FACE RECOGNITION USING MARGIN-ENHANCED CLASSIFIER IN GRAPH-BASED SPACE

Ju-Chin Chen, Shang-You Shi and Jenn-Jier James Lien
Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

Keywords: Dimensionality Reduction, Distance Metric Learning, Face Recognition.

Abstract: In this paper, we develop a face recognition system with the derived subspace learning method, i.e. classifier-concerning subspace, where not only the discriminant structure of data can be preserved but also the classification ability can be explicitly considered by introducing the Mahalanobis distance metric in the subspace. Most of graph-based subspace learning methods find a subspace with the preservation of certain geometric and discriminant structure of data but not explicitly include the classification information from the classifier. Via the distance metric, which is constrained by k-NN classification rule, the pairwise distance relation can be locally adjusted and thus the projected data in the *classifier-concerning* subspace are more suitable for k-NN classifier. In addition, an iterative procedure is derived to get rid of the overfitting problem. Experimental results show that the proposed system can yield the promising recognition results under various lighting, pose and expression conditions.

1 INTRODUCTION

Face recognition has been an important issue over the last decades, which has created a wide range of applications, such as surveillance, security systems, etc. Among those appearance-based methods (Murase et al., 1999; Turk et al., 1991), the most well-known algorithms are Eigenface (Turk et al., 1991) and Fisherface (Belhumeur et al., 1997). The former is based on principal component analysis (PCA) to obtain the linear transformation by maximizing the variance of training images but the class information is excluded; while the latter applied linear discriminant analysis (LDA) in (Etemad et al., 1997) which includes the class label information and the discriminant projection is obtained by maximizing the ratio of the between-class and within-class distance.

For non-linearly distributed data such as those associated with non-frontal facial images and under different lighting conditions, the classification performances of the PCA and LDA are somewhat limited due to their assumption of an essentially linear data structure. To resolve the problem, the graph-based dimensionality reduction methods are recently developed by investigating the local information and the essential structure of data manifold which are important for classification purpose, including isometric feature mapping (ISOMAP) (Tenebaum et al., 2000), locally linear embedding (LLE) (Roweis et al., 2000), and Laplacian eigenmap (LE) (Belkin et al., 2003). In (Yan et al., 2007), a unified framework for dimensionality reduction algorithms, i.e. graph embedding, is proposed and most dimensionality reduction algorithms such as PCA, LDA, LPP (He et al., 2003), ISOMAP, LLE and LE can be re-formulated via specific graph Laplacian matrix design. Moreover, it provides a platform such that the new algorithm for dimensionality reduction can be developed with a specific motivation.

Mostly, after obtaining the desired low-dimensional subspace via the above dimensional reduction algorithm, k-NN classifier, which is simpler and more flexible for the multiple-class extension, are often applied based on Euclidean distance metric for recognition. However, Euclidean distance metric ignores the statistical properties of data (Weinberger et al., 2009). Instead of Euclidean distance metric, several distance metric algorithms (Bar-Hillel et al., 2005; Goldberger et al., 2004; Weinberger et al., 2009; Xing et al., 2003) are proposed to obtain a new distance metric to investigate the data properties from class labels in order that the classification accuracy can be improved for k-NN classifier. Among them, Mahala-

Chen J., Shi S. and James Lien J. (2010).
FACE RECOGNITION USING MARGIN-ENHANCED CLASSIFIER IN GRAPH-BASED SPACE.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 382-388
DOI: 10.5220/0002831903820388

