Information **2015**, *6*, 152-161; doi:10.3390/info6020152

OPEN ACCESS

information

ISSN 2078-2489

www.mdpi.com/journal/information

Article

Graph Regularized Within-Class Sparsity Preserving Projection for Face Recognition

Songjiang Lou *, Xiaoming Zhao, Wenping Guo and Ying Chen

Institute of Image Processing & Pattern Recognition, Tai Zhou University, Taizhou 318000, China;
E-Mails: tzxyzxm@yahoo.com.cn (X.Z.); guowp@tzc.edu.cn (W.G.); ychen222@tzc.edu.cn (Y.C.)

* Author to whom correspondence should be addressed; E-Mail: lousongjiangac@163.com;
Tel.: +86-576-85137063.

Academic Editor: Willy Susilo

Received: 18 December 2014 / Accepted: 10 April 2015 / Published: 24 April 2015

Abstract: As a dominant method for face recognition, the subspace learning algorithm shows desirable performance. Manifold learning can deal with the nonlinearity hidden in the data, and can project high dimensional data onto low dimensional data while preserving manifold structure. Sparse representation shows its robustness for noises and is very practical for face recognition. In order to extract the facial features from face images effectively and robustly, in this paper, a method called graph regularized within-class sparsity preserving analysis (GRWSPA) is proposed, which can preserve the within-class sparse reconstructive relationship and enhances separatability for different classes. Specifically, for each sample, we use the samples in the same class (except itself) to represent it, and keep the reconstructive weight unchanged during projection. To preserve the manifold geometry structure of the original space, one adjacency graph is constructed to characterize the interclass separability and is incorporated into its criteria equation as a constraint in a supervised manner. As a result, the features extracted are sparse and discriminative and helpful for classification. Experiments are conducted on the two open face databases, the ORL and YALE face databases, and the results show that the proposed method can effectively and correctly find the key facial features from face images and can achieve better recognition rate compared with other existing ones.

Keywords: dimensionality reduction; sparse representation; graph embedding; face recognition

1. Introduction

Face recognition is an important but complicated problem in computer vision. It has broad applications ranging from computer interface to surveillance. Many algorithms have been proposed in literature, including two-dimensional face recognition and three-dimensional face recognition methods [1–4]. Three-dimensional face recognition, which tries to use 3D geometry of the face for identification, proves to be more robust to illumination, pose and disguise. However, the problem of facial expressions is a major issue in 3D face recognition, since the geometry of the face significantly changes with different facial expressions. Most of the images can be seen as two-dimensional matrices, so 2D face recognition also received tremendous attention in computer vision and pattern recognition. Subspace learning methods, such as principle component analysis (PCA) [5] and linear discriminant analysis (LDA) [6], have been extensively studied. Both of them seek to find the low-dimensional representation for the original high-dimensional data, and to preserve some kind of intrinsic structure.

PCA is an unsupervised method and the projections are obtained by maximizing the total scatter of the data. While LDA is a supervised method and it tries to maximize the ratio of between-class scatter to within-class scatter. Experiments show that LDA outperforms PCA in face recognition. However, it is reported that the face images probably reside in some sort of manifolds [7]. One problem of these two algorithms is that they only exploit the linear global Euclidean structure and ignores the local geometry structure. Although they have been extended to nonlinear methods like KPCA [8] and KLDA [9] by kernel trick, it is hard to choose a perfect kernel function and the computation is expensive.

Manifold learning tries to find an embedding that projects the high dimensional data onto low dimensional data while preserving the intrinsic geometry of data, especially the local geometry. The representatives are Isomap [10], locally linear embedding (LLE) [11] and Laplacian eigenmaps (LE) [12]. However, the manifold learning algorithms are affected by two critical problems [13]: (i) the construction of the adjacency graph, (ii) the embedding of new test data, which is also called the out of sample problem. As for the later problem, He proposed a linear method named locality preserving projections (LPP) [14] to approximate the eigenfunctions of the Laplace–Beltrami operator on the manifold, that is to say, LPP is a linear version of LE. By considering the local information and class label information, many variants [15–18] were proposed and can achieve good performance. One critical step in these methods is to construct the adjacency graph; however, how to define the sparse adjacency weight matrices is still an open problem.

