Face Recognition and Unseen Subject Rejection in Margin-Enhanced Space

Ju-Chin Chen, Shang-You Shi, and Jenn-Jier James Lien
Dept. of Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan
joan@csie.ncku.edu.tw, elvise_818@gmail.com, jjlien@csie.ncku.edu.tw

***Abstract*—In this paper, we develop a face recognition system with a rejection mechanism for imposter or unseen subjects. In order to boost the recognition rate and provide the promising rejection accuracy, a margin-enhanced space is derived by reweighting the LSDA space via explicitly imposing the constraint of the k-NN classification rule. In this space, not only the local discriminant structure of data can be extracted but the enhanced pairwise distance can be used to model the acceptance and rejection likelihood probability. According to the Bayes decision rule, the unseen subject can be rejected if the likelihood ratio is smaller than the estimated threshold. Note that the rejection performance based on the likelihood ratio is more tolerable than the pre-defined distance only. Experimental results show that the proposed system not only yields the higher recognition rate than other subspace learning methods but also provides the promising rejection accuracy on the challenging databases of various lighting conditions and facial expression.**

***Keywords*—face recognition, graph-based subspace learning, margin-enhanced space**

I. INTRODUCTION

Automatic face recognition is an essential requirement in a wide range of applications, including the surveillance system, security systems, access control systems, etc. For these applications, not only high recognition performance but the promising accuracy, i.e. the rejection of imposter or unseen subjects, are required.

Among those appearance-based face recognition methods, the most well-known algorithms are Eigenface [10] and Fisherface [16]. However, 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 essential assumption of the linear data structure. To resolve this problem, the kernel-based algorithms, such as Kernel PCA and Kernel Fisherface [9], [13], [19], are explored by extending the linear dimensionality reduction algorithms into non-linear ones by performing the algorithms on higher or infinite dimensional feature spaces. However, for those algorithms, the local structure of data is not explicit considered, which is important for classification purpose. Unlike kernel methods, manifold learning or subspace learning methods are recently developed to investigate the local information and the essential structure of data manifold, including isometric feature mapping (ISOMAP) [14], locally linear embedding (LLE) [11], Laplacian eigenmap (LE) [2], locality preserving projections (LPP) [7], and marginal Fisher analysis (MFA) [20]. These methods can provide the promising recognition performance by using *k*-NN classifier in the corresponding low-dimensional subspace but imposter classification is not further discussed.

To improve the classification performance for *k*-NN classifier, several distance metric algorithms [1], [8], [17], [18] are proposed to investigate the data properties from class labels recently. Instead of using Euclidean distance which ignores the statistical properties of data, Mahalanobis distance metric is learned based on various object function [1], [17], [8], [15], [18]. Among them, Large Margin Nearest Neighbor (LMNN) [8], [17] learn a Mahalanobis distance metric by imposing the constraint of *k*-NN classification rule. Thus, via the learned metric, *k*-nearest neighbors always belong to the same class while data from different classes are separated by a large margin.

Inspired from above studies, we propose a space learning method with the goal that in the proposed space, i.e. margin-enhanced space, not only the local geometric and discriminative structure of data can be preserved but also the enhanced distance can be used for the imposter classification. Moreover, rather than using a pre-defined distance threshold to reject imposter or unseen subject, which is hard to be estimated and less flexible for various kinds of application, the acceptance and rejection likelihood probabilities are modeled by Gaussian-like distribution based on the nearest neighbor (1-NN) distance information. Then the likelihood ratio can be applied for the imposter classification to reject or accept the test facial image before face recognition.

II. SYSTEM FLOWCHART

Fig. 1 shows the flowchart of the proposed face recognition system with the imposter classification. In the training process, the training data $\mathbf{X}=(\mathbf{X}_F, \mathbf{X}_I)$ are composed of two facial image sets: the recognition face set $\mathbf{X}_F$ and the imposter face set $\mathbf{X}_I$. The recognition face set $\mathbf{X}_F=\{x_1, x_2, ..., x_N\}$ contains total N ($=c\times m$) facial images and each image $x_i \in \mathbf{X}_F$ has the corresponding class label $\ell(x_i) \in \{1,2,...,c\}$, where c is the number of classes (subjects) and each class is complied

