Tensor Distance Based Multilinear Locality-Preserved Maximum Information Embedding

Yang Liu, Yan Liu, and Keith C. C. Chan

***Abstract*— This brief paper presents a unified framework for tensor-based dimensionality reduction (DR) with a new tensor distance (TD) metric and a novel multilinear locality-preserved maximum information embedding (MLPMIE) algorithm. Different from traditional Euclidean distance, which is constrained by the orthogonality assumption, TD measures the distance between data points by considering the relationships among different coordinates. To preserve the natural tensor structure in low-dimensional space, MLPMIE directly works on the high-order form of input data and iteratively learns the transformation matrices. In order to preserve the local geometry and to maximize the global discrimination simultaneously, MLPMIE keeps both local and global structures in a manifold model. By integrating TD into tensor embedding, TD-MLPMIE performs tensor-based DR through the whole learning procedure, and achieves stable performance improvement on various standard datasets.**

***Index Terms*— Dimensionality reduction, manifold learning, multilinear embedding, tensor distance.**

I. Introduction

Pattern analysis tasks often suffer from a high-dimensional feature space, which leads to low recognition accuracy and expensive computational costs. Dimensionality reduction (DR) techniques provide a means to solve this problem by generating a low-dimensional equivalence of the original feature space for the given observations and targets.

However, traditional DR algorithms unfold input data to vectors before embedding, even though the data are naturally high-order tensors. Such kind of vectorization largely increases the computational costs of data analysis and seriously destroys the intrinsic tensor structure of high-order data [1]–[3]. To address these problems, Vasilescu and Terzopoulos have introduced multilinear algebra into DR for high-order data analysis [4]. Since then, many multilinear DR techniques, also referred as tensor-based DR techniques, have been proposed. Typical multilinear DR algorithms include multilinear principal component analysis (MPCA) [5], [1], [2], [6], [7], multilinear discriminant analysis (MLDA) [8], [3], [9], [10], multilinear independent component analysis [11], tensor subspace analysis [12], tensor-based locality preserving projections (TLPP) [13], tensor linear Laplacian discrimination [14], and so on [15]–[18].

Most of the existing tensor-based techniques intend to preserve the relationships among high-order data, which are generally measured by the distances between different data points. Therefore, the performance of tensor-based DR does not only depend on embedding strategies but also closely relate to distance metrics. Current tensor-based techniques simply use the Euclidean distance to measure relationships among different data points. However, in Euclidean space, a high-order data point $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is represented by the coordinates $x_1, x_2, \ldots, x_{I_1 \times I_2 \times \cdots \times I_N}$ under corresponding bases $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{I_1 \times I_2 \times \cdots \times I_N}$, where $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ $(i \neq j)$. It means that any two bases $\mathbf{e}_i$ and $\mathbf{e}_j$ in the Euclidean space are assumed to be mutually perpendicular, so the coordinates x_i and x_j are independent of each other. Unfortunately, this orthogonality assumption ignores the relationships among different coordinates for high-order data, such as the spatial relationships of pixels in images, and thus limits the performance of further tensor based embedding.

To release this orthogonality assumption, we propose a new tensor distance (TD) metric to measure the relationships between high-order data points. Based on this distance metric, we further present a unified DR framework called TD based multilinear locality-preserved maximum information embedding (TD-MLPMIE), according to a vector based-algorithm called locality-preserved maximum information projection (LPMIP) [19]. Inheriting the properties from LPMIP, TD-MLPMIE preserves local geometry and maximizes global discrimination by balancing both *within locality* and *between locality* simultaneously in a manifold model. Moreover, as a multilinear technique, TD-MLPMIE improves data analysis accuracy while reducing computational complexity.

The rest of this paper is organized as follows. Section II proposes the TD metric. Section III presents TD-MLPMIE. Theoretical analysis of the proposed method is given in Section IV. Experimental results are reported in Section V. This paper ends with the conclusion in Section VI.

II. Tensor Distance

For some kinds of high-order data, traditional Euclidean distance may not reflect the real distance between two data points because of the orthogonality assumption discussed previously. In this section, we propose a new distance metric called TD to model the correlation among different coordinates of data with arbitrary number of orders.

Given a data point $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} (N > 1)$, we use $\mathbf{x}$ to denote the vector form representation of $\mathcal{X}$. Therefore, the element $\mathcal{X}_{i_1 i_2 \ldots i_N} (1 \leq i_j \leq I_j, 1 \leq j \leq N)$ in $\mathcal{X}$ is corresponding to x_l, i.e., the lth element in $\mathbf{x}$, where $l = i_1 + \sum_{j=2}^{N} (i_j - 1) \prod_{o=1}^{j-1} I_o (2 \leq j \leq N)$. Then TD between two tensors $\mathcal{X}$ and $\mathcal{Y}$ can be represented as

$$d_{TD} = \sqrt{\sum_{l,m=1}^{I_1 \times I_2 \times \cdots \times I_N} g_{lm}(x_l - y_l)(x_m - y_m)} = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{G} (\mathbf{x} - \mathbf{y})} \quad (1)$$

where g_{lm} is the metric coefficient and $\mathbf{G}$ is the metric matrix. To reflect the intrinsic relationships between different coordinates for high-order data, a natural consideration is that

Manuscript received February 12, 2010; revised May 28, 2010 and July 20, 2010; accepted July 29, 2010. Date of current version November 3, 2010. This work was supported in part by the Hong Kong Research Grants Council General Research Fund PolyU 5204/09E.

The authors are with the Department of Computing, the Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: csygliu@comp.polyu.edu.hk; csyliu@comp.polyu.edu.hk; cskcchan@comp.polyu.edu.hk).

