Research on Image Recognition for License Plate from Classifier Algorithm

2021 3rd International Conference on Robotics and Computer Vision (ICRCV) | 978-1-6654-3628-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICRCV52986.2021.9546962

Lijuan Qin
School of Information Science and Engineering,
Shenyang Ligong University
Shenyang, Liaoning, P.R. China
qinlijuan06@163.com

Xiaofang Liu
School of Information Science and Engineering,
Shenyang Ligong University
Shenyang, Liaoning, P.R. China

***Abstract*—Classification algorithm based on Adaboost and Harr-like features is quite mature to be used in the field of face detection. In this paper, the traditional Adaboost algorithm is improved and the vehicle positioning module is added. At the same time, this algorithm is applied to vehicle license plate location and it greatly improves the accuracy of license plate location. License plate location is to distinguish the plate under the complex background environment. In license plate recognition, it is the most critical step. Its accuracy directly affects the accuracy of license plate recognition. In this aritical, a new algorithm is proprosed, and two simulation softwares are combined to make results of license plate location and recognition accuracy improved significantly.**

Keywords—License plate, Classifier algorithm, Recognition and location

I. INTRODUCTION

License plate recognition system highlights its importance in the modern traffic day by day, and it plays a key role in the license plate recognition process. It directly affects the accuracy of the license plate recognition. Many methods have been applied to license plate location technology. Since the end of last century, there have been more and more scholars begin to research license plate recognition system, and many excellent research achievements have been made. However, a large number of complex external factors can lead to the effect unsatisfactory, and make the license plate localization extremely difficult. For example, more vehicles will cause the vehicle's license plate obscured, the turning of the vehicle can result in such as license plate tilt condition will affect the recognition of license plate, thus how to improve the accuracy of the license plate localization has become a research hot spot issues.

There are already a lot of license plate location technology domestic and abroad, some are more mature and have practical use. The main existing foreign excellent license plate recognition system products are: The VLPRS products of the Singapore company Optasia. The ARTEM7S system of Germany's Siemens. The See/Car System series products of Hi-Tech and so on. Due to the differences of the development of the economic and the modern technology at home and abroad, the license plate format differences, and the differences in driving and traffic rules. So although foreign license plate recognition technology developing rapidly, but it can only be used as a reference. The domestic research of license plate recognition system started in the 1990s, After more than 20 years of development, engaged in license plate recognition research in domestic well-known institutions of higher learning and scientific research institutions are: Sichuan zhisheng, Tsinghua university, etc. The research achievements of xi 'an Jiaotong university in China to promote the rapid development of China's highway traffic, at the same time, Shanghai MrCote license plate automatic recognition is done very well. But in our current level of science and technology to realize the license plate of high recognition rate and has the certain difficulty, one can not be ignored because the license plate localization is not accurate. This paper researches on the existing problems of license plate location technology, in order to improve the accuracy of license plate recognition under complex conditions.

In terms of the license plate positioning mainly adopts three methods of license plate location algorithm is divided into three categories, namely based on edge, based on color and algorithm based on machine learning. Basing on edge is common, and its basic principle is to judge the license plate area, with the aid of edge detection operator is suitable for the ideal background conditions of the license plate recognition. The static traffic moving vehicles or normal linear recognition rate is higher. But under complex scene, mistakenly identified will be more, false alarm rate is higher. On the basis of the localization algorithm using color image edge color or gray scale image edge color combination to locate. Its advantage is that: using the color character difference can be easily license plate edge detection, and image gray level will be affected by the light intensity, and the color is not sensitive to light. But with a greater difference between the human eye perception center for two kinds of color and class have the same color (center) similarity. This kind of phenomenon will cause the false judgment of the color, the drop of the system's ability to adapt to the environment, at the same time directly with color images will inevitably cause lower efficiency and memory resources waste. So, the first two is more suitable for ideal conditions of the license plate positioning, disturbance in tilt, dark, AD positioning accuracy greatly reduced. Based on machine learning algorithms, mainly through the extraction of characteristic value, and establish the cascade classifier [1]. The algorithm and training algorithm for image features is easy to cause the license plate area uncomplete detection.

