FAST OBJECT DETECTION USING BOOSTED CO-OCCURRENCE HISTOGRAMS OF ORIENTED GRADIENTS

Haoyu Ren[1,2] *Cher-Keng Heng*[3] *Wei Zheng*[1,2] *Luhong Liang*[1,2] *Xilin Chen*[1,2]
[1]Key Lab of Intelligent Information Processing, Chinese Academy of Sciences (CAS), Beijing, 100190, China
[2]Institute of Computing Technology, CAS, Beijing, 100190, China
[3]Panasonic Singapore Laboratories Pte Ltd
{hyren, wzheng, lhliang, xlchen}@jdl.ac.cn CherKeng.Heng@sg.panasonic.com

ABSTRACT

Co-occurrence histograms of oriented gradients (CoHOG) are powerful descriptors in object detection. In this paper, we propose to utilize a very large pool of CoHOG features with variable-location and variable-size blocks to capture salient characteristics of the object structure. We consider a CoHOG feature as a block with a special pattern described by the *offset*. A boosting algorithm is further introduced to select the appropriate locations and offsets to construct an efficient and accurate cascade classifier. Experimental results on public datasets show that our approach simultaneously achieves high accuracy and fast speed on both pedestrian detection and car detection tasks.

Index Terms—Object Detection, CoHOG, Boosting, Cascade Classifier

1. INTRODUCTION

In object detection, the high accuracy and fast speed are the general targets. On one hand, to achieve better accuracy, some researchers propose discriminative descriptors, e.g., HOG feature [1], covariance matrix [10] or their combinations, e.g., heterogeneous feature [14], HOG-LBP [16]. Others focus on designing more effective classifiers, e.g., vector boosted tree [5], multiple kernels SVM [13] and multiple instance learning [11], etc. On the other hand, the speed issue is also important for many real-time applications. The cascade framework proposed by Viola and Jones [12] performs well on face detection. The recent works have proved its efficiency and effectiveness on other object categories such as pedestrians [10][17] and cars [14]. In most of the cases, the performance and speed of a boosted classifier mainly depend on the features. To address the accuracy and speed issue simultaneously, boosting with appropriate features to construct the cascade classifier is the key step.

Recently, T. Watanabe et al. presented a pedestrian detection algorithm with promising detection results [15]. Their method uses dense grid of the CoHOG features with 31 offsets (each offset corresponds to a pixel-pair that describes a special local pattern), computed 72 fixed blocks of a 64 × 128 size patch. This representation is proved to be powerful enough on pedestrian detection under linear SVM framework. Unfortunately, the detection speed is relative slow due to the high-dimensional (138,816 dimensions) feature vectors.

To improve the efficiency of the CoHOG feature, M. Hiromoto and R. Miyamoto [6] divided the trained SVM classifier into fixed blocks and rearrange them to a cascade classifier, which achieved about 2.5 times speed up. This approach considers the accuracy and the speed in different steps (dense extraction for accuracy, block division for speed). The CoHOG features in this approach are regularized to fixed-location and fixed-size rectangular blocks, which might mismatch the typical structures of the objects. We still need to calculate all the 31 offsets in each block. So there are extra computation costs due to the redundant offsets calculation.

In this paper, we achieve the target of the high accuracy and fast speed simultaneously. We consider a single CoHOG feature as a combination of a variable-size and variable-location block with a fixed offset, which is capable of describing the structural characteristics of the objects. The boosting algorithm is further used to construct a cascade classifier from a large feature pool by selecting both the locations and the offsets. The experimental results on several public datasets show that our approach achieves comparable performance with the state-of-the-art methods on both pedestrian detection and car detection, while the detection speed is about 30 times faster than the SVM approach [15].

