Traffic Sign Recognition Based on HOG Feature and SVM

Jialin Tang
Beijing Institute of Technology, Zhuhai, 519088, China; City University of Macau, Macau, China
Zhuhai,China
email:01068@zhbit.com

Qinglang Su
City University of Macau, Macau, China
Macau,China
email:sonnysu@cityu.mo

*Chenyu Lin
Beijing Institute of Technology, Zhuhai, 519088, China; Beijing Institute of Technology, 10081, China
Zhuhai,China
email:325002702@qq.com

Yangjun Wen
Beijing Institute of Technology, Zhuhai, 519088, China
Zhuhai,China

Binghua Su
Beijing Institute of Technology, Zhuhai, 519088, China; Beijing Institute of Technology, 10081, China
Zhuhai,China

Juqing Yang
Beijing Institute of Technology, Zhuhai, 519088, China
Zhuhai,China

ABSTRACT

Traffic sign recognition is an important key technology in the unmanned driver technology. This paper uses TT100K traffic sign dataset to realize intelligent detection and automatic recognition of multi category natural road traffic signs. First we enhanced the image based on YUV color space and histogram equalization which solve the blurry or low brightness problem of the image; secondly, located the traffic sign area with Hough transformation based on the spatial characteristics of the image; then, using the SVM classifier to get the training model with HOG features extracted from traffic signs; finally, the test set and the number of ownership in TT100K dataset are used .Validated the reliability of the training model with test set and own data sample set in TT100K dataset, test results prove the reliability of training samples.

CCS CONCEPTS

•Computing methodologies •Computer graphics •Image manipulation •Image processing

EITCE 2020, November 6–8, 2020, Xiamen, China

ACM ISBN 978-1-4503-8781-1/20/11…$15.00
https://doi.org/10.1145/3443467.3443811

KEYWORDS

Traffic Sign Recognition, Machine Learning, SVM, HOG Features

1 Introduction

Traffic accidents have always been a major threat to people's safe and reliable assisted driving and unmanned driving technologies can greatly improve the safety, efficiency and comfort of people's travel. Reliable identification of natural road traffic signs provides an indispensable guarantee for the safety of auto-driving technology[3].

At present, experts have done a lot of research on traffic sign detection based on deep convolution neural network, support vector machine (SVM), and other methods. In 2012, Greenhalgh et al. proposed to extract the candidate areas of German traffic signs using the maximum stable extreme (MSER) classifier, and then trained the SVM classifier with the HOG feature extracted from the dataset of the candidate extreme areas[6]. Ultimately, their algorithm has achieved good results on GTSRB; after the advent of deep learning technology, German scientists used deep neural network to train traffic sign recognition model[7], which have got the accuracy exceeds the average level of human recognition. Support Vector Machine (SVM) with its excellent performance in classification and learning ability still has a place in the field of traffic sign recognition[2]. Local texture feature LBP and local gradient feature HOG are the common features in traffic sign recognition[1][5], HOG feature with the excellent gradient description ability has become a common recognition feature[4].

This paper presents a traffic sign recognition method based on HOG features and Support Vector Machine (SVM).

2 Segmentation and Enhancement of Traffic sign Image

2.1 Original Image Enhancement Based on YUV Color Space

2.1.1 Conversion of RGB and YUV color space. In order to enhance the image in YUV color space, the image obtained from the camera needs to be transformed to YUV color space according to formula (1) ~ (3).

$$Y = 0.299 * R + 0.587 * G + 0.114 * B \tag{1}$$

$$U = -0.147 * R - 0.289 * G + 0.436 * B \tag{2}$$

$$V = 0.615 * R - 0.515 * G - 0.100 * B \tag{3}$$

2.1.2 Histogram equalization of y-channel. After converting the original image into YUV image, extract Y channel to get gray image of the original image, and then get the gray histogram. From the graph, we can see that gray levels are mainly distributed in 40 to 180, which is very uneven, therefore histogram equalization is necessary to use.

