

Pedestrian detection based on gradient and texture feature integration

Chun-Hou Zheng^{a,c,*}, Wen-Juan Pei^a, Qing Yan^a, Yan-Wen Chong^b

^a College of Electrical Engineering and Automation, Anhui University, Hefei, China

^b State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

^c Center of Information Support & Assurance Technology, Anhui University, China

ARTICLE INFO

Keywords:

Pedestrian detection
Local binary patterns
Histogram of oriented gradient
Sparse representation
K-SVD

ABSTRACT

In this paper, based on feature integration, we proposed a new method for pedestrian detection. Firstly, we extracted the histogram of oriented gradients (HOG) feature and local binary pattern (LBP) feature from the original images respectively. Secondly, K-singular value decomposition (K-SVD) was used to extract sparse representation features from the HOG and LBP features. Moreover, PCA was used to reduce the dimension of HOG and LBP. Finally, we combined the PCA based features and the K-SVD based sparse representation features directly for fast pedestrian detection in still images. Experimental results on two databases show that the proposed approach is effective for pedestrian detection.

1. Introduction

Pedestrian detection is an important branch of pattern recognition which can be applied in monitoring, intelligent machines, aircraft, image retrieval, smart cameras [7], etc. But real-time pedestrian detection from static images is still highly challenging at present due to the complicated background and the wide range of the pedestrians' states such as illumination condition, occlusion, camera angle and the variability in pose, etc. Usually, pedestrian detection methods firstly extract the features from the images, then train the corresponding classifiers which is applied to detect the interested areas through the sliding window technology. Until now, many feature extracting methods have been proposed [32–35]. For pedestrian detection, the locally normalized histogram of oriented gradients (HOG) descriptor recently proposed by Dalal and Triggs [2] is one of the most popular used feature extraction methods. By using HOG to extract the gradient features of the Gabor preprocessed images, the improved method HoGG is presented [21] for human detection. Zhu et al. [18] proposed a method for fast human detection by integrating a cascade-of-rejecters with HOGs of variable-size blocks to extract salient features. In the implementation, the best blocks are selected by using AdaBoost [19,20,22] which has been generally used for face detection.

Another feature extracting method, i.e., the local binary pattern (LBP) texture descriptors [10], has been widely used for texture classification and face recognition for its invariance to gray level changes and good anti-noise performance. In recent years, plenty of improved LBP feature extraction methods have been proposed [3,4,8,9,11–13]. Tan and Triggs proposed the local ternary pattern

(LTP) [13] to apply two parameters as the center pixel thresholds instead of a single in the original LBP, which improves the operator's ability to resist noise. Wu et al. [4] proposed the semantic LBP and Fourier LBP which capture more about the local features and global statistics as the pedestrian detection region descriptors. Experiments on the INRIA dataset show that it outperforms other gradient-based features, for example the HOG features. To effectively represent missed information in the original LBP and achieve better texture classification, the completed LBP (CLBP) [12] is proposed from the viewpoint of a local difference sign-magnitude transform (LDSMT).

Wang et al. [6] proposed a texture descriptor (HOG-LBP features) which combines edge information (i.e., HOG) and the texture information (i.e., cell-structured LBP) for human detection. By using the linear SVM as the classifier, a detection rate of 91.3% at 10^{-6} FPPW and 94.7% at 10^{-5} FPPW on the INRIA dataset is gotten. Zhang et al. [3] proposed Local Structured HOG (LSHOG), Local Structured LBP (LSLBP) and the boosted Local Structures HOG-LBP for object localization. Zeng et al. [15] used a mi-SVM (Support Vector Machine for multiple instance learning) and a cascade of HOG-LBP feature for human detection, and achieved a near real-time detection speed and similar accuracy to the HOG-LBP features [6]. Won-Jae Park et al. [17] extract the HOG-LBP features from the selected regions (by training the AdaBoost classifier using the block SVM response as the features [19]) and then reevaluate the candidates which are not rejected.

Recently, Wang et al. [14] use the principal components analysis (PCA) to reduce the dimension of HOG features from 3780 to 30 while keeping the recognition rate essentially the same on INRIA dataset. Lu et al. [5] use PCA to project the HOG feature to a linear subspace for

* Corresponding author at: College of Electrical Engineering and Automation, Anhui University, Hefei, China.
E-mail address: zhengch99@126.com (C.-H. Zheng).

tracking and action recognition. A hybrid approach [24] of PCA and LBP is also proposed for facial expression feature extraction. Zeng et al. [23] combine the multilevel HOG with the multilevel LBP as the feature vectors of a robust and rapid head-shoulder detector.

Sparse representation methods have been extensively used in image denoising, document analysis, image reconstruction, object recognition and detection [25–27,31], etc. For sparse representation method, the dictionary updating method K-SVD [16,28,29] can be used to better represent the data. E.g., Ren et al. [30] proposed a sparse representation Histograms of Sparse Codes (HSC) which extract the local histograms by aggregating sparse codes based on K-SVD of each pixel. The proposed sparse representation HSC features can represent more than gradient structures and the detection performance on the PASCAL benchmark data is greatly improved compared with traditional methods.

In this paper, we explore the performance of several combined features for pedestrian detection. Firstly, we study the performance of the features extracted by using PCA from HOG and the cell-structured Local Binary Pattern (LBP) features, which denoted as PCA-HOG, PCA-LBP, PCA-HOG and PCA-LBP respectively. Then we study the performance of the sparse features extracted using the dictionary learned from K-SVD from the HOG and LBP features, which denoted as K-SVD-HOG, K-SVD-LBP, K-SVD-HOG and K-SVD-LBP respectively. The combined original HOG-LBP [6] feature is used for comparison. We also combined K-SVD-HOG and PCA-HOG for pedestrian detection, the experimental results show that this method can reduce the dimension of the feature set and time consumption while not sacrificing detection performance significantly.

The rest of the paper is organized as follows: In Section 2, we introduce of the HOG, the cell-structure LBP, and the K-SVD. Section 3 gives the detailed description of our algorithm. Section 4 shows the experimental results on INRIA dataset and Daimler pedestrian dataset. Section 5 concludes this paper and outlines the future works.

2. Image feature extraction methods

2.1. Local binary patterns of image

The original local binary patterns (LBP) [10] is proposed by Ojala et al. to measure the differences of a pixel and its surrounding pixels for texture classification.

