LPP-HOG: A New Local Image Descriptor for Fast Human Detection

Qing Jun Wang
Department of computer science and technology
Harbin Engineering University
Harbin, China
Qingjunwangfast@163.com

Ru Bo Zhang
Department of computer science and technology
Harbin Engineering University
Harbin, China
zrbzrb@hrbeu.edu.cn

***Abstract*—LPP(Locality Preserving Projection), as a linear version of manifold learning algorithm, has attracted considerable interests in recent years. For real time applications, the response time is required to be as short as possible. In this paper, a new Local image descriptor—LPP-HOG (Histograms of Oriented Gradients) for fast human detection is presented. We employ HOG features extracted from all locations of a grid on the image as candidates of the feature vectors. LPP is applied to these HOG feature vectors to obtain the low dimensional LPP-HOG vectors. The selected LPP-HOG feature vectors are used as an input of linear SVM to classify the given input into pedestrian/non-pedestrian. We also present results showing that using these descriptors in human detection application results in increased accuracy and faster matching.**

Keywords-manifold learning; LPP;HOG;human detection

I. INTRODUCTION

Detecting humans in still images is a challenging task owing to their variable appearance and the wide range of poses that they can adopt. The first need is a robust feature set that allows the human form to be discriminated cleanly, even in cluttered backgrounds under difficult illumination.

The proposed descriptors are reminiscent of SIFT descriptors [1], shape contexts [2], HOG descriptors [3] [4] [5]. Dalal [3] extracted the HOG features from all locations of a dense grid on a image region and the combined features are classified by using linear SVM. They showed that the grids of HOG descriptors significantly outer-performed existing feature sets for human detection. The second need is an effective method for real time human detection. We should extract a set of feature vectors from all locations in an image grid and are used for classification. The total number of features becomes over ten thousands when the features extracted from all locations on the grid. These features are probably too many and are redundant. Ke [6] applied Principal Components Analysis (PCA) in SIFT features to reduce the dimensionality of the feature vectors. Kobayashi [8] applied PCA in HOG features to reduce the dimensionality of the feature vectors. Locality Preserving Projection (LPP), as a linear version of manifold learning algorithm, has attracted considerable interests in recent years. As an effective way of reducing dimensionality of datasets, LPP can also be used in human detection. This paper focus on applying LPP to reduce the dimensionality of the HOG features vectors.

The remainder of this paper is organized as follows. The second section introduced the HOG image descriptor. The third section explained the linear manifold learning—LPP. And then a new image local descriptor was presented in the fourth section. The experimental results are described in section 5.And the performance of the LPP-HOG descriptor and the future work are given finally.

978-1-4244-3531-9/08/$25.00©2008 IEEE

II. HOG FEATURES

Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge direction. HOG features are calculated by taking orientation histograms of edge intensity in local region.

Specially, we can divide an image into N local regions called"block". And then this local region is divided into small spatial area called "cell". The spatial relation among cell, block and input image is shown in *Fig. 1*.And Extraction Process of HOG features are shown in *Fig. 2*. The HOG features are extracted from local regions with 16 × 16 pixels. Histograms of oriented gradients with 8 orientations are calculated from each 4 × 4 local cells.

Each feature is defined by its cell position $\mathrm{C}(x_c, y_c, w_c, h_c)$, the parent local region position $\mathrm{B}(x_b, y_b, w_b, h_b)$ and the orientation bin number k. So each feature f is denoted by $f(C, B, k)$. The gradients at the point (x, y) of image I can be found by convolving gradient operator with the image:

$$G_x(x,y) = \begin{bmatrix}-1 & 0 & 1\end{bmatrix} * I(x,y) \tag{1}$$

$$G_y(x,y) = \begin{bmatrix}-1 & 0 & 1\end{bmatrix}^T * I(x,y) \tag{2}$$

The strength of the gradient at the point (x, y) is:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{3}$$

The orientation of the edge at the point (x, y) is:

$$\theta(x,y) = \arctan\left[\frac{|G_y(x,y)|}{|G_x(x,y)|}\right] \tag{4}$$

We divide the orientation range into k bins and denote the value of k_{th} bin to be:

$$\varphi_k(x,y)=\begin{cases} G(\text{x,y}) & \text{if } \theta(x,y)\in bin_k \\ 0 & otherwise \end{cases} \quad (5)$$

Then the feature value is defined as:

$$f(C,B,k)=\frac{\sum_{(x,y)\in C}\varphi_k(x,y)}{\sum_{(x,y)\in B}G(x,y)} \quad (6)$$

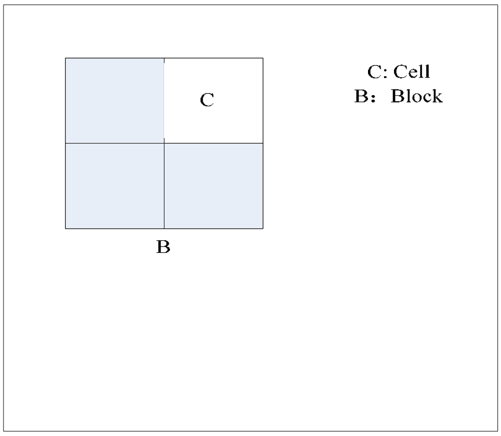

Fig. 1: the spatial relationship among cell, block and input image.

