Atlantis Highlights in Engineering, volume 3
3rd Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC 2018)

A Novel Pedestrian Detection Method Based on Histogram of Oriented gradient and Support Vector Data Description

Lei Ke, Yanping Dai
College of Transportation and Automobile Engineering, Panzhihua University, Sichuan Province
Panzhihua, China
78622976@qq.com, 253906676@qq.com

***Abstract*—Pedestrian detection is an important research field of computer vision, and is a key technology of automatic drive. One category of pedestrian detection methods is statistics learning methods. But these methods cannot guarantee that the negative samples include all real scenes in practice. And big negative data lead to a complex classifier. Compared to negative samples, positive samples are more easily to be complete. To solve the above problem, we propose a novel pedestrian detection method based on Histogram of Oriented Gradient (HOG) and support vector data description (SVDD). First, create multi-scale samples by shift window. Second, samples are transformed by HOG method to provide features for classifiers. Third, a single classifier is trained with complete positive samples. The single classifier is based on SVDD. A series of experiments show that the proposed method needs more positive samples to train the classifier compared to previous works, but the number of total samples is less than that of previous works. And the false detection of the proposed method is less than that of classical method.**

Keywords—histogram of oriented gradient; support vector data description; pedestrian detection; statistics learning

I. INTRODUCTION

Pedestrian detection is an important research direction in computer vision, especially in the field of intelligent city and digital traffic. With the development of autonomous driving and intelligent driving, automatic pedestrian detection algorithm has attracted extensive attention and research.

It is very difficult to detect pedestrians. The difficulty of detecting objects mainly depends on the difference between these objects. Pedestrian detection is a unique branch of object detection. It is unique for three reasons. Firstly, it has great potential application value. Secondly, the human body has good flexibility and a certain degree of rigidity. Pedestrian detection studying can help us better study other objects. Thirdly, people dress differently, and from the perspective of color, the differences of pedestrians are extremely rich. Therefore, to study pedestrian detection is to study how to grasp essential features in the case of huge difference in appearance.

The pedestrian detection algorithm was mainly completed by two technical means: Background-based approach and statistical- based learning method[1][2][3].The method based on background modeling was greatly influenced by external factors such as illumination and scene, but it is not very robust [4].Papageorgiou[5] proposed the multi-scale Haar wavelet[6] and Support Vector Machine (SVM)[7] for pedestrian detection, and proposed the sliding window method. Viola et al.[8] used the integral graph for detection, and adopted the algorithm of Adaboost for classification. Dalal and Triggs[9] proposed the Histogram of Oriented Gradient (HOG) method to describe image samples and classify them by SVM. All of the above classifiers required negative samples for training. However, in the actual situation, the negative sample used for training cannot contain all the actual scenarios, which limited the generalization ability of the above algorithm. The current solution is to increase the number of negative sample training, which will make the model more complex, the training and solution more difficult.

In practice, it is relatively easy to guarantee the completeness of positive samples. Therefore, we solve the problem of training negative sample incompleteness in statistical learning algorithm from the Angle of mining complete positive sample similarity. This paper proposed a pedestrian detection algorithm based on HOG[9] and support vector data description[10]. Firstly, a multi-scale sliding window was designed to generate multi-scale candidate region samples. Secondly, the above samples were sampled into uniform size and histogram statistics of gradient direction were conducted. Thirdly, based on the data description of the support vector domain, a single classifier was developed by using the complete positive sample training to carry out pedestrian detection. The advantage of the method was that there was no need to introduce the negative sample set for training, thus avoiding the impact of the incomplete negative sample on the classifier.

Finally, a series of comparative experiments were conducted in this paper. Compared with the classic HOG and SVM algorithms, the algorithm in this paper required more positive samples for training when the performance was equivalent, but the total number of samples required was less. Furthermore, the algorithm was easier to avoid error detection.

II. ALGORITHM FLOW CHART

The core idea of this algorithm was to fully explore the similarity features of positive samples under the premise of ensuring positive samples completeness. The algorithm was divided into two modules: feature extraction and clustering.

From the perspective of algorithm execution, it was divided into training part and application part. The training of the model required a manually marked positive sample set, samples were sampled to the normalized size (64 x 64), and a single classifier was trained based on SVDD by using its HOG feature.

In the application part of the algorithm, the variable size sliding window was applied to the input image, the candidate sample was proposed, it's HOG feature was extracted, and input into the trained single classifier for classification. The overall flow chart of this algorithm was shown in fig.1.

