Pattern Recognition 43 (2010) 3298–3306

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Feature extraction by learning Lorentzian metric tensor and its extensions ☆

Risheng Liu [a], Zhouchen Lin [b], Zhixun Su [a,*], Kewei Tang [a]

[a] Dalian University of Technology, Ganjingzi District, Dalian 116024, PR China
[b] Microsoft Research, Asia, Zhichun Road #49, Haidian District, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 20 September 2009
Received in revised form
8 April 2010
Accepted 4 May 2010

Keywords:
Feature extraction
Dimensionality reduction
Lorentzian geometry
Metric learning
Discriminant analysis

ABSTRACT

We develop a supervised dimensionality reduction method, called Lorentzian discriminant projection (LDP), for feature extraction and classification. Our method represents the structures of sample data by a manifold, which is furnished with a Lorentzian metric tensor. Different from classic discriminant analysis techniques, LDP uses distances from points to their within-class neighbors and global geometric centroid to model a new manifold to detect the intrinsic local and global geometric structures of data set. In this way, both the geometry of a group of classes and global data structures can be learnt from the Lorentzian metric tensor. Thus discriminant analysis in the original sample space reduces to metric learning on a Lorentzian manifold. We also establish the kernel, tensor and regularization extensions of LDP in this paper. The experimental results on benchmark databases demonstrate the effectiveness of our proposed method and the corresponding extensions.

1. Introduction

Feature extraction has been studied by researchers in machine learning, pattern recognition and computer vision for long time. There are many approaches for this task. One of the most successful and well-studied techniques is dimensionality reduction. We devote this paper to addressing the supervised dimensionality reduction from the perspective of Lorentzian geometry which is extensively used in general relativity, as a basic geometric tool for modeling the space–time in physics.

1.1. Related work

Principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [1] are two most popular linear dimensionality reduction techniques. PCA projects the data points along the directions of maximal variances and aims to preserve the Euclidean distances between samples. Unlike PCA which is unsupervised, LDA is supervised. It searches for the projection axes on which the points of different classes are far from each other and at the same time the data points of the same class are close to each other. However, these linear models may fail to discover nonlinear data structures.

During the recent years, a number of nonlinear dimensionality reduction algorithms called manifold learning have been developed to address this issue [17,7,14,8,10,13]. However, these nonlinear techniques might not be suitable for real world applications because they yield maps that are defined only on the training data points. To compute the maps for the new testing points requires extra effort.

Along this direction, there is considerable interest in using linear methods, inspired by the geometric intuition of manifold learning, to find the nonlinear structure of data set. Some popular ones include locality preserving projection (LPP) [19,12], neighborhood preserving embedding (NPE) [18], marginal Fisher analysis (MFA) [11], maximum margin criterion (MMC) [20], average neighborhood margin maximization (ANMM) [21], semi-Riemannian discriminant analysis (SRDA) [5] and unsupervised discriminant projection (UDP) [24].

In addition, the kernel trick [3] has been widely applied to extend linear dimensionality reduction algorithms to nonlinear ones by mapping the data to a high-dimensional (usually infinite-dimensional) feature space. It is worth noting that most of the existing dimensionality reduction methods are vector based, but in many real world tasks, the data are more naturally represented as higher-order tensors. For example, a captured image is an order-2 tensor, i.e. matrix, and the LBP or Gabor feature of an image is in the form of order-3 tensor [26]. Thus a number of algorithms [28,29,31] have been proposed to handle the data as tensors directly. Cai et al. [43] also proposed a regularized subspace learning framework which explicitly considers the spatial relationship between the pixels in images.

1.2. Our approach

Recently, Yang et al. [24] adapted both local and global scatters to unsupervised dimensionality reduction. They maximized the

☆ A preliminary version of this paper appeared on ACCV'09 [25].
* Corresponding author. Tel.: +86 411 84708351x8020; fax: +86 411 84708354.
E-mail addresses: rsliu0705@gmail.com (R. Liu), zhoulin@microsoft.com (Z. Lin), zxsu@dlut.edu.cn (Z. Su), tkwliaoning@gmail.com (K. Tang).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.05.009

ratio of the global scatters to the local scatters. Zhao et al. [5] first applied the semi-Riemannian geometry to classification [5]. Inspired by prior work, in this paper, we propose a novel method, called Lorentzian discriminant projection (LDP), which focuses on supervised dimensionality reduction. Its goal is to discover both local class discriminant and global geometric structures of the data set from the perspective of Lorentzian geometry. We first construct a manifold to model the local class and the global data structures. In this way, both the local discriminant and the global geometric structures of the data set can be accurately characterized by learning a special Lorentzian metric tensor on the newly built manifold. In fact, the role of Lorentzian metric tensor in LDP is to transfer the geometry from the sample space to the feature space.

To our knowledge, this is the first time to introduce Lorentzian geometry to feature extraction. Compared with traditional algorithms, our method has the following advantages:

(1) The solution to many popular dimensionality reduction algorithms, such as LPP, NPE, LDA, MFA and UDP is to pose a trace ratio optimization problem, which however does not have a closed-form solution [6]. While LDP avoids this problem since it only needs to compute a simple eigen-decomposition problem.
(2) In general, the amount and the prior distribution of the training data, and the type of problem all influence the classification performance. Our LDP proposes a Lorentzian metric learning framework to deform feature space towards the optimization of both local within-class compactness and global structure diversity. For different data set, we can learn their specific discriminant structure from the original sample space and apply it to the feature space. Therefore, our "learning strategy" is more natural than the traditional "design-based strategy", i.e. design a weighted graph directly [11,19]. The experimental results also indicate that LDP is more effective than traditional methods in extracting discriminant features.

The rest of this paper is organized as follows. In Section 2, we introduce the algorithm details of Lorentzian discriminant projection (LDP). Section 3 builds the kernel, tensor and regularization extension of LDP, respectively. The experimental results of LDP applied to real-world face analysis and handwriting digits classification are presented in Section 4. Finally, we conclude the paper along with some directions for further research in Section 5.

2. Lorentzian discriminant projection

2.1. Fundamentals of Lorentzian manifold

Lorentzian geometry is an active field of mathematical research that can be seen as part of differential geometry as well as mathematical physics. It represents the mathematical foundation of the general relativity which is probably one of the most successful and beautiful theories of physics.

In differential geometry, a semi-Riemannian manifold is a generalization of a Riemannian manifold. It is furnished with a non-degenerate and symmetric metric tensor called the semi-Riemannian metric tensor. The metric matrix on the semi-Riemannian manifold is diagonalizable and the diagonal entries are non-zero. We use the metric signature to denote the number of positive and negative ones. Given a semi-Riemannian manifold $\mathbb{M}$ of dimension n, if the metric has p positive and q negative diagonal entries, then the metric signature is (p,q), where $p+q=n$.