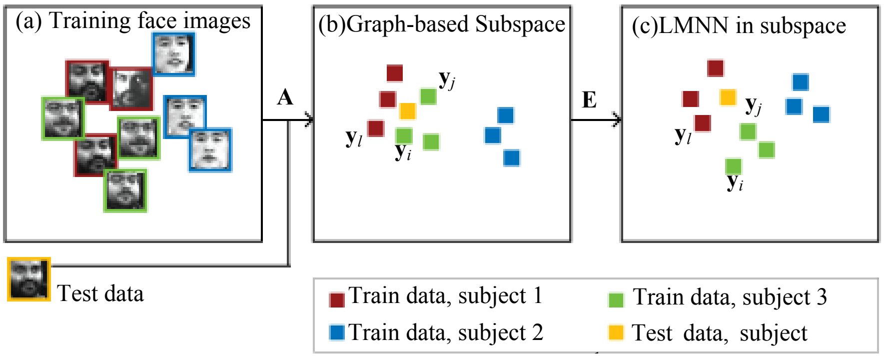

Figure 1: Schematic illustration of proposed subspace learning method, classifier-concerning subspace. (a) The distribution of training images in original input space. (b) Projecting face images to graph-based subspace using the linear transformation matrix A. (c) The classifier-concerning subspace is obtained by applying Large Margin Nearest Neighbour in subspace (b). The pairwise distance relation are locally adjusted to be more suitable for k-NN classifier.

nobis distance metric is learned based on various object function. Relevant Component Analysis (RCA) (Bar-Hillel et al., 2005) learns a full ranked Mahalanobis distance metric using equivalence constraints and the linear transformation can be obtained via solving eigen-problem. In (Xing et al., 2003), a Mahalanobis metric based on the unimodel assumption is proposed. Different from these studies, Large Margin Nearest Neighbour (LMNN) (Weinberger et al., 2009) learn a Mahalanobis distance metric by the constraint on the distance metric imposed by accurate k-NN classification. Thus, via the metric, k-nearest neighbours always belong to the same class while examples from different classes are separated by a large margin.

Inspired from above studies, we propose a subspace learning method with the goal that in the obtained subspace, defined as *classifier-concerning* subspace, not only the local geometric and discriminative structure of data can be preserved but also the projected data are suitable for using k-NN classifier. Fig. 1 shows the schematic illustration for our proposed method. Firstly, in order to analyze the high-dimensional input images in a compact low-dimensional subspace, the graph-based subspace learning method are applied that the discriminant structure of data can be preserved. As shown in Fig. 1(b), most data in the subspace can be well-separated by preserving the local discriminant structure but there exists the data that can not be correctly classified by the k-NN classifier, i.e. y_i and y_k in Fig. 1(b). Then, in order to make them separable, the Mahalanobis distance metric is learned with the constraint on k-NN classification rule. Thus, via the learned distance metric, the pairwise distance of bad-separated points can be locally adjusted while relations of well-separated ones are kept separable (Fig. 1(c)). Moreover, in order to cope with the overfitting problem, an iterative optimization process is derived to obtain a more general subspace than overfitting one for k-NN classifier.

2 GRAPH-BASED SUBSPACE LEARNING

The essential task of subspace learning is to find a mapping function $\boldsymbol{F}$: $x \rightarrow y$ that transforms each image $x \in R^D$ into the desired low-dimensional representation $y \in R^d$ (d<<D) such that y can represent x well in terms of various optimal criteria and each of which corresponds to a specific graph design (Yan et al., 2007). Suppose we have the face image set $\mathbf{X}=[x_1, x_2,\ldots, x_N]$. and each image x_i has the corresponding subject (class label) $l_i \in \{1,2,\ldots,c\}$, where N is number of training images and c is the number of subjects.

2.1 Dimensionality Reduction using Graph Embedding

Given the set $\mathbf{X}$, which could be represented by an intrinsic graph $G=\{\mathbf{X},\mathbf{W}\}$, where the vertices $\mathbf{X}$ corresponds to all facial data and each element W_{ij} of similarity matrix $\mathbf{W} \in R^{N\times N}$ measures the pairwise similarity between vertex x_i and x_j. The diagonal matrix $\mathbf{D}$ and the Laplacian matrix $\mathbf{L}$ of graph G are defined as

$$\mathbf{L}=\mathbf{D}-\mathbf{W}, \quad D_{ii}=\sum_j W_{ij}\ \forall i \tag{1}$$

Let $G_c = \{\mathbf{X}, \mathbf{C}\}$ be a constraint graph with the same vertex set **X** as that of intrinsic graph G and constraint matrix **C**. Note that the similarity matrix **W** and the constraint matrix **C** are designed to capture certain geometric or statistical properties of the data set.