Color versions of one or more of the figures in this brief are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2010.2066574

the metric coefficients should be related to the element distances. Wang *et al.* [20] have already demonstrated that, for image data, i.e., the second-order tensors, if the metric coefficients depend properly on distances of pixel locations, the obtained distance metric can effectively reflect the spatial relationships between pixels. Inspired by this paper, we design the following metric matrix $\mathbf{G}$:

$$g_{lm} = \frac{1}{2\pi\sigma_1^2}\exp\left\{\frac{-\|\mathbf{p}_l - \mathbf{p}_m\|_2^2}{2\sigma_1^2}\right\} \tag{2}$$

where σ_1 is a regularization parameter and $\|\mathbf{p}_l - \mathbf{p}_m\|_2$ is the location distance between $\mathcal{X}_{i_1 i_2 \ldots i_N}$ (corresponding to x_l) and $\mathcal{X}_{i'_1 i'_2 \ldots i'_N}$ (corresponding to x_m), which is defined as

$$\|\mathbf{p}_l - \mathbf{p}_m\|_2 = \sqrt{(i_1 - i'_1)^2 + (i_2 - i'_2)^2 + \cdots + (i_N - i'_N)^2}. \tag{3}$$

Then, d_{TD} can be rewritten as

$$d_{TD} = \sqrt{\frac{1}{2\pi\sigma_1^2}\sum_{l,m=1}^{I_1\times I_2\times\cdots\times I_N}\exp\left\{-\frac{\|\mathbf{p}_l - \mathbf{p}_m\|_2^2}{2\sigma_1^2}\right\}(x_l - y_l)(x_m - y_m)}. \tag{4}$$

Actually, the Euclidean distance can be viewed as a special case of the proposed TD. If we let the metric matrix to be the identity matrix, i.e., $\mathbf{G} = \mathbf{I}$, which means that we only consider the distance between corresponding coordinates of two high-order data in tensor space, then TD is reduced to Euclidean distance.

Since $\mathbf{G}$ is a real symmetric positive-definite matrix, we can easily decompose it as follows:

$$\mathbf{G} = \mathbf{G}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}} \tag{5}$$

where $\mathbf{G}^{1/2}$ is also a real symmetric matrix defined as

$$\mathbf{G}^{\frac{1}{2}} = \mathbf{U}_G \Lambda_G^{\frac{1}{2}} \mathbf{U}_G^T. \tag{6}$$

Here, Λ_G is a diagonal matrix whose elements are the eigenvalues of $\mathbf{G}$, and $\mathbf{U}_G$ is an orthogonal matrix whose column vectors are eigenvectors of $\mathbf{G}$. Applying the transformation $\mathbf{G}^{1/2}$ to the vector form representations $\mathbf{x}$ and $\mathbf{y}$, i.e., $\mathbf{x}' = \mathbf{G}^{1/2}\mathbf{x}$, $\mathbf{y}' = \mathbf{G}^{1/2}\mathbf{y}$, the proposed TD between $\mathbf{x}$ and $\mathbf{y}$ is then reduced to the traditional Euclidean distance between $\mathbf{x}'$ and $\mathbf{y}'$

$$\begin{aligned}\sqrt{(\mathbf{x}-\mathbf{y})^T\mathbf{G}(\mathbf{x}-\mathbf{y})} &= \sqrt{(\mathbf{x}-\mathbf{y})^T\mathbf{G}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}}(\mathbf{x}-\mathbf{y})} \\ &= \sqrt{(\mathbf{x}'-\mathbf{y}')^T(\mathbf{x}'-\mathbf{y}')}.\end{aligned} \tag{7}$$

So it is easy to embed TD to general learning procedures, we simply need to perform the transformation $\mathbf{G}^{1/2}$ on original data and then use transformed data in the following procedures.

III. Tensor Distance Based Multilinear Locality-Preserved Maximum Information Embedding

Based on the new distance metric TD, we present a novel DR method TD-MLPMIE according to a vector-based method LPMIP [19].

Given n data points $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ in the tensor space $\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, TD-MLPMIE aims to find N transformation matrices $\mathbf{V}_k \in \mathrm{R}^{I_k\times I'_k}$ ($I'_k \ll I_k$, $k = 1, \ldots, N$) such that n low-dimensional data points $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_n$ in the space $\mathbb{R}^{I'_1\times I'_2\times\cdots\times I'_N}$ can be obtained by $\mathcal{Y}_j = \mathcal{X}_j \times_1 \mathbf{V}_1^T \times_2 \cdots \times_N \mathbf{V}_N^T$ ($j = 1, \ldots, n$), where $\mathcal{X}\times_k\mathbf{V}_k^T$ denotes the k-mode product of $\mathcal{X}$ by $\mathbf{V}_k^T$. Refer to [21]–[23] for details of multilinear algebra.

