Lorentzian Discriminant Projection and Its Applications

Risheng Liu[1], Zhixun Su[1], Zhouchen Lin[2], and Xiaoyu Hou[1]

[1] Dalian University of Technology, Dalian 116024, China
[2] Microsoft Research Asia, Beijing 100080, China
zxsu@dlut.edu.cn

Abstract. This paper develops a supervised dimensionality reduction method, Lorentzian Discriminant Projection (LDP), for discriminant analysis 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. The experimental results on benchmark databases demonstrate the effectiveness of our proposed method.

1 Introduction

In recent years, the computer vision and pattern recognition community has witnessed a growing interest in dimensionality reduction. One of the most successful and well-studied techniques is the supervised discriminant analysis. We devote this paper to addressing the discriminant analysis from the perspective of Lorentzian geometry.

1.1 Related Work

Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] 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 [14][5][11][6] [7][10]. 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) [16][9], Neighborhood Preserving Embedding (NPE) [15], Marginal Fisher Analysis (MFA) [8], Maximum Margin Criterion (MMC) [17], Average Neighborhood Margin Maximization (ANMM) [18], Semi-Riemannian Discriminant Analysis (SRDA) [4] and Unsupervised Discriminant Projection (UDP) [19].

1.2 Our Approach

Yang *et al* . [19] adapted both local and global scatters to unsupervised dimensionality reduction. They maximized the ratio of the global scatters to the local scatters. Zhao *et al* . [4] first applied the semi-Riemannian geometry to classification [4]. 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. We first construct a manifold to model the local class and the global data structures. In this way, both of the local discriminant and the global geometric structures of the data set can accurately be characterized by learning a special Lorentzian metric tensor on the newly built manifold. In fact, the role of the Lorentzian metric learning is equivalent to the media of transferring geometry from the sample space to the feature space.

The rest of this paper is organized as follows. In Section 2, we provide the Lorentzian Discriminant Projection algorithm. The experimental results of LDP approach to real-world face analysis and handwriting digits classification are presented in Section 3. Finally, we summarize our work and conclude the paper in Section 4.

2 Lorentzian Discriminant Projection

2.1 Fundamentals of Lorentzian Manifold

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 $p + q = n$. This concept is extensively used in general relativity, as a basic geometric tool for modeling the space-time in physics.

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 nondegeneracy 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 $\mathbf{r} \neq 0$). One may refer to [3] for more details.

2.2 The Motivation of LDP

The theory and algorithm in this paper are based on the perspective that the discrimination is tightly related to both local class and global data structures. Our motivation of LDP are twofold: 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 4-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 use the relationship (e.g., distances) between the sample and its within-class neighbors to model the local class structure. On the other hand, we characterize the global data structure by the dissimilarities between each point and the global geometric centorid. By performing discrepancies of within-class quantities and global quantities, we obtain an ambient space with Lorentzian metrics where coordinates are characterized by dissimilarities between sample pairs (each point with its within-class neighbors and the global geometric centorid). Therefore, the discriminant structure of the data set is initially modeled as a Lorentzian manifold.

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 applied it to the discriminant subspace.

2.3 Modeling Features as a Lorentzian Manifold

For supervised dimensionality reduction task, the samples can be represented as a point set $\mathcal{S}_x = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$, $\mathbf{x}_i \in \mathbb{R}^n$. 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 $\mathbf{y}_i$, $\mathcal{S}_{y_i} = \{\mathbf{y}_i, \mathbf{y}_1^i, ..., \mathbf{y}_{m_i-1}^i\}$ (points share the same class label with $\mathbf{y}_i$) and $\bar{\mathbf{y}}$ (the geometric centroid of $\mathcal{S}_y$, i.e., $\bar{\mathbf{y}} = \frac{1}{m}\sum\limits_{i=1}^{m}\mathbf{y}_i$), a new point $\mathbf{d}_{y_i}$ is defined as:

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

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

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

where the metric matrix $\mathbf{G}_i^l$ is real diagonal and the signature of the metric is $(m_i - 1, 1)$, $\mathbf{D} = [\mathbf{e}_{m_i}, -\mathbf{I}_{m_i, m_i}]^T$ ($\mathbf{I}_{m_i, m_i}$ 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 $\mathbf{Y}_i = [\mathbf{y}_i, \mathbf{y}_1^i, ...\mathbf{y}_{m_i-1}^i, \bar{\mathbf{y}}]$.