Traditional method for adjacency graph is to use the $k-nearest$ neighborhood graph or ε-neighborhood graph. However, these two methods are all parametric and the performances of the algorithms are highly dependent on the choice of its parameter. In [19], it is reported that the adjacency graph structure and graph weights are highly interrelated and should not be separated. So, it is desirable to design a method that can simultaneously carry out these two tasks in one step. To this end, recently two l_1 graph construction methods [20,21] have been proposed, which complete the adjacency graph and graph weight calculation within one step.

Recently, the sparse representation (SR) [22] has been extensively studied and found wide applications in computer vision and image processing problems. The main idea of SR is that a query image can be sparsely represented as a linear combination of all the training samples, its non-zero representation coefficients are naturally sparse and the representations are mostly from the same class of the query image, SR is an unsupervised method but it exploits the discriminant nature of sparse representation for classification. Based on this idea, Qiao proposed sparsity preserving projection (SPP) [23] for feature extraction, which tries to preserve the sparse reconstructive relationship of samples in the low-dimensional data by minimizing the distance between sparsely reconstructed samples and the original sample. However, there are still some issues to be solved. First, SPP is an unsupervised method and does not make use of the class information. Second, when the dictionary is large, SPP is very time-consuming.

To this end, in this paper, a method called graph regularized within-class sparsity preserving projection analysis (GRWSPA) is proposed, which aims at preserving the within-class sparse reconstructive relationship by minimizing the distance between sparsely reconstructed samples in the same class (within class) and their corresponding original samples like SPP, which can reduce the computation time, as the number of samples in each class is usually much smaller than the total number of training samples. At the same time, by assuming samples in different classes lie on different sub-manifolds, it tries to maximize the scatter of inter-class samples by constructing a between-class adjacency graph, and pulls samples from different classes as far as possible.

The rest of the paper is organized as follows. In Section 2, SPP is briefly reviewed. The proposed algorithm is presented in Section 3. In Section 4, experiments are carried out to evaluate the proposed algorithm. Finally, the conclusions are drawn in Section 5.

2. Sparsity Preserving Projections

Let $X = \{x_1, x_2, \cdots, x_n\}, x_i \in R^D, i = 1, 2, \cdots, n$ be the training samples. In real applications, the dimensionality D is often very high. One reasonably way is to use dimensionality reduction to map the data from the high-dimensional space to a low dimensional one, which can be expressed mathematically as $y_i = A^T x_i \in R^d$ for any x_i, usually $d << D$, here A is called the transformation matrix.

The idea of SPP is that every sample in the training samples can be represented as a linear combination of the remaining samples. That is, $x_i = X\alpha_i$, here α_i has the form of $\alpha_i = \{a_1, a_2, \cdots, a_{i-1}, 0, a_{i+1}, \cdots, a_n\}$, and most elements of α_i are zero. This can be formulated as

$$\begin{aligned} &\min_{\alpha} \|\alpha_i\|_0 \\ s.t.\quad & x_i = X\alpha_i \end{aligned} \tag{1}$$

where $\|\alpha_i\|_0$ denotes the l_0 norm, meaning the number of non-zero entries in α_i. However, this problem is NP-hard. If α_i is sparse enough, the above optimization can be replaced as

$$\begin{aligned} &\min_{\alpha} \|\alpha_i\|_1 \\ s.t.\quad & x_i = X\alpha_i \end{aligned} \tag{2}$$

This can be solved by standard convex programming method [24]. Suppose $\tilde{\alpha}_i$ is the optimal solution to the above optimization, SPP then tries to preserve the sparse reconstruction relationship, which can be expressed as the following optimization:

$$\begin{aligned} &\min \sum_{i}^{n} \left\| A^T x_i - A^T X \tilde{\alpha}_i \right\| \\ &s.t. \quad A^T X X^T A = I \end{aligned} \tag{3}$$

which can be simplified by simple algebra:

$$\begin{aligned} &\min \sum_{i}^{n} \left\| A^T x_i - A^T X \tilde{\alpha}_i \right\| \\ &= A^T \sum_{i=1}^{n} (x_i - X\tilde{\alpha}_i)(x_i - X\tilde{\alpha}_i)^T A \\ &= A^T X (\sum_{i=1}^{n} (e_i - \tilde{\alpha}_i)(e_i - \tilde{\alpha}_i)^T) X^T A \\ &= A^T X (I - S - S^T + S^T S) X^T A \end{aligned} \tag{4}$$

So the optimal projections are the eigenvectors of the following generalized eigenvalue problem

$$X(I - S - S^T + S^T S) X^T A = \lambda X X^T A \tag{5}$$

where $S = \{\tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_n\}$.