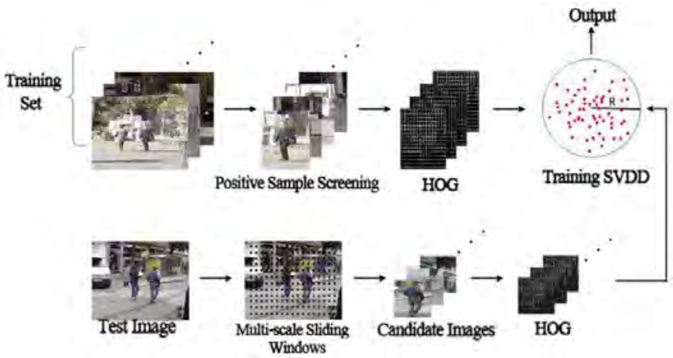

Fig. 1. Algorithm Flow Chart

III. HISTOGRAM OF GRADIENT DIRECTION

The histogram of gradient direction is divided into four stages: preprocessing, gradient calculation, gradient statistics and HOG descriptor generation.

The pre-processing stage was to use the Gamma algorithm to adjusted the impact of illumination and contrast on the image. The formula of the Gamma algorithm was as follows:

$$I(x,y)' = I(x,y)^{Gamma} \tag{1}$$

In the formula, I(x, y) Represents the pixels in the image, Gamma generally takes 0.5.

Formula (2) is the calculation formula of gradient

$$\begin{aligned} &G_x(x,y) = I(x+1,y) - I(x-1,y) \\ &G_y(x,y) = I(x,y+1) - I(x,y-1) \\ &G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \\ &\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \end{aligned} \tag{2}$$

In the formula, G(x, y) represents the gradient, α (x, y) represents the gradient direction。

After calculated the gradient of each point of the image, the local image (6x6) was encoded, and the number of quantization directions in this paper was 9. In this way, each local region generated an encoding that characterizes the image. The description of the whole image was generated by integrate the directional gradient statistics of each local region. The HOG feature was operated on the local unit area of the image, which can maintain certain stability for the geometric and optical deformation of the image. A schematic diagram of HOG feature extraction was showed in fig.2.

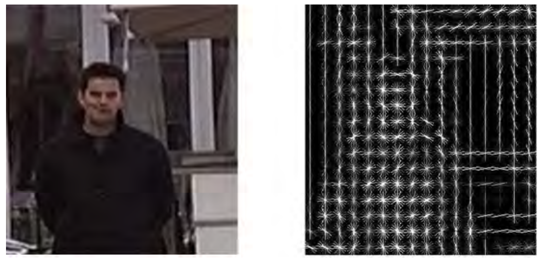

Fig. 2. HOG Feature Extraction Schematic

IV. SUPPORT VECTOR DOMAIN DATA DESCRIPTION

The support vector domain[10] can be regarded as a clustering algorithm, and the "spherical" hyperplane can be trained according to the training samples, the high-dimensional space can be divided into two parts, so it can be used for classification. Unlike support vector machines, support vector field data representation does not require negative samples to participate in training, but to train the boundaries of positive samples. Therefore, the key to the application of support vector field data description in practice is to ensure the completeness of positive samples.

A. *Training of data description in support vector domain*

Supposing sample set is described by $X=\{x_i, i=1,2,...,M\}, x_i \in R^N$, dimension of the sample is N. Suppose the center of the "spherical" hyperplane is α , the radius is R, and this hyperplane satisfied equation (3).

$$(x_i - a)(x_i - a)^T \le R^2 \tag{3}$$

The training support vector domain can be regarded as solving a constrained optimization equation, as shown in equation (4).

$$\begin{aligned} \min \quad & F(R,\ a) = R^2 \\ s.t. \quad & R^2 \ge (x_i - a)(x_i - a)^T \end{aligned} \tag{4}$$

However, in practical application, we need to allow the influence of noise. Therefore, on the basis of equation (4), the penalty term ε_i is introduced, such as equation (5).

$$\begin{aligned} \min \quad & F(R,\ a,\ \varepsilon) = R^2 + C\sum_{i=1}^{M}\varepsilon_i \\ s.t. \quad & R^2 + \varepsilon_i \ge (x_i - a)(x_i - a)^T \\ & \varepsilon_i > 0 \end{aligned} \tag{5}$$

Where, C is the penalty coefficient. Using KKT algorithm[11], the Lagrangian operator was introduced, and formula (5) can be converted to formula (6).

$$\begin{aligned} &L(R,\ a,\ \varepsilon,\ \alpha) = R^2 + C\sum_{i=1}^{M}\varepsilon_i \\ &-\sum_{i=1}^{M}\alpha_i[R^2 + \varepsilon_i - (x_i^2 - 2ax_i + a^2)] - \sum_{i=1}^{M}\gamma_i\varepsilon_i \end{aligned} \tag{6}$$

Take the partial derivative of the above equation, substitute the result into the above equation, and equation (7) can be obtained.