2. COHOG FEATURES

2.1 COHOG FEATURES

The CoHOG features are powerful region descriptors [15] that reflect the co-occurrence frequency of the oriented gradients between the pixel-pairs. A pixel-pair can be represented by an offset (Δx,Δy), which reflects the spatial relationship of the two points. As shown in Fig.1a, we define 31 offsets for a given point specifically. For each pixel in the image, we quantize the gradient orientation into 8 directions. Therefore, there are totally 8×8=64 elements in the co-occurrence matrix as in Fig.1b.

Different from the previous approaches [6][15] that extracting the feature vector from a fixed region using all

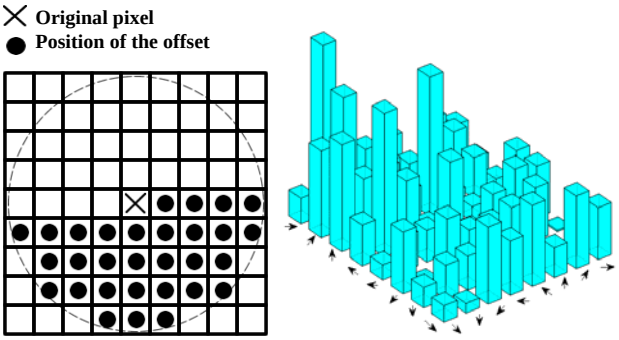


Fig.1. CoHOG features

the 31 offsets, our CoHOG feature is an $m \times n$ pixels block $\boldsymbol{B}$ with a fixed offset (p, q). Each element in the co-occurrence matrix is formulated as

$$C^{B}_{i,j,p,q} = \sum_{(x,y)\in B} I_{i,j}(x, y, p, q) \quad i, j \in \{0,...,7\}, \tag{1}$$

where (x,y) is a pixel in B and $I_{i,j}(x, y, p, q)$ is an indicator function of the gradient orientation $O(x,y)$

$$I_{i,j}(x, y, p, q) = \begin{cases} 1 \; iff \; O(x, y) = i, O(x + p, y + q) = j \\ 0 \; otherwise \end{cases}. \tag{2}$$

The co-occurrence matrix is directly used as a 64-D feature vector.

We build the CoHOG feature pool that consists of variable-location and variable-size blocks. To generate the feature pool, we sample the locations of the block centers every 4 pixels. In our case, we use 64×128 detection window with 7,997 blocks from 8×8 to 48×96 in pedestrian detection, and 80×40 detection window with 1,751 blocks from 8×8 to 64×32 in car detection.

2.2 BOOSTING WITH COHOG FEATURES

In this section, we introduce how to build the cascade classifier using CoHOG feature. We follow the cascade structure in [12] and use the RealBoost [2] as the example. In RealBoost, the weak classifier is a function from the feature vector $\mathbf{x}$ to a real valued object/non-object classification confidence $f(\mathbf{x})$. For the binary classification problem, suppose the training data as $(\mathbf{x}_1, y_1),\ldots,(\mathbf{x}_N, y_N)$, where $\mathbf{x}_i$ is the training sample and $y_i \in (-1,1)$ is the class label, we first divide the sample space into N_b equal sized sub-ranges R_j

$$X_j = \{\mathbf{x} \mid f(\mathbf{x}) \in R_j\}, \; j = 1,..., N_b. \tag{3}$$

The weak classifier is defined as a piecewise function

$$h(\mathbf{x}) = \frac{1}{2}\ln\left(\frac{W^j_+ + \varepsilon}{W^j_- + \varepsilon}\right), \; j = 1,..., N_b, \tag{4}$$

where ε is the smoothing factor, and $W_{\pm}$ is the probability distribution of the feature value for positive/negative samples

Parameters:
- N number of training samples
- N_f number of randomly evaluated features in each iteration
- T maximum number of weak classifiers

Input: Training set

$\{(x_i, y_i)\}\; i = 1,..., N, \mathbf{x}_i \in \mathrm{R}^d, y_i \in \{-1,1\}\}$

