The Method of Cutting Image of Vehicle Face based on Haar Feature and Improved Cascade Classifier

ZHANG Tong\textsuperscript{1} and MAO Li\textsuperscript{1}

\textsuperscript{1} Network information security department of GuangDong police officer college, Guangzhou, 510230, China

Email: susanzhangtong@126.com

Abstract: A method of car face image segmentation based on haar-like features and improved cascade classifier was proposed to solve the problems of insufficient classification recognition performance and too long training time of the current classifier. First, the haar-like features of the image are calculated based on integral graph. Then the extracted feature values are input into the cascade classifier for training. Finally, the trained classifier was used to detect the vehicle face and segment the image of the vehicle face. In this paper an adaptive threshold setting method is proposed. The experimental results show that the method has the better performance than the traditional methods, and has better application prospects.

1. Introduction

In China the automobile is increasingly becoming the most important means of transportation in People's Daily life. While bringing convenience to people, cars have also become the preferred way for criminals to commit crimes and escape, which brings great difficulties to the investigation and inspection work of public security. In many cases the lucky escape of criminals often resulted in the descending detection speed of cases because criminals are not captured in time. How to lock the target suspect vehicle in the shortest time is an urgent problem to be solved by the public security system. The most prominent feature of the target suspect vehicle is the license plate number. Once the license plate number of the escaping vehicle is obtained, the suspect vehicle can be quickly tracked down using a strategic pass system that covers the city. However, in real life, due to the weather, light, time and other reasons, many informers failed to record the number of escape vehicle license plate. Searching a suspect car among the city's millions of vehicles is like looking for a needle in sea. Although many informer have not been able to record the number of the escape vehicle license plate, but generally they know the suspect vehicle color, brand, additionally considering when and where the case happen, combined with the vehicle type recognition technology, police officers will be able to lock the suspect vehicle in a very small range, which brings great convenience to detect the case. The technology of vehicle type recognition aims at finding out the brand which the vehicle belongs to i.e. Audi, BMW and Volkswagen etc. based on vehicle face features.

The face of vehicle is made up of the logo, grid and headlights. The feature of face can distinguish different type of vehicle, so the vehicle can be identified only by the image of its face [1]. In some extent this method can reduce the number of feature, and thus reducing the quantity and complexity of calculation. The face of vehicle is detected from traffic monitoring video, and then images of vehicle face are cut from video for further feature extraction. The detection of car face area is similar to that of human detection. There are many researches on face detection [2].
For face detection research there are three classes. The first class is the method based on the geometric features, the second class is the method based on skin color model, and the third class is the method based on statistical theory. The geometrical characteristics of the human face refer to the person facial features and the location of the relationship between them. Because the color of person face skin is different from background color, feature of color can be used to distinguish between face and non-face. The statistical method refers to the statistical features of face and non-face samples respectively, and then uses the method of statistical analysis and machine learning to distinguish. For the car face area, the position of the car logo varies, and the relative position is not easy to determine, and the color of the car face are a is consistent with the color of the car body in many cases, so the method of statistical theory is adopted to detect the car face area. Support vector machine (SVM), artificial neural network (ANN) and AdaBoost are the most widely used methods based on statistical theory. This paper adopted Harr-like features which can be figure out quickly and cascade classifier which was improved to adjust to real-time environment.

2. Integral Image
The integral image, as the name implies, is the integral of the pixel value of each point, which is simply the sum of all the previous pixel values (shown in Figure 1).

![Figure 1. the image and integral image](image1)

![Figure 2. the computation of the sum of rectangle based integral image](image2)

Shown by the following formula:

$$\text{II}(x,y) = \sum_{i\leq x, j\leq y} \text{I}(i,j)$$

where $\text{I}(x',y')$ is gray image value and $\text{II}(x,y)$ is integral image value.

There is a special property in integral image. If we want to get the sum of pixel of area yellow rectangle in Figure 2, we can use formula (2) to figure out quickly.

$$\text{sum}(m,n)=\text{II}(x,y)+\text{II}(u,v)-\text{II}(x,v)-\text{II}(u,y)$$

where $m$ and $n$ is the width and height of yellow rectangle. $m= x-u$, $n= y-v$

The integral image is used to calculate the value of any rectangular box. Generally speaking, it is simple to do 4 times of table lookup, and then simple "-" and "+" operation can get the sum of the grayscale in the rectangular area.

3. Scale spaces
Scale spaces are built like image pyramids. The image pyramid is a collection of images from the same original image whose resolutions are progressively reduced in the upper layer of pyramid. It is obtained by means of cascade downward sampling, and the sampling is stopped until a certain termination condition is reached. The bottom of the pyramid is a high-resolution representation of the image to be processed, while the top is a low-resolution approximation. We compare layers of images to a pyramid. Gaussian pyramid is obtained from the Gaussian smoothing and down-sampling for a series of images, that is to say, the layer $K$ of Gaussian pyramid was smoothed, and down-sampled for gaining the $K+1$ layer of Gaussian image, Gaussian pyramid contains a series of low-pass filter, its ending frequency from one layer to the next layer is the factor 2 gradually increase, so the gaussian pyramid can span a large frequency range. The pyramid looks like this in Figure 3.
In order to obtain the image pyramid with the level of $G_{k+1}$, the method steps are as follows:

- carries out gaussian kernel convolution for image $G_k$ and Gaussian blur.
- divides all even rows and columns.

The resulting image is $G_{k+1}$, which is obviously only a quarter of the original image. By iterating over the input image $G_k$ (the original image), you get the whole pyramid. At the same time, we can see that the downward sampling will gradually lose the information of the image. So that's the down sampling of the image, the narrowing of the image.

