

# Pedestrian detection based on the improved HOG features

JianliangMeng, Shujin Li<sup>a</sup>

School of North China Electric Power University, Baoding 071000, China.

<sup>a</sup>kathycrystal@126.com

**Keywords:** Pedestrian detection, MultiHOG, LBP, additive kernel SVM.

**Abstract.** For HOG features' characteristics of high accuracy and large amount of calculation, selected MultiHOG features instead of traditional HOG by means of adjusting the structure of HOG features and using Fisher selection criteria. For the further detection effect, coalesced LBP feature which good at texture based on MultiHOG. The algorithm combining additive cross the SVM classifier to reduce the test time, improved the efficiency of detection and detected pedestrians sliding window. Finally, tested by INRIA standard data sets. The results showed that the algorithm has better feature detection and detection time than traditional one.

## 1. Introduction

Pedestrian detection can be defined as pedestrian location will be given if the image or video was determined to contain the pedestrian[1]. Literature [2] presented the combination of SVM method and HOG features shows good performance in pedestrian detection, but there are also large amount of calculation and the method computing time. Therefore, this paper makes an improvement on the traditional HOG feature method in three aspects. Firstly, use multi-scale block to replace the original block, and carries it on the choice, reduce the feature dimension from the traditional HOG 3780 d to 360 d; Secondly, add the LBP texture which has good recognition effect characteristics to improve the detection effect; Thirdly, the detection accuracy can be significantly improved and detection time basic remain unchanged by using additive cross nuclear instead of linear SVM.

## 2. MultiHOG feature

It just can measure the local detail features of human body due to the original HOG[2] influenced by the original block designing. This paper based on the improved HOG features, used three linear interpolation methods to vote each cell in the gradient direction, and used a variety of scale block to adjust the structure of HOG, finally, choose the 14 block of multi-scale HOG of 21 generated block feature selection.

HOG different size of block and block overlap structure on pedestrian detection show that tissue block structure rationally had a good effect to the accuracy of pedestrian detection and real-time performance. Combining with the linear interpolation and the block size multiscale design which can fully said details and the overall characteristics of the pedestrian contours.

This paper proposed a MultiHOG (Multi Scale Block HOG) characteristic of the composition is as follows:

(1) Diversification design of Block size, mainly aims at the human body's macro gradient features and local detail features together. The use of sample size is  $64 \times 128$  of the human body to extract features, and size of blocks is  $64 \times 128$ ,  $32 \times 64$  and  $16 \times 8$ , while the long and wide of cells size is half of block size. There is no overlap between block and block. There will be 21 blocks, and HOG feature dimension will be reduced from the original 3780 d to 756 d.

(2) Three linear interpolation is used to realize the statistics of the gradient direction in block. The so-called three linear interpolation is interpolation in the three parameter space,  $(x, y, \theta)$ , namely the X direction, Y direction and gradient angle of space. When the pixel  $(x, y, \theta)$  vote with gradient magnitude as the weight, the distance from this pixel to other block center are weighted, and gradient of the pixels is interpolated in the adjacent area.

The three linear interpolation uses the long and wide ratio of the distance between the pixels and each cell center point, and the angle value range of the pixels in the projection range, then projected to the cell center of each block. Finally, the block is obtained.

By using three linear interpolation and multi scale, the advantages of high accuracy of block and block overlap in the original HOG are achieved. The multi scale selection avoids the disadvantages of large block and small block in the original HOG, which makes the multi dimension into an advantage.

MultiHOG constructs 21 blocks, which is 756 dimension vector. Although some of them are very good recognition effect, some of them affect the accuracy of pedestrian contour features. In view of this, this paper proposed feature blocks with distinguishing ability which is selected from the 21 blocks using Fisher selection criteria.

Fisher criterion[3] is used to measure distinguishing ability of feature blocks according to the ratio of the inter class scatter matrix and the inside class scatter matrix, which makes the discrete degree between different sample as large as possible, and the discrete degree between same sample as small as possible. A feature block with bigger ratio has better distinguishing ability. Compared to other feature selection methods, Fisher feature selection is the whole sample projection classification, and each projection is a classification.

The feature vector  $v(j)$  is arranged according to the value of  $F(j)$ , and then the classifier is trained with the sequence of each feature block. Without the requirement of classification accuracy, a feature vector is combined with the classification and prediction, until the classification results are achieved.

Only the first 10 feature blocks can achieve better classification results. So far, we have selected the block with good classification results, and obtained the final MultiHOG feature vector, which has 360 dimensions. The feature blocks picked out by Fisher is with some semantic features, such as head and shoulder part of the body, legs, hands, etc.

### 3. LBP extraction in pedestrian detection

Like cells and blocks in the HOG feature extraction [4], we have set up a LBP-Cell LBP feature extraction for human images. The size of each LBP-Cell is 16x16 pixels, and they are not overlapping. Each 64x128 pixels size detection window can be divided into 32 LBP-Cell. In each LBP-Cell, we do histogram statistics based on the number of pixels as the weight into 59  $LBP_8^1$  uniform pattern mode characteristics channel, and normalize with L1-sqrt norm. Finally, LBP feature histograms of different LBP-Cells are combined to form a dimension of 1888d, which can describe the texture features of the current detection window.

For each training window, the extracted LBP feature and MultiHOG feature are fused, then inputed into the SVM classifier for training. The MultiHOG and LBP features of the pedestrian detection classifier are obtained.