Given a pixel of a grayscale image, the basic mathematical formula of the LBP operator is expressed as:

$$LBP_{P,R} = \sum_{i=0}^{P-1} \delta_i 2^i, \delta_i = \delta(h_i - h_c) = \begin{cases} 0, & h_i \geq h_c \\ 1, & h_i < h_c \end{cases} \quad (1)$$

here h_c is the center pixel, h_i is the circularly symmetric neighborhoods. The parameter P controls the quantization level, R determines the radius of the operator. To get the number of the bitwise 0–1 transitions, a spatial uniform LBP is calculated in the pattern. An uniform LBP patterns have limited the transition numbers ($U(LBP_{P,R}) \geq 2$) to 2. For instance, 11111111(0 transitions), 10001111(2 transitions) are uniform which can be mapped as $LBP_{P,R}^2$ (superscript means the 0–1 transitions $U(t)$ and the other situations are non-uniforms).

$$U(LBP_{P,R}) = \sum_{i=1}^{P-1} |\delta_i - \delta_{i-1}| + |\delta_0 - \delta_{P-1}| \quad (2)$$

It has been observed by Ojala et al. that plenty of uniform LBP can be categorized in texture image and then are cast into histogram bins to produce the final feature vectors. Note that the non-uniform LBP is marked with a single label simultaneously.

Wang et al. [6] proposed a cell-structured LBP which is suitable for pedestrian detection. Firstly, the image is segmented into the same size non-overlapping cells for creating pattern histograms. The LBP pattern

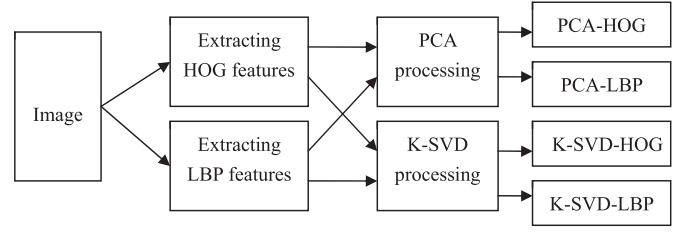


Fig. 1. An overview of feature extracting method.

histograms from the whole cells are concatenated to represent the texture features of the image.

2.2. Histograms of oriented gradients of image

Histograms of oriented gradients (HOG) [2] is well known to extract object gradient histogram, which is one of the widely accepted descriptors to calculate the edge and local shape information of image. It has been applied to pedestrian detection and object recognition successfully [2].

To reduce the influence of light and noisy, the HOG descriptor manages the images with color normalization and gamma normalization. Then computing the direction gradient using the template operator (a 1-D gradient filter, i.e., $[-1, 0, 1]$). The gradient is divide into 9 orientation bin levels linearly ranging in 0–180°. The gradient amplitudes in each direction are the weights. The image is divided into 16×16 pixel block (containing four 8×8 pixel cells, rectangular blocks(R-HOG) or circular blocks(C-HOG)) and the histograms of the cells are concatenated to produce the feature vector of the block. To reduce the effect of the local illumination variation and the transformation of the visual angle, the Lowe-style clipped L_2 norm normalization is used to the feature vector. Finally, concatenating all the feature vectors of the blocks (8 pixel spacing stride) into a vector to generate the HOG feature vector of a 2D picture.

Given a 128×64 sliding detection window [2], (the cell size is 8×8 , each block contains 4 cells, the gap between the blocks is 8 pixel and 4 pixel coverage of each cell) we can get a 3780 dimensions HOG feature vector, which is represented by 7×15 blocks.

2.3. The K-SVD algorithm

The K-SVD [16] is a generalization of the k-means clustering algorithm which learns a dictionary for sparse and redundant representation. It's an iterative implementation containing two stages. The updating computations of the sparse coding coefficient based on current dictionary and atoms column by column in the dictionary are separated. Given a set of signals $Y = \{y_1, y_2, \dots, y_N\} \in R^{M \times N}$ ($M < N$) and an overcomplete dictionary $D = \{d_1, d_2, \dots, d_K\} \in R^{M \times K}$ (assuming $M < K$, thus the dictionary is a full-rank matrix), a given signal $y_i \in R^{M \times 1}$ can be represented as a sparse combination of the K atoms $\{d_k\}$. The sparse representation coefficients $X = \{x_1, x_2, \dots, x_N\} \in R^{K \times N}$ of the signal should contain zero coefficients as much as possible. Which can be solved by getting the smallest l_0 norm of X , i.e., the solution of $\min \|x_i\|_0$. Moreover, the sparse representation and the overcompleted dictionary should satisfy the precise condition $Y = DX$ or approximate condition where the deviation is measured by l_2 norm $\|y_i - Dx_i\|_2 \leq \epsilon$ ($i = 1, 2, \dots, N$).

The k-means clustering iteratively alternates between extracting a sparse representation by the closest dictionary atom of the signal itself and updating the dictionary atom. Under l^2 -norm distance, that is to find the nearest neighbor to represent the signals Y by solving

$$\min_{D, X} \left\{ \sum_{i=1}^N \|y_i - Dx_i\|_2^2 \right\} \quad s. t. \quad \forall i, x_i = e_k \text{ or } \forall i, \|x_i\|_0 = 1 \quad (3)$$



Fig. 2. Some sample images from the INRIA database.

Here $x_i = e_k$ is a basis vector with a one in the k th position. That means only one column in dictionary D is used to represent the signal.

The K-SVD algorithm is a generalization of K-means algorithm to represent signal as a linear combination of more than one but less than L_0 atoms in D . So, the objective could be written as:

$$\min_{D, X} \left\{ \sum_{i=1}^N \|y_i - Dx_i\|_2^2 \right\} \quad \text{s.t.} \quad \forall i, \|x_i\|_0 \leq L_0 \quad (4)$$

At the first stage, the dictionary D is fixed and then find the approximately optimal coefficients x_i to represent the signal y_i . An approximation pursuit algorithm named the orthogonal matching pursuit (OMP) can be used for the calculation. At the second stage, the sparse representation X is fixed, and then updating all column (d_k) one by one in dictionary D . Assuming the k th row sparse coding in X is x_T^k which is corresponding to the dictionary d_k . The objective of updating d_k can be alternatively rewritten as the following minimization problem:

$$\sum_{i=1}^N \|y_i - Dx_i\|_2^2 = \|Y - DX\|_F^2 = \left\| \left(Y - \sum_{i \neq k} d_i x_T^i \right) - d_k x_T^k \right\|_F^2 \quad (5)$$

Assuming the first item of the right side formula is fixed, we can solve the problem of updating d_k and x_T^k by approximately using singular value decomposition (SVD) [16]. The nonzero representation $x_T^k(i)$ of the signals Y using the atom d_k is defined as:

$$Z_k = \{i | 1 \leq i \leq M, x_T^k(i) \neq 0\} \quad (6)$$

Then, a matrix $\Omega \in R^{N \times Z_k}$ is defined with ones on the $(Z_k(i), i)$ position and zeros otherwise. The original minimization problem is equivalent to:

$$\left\| \left(Y - \sum_{i \neq k} d_i x_T^i \right) \Omega - d_k x_T^k \Omega \right\|_F^2 = \|E - d_k x_R^k\|_F^2 \quad (7)$$

Here E is decomposed into $U\Delta V^T$ by using singular value decomposition directly. The first column of U is defined as the solution for d_k and the first column of $V \times \Delta(1, 1)$ is defined as the coefficient vector x_R^k .

