Boosted Sigma Set for Pedestrian Detection

Xiaopeng Hong[1,2] Hong Chang[2] Xilin Chen[2] Wen Gao[3,1]
[1] School of Computer Science and Technology, Harbin Institute of Technology
[2] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences, Institute of Computing Technology, CAS, China
[3] Institute of Digital Media, Peking University, China
{xphong, hchang, xlchen, wgao}@jdl.ac.cn

Abstract—**This paper presents a new method to detect pedestrian in still image using Sigma sets as image region descriptors in the boosting framework. Sigma set encodes second order statistics of an image region implicitly in the form of a point set. Compared with the covariance matrix, the traditional second order statistics based region descriptor, which requires computationally demanding operations based on Riemannian manifold, Sigma set preserves similar robustness and discriminative power more efficiently because the classification on Sigma sets can be directly performed in vector space. Experimental results on the INRIA and the Daimler Chrysler pedestrian datasets show the effectiveness and efficiency of the proposed method.**

Keywords-pedestrian detection; boosting; covariance matrix; Sigma set

1. Introduction

Pedestrian detection plays crucial roles in many computer vision tasks, such as visual surveillance, smart cars, and image retrieval. It is a challenging task due to the variations in pose, shape, motion, and appearance of pedestrians.

This paper concentrates on one of the fundamental problems of pedestrian detection: using what kind of descriptors to represent image regions effectively and efficiently. To this end, we discuss on the holistic modeling approaches, which describes the human body as a unified model [1, 4, 7, 9-11, 13-14, 17-18], in consideration that good descriptors can also be applied to the part based approaches, which represent the human body as a collection of several part models [3, 6, 8, 16].

Plenty of descriptors have been proposed for pedestrian detection. Some early works include Haar-like wavelet [11] and its space-time differences extension [14], local receptive fields [9], and edge-let [16], *etc*.

Using gradient-orientation-histogram based descriptors is a popular choice. One seminal work is the Histograms of Oriented Gradients (HOG) proposed by Dalal and Triggs [1]. The boosted HOG [18], multi-level oriented edge energy features [7], and adaptive contour features [4], *etc*., also belong to this kind of descriptors.

Recently, using 2^{nd} order statistics (typically, the covariance matrix (COV) [10][13][17]) to represent an image region attracts much attention. COVs capture the co-occurrence of (the 2^{nd} order statistics of) a few elementary features. COVs have good discriminative power and robustness, *e.g.*, as been verified in [17], COV outperforms the features of HOG and edge-let in pedestrian detection. In addition, compared with histograms when covering the same number of elementary features, the dimensionality of COVs is relatively lower. However, since COVs do not lie in Euclidean space, the time-consuming operations based on Riemannian manifold are required to measure the distance between COVs accurately. In addition, what is worse, an iterative procedure is necessary to calculate the mean of COVs on Riemannian manifold when learning the weak classifier in each iteration of boosting [13].

This paper addresses the efficiency problem of using traditional 2^{nd} order statistics based region descriptors by boosting the Sigma set descriptors, which are recently proposed in [5]. Instead of representing 2^{nd} order statistics directly as COVs, Sigma set encodes the information in an implicit manner by a set of vectors, which owns the same 2^{nd} order statistics as the given image region. The form of Sigma set leads to more efficient between-set metrics of Sigma sets [5], compared with the Riemannian manifold based metrics of COVs [12]. As we will mentioned in Section 2, the distance metric of Sigma set can be implemented by a norm in vector space. This observation implies that adopting linear classifier directly would be appropriate for Sigma set. As a result, we propose to adopt Sigma set in the LogitBoost framework (as illustrated in Section 3) and name this method as *boosted Sigma set* in this paper.

To demonstrate the effectiveness and efficiency of boosted Sigma set, we carry out experiments on the INRIA and the Daimler Chrysler pedestrian datasets in Section 4.

2. Sigma Set Descriptor

Recently, Sigma set region descriptor is developed and successfully applied to texture classification and object tracking in [5]. Sigma set is formed by a set of constructed points, which owns the same 2^{nd} order statistics of the original image region. In particular, suppose the dimensionality of the elemental feature vectors of a given

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.739

image region $\Re$ is d. The Sigma set descriptor S of $\Re$ can be constructed [5] as follows:

1) Calculate the covariance matrix $\mathbf{C}$ of $\Re$ [10].
2) Perform Cholesky decomposition on $\mathbf{C}$ and obtain a lower triangular matrix $\mathbf{L}$.
3) Build S by the columns of $\mathbf{L}$ as (1):

$$S = \{\mathbf{L}_1, \cdots, \mathbf{L}_d\}, \quad (1)$$

where $\mathbf{L}_i$, i=1…,d are the d columns of the lower triangular matrix $\mathbf{L}$. The set, which consists of $\mathbf{L}_i$ forms the Sigma set region descriptor of $\Re$.

