Semi-Supervised Bilinear Subspace Learning

Dong Xu and Shuicheng Yan

Abstract—Recent research has demonstrated the success of tensor based subspace learning in both unsupervised and supervised configurations (e.g., 2-D PCA, 2-D LDA, and DATER). In this correspondence, we present a new semi-supervised subspace learning algorithm by integrating the tensor representation and the complementary information conveyed by unlabeled data. Conventional semi-supervised algorithms mostly impose a regularization term based on the data representation in the original feature space. Instead, we utilize graph Laplacian regularization based on the low-dimensional feature space. An iterative algorithm, referred to as adaptive regularization based semi-supervised discriminant analysis with tensor representation (ARSDA/T), is also developed to compute the solution. In addition to handling tensor data, a vector-based variant (ARSDA/V) is also presented, in which the tensor data are converted into vectors before subspace learning. Comprehensive experiments on the CMU PIE and YALE-B databases demonstrate that ARSDA/T brings significant improvement in face recognition accuracy over both conventional supervised and semi-supervised subspace learning algorithms.

Index Terms—Adaptive regularization, dimensionality reduction, face recognition, semi-supervised learning.

I. INTRODUCTION

Dimensionality reduction algorithms [e.g., principal component analysis (PCA), linear discriminant analysis (LDA), and tensorface] have been successfully used in the face recognition task. PCA [14] is an unsupervised method, which pursues the directions of maximum variance for optimal image reconstruction. As a supervised method, LDA [1] aims to maximize the between-class scatter and at the same time minimize the within-class scatter. Tensorface [16] organizes the image ensembles of different persons under different poses, illuminations, and expressions into a high-order tensor, and then multilinear analysis is performed on the tensor data. PCA, LDA, and tensorface commonly unfold each image into a single column vector before dimensionality reduction, which is referred to as an image-as-vector representation as in [19].

However, image objects are intrinsically in the form of second or higher order tensors. For example, gray-level images represent second-order tensor data (matrices), and can be expanded to third-order tensors by representing sets of images after Gabor filtering [19]. It is often helpful to process the data in their original form and order [11]–[13], [19]–[23], which for the case of images is referred to as an image-as-matrix representation [19]. Several different groups [11], [19], [22] have experimentally reported that the image-as-matrix representation can lead to good classification performance for different applications (e.g., face recognition and human gait recognition), especially when the number of training samples is small. A possible explanation is that an image-as-matrix representation can utilize correlations among pixels within different dimensions (i.e., rows, columns, and so on). Moreover, the smaller number of data entries along each data dimension facilitates subspace learning with limited training data.

Supervised learning algorithms generally outperform unsupervised learning techniques because of utilization of label information. For example, LDA generally outperforms PCA when sufficient labeled training data are available [1]. But it is usually difficult, expensive, and time-consuming to collect sufficient labeled data. Meanwhile, unlabeled data are relatively much easier to obtain. Semi-supervised learning techniques [8], [17], [18], [24], [26], [28] aim to utilize a large amount of unlabeled data as well as relatively limited labeled data for better classification. The classical algorithms include transductive support vector machines (T-SVM) [15] and co-training [4].

There is an increasing interest in graph based semi-supervised classification techniques [3], [10], [25], [27]. These techniques aim to estimate a function on the graph with respect to two factors: 1) the function output is equal to the given labels for the labeled data; and 2) the function is smooth for all labeled and unlabeled samples. These algorithms not only consider the label information, but also utilize a consistency assumption, namely, nearby points are likely to have the same label in classification tasks [25], [27]. For semi-supervised dimensionality reduction, Cai et al. [5] recently proposed semi-supervised discriminant analysis (SDA) by adding a smoothness constraint into the objective function of LDA. This constraint based on graph Laplacian regularization [2] aims to enforce nearby points to have similar representations in the low-dimensional feature space.