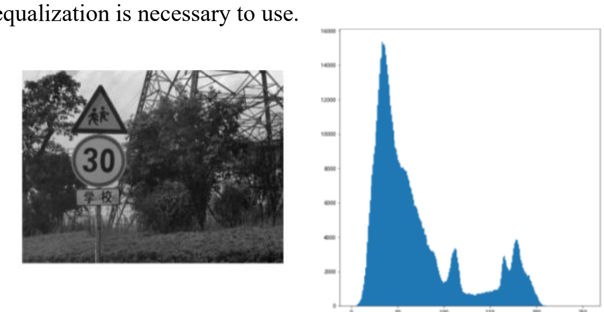

Figure 1: Y-channel Grayscale and Histogram of Original Image

Global histogram equalization probably results in loss of gray level of the image and affect accuracy, thus an adaptive equalization is used. First, divide the image evenly into many small blocks, than histogram equalization is performed for each block with its contrast limited. So that we can improving the image contrast without losing image details.

Figure 2: Grayscale and Histogram After Adaptive Equalization

Finally, the equalized Y-channel is merged with U-channel and V-channel than converted into RGB image to obtain the enhanced color image. As shown in Fig.3, the brightness and detail of the enhanced image are improved.

Figure 3: Image Before and After Processing (Left :Before Right :After)

2.2 Image Segmentation Based on HSV Color Space

2.2.1 Conversion of RGB and HSV Color Space. Normalize the values of R, G, B and convert the image from RGB space into HSV space by the formula (4) ~ (7).

$$V = \frac{(R+G+B)}{\sqrt{3}} \tag{4}$$

$$S = 1 - \frac{\sqrt{3}}{V}\min(\mathrm{R, G, B}) \tag{5}$$

$$H = \begin{cases} \theta, B \leq G \\ 360° - \theta, B \geq G \end{cases} \tag{6}$$

$$\theta = \arccos\left[\frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right] \tag{7}$$

2.2.2 Color Image Segmentation Based on Image Mask Technology. The image in HSV space is composed of three components: H, S and V. By restricting the H component and using image mask technology, we can obtain the color threshold segmented image of blue, yellow and red.

Based on the color distribution in HSV color space, the blue, yellow and red components in the "H" channel in the HSV color space are separated to obtain the image mask. And after removing the mask part, the binary image with color threshold segmentation is obtained. As shown in Fig.4, the top left is blue channel, the top right is yellow channel, the bottom left is red channel 1, and the bottom right is red channel 2

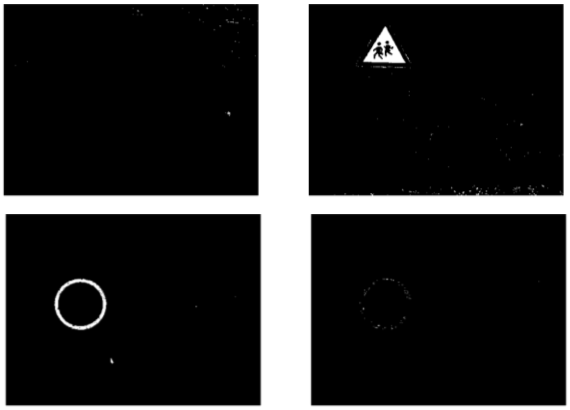

Figure 4: Image after Mask Removal

2.3 Locate and Intercept Traffic Signs

For circular traffic signs, Hough transformation is used. Hough transformation is an important method in graphic detection field. It transforms the detection problem in image space into the detection problem in parameter space based on the duality of point-line in image space and parameter space. It can effectively extract the edges of straight lines, curves, circles or even any other complex shape. It is widely used in various fields because of the high efficiency, high speed, low cost and reusability.

For triangle and rectangle traffic signs, because their edges are composed of straight lines, we first extracting the contour of all figures in binary image; secondly, In order to eliminate the noise produced by shooting angle, object occlusion, and some other factors in the contour， do polygon fitting on the contour. Finally filter the contours that the contour with 3 vertices is determined as a triangle, and the contour with 4 vertices is rectangular, The identification results are shown in Fig.5.

Figure 5: Result

3 Detection and Recognition of Traffic Signs

3.1 Construction and Enhancement of Data Sets

This paper chooses TT100K dataset as training set and test set. Because the number of flags in this dataset is single, before starting training, use translation, rotation and affine transformation to enhance the dataset, increase and balance the number of flags.

Finally, 22 common positive samples were selected from the enhanced dataset, including five classes of indicator signs (including straight, left, right turns, etc.), 12 classes of prohibition signs (including speed limits, no parking signs, no traffic signs), 4 classes of warning signs (including intersections, zebra lines, school sections, etc.), and also background maps without traffic signs was selected as negative sample.