3. Feature combination method

Our experimental process is shown in Fig. 1. As is known to all, PCA is a widely used method for dimension reduction. By finding a linear transformation that maps the high dimensional features to a lower space, PCA could expose the latent structures of the original features. In [14], the authors use PCA to reduce the HOG feature vector dimensions while keeping the recognition rate essentially the same as on original features and gain a practical filter tracking. Inspired by them, we firstly survey the effects of combining the PCA with the LBP and HOG features respectively, thus obtaining the low dimensional feature, i.e., PCA-HOG and PCA-LBP. The LBP and the HOG features will be used as the comparisons group respectively. Secondly, we apply

the efficient K-SVD algorithm [16] to produce the sparse representation of the HOG and LBP features set separately by learning the dictionary, we denoted them as K-SVD-HOG and K-SVD-LBP. In addition, we combine each of two feature set respectively, i.e., K-SVD-LBP and PCA-LBP, K-SVD-HOG and PCA-HOG, K-SVD-HOG and K-SVD-LBP, PCA-HOG and PCA-LBP, for pedestrian detection.

3.1. Dimension reduction by PCA

In our implementation, we follow the procedure used in [2,6] to extract the traditional HOG feature and the cell-structured LBP feature from the dataset at first. We use $A = [a_1, a_2, \dots, a_M] \in R^{N \times M}$ to describe the HOG feature or the LBP feature of the training samples, and $B = [b_1, b_2, \dots, b_L] \in R^{N \times K}$ for the testing samples (where M is the number of training samples, K is the number of testing samples, N is the feature dimension). As well known, PCA is the simplest of the true eigenvector-based multivariate analyses. PCA dimension reduction methods can be concisely described as:

$$A = WH \quad (8)$$

where $W \in R^{N \times J}$, and $H \in R^{J \times M}$ is the reduced feature matrix. If the dimension reduction target is J , the projection matrix W is composed of J eigenvector groups corresponding to the J maximal eigenvalues of covariance matrix of A .

3.2. Sparse representation using K-SVD

The K-SVD algorithm is a simple and effective implementation for learning an over-completed dictionary to obtain the sparse representation of the sample. Due to the purpose of K-SVD is to find the common direction of energy distribution of the dataset, the feature set of sparse representation mapped by the over-completed matrix can be used for classification and compression, etc. In this study, firstly, we establish the over-completed dictionary from a portion of training feature set, and then produce projection coefficients for training and testing set in the sparse space spanned by the dictionary. After getting the original feature (HOG and LBP) $A = [a_1, a_2, \dots, a_M] \in R^{N \times M}$ of the training set, We use all the positive feature set to train the dictionary. Given the feature set (HOG or LBP) $A^+ = [a_1, a_2, \dots, a_{1000}] \in R^{N \times M^+}$ of the training sample, thus the number of dictionary elements to train is M^+ , the iterative implementation K-SVD try to find the best dictionary $D \in R^{N \times M^+}$ by solving:

$$\min_{D, S} \{ \|A^+ - DS\|_F^2 \} \quad \text{s.t.} \quad \forall i, \|s_i\|_0 \leq L_0 \quad (9)$$

In this paper, we used the method in [16] to yield the dictionary D . The sparse representations of original feature set can be simply presented as:

$$\hat{A} = A^T D = [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_M] \in R^{M^+ \times M} \quad (10)$$

$$\hat{B} = B^T D = [\hat{b}_1, \hat{b}_2, \dots, \hat{b}_K] \in R^{M^+ \times K} \quad (11)$$

where \hat{a}_i and \hat{b}_i are the sparse coding coefficients.

