

Research Article

Robust Tensor Preserving Projection for Multispectral Face Recognition

Shaoyuan Sun,¹ Haitao Zhao,² and Bo Jin³

¹ Automation Department, Donghua University, Shanghai 201620, China

² Automation Department, East China University of Science and Technology, Shanghai 200237, China

³ School of Aeronautics and Astronautics, Shanghai Jiaotong University, Shanghai 200240, China

Correspondence should be addressed to Haitao Zhao; ht.zhao.ecust@gmail.com

Received 14 April 2014; Accepted 21 July 2014; Published 12 August 2014

Academic Editor: Haipeng Peng

Copyright © 2014 Shaoyuan Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple imaging modalities based face recognition has become a hot research topic. A great number of multispectral face recognition algorithms/systems have been designed in the last decade. How to extract features of different spectrum has still been an important issue for face recognition. To address this problem, we propose a robust tensor preserving projection (RTPP) algorithm which represents a multispectral image as a third-order tensor. RTPP constructs sparse neighborhoods and then computes weights of the tensor. RTPP iteratively obtains one spectral space transformation matrix through preserving the sparse neighborhoods. Due to sparse representation, RTPP can not only keep the underlying spatial structure of multispectral images but also enhance robustness. The experiments on both Equinox and DHUFO face databases show that the performance of the proposed method is better than those of related algorithms.

1. Introduction

Multibiometrics can be considered the fusion of different sensor modalities in a single recognition system. The reason of using two or more sensor modalities is to improve the recognition accuracy. Multiple imaging modalities based face recognition has become a hot research topic [1–7]. Recent studies have shown that multispectral face recognition offers several advantages, such as invariance to illumination changes [8, 9]. Multispectral image also reveals anatomical information of a subject [1]. Socolinsky and Selinger [3, 10] developed different recognition algorithms on visible and thermal infrared face image databases and obtained good performances. Chen et al. [11] tested the effect of illumination, facial expression, and passage of time between the training and testing images. Wang et al. [12] showed that color space combination represents a viable approach for improving face recognition performance. The image-based fusion designed in the wavelet domain and the feature-based fusion developed in the eigenspace domain were shown in [13]. Heo et al. [14] proposed to fuse visual and thermal images for robust

face recognition. Multisensory biometric fusion algorithms were investigated for personal identification [15]. Pan et al. [4, 5] analyzed the facial tissue spectral measurements in the near-infrared spectral range ($0.7\ \mu\text{m}$ – $1.0\ \mu\text{m}$) for face recognition. Denes et al. [6] tested the spectral asymmetry with three visible bands ($0.6\ \mu\text{m}$, $0.7\ \mu\text{m}$, and $0.8\ \mu\text{m}$). Chang et al. [16] fused the multispectral images in the visible spectrum ($0.4\ \mu\text{m}$ – $0.72\ \mu\text{m}$) into a single image to enhance face recognition accuracies. Chou and Bajcsy [7] preprocessed the multispectral images (visible: $0.4\ \mu\text{m}$ – $0.72\ \mu\text{m}$ and near-infrared: $0.65\ \mu\text{m}$ – $1.1\ \mu\text{m}$) by principal component analysis (PCA) to perform face detection. Based on visible images, Wong and Zhao [17] adopted kernel PCA to remove eyeglasses of thermal face images.

The above algorithms are mainly developed to preserve the global structure information of the multispectral data. They do not clearly treat the manifold structure of the data. However, research results of manifold learning algorithms presented in the past decade demonstrate that the local geometric structure is more important than the global structure since the high-dimensional data often lies on

the low-dimensional manifold [18]. Due to the low-dimensional manifold structure of the face images, the manifold-learning-based linear dimension reduction algorithms [4–9, 12–14, 19–24] become popular.

Since these linear feature extraction algorithms cannot deal with high-order tensor data, some of these algorithms were further extended to be multilinear cases, and a lot of tensor-based manifold learning algorithms were proposed by using higher order tensor decomposition [25–27]. Within the past ten years, there has been great interest in high-order tensor feature extraction, and the tensor-based methods have been popular in computer vision and pattern recognition [28–31]. For example, Igarashi et al. proposed tensor subspace analysis (TSA) [32] for second-order learning. Dai and Yeung proposed tensor NPE (TNPE) [19]. Recently, orthogonal tensor neighborhood preserving embedding (OTNPE) was proposed for facial expression recognition. Some variations were also proposed for gait recognition, action recognition, and so forth. For more details, please see the latest survey of multilinear subspace learning [1–3].

Recent research demonstrates that the high-order tensor based manifold learning algorithms, such as tensor preserving embedding (tensor NPE) [20], can obtain better performance than the classical feature extraction algorithms on tensor data set. Unfortunately, the tensor data contain large quantities of information redundancy and thus not all the features/variables are important to feature extraction and classification [21–23]. It was shown that integrating sparse representation and manifold learning for feature extraction may obtain better performance [24]. It has been shown that the sparse representation methods can obtain better performance than their corresponding nonsparse methods in the real data. And these sparse methods can give an intuitionistic or semantic interpretation for the transformed features [25].

Till now, the field in high-order tensor data embedding with sparse manner has not been widely investigated and how to extend the manifold learning algorithms integrating sparseness and manifold structure for multispectral face recognition is unsolved. In this paper, motivated by tensor data embedding and sparse representation, we propose a novel method called robust tensor preserving projection (RTPP) for multispectral image feature extraction. The multispectral image is considered a third-order tensor. The aim of RTPP is to obtain transformation matrices through preserving the sparse information of the third-order tensors.

The rest of the paper is organized as follows. In Section 2, we give the related definitions to tensor. In Section 3, the introduction of tensor locality preserving projection is provided. In Section 4, a novel sparse tensor embedding method is presented. Experiments are carried out to evaluate the proposed tensor learning method in Section 5, and conclusions are given in Section 6.