The purpose of graph embedding is to find the low-dimensional representation, i.e. Let **Y**=[y_1,y_2,…,y_N] for data set **X** such that the pairwise similarities measured by **W** can be preserved and the similarities measured by **C** can be suppressed. Therefore, the optimal **Y** can be obtained by

$$\begin{aligned}\mathbf{Y}^* &= \arg\min_{\mathbf{Y}^T\mathbf{CY}=\mathbf{I}} \sum_{i,j} \left\| y_i - y_j \right\|^2 W_{ij} \\ &= \arg\min_{\mathbf{Y}^T\mathbf{CY}=\mathbf{I}} 2tr(\mathbf{Y}^T\mathbf{LY}) = \arg\min tr(\frac{\mathbf{Y}^T\mathbf{LY}}{\mathbf{Y}^T\mathbf{CY}})\end{aligned} \quad (2)$$

In order to minimize the object function, for larger similarity between samples x_i and x_j, the distance between y_i and y_j should be smaller; while smaller similarity between x_i and x_j should lead to larger distances between y_i and y_j (Yan et al., 2007). The optimization of Eq. (2) can be solved by a generalized eigenvalue problem as **LY**=λ**CY**.

2.2 Linear Transformation

Assume that the low-dimensional representation y can be obtained from a linear transformation vector **a**, i.e. y_i=$\mathbf{a}^T x_i$. Thus the Eq. (2) can be rewritten as

$$\mathbf{a}^* = \arg\min_{\mathbf{a}^T\mathbf{XCX}^T\mathbf{a}=\mathbf{I}} 2\,tr(\mathbf{a}^T\mathbf{XLX}^T\mathbf{a}) = \arg\min tr(\frac{\mathbf{a}^T\mathbf{XLX}^T\mathbf{a}}{\mathbf{a}^T\mathbf{XCX}^T\mathbf{a}}) \quad (3)$$

The transformation **A**=[$\mathbf{a}_1$,$\mathbf{a}_2$,…,$\mathbf{a}_d$] can be solved as $\mathbf{XLX}^T\mathbf{a} = \lambda\mathbf{XCX}^T\mathbf{a}$ by selecting the eigenvectors corresponding to d smallest eigenvalue. Different design on intrinsic and constraint graphs will lead to many popular linear dimensionality reduction algorithms (Cai et al., 2007; Etemad et al., 1997; He et al., 2003; Yan et al., 2007). For example, the graphs of the locality sensitive discriminant analysis (LSDA) are defined by

$$\begin{aligned} \mathrm{W}_{w,ij} &= \begin{cases} 1 & , \text{ if } x_i \in N_w(x_j) \text{ and } x_j \in N_w(x_i) \\ 0 & , \text{ otherwise} \end{cases} \\ \mathrm{W}_{b,ij} &= \begin{cases} 1 & , \text{ if } x_i \in N_b(x_j) \text{ and } x_j \in N_b(x_i) \\ 0 & , \text{ otherwise} \end{cases} \end{aligned} \quad (4)$$

where the subject of each element in $N_w(x_i)$ is as same as x_i and $N_b(x_i)$ is as different as x_i. The detailed graph design and objective function can refer to (Cai et al., 2007).

3 CLASSIFIER-CONCERNING SUBSPACE LEARNING

After finding a subspace, k-NN classifier is widely applied for classification in the desire low-dimensional subspace. Inspired by this situation, in Section 3.1, our face recognition system attempts to find a subspace, i.e. *classifier-concerning* subspace, where not only the structure of data can be preserved but also the classification ability can be explicitly considered by introducing the Mahalanobis distance metric in the subspace. In addition, an iterative optimization for obtaining the subspace is derived in Section 3.2.

