

HOG-NPE: A Novel Local Description Operator for Face Image Recognition

Hongfeng Wang*

Dezhou University, Dezhou, Shandong 253023, China

Received 20 June 2014; accepted (in revised version) 28 August 2014; available online 23 September 2014

Abstract

The method of extracting robust feature sets from an image is a crucial issue for the areas of computer vision and pattern recognition. The nature of real data is a very high dimensional data. However, the hidden structure can be well characterized by a small number of features in most cases. As a result, the method of extracting a small number of good features is an important question in computer vision and pattern recognition, etc. We employ the Histograms of Oriented Gradient (HOG) to extract the robust feature sets of facial images, which is a local description operator that possesses a certain degree of invariance against geometric and photometric deformations. Neighborhood Preserving Embedding (NPE) which is a subspace learning algorithm is adopted to extract a small number of good features on the local description operators. We use the novel local description operator - HOG-NPE for facial image recognition, and several experiments on well-known facial databases are conducted, which demonstrate good performance and effectiveness of this novel local description operator.

Keywords: Good Features; Histograms of Oriented Gradient; Subspace Learning; Neighborhood Preserving Embedding

1 Introduction

The method of extracting robust feature sets has been an issue of concern in the field of machine learning, computer vision, pattern recognition, etc. Facial image recognition is also a challenging task when encountering practical problems such as illumination, occlusion, expression and deformation, etc. The method, *Histograms of Oriented Gradient* (HOG), is a kind of local description operator, which mainly possesses two advantages. Firstly, the local shape and appearances are well presented by capturing the edge or its gradient structure. Secondly, it possesses a good invariance against image geometry and optical deformation [1].

The nature of real data is a very high dimensional data. Practical algorithms do not perform well, and possess too many features which are not necessary or are redundant. Therefore, the

*Corresponding author.

Email address: dzuwhf@163.com (Hongfeng Wang).

method of extracting a small number of good features plays an important role in the fields of machine learning, computer vision, pattern recognition, etc. The underlying structure is well characterized by a small number of features in many cases [2–5]. As a result, the linear subspace learning will be appropriate for the exploration or capturing the intrinsic structure and, at the same time, obtaining the small computational cost [6–8]. One of the most popular linear subspace learning techniques is the *Principal Component Analysis* (PCA). PCA aims at preserving the global Euclidean structure on data manifold and only performs well when the manifold is embedded linearly in ambient space. While *Neighborhood Preserving Embedding* (NPE) aims at preservation of the local neighborhood (manifold) structure, which is less sensitive to outliers than PCA. In this paper, the HOG features were extracted and then a linear subspace with NPE was learned. And finally, we are able to gain a novel local description operator HOG-NPE [2].

The rest of this paper is organized as follows. In Section 2, the local description operator Histograms of Oriented Gradient (HOG) will be introduced. *Neighborhood Preserving Embedding* (NPE) is reviewed in Section 3. In Section 4, our novel local description operator HOG-NPE will be proposed, then, the experiment setup will be presented in Section 5 followed by the conclusion in Section 6.

2 Local Description Operator Histograms of Oriented Gradient (HOG)

Histograms of Oriented Gradient (HOG) algorithm is a local description operator, which is operated on small unit grids ('cells'). In actual operation, the image is divided into small regions of space, where a local one-dimensional histogram of gradient directions over pixels of the cell is accumulated. An example is demonstrated in the third image 'Cells of face' in Fig. 1. Before forming the one-dimensional histogram of gradient directions, it is necessary to implement an effective local contrast normalization over 'block' which includes, 2×2 cells, as shown in the last image in Fig. 1. Finally, the final description is constituted by the combined histogram entries. The method which involves normalizing the local histograms of gradient directions in a dense grid has two advantages: the first one is that the local appearance and shape of an object can be well characterized by the capturing gradient directions structure. The second advantage is that, it possesses a good invariance against image geometry and optical deformations. The steps are as follows [1, 9, 10].