1. Initialization
   $\varpi_i = 1/N, H(\mathbf{x}_i) = 0, i = 1,..., N$
2. Repeat for t = 1,…,T
   2.1 Update the sample weight ϖ_i using the t^{th} weak classifier output $h_t(\mathbf{x}_i)$ (4) as $\varpi_i = \varpi_i e^{-y_i h_t(\mathbf{x}_i)}$
   2.2 For m = 1 to N_f
   For n = 0 to 30
   2.2.1 Calculate the feature vectors $\mathbf{x}_i$ according to the mth block and nth offset
   2.2.2 Train a classify plane $\mathbf{w}^*$ in the feature space and calculate the feature response $f(\mathbf{x}_i)$ by (7)
   2.2.3 Choose the best block and offset according to the classification error (7)
   2.3 Calculate the weak classifier response $h(\mathbf{x}_i)$, update $H(\mathbf{x}_i)$ as $H_{t+1}(\mathbf{x}_i) = H_t(\mathbf{x}_i) + h_t(\mathbf{x}_i)$
3. Output classifier $H(\mathbf{x}) = sign[\sum_{j=1}^{T} h_j(\mathbf{x})]$

Fig. 2 RealBoost training with CoHOG features

$$W^j_{\pm} = P(\mathbf{x} \in X_j, y = \pm 1), \; j = 1,..., N_b. \tag{5}$$

The best weak classifier is selected according to the classification error Z of the piecewise function (6).

$$Z = 2\sqrt{\sum_{j=1}^{N_b} W^j_+ W^j_-}, \tag{6}$$

We use the proposed CoHOG blocks as the feature pool of RealBoost training. In order to learn the best weak classifiers, the most intuitive way is to look through the whole feature pool, which is rather time consuming. We resort to a sampling method to speed up the feature selection process similar as [17]. More specifically, a random sub-sample of size $\log 0.05 / \log 0.95 \approx 59$ will guarantee that we can find the best 5% features with a probability of 95%. In practice, 60 blocks are evaluated each iteration.

For each candidate block, the 64-D feature vectors are extracted according to the best offset selected from 0 to 30 and further train a linear classify plane $\mathbf{w}^*$ using least square. Then the feature response $f(\mathbf{x})$ is calculated by

$$f(\mathbf{x}) = \mathbf{w}^* \bullet \mathbf{x} + b, \tag{7}$$

where b is the bias. Fig. 2 illustrates more details.