3.1 Margin Enhancement in Subspace

As known, the face images under different poses, lighting conditions and facial expressions are non-linearly distributed in high-dimensional input space. Hence, at the first step of our proposed method (Fig. 1), an projection matrix **A**=[$\mathbf{a}_1$,$\mathbf{a}_2$,…,$\mathbf{a}_d$], learned by the graph-based subspace learning method, such that most of the desired low-dimensional data $\{y_i\}_{i=1}^n \in R^d$, i.e. $\mathbf{Y}=\mathbf{A}^T\mathbf{X}$, can be well-separated in that subspace.

Although the discriminant structure has been discovered in the subspace **A**, there exists some bad-separated data, i.e. the distance of y_i and y_j of the same subject (class) is larger than the distance y_i and y_j of different subjects (shown in Fig. 1(b)). Therefore, in the second step, a Mahalanobis distance metric is applied to enhance the margin between data in the subspace **A** by applying LMNN (Weinberger et al., 2009) which is according to the k-NN classification rule. Thus, via the learned distance metric, the distances for those bad-separated data can be locally adjusted and meanwhile the distance for well-separated data can be kept in the resulting subspace, named as *classifier-concerning* subspace (Fig. 1(c)). As shown, the k-nearest neighbours always belongs to the same class while data from different classes are separated by a large margin and thus the projected data in the *classifier-concerning* subspace would have better distance relation for the k-NN classifier for recognition.

Let each data y_i in the subspace **A** with the class label $l_i \in \{1,2,...,c\}$. The Mahalanobis distance metric $\mathbf{M} \in R^{d\times d}$ can be expressed in terms of the square matrix $\mathbf{M} = \mathbf{E}^T\mathbf{E}$, where $\mathbf{E} \in R^{d\times d}$ represents a linear transformation. The square distance between two low-dimensional embeddings y_i and y_j is computed as:

$$D(y_i, y_j) = (y_i - y_j)^T \mathbf{M} (y_i - y_j) = \left\| \mathbf{E}(y_i - y_j) \right\|^2 \quad (5)$$

According to the k-NN classification rule, the cost function can be defined as (Weinberger et al., 2009):

$$\varepsilon(\mathbf{E}) = (1-\alpha)\sum_{ij}\eta_{ij}\left\|\mathbf{E}(y_i - y_j)\right\|^2 + \alpha\sum_{ijk}\eta_{ij}(1-\ell_{ik})[1+\left\|\mathbf{E}(y_i - y_j)\right\|^2 - \left\|\mathbf{E}(y_i - y_k)\right\|^2]_+ \quad (6)$$

where $\eta_{ij} \in \{0,1\}$ indicate whether y_j is a target neighbour of y_i, $\ell_{ik} \in \{0,1\}$ indicate whether y_i and y_k share the same class label, and $[z]_+ = \max(z,0)$ denotes the standard hinge loss function. Note that the first term of cost function only penalizes large distances between each input y_i and its target neighbour. While the second term penalizes small distances between each input and all other inputs with different class label. The scalar α can be tuned the importance between two terms. By introducing a nonnegative slack variable δ_{ijk} , Eq.(6) can be reformulate to a semi-definite programming (SDP) problem, as well as SVM (Weinberger et al., 2009).

$$\begin{aligned} \min \quad & \sum_{ij}\eta_{ij}(y_i - y_j)^T\mathbf{M}(y_i - y_j) + \alpha\sum_{ijk}\eta_{ij}(1-l_{ik})\delta_{ijk} \\ \text{s.t.} \quad & (y_i - y_k)^T\mathbf{M}(y_i - y_k) - (y_i - y_j)^T\mathbf{M}(y_i - y_j) \geq 1 - \delta_{ijk} \\ & \delta_{ijk} \geq 0 \\ & \mathbf{M} \succeq 0 \end{aligned} \quad (7)$$

Thus, for the input image x_i, the desired low-dimensional representation contained the classification information, $\tilde{y}$, can be obtained by

$$\tilde{\mathbf{Y}} = \mathbf{EY} = \mathbf{E}(\mathbf{A}^T\mathbf{X}) = \mathbf{P}^T\mathbf{X} \quad (8)$$

Note that via the transformation matrix **P**, the pairwise distance between low-dimensional data $\tilde{y}$ in the obtained subspace, named as *classifier-concerning* subspace, is constraint on the k-NN classification rule. Thus, not only the desired data structure in the high-dimensional space can be preserved but also the low-dimensional data are suitable for using k-NN classifier for further classification.

