Semi-supervised Neighborhood Preserving Discriminant Embedding: A Semi-supervised Subspace Learning Algorithm

Maryam Mehdizadeh[1], Cara MacNish[1], R. Nazim Khan[2], and Mohammed Bennamoun[1]

[1] Department of Computer Science and Software Engineering, University of Western Australia
[2] Department of Mathematics of the University of Western Australia

Abstract. Over the last decade, supervised and unsupervised subspace learning methods, such as LDA and NPE, have been applied for face recognition. In real life applications, besides unlabeled image data, prior knowledge in the form of labeled data is also available, and can be incorporated in subspace learning algorithm resulting in improved performance. In this paper, we propose a subspace learning method based on semi-supervised neighborhood preserving discriminant learning, which we call Semi-supervised Neighborhood Preserving Discriminant Embedding (SNPDE). The method preserves the local neighborhood structure of face manifold using NPE, and maximizes the separability of different classes using LDA. Experimental results on two face databases demonstrate the effectiveness of the proposed method.

1 Introduction

Biometric face data are data of high dimensions and are susceptible to the well-known problem of the *curse of dimensionality* when using machine learning techniques. A common approach is to transform the high dimensional data into a lower dimensional subspace which preserves the perceptually meaningful structure of these images. Fisherface [1], and NPEface [2] are two face subspace learning methods. Fisherface, which is a supervised method based on LDA [3], projects the data points along the directions with optimal class separability, and performs subspace learning based on global Euclidean properties of the image data. NPE on the other hand, is an unsupervised subspace learning method, which performs subspace learning based on local neighborhood properties of the high dimensional image data. In this method, an image is considered as a high dimensional vector, that is, a point in a high dimensional vector space, and the set of all faces are assumed to lie on or near a lower dimensional manifold. The aim of NPE is to discover this manifold structure and perform subspace learning with the objective of best preserving the manifold structure.

R. Kimmel, R. Klette, and A. Sugimoto (Eds.): ACCV 2010, Part III, LNCS 6494, pp. 199–212, 2011.

The assumption of NPE is that nearby points share class information, and recognition of points are based on their closest neighbors in the reduced face subspace. However, in face recognition, variability in illumination and expression makes it hard to discern identities based solely on similarity of images. In other words, images in a small neighborhood might belong to different identities. Therefore, in addition to the neighborhood preserving criteria, there is also a need for discriminant analysis of data, so that the projection of two similar images that belong to different identities is not close in the reduced subspace.

In recent years graph-based subspace learning methods have been studied, which encode discriminant information or manifold structure of image data as graphs and perform subspace learning based on *graph preserving criterion*. Graph Embedding (GE) [4] was introduced as a general framework for dimensionality reduction enabling popular methods of subspace learning to be interpreted and implemented as graph based methods. In addition, Cai et al. [5] provided a general framework for subspace learning, and discussed the possibility of constructing multiple graphs to learn the intrinsic discriminant structure of the image data. In addition, they showed that their framework follows the GE view of subspace learning.

In this paper, along the framework introduced by Cai et al. [5] for content-based image retrieval, we propose a semi-supervised subspace learning method for face recognition which uses two graphs that are constructed to encode the necessary information of image data. We call this Semi-supervised Neighborhood Preserving Discriminant Embedding (SNPDE) for face representation and recognition. Our method is constructed based on: (i) graph view of NPE, which builds an adjacency graph that best reflects the geometry of the face manifold; and (ii) graph view of LDA, which builds a graph with edge weights that reflect the discriminant structure of data. The projection function then consists of a set of basis vectors obtained based on a unified objective function incorporating the graph preserving of NPE and LDA. Since SNPDE combines the objective of NPE with discriminative objective of LDA, it is expected to perform better than NPE for face recognition, and this is demonstrated in our results section.

The rest of the paper is organized as follows. In Section 2. We review GE view of subspace learning and discuss the graph view of NPE and LDA. The SNPDE method is described in Section 3. The experimental results are discussed and compared with other methods in Section 4, followed by concluding remarks in Section 5.