(1) Compute the gradient. We employ a template $[-1 \ 0 \ 1]$ and $[-1 \ 0 \ 1]^T$ to convolve with the original image for each pixel, respectively in the x-axis and y-axis directions.

(2) Form the one-dimensional histogram of gradient directions over the pixels of the cell. The image is divided into small spatial regions called "cells". The size, for example, 8×8 pixels, with the number of bins of orientations 9, and the gradient direction is 180° or 360° . One-dimensional histogram of gradient directions over the pixels of each cell will be formed by collecting the statistical gradient information.

(3) Put cells into blocks. Along with the local variations in illumination and face image contrast, gradient strength also varies over a wide range. As a result, it is necessary to implement an effective local contrast normalization. In order to achieve better results, a Gaussian spatial window is applied for each block.

(4) Finally, the normalization scheme L2-Hys is adopted for each block. Then, the final descriptor is constituted by the well-normalized local histograms of face image gradient directions in a dense grid.

To understand these steps, please see Fig. 1 for the process of extracting the HOG features.

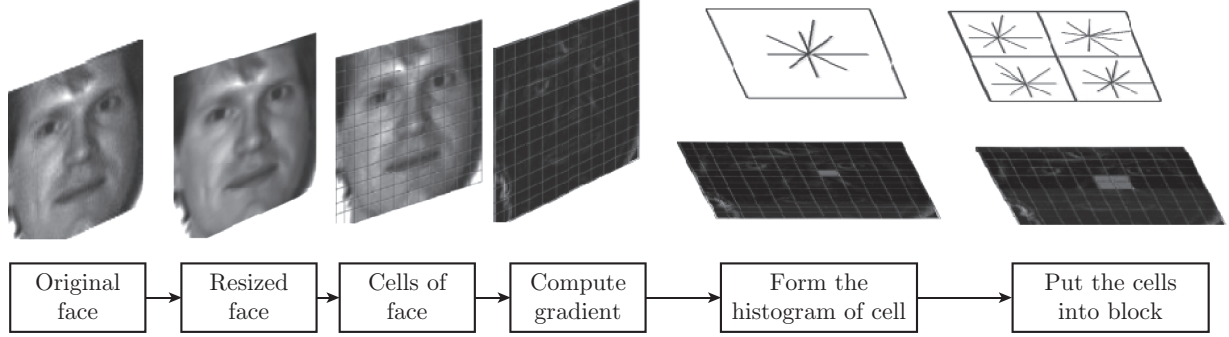


Fig. 1: The specific process of extracting HOG features

3 Neighborhood Preserving Embedding (NPE)

Neighborhood Preserving Embedding (NPE) a subspace learning algorithm, and is adopted to extract a small number of good features on the samples data. NPE aims at preserving the local neighborhood (manifold) structure, which will be less sensitive to outliers than PCA. The main idea of NPE is that for each data point, each data point will be represented as a linear combination for the other neighboring points. The rough steps are as follows [2].

Let $X = \{X_1, X_2, \dots, X_N\}$ denote data set and N is the number of the data set.

(1) Build an affinity graph. Let G denote the graph with N nodes. The KNN method was adpted to put an edge between node i and node j , if X_i is among the K nearest neighbors of X_j , or X_j is among the K nearest neighbors of X_i .