### 4. Additive kernel SVM classifier

Efficient classifier is the key to the pedestrian detection system. Linear SVM is widely used in real-time applications. However, the nonlinear kernel SVM has better classification performance. Maji et al. [5] proposed the method of adding cross kernel SVM, which is close to the nonlinear SVM, and the time complexity is similar to linear SVM.

Additive kernel SVM (IKSVM) speeds the classification process by means of a number of special additive kernel. According feature of IKSVM, the complexity of the algorithm is  $O(n \log m)$ , therefore the total classification algorithm complexity is  $O(n \log m)$  and comparing with  $O(nm)$  it is obviously improved. If the piecewise polynomial function is used to approximate every 1 dimensional function, with the the polynomial parameter preexisted in lookup table, classification complexity can be further reduced to  $O(n)$ , which has been approximated by linear SVM.

## 5. Experiment and analysis

In order to test the performance of the proposed features, the INRIA database [6] is used as positive samples and negative samples. Sample composition is random, and the size of the training sample size is  $64 \times 128$ . The experimental environment is Inter (R) core (TM) i7, 2.50GHz, memory 4G, 64 bit Windows7 system. The parameters such as the ROC curve combined by feature extraction time, detection rate, accuracy rate and false alarm rate were used to evaluate the performance.

HOG feature has 3780 dimensions, and detective time of a  $64 \times 128$  image is 0.028s. The feature proposed in this paper are 2248 dimensional (MultiHOG feature 360 dimension and LBP feature 1888 dimension), while feature extraction time is as long as 0.021s, which is 25% faster than before, as shown in table 1:

TABLE 1 Feature compare

Features	Dimensional	Computation time
HOG	3780	0.028s
The proposed feature	2248	0.021s

Detection rate, accuracy rate, false alarm rate and detective time of tow method are shown in table2:

TABLE 2The compare of method in this paper and HOG+SVM

Methods	Detection rate	Accuracy rate	False alarm rate	Detective time
HOG+SVM	95.03%	93.44%	5.83%	17.74s
This paper	98.45%	96.79%	3.79%	13.08s

As can be seen from table II, the detection rate and accuracy of this method are 4% higher than the HOG+SVM method. For a  $640 \times 480$  image, using a trained model, this paper saves 1/4 of the time compared with HOG+SVM. Fig.1 shows the ROC curves of two methods in INRIA data set:

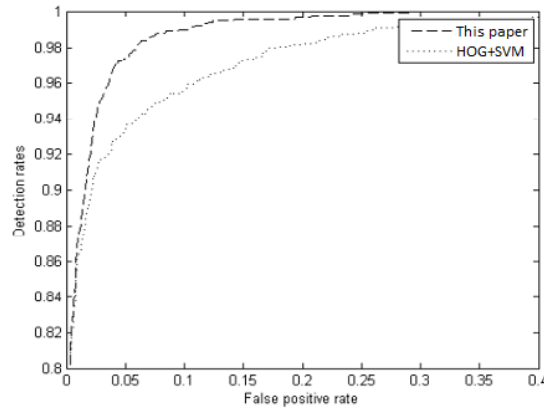


Fig.1 ROC curves of two methods



Fig.2 Windows combination

After detection, there will appear some overlapping windows in the image, so the combination of the windows is necessary. Windows which have overlap area over 60% are combined into one class, then the average size of the window and the average position is taken as the final location of the pedestrian. As shown in fig. 2.

Some test results are shown in fig. 3. On the left are the HOG+SVM method detection results, while the right is the result of this method. It can be seen that the detection effect of this method is more excellent than the HOG+SVM method. Fig. 4 shows the missed and false detection.



Fig.3 Detective result of HOG+SVM and method in this paper Fig.4 Missed detective result and false detective result

## 6. Conclusion

In this paper, the HOG feature was improved into the MultiHOG feature. Because of the limitations of HOG, LBP feature was added to improve the detection effect of the texture features. In order to balance the time consumption caused by the addition of LBP features, additive kernel SVM which has the similar time complexity to linear SVM classifier was used to approximate the effect of the nonlinear SVM classifier. This method compared with the traditional hog method has better detection effect, also have certain advantages in the detection time, at the same time it can reduce the missed detection rate and false alarm rate. But this method is not perfect, and it is difficult to identify the pedestrian in the distance or pedestrian covered, which is mainly because of the detection window. After the research, we can continue to carry out the two aspects of the detection of complex environment and improve the detection time.

## References

- [1].Su Songzhi, Li Shaozi, Chen Shuyuan, Cai Guorong, Wu Yundong. A Survey on Pedestrian Detection[J]. Acta Electronica Sinica,2012,04:814-820.
- [2].Dalal N and Triggs B. Histograms of oriented gradients for human detection[C]. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition,San Diego, USA, 2005: 886-893.
- [3].Hu Xuejuan, Ruan Shuangchen, Guo Chunyu, Liu Chengxiang. Improved histograms of oriented gradients for Chinese RMB currency recognition[J]. Journal of Shenzhen University Science and Engineering,2014,05:487-492.
- [4].Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern recognition, 1996, 29(1): 51-59.
- [5].Maji S, Berg A C, Malik J. Efficient classification for additive kernel SVMs[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(1): 66-77.
- [6].INRIA Database[OL].<http://pascal.inrialpes.fr/data/human/>