General Solution for Supervised Graph Embedding

Qubo You, Nanning Zheng, Shaoyi Du, and Yang Wu

Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University,
Xi'an,Shaanxi Province 710049 P.R. China
youqubo@gmail.com, nnzheng@mail.xjtu.edu.cn, sydu@aiar.xjtu.edu.cn,
ywu@aiar.xjtu.edu.cn

Abstract. Recently, Graph Embedding Framework has been proposed for feature extraction. However, it is an open issue that how to compute the robust discriminant transformation. In this paper, we first show that supervised graph embedding algorithms share a general criterion (Generalized Rayleigh Quotient). Through novel perspective to Generalized Rayleigh Quotient, we propose a general solution, called *General Solution for Supervised Graph Embedding (GSSGE)*, for extracting the robust discriminant transformation of Supervised Graph Embedding. Finally, extensive experiments on real-world data are performed to demonstrate the effectiveness and robustness of our proposed GSSGE.

1 Introduction

Distance metric learning plays an important role in the field of machine learning. Actually, learning a robust distance metric is equivalent to looking for a robust transformation which transforms the original space into the feature space. Hence, those traditional methods for feature extraction can be viewed as metric learning algorithms, such as Linear Discriminant Analysis (LDA) [1], Local Discriminant Embedding (LDE) [2] and Locality Sensitive Discriminant Analysis (LSDA) [3]. In 2007, Yan et al. [4] presented a general framework (Graph Embedding) to unify the above algorithms. In this framework, however, it is an open issue that how to compute the robust discriminant transformation. Since this framework uses a general criterion, it is natural that a general solution can be used for extracting the robust discriminant transformation. Unfortunately, most of algorithms don't thoroughly consider the problem that how to compute the robust discriminant transformation. Similar to classical LDA [1], LDE and LSDA first use Principal Component Analysis (PCA) to reduce dimension for dealing with small sample size problem (SSS) where the data dimension is much larger than the sample size. However, this method may result in the loss of important discriminative information [5]. In the literature of LDA, different variations have been proposed to deal with the SSS problem [6,7,8,9]; however, they aim at LDA.

In this paper, we focus on the problem that how to compute the robust discriminant transformation of Supervised Graph Embedding. We first show that supervised graph embedding algorithms share a general criterion (Generalized

J.N. Kok et al. (Eds.): ECML 2007, LNAI 4701, pp. 782–789, 2007.

Rayleigh Quotient). Then, we propose a general solution, called *General Solution for Supervised Graph Embedding (GSSGE)*, to extract the robust discriminant transformation. Finally, experimental results on real-world data demonstrate the robustness of GSSGE.

The rest of the paper is organized as follows: Supervised Graph Embedding framework is briefly reviewed in Section 2. In Section 3, General Solution for Supervised Graph Embedding (GSSGE) is described. Extensive experiments are performed to demonstrate the effectiveness and robustness of our proposed GSSGE in Section 4. Finally, conclusions are summarized in Section 5.

2 Supervised Graph Embedding

For the convenience of understanding, in the following, the small *italic* letters denote scalars, such as a, b, c; the small **bold** non-italic letters denote vectors, such as $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$; and the capital **bold** non-italic letters denote matrices, such as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$. Let we have n samples $\{\mathbf{x}_i | \mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ belonging to c classes.

Let $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$ be an undirected weight graph with vertex set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ and the similarity matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$. The element of similarity matrix $\mathbf{W}$ measures the similarity of the vertex pair. Graph Embedding is to extract the optimal low dimensional vector representation for each vertex of the graph $\mathbf{G}$. Assume that the low dimensional vector representation of each vertex can be obtained from linear projections. For simplicity, we consider the one dimensional case. It is easy to extend to multi-dimensional cases. Let $\mathbf{q} \in \mathbb{R}^{d\times 1}$ be the linear mapping from the d-dimensional space to a line, and $\{y_i = \mathbf{q}^T\mathbf{x}_i\}_{i=1}^n$ are the low dimensional representation of the vertex set $\mathbf{X}$. In order to preserve the similarity of the graph $\mathbf{G}$, we should minimize the *graph preserving criterion* as follows [4]:

$$\mathbf{q}^* = \arg \min_{\substack{\mathbf{q} \\ \mathbf{q}^T\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{q}=c}} \mathbf{q}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{q} \tag{1}$$

where c is a constant, $\mathbf{B}$ is the constraint matrix and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix. $\mathbf{D}$ is the diagonal matrix with diagonal element $\mathbf{D}_{ii} = \sum_{j\neq i} \mathbf{W}_{ij} \forall i$.