$$L(R,\ a,\ \varepsilon,\ \alpha)=\sum_{i=1}^{M}\alpha_i(x_i g x_i)-\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j(x_i g x_j) \tag{7}$$

In the practical application, we introduced the kernel function to map the linear space of low dimension to the nonlinear space of high dimension, and formula (7) was rewritten as formula (8).

$$L(R,\ a,\ \varepsilon,\ \alpha)=\sum_{i=1}^{M}\alpha_i K(x_i g x_i)-\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j K(x_i g x_j) \tag{8}$$

By solving the above equation, the optimal α_i can be obtained which can minimize formula (8), and the "spherical" hyperplane can be obtained. And the radius R was obtained from α_i . as showed in fig.3, the radius R was the final training result.

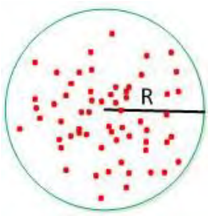

Fig. 3. Diagram of Training Results

In the actual application, we extracted the HOG feature from the positive sample set and get the training sample set $X=\{x_i, i=1,2,...,M\}$. In the actual situation, there was no zero sample. According to equation (8), obtained $A=\{\alpha_1,\alpha_2,...,\alpha_M\}$,where the element corresponds to the element of X. Suppose $X'=\{x_1,x_2,...,x_N\}$, satisfied $\forall x_i \in X', 0<\alpha_i<C$, substitute the element of X' into equation (8), and the radius R can be calculated as:

$$R=1+\frac{1}{N}\sum_{n=1}^{N}(-2\sum_{i=1}^{M}(\alpha_i K(x_n,x_i)))+\sum_{i}^{M}\sum_{j}^{M}\alpha_i\alpha_j K(x_i,x_j) \tag{9}$$

B. Support vector domain data description is used for classification

For the samples x_u to be classified, calculated the distance to the center of the hyperplane "spherical". If equation (10) is satisfied, judged as a positive sample; otherwise, as a negative sample.

$$(x_u-a)(x_u-a)^T \le R^2 \tag{10}$$

In practice, equation (10) was modified and obtained equation (11).

$$K(x_u,x_u)-2\sum_{i=1}^{M}\alpha_i K(x_u,x_i)+\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j K(x_i,x_j)-R^2\le 0 \tag{11}$$

V. EXPERIMENTAL

A series of experiments were carried out to verify the effectiveness of the above algorithm. The experiment was conducted under MatlabR2013a, and the computer performance parameters were: Intel(R) Core(TM) i5-3470 CPU, 4.gb RAM, 64-bit operating system. The images used in the experimental simulation were derived from data sets INRIA and MIT pedestrian detection data sets.

A. Algorithm results

Two typical results were showed in fig.4 and fig.5. Red box was the correctly detected pedestrian and white box was the error detection. The training positive sample was 5000 and the test positive sample was 768. The detection results of the algorithm in this paper were showed in table Ⅰ.

TABLE I. ALGORITHM TEST SET OUTPUT RESULTS

	Detection rate	Error detection rate
HOG+SVDD	88.5%	1.1%

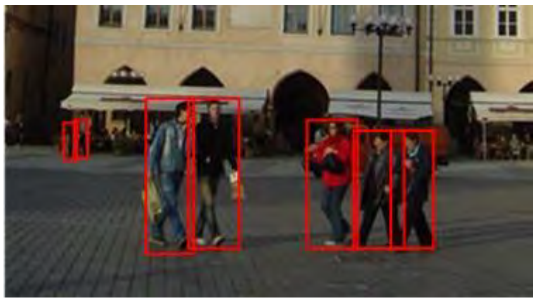

Fig. 4. Algorithm Test Result I

For regular pedestrians in simple scenes, the algorithm in this paper can correctly detect and locate them relatively accurately. Moreover, this algorithm can also adapt to the influence of target distance and scale.