We present a new semi-supervised subspace learning algorithm by integrating the tensor representation and the complementary information conveyed by unlabeled data. Similar to [5], we use graph Laplacian regularization [2] for constraining the consistency. The prior work [5], [25], [27] assumed that the consistency constraint is defined on the original feature space. For dimension reduction or subspace learning, the main target is to search for a better lower dimensional representation for subsequent classification, and, thus, the original feature space is not the best for defining consistency. A natural way to define consistency in a subspace learning problem is to directly define consistency (e.g., neighborhood relationship) based on the data representation in the expected low-dimensional feature space. We, therefore, refer to the new consistency definition as adaptive regularization in this work. Since expected neighborhood relationship (ENR) is generally indeterminable from examination of the original feature space, we propose an iterative procedure, referred to as adaptive regularization based semi-supervised discriminant analysis with tensor representation (ARSDA/T), for the adaptive refinement of ENR. More specifically, this procedure initially utilizes ENR computed in the original feature space, and then iteratively adapts ENR according to the dimension-reduced data representation from the most recently computed projection matrices. With the incorporation of a tensor representation and adaptive regularization technique, significant improvements can be gained in face recognition performance when compared with prior work [1], [5], [22]. In addition to handling tensor data, vector-based variant ARSDA/V is also developed in this work, in which the tensor data are first converted into vectors before subspace learning.

The main aspects of the proposed approach are listed as follows.

- To the best of our knowledge, ARSDA/T is the first semi-supervised subspace learning algorithm, which deals with tensor data and adaptive graph regularization.
- ARSDA/V outperforms LDA [1] and the recently proposed semi-supervised technique SDA [5].
- In this work, we primarily focus our general formulation on second-order tensors (namely matrices) to enhance the read-
The low-dimensional data representation high-order input. In this work, we briefly review 2-D LDA to enhance learnability and preserve structural information, 2-D LDA were recently proposed to directly perform dimensionality reduction. We then denote the sample set with vector representation as $X = \{x_1, x_2, \ldots, x_N\} \subseteq \mathbb{R}^{m \times n}$, and $x_i \in \mathbb{R}^{m \times n}$, where $m$ and $n$ are the dimensions of the lower dimensional matrices. The labels of $X_i$ are denoted as $l_i \in \{1, 2, \ldots, N\}$ with $N$ being the total number of classes. The objective function of 2-D LDA is written as

$$
(U^*, V^*) = \arg \max_{U, V} \frac{S_b}{S_w} = \arg \max_{U, V} \frac{\sum_{i=1}^{N} n_i ||(U^T \tilde{X}_i V - U^T X_i V)^T||^2}{\sum_{i=1}^{N} ||U^T X_i V - U^T X_i \tilde{X}_i V||^2}
$$

where $n_i$ is the total number of labeled samples in the $i$-th class, $\tilde{X}_i$ is the mean image of labeled samples from the $i$-th class, and $X_i$ is the mean image of all labeled samples.

Let us define two similarity matrices $\tilde{S}_1, \tilde{S}_2 \in \mathbb{R}^{N \times N}$, where $\tilde{S}_1 = \delta_{l_i l_j} / n_i$, $\tilde{S}_2 = (1/N) - \tilde{S}_1$. According to the graph embedding framework in [21], the interclass scatter $S_i$ and the intraclass scatter $S_w$ in 2-D LDA (or DATER) can be rewritten as

$$
\tilde{S}_w = \sum_{i=1}^{N} \sum_{j=1}^{N} ||(U^T X_i V - U^T X_j V)^T||^2 \tilde{S}_1
$$

$$
\tilde{S}_b = \sum_{i=1}^{N} \sum_{j=1}^{N} ||U^T X_i V - U^T X_j V||^2 \tilde{S}_2. \tag{2}
$$

B. Semi-Supervised Discriminant Analysis (SDA)

Both LDA and 2-D LDA are supervised learning algorithms. In real applications, it is usually difficult to collect sufficient labeled data. To utilize large amounts of unlabeled data as well as relatively limited labeled data, Cai et al. [5] recently extended LDA to Semi-supervised Discriminant Analysis (SDA) for semi-supervised dimensionality reduction. Face images are firstly converted into vectors before dimensionality reduction. We then denote the sample set with vector representation as $X = \{x_1, x_2, \ldots, x_N, x_{N+1}, \ldots, x_M\} \subseteq \mathbb{R}^{m \times n}$, where $x_i \in \mathbb{R}^{m \times n}$, and $x_{N+1}, \ldots, x_M$ are labeled and unlabeled data, respectively. We also use $X_i = \{x_1, x_2, \ldots, x_N\}$ to represent labeled data. Let us denote the projection matrix as $P \in \mathbb{R}^{m \times d}$, where $d$ is the dimension of the reduced feature space. For any similarity matrix $S$, the corresponding diagonal matrix $D$ and Laplacian matrix $L$ are defined as follows [2], [5], [21]:

