More on Weak Feature: Self-correlate Histogram Distances

Sheng Wang, Qiang Wu, Xiangjian He, and Wenjing Jia

Research Centre for Innovation in IT Services and Applications (iNEXT)
University of Technology, Sydney, Broadway 2007, Australia

Abstract. In object detection research, there is a discussion on weak feature and strong feature, feature descriptors, regardless of being considered as 'weak feature descriptors' or 'strong feature descriptors' does not necessarily imply detector performance unless combined with relevant classification algorithms. Since 2001, main stream object detection research projects have been following the Viola Jone's weak feature (Haar-like feature) and AdaBoost classifier approach. Until 2005, when Dalal and Triggs have created the approach of a strong feature (Histogram of Oriented Gradient) and Support Vector Machine (SVM) framework for human detection.

This paper proposes an approach to improve the salience of a weak feature descriptor by using intra-feature correlation. Although the intensity histogram distance feature known as Histogram Distance of Haar Regions (HDHR) itself is considered as a weak feature and can only be used to construct a weak learner to learn an AdaBoost classifier. In our paper, we explore the pairwise correlations between each and every histograms constructed and a strong feature can then be formulated. With the newly constructed strong feature based on histogram distances, a SVM classifier can be trained and later used for classification tasks. Promising experimental results have been obtained.

Keywords: Weak feature, Pairwise correlations, Histogram distances, SVM classifier.

1 Introduction

In computer vision research, it is widely recognized that good features are crucial for object detection tasks, there is abundant literature introducing state-of-the-art feature extraction algorithms [1][2][3]. Another research direction is the introduction of new object detection frameworks or improved feature extraction algorithm(s) [4][5]. In this paper, in addition to proposing a new feature based on correlate histograms, we are more interested in introducing a way to extract more information from an existing weak feature, we use the Histogram Distance of Haar Regions (HDHR) feature as an example.

In [4], the authors proposed the HDHR feature, the HDHR feature is defined as the intensity histogram distance between two adjacent Haar regions. Comparing with the simple Haar-like feature used by [6], the HDHR feature contains more

Y.-S. Ho (Ed.): PSIVT 2011, Part I, LNCS 7087, pp. 214–223, 2011.

information (hence should be able to better distinguish positive samples from negative samples) and can be calculated efficiently with the Integral Histogram framework proposed in [7][8]. An AdaBoost classifier is used in [4] to perform the object detection task of separating image regions that contains airplane from those regions which do not contain airplane.

In [9], the authors introduced the Shape Context feature descriptor, the shape context feature extraction algorithm is composed of three steps, the first step is to extract sample points from the edge map of the input image; the second step is to calculate the distance and orientation difference between the current sample point and every other sample point; the third step is to quantize those distances and orientation differences in to predefined number of bins. [9] is an early approach of feature extraction algorithms which are based on measuring object similarities with regard to certain distance metrics.

In [1], the authors introduced an approach to measure similarities between objects with a local descriptor, the descriptor is called Local Self-similarities (LSS). The LSS is based on matching internal self-similarities. That is, only the internal layout is correlated across images (or video sequences). Because the attributes for visual tasks (color, texture and illumination) within an image is relatively uniform compared to that of other images, exploring internal self-similarities can better capture the pattern of the visual entity. The LSS feature extraction process can be regarded as two steps, the first step is calculating correlation surface, this step is achieved by matching a smaller image patch from an image with a larger image region within the same image; the second step is translating the correlation surface into*a binned log-polar representation*, this step is similar to the final step of the Shape Context feature extraction. In [1], the CIE L*a*b space is used instead of the RGB color space to calculated the Sum of Squared Distances (SSDs) between patch colors. The LSS is a state-of-the-art feature descriptor based on self-similarity.