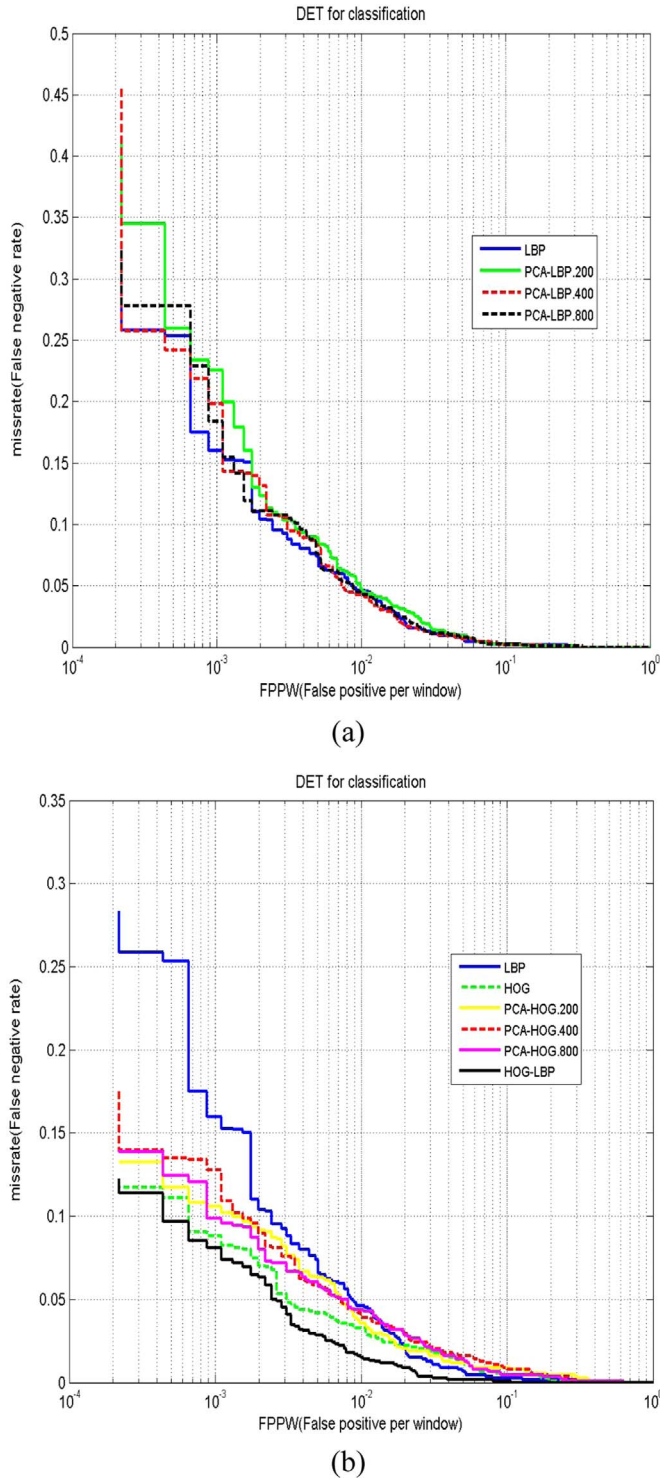


Fig. 3. (a) Performance comparison of PCA-LBP features in different dimensions on the INRIA dataset. The performance of different PCA-LBP dimension is not obviously differences from the original LBP. (b) The performance comparison of HOG feature with different dimensions projected by PCA. The performance of the original HOG is also better than PCA-HOG. The performance of the combined feature, i.e., HOG-LBP is the best.

3.3. Combination of the feature set

As is well known, PCA pursues keeping the maximize data information after dimension reduction. By calculating the variance of the data on the projection direction, PCA can measure the importance of the direction. However, the projection direction often

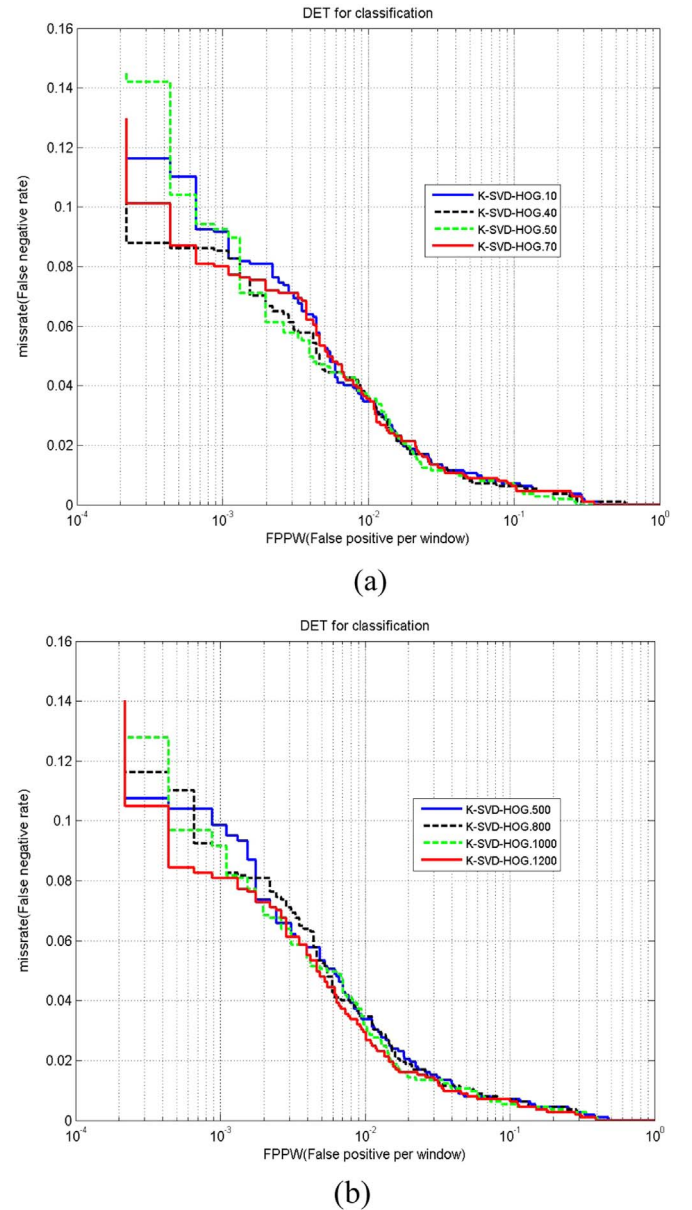


Fig. 4. The DET curves of K-SVD-HOG with different random samples and iterations for training the dictionary. (a) For obtaining the spare representation of HOG feature using K-SVD, we training the dictionary with iterations 10, 30, 40, 70, through the random selected 1000 samples. (b) The performance of K-SVD-HOG with dimension 500, 800, 1000, 1200 respectively, with 40 iterations.

does not benefit for classification. On the contrary, the projection data points may be mixed together and indistinguishable.

K-SVD uses an over-complete dictionary to described the data by sparse linear combinations and could be applied to compression and feature extraction, etc. By combining PCA and K-SVD, we can make full use of the advantages of the two methods. The combined features may be eliminate the redundancy in the feature set and obtain a more stable feature set for classification.

4. Experiment results

For the image dataset, we extract the original histogram feature using HOG and texture feature using cell-structure LBP firstly. Then reducing the dimension of the two feature sets by using PCA respectively, and getting the PCA-HOG and PCA-LBP feature of low dimensions (200/400/800, The performances are shown in Fig. 4(a) and (b)