To preserve both global information and local geometry in the low-dimensional subspace in terms of TD, the objective function of TD-MLPMIE is given as follows:

$$\begin{aligned}\max\ & J(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \alpha J_b(\mathbf{V}_1, \ldots, \mathbf{V}_N) \\ & - (1-\alpha)J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N) \\ & \text{s.t.}\mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_{I'_k}(k = 1, 2, \ldots, N)\end{aligned} \tag{8}$$

where $J_b(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \sum_{i=1}^n \sum_{j\notin O(i;\varepsilon)} \|\mathcal{Y}_i - \mathcal{Y}_j\|_F^2 \mathbf{W}(i,j)$ and $J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \sum_{i=1}^n \sum_{j\in O(i;\varepsilon)} \|\mathcal{Y}_i - \mathcal{Y}_j\|_F^2 \mathbf{W}(i,j)$. Let $O(\mathcal{X}_i;\varepsilon)$ denote the set of neighborhood points of $\mathcal{X}_i$ with radius ε. Then, $O(i;\varepsilon)$ denotes the indexes (subscripts) of the data points falling into the ε-neighborhood of $\mathcal{X}_i$. $\alpha \in [0, 1]$ is a trade-off parameter to balance *within locality* and *between locality* of the dataset, and $\mathbf{W}(i, j)$ is the weight on the edge that connects data points $\mathcal{X}_i$ and $\mathcal{X}_j$, which is defined using the proposed TD

$$\mathbf{W}(i,j) = \mathrm{e}^{\frac{-d_{TD}^2(\mathcal{X}_i,\mathcal{X}_j)}{\sigma_2}} \tag{9}$$

where $\sigma_2 > 0$ is a regularization parameter.

Maximizing $J(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ in (8) is equivalent to minimizing $J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ and maximizing $J_b(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ simultaneously. $J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ reveals the *within locality* of the dataset, and $J_b(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ reveals the *between locality* of the dataset. By maximizing the difference between the *between locality* and *within locality*, the embeddings corresponding to the same manifold are close to each other while the embeddings corresponding to different manifolds are far away from each other.

We define an adjacency matrix $\mathbf{A}$

$$\mathbf{A}(i,j) = \begin{cases}\mathbf{W}(i,j), & \mathcal{X}_j \in O(\mathcal{X}_i;\varepsilon) \\ 0, & \text{otherwise.}\end{cases} \tag{10}$$

The objective function of TD-MLPMIE can be rewritten as

$$\begin{aligned}&\max J(\mathbf{V}_1, \ldots, \mathbf{V}_N) \\ &\quad = \sum_{i=1}^n\sum_{j=1}^n \left\|\mathcal{Y}_i - \mathcal{Y}_j\right\|_F^2 (\alpha\mathbf{W}(i,j) - \mathbf{A}(i,j)) \\ &\text{s.t. } \mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}_{I'_k}\ (k = 1, 2, \ldots, N).\end{aligned} \tag{11}$$

There is no closed-form solution for this nonlinear-programming problem. So we employ an iterative strategy

TABLE I
TENSOR DISTANCE BASED MULTILINEAR LOCALITY-PRESERVED MAXIMUM INFORMATION EMBEDDING

Input: Training dataset $\mathcal{X}_1, \ldots, \mathcal{X}_n \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$; Iteration number T_{max}; Trade-off parameter α; Low dimensions $I'_1, I'_2, \ldots, I'_N$

Output: Transformation matrices $\mathbf{V}_k = \mathbf{V}_k^t \in \mathbb{R}^{I_k \times I'_k} (k = 1, \ldots, N)$

1. Construct $\mathbf{W}$ according to (9);
2. Construct $\mathbf{A}$ according to (10);
3. Initialize $t = 0$; $\mathbf{V}_k^0 = \mathbf{I}_{I_k} (k = 1, \ldots, N)$;
4. **for** $t = 1, \ldots, T_{max}$ **do**
5. **for** $k = 1, \ldots, N$ **do**
6. $\mathcal{X}_i^k = \mathcal{X}_i \times_1 (\mathbf{V}_1^t)^T \ldots \times_{k-1} (\mathbf{V}_{k-1}^t)^T \times_{k+1} (\mathbf{V}_{k+1}^{t-1})^T \cdots \times_N (\mathbf{V}_N^{t-1})^T$
7. $\mathbf{X}_i^k \Leftarrow_k \mathcal{X}_i^k$
8. $\mathbf{H}^k = \sum_{i=1}^n \sum_{j=1}^n (\mathbf{X}_i^k - \mathbf{X}_j^k)(\mathbf{X}_i^k - \mathbf{X}_j^k)^T (\alpha \mathbf{W}(i, j) - \mathbf{A}(i, j))$
9. $\mathbf{H}^k \mathbf{V}_k^t = \mathbf{V}_k^t \Lambda_k (\mathbf{V}_k^t \in \mathbb{R}^{I_k \times I'_k})$
10. **end for**
11. **if** $\| \mathbf{V}_k^t - \mathbf{V}_k^{t-1} \|_F < \varepsilon_1$ and $|(tr(\mathbf{V}_k^t)^T \mathbf{V}_k^{t-1})/I'_k - 1| < \varepsilon_2$ (k = 1, ..., N)
12. **break;**
13. **end if**
14. **end for**

[1], [2], [3], [22] to find a local optimal solution. Actually, if we assume that $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{k-1}, \mathbf{V}_{k+1}, \ldots, \mathbf{V}_N$ are fixed, then the optimal transformation matrix $\mathbf{V}_k$ maximizing $J(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ is composed of the first I'_k eigenvectors of the following matrix:

$$\mathbf{H}^k = \sum_{i=1}^n \sum_{j=1}^n \left(\mathbf{X}_i^k - \mathbf{X}_j^k\right) \left(\mathbf{X}_i^k - \mathbf{X}_j^k\right)^T (\alpha \mathbf{W}(i, j) - \mathbf{A}(i, j)) \tag{12}$$

corresponding to the first I'_k largest eigenvalues. $\mathbf{X}_i^k$ is the k-mode unfolding of the tensor $\mathcal{X}_i^k$, i.e., $\mathbf{X}_i^k \Leftarrow_k \mathcal{X}_i^k$, and $\mathcal{X}_i^k = \mathcal{X}_i \times_1 \mathbf{V}_1^T \times_2 \cdots \times_{k-1} \mathbf{V}_{k-1}^T \times_{k+1} \mathbf{V}_{k+1}^T \times_{k+2} \cdots \times_N \mathbf{V}_N^T$.