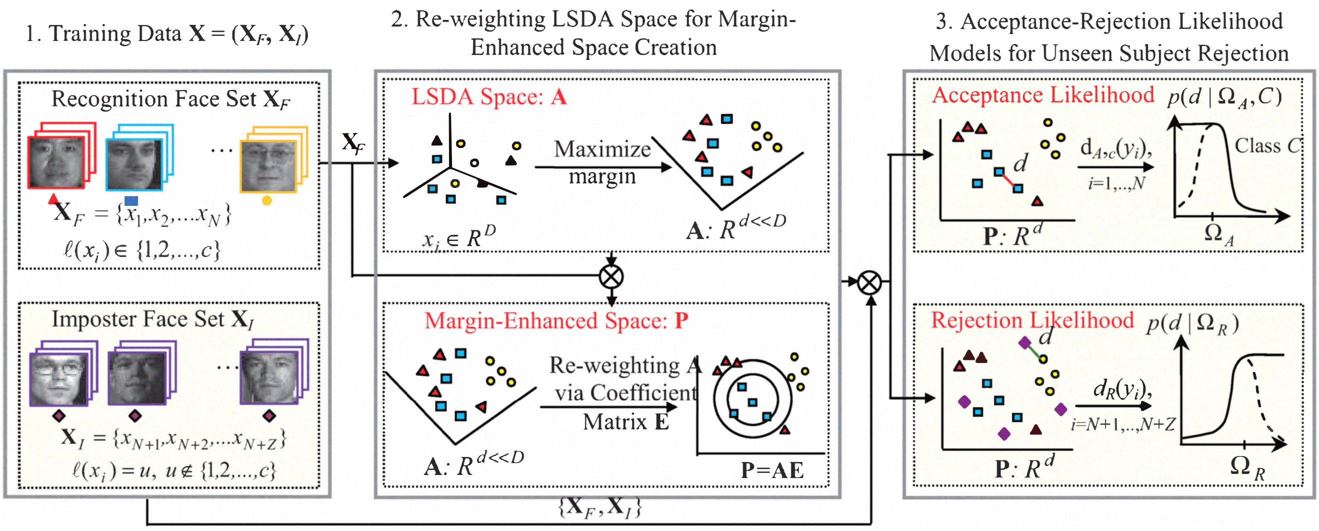


Figure 1. Flowchart of the proposed face recognition system with imposter classification.

m D-dimensional (32×32-pixel) facial images of arbitrary slight head motions and various lighting conditions. Besides, different from other recognition systems, the imposter face set $\mathbf{X}_I = \{x_{N+1}, x_{N+2}, ..., x_{N+Z}\}$ is incorporated for the development of imposter classification, which consists of Z facial images and the label of each image $x_i \in \mathbf{X}_I$ is $\ell(x_i) = u$, i.e. imposter class and $u \notin \{1,2,..., c\}$.

As shown in Fig. 1, the face set $\mathbf{X}_F$ is used for margin-enhanced space creation. To preserve the local discriminant structure of facial images, local sensitive discriminant analysis (LSDA) [5] is applied to extract the low-dimensional embeddings for each image. Although most data of different labels in the LSDA space can be well-separated, there exists data that cannot be correctly classified. Besides, no explicit constraint on margin distance is imposed for the LSDA function and thus in the LSDA space the threshold is hard to be estimated for the imposter classification. Therefore, the transformation **A** is reweighted by a matrix **E**, i.e. **P=AE**, and in the resulting margin-enhanced space, not only the recognition performance is improved but also the enhanced margin distance can be used for imposter classification. Subsequently, both facial images in the recognition face set $\mathbf{X}_F$ and imposter face set $\mathbf{X}_I$ are projected to the margin-enhanced space and then the 1-NN distance samples obtained from $\mathbf{X}_F$ and $\mathbf{X}_I$ are used to model the acceptance and rejection likelihood probability, respectively.

For the test process, the k-NN classifier is applied for the face recognition (k=3 in our system). Note that before outputting the final recognition result, the Bayes decision rule is applied for the imposter classification to reject or accept the test image.

III. RE-WIGHTING LSDA SPACE FOR MARGIN-ENHANCED SPACE

A margin-enhanced space is introduced by reweighting the LSDA space via imposing k-NN classification rule and unit margin constraint with a SDP formulation [17]. Not only the local geometric and discriminative structure of data can be extracted for the face recognition but also the enhanced pairwise distance in the margin-enhanced space is more suitable for imposter classification.