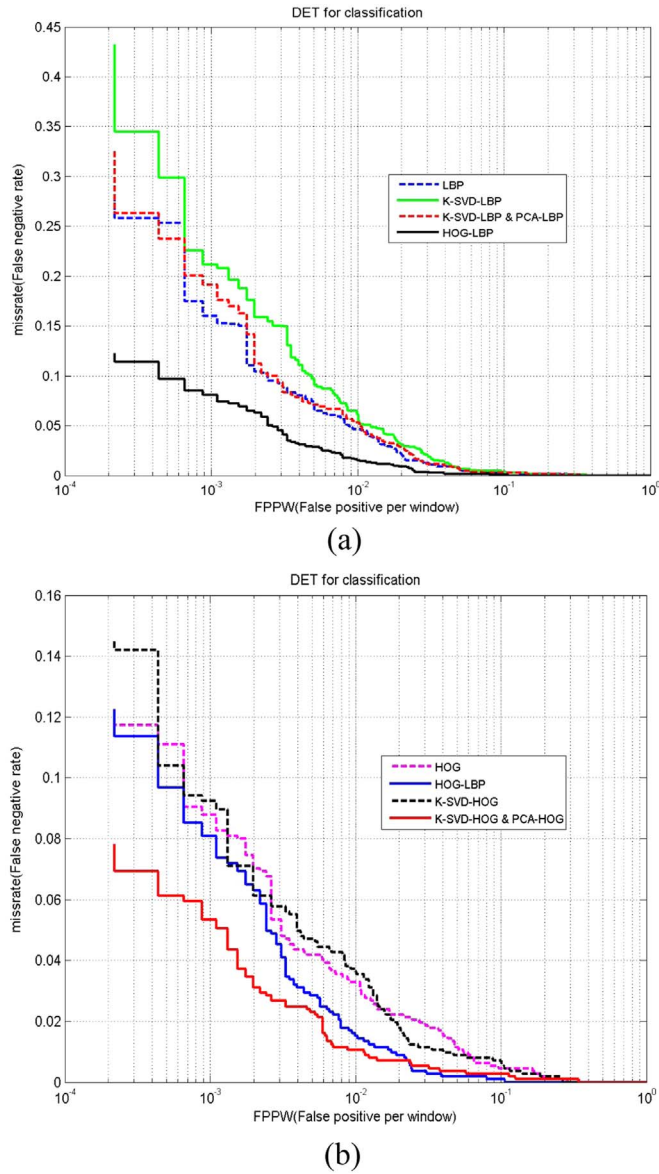


Fig. 5. The DET curves of using K-SVD-LBP and K-SVD-HOG. (a) Compared with the original features LBP, the performance of the K-SVD-LBP is a little bit poorer. The combined K-SVD-LBP and PCA-LBP performance is also not affirmative. (b) The performance of K-SVD-HOG and PCA-HOG is encouraging and better than HOG-LBP.

in the experiments section). Thirdly, we produce sparse representation of the two original feature sets by using K-SVD, named K-SVD-HOG and K-SVD-LBP. At last, we adopt the lib-SVM [1] classifier to classify the samples based on the combined features.

4.1. Experiments on INRIA dataset

In this subsection, we perform the experiments on the INRIA database. This database [4] includes training set (2416 positive samples of fixed pixel and 1218 initial negative samples of various pixel and format) and testing set (1126 positive samples and 453 full negative samples, similar basic information to training set). Except for some partial situations, such as occlusions, the pedestrians are always upright. The state (pose, appearance, clothing, illumination) and background are in large variation range, as shown in Fig. 2. We fed the central 64×128 pixels of the pedestrian images to the descriptor to capture positive features vectors. In this paper, we collected 10 negative feature vectors randomly from each person-free images with a fixed 64×128 window. Classification performances are evaluated on

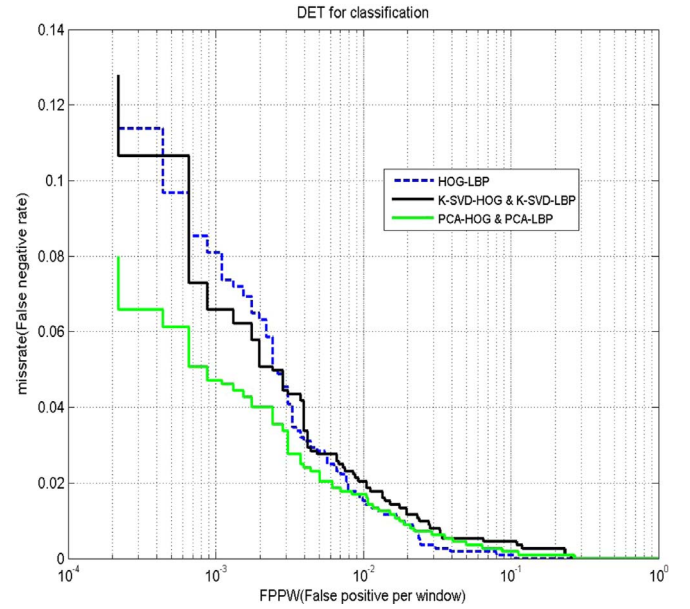


Fig. 6. Miss rate under varying combined features. Experimental results show that the performance of PCA-HOG and PCA-LBP is the best one, and K-SVD-HOG and K-SVD-LBP does not decrease the miss rate obviously.

1126 positive and 4530 negative patches.

In our experiment, the widely used lib-SVM (support vector machine) is trained as the classifier. To quantify the classification performance of the detection algorithm, we choose the Detection Error Tradeoff (DET) curves (a log-log scale, miss rate versus FPPW (False Positives Per Window)) [2] as the criteria. All the experiment results are simulated in matlab2012b environment, computer environment is 2.70 GHz CPU and 6.00 GB internal storage memory.

Firstly, the HOG feature and the LBP feature are extracted. Then we combine the two kind of original features to get the HOG-LBP features directly as the compared group. Thirdly, we reduce the dimensions of the two kind of original features using the PCA and K-SVD respectively. All the experiments are based on these three kinds of features, i.e., original HOG or LBP feature, combined HOG-LBP feature, and the reduced feature by using PCA or KSVD.

4.1.1. Performance on PCA based features

In this subsection, we extract feature with three different dimensions (200/400/800) from the original features (HOG feature and LBP feature) by using PCA (denoted as PCA-LBP, PCA-HOG), and use lib-SVM to classify the dataset. The cell-structure LBP, HOG and HOG-LBP features are used as the compared groups. The relationships between the classification results and the characteristic dimensions are shown in Fig. 3. From Fig. 3(a) we can see that the performances of different dimensions of PCA-LBP features are undistinguished from the original LBP feature. The higher the dimension is, the more similar the performance of PCA-LBP to the original features.

As shown in Fig. 3(b), the performances of the PCA-HOG features are better than the PCA-LBP. We can also find from Fig. 3(b) that, the performance of PCA-HOG with different dimensions are similar. In other words, for the PCA-HOG features, it is not sensitive to the dimensions. However, the performances of PCA-HOG features are not better than the original HOG feature. Moreover, it can be found from Fig. 3(b) that the performance of the combined feature HOG-LBP is the best one.

