Minimal Local Reconstruction Error Measure Based Discriminant Feature Extraction and Classification

Jian Yang Zhen Lou Zhong Jin Jing-yu Yang
School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, 210094, P.R. China
{csjyang,louzhen,zhongjin,yangjy}@mail.njust.edu.cn

Abstract

This paper introduces the minimal local reconstruction error (MLRE) as a similarity measure and presents a MLRE-based classifier. From the geometric meaning of the minimal local reconstruction error, we derive that the MLRE-based classifier is a generalization of the conventional nearest neighbor classifier and the nearest neighbor line and plane classifiers. We further apply the MLRE measure to characterize the within-class and between-class local scatters and then develop a MLRE measure based discriminant feature extraction method. The proposed MLRE-based feature extraction method is in line with the MLRE-based classification method in spirit, thus the two methods can be seamlessly combined in applications. The experimental results on the CENPARMI handwritten numeral database and the FERET face image database show effectiveness of the proposed MLRE-based feature extraction and classification method.

1. Introduction

The nearest neighbor (NN) classifier is widely-used in pattern classification due to its simplicity and effectiveness. Cover and Hart have shown that in large sample cases, the error rate of the NN classifier is bounded above by twice the Bayes error rate [1]. In practice, the performance of the NN-classifier depends on the representational capacity of prototypes as well as on how many prototypes are available. Li and Lu [2] proposed the nearest feature line (NFL) method to generalize the representational capacity of the available limited prototypes. In a feature space, the NFL method uses a feature line to interpolate and extrapolate each pair of prototype feature points belonging to the same class. The feature line virtually provides an infinite number of prototype feature points of the class. The representational capacity of the prototypes is thus expanded. Chien and Wu [3] further extended Li and Lu's work on NFL and proposed the nearest feature plane (NFP) and the nearest feature space (NFS) methods for pattern classification.

Since the NFL method is conducted for each pair of available prototypes, it faces the large computation complexity problem when there are many prototypes in each class. The NFP method also faces the similar problem due to its increased computational requirement. The nearest neighbor line (NNL) and the nearest neighbor plane (NNP) [4] methods were suggested to alleviate the computation complexity of the NFL and NFP methods. Because only one feature line or a feature plane in each class need to be computed, The NNL and NNP methods are computationally more efficient than the NFL and NFP methods. Recently, the hit-distance based nearest neighbor classifiers were proposed to enhance the generalization power of the NNL and NNP methods [5].

Motivated by the NNL and NNP methods, we propose a minimal local reconstruction error (MLRE) based classification method. From the geometric meaning of the minimal local reconstruction error, it is easy to know that the MLRE-based classifier is a generalization of the NN and NNL/NNP classifier. Specifically, when the nearest neighborhood parameter K=1, the MLRE-based classifier is equivalent to the nearest neighbor classifier. When K = 2, the MLRE-based classifier is the nearest neighbor line classifier [4], and when K = 3, the MLRE-based classifier is the nearest neighbor plane classifier [4]. When K>3, the MLRE-based classifier is the nearest neighbor "space" classifier, which is a further generalization of the NNL/NNP classifier.

If the dimension of the input space is very high, it must be time-consuming to perform classification directly in the input space. In addition, performing classification in the high-dimensional space might encounter the so-called "dimensionality curse". Therefore, to avoid these problems, we would rather perform feature extraction first before the classification step. In this paper, we will further apply the idea of the MLRE measure to feature extraction and develop a MLRE measure based discriminant feature extraction method.

Regarding discriminant feature extraction, the most well-known method is Fisher linear discriminant analysis (FLD or LDA). LDA seeks to find a projection axis such that the Fisher criterion (i.e., the ratio of the between-class scatter to the within-class scatter) is maximized after the

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

projection of samples. The between-class scatter and the within-class scatter of LDA are characterized based on the Euclidian distance measure. Recently, the canonical angle was introduced as a measure to characterize the between-class set similarities and the within-class set similarities, based on which a discriminative learning method was developed for the recognition of the image set classes [6]. In addition, the locality characterization related discriminant feature extraction techniques are becoming popular following the development of manifold learning. The representative techniques include: Locality Preserving Projections (LPP) [7], Neighbor Preserving Embedding (NPE) [8], Marginal Fisher Analysis (MFA) [9], and Unsupervised Discriminant Projection [10], etc.

