

A Sparse Representation Method with Maximum Probability of Partial Ranking for Face Recognition

Yi-Haur Shiau

Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan

Chaur-Chin Chen

Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan

Abstract—Face recognition is a popular topic in computer vision applications. Compressive sensing is a novel sampling technique for finding sparse solutions to underdetermined linear systems. Recently, a sparse representation-based classification (SRC) method based on compressive sensing is presented. It has been successfully applied in face recognition. In this paper, we proposed a maximum probability of partial ranking method based on the framework of SRC, called SRC-MP, for face recognition. Eigenfaces, fisherfaces, 2DPCA and 2DLDA are used for feature extraction. Experiments are implemented on two public face databases, Entended Yale B and ORL. In order to show our proposed method is robust for face recognition in the real world, experiment is also implemented on a web female album (WFA) face database. We utilize AdaBoost method to automatically detect human face from web album images with complex background, illumination variation and image misalignment to construct WFA database. Furthermore, we compare our proposed method with the classical projection-based methods such as principal component analysis (PCA), linear discriminant analysis (LDA), 2DPCA and 2DLDA. The experimental results demonstrate our proposed method not only is robust for varied viewing angles, expressions, and illumination, but also has higher recognition rates than other methods.

Keywords—Compressive sensing; Face recognition; Sparse representation classification; AdaBoost.

I. INTRODUCTION

Face recognition is a hot research area in recent years. Although many papers reported face recognition methods, researchers have focused primarily on projection-based methods rather than other methods [1]. As to the advantages of the projection-based methods, face images are reconstructed promptly and image features are extracted instantly, such as Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3]. Besides, the projection-based methods have achieved high recognition rates for several public face image databases. However, the disadvantage of the linear dimensionality reduction algorithms is that the projections are linear combination of all the original features. Meanwhile, all weighting coefficients in the linear combination are non-zero. Fortunately, compressive sensing theorem [4, 5, 6], a novel sampling technique, is proved to overcome the drawback. According to sparsity principle of compressive sensing, it is possible to recover certain signals and images exactly from far fewer samples of measurements beyond Nyquist rates [7]. A sparse representation-based classification

(SRC) method based on compressive sensing is presented [8, 9]. It has been successfully applied in face recognition. In this paper, we propose a maximum probability of partial ranking method based on the framework of SRC, called SRC-MP, for face recognition. PCA (eigenfaces), LDA (fisherfaces), 2DPCA [10] and 2DLDA [11] are used for feature extraction. Experiments are implemented on three face databases: Extended Yale B, ORL, and web female album (WFA). The images on WFA database are obtained by using AdaBoost [12] to implement human face detection automatically from web album images in the real world. By applying our proposed method, it is robust for face recognition with varied faces in the real world, and we enable to gain the higher recognition rate than classical projection-based methods.

The rest of this paper is organized as follows: Section 2 briefly reviews SRC method [8]. Section 3 proposes our method based on the framework of SRC. Section 4 depicts experiment results and the conclusion is drawn in Section 5.

II. SPARSE REPRESENTATION BASED CLASSIFICATION

A. Compressive Sensing

Compressive sensing is a sampling technique for finding sparse solutions to underdetermined linear systems [7, 8]. A K -sparse signal is a signal that owns at most K nonzero coefficients where $K \ll N$, N is the size of signal. The compressive sensing theorem adopts the sparsity property, and is performed under the following optimization method based on l_1 -norm:

$$\min_{\hat{A} \in \mathbb{R}^N} \|\hat{A}\|_1 \text{ subject to } Y = \Phi^T \hat{A} \quad (1)$$

Where

Y : an observed M -dimensional signal (M -dimensional column vector, $M < N$);

\hat{A} : an N -dimensional sparse signal (a column vector of N components);

$\|\hat{A}\|_1$: the l_1 -norm of \hat{A} ;

Φ : an $N \times M$ sensing matrix.

B. Sparse Representatin-based Calssification

An SRC method based on compressive sensing theorem is provided for face recognition [8]. The basic idea of SRC is to

represent a testing image as a sparse linear combination of all training images. In order to obtain a sparse solution, the feature dimensions must be much smaller than the number of all training images.