$$
L = D - S, \quad D_{ii} = \sum_{j \neq i} S_{ij}, \quad \forall i. \tag{3}
$$

For the similarity matrices $\tilde{S}_1$ and $\tilde{S}_2$ used in (2), we denote their corresponding diagonal matrices as $\tilde{D}_1$ and $\tilde{D}_2$, and their Laplacian matrices as $\tilde{L}_1$ and $\tilde{L}_2$. The intraclass scatter $S_i$ in LDA [1] is defined as $S_i = \sum_{j=1}^{N} P^T (x_i - \tilde{x}_i)(x_j - \tilde{x}_i)^T P$, where $\tilde{x}_i$ is the mean of labeled samples in the $i$-th class. As shown in the graph embedding framework in [21] and [5], $S_w$ can be rewritten as $S_w = \text{Tr}[P^T X_i \tilde{D}_1^{-1} \tilde{L}_1^{-1} X_i^T P] = \text{Tr}[S_i X_i \tilde{D}_1^{-1} \tilde{L}_1^{-1}]$. The interclass scatter $S_i$ in LDA [1] can be similarly rewritten as $S_{ij} = \text{Tr}[P^T X_i (\tilde{D}_2^{-1} - \tilde{S}_1) X_j^T P] = \text{Tr}[P^T X_i \tilde{L}_2^{-1} X_j^T P]$.

The core assumption in SDA is that nearby points will have similar representations in the lower dimensional space. To enforce this assumption, SDA additionally introduced similarity matrix $S \in \mathbb{R}^{m \times m}$ with the graph Laplacian criterion [2]

$$
S_{ij} = \begin{cases} 
1, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\
0, & \text{otherwise}
\end{cases} \tag{4}
$$

where $N_k(x_i)$ denotes the set of $k$ nearest neighbors of $x_i$, based on the Euclidean distance in the original feature space. But for dimensionality reduction, the original feature space is considered to be not the best. Ideally, $k$ nearest neighbors of $x_i$ should be decided based on the data representation in the expected low-dimensional space, which is one of the main motivations for this work. The objective function in SDA is formulated as

$$
P^* = \arg \max_{P} \frac{S_b}{S_w + \alpha \text{Tr}[P^T X_i (D - S) X_i^T P]} = \arg \max_{P} \frac{S_b}{\text{Tr}[P^T X_i \tilde{D}_1^{-1} \tilde{L}_1^{-1} X_i^T P] + \alpha \text{Tr}[P^T X_i \tilde{L}_2^{-1} X_i^T P]} \tag{5}
$$

where $\alpha$ is a parameter to balance within-class scatter from labeled data and the consistency from the whole set of samples [5].

III. ADAPTIVE REGULARIZATION BASED SEMI-SUPERVISED DISCRIMINANT ANALYSIS WITH TENSOR REPRESENTATION (ARSDA/T)

Inspired by the recent success of algorithms based on the image-as-matrix representation, we represent the sample set as second-order tensors $\{X_1, X_2, \ldots, X_N, X_{N+1}, \ldots, X_M\} \subseteq \mathbb{R}^{m \times n \times d}$, where $X_i \in \mathbb{R}^{m \times n \times d}$ are labeled and unlabeled data respectively. Let us denote two projection matrices as $U \in \mathbb{R}^{p \times m \times d}$ and $V \in \mathbb{R}^{q \times n \times d}$. The low-dimensional data representation $Y_i$ is then obtained by $Y_i = U^T X_i V; i = 1, \ldots, M$. To more effectively impose data consistency regularization, we define the graph for Laplacian regularization on data in the expected low-dimensional feature space. More specifically speaking, we denote a new similarity matrix $S^{out}(U, V)$ related to the projection matrices $U$ and $V$

$$
S^{out}_{ij}(U, V) = \begin{cases} 
1, & \text{if } X_i \in N_k(X_j, U, V) \text{ or } X_j \in N_k(X_i, U, V) \\
0, & \text{otherwise}
\end{cases} \tag{6}
$$

where $N_k(X_i, U, V)$ represents the $k$ nearest neighbors of $X_i$ based on the Euclidean distances in the expected low-dimensional feature space projected by $U$ and $V$.

