# Multiple Instance Boost Using Graph Embedding Based Decision Stump for Pedestrian Detection

Junbiao Pang[1,2,3], Qingming Huang[1,2,3], and Shuqiang Jiang[2,3]

[1] Graduate university of Chinese Academy of Sciences, Beijing, 100190, China
[2] Key Lab. of Intelligent Information Processing, Chinese Academy of Sciences (CAS)
[3] Institute of Computing Technology, CAS, Beijing, 100190, China
{jbpang,qmhuang,sqjiang}@jdl.ac.cn

**Abstract.** Pedestrian detection in still image should handle the large appearance and stance variations arising from the articulated structure, various clothing of human as well as viewpoints. In this paper, we address this problem from a view which utilizes multiple instances to represent the variations in multiple instance learning (MIL) framework. Specifically, logistic multiple instance boost (LMIBoost) is advocated to learn the pedestrian appearance model. To efficiently use the histogram feature, we propose the graph embedding based decision stump for the data with non-Gaussian distribution. First the topology structure of the examples are carefully designed to keep between-class far and within-class close. Second, K-means algorithm is adopted to fast locate the multiple decision planes for the weak classifier. Experiments show the improved accuracy of the proposed approach in comparison with existing pedestrian detection methods, on two public test sets: INRIA and VOC2006's person detection subtask [1].

## 1 Introduction

Pedestrian detection is a practical requirement of many today's automated surveillance, vehicle driver assistance systems and robot vision systems. However, the issue of large appearance and stance variations accompanied with different viewpoints makes pedestrian detection very difficult. The reasons can be multifold, such as variable human clothing, articulated human structure and illumination change, etc. The variations bring various challenges including misalignment problem, which is often encountered in non-rigid object detection.

There exist a variety of pedestrian detection algorithms from the different perspectives, directly template matching [2], unsupervised model [3], *traditional* supervised model [4,5,6] and so on. Generally, these approaches cope with "mushroom" shape – the torso is wider than the legs, which dominates the frontal pedestrian, and deal with "scissor" shape – the legs are switching in walk, which dominates the lateral pedestrian. However, for some uncommon stances, such as mounting on bike, they incline to fail. In these conditions, the variations often impair the performance of these conventional approaches. Fig. 1 shows some false negatives generated by Dalal et al [4]. These false negatives are typically non-"mushroom" or non-"scissor" shape, and have large variations between each other.

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 541–552, 2008.

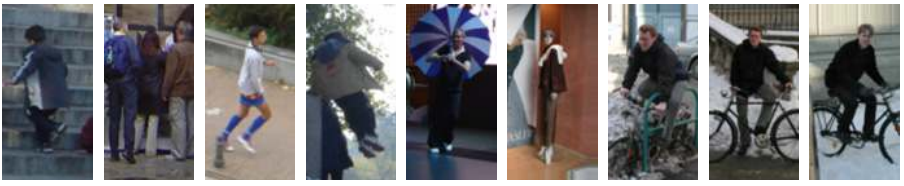

**Fig. 1.** Some detection results in our method producing fewer false negatives than Dalal et al do [4]

The key notion of our solution is that the variations are represented within multiple instances, and the "well" aligned instances are automatically selected to train a classifier via multiple instance learning (MIL) [7,8]. In MIL, a training example is not singletons, but is represented as a "bag" where all of the instances in a bag share the bag's label. A positive bag means that at least one instance in the bag is positive, while a negative bag means that all instances in the bag are negative. To pedestrian detection, the standard scanning window is considered as the "bag", a set of sub-images in window are treated as instances. If one instance is classified as pedestrian, the pedestrian is located in detection stage. The logistic multiple instance boost (LMIBoost) [9] is utilized to learn the pedestrian appearance, which assumes the average relationship between bag's label and instance's label.

