Research on Numeral Recognition Method Based on HOG Features

Yichun Jiang[1,a], Ziwei Cui[1,b] and Jingqi Fu[1,c]

[1]Changchun University of Science and Technology National Demonstration Center for Experimental Electrical and Electronic Technology ChangChun, JiLin ,130022, China.

[a]931645699@qq.com, [b]525706946@qq.com, [c]321918448@qq.com

Abstract

The current numeral recognition method has high computational complexity and large amount of parameter, which cannot meet the needs of real-time detection. Therefore, a numeral recognition method based on machine learning is designed. The method uses canny operator and morphological method to extract the numeral target in the image for obtaining the candidate box. After the candidate box is filtered, the HOG feature of the key region are calculated. Then,the feature vectors are sent to the support vector machine for classification to realize the location and recognition of numbers in the image. The experimental results show that the method has better real-time performance and good detection accuracy, and the method can meet the real-time requirements of numeral detection such as license plate recognition and handwriting recognition.

Keywords

Target Recognition, HOG, SVM, Numeral Recognition.

1. Introduction

At present, the application range of numeral recognition technology is very wide, such as handwriting function of smart phones and license plate detection.The numeral reconiton is always numeral recognition one of the research hotspots in computer vision. Although the current numeral recognition method based on neural network has made a breakthrough [1,2], its large amount of parameters and high computational complexity lead to slower training speed and computational speed. The traditional algorithm extracts the artificially features in the image, and then performs classification or linear regression through support vector machines or similar classifiers. The accuracy of traditional algorithm can meet the needs of daily applications, and it has more advantages in training time and computing time.

In the context, this paper has launched a research on HOG feature-based numeral recognition methods.The research focuses on the analysis of HOG feature extraction methods and parameters. Then it analyzes the influence of HOG feature parameters on SVM classification results and selects the best parameters. Finally, the method is implemented on matlab for numeral recognition in actual captured images.

2. Basic Principles

2.1 HOG Feature Descriper

Histograms of Oriented Gradients (HOG) is a feature descriper proposed by Navneet Dalal and Bill Triggs, which is often used for feature extraction and classification of various targets [3,4]. The core idea is to calculate the statistical value of the direction information of the partial image, use these direction information as the feature of the image, and finally combine the features into a feature vector and input the classifier for classification.

The HOG algorithm performs grayscale processing on the image firstly, and eliminates the effects of local shadows and light changes by using gamma correction. Then calculate the horizontal gradient value $G_x(i,j)$ and the vertical gradient value $G_y(i,j)$ of each pixel [5].

$$G_x(i,j) = I(i+1,j) - I(i-1,j) \tag{1}$$

$$G_y(i,j) = I(i,j+1) - I(i,j-1) \quad (2)$$

Then calculate the gradient magnitude $G(i,j)$ and gradient direction $\theta(i,j)$ of the pixel by the following formula:

$$G(i,j) = \sqrt{G_x(i,j)^2 + G_y(i,j)^2} \quad (3)$$

$$\theta(i,j) = tan^{-1}(\frac{G_y(i,j)}{G_x(i,j)}) \quad (4)$$

HOG divides the image into 8x8 cells, 2x2 cells are combined into a block, and the gradient direction is quantified within 0°-180° (Regardless of gradient direction) or 0°-360° (considering the gradient direction), then the gradient direction of all pixels in each block is weighted by the gradient amplitude to form a gradient histogram. Finally, the gradient histogram of each block is normalized to balance the histogram distribution of different blocks in the entire detection window and reduce the influence of factors such as light and shadow. The HOG descriptors of all blocks in the detection window are combined to form the final feature vector. The extracted feature vector is input into the SVM classifier for classification operation.

2.2 Support Vector Machines

Support Vector Machine (SVM) is usually used for pattern recognition, classification, and regression analysis [6,7]. Its basic model is a linear classifier with the largest interval defined in the feature space. The largest interval makes it different from the perceptron. Since the distance between the finally obtained optimal hyperplane and the data on both sides is the largest, the use of this hyperplane to classify the two-class samples can maximize the anti-interference ability and obtain better generalization ability. The hyperplane solved by the support vector machine has nothing to do with most data, but it is related to a few support vectors. This is also the source of the name of the support vector machine.