A. Graph-based Space Creation Using LSDA

For the face recognition, the local structure of data is important especially for facial images with appearance variations in image space. In this paper, LSDA [5] is applied to extract the low-dimensional discriminant feature to preserve both data discriminant and geometrical structure. Let $N(x_i) = \{x_i^1, ..., x_i^k\}$ be the set of k nearest neighbors for each image $x_i \in \mathbf{X}_F$, and the set $N(x_i)$ can equal to sum of two subsets $N_b(x_i)$ and $N_w(x_i)$, each of which contain the data with the same and different class label from x_i as

$$\begin{aligned} N_w(x_i) &= \{x_j \mid l(x_j) = l(x_i), 1 \le j \le k\} \\ N_b(x_i) &= \{x_j \mid l(x_j) \ne l(x_i), 1 \le j \le k\} \end{aligned} \tag{1}$$

where $N_b(x_i) \cup N_w(x_i) = N(x_i)$ and $N_b(x_i) \cap N_w(x_i) = \emptyset$. According to (1), the within-class graph $G_w=\{\mathbf{X}, \mathbf{W}_w\}$ and between-class graph $G_b=\{\mathbf{X}, \mathbf{W}_b\}$ are constructed where the vertices **X** correspond to all face images in $\mathbf{X}_F$ and the weight matrices $\mathbf{W}_w$ and $\mathbf{W}_b$ represent the connection between each image x_i to its local neighbors of same and different classes as

$$W_{w,ij} = \begin{cases} 1 & , \ if\, x_i \in N_w(x_j) \ and\ x_j \in N_w(x_i) \\ 0 & , \ otherwise \end{cases} \tag{2}$$

$$W_{b,ij} = \begin{cases} 1 & , \ if\, x_i \in N_b(x_j) \ and\ x_j \in N_b(x_i) \\ 0 & , \ otherwise \end{cases} \tag{3}$$

Then, the low-dimensional embeddings $\mathbf{Y}^T = \{y_1, ..., y_N\}$ (1-D case for example) are obtained according to the criterions that the distance of same class should be minimized while the distance between different classes should be maximized:

$$\min \textstyle\sum_{i,j} \|y_i - y_j\|^2 W_{w,ij} = \min \mathbf{Y}^T \mathbf{L}_w \mathbf{Y} \tag{4}$$

$$\max \textstyle\sum_{i,j} \|y_i - y_j\|^2 W_{b,ij} = \max \mathbf{Y}^T \mathbf{L}_b \mathbf{Y} \tag{5}$$

where $\mathbf{L}_w = \mathbf{D}_w - \mathbf{W}_w$ and $\mathbf{L}_b = \mathbf{D}_b - \mathbf{W}_b$ are Laplacian matrix of the graph G_w and G_b, respectively; $\mathbf{D}_w$ and $\mathbf{D}_b$ are diagonal matrix to represent the connection degrees. By combining the criterions in (4) and (5), the local margin between same and different classes can be maximized by

$$\max_{\mathbf{Y}^T\mathbf{D}_w\mathbf{Y}=1} \mathbf{Y}^T\mathbf{L}_b\mathbf{Y} + \mathbf{Y}^T\mathbf{W}_w\mathbf{Y} \tag{6}$$

Note that LSDA imposes a constraint $\mathbf{Y}^T\mathbf{D}_w\mathbf{Y}=1$ for the scale invariance [4], [5], [8]. In order to process the new input data, a linear transformation vector $\mathbf{a}$ is assumed that the low-dimensional embedding y_i can be obtained from x_i by $y_i = x_i^T\mathbf{a}$ and thus (6) can be rewritten as

$$\max_{\mathbf{a}^T\mathbf{X}\mathbf{D}_w\mathbf{X}^T\mathbf{a}=1} \mathbf{a}^T\mathbf{X}(\alpha\mathbf{L}_b + (1-\alpha)\mathbf{W}_w)\mathbf{X}^T\mathbf{a} \tag{7}$$

where $\alpha \in [0,1]$ is a constant to weight the distance between same and different classes. In general case, the D-by-d transformation matrix $\mathbf{A}=[\mathbf{a}_1,\ldots,\mathbf{a}_d]$ is given by the eigenvectors corresponding d biggest eigenvalue to generalized eigenvalue problem $\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}\mathbf{D}_w\mathbf{X}^T\mathbf{a}$, where $d<<D$.