Sigma set in (1) is slightly different with that in [5] by removing the redundant d elements with negative signs and the normalization constant. Nevertheless, the differences between that in [5] and ours are trivial in classification.

Since Sigma Set is derived from COV uniquely, Sigma set is discriminative and robust as COV [5]. In addition, in the case of a rectangle region, the calculation of Step 1) of the above-mentioned algorithms can be accelerated through the method of *integral image* [12]. Furthermore, distances between Sigma Sets can be calculated more efficiently, compared with the Riemannian manifold based distances of COVs [12].

This paper utilizes the point restricted distance of Sigma sets defined in [5], as formulated in (2):

$$d(S_A, S_B) = \sum_{i=1}^{d} d_{\mathrm{E}}(\mathbf{L}_i^A, \mathbf{L}_i^B), \quad (2)$$

where $\mathbf{L}_i^A$ and $\mathbf{L}_i^B$, $i = 1, \cdots, d$ are elements from two Sigma sets S_A and S_B respectively. It is worth to notice that when exploiting Manhattan distance as the element-wise distance d_{E}, Equation (2) becomes:

$$d(S_A, S_B) = \|\mathbf{vec}_A - \mathbf{vec}_B\|_1, \quad (3)$$

where $\mathbf{vec}_A$ and $\mathbf{vec}_B$ are the vectors including the $d(d+1)/2$ lower triangular entries in $\mathbf{L}_A$ and $\mathbf{L}_B$ related to S_A and S_B respectively. Specifically, $\mathbf{vec}_A$ of $\mathbf{L}_A$ can be calculated in the following manner:

$$\mathbf{vec}_A = [l_{1,1}, l_{2,1}, l_{2,2}, \cdots, l_{d,1}, \cdots, l_{d,d}]^T, \quad (4)$$

where $l_{i,j}$ is the $(i, j)^{\text{th}}$ entry of $\mathbf{L}_A$. For clarity, we denote the space of $\mathbf{vec}$ as $\mathbf{S}_v$ hereinafter.

Equation (3) implies that Sigma sets lie in vector space and adopting linear classifiers directly on $\mathbf{S}_v$ is a possible way to get acceptable performance. As a result, we utilize linear least square [13] to learn linear weak classifiers on Sigma sets and exploit the cascade LogitBoost framework [13] to learn the strong classifier, as we will describe in the next section.

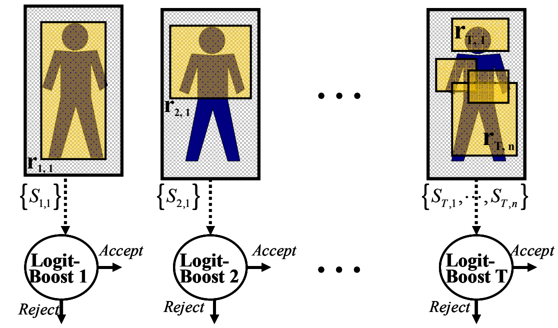


Figure 1. Illustration of a T-stage LogitBoost classifier, where Sigma set $S_{k,i}$ corresponds to an image region $\Re_{k,i}$.

Algorithm. 1. LogitBoost Weak Classifier Learning

Input:
1. Training set: $\{(\mathbf{x}_i, y_i)\}$, i=1,…,N, $\mathbf{x}_i \in \mathbf{S}_v$, $y_i \in \{0,1\}$.
2. Strong classifier $F_{t-1}(\mathbf{x})$ by the previous t-1 iterations.

Output: the updated strong classifier $F_t(\mathbf{x})$.

1. Calculate the probability: $p(\mathbf{x}_i) = e^{F_{t-1}(\mathbf{x}_i)} / \left(e^{F_{t-1}(\mathbf{x}_i)} + e^{-F_{t-1}(\mathbf{x}_i)}\right)$.
2. Calculate the response value and the weight as $z_i = \dfrac{y_i - p(\mathbf{x}_i)}{p(\mathbf{x}_i)(1 - p(\mathbf{x}_i))}$ and $w_i = p(\mathbf{x}_i)(1 - p(\mathbf{x}_i))$.
3. Fit a function $f_t(\mathbf{x})$ of z_i to $\mathbf{x}_i$ by weighted least-square using weights w_i on training data.
4. $F_t(\mathbf{x}) = F_{t-1}(\mathbf{x}) + 0.5 f_t(\mathbf{x})$.

Figure 2. Illustration LogitBoost Weak Classifier Learning.