2. Tensor Fundamentals

A tensor is a multidimensional array. It is the higher order generalization of scalar (zero-order tensor), vector (1st-order

tensor), and matrix (2nd-order tensor). In this paper, lowercase letters (i.e., a, b, c) denote scalars, bold lowercase letters (i.e., $\mathbf{a}, \mathbf{b}, \mathbf{c}$) denote vectors, uppercase letters (i.e., A, B, C) denote matrices, and bold uppercase letters (i.e., $\mathbf{A}, \mathbf{B}, \mathbf{C}$) denote the tensors. It is assumed that the training samples are represented as the n th-order tensor $\{\mathbf{A}_i \in R^{m_1 \times m_2 \times \dots \times m_n}, i = 1, 2, \dots, N\}$, where N denotes the total number of training samples.

Definition 1. The inner product of two tensors $\mathbf{A}, \mathbf{B} \in R^{m_1 \times m_2 \times \dots \times m_n}$ is defined as $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_n} \mathbf{A}_{i_1, i_2, \dots, i_n} \mathbf{B}_{i_1, i_2, \dots, i_n}$. The Frobenius norm of a tensor $\mathbf{A} \in R^{m_1 \times m_2 \times \dots \times m_n}$ is then defined as $\|\mathbf{A}\| = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}$. And the distance between two tensors $\mathbf{A}, \mathbf{B} \in R^{m_1 \times m_2 \times \dots \times m_n}$ is defined as $D(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|$.

Definition 2. The k -mode flattening of the n th-order tensor $\mathbf{A}_i \in R^{m_1 \times m_2 \times \dots \times m_n}$ ($i = 1, 2, \dots, N$) into matrix $A^{(k)} \in R^{m_k \times \prod_{i \neq k} m_i}$, that is, $A^{(k)} \leftarrow_k \mathbf{A}$, is defined as $A_{i_k, j}^{(k)} = \mathbf{A}_{i_1, i_2, \dots, i_n}$, $j = 1 + \sum_{q=1, q \neq k}^n (i_q - 1) \prod_{p=q+1, p \neq k}^n m_p$.

Definition 3. The k -mode product of a tensor $\mathbf{A} \in R^{m_1 \times m_2 \times \dots \times m_n}$ by a matrix $U \in R^{m'_k \times m_k}$, denoted by $\mathbf{B} = \mathbf{A} \times_k U$, is an $(m_1 \times m_2 \times \dots \times m_{k-1} \times m'_k \times m_{k+1} \times \dots \times m_n)$ -tensor of which the entries are given by $\mathbf{B}_{i_1, \dots, i_{k-1}, i, i_{k+1}, \dots, i_n} = \sum_{j=1}^{m_k} \mathbf{A}_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n} U_{i, j}$ ($j = 1, 2, \dots, m'_k$).

The aim of tensor learning algorithm is to obtain a set of projection matrices $\{U_i \in R^{d_i \times m_i}, d_i \leq m_i, i = 1, 2, \dots, n\}$ and map the original tensor into a new tensor:

$$\mathbf{B}_i = \mathbf{A}_i \times_1 U_1 \times_2 U_2 \cdots \times_n U_n. \quad (1)$$

3. Tensor Neighborhood Preserving Embedding

Let $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N$ be the multispectral face images in a high-order tensor form and $\mathbf{A}_i \in R^{m_1 \times m_2 \times \dots \times m_n}$ ($i = 1, 2, \dots, N$), N , the number of individuals. Assume that $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N$ are from an unknown manifold \mathcal{M} embedding in a tensor space $R^{d_1 \times d_2 \times \dots \times d_n}$. The aim of tensor NPE is to find optimal transformation matrices U_1, U_2, \dots, U_n such that the local topological structure of \mathcal{M} is preserved and the intrinsic geometric property is effectively captured. The optimal transformation matrices $U_j \in R^{d_j \times m_j}$ ($d_j \leq m_j, j = 1, 2, \dots, n$) project high-dimensional \mathbf{A}_i into low-dimensional representation \mathbf{B}_i , where $\mathbf{B}_i = \mathbf{A}_i \times_1 U_1 \times_2 U_2 \cdots \times_n U_n$.

We construct a neighborhood graph to represent the intrinsic geometric structure of \mathcal{M} and apply the heat kernel to define the affinity matrix $W = [w_{ij}]_{N \times N}$ as

$$w_{ij} = \begin{cases} \exp\left(\frac{-\|\mathbf{A}_i - \mathbf{A}_j\|^2}{t}\right), & \text{if } \mathbf{A}_j \in \mathcal{O}(K, \mathbf{A}_i), \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where $O(K, \mathbf{A}_i)$ denotes the set of K nearest neighbors of \mathbf{A}_i and t is a positive constant. The affinity matrix W is then normalized such that each row sums to one. In order to preserve the geometric structure explicitly, we define the following objective function based on the Frobenius norm of a tensor:

$$\begin{aligned} \arg \min J(U_1, U_2, \dots, U_n) &= \sum_i \left\| \mathbf{B}_i - \sum_j w_{ij} \mathbf{B}_j \right\|^2 \\ &= \sum_i \left\| \mathbf{A}_i \times_1 U_1 \cdots \times_n U_n - \sum_j w_{ij} \mathbf{A}_j \times_1 U_1 \cdots \times_n U_n \right\|^2. \end{aligned} \quad (3)$$

To eliminate an arbitrary scaling factor in the projection matrices, we impose the following constraint: $\sum_i \|\mathbf{B}_i\|^2 = \sum_i \|\mathbf{A}_i \times_1 U_1 \cdots \times_n U_n\|^2 = 1$. Then the optimization problem for tensor NPE can be expressed as

$$\begin{aligned} \arg \min \quad & J(U_1, U_2, \dots, U_n) \\ &= \sum_i \left\| \mathbf{A}_i \times_1 U_1 \cdots \times_n U_n - \sum_j w_{ij} \mathbf{A}_j \times_1 U_1 \cdots \times_n U_n \right\|^2, \\ \text{s.t.} \quad & \sum_i \|\mathbf{A}_i \times_1 U_1 \cdots \times_n U_n\|^2 = 1. \end{aligned} \quad (4)$$