Therefore, this paper added the license plate thick locating module before processing plates to avoid the loss of the edge of the plate and greatly reduces the error rate. The machine learning algorithms adopted the Adaboost method. The algorithm extracts the Haar-like characteristics of a large number of positive and negative samples [2], then do the multi-level training. Finally we get a Adaboost classifier of strong ability. This article uses the method of determining the alternative of the license plate area based on edge information and using the Adaboost classifier to eliminate the license plate area, results show that the detection rate of older algorithm is greatly increased.

II. LICENSE PLATE RECOGNITION SYSTEM

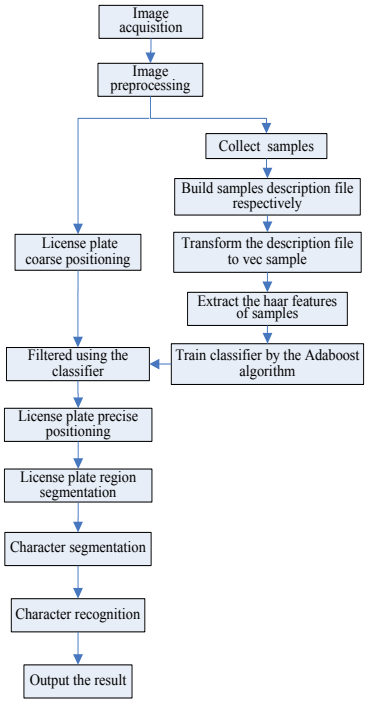

Fig. 1. Overall flow chart for license plate recognition.

To collect the positive and negative samples, different background conditions should be built by artificial structures, included normal environment and complex conditions. Camera collected the license plate image plates under ideal conditions and special conditions. Secondly to pretreat the collected samples and license plate image to be processed, image preprocessing is the key to license plate processing, including basic pretreatment and preprocessing based on Adaboost and Haar - like algorithm. Basic pretreatment are usually in order to implement image denoising and enhancement (including gray scale transformation, binarization, gauss transformation, smoothing denoising, image sharpening). In view of the positive and negative samples, using Haar - Like character description method to extract the characteristic value, and then through a large number of training process, construct Adaboost cascade classifier. First, for the target image, using edge detection operator, rough positioning license plate area [3], to avoid the loss of edge, to provide the training of the classifier inspected images and more accurately. And then extract the image of Haar-Like features. After coarse positioning based on edge detection and precise localization based on Adaboost, determine the part for segmentation of license plate. After the positioning of the complete license plate rea, characters can be cut and identified, that is, the license plate area divided into seven single characters, and then identified. Vertical projection is generally used in the character segmentation, and the character recognition are mainly based on template matching algorithm and artificial neural network at present. License plate recognition flow diagram is as shown in the Figure 1.

III. FEATURE EXTRACTION OF THE LICENSE PLATE IMAGE BASED ON HAAR-LIKE DESCRIPTION METHOD

A. License Plate Coarse Locating

The rough positioning uses the most basic of the localization algorithm [4]. First of all, use edge detection operator Sobel operator for edge detection, then the corrosion and smooth processing. Finally using the theory of horizontal and vertical projection to get the license plate area. Different from the traditional edge detection, rough localization process effectively expand the theory of plate region (known as the coarse location area), ensure that license plate area is not larger than the coarse location area.

Fig. 2. The gray image.

Fig. 3. Traditional positioning plate region(missing important information).

Fig. 4. Coarse positioning results figure (keep important information).

Figure 2 is the gray image. Figure 3 is the traditional positioning plate region (missing important information). Figure 4 is coarse positioning results figure (keep important information). By the above, we can come to the conclusion that coarse positioning can retain the license plate of important information, also reduce the pixel values of the license plate, improve the application rate. Coarse positioning make up for the shortcomings of traditional Adaboost algorithm, which is easy to lose the edge. The method can make license plate area to reduce effectively and improve the recognition efficiency.

B. Haar-like Character Description

Haar-like features is a kind of character description operator used in the field of computer vision commonly [2].

Assumption in the image window size hypothesis training or area consists of W * H pixels. Any rectangle window can be consists of an array of 5-bit integers. $r=(x,y,w,h,\alpha)$. (x, y) is the vertex of the left upper corner of the rectangular coordinates. W, h are the rectangular length and width respectively, α is the rectangular rotation angle alpha.