As opposed to LDA, the proposed MLRE-based feature extractor uses the local reconstruction error measure instead of the Euclidean distance to characterize the within-class and between-class scatters. It can thus preserve the locality characteristics of the data. In addition, the proposed MLRE-based feature extraction method is in line with the MLRE-based classification method in spirit. The proposed feature extractor and classifier can be seamlessly combined in their applications. In contrast, the other locality characterization based feature extraction methods do not have this advantage because it is difficult to find a classifier matching the method itself well.

2. Minimal Local Reconstruction Error Measure Based Classification

2.1. Minimal Local Reconstruction Error (MLRE)

Let $\mathfrak{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ be a set of N points in a high-dimensional observation space $\mathfrak{R}^D$. For a given point $\mathbf{x}$, we find its K nearest neighbors from $\mathfrak{X}$. Let $\Omega = \{j \mid \mathbf{x}_j$ belongs to the set of K nearest neighbors of $\mathbf{x}\}$. We try to find a set of reconstruction weights which minimize the following reconstruction error:

$$\varepsilon = \left\| \mathbf{x} - \sum_{j \in \Omega} w_j \mathbf{x}_j \right\|^2, \text{ subject to } \sum_{j \in \Omega} w_j = 1, \qquad (1)$$

where $\|\cdot\|$ denotes the Euclidean norm.

The optimal reconstruction weights w_j^* can be calculated using the algorithm suggested in [11]. The corresponding minimal local reconstruction error is $\varepsilon_{\min} = \left\| \mathbf{x} - \sum_{j \in \Omega} w_j^* \mathbf{x}_j \right\|^2$.

2.2. Geometric Meaning of the MLRE Measure

The minimal local reconstruction error $\varepsilon_{\min}$ actually defines a distance measure from $\mathbf{x}$ to the data set $\mathfrak{X}$. This subsection will expatiate on the geometric meaning of $\varepsilon_{\min}$ from the distance point of view.

When K = 1, let $\mathbf{x}_k$ be the 1-nearest neighbor of data point $\mathbf{x}$. Here, $w_k = 1$ since $\sum_{j \in \Omega} w_j = 1$. Thus, $\varepsilon_{\min}$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_k$, that is, $\varepsilon_{\min} = \|\mathbf{x} - \mathbf{x}_k\|^2$.

When K = 2, let $\mathbf{x}_k$ and $\mathbf{x}_l$ be the 2-nearest neighbors of data point $\mathbf{x}$. The two points $\mathbf{x}_k$ and $\mathbf{x}_l$ determine a straight line $\mathfrak{L}$ in the observation space, which is called nearest neighbor line of $\mathbf{x}$ [4]. The linear combination of $\mathbf{x}_k$ and $\mathbf{x}_l$

$$\begin{aligned} \mathbf{z} &= \sum_{j \in \Omega} w_j \mathbf{x}_j = w_k \mathbf{x}_k + w_l \mathbf{x}_l \\ &= (1 - w_l)\mathbf{x}_k + w_l \mathbf{x}_l = \mathbf{x}_k + w_l(\mathbf{x}_l - \mathbf{x}_k) \end{aligned} \qquad (2)$$

is an arbitrary point on the line $\mathfrak{L}$, as shown in Figure 1 (a). Obviously, when $0 \le w_{il} \le 1$, $\mathbf{z}$ lies on the line segment $\overline{\mathbf{x}_k \mathbf{x}_l}$; when $w_{il} < 0$, $\mathbf{z}$ lies on the left side of $\mathbf{x}_k$; when $w_{il} > 1$, $\mathbf{z}$ lies on the right side of $\mathbf{x}_l$.