3. Graph Regularized Within-Class Sparsity Preserving Analysis

From the above section, we can see that SPP is an unsupervised method, and does not use the label information properly. Moreover, the sparse representations are obtained from the whole training samples. If the number of training samples is large, the process is very computation-expensive. In this section we will present an improved SPP algorithm.

3.1. Preserve the Sparsity Structure for Within-Class Samples

In sparse representation, a test sample x_i can be represented as a linear combination of all training samples, the non-zero sparse representation coefficients w_i^j can reflect the contribution of x_j while reconstructing x_i. The higher value w_i^j is, the more similar x_j and x_i are, and are supposed to concentrate on the training samples within the same class as the test sample. While the small value w_i^j means that x_j has little contribution for reconstructing x_i, and is probably from different classes. However, SPP does not consider the class information, and its adjacency graph weights are based on sparse representation and take the whole training samples as dictionary. However, it is very time-consuming if the number of training samples is large. One solution to this problem might be that we can take the samples in the same class as the dictionary to reconstruct x_i, like SPP, it can be represented as:

$$\begin{aligned} &\min \left\| w_{k,i} \right\|_1 \\ s.t. \quad &x_{k,i} = X_{k.i} w_{k,i} \\ &\sum_j w_{k.i}^j = 1 \end{aligned} \tag{6}$$

where $x_{k,i}$ is denoted as the i^{th} sample in the k^{th} class, $i = 1, 2, ..., n_k$, $k = 1, 2, ..., c$, here n_k means the number of samples in the k^{th} class, c means the number of classes. $X_{k.i}$ denotes the whole samples in the k^{th} class. The sparse representation coefficients $w_{k,i}$ have the form of $w_{k,i} = (w_{k,i}^{1}, w_{k,i}^{2}, w_{k,i}^{3}, \cdots, w_{k,i}^{i-1}, 0, w_{k,i}^{i+1}, \cdots, w_{k,i}^{n_k})^T$. Suppose $\tilde{\tilde{w}}_{k,i}^{'} = (0, 0, 0, \cdots, 0, w_{k,i}^{T}, 0, \cdots, 0)^T$, then the weight matrix has the form of $W = (w_{1,1}^{'} \cdots, w_{1,N_1}^{'}, w_{2,1}^{'} \cdots, w_{2,N_2}^{'} \cdots w_{c,1}^{'} \cdots, w_{c,n_c}^{'})$.

Like SPP, we hope that the sparse structure can be well preserved, which can be solved by the following formulation:

$$\begin{aligned} &\min \sum_{k=1}^{c} \sum_{i=1}^{n_k} \left\| A^T x_{k,i} - A^T X w_{k,i}^{'} \right\|^2 \\ s.t. \quad & A^T X X^T A = I \end{aligned} \tag{7}$$

The above optimization can be reduced to the following problem:

$$\begin{aligned} &\max tr(A^T X S X^T A) \\ s.t. \quad & A^T X X^T A = I \end{aligned} \tag{8}$$

where $S = W + W^T - W^T W$.

3.2. Discover the Discriminant Structure for between Class Samples

It is supposed that samples from different classes lie on different sub-manifolds; one reasonable way for classification is to map these sub-manifolds as far as possible. We construct an adjacency graph $G = (X, B)$ over the training data X to characterize the relationship for different classes. The elements of the weight matrix B can be defined as follows:

$$B_{ij} = \begin{cases} (1 - \dfrac{x_i \cdot x_j}{\|x_i\| \|x_j\|}) & \text{if } \mathrm{x_i} \text{ and } \mathrm{x_j} \text{ are k nearest neighbors but have different labels} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

In order to guarantee the discriminant ability in low dimensional representation, like Unsupervised Discriminant Projection (UDP) [25], we hope that the connected points in the adjacency graph should stay as distant as possible, which can be expressed as the following optimization:

$$\max \frac{1}{2nn} \sum_{ij} (y_i - y_j)^2 B_{ij} \tag{10}$$

where y_i is the low dimensional representation of x_i . The above objective incurs a heavy penalty if nearby points x_i and x_j are mapped close while they are belonging to different classes, which is an attempt to ensure that if points x_i and x_j are close but are from different classes, then y_i and y_j are far apart, which can encode the local discriminant information and helpful for classification.