According to the boundary conditions to satisfy the following formula:

$0\le x,x+w\le \mathrm{W};0\le y,y+h\le H;x,y\ge 0;w,h>0;\alpha\in\{0^{\circ},45^{\circ}\}$

Calculate rectangular characteristic value in the the following formula:

$$feature_I=\sum_{i\in I=\{1,\ldots,N\}}\omega_i \operatorname{Re}cSum(r_i) \tag{1}$$

Among them, the ω_i for the value of the first rectangle for the sum of all pixels in the ith rectangle. {1, ..., N} is the composition characteristics of the number of rectangles.

As the training samples are usually more than one thousand, and the number of rectangle features is very large, If every time calculating eigenvalues are statistics and the sum of all pixels in the rectangle will be greatly reduced the speed of training and testing. Therefore, introduces a new method for image—integral image, rectangular features of value calculation, only related to the characteristics of the rectangular endpoint integral figure. So no matter the characteristics of the scale of the rectangular how to transform, the consumption of eigenvalue computing time is constant. So as long as the traversal time, image can obtain all the eigenvalues of the child window.

Integral figure is a rectangular area of the calculation of the sum of all the pixels [3]. Integral figure $\mathrm{I}_{\Sigma}(\mathrm{p})$ in the position $\mathrm{p}=(\mathrm{x},\mathrm{y})^{\mathrm{T}}$ is defined as follows:

$$\mathrm{I}_{\Sigma}(\mathrm{p})=\sum_{i=0}^{i\le x}\sum_{j=0}^{j\le y}I(x,y) \tag{2}$$

First of all, calculate the sum of all pixel in the image of arbitrary rectangular can be convenient and fast with the aid of image of integral figure calculation by an example. As shown in the Figure 5 below, the top left corner of the image I is assumed to be the origin of coordinates, rectangular area D is any area of the image I. Four vertices are 1, 2, 3, 4, Obviously, by definition, a rectangular area location 1 integral image I1 has a value of $I_1=sum(A)$,By the same token, the position of 2, 3, 4 integral image are:

$$\mathrm{I}_2=\mathrm{Sum}(A)+\mathrm{Sum}(B) \tag{3}$$

$$I_3=Sum(A)+Sum(C) \tag{4}$$

$$I_4=Sum(A)+Sum(B)+Sum(C)+Sum(D) \tag{5}$$

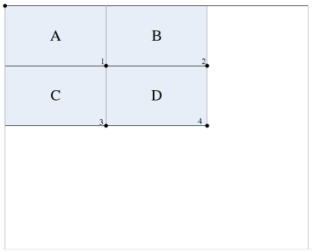

Fig. 5. Integral image calculation.

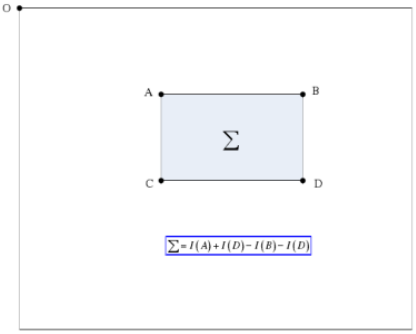

Fig. 6. Sum of pixels for arbitrarily rectangular in image.

Through the four points of integral figure, we can get the total all the pixel in the rectangular area D: $Sum(D)=I_1+I_4-I_2-I_3$, Once a picture of the integral image computation, then according to the previous example, the image of a rectangular sum of pixels in the area of operation only need three plus (minus) method can get. And operator is independent of the size of the rectangular area.

Solve the problem of the summation of the rectangular area, the rectangle characteristic value calculation is just based on this the sum of the computation, this is as seen in Figure 6.

This is a strategy to take space in time. For the input image, prior to it at each pixel position of integral figure. Save the integral figure of each pixel location in advance. So, when need to compute the rectangular characteristic value, just need to take four endpoints rectangular integral figure for simple addition and subtraction operation. And, no matter how the scale and location of the characteristics change. To determine the type of rectangular characteristics (2 - rectangular), the cost calculation of the eigenvalues of the time is a constant. Integral graph traversal process as shown in the Figure 7 below:

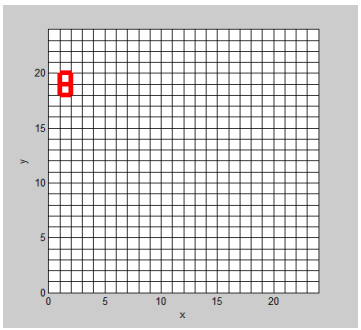

Fig. 7. Haar-like traversal graph.