4.1.2. Performance on K-SVD based features

How to choose the iteration and training samples are two problems need to be solved for training the dictionary for sparse representation by using K-SVD algorithm. To solve these problems, firstly, we

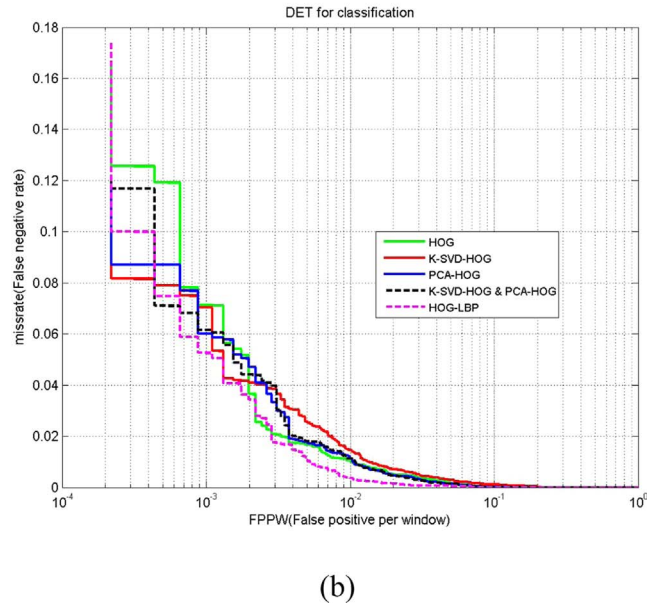
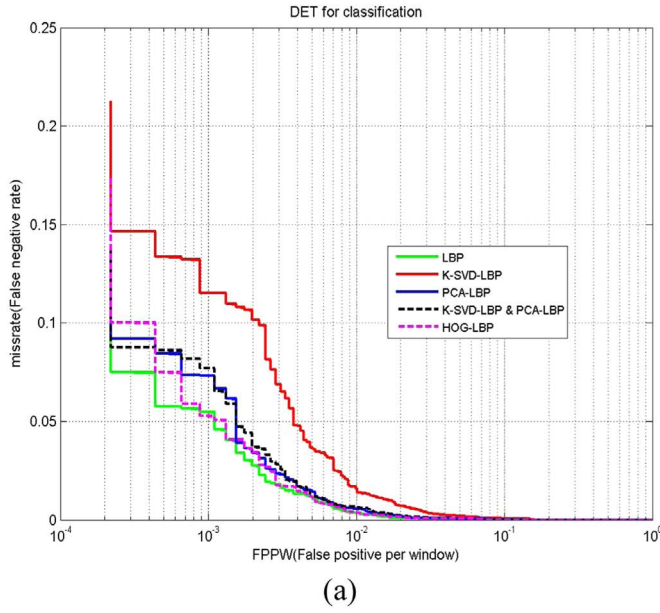


Fig. 7. The DET curves of LBP, HOG, and the extracted features using KSVD and PCA respectively. The performance of HOG-LBP feature is used as the comparison group. (a) Compared with the original LBP feature, the performances of K-SVD-LBP, PCA-LBP and the combined K-SVD-LBP and PCA-LBP are not well. (b) The performances of K-SVD-HOG, PCA-HOG and the combined K-SVD-HOG and PCA-HOG are not distinguished.

calculate the dictionary with 1000 random positive training samples of HOG feature (K-SVD-HOG) through different iterations while keep the total error within the scope of 0.065, as shown in formula (7). In our experiments, we extract the K-SVD-HOG with the training iterations of 10, 30, 40, 70 respectively. As shown in Fig. 4(a), the performances of K-SVD-HOG features with all the different iterations are similar. So in the experiments of this subsection, the iteration number of calculating the dictionary using K-SVD is fixed as 40. We compute the miss rate using the K-SVD-HOG features with dimension 500, 800, 1000 and 1200 respectively, the results are shown in Fig. 4(b). The performance of K-SVD-HOG with 1200 dimension is the best one, but it is not obvious compared with other dimensions. In the following experiments, for easy to compare with the PCA-HOG feature, the dimension of K-SVD-HOG feature is set as 800.

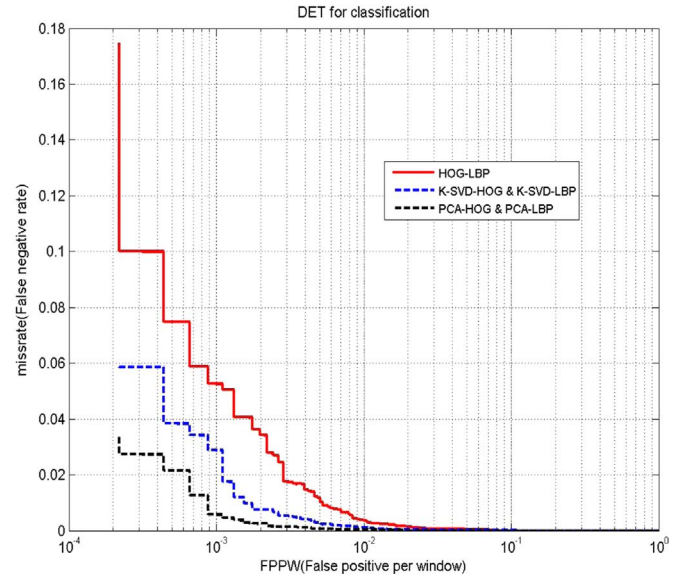


Fig. 8. Miss rate under varying combined features.

Table 1

Statistic comparison of the detection results of different combined features. The processing time is obtained by detecting the INRIA and Daimler testing data with lib-SVM. Miss rate at 10^{-3} FPPW and 10^{-2} FPPW of different groups.

Features	INRIA person dataset			Daimler pedestrian dataset		
	Processing time/s	Miss rate at 10^{-3} FPPW	Miss rate at 10^{-2} FPPW	Processing time/s	Miss rate at 10^{-3} FPPW	Miss rate at 10^{-2} FPPW
HOG	157	0.088	0.033	128	0.071	0.011
PCA-HOG	200	7	0.102	0.039	×	×
× 400	13	0.128	0.043	×	×	×
800	32	0.098	0.044	48	0.060	0.012
LBP	258	0.159	0.052	372	0.055	0.004
PCA-LBP	200	8	0.226	0.050	×	×
L- × 400	13	0.198	0.043	×	×	×
800	37	0.183	0.043	34	0.073	0.007
K-SVD-HOG	55	0.092	0.035	76		
0.07-0	0.015					
K-SVD-LBP	57	0.211	0.065	59	0.115	
0.01-6						
HOG-LBP	533	0.080	0.015	427	0.053	0.004
K-SVD-HOG and K-SVD-LBP	102	0.066	0.010	59	0.029	0.001
PCA-HOG and PCA-LBP	67	0.047	0.016	61	0.007	0.00
K-SVD-HOG and PCA-HOG	93	0.053	0.011	106	0.062	0.011
K-SVD-LBP and PCA-LBP	96	0.191	0.051	81	0.077	0.007

4.1.3. Performance on combined features of different projections

To evaluate the performance of sparse coding techniques based on gradient feature and texture feature, in this subsection, we combined several groups of features for pedestrian detection. The experimental

results are shown in Fig. 5. The computational process of K-SVD-LBP is the same as K-SVD-HOG. The dictionary size is 800 and the train iterations are 40. Fig. 5(a) shows that the performance of K-SVD-LBP is worse than the original LBP. As a comparison group, we do experiment using K-SVD-LBP and PCA-LBP and the performance is also not nice. Fig. 5(b) shows that the performance of K-SVD-HOG is not well. We then investigate whether the simple combined K-SVD-HOG and PCA-HOG can performance better. It is inspiring to see that error detection rate of the combined feature greatly decrease and the performance is better than HOG-LBP feature.