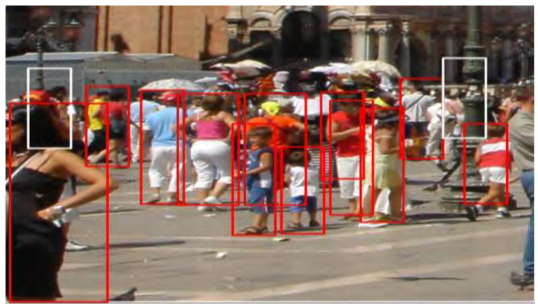

Fig. 5. Algorithm Test Result II

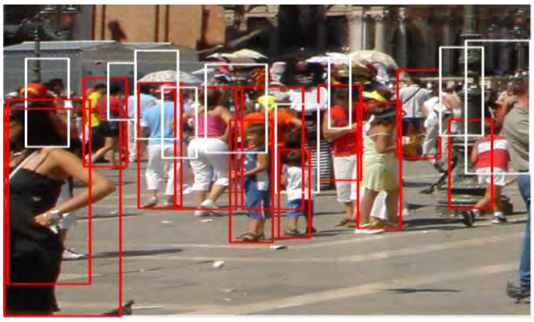

Fig. 6. HOG+SVM Algorithm Test Result II

The relatively crowded scene was showed in fig.5, the test results of the classic algorithm of the same image was showed in fig.6. In this paper, most pedestrians can be detected with less error detection in the algorithm. However, in densely populated areas, especially overlapping areas, the same pedestrian and the overlapped pedestrian has been detected for many times. In addition, for the partially obstructed pedestrians, the algorithm in this paper also has leakage detection. In the case that the training set was relatively complete, the algorithm in this paper has error detection, but the error detection rate was relatively ideal.

B. Comparison test

In the case of using the same data set, we compared the algorithm in this paper with the classical HOG+SVM method. The purpose of this experiment was to compare the training samples required by the two algorithms in the case of equivalent performance.

TABLE II. EXPERIMENTAL COMPARISON RESULTS

	Dete ction rate	Error detection rate	Training positive sample	Train ing sample
HOG+SVM	70.2%	7.8%	1000	6000
HOG+SVDD	72.5%	5.8%	2000	2000
HOG+SVM	83.2%	4.8%	2000	22000
HOG+SVDD	82.7%	3.1%	3500	3500
HOG+SVM	89.2%	2.8%	2400	23000
HOG+SVDD	88.5%	1.1%	5000	5000

The comparison of experimental results were showed in table Ⅱ. According to the table, the algorithm in this paper performed well in the control of error detection, and the error detection rate was lower when the detection rate was equal to the classical algorithm. Furthermore, in the case of comparable performance, more positive samples were needed for the algorithm in this paper, but less training samples are needed overall.

VI. CONCLUSION

Pedestrian detection is mainly used to determine whether the input image contains pedestrians and give corresponding location information, which is an important step in the completion of intelligent identification. Due to the influence of wearing, scale, occlusion, posture and visual angle on the appearance of pedestrians, pedestrian detection has become a research hotspot and difficulty in computer vision.

This paper proposed a pedestrian detection algorithm based on HOG and support vector domain data description, which can better solve the problem of poor generalization ability caused by incomplete training negative samples in previous pedestrian detection algorithms based on statistical learning. Firstly, multi-scale candidate region samples were generated, and then histogram statistics of gradient direction were conducted. Finally, a single classifier was trained by using complete positive samples to complete pedestrian detection. Compared with the classic HOG+SVM algorithm, this algorithm has a low error detection rate and a small number of required samples, which was conducive to the rapid and accurate detection of pedestrian target.

In this algorithm, the dimensions of image HOG feature were too large, and there were a large number of redundant features, which made subsequent SVDD training more difficult and reduced the efficiency of the algorithm. Therefore, in future work, it is necessary to find and design more efficient and concise feature extraction methods, further reduce the training difficulty of the algorithm and improve its application efficiency.

PROJECT

National natural science foundation (NO.61171194)

REFERENCES

[1] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: An evaluation of the state of the art[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012, 34(4): 743-761.

[2] SU Song-zhi, LI Shao-zi ,CHEN Shu-yuan ,CAI Guo-rong, WU Yun-dong. A Survey on Pedestrian Detection[J]. Acta Electronica Sinica , 2012, 40(4): 814-820. (in chinese)

[3] JIA Hui-Xing ZHANG Yu-Jin. A Survey of Computer Vision Based Pedestrian Detection for Driver Assistance Systems [J], Acta Automatica Sinica, 2007, 33(1): 84-90. (in chinese)

[4] M. Enzweiler, and D.Gavrila. Monocular pedestrian detection: survey and experiments [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2179-2195.

[5] Papageorgiou C, Poggio T. A trainable system for object detection[J]. International Journal of Computer Vision, 2000, 38(1): 15-33.

[6] Struzik Z R, Siebes A. The Haar wavelet transform in the time series similarity paradigm[M]//Principles of Data Mining and Knowledge Discovery. Springer Berlin Heidelberg, 1999: 12-22.

[7] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3): 293-300.

[8] Viola P, Jones M. Robust real-time object detection[J]. International Journal of Computer Vision, 2001, 4: 51-52.

[9] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[10] Tax D M J, Duin R P W. Support vector data description[J]. Machine learning, 2004, 54(1): 45-66.

[11] Qi L, Jiang H. Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations[J]. Mathematics of Operations Research, 1997, 22(2): 301