The above observation can be proved by using the properties of multilinear algebra and matrix trace [2], [3]. Based on this observation, the iterative strategy can then be presented. First we fix $\mathbf{V}_2, \ldots, \mathbf{V}_N$, and obtain optimal $\mathbf{V}_1$ by finding the leading eigenvectors of matrix $\mathbf{H}^1$. Then we fix $\mathbf{V}_1, \mathbf{V}_3, \ldots, \mathbf{V}_N$, and obtain the optimal $\mathbf{V}_2$. The rest can be deduced by analogy. Finally, we fix $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{N-1}$, and obtain the optimal $\mathbf{V}_N$. We repeat above steps until algorithm converges. The convergence proof of proposed algorithm is included in the Appendix. Table I describes the procedure of TD-MLPMIE.

Table II lists the training and test time costs of TD-MLPMIE and LPMIP. Here r is the rank of matrix $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, where $\mathbf{x}_i$ is the vectorized representation of tensor $\mathcal{X}_i$. The training time cost of TD-MLPMIE is lower than that of original LPMIP, but it is higher than that of LPMIP/QR, which is a QR factorization based LPMIP [19]. The test time cost of TD-MLPMIE is much less than that of both LPMIP and LPMIP/QR. Here we assume that sample tensors and embedded tensors are of uniform sizes in each order, respectively, i.e., $I_1 = \cdots = I_N = I$ and $I'_1 = \cdots = I'_N = I'$.

IV. THEORETICAL ANALYSIS

In this section, we show the generalization of TD-MLPMIE by analyzing its relation with LPMIP [19] and some representative multilinear DR techniques.

A. Relation with LPMIP

When $\mathbf{G} = \mathbf{I}$ and $N = 1$, TD-MLPMIE is reduced to LPMIP [19].

B. Relation with MPCA

Unlike TD-MLPMIE, MPCA [2] only preserves the global information of the dataset. The objective function of MPCA is

$$\begin{aligned} \max P(\mathbf{V}_1, \ldots, \mathbf{V}_N) &= \sum_{i=1}^n \left\| \mathcal{Y}_i - \bar{\mathcal{Y}} \right\|_F^2 \\ \text{s.t.} \mathbf{V}_k^T \mathbf{V}_k &= \mathbf{I}_{I'_k} \ (k = 1, 2, \ldots, N) \end{aligned} \tag{13}$$

where $\overline{\mathcal{Y}} = \left(\sum_{i=1}^n \mathcal{Y}_i\right)/n$. When $\mathbf{G} = \mathbf{I}$, $O(i; \varepsilon) = \varnothing$, and $\mathbf{W}(i, j) = 1$ for all i and j, TD-MLPMIE is reduced to MPCA. If we assign $N = 1$, PCA [24] can be deduced from TD-MLPMIE. If we assign $N = 2$ and use the ride-side transformation matrix, TD-MLPMIE is reduced to 2-D PCA [5].

C. Relation with MLDA

The objective function of MLDA can be expressed as

$$\max \ Q(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \frac{Q(b)}{Q(w)} = \frac{\sum_{m=1}^c n_m \left\| \overline{\mathcal{Y}}_m - \overline{\mathcal{Y}} \right\|_F^2}{\sum_{i=1}^n \left\| \mathcal{Y}_i - \overline{\mathcal{Y}}_{m_i} \right\|_F^2} \tag{14}$$

where $\overline{\mathcal{Y}}_m$ is the average tensor of embedded samples belonging to class m, and n_m is the sample number of class m. Supposing that $\mathbf{G} = \mathbf{I}, \mathbf{W}(i, j) = 1$ for all i, j, and $\alpha = n_0/n$, where n_0 is the number of samples in each class (we assume that c classes have same number of samples, i.e., $n_0 = n/c$). Moreover, we assign $O(\mathcal{X}_i; \varepsilon) = \{\mathcal{X} | \mathcal{X}$ and $\mathcal{X}_i$ belong to the same class$\}$, then TD-MLPMIE is reduced to MLDA [3] after introducing the constraint $Q(w) = 1$. If we assign $N = 1$ or $N = 2$, TD-MLPMIE is reduced to original LDA [25] or 2-D LDA [8], respectively.

D. Relation with TLPP

Unlike TD-MLPMIE, TLPP only considers *within locality*. The objective function of TLPP can be expressed as

$$\begin{aligned} \min R(\mathbf{V}_1, \ldots, \mathbf{V}_N) &= \sum_{i=1}^n \sum_{j=1}^n \left\| \mathcal{Y}_i - \mathcal{Y}_j \right\|_F^2 \mathbf{A}(i, j) \\ \text{s.t.} tr(\mathbf{V}_k^T \mathbf{V}_k) &= 1 \ (k = 1, 2, \ldots, N) \end{aligned} \tag{15}$$

when $\mathbf{G} = \mathbf{I}$ and $\alpha = 0$, maximizing $J(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ is reduced to maximizing $-J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N)$, i.e., minimizing $J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N)$. This is also the objective function of TLPP since $R(\mathbf{V}_1, \ldots, \mathbf{V}_N) = J_w(\mathbf{V}_1, \ldots, \mathbf{V}_N)$. After introducing

TABLE II
TRAINING AND TEST TIME COSTS OF TD-MLPMIE, LPMIP, AND LPMIP/QR

Algorithms	TD-MLPMIE	Original LPMIP	LPMIP/QR
Training time costs	$O(n^2 I^{2N} + n^2 N I^{N+1} + N I^3)$	$O(I^{3N})$	$O(r^2 n + r^3)$
Test time costs	$O(n(I')^{2N} + \sum_{i=1}^{N} (I')^i I^{N+1-i})$	$O((I')^N I^N)$	

some constraints, TD-MLPMIE is exactly reduced to TLPP. If we further assign the tensor order $N = 1$ or $N = 2$, TD-MLPMIE is reduced to LPP [26], [27], or 2-D LPP [12], respectively.