Suppose there are K individuals in the face database, and let $B = [B_1, B_2, \dots, B_K]$ be the concatenation of the N training images from all of the K individuals, where $N = n_1 + n_2 + \dots + n_K$. $B_i = [s_1^{(i)}, s_2^{(i)}, \dots, s_{n_i}^{(i)}] \in \mathbf{R}^{m \times n_i}$, is the set of training images of the i^{th} individual, where $s_j^{(i)}$, $j = 1, 2, \dots, n_i$, is an m -dimensional vector stretched by the j^{th} image of the i^{th} individual. A new testing image $\mathbf{y} \in \mathbf{R}^m$ of the i^{th} individual could be represented as a linear combination of the training images in B_i , i.e. $\mathbf{y} = \sum_{j=1}^{n_i} \alpha_j^{(i)} s_j^{(i)} = B_i \boldsymbol{\alpha}^{(i)}$, where $\boldsymbol{\alpha}^{(i)} = [\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_{n_i}^{(i)}]^T \in \mathbf{R}^{n_i}$ are weighting coefficients. Let $\mathbf{y} = B\boldsymbol{\alpha}$ represent the testing image \mathbf{y} by using B , where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}^{(1)}; \boldsymbol{\alpha}^{(2)}; \dots; \boldsymbol{\alpha}^{(K)}]$. Due to \mathbf{y} belongs to the i^{th} individual and $\mathbf{y} = B_i \boldsymbol{\alpha}^{(i)}$, only the coefficients in $\boldsymbol{\alpha}^{(i)}$ have significant values in a noiseless case to $\boldsymbol{\alpha}$, and all the coefficients in $\boldsymbol{\alpha}^{(j)}$, $j=1,2,\dots,K$ and $j \neq i$, are nearly zero. The SRC algorithm is listed as follows [8].

- 1) Normalize the columns of B to have unit l_2 -norm.
- 2) Solve the following l_1 -norm minimization problem:

$$\hat{\boldsymbol{\alpha}}_1 = \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|B\boldsymbol{\alpha} - \mathbf{y}\|_2 \leq \varepsilon. \quad (2)$$
- 3) Compute the residuals

$$r_i(\mathbf{y}) = \|\mathbf{y} - B\delta_i(\hat{\boldsymbol{\alpha}}_1)\|_2 \text{ for } i = 1, \dots, K. \quad (3)$$
- 4) Get output result by $\text{identity}(\mathbf{y}) = \arg\{\min r_i(\mathbf{y})\}$.

III. MAXIMUM PROBABILITY OF PARTIAL RANKING METHOD

In the noiseless case, all the non-zero coefficients of $\hat{\boldsymbol{\alpha}}_1$ will completely be associated with the columns in B from a single individual. The testing image \mathbf{y} can be easily assigned to the correct individual. As to the noise case, however, these non-zero weighting coefficient are not concentrated on any one individual and instead spread widely across the entire training set. \mathbf{y} is difficult to be represented as which one individual. Some classifiers are used to solve this problem. An SRC method classifies \mathbf{y} to one individual by minimizing the residuals.

A simple, rapid method classifies \mathbf{y} by using only the largest weighting coefficient value of $\hat{\boldsymbol{\alpha}}_1$, called SRC-LV. However, such heuristics do not harness the subspace structure associated with images in face recognition. In this paper, we propose a maximum probability of partial ranking method as a classifier, called SRC-MP. It is found by experiments that the largest weighting coefficient may not belong to the correct individual, however, the first largest weighting coefficients concentrate mostly on the correct individual.

Thus, we convert and normalize the weighting coefficient $v_j^{(i)}$ into the probability value $p_j^{(i)} = \frac{v_j^{(i)}}{\sum_{i=1}^K \sum_{j=1}^{n_i} v_j^{(i)}}$, where $v_j^{(i)}$ is the j^{th} non-zero coefficient greater than zero of the i^{th} individual of $\hat{\boldsymbol{\alpha}}_1$. Then, we assign a partial ranking value γ (first largest coefficients), and sum up these largest γ coefficients to obtain a new probability value for each of the individuals, respectively. Moreover, we employed the new maximum probability as the classifier. Figure 1 shows the weighting coefficients of the

testing image \mathbf{y} . The green box represents the correct individual that \mathbf{y} should be assigned, and the blue box represents the wrong individual.