Based on the results of the experiments as shown in Fig. 5, we can draw the following summaries. First, for both the texture and gradient features, the sparse representations can not improve the performances directly. Yet the sparse representation of the gradient feature combined with the PCA based gradient feature can improves the performance greatly. However, it is not suit for cell-structure LBP feature. The combination of different projections of HOG feature makes it more distinguish and even better than the simply combined HOG-LBP feature.

4.1.4. Performance on the combined features of texture and gradients

In this subsection, we research the ability of the combined feature with the same projection of the texture feature and the gradients feature. We compare the performances of the combined K-SVD-HOG and K-SVD-LBP feature and PCA-HOG and PCA-LBP feature. The results are shown in Fig. 6. From this experiment it can be found that the average miss rate between 10^{-3} and 10^{-2} FPPW of K-SVD-HOG and K-SVD-LBP is lower than HOG-LBP. And the performance of PCA-HOG and PCA-LBP feature is the best one. It means that the combined projection feature of HOG and LBP using PCA can improve the performance.

4.2. Experiments on daimler pedestrian dataset

In this subsection, we studied the performance of the combined features on Daimler pedestrian dataset. The Daimler mono pedestrian dataset contains 15,560 pedestrian samples of 48×96 resolution. We select 7830 pedestrian samples randomly as the positive training samples. Additionally, we collected 10 samples randomly from each person-free images of the INRIA training data set with a fixed 48×96 window. Thus 12,180 negative samples along with 7830 pedestrian samples for training. Similarly, we collect 4530 negative samples from the INRIA testing set along with the remaining 7830 pedestrian samples for testing.

The experiment results are shown in Fig. 7. It can be seen that the performance of K-SVD-HOG, K-SVD-LBP, PCA-HOG and PCA-LBP do not performance better than the original HOG and LBP on this dataset. The performance of the combined HOG-LBP is best as studied in [6]. But the performance of the combined K-SVD-HOG and PCA-HOG, K-SVD-HOG and PCA-LBP and PCA-HOG and PCA-LBP demonstrates a tiny better than HOG-LBP. During these experiments, the dimensions of K-SVD-HOG, K-SVD-LBP, PCA-HOG and PCA-LBP are set as 800, and we evaluate the performances of all the features by using lib-SVM as the classifier.

Fig. 8 shows the performances of combined K-SVD-HOG and K-SVD-LBP and PCA-HOG and PCA-LBP. As shown in Fig. 8, similar to the experiment results on INRIA dataset, these two combined features also performing much better than HOG-LBP on Daimler pedestrian dataset. Thus we can conclude that, although K-SVD and PCA do not improve the performances of HOG and LBP under simply dimension reducing form, their combined features are more robust for pedestrian detection.

4.3. Performance evaluation

To further intuitively evaluate the performance of different combi-

nation of features, we also the compared the performance of them statistically under three criteria, i.e., the processing time on testing set by using lib-SVM, the miss rates at 10^{-3} FPPW and 10^{-2} FPPW. The results are shown in Table 1.

From Table 1 we can find that, first, both on the two datasets, the processing times of the original HOG feature are reduced greatly by using K-SVD-HOG and PCA-HOG. And it is the same for cell-structure LBP. In addition, compared to the performances of HOG and LBP features on INRIA dataset and Daimler dataset, the performances of PCA-HOG, K-SVD-LBP, PCA-HOG and PCA-LBP features are undisputedly improved. Second, for INRIA dataset, beside the combined K-SVD-LBP and PCA-LBP, the performance of other three combined feature sets, i.e., K-SVD-HOG and K-SVD-LBP, PCA-HOG and PCA-LBP, K-SVD-HOG and PCA-HOG, are all well. On Daimler dataset, all of the four combined features perform well. Thus, we can conclude that compared with the original HOG-LBP, the combined features are more robust. Moreover, the processing time is decreased drastically.

5. Conclusions

In this paper, we demonstrated the improvement to the well known features, i.e., HOG and LBP, by utilizing K-SVD and PCA to extracted new features and combining them for pedestrian detection. Firstly, we generated the original HOG and LBP features from the images. Then, we used PCA and K-SVD to extract the new features from them. Finally, we combined these new features for pedestrian detection. We test our methods on INRIA and Daimler dataset, the experimental results show that our method is effective and outperform the combined original HOG-LBP feature. In summary, by using effective dictionary, PCA based feature extraction and effective feature combination, we may construct new features with richer structures than the original combined HOG-LBP.

However, how to automatic but effective confirm the number of dictionary and the dimension for pedestrian detection under real-world conditions is still a challenging topic. Moreover, there are plenty of sophisticated implementations for sparse representations and dimension reduction of texture feature and gradient feature. More study should be carried out in future to explore more effective feature extracting and combining method for pedestrian detection.

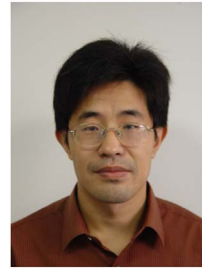
Acknowledgements

This work was supported by the National Science Foundation of China under Grant no. 61272339, and the Natural Science Foundation of Hubei Province under Grant no. 2012FFB04204.