3.2 Iterative Optimization

In this subsection, a procedure is derived to optimize the *classifier-concerning* subspace **P** in Eq. (8) via iteratively optimizing the graph-embedding projection **A** and the distance metric **M**. From Eq. (3), learning a subspace via specific graph Laplacian matrix, the better graph embedding projection can be obtained if *classifier-concerning* information could be acquired from distance metric **M**. And thus the desire low-dimensional subspace not only preserve the data structure but are also suitable for k-NN classifier, i.e. the ideal case in *classifier-concerning* subspace learning is **P**=**A**, and **M** is an identity metric. However, it is not intuitive to know the classification results in advance but via the Mahalanobis distance metric learned in Eq. (7), we can obtain the pairwise data relation, which embeds the classification information, in the low-dimensional subspace.

To obtain a virtual high-dimensional distance metric which can pass the low-dimensional pairwise distance relation into the original input space, we first inspect the pairwise distance in the desire low-dimensional subspace

$$\begin{aligned} d_{ij} &= (y_i - y_j)^T\mathbf{M}(y_i - y_j) \\ &= (x_i - x_j)^T\mathbf{AMA}^T(x_i - x_j) \\ &= (x_i - x_j)^T\hat{\mathbf{M}}(x_i - x_j) \end{aligned} \quad (9)$$

From Eq. (9), we know Mahalanobis distance metric **M** can be re-projecting to the original space via the projection A. i.e. $\hat{\mathbf{M}} = \mathbf{AMA}^T$ .As the metric can be represented as $\mathbf{M} = \mathbf{E}^T\mathbf{E}$, the metric $\hat{\mathbf{M}}$ can be as

$$\hat{\mathbf{M}} \overset{SVD}{=} U\Lambda U^T = U\Lambda^{1/2}\Lambda^{1/2}U^T = \hat{\mathbf{E}}^T\hat{\mathbf{E}} \quad (10)$$

$\hat{\mathbf{E}}: R^m \to R^m$ is a linear transformation. Thus, through $\hat{\mathbf{E}}$, each original input data x_i can be represented as $\hat{\mathbf{x}}_i = \hat{\mathbf{E}}\mathbf{x}_i$ such that the all input data $\hat{\mathbf{X}} = \hat{\mathbf{E}}\mathbf{X}$ would implicitly have k-NN classification information. Therefore, by using the data with classification information, i.e. $\hat{\mathbf{X}}$, the new graph-embedding projection $\mathbf{A}=[\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_d]$ can be obtained based on the data structure as:

$$\mathbf{a}^* = \arg\min_a \sum_{ij} \left\|\mathbf{a}^T(\hat{x}_i - \hat{x}_j)\right\|^2 w_{ij} \quad (11)$$

By introducing the Laplacian matrix **L** and constraint matrix **C** according to the original data structure, Eq. (11) can be derived as

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} tr(\frac{\mathbf{a}^T\hat{\mathbf{X}}\mathbf{L}\hat{\mathbf{X}}^T\mathbf{a}}{\mathbf{a}^T\hat{\mathbf{X}}\mathbf{C}\hat{\mathbf{X}}^T\mathbf{a}}) = \arg\min tr(\frac{\mathbf{a}^T\hat{\mathbf{E}}\mathbf{X}\mathbf{L}\mathbf{X}^T\hat{\mathbf{E}}^T\mathbf{a}}{\mathbf{a}^T\hat{\mathbf{E}}\mathbf{X}\mathbf{C}\mathbf{X}^T\hat{\mathbf{E}}^T\mathbf{a}}) \quad (12)$$