2 Graph Embedding and Graph Based NPE and LDA

2.1 Graph Embedding View of Subspace Learning

A given set $\{\mathbf{x_i}\}_{i=1}^{N} \subset \mathbf{R}^n$ of N images can be represented as an image matrix $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$. The essential task of subspace learning is to find an optimal mapping function that projects the high dimensional face data into a lower dimensional face space $Y = \{\mathbf{y_i}\}_{i=1}^{N} \subset R^d$,where $d << n$. That is,

$$Y = \mathbf{A}^T X \tag{1}$$

where A is an $n \times d$ matrix consisting of a set of basis vectors $\mathbf{a}^* = [\mathbf{a}_1, \ldots, \mathbf{a}_d]$, (where $\mathbf{a}_i \in \mathbf{R}^n$ for $i = 1, \ldots, d$). In graph based subspace learning methods, the data vector is represented as a graph G, such that vertex i of the graph represents vector $\mathbf{x}_i$. An $n \times n$ weight matrix W is then defined such that each edge weight W_{ij} reflects the relationship between data points $\mathbf{x}_i$ and $\mathbf{x}_j$. The objective of Graph Embedding (GE) is to represent each vertex of a graph as a low dimensional vector where the relationship between vertex pairs (i, j) is best preserved.

The GE formulation of subspace learning is as follows

$$\mathbf{a}^* = \underset{a:\ y^T D y=1}{\arg\min} \sum_{i \neq j} (\mathbf{a}^T \mathbf{x}_i - \mathbf{a}^T \mathbf{x}_j)^2 W_{ij} \tag{2}$$

$$= \underset{a:\ y^T D y=1}{\arg\min}\ \mathbf{a}^T X L X^T \mathbf{a} \tag{3}$$

where $L = D - W$ is the *graph Laplacian* [6], and D is a diagonal matrix whose entries are column (or, since W is symmetric, row) sums of W. The optimization in (3) can also be written as

$$\mathbf{a}^* = \underset{a}{\arg\min}\ \frac{\mathbf{a}^T X L X^T \mathbf{a}}{\mathbf{a}^T X D X^T \mathbf{a}} \tag{4}$$

which reduces to solving the general eigenvalue problem

$$X L X^T \mathbf{a} = \lambda X D X^T \mathbf{a}, \tag{5}$$

or equivalently as

$$\mathbf{a}^* = \underset{a}{\arg\max}\ \frac{\mathbf{a}^T X W X^T \mathbf{a}}{\mathbf{a}^T X D X^T \mathbf{a}}, \tag{6}$$

which reduces to solving the general eigenvalue problem

$$X W X^T \mathbf{a} = \lambda X D X^T \mathbf{a}. \tag{7}$$

In sections 2.2 and 2.3, we will show that graph view of NPE and LDA can be derived from the GE formulations in (4) and (6).

2.2 Graph View of NPE

In this section, we will review the NPE algorithm and discuss the graph G^{NPE} which can model the learnt manifold structure based on NPE algorithm. We will also reformulate the NPE method based on the GE formulation of subspace learning in (4), and this will help us develop our SNPDE algorithm.

NPE [2] is an unsupervised subspace learning method that inherits the local linear but global nonlinear learning characteristics of Locally Linear Embedding (LLE) [7], a well-known nonlinear dimensionality reduction method. Unlike LLE, which is only applicable to input training data, NPE obtains a linear mapping function based on the training data, that is also applicable to unseen test data. NPE assumes data to lies on a nonlinear manifold and obtains its linear mapping with the aim that the local neighborhood characteristics of the manifold are best preserved. Similar to LLE, NPE characterizes the local neighborhood structure of each data point by linear coefficients W_{ij} such that each data point $\mathbf{x}_i$ can be (approximately) reconstructed from its k-neighbors $\{\mathbf{x}_j\}_{j=1}^{k}$ by $\hat{\mathbf{x}}_i = \sum_{j=1}^{k} W_{ij}\mathbf{x}_j$. NPE then obtains a linear mapping function such that the local linear characteristics identified by W_{ij} are best preserved in the lower dimensional subspace. The actual computations for obtaining the linear mapping by NPE involve solving a generalized eigenvalue problem derived from the cost function of LLE.