Note that this optimization problem is a high-order non-linear programming problem with a high-order nonlinear constraint, making direct computation of the projection matrices infeasible. In general, this type of problems can be solved approximately by employing an iterative scheme which was proposed for low-rank approximation. The optimization problem in (4) can be solved by such an iterative scheme. Assuming that $U_1, U_2, \dots, U_{k-1}, U_{k+1}, \dots, U_n$ are known, let $\mathbf{B}_i^k = \mathbf{A}_i \times_1 U_1 \cdots \times_{k-1} U_{k-1} \times_{k+1} U_{k+1} \cdots \times_n U_n$. In addition, since $B_i^{(k)} \leftarrow_k \mathbf{B}_i^k$ and based on the properties of tensor and trace, we rewrite the optimization function and the constraint in (4) as follows:

$$\begin{aligned} \arg \min J_k(U_k) &= \sum_i \left\| \mathbf{B}_i^k \times_k U_k - \sum_j w_{ij} \mathbf{B}_j^k \times_k U_k \right\|^2 \\ &= \sum_i \left\| U_k B_i^{(k)} - \sum_j w_{ij} U_k B_j^{(k)} \right\|^2 \\ &= \sum_{i,j} \text{tr} \left\{ U_k \left((B_i^{(k)} - w_{ij} B_j^{(k)}) \right. \right. \\ &\quad \left. \left. \times (B_i^{(k)} - w_{ij} B_j^{(k)})^T \right) U_k^T \right\} \end{aligned}$$

$$= \text{tr} \left\{ U_k \left(\sum_{i,j} (B_i^{(k)} - w_{ij} B_j^{(k)}) \times (B_i^{(k)} - w_{ij} B_j^{(k)})^T \right) U_k^T \right\},$$

$$\begin{aligned} \sum_i \|\mathbf{B}_i^k \times_k U_k\|^2 &= \sum_i \|U_k B_i^{(k)}\|^2 \\ &= \text{tr} \left\{ U_k \left(\sum_i B_i^{(k)} B_i^{(k)T} \right) U_k^T \right\}. \end{aligned} \quad (5)$$

Thus, the optimization problem in (4) can be reformulated as

$$\begin{aligned} \arg \min \quad & J_k(U_k) \\ &= \text{tr} \left\{ U_k \left(\sum_{i,j} (B_i^{(k)} - w_{ij} B_j^{(k)}) \right. \right. \\ &\quad \left. \left. \times (B_i^{(k)} - w_{ij} B_j^{(k)})^T \right) U_k^T \right\}, \\ \text{s.t.} \quad & \text{tr} \left\{ U_k \left(\sum_i B_i^{(k)} B_i^{(k)T} \right) U_k^T \right\} = 1. \end{aligned} \quad (6)$$

The unknown transformation matrix U_k can be obtained by solving the eigenvectors corresponding to the k th smallest eigenvalues in the generalized eigenvalue equation

$$\begin{aligned} & \left(\sum_{i,j} (B_i^{(k)} - w_{ij} B_j^{(k)}) (B_i^{(k)} - w_{ij} B_j^{(k)})^T \right) \mathbf{u} \\ &= \lambda \left(\sum_i B_i^{(k)} B_i^{(k)T} \right) \mathbf{u}. \end{aligned} \quad (7)$$

The other transformation matrices can be obtained in a similar manner.

3.1. Tensor Locality Preserving Projection. Different from tensor NPE, the optimization problem for tensor LPP can be expressed as

$$\begin{aligned} \arg \min \quad & J(U_1, U_2, \dots, U_n) = \sum_{i,j} \|\mathbf{B}_i - \mathbf{B}_j\|^2 w_{ij} \\ &= \sum_{i,j} \|\mathbf{A}_i \times_1 U_1 \cdots \times_n U_n - \mathbf{A}_j \times_1 U_1 \cdots \times_n U_n\|^2 w_{ij}, \\ \text{s.t.} \quad & \sum_i \|\mathbf{A}_i \times_1 U_1 \cdots \times_n U_n\|^2 d_{ii} = 1. \end{aligned} \quad (8)$$

In general, the larger the value of $d_{ii} = \sum_j w_{ij}$ is, the more important the tensor \mathbf{B}_i is in the embedded tensor space for

representing the original tensor \mathbf{A}_i . It is easy to see that the objective function will give a high penalty if neighboring tensors \mathbf{A}_i and \mathbf{A}_j are mapped far apart. Thus if two tensors \mathbf{A}_i and \mathbf{A}_j are close to each other, then the corresponding tensors \mathbf{B}_i and \mathbf{B}_j in the embedded tensor space are also expected to be close to each other.

The optimization function of tensor LPP can be formulated as follows:

$$\begin{aligned}
\text{argmin } J_k(U_k) &= \sum_{i,j} \left\| \mathbf{B}_i^k \times_k U_k - \mathbf{B}_j^k \times_k U_k \right\|^2 w_{ij} \\
&= \sum_{i,j} \left\| U_k B_i^{(k)} - U_k B_j^{(k)} \right\|^2 w_{ij} \\
&= \sum_{i,j} \text{tr} \left\{ U_k \left((B_i^{(k)} - B_j^{(k)}) \right. \right. \\
&\quad \left. \left. \times (B_i^{(k)} - B_j^{(k)})^T w_{ij} \right) U_k^T \right\} \\
&= \text{tr} \left\{ U_k \left(\sum_{i,j} (B_i^{(k)} - B_j^{(k)}) \right. \right. \\
&\quad \left. \left. \times (B_i^{(k)} - B_j^{(k)})^T w_{ij} \right) U_k^T \right\}, \\
\text{s.t. } \text{tr} \left\{ U_k \left(\sum_i B_i^{(k)} B_i^{(k)T} d_{ii} \right) U_k^T \right\} &= 1.
\end{aligned} \tag{9}$$

Moreover the transformation matrix U_k can be computed by solving the eigenvectors corresponding to the k th smallest eigenvalues in the generalized eigenvalue equation

$$\begin{aligned}
&\left(\sum_{i,j} (B_i^{(k)} - B_j^{(k)}) (B_i^{(k)} - B_j^{(k)})^T w_{ij} \right) \mathbf{u} \\
&= \lambda \left(\sum_i B_i^{(k)} B_i^{(k)T} d_{ii} \right) \mathbf{u}.
\end{aligned} \tag{10}$$

4. Robust Tensor Preserving Projection

Sparse representation algorithms have been widely studied in signal processing, computer vision, and pattern recognition. Wright et al. [25] used sparse representation for robust face reconstruction and recognition, Qiao et al. [26] proposed sparse preserving projections, and Cheng et al. [27] used the L_1 graph for image clustering. As demonstrated in [26, 27], the graphs constructed by the L_1 norm have the advantages of greater robustness to noise and information redundancy. In the following, we fuse the sparse representation with tensor feature extraction.