(2) Obtain the weights. Let S denote the weight matrix with S_{ij} computing the weights of the edge between node i and node j . S can be obtained by minimizing the distance between node i and the linear combination of this node's neighboring points.

$$\min \sum_i \|X_i - \sum_j S_{ij} X_j\|^2, \quad \text{subject to equ (2)} \quad (1)$$

$$\min \sum_j S_{ij} = 1, \quad j = 1, 2, \dots, K \quad (2)$$

(3) Work out the objective function of NPE. By solving the generalized eigenvector problem, the projections of matrix W can be obtained, and then the NPE embedding can be performed. The objective function is as follows,

$$\arg \min_{W^T X X^T W = 1} W^T X L X^T W \quad X L X W = \lambda X X W \quad (3)$$

where

$$X = (X_1, \dots, X_N), \quad L = (I - S)^T (I - S), \quad I = \text{diag}(1, \dots, 1) \quad (4)$$

In this section, the novel local description operator, HOG-NPE will be introduced. First, the robust feature sets of facial images is extracted by the HOG, because it is a local description operator that possesses a certain degree of invariance against geometric and photometric deformations. The underlying structure will be well characterized by a small number of features in most cases. Therefore, the linear subspace learning will be appropriate to explore or capture the intrinsic structure and at the same time, obtain a small computational cost. As a result the subspace learning algorithm NPE can be adopted to extract a small number of good features on the local description operators.

3.1 Extracting the Local Description Operator HOG Features for Facial Images

The steps for extracting local description operator HOG features are presented in Section 2. The detailed processes are described as follows.

(1) As is described in Section 2. (1) the gradient, the magnitude of gradient and gradient orientation at the point (x, y) are as follows,

$$\text{Grad}_x(x, y) = \text{Img}(x + 1, y) - \text{Img}(x - 1, y) = [-101] * \text{Img}(x, y) \quad (5)$$

$$\text{Grad}_y(x, y) = \text{Img}(x, y + 1) - \text{Img}(x, y - 1) = [-101] * \text{Img}(x, y) \quad (6)$$

The magnitude of gradient and gradient orientation at the point (x, y) is:

$$\text{Grad}(x, y) = \sqrt{\text{Grad}_x(x, y)^2 + \text{Grad}_y(x, y)^2} \quad (7)$$

$$\phi(x, y) = \arctan(\text{Grad}_y(x, y) / \text{Grad}_x(x, y)) \quad (8)$$

(2) First, resize the facial image into an appropriate size, such as resizing of the original Yale face 100×100 into 96×96 , and the resizing the original Extended Yale-B face 192×168 into 192×160 . Then, we set cell size as 8×8 pixels, the number of bins of orientations as 9, and the gradient direction as 180° or 360° , overlap be 0 or 0.5. In the Section 5, the influence of these parameters will be explored. An one-dimensional histogram of gradient directions over the pixels of each cell is thus formed by collecting the statistical gradient information.

(3) Cells are collected into larger blocks, such as, 2×2 or 4×4 cells for each block. Then, implement local contrast normalization for each block, with methods such as the L2-Hys normalization scheme.

(4) These one-dimensional histograms of gradient directions was are thus collected over the pixels of the cell into a final description vector. The specific steps are illustrated in Fig. 1.

3.2 Adopting NPE to Learn a Linear Subspace Through the HOG Features

The underlying structure is well characterized by a small number of features in most cases. As a result, the NPE was adopted to learn a linear subspace through the HOG features. The process of learning this linear subspace on the HOG features is as follows.

$X = \{X_1^{\text{hog}}, X_2^{\text{hog}}, \dots, X_N^{\text{hog}}\}$ denotes the descriptor vector set and N is the number of the descriptor vector set.

Step 1. Build an affinity graph for these descriptor vector sets. Let G^{hog} denote the graphs with N nodes and the point X_i just is the i -th node. We use the KNN method to find the K nearest neighbors and connect the corresponding edges.