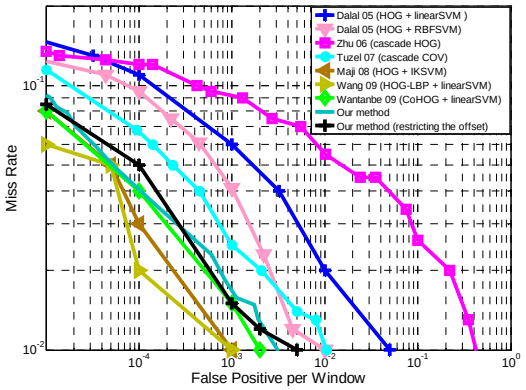


Fig. 3 Per-image performance evaluation on INRIA dataset

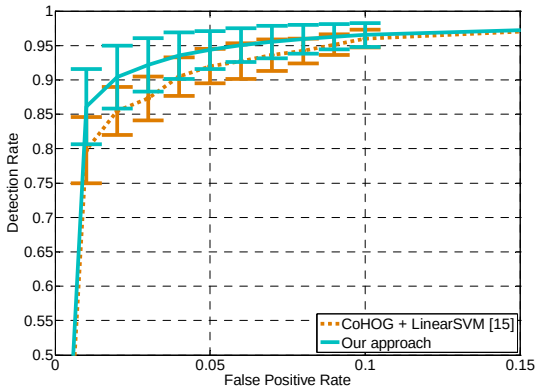


Fig. 4 Performance evaluation on DC dataset

3. EXPERIMENTS

3.1. INRIA datasets

We evaluate the proposed method on the commonly used INRIA pedestrian dataset following the training and testing protocols at patch level proposed by Dalal and Triggs [1]. In Fig.3, we plot the miss rate tradeoff False Positive rate Per Window (FPPW) curves on a log-log scale by tuning the rejection threshold of the classifiers. We compare our method with the state-of-the-art classifiers. Our approach (blue curve) shows comparable result with most of the algorithms, which achieves similar accuracy with the CoHOG method using linear SVM classifier [15]. The performance is potentially to be improved using more complicate features (e.g., composite features [17]) or classifiers (e.g., using kernel methods instead of linear training [8]). Our cascade classifier evaluates 76 blocks in average while processing a patch. The average feature dimension is 76×64=4,864, which is much efficient compared with the 138,816 dimensions CoHOG feature in SVM approach. We test these two algorithms on a computer with 3.0GHz CPU and 4G memory. Our cascade classifier achieves about 30 times acceleration compared with the SVM approach [15]. This result seems to be more efficient than the cascade classifier [6], which achieves about 2.5 times acceleration than [15].

3.2. Daimler Chrysler pedestrian benchmark datasets

We use the Daimler Chrysler (DC) pedestrian benchmark dataset created by Munder and Gavrila [9]. The dataset is split into five disjoint sets, three for training and two for testing. It also provides negative image set containing 3,600 images without pedestrians. We bootstrap about 50,000 negative samples from the negative image set for both SVM and boosting methods. We report the ROC curves by training on two out of three training sets at a time and testing on each of the test sets to obtain six curves. The average curve with the standard error is used for evaluation as shown in Fig.4. Our cascade CoHOG classifier achieves better result than the SVM approach, which has about 5% improvement on the detection rate at low false positive rate. The improvement might come from our larger feature pool, which captures more informative structures in objects.

3.3. VOC2006 datasets

In this section, we evaluate the proposed method on the multi-view car detection task. The evaluation follows the training and detection protocol of PASCAL VOC2006 challenge [2]. We collect about 500 out of 854 cars from the *train & val* set for training the cascade classifier. Detection results are considered as true positive when the area of overlap a_0 (8) between the predicted bounding box B_p and ground truth bounding box B_{gt} exceed 50%

$$a_0 = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}. \qquad (8)$$

We plot the precision-recall curve in Fig.5. The reported result of PASCAL VOC2006 [2] is also drawn in the figure. We can see that the proposed approach achieves 0.515 AP (Average Precision rate), which outperforms all the results in [3]. In the later, [4] reported a much better AP at 0.635 via considering the occlusion and deformation in the training process. Our approach is also expected to be improved using the similar way.

3.4. Feature Analysis

In this section, we give some analysis on the selected features and offsets. Fig.6 shows the first 3 selected features of the pedestrians and cars. These selected features could well capture the typical structures of the objects. We also show the frequency of the selected offsets in our cascade classifier in Fig.7. We can see that the key offset patterns of different object categories are different. For pedestrian detection, the selected offsets mainly lay on the edge of the offset circle, which corresponds to the key structures in pedestrian described by large offset, e.g., arms and legs. The selected offsets of the cars focus on the center of the circle, which emphasizes more on local pattern, e.g., wheels and

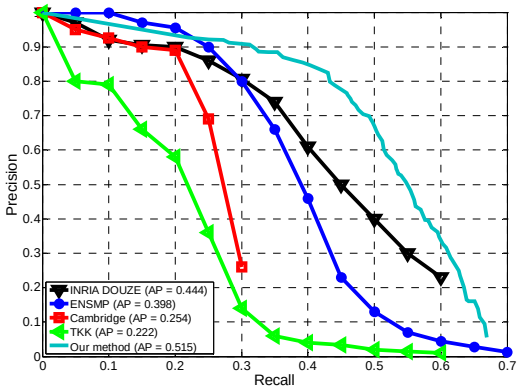
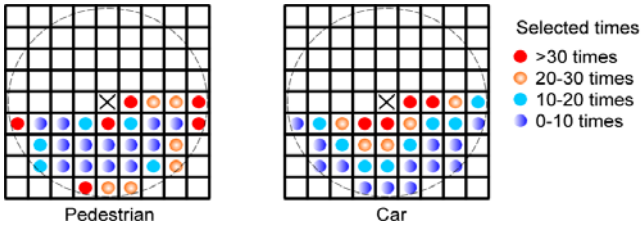