By using SRC-LV classifier, the testing image \mathbf{y} is classified to the wrong individual due to the blue box has the largest weighting coefficient. SRC-MP uses the maximum probability of partial ranking as the new weighting value. The maximum probability value of the green box is larger than that of the blue box, and \mathbf{y} is classified to the correct individual.

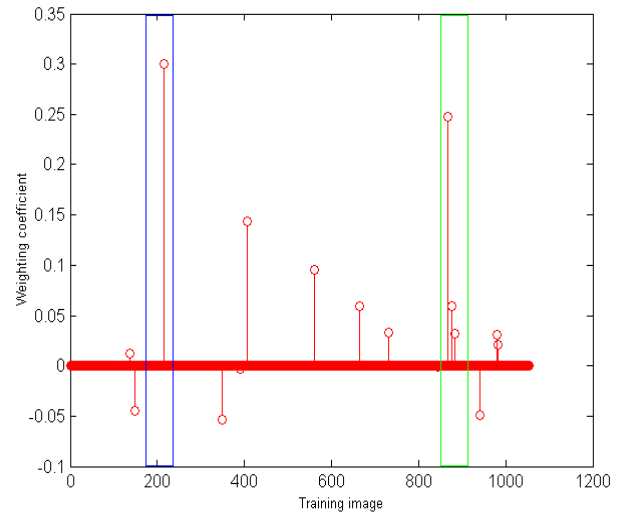


Figure 1. Weighting coefficients of the testing image \mathbf{y} . The green box indicates the correct individual that \mathbf{y} is assigned, and blue box indicates the wrong individual.

The complete method we proposed is summarized as below,

- 1) Set $B_i = [s_1^{(i)}, s_2^{(i)}, \dots, s_{n_i}^{(i)}] \in \mathbf{R}^{m \times n_i}$ as a matrix of the training images for K individuals, and a testing image $\mathbf{y} \in \mathbf{R}^m$, as input data.
- 2) Solve the l_1 -norm minimization problem.

$$\hat{\boldsymbol{\alpha}}_1 = \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|B\boldsymbol{\alpha} - \mathbf{y}\|_2 \leq \varepsilon.$$
- 3) Compute the probability value $p_j^{(i)} = \frac{v_j^{(i)}}{\sum_{i=1}^K \sum_{j=1}^{n_i} v_j^{(i)}}$ for all non-zero values greater than zero.
- 4) Assign a partial ranking value γ , and compute new probability value for each individual $w_i(\mathbf{y})$, respectively.
for $k \leq \gamma$, $w_i(\mathbf{y}) = w_i(\mathbf{y}) + p_k^{(i)}$ for $i = 1, \dots, K$, where $p_k^{(i)}$ is the k^{th} largest probability value that belongs to the i^{th} individual.
- 5) Label \mathbf{y} by $\text{identity}(\mathbf{y}) = \arg\{\max_i w_i(\mathbf{y})\}$.

IV. EXPERIMENTAL RESULTS

We evaluate the performance of our proposed method (SRC-MP) on Extended Yale B [13] and ORL [14] face databases, and PCA (eigenfaces), LDA (fisherfaces), 2DPCA and 2DLDA are used for feature extraction, respectively. We compare SRC-MP with classical projection-based methods such as PCA, LDA, 2DPCA and 2DLDA that adopt the nearest decision rule as the classifier. We also compare the recognition rates with different parameters γ introduced in Section 3. In

order to show SRC-MP method is robust for face recognition in the real world, experiment is also implemented on WFA face database.

A. Extended Yale B Face Image Database

The Extended Yale B database has about 2,500 images of 39 different individuals. We use 34 individuals because there are some images missing. Our database consists of 2,108 face cropped and normalized images of 192 rows and 168 columns in PGM file format. There are 34 persons individually contributed 62 frontal-images by capturing under various laboratory-controlled lighting conditions. The first 10 images of individual 1 are shown in Figure 2. As for each subject, 31 images for training and the rest 31 images for testing are randomly selected.



Figure 2. The 10 face images of the 1th individual on the Extended Yale B face database.