Lorentzian manifold is the most important subclass of semi-Riemannian manifold in which the metric signature is $(n-1,1)$. The metric matrix on the Lorentzian manifold $\mathbb{L}_1^n$ is of form

$$\mathbf{G}=\begin{bmatrix} \hat{\Lambda}_{(n-1)\times(n-1)} & 0 \\ 0 & -\check{\lambda} \end{bmatrix}, \tag{1}$$

where $\hat{\Lambda}_{(n-1)\times(n-1)}$ is diagonal and its diagonal entries and $\check{\lambda}$ are positive. Suppose that $\mathbf{r}=[\hat{\mathbf{r}}^T,\check{r}]^T$ is an n-dimensional vector, then a metric tensor $g(\mathbf{r},\mathbf{r})$ with respect to $\mathbf{G}$ is expressible as

$$g(\mathbf{r},\mathbf{r})=\mathbf{r}^T\mathbf{G}\mathbf{r}=\hat{\mathbf{r}}^T\hat{\Lambda}\hat{\mathbf{r}}-\check{\lambda}(\check{r})^2. \tag{2}$$

Because of the non-degeneracy of the Lorentzian metric, vectors can be classified into space-like ($g(\mathbf{r},\mathbf{r})>0$ or $\mathbf{r}=0$), time-like ($g(\mathbf{r},\mathbf{r})<0$) or null ($g(\mathbf{r},\mathbf{r})=0$ and $\mathbf{r}\neq 0$). Fig. 1 shows the three-dimensional Lorentzian space–time with the signature (2,1). One may refer to [4] for more details.

2.2. The motivation of LDP

The theory and algorithm in this paper are based on the perspective that the discrimination power is tightly related to both local class and global data structures. Our LDP is inspired by two factors: the viewpoint of Lorentzian manifold applied to general relativity and the success of considering both local and global structures for dimensionality reduction.

The Lorentzian geometry has been successfully applied to Einstein's general relativity to model the space–time as a four-dimensional Lorentzian manifold of signature (3,1). And as will be shown later, this manifold is also convenient to model the structures of a group of classes. On one hand, we model the local class structure by the distances between each sample and its within-class neighbors. We also characterize the global data structure by the distances between each point and the global geometric centroid. Combining both local and global distances together, we naturally form a new manifold to preserve the discriminant structure for data set. On the other hand, to optimize both local and global structures at the same time, we need to perform discrepancies of within-class quantities and global quantities. To do so, we introduce Lorentzian metrics which are the unique tools to integrate such kinds of dual quantities from mathematical point of view. Therefore, the discriminant structure

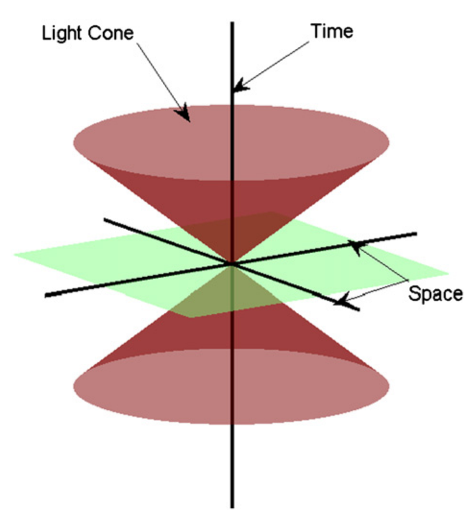

Fig. 1. An illustration of a three-dimensional Lorentzian space–time with the signature (2,1). Inside the light cone is the time-like space–time and outside the space-like space–time.

of the data set is initially modeled as a Lorentzian manifold where coordinates are characterized by the distances between sample pairs (each point with its within-class neighbors and the global geometric centroid). Furthermore, we use the positive part $\hat{\Lambda}$ to handle the local class structure and the negative part $-\check{\lambda}$ to model the global data structure.

To this end, learning a discriminant subspace reduces to learning the geometry of a Lorentzian manifold. Thus, supervised dimensionality reduction is coupled with Lorentzian metric learning. Moreover, we present an approach to optimize both the local discriminant and global geometric structures by learning the Lorentzian metric in the original sample space and applying it to the discriminant subspace.

2.3. Modeling feature space as a Lorentzian manifold

For supervised dimensionality reduction task, the samples can be represented as a point set $\mathcal{S}_x = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$, $\mathbf{x}_i \in \mathbb{R}^n$. The class label of $\mathbf{x}_i$ is denoted by C_i and m_i is the number of points which share the same label with $\mathbf{x}_i$. As we have previously described, the goal of the proposed algorithm is to transform points from the original high-dimensional sample space to a low-dimensional discriminant subspace, i.e. $\mathcal{S}_y \subset \mathbb{R}^d$ where $d \ll n$. In this subspace, feature points belonging to the same class should have higher within-class similarity and more consistent global geometric structure. To achieve this goal, we introduce a Lorentzian manifold to model the structure of features in a low dimensional discriminant subspace.

With $\mathbf{y}_i$, $\mathcal{S}_{y_i} = \{\mathbf{y}_i, \mathbf{y}_1^i, \ldots, \mathbf{y}_{m_i-1}^i\}$ (points share the same class label with $\mathbf{y}_i$) and $\overline{\mathbf{y}}$ (the geometric centroid of $\mathcal{S}_y$, i.e., $\overline{\mathbf{y}} = (1/m)\sum_{i=1}^m \mathbf{y}_i$), a new point $\mathbf{d}_{y_i}$ is defined as

$$\mathbf{d}_{y_i} = [d(\mathbf{y}_i, \mathbf{y}_1^i), \ldots, d(\mathbf{y}_i, \mathbf{y}_{m_i-1}^i), d(\mathbf{y}_i, \overline{\mathbf{y}})]^T \equiv [\hat{\mathbf{d}}_{y_i}^T, d(\mathbf{y}_i, \overline{\mathbf{y}})]^T, \tag{3}$$

where $\mathbf{y}_j^i \in \mathcal{S}_{y_i}$ and $d(\mathbf{y}_p, \mathbf{y}_q)$ is the distance between $\mathbf{y}_p$ and $\mathbf{y}_q$. It is easy to see that this coordinate representation can contain both local within-class similarity and global geometric structure. We consider these m_i-tuple coordinate representations as points sampled from a new manifold $\mathbb{L}_1^{m_i}$ furnished with a Lorentzian metric tensor g_l. It is straightforward to see that $g_l(\mathbf{d}_{y_i}, \mathbf{d}_{y_i})$ can be written as

$$g_l(\mathbf{d}_{y_i}, \mathbf{d}_{y_i}) = \mathbf{d}_{y_i}^T \mathbf{G}_i^l \mathbf{d}_{y_i} = \mathrm{tr}((\mathbf{Y}_i \mathbf{D}_i) \mathbf{G}_i^l (\mathbf{Y}_i \mathbf{D}_i)^T), \tag{4}$$

where the metric matrix $\mathbf{G}_i^l$ is real diagonal and the signature of the metric is $(m_i^{-1}, 1)$, $\mathbf{D}_i = [\mathbf{e}_{m_i}, -\mathbf{I}_{m_i \times m_i}]^T$ ($\mathbf{I}_{m_i \times m_i}$ is an identity matrix of size $m_i \times m_i$ and $\mathbf{e}_{m_i}$ is an all-one column vector of length m_i) and $\mathbf{Y}_i = [\mathbf{y}_i, \mathbf{y}_1^i, \ldots \mathbf{y}_{m_i-1}^i, \overline{\mathbf{y}}]$.

