IEICE TRANS. INF. & SYST., VOL.E93–D, NO.9 SEPTEMBER 2010

LETTER

Feature Interaction Descriptor for Pedestrian Detection

Hui CAO[†a)], ***Nonmember***, **Koichiro YAMAGUCHI**[†], ***Member***, **Mitsuhiko OHTA**[†], ***Nonmember***, **Takashi NAITO**[†], ***and*** **Yoshiki NINOMIYA**[††], ***Members***

SUMMARY We propose a novel representation called Feature Interaction Descriptor (FIND) to capture high-level properties of object appearance by computing pairwise interactions of adjacent region-level features. In order to deal with pedestrian detection task, we employ localized oriented gradient histograms as region-level features and measure interactions between adjacent histogram elements with a suitable histogram-similarity function. The experimental results show that our descriptor improves upon HOG significantly and outperforms related high-level features such as GLAC and CoHOG.

key words: *pedestrian detection, feature interaction, HOG, co-occurrence*

1. Introduction

Pedestrian detection is one of the most active research topics in the field of computer vision. Generally, pedestrian detection is considered a classification problem, and a classifier is trained to distinguish the pedestrian from the background. Feature vectors are extracted from image patches containing pedestrians, and a classifier, e.g. linear SVM or Adaboost, is trained to separate these feature vectors from those extracted from background images.

A wide variety of image representations for pedestrian detection have been proposed in the literature [1]–[5]. Recently, Dalal and Triggs' Histograms of Oriented Gradients (HOG) [1] have received increasingly widespread attention. The HOG descriptor and its variants [6], [7], [12] have shown excellent performances on many pedestrian datasets. Unfortunately, the HOG descriptor does not capture high-level properties of object appearance, and it would be incapable of dealing with non-pedestrians of pedestrian-like edge gradients statistics.

In this paper, we propose a general high-level representation, called Feature Interaction Descriptor (FIND), which computes interaction informations between adjacent region-level features. In order to deal with pedestrian detection task, we employ oriented gradient histograms (inspired by HOG) and measure interactions of adjacent histogram elements with a suitable histogram-similarity function. The resulting descriptor inherits the advantages of HOG, such as invariance to geometric and photometric transformations, and additionally it somewhat captures high-level properties of object appearance. Our experimental results demonstrate that the proposed descriptor improves upon HOG significantly and outperforms other related high-level features.

The rest of paper is organized as follows: Sect. 2 reviews current image representations used in pedestrian detection; Sect. 3 presents the implementation detail of the Feature Interaction Descriptor; experimental results are shown in Sect. 4 and Sect. 5 concludes the paper with some final remarks.

2. Related Works

There is an extensive literature on human detection, and here we focus on reviewing some image representations that have been recently applied to human detection.

In the seminal work of Viola and Jones [2], the authors proposed a rapid object detection framework based on Haar wavelet features and an Adaboost classification cascade. Since then many other features have been integrated into the cascade Adaboost framework to improve performance, including HOG [4], [7], Edgelet [3], Covariance [5] and mixing of these features [8], [9].

The methods based on above image representations can be further improved by considering the articulated parts structure of human body. Papageorgiou and Poggio learned a polynomial SVM classifier using Haar wavelets per body part and integrated individual classification scores using a second-stage SVM [10]. Shashua et al. learned a set of linear classifiers using HOG-like gradient features for different portions of human body and integrated these classification scores using a second-stage Adaboost [11]. More recently, Felzenszwalb et al. [12] proposed a multi-scale and deformable SVM-based part model using HOG features which exhibited best performance on the PASCAL VOC dataset.

Gradient Local Auto-Correlations (GLAC) [13] and Co-occurrence Histograms of Oriented Gradients (CoHOG) [14] are two image representations which intend to capture high-level properties of object appearance. They share similar implementation technique which counts co-occurrences of gradient orientations for various patterns of adjacent pixels, such as up-down,left-right, etc.. Different from their implementations based on co-occurrences of pixel-level features, the descriptor proposed in this paper

Manuscript received April 2, 2010.
Manuscript revised May 11, 2010.

[†]The authors are with the Road Environment Recognition Laboratory, Safety & Information Systems Division, Toyota Central R&D LABS., INC., Aichi-ken, 480–1192 Japan.
[††]The author is with the Safety & Information Systems Division, Toyota Central R&D LABS., INC., Aichi-ken, 480–1192 Japan.
a) E-mail: caohui@mosk.tytlabs.co.jp

DOI: 10.1587/transinf.E93.D.2656

measures interactions of region-level features.