Step 2. Set weights for Graph G^{hog} . Let S_{ij}^{hog} denote the weight matrix. We minimize the following objective function to compute the weights. The details to solve this minimization problem is shown in [11].

$$\min \sum_i \|X_i^{\text{hog}} - \sum_j S_{ij}^{\text{hog}} X_j^{\text{hog}}\|^2, \quad \text{subject to equ (10)} \quad (9)$$

$$\min \sum_j S_{ij}^{\text{rmhog}} = 1, \quad j = 1, 2, \dots, K \quad (10)$$

Step 3. Getting the objective function of NPE and compute the projections. First, we solve the generalized eigenvector problem of the objective function:

$$\arg \min_{W^T X^{\text{hog}} X^{\text{hog}T} W = 1} W^T X^{\text{hog}} L X^{\text{hog}T} W \quad X^{\text{hog}} L X^{\text{hog}} W = \lambda X^{\text{hog}} X^{\text{hog}} W \quad (11)$$

where

$$X^{\text{hog}} = (X_1^{\text{hog}}, \dots, X_N^{\text{hog}}), \quad L = (I - S^{\text{hog}})^T (I - S^{\text{hog}}), \quad I = \text{diag}(1, \dots, 1) \quad (12)$$

Step 4. NPE Embedding.

$$X_i^{\text{hog}} \rightarrow Y_i^{\text{hog}} = W^T X_i^{\text{hog}} \quad W = (W_0, \dots, W_{d-1}) \quad (13)$$

where we let the column vectors W_0, \dots, W_{d-1} (d is the dimensionality to be reduced) be the solutions of equation (11), ascending ordered according to their eigenvalues, that is $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{d-1}$.

4 Experiments Setup

4.1 Experiment Results on Yale Face Database

Yale face database has 15 individuals, and there are 11 images per person in *bmp* format. There is a total of 165 grayscale images containing the change of facial expressions, illumination and posture [12]. In this experiment, there are 7 groups of different training and testing sets, (G2/P9, \dots , G8/P3). Gm/Pn denotes that we randomly choose m images for training and n images for testing of each person. After the novel local description operator is obtained, HOG-NPE, we will iterate the same experiment 50 times for each group and set the average as the final result.

The recognition rate of Yale face database of our algorithms is demonstrated in Table 1 and Fig. 6 (a), and the parameters we chose are cell 8×8 , block 2×2 cells, bins 9, overlap 0.5.

Table 1: The recognition rate on Yale ; recognition rates (%) \pm std (%) [dim]

Algorithm	G2/P9	G3/P8	G4/P7	G5/P6	G6/P5	G7/P4	G8/P3
NPE	61.76 \pm 7.33 [14]	71.42 \pm 3.65 [17]	76.15 \pm 2.67 [26]	77.00 \pm 2.81 [35]	78.29 \pm 3.93 [23]	80.03 \pm 3.86 [50]	79.82 \pm 5.6 [56]
HOG-NPE	72.10 \pm 7.11 [14]	85.52 \pm 3.51 [17]	88.29 \pm 2.16 [14]	91.64 \pm 3.14 [14]	94.56 \pm 3.08 [11]	96.87 \pm 2.22 [14]	98.40 \pm 1.80 [14]

4.1.1 Explore the Influence of HOG Parameters

In the process of extracting HOG features, these HOG parameters, such as the cell, block and overlap, will be involved. In order to explore the influence of these parameters, we have designed several groups of experiments on the Yale face database. The result is shown in Fig. 2.