The computations for NPE can be divided into the following three steps.

1. **Construct the neighborhood graph**
   The neighborhood weights are obtained by the following optimization:

$$\min \sum_i \left\| \mathbf{x}_i - \sum_j W_{ij}\mathbf{x}_{N(j)} \right\|^2 \tag{8}$$

   subject to the constraints:

$$\sum_{j=1}^{k} W_{ij} = 1, \text{for each } j = 1, \ldots, k. \tag{9}$$

   The neighborhood weight matrix W forms a $k \times n$ matrix, where k is the number of neighboring points for each image data and n is the number of data points in the image database. The $n \times n$ weight matrix W^{NPE} of graph G^{NPE} is obtained as

$$W_{ij}^{NPE} = \begin{cases} (W + W^T - W^T W)_{ij} & \text{if i} \neq \text{j} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

   The structure of the similarity weight matrix W_{ij}^{NPE} was first introduced in [4].

2. **Compute the optimal linear projections**
He et al. [2] obtained the neighborhood preserving mapping of NPE, that is, they obtained the matrix A such that the mapping $X \rightarrow Y$ where $Y = A^T X$ preserves the neighborhood characteristics of the data manifold. Thus similar to (8), we select $\mathbf{y}^* = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ as:

$$\mathbf{y}^* = \arg\min_{y} \sum_i \left\| \mathbf{y}_i - \sum_j W_{ij} \mathbf{y}_{N(j)} \right\|^2 \tag{11}$$

$$= \arg\min_{y:\ y^T y = 1} \mathbf{y}^T (I - W)^T (I - W) \mathbf{y} \tag{12}$$

$$= \arg\min_{a:\ a^T X X^T a = 1} \mathbf{a}^T X (I - W)^T (I - W) X^T \mathbf{a} \tag{13}$$

$$= \arg\min_{a:\ a^T X X^T a = 1} \mathbf{a}^T X M X^T \mathbf{a} \tag{14}$$

Here I is the $n \times n$ identity matrix and M is the $n \times n$ matrix given by

$$M = (I - W)^T (I - W). \tag{15}$$

The optimal linear projections in (14) are the eigenvectors associated with smallest eigenvalues of the generalized eigenvalue problem

$$X M X^T \mathbf{a} = \lambda X X^T \mathbf{a} \tag{16}$$

When NPE is applied on face image data, the eigenvectors $\mathbf{a}$ are called *NPE-faces*.

Yan et al. [4] discussed that the LLE algorithm can be considered as the direct GE formulation in (3). The matrix M in (14) can be considered as the Laplacian matrix L^{NPE} of the graph G^{NPE}, that is $M = D - W^{NPE} = L^{NPE}$, giving

$$\mathbf{y}^* = \arg\min_{a:\ a^T X X^T a = 1} \mathbf{a}^T X L^{NPE} X^T \mathbf{a} \tag{17}$$

This GE formulation of NPE in (17) will help us develop our algorithm.

2.3 Graph View of LDA

LDA is a supervised linear subspace learning algorithm that obtains a discriminant projection function according to class label information of the input data.

The aim of LDA is to find projection directions that maximize the separability of data points belonging to different classes while simultaneously minimizing the distance between data of the same class. Suppose we have N high dimensional image vectors belonging to l classes of faces. LDA maximizes the ratio of the between-class scatter S_b to the within-class scatter S_w, where

$$S_b = \sum_{k=1}^{c} l_k (\overline{\mathbf{x}}^{(k)} - \overline{\mathbf{x}})(\overline{\mathbf{x}}^{(k)} - \overline{\mathbf{x}})^T \tag{18}$$

and

$$S_w = \sum_{k=1}^{c} (\sum_{i=1}^{l_k} (\mathbf{x}_i{}^{(k)} - \overline{\mathbf{x}}^{(k)})(\mathbf{x}_i{}^{(k)} - \overline{\mathbf{x}}^{(k)})^T), \tag{19}$$

where $\overline{\mathbf{x}}$ is the total sample mean vector, l_k is the number of samples in the k-th class and $\mathbf{x}_i^{(k)}$ is the i-th sample in the k-th class. That is, LDA selects optimal $\mathbf{a}$'s as

$$\mathbf{a}^* = \arg\max_a \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_w \mathbf{a}}. \tag{20}$$