4.1. Sparse Tensor Representation. In this part, we present the sparse representation for the tensor data $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N$. Let $Z = [z_{ij}]_{N \times N}$ be the optimal sparse representation coefficients. Sparse representation assumes that the training

sample \mathbf{A}_i ($i = 1, 2, \dots, N$) can be sparsely represented as a linear combination of the other data. Based on this assumption, the following sparse optimization problem was proposed in [25]:

$$\begin{aligned}
\min \quad & \|Z_{i,:}\|_0 \\
\text{s.t.} \quad & \left\| \mathbf{A}_i - \sum_{j,j \neq i} z_{ij} \mathbf{A}_j \right\|^2 \leq \varepsilon,
\end{aligned} \tag{11}$$

where $N \times N$ matrix $Z = [z_{ij}]_{N \times N}$ is the representation coefficient matrix satisfying $\text{diag}(Z) = 0$ and $Z_{i,:}$ denotes the i th row vector of Z . The parameter ε is a small constant set by users. However, the above optimization problem is NP hard. One can use the convex relaxation method to the NP-hard problem and solve the following optimization problem:

$$\begin{aligned}
\min \quad & \|Z_{i,:}\|_1 \\
\text{s.t.} \quad & \left\| \mathbf{A}_i - \sum_{j,j \neq i} z_{ij} \mathbf{A}_j \right\|^2 \leq \varepsilon.
\end{aligned} \tag{12}$$

Due to the sparseness of z_{ij} , then only a few $z_{ij} \neq 0$ ($i, j = 1, 2, \dots, N$). It means that, for tensor \mathbf{A}_i ($i = 1, 2, \dots, N$), not all the other tensors are used in the representation. Let z_{ij_k} ($i = 1, 2, \dots, N; k = 1, 2, \dots, K_i$) be the nonzero coefficients. Then the representation error is as follows:

$$d(\mathbf{A}_i) = \left\| \mathbf{A}_i - \sum_{j,j \neq i} z_{ij} \mathbf{A}_j \right\|^2 \quad (i = 1, 2, \dots, N). \tag{13}$$

In this paper, we also use sparse representation classifier (SRC) for RTPP. SRC classifies the test sample to the class with the least within-class reconstruction error. For more details of SRC, please refer to [25].

Define the affinity matrix $S = [s_{ij}]_{N \times N}$. Let $s_{ij} = 0$, if $z_{ij} = 0$. We can compute the s_{ij_k} ($i = 1, 2, \dots, N; k = 1, 2, \dots, K_i$) as follows:

$$\begin{aligned}
\min \quad & \left\| \mathbf{A}_i - \sum_{k=1}^{K_i} s_{ij_k} \mathbf{A}_{j_k} \right\|^2 \\
\text{s.t.} \quad & \sum_{k=1}^{K_i} s_{ij_k} = 1 \quad s_{ij_k} \geq 0 \quad (k = 1, 2, \dots, K_i).
\end{aligned} \tag{14}$$

Here we obtain the weights in a similar way as in LLE, except the constraints $s_{ij_k} \geq 0$ ($k = 1, 2, \dots, K_i$). The nonnegative constraints lead to a parts-based representation because they allow only additive, not subtractive, combinations. Previous studies have shown that there is psychological and physiological evidence for parts-based representation in human brain [11, 13, 17]. The sum-to-one constraint $\sum_{k=1}^{K_i} s_{ij_k} = 1$ is used to make the weights invariant to translation.

Discriminant information can be naturally preserved in the weights, even if no class information is available. In face recognition, one particularly simple but reasonable

assumption is that the samples from the same class lie on a linear subspace. In other words, the nonzero weights mostly correspond to the samples from the same class, which implies that the nonzero weights may help distinguish that class from the others. Therefore, the weights tend to include potential discriminant information.

In the design of the proposed RTPP, we use $S = [s_{ij}]_{N \times N}$ instead of the similarities used in tensor NPE. One advantage of the proposed technique is that the difficulty in selecting the size K of the local neighborhood can be avoided in tensor NPE. Moreover, the similarities can give intuitionistic or semantic interpretation of the represented tensor data. Another advantage is that sparse representation has the potential discriminative ability since most nonzero sparse representation coefficients are located on the samples in the same class as the represented sample.

4.2. Algorithm of Robust Tensor Preserving Projection. For convenience, in this section, we use the notations in Section 3 to derive RTPP. Assuming that $U_1, U_2, \dots, U_{k-1}, U_{k+1}, \dots, U_n$ are known, now we want to compute the projection matrix U_k . Using the similarities in (13), we have the optimization function as follows:

$$\begin{aligned} \text{argmin} \quad J_k(U_k) &= \sum_i \left\| U_k B_i^{(k)} - \sum_j s_{ij} U_k B_j^{(k)} \right\|^2 \\ &= \text{tr} \left\{ U_k \left(\sum_{i,j} (B_i^{(k)} - s_{ij} B_j^{(k)}) \right. \right. \\ &\quad \left. \left. \times (B_i^{(k)} - s_{ij} B_j^{(k)})^T \right) U_k^T \right\} \\ \text{s.t.} \quad \text{tr} \left\{ U_k \left(\sum_i B_i^{(k)} B_i^{(k)T} \right) U_k^T \right\} &= 1. \end{aligned} \quad (15)$$