To utilize the tensor representation as well as effectively enhance consistency for both labeled and unlabeled data, we propose the objective function shown in (7), at the bottom of the next page, where $\alpha$ is a parameter to balance within-class scatter from labeled data and the consistency of the whole set of samples.
It can be seen that this criterion poses a chicken-and-egg problem, since $U$ and $V$ need to be known for determining $S^{opt}(U, V)$ and vice versa. To the best of our knowledge, no closed-form solution exists for this objective function. To address this problem, we present a procedure to search for the solution in an iterative manner. Given the projection matrices $U^{t-1}$ and $V^{t-1}$ at the $(t-1)$th step, the projection matrices $U^t$ and $V^t$ at the $t$th step are computed based on $S^{opt}(U^{t-1}, V^{t-1})$ as an approximation of $S^{opt}(U, V)$. Let us denote two new matrices $S^1$ and $S^2$ in $\mathbb{R}^{m \times M}$ based on similarity matrices $\hat{S}^1$ and $\hat{S}^2$:

$$
S^1 = \begin{bmatrix} \hat{S}^1 & 0 \\ 0 & 0 \end{bmatrix}, \quad S^2 = \begin{bmatrix} \hat{S}^2 & 0 \\ 0 & 0 \end{bmatrix}.
$$

(8)

Then we have $\hat{S}_u = \sum_{i=1}^{M} \sum_{j=1}^{M} ||U^{t-1} X_i V^t - U^{t-1} X_j V^t||^2 S^2_{ij}$ and $\hat{S}_v = \sum_{i=1}^{M} \sum_{j=1}^{M} ||U^{t-1} X_i V^t - U^{t-1} X_j V^t||^2 S^1_{ij}$. By utilizing $S^{opt}(U^{t-1}, V^{t-1})$ as an approximation of $S^{opt}(U, V)$, we have the following formulation (see (9), shown at the bottom of the page).

As shown in [20] and [22], $U^t$ (or $V^t$) can be solved iteratively by fixing the other projection matrices $V^t$ (or $U^t$). Specifically, for a given $U^{t-1} \in \mathbb{R}^{m \times m}$, $\hat{S}_u$ and $\hat{S}_v$ can be rewritten as

$$
\hat{S}_u = \text{Tr}(U^{t-1} \hat{S}_u V^t), \quad \hat{S}_v = \text{Tr}(U^{t-1} \hat{S}_v V^t).
$$

(10)

$\hat{S}_u$ and $\hat{S}_v$ are defined as follows:

$$
\hat{S}_u = \sum_{i=1}^{M} \sum_{j=1}^{M} (X^u_i - X^u_j)^T (X^u_i - X^u_j) S^1_{ij} + \alpha S^2_{ij}(U^{t-1}, V^{t-1}),
$$

$$
\hat{S}_v = \sum_{i=1}^{M} \sum_{j=1}^{M} (X^{v^T}_i - X^{v^T}_j)^T (X^{v^T}_i - X^{v^T}_j) S^2_{ij},
$$

where $X^u_i = (U^{t-1})^T X_i$ and $X^{v^T}_i = (U^{t-1})^T X_i$. We use generalized eigenvalue decomposition method to compute an approximate solution as

$$
\hat{S}_u^{v^{T^t}} i = \tau_i^{v^{T^t}} \hat{S}_u^{v^{T^t}} i,
$$

(11)

where $\tau_i^{v^{T^t}} \geq \tau_1^{v^{T^t}} \geq \cdots \geq \tau_{M'N+1}^{v^{T^t}}$ are the $M'$ largest eigenvalues, and $v_i^{v^{T^t}}$ is the eigenvector corresponding to eigenvalue $\tau_i^{v^{T^t}}$, which constitutes the $i$th column of the solution $V^t$.