Define the total scatter matrix S_t as [8]

$$S_t = \sum_{i=1}^{l} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T. \tag{21}$$

It can be easily shown that

$$S_t = S_b + S_w. \tag{22}$$

Therefore, the optimization (20) can be rewritten as

$$\mathbf{a}^* = \arg\max_a \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_t \mathbf{a}}. \tag{23}$$

That is, the optimal $\mathbf{a}$'s are the eigenvectors corresponding to the largest non-zero eigenvalues of the generalized eigenvalue problem

$$S_b \mathbf{a} = \lambda S_t \mathbf{a} \tag{24}$$

According to Yan et al. [4], the LDA algorithm can also be reformulated as a direct GE by constructing c complete subgraphs $\{G_k\}_{k=1}^c$ each representing data belonging to the corresponding class. Assuming each subgraph G_k has l_k vertices, the weights of each subgraph are defined as an $l_k \times l_k$ weight matrix $W^{(K)}$ with each element equal to $1/l_k$.

Assuming that $X^{(k)} = [X_1^{(k)}, \ldots, X_{l_k}^{(k)}]$ is the data matrix of the k-th class, the between class scatter S_b can be written as

$$S_b = \sum_{k=1}^{c} X^{(k)} W^{(k)} (X^{(k)})^T \tag{25}$$

and the total scatter matrix S_t can be written as

$$S_t = XX^T \tag{26}$$

where X is the data matrix. If the data are ordered based on their class labels, so $X = [X^{(1)}, \ldots, X^{(c)}]$, then the $l \times l$ weight matrix $W_{l\times l}$ of the graph G^{LDA} consisting of all the c subgraphs is defined as

$$W_{l\times l} = \begin{bmatrix} W^{(1)} & 0 & \ldots & 0 \\ 0 & W^{(2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & W^{(c)} \end{bmatrix} \tag{27}$$

Then the optimization in (23) can be rewriteen as

$$\mathbf{a}^* = \underset{a}{\operatorname{argmax}} \frac{\mathbf{a}^T X W_{l\times l} X^T \mathbf{a}}{\mathbf{a}^T X X^T \mathbf{a}} \tag{28}$$

This formulation of LDA in (28) was first introduced in [9] and will help us develop our semi-supervised learning method.

3 Semi-supervised Neighborhood Discriminant Embedding

In this section we develop a semi-supervised subspace learning algorithm which incorporates the manifold structure provided by unlabeled data and the discriminant structure provided by labeled data. Cai et al. [5] provided a general framework for semi-supervised subspace learning for Content Based Image Retrieval (CBIR) and discussed the possibility of constructing multiple graphs to learn the intrinsic discriminant structure of the image data. Following the general framework in [5], we construct two graphs; one to encode the neighborhood preserving information based on the NPE method and the other to encode discriminant class label information based on the LDA method. We exploit the information encoded by the two graphs by formulating a constrained optimization problem consisting of the GE objectives of NPE and LDA. The computation of the projection function reduces to solving a general eigenvalue problem. The Semi-supervised Neighborhood Preserving Discriminant Embedding (SNPDE) algorithm enables us to introduce a new image representation and an improved precision for subspace learning and classification of face images.