In [2], the authors introduced a new feature termed as Color Self-similarity (CSS), the CSS is based on the observation that objects such as a human do exhibit some structure in which colors are locally similar (*e.g.* the skin color of a specific person is similar on their two arms and face). In CSS, a positive sample (*i.e.* sample images which tightly bounds the object of interest) is first labeled with different semantic patches, such as arms, legs, upper body and background, then each semantic patch (of size 8×8 pixels) is used to measure the color similarity between the patch and the whole sample, the authors used HSV color space because it works best compared to RGB, HLS, CIE Luv, and etc. Each semantic patch will generate a similarity sample, in such similarity samples, the homogeneous region (for its corresponding similarity patch) will have a higher similarity score. Self-similarities between those similarity samples are then explored and utilized to construct a SVM classifier. In [2], the CSS is integrated with other features for object detection. It is one of the latest object detection approach using self-similarity measurement.

Motivated by the self-similarity feature being introduced in [1] and [2]. We propose a method that is capable of bring significant improvement over the

saliency of the original weak feature such that a SVM classifier can be used to substitute the original AdaBoost classifier. Our feature extraction algorithm is composed of three steps, sub blocking, histogram binning, and correlating. Details will be given in Section 2.

Our contributions in this paper can be summarized as follows.

Firstly, by exploring its self-correlation, we transform a weak feature (HDHR) into a strong feature, we term it Correlation based Histogram Distance (COHD), this transformation is similar to the self-similarity features being proposed in [1]. As a strong feature, COHD enables the use of a SVM classifier for object detection, this saves a lot of time in comparison with having to train an AdaBoost classifier for the original weak HDHR feature.

Secondly, the newly proposed self-correlation feature based on histogram distances can be quickly calculated with the method proposed in [7], this is a precious computational advantage.

Thirdly, different from [1], which explores self-similarities from raw image level, we seek self-correlations from feature descriptor (*i.e.* Intensity Histogram) level, this can greatly reduce the computational cost and still well preserve the feature saliency.

The rest of this paper will be organized as follows, Section 2 introduces the formulation of a strong feature, we follow a typical object detection framework by replacing the original feature with the newly formed feature. Section 3 gives experimental results. Section 4 concludes this paper.

2 Weak Feature and Self-correlations

In this section, we will first introduce two types of weak feature, they are Density Variance feature and Histogram Distance of Haar Regions (HDHR) feature (neither of them can be directly combined with a SVM classifier for object detection task due to their weak saliency), then we introduce our proposed correlation feature derived from those two features mentioned above.

The Density Variance Feature was introduced in [5], such feature can be represented by

$$V_G = \frac{\sum_{i=1}^{n} |G_i - G|}{n \cdot G} \tag{1}$$

In (1), i is the index for the sub blocks as illustrated in Fig. 1, G is defined as the mean value of the gradient strength for the whole sample, and G_i is the mean value of the gradient strength for sub block i, n is the total number of sub blocks in a sample. In [5], the Density Variance feature was simply used as a global statistical filter to speed up the detection process for a license plate detector.

The Histogram of Haar Regions (HDHR) feature was first proposed in [4], the HDHR feature was introduced because of two reasons. Firstly, in order to differentiate two adjacent regions in a more suitable way, histograms provides more detailed information than classical Haar features. Secondly, Histograms can

be computed linearly, which is a precious computational advantage. The HDHR feature descriptor is represented by

$$D(f,g) = \frac{\sum_{j=1}^{N} (f[j] - g[j])^2}{\sum_{j=1}^{N} (f^2[j] + g^2[j])} \tag{2}$$

In (2), D is defined as the Distance between the histogram $f[\cdot]$ and histogram $g[\cdot]$, as $f[\cdot]$ and $g[\cdot]$ each corresponding to a histogram constructed from image regions f and g, respectively. The number of bins in $f[\cdot]$ equals to the number of bins in $g[\cdot]$ and both equal to N, hence the distance calculation is a division of two summations over the bin index j. In [4], the HDHR feature was used together with AdaBoost supervised learning algorithm for airplane detection.

As mentioned in Section 1, our feature extraction method is composed of sub blocking, histogram binning, and correlating. Our sub blocking method was motivated by [5], our histogram binning method was motivated by [4], and motivated by [2], we use correlating to increase the feature salience.