References

- [1] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.*, 2:27:1–27:27, 2011.
- [2] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, *IEEE Conf. Comput. Vis. Pattern Recognit.* (2005) 886–893.
- [3] J. Zhang, K. Huang, Y. Yu, T. Tan, Boosted local structured HOG-LBP for object localization, *IEEE Conf. Pattern Anal. Mach. Intell.* (2011) 1393–1400.
- [4] Y. Mu, S. Yan, Y. Liu, T. Huang, B. Zhou, Discriminative local binary patterns for human detection in personal album, *IEEE Conf. Comput. Vis. Pattern Recognit.* (2008) 1–8.
- [5] W.L. Lu, J.J. Little, Simultaneous tracking and action recognition using the PCA-HOG descriptor, in: *Proceedings of the 3rd Canadian Conference Computer and Robot Vision*, June, 2006, p. 6.
- [6] X. Wang, T.X. Han, S. Yan, An HOG-LBP human detector with partial occlusion handling, *IEEE Conf. Comput. Vis.* (2009) 32–39.
- [7] P. Dollár, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: an evaluation of the state of the art, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (2011) 743–761.
- [8] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns: application to Face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (12) (2006) 2037–2041.
- [9] G. Zhao, M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (6) (2007) 915–928.
- [10] T. Ojala, M. Pietikainen, T. Maenpää, Multiresolution gray-scale and rotation

- invariant texture classification with local binary patterns, *IEEE Trans. Comput. Vis. Pattern Recognit.* 24 (7) (2002) 971–987.
- [11] S. Liao, M.W.K. Law, A.C.S. Chung, Dominant local binary patterns for texture classification, *IEEE Trans. Image Process.* 18 (5) (2009) 1107–1118.
- [12] Z. Guo, L. Zhang, D. Zhang, A Completed modeling of local binary pattern operator for texture classification, *IEEE Trans. Image Process.* 19 (6) (2010) 1657–1663.
- [13] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, *IEEE Trans. Image Process.* 19 (6) (2010) 1635–1650.
- [14] C. Wang, J. Zhou, S. Huang, Motion human detection based on mixture of Gaussians and PCA-HOG, *Appl. Res. Comput.* 29 (6) (2012) 2156–2160.
- [15] C. Zeng, H. Ma, Anlong Ming, Fast human detection using Mi-SVM and a cascade of HOG-LBP features, *IEEE Conf. Image Process.* (2010) 3845–3848.
- [16] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing over-complete dictionaries for sparse representation, *IEEE Trans. Signal Process.* 54 (11) (2006) 4311–4322.
- [17] W. Park, D. Kim, Suryanto, Ch Lyuh, T. Roh, S. Ko, Fast human detection using selective block-based HOG-LBP, *IEEE Conf. Image Process.* (2012) 601–604.
- [18] Q. Zhu, S. Avidan, M.C. Yeh, K.T. Cheng, Fast human detection using a cascade of histograms of oriented gradients, *IEEE Conf. Comput. Vis. Pattern Recognit.* 2 (2006) 1491–1498.
- [19] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, *IEEE Conf. Comput. Vis. Pattern Recognit.* 1 (2001) 511–518.
- [20] H. Schneiderman, Feature-centric evaluation for efficient cascaded object detection, *IEEE Conf. Comput. Vis. Pattern Recognit.* (2004) 29–36.
- [21] C. Conde, D. Moctezuma, I. Martín, E. Cabello, HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments, *Neurocomputing* 100 (2013) 19–30.
- [22] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, *Ann. Stat.* 28 (2) (2000) 337–407.
- [23] C. Zeng, H. Ma, Robust head-shoulder detection by PCA-based multilevel HOG-LBP detector for people counting, *IEEE Conf. Pattern Recognit.* (2010) 2069–2072.
- [24] Y. Luo, C. Wu, Y. Zhang, Facial expression feature extraction using hybrid PCA and LBP, *J. China Univ. Posts Telecommun.* 20 (2) (2013) 120–124.
- [25] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [26] K. Labusch, E. Barth, Ts Martinetz, Soft-competitive learning of sparse codes and its application to image reconstruction, *Neurocomputing* 74 (2011) 1418–1428.
- [27] J. Liu, B. Li, W. Zhang, Feature extraction using maximum variance sparse mapping, *Neural Comput. Appl.* 21 (8) (2012) 1827–1833.
- [28] R. Rubinstein, T. Peleg, M. Elad, Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model, *IEEE Trans. Signal Process.* 61 (3) (2013) 661–677.
- [29] M. Aharon, M. Elad, A.M. Bruckstein, K-SVD and its non-negative variant for dictionary design, *Soc. Photo-Opt. Instrum. Eng. Conf. Ser.* 5914 (2005) 327–339.
- [30] X. Ren, D. Ramanan, Histograms of sparse codes for object detection, *IEEE Conf. Comput. Vis. Pattern Recognit.* (2013) 3246–3253.
- [31] X. Song, Z. Liu, X. Yang, J. Yang, A parameterized fuzzy adaptive K-SVD approach for the multi-classes study of pursuit algorithms, *Neurocomputing* 123 (2014) 131–139.
- [32] B. Li, D.S. Huang, Locally linear discriminant embedding: an efficient method for face recognition, *Pattern Recognit.* 41 (12) (2008) 3813–3821.
- [33] D.S. Huang, C.H. Zheng, Independent component analysis based penalized discriminant method for tumor classification using gene expression data, *Bioinformatics* 22 (15) (2006) 1855–1862.
- [34] B. Li, D.S. Huang, C. Wang, K.H. Liu, Feature extraction using constrained maximum variance mapping, *Pattern Recognit.* 41 (11) (2008) 3287–3294.
- [35] Z.Q. Zhao, D.S. Huang, B.Y. Sun, Human face recognition based on multiple features using neural networks committee, *Pattern Recognit. Lett.* 25 (12) (2004) 1351–1358.



Chun-Hou Zheng received the B.Sc degree in Physics Education in 1995 and the M.Sc. degree in Control Theory and Control Engineering in 2001 from QuFu Normal University, and the Ph.D degree in Pattern Recognition and Intelligent System in 2006, from University of Science and Technology of China. From Feb. 2007 to Jun. 2009 he worked as a Postdoctoral Fellow in the Hefei Institutes of Physical Science, Chinese Academy of Sciences. From Jul. 2009 to Jul. 2010 he worked as a Postdoctoral Fellow in the Dept. of Computing, The Hong Kong Polytechnic University. He is currently an Professor in the College College of Electrical Engineering and Automation, Anhui University, China. His research interests include Pattern Recognition and Bioinformatics.



Wen-Juan Pei received the BSc degree in automation from Anhui University in 2012. She is now a master student in pattern recognition and intelligent system at the College of Electrical Engineering and Automation, Anhui University. Her research interests cover pedestrian recognition and deep learning.



Yan-Wen Chong received the B.S. degree from the Qufu Normal University, China, in 1995, M.S. and Ph.D. degrees from Wuhan University, China. He is an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. His research interests include video processing, intelligent transportation system and pat tern recognition.