We compute the recognition rates with the feature space dimensions $d = 20, 30, 60, 120, 150$, respectively. For SRC-MP method, we assign the partial ranking value $\gamma = 10$. Table I shows the recognition rates of all methods: (1) PCA, (2) Eigen + SRC-LV, (3) Eigen + SRC-MP, (4) LDA, (5) Fisher + SRC-LV and (6) Fisher + SRC-MP.

The recognition rates of SRC-LV are higher than classical projection-based methods, and SRC-MP obtains higher recognition rates than SRC-LV. In particular, the bold values indicate the best recognition rate accomplished by our proposed method. The curves of recognition rate versus the dimension of features are illustrated in Figure 3.

TABLE I. THE RECOGNITION RATES (%) OF ALL METHODS ON THE EXTENDED YALE B DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 20	d = 30	d = 60	d = 120	d = 150
(1)	51.04	59.58	70.59	76.85	77.61
(2)	79.13	89.47	93.26	94.97	95.73
(3)	80.74	90.99	94.40	95.92	96.20
(4)	92.60	94.59	92.88	88.99	89.66
(5)	93.55	94.59	96.77	96.58	96.68
(6)	94.59	95.16	97.25	97.25	97.34

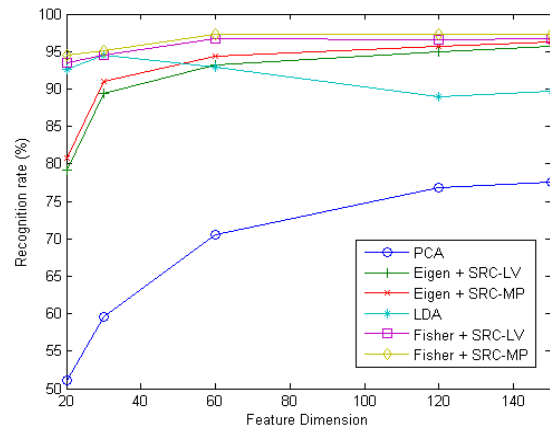


Figure 3. Recognition rates of all methods versus feature dimension on Extended Yale B database.

We also compare SRC-MP with 2DPCA and 2DLDA. Due to the feature dimension must be smaller than the number of training samples, we convert the images on Extended Yale B database into the size of 84x96. We compute the recognition rates with the feature space dimensions $96 \times d$ where $d = 2, 3, 4$, respectively. Table II shows the recognition rates of all methods: (1) 2DPCA, (2) 2DPCA + SRC-MP ($\gamma = 10$), (3) 2DLDA, (4) 2DLDA + SRC-MP ($\gamma = 10$). The curves of recognition rate versus the feature dimensions are illustrated in Figure 4.

TABLE II. THE RECOGNITION RATES (%) OF ALL METHODS VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 2	d = 3	d = 4
(1)	59.01	64.52	66.89
(2)	95.45	95.54	95.16
(3)	83.02	81.97	81.02
(4)	95.83	95.92	96.02

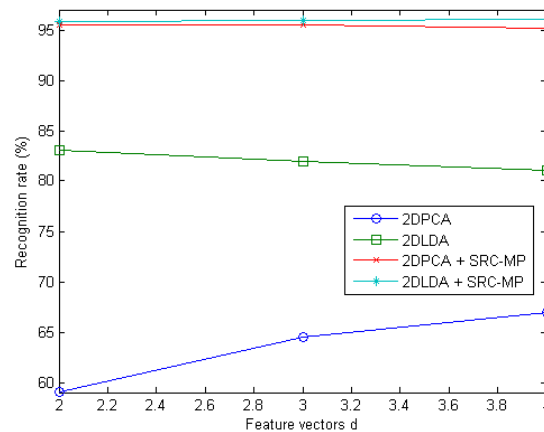


Figure 4. Recognition rates of all methods versus feature dimension on Extended Yale B database.

For SRC-MP method, a partial ranking value γ needs to be assigned to compute the new maximum probability value. We compare the recognition rates with different parameter γ values. Table III shows the recognition rates of different γ values. The curves of recognition rate versus the different γ values are illustrated in Figure 5. It shows that the larger the parameter γ , the higher the recognition rate when γ is in a certain range.