Geometrically, minimizing the reconstruction error in Eq. (1) is to find a point $\hat{\mathbf{x}}$ on the line $\mathfrak{L}$ that is closest to $\mathbf{x}$. This point must be the projection of $\mathbf{x}$ onto the line $\mathfrak{L}$. So, $\hat{\mathbf{x}}$ is actually the projection of $\mathbf{x}$ onto its nearest neighbor line $\mathfrak{L}$. The minimal reconstruction error $\varepsilon_{\min}$ is actually the distance from $\mathbf{x}$ to its nearest neighbor line $\mathfrak{L}$, that is, the length of the line segment $\overline{\mathbf{x}\hat{\mathbf{x}}}$.

When K = 3, let $\mathbf{x}_k$, $\mathbf{x}_l$ and $\mathbf{x}_m$ be the 3-nearest neighbors of data point $\mathbf{x}$. Assume that $\mathbf{x}_k$, $\mathbf{x}_l$ and $\mathbf{x}_m$ are linearly independent. The three points determine a plane $\mathfrak{P}$ in the observation space, which is called nearest neighbor plane of $\mathbf{x}$ [4]. The linear combination of $\mathbf{x}_k$, $\mathbf{x}_l$ and $\mathbf{x}_m$

$$\begin{aligned} \mathbf{z} &= \sum_{j \in \Omega} w_j \mathbf{x}_j = w_k \mathbf{x}_k + w_l \mathbf{x}_l + w_m \mathbf{x}_m \\ &= (1 - w_l - w_m)\mathbf{x}_k + w_l \mathbf{x}_l + w_m \mathbf{x}_m \\ &= \mathbf{x}_k + w_l(\mathbf{x}_l - \mathbf{x}_k) + w_m(\mathbf{x}_m - \mathbf{x}_k) \end{aligned} \qquad (3)$$

is an arbitrary point on the plane $\mathfrak{P}$, as shown in Figure 1 (b).

Geometrically, minimizing the reconstruction error in Eq. (1) is to find a point $\hat{\mathbf{x}}$ on the plane $\mathfrak{P}$ that is closest to $\mathbf{x}$. This point must be the projection of $\mathbf{x}$ onto the plane $\mathfrak{P}$. So, $\hat{\mathbf{x}}$ is actually the projection of $\mathbf{x}$ onto its nearest neighbor plane $\mathfrak{P}$. The reconstruction error $\varepsilon_{\min}$ is actually the distance from $\mathbf{x}$ to its nearest neighbor plane $\mathfrak{P}$, that is, the length of the line segment $\overline{\mathbf{x}\hat{\mathbf{x}}}$.

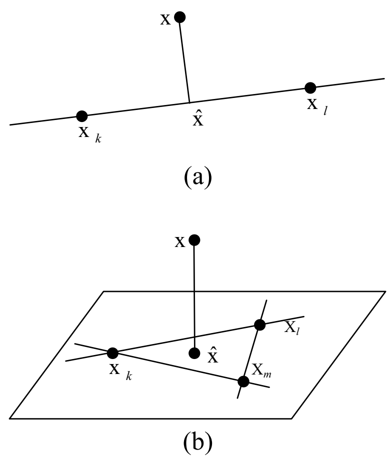

Figure 1: The geometric meaning of the minimal local reconstruction error (MLRE) when the nearest neighbor parameter K=2 and K=3. (a) the geometric meaning of MLRE when K = 2, (b) the geometric meaning of MLRE when K = 3.