3.1 The Objective Function

Let $X_l = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l]$ be the labeled data set and $X_{l+1} = [\mathbf{x}_{l+1}, \ldots, \mathbf{x}_n]$ be the unlabeled data set, where each sample $\mathbf{x}_i$ $(i = 1, \ldots, n)$ is from one of c classes. Let l_k be the number of labeled samples in class k, $(k = 1, \ldots, c)$, so $(\sum_{k=1}^{c} l_k = l)$. Put $X = [\mathbf{X}_l, \mathbf{X}_{l+1}]$ and

$$W^{LDA} = \begin{bmatrix} W_{l\times l} & 0 \\ 0 & 0 \end{bmatrix} \tag{29}$$

where the $W_{l\times l}$ is defined in (27). The SNPDE objective consists of two parts corresponding to the objectives of the graph views of LDA and NPE. Put

$$O_{LDA} = \arg\max_{a} \frac{\mathbf{a}^T X W^{LDA} X^T \mathbf{a}}{\mathbf{a}^T X X^T \mathbf{a}} \tag{30}$$

and

$$O_{NPE} = \arg\min_{a} \mathbf{a}^T X L^{NPE} X^T \mathbf{a}. \tag{31}$$

When sufficient labeled data is not available, the LDA algorithm tends to overfit the objective function. In order to avoid overfitting, we use the regularized version of LDA [3]:

$$\arg\max_{a} \frac{\mathbf{a}^T X W^{LDA} X^T \mathbf{a}}{\mathbf{a}^T X X^T + \alpha J(\mathbf{a})} \tag{32}$$

where $J(a)$ is the regularization term. This term provides us the flexibility to incorporate graph objective of NPE in the GE objective of LDA. The combination of LDA objective with other graph based objective function for subspace learning are discussed and applied in [10] and [5]. We append the graph embedding criteria of NPE as a regularization term to LDA. That is, define

$$O_{SSNPE} = O_{LDA} \text{“}+\text{”} O_{NPE} \tag{33}$$

$$= \arg\max_{a} \frac{\mathbf{a}^T X W^{LDA} X^T \mathbf{a}}{\mathbf{a}^T X X^T \mathbf{a} + \alpha \mathbf{a}^T X L^{NPE} X^T \mathbf{a}} \tag{34}$$

$$= \arg\max_{a} \frac{\mathbf{a}^T X W^{LDA} X^T \mathbf{a}}{\mathbf{a}^T X (\tilde{I} + \alpha L^{NPE}) X^T \mathbf{a}} \tag{35}$$

where $\tilde{I}$ is defined as

$$\tilde{I} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \tag{36}$$

and I is the $l \times l$ identity matrix.

The objective function in (35) reduces to solving the maximum eigenvalue solution to the generalized eigenvalue problem

$$XW^{LDA}X^T\mathbf{a} = \lambda X(\tilde{I} + \alpha L^{NPE})X^T\mathbf{a} \tag{37}$$

To get a stable solution of (37), the matrix $X(\tilde{I} + \alpha L^{NPE})X^T$ is required to be non-singular [11], which is not the case when the image dimension is larger than the number of image samples. We apply Tikhonov Regularization [12], a well studied solution to ill-posed problems in statistics, to solve the singularity problem. The generalized eigenvalue problem in (37) then becomes

$$XW^{LDA}X^T\mathbf{a} = \lambda \left(X(\tilde{I} + \alpha L^{NPE})X^T + \beta I)\right)\mathbf{a} \tag{38}$$

which has stable solutions for $\beta > 0$.

3.2 The Algorithm

The SNPDE algorithm consists of the following steps.

1. **Construct the labeled graph G^{LDA}:** Construct the $n \times n$ weight matrix W^{LDA} of the labeled graph.
2. **Construct the unlabeled graph G^{NPE}:** Construct the k-nearest neighbor graph matrix W^{NPE} based on (10) and calculate the graph laplacian $L^{NPE} = D - W^{NPE}$, where D is a diagonal matrix with entries the column (since W^{NPE} is symmetric, or row) sums of W^{NPE} that is, $D_{ii} = \sum_j W^{NPE}_{ij}$.
3. **Compute the projection matrix:** The $n \times c$ transformation matrix $A = [\mathbf{a}_1, \ldots, \mathbf{a}_c]$ consists of eigenvectors corresponding to the largest non-zero eigenvalues of the generalized eigenvalue problem in (38). Since W^{LDA} is of rank c, we will have exactly c eigenvectors corresponding to the nonzero eigenvalues.