Similarly, for a given $V^t \in \mathbb{R}^{n \times n'}$, $\hat{S}_u$ and $\hat{S}_v$ can be rewritten as

$$
\hat{S}_u = \text{Tr}(U^T \hat{S}_u U^t), \quad \hat{S}_v = \text{Tr}(U^T \hat{S}_v U^t).
$$

(12)

$$
(U^*, V^*) = \arg \max_{u, v} \frac{\hat{S}_u}{\alpha \sum_{i=1}^{M} \sum_{j=1}^{M} ||U^T X_i V - U^T X_j V||^2 S^{opt}_{ij}(U, V)}
$$

$$
= \arg \max_{u, v} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} ||U^T X_i V - U^T X_j V||^2 S^1_{ij} + \alpha \sum_{i=1}^{M} \sum_{j=1}^{M} ||U^T X_i V - U^T X_j V||^2 S^2_{ij}}
$$

(7)

$$
(U^{t*}, V^{t*}) = \arg \max_{u, v} \frac{\hat{S}_u}{\alpha \sum_{i=1}^{M} \sum_{j=1}^{M} ||U^T X_i V - U^T X_j V||^2 S^{opt}_{ij}(U, V)}
$$

$$
= \arg \max_{u, v} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} ||U^T X_i V - U^T X_j V||^2 S^1_{ij} + \alpha \sum_{i=1}^{M} \sum_{j=1}^{M} ||U^T X_i V - U^T X_j V||^2 S^2_{ij}}
$$

(9)
Then we have $X_i(\hat{D}^i - \hat{S}^i)X^T = X_i\hat{L}^iX^T = X_{i}L_{i}X^T$ and $X_i(\hat{D}^2 - \hat{S}^2)X^T = X_i\hat{L}^2X^T = X_{i}L_{2}X^T$. By utilizing $S^{\text{uns}(P^{d^1})}$ as an approximation of $S^{\text{uns}(P)}$, the objective function in (15) can be rewritten as

$$P^* = \arg \max_{P^d} \frac{\text{Tr}[P^T X L^2 X^T P^d]}{\text{Tr}[P^T X L^1 X^T P^d] + \alpha \text{Tr}[P^T X S^{\text{uns}(P^{d^1})}X^T P^d]}$$

(17)

where $L^{\text{uns}(P^{d^1})}$ is the corresponding Laplacian matrix of $S^{\text{uns}(P^{d^1})}$.

We then use the generalized eigenvalue decomposition method to compute the approximate solution $P^d$:

$$X L^2 X^T p_i = \lambda_i^d X L^1 + \alpha L^{\text{uns}(P^{d^1})} X^T p_i$$

(18)

where $\lambda_1^d \geq \lambda_2^d \geq \ldots \geq \lambda_{n^d}^d$ are the $d$ largest eigenvalues, and $p_i^d$ is the eigenvector corresponding to eigenvalue $\lambda_i^d$, which constitutes the $i$th column of the solution $P^d$. The details of ARSDA/V are listed in Procedure 2.

**Procedure-2: The procedure for ARSDA/V**

Given the labeled data $x_i|_{i=1}^{n_l} \in \mathbb{R}^{n \times m}$ and unlabeled data $x_i|_{i=n_l+1}^{n} \in \mathbb{R}^{n \times m}$, the final lower dimensions $d$, and the iteration number $T_{\text{max}}$.

1. Initialize $P^0 = I_{m \times m}$.
2. For $t = 1, 2, \ldots, T_{\text{max}}$ Do
   • Compute the similarity matrix $S^{\text{uns}(P^{d^t-1})}$ based on $P^{d^t-1}$ with (14).
   • Compute the optimal project matrix $P^d \in \mathbb{R}^{n \times n \times d}$ with (18).
3. Output the project matrix $P = P^t$.