3. Feature Interaction Descriptor

This section describes the implementation detail of the proposed Feature Interaction Descriptor. The FIND is a general high-level image representation which computes pairwise interactions of adjacent region-level features. The format of region-level feature and the metric of interaction can vary to suit different tasks. For the pedestrian detection task we are interested in, localized oriented gradient histograms are employed as region-level features and a suitable histogram-similarity function is used to measure interactions between adjacent histogram elements. Therefore, computing FIND involves two key components: (1) constructing localized oriented gradient histograms; (2) measuring interactions of adjacent histogram elements. The detail of computation is described step by step as follows.

3.1 Localized Oriented Gradient Histograms

The first step is the computation of the edge gradients from an image by 1-D derivatives or 2-D derivatives like sobel masks. The second step is counting occurrences of gradient orientation in localized portions of the image. Specifically, the image is divided into spatial grids, called cells (following the naming rules of HOG in [1]). For each cell, a gradient orientation histogram is constructed by casting, for each pixel belonging to the cell, a gradient-magnitude weighted vote onto a histogram bin according to the gradient orientation. The orientations of histogram are evenly spaced over $0 \sim 180^\circ$ or $0 \sim 360^\circ$ depending on whether the signs of gradient are informative or not.

3.2 Interactions of Histogram Elements

After oriented gradient histograms are computed, the third step is measuring interactions of adjacent histogram elements with a histogram-similarity function. We compute interactions within a larger region, called block, which contains spatially adjacent cells. Pairwise interactions are computed for all possible combinations of histogram elements belonging to member cells (illustrated in Fig. 1).

Denote for a block the concatenation of oriented gradient histograms as $[h_1, \ldots, h_m]$, the feature interaction vector is then represented by

$$[f(h_1, h_1), \ldots, f(h_1, h_m), f(h_2, h_2), \ldots, f(h_m, h_m)],$$

where many redundant pairs are discarded due to using symmetric histogram-similarity function, i.e. $f(h_i, h_j) = f(h_j, h_i)$. Three histogram-similarity functions are considered in this work, including: (1) *harmonic mean*, (2) *min*, (3) *product*. Finally, block normalization is applied to the feature interaction vector.

The FIND is the concatenation of feature interaction vectors obtained on all blocks.

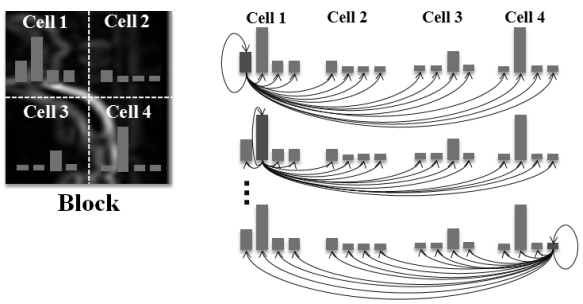

Fig. 1 Interactions of histogram-elements within a block.

4. Experimental Validation

The proposed descriptor is evaluated on Daimler-Chrysler [15] and Near-Infrared pedestrian datasets. The Daimler-Chrysler dataset is a public benchmark dataset which has been widely used for performance comparison. The Near-Infrared dataset consists of pedestrian and non-pedestrian images which were collected at night by us with a near-infrared camera. Table 1 shows the details of two datasets. Some samples are shown in Fig. 2.

The proposed descriptor is compared to three related descriptors introduced in previous sections, including HOG, GLAC and CoHOG. All the experiments used the following parameter setting:(1) roberts gradient filter; (2) 8 orientation bins in $0 \sim 360$ degrees;(3) L2-norm block normalization; (4) block containing 2×2 cells for HOG and FIND; (5) non-overlapping and overlapping block schemes, excepting CoHOG for which only the non-overlapping scheme was considered due to memory restriction.

The Daimler-Chrysler dataset has five disjoint sets: three for training and two for testing. Each subset consists of 4,800 pedestrian examples and 5,000 non-pedestrian examples. Any two out of three training subsets were trained using linear SVM [16] and the remaining training subset was used for parameter tuning. Applying three trained SVM classifiers to two test subsets generated six ROC curves from which the mean ROC curve was computed.

The Near-Infrared dataset has one training set and one

Table 1 The specification of pedestrian datasets.

