A Sample Pre-mapping Method Enhancing Boosting for Object Detection

Haoyu Ren[1] *Xiaopeng Hong*[2,1] *Cher-Keng Heng*[3] *Luhong Liang*[1] *Xilin Chen*[1]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences,
Institute of Computing Technology, CAS, China
[2]School of Computer Science and Technology, Harbin Institute of Technology
[3]Panasonic Singapore Laboratories Pte Ltd
{hyren, xphong, lhliang, xlchen}@jdl.ac.cn CherKeng.Heng@sg.panasonic.com

Abstract

We propose a novel method to improve the training efficiency and accuracy of boosted classifiers for object detection. The key step of the proposed method is a sample pre-mapping on original space by referring to the selected 'reference sample' before feeding into weak classifiers. The reference sample corresponds to an approximation of the optimal separating hyper-plane in an implicit high-dimensional space, so that the resulting classifier could achieve the performance similar to kernel method, while spending the computation cost of linear classifier in both training and detection. We employ two different non-linear mappings to verify the proposed method under boosting framework. Experimental results show that the proposed approach achieves performance comparable with the common-used methods on public datasets in both pedestrian detection and car detection.

1. Introduction

Object detection of a special class is a basic vision task, where SVM (Support Vector Machine) approaches [2, 8, 10, 11, 15] and boosting methods [1, 3, 4, 5, 6, 7, 13, 14] are widely used. Due to the efficient cascade framework proposed by Viola and Jones [1], boosting family algorithms achieve satisfactory performance and fast speed on many detection tasks. The cascade classifier [1] performs well on some simple object categories, e.g., the frontal-view faces and frontal/rear-view pedestrians. However, for those more complicated tasks such as multi-view and multi-pose pedestrian detection, classifiers with more discriminative power are required. Some previous approaches use more powerful weak classifiers, e.g., linear SVM [3], Least Square [4] and Fisher linear discriminative analysis [6]. Other approaches follow the divide-and-conquer strategy to build strong classifiers with more complex structures, e.g., the vector boosted tree [5] and cluster boosted tree [7]. Most of the above algorithms utilize linear (weak) classifiers for efficient training and classification.

The discriminative powers of classifiers often depend on measuring the differences between distributions of features. In most of the cases, using appropriate kernel functions to measure the similarity between features is more effective than linear functions. However, owing to the high computation cost of kernel operations, the speed is relatively slow in both training and classification.

To solve this problem, we propose a novel pre-mapping approach as an enhancing step in the weak classifier selection of the boosting framework. The pre-mapping maps the original space to a new space referring to the selected 'reference samples' by the boosting framework. Linear classification in the mapped space is an efficient approximation of the kernel methods mentioned above. As a result, using the proposed pre-mapping method could consistently improve both the training efficiency and the accuracy of the boosted classifier, compared with using linear weak classifiers directly. In our case, the mapping can be realized by some kernel functions such as the proposed Minimum Intersection Mapping (MIM) or the Chi-Square Mapping (CSM).

Plenty of experiments on public datasets are used to evaluate our method. The results show that our approach apparently improves both the training efficiency and the performance of the boosted classifier, which achieves comparable results with the commonly used approaches in both pedestrian and car detection tasks.

2. Sample Pre-mapping

2.1 Classification in pre-mapped space

Many of the kernel methods, which achieve better performance than the linear methods on many tasks, map the original observations into an implicit high-dimensional space, where the linear classification is subsequently applied. Generally, linear classification in the implicit space can be implemented in the

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.736

IEEE computer society

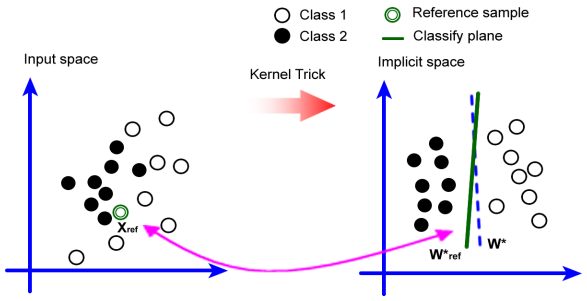


Fig. 1 Reference Sample

original space through the so-called kernel trick. More specifically, given two *d*-dimensions sample vectors **x, z** in the original space and a kernel function ***K*(x,z)** that satisfies the Mercer's Condition, there exists a function $\varphi: \mathrm{R}^d \rightarrow \mathrm{R}^m$, such that

$$K(\mathbf{x},\mathbf{z}) = \varphi(\mathbf{x}) \bullet \varphi(\mathbf{z}), \quad (1)$$

where $\bullet$ denotes the dot product of two vectors.