Considering the non-Gaussian distribution (which dominates the positive and negative examples) and aims of detection (which are accurate and fast), a graph embedding based weak classifier is proposed for histogram feature in boosting. The graph embedding can effectively model the non-Gaussian distribution, and maximally separate the pedestrians from negative examples in low dimension space [10]. After feature is projected onto discriminative one dimension manifold, K-means is utilized to fast locate the multiple decision planes for the decision stump. The proposed weak classifier has the following advantages: 1) it handles training examples with any distribution; and 2) it not only needs less computation cost, but also results in robust boosting classifier. The main contributions of the proposed algorithm are summarized as following:

- The pose variations are handled by multiple instance learning. The variations between examples are represented within the instances, and are automatically reduced during learning stage.
- Considering the boost setting, graph embedding based decision stump is proposed to handle training data with non-Gaussian distribution.

In the next section, related work is briefly summarized. Section 3 introduces the LMIBoost for solving the variations. Section 4 first introduces the graph embedding based discriminative analysis, and then presents the multi-channel decision stump. In section 5, we describe the experimental settings for pedestrian detection. Finally the experiment and conclusion sections are provided, respectively.

## 2 Related Work

Generally, the "mushroom" or "scissor" shape encourages the use of template matching and traditional machine learning approach as discussed in section 1. The contour templates are hierarchically matched via Chamfer matching [2]. A polynomial support vector machine (SVM) is learned with Haar wavelets as human descriptor [5] (and variants are described in [11]). Similar to still images, a real-time boosted cascade detector also uses Haar wavelets descriptor but extracted from space-time differences in video [6]. In [4], an excellent pedestrian detector is described by training a linear SVM classifier using densely sampled histogram of oriented gradients (HOG) feature (this is a variant of Lowe's SIFT descriptor [12]). In a similar approach [13], the near real-time detection performance is achieved by training a cascade detector using SVM and HOG feature in AdaBoost. However, their "fixed-template-style" detectors are sensitive to pose variations. If the pose or appearance of the pedestrian has large change, the "template"-like methods are doomed to fail. Therefore, more robust feature is proposed to withstand translation and scale transformation [14].

Several existing publications have been aware of the pose variation problem, and have handled it by "divide and conquer"– the parts based approach. In [15], the body parts are explicitly represented by co-occurrences of local orientation features. The separate detector is trained for each part using AdaBoost. Pedestrian location is determined by maximizing the joint likelihood of the part occurrences according to the geometric relations. Codebook approach avoids explicitly modeling the body segments or the body parts, and instead uses unsupervised methods to find part decompositions [16]. Recently, the body configuration estimation is exploited to improve pedestrian detection via structure learning [17]. However, parts based approaches have two drawbacks. First, different part detector has to be applied to the same image patch. This reduces the detection speed. Second, labeling and aligning the local parts are tedious and time-costing work in supervised learning. Therefore, the deformable part model supervised learns the holistic classifier to coarsely locate the person, and then utilizes part filters to refine body parts in unsupervised method [18].

The multiple instance learning(MIL) problem is first identified in [8], which represents ambiguously labeled examples using axis-paralled hyperrectangles. Previous applications of MIL in vision have focused on image retrieval [19]. The seemingly most similar work to ours may be the upper-body detection [20]. Viola et al use Noisy-OR boost which assumes that only sparse instances are upper-body in a positive bag. However, in our pedestrian detection setting, the instances in a positive bag are all positive, and this facilitates to simply assume that every instance in a bag contributes equally to the bag's class label.

In pedestrian detection, the histogram feature (such as SIFT, HOG) is typically used. The histogram feature can be computed rapidly using an intermediate data representation called "Integral Histogram" [21]. However, the efficient use of the histogram feature is not well discussed. In [13], the linear SVM and HOG feature is used as weak classifier. Kullback-Leibler (K-L) Boost uses the log-ratio between the positive and negative projected histograms as weak classifier. The

projection function is optimized by maximizing the K-L divergence between the positive and negative features [22]. SVM has high computational cost and hence reduces the detection speed. Optimizing the projection function in K-L Boost is also computationally costly and numerically unstable. Fisher linear discriminative analysis (FLDA) is used as weak classifier for histogram feature [23]. Despite the success of FLDA for building weak classifier, it still has the following limitations: it is optimal only in the case that the data for each class are approximate Gaussian distribution with equal covariance matrix.