Then the total Lorentzian metric tensor can be given as

$$\sum_{i=1}^{m} g_l(\mathbf{d}_{\mathbf{y}_i}, \mathbf{d}_{y_i}) = \mathrm{tr}(\mathbf{Y}\mathbf{L}\mathbf{Y}^T), \tag{5}$$

where $\mathbf{L} = \sum_{i=1}^m \mathbf{B}_i \mathbf{D}_i \mathbf{G}_i^l \mathbf{D}_i^T \mathbf{B}_i^T$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m, \overline{\mathbf{y}}]$ and $\mathbf{B}_i$ is a binary selection matrix of size $(m+1) \times (m_i+1)$ which satisfies $\mathbf{Y}_i = \mathbf{Y}\mathbf{B}_i$.[1]

If there is a linear isometric transformation between the low dimensional feature $\mathbf{y}$ and the original sample $\mathbf{x}$, i.e., $\mathbf{y} \to \mathbf{U}\mathbf{y} = \mathbf{x}$, we can have an optimization model:

$$\begin{cases} \arg\min_{\mathbf{U}} \mathrm{tr}(\mathbf{U}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{U}), \\ \text{s.t. } \mathbf{U}^T\mathbf{U} = \mathbf{I}_{d\times d}, \end{cases} \tag{6}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \overline{\mathbf{x}}]$ and $\overline{\mathbf{x}}$ is the geometric centroid of $\mathcal{S}_x$. The linear transformation $\mathbf{U}$ that minimizes the objective function in (6) can be found as being composed of the eigenvectors associated with the d smallest eigenvalues of the following problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{u} = \lambda \mathbf{u}. \tag{7}$$

It is sufficient to note that the Lorentzian metric tensor forms the geometry of the feature structure. Thus a question naturally arises: How to learn a special Lorentzian metric tensor to furnish the newly built manifold? This is discussed in the next subsection.

2.4. Learning the Lorentzian metric tensors

The Lorentzian metric matrices $\mathbf{G}_i^l$ are key to the proposed dimensionality reduction algorithm. The role of $\mathbf{G}_i^l$ in our model is similar to that of weights in graph based models [11]. In these algorithms, one should design a weighted graph based on some similarity criteria, such as Gaussian similarity from Euclidean distance as in [19] and prior class information in supervised learning algorithms as in [1]. The performance of these algorithms strongly depends on such human designed graph weight matrix. In contrast, our LDP proposes a novel method to learn Lorentzian metric matrices from the sample set $\mathcal{S}_x$ and then apply it to the feature set $\mathcal{S}_y$. In this way, LDP can transfer both local compactness and global structure diversity from the sample space to the feature space for specific data set. The metric $\mathbf{G}_i^l$ consists of two parts: the positive-definite part $\hat{\Lambda}_i$ and the negative-definite part $-\check{\lambda}_i$. In this subsection, we introduce an efficient way to learn $\hat{\Lambda}_i$ and $\check{\lambda}_i$ successively.

The positive part $\hat{\Lambda}_i$ of the Lorentzian metric tensor is used to measure the local structure of $\mathcal{S}_{y_i}$ in low-dimensional discriminant subspace. We can characterize the within-class similarity and local geometry by learning $\hat{\Lambda}_i$ from $\mathcal{S}_{x_i}$ and then apply it to $\mathcal{S}_{y_i}$. $\hat{\Lambda}_i$ in the original sample space can be given as

$$g_l^p(\hat{\mathbf{d}}_{x_i}, \hat{\mathbf{d}}_{x_i}) = \hat{\mathbf{d}}_{x_i}^T \hat{\Lambda}_i \hat{\mathbf{d}}_{x_i} = \mathbf{g}_i^T \hat{\mathbf{D}}_{x_i} \mathbf{g}_i, \tag{8}$$

where

$$\mathbf{g}_i = \left[\sqrt{\hat{\Lambda}_i(1,1)}, \ldots, \sqrt{\hat{\Lambda}_i(m_i-1, m_i-1)}\right]^T$$

and

$$\hat{\mathbf{D}}_{x_i} = \mathrm{diag}(d(\mathbf{x}_i, \mathbf{x}_1^i)^2, \ldots, d(\mathbf{x}_i, \mathbf{x}_{m_i-1}^i)^2).$$

For the purpose of classification, we try to find $\mathbf{g}_i$ which will draw the within-class samples closer together. Therefore, for each $\mathcal{S}_{x_i}$, we may minimize this metric and obtain the following optimization problem:

$$\begin{cases} \arg\min_{\mathbf{g}_i} \mathbf{g}_i^T \hat{\mathbf{D}}_{x_i} \mathbf{g}_i, \\ \text{s.t. } \mathbf{e}_{m_i-1}^T \mathbf{g}_i = 1. \end{cases} \tag{9}$$

Imposing the sum-to-one constraint $\mathbf{e}_{m_i-1}^T \mathbf{g}_i = 1$ leads to the symmetries of the objective function, say, invariants to translations, rotations, and scalings [9]. It is easy to check that the solution to the above problem is

$$\mathbf{g}_i = \frac{(\hat{\mathbf{D}}_{x_i})^{-1} \mathbf{e}_{m_i-1}}{\mathbf{e}_{m_i-1}^T (\hat{\mathbf{D}}_{x_i})^{-1} \mathbf{e}_{m_i-1}}. \tag{10}$$

Thus the positive-definite part $\hat{\Lambda}_i$ can be obtained as

$$\hat{\Lambda}_i(p,q) = \begin{cases} (\mathbf{g}_i(p))^2 & \text{if } p = q, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

It is easy to check that LDP coincides with the PCA algorithm if $\hat{\Lambda}_i = \mathbf{0}$ and $\check{\lambda}_i = 1$, $i = 1, 2, \ldots, m$. From this point of view, the negative part $\check{\lambda}_i$ of the Lorentzian metric tensor is exactly a special weight used to measure the global geometric structure

[1] It means $(\mathbf{B}_i)_{pq} = 1$ if the q-th vector in $\mathbf{Y}_i$ is the p-th vector in $\mathbf{Y}$ [16,15].

Table 1
Algorithm of LDP.

Input: Sample point set $\mathcal{S}_x$ and the labels $\{C_1, C_2, \ldots, C_m\}$.
Output: Feature point set $\mathcal{S}_y$ and the projection matrix $\mathbf{U}$.
1. Compute the metric matrix $\mathbf{G}_i^l$ using Eqs. (10) and (11).
Form $\mathbf{L}$ using $\mathbf{L}=\sum_{i=1}^m \mathbf{B}_i\mathbf{D}_i\mathbf{G}_i^l\mathbf{D}_i^T\mathbf{B}_i^T$.
2. Obtain $\mathbf{U}$ by Eq. (7), and project samples: $\mathbf{y}=\mathbf{U}^T\mathbf{x}$.
3. Choose an optimal γ in [0, 1.5] with the adaptation $\check{\lambda}_i \leftarrow \gamma\check{\lambda}_i$.