B. Margin-Enhanced Space Creation

Although the discriminant structure has been discovered in the LSDA space and most projected data can be well-separated, there exists some bad-separated data especially for the non-linear facial distribution of various pose and lighting conditions. Besides, the margin between different classes might be discordant in LSDA space because no explicit constraint is imposed on margin distance in (7). LSDA can still provide high recognition results because the discordant margin does not account for recognition performance; however, the margin distance in LSDA space can not be used for the imposter classification to reject the unseen subjects.

Considering the above problem, a transformation matrix $\mathbf{P} \in R^{D\times d}$ is proposed by reweighting the LSDA transformation $\mathbf{A}$ with a coefficient vector w that $\mathbf{P}=[w_1\mathbf{a}_1, w_2\mathbf{a}_2,\ldots, w_d\mathbf{a}_d]$. For the generality, the transformation $\mathbf{P}$ can be rewritten via the matrix form as

$$\mathbf{P}=\mathbf{A}\mathbf{E} \tag{8}$$

where $\mathbf{E} \in R^{d\times d}$ is a coefficient matrix. Via the learned transformation matrix $\mathbf{P}$, the pairwise distance of bad-separated data can be locally adjusted meanwhile the distance between well-separated ones can keep separable in the resulting space, i.e. margin-enhanced space. Furthermore, the discordant margin can be repaired such that the nearest neighbor distance can be used for the imposter classification. Note that in our system, only the matrix $\mathbf{E}$ need to be estimated.

According to the k-NN classification rule, the correct distance order between different label data should be:

$$\begin{aligned}&\left\|\mathbf{P}^T(x_i - x_l)\right\|^2 - \left\|\mathbf{P}(x_i - x_j)\right\|^2 > 1 - \delta_{ijl}\\ &\textit{where}\quad \ell(x_i) \neq \ell(x_l) \textit{ and } \ell(x_i) = \ell(x_j)\end{aligned} \tag{9}$$

where $\left\|\mathbf{P}^T(x_i - x_l)\right\|^2$ is the distance between different labeled data x_i and x_l, while $\left\|\mathbf{P}^T(x_i - x_j)\right\|^2$ is the distance between same labeled data x_i and x_j, scalar 1 represents a unit margin and a nonnegative slack variable δ_{ijl} is introduced to allow misclassification for the non-linearly distributed data. Although the transformation $\mathbf{P}$ is introduced to enhance the margin between different labeled data, the distance between data and the corresponding neighbors of same class should be keep close. Thus the cost function can be defined as [17]

$$\begin{aligned}\varepsilon(\mathbf{P}) = {} & (1-\beta)\sum_{ij}\eta_{ij}\left\|\mathbf{P}^T(x_i - x_j)\right\|^2 + \\ & \beta\sum_{ijl}\eta_{ij}(1-\ell_{il})[1 + \left\|\mathbf{P}^T(x_i - x_j)\right\|^2 - \left\|\mathbf{P}^T(x_i - x_l)\right\|^2]_+\end{aligned} \tag{10}$$

where $\eta_{ij} \in \{0,1\}$ indicates whether x_j is the same labeled neighbor of x_i, $\ell_{il} \in \{0,1\}$ indicate whether x_i and x_l share the same class label, the scalar β can tune the importance between two terms and $[z]_+ = \max(z,0)$ denotes the standard hinge loss. Note that the first term only penalizes large distances between each input x_i and its same labeled neighbors while the second term penalizes small distances between each data and its corresponding different labeled neighbors. As the definition of $\mathbf{P}$ in (8), the distance between data x_i and x_j in margin-enhanced space becomes

$$\begin{aligned}D(x_i, x_j) &= \left\|\mathbf{P}^T(x_i - x_j)\right\|^2\\ &= (x_i - x_j)^T\mathbf{A}\mathbf{E}\mathbf{E}^T\mathbf{A}^T(x_i - x_j)\\ &= \left\|\mathbf{E}^T(y_i - y_j)\right\|^2\end{aligned} \tag{11}$$

where y_i and y_j are the low-dimensional embeddings in the LSDA space. Then the cost function in (10) can be rewritten as

$$\begin{aligned}\varepsilon(\mathbf{E}) = {} & (1-\beta)\sum_{ij}\eta_{ij}\left\|\mathbf{E}^T(y_i - y_j)\right\|^2 + \\ & \beta\sum_{ijl}\eta_{ij}(1-\ell_{il})[1 + \left\|\mathbf{E}^T(y_i - y_j)\right\|^2 - \left\|\mathbf{E}^T(y_i - y_l)\right\|^2]_+\end{aligned} \tag{12}$$