When K > 3, the K-nearest neighbors of data point $\mathbf{x}$ form a subset of K-dimensional Euclidean space, that is,

$$\mathfrak{F} = \{\mathbf{z} = \sum_{j\in\Omega} w_j \mathbf{x}_j \mid \sum_{j\in\Omega} w_j = 1\}, \quad (4)$$

which is called the K-nearest neighbor "space"[1] of $\mathbf{x}$. Minimizing the reconstruction error in Eq. (1) is to find a point $\hat{\mathbf{x}}$ in the subset $\mathfrak{F}$ that is closest to $\mathbf{x}$. This point must be the projection of $\mathbf{x}$ onto the "space" $\mathfrak{F}$. So, $\hat{\mathbf{x}}$ is actually the projection of $\mathbf{x}$ onto its K-nearest neighbor "space". The reconstruction error $\varepsilon_{\min}$ is actually the distance from $\mathbf{x}$ to its K-nearest neighbor "space".

2.3. MLRE Measure based Classifier

Suppose there are c known pattern classes. Let $\mathfrak{X}_i = \{\mathbf{x}_{ij}\}$ be the training sample set of Class i, which contain M_i points. For a given new sample $\mathbf{x}$, let us first find its K-nearest neighbors in each class. Suppose its K-nearest neighbors in Class i are $\mathbf{x}_{ij}$ ($j = 1,\cdots,\mathrm{K}$). The MLRE distance from $\mathbf{x}$ to Class i is defined by

$$\varepsilon_i(\mathbf{x}) = \min_{\sum_{j=1}^{K} w_{ij}=1} \left\| \mathbf{x} - \sum_{j=1}^{K} w_{ij}\mathbf{x}_{ij} \right\|^2. \quad (5)$$

Assume that the distance between $\mathbf{x}$ and Class l is minimal, that is,

$$\varepsilon_l(\mathbf{x}) = \min_i \varepsilon_i(\mathbf{x}). \quad (6)$$

The decision rule of the MLRE-based classifier is that $\mathbf{x}$ belongs to Class l.

For robustness, before classification, we generally need to normalize feature vectors, making the length of each feature vector to be 1, that is, $\mathbf{x} \leftarrow \mathbf{x}/\|\mathbf{x}\|$. Specially, when K = 1, the MLRE-based classifier is equivalent to the nearest neighbor classifier using the normalized Euclidean distance. Since the normalized Euclidean distance is equivalent to the cosine distance, the MLRE-based classifier is actually the nearest neighbor classifier using the cosine distance when K = 1. When K = 2, the MLRE-based classifier is actually the nearest neighbor line classifier [4], and when K = 3, the MLRE-based classifier is the nearest neighbor plane classifier [4]. Thus, the MLRE-based classifier can be viewed as a generalization of K-nearest neighbor line and plane classifiers.

3. Minimal Local Reconstruction Error Measure Based Discriminant Feature Extraction

If the dimension of the input space is very high, it is time-consuming to perform classification directly in the input space. In addition, performing classification in the high-dimensional space might encounter the so-called "dimensionality curse". Therefore, to avoid these problems, we would rather perform feature extraction first before the classification step. In this section, we will develop a MLRE measure based discriminant feature extraction method.

3.1. Basic Idea

Suppose there are c known pattern classes. Let $\mathfrak{X} = \{\mathbf{x}_{ij}\}$ be the training sample set, where $i = 1,\cdots,c$ and $j = 1,\cdots,M_i$. For each sample $\mathbf{x}_{ij}$, we can find its L-nearest neighbors in every class. Let ε_{ij}^s be the minimal local reconstruction error of $\mathbf{x}_{ij}$ from its K-nearest neighbors in Class s. The reconstruction weights can be calculated by solving the following optimization problem:

$$\varepsilon_{ij}^s = \min \left\| \mathbf{x}_{ij} - \sum_t w_{st}^{ij}\mathbf{x}_{st} \right\|^2, \quad (7)$$

subject to $\sum_t w_{st}^{ij} = 1$ and $w_{st}^{ij} = 0$ if $\mathbf{x}_{\mathrm{st}}$ does not belong to the set of K-nearest neighbors of $\mathbf{x}_{ij}$ in Class s. Based on the obtained optimal reconstruction weights w_{st}^{ij}, we can define the within-class local scatter of samples in the input space as follows