(global scatter) of $\mathcal{S}_y$. As introduced in Section 2.1, a null (or light-like) vector $\mathbf{r}$ is the vector that vanishes the metric tensor: $g(\mathbf{r},\mathbf{r})=0$. Inspired by this physical property used in general relativity, we make $\mathbf{G}_i^l$ satisfy the following *simplified* local null property for discriminant analysis:

$$g(\mathbf{e}_{m_i},\mathbf{e}_{m_i}) = \sum_{j=1}^{m_i-1} \hat{\Lambda}_i(j,j) - \check{\lambda}_i = 0. \tag{12}$$

So the negative definite part of $\mathbf{G}_i^l$ can be determined by $\check{\lambda}_i = \sum_{j=1}^{m_i-1} \hat{\Lambda}_i(j,j)$. We empirically find that the discriminability will be enhanced if we choose a positive factor $\gamma \in [0,1.5]$ to multiply the negative part i.e., $\check{\lambda}_i \leftarrow \gamma\check{\lambda}_i$. This parameter actually plays the role of adjusting the trade-off between local compactness and global structure diversity. The value of γ can be determined by cross-validation.

To summarize, the main procedure of LDP is shown in Table 1.

2.5. Comparisons with the work in [5]

In [5], the authors proposed a geometric framework for classification. This work and LDP are both supervised feature extraction techniques. Their criteria, however, are quite different. The work in [5] only considers the local structures and uses the distances between each sample and its local interclass/intraclass neighbors to model the intrinsic structure of a group of classes. While, LDP constructs an $(m_i,1)$ Lorentzian manifold to model both local compactness and global structure diversity for feature extraction. Therefore, the newly built Lorentzian manifold has a more transparent link to discriminant analysis than general semi-Riemannian manifold.

Second, in [5], the authors propose an alternative way to learn the general semi-Riemannian metric matrix based on the "smoothing" criteria. However, this criterion is not directly linked to classification. In contrast, our LDP proposes a more natural criterion for metric learning: we try to find the positive part of the Lorentzian metric matrix which draws the within-class samples closer together while simultaneously determines the negative elements to preserve the global data structure.

3. Extensions

In this section, we introduce three useful extensions of Lorentzian discriminant projection, kernel LDP (KLDP), tensor LDP (TLDP) and smooth LDP(SLDP), which have their own advantages under different circumstances.

3.1. Kernel LDP

We describe a method to conduct LDP in the reproducing kernel Hilbert space into which the data points are mapped. This gives rise to Kernel LDP.

Suppose that we map $\mathcal{S}_x$ to some high (usually infinite) dimensional feature space $\mathcal{F}$ through a nonlinear mapping $\Phi : \mathbb{R}^n \rightarrow \mathcal{F}$, and apply linear LDP there.

Assume the kernel Gram matrix is $\mathbf{K}$ with $\mathbf{K}_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. Let the projection be $\mathbf{u} = \sum_{i=1}^m \alpha_i \Phi(\mathbf{x}_i) + \alpha_{m+1}\Phi(\overline{\mathbf{x}}) = \Phi(\mathbf{X})\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_m, \alpha_{m+1}]^T$. Then the optimal $\boldsymbol{\alpha}$ can be obtained by solving

$$\begin{cases} \arg\min\limits_{\boldsymbol{\alpha}} \boldsymbol{\alpha}^T\mathbf{KLK}\boldsymbol{\alpha}, \\ \text{s.t. } \boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha} = 1. \end{cases} \tag{13}$$

3.2. Tensor LDP

In order to match the tensor nature of data, we further extend vector-based LDP to tensor form.

An order-n tensor is an element of the space $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$. The scalar product of tensors A and B with the same dimensions is $\langle A,B \rangle = \sum_{i_1=1}^{n_1} \cdots \sum_{i_N=1}^{n_N} A(i_1,\ldots,i_N)\ B(i_1,\ldots,i_N)$. The Frobenius-norm of a tensor A is given by $\|A\|_F = \langle A,A \rangle$. The j-mode product of a tensor A and a matrix $\mathbf{V} \in \mathbb{R}^{n_j \times d_j}$ is an $n_1 \times n_2 \times \cdots \times n_{j-1} \times d_j \times n_{j+1} \times \cdots \times n_N$ tensor denoted as $A \times_j \mathbf{V}$. The j-mode unfolding of A is denoted by $\mathbf{A}^{(j)} \in \mathbb{R}^{n_j \times (n_{j+1} \ldots n_N n_1 \ldots n_{j-1})}$, where the element $A(i_1,\ldots,i_N)$ of the original tensor appears at the i_j-th row and the u_j-th column of $\mathbf{A}^{(j)}$, in which $u_j = (i_{j+1}-1)\ n_{j+2}n_{j+3} \ldots n_N n_1 n_2 \ldots n_{j-1} + (i_{j+2}-1) n_{j+3} \ldots n_N n_1 n_2 \ldots n_{j-1} + \cdots + (i_N-1) n_1 n_2 \ldots n_{j-1} + (i_1-1) n_2 n_3 \ldots n_{j-1} + (i_2-1) n_3 \ldots n_{j-1} + \cdots + i_{j-1}$.

Given $\mathcal{S}_X = \{X_1, X_2, \ldots, X_m\}$, $X_i \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, our objective is to find N optimal interrelated projection matrices $\mathbf{U}_j \in \mathbf{R}^{n_j \times d_j}$, such that the projected low-dimensional tensors can be represented as

$$Y_i = X_i \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_N \mathbf{U}_N, \quad i = 1,2,\ldots,m.$$

We adopt an iterative scheme to obtain the projections [23,22]. Given

$$\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_{j-1}, \mathbf{U}_{j+1}, \ldots, \mathbf{U}_N,$$

let

$$Y_i^{(j)} = X_i \times_1 \mathbf{U}_1 \times \cdots \times_{j-1} \mathbf{U}_{j-1} \times_{j+1} \mathbf{U}_{j+1} \times \cdots \times_N \mathbf{U}_N.$$

Then, by the corresponding j-mode unfolding, we can get $Y_i^{(j)} \Rightarrow \mathbf{Y}_i^{(j)}$. Therefore, the optimization model (6) can be rewritten as

$$\begin{cases} \arg\min\limits_{\mathbf{U}_j} \mathrm{tr}(\mathbf{U}_j^T \mathbf{L}^* \mathbf{U}_j), \\ \text{s.t. } \mathbf{U}_j^T \mathbf{U}_j = \mathbf{I}_{d_j \times d_j}, \end{cases} \tag{14}$$

where $\mathbf{L}^* = \sum_{i=1}^m (\mathbf{Y}^{(j)} \mathbf{B}_i^* \mathbf{D}_i^*)(\mathbf{G}_i^l)^* (\mathbf{Y}^{(j)} \mathbf{B}_i^* \mathbf{D}_i^*)^T$ and $\mathbf{Y}^{(j)} = [\mathbf{Y}_1^{(j)}, \mathbf{Y}_2^{(j)}, \ldots, \mathbf{Y}_m^{(j)}, \overline{\mathbf{Y}}^{(j)}]$. The matric matrix $(\mathbf{G}_i^l)^*$ can be obtained by replacing each element in $\mathbf{G}_i^l$ by a $\overline{d}_j \times \overline{d}_j$ matrix:

$$(\mathbf{G}_i^l)^* = \begin{pmatrix} \mathbf{G}_i^l(1,1)\mathbf{I}_{\overline{d}_j \times \overline{d}_j} & & \\ & \ddots & \\ & & \mathbf{G}_i^l(m_i,m_i)\mathbf{I}_{\overline{d}_j \times \overline{d}_j} \end{pmatrix},$$

where $\overline{d}_j = d_{j+1} d_{j+2} \ldots d_N d_1 \ldots d_{j-1}$. We can also obtain $\mathbf{B}_i^*$ and $\mathbf{D}_i^*$ in the same way, respectively.