3.2 Calculation of HOG Features

HOG is a feature descriptor commonly used for target detection in computer vision and image processing. It has the stability of less offset to illumination variation and gradient feature, and can well represent the outline and edge information of traffic signs.

In the process of extracting HOG features, the image is first color normalized, and then the gradient of the traffic sign image in the horizontal and vertical directions of the pixel points (x, y) as the formula (8) ~ (11)

$$G_x(x,y) = I(x+1,y) - I(x-1,y) \tag{8}$$

$$G_y(x,y) = I(x,y+1) - I(x,y-1) \tag{9}$$

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{10}$$

$$a(x,y) = \arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{11}$$

$G_x(x,y)$, $G_y\,(x,y)$ represents the horizontal and vertical gradients of the image at the pixel points (x,y), and $G\,(x,y)$ and $\alpha\,(x,y)$ represent the gradient magnitude and gradient direction of the pixel points (x,y).

Divided the image into several cells, collect the gradient direction histogram of each pixel point in each cell, then the cells are blocked and normalized, and then the feature vectors in all the blocks are concatenated to get the final HOG feature vectors, whose feature dimension N is:

$$N = \left(\frac{I_w}{C_w} - 1\right) \times \left(\frac{I_h}{C_h} - 1\right) \times B \times H \tag{12}$$

I_w and I_h are the length and width of the traffic sign image, C_w and C_h is the length and width of the cell. B is the number of cells contained in each block.

In this paper, the image size in the training sample is modified to 64x64, and every 8x8 pixels in the image is selected as a cell. 4 adjacent cells are used as blocks to calculate the gradient information in 9 directions. Therefore, respectively $C_w = C_h = 8$, $B = 4$, $H = 9$. HOG feature dimension is 1764.

3.3 SVM Classification Training

Support vector machine (SVM) is a pattern recognition method based on statistical learning theory. Its basic idea is to use kernel function to map the data in the input space into a high-dimensional feature space, so that the linearly inseparable data in the original space can be classified linearly in the high-dimensional space, and then then by minimizing the structural risk, an optimal hyperplane is constructed in high-dimensional space, so that the interval boundary has sufficient space to contain test samples, to minimize the empirical risk, and to improve the learning ability of support vector machines. Therefore, in the case of a small number of statistical samples, good classification results can also be achieved.

3.3.1 Construction of Optimal Hyperplane. The optimal hyperplane is constructed by the following method. In training data, taking n-dimensional space as an example, each data has n attributes and positive or negative category labels. The whole training data is divided into positive and negative categories through n attributes, and the separated data belongs to different categories. There are many classified hyperplanes to solve this problem. In order to get a better classification result, we need to find a classification surface so that points of different classes that are close to the classification surface have the largest separation from the surface (optimal classification hyperplane)

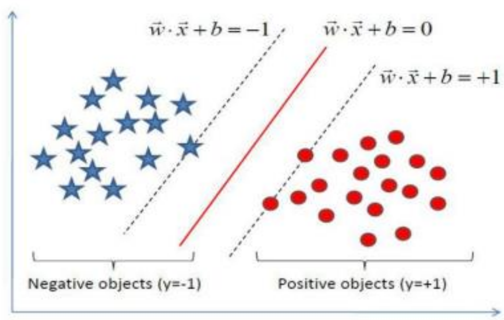

Figure 6: Optimal Classified Hyperplane Diagram

The optimal classification hyperplane is shown in Fig.6, where the lower circle points represent positive classes and the upper asterisk represents negative classes.

So the training process of support vector machines (SVM) can be viewed as an optimization problem, that is, to minimize the function 13.

$$\frac{\|\omega^2\|}{2} + C\sum_{i=1}^{l}\varepsilon_i \tag{13}$$

ω is the normal vector of the sample separation surface, ε is the relaxation parameter, l is the number of samples, and C is the penalty factor.When the ε is constant, the larger the C is, the greater the impact on the objective function and the higher the accuracy of training samples, but the possibility of over fitting is greater. Conversely, lowering the C value appropriately can result in some misclassified samples and improve the generalization ability of the classifier. Fig. 7 shows the effect of penalty factors C on the accuracy of the test in this experiment.