In our approach, instead of considering the distance between two adjacent Haar-like Regions, we divide the sample image region into sub blocks of $p \times q$, in each sub block, a histogram can be constructed, hence the total number of histograms can be used to calculate D is $p \cdot q$. Given $p \cdot q$ histograms, we will consider the pairwise correlation between each pair of histograms, hence the total number of histogram distances can be measured is represented by

$$C_{p \cdot q}^2 = \frac{(p \cdot q) \times (p \cdot q - 1)}{2} \tag{3}$$

Finally, the Correlation based Histogram Distance feature, we term it Correlation Histogram Distance (COHD) feature descriptor is represented by

$$\mathbf{S}_D = \{D(f,g)\} \tag{4}$$

which is a vector of length $C_{p \cdot q}^2$.

With COHD feature, an object detection framework can be easily constructed by train a Support Vector Machine (SVM) Classifier.

Moreover, we propose two variants based on different normalization schemes, the $L1 - norm$ for COHD feature is represented by

$$E_1(f,g) = \sum_{j=1}^{N} |f[j] - g[j]| \tag{5}$$

The corresponding $L2 - norm$ is represented by

$$E_2(f,g) = \sqrt{\sum_{j=1}^{N} (f[j] - g[j])^2} \tag{6}$$

In (5) and (6), the definitions for $f[\cdot]$, $g[\cdot]$, j and N are the same as those of (2).

By substitute D with E_1, the COHD $L1-norm$ feature descriptor is represented by

$$\mathbf{S}_{E_1} = \{E_1(f,g)\} \tag{7}$$

Similarly, the COHD $L2-norm$ feature descriptor is represented by

$$\mathbf{S}_{E_2} = \{E_2(f,g)\} \tag{8}$$

Details of the Correlation of Histogram Distance features (*i.e.* S_D, S_{E_1}, and S_{E_2}) are illustrated in Fig. 1. In Fig. 1, f corresponding to the sub block from where histogram $f[\cdot]$ is constructed, and g corresponding to the sub block from where histogram $g[\cdot]$ is constructed.

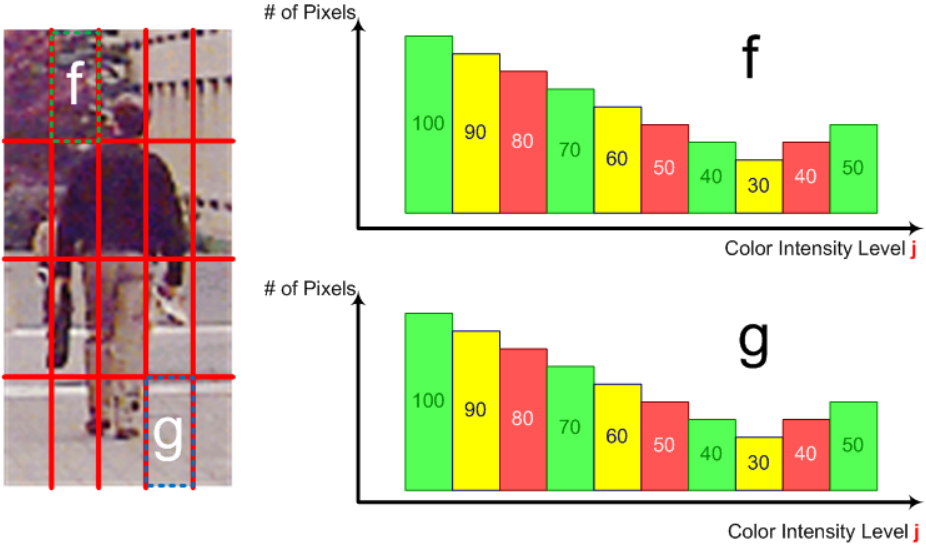


Fig. 1. Extracting Correlation of Histogram Distance features

The input image is first divided into $p \cdot q$ sub blocks, for each sub block f, a histogram $f[\cdot]$ can be obtained, $f[\cdot]$ is then compared with another histogram $g[\cdot]$ resulted from region g. The distance between $f[\cdot]$ and $g[\cdot]$ is one dimension of the $C^2_{p\cdot q}$-Dimensional feature vector.