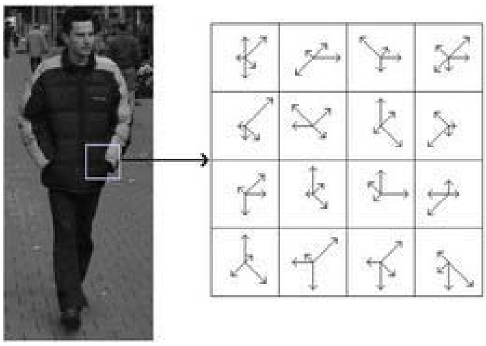

Fig. 2: Extraction Process of HOG features, The HOG features are extracted from local regions with 16 × 16 pixels. Histograms of oriented gradients with 8 orientations are calculated from each 4 × 4 local cells.

III. LPP ALGORITHM

LPP is originally derived by the linear approximation of the Laplace Beltrami operator on a compact Riemannian manifold; its aim is to preserve the similarity between the data, that is, data which are close to each other in the original space remain close after the projection. Assume the data set has n samples $X=\{x_1,x_2,...,x_n\}$, $x_i\in R^D$, each of which is on or near the sub-manifold whose intrinsic dimension is $d(d<<D)$. The algorithm of linear dimensionality reduction is to find a transformation matrix W that maps the original data to the objective space so that $Y=W^TX$.

The objective function of LPP is as follows:

$$\min_W \sum_{ij}(y_i-y_j)^2 S_{ij} \quad (7)$$

Here S_{ij} is the weight which measures the similarity of two data x_i, x_j; one way of defining the weight is to use the heat kernel:

$$S_{ij}=\begin{cases} \exp(-\|x_i-x_j\|^2)/t & x_i(or\ x_j)\ is\ among\ k\ nearest \\ & neighborhood\ of\ x_j(or\ x_i) \\ 0 & otherwise \end{cases} \quad (8)$$

The objective function can be simplified as follows:

$$\begin{aligned} W_{opt} &= \arg\min_W \sum_{ij}(y_i-y_j)^2 S_{ij} \\ &= \arg\min_W \sum_{ij}(W^Tx_i-W^Tx_j)^2 S_{ij} \quad (9) \\ &= \arg\min_W W^TXLX^TW \end{aligned}$$

In addiction, to remove the arbitrary scaling factor, a constraint is imposed $W^TXDX^TW=1$, here, $X=\{x_1,x_2,...,x_n\}$, $L=D-S$ is the Laplacian matrix, D is a diagonal matrix whose entries are row sum (or column, since S_{ij} is symmetric) of S, $D_{ij}=\sum_j S_{ij}$. So the final objective function becomes:

$$\begin{cases} \arg\min_W W^TXLX^TW \\ s.t: W^TXDX^TW=1 \end{cases} \quad (10)$$

The vectors that minimize the above objection can be gained through the following generalized eigenvalue problem:

$$XLX^Tw=\lambda XDX^Tw \quad (11)$$

Suppose $\lambda_0,\lambda_1,...,\lambda_{d-1}$ are the d smallest eigenvalues, and $w_0,w_1,...,w_{d-1}$ are the corresponding eigenvectors, then we can get the embedding: $x_i\rightarrow y_i=W_{LPP}{}^Tx_i$, y_i is a d dimensional vector, W_{LPP} is the transformation matrix. So far, the dimensionality reduction is obtained from $X\subset R^D$ to $Y\subset R^d$.

IV. LPP-HOG DESCRIPTOR

Our algorithm for local descriptors (termed LPP-HOG) accepts the same input as the standard HOG descriptor. The total number of features becomes over ten thousands when the HOG features extracted from all locations on the grid. These features are probably too many and are redundant. The HOG features extracted from regions without edges are redundant and not effective for classification because they are based on the information on edges.

Because LPP aims to preserve the intrinsic geometry of the data and local structure, it can get better result than PCA. With 4×4 cell size and 8×8 block size, the feature number of a block is 2×2×8=32.And then the feature dimension of a 64 × 128 image is 8 × 16 × 32=4096.We can find that the dimension of the feature is too great and the computational complexity is large.

With linear LPP algorithm, we can get a transformation vector W_{LPP}.When a test image arrives; we compute its HOG descriptor and then use W_{LPP} to get a low dimensional descriptor. Through above works, the computational complexity will be greatly reduced and also most redundancies of input descriptor will be thrown away.