4. **Embed sample images into $c-$dimensional subspace:** Each image sample can be embedded into $c-$dimensional subspaces by

$$\mathbf{x}_i \rightarrow \mathbf{y}_i = A^T\mathbf{x}_i$$

4 Experiments and Discussions

We present experiments and comparisons to demonstrate the effectiveness of our proposed semi-supervised subspace learning algorithm. In section 4.1 we describe the face image datasets that we used in our experiments. In section 4.2 we illustrate the face representations in lower dimensional subspace. The implementation details and recognition error rates are reported in section 4.3.

4.1 Data Sets

We tested our proposed method on two face databases of CMU PIE [13], and ORL [14]. The CMU PIE database contains 68 subjects with 41,368 images of varying poses, lighting and expressions. The ORL database includes 400 images of 40 individuals under different poses and expressions. In our experiments on the PIE database, we chose the frontal pose $C27$ with varying lighting and illumination which leaves us with 43 images for each subject. In our experiments on ORL database, we used all of the available 400 images in the dataset. Figure 1 and 2 show a sample of images from PIE and ORL databases respectively.

The original images from the CMU PIE database were cropped (The ORL images were already cropped) and the cropped images from both databases were then resized to 32×32 pixels. Each image was represented by a 1024-dimensional vector in the original image space. The training dataset which included labeled and unlabeled data was used to learn a projection matrix to project the high dimensional face images to a lower dimensional subspace. We then applied the nearest neighbor classifier in the subspace to determine the recognition error rate of the unlabeled data and the unseen test data. In all cases the training and the test datasets were randomly selected from the database without mixing between the training and testing data points. The results were averaged over 20 different runs.

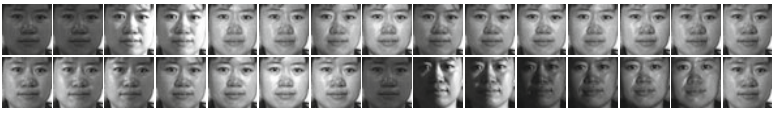
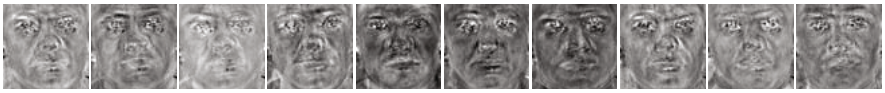


Fig. 1. Sample face images of the CMU PIE face database. Each subject has 43 different images of frontal poses under different lighting conditions.

Fig. 2. Sample face images of ORL face database. Each subject has 10 face images with a different pose and expression.

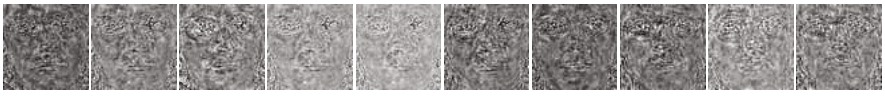
4.2 Face Representation

As mentioned earlier, a high dimensional vector such as the face image vector is prone to the curse of dimensionality and is better studied in lower dimensional subspaces. We compare three algorithms - NPE, SDA, and SNPDE for face representation. In each of these methods, basis functions are thought of as basis images, where each sample image is constructed as a linear combination of the basis images. In Figure. 3, we illustrate first 10 $SNPDEfaces$ together with $NPEfaces$ and $SDAfaces$.



(a) NPEfaces

(b) SDAfaces



(c) SNPDEfaces

Fig. 3. The first 10 NPEfaces, SDAfaces, and SNPDEfaces obtained from samples from the PIE database

4.3 Face Recognition

Table 1 and Table 2 summarize the recognition error rates of four different algorithms. The baseline approach is simply the nearest neighbor approach on the original image space. For the other approaches, training images (labeled and unlabeled) are used to learn a subspace - in NPE approach the training data is constructed in a similar way to the SDA and SNPDE approach, only NPE considers labeled training data as unlabeled. After learning the projection function and projecting the high dimensional data to the image subspace, nearest neighbor classification is performed for recognition purposes. There are two kinds of error rates reported here; the unlabeled error rate, and the test error rate. Although the unlabeled data are used in the training stage, their labels still need to be recognized by the subspace learning algorithm. Therefore, the unlabeled