The optimal projection **A** in Eq. (12) can be solved by the generalized eigenvalue problem, and the newly low-dimensional embedding are $\widetilde{\mathbf{Y}}' = \mathbf{A}^T\widetilde{\mathbf{X}}$ .Though the iterative process, the new low-dimensional distance metric **M** and projection **A** are obtained alternatively by using the information from the other until $\mathbf{M} \approx \mathbf{I}$, e.g. all the classification information can be embedded to the project **A**, i.e. **P**=**A**. The proposed iteratively updating procedure is summarized in Fig. 2.

Input: All training images $\mathbf{X} = \{x_1, x_2, ..., x_N\} \in R^m$ with the class label $\{l_i\}_i^N$ and the graph for **L** and **C**.
Output: Subspace **P**

1. Initialize: $\hat{\mathbf{E}}^{(1)} = \mathbf{I}$
2. For t=1: iter
3. $\hat{\mathbf{X}}^{(t)} = \hat{\mathbf{E}}^{(t)}\mathbf{X}$
4. Compute $\mathbf{A}^{(t)} = \arg\min tr(\frac{\mathbf{a}^T\hat{\mathbf{X}}^{(t)}\mathbf{L}\hat{\mathbf{X}}^{(t)^T}\mathbf{a}}{\mathbf{a}^T\hat{\mathbf{X}}^{(t)}\mathbf{C}\hat{\mathbf{X}}^{(t)^T}\mathbf{a}})$
5. For all data i, $y_i^{(t)} = \mathbf{A}^{(t)^T} x_i^{(t)}$
6. Compute $\mathbf{M}^{(t)} = \text{LMNN}(\{y_i^{(t)}\}_{i=1}^N, \{l_i\}_{i=1}^N)$
7. $\mathbf{M}^{(t)} = \mathbf{E}^{(t)^T}\mathbf{E}^{(t)}$
8. Compute high-dimensional metric $\hat{\mathbf{M}}^{(t)}$: $\hat{\mathbf{M}}^{(t)} = \mathbf{A}^{(t)}\mathbf{M}^{(t)}\mathbf{A}^{(t)^T}$
9. Perform SVD on $\hat{\mathbf{M}}^{(t)}$: $\hat{\mathbf{M}}^{(t)} = \hat{\mathbf{E}}^{*^T}\hat{\mathbf{E}}^*$
10. $\hat{\mathbf{E}}^{(t+1)} \leftarrow \hat{\mathbf{E}}^*\hat{\mathbf{E}}^{(t)}$
11. Check convergence condition: $\left\|\mathbf{M}^{(t)} - \mathbf{I}\right\| < \varepsilon$
12. End
13. $\mathbf{P} = \mathbf{A}^{(t)}\mathbf{E}^{(t)^T}$

Figure 2: Procedure of iteratively optimizing the subspace **P**.

4 EXPERIMENTAL RESULTS

In this section, we investigate the performance of the proposed subspace learning method for face recognition under different lightings, poses, and expressions. Subspace learning methods: supervised-LPP (Zheng et al., 2007) (designated as SLPP) and LSDA (Cai et al., 2007) are compared with our proposed model, e.g. supervised-LPP cascaded by LMNN model (designated as SLPP+LMNN) and LSDA cascaded by LMNN model (designated as LSDA+LMNN), respectively. In addition, Eigenface (PCA) (Murase et al., 1995) and RLDA (Ye et al., 2006) which give impressive results (Cai et al. 2007) are compared as well. Note that k-nearest neighbours (*k*=3) are applied in the following experiments. We use the source code kindly proved the authors of (Cai et al., 2007; Weinberger et al., 2009).