Although the histogram feature is projected into one dimension manifold using the projection functions, the learned manifold does not directly supply classification ability. The widely used decision stump is a kind of threshold-type weak classifier, but a lot of discriminative information is lost [24]. Therefore, the single-node, multi-channel split decision tree is introduce to exploit the discriminative ability. In face detection [25], Huang et al use the histogram to approximate the distributions of the real value feature by dividing the feature into many sub-regions with equal width in RealBoost. Then a weak classifier based on a look up table (LUT) function is built by computing the log-ratio on each sub-bins. However, the equal regions unnecessarily waste decision stump in low discriminative region. In [26], the unequal regions are obtained by exhaustively merging or splitting the large number of histogram bins via Bayes decision rule. In this paper, we avoid exhaustive searching and emphasize on fast designing the multi-channel decision stump via K-means clustering.

## 3 Logistic Multiple Instance Boost

If pedestrian have uncommon stance, human-centering normalization often produces miss-aligned examples as illustrated in Fig. 1. Intuitively, some parts of human can be aligned by shifting the normalization window. Therefore, we augment the training set by perturbing the training examples. The created instances can take advantage of all information of the "omega" heads and the rectangle bodies. Moreover, the augmented training set should cover the possible pose variations for MIL. Fig. 2 illustrates the proposed approach.

Compared with traditionally supervised learning, an instance in MIL is indexed with two indices: $i$ which indexes the bag, and $j$ which indexes the instance within the bag. Given a bag $\mathbf{x}_i$, the conditional probability of the bag-level class $\mathbf{y}_i$ is

$$p(\mathbf{y}_i|\mathbf{x}_i) = \frac{1}{n_i}\sum_{j=1}^{n_i} p(y_{ij}|x_{ij}), \tag{1}$$

where $n_i$ is the number of the instances in the $i$-th bag, $y_{ij}$ is the instance-level class label for the instance $x_{ij}$. Equation.(1) indicates that every instance contributes equally to the bag's label. This simple assumption is suitable for the instances generated by perturbing around the person. Because the generated every instance is positive pedestrian image.

The instance-level class probability is given as $p(y|x) = 1/(1 + e^{\beta x})$, where $\beta$ is the parameter to be estimated. Controlling the parameter $\beta$ gives different

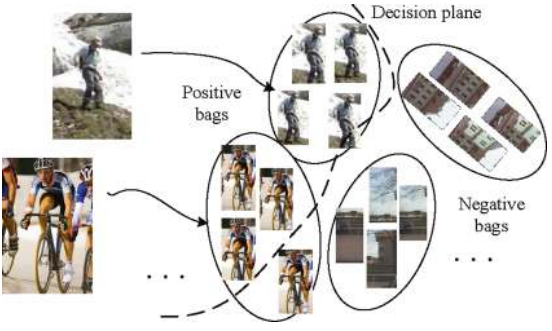


**Fig. 2.** Overview of the multiple instance learning process. The training example is first converted into a bag of instances. Note that we only generate the instances spatially, and the instances can also be generated at different scales. Therefore, the resulting classifier will withstand the translation and scale transformation.

instance-level class probability, which gives different contribution to bag-level probability. Ideally, the "well" aligned instances should be assigned higher probability than the non-aligned. Given a collection of $N$ *i.i.d* bags $\mathbf{x}_1, \ldots, \mathbf{x}_N$, the parameter $\beta$ can be estimated by maximizing the bag-level binomial log-likelihood function

$$L = \sum_{i}^{N} \left[ \mathbf{y}_i \log p(\mathbf{y}_i = 1|\mathbf{x}_i) + (1 - \mathbf{y}_i) \log p(\mathbf{y}_i = 0|\mathbf{x}_i) \right]. \tag{2}$$