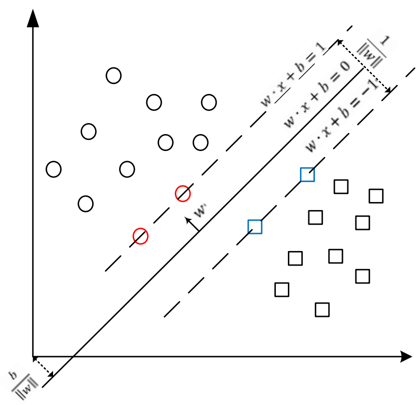


Fig. 1 Schematic diagram of linear SVM

3. Numeral Recognition Method

The paper proposes and implements a numeral recognition method based on HOG features. First, the RGB image is changed to a grayscale image, and then the edge is extracted by the canny operator [8]. After filling, the open operation in the morphological operation [9] is used to filter out the thinner edges in the picture. After calculating the image mask to obtain the candidate frames, the images in each candidate frames are uniformly down-sampled to 28x28 resolution to obtain their HOG features. Then the HOG feature vector is used as the input of each SVM classifier to classify each SVM. The classification result of the detector is processed, and the image recognition result is finally obtained.

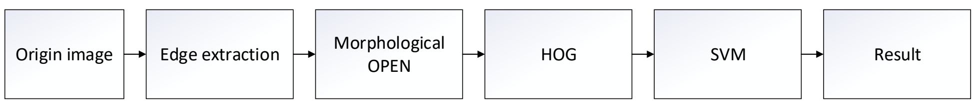


Fig. 2 Algorithm flow chart

3.1 Edge Extraction and Opening Operation

Since there is background part without targets in the image other than the image with numeral, it will consume a lot of time if the commonly used window sliding method is used to traverse the image through the window. Considering that the numbers are usually located in a simple background, such as walls, paper, license plates, edge extraction algorithms are used to extract the outlines of the numbers. Figure 2 shows the effect test of various edge extraction operators.

(a) (b) (c) (d) (e)

Fig. 3 Comparison of edge detection experiment results:(a)Origin(b)LoG(c)prewitt(d)sobel(e)canny

It can be seen through experiments that the canny operator has the best detection effect . The canny operator can extract the complete contour of the number, so it is used to extract the target contour in the method. But the canny operator will also extract some of the edges that are not belong to numbers, and these edges need to be filtered out, or it will increase unnecessary calculations later. Therefore, the method first fills the closed area, and then performs an opening operation on the image, which can eliminate the thinner edges. Figure 4 shows the result of the open operation.

(a) (b)

Fig. 4 Comparison of open operation processing results:(a)before(b)after

After obtaining the mask as Fig. 4, by dividing the regions, different candidate frames can be calculated and the images in the candidate frames will be resized to 28x28 for subsequent recognition work.

3.2 SVM kernel function and multi-class strategies

This method uses the SVM classifier. The SVM method requires inner product calculation in the feature space, and what the kernel function actually does is an inner product operation.The choice of kernel function is directly related to the performance of SVM [10], Especially when the data is linearly inseparable. Therefore, the kernel function needs to be selected before SVM training.

SVM generally uses linear kernel, Gaussian kernel or sigmoid. If the sigmoid kernel function is used, the SVM actually implements a multilayer neural network, so the method does not analyze the sigmoid kernel function. The method focuses on the analysis of the linear kernel and the Gaussian kernel, using 20000 samples in MNIST dataset for training, and collecting the data such as training time, average test time and recognition accuracy of the two SVM classifiers. The specific data is shown in Table 1.

Table 1 Performance comparison of kernel functions

Kernel	Linear	Gaussian
Training time(s)	136.39	770.24
Ave. testing time(ms)	18	510
Correct rate	95.6%	97.2%

Based on the above research, It can be concluded that the feature space composed of HOG features is more complex, and the feature space is not linearly separable. Therefore, the use of Gaussian kernel can improve the accuracy of recognition to a certain extent. However, because the Gaussian kernel has more parameters and more complex calculations, the training time is almost 6 times that of the linear kernel function, and the average detection speed is even more difficult to meet the needs of real-time detection. Although the detection accuracy of the linear kernel is a little lower than that of the Gaussian kernel function, and it also has better training and test time performance. Therefore, the method uses the linear kernel.