Then the total Lorentzian metric tensor can be given as:

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

where $\mathbf{L} = \sum_{i=1}^{m} \mathbf{B}_i \mathbf{D} \mathbf{G}_i^l \mathbf{D}^T \mathbf{B}_i^T$, $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_m, \bar{\mathbf{y}}]$ and $\mathbf{B}_i$ is a binary selection matrix of size $m \times (m_i + 1)$ which satisfies $\mathbf{Y}_i = \mathbf{Y}\mathbf{B}_i$ [13][12].

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:

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

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 Manifold

The Lorentzian metric matrices $\mathbf{G}_i^l$ are key to the proposed dimensionality reduction problem. We give a novel method to learn it from the sample set $\mathcal{S}_x$ and then apply it to the feature set $\mathcal{S}_y$. 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 of the Lorentzian metric tensor 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

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

and

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

We may minimize this metric and obtain the following optimization problem:

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

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

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

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 can make the metric locally unbiased. So the negative definite part $\check{\lambda}_i$ of $\mathbf{G}_i^l$ can be determined by:

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

We empirically find that the discriminability will be enhanced if we choose a positive factor $\gamma \in [0, 1]$ to multiply the negative part i.e., $\check{\lambda}_i \to \gamma\check{\lambda}_i$. The value of $\check{\lambda}$ can be determined by cross validation.

3 Experimental Results

Experiments are conducted on Yale [3], FRGC [20] and USPS [4] databases to test the performance of LDP against existing algorithms. For these databases, the image set of each subject is split into different gallery and probe sets, where Gm/Pn means m images are randomly selected for training and the remaining n images are for testing. Such a trial is repeated 20 times.

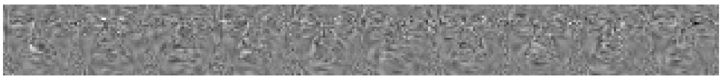
In the face analysis (representation and recognition) problem, we want to use LDP to learn an optimal discriminant subspace which is spanned by the columns of $\mathbf{U}$ in (6) for face representation. The eigenvectors can be displayed as images, called the Lorentzianfaces in our approach. Using the facial images in experiment 4 of FRGC version 2 as the training set, we present the first 10 Lorentzianfaces in Fig 1, together with Eigenfaces [1], Laplacianfaces [9] and Fisherfaces [2].

We perform the discriminant subspace learning on the expressive features yielded by PCA which is classic and well-recognized preprocessing. For the PCA-based two step strategy, the number of principal components is a free parameter. In our experiments, we choose the percentage of energy retained in PCA preprocessing step 99%.

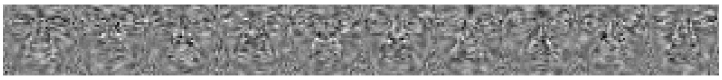
[3] Available at http://cvc.yale.edu/projects/yalefaces/yalefaces.html

[4] Available at http://www.cs.toronto.edu/ roweis/data.html

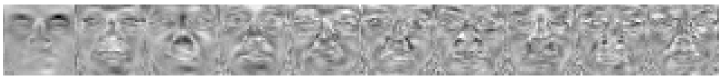
(a) Eigenfaces



(b) Laplacianfaces



(c) Fisherfaces



(d) Lorentzianfaces

Fig. 1. Eigenfaces, Laplacianfaces, Fisherfaces and our proposed Lorentzianfaces.

Fig. 2. Some cropped Yale facial images.

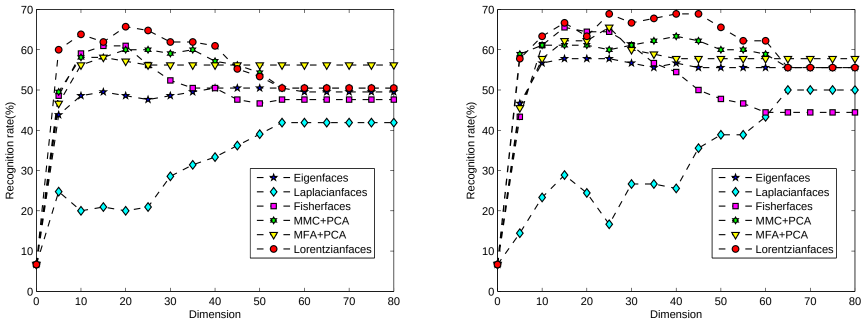
3.1 Experiments on Yale

The Yale face database was constructed at the Yale Center for the Computational Vision and Control. It contains 165 gray-scale images of 15 subjects under various facial expressions and lighting conditions such as center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and winking. In our experiment, we cropped each image to a size of 32×32. Figure 2 shows some cropped images in Yale database.