On the other hand, the d-by-d matrix $\mathbf{M}=\mathbf{E}\mathbf{E}^T$ (11) can be viewed as a Mahalanobis distance metric and (12) can be reformulated to a semi-definite programming (SDP) problem as [17]

$$\begin{aligned}\min\ & \textstyle\sum_{ij}\eta_{ij}(y_i - y_j)^T\mathbf{M}(y_i - y_j) + \beta\sum_{ijl}\eta_{ij}(1-\ell_{il})\delta_{ijl}\\ s.t.\ & (y_i - y_l)^T\mathbf{M}(y_i - y_l) - (y_i - y_j)^T\mathbf{M}(y_i - y_j) \geq 1 - \delta_{ijl}\\ & \delta_{ijl} \geq 0\\ & \mathbf{M} \succeq 0\end{aligned} \tag{13}$$

The coefficient matrix $\mathbf{E}$ can be solved directly as in (12) [1], [6], which is prone to local minimum [8], or via separating the matrix $\mathbf{M}$ as $\mathbf{M}=\mathbf{E}\mathbf{E}^T$ [7]. Note that the matrix $\mathbf{E}$ can be restricted to a diagonal matrix by adding the additional constraint or other regularization terms in (12) and (13) to avoid overfitting [8].

IV. Acceptance-Rejection Likelihood Modeling For Imposter Classification

The discordant margin distance can be avoided in the proposed margin-enhanced space and the nearest neighbor distance (1-NN distance) can be used for the imposter classification that the unseen subjects whose facial images are excluded from the training recognition face set $\mathbf{X}_F$ will be rejected. Note that $\mathbf{X}_I$ is only composed of some facial images of unseen subjects for the rejection likelihood modeling.

Rather than using a threshold based on the 1-NN distance for the imposter classification, the posterior functions are applied to estimate the acceptance and rejection probability:

$$p(\Omega \mid d) = \frac{p(d \mid \Omega)P(\Omega)}{p(d)}, \ \Omega \in \{\Omega_A, \Omega_R\} \tag{14}$$

where $p(\Omega_A)$ and $p(\Omega_R)$ are the prior probabilities of the acceptance class Ω_A whose facial images are included in the recognition face set $\mathbf{X}_F$, and the rejection class Ω_R whose facial images are excluded from $\mathbf{X}_F$. Both prior probabilities are assumed to be uniformly distribute and $p(\Omega_A) + p(\Omega_R) = 1$. Note that the subjects whose images in the imposter face set $\mathbf{X}_I$ are only part of Ω_R ($X_I \subset \Omega_R$); $p(d \mid \Omega_A)$ and $p(d \mid \Omega_R)$ are the acceptance and rejection likelihood probability, respectively, and d is the 1-NN distance between the image data to its nearest neighbor in the margin-enhanced space. Because the acceptance class Ω_A is composed of c classes, the posterior probability $p(\Omega_A|d)$ can be represented as a marginal probability $p(\Omega_A|d)=\sum_C p(\Omega_A,C|d)$, where $C \in \{1,2,..c\}$ is an index of class. Thus, (14) for acceptance class Ω_A becomes

$$\begin{aligned} p(\Omega_A \mid d) &= \sum_C p(\Omega_A, C \mid d) \\ &= \frac{\sum_C p(\mathrm{d} \mid \Omega_A, C) p(C \mid \Omega_A) p(\Omega_A)}{p(d)} \end{aligned} \tag{15}$$

where the probability $p(C \mid \Omega_A)$ measures the probability of data generated from class C when data is classified as the acceptance class. $p(C \mid \Omega_A)$ can be estimated from the k-NN classification results without additional probability modeling [3], i.e. $p(C=s|\Omega_A)=k_s/k$, where k and k_s is the total number of neighbors and the number of neighbors belonging to class s, respectively. Therefore, based on the 1-NN distance d, an image data x_i can be classified as either acceptance class Ω_A or rejection class Ω_R in accordance with the following Bayes decision rule:

$$x_i \in \begin{cases} \Omega_A, & \text{if } p(\Omega_A \mid d) \geq p(\Omega_R \mid d) \\ \Omega_R, & \text{otherwise} \end{cases} \tag{16}$$

Generally, the prior probability $p(\Omega_A)$ and $p(\Omega_R)$ are unknown, and hence the threshold value of the likelihood ratio θ is employed [12]. Equation (16) is rewritten as

$$\frac{p(d \mid \Omega_A)}{p(d \mid \Omega_R)} \underset{\text{Rejection class}}{\overset{\text{Acceptance class}}{\gtrless}} \frac{p(\Omega_R)}{p(\Omega_A)} = \theta \tag{17}$$

If the likelihood ratio is lower than the threshold value θ, the image data is classified as an imposter (unseen subject) and rejected for further recognition process.