[1] Note that $\mathfrak{F}$ is not a vector space in a strict mathematic sense because is not closed under vector addition and scalar multiplication.

$$\sum_{\substack{i,j\\s=i}} \varepsilon_{ij}^{s} = \sum_{\substack{i,j\\s=i}} \left\| \mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{st} \right\|^2 , \tag{8}$$

and the between-class local scatter of samples in the input space as follows

$$\sum_{i,j} \sum_{s\neq i} \varepsilon_{ij}^{s} = \sum_{i,j} \sum_{s\neq i} \left\| \mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{st} \right\|^2 . \tag{9}$$

The within-class local scatter characterizes the compactness of the nearby samples of the same class, and the between-class local scatter characterizes the separability of the nearby samples belonging to different classes. It is obvious that larger between-class local scatter and smaller within-class local scatter will lead to better classification results if samples are classified in the input space.

Our goal is to find a low-dimensional linear embedding of the data by virtue of the linear transformation

$$\mathbf{y} = \mathbf{P}^T \mathbf{x}, \text{ where } \mathbf{P} = (\varphi_1, \cdots, \varphi_d) \tag{10}$$

such that the data points in the embedding space have the following properties:

(i) The local reconstruction weights are preserved;

(ii) The between-class local scatter of samples is maximized while at the same time the within-class local scatter of samples is minimized.

The first property is to guarantee that the nearby points in the high-dimensional space remain nearby and similarly co-located with respect to one another in the low-dimensional space. The second property aims to make the nearby samples of the same class become as compact as possible and simultaneously the nearby samples belonging to different classes become as far as possible. The second property is therefore closely related to classification.

3.2. MLRE-based Feature Extractor

Under the linear transformation, each data point $\mathbf{x}_{ij}$ in observation space is mapped into $\mathbf{y}_{ij} = \mathbf{P}^T \mathbf{x}_{ij}$ in d-dimensional embedding space. The within-class local scatter of samples in the embedding space is

$$\begin{aligned}
\sum_{\substack{i,j\\s=i}} \left\| \mathbf{y}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{y}_{st} \right\|^2 &= \sum_{\substack{i,j\\s=i}} (\mathbf{y}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{y}_{st})^{\mathrm{T}} (\mathbf{y}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{y}_{st}) \\
&= \sum_{\substack{i,j\\s=i}} (\mathbf{P}^T \mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{P}^T \mathbf{x}_{ij})^{\mathrm{T}} (\mathbf{P}^T \mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{P}^T \mathbf{x}_{ij}) \\
&= \sum_{\substack{i,j\\s=i}} [\mathbf{P}^T (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})]^{\mathrm{T}} [\mathbf{P}^T (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})] \\
&= \mathrm{tr}\{\mathbf{P}^T [\sum_{\substack{i,j\\s=i}} (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})(\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})^{\mathrm{T}}]\mathbf{P}\} \\
&= \mathrm{tr}\{\mathbf{P}^T \mathbf{S}_{\mathrm{w}}^{\mathrm{L}} \mathbf{P}\} ,
\end{aligned} \tag{11}$$

where $\mathrm{tr}(\cdot)$ is the notation of trace operator, and

$$\mathbf{S}_{\mathrm{w}}^{\mathrm{L}} = \sum_{\substack{i,j\\s=i}} (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})(\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})^{\mathrm{T}} \tag{12}$$

is called the within-class local scatter matrix. It is easy to show that $\mathbf{S}_{\mathrm{w}}^{\mathrm{L}}$ a nonnegative definite matrix.