IV. License Plate Image Recognition and Location Based on Improved Adaboost Classifier Algorithm

A. Adaboost Classifier

Adaboost algorithm is proposed an adaptive Boosting algorithm by Freund and Schapire as early as 1997 [4]. Its basic idea is constituting a strong classifier from a set of weak classifier through a certain method of superposition. The algorithm itself is implemented by changing the data distribution. It is according to the classification of the training set in each sample is correct, and the last time the overall classification accuracy, to determine the weights of each sample. Modified classifier of the new data set for the lower weight training, finally, each trained classifier finally

together as the final classifier. Using adaboost classifier can eliminate some unnecessary training data characteristics, and put the key in the key of training data. The initial weak classifier may be just one of the most basic characteristics of Haar-like, Calculation of the input image Haar - like eigenvalues, and compared it with the eigenvalues of the original weak classifier, in order to determine if the input image plates. Then training weak classifiers to be optimal weak classifier, namely a error is relatively lower weak classifier, the training weak classifier is actually setting classifier process. The license plate recognition classifier is finally determined through screening of license plate location.

Algorithm description [5]: given the training sample set $\{(x_1,y_1),(x_2,y_2),(x_3,y_3),\cdots\cdots,(x_n,y_n)\}$, Among them $y_i=0,1$. Corresponding negative samples and positive samples. Sample concentration of positive and negative sample l, m, respectively. Initial weight of the sample I:

$$\omega_{l,i}=\begin{cases}\dfrac{1}{2m}, y_i=0\\ \dfrac{1}{2l}, y_i=1\end{cases} \tag{6}$$

The basic steps:

1) t = 1, sample normalized weights expressed as:

$$\omega_{t,i}\leftarrow\frac{\omega_{t,i}}{\sum_{i=1}^{n}\omega_{t,i}} \tag{7}$$

2) For feature j, training a weak classifier $h_{t,j}(x_i)$, the weighted error for its classification $\varepsilon_{t,j}$ is:

$$\varepsilon_{t,j}=\sum_{i=1}^{n}\left[\omega_{t,j}\left|h_{t,j}(x_i)-y_i\right|\right] \tag{8}$$

3) Classification of weighted error minimum classifier as the iteration of the classifier, record as $h_t(x_i)$. The classification of the corresponding weighted error remember as ε_t , sample weights updated as:

$$\omega=\begin{cases}\dfrac{\omega_{t,i}\varepsilon_t}{1-\varepsilon_t} & h_t(x_i)=y_i\\ \omega_{t,i} & \text{others}\end{cases} \tag{9}$$

4) t= t + 1, repeat steps 2 and 3, until t with a given equal the number of iteration.

Assuming that the intermediate variables α_t is associated with classified weighted ε_t , represented by:

$$\alpha_t=\lg(\frac{\varepsilon_t-1}{\varepsilon_t}) \tag{10}$$

End up with a strong classifier $H(z)$ for:

$$H(z)=\begin{cases}1 & \sum_{t=1}^{T}\alpha_t h_t(z)\ge\frac{1}{2}\sum_{t=1}^{T}\alpha_t\\ 0 & \text{others}\end{cases} \tag{11}$$

z to be classified data characteristics.

B. Training Process

To train a Haar classifier [6], generally includes three steps: 1) Prepare positive and negative samples: collected 500 copies of license plate images, through gray and size adjustment, and ultimately are samples of the same size. Collect 800 images from the body or vehicles tailored background images. As the negative samples (including lights, logo, slogan, etc.), negative sample size does not require the same. 2) established positive sample set by using build sample function program: Firstly, build the positive and negative sample description file (txt file), modify the file parameter, add the file name and path (the positive sample description file also add the picture parameters). Sample description file is converted to 'vec' file used for computer recognition. Using the CreateSamples program of Opencv and the command CMD [7] to create the sample files. 3) Using Adaboost algorithm to classify Haar features of training, using the HaarTraining of Opencv to get the final classifier model (XML file), the sample size increased in the process of the training. Figure.8 is the original figure, Figure 9 is the positive sample and Figure10 are samples of unified size (pixel size 60 x 20). Figure 11 are negative samples.