The average recognition rate of each method and the corresponding dimension are given in Table 1. The recognition rate curves versus the variation of dimensions are illustrated in Figure 3. As can be seen, the proposed Lorentzianfaces outperforms other methods involved in this experiment.

Table 1. The average recognition results on Yale database (mean $\pm$ std %). The optimal dimension of face subspace are given in the brackets. U and S mean unsupervised and supervised methods

Method	G4/P7	G6/P5	Type
Eigenfaces (PCA)	57.62 $\pm$ 3.69 (40)	59.56 $\pm$ 5.22 (15)	U
Laplacianfaces (LPP)	46.24 $\pm$ 5.20 (55)	46.44 $\pm$ 6.96 (65)	U
Fisherfaces (LDA)	70.38 $\pm$ 4.83 (15)	72.72 $\pm$ 6.41 (15)	S
MMC + PCA	70.05 $\pm$ 4.48 (20)	70.50 $\pm$ 5.63 (40)	S
MFA + PCA	67.10 $\pm$ 6.04 (15)	70.06 $\pm$ 6.48 (25)	S
Lorentzianfaces (LDP)	**72.90** $\pm$ **5.28 (20)**	**74.28** $\pm$ **6.19 (15)**	S



(a) G4/P7 (b) G6/P5

Fig. 3. The recognition rate curves versus the variation of dimensions on Yale database. The left figure shows the G4/P7 results and the right one shows the G6/P5 results.

3.2 Experiments on FRGC

Experiments are also conducted on a subset of facial data in experiment 4 of FRGC version 2 [20] 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 subjects in the query set. We choose the first 30 images of each subject in this set if the number of images is not less than 30. Thus we get 2850 images of 95 subjects. The images are all cropped to a size of 32×32. Figure 4 shows facial images of one subject in our experiment.

Table 2 shows the average recognition results on experiment 4 of FRGC version 2. Figure 5 displays the recognition rate curves versus the feature space dimensions when performing these methods. One can see that Lorentzianfaces is significantly better than other methods in comparison.

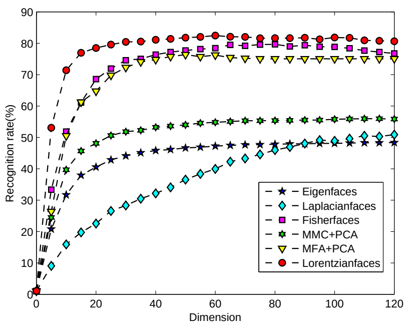
3.3 Experiments on USPS

The USPS handwriting digital data includes 10 classes from "0" to "9". Each class has 1100 samples. The first 200 images of each class are chosen for our experiments. We

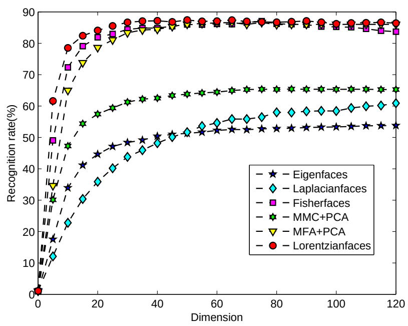
Fig. 4. Some cropped FRGC version 2 facial images.

Table 2. The average recognition results on experiment 4 of FRGC version 2 database (mean $\pm$ std %). The optimal dimension of face subspace are given in the brackets. U and S mean unsupervised and supervised methods

Method	G3/P27	G5/P25	Type
Eigenfaces (PCA)	48.31 $\pm$ 4.32 (120)	58.61 $\pm$ 3.47 (120)	U
Laplacianfaces (LPP)	46.61 $\pm$ 3.54 (120)	57.74 $\pm$ 2.08 (120)	U
Fisherfaces (LDA)	74.21 $\pm$ 5.35 (80)	87.49 $\pm$ 1.71 (75)	S
MMC + PCA	61.08 $\pm$ 7.09 (110)	77.97 $\pm$ 6.42 (85)	S
MFA + PCA	72.93 $\pm$ 5.08 (50)	85.72 $\pm$ 3.80 (65)	S
Lorentzianfaces (LDP)	**80.63 $\pm$ 5.19 (60)**	**89.41 $\pm$ 2.24 (50)**	S



(a) G3/P27



(b) G5/P25

Fig. 5. The recognition rate curves versus the variation of dimensions on experiment 4 of FRGC version 2 database. The left figure shows the G3/P27 results and the right one shows the G5/P25 results.