The between-class local scatter of samples in the embedding space is

$$\begin{aligned}
&\sum_{i,j} \sum_{s\neq i} \left\| \mathbf{y}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{y}_{st} \right\|^2 \\
&= \sum_{i,j} \sum_{s\neq i} (\mathbf{y}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{y}_{st})^{\mathrm{T}} (\mathbf{y}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{y}_{st}) \\
&= \sum_{i,j} \sum_{s\neq i} (\mathbf{P}^T \mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{P}^T \mathbf{x}_{ij})^{\mathrm{T}} (\mathbf{P}^T \mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{P}^T \mathbf{x}_{ij}) \\
&= \sum_{i,j} \sum_{s\neq i} [\mathbf{P}^T (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})]^{\mathrm{T}} [\mathbf{P}^T (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij}) \\
&= \mathrm{tr}\{\mathbf{P}^T [\sum_{i,j} \sum_{s\neq i} (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})(\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})^{\mathrm{T}}]\mathbf{P}\} \\
&= \mathrm{tr}\{\mathbf{P}^T \mathbf{S}_{\mathrm{b}}^{\mathrm{L}} \mathbf{P}\} ,
\end{aligned} \tag{13}$$

where

$$\mathbf{S}_{\mathrm{b}}^{\mathrm{L}} = \sum_{i,j} \sum_{s\neq i} (\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})(\mathbf{x}_{ij} - \sum_{t} w_{st}^{ij} \mathbf{x}_{ij})^{\mathrm{T}} \tag{14}$$

is called the between-class local scatter matrix. It is easy to show that $\mathbf{S}_{\mathrm{b}}^{\mathrm{L}}$ is a nonnegative definite matrix.

To maximize the between-class scatter and simultaneously minimize the within-class scatter, we can choose to maximize the following criterion:

$$J(\mathrm{P}) = \frac{\mathrm{tr}\{\mathbf{P}^T \mathbf{S}_{\mathrm{b}}^{\mathrm{L}} \mathbf{P}\}}{\mathrm{tr}\{\mathbf{P}^T \mathbf{S}_{\mathrm{w}}^{\mathrm{L}} \mathbf{P}\}} \tag{15}$$

Specially, when P is one-dimensional vector, i.e., $\mathbf{P} = \varphi$, the criterion becomes

$$J(\varphi) = \frac{\varphi^T \mathbf{S}_{\mathrm{b}}^{\mathrm{L}} \varphi}{\varphi^T \mathbf{S}_{\mathrm{w}}^{\mathrm{L}} \varphi} \tag{16}$$

The optimal solution of the criterion in Eq. (16) is actually the generalized eigenvectors of $\mathrm{S}_{\mathrm{b}}^{\mathrm{L}} \mathbf{X} = \lambda \mathrm{S}_{\mathrm{w}}^{\mathrm{L}} \mathbf{X}$ corresponding to the largest eigenvalue. Calculate the generalized eigenvectors $\varphi_1, \cdots, \varphi_d$ of $\mathrm{S}_{\mathrm{b}}^{\mathrm{L}} \mathbf{X} = \lambda \mathrm{S}_{\mathrm{w}}^{\mathrm{L}} \mathbf{X}$ corresponding to the d largest eigenvalues and let $\mathbf{P} = (\varphi_1, \cdots, \varphi_d)$. It is easy to show that $\mathbf{P}$ is the matrix maximizing the criterion in Eq. (15). The corresponding linear transformation $\mathbf{y} = \mathbf{P}^T \mathbf{x}$ is called MLRE-based feature extractor.

In the small sample size cases, the within-class local scatter matrix $\mathbf{S}_{\mathrm{w}}^{\mathrm{L}}$ is singular because the training sample size is smaller than the dimension of the image vector space. To address this issue, we first use PCA to reduce the dimension of the input space such that $\mathbf{S}_{\mathrm{w}}^{\mathrm{L}}$ is nonsingular in the PCA-transformed space. We then use MLRE-based feature extractor for the second dimensionality reduction.

4. Experiments

In this section, the performance of the MLRE-based feature extractor and classifier is evaluated on the CENPARMI handwritten numeral database and the FERET face image database and compared with the performances of PCA and LDA. In the following experiments, the proposed MLRE-based feature extraction method and classification method are combined and used in a unified framework: the MLRE-based feature extractor is first performed and the MLRE-based classifier is then employed for classification. Despite this, the neighbor parameters can be chosen differently in the feature extraction and classification phases. For example, we can choose L neighbors of a sample point in the feature extractor and K neighbors in the classifier.