Fig. 8. The original figure.

Fig. 9. The positive sample.

Fig. 10. Samples of unified size.

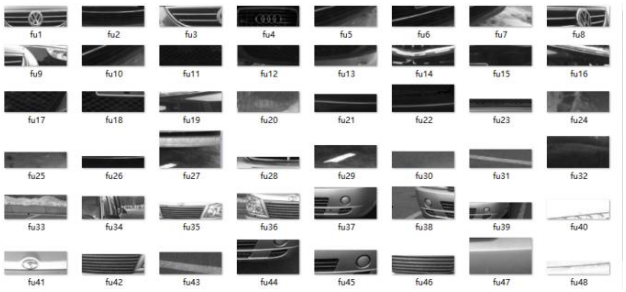

Fig. 11. The negative samples.

The results: 500 positive samples,600 negative samples, the accuracy is low. The result is as seen in Figure 12.

Fig. 12. License plate location results 1.

1000 positive samples, 1500 negative samples, the result accuracy is better but it is still not ideal. The result is as seen in Figure 13.

Fig. 13. License plate location results 2.

1500 positive samples, 2000 samples negative, the result accuracy is higher. The result is as seen in Figure 14.

Fig. 14. License plate location results 3.

V. Experimental Analysis

This paper puts forward a new research about License plate location problem under complex condition, it provide a better study environment for the subsequent license plate recognition process. Also solved the shortcomings of traditional license plate recognition under complex conditions, such as high error rate, long recognition time, long algorithm.

Image analysis method of MATLAB and OPenCV are powerful image analysis and processing tools [8]. MATLAB can quickly realize the MATLAB image processing, facilitate validation and analysis of the algorithm, the characteristics of MATLAB language is different from other senior language, known as the fourth generation of computer language. But the MATLAB to feature extraction and classification training, processing speed is slow. So this paper also adopted the processing method of OPenCV, it is mainly used for Haar-like feature extraction and Adaboost classifier training.

VI. Conclusions

This paper is for license plate location problem under the complex conditions. Because the external environment and equipment will bring image noise, and affect the image quality, so the basic image preprocessing is necessary. Then it is carried on the license plate coarse positioning, the addition of the module improved the defects of traditional Adaboost algorithm easy to loss of edge information, and improve the classification efficiency of the classifier, shorten the program running time. Feature extraction and classifier training, by traversing the graph extraction of characteristic value, detecting the target compared with positive and negative samples, after multiple classifier of screening plate region. The new algorithm makes the license plate positioning accuracy greatly increased, so that the character recognition rate is also greatly improved.

References

[1] Kruizinga, P. 2008.Nonlinear operator for oriented texture, *IEEE Transactions on image processing*, 8, 10, 395-407.

[2] Zhao, Q.Y., Pan, B.C., Zheng, S.L. 2008. Keep the color image enhancement new algorithm, *Journal of computing applications*, 28, 2, 448-451.

[3] Wei, W.W., Chen, Y.S. 2008. AdaBoost face detection method based on support vector machine, Computer simulation, 25, 6, 210-213.

[4] Ni, K., Tao, X.Z., Zhang, F. 2011. Pedestrian detection method based on gradient direction histogram study, *Journal of TV technology*, 35, 5, 96-99.

[5] Ma, H.X. 2012. License plate tilt correction based on sub area projection analysis, *Journal of computer applications and software*, 29, .6,253-255.

[6] Wang, P.,Cheng, H., Luo Y.X. 2007. Color image enhancement based on the color the same, *Journal of image and graphics*, 12, 7, 1173-1177.

[7] Zhang, L. L. , Zhang, L. , Ma, J.X. 2008. Marine environment of the durability of the carbon fiber reinforced composite sheet, *Journal of building materials*, 11, 6, 732-735.

[8] Tian, T., Lei, J. L. 2009. Improved face detection based on skin color and AdaBoost algorithm, *Journal of computer applications and software*, 26, 12, 79-80.