We can simplify the above optimization as follows:

$$J_B = \frac{1}{2nn} \sum_{ij}^{n} (y_i - y_j) B_{ij} = \frac{1}{nn} A^T S_B A \tag{11}$$

where S_B is called the Laplacian difference scatter matrix.

$$\begin{aligned} S_B &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}(x_i - x_j)(x_i - x_j)^T \\ &= \frac{1}{2}(\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}x_i x_i^T - 2\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}x_i x_i^T + \sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}x_j x_j^T) \\ &= \sum_{i=1}^{n} D_{ii}x_i x_i^T - \sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}x_j x_j^T \\ &= XDX^T - XBX^T \\ &= XLX^T \end{aligned} \tag{12}$$

where D is a diagonal matrix, as $D_{ii} = \sum_j B_{ij}$, $L = D - B$ is the Laplacian matrix.

3.3. GRWSPA

To take the within-class reconstruction relationship and between-class separability into account, it is desirable to keep the reconstruction weights in the same class as SPP while maximize the local discriminant information. By combining 3.1 and 3.2, it can easily form the following optimization

$$\begin{aligned} &\max\ tr(A^T XSX^T A) + \mu tr(A^T XLX^T A) \\ s.t.\quad & A^T XX^T A = I \end{aligned} \tag{13}$$

where μ is a factor to balance the sparse representation and the discriminant ability.

For compact expression, the maximization problem can further be transformed to the following problem:

$$\max_{A} \frac{A^T XSX^T A + \mu A^T XLX^T A}{A^T XX^T A} \tag{14}$$

Then the optimal A is the eigenvectors corresponding to the largest d eigenvalues of the following generalized eigenvalue problem:

$$(XSX^T + \mu XLX^T)A = \lambda XX^T A \tag{15}$$

4. Experimental Section

In this section, several experiments are carried out to show the effectiveness of the proposed algorithm on the ORL and YALE databases. We compare our method with some classic methods including LDA, LPP, UDP and SPP. For classification, we use the nearest neighbor classifier for its easy implementation. There is a parameter μ, here we set it to $\mu = \lambda_{\max}(XSX^T)/\lambda_{\max}(XLX^T)$, where $\lambda_{\max}(XSX^T)$ means the maximum eigenvalue of XSX^T. Note that, during the feature extraction, we will encounter that some matrices are singular, so here PCA is employed as a preprocessing step and keep 98% energy of images. For UDP, the neighborhood size needs to be determined, here we set it to $k = n_i - 1$, where n_i is the number of samples in the i^{th} class.

The ORL database contains 40 individuals; each has 10 sample images with some variations in poses, facial expressions and some details. For each image, it is taken at different times and has different variations including expressions like open or closed eyes, smiling or non-smiling. Some are captured with a tolerance for some tilting and rotation of the face up to 20 degrees. Figure 1 shows some samples of one subject from ORL database.

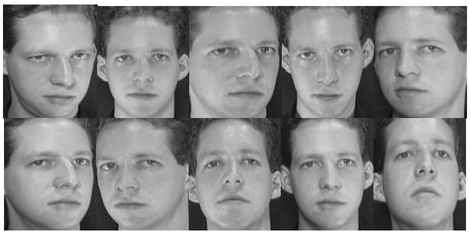

Figure 1. Samples of one subject from ORL database.

We randomly choose $l = 3, 4, 5, 6, and\ 7$ images from each class for training and the remaining for test. For each l, we run 10 times for each algorithm and obtain the average rate as the recognition rate. Table 1 gives the classification accuracy rates (%) for each algorithm under different sizes of training.

Table 1. Recognition Rates on ORL.

Training	PCA	LDA	UDP	SPP	GRWSPA
3	78.2	84.7	82.8	83.2	82.8
4	83.7	90.8	88.2	88.8	89.9
5	86.8	93.7	88.7	90.4	95.2
6	89.1	95.6	93.8	91.5	96.8
7	92.4	96.9	94.7	94.8	98.1

To see how the dimensionality affects recognition rate, Figure 2 shows the recognition rates for different method with respect to different dimensionality on ORL database with four training samples per person.