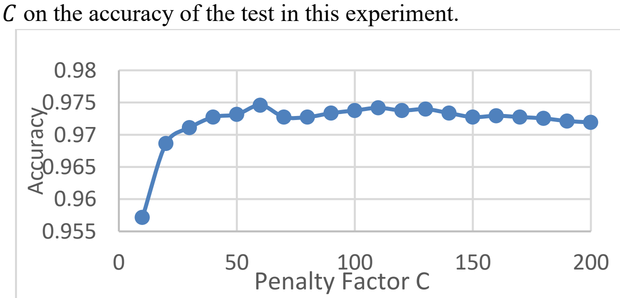

Figure 7: The influence of penalty factor (C) on test accuracy

As can be seen from Fig.7, the SVM classifier has the highest accuracy when C =60.

3.3.2 Selection of Kernel Functions. As mentioned above, the kernel function is used to transform the linear nonseparable data in low-dimensional space into the linear separable data in high-dimensional space and to find the optimal hyperplane. Kernel functions should be chosen according to specific problems. The common kernel functions are: Linear Kernel Function, Polynomial Kernel Function, Radical Basis Function (RBF) and Sigmoid Kernel Function.

Table 1 shows the accuracy of using different kernel functions when C = 60.

Table 1: Accuracy of Different Kernel Functions

Kernel Function	Linear	Polynomial	RBF	Sigmoid
Accuracy (%)	97.233	20.578	97.458	97.110

Therefore, the training classifier used in this paper selects Radical Basis Function with a penalty factor C of 60 and the accuracy test score is 97.458%.

4 Conclusion

This section selects some images from CCTSDB dataset and uses the SVM classifier based on TT100K dataset to detect and identify traffic signs. The results are as follows.

Figure 8: Results

Fig.8 shows some examples of good recognition results, where traffic signs can be recognized clearly in sunny, rainy, and foggy weather.

Figure 9: Results with Occlusion

Fig.9 shows an example of successful identification with partial occlusion, where a stop sign can still be identified after a small part of the non-main body is covered by a branch.

Figure 10: Defect Results

Fig.10 shows some recognition errors or unrecognized results. The speed limit 50 traffic sign on the left is incorrectly identified as other signs when it is obscured. The stop sign on the right is unrecognized due to severe reflections.

Therefore, it can be concluded that the SVM classifier has higher recognition accuracy when the color difference between background image and flag image is large; when part of the flag is occlusion, the system also has the ability to identify the flag. The speed limit 50 traffic flag in Fig.10 which contains more detailed information failed to identified, while the successful recognition of the stop-stop traffic flag in Fig.9 contains less detailed information, so it can be concluded that traffic signs with more details are easier to identify errors when blocked.

It can be concluded from the error identification sample that the major factors influencing traffic sign detection are small target area, interference from other objects and motion blurring. In addition, areas with the same color as traffic signs and objects with the same color and shape of traffic signs will affect the reliability of detection results.

ACKNOWLEDGEMENT

This work is supported by the project of Macao Higher Education Fund (OTH1903)，the Key Program of Guangdong University (NO. 2019KZDXM060) and a grant from the Innovative Youth Program of Guangdong University (NO. 2019KQNCX194).

REFERENCES

[1] Jie W ,Li W X ,Li W. Traffic signs real-time classification and recognition based on multi - feature fusion [J]. Modern Electronics Technique ,2019,42(11):50-53+58.

[2] Jiang J H ,Bao S L ,Shi W D ,Wei Z K . Improved traffic sign recognition algorithm based on YOLOv3 algorithm [J/OL]. Journal of Computer Applications:1-8[2020-05-28].

[3] Zhu Z , Liang D , Zhang S , et al. Traffic-Sign Detection and Classification in the Wild[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.

[4] Wang B ,Chang F L ,Liu C S .Rapid traffic sign detection based on MSER and SVM [J].Journal of Optoelectronics·Laser,2016,27(06):625-632.

[5] Han X X ,Wei M ,Xu X Y ,Li Q Y,Chen X ,Zhu H C . Traffic Sign Recognition Algorithm Based on Multiple-Feature Fusion [J]. Computer Engineering and Applications,2019,55(18):195-200.

[6] Chung J H , Kim D W , Kang T K , et al. Traffic Sign Recognition in Harsh Environment Using Attention Based Convolutional Pooling Neural Network[J]. Neural Processing Letters, 2020(3).

[7] Khan J A , Chen Y , Rehman Y , et al. Performance enhancement techniques for traffic sign recognition using a deep neural network[J]. Multimedia Tools and Applications, 2020(8).