Denote the optimal classification hyper-plane in the implicit *m*-dimensions space by $\mathbf{w}^*$, given a sample in the original *d*-dimensions space by $\mathbf{x} = [\mathbf{x}_1,\ldots,\mathbf{x}_d]$, the optimal classification function $F^*(\mathbf{x})$ is the dot product of the mapped vector $\varphi(\mathbf{x})$ and $\mathbf{w}^*$ as

$$F^*(\mathbf{x}) = \mathbf{w}^* \bullet \varphi(\mathbf{x}). \quad (2)$$

In the extreme case, if there is a vector $\mathbf{x}^* \in \mathrm{R}^d$ satisfies $\varphi(\mathbf{x}^*) = \mathbf{w}^*$, (2) can be implemented by (3)

$$\mathbf{w}^* \bullet \varphi(\mathbf{x}) = \varphi(\mathbf{x}^*) \bullet \varphi(\mathbf{x}) = K(\mathbf{x},\mathbf{x}^*). \quad (3)$$

The only problem is to find out such an $\mathbf{x}^*$ in this case.

However, in most of the cases, because φ is *not invertible* or even φ itself could not be explicitly described, it seems impossible to find such an $\mathbf{x}^*$. Fortunately, as illustrated in Fig. 1, in boosting framework, we could approximate $\mathbf{x}^*$ efficiently by selecting a '*reference sample*' $\mathbf{x}_{ref}$ from the training samples in each iteration. $\mathbf{x}_{ref}$ is such a vector, whose corresponding hyper-plane $\mathbf{w}_{ref}^* = \varphi(\mathbf{x}_{ref})$ has the minimum classification error compared with $\mathbf{w}^*$. The optimal F^* is then approximated using the classification function F defined by (4)

$$F(\mathbf{x}) = \mathbf{w}_{ref}^* \bullet \varphi(\mathbf{x}) = K(\mathbf{x},\mathbf{x}_{ref}), \quad (4)$$

which implies that by referring to an appropriate reference sample, using the kernel function is equal to linear classification in the implicit space.

Generally, we could use any kernel for the *K* in (4). In consideration of the efficiency, we focus on the additive kernels with simple structure as

$$K(\mathbf{x},\mathbf{x}_{ref}) = \sum_{i=1}^{d} k(x_i, x_{ref,i}). \quad (5)$$

We further fit a linear weak classifier using (5) as

$$F(x) = \sum_{i=1}^{d} a_i k(x_i, x_{ref,i}) + b, \quad (6)$$

and integrate it into the boosting framework, where $\mathbf{a}=[a_1,\ldots,a_d]$ is a *d*-dimensional weight vector. In consideration that the coefficients in (6) are actually learnt in the *d*-dimensions space formed by $k(\mathbf{x}_i, \mathbf{x}_{ref,i})$, to utilize this observation, we propose the sample pre-mapping $\phi: \mathrm{R}^d \times \mathrm{R}^d \rightarrow \mathrm{R}^d$, which is formulated as

$$\Phi(\mathbf{x}) = \phi(\mathbf{x}, \mathbf{x}_{ref}) = [k(x_1, x_{ref,1}),\ldots, k(x_d, x_{ref,d})]^T. \quad (7)$$

We give two examples, the Minimum Intersection Mapping (MIM) and Chi-Square Mapping (CSM) based on two commonly used additive kernels, Histogram intersection kernel and Chi-Square Kernel respectively as (8) and (9) [10].

$$k_{MIN}(\mathbf{x},\mathbf{x}_{ref}) = \min(x_i, x_{ref,i}) \quad (8)$$

$$k_{CSM}(\mathbf{x},\mathbf{x}_{ref}) = (x_i - x_{ref,i})^2 / (x_i + x_{ref,i}). \quad (9)$$

In fact, differences between the kernels tend not to matter on large training sets [11]. So the discrimination power of the trained boosted classifiers with different sample pre-mappings is similar, as the experimental result in section 3.1.

2.2 Boosting with sample pre-mapping

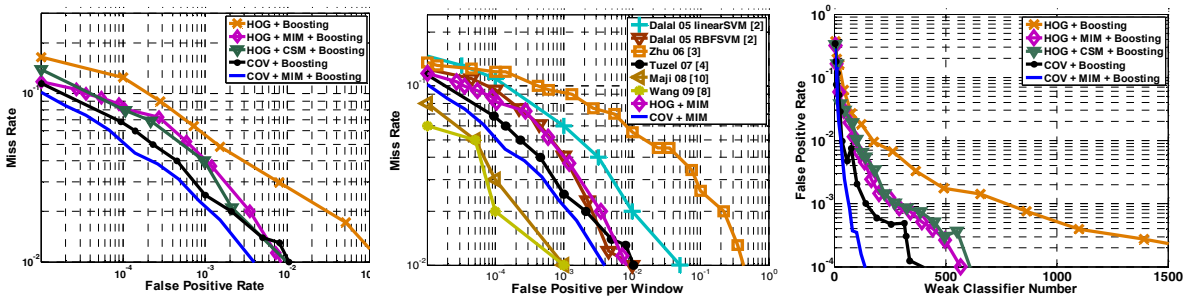
Parameters:
- N Number of training samples
- N_r Number of reference samples evaluated in each iteration
- N_f Number of features evaluated in each iteration
- T Maximum number of weak classifiers