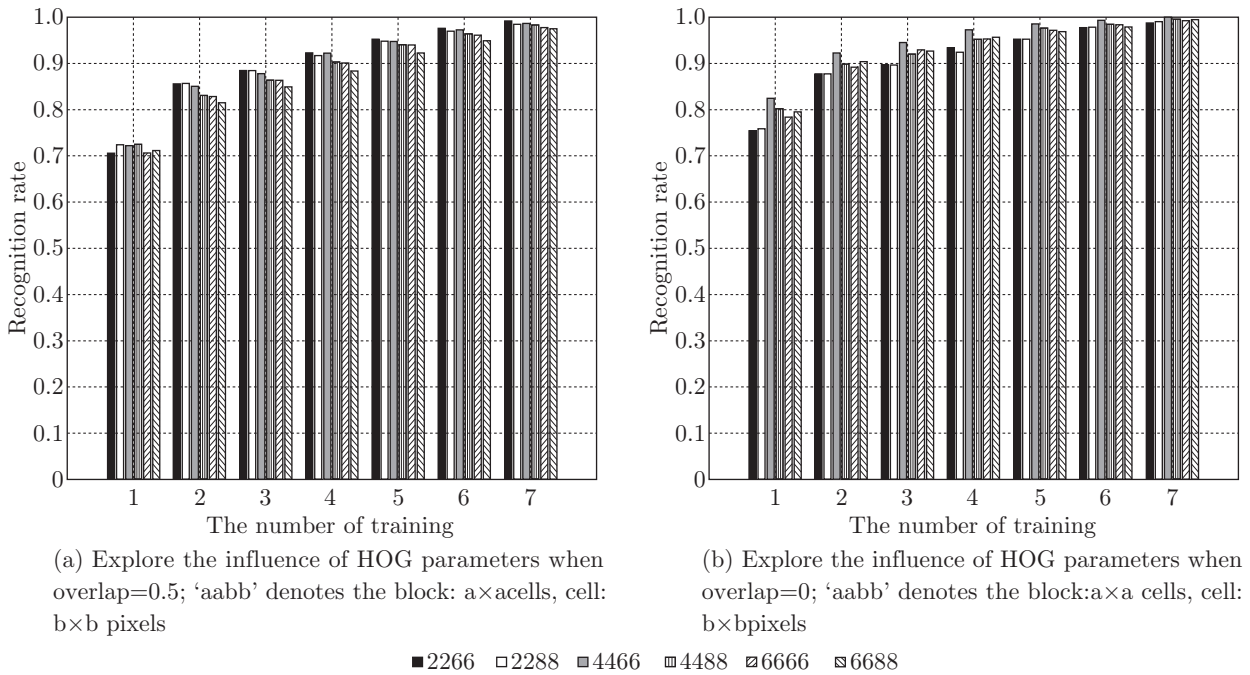


Fig. 2: Explore the influence of HOG parameters

From Fig. 2 (a), we can see that '2266' has the best result. In order to acquire the high recognition rate we should employ the smaller block and cell. From the Fig. 2 (b), we can see that '4466' has the best result. In order to acquire a high recognition rate we should employ the appropriate combination of block and cell. Comparing the figures of with Fig. 2 (a) and Fig. 2 (b), we find that we can gain better result when the overlap is 0.

4.1.2 Robust Test Against Geometric Deformations

Geometric deformations may occur during image acquisition and processing, these would result in a sharp decline in the recognition rate. In this subsection, we explore two ways that geometric deformations occur. The first is the "translation", we will shift the (5, 5) pixels from the top left to the bottom right for the 1st, 5th, and 9th images of each person, as shown in Fig. 3. The

other is to carry out the “shear” geometric deformation for the 1th, 5th, and 9th images of each person, and as is shown in Fig. 4., the recognition rate on Yale facial database that is suffering from geometric deformation is demonstrated in Table 2 and Fig. 6 (b).



Fig. 3: The first person face images with translation on Yale database



Fig. 4: The first person face images with shear on Yale database

Table 2: The recognition rate on Yale suffering geometric deformation, the parameters we set are block 2×2 , cell 8×8 , overlap 0.5

Algorithm	G2/P11	G3/P10	G4/P9	G5/P8	G6/P7	G7/P6	G8/P5
Original NPE	61.76%	71.42%	76.15%	77.00%	78.29%	80.03%	79.82%
Original NPE with translation	37.04%	49.45%	55.01%	59.36%	63.17%	66.27%	68.53%
HOG-NPE with translation	40.62%	57.45%	64.42%	69.76%	72.13%	76.13%	79.11%
Original NPE with shear	37.76%	48.80%	53.77%	58.49%	61.31%	64.57%	66.04%
HOG-NPE with shear	55.45%	72.82%	78.00%	80.53%	82.80%	85.90%	89.20%

4.2 Experiment Results on Extended Yale-B Face Database

The Extended Yale-B face database contains 16128 face images of 38 human subjects under 9 poses and 64 illumination conditions. We acquire Extended Yale-B face from [13], in which we choose the top 20 individual facial pictures and choose 16 images per human subject (we choose one every four images in this database, 64 images per human subject), for a total 320 images in the *pgm* format. We use this sub Extended Yale-B for our experiments.