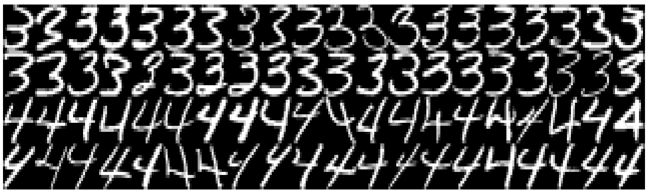


Fig. 6. Some cropped USPS digital images.

directly apply all algorithms to the normalized data without using PCA as preprocessing. The average classification results are shown in Table 3. The performance of LDP is again better than other methods under consideration.

Table 3. The average classification results on USPS database (mean $\pm$ std %). The optimal dimension of feature space are given in the brackets. U and S mean unsupervised and supervised methods

Method	G30/P170	G40/P160	Type
PCA	99.27 $\pm$ 0.21 (10)	99.42 $\pm$ 0.24 (20)	U
LPP	76.29 $\pm$ 4.65 (25)	87.60 $\pm$ 10.54 (25)	U
LDA	80.23 $\pm$ 8.18 (5)	88.70 $\pm$ 7.96 (25)	S
MMC	95.04 $\pm$ 1.57 (20)	95.47 $\pm$ 0.94 (15)	S
MFA	80.81 $\pm$ 2.89 (5)	87.63 $\pm$ 2.98 (15)	S
LDP	**99.38 $\pm$ 0.18 (20)**	**99.50 $\pm$ 0.21 (15)**	S

3.4 Discussions

By conducting experiments systematically, we can find that: as in LPP and MFA, the number of neighbors (e.g., k, $\hat{k}$ and $\check{k}$) is the most important parameter. The parameter k in LPP is set to 5 in all experiments. For face recognition, the parameters $\hat{k}$ and $\check{k}$ in MFA are set to $m-1$ (m is the number of images in the gallery set) and 210. For handwriting digits classification, we set the parameters $\hat{k}$ and $\check{k}$ to 7 and 40. The Gaussian kernel $\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{t}\right)$ is used in LPP. We set the parameter t to 250 in our experiments.

4 Conclusions

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 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. The experimental results have shown that our proposed LDP is promising.

References

1. Turk, M.A., Pentland, A.P.: Face Recognition Using Eigenfaces. In: Proc. CVPR (1991)
2. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenface vs. Fisherfaces:Recognition Using Class Specific Linear Projection. IEEE TPAMI, 19, 711–720 (1997)
3. O'Neill, B.: Semi-Riemannian Geometry with Applications to Relativity. Academic Press, New York (1983)
4. Zhao, D., Lin, Z., Tang, X.: Classification via semi-Riemannian spaces. In: Proc. CVPR (2008)
5. Donoho, D.L., Grimes, C.: Hessian Eigenmaps: New Local Linear Embdedding Techniques for High-dimensioanl data. PNAS, 102, 7426–7431 (2005)
6. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science, 290, 2323–2326 (2000)
7. Tenenbaum, J.B., Silva, V.D., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science, 290, 2319–2323 (2000)
8. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE TPAM, 27, 40–51 (2007)
9. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face Recognition Using Laplacianfaces. IEEE TPAM, 27, (2005)
10. Zhang, Z., Zha, H.: Principal manifolds and nonlinear dimensionality reduction by local tangent space alignment. SIAM Journal of Scientific Computing, 26, 313–338 (2004)
11. Lafon, S., Lee, A.B.: Diffusion Maps and Coarse-Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Data Set Parameterization. IEEE TPAM, 28, 1393–1403 (2006)
12. Zhao, D.: Formulating lle using alignment technique. Pattern Recogition, 39, 2233–2235 (2006)
13. Zhao, D., Lin, Z., Tang, X.: Laplacian PCA and its applications. In: Proc. ICCV (2007)
14. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Proc. NIPS (2001)
15. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood Preserving Embedding. In: Proc. ICCV (2005)
16. He, X., Niyogi, P.: Locality Preserving Projections. In: Proc. NIPS (2003)
17. Li, H., Jiang, T., Zhang,K.: Efficient Robust Feature Extraction by Maximum Margin Criterion. In: Proc. NIPS (2003)
18. Wang, F., Zhang, C.: Feature Extraction by Maximizing the Average Neighborhood Margin. In: Proc. CVPR (2007)
19. Yang, J., Zhang, D., Yang, J.Y., Niu, B.: Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. IEEE TPAM, 29, 650–664 (2007)
20. Phillips, P.J., Flynn, P.J., Scruqqs, T., Bowyer, K.W., Jin, C., Hoffman, K., Marques, J., Jaesik, M., Worek, W.: Overview of The Face Recognition Grand Challenge. In: Proc. CVPR (2005)