Input: Training set $\{(x_i, y_i)\}\ i = 1,\ldots, N$, $\mathbf{x}_i \in \mathrm{R}^d$, $y_i \in \{-1,1\}\}$

1. Initialization $\omega_i = 1/N$, $H(\mathbf{x}_i) = 0$, $i = 1,\ldots, N$
2. Repeat for $t = 1,\ldots,T$
   - 2.1 Update the sample weight ω_i using the t^{th} weak classifier output $F_t(\mathbf{x}_i)$
   - 2.2 For $m = 1$ to N_f (Choose the best feature)
     - For $n = 1$ to N_r (Choose the best reference sample)
       - 2.2.1 Randomly select a reference sample $\mathbf{x}_{ref}$ from the training set
       - 2.2.2 Calculate the feature vectors in the mapped space according to ϕ in (7)
       - 2.2.3 Train a classification hyper-plane **a** in the mapped space and select the best feature according to classification error of (6)
   - 2.3 Update the strong classifier $H(\mathbf{x}_i)$ using $F_t(\mathbf{x}_i)$
3. Output classifier $H(\mathbf{x}) = sign[\sum_{t=1}^{T} F_t(\mathbf{x})]$

Fig. 2 RealBoost training with sample pre-mapping

In this section, we introduce how to utilize the proposed method into boosting framework. We integrate the sample pre-mapping as an individual step in each boosting round, as shown in Step 2.1~2.3 of Fig. 2. In the boosting training procedure, we evaluate several training samples to select the reference sample in each iteration. For each candidate reference sample, we train a linear classification hyper-plane **a** in the mapped space using Least Square and further choose the best with the minimum classification error. In boosting framework, the approximation error of using



(a) (b) (c)

Fig. 3 Per-patch performance evaluation on INRIA dataset

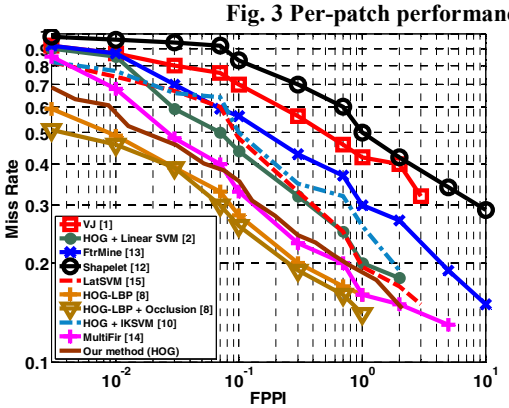


Fig. 4 Per-image performance evaluation on INRIA dataset

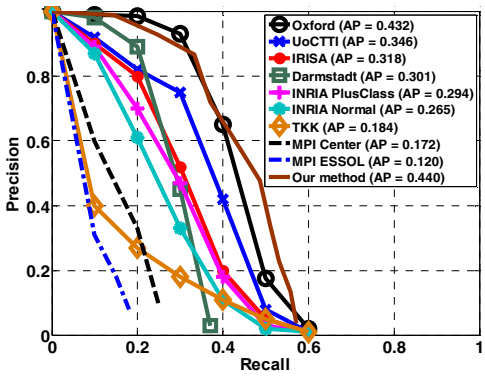


Fig. 5 Performance evaluation on PASCAL car dataset

$\mathbf{x}_{ref}$ instead of the optimal hyper-plane for classification could be minimized by averaging over several weak classifiers [4].

Fig. 2 uses RealBoost [1] [5] as an example to illustrate more details. The updating strategies of the sample weights (Step 2.1) and the classifier response (Step 2.3) are different in other boostings.

3. Experiment

3.1 Experiment on INRIA pedestrian dataset

We evaluate the proposed pre-mapping method on the commonly used INRIA pedestrian dataset [2]. The multi-size HOG descriptors [3] and covariance matrix descriptor [4] are used to train cascade classifiers following the algorithm described in Fig. 2. 4,496 HOG descriptors and 13,440 COV descriptors are generated as the feature pool for the 64×128 window.