Fig.5 Performance evaluation on VOC2006 car dataset

Fig. 6 The selected features of the cascade classifier



(a) **(b)**

Fig. 7 The selected offsets of the cascade classifier

lamps. The experiments have proved that the boosted classifier trained by restricting the feature pool to these key offsets could get similar performance with the original one. As shown in Fig. 1, the detection rate (black curve) is similar to the boosted classifier using full feature pool (blue curve).

4. CONCLUSION

In this paper, we proposed an object detector that achieves both high accuracy and fast speed. We built a large pool of CoHOG features with variable-location and variable-size blocks, which could capture the characteristics in object structure. We further utilized a boosting algorithm to select the more informative CoHOG features by evaluating different blocks and offsets. Experimental results show that our approach achieves high accuracy on both pedestrian and car detection tasks. The detection speed is about 30 times acceleration than the SVM approach.

5. ACKNOWLEDGEMENT

This paper is partially supported by NSFC under contracts U0835005, 60832004, 60872124; National Basic Research Program of China (973 Program) under contract 2009CB320902; and Grand Program of International S&T Cooperation of Zhejiang Province S&T Department under contract No. 2008C14063. Especially, we would like to thank Xiaopeng Hong for the helpful discussion and suggestion.

6. REFERENCES

[1] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR*, 2005.

[2] R. Schapire and Y. Singer. Improved Boosting Algorithms Using Confidence-rated Predictions. In *Machine Learning*, Vol. 37, Page(s):297-336, 1999.

[3] M. Everingham and A. Zisserman. The Pascal Visual Object Classes Challenge 2006 (VOC2006) Results.

[4] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. To appear in *PAMI*.

[5] C. Huang, H. Ai, Y. Li, and S. Lao. Vector Boosting for Rotation Invariant Multi-View Face Detection. In *ICCV*, 2005.

[6] M. Hiromoto and R. Miyamoto. Cascade Classifier Using Divided CoHOG Features for Rapid Pedestrian Detection. In *ICVS*, 2009.

[7] I. Laptev. Improvements of Object Detection using Boosted Histograms. In *BMVC*, 2006.

[8] S. Maji, A. Berg, and J. Malik. Classification Using Intersection Kernel Support Vector Machines is Efficient. In *CVPR*, 2008.

[9] S. Munder and D. Gavrila. An Experimental Study on Pedestrian Classification. In *PAMI*, Vol. 28, Page(s):1863-1868, 2006.

[10] O. Tuzel, F. Porikli, and P. Meer. Pedestrian Detection via Classification on Riemannian Manifolds. In *PAMI*, Vol. 30, Page(s):1713-1727, 2008.

[11] P. Viola, J. Platt, and C. Zhang. Multiple Instance Boosting for Object Detection. In *NIPS*, 2005.

[12] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In *CVPR*, 2001.

[13] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple Kernels for Object Detection. In *ICCV*, 2009.

[14] B. Wu and R. Nevatia. Optimizing Discrimination-Efficiency Tradeoff in Integrating Heterogeneous Local Features for Object Detection. In *CVPR*, 2008.

[15] T. Watanabe, S. Ito, and K. Yokoi. Co-occurrence histograms of oriented gradients for pedestrian detection. In *PSIVT*, 2009.

[16] X. Wang, T. Han, and S. Yan. An HOG-LBP Human Detector with Partial Occlusion Handling. In *ICCV*, 2009.

[17] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng. Fast Human Detection using a Cascade of Histograms of Oriented Gradients. In *CVPR*, 2006.