E. Relation with Tensor Local Discriminant Embedding (TLDE)

The TLDE [13] can be viewed as the supervised version of TLPP. The objective function of TLDE can be expressed as

$$\min S(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathcal{Y}_i - \mathcal{Y}_j\|_F^2 \mathbf{B}(i, j)$$
$$\text{s.t.} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathcal{Y}_i - \mathcal{Y}_j\|_F^2 \overline{\mathbf{B}}(i, j) = 1 \tag{16}$$

where $\mathbf{B} + \overline{\mathbf{B}} = \mathbf{A}$, and $\mathbf{B}(i, j) = \mathbf{A}(i, j)$ if $\mathcal{X}_i$ and $\mathcal{X}_j$ belong to the same class; $\mathbf{B}(i, j) = 0$ otherwise. Similar to TLPP, TLDE only takes the *within locality* into account. When $\mathbf{G} = \mathbf{I}$, $\alpha = 0$, and $O(\mathcal{X}_i; \varepsilon) = \{\mathcal{X} | \mathcal{X}$ is one of $\mathcal{X}_i$'s nearest neighbors and $\mathcal{X}, \mathcal{X}_i$ belong to the same class$\}$, the objective function of TD-MLPMIE is consistent with that of TLDE after introducing some constraints.

F. Relation with Tensor-Neighborhood-Preserving Embedding (TNPE)

The TNPE [13] is the multilinear extension of locally linear embedding [28]. It assumes that each data point can be represented by the linear combination of its nearest neighbors, and aims to recover this linear relationship in the low-dimensional tensor space. The objective function of TNPE can be expressed as

$$\min T(\mathbf{V}_1, \ldots, \mathbf{V}_N) = \sum_{i=1}^{n} \left\| \mathcal{Y}_i - \sum_{j \in O(i;\varepsilon)} \mathbf{M}(i, j) \mathcal{Y}_j \right\|_F^2$$
$$\text{s.t. } tr(\mathbf{V}_k^T \mathbf{V}_k) = 1 \ (k = 1, 2, ..., N) \tag{17}$$

where $\mathbf{M}(i, j)$ is the weight on the edge that connects data points $\mathcal{X}_i$ and $\mathcal{X}_j$, which is calculated by solving the following optimization problem:

$$\min \left\| \mathcal{X}_i - \sum_{j \in O(i;\varepsilon)} \mathbf{M}(i, j) \mathcal{X}_j \right\|_F^2$$
$$\text{s.t.} \sum_{j \in O(i;\varepsilon)} \mathbf{M}(i, j) = 1. \tag{18}$$

According to [15], if we set $\mathbf{W} = \mathbf{M} + \mathbf{M}^T - \mathbf{M}^T \mathbf{M}$, and suppose that $\mathbf{G} = \mathbf{I}$ and $\alpha = 0$, then TD-MLPMIE is reduced to TNPE after introducing the scaling constraints.

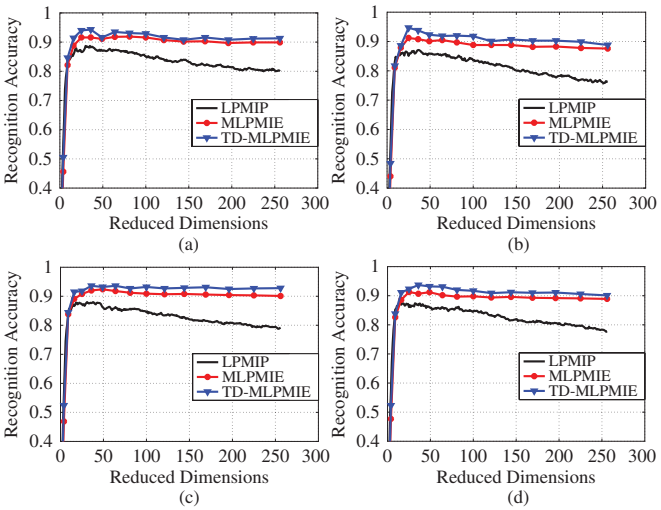

Fig. 1. Recognition accuracy of LPMIP, MLPMIE, and TD-MLPMIE on the USPS database, with various values of k and fixed $\alpha = 0.01$. (a) $k = 1$. (b) $k = 2$. (c) $k = 3$. (d) $k = 4$.

V. EXPERIMENTS

In this section, we evaluate proposed methods using pattern classification tasks on the following four databases: The United States Postal Service (USPS) digit database [29], [30], Carnegie Mellon University (CMU) Pose, Illumination, and Expression (PIE) face database [31], [32], Olivetti and Oracle Research Laboratory (ORL) face database [30], and Honda/University of California, San Diego (UCSD) video database [33]. Images in the first three databases are naturally second-order tensors, and videos in the last database are third-order tensors.

The recognition process is composed of three steps. First, the subspace is calculated from the training dataset. Second, for three image databases, test images are embedded into d-dimensional subspace (vector-based methods) or $(d \times d)$-dimensional subspace (tensor-based methods), for the last video database, test videos are embedded into $(d_1 \times d_2 \times d_3)$-dimensional subspace. Finally, the k nearest neighbor algorithm is applied in low-dimensional subspace for classification. In all the experiments, the range of pixel values of images and videos is normalized to [0, 1]. To guarantee the fair comparison, the regularization parameters used in different algorithms, such as σ in LPMIP, σ_1 and σ_2 in TD-MLPMIE, and t in TLPP, LPP, TLDE, and LDE, are set to be one in all experiments. We repeat each experiment ten times on randomly selected training and test datasets and report the average recognition accuracy.