Dataset	Daimler-Chrysler	Near-Infrared
Training data	14,400 pedestrian 15,000 non-ped.	5,000 pedestrian 5,000 non-ped.
Test data	9,600 pedestrian 10,000 non-ped.	5,000 pedestrian 5,000 non-ped.
Image size	18×36 pixels	30×60 pixels

Fig. 2 Left six columns are samples of Daimler-Chrysler dataset and the right six columns are samples of our Near-Infrared dataset (Top: pedestrian, Bottom: non-pedestrian).

test set, each consisting of 5000 pedestrian and 5000 non-pedestrian examples. The training subset was trained using linear SVM and the parameter of SVM was selected by cross validation. Applying the trained SVM classifier to test subset generated a ROC curve.

The results on two datasets are illustrated in Fig. 3, in which the mean ROC curves for Daimler-Chrysler dataset is shown on the left and the ROC curves for Near-Infrared dataset shown on the the right. It can be seen that FIND (using harmonic mean function) improves HOG significantly and outperforms GLAC and CoHOG on both overlapping and non-overlapping cases in terms of overall performances. To the best of our knowledge, the accuracy obtained by FIND on Daimler-Chrysler dataset is the best one that has ever been published to date.

Since the low false positive region is more interesting for detection task, ROC statistics at low false-positive-rates

Table 2 Dimensions of four descriptors in experiment.

	FIND	HOG	GLAC	CoHOG
Non-overlapping	9,404	576	4,608	34,704
Overlapping	29,040	1,760	14,080	106,040

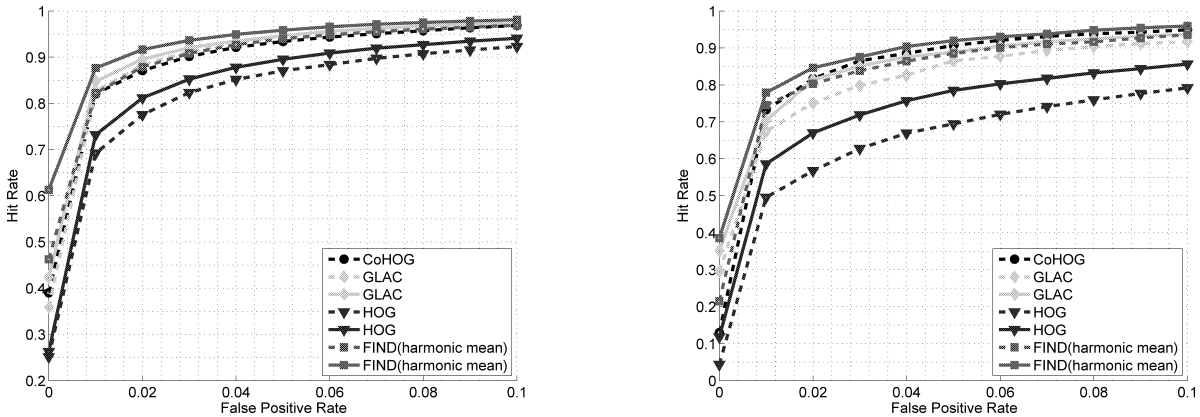

Fig. 3 ROC curves of different descriptors on Daimler-Chrysler (Left) and Near-Infrared (Right) datasets. Solid lines and dashed lines indicate results obtained under overlapping and non-overlapping block schemes, respectively.

Table 3 ROC statistics of different descriptors on Daimler-Chrysler and Near-Infrared datasets. Each column lists true-positive-rates obtained under overlapping (non-overlapping) block schemes when the false-positive-rate is 0, 0.01 and 0.05.

Dataset	Daimler-Chrysler			Near-Infrared		
ROC statistics	tp@fp=0	tp@fp=.01	tp@fp=.05	tp@fp=0	tp@fp=.01	tp@fp=.05
COHOG	(.390)	(.821)	(.934)	(.130)	(.732)	(.906)
GLAC	.423(.359)	.847(.827)	.946(.944)	.351(.295)	.703(.675)	.889(.865)
HOG	.262(.251)	.732(.692)	.895(.870)	.118(.044)	.586(.495)	.785(.694)
FIND (harmonic mean)	.613(.463)	.877(.823)	.958(.940)	.386(.215)	.779(.745)	.920(.885)