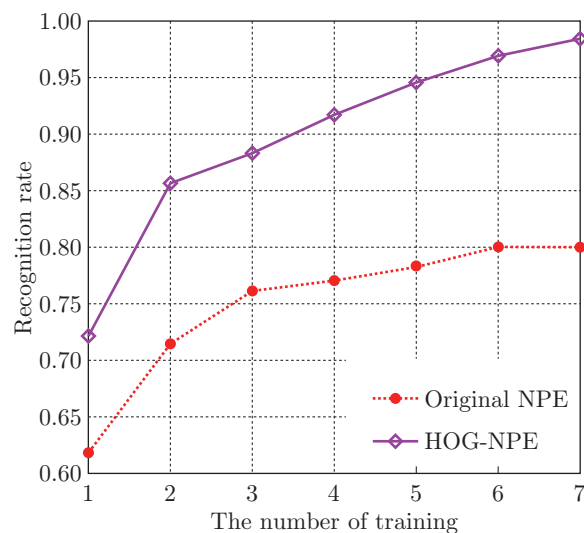
In this experiment, we have 7 groups of different training and testing set, ($G3/P13, \dots, G9/P7$). Gm/Pn denotes that we randomly choose m images for training and n images for testing of each person. After we gain the novel local description operator, HOG-NPE, we will iterate the experiments for each group 50 times and set the average as the final result. The first person containing 16 images is demonstrated in Fig. 5. The recognition rate of our algorithms on our sub Extended Yale-B face database is demonstrated in Table 3 and Fig. 6 (c).



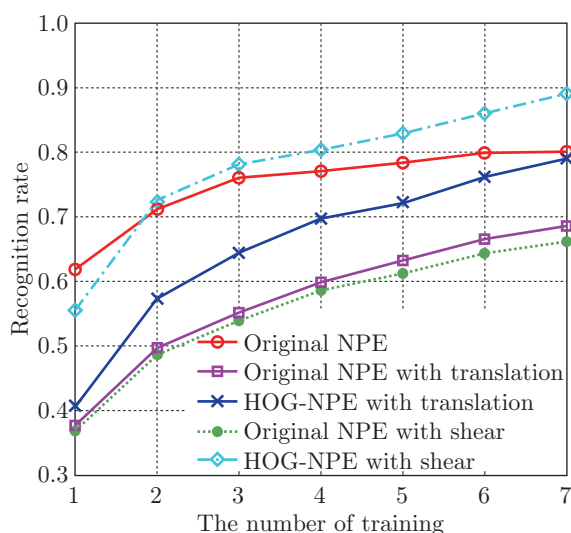
Fig. 5: The 16 face images of the first person on our sub Extended Yale-B

Table 3: The recognition rates on our sub Extended Yale-B; recognition rates (%) \pm std (%) [dim]