TABLE III
COMPARISON OF RECOGNITION ACCURACY (%) AS WELL AS CORRESPONDING OPTIMAL REDUCED DIMENSIONS ON USPS, PIE, AND ORL DATABASES

		TD-MLPMIE	MLPMIE	TD-LPMIP	LPMIP	MLDA	LDA	MPCA	PCA	TLPP	LPP	TLDE	LDE
USPS	Recog.	**93.8**	**92.7**	90	88.2	91.8	89.1	87.4	82.9	91	85.2	91.4	87.3
	Dims	6^2	7^2	23	21	6^2	20	12^2	29	13^2	38	$\mathbf{5^2}$	15
PIE	Recog.	**90.7**	**89.2**	82.8	82.1	87.5	81.8	81.6	74.1	83.2	80.5	87	81.4
	Dims	9^2	10^2	42	45	$\mathbf{7^2}$	49	23^2	61	12^2	63	10^2	51
ORL	Recog.	**98.4**	**96.5**	95.5	93.2	96.3	95.8	90.1	92.6	92.8	90	96.1	92.8
	Dims	$\mathbf{8^2}$	$\mathbf{8^2}$	39	50	$\mathbf{8^2}$	52	26^2	79	13^2	161	$\mathbf{8^2}$	45

A. USPS Digit Database

The USPS database of handwritten digital characters contains 11 000 normalized grayscale images of size 16 × 16, with 1100 images in each class [29], [30].

We conduct two experiments in this database. In the first experiment, we split the whole dataset into 11 subgroups, and each subgroup includes 1000 images (100 images for each class). Then we randomly select two subgroups, one for training and another one for test. In this experiment, we compare the classification accuracy of LPMIP, MLPMIE, and TD-MLPMIE with the fixed trade-off parameter $\alpha = 0.01$ and different neighborhood sizes k. For MLPMIE and TD-MLPMIE, we only show their performance in $(d \times d)$-dimensional subspaces, i.e., 1, 4, 9, etc. From Fig. 1, we can see that both MLPMIE and TD-MLPMIE achieve better results than LPMIP for different sizes of subspaces under various values of k. Moreover, TD-MLPMIE outperforms MLPMIE slightly.

In the second experiment, we compare MLPMIE and TD-MLPMIE with other nine representative DR algorithms: LPMIP, MLDA, LDA, MPCA, PCA, TLPP, LPP, TLDE, and local discriminant embedding (LDE) [34]. We fix $k = 4$ and $\alpha = 0.01$. For each digit, 100 images are randomly selected for training and the remaining 1000 images are used for test. Table III lists the best recognition results and the corresponding optimal reduced dimensions of all algorithms. Obviously, TD-MLPMIE performs better than the other algorithms. To further demonstrate that the improvements in recognition accuracy are indeed due to the combination of the TD and multilinear embedding, rather than TD or MLPMIE alone, we also apply TD on LPMIP, and we called this algorithm TD-LPMIP. From Table III, we can see that TD-MLPMIE performs much better than both TD-LPMIP and MLPMIE, which indicates that the combination of TD metric and multilinear strategy effectively builds a unified tensor based DR framework.

B. CMU PIE Face Database

The CMU PIE database contains 41 368 images of 68 people with different poses, illumination conditions, and facial expressions [31]. Currently, 11 560 images in PIE are downloadable [32]. We evaluate the proposed algorithms on the subset of PIE, which is generated by randomly selecting 1700 images of 10 persons (170 for each). Images are grayscale and normalized to a resolution of 32 × 32 pixels.

We conduct two experiments on this database. First, we randomly choose 85 training images and 85 test images for each individual. We fix $k = 4$ and calculate the recognition accuracy under different values of α. As shown in Fig. 2, for various values of α, MLPMIE performs better than LPMIP because of the tensor embedding, and TD-MLPMIE achieves further improvement by integrating the proposed TD metric.

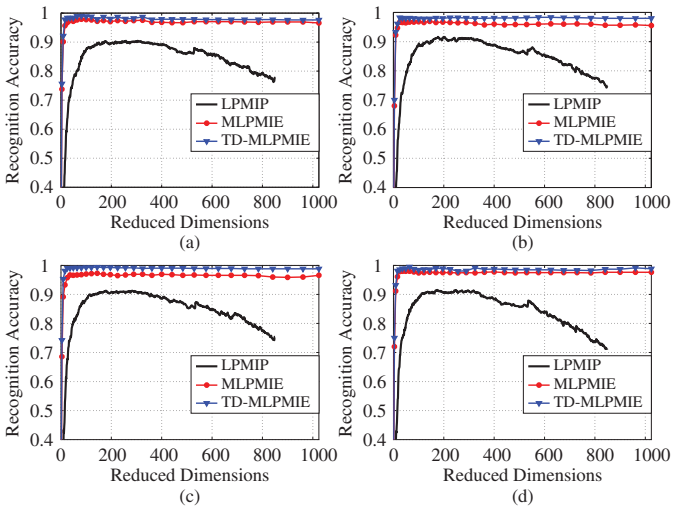

Fig. 2. Recognition accuracy of LPMIP, MLPMIE, and TD-MLPMIE on CMU PIE database, with various values of α and fixed $k = 4$. (a) $\alpha = 0.5$. (b) $\alpha = 0.1$. (c) $\alpha = 0.01$. (d) $\alpha = 0.001$.