Let $B^{(k)} = [B_1^{(k)}, B_2^{(k)}, \dots, B_N^{(k)}]$, ϵ_i be a N -dimensional unit vector with the i th element 1, 0 otherwise, and $S_{i,:}$ denotes the i th row vector of S . With simple formulation, we can get

$$\begin{aligned} &\sum_i \left\| U_k B_i^{(k)} - \sum_j s_{ij} U_k B_j^{(k)} \right\|^2 \\ &= \text{tr} \left\{ U_k \left(\sum_{i,j} (B_i^{(k)} - s_{ij} B_j^{(k)}) (B_i^{(k)} - s_{ij} B_j^{(k)})^T \right) U_k^T \right\} \\ &= \text{tr} \left\{ U_k \left(\sum_i (B_i^{(k)} - B^{(k)} S_{i,:}) (B_i^{(k)} - B^{(k)} S_{i,:})^T \right) U_k^T \right\} \\ &= \text{tr} \left\{ U_k \left(\sum_i (B^{(k)} \epsilon_i - B^{(k)} S_{i,:}^T) \right. \right. \\ &\quad \left. \left. \times (B^{(k)} \epsilon_i - B^{(k)} S_{i,:}^T)^T \right) U_k^T \right\} \end{aligned}$$

$$\begin{aligned} &= \text{tr} \left\{ U_k \left(\sum_i (B^{(k)} \epsilon_i - B^{(k)} S_{i,:}^T) \right. \right. \\ &\quad \left. \left. \times (B^{(k)} \epsilon_i - B^{(k)} S_{i,:}^T)^T \right) U_k^T \right\} \\ &= \text{tr} \left\{ U_k B^{(k)} \left(\sum_i (\epsilon_i - S_{i,:}^T) (\epsilon_i - S_{i,:}^T)^T \right) U_k^T \right\} \\ &= \text{tr} \left\{ U_k B^{(k)} (I - S)^T (I - S) B^{(k)T} U_k^T \right\}, \end{aligned} \quad (16)$$

where I is $N \times N$ identity matrix. We can also obtain

$$\text{tr} \left\{ U_k \left(\sum_i B_i^{(k)} B_i^{(k)T} \right) U_k^T \right\} = \text{tr} \left\{ U_k B^{(k)} B^{(k)T} U_k^T \right\}. \quad (17)$$

Then the optimization problem in (14) can be rewritten as

$$\begin{aligned} \text{argmin} \quad J_k(U_k) &= \text{tr} \left\{ U_k B^{(k)} (I - S)^T (I - S) B^{(k)T} U_k^T \right\} \\ \text{s.t.} \quad \text{tr} \left\{ U_k B^{(k)} B^{(k)T} U_k^T \right\} &= 1. \end{aligned} \quad (18)$$

Then the transformation matrix U_k can be obtained by solving the eigenvectors corresponding to the d_k smallest eigenvalues in the generalized eigenvalue equation

$$\left(B^{(k)} (I - S)^T (I - S) B^{(k)T} \right) \mathbf{u} = \lambda \left(B^{(k)} B^{(k)T} \right) \mathbf{u}. \quad (19)$$

The other transformation matrices can be obtained in a similar manner.

5. Experimental Results

In this section, experiments on Equinox data set and DHUFO data set are presented to evaluate RTPP in Algorithm 1 for recognition tasks. In the experiments, we compare the RTPP algorithm with the tensor NPE method. Besides tensor NPE, we also perform NPE directly on the serial combined data. The serial combined data z_{serial} is a super vector by combining $x_1, x_2, \dots, x_s \in R^m$ as $z_{\text{serial}} = [x_1^T, x_2^T, \dots, x_s^T]^T$. For visible and thermal infrared data, such as x_{visible} and x_{IR} , the serial combined data $z_{\text{serial}} = [x_{\text{visible}}^T, x_{\text{IR}}^T]^T$.

For the purpose of evaluating the performance of RTPP, we used face verification rate as the criteria. The FERET Verification Testing Protocol [28] recommends using the receiver operating characteristic (ROC) curves to depict the relations between the face verification rate (FVR) and the false accept rate (FAR). The ROC curves were plotted by using the Statistical Learning Toolbox according to the obtained score matrix. For tensor operations, we used the tensor toolbox developed by Bader and Kolda in MATLAB [29]. The sparse representations were obtained by Friedman et al. [30]. In the following experiments, we set $\eta = 0.1$ for both data sets.

```

Input:  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N$  ( $\mathbf{A}_i \in R^{m_1 \times m_2 \times \dots \times m_n}$ ) and  $d_1 \times d_2 \times \dots \times d_n$ ;
(1) Construct the similarity matrix  $S$  by (12) and (14);
(2) Compute the embedding as follows:
Initialize  $U_1^0 = I_{d_1 \times m_1}, I_{d_2 \times m_2}, \dots, U_n^0 = I_{d_n \times m_n}$ ;
for  $t = 1, 2, \dots, T_{\max}$  do
  for  $k = 1, 2, \dots, n$  do
     $\mathbf{B}_i^k = \mathbf{A}_i \times_1 U_1 \cdots \times_{k-1} U_{k-1} \times_{k+1} U_{k+1} \cdots \times_n U_n$ ;
     $B_i^{(k)} \leftarrow_k \mathbf{B}_i^k$ ;
     $H_1 = B^{(k)}(I - S)^T(I - S)B^{(k)T}$ ;
     $H_2 = B^{(k)}B^{(k)T}$ ;
    Compute  $U_k^t \in R^{d_k \times m_k}$  by solving the eigen function:  $H_1 U_k^t = H_2 U_k^t \Lambda_k$ ;
    if  $\|U_k^t - U_k^{t-1}\| < \varepsilon$  for each  $k$  then
      break;
    end if
  end for
end for
Output:  $U_i = U_i^t \in R^{d_i \times m_i}$  ( $i = 1, 2, \dots, n$ ).

```

ALGORITHM 1: The RTPP algorithm.