4. Haar-like feature

In 1910, Haar proposed the Haar wavelet function [3], one hundred years later in 2001, two scholars Viola and Jones were inspired by this function and published Rapid Object Detection using the method of Haar-like wavelet and integral image for detecting human faces [4].

The haarr wavelet function is the wave graph of the difference between two adjacent regions (see in Figure 4). The difference in the x direction is equivalent to the partial derivative of the x axis, and the difference in the y direction is equivalent to the partial derivative of the y axis.

$$F_{\text{Haar}} = E(R_{\text{white}}) - E(R_{\text{black}})$$

(3)

Where $E(R_{\text{white}})$ is the sum of the gray value in white area of the detection window and $E(R_{\text{black}})$ is the one of black area (shown in Figure 5).

Considering different lighting of the image, normalize the haar feature as follows

$$F_{\text{linear}} = \frac{E(R_{\text{white}}) - E(R_{\text{black}})}{w \cdot h \cdot \sqrt{E(R_{\text{white}})^2 - E(R_{\text{black}})^2}}$$

(4)

Where $w$ and $h$ is wide or height of haar window, $E(R_{\mu})$ is the sum of the gray value and $E(R_{\mu}^2)$ is the sum of gray value square of the image in the detection window.

The haar-like features [5] are obtained by extending the basic features of Haar (shown in Figure 6)
Haar like feature has the advantage
- High variability of different class
  The feature vary greatly between different classes
- Low variability of same class
  The feature of the same class coincide with each other
- Local strength difference
  The subtraction of gray value in two adjacent areas reflects the difference of local strength.
- Different scales space
  The haar window can be $2 \times 2$, $4 \times 4$, $8 \times 8$, $16 \times 16$, $24 \times 24$ to satisfy different scales space.
- High computational efficiency
  The feature is calculated out by integral image so it is independent of the window radius.

For example we choose haar like features to detect the vehicle in the following figure 7.

![Figure 7](image)

Figure 7. haar like features to detect the vehicle

5. Principle of Cascade Classifier
After selecting the features, the feature values of the sample data are input into the classifier to determine whether the car face is detected. Cascade classifiers are selected because of the large number of feature values. Cascade classifier is made up of some weak classifier. The weak classifier uses a single Haar feature to divide the sample data. For an example see in figure 8, we can get a line between positive and negative samples to separate the two categories. The line is the first weak classifier. The weight of the misclassified sample increases, and then another feature is chosen. The second weak classifier can be got and at this time the misclassified sample is right. But others sample is misclassified, then a new feature is chosen, the third weak classifier is obtained. The three weak classifiers are connected to form a strong classifier as following figure 9.

![Figure 8](image)

Figure 8. three weak classifiers

![Figure 9](image)

Figure 9. a strong classifier formed with three weak classifiers
6. Results of Test

In experiment we process video and image by vs2015 supported with OpenCV open-source library. We collect a large number of samples. There are now 2838 positive samples (shown in Figure 10), all normalized, 33×33 images and 7708 negative sample (shown in Figure 11), which were free size.

![Figure 10. positive samples](image1)

![Figure 11. negative samples](image2)

Training the weak classifier

- For each feature, calculate the feature values of the samples.
- Sort the eigenvalues
- For each element that is sorted
  - Calculate the weight sum of all positive examples $T^+$.
  - Calculate the weight sum of all negative examples $T^-$.
  - Calculate the weight sum of a positive example before some element $S^+$.
  - Calculate the weight sum of the negative examples before some element $S^-$.
- A certain feature value is selected as the threshold value to minimize the error

\[
e = \min (S^+ + (T^- - S^-), S^- + (T^+ - S^+))
\]  

(5)

Train strong classifiers

Set the weight for each weak classifier, calculate the weighted error rate of each weak classifier, and find the weight with the minimum error rate.

After get the cascade classifier, we test the detecting result by traffic video. The frame pictures in Figure 12 were captured from the traffic video. The detected vehicles are outlined in red rectangles.

![Figure 12. the results of detecting car from traffic video](image3)

Then we cut the image include the vehicle, for a 2 minute and 18 second video, a total of 2,048 car face images were segmented as shown in Figure 13. The number of vehicles in video was 115. Two buses were not successfully segmented, and the remaining vehicles were all present, with a success rate of 98%. The low success rate of buses is due to the lack of pictures of the training samples. From the experimental results, we can see the effectiveness of the car face segmentation method of the detection vehicle, which can meet the needs of real-time.
7. Conclusion
This paper presents an improved cascade classification algorithm based on haar-like features. Firstly, haar-like features of the image were calculated based on the idea of integral graph. Then the extracted haar-like feature value set is applied to train the cascading classifier. In this paper an adaptive threshold setting method is proposed. The experimental results show that the method proposed achieves good results in terms of detecting performance. Compared with the traditional method, it has better effect and better application prospect.

Acknowledgements
The authors gratefully acknowledge the support from Key research platforms and projects of universities in guangdong province (2018KTSCX157) and Guangdong Science and Technology Plan Project(2018KJYZ009) and the young teachers training of Guangdong police officer college(2018QNGG06)

References
[1] Zhang Tong, Peng Han, Yan Donsheng. The method of Vehicle type recognition based on improved SURF feature[C]. International Conference on Computer Graphis and Digital Image Processing.2017,7.
[2] Viola P, Jones M. Robust real-time face detection[J]. International Journal of Computer Vision, Published by Springer, 2004, 57(2): 137- 154.
[3] Haar, Alfréd. Zur Theorie der orthogonalen Funktionensysteme[J]. Mathematische Annalen[J], 1910, 69 (3), pp: 331-371
[4] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features[C]. in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp: 511-518.
[5] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection[C]. The IEEE International Conference on Image Processing. New York, USA, 2002, pp: 900-903.