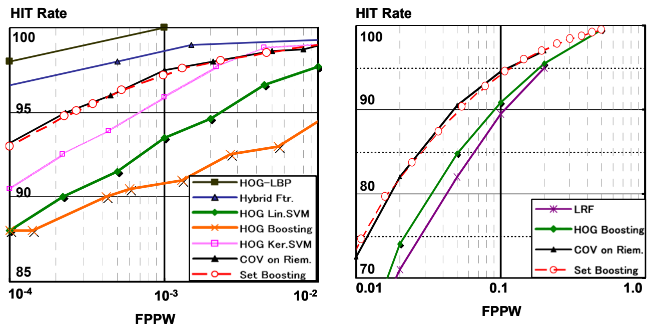
Feature vectors of 8 elementary features, which achieve promising performance in [13] are adopted to calculate Sigma set in this paper:

$$\left[|I_x|, |I_y|, |I_{xx}|, |I_{yy}|, \sqrt{I_x^2 + I_y^2}, \arctan\left(\frac{|I_x|}{|I_y|}\right), x, y\right]^T, \quad (5)$$

where I_x and I_y, and I_{xx} and I_{yy} refer to the 1^{st} order and the 2^{nd} order derivatives at pixel (x, y) respectively.

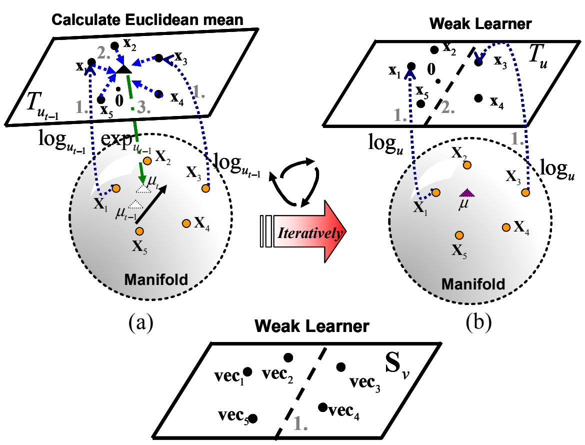
3. Cascade LogitBoost Framework

To learn an effective pedestrian detection classifier, we apply Sigma set descriptor to the cascade LogitBoost framework as the feature extraction module in each round of boosting. The descriptor normalization and feature pool generation are consistent to that in [13]. A T-stage LogitBoost classifier is illustrated in Fig. 1. In addition, Fig. 2 summarizes the weak classifier learning algorithm under LogitBoost. Due to space limitations, readers are referred to [13] for more details of learning the whole cascade of LogitBoost classifier.



(a) (b)

Figure 3. Comparisons between our method with the baseline algorithms on (a) the INRIA dataset and (b) the DC dataset using Hit Rate versus FPPW. See text for more details.



(a) (b)

(c)

Figure 4. Weak learning in the cases of the covariance matrix and Sigma set respectively. (a) One iteration of updating the mean of covariance matrices. (b) Weak learning on the tangent space at the mean of covariance matrices. (c). Weak learning on the space of Sigma sets. See text for more details.

4. Experiments And Discussions

We verify the proposed method on two public datasets using the criterion of patch level test, i.e., Hit Rate of positive samples versus False Positive Per Window (FPPW) of negative samples. By default, experiments are carried out on desktops of Intel Core 2 Duo 3.0 GHz CPU and 4.0G memory.

4.1 Experiments on the INRIA dataset

The *first* one is the INRIA pedestrian dataset [1], which includes 2,416 and 1,132 pedestrian samples for training and testing respectively. INRIA dataset has advanced the research on pedestrian detection greatly, though it is not without shortcomings [2], *e.g.*, the risk of suffering "boundary effect". To avoid it, we use both cropped positive and negative samples, each of them with an additional margin for gradient calculation.

Fig. 3 (a) illustrates the comparisons between boosted Sigma set with some baseline algorithms. Boosted Sigma set achieves nearly the same performance with COVs on Riemannian manifold [13] ("COV on Riem." in Fig. 3). Besides, boosted Sigma set outperforms the HOG based detectors, including HOG using linear SVM and kernel SVM (denoted as "HOG Lin.SVM" and "HOG Ker.SVM" respectively) [1], and the boosted HOG ("HOG Boosting") [18]. This is mainly because that Sigma set as well as COV capture the co-occurrence of more elementary features of a given region, compared with HOG in a comparable dimensionality.

It is easy to understand that the state-of-the-art methods, *e.g.*, HOG-LBP detector [1] ("HOG-LBP") [15] and the hybrid features method ("Hybrid Ftr.") [17], get higher performance than our method due to the usage of multiple features, which always contains more information but in higher dimensionality.