TABLE III. THE RECOGNITION RATES (%) OF DIFFERENT γ VALUES ON THE EXTENDED YALE B DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 20	d = 30	d = 60	d = 120	d = 150
$\gamma = 0$	79.13	89.46	93.26	94.97	95.73
$\gamma = 5$	80.27	90.89	94.21	95.64	96.20
$\gamma = 10$	80.74	90.99	94.40	95.92	96.20
$\gamma = 20$	80.65	90.99	94.50	96.02	96.77

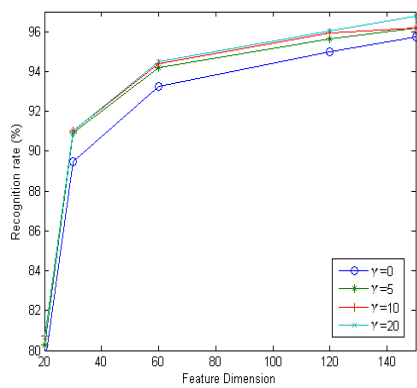


Figure 5. Recognition rates with different parameters γ on Extended Yale B database.

B. ORL Face Image Database

The ORL face database contains 400 8-bit gray level images of 112 rows and 92 columns in PGM file format. There are 40 persons individually contributed 10 images at different times, lightings, facial expressions, and some details on face. The 10 images of individual 17 are shown in Figure 6. As for each individual, the first 5 images for training and the next 5 images for testing were selected.



Figure 6. The face images of the 17th individual on the ORL face database.

We compute the recognition rates with the feature space dimensions $d = 16, 30, 60$, respectively. Table IV shows the recognition rates of all methods: (1) PCA, (2) Eigen + SRC-MP

($\gamma = 10$), (3) LDA and (4) Fisher + SRC-MP ($\gamma = 10$). The bold values indicate the best recognition rate accomplished by our proposed method. The curves of recognition rate versus the dimension of features are illustrated in Figure 7.

TABLE IV. THE RECOGNITION RATES (%) OF ALL METHODS ON ORL DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 16	d = 30	d = 60
(1)	83.0	87.5	89.0
(2)	87.0	89.0	90.0
(3)	88.0	86.0	89.5
(4)	90.0	91.5	90.0

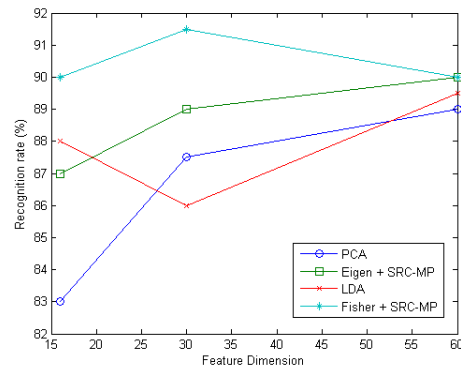


Figure 7. Recognition rates of all methods versus feature dimension on ORL database.

C. Web Female Album Face Image Database

Face detection in the real world is a difficult problem because faces are non-rigid objects and the complex background. In this paper, we collect some web female albums images with complex background, illumination variation, and image misalignment. AdaBoost method [12] is adopted to automatically detect human face on these images to construct WFA face image database. Figure 8 shows the workflow of the WFA database construction.



Figure 8. Workflow of the WFA database construction.

The WFA database contains 1320 images of 192 rows and 168 columns in JPEG file format. There are 33 persons individually contributed 40 images with varied viewing angles, expressions, and illumination. As for each individual, 20 images for training and the rest 20 images for testing are randomly selected. We compute the recognition rates with the feature space dimensions $d = 12, 16, 20, 30, 40$, respectively. Table V shows the recognition rates of all methods: (1) Eigen + SRC-LV, (2) Eigen + SRC-MP ($\gamma = 10$), (3) Fisher + SRC-LV and (4) Fisher + SRC-MP ($\gamma = 10$). The bold values indicate the best recognition rate accomplished by our proposed method. The curves of recognition rate versus the dimension of features are illustrated in Figure 9.