In supervised learning problem, Supervised Graph Embedding is to extract the graph embeddings which best characterize the compactness of the intra-class graph and the separability of the inter-class graph. Therefore, in order to provide the robust graph embeddings, we should maximize the following criterion:

$$\mathbf{q}^* = \arg \max_{\mathbf{q}} \frac{\mathbf{q}^T\mathbf{X}\mathbf{L}^b\mathbf{X}^T\mathbf{q}}{\mathbf{q}^T\mathbf{X}\mathbf{L}^w\mathbf{X}^T\mathbf{q}} \tag{2}$$

where $\mathbf{L}^b = \mathbf{D}^b - \mathbf{W}^b$ and $\mathbf{L}^w = \mathbf{D}^w - \mathbf{W}^w$ are the Laplacian matrices of the *inter-class graph* $\mathbf{G}^b$ and the *intra-class graph* $\mathbf{G}^w$ respectively. Both $\mathbf{D}^b$ and $\mathbf{D}^w$ are the diagonal matrices with diagonal element $\mathbf{D}^b_{ii} = \sum_{j\neq i} \mathbf{W}^b_{ij} \forall i$ and $\mathbf{D}^w_{ii} = \sum_{j\neq i} \mathbf{W}^w_{ij} \forall i$ respectively.

3 General Solution for Supervised Graph Embedding

As we know in Section 2, the criterion of Supervised Graph Embedding is:

$$\mathrm{J}(\mathbf{q}) = \frac{\mathbf{q}^T\mathbf{X}\mathbf{L}^b\mathbf{X}^T\mathbf{q}}{\mathbf{q}^T\mathbf{X}\mathbf{L}^w\mathbf{X}^T\mathbf{q}} \tag{3}$$

Thus, the optimal vector is $\mathbf{q}^* = \arg\max_{\mathbf{q}} \mathrm{J}(\mathbf{q})$. From Eq. (3), we can find that the criterion of Supervised Graph Embedding is a Generalized Rayleigh Quotient. The optimal $\mathbf{q}^*$ is the top eigenvector of the generalized eigenvalue problem

$$(\mathbf{X}\mathbf{L}^b\mathbf{X}^T)\mathbf{q} = \lambda(\mathbf{X}\mathbf{L}^w\mathbf{X}^T)\mathbf{q} \tag{4}$$

For the convenience of description, we define that $\mathbf{M}_b = \mathbf{X}\mathbf{L}^b\mathbf{X}^T$, $\mathbf{M}_w = \mathbf{X}\mathbf{L}^w\mathbf{X}^T$ and $\mathbf{M}_t = \mathbf{M}_b + \mathbf{M}_w$. Since both $\mathbf{L}^b$ and $\mathbf{L}^w$ are the Laplacian matrices, $\mathbf{M}_b$, $\mathbf{M}_w$ and $\mathbf{M}_t$ are symmetric positive semi-definite.

Since the null space of $\mathbf{M}_t = \mathbf{M}_b + \mathbf{M}_w$ is the intersection of the null space of $\mathbf{M}_b$ and the null space of $\mathbf{M}_w$, the samples are projected onto the range space of $\mathbf{M}_t$ without loss of any discriminant information. Therefore, the criterion of Supervised Graph Embedding can be changed to:

$$\mathrm{J}_1(\mathbf{p}) = \frac{\mathbf{p}^T\tilde{\mathbf{M}}_b\mathbf{p}}{\mathbf{p}^T\tilde{\mathbf{M}}_w\mathbf{p}} \tag{5}$$

where $\tilde{\mathbf{M}}_b = \mathbf{U}_t^T\mathbf{M}_b\mathbf{U}_t$, $\tilde{\mathbf{M}}_w = \mathbf{U}_t^T\mathbf{M}_w\mathbf{U}_t$ and $\mathbf{U}_t$ is the set of eigenvectors, corresponding to all the nonzero eigenvalues, of $\mathbf{M}_t$. Then, the optimal discriminant transformation vector is $\mathbf{q}^* = \mathbf{U}_t\mathbf{p}^*$ where $\mathbf{p}^* = \arg\max_{\mathbf{p}} \mathrm{J}_1(\mathbf{p})$.