In the second experiment, we provide comprehensive comparison between the proposed algorithms and the ten aforementioned DR algorithms. We fix $k = 4$ and $\alpha = 0.01$. For each class, ten images are randomly selected for training, and the remaining 160 images are used for test. Table III shows that the proposed algorithms achieve better recognition accuracy.

C. ORL Face Database

The ORL database contains 400 images of 40 different individuals (10 for each) [30]. All images are grayscale and normalized to a resolution of 64×64 pixels in our experiments. In the first experiment, we compare LPMIP, MLPMIE, and TD-MLPMIE under different sizes of training datasets. For each individual, p images are randomly selected for training and the rest are used for test. We fix $k = 4$ and $\alpha = 0.01$. Fig. 3 shows the recognition accuracy. By employing tensor representation, MLPMIE and TD-MLPMIE obtain higher recognition accuracy than LPMIP under small training datasets. Furthermore, the performance of TD-MLPMIE is better than that

TABLE IV
COMPARISON OF RECOGNITION ACCURACY (%) AS WELL AS CORRESPONDING OPTIMAL REDUCED DIMENSIONS ON HONDA/UCSD VIDEO DATABASE

	TD-MLPMIE	MLPMIE	MLDA	MPCA	TLPP	TLDE
Recog.	**95.4**	**93.9**	92.6	89.2	91.5	93.5
Dims	$5 \times 3 \times 1$	$7 \times 2 \times 1$	$3 \times 3 \times 2$	$6 \times 6 \times 2$	$10 \times 5 \times 1$	$4 \times 10 \times 1$

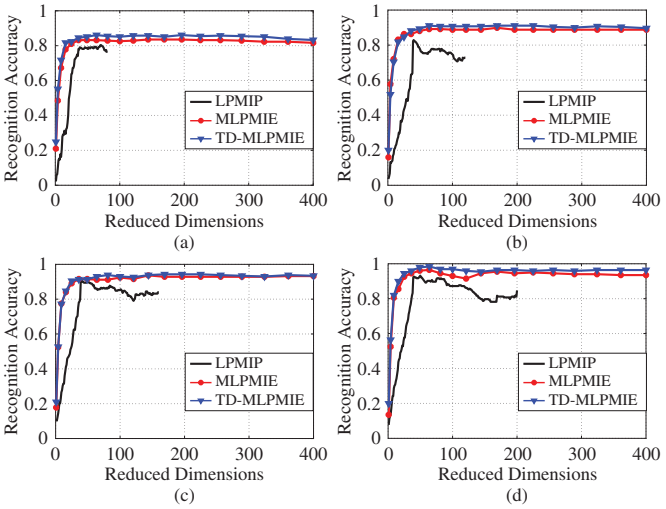

Fig. 3. Recognition accuracy of LPMIP, MLPMIE, and TD-MLPMIE on ORL database, with different training data numbers p. (a) $p = 2$. (b) $p = 3$. (c) $p = 4$. (d) $p = 5$.

of MLPMIE since the TD metric provides a more faithful measure of distances between high-order data points.

In the second experiment, we compare the proposed algorithms with the ten aforementioned DR algorithms. For each individual, we fix $p = 54$, $k = 4$, and $\alpha = 0.01$. As shown in Table III, TD-MLPMIE achieves the highest recognition accuracy. Moreover, TD-MLPMIE and MLPMIE, as well as MLDA and TLDE, significantly reduce the dimensionality of original data.

D. Honda/UCSD Video Database

In this section, we use the first dataset of Honda/UCSD video database [33] to test the performance of the proposed algorithms. This dataset contains 75 videos from 20 human subjects. Each video sequence is recorded in an indoor environment at 15 frames/s, and each lasting for at least 15 s. The resolution of each video sequence is 640×480. In our experiment, the original videos are downsampled into 64×48 pixels. In order to collect more training and test data, we cut each original video to several shorter ones of uniform length of 3 s, i.e., 45 frames. Therefore, the input data are third-order tensors of size $64 \times 48 \times 45$. Note that even though the data are downsampled, the size of $\mathbf{G}$ is still very large. To avoid storing this large matrix in the memory directly, we calculate the TD between two high-order data points using the following formula:

$$d_{TD} = \sqrt{\sum_{l,m=1}^{I_1 \times I_2 \times \cdots \times I_N} g_{lm}(x_l - y_l)(x_m - y_m)}. \quad (19)$$

The variables in (19) are defined similarly as in (1). By summating g_{lm} one by one, we can obtain d_{TD} without storing the whole matrix $\mathbf{G}$.

We compare MLPMIE and TD-MLPMIE with the four tensor-based algorithms MLDA, MPCA, TLPP, and TLDE. For each individual, we randomly select ten videos, five for training and five for test. We fix $k = 4$ and $\alpha = 0.01$. The recognition accuracy and corresponding optimal reduced dimensions of these algorithms are reported in Table IV. The proposed algorithms achieve better results in the embedded subspace.

E. Discussion

As shown on various datasets, the TD, which considers spatial relationship of image data and spatial-temporal relationship of video data, is really helpful to improve the recognition accuracy. However, this performance improvement is not distinguishable. Actually, the most significant advantage of TD-based algorithms is that they are robust and perform better than Euclidean-distance-based algorithms on most cases. Regarding how much the performance can be improved further, it actually depends on the individual dataset and learning task.

VI. CONCLUSION AND FUTURE WORK

This brief proposed a unified framework TD-MLPMIE for tensor-based DR. By integrating the advantages of the TD metric, high-order data representation, and locality-preserved maximum information projection, TD-MLPMIE keeps the intrinsic structure of high-order data as well as the local and global relationships between different data points in the whole learning procedure. Theoretical analysis and empirical evaluation validate the effectiveness of the proposed techniques.