Equation.(2) can not be solved analytically. Xu el al [9] propose an boosting method to maximize the log-likelihood function. We need to learn a bag-level function $\mathbf{F}(\mathbf{x}) = \sum_m c_m \mathbf{f}_m(\mathbf{x})$ and the corresponding *strong* classifier $H = \text{sign}(\mathbf{F}(\mathbf{x}))$, where weights $c_1, \ldots, c_M \in \mathbb{R}$, the $\mathbf{f}$ is the bag-level weak classifier. The expected empirical loss is

$$E[I(\mathbf{F}(\mathbf{x}) \neq \mathbf{y})] = -\frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{F}(\mathbf{x}_i), \tag{3}$$

where $I(\cdot)$ is the indicator function. We are interesting in wrapping the bag-level weak classifier $\mathbf{f}$ with the instance-level weak classifier $f$. Using the Equation.(1), Equation.(3) is converted into the instance-level's exponential loss $E_{\mathbf{x}} E_{\mathbf{y}|\mathbf{x}}[e^{-yf}]$ as $e^{-\mathbf{yH}} \geq I(H(\mathbf{x}) \neq \mathbf{y}), \forall M$. One searches for the optimal update $c_m f_m$ such that minimizes

$$E_{\mathbf{x}} E_{\mathbf{y}|\mathbf{x}} \left[ e^{-y_{ij} F_{m-1}(x_{ij}) - c_m y_{ij} f_m(x_{ij})} \right] = \sum_i w_i e^{[(2\epsilon_i - 1)c_m]}, \tag{4}$$

where $\epsilon_i = \sum_j 1_{f_m(x_{ij}) \neq y_{ij}} / n_i$, $w_i$ is the example's weight. The error $\epsilon_i$ describes the discrepancy between the bag's label and instance's label. The instance in positive bags with higher score $f(x_{ij})$ gives higher confidence to the bag's label, even though there are some negative instances occurring in the positive bag.

**Algorithm. 1** Graph embedding based decision stump

**Input:**

The training data $\{h_i, y_i\}, i = 1, \ldots, n$

**Training:**

1. Learn the projection matrix $P \in \mathbb{R}^{1\times D}$ by Equation. (4), and project the data into one dimension manifold $\{\hat{h}_i, y_i\}, \hat{h}_i = Ph_i$.
2. Calculate the clustering center $Pc = \{C_1^p, \ldots, C_{N_p}^p\}$ and $Nc = \{C_1^n, \ldots, C_{N_n}^n\}$ for the positive and negative data via K-means, where $N_p$ and $N_n$ is the number of clustering center.
3. Sort the clustering center $C = \{Pc, Nc\}$ with ascendent order, and find the middle value $r_k = (C_k + C_{k+1})/2$ as the rough decision plane.
4. Generate the histogram with the intervals $\sigma_k = (r_k, r_{k+1}]$, and produce the class label $\omega_c$ for each interval via Bayesian decision rule.
5. Iteratively merge adjacent intervals with same decision label $\omega_c$ to produce a set of consistent intervals $\hat{\sigma}_k$.

**Output:**

A LUT function lup($k$) on the merged intervals $\hat{\sigma}_k, k = 1, \ldots, K$.

Therefore, the final classifier often classifies these bags as positive. The variations problem in training examples will be reduced.

## 4 Graph Embedding Based Decision Stump

### 4.1 Supervised Graph Embedding

Let $h_i \in \mathbb{R}^D (i = 1, 2, \ldots, n)$ be the $D$-dimensional histogram feature and $y_i \in \{\omega_c\}_{c=1}^2$ be the associated class label. The feature is written as matrix form: $H = (h_1|h_2|\ldots|h_n)$. Let $G = \{\{h_i\}_{i=1}^n, S\}$ be an undirected weighted graph with vertex set$\{h_i\}_{i=1}^n$ and the similarity matrix $S \in \mathbb{R}^{n\times n}$. The element $s_{i,j}$ of matrix $S$ measures the similarity of vertex pair $i$ and $j$. The unsupervised graph embedding is defined as the optimal low dimension vector representations for the vertices of graph $G$

$$P^* = \min_{P^T HMH^T P=I} \sum_{i,j} ||Ph_i - Ph_j||^2 s_{i,j} = \min_{P^T HMH^T P=I} 2\mathrm{tr}(P^T HLH^T P), \quad (5)$$

where projection $P \in \mathbb{R}^{d\times D}, (d < D)$ maps feature $h$ from high dimension space $\mathbb{R}^D$ to low dimension space $\mathbb{R}^d$. The elements in the diagonal matrix $M$ is $m_{i,j} = \sum_{i\neq j} s_{i,j}$, and the Laplacian matrix $L$ is $M - S$.