4.1. Experiment using the CENPARMI handwritten numeral database

The experiment was done on Concordia University CENPARMI handwritten numeral database. The database contains 6000 samples of 10 numeral classes (each class has 600 samples). In our experiment, we choose the first 100 samples of each class for training, the remaining 500 samples for testing.

The PCA, LDA, and the proposed MLRE-based feature extractor are, respectively, used for feature extraction based on the original 121-dimensional Legendre moment features. For PCA and LDA, we utilize the Euclidean distance based nearest neighbor classifier, since the PCA and LDA models are both developed in the Euclidean space. For the MLRE-based feature extractor, we utilize the MLRE-based classifier for measure consistency. The nearest neighbor parameters are chosen as L=10 in the feature extractor and K=6 in the classifier.

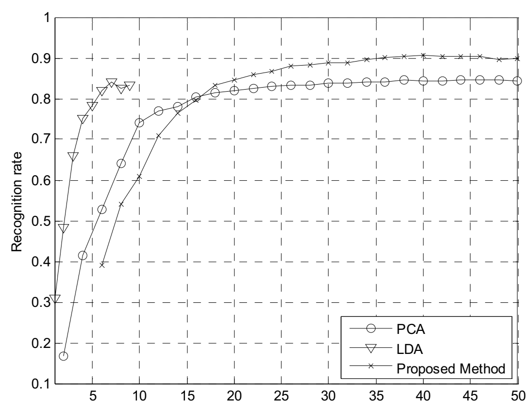

Figure 2: The recognition rates of PCA, LDA and the proposed method versus the dimensions on the CENPARMI handwritten numeral database

Table 1: The maximal recognition rates (%) of PCA, LDA and the proposed method on the CENPARMI handwritten numeral database and the corresponding dimensions

Method	PCA	LDA	Proposed method
Recognition rate	84.6	84.0	90.6
Dimension	44	7	40

The classification results are shown in Figure 2 and Table 1. When the dimension is relatively small, LDA features are more powerful than others. However, LDA can only extract c-1=9 features. It is obvious that this small amount of features is not enough to represent digital pattern for recognition purposes. Like PCA, the proposed MLRE-based feature extractor can yield more features for pattern representation. Our experimental results show the MLRE-based features and classifier are more effective than others when the dimension is over 20. The maximal recognition rate of the proposed method is 6% higher than that of PCA and LDA.

4.2. Experiment Using the FERET Database

The FERET face image database has become a standard database for testing and evaluating state-of-the-art face recognition algorithms [12, 13]. The proposed method was tested on a subset of the FERET database. This subset includes 1000 images of 200 individuals (each one has 5 images). It is composed of the images whose names are marked with two-character strings: "*ba*", "*bj*", "*bk*", "*be*", "*bf*". This subset involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original image was automatically cropped based on the location of eyes and mouth, and the cropped image was resized to 80×80 pixels and further pre-processed by histogram equalization.

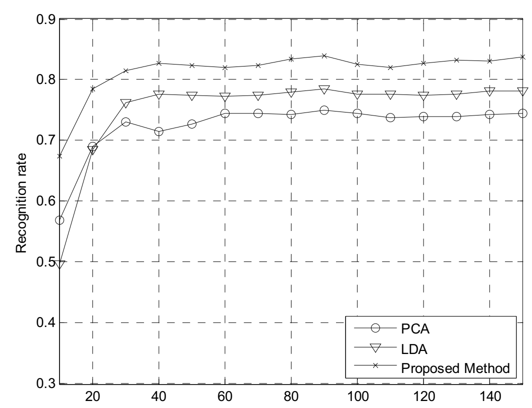

Figure 3: The recognition rates of PCA, LDA and the proposed method versus the dimensions on the FERET face image database

Table 2: The maximal recognition rates (%) of PCA, LDA and the proposed method on the FERET face image database and the corresponding dimensions