In future work, we will work on reducing the time cost of TD-based DR algorithms. The computation of TD increases rapidly with respect to the size of metric matrix $\mathbf{G}$. How to reduce the computational cost is of great importance in making DR algorithms practical for more real-world applications.

APPENDIX

A. Convergence Proof of the Proposed Algorithm

Proof: To prove the convergence of proposed algorithm, we need to show that $J(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ is nondecreasing and has an upper bound. Actually, in each iteration of TD-MLPMIE, $J(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ is nondecreasing

$$\begin{aligned} J(\mathbf{V}_1^t, \mathbf{V}_2^t, ..., \mathbf{V}_N^t) &\le J(\mathbf{V}_1^{t+1}, \mathbf{V}_2^t, ..., \mathbf{V}_N^t) \\ &\le J(\mathbf{V}_1^{t+1}, \mathbf{V}_2^{t+1}, ..., \mathbf{V}_N^t) \\ &\le \cdots \le J(\mathbf{V}_1^{t+1}, \mathbf{V}_2^{t+1}, ..., \mathbf{V}_N^{t+1}). \end{aligned} \quad (20)$$

This is because that each update of transformation matrix $\mathbf{V}_k$ maximizes current objective function $J(\mathbf{V}_k)$ while other matrices $\mathbf{V}_i$ ($i = 1, 2, \ldots, k\text{-}1, k\text{+}1, \ldots, N$) are fixed.

On the other hand, for $\forall\, i$ and j, $\|\mathcal{Y}_i - \mathcal{Y}_j\|_F^2 \geq 0$. Furthermore, $\max[\alpha\mathbf{W}(i, j) - \mathbf{A}(i, j)] \geq 0$. Therefore, we have the following:

$$\begin{aligned} J(\mathbf{V}_1, ..., \mathbf{V}_N) &\leq \left(\sum_{i,j} \|\mathcal{Y}_i - \mathcal{Y}_j\|_F^2 \right) \\ &\quad \times \max[\alpha\mathbf{W}(i, j) - \mathbf{A}(i, j)] \\ &\leq \left(\sum_{i,j} \|\mathcal{X}_i - \mathcal{X}_j\|_F^2 \right) \\ &\quad \times \max[\alpha\mathbf{W}(i, j) - \mathbf{A}(i, j)] \end{aligned} \tag{21}$$

where the second inequality results from the fact that the distance between embedded data points is not larger than that between corresponding original data points.

Therefore, $J(\mathbf{V}_1, \ldots, \mathbf{V}_N)$ is nondecreasing and has an upper bound, i.e., TD-MLPMIE will finally converge.

REFERENCES

[1] D. Xu, S. Yan, L. Zhang, H.-J. Zhang, Z. Liu, and H.-Y. Shum, "Concurrent subspaces analysis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 203–208.

[2] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[3] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.

[4] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2003, pp. 93–99.

[5] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "2-D PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[6] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning," *IEEE Trans. Neural Netw.*, vol. 20, no. 11, pp. 1820–1836, Nov. 2009.

[7] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 576–588, Mar. 2010.

[8] J. Ye, R. Janardan, and Q. Li, "2-D linear discriminant analysis," in *Advances in Neural Information Processing Systems 17*. Cambridge, MA: MIT Press, 2004, pp. 1569–1576.

[9] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 103–123, Jan. 2009.

[10] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[11] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear independent components analysis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 547–553.

[12] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Advances in Neural Information Processing Systems 18*. Cambridge, MA: MIT Press, 2005, pp. 1–8.

[13] G. Dai and D.-Y. Yeung, "Tensor embedding methods," in *Proc. 21st AAAI Conf. Artificial Intell.*, 2006, pp. 330–335.

[14] W. Zhang, Z. Lin, and X. Tang, "Tensor linear Laplacian discrimination for feature extraction," *Pattern Recognit.*, vol. 42, no. 9, pp. 1941–1948, 2009.

[15] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[16] Y. Fu and T. S. Huang, "Image classification using correlation tensor analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 226–234, Feb. 2008.

[17] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.

[18] Y. Liu, Y. Liu, and K. C. C. Chan, "Multilinear isometric embedding for visual pattern analysis," in *Proc. 12th IEEE Int. Conf. Comput. Vis., Workshop Subspace Methods*, 2009, pp. 212–218.

[19] H. Wang, S. Chen, Z. Hu, and W. Zheng, "Locality-preserved maximum information projection," *IEEE Trans. Neural Netw.*, vol. 19, no. 4, pp. 571–585, Apr. 2008.

[20] L. Wang, Y. Zhang, and J. Feng, "On the Euclidean distance of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1334–1339, Aug. 2005.

[21] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Mar.–May 2000.

[22] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, Mar.–May 2000.

[23] T. Kolda, "Orthogonal tensor decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 243–255, 2001.

[24] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Edu. Psychol.*, vol. 24, no. 7, pp. 498–520, Oct. 1933.

[25] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.

[26] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems 16*. Chicago, IL: Univ. Chicago Press, 2003, pp. 1–8.

[27] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

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

[29] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.

[30] *Data for MATLAB Hackers* [Online]. Available: http://www.cs.toronto.edu/roweis/data.html

[31] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[32] *Popular Face Data Sets in MATLAB Format* [Online]. Available: http://www.zjucadcg.cn/dengcai/Data/FaceData.html

[33] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Comput. Vis. Image Understanding*, vol. 99, no. 3, pp. 303–331, Sep. 2005.

[34] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 846–853.