The similarity $s_{i,j}$ connects the relationship between high dimension and low dimension space. If two vertexes $h_i$ and $h_j$ are close, $s_{i,j}$ will be large, and vice versa. To classification, the projection $P$ should keep the between-class far and within-class close. The similarity matrix $S$ should reflect the separable ability. The between-class similarity $s_{i,j}^b$ and within-class similarity $s_{i,j}^w$ can be defined as[1]

[1] We refer the interested reader to [10] for more details.

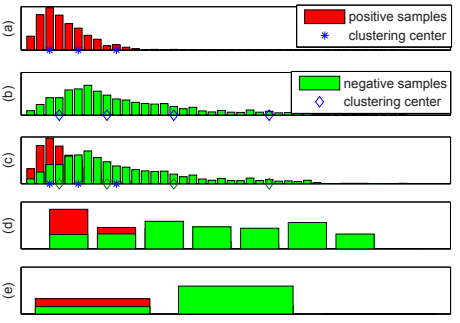


**Fig. 3.** A demonstration of generating the multi-channel decision stump. (a)-(b) Cluster on positive and negative examples, respectively. (d)Generate the decision stumps via histogram. (e)Merge the consistent decision stumps.

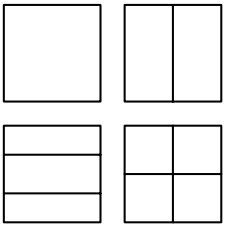

**Fig. 4.** 4 type block feature

$$s^b_{i,j} = \begin{cases} 1/n - 1/n_c \text{ if } & y_i = y_j = \omega_c, \\ 1/n \quad \text{if} & y_i \neq y_j, \end{cases} \qquad s^w_{i,j} = \begin{cases} 1/n_c \text{ if } & y_i = y_j = \omega_c, \\ 0 \quad \text{if} & y_i = y_j, \end{cases}, \tag{6}$$

where $n_c$ is the cardinality of the $\omega_c$ class. The pairwise $s^b_{i,j}$ and $s^w_{i,j}$ try to keep within-class sample close (since $s^w_{i,j}$ is positive and $s^b_{i,j}$ is negative if $y_i = y_j$) and between-class sample pairs apart (since $s^b_{i,j}$ is positive if $y_i \neq y_j$). The projection matrix $P$ can be calculated by Fisher criterion

$$P^* = \max_{P^T H(M_w - S_w)H^T P = I} \text{tr}(P^T H(M_b - S_b)H^T P). \tag{7}$$

The projection matrix $P = [p_1, p_2, \ldots, p_l]$ are solved by generalized eigenvectors corresponding to the $l$ largest eigenvalues $p$ in $H(M_w - S_w)H^T p_l = \lambda H(M_b - S_b)H^T p_l$.

### 4.2 Multi-channel Decision Stump

According to Bayesian decision theory, if class conditional probability $p(\omega_1|x) > p(\omega_2|x)$ we would naturally incline to decide that the true label of $x$ is $\omega_1$, and vice versa. Using Bayes rule $p(\omega|x) = p(x|\omega)p(\omega)$, the optimal decision plane is located at where $p(x|\omega_1) = p(x|\omega_2)$ with $p(\omega_1) = p(\omega_2)$. We obtain the Bayes error $p(error|x) = \int min[p(x|\omega_1), p(x|\omega_2)]dx$. However, the $p(x|\omega_c)$ is not directly available. To accurately estimate the $p(x|\omega_c)$, histogram needs large numbers of bins via uniform sampling in [25,26]. We avoid estimating the $p(x|\omega_c)$ with uniform sampling or rejection sampling. As demonstrated in Fig. 3(c), we consider the local region of feature space, and the location at the middle of two modal is a natural decision plane. The decision plane would approximately minimize Bayes error, if $p(\omega_1) = p(\omega_2)$. Algorithm. 1 shows the graph embedding based decision stump. Note that the number of decision planes is automatically decided.