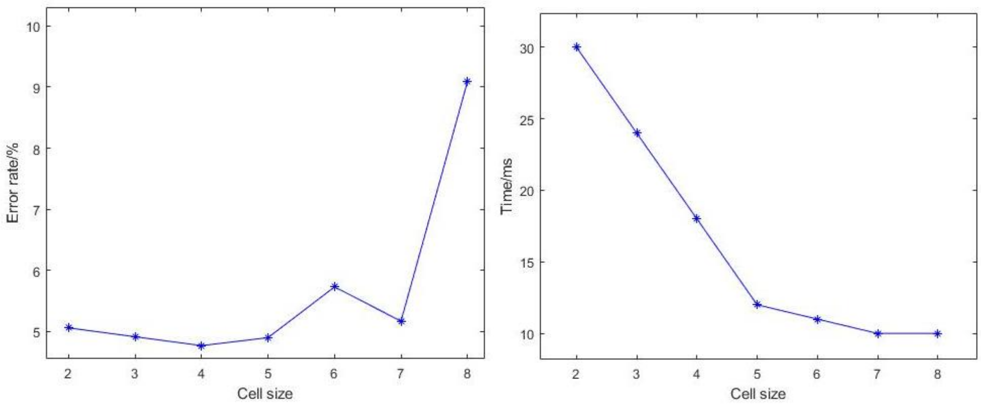
The multi-class strategy of SVM is mainly divided into two types: OVO (One VS One) and OVR (One VS Rest) [11]. The OVO strategy will train a one-to-one classifier for every two different numbers in the 10 numbers, so 45 classifiers are needed.The OVR strategy takes all numerals except the target numeral as negative samples, and trains a classifier, so that only 10 classifiers are needed to classify and recognize 10 numerals. According to the test, the two combination strategies have similar or even the same accuracy performance. The detection time of the OVR combination strategy is shorter. Therefore, this method uses the OVR multi-class strategy to train the classifiers. If multiple classifiers judge that a numeral belongs to the category they are responsible for at the same time, the method will use the classifier with the lowest error rate as the standard.

3.3 Analysis and improvement of HOG features

The method selects HOG features to extract the numeral features in the image, and it needs to analyze and improve the numeral image features. Since the magnitude and direction of the gradient between the number and the background in the actual image are uncertain, it should not be distinguished between positive and negative when calculating the gradient direction.Moreover, the original HOG cell size is 8x8, which is based on image pixels used in pedestrian detection. But this method uses the MNIST handwritten digit training set of the National Institute of Standards and Technology. A total of 60,000 sheets, each with a size of 28*28 pixels grayscale images. The 8x8 size cell can only obtain 144-dimensional feature vectors, so the detection error increases accordingly. Therefore, it is necessary to study the influence of different size cells on the SVM classification result to determine the optimal cell size and obtain better accuracy performance. The test results for different cell sizes are shown in Table 2. Figure 5 is also used to show the relationship between cell size and detection accuracy and detection time more intuitively.

Table 2 The influence of different cell size on test results

Cell size(pixel)	2x2	4x4	5x5	6x6	7x7	8x8
Error rate	5.06%	4.77%	4.90%	5.73%	5.17%	9.09%
Time(ms)	30	18	12	11	10	10



(a) (b)

Fig. 5 The influence of cell size on the detection result: (a) the influence on the error rate (b) the influence on the detection time

It can be seen from the figure that when the pixel size is 4x4 and 7x7, its error rate is often lower than other adjacent size pixels, and obviously 4 and 7 are factors of 28. This is mainly because the image is completely segmented by the unit, and no information is lost. When the cells is smaller, the number of gradient histogram statistics is more and more feature vectors are generated, which usually reduces the error rate. Therefore, when the size of cells is 4x4, the method has the best error rate performance, and its calculation time is also at a relatively low level, so 4x4 cells are used to calculate the HOG feature vector. As shown in Figure 6, HOG features are shown here, where HOG features with appropriate size of cell on the surface can better describe the outline of the number.

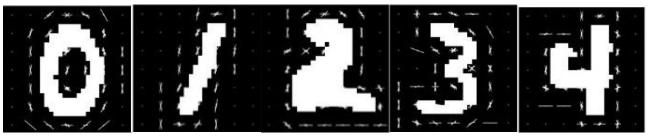


Fig. 6 Visualized HOG features of numeral images

In order to further improve the accuracy of HOG features for feature classification, the paper introduces the pixel information of the image as a additional feature vector. However, if the gray values of all pixels are used as the feature vector, the dimensionality of the feature vector will be greatly increased, and more useless information will slow down the calculation speed. Therefore, in the method, the average value of gray values of each row and each column is obtained separately, and these average values are useed as feature vectors. In this way, the dimension of the feature vector is only increased by 56. Experimental results show that the error rate can be reduced by about 0.5%, and the calculation time is still maintained at about 18ms.

4. Experimental Results and Analysis

In order to verify the effectiveness of the method in the paper, the paper uses the SVHN dataset [12] to verify the numeral recognition method. The experiment is carried out on a computer equipped with Intel Core i7-10875H CPU, 16GB RAM, and Windows10 operating system. The simulation software uses MATLAB2019a.