Human detection work becomes difficult as a result of the variable size of human in images. And then we should use too many detection windows from all sub regions with variable size. And then the computational complexity will be too large for real-time human detection. To accelerate the detection process we use a detection window with a certain size. We give a frame of Pyramid-shaped for detecting variable size of human with a certain size detection window in images. The frame is shown in *Fig. 3*.

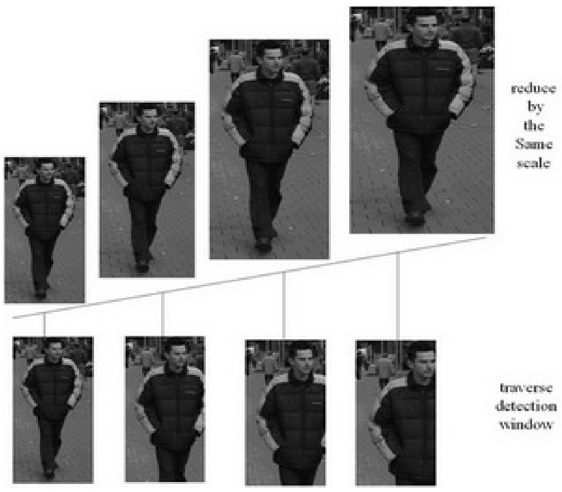

Fig. 3: a Pyramid-shaped frame for human detection

V. EXPERIMENTS

The proposed algorithm was evaluated by using MIT CBCL pedestrian database which contains 1000 images of pedestrians in city scenes. The size of the image is 64 × 128 pixels. These images were used for positive samples in the following experiments. The negative samples were originally collected from images of sky, mountain, airplane, building etc. The number of negative images is 2000. About 100 pedestrian images and 200 negative samples were used as test samples to evaluate the recognition performance of the local descriptor.

When we implemented Dalal algorithm using that dataset, the recognition rate for test dataset is 94.6%. We applied LPP-HOG feature selection to improve its result. After low dimensional LPP-HOG feature vectors were selected, we apply SVM classifier to get the recognition result. The cell size is 4×4 and the block size is 8×8. Some examples of human detection are shown in *Fig. 4*. The dimensional d and detection rate of PCA-HOG features and LPP-HOG features are shown in *Fig. 5*. As we can see, LPP-HOG works better than PCA-HOG when dimension is less than 25.The best recognition rate 95.1% with 2.9% false alarm rate was obtained at 24 LPP-HOG features which is better than both PCA-HOG and conventional HOG features.

Fig. 4: Two examples of human detection

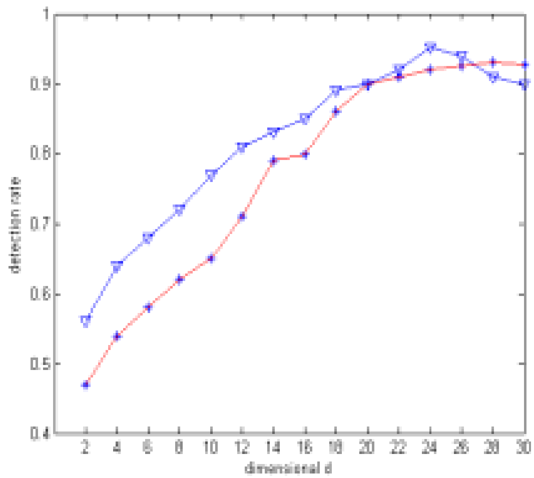

*Fig. 5: The dimensional d and detection rate of PCA-HOG features (labeled' *') and LPP-HOG features (labeled' ∇ ')*

VI. CONCLUSIONS

In this paper, we present a novel real-time human detection algorithm based on image local descriptor, called

LPP-HOG. We evaluated the performance of the LPP-HOG feature vectors for pedestrian detection. As a result, we could greatly reduce the number of features without lowering the performance. As other nonlinear manifold learning algorithm, LPP has two parameters: neighborhood size k and d. How to determine the size of k and d is our future work.

[1] D. G. Lowe. ”Distinctive image features from scale-invariant key points.” IJCV, 60(2):91–110, 2004.

[2] S. Belongie, J. Malik, and J. Puzicha. “Matching shapes.” The 8th ICCV, Vancouver, Canada, pages 454–461, 2001.

[3] Navneet Dalal and Bill Triggs. “Histograms of Oriented Gradients for Human Detection” Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[4] Navneet Dalal, Bill Triggs and Cordelia Schmid. “Human Detection Using Oriented Histograms of Flow and Appearance. “ECCV 2006, Part II, LNCS 3952, pp. 428–441, 2006.

[5] Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng. “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients.” 2006.

[6] Ke, Y., Sukthankar, R.” PCA-SIFT: A more distinctive representation for local image descriptors. “In: Proc. of Computer Vision and Pattern Recognition, Washington, pp. 66–75 .2004.

[7] Takuya Kobayashi, Akinori Hidaka, and Takio Kurita. “Selection of Histograms of Oriented Gradients Features for Pedestrian Detection.” ICONIP 2007, Part II, LNCS 4985, pp. 598–607, 2008