Table 1. Comparison of recognition error rates on PIE database

Number of Labeled Samples	**Error Rate**(%)							
	Baseline(1024)		**SDA(68)**		**NPE(30)**		**SNPDE(68)**	
	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
1	68.10	68.80	61.12	61.23	55.02	57.06	**49.77**	**49.75**
2	56.40	56.49	43.26	43.78	39.93	45.58	**31.45**	**33.61**
3	51.36	46.43	32.74	29.21	36.30	36.19	**23.72**	**22.45**
4	45.13	46.32	25.03	25.26	30.31	33.81	**18.98**	**20.86**
5	39.11	38.47	19.18	18.04	25.55	28.66	**10.80**	**12.70**
6	30.67	33.17	15.49	14.58	20.85	23.37	**8.20**	**8.22**
7	26.89	27.68	12.66	9.64	17.34	20.52	**5.24**	**4.69**
8	29.50	27.25	11.02	9.48	18.85	20.66	**5.92**	**5.88**
9	26.50	24.07	8.81	7.18	16.00	17.88	**4.51**	**4.42**
10	**18.02**	**17.53**	**4.39**	**4.33**	**11.36**	**14.37**	**1.97**	**2.95**

Table 2. Comparison of recognition error rates on ORL database

Number of Labeled Samples	Error Rate(%)							
	Baseline(1024)		SDA(68)		NPE(30)		SNPDE(68)	
	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
1	31.03	30.38	27.67	29.38	31.34	39.00	**21.27**	**21.00**
2	17.61	17.38	17.34	22.75	20.32	38.13	**13.80**	**15.13**
3	10.90	11.25	11.50	15.25	11.88	30.88	**7.73**	**10.25**
4	7.10	7.50	9.93	12.50	8.78	30.25	**5.88**	**9.00**
5	**5.34**	**5.25**	**8.00**	**9.00**	**7.47**	**28.50**	**4.69**	**5.75**

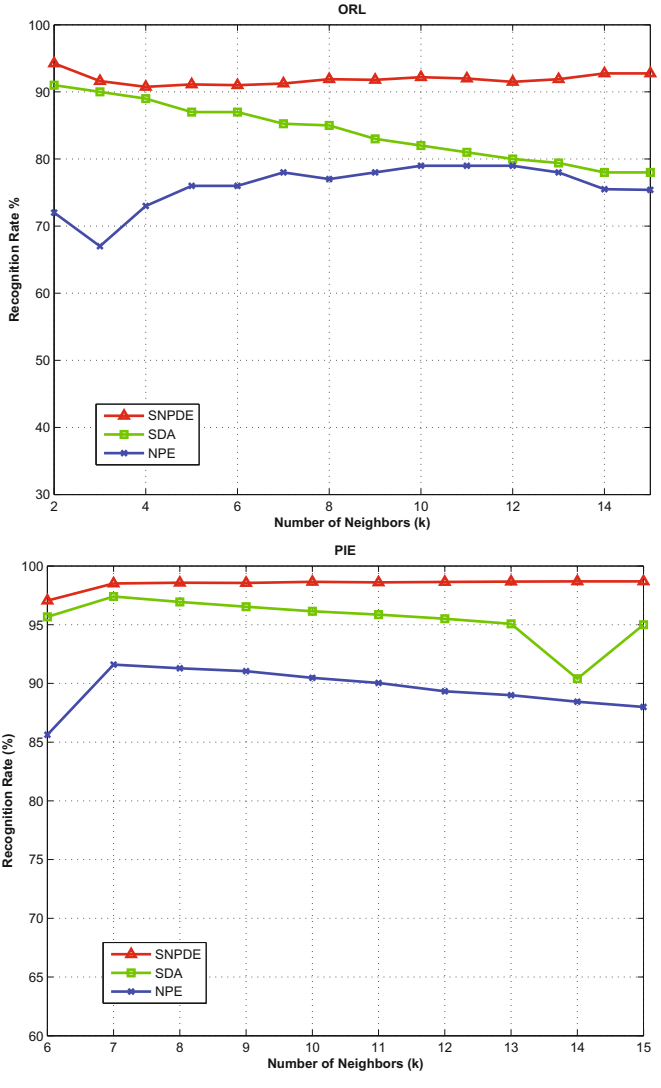


Fig. 4. The effect of the number of neighbors (k) on the performance of the three subspace learning algorithms discussed in this paper