## 5 Pedestrian Detection

To achieve the fast pedestrian detection, we adopt the cascade structure of detector [6]. Each stage is designed to achieve high detection rate and modest false positive rate. We combine $K = 30$ LMIBoost on HOG feature with rejection cascade. To exploit the discriminative ability of HOG feature, we design 4 type block feature as showed in Fig.4. In each cell, 9-bins HOG feature is extracted and concatenated into a single histogram to represent the block feature. To obtains a modicum of illumination invariance, the feature is normalized with L2 norm. The dimension of the 4 different type feature are 9, 18, 27 and 36, respectively. The $453 \times 4$ number of block HOG feature can be computed from a single detection window.

Assuming that the $i$-th cascade stage is trained, we classify all the possible detection window on the negative training images with the cascade of the previous $k$-1 LMIBoost classifiers. The examples which are misclassified in scanning window form the possible new negative training set. While, the positive training samples do not change during bootstrap. Let $N_{pi}$ and $N_{ni}$ be cardinality of the positive and negative training examples at $i$-th stage. Considering the influence of asymmetric training data on the classifier and computer RAM limitations, we constrain $N_{pi}$ and $N_{ni}$ to be approximately equal.

According to "There is no free lunch" theorem, it is very important to choose suitable number of instances in a bag for training and detection. More instances in a bag will represent more variations and improve the detection results, but will also reduce the training and detection speed. We experimentally set 4 instances for training and detection, respectively. Each level of cascade classifier is optimized to correctly detect at least 99% of the positive bags, while reject at least 40% of the negative bags.

## 6 Experiments

To test our method, we perform the experiments on two public dataset: INRIA [4] and VOC2006 [1]. The INRIA dataset contains 1239 pedestrian images (2478 with their left-right reflections) and 1218 person-free images for training. In the test set, there are 566 images containing pedestrians. The pedestrian images provided by INRIA dataset have large variations (but most of them have standing pose), different clothing and urban background. This dataset is very close to real-life setting. The VOC2006's *person* detection subtask supplies 319 images with 577 person as training set, and 347 images with 579 person as validation set. 675 images with 1153 person is supplied as test data. Note that the VOC2006's person detection dataset contains various human activities, different stances and clothing. Some examples of the two different datasets are showed in Fig. 8.

### 6.1 Performance Comparisons on Multiple Datasets

We plot the detection error tradeoff curves on a log-log scale for INRIA dataset. The $y$-axis corresponds to the miss rate, $FalseNeg/(FalseNeg+TruePos)$ and the

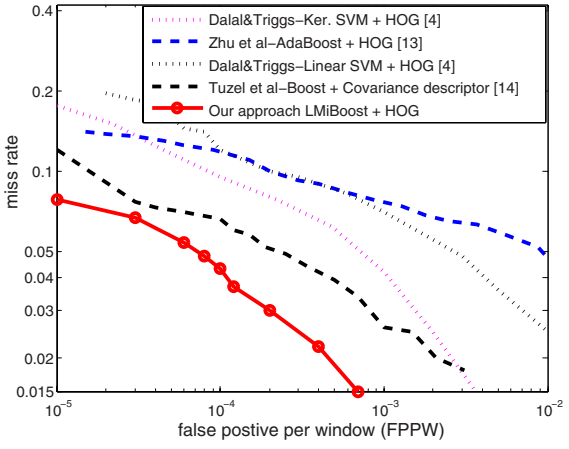


**Fig. 5.** Comparison results on INRIA dataset. Note that the curve of our detector is generated by changing the number of cascade stage used.