Before modeling the likelihood probabilities $p(d|\Omega_A,C)$ and $p(d|\Omega_R)$, all training face images $\mathbf{X}=(\mathbf{X}_F, \mathbf{X}_I)$ are projected to the margin-enhanced space via the transformation $\mathbf{P}$ (8) and data $\mathbf{Y}=(\mathbf{Y}_F,\mathbf{Y}_I)=(\mathbf{P}^T\mathbf{X}_F, \mathbf{P}^T\mathbf{X}_I)$ are obtained, where $\mathbf{Y}_F = \{y_1,\ldots,y_N\}$ and $\mathbf{Y}_I = \{y_{N+1},\ldots,y_{N+Z}\}$. The label of y_i, $l(y_i)$, is the same as the corresponding data x_i. To build the likelihood model $p(d|\Omega_A,C)$, each data $y_i \in \mathbf{Y}_F$ searches the nearest neighbor $n^1(y_i)$ among the set $\mathbf{Y}_F$ with the 1-NN distance $d_{A,C}(y_i) = \|y_i - n^1(y_i)\|^2$ and the label of $n^1(y_i)$ is $l(n^1(y_i)) \in \{1,2,\ldots,c\}$. If the 1-NN distance is small, the probability of accepting the data is high; otherwise the data might be an imposter. The likelihood probability should decrease or increase smoothly as the 1-NN distance is getting large or small. Hence, for each class $C \in \{1,2,..c\}$, the acceptance likelihood probability $p(d|\Omega_A,C)$ is defined as (Fig. 1):

$$p(d \mid \Omega_A, C) = \begin{cases} 1 & \text{if } d < \mu_{A,C} \\ \dfrac{1}{z_A} \exp\left(\dfrac{-(d - \mu_{A,C})}{\sigma_{A,C}}\right)^2, & \text{otherwise} \end{cases} \tag{18}$$

Where Z_A is the normalization term, $\mu_{A,C}$ and $\sigma_{A,C}$ is the sample mean and sample standard deviation for each class C, which are estimated from samples $d_{A,C}(y_i)$, $\forall y_i = \{y_i \mid y_i \in \mathbf{Y}_F, l(y_i) = C\}$.

As the similar process, each data $y_i \in \mathbf{Y}_I$ searches the nearest neighbors among the set Y_F and all 1-NN distance samples $d_R(y_i) = \|y_i - n^1(y_i)\|^2$ are applied to model the rejection likelihood probability $p(d|\Omega_R)$ as

$$p(d \mid \Omega_R) = \begin{cases} 1 & \text{if } d > \mu_R \\ \dfrac{1}{Z_R} \exp\left(\dfrac{-(d - \mu_R)}{\sigma_R}\right)^2, & \text{otherwise} \end{cases} \tag{19}$$

where Z_R is the normalization term, μ_R and σ_R is the sample mean and sample standard deviation of all 1-NN distance $d_R(y_i)$, $\forall y_i \in \mathbf{Y}_I$. Different from $p(d|\Omega_A,C)$, the large 1-NN distance should be with high rejection probability $p(d|\Omega_R)$ vice versa for small distance. Furthermore, the threshold value θ (17) is set when the sum of recognition rate and rejection rate can be maximum for training data, which is the intersection point as shown in Fig. 2.c. Note that in Fig. 2.c the intersection point (recognition rate equals rejection rate) closer to the top-right corner implies the better system performance.