error rate is the error rate associated to the unlabeled data used at the training stage, and the test error rate is the error rate associated with the unseen test images.

The nearest neighbor approach does not consider the manifold structure, and since its decision making is only based on Euclidean distance between images, it provides a very poor performance due to illumination and pose changes. The other approaches learn from the manifold structure, and their difference in performance is due to whether or not they take into account labeled information in their algorithm, and the way the manifold structure is modeled by graphs. SDA is a subspace learning algorithm that considers both labeled and unlabeled data, but since its graph cannot model the manifold structure as accurately as the NPE algorithm does, its performance is inferior to SNPDE.

The error rate of NPE decreases, by increasing the amount of data used in its training stage. The error rate of SDA and SNPDE decreases by increasing the amount of labeled data used at their training stage.

Figure 4 illustrates the sensitivity of three graph-based subspace learning algorithms - NPE, SDA, and SNPDE - to the number of nearest neighbors k in the construction of graphs. The performance of graph-based subspace learning algorithms depend on whether a data point and its nearest neighbors belong to the same class. Therefore, when the number of points in each class in the training data is less than the number of nearest neighbors k, then the possibility of nearest neighbors belonging to different classes increases, consequently reducing the performance of these graph-based methods. This is the case with ORL, a small data set. In contrast, PIE is a large dataset, so in this case all the methods are less sensitive to k. However, the SNPDE still maintains the highest recognition rate of all three algorithms and also is less sensitive to k for both datasets.

5 Conclusion

In this paper, we propose a new linear subspace learning algorithm called Semi-supervised Neighborhood Discriminant Embedding. It can learn from both labeled and unlabeled data to optimize the projection matrix based on both discriminant and geometrical information of high dimensional data. The experimental results on PIE and ORL database demonstrate the effectiveness of our algorithm. As in real applications of biometric face recognition, data becomes available to the system in incremental fashion, we will consider incremental semi-supervised learning based on SNPDE in our future work.

Acknowledgement. We would like to thank Professor Gordon Royle of the Department of Mathematics at the University of Western Australia and Dr. Ashraf Daneshkhah of the Department of Mathematics of the University of Bu-Ali Sina for their valuable discussions and comments. The authors would like to acknowledge the financial support of the Australian Research Council (ARC). This paper is related to ARC DP0771294.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Anal. Mach. Intell. 19, 711–720 (1997)
2. He, X., Cai, D., Yan, S., Zhang, H.: Neighborhood preserving embedding. In: ICCV, pp. 1208–1213 (2005)
3. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York (2001)
4. Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., Lin, S.: Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. 29, 40–51 (2007)
5. Cai, D., He, X., Han, J.: Spectral regression: a unified subspace learning framework for content-based image retrieval. In: ACM Multimedia, pp. 403–412 (2007)
6. Chung, F.R.K.: Spectral Graph Theory. The Regional Conference Series in Mathematics, vol. 92. AMS, Providence (1997)
7. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
8. Fukunnaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press, London (1991)
9. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face recognition using laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. 27, 328–340 (2005)
10. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: ICCV, pp. 1–7 (2007)
11. Watkins, D.S.: Fundamentals of matrix computations. John Wiley & Sons, Inc., New York (1991)
12. Lauter, H., Liero, H.: Ill-posed inverse problems and their optimal regularization (1997)
13. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces. Technical Report CMU-RI-TR-01-02, Robotics Institute, Pittsburgh, PA (2001)
14. ORL database: Cambridge University Computer Laboratory (2002), `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`