4.1 Database and Image Preprocess

Both Extended Yale-B[1] and CMU PIE[2] database are considered as a tough task for face recognition problem because they are in a complex environmental setting. Hence, we use both databases to evaluate our proposed method. The Extended Yale-B face database contains 16128 images of 38 human subjects under 9 poses and 64 illumination conditions. We choose the frontal pose and use all the images under different illumination, thus we get 64 images for each person. The CMU PIE face data base contains 68 human subjects with 41,368 face images. The face images were acquired across different poses, illumination conditions, and expressions. In our experiment, five near frontal poses (C05, C07, C09, C27, C29) and all the images under different illuminations and expressions are used, thus we get 170 images for each individual.

For both databases, all these face images are manually aligned and cropped to 32x32 pixels, and the pixel values are then normalized to the range [0, 1] (divided by 256). Each database is then partitioned into the gallery and probe set, denoted as G_p / P_q ,where p images per person are randomly selected for training and the remaining q images are used for test. Note that in order to reduce the noise, the data are processed by PCA and 98% energy is saved. The dimensionality of feature subspace is set to c-1 for all subspace learning methods, where c is the number of individuals.

4.2 Comparison of Subspace Learning Methods

The recognition results conducted on PIE and Extend Yale-B database are listed in Table 1 and 2, respectively. For each G_p / P_q , we have 35 random splits but exclude possible over-fitting splits, then average the remaining splits, report the mean recognition rate as well as the standard deviation in the table. Through Table 1 and 2, it can be seen that our proposed method can further improve the results

Table 1: Recognition accuracies of different algorithms on Extended Yale-B database.

	G30/P34	G40/P24
PCA (98% energy)	38.63±1.31	33.78±1.03
RLDA	8.60±0.89	6.68±0.84
SLPP	8.60±0.89	6.68±0.84
SLPP+LMNN	8.13±0.64	6.63±0.71
LSDA	8.63±0.85	6.92±0.74
LSDA+LMNN	8.03±0.87	6.62±0.64

than RLDA, SLPP and LSDA, respectively, especially for the PIE database. This indicates that embedding the classification information to the subspace via a trained Mahalanobis distance metric can discover a more discriminative structure of the face manifold and hence improve the recognition rate.

In order to investigate the stability of the proposed method, we compare the results of projecting original data **X** into various *d*-dimensionality feature subspaces via LSDA and LSDA+LMNN subspaces. Fig. 3 shows the results on Yale-B and PIE database, respectively. As Fig 3 shown, metric learning needs enough desired low-dimension to learn the Mahalanobis distance metric for further improvement recognition result. Ideally desired low-dimension set *c*-1 gives convenience and efficiency from the result.

4.3 Recognition Results of Classifier-Concerning Subspace with Iterative Optimization

Table 2: Recognition accuracies of different algorithms on PIE database.

	G30/P34	G40/P24
PCA (98% energy)	43.32±0.62	36.02±0.73
RLDA	8.78±0.34	6.45±0.32
SLPP	8.78±0.34	6.45±0.32
SLPP+LMNN	6.60±0.21	4.94±0.20
LSDA	8.87±0.36	6.52±0.32
LSDA+LMNN	6.55±0.25	4.91±0.20

As mentioned in 4.2, some splits cause overfitting result. The main reason is that the model fit the training data too much to lose the generality of the model. From Table 3, it can be seen that the overfitting problem can be overcome by iterative way. Interestingly, the number of PIE training data is enough to cause nearly no over-fitting result.

5 CONCLUSIONS

Table 3: Recognition accuracies of the proposed iteratively updating framework on extend Yale-B database.