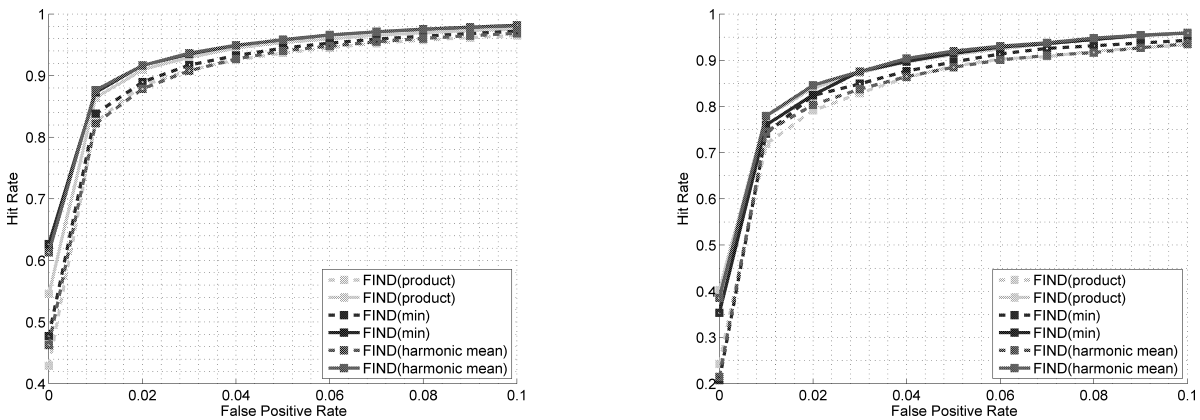

Fig. 4 Performance comparison of the proposed descriptor with different interaction measures on Daimler-Chrysler (Left) and Near-Infrared (Right) datasets. Solid lines and dashed lines indicate results obtained under overlapping and non-overlapping block schemes, respectively.

are listed in Table 3. The FIND exhibits evident advantage at low false-positive-rates and particularly when false-positive-rate is 0.

We next investigated FIND's performances with respect to different forms of measuring interaction. The performance comparison is shown on Fig. 4. Overall, the harmonic-mean function seems to be superior to min and product functions. It suggests the necessity of choosing a suitable interaction measure, which will be further investigated in following works.

Table 2 lists the dimension of four descriptors in experiment. The dimension of FIND is roughly 16 times larger than that of HOG, twice larger than that of GLAC and less than one third of that of CoHOG.

5. Conclusion

The high-level properties of object appearance are important cues for discriminating them from other objects. In this work we present a Feature Interaction Descriptor (FIND) to describe high-level properties by interactions of adjacent region-level features. This descriptor can inherit the advantages of region-level features, and moreover, it somewhat captures high-level properties by computing pairwise feature interactions. To apply this representation to pedestrian detection, we utilize the oriented gradient histograms as region-level features and their interactions are measured with a histogram-similarity function. Experiments on two pedestrian datasets have proven that the proposed descriptor significantly improves HOG and outperforms related high-level features.

We plan to apply this high-level representation to other applications like scene classification, which may require us to employ other interaction measures to suit different tasks.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," CVPR, pp.886–893, 2005.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," CVPR, 2001.

[3] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," ICCV, 2005.

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

[5] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," CVPR, 2007.

[6] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," CIVR, 2007.

[7] I. Laptev, "Improving object detection with boosted histograms," Image Vis. Comput., vol.27, no.5, pp.535–544, 2009.

[8] D. Geronimo, A. Lopez, D. Ponsa, and A.D. Sappa, "Haar wavelets and edge orientation histograms for on-board pedestrian detection," Proc. Iberian Conference on Pattern Recognition and Image Analysis, 2007.

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

[10] C. Papageorgiou and T. Poggio, "A trainable system for object detection," IJCV, vol.38, no.1, 15.33, 2000.

[11] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: Single-frame classification and system level performance," IEEE Intelligent Vehicles Symposium, 2004.

[12] P. Felzenszwalb, D. McAllester and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," CVPR, 2008.

[13] T. Kobayashi and N. Otsu, "Image feature extraction using gradient local auto-correlations," ECCV, pp.346–358, Oct. 2008.

[14] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection," Proc. 3rd Pacific Rim Symposium on Advances in Image and Video Technology, pp.37–47, 2009.

[15] S. Munder and D.M. Gavrila, "An experimental study on pedestrian classification," IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.11, pp.1863–1868, 2006.

[16] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," J. Machine Learning Research, vol.9, pp.1871–1874, 2008.