5.1. *Experiments on Equinox Data Set.* The National Institute of Standards and Technology and Equinox Corporation have developed a database (<http://www.equinoxsensors.com/products/HID.html>) of face images using registered broadband-visible/IR camera sensors for experimentation and performance evaluations [10]. Since the registration of the thermal images and the corresponding visible images is fulfilled by camera sensors, in our experiments, we did not need to do these procedures.

We used the long-wave infrared (LWIR) (i.e., $8 \mu\text{m}$ – $12 \mu\text{m}$) and the corresponding visible spectrum images from this database. The data were collected during a two-day period. Each pair of LWIR and visible light images was taken simultaneously and coregistered with 1/3 pixel accuracy. The LWIR images were radiometrically calibrated and stored as grayscale images with 12 bits per pixels. The visible images were also grayscale images represented with 8 bits per pixel [10].

The database contains frontal faces under the following scenarios: (1) three different light directions: frontal and lateral (right and left); (2) three facial expressions: frown, surprise, and smile; (3) vocals pronunciation expressions: subjects were asked to pronounce several vocals from which three representative frames were chosen; and (4) presence of glasses: for subjects wearing glasses, all of the above scenarios were repeated with and without glasses.

In our experiments, 1320 images (660 thermal images and 660 corresponding registered visible images) were used. These images belonged to 33 individuals. For each individual, we had 20 thermal images and 20 corresponding visible images. Original 12-bit gray level thermal images were converted into 8 bits. All images (including thermal images and visible images) were cut off the background, aligned, and then normalized with a resolution of 28×24 . The goal of the preprocessing was to remove background and scale the faces.

Figure 1 shows sample images of one person in the Equinox data set.

For any thermal image and its corresponding visible image, the tensor sample was represented in the size of $28 \times 24 \times 2$ pixels. In the experiments, 10 tensor samples (10 thermal images and 10 corresponding visible images) of each individual were randomly selected and used as training set and the remaining 10 tensor samples as test set. The experiments were independently performed 20 times and the average results were calculated.

For our proposed RTPP algorithm and tensor NPE algorithm, the reduced dimensions $d_1 \times d_2 \times d_3$ of the extracted features were $14 \times 12 \times 1$ and $10 \times 8 \times 1$, respectively. For NPE algorithms performed on IR feature, visible feature, and serial combined feature, the corresponding reduced dimensions were $d = 168$ and $d = 80$, respectively. For tensor NPE algorithm, we performed experiments to obtain the best parameter K (the number of nearest neighbors) for Equinox data set. Figures 2 and 3 showed the ROC curves of the proposed RTPP algorithm and tensor NPE algorithm using different K 's ($K = 5, 10, 15, 20, 25, 30, 35$). From the ROC curves, we can find that the best performance could be obtained when $K = 5$ (for both $d = 168$ and $d = 80$). And the performance of the proposed RTPP algorithm is much better than the tensor NPE algorithm, no matter which K was selected.

The ROC curves of the different methods were shown in Figures 4 and 5. The NPE algorithm was also separately performed on visible data and thermal infrared data. In the NPE algorithm, we set $K = 5$. The results indicate that the performance of the proposed RTPP algorithm is better than other algorithms.

5.2. *Experiments on DHUFO Data Set.* DHUFO is a database of face images using registered visible/IR camera sensors for

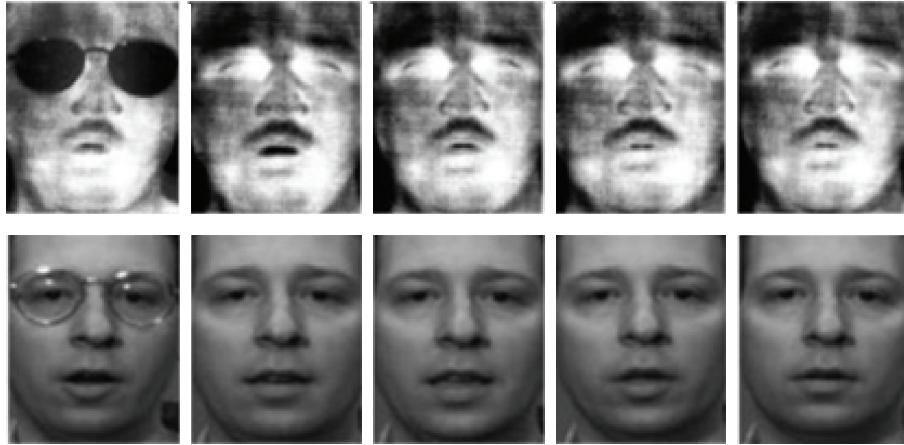


FIGURE 1: Sample images of one individual from Equinox data set.

experimentation and performance evaluations. The data set was designed by the researchers. In our experiments, the long-wave infrared (LWIR) (i.e., $8\ \mu\text{--}12\ \mu$) sensor was used. The registration of the thermal images and the corresponding visible images was fulfilled by the camera sensors. Face image variations in the DHUFO database included illumination, facial expression, and glasses. In our experiments, 1020 images, which involved variations in illumination and facial expressions, were selected. We manually cropped the face portion of the images. These images belonged to 17 individuals. For each individual, there were 30 thermal images and 30 corresponding visible images. All images (including thermal images and visible images) were cut off the background, aligned, and then normalized with a resolution of 28×24 .

For any thermal image and its corresponding visible image, the tensor sample was represented in the size of $28 \times 24 \times 2$ pixels. In the experiments, 15 tensor samples (15 thermal images and 15 corresponding visible images) of each individual were randomly selected and used as training set and the remaining 15 tensor samples as test set. The experiments were independently performed 20 times and the average results were calculated.

For our proposed RTPP algorithm, tensor NPE algorithm, and tensor LPP algorithm, the reduced dimensions $d_1 \times d_2 \times d_3$ of the extracted features were $14 \times 12 \times 1$ and $10 \times 8 \times 1$, respectively. For both NPE and LPP performed on the serial combined feature, the corresponding reduced dimensions were $d = 168$ and $d = 80$, respectively. The ROC curves of the different methods were shown in Figures 6 and 7. In both NPE and LPP algorithm, we set $K = 5$. The results indicate that the performance of the proposed RTPP algorithm is better than other algorithms.