TABLE V. THE RECOGNITION RATES (%) OF ALL METHODS ON WAF DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 12	d = 16	d = 20	d = 30	d = 40
(1)	59.70	69.39	76.06	78.33	81.36
(2)	60.91	73.18	78.48	80.15	84.24
(3)	56.52	64.24	71.21	74.39	76.82
(4)	59.24	68.64	75.00	80.45	81.67

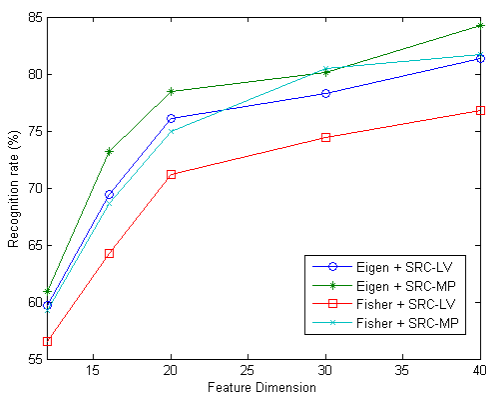


Figure 9. Recognition rates by different methods versus feature dimension on WFA database.

V. CONCLUSIONS

In this paper, we presented a maximum probability of partial ranking method (SRC-MP) based on the framework of sparse representation-based classification (SRC) for face recognition. PCA (eigenfaces), LDA (fisherfaces), 2DPCA and 2DLDA are utilized for feature extraction. We compared our proposed method with classical projection-based methods such as PCA, LDA, 2DPCA and 2DLDA. Our experiments on Extended Yale B and ORL face databases demonstrated that our proposed method achieves higher recognition rate than classical projection-based methods under the same

dimensionality. The experimental result showed that SRC-MP can obtain higher recognition rate than SRC-LV in all cases. It also demonstrated that the larger the parameter γ , the higher the recognition rate when γ is in a certain range. We also adopted AdaBoost method to automatically detect human face from web female album images in the real world to construct a web female album (WFA) face database. The experimental was implemented on WFA database for face recognition in the real world. The experimental result showed it was robust for varied viewing angles, expressions, and illumination, and enabled to achieve high recognition accuracy.

REFERENCES

- [1] C.C. Chen, Y.S. Shieh, and H.T. Chu, "Face image retrieval by projection-based features," The 3rd International Workshop on Image Media Quality and its Applications, pp. 138-144, 2008.
- [2] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991.
- [3] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers, "Fisher Discriminant Analysis with Kernels," IEEE International Workshop on Neural Networks for Signal Processing, vol.9, pp. 41-48, 1999.
- [4] E.J. Candes and M.B. Wakin, "An Introduction to Compressive Sampling," IEEE Signal Processing Magazine, vol.5, no.2, pp. 21-30, 2008.
- [5] M. Stojnic, "L1 Optimization and its Various Thresholds in Compressed Sensing," International Conference on Acoustics, Signal and Speech Processing, pp. 3910-3913, 2010.
- [6] Q. Zhang and B. Li, "Joint Sparsity Model with Matrix Completion for an Ensemble of Face Images," IEEE International Conference on Image Processing, pp. 1665-1668, 2010.
- [7] C. Moler, "'Magic' reconstruction: compressive sensing," Cleves Corner, Mathworks News&Notes, pp. 1-4, 2010, <http://www.mathworks.com>
- [8] J. Wright, A.Y. Yang, and A. Ganesh, "Robust Face Recognition via Sparse Representation," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.31, no.2, pp. 210-227, 2009.
- [9] L. Zhang, M. Yang, Z. Feng, and D. Zhang, "On the Dimensionality Reduction for Sparse Representation based Face Recognition," International Conference on Pattern Recognition, pp. 1237-1240, 2010.
- [10] J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.1, pp. 131-137, 2004.
- [11] J. Yang, D. Zhang, and J.Y. Yang, "Two-dimensional discriminant transform for face recognition," Pattern recognition, vol.38, pp. 1125-1129, 2005.
- [12] L. Ming, and Y. Baozong, "2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix," Pattern Recognition Letters, vol.26, pp.527-532, 2005.
- [13] P. Viola, and M. J. Jones, "Robust Real-Time Face Detection. Internal," Jour. Computer Vision, vol.57, no.2, pp. 137-154, 2004.
- [14] Extended Yale B face database: <http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html>, last access on Mar. 6, 2012.
- [15] ORL face database: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>, last access on Mar. 6, 2012.