3 Experimental Results

As one of the most representative strong feature, Histogram of Oriented Gradient (HOG) has attracted numerous attention of various researchers. As a result, we compare the descriptive power of HOG with our newly proposed correlation feature by replacing the HOG feature within the HOG and SVM framework with the correlation feature [10].

We use the MIT CBCL Dataset for our experiments, in particular, we evaluate the performance of the framework using Face, Human and Car [11][12][13]. The MIT CBCL Dataset is composed of four types of Data, they are, face, human, car, and scenario. More details of the Dataset can be found from Table 1.

Table 1. Details of the MIT CBCL Dataset

	Face	Human	Car
# of Positive Training Samples	2429	924	516
# of Negative Training Samples	4548	-	-
# of Positive Testing Samples	472	-	-
# of Negative Testing Samples	23573	-	-
Sample Size(Width×Height)	19×19	64×128	128×128

Some of the examples being used in our experiments can be found from Fig. 2.

(a) Face (b) Human (c) Car

Fig. 2. Some Examples from MIT CBCL Dataset

Detailed parameter settings can be found from Table 2.

As mentioned in [2], block normalization proven to be crucial, we use the same normalization scheme as provided in the MATLAB implementation of HOG and SVM framework by [10] to normalize the COHD feature descriptor.

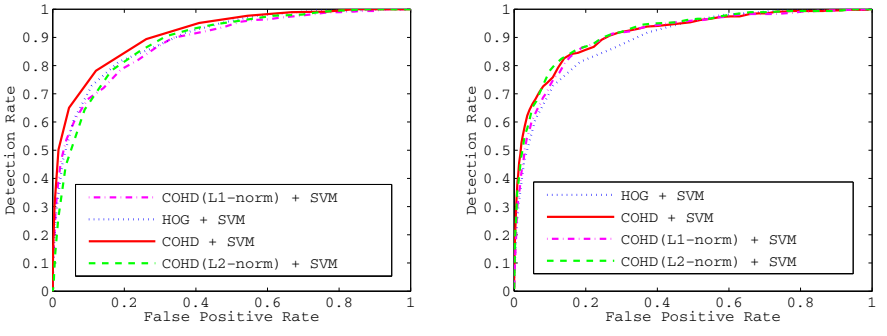
A quantitative measure of the experimental results can be observed from Fig. 3. From Fig. 3, we can see that before sub block normalization, the newly proposed correlation feature based on HDHR can out perform HOG by approximately 4% on the MIT CBCL Face Dataset. However, the HOG feature remains extremely competitive on the MIT CBCL Human Dataset and MIT CBCL Car Dataset. Those results can be observed from Fig. 4 and Fig. 5, respectively. Yet our newly proposed feature (COHD with *L1-norm*) can achieve a detection rate of 97% at a false positive rate of approximately 2% on the Human dataset and 90% detection rate at 2% false positive rate on the Car dataset.

As we can see from Fig. 3b, Fig. 4b, and Fig. 5b, normalization can significantly improve the experimental results. The ROC curve for the human dataset

Table 2. Detailed parameter settings in our experiments[1]

	Face	Human	Car
# of Sub blocks (W×H)[2]	3 × 3	5 × 4	5 × 5
Scaled sample size (Width×Height)	19 × 19	32 × 64	32 × 32
# of Bins for COHD	32	32	32
# of Bins for COHD(*L1*)	32	32	32
# of Bins for COHD(*L2*)	32	32	32
# of Bins for HOG	9	9	9
# of Training Positive	2429	800	400
# of Training Negative	4548	1600	881
# of Testing Positive	472	124	116
# of Testing Negative	23573	195	160

[1]The negative samples for Human and Car Dataset was randomly cropped from background images which contains neither human nor car.
[2]W: the number of sub blocks in each row, H: the number of sub blocks in each column.