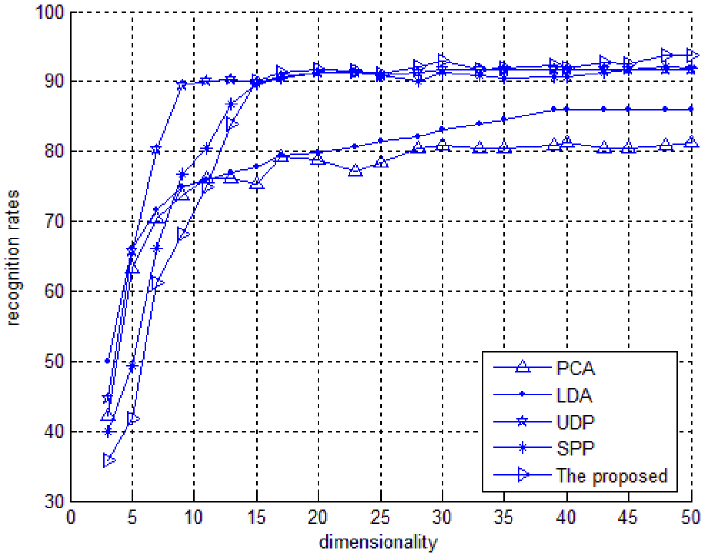

Figure 2. Recognition rates *vs*. dimensionality on ORL database.

The YALE database contains 165 images from 15 subjects, with each 11 images. The images are captured with variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). Figure 3 shows some samples of one subject from YALE database.

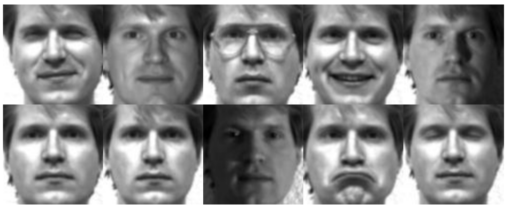

Figure 3. Samples of one subject from YALE database.

We randomly choose $l = 3, 4, 5, 6, \mathit{and}\ 7$ images from each class for training and the remaining for test. For each l, we runs 10 times for each algorithm and obtain the average rate as the recognition rate. Table 2 gives the classification accuracy rates (%) for each algorithm under different sizes of training.

Table 2. Recognition Rates on YALE.

Training	PCA	LDA	UDP	SPP	GRWSPA
3	70.9	72.9	74.8	75.7	76.5
4	73.4	74.5	75.8	77.4	75.9
5	73.9	75.9	78.2	79.3	82.6
6	75.3	76.2	80.5	81.4	85.5
7	76.8	78.1	82.4	83.6	87.8

From Figure 2 and the tables above, we can see that all the algorithms perform better on ORL than YALE database. This is probably on ORL the images have less variation than the images on YALE. LDA and UDP outperform PCA, this is probably PCA is representative in the low dimensional space and helpful for reconstruction, while LDA is a supervised method and takes the class information into account. UDP, as a manifold learning algorithm, makes use of the local and non-local information of the face image, demonstrates its effectiveness in feature extraction. SPP is based upon sparse representation, which preserves the sparse reconstructive relationship of the data and contains natural discriminant information even if it is unsupervised. The proposed algorithm, on one hand, preserves the within-class sparse reconstructive relationships like SPP, on the other hand, maximizes the scatter of samples from different classes. So after projection, data from the same class are compact while data from different classes are far apart. So the proposed algorithm has much better performance than PCA, LDA, LPP, UDP and SPP.

5. Conclusions

In this paper, based on sparsity preserving projection, we propose a new algorithm called Graph Regularized Within-class Sparsity Preserving Analysis (GRWSPA). GRWSPA preserves the within-class sparse reconstruction weights so as to discover the intrinsic information, while maximizing the between-class scatter so that after projection the samples from different classes are far apart. Experiments were carried out on the ORL and YALE face databases, and the results demonstrate the performance advantage of the proposed algorithm over others.

Acknowledgments

This research was supported by Zhejiang Provincial Natural Science Foundation of China (Nos. LQ15F020001, LY14F020036), Scientific Research Fund of Zhejiang Provincial Education Department (Y201223744) and National Science Foundation of China under Grant No. 6127226.