$x$-axis corresponds to false positives per window(FPPW), $FalsePos/(TrueNeg+FalsePos)$. We compare ours results with [4,13,14] on INRIA dataset. Although it has been noted that kernel SVM is computationally expensive, we consider both the kernel and linear SVM method of [4]. Only the best performing result, the L2-norm in HOG feature, is considered. Covariance descriptor [14] is also compared. Fig. 5 shows that the performance of our method is comparable to the state-of-art approaches. We achieve 4.3% miss rate at $10^{-4}$ FPPW. Notice that all the results by other methods are quoted directly from the original papers, since we perform the same separation of training-testing sets.

The Fig.7 shows the precision-recall cure on VOC2006 person detection subtask for comp3 [1]. The protocol of the comp3 is that the training data is composed of the training set and validation set. The non-normalized examples are first approximately aligned, and then be converted into a bag of instances. Some truncated and difficult examples in training data are discarded. The standard scanning window technique is adopted for detection, although the scanning window may be not suitable for VOC2006 detection subtask. The average precision scores is 0.23, which is better than the best results 0.164 reported by INRIA_Douze [1]. In Fig. 8, several detection results are showed for different scenes with human having variable appearance and pose. Significantly overlapping detection windows are averaged into a single window.

### 6.2 Analysis of the Weak Classifiers

For our next experiment, we conduct experiments to compare the performance of different weak classifiers. A common set of parameters (such as, false positive rate for every stage)are controlled equally for cascade training. Two detectors are trained with different weak classifiers, including FLDA and graph embedding based decision stump.

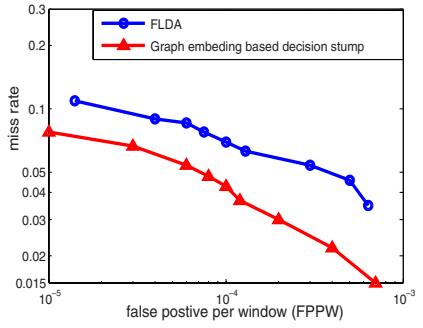


**Fig. 6.** Comparison on different weak classifier

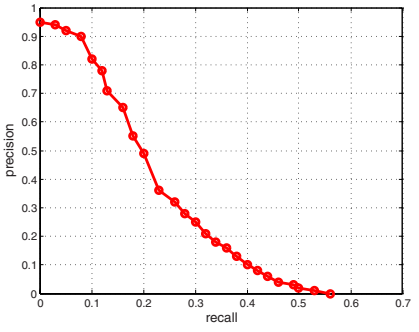


**Fig. 7.** Performance on VOC2006 dataset

(a) INRIA dataset (b) VOC2006 dataset

**Fig. 8.** Some detection samples on INRIA and VOC2006 datasets

The performance results on INRIA show that the detectors based on graph embedding decision stump outperforms the detector based on FLDA in Fig. 6. Unlike the other LUT weak classifier [26,25], the bins of decision stumps are automatically decided by algorithm.

### 6.3 Analysis of the Detection Speed

There are 90% of the negative examples are rejected at first five stage. The speed of the cascaded detector is directly related to the number of feature evaluated per scanned sub-window. For INRIA dataset, on average our method requires to evaluate 10.05 HOG feature per negative detection window. Densely scanning at 0.8 scale and 4 pixel step in a $320 \times 240$ image needs average 150ms under PC with 2.8GHz CPU and 512RAM. While, 250ms for $320 \times 240$ image is reported in Zhu et al's detector [13].

## 7 Conclusion and Future Work

We introduce the multiple instance learning into the pedestrian detection for solving pose variations. The training example does not need to be well aligned, but to be represented as a bag of instances. To efficiently utilizing histogram feature, a graph embedding based decision stump is proposed. The weak classifier guarantees the fast detection and better discriminative ability. The promising performances of the approach are shown on INRIA and VOC2006's person detection subtask.

Using multiple instance learning has enabled detector robust to the pose and appearance variations. Theoretically, the more instances are supplied, the more variations would be learned. Modeling the average relationship between the instance's label and bag's label may be unsuitable when there are large numbers of instances in a positive bag. In future, more experiments will be carried out to compare the different way to model the relationship.

## Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grant 60773136 and 60702035, in part by National Hi-Tech Development Program (863 Program) of China under Grant 2006AA01Z117 and 2006AA010105. We would also thank the anonymous reviewers for their valuable comments.

## References

1. Everingham, M., Zisserman, A., Williams, C.K.I., Gool, L.V.: The PASCAL Visual Object Classes Challenge (VOC 2006) Results (2006), `http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf`
2. Gavrila, D.M.: Pedestrian detection from a moving vehicle. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 37–49. Springer, Heidelberg (2000)
3. Bissacco, A., Yang, M., Soatto, S.: Detection human via their pose. In: Proc. NIPS (2006)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR, vol. I, pp. 886–893. IEEE, Los Alamitos (2005)
5. Papageorgiou, P., Poggio, T.: A trainable system for object detection. IJCV, 15–33 (2000)
6. Viola, P., Jones, M., Snow, D.: Detecing pedestrians using patterns of motion and appearance. In: Proc. ICCV (2003)
7. Maron, O., Lozanno-Perez, T.: A framework for multiple-instance learning. In: Proc. NIPS, pp. 570–576 (1998)
8. Dietterich, T., Lathrop, R., Lozano-Perez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artifical intelligence, 31–71 (1997)
9. Xu, X., Frank, E.: Logistic regression and boosting for labeled bags of instances. In: Dai, H., Srikant, R., Zhang, C. (eds.) PAKDD 2004. LNCS (LNAI), vol. 3056, pp. 272–281. Springer, Heidelberg (2004)

10. Sugiyama, M.: local fisher discriminat analysis for supervised dimensionality reduction. In: Proc. ICML (2006)
11. Monhan, A., Papageorgiou, C., Poggio, T.: Example-based object detection in images by componets. IEEE Trans. PAMI 23, 349–360 (2001)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints, 91–110 (2004)
13. Zhu, Q., Avidan, S., Yeh, M.C., Cheng, K.T.: Fast human detection using a cascade of histograms of oriented gradients. In: Proc. CVPR, vol. 2, pp. 1491–1498. IEEE, Los Alamitos (2006)
14. Tuzel, O., Porikli, F., Meer, P.: Human detection via classification on riemannina manifolds. In: Proc. CVPR. IEEE, Los Alamitos (2007)
15. Zisserman, A., Schmid, C., Mikolajczyk, K.: Human detection based on a probabilistic assembly of robust part detectors. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 69–82. Springer, Heidelberg (2004)
16. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowed scenes. In: Proc. CVPR, pp. 878–885. IEEE, Los Alamitos (2005)
17. Tran, D., Forsyth, D.A.: Configuration estimates improve pedestrian finding. In: Proc. NIPS (2007)
18. Felzenszwalb, P., Mcallester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: Proc. CVPR. IEEE, Los Alamitos (2008)
19. Maron, O., Ratan, A.: Multiple-instance learning for natural scene classification. In: Proc. ICML (1998)
20. Viola, P., Platt, J.C., Zhang, C.: Multiple instance boosting for object detection. In: Proc. NIPS (2006)
21. Porikli, F.M.: Integral histogram: a fast way to extract histogram in cartesian space. In: Proc. CVPR, pp. 829–836. IEEE, Los Alamitos (2005)
22. Liu, C., Shum, H.Y.: Kullback-leibler boosting. In: Proc. CVPR, pp. 587–594 (2003)
23. Laptev, I.: Improvements of object detection using boosted histograms. In: BMVC (2006)
24. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. CVPR. IEEE, Los Alamitos (2001)
25. Huang, C., Ai, H., Wu, B., Lao, S.: Boosting nested cascade detector for multi-view face detection. In: Proc. ICPR. IEEE, Los Alamitos (2004)
26. Xiao, R., Zhu, H., Sun, H., Tang, X.: Dynamic cascades for face detection. In: Proc. ICCV. IEEE, Los Alamitos (2007)