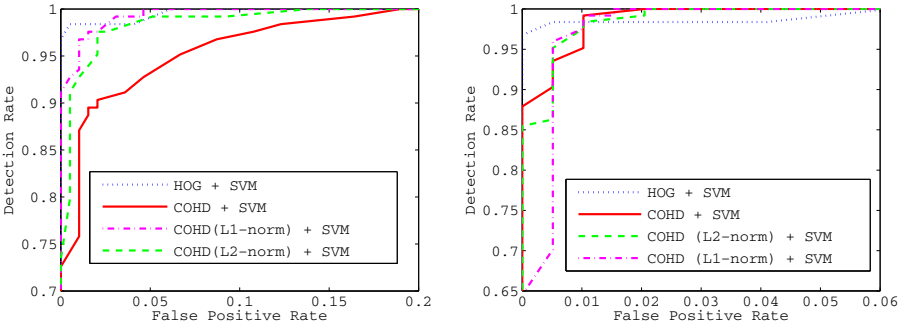


(a) Before Sub block Normalization (b) After Sub block Normalization

Fig. 3. ROC Curves on MIT CBCL Face Dataset

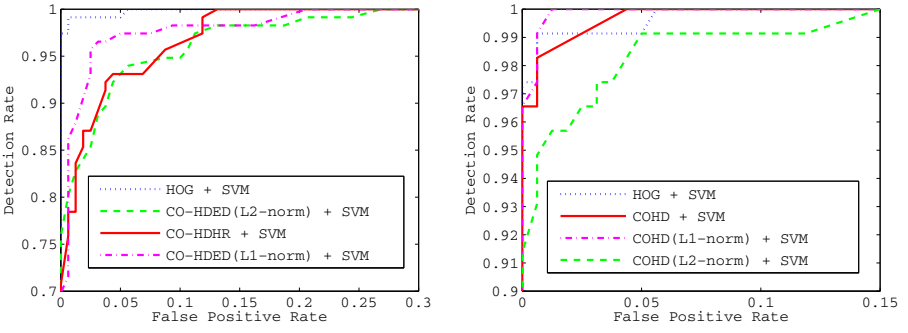
and car dataset is more rough than than that of the face dataset due to a smaller number of testing samples.

Those experimental results indicate that by exploring self-correlations, an original weak feature can be significantly improved to a strong feature, this approach of exploring intra-feature self-correlations is similar to the self-similarity features being proposed in [1][2][3], the difference is that the self-correlation is extracted from the feature descriptor level instead of the raw image data level, hence there will be some information loss to degrade the quality of the feature descriptor, but the computational complexity is also greatly reduced compared to that of self-similarity features and there is always a weight of balance between the computational cost and performance gain.



(a) Before Sub block Normalization (b) After Sub block Normalization

Fig. 4. ROC Curves on MIT CBCL Human Dataset



(a) Before Sub block Normalization (b) After Sub block Normalization

Fig. 5. ROC Curves on MIT CBCL Car Dataset

The proposed correlation method does not require matrix convolution during the feature extraction process, comparing with HOG, which needs gradient magnitude computation and arc tangent computation, the feature extraction process is much simpler. Although to extract Haar-like feature is also very simple, the computational cost (especially time complexity) for AdaBoost training is very high, this computational advantage is especially important for devices with limited computational power, such as wireless sensors.

The computational cost (in terms of time complexity) to measure a pairwise histogram distances for a detection window that is partitioned into k sub windows is $\frac{k\times(k-1)}{2}$. Without using Integral Histogram, the computational cost needed to calculate the histogram feature of a detection window of size $n \times n$ is $O(n^2)$, the Integral Histogram can reduce this cost to $O(1)$. As reported by [3], to calculate the LSS descriptor for one pixel with patch size $\omega \times \omega$ and block size $N \times N$ requires $N^2\omega^2$ operations, the authors for [3] also mentioned that although Fast Fourier Transform(FFT) can speed up the process with $3N^2 log N^2 + N^2$ operations, the speed up is marginal as $N > \omega$.