We firstly follow the training and testing criteria at patch level proposed by Dalal & Triggs [2]. In Fig. 3, we plot the ROC curves on a log-log scale by tuning the rejection threshold of the classifiers. In Fig. 3(a), we compare the performances of the boosted classifiers with different sample pre-mappings and without sample pre-mapping. It can be seen that all the algorithms with sample pre-mapping reach better performance compared with the algorithms without sample pre-mapping on both HOG feature (~4% at 10e-4 FP) and COV feature (~2% at 10e-4 FP). In Fig. 3(b), we show that our methods reach similar accuracy with the commonly used boosted algorithms [3, 4] in pedestrian detection. The detection rate is expected to be better using more complicate features such as pyramid HOG [10] or HOG-LBP [8]. We also list the training and detection speed of these methods in Table 1. All the algorithms are tested on an P4 Intel 3.0GHz CPU with 4.0G memory. Both the training speed and detection speed are improved using the sample pre-mappings.

Table 1 Comparison of the training and detection speed

Algorithm	Training time to 10e^{-5} FP	Patches per sec. in detection
HOG w/o pre-mapping	72+ hours	~85k
HOG + MIM	~20 hours	~95k
HOG + CSM	~20 hours	~90k
COV w/o pre-mapping	~72 hours	~8.2k
COV + MIM	~24 hours	~10k

The experimental results above show the efficiency and the effectiveness of the proposed method. Benefit from the pre-mapping procedure that enhances the classification power to the level of kernel method, the boosted classifier achieves better accuracy. The training time of the boosted classifier mainly depends on the weak classifier number. We can find that the algorithms with sample pre-mapping use only 1/3 weak classifiers comparing the original version as in

[1] In RealBoost, the weak classifier is a piecewise function based on the probability distribution of $F(\mathbf{x})$ in practice.

Fig. 3(c). Less weak classifier number will also contribute to the detection speed. Therefore, using the sample pre-mapping is more efficient. So we can conclude that the proposed method could consistently improve both the training efficiency and the detection accuracy of boosted classifiers.

Furthermore, we evaluate our method with HOG feature under the criteria of the detection rate versus False Positive rate Per Image (FPPI) [9]. Fig. 4 gives the comparison with the state-of-the-art detectors [1, 2, 8, 10, 12, 13, 14, 15] on INRIA dataset. Our algorithm shows comparable result, which is expected to be better using more complicated feature [8] or feature combination [14]. Our cascade framework is also an advantage compared with SVM algorithms [2, 8, 10, 15] on the detection speed.

3.2 Experiment on PASCAL VOC car dataset

In this section, we evaluate the sample pre-mapping with HOG feature on the multi-view car detection task. We follow the training and detection criteria of PASCAL VOC 2007 challenge [16]. The positive training samples with 'hard' label, strong occlusion or weak illumination are dropped. We plot the recall tradeoff 1-precision curve in Fig. 5. The reported result of PASCAL 2007 is also drawn in the chart. It can be seen that the proposed method achieves the best result, with the AP (average precision, the arithmetical average of the precision points with 0, 0.1, … , 1.0 recall) at 0.440.

4. Conclusion

In this paper, we proposed a sample pre-mapping method that is capable of improving training speed and accuracy of boosted classifiers. We mapped the original samples to a space defined by identified reference samples. Classification in the mapped space is capable of approximating the classification power of kernel method while maintaining the speed of linear classifier. In this paper, we prove the effectiveness of our methods through two examples of the many possible non-linear mappings, namely MIM and CSM. Both the two pre-mappings contribute to a considerable reduction in miss rate at all FPPW and three-time acceleration in convergence speed during training, compared to the traditional method. Experimental results show that our method achieves performance comparable to the commonly used methods in both pedestrian and car detection.

The approach proposed in this paper is promising to be further studied. We will apply our sample pre-mapping to other descriptors and machine learning algorithms in our future work.

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.

References

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

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

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

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

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

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

[7] B. Wu and R. Nevatia. Cluster Boosted Tree Classifier for Multi-View, Multi-Pose Object Detection. In *ICCV*, 2007.

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

[9] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: A Benchmark. In *CVPR*, 2009.

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

[11] S. Maji and A. Berg. Max-Margin Additive Classifiers for Detection. In *ICCV*, 2009.

[12] P. Sabzmeydani and G. Mori. Detecting Pedestrians by Learning Shapelet Features. In *CVPR*, 2007.

[13] P. Dollar, Z. Tu, H. Tao, and S. Belongie. Feature Mining for Image Classification. In *CVPR*, 2007.

[14] C.Wojek and B. Schiele. A Performance Evaluation of Single and Multi-feature People Detection. In *DAGM*, 2008.

[15] P. Felzenszwalb, D. McAllester, and D. Ramanan. A Discriminatively Trained, Multiscale, Deformable Part Model. In *CVPR*, 2008.

[16] http://pascallin.ecs.soton.ac.uk/challenges/VOC/