3.3. Smooth regularized LDP

Learning the spatial relationship between the pixels in images is important for dimensionality reduction, especially in face recognition, clustering and image retrieval applications. In [43], a Laplacian penalized functional was introduced as a smooth regularization for dimensionality reduction. This prior information significantly improves the performance of traditional methods. By incorporating this Laplacian regularization, we propose another extended LDP (SLDP) for spatially smooth subspace learning:

$$\begin{cases} \arg\min_{\mathbf{U}} \mathrm{tr}(\mathbf{U}^T\mathbf{L}^s\mathbf{U}), \\ \text{s.t. } \mathbf{U}^T\mathbf{U} = \mathbf{I}_{d\times d}, \end{cases} \tag{15}$$

where $\mathbf{L}^s = ((1-\delta)\mathbf{XLX}^T + \delta\Delta^T\Delta)$ and $\Delta^T\Delta$ is the discretized Laplacian regularization [43] and $\delta \in [0,1]$ controls the smoothness of the estimator.

4. Experimental results

To evaluate our proposed LDP and its kernel, tensor and smooth extensions, four groups of experiments are conducted on different kinds of benchmark databases (CMU PIE, FRGC v2 [32] and MNIST[2]).

(1) *Linear* techniques: The performance of LDP is compared with PCA, LPP, LDA, MMC and MFA.
(2) *Kernel* techniques: The performance of KLDP is compared with KPCA, KLPP, KDA [35], KMMC and KMFA. We all adopt the Gaussian kernel, and the variance of the Gaussian kernel were set by cross-validation.
(3) *Tensor* techniques: The performance of TLDP is compared with TPCA [30], TLPP [36],[3] TLDA (DATER) [28], TMMC and TMFA.
(4) *Smooth regularization* techniques: The performance of SLDP is compared with SLPP, SMFA and SLDA.[4]

We use original one-dimensional (vector) and two-dimensional (matrix) image data and the expressive features yielded by LBP [33] and Gabor filter [34] for our experiments, respectively.

The generalized eigen-analysis based methods (e.g. LDA, LPP and MFA) encounter the computational trouble as they need to compute the matrix inverse. This *small sample size* problem [1] frequently occurs in computer vision and pattern recognition since samples have large dimensions whereas the number of classes is usually small. The PCA preprocessing is a classic and well-recognized method to solve this problem. For a fair comparison with other algorithms, we perform the PCA-based two step strategy in all experiments. Here, we choose the percentage of the energy retained in the PCA preprocessing step between 97% and 100% along with all possible dimensions.

4.1. Face analysis

In this subsection, we demonstrate the effectiveness of LDP (*linear*, *kernel*, *tensor* and *smooth regularized* forms) with real-world face analysis (representation and recognition). We show as follows the comprehensive performance comparisons between our proposed algorithms and the other state-of-the-art methods.

4.1.1. Face representation

In the face representation problem, we want to use LDP to learn an optimal discriminant subspace which is spanned by the columns of $\mathbf{U}$ in (6). The eigenvectors can be displayed as images, called the Lorentzianfaces in our approach. Using the facial images in experiment 4 of FRGC v2 as the training set, we present the Lorentzianfaces in Fig. 2, together with Eigenfaces [2] and Fisherfaces [1]. We can find that the tailing Lorentzianfaces contain most discriminant facial features (i.e. eyes, nose and mouth) which are insensitive to variations in both lighting direction and facial expression [37], while the leading Lorentzianfaces retain unwanted variations due to lighting and facial expression. It is also interesting to see that the tailing Lorentzianfaces (c.2 in Fig. 2) share similar patterns with the leading Fisherfaces (b.1 in Fig. 2) when $\gamma = 0.1$. But if we choose $\gamma = 1.5$, the tailing Lorentzianfaces (d.2 in Fig. 2) are somehow similar to the leading Eigenfaces (a.1 in Fig. 2). Thus the parameter γ in our LDP has its own advantages for different circumstances. Hence, LDP is capable of resolving a wide range of problems.

4.1.2. Face recognition experiments on CMU PIE

The CMU PIE database contains 68 persons with 41,368 face images as a whole. The face images were captured under varying pose, illumination and expression. We choose the five near frontal pose (C05, C07, C09, C27 and C29) and illumination indexed as 10 and 13 such that each person has 10 images. All the face images are manually aligned and cropped. The cropped images are 32 $\times$ 32 pixels, with 256 gray levels per pixel. We randomly select three images of each person for training and the remaining seven images are for testing. The top row of Fig. 3 shows facial images of one person.

The recognition rate curves of linear methods versus the variation of dimensions are illustrated in Fig. 4. The recognition rate of each method and the corresponding dimension are given in Table 2. As can be seen, the proposed LDP outperforms other algorithms involved in all four experiments.

LBP is a new approach which is proved effective for feature extraction. In our experiments, we subdivide each image by 4 $\times$ 4 grids and perform the $\mathrm{LBP}_8^{u\,2}$ on 16 evenly partitioned sub-blocks. Thus the LBP feature of one image is a 59 $\times$(4 $\times$ 4) order-3 tensor. We compare *tensor* methods on LBP features of CMU PIE to test the discriminative power of different methods on order-3 tensor data. Table 3 shows the recognition results. One can see that our proposed TLDP is the best among them.

4.1.3. Face recognition experiments on FRGC v2

Experiments are also conducted on a subset of facial data in experiment 4 of FRGC v2 that measures the recognition performance from uncontrolled images. Experiment 4 is the most challenging FRGC experiment which has 8014 single uncontrolled still images of 466 persons in the query set. We choose the first 10 images of each person in this set if the number of images is not less than 10. Then we collect 800 images of first 80 persons. The images are all cropped to a size of 32 $\times$ 32. The bottom row of Fig. 3 shows the facial images of one person in our experiment.

We randomly select two images of each person as the training set and the rest images are used as the testing set. Table 4 shows the recognition results on original raw data of experiment 4 of FRGC v2. The recognition rate curves versus the variation of dimensions are illustrated in Fig. 4. One can find that our proposed LDP is superior to other methods on uncontrolled facial data. We also compare *tensor* methods on the LBP feature yielded from FRGC v2. Again, the results presented in Table 3 show that LDP is better than other methods in comparison.

4.2. Handwriting digits classification

The handwriting digits classification experiments are designed to test the performance of feature extraction on multi-resolution images, which are widely used for image processing. Since Gabor filter is the most popular multi-resolution operator which has been frequently used in texture analysis [34] and digit recognition

[2] http://yann.lecun.com/exdb/mnist/

[3] TPCA and TLPP were designed for matrices only, so we just test it on order-2 tensor data.

[4] Three compared smooth regularization algorithms are all proposed in [43].