V. EXPERIMENTS

We compare ARSDA/T and ARSDA/V with two classical supervised learning algorithms Fisherface [1] and 2-D LDA [22], in which only the labeled data are used. We also compare our algorithms with the recently proposed semi-supervised dimensionality reduction algorithm SDA [5], which outperforms several semi-supervised algorithms [3], [25] and unsupervised algorithms (e.g., PCA [14] and LPP [7]) for face recognition. To demonstrate the advantage of adaptive regularization, we also report the results from tensorized SDA (referred to as SDA/T here) by extending SDA with our tensorization framework introduced in [21]. ARSDA/T differs from SDA/T in that the neighborhood relationship in SDA/T is decided based on the original sample representation, while it is determined based on the data representation in the expected low-dimensional feature space in ARSDA/T. To clearly denote algorithms operating on an image-as-vector representation, we refer to Fisherface, SDA, 2-D LDA as LDA/V, SDA/V, and LDA/T, respectively. For all the experiments, the nearest neighbor classifier is applied after dimensionality reduction for a final decision.

There are two settings referred to as the transductive setting [25], [27] and the semi-supervised setting [5] in utilizing the unlabeled data. In the transductive setting [25], [27], both training and test samples (without label information) are used during the learning stage. For the semi-supervised setting, only training samples (including both labeled and unlabeled data) are used for model training, and the test samples
are not available during the training process. In this work, we choose the semi-supervised setting because it is a more natural setting to compare different dimensionality reduction algorithms for face recognition. To improve the performance over classical supervised learning methods, semi-supervised learning algorithms [5], [28] usually require sufficient unlabeled samples. We choose the CMU PIE database [9] and YALE-B database (Extended Yale Face Database) [6] for experiments because each subject has more than 40 images in these two databases. The databases are downloadable from http://www.cs.uiuc.edu/homes/dengcai2/Data/FaceData.html.

A. Experiments on CMU PIE Database

The CMU pose, illumination, and expression (PIE) database [9] contains more than 40,000 facial images of 68 people. The images were acquired over different poses, under variable illumination conditions, and with different facial expressions. In this experiment, we choose the images from the frontal pose and each subject has around 49 images from varying illuminations and facial expressions. The images are cropped by fixing the positions of the two eyes, and then resizing to 64 × 64 pixels. Histogram equalization is used for preprocessing. The images after preprocessing are shown in Fig. 1. For each subject, 30 images are randomly selected as the training set, and the remaining images are used as the test set. We consider two configurations to divide training samples. \( \text{L2} \) means two images per subject are randomly selected as labeled samples and the remaining 28 images per subject are used as unlabeled data. \( \text{L3} \) means three images per subject are randomly selected as labeled data and the remaining 27 images per subject are used as unlabeled data.

For the face recognition experiments, several parameters need to be decided beforehand. For LDA/V, SDA/V, and ARSDA/V, we use PCA as a first step dimensionality reduction algorithm to avoid the singularity problem. The regularization technique in [5] is not used. As in [1], the dimension of the PCA step is fixed as \( T \). In terms of the optimal dimension after dimensionality reduction, we run all the possible dimensions and report the best results for LDA/V, SDA/V, and ARSDA/V. For LDA/T, SDA/T, and ARSDA/T, we set \( m', n' \) from 2 to 40 at intervals of 2 for the CMU PIE database. For SDA/V, ARSDA/V, SDA/T, and ARSDA/T, we need to decide the optimal \( \lambda \).

B. Experiments on YALE-B Database

For the YALE-B database [6], 38 subjects are used in this work, with each person having around 64 near frontal images under different illu-