5.3. Discussion. Based on the experimental results, the following observations are obtained.

- (1) From the ROC curves of different methods on Equinox data set and DHUFO data set, the proposed RTPP algorithm obtained the best performance. The experimental results indicate that combining the L_1

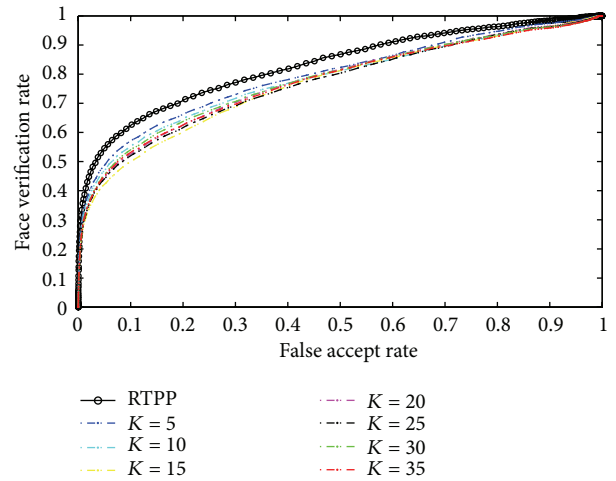


FIGURE 2: ROC curves of RTPP and NPE (NPE was performed on serial combined data with different K 's) on Equinox data set with $d_1 \times d_2 \times d_3 = 10 \times 8 \times 1$ and $d = 80$.

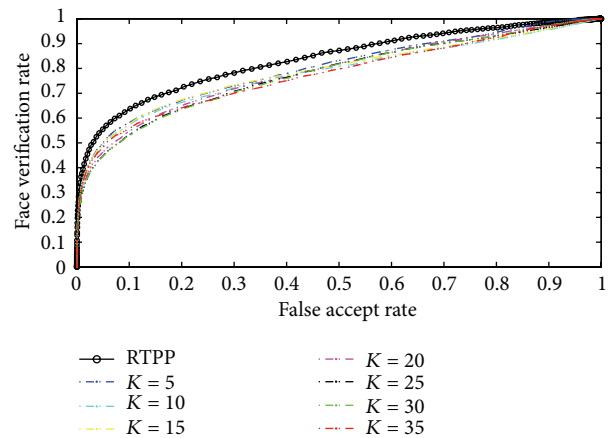


FIGURE 3: ROC curves of RTPP and NPE (NPE was performed on serial combined data with different K 's) on Equinox data set with $d_1 \times d_2 \times d_3 = 14 \times 12 \times 1$ and $d = 168$.

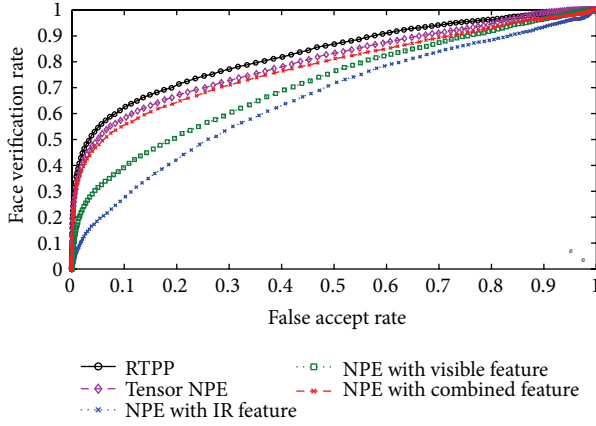


FIGURE 4: ROC curves of RTPP and NPE on Equinox data set with $d_1 \times d_2 \times d_3 = 10 \times 8 \times 1$ and $d = 80$.

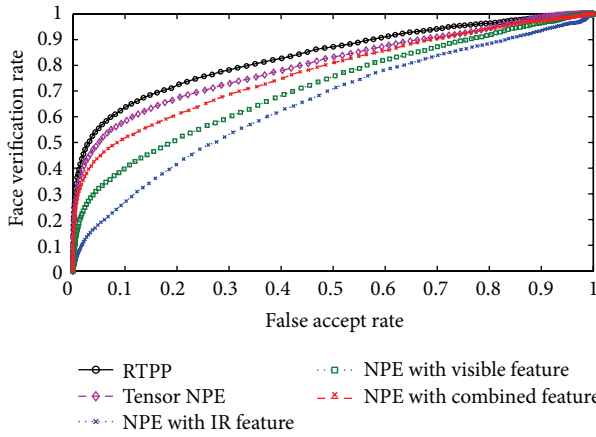


FIGURE 5: ROC curves of RTPP and NPE on Equinox data set with $d_1 \times d_2 \times d_3 = 14 \times 12 \times 1$ and $d = 168$.

norm for sparse tensor learning is a better way than using local information reconstruction.

- (2) RTPP does not introduce the local neighborhood parameter K and thus there is essential difference. In RTPP, the L_1 norm is combined for the reconstruction coefficients with sparse properties; thus the advantages of robustness to data distortion and the potential discriminative ability proven in [25–27] are encoded in the representation coefficients, which are preserved in the low-dimensional subspace. These are the essential reasons for RTPP to achieve good performance.
- (3) Since RTPP well preserves the spatial structure of the original multispectral face images, RTPP outperforms serial combined feature extraction algorithms.

6. Conclusion

We have proposed in this paper a novel tensor learning algorithm, called robust tensor preserving projection (RTPP),

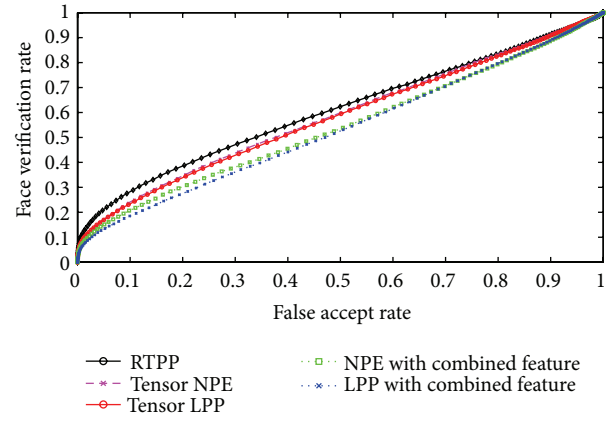


FIGURE 6: ROC curves of RTPP, NPE, and LPP on DHUFO data set with $d_1 \times d_2 \times d_3 = 10 \times 8 \times 1$ and $d = 80$.