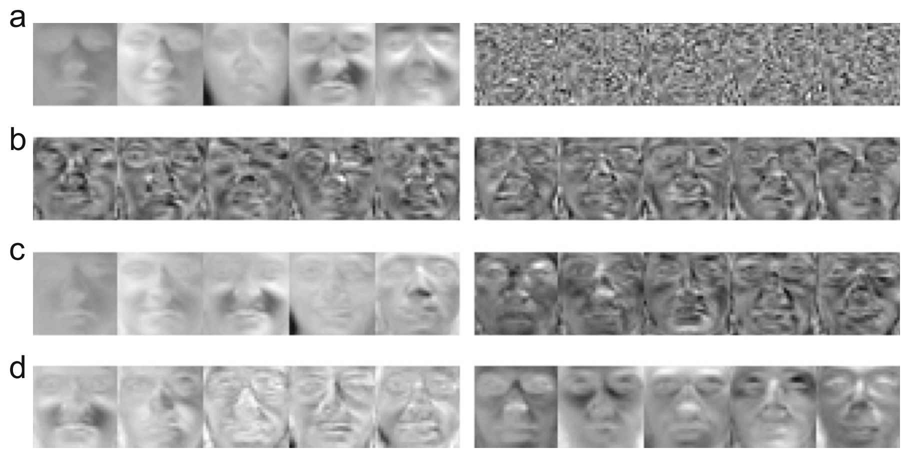

Fig. 2. Eigenfaces, Fisherfaces and Lorentzianfaces calculated from the facial images in the FRGC v2 database. For each row, the first five faces are spanned by the leading eigenvectors and the last five faces are spanned by the tailing eigenvectors. (a) Eigenfaces, (b) Fisherfaces, (c) Lorentzianfaces (γ=0.1), and (d) Lorentzianfaces (γ=1.5).

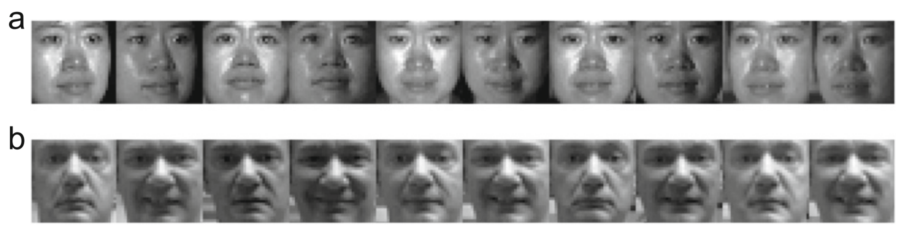

Fig. 3. Some facial images used in our experiments. All images are 32×32 pixels in size. (a) CMU PIE, (b) FRGC v2

[27], we perform experiments on Gabor features of MNIST database. Firstly, we choose the first 20 images of each class for the experiments. Then the images are all cropped to a size of 28 $\times$ 28 (Fig. 5). For each image, we extract 24 Gabor features in four different scales and six different directions and down-sample them to 7 $\times$ 7 images [26]. Then we get order-3 tensor features of size 24 $\times$(7 $\times$ 7). We randomly take five images as the training set and the remaining 15 images as the testing set. The classification results, listed in Table 5 and showed in Fig. 4, demonstrate that our proposed LDP and TLDP perform better than other methods on the multi-resolution image set, respectively.

4.3. Discussions

We find the free parameters for the tested methods in the following way. The number of K-nearest neighborhoods in LPP and the intraclass neighbor parameter $\hat{K}$ in MFA are chosen as $l-1$, where l denotes the number of training samples per class. The interclass neighbor parameter $\check{K}$ in MFA, the values of the Gaussian kernel parameter t in LPP and the value of γ in LDP are all tuned optimally in the training phase.

By conducting experiments systematically, we find that our proposed LDP and its extensions can perform better than those traditional methods on the three databases. It can also be seen that the kernel and tensor approaches outperform vector-based methods in some databases, but the vector-based methods have their own advantages under some circumstances (Table 4). In addition, the results demonstrate that, when the training set is not enough to characterize the data distribution (only three training images for CMU PIE or two training images for FRGC v2), *discrepancy criterion* based MMC and its tensor extension appear to be less effective than other methods (Tables 2 and 4). Fortunately, the kernel trick can significantly improve the performance of MMC. If the training set adequately characterizes the data distribution as the case of five training images for MNIST, MMC has the potential to outperform other methods (Table 5). But all experiments show that MMC does not perform better than LDP.

The face recognition experiments also demonstrate the power of smooth regularization for dimensionality reduction. By using 2-D Laplacian smoothing regularization technique, the regularized algorithms significantly outperform the corresponding ordinary versions. From Tables 2 and 4, we can see the performance of traditional algorithms is significantly improved by smooth regularization (e.g., the recognition rate of LPP is improved from 70.0% to 77.9% on CMU PIE and from 81.3% to 87.5% on FRGC v2, respectively). SLDP also outperforms the original LDP on both two face data. This is because that smooth regularization can explicitly take into account the spatial relationship between the pixels in an image and the projection vectors can be smoother than those obtained by the ordinary dimensionality reduction algorithms.

For handwriting digits data, the dimension of the embedding subspace significantly affects the performance of some feature extraction algorithms. This is because when the dimension increases, the noise begins to appear in the embedding subspace and starts to affect the accuracy. This drawback is not only for LDP, but also for most other feature extraction algorithms (e.g., LDA and MMC). However, solving this problem is clearly beyond the scope of this paper.

5. Conclusions and future work

This paper presents a novel discriminant analysis method called Lorentzian discriminant projection (LDP). In the first step, we construct a Lorentzian manifold to model both local and global discriminant and geometric structures of the data set. Then, an approach to Lorentzian metric learning is proposed to learn metric tensor from the original high-dimensional sample space