---

**TABLE I**

TOP-1 RECOGNITION RATES (%) ON THE CMU PIE AND YALE-B DATABASES. NOTE THAT THE LAST NUMBERS IN PARENTHESES ARE THE OPTIMAL DIMENSIONS AFTER DIMENSIONALITY REDUCTION. THE FIRST NUMBERS IN ARSDA/V, LDA/T, SDA/T, AND ARSDA/T ARE THE OPTIMAL ITERATION NUMBERS.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>CMU PIE database (L2)</th>
<th>CMU PIE database (L3)</th>
<th>YALE-B database (L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA/V [1]</td>
<td>64.1 (36)</td>
<td>72.7 (64)</td>
<td>58.7 (23)</td>
</tr>
<tr>
<td>SDA/V [5]</td>
<td>63.4 (38)</td>
<td>72.8 (67)</td>
<td>56.8 (25)</td>
</tr>
<tr>
<td>ARSDA/V</td>
<td>\textbf{68.7} (10, 35)</td>
<td>\textbf{77.1} (10, 65)</td>
<td>\textbf{61.8} (10, 24)</td>
</tr>
<tr>
<td>LDA/T [23]</td>
<td>66.3 (10, 40 × 12)</td>
<td>77.6 (6, 36 × 4)</td>
<td>73.4 (10, 26 × 6)</td>
</tr>
<tr>
<td>SDA/T</td>
<td>69.4 (5, 36 × 12)</td>
<td>77.6 (9, 30 × 10)</td>
<td>67.9 (5, 28 × 4)</td>
</tr>
<tr>
<td>ARSDA/T</td>
<td>\textbf{71.2} (7, 36 × 12)</td>
<td>\textbf{79.0} (10, 32 × 4)</td>
<td>\textbf{76.2} (10, 30 × 4)</td>
</tr>
</tbody>
</table>

Fig. 4. Top-1 recognition rates (%) with different feature dimensions on the CMU PIE and YALE-B databases with 1-D vector input (top) and 2-D matrix input (bottom). For tensor-based algorithms, the feature dimension is the product of the reduced dimensions. (a) CMU PIE database (L2), (b) CMU PIE database (L3), (c) YALE-B database (L2).
minations. The images are cropped by fixing the positions of two eyes and then resizing to \(32 \times 32\) pixels. Histogram equalization is used for preprocessing. The images after preprocessing is also shown in Fig. 1. For each subject, 30 images are randomly selected as the training set, and the remaining images are used as the test set. Among the training images, 2 images per subject are randomly selected as labeled data and the other 28 images per subject are used as unlabeled data. We use the same parameter configuration as in the CMU PIE database, except that we set \(n\) and \(n\)' from 1 to 32 at intervals of 1 for LDA/T, SDA/T and ARSDA/T. The top-one recognition rates are also listed in Table I. Fig. 4(c) displays recognition accuracy when using different feature dimensions.

C. Discussion

From the above experiments, the following observations can be made.

- LDA/T, SDA/T, and ARSDA/T outperform LDA/V, SDA/V, and ARSDA/V, respectively, which demonstrates the advantage of the image-as-matrix representation. The limited training data are generally insufficient for representing such complex data distributions in the CMU PIE and YALE-B databases, so algorithms based on the image-as-matrix representation outperform their corresponding versions based on the image-as-vector representation. These findings are consistent with those of previous studies [11], [20], [22].

- SDA/V outperforms LDA/V in the CMU PIE database (L3) and SDA/T is better than LDA/T on the CMU PIE database (L2), which demonstrates that unlabeled data can be used to improve face recognition performance. However, we also observe that SDA/V and SDA/T are worse than LDA/V and LDA/T, respectively in other cases. A possible explanation is that the consistency assumption based on data representation in the original feature space cannot ensure that nearby points will have the same class labels, and, hence, the graph regularization based on the original feature space may even bring negative effects, which on the other hand illustrates the necessity of our proposed algorithm for semi-supervised subspace learning.

- By enforcing consistency on the data in the expected low-dimensional space, ARSDA/V outperforms LDA/V, and SDA/V in all cases.

- With the incorporation of the image-as-matrix representation as well as the adaptive regularization technique, ARSDA/T achieves the best results in all the cases.

VI. Conclusion

In this correspondence, we proposed a new semi-supervised dimensionality reduction algorithm called adaptive regularization based semi-supervised discriminant analysis with tensor representation (ARSDA/T). ARSDA/T represents face images as matrices and enforces graph Laplacian regularization based on the data representation in the expected low-dimensional feature space. The core idea of this work, semi-supervised subspace learning with adaptive graph Laplacian regularization, is general, and can be used for converting any new supervised subspace learning algorithm (e.g., [13]) into a semi-supervised configuration.

REFERENCES