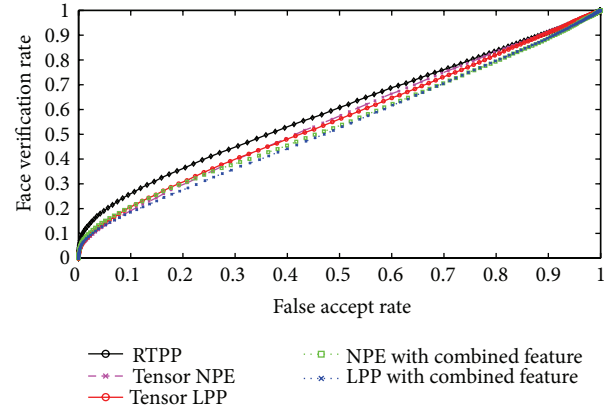


FIGURE 7: ROC curves of RTPP, NPE, and LPP on DHUFO data set with $d_1 \times d_2 \times d_3 = 14 \times 12 \times 1$ and $d = 168$.

for multispectral face recognition. The STE algorithm incorporates tensor manifold criterion to learn multiple subspaces in high-order tensor space by preserving the sparse representation information of the multispectral images. RTPP cannot only keep the underlying spatial structure of multispectral images but also enhance robustness. Experimental results demonstrate the excellent performance of RTPP.

Since RTPP is an unsupervised learning algorithm, one of our future works will be supervised tensor learning algorithms. We also plan to enforce the sparsity on the projection matrix/vector and investigate the sparse projection learning methods for tensor recognition.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 61375007) and the Shanghai Pujiang Program (Project no. 12PJ1402200).

References

- [1] S. G. Kong, J. Heo, F. Boughorbel et al., "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," *International Journal of Computer Vision*, vol. 71, no. 2, pp. 215–233, 2007.
- [2] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition—a review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005.
- [3] D. A. Socolinsky and A. Selinger, "A comparative analysis of face recognition performance with visible and thermal infrared imagery," in *Proceedings of the 16th International Conference on Pattern Recognition*, pp. 217–222, IEEE, 2002.
- [4] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003.
- [5] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Illumination-invariant face recognition in hyperspectral images," in *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, vol. 5093, pp. 275–282, 2003.
- [6] L. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," Tech. Rep. CMU-RI-TR-02-25, 2002.
- [7] Y. Chou and P. Bajcsy, "Toward face detection, pose estimation and human recognition from hyperspectral imagery," Tech. Rep. NCSA-ALG-04-0005, 2004, <http://isda.ncsa.uiuc.edu/peter/>.
- [8] S. Wang, Z. Liu, S. Lv et al., "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 682–691, 2010.
- [9] W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Transactions on Systems, Man, and Cybernetics A: Systems and Humans*, vol. 40, no. 6, pp. 1354–1361, 2010.
- [10] A. Selinger and D. A. Socolinsky, "Appearance-based facial recognition using visible and thermal imagery: a comparative study," Tech. Rep., Equinox Corporation, 2000.
- [11] X. Chen, P. Flynn, and K. Bowyer, "PCA-based face recognition in infrared imagery: baseline and comparative studies," in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
- [12] S. Wang, J. Yang, M. Sun, X. Peng, M. Sun, and C. Zhou, "Sparse tensor discriminant color space for face verification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 876–888, 2012.
- [13] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, "Infrared and visible image fusion for face recognition," in *Biometric Technology for Human Identification*, vol. 5404 of *Proceedings of the SPIE*, pp. 585–596, April 2004.
- [14] J. Heo, S. G. Kong, B. R. Abidi, and M. A. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *Proceedings of IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, pp. 94–99, 2004.
- [15] O. Arandjelovic, "Multi-sensory face biometric fusion (for personal identification) method details," in *Proceedings of IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, pp. 1–8, 2006.
- [16] H. Chang, H. Harishwaran, M. Yi, A. Koschan, B. Abidi, and M. Abidi, "An indoor and outdoor, multimodal, multispectral and multi-illuminant database for face recognition," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '06)*, p. 54, June 2006.
- [17] W. K. Wong and H. Zhao, "Eyeglasses removal of thermal image based on visible information," *Information Fusion*, vol. 14, no. 2, pp. 163–176, 2013.
- [18] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [19] G. Dai and D. Y. Yeung, "Tensor embedding methods," in *Proceedings of the 21st AAAI Conference on Artificial Intelligence*, vol. 1, pp. 330–335, 2005.
- [20] D. Zheng, X. Du, and L. Cui, "Tensor locality preserving projections for face recognition," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '10)*, vol. 1, pp. 2347–2350, October 2010.
- [21] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.
- [22] H. Lu, K. N. K. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning," *IEEE Transactions on Neural Networks*, vol. 20, no. 11, pp. 1820–1836, 2009.
- [23] D. Xu, S. Yan, L. Zhang, S. Lin, H. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 36–47, 2008.
- [24] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: a unified framework for sparse dimension reduction," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 340–371, 2011.
- [25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [26] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.
- [27] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with l^1 -graph for image analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 858–866, 2010.
- [28] H. Moon and P. Phillips, "The FERET verification testing protocol for face recognition algorithms," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 48–53, 1998.
- [29] B. Bader and T. G. Kolda, *Tensor toolbox version 2.3*, Sandia National Laboratories, 2009, <http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.3.html>.

- [30] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [31] M. Rosen and X. Jiang, "Lippmann2000: a spectral image database under construction," in *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, pp. 117–122, 1999.
- [32] T. Igarashi, K. Nishino, and S. Nayar, "The appearance of human skin," Tech. Rep. CUCS-02405, Columbia University, 2005.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