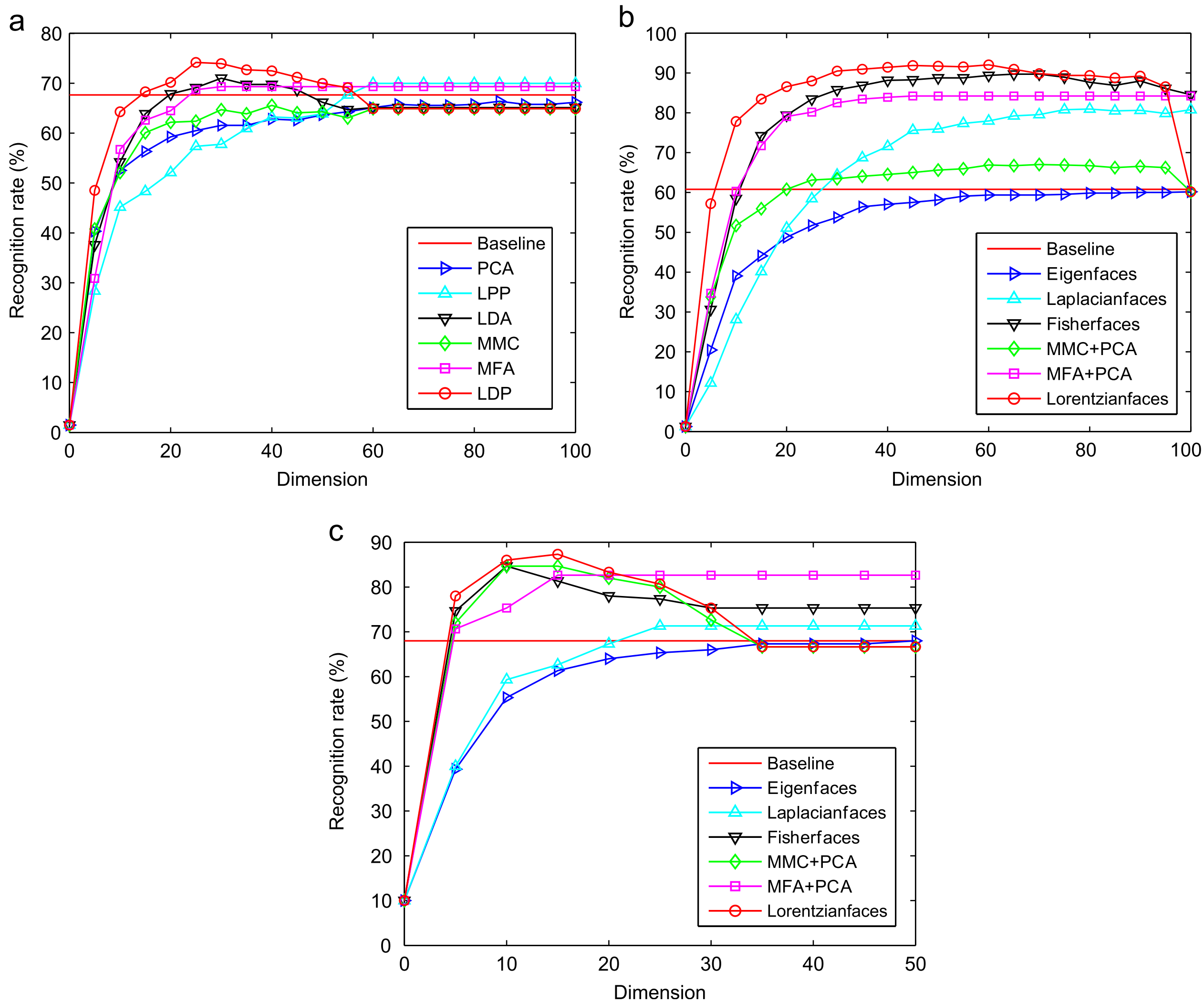

Fig. 4. The recognition rate curves of linear methods versus the variation of dimensions. (a) CMU PIE. (b) FRGC v2. (c) MNIST.

Table 2
The maximal recognition results (%) on the original CMU PIE facial data (vector and matrix images). Using the original data directly without dimensionality reduction is the baseline. The percentage of energy retained in the PCA step is 97%. The optimal dimensions of feature space are given in the brackets.

Method	*Linear*	*Kernel*	Order-2 *Tensor*	*Smooth regularized*
Baseline	67.65	–	–	–
PCA	66.6 (86)	67.7 (188)	68.5 (14, 3)	–
LPP	70.0 (60)	62.4 (30)	69.5 (18, 27)	77.9 (55)
LDA	71.2 (27)	81.9 (67)	74.8 (25, 5)	76.9 (69)
MMC	66.0 (39)	83.4 (80)	73.1 (31, 26)	–
MFA	70.8 (28)	80.5 (27)	73.5 (28, 5)	73.1 (98)
LDP	**74.8 (26)**	**84.0 (115)**	**79.4 (15, 14)**	**79.2 (69)**

and apply it to the low-dimensional discriminant subspace. In this way, both the local class and the global data structures can be well preserved in the reduced low-dimensional discriminant subspace. We also derive the kernel, tensor and smooth regularized extension of LDP for nonlinear and multi-linear data, respectively. The experimental results have shown that our proposed LDP, KLDP, TLDP and SLDP are all promising.

For future work, we are considering the sparsity of the data set. For example, our LDP only model the Lorentzian manifold by combining the L_2 distances as the coordinates. One of the disadvantages of this approach is that the learnt projective maps

Table 3
The maximal recognition results (%) on the LBP features of CMU PIE and FRGC v2 facial data. Using the LBP features directly without dimensionality reduction is the baseline. The optimal dimensions of feature space are given in the brackets.

Method	CMU PIE	FRGC v2
Baseline	85.3	75.2
TLDA	90.6 (34, 4, 4)	82.3 (33, 4, 3)
TMMC	90.1 (53, 3, 4)	80.6 (53, 4, 3)
TMFA	89.5 (48, 4, 4)	81.1 (57, 4, 3)
TLDP	**91.6 (52, 4, 3)**	**82.5 (42, 3, 3)**

Table 4
The maximal recognition results (%) on the original FRGC v2 facial data (vector and matrix images). Using the original data directly without dimensionality reduction is the baseline. The percentage of energy retained in the PCA step is 99%. The optimal dimensions of feature space are given in the brackets.

Method	*Linear*	*Kernel*	Order-2 *Tensor*	*Smooth regularized*
Baseline	60.8	–	–	–
PCA	60.8 (151)	60.8 (144)	60.9 (32, 22)	–
LPP	81.3 (94)	78.1 (29)	75.2 (24, 17)	87.5 (154)
LDA	90.2 (63)	91.7 (77)	85.5 (13, 15)	90.3 (107)
MMC	67.2 (74)	90.5 (115)	64.7 (27, 29)	–
MFA	84.4 (42)	89.7 (28)	83.8 (15, 15)	89.4 (107)
LDP	**92.0 (59)**	**92.7 (79)**	**87.2 (20, 32)**	**92.5 (79)**

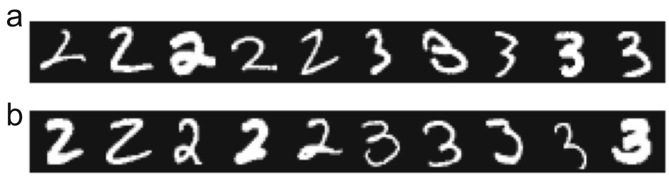

Fig. 5. Some handwriting digits in the MNIST database. All images are 28 $\times$28 pixels in size. (a) Training, (b) Testing.

Table 5
The maximal classification results (%) on the Gabor features of MNIST data. Using the Gabor features directly without dimensionality reduction is the baseline. For linear methods, the percentage of energy retained in the PCA step is 98%. The optimal dimensions of feature space are given in the brackets.

Method	*Linear*	Order-3 *Tensor*
Baseline	68.0	–
PCA	68.0 (47)	–
LPP	71.3 (22)	–
LDA	84.7 (10)	86.0 (13, 6, 4)
MMC	84.7 (9)	79.3 (21, 5, 7)
MFA	84.0 (13)	86.0 (16, 5, 5)
LDP	**87.3 (15)**	**87.4 (24, 5, 4)**

are linear combination of *all* the original features. But recent psychological and physiological evidence have shown that the representation of objects in human brain may be sparse [38,39]. How to utilize the sparsity for the Lorentzian metric learning framework effectively is an interesting direction. Another open problem in LDP is that the dense matrix eigenvalue problem is computationally expensive to solve especially for large-scale problems. Recently, Cai et al. [40–42] propose a new regularization framework for linear dimensionality reduction called spectral regression (SR). With this framework, different kinds of regularizers can be naturally incorporated in dimensionality reduction algorithms which make them more flexible. Furthermore, SR only needs to solve a set of regularized least squares problems and computational analysis shows that it has only linear-time complexity which is huge speed up comparing to the cubic-time complexity of the ordinary approaches. We intend to further investigate regularization and least squares formulation for our LDP model.