	G30/P34 (train error / testerror)	G40/P24 (train error / testerror)
LSDA	6.04±0.49/ 8.60±1.0	4.75±0.42/ 6.60±0.80
LSDA+LMNN	2.91±0.42/ 8.79+0.96	2.54±0.39/ 6.70±0.95
LSDA+LMNN+Iter.	5.41±0.6/ 8.03+1.06	4.60±0.36/ 6.18±0.82

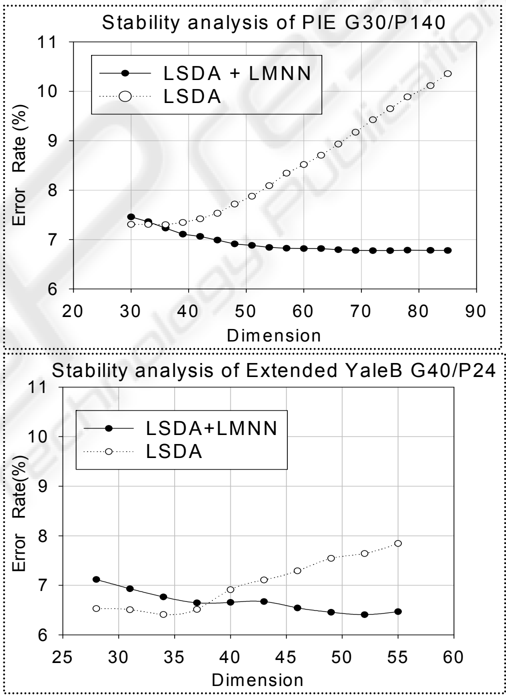

Figure 3: Recognition results on various *d*-dimensionality feature subspaces obtained via LSDA and the proposed method which embed the classification information additionally.

In this paper, we have shown how to learn a *classifier-concerning subspace* for face recognition and that can provide the promising recognition results under various lighting, pose and expression condition. The *classifier-concerning* subspace preserves certain data characteristic by specifying a graph and the low-dimensional projected data are suitable for the usage of k-NN classifier for recognition task. Because the proposed method consists of two learning process, i.e. the initial subspace and the distance metric learning, the model would suffer from the overfitting to data if the number of training data is insufficient or ineffective. Hence, we propose an iterative solution via a virtual Mahalanobis distance in original high-dimensional input space that can pass low-dimensional pairwise distance relation. Based on this virtual Mahalanobis distance the pairwise distance relation of data in input space can be adjusted and then the *classifier-concerning* subspace can be updated. Ongoing work is to circumvent overfitting via adding some mechanics for choosing effective training data or applying regularization information.

REFERENCES

Bar-Hillel, A., Hertz, T., Shental, N., and Weinshall, D. (2005). Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, 6: 937-965.

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Journal of Neural Computation*, 15(6): 1373-1396.

Belhumeur, P.N., Hespanha, J.P., and Kriegman, D.J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 711-720.

Cai, D., He, X., and Zhou, K. (2007). Spectral regression for efficient regularized subspace learning. In *International Conference on Computer Vision*.

Cai, D., He, X., Zhou, K., Han, J., and Bao, H. (2007). Locality sensitive discriminant analysis. In *International Joint Conference on Artificial Intelligence*.

Etemad, K. and Chellapa, R. (1997). Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, 14(8): 1724-1733.

Goldberger, J., Roweis, S. T., Hinton, G.E., and Salakhutdinov, R. (2004). Neighbourhood components analysis. In *Advances in Neural Information Processing Systems*.

He, X. and Niyogi, P. (2003). Locality preserving projections. In *Advances in Neural Information Processing Systems*.

Murase, H. and Nayar, S. K. (1995). Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14(1): 5-24.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500): 2323-2326.

Tenebaum, J. B., de Silva, V. and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500): 2319-2323.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1): 71-86.

Weinberger, K. Q. and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10: 209-244.

Xing, E., Ng, A., Jordan, and M. Russell, S. (2003). Distance metric learning, with application to cluster with side information. In *Advances in Neural Information Processing Systems*.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., and Lin, S. (2007). Graph embedding and extension: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 29(1): 40-51.

Ye, J. P. and Wang, T. (2006). Regularized discriminant analysis for high dimensional, low sample size data. In *International Conference. on Knowledge Discovery and Data Mining*.

Zheng, Z., Yang, F., Tan, W., Jia, J. and Yang, J. (2007). Gabor feature-based face recognition using supervised locality preserving projection. *Signal Processing*, 87: 2473-2483.