4.2 Experiments on the Daimler Chrysler dataset

The *second dataset* is the Daimler Chrysler (DC) benchmark dataset [9], which consists of three training sets and two test sets, each of them having 4,800 positive and 5,000 negative samples. The small size of samples makes detection (or classification) on this dataset challenging. Since the experiment is carried out at patch level and the number of negative training samples is limited, we do not adopt the cascade structure and only use one stage of LogitBoost. We compare our method with the results reported in [13] in Fig. 3 (b). The results are consistent to that on the INRIA dataset.

In summary, by comparing several baseline methods on the INRIA and DC dataset, the proposed boosted Sigma set method is an effective pedestrian detection method. It has similar performance to that of using COVs on manifold and has more discriminative power than the baseline detectors based on HOG.

4.3 Computational Analysis

The efficiency superiority our method in the training stage, compared with the method of [13] is illustrated in Fig. 4. For COVs, a gradient descent procedure is required to obtain the mean of COVs (One iteration is shown in Fig. 4 (a)). Then a linear (weak) classifier is learnt using all projections of COVs on the tangent space T_{μ} at the mean u, as shown in Fig. 4 (b). It is worth to notice that all steps above are based on the matrix logarithm and exponential operations [13], which are computationally demanding. Instead, for Sigma set, the classifier learning can be performed *directly* in vector space $\mathbf{S}_v$, as shown in Fig. 4 (c). Besides, the superiority in the detection efficiency of our method is derived from the training procedure. For each weak classifier, an additional projection to the tangent space at the learnt mean is required in the case of using

[1] To our best knowledge, the HOG-LBP detector achieves the best performance on the INRIA dataset by far.

COVs, whereas it is not required any more in the case of using Sigma set.

To show a quantitative example, we evaluate the time of descriptor calculation in learning one weak classifier on a collection of 2,416 positive patches and randomly sampled 12,180 negative patches on the INRIA training set. We aggregate the duration for all 200 random blocks, which lasts from the time after preparation of all COVs (required in both cases of using COVs and Sigma sets), to the time just before executing Algorithm 1 in Fig. 2. For Sigma set, the total time is about 1.9 seconds, whereas for COV, in the fastest case with the number of iteration fixed to 1, the total time is about 85 seconds, which is at least about 45 times more than Sigma set's.

4. Conclusion

This paper presents a new pedestrian detection approach utilizing Sigma sets as region descriptors, which can be efficiently trained and evaluated directly in vector space. Experimental results show that our method achieves consistent performance with the traditional covariance matrix on Riemannian manifold and outperforms the baseline HOG based methods on both the INRIA and the Daimler Chrysler pedestrian datasets.

In future work, we will combine Sigma set with other features and compare it with the state-of-the-art detectors.

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]. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Proc. of CVPR, 2005.

[2]. P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: a benchmark. In Proc. of CVPR, 2009.

[3]. P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. IJCV, Vol. 61, No. 1, pp. 55-79, 2005.

[4]. W. Gao, H. Ai and S. Lao. Adaptive contour features in oriented granular space for human detection and segmentation, In Proc. of CVPR, 2009.

[5]. X. Hong, H. Chang, S. Shan, X. Chen, and W. Gao. Sigma set: a small second order statistical region descriptor. In Proc. of CVPR, 2009.

[6]. B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In Proc. of the ECCV Workshop on Statistical Learning in Computer Vision, 2004.

[7]. S. Maji, A. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In Proc. of CVPR, 2008.

[8]. K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In Proc. of ECCV, 2004.

[9]. S. Munder and D. Gavrila. An experimental study on pedestrian classification. TPAMI, Vol. 28, No. 11, pp. 1863-1868, 2006.

[10]. S. Paisitkriangkrai, C. Shen, and J. Zhang. Fast pedestrian detection using a cascade of boosted covariance features. TCSVT, Vol. 18, No. 8, pp. 1140-1151, 2008.

[11]. C. Papageorgiou and T. Poggio. A trainable system for object detection. IJCV, Vol. 38, No. 1, pp. 15-33, 2000.

[12]. O. Tuzel, F. Porikli, and P. Meer. Region covariance: a fast descriptor for detection and classification. In Proc. of ECCV, 2006.

[13]. O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on Riemannian Manifolds. TPAMI, Vol. 30, No. 10, pp. 1713-1727, 2008.

[14]. P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In Proc. of CVPR, 2003.

[15]. X. Wang, X. Han, and S. Yan. An HOG-LBP human detector with partial occlusion handling. In Proc. of ICCV, 2009.

[16]. B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by Bayesian combination of Edgelet part detectors. In Proc. of ICCV, 2005.

[17]. B. Wu and R. Nevatia. Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection. In Proc. of CVPR, 2008.

[18]. Q. Zhu, S. Avidan, M. Yeh, and K. Cheng. Fast human detection using a cascade of histograms of oriented gradients. In Proc. of CVPR, 2006.