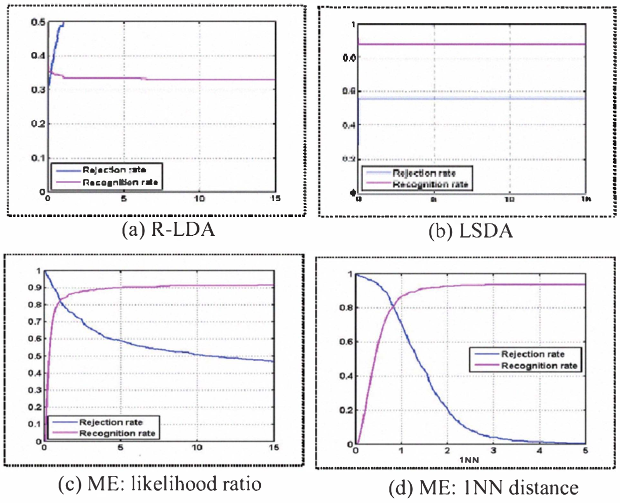
V. Experimental Result

We compare the performance of the proposed margin-enhanced space with other subspace learning methods, including Eigenface (PCA [9, 11]), R-LDA [21], supervised-LPP [22] (designated as SLPP), MFA [20] and LSDA [5], for face recognition and unseen subject rejection under different lightings, poses, and expressions. Note that 3-NN classifier is applied in the following experiments.

Both Extended Yale-B [23] and CMU PIE [24] database are used to evaluate our proposed system. The Extended Yale-B face database contains the facial images of 38 human subjects under 9 poses and 64 illumination conditions. We choose 64 images of frontal pose under all illumination conditions. The CMU PIE face database contains the facial images of 68 human subjects under different poses, illumination conditions, and expressions. In our experiment, five near frontal poses (C05, C07, C09, C27, C29) and the corresponding images under different illuminations and expressions are used; thus we get

TABLE I. Recognition rate in different subspaces of Extended Yale-B and PIE databases. ME: the proposed margin-enhanced space.

	YaleB G30/P34	YaleB G40/P24	PIE G30/P140	PIE G40/P130
PCA	84.6±0.8	86.6±0.8	56.7±0.6	63.9±0.7
R-LDA	90.8±0.8	92.8±0.8	91.2±0.4	93.5±0.3
SLPP	79.4±5.8	82.8±1.1	75.0±8.5	81.1±1.1
MFA	84.1±0.9	85.3±0.9	89.5±0.5	92.2±0.5
LSDA	89.4±0.7	91.3±0.6	91.1±0.4	93.4±0.4
ME	91.6±0.7	93.5±0.6	93.4±0.3	95.0±0.2



(a) R-LDA (b) LSDA

(c) ME: likelihood ratio (d) ME: 1NN distance

Figure 2. Recognition rate (magenta line) vs. rejection rate (blue line) on Extended Yale-B database (G30/P34) under various threshold values θ of the likelihood ratio in spaces: (a) LDA (b) LSDA (c) ME (margin-enhanced space), and (d) the threshold value θ' of 1-NN distance in the margin-enhanced space.

170 images for each subject. For both databases, all selected face images are manually aligned and cropped to 32×32 pixels and PCA is applied to save 98% energy to reduce the noise. The dimensionality of feature subspace is set to c-1 for all subspace learning methods, where c is the number of subject. Especially for Extend Yale-B database the first 5 eigenvalues with corresponding eigenvectors are discarded [5]. In order to evaluate the performance of face recognition and imposter classification, the Extend Yale-B database is partitioned that face images of 20 subjects are used to train the margin-enhanced space and images of the remaining 18 subjects are used to evaluate the rejection rate of imposter classification. Note that the images of 5 subjects among the remaining 18 ones are used as the training imposter data $\mathbf{X_I}$ to model the rejection likelihood probability. The similar partition is applied for the PIE database that face images of 40 subjects and the remaining 28 subjects are used to train the margin-enhanced space and evaluate the rejection performance, respectively. Images of 5 subjects among 28 subjects are used to model the rejection likelihood probability.

The face recognition results performed in different spaces are listed in Table I. For each training and test combination G_p / P_q, where p images per training subject are randomly selected for the training process and the remaining q images are used to evaluate the recognition performance, the mean and the standard deviation of recognition rate performed by 40 random splits are reported. It can be seen that our proposed margin-enhanced space (ME) has higher recognition rate than R-LDA, SLPP, MFA and LSDA, especially for the PIE database. This indicates that by reweighting LSDA space via a trained distance metric can discover a more discriminative structure of the face manifold and improve the recognition performance.

TABLE II. Rejection rate using the proposed likelihood ratio and 1-NN distance in the margin-enhanced space.