Acknowledgments

The first author would like to thank Wei Zhang and Yanqi Liu for valuable discussions. This work was supported by the grants of the National Science Foundation of China, Nos. 60673006, 60533060 and U0935004; New Century Excellent Talents in University of China, No. NCET-05-0275.

References

[1] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenface vs Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[2] M. Turk, A. Pentland, Face recognition using Eigenfaces, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 1991.
[3] K. Muller, S. Mika, G. Riitsch, K. Tsuda, B. Scholkopf, An introduction to kernel-based learning algorithms, IEEE Transactions on Neural Networks 12 (2) (2001) 181–201.
[4] B. O'Neill, Semi-Riemannian Geometry with Applications to Relativity, Academic Press, New York, 1983.
[5] D. Zhao, Z. Lin, X. Tang, Classification via semi-Riemannian spaces, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
[6] H. Wang, S. Yan, D. Xu, X. Tang, T. Huang, Trace ratio vs. ratio trace for dimensionality reduction, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2007.
[7] D.L. Donoho, C. Grimes, Hessian eigenmaps: new local linear embedding techniques for high-dimensional data, Proceedings of the National Academy of Sciences 100 (10) (2005) 5591–5596.
[8] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[9] L. Saul, S. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, Journal of Machine Learning Research 4 (2003) 119–155.
[10] J. Tenenbaum, V. Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[11] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 40–51.
[12] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
[13] Z. Zhang, H. Zha, Principal manifolds and nonlinear reduction via local tangent space alignment, SIAM Journal of Scientific Computing 26 (2002) 313–338.
[14] S. Lafon, A. Lee, Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set, IEEE Transactions on Pattern Analysis and Machine Intelligence Parameterization 28 (9) (2006) 1393–1403.
[15] D. Zhao, Formulating LLE using alignment technique, Pattern Recognition 39 (11) (2006) 2233–2235.
[16] D. Zhao, Z. Lin, X. Tang, Laplacian PCA and its applications, in: IEEE International Conference on Computer Vision (ICCV), 2007.
[17] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Advances in Neural Information Processing System, vol. 14, Vancouver, British Columbia, Canada, 2001.
[18] X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, in: IEEE International Conference on Computer Vision (ICCV), 2005.
[19] X. He, P. Niyogi, Locality preserving projections, in: Advances in Neural Information Processing Systems (NIPS), vol. 16, 2003.
[20] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, in: Neural Information Processing Systems (NIPS), 2003.
[21] F. Wang, C. Zhang, Feature extraction by maximizing the average neighborhood margin, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2007.
[22] H. Wang, Q. Wu, L. Shi, Y. Yu, N. Ahuja, Out-of-core tensor approximation of multi-dimensional matrices of visual data, in: ACM SIGGRAPH, 2005.
[23] J. Ye, Generalized low rank approximations of matrices, in: International Conference on Machine Learning (ICML), 2004.
[24] J. Yang, D. Zhang, J.-Y. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (8) (2007) 1503–1504.

[25] R. Liu, Z. Su, Z. Lin, X. Hou, Lorentzian discriminant projection and its applications, in: Asian Conference on Computer Vision (ACCV), 2009.
[26] W. Zhang, Z. Lin, X. Tang, Tensor linear Laplacian discrimination (TLLD) for feature extraction, Pattern Recognition 42 (9) (2009) 1941–1948.
[27] D. Xu, S. Yan, H.J. Zhang, Z. Liu, H.Y. Shum, Concurrent subspace analysis, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[28] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H. Zhang, Discriminant analysis with tensor representation, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[29] J. Yang, D. Zhang, A. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence Parameterization 26 (1) (2004) 131–137.
[30] D. Cai, X. He, J. Han, Subspace learning based on tensor analysis, Technical Report No. 2572, Department of Computer Science, University of Illinois at Urbana-Champaign (UIUCDCS-R-2005-2572), 2005.
[31] J. Ye, R. Janardan, O. Li, Two-dimensional linear discriminant analysis, in: Advances in Neural Information Processing Systems, 2005.
[32] P. Philips, P. Flynn, T. Scruggs, K. Bowyer, Overview of the face recognition grand challenge, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[33] T. Ojala, M. Pietikainen, T. Maenpaa, Gray scale and rotation invariant texture classification with local binary patterns, in: European Conference on Computer Vision (ECCV), 2000.
[34] S. Arivazhagan, L. Ganesan, S. Priyal, Texture classification using Gabor wavelets based rotation invariant features, Pattern Recognition Letters 27 (16) (2005) 1976–1982.
[35] M.H. Yang, Kernel Eigenfaces vs. kernel Fisherfaces: face recognition using kernel methods. in: IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
[36] X. He, D. Cai, P. Niyogi, Tensor subspace analysis, in: Advances in Neural Information Processing Systems, 2005.
[37] T. Kandade, Computer Recognition of Human Faces, Birkhauser Verlag, Basel, Switzerland, 1977.
[38] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[39] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence Parameterization 31 (2) (2009) 210–227.
[40] D. Cai, X. He, J. Han, Spectral regression for efficient regularized subspace learning, in: IEEE International Conference on Computer Vision (ICCV), 2007.
[41] D. Cai, X. He, J. Han, SRDA: an efficient algorithm for large scale discriminant analysis, IEEE Transactions on Knowledge and Data Engineering 20 (1) (2008) 1–12.
[42] D. Cai, X. He, J. Han, Efficient kernel discriminant analysis via spectral regression, in: International Conference on Data Mining (ICDM'07), 2007.
[43] D. Cai, X. He, Y. Hu, J. Han, T. Huang, Learning a spatially smooth subspace for face recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

About the Author—RISHENG LIU received the B.Sc. degree in Mathematics from Dalian University of Technology in 2007. He is currently a Ph.D. candidate in the School of Mathematical Sciences of Dalian University of Technology. His research interests include machine learning, pattern recognition, computer vision and manifold learning.

About the Author—ZHOUCHEN LIN received the Ph.D. degree in Applied Mathematics from Peking University in 2000. He is currently a researcher in Visual Computing Group, Microsoft Research, Asia. His research interests include computer vision, computer graphics, pattern recognition, statistical learning, document processing, and human computer interaction. He is a senior member of the IEEE.

About the Author—ZHIXUN SU received the B.Sc. degree in Mathematics from Jilin University in 1987 and M.Sc. degree in Computer Science from Nankai University in 1990. He receive his Ph.D. degree in 1993 from Dalian University of Technology, where he has been a professor in the School of Mathematical Sciences since 1999. His research interests include computer graphics and image processing, computational geometry, computer vision, etc.

About the Author—KEWEI TANG received the B.Sc. degree in Mathematics from Liaoning Normal University in 2008. He is currently an M.Sc. candidate in the School of Mathematical Sciences of Dalian University of Technology. His research interests include pattern recognition and computer vision.