In our experiments, we also compared the execution speed of the COHD feature extraction algorithms with that of the HOG feature extraction algorithm on the same platform, details are listed in Table 3. For the implementation of HOG feature extraction, we use the code provided by [10]. Depending on each particular sample, the speed of feature extraction varies, hence we compare the total time needed to convert the entire training dataset to corresponding feature descriptors. Details about each dataset is given in Table 2. Using Matlab 2009b with a Windows XP(32bit) environment, on a computer with 3.16GHz CPU and 3.25GB of RAM, we obtained the results in Table 3.

Table 3. Speed Comparison for Feature Extraction

	Face	Human	Car
HOG [10]	16.42 seconds	10.88 seconds	11.70 seconds
COHD	11.98 seconds	7.85 seconds	6.28 seconds
COHD(*L1*)	11.99 seconds	7.84 seconds	6.28 seconds
COHD(*L2*)	12.04 seconds	7.98 seconds	6.38 seconds

4 Conclusion

In this paper, we have proposed a self-correlation method to improve the saliency of a weak feature, by dividing the detection window into sub blocks, we have proposed three different normalization schemes for self-correlated features derived from intensity histograms. The experimental results on MIT CBCL Dataset proved that those self-correlated features can dramatically increase the feature saliency. In particular, for MIT CBCL Face Dataset, the self-correlated feature outperform one classical strong feature object detection framework. However, this method is not limited to one particular type of feature, other weak features can be enhanced by this self-correlation method as well.

References

1. Shechtman, E., Irani, M.: Matching Local Self-similarities across Images and Videos. In: Proc. CVPR, Minneapolis, pp. 1–8 (2007)
2. Walk, S., Majer, N., Schindler, K., Schiele, B.: New Features and Insights for Pedestrian Detection. In: Proc. CVPR, San Francisco, pp. 1030–1037 (2010)
3. Deselaers, T., Ferrari, V.: Global and Efficient Self-similarity for Object Classification and Detection. In: Proc. CVPR, San Francisco, pp. 1633–1640 (2010)
4. Perrotton, X., Sturzel, M., Roux, M.: Automatic Object Detection on Aerial Images Using Local Descriptors and Image Synthesis. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 302–311. Springer, Heidelberg (2008)
5. Zhang, H., Jia, W., He, X., Wu, Q.: Learning-Based License Plate Detection Using Global and Local Features. In: Proc. ICPR, Hong Kong, pp. 1102–1105 (2006)

6. Viola, P., Jones, M.: Rapid Object Detection Using A Boosted Cascade of Simple Features. In: Proc. CVPR, Kauai, pp. 511–518 (2001)
7. Porikli, F.: Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces. In: Proc. CVPR, San Diego, pp. 829–836 (2005)
8. Kovesi, P.: University of Western, Australia, http://www.csse.uwa.edu.au/
9. Belongie, S., Malik, J.: Matching with Shape Contexts. In: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, Hilton Head, pp. 20–26 (2000)
10. Ludwig, O., Delgado, D., Goncalves, V., Nunes, U.: Trainable Classifier-Fusion Schemes: An Application to Pedestrian Detection. In: Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, pp. 432–437 (2009)
11. Weyrauch, B., Huang, J., Heisele, B., Blanz, V.: Component-based Face Recognition with 3D Morphable Models. In: Proceedings of the First IEEE Workshop on Face Processing in Video, Washington, D.C, pp. 85–89 (2004)
12. Papageorgiou, C., Evgeniou, T., Poggio, T.: A Trainable Pedestrian Detection System. In: Proceedings of the IEEE International Conference on Intelligent Vehicles, Stuttgart, pp. 241–246 (1998)
13. Oren, M., Papageorgiou, C.P., Sinha, P., Osuna, E., Poggio, T.: Pedestrian Detection Using Wavelet Templates. In: Proc. CVPR, San Juan, pp. 193–199 (1997)
14. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: Proc. CVPR, San Diego, pp. 886–893 (2005)