	Likelihood Ratio	1-NN Distance
YaleB G30/P34	87.2±1.8	85.6±3.2
YaleB G40/P24	87.6±1.9	86.6±2.3
PIE G30/P140	83.6±1.7	81.6±3.5
PIE G40/P130	85.7±1.7	85.2±3.8

Fig. 2.a-2.c show the recognition and rejection rate with various threshold values θ in different spaces. It can be observed that in the proposed space the system can yield higher recognition rate and rejection rate than in LSDA space (Fig. 2.b and Fig. 2.c). Moreover, to make the comparison with Fig. 2.c and Fig. 2.d, the proposed rejection mechanism using likelihood ratio, has more tolerance for the setting of threshold value and has promising rejection performance than the threshold of using 1-NN distance only (i.e. the subject will be rejected if 1-NN distance>θ'). As Fig. 2.d shown, the 1-NN distance method is sensitive to the threshold value and hence the imprecise setting of the threshold value will result in a very low rejection rate. In addition, Table II lists the rejection rates using the proposed likelihood ratio (17) and 1-NN distance on test database. The threshold value θ is set as discussed in section 4. The proposed likelihood ratio can reject more imposter data than the use of 1-NN distance only.

VI. CONCLUSIONS

In this paper, a margin-enhanced space is proposed by reweighting the LSDA space based on the k-NN classification rule and unit margin constraint. In the proposed space, the local discriminant structure of data can be extracted and the pairwise distance applies for the imposter classification. Moreover, the acceptance and rejection likelihood probability are modeled using the Gaussian distribution and the likelihood ratio can be applied to reject the imposter (the unseen subjects). Experiments on Extended Yale-B and PIE databases have been conducted to demonstrate that the proposed system not only provides higher recognition performance than other subspace learning methods but has the promising rejection performance for imposter classification.

REFERENCES

[1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a Mahalanobis metric from equivalence constraints," Journal of Machine Learning Research, vol. 6, pp. 937-965, 2005.

[2] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Journal of Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2006.

[3] C.M. Bishop, Pattern Recognition and Machine Learning, Springer Science+Business Media, pp. 124-125, 2006.

[4] D. Cai, X. He, and K. Zhou, "Spectral regression for efficient regularized subspace learning," International Conf. on Computer Vision, pp. 1-8, 2007.

[5] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," International Joint Conference on Artificial Intelligence, pp. 708-713, 2007.

[6] J. Goldberger, S.T. Roweis, G.E. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," Advances in Neural Information Processing Systems, pp. 13-18, 2004.

[7] X. He and P. Niyogi, "Locality preserving projections," Advances in Neural Information Processing Systems, pp. 585-591, 2003.

[8] M.P. Kumar, P.H.S. Torr, and A. Zisserman, "An invariant large margin nearest neighbour classifier," International Conf. on Computer Vision, pp. 1-8, 2007.

[9] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, A. Smola, and K.-R. Muller, "Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 623-628, 2003.

[10] H. Murase and S.K. Nayar, "Visual learning and recognition of 3-d objects from appearance," International Journal of Computer Vision, vol. 14, pp.5-24, 1995.

[11] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, no. 5500, pp.2323-2326, 2000.

[12] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," IEEE Conf. on Computer Vision and Pattern Recognition, pp. 746–751, 2000.

[13] B. Sholkopf, A. Smola and K.R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computing, vol. 10, no. 5, pp.1299-1319, 1998.

[14] J.B. Tenebaum, V. de Silva and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science, vol. 290, no. 5500, pp.2319-2323, 2000.

[15] D. Tran and A. Sorokin, "Human activity recognition with metric learning," European Conference on Computer Vision, pp. 548-561, 2008.

[16] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[17] K.Q. Weinberger and L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," Journal of Machine Learning Research, vol. 10, pp.209-244, 2009.

[18] E. Xing, A. Ng, M. Jordan, and S. Russell, "Distance metric learning with application to cluster with side information," Advances in Neural Information Processing Systems, pp. 505-512, 2003.

[19] M. Yang, "Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods," IEEE Conf. on Automatic Face and Gesture Recognition, pp. 215-220, 2002.

[20] S.Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extension: A general framework for dimensionality reduction," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 40-51, 2007.

[21] J.P. Ye and T. Wang, "Regularized discriminant analysis for high dimensional, low sample size data," International Conference on Knowledge Discovery and Data Mining, pp. 454-463, 2006.

[22] Z. Zheng, F. Yang, W. Tan, J. Jia, and J. Yang, "Gabor feature-based face recognition using supervised locality preserving projection," Signal Processing, pp. 2473-2483, 2007.

[23] http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html.

[24] http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261.