Method	PCA	LDA	Proposed method
Recognition rate	75.0	78.5	84.0
Dimension	90	90	90

In our test, we use the first three images (i.e., "*ba*", "*bj*" and "*bk*") per class for training, and the remaining two images (i.e., "*be*" and "*bf*") for testing. PCA, LDA, and the proposed MLRE-based feature extractor are, respectively, used for feature extraction based on the original 6400-dimensional image vectors. In the PCA phase of LDA and the MLRE-based feature extractor (Note that LDA also needs a PCA process in its implementation [14]), the number of principal components is set as 150. In the MLRE-based feature extractor and classifier, we choose the nearest neighbor parameters L=1 and K=1. In the PCA and LDA methods, we use the nearest neighbor classifier with cosine distance.

The maximal recognition rate of each method and the corresponding dimension are given in Table 2. The recognition rate curve versus the variation of dimensions is shown in Figure 3. Figure 3 indicates that the proposed method consistently performs better than PCA and LDA, irrespective the variation of dimensions. Table 2 indicates that the maximal recognition rate of the proposed method is 5.5% higher than that of PCA and LDA.

5. Conclusions

In this paper, the minimal local reconstruction error (MLRE) is introduced as a similarity measure and a MLRE-based classifier is presented. From the geometric meaning of the minimal local reconstruction error, we know that the MLRE-based classifier is a generalization of the conventional nearest neighbor classifier and the nearest neighbor line and plane classifiers. We further apply the MLRE measure to characterize the within-class and between-class local scatters and then develop a MLRE measure based discriminant feature extraction method. The proposed MLRE-based feature extraction method is in line with the MLRE-based classification method in spirit so that they can be seamlessly combined into a pattern recognition system. The proposed feature extraction and classification method is evaluated using the CENPARMI handwritten numeral database and the FERET face image database. The experimental results indicate that the proposed method is more effective than the PCA and LDA methods.

References

[1] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[2] S. Z. Li and Juwei Lu, "Face recognition using the nearest feature line method", *IEEE Transactions on Neural Networks*, 10(2):439–443, 1999.

[3] J.-T. Chien and C.-C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition", *IEEE Trans. Pattern Anal. Machine Intell.*, 24(12):1644–1649, 2002.

[4] W. Zheng, L. Zhao, and C. Zou, "Locally nearest neighbour classifiers for pattern classification", *Pattern Recognition*, 37(6):1307–1309, 2004.

[5] Z. Lou and Z. Jin, "Novel adaptive nearest neighbor classifiers based on hit-distance", In Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006), Vol. 3, pages 87-90. Aug. 20-24, 2006.

[6] T.-K. Kim, J. Kittler and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations", *IEEE Trans. Pattern Anal. Machine Intell.,* 2007, 29(6),1005-1018.

[7] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, "Face Recognition Using Laplacianfaces", *IEEE Trans. Pattern Anal. Machine Intell.,* 2005, 27(3), pp. 328- 340.

[8] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood Preserving Embedding", *IEEE International Conference on Computer Vision* (ICCV), Beijing, China, 2005.

[9] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang and S. Lin. "Graph Embedding and Extension: A General Framework for Dimensionality Reduction", *IEEE Trans. Pattern Anal. Machine Intell.,* VOL. 29, NO. 1, 40-51, 2007.

[10] J. Yang, D. Zhang, J.-y. Yang, B. Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics", *IEEE Trans. Pattern Anal. Machine Intell.,* 2007,29(4),650-664.

[11] L. K. Saul and S. T. Roweis, "Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds", *Journal of Machine Learning Research*, vol. 4, pp. 119-155, 2003.

[12] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE Trans. Pattern Anal. Machine Intell.*, 2000, 22 (10), pp.1090-1104.

[13] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, 1998, 16 (5), pp. 295-306.

[14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriengman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Trans. Pattern Anal. Machine Intell.* 1997, 19 (7), pp. 711-720.