorithm using Visual C++ 2010 with OpenCV 2.4.10 and tested the data sets on a Core Duo PC with 2.8 GHz CPU and 4GB RAM.

Discussions. Compared with the other methods, the experimental results show the proposed method is more robust on the testing data. The FIR of the training based methods without segmentation on the testing data is higher because of many untrained negatives. The license plate location cannot guarantee whether the license plate region is the minimum bounding box or not. It is a challenge for machine learning to deal with hundreds of license plate styles in the United States. Besides, the testing set is captured by a moving camera, so the scene and illuminations are always changing, and the license plates occur at uncertain angles. However, character segmentation removes background patterns and reduces the load of training and classifying.

From Fig. 4, we can notice the algorithm can classify multi-style license plates from the negatives under various illuminations. One limit of the proposed method is the algorithm cannot distinguish text from license plates well (Fig. 4c), because of the character classification theory.

Conclusion

In order to improve the accuracy of verifying the multi-style license plate candidates in a video-based LPR system, we proposed a license plate verification method based on isolated character classification using PCA and SVM. The experimental results show it can deal with multi-style license plates without using the style information under various illumination conditions, and achieve a high identification rate. Unfortunately, it is difficult to distinguish license plates from scene texts, which will be the research focus in the future. Besides, the proposed method needs more preprocessing and segmentation; hence it will take more time than other methods. However, we can improve performance by optimizing the code, taking advantage of multithreading technologies and the hardware acceleration. Furthermore, the proposed method will expand the scope of applying the LPR system under uncontrolled environment.

References

[1] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, Automatic License Plate Recognition (ALPR): A State-of-the-Art Review, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 23 (2013) No. 2, p.322.

[2] H. Gao, N. Liu, and Z. Zhao, A Combination Method of Edge Detection and SVM Filtering for License Plate Extraction, *Proceedings of 8th International Conference on Image and Graphics, ICIG 2015*, Part I, (Tianjin, China, Aug. 13-16, 2015), p.21.

[3] D. Lee and J. H. Choi, Precise Detection of Car License Plates by Locating Main Characters, Journal of the Optical Society of Korea, Vol. 14 (2010) No. 4, p.376.

[4] Y. Liu and H. Huang, Car plate character recognition using a convolutional neural network with shared hidden layers, *Chinese Automation Congress*, (Wuhan, China, Nov. 27-29, 2015), p. 638.

[5] S. Yu, B. Li, Q. Zhang, and C. Liu, A novel license plate location method based on wavelet transform and EMD analysis, Pattern Recognition, Vol. 48 (2015) No. 1, p.114.

[6] R. Al-Hmouz and S. Challa, License Plate Localization based on A Probabilistic Model, Machine Vision and Applications, Vol. 21 (2010) No. 3, p.319.

[7] I. Giannoukos, C.-N. Anagnostopoulos, V. Loumos, and E. Kayafas, Operator context scanning to support high segmentation rates for real time license plate recognition, Pattern Recognition, Vol. 43 (2010) No. 11, p.3866.

[8] B. Epshtein, E. Ofek, and Y. Wexler, Detecting Text in Natural Scenes with Stroke Width Transform, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (San Francisco, USA, Jun. 13-18, 2010), p.2963.

[9] H. Xia and D. Liao, The study of license plate character segmentation algorithm based on vertical projection, *Int'l Conf. on Consumer Electronics, Communications and Networks (CECNet)*, (Xianning, China, Apr. 16-18, 2011), p.4583.

[10] N. F. Gazcón, C. I. Chesñevar, and S. M. Castro, Automatic vehicle identification for Argentinean license plates using intelligent template matching, Pattern Recognition Letters, Vol. 33 (2012) No. 9, p.1066.

[11] I. T. Jolliffe, *Principal Component Analysis*, (Springer Verlag, 1986).

[12] C. J. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, Vol. 2 (1998) No. 2, p.121.

[13] N. Thome, A. Vacavant, L. Robinault, and S. Miguet, A Cognitive and Video-based Approach for Multinational License Plate Recognition, Machine Vision And Applications, Vol. 22 (2011) No. 2, p.389.