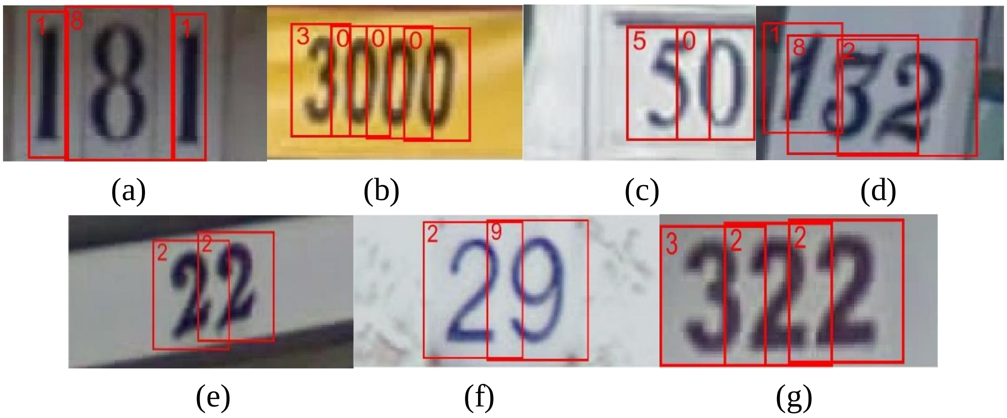


Fig. 7 Numeral detection results in different Scenes

It can be seen from Figure 7 that the method has generally achieved good detection accuracy. In most scenes, it can detect and recognize the numbers in the image, even when the background has some interference. For example, The door frame in Figure 7(a), (c), (d) and the pattern in Figure 7 (f) did not affect the identification and detection of numbers, all of the numerals in images are completely and correctly detected. It can be seen from Figure 7(d) that the actual number in the figure is 3 and the predicted result of this method is 8. Therefore, in some cases, due to the problem of different font shapes at the angle of shooting, this method may be in the recognition process An error occurred in.

5. Conclusion

In order to meet the real-time detection requirements of license plate detection and handwriting detection, a numeral recognition method based on HOG features is proposed. The paper uses canny edge extraction and morphological operations to extract the key regions that may contain numeral targets in the image. Then it uses improved HOG feature extraction and support vector machines to classify and recognize the extracted regions. Experimental results show that the method can efficiently and accurately extract the location information and category of the target numerals in the image. The method has a good detection accuracy and detection speed. In the future work, we will continue to carry out research, and establish a special data set for training to solve the problem of false detection between similar numerals. And we will also conduct a more in-depth analysis of the HOG feature space, and study the methods to more effectively segment positive and negative samples method.

References

[1] Gupta D, Bag S. CNN-based multilingual handwritten numeral recognition: A fusion-free approach[J]. Expert Systems with Applications, 2020: 113784.

[2] Zhao H, Liu H. Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition[J]. Granular Computing, 2020, 5(3): 411-418.

[3] Nongmeikapam K, Kumar W K, Meetei O N, et al. Increasing the effectiveness of handwritten Manipuri Meetei-Mayek character recognition using multiple-HOG-feature descriptors[J]. Sādhanā, 2019, 44(5): 104.

[4] Challa A. Automatic Handwritten Digit Recognition On Document Images Using Machine Learning Methods[J]. 2019.

[5] Singh G, Chhabra I. Effective and fast face recognition system using complementary OC-LBP and HOG feature descriptors with SVM classifier[J]. Journal of Information Technology Research (JITR), 2018, 11(1): 91-110.

[6] Salam M, Hassan A A. Offline isolated arabic handwriting character recognition system bsed on SVM[J]. Int. Arab J. Inf. Technol., 2019, 16(3): 467-472.

[7] Deore S P, Pravin A. Histogram of Oriented Gradients Based Off-Line Handwritten Devanagari Characters Recognition Using SVM, K-NN and NN Classifiers[J]. Revue d'Intelligence Artificielle, 2019, 33(6): 441-446.

[8] Yuan W, Zhang W, Lai Z, et al. Extraction of Yardang Characteristics Using Object-Based Image Analysis and Canny Edge Detection Methods[J]. Remote Sensing, 2020, 12(4): 726.

[9] Nair V, Ram P G K, Sundararaman S. Shadow detection and removal from images using machine learning and morphological operations[J]. The Journal of Engineering, 2019, 2019(1): 11-18.

[10] Abdollahi S, Pourghasemi H R, Ghanbarian G A, et al. Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions[J]. Bulletin of Engineering Geology and the Environment, 2019, 78(6): 4017-4034.

[11] Chandra M A, Bedi S S. Survey on SVM and their application in image classification[J]. International Journal of Information Technology, 2018: 1-11.

[12] Guo C, Wu R, Weinberger K Q. TrojanNet: Embedding Hidden Trojan Horse Models in Neural Networks[J]. arXiv preprint arXiv: 2002. 10078, 2020.