Author Contributions

Songjiang Lou designed and wrote the paper, Ying Chen did the experiments and data analysis, Wenping Guo surveyed the related works, Xiaoming Zhao supervised the work. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Sandbach, G.; Zafeiriou, S.; Pantic, M.; Yin, L. Static and Dynamic 3D Facial Expression Recognition: A Comprehensive Survey. *Image Vis. Comput*. **2012**, *30*, 683–697.
2. Vezzetti, E.; Marcolin, F. 3D human face description: Landmarks measures and geometrical feature. *Image Vis. Comput*. **2012**, *30*, 698–712.
3. Pears, N.; Heseltine, T.; Romero, M. From 3D point clouds to pose-normalised depth maps. *Int. J. Comput. Vis.* **2010**, *89*, 152–176.
4. Vezzetti, E.; Marcolin, F.; Stola, V. 3D human face soft tissues landmarking method: An advanced approach. *Comput. Ind.* **2013**, *64*, 1326–1354.
5. Turk, M.; Pentland, A. Eigenfaces for recognition. *Cogn. Neurosci*. **1991**, *3*, 71–86.
6. Belhume, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces *vs.* Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720.
7. Seung, H.S.; Lee, D.D. The Manifold Ways of Perception. *Science* **2000**, *290*, 2268–2269.
8. Scholkopf, B.; Smola, A.; Smola, E.; Müller, K.-R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Comput*. **1998**, *10*, 1299–1399.
9. Baudat, G.; Anouar, F. Generalized Discriminant Analysis Using Kernel Approach. *Neural Comput.* **2000**, *12*, 2385–2404.
10. Tenenbaum, J.B.; de Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.
11. Rowies, S.T.; Saul, L.K. Nonliear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326.
12. Belkin, M.; Niyogo, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput*. **2003**, *15*, 1373–1396.

13. Raducanu, B.; Dornaika, F. Embedding new observations via sparse-coding for non-linear manifold learning. *Pattern Recognit*. **2014**, *47*, 480–492.
14. He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H.-J. Face recognition using Laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell*. **2005**, *27*, 328–340.
15. Chen, H.-T.; Chang, H.-W.; Liu, T.-L. Local Discriminant Embedding and Its Variants. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; IEEE Computer Society: Washington, DC, USA, 2005; pp. 846–853.
16. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.-J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51.
17. Fu, Y.; Yan, S.; Huang, T.S. Classification and feature extraction by simplexization. *IEEE Trans. Inf. Forensics Secur*. **2008**, *3*, 91–100.
18. Zhang, T.; Yang, J.; Wang, H.; Du, C.; Zhao, D. Maximum variance projections for face recognition. *Opt. Eng*. **2007**, *46*, 067206:1–067206:8.
19. Wang, Y.; Zhao, Y.; Zhang, L.; Liang, J.; Zeng, M.; Liu, X. Graph Construction Based on Re-weighted Sparse Representation for Semi-supervised Learning. *J. Inf. Comput. Sci*. **2013**, *10*, 375–383.
20. Cheng, H.; Liu, Z.; Yang, J. Sparsity induced similarity measure for label propagation. In Proceedings of IEEE 12th International Conference on Computer Vision (ICCV), Kyoto, Japan, 27 September–4 October 2009; pp. 317–324.
21. Yan, S.; Wang, H. Semi-supervised Learning by Sparse Representation. In Proceedings of the 9th SIAM International Conference on Data Mining (SDM 09), Sparks, NV, USA, 30 April–2 May 2009; pp. 792–801.
22. Wright, J.; Yang, A.Y.; Ganesh, A.; Shankar, S.S.; Ma, Y. Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell*. **2009**, *31*, 210–227.
23. Qiao, L.; Chen, S.; Tan, X. Sparsity Preserving Projections with Applications to Face Recognition. *Pattern Recognit*. **2010**, *43*, 331–341.
24. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM Rev.* **2001**, *43*, 129–159.
25. Yang, J.; Zhang, D.; Yang, J.-Y.; Niu, B. Globally maximizing, locally minimizing: Unsupervised discriminant projection with application to face and palm biometrics. *IEEE Trans. Pattern Anal. Mach. Intell*. **2007**, *29*, 650–664.