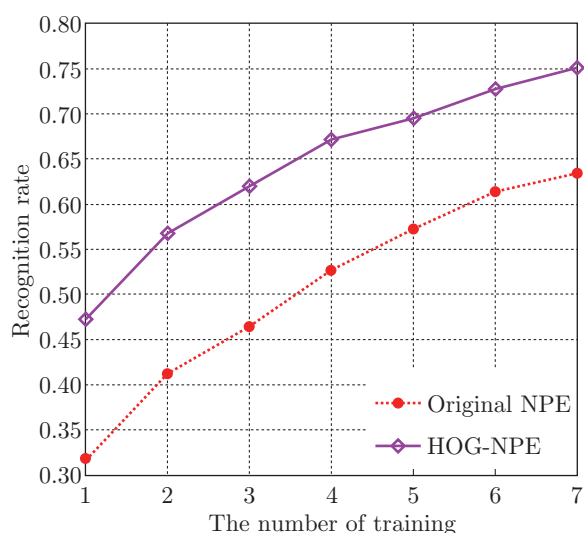
Algorithm	G3/P13	G4/P12	G5/P11	G6/P10	G7/P9	G8/P8	G9/P7
NPE	31.82 \pm 5.57 [26]	41.16 \pm 4.90 [38]	46.40 \pm 4.88 [44]	52.68 \pm 4.62 [59]	57.24 \pm 4.00 [65]	61.37 \pm 3.86 [77]	63.47 \pm 3.94 [89]
HOG-NPE	47.02 \pm 4.14 [26]	56.73 \pm 3.55 [35]	61.61 \pm 2.66 [44]	67.08 \pm 2.98 [53]	69.41 \pm 3.76 [62]	72.90 \pm 3.30 [74]	75.07 \pm 3.49 [74]



(a) The recognition rates for two different algorithms on yale facedatabase



(b) The recognition rates against distortion on yale face database



(c) The recognition rates for two different algorithms on our sub extended yale-B face database

Fig. 6: The recognition rate of the 7 groups data on Yale and our sub Extended Yale-B database

5 Conclusion

In this paper, the HOG was employed to extract robust feature sets of facial images because of its unique advantages as a local description operator. However, the dimensionality of the final

descriptor is high which results in large computational cost and will contain redundant information. As a result, the NPE was used to learn a linear subspace. The underlying structure is well characterized by a small number of features in many cases. Finally, a novel local description operator HOG-NPE was obtained. The experiment results show that our novel HOG-NPE operator has a certain degree of robust against illumination and geometric deformations, etc.

Acknowledgement

Project supported by Natural Science Foundation of Shandong (No. ZR2012AM008).

References

- [1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, 1: 886-893.
- [2] He XF, Cai D, Yan SC, Zhang HJ. Neighborhood preserving embedding. *IEEE International Conference on Computer Vision, ICCV 2005*, 2005, 2: 1208-1213.
- [3] Jiang XY, Zhou XH, Weng M, Zheng JJ, Jiang YX. Image processing techniques and its application in water transport through fabrics. *J. Fiber Bioengineering Informatics*, Vol. 3, No. 2, Sept 2010. pp. 88-93
- [4] Sun C, Zheng R, Li J, Huang HQ, Cui MH. Image perception of female breast beauty and its relation to 3D anthropometric measurements. *J. Fiber Bioengineering Informatics*, Vol. 4, No. 1, Apr. 2011. pp. 23-34
- [5] Li PF, Jiao Y, Jing JF, Li JN. The high-speed fabric defect detection algorithm based on the image layered model. *Journal of Fiber Bioengineering & Informatics* 6: 2 (2013) 161-173.
- [6] Cai D, He XF, Han JW, Zhang HJ. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, 2006, 15(11): 3608-3614.
- [7] Cai D, He XF, Zhou K. Locality sensitive discriminant Analysis. *IJCAI*. 2007: 708-713.
- [8] Cai D, He XF, Han JW. Semi-supervised discriminant analysis. *IEEE 11th International Conference on Computer Vision, ICCV 2007*, 2007: 1-7.
- [9] Wang QJ, Zhang RB. LPP-HOG: A new local image descriptor for fast human detection. *IEEE International Symposium on Knowledge Acquisition and Modeling Workshop*, 2008: 640-643.
- [10] Zhu, Qiang, et al. Fast human detection using a cascade of histograms of oriented gradients. *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Vol. 2. IEEE, 2006.
- [11] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(5500): 2323-2326.
- [12] <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>
- [13] <http://vision.ucsd.edu/leekc/ExtYaleDatabase/ExtYaleB.html>