Let λ_i^t and $\mathbf{u}_i^t$ be the ith eigenvalue and the corresponding eigenvector of $\mathbf{M}_t$, $\lambda_1^t \geq \cdots \geq \lambda_{\ell_1^t}^t$, $\mathbf{\Sigma}_t = diag(\lambda_1^t, \cdots, \lambda_{\ell_1^t}^t)$, $\mathbf{U}_t = [\mathbf{u}_1^t, \cdots, \mathbf{u}_{\ell_1^t}^t]$ and $\ell_1^t = rank(\mathbf{M}_t)$. Then, $\tilde{\mathbf{M}}_t = \mathbf{U}_t^T\mathbf{M}_t\mathbf{U}_t = \mathbf{\Sigma}_t$. Let $\bar{\lambda}_i^b$ and $\bar{\mathbf{u}}_i^b$ be the ith eigenvalue and the corresponding eigenvector of $\mathbf{\Sigma}_t^{-1/2}\tilde{\mathbf{M}}_b\mathbf{\Sigma}_t^{-1/2}$, $\bar{\lambda}_1^b \geq \cdots \geq \bar{\lambda}_{\ell_1^t}^b \geq 0$, $\bar{\mathbf{\Sigma}}_b = diag(\bar{\lambda}_1^b, \cdots, \bar{\lambda}_{\ell_1^t}^b)$ and $\bar{\mathbf{U}}_b = [\bar{\mathbf{u}}_1^b, \cdots, \bar{\mathbf{u}}_{\ell_1^t}^b]$.

Theorem 1. $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b$ *simultaneously diagonalizes* $\tilde{\mathbf{M}}_b$, $\tilde{\mathbf{M}}_w$ *and* $\tilde{\mathbf{M}}_t$.

Proof. Due to $\bar{\mathbf{U}}_b^T\bar{\mathbf{U}}_b = \mathbf{I}$, we can obtain

$$(\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b)^T\tilde{\mathbf{M}}_t\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b = \bar{\mathbf{U}}_b^T\bar{\mathbf{U}}_b = \mathbf{I} \tag{6}$$

Because of $\tilde{\mathbf{M}}_t = \tilde{\mathbf{M}}_b + \tilde{\mathbf{M}}_w$, we can rewrite Eq. (6) as:

$$(\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b)^T\tilde{\mathbf{M}}_t\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b = \bar{\mathbf{\Sigma}}_b + (\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b)^T\tilde{\mathbf{M}}_w\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b = \mathbf{I}$$

Then, we can obtain

$$(\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b)^T\tilde{\mathbf{M}}_w\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b = \mathbf{I} - \bar{\mathbf{\Sigma}}_b = \bar{\mathbf{\Sigma}}_w \tag{7}$$

where $\bar{\mathbf{\Sigma}}_w = diag(\bar{\lambda}_1^w, \cdots, \bar{\lambda}_{\ell_1^t}^w)$. Due to $\mathbf{I} - \bar{\mathbf{\Sigma}}_b = \bar{\mathbf{\Sigma}}_w$, we can find that $\bar{\mathbf{\Sigma}}_w = diag(1 - \bar{\lambda}_1^b, \cdots, 1 - \bar{\lambda}_{\ell_1^t}^b)$, $1 \geq \bar{\lambda}_i^b, \bar{\lambda}_i^w \geq 0$ and $\bar{\lambda}_1^w \leq \cdots \leq \bar{\lambda}_{\ell_1^t}^w$. □

Theorem 2. $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b$ *is the optimal discriminant transformation which maximizes the following criterion:*

$$\mathrm{J}_2(\mathbf{p}) = \frac{\mathbf{p}^T\tilde{\mathbf{M}}_b\mathbf{p}}{\mathbf{p}^T\tilde{\mathbf{M}}_t\mathbf{p}} \tag{8}$$

Proof. From Eq. (8), we can find that the optimal $\mathbf{p}^*$, maximizing J_2, is the top eigenvector of the eigenvalue problem

$$\tilde{\mathbf{M}}_b\mathbf{p} = \lambda\tilde{\mathbf{M}}_t\mathbf{p} \overset{\tilde{\mathbf{M}}_t=\mathbf{\Sigma}_t}{\Rightarrow} \mathbf{\Sigma}_t^{-1/2}\tilde{\mathbf{M}}_b\mathbf{p} = \lambda\mathbf{\Sigma}_t^{1/2}\mathbf{p} \tag{9}$$

Let $\mathbf{p}' = \mathbf{\Sigma}_t^{1/2}\mathbf{p}$, we can obtain

$$\mathbf{\Sigma}_t^{-1/2}\tilde{\mathbf{M}}_b\mathbf{\Sigma}_t^{-1/2}\mathbf{p}' = \lambda\mathbf{p}' \tag{10}$$

Since $\bar{\lambda}_i^b$ and $\bar{\mathbf{u}}_i^b$ are the ith eigenvalue and the corresponding eigenvector of Eq. (10), $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b$ maximizes the criterion J_2. □

Theorem 3. *When* $\tilde{\mathbf{M}}_w$ *is nonsingular,* $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b$ *is the optimal discriminant transformation which maximizes the criterion* J_1 *in Eq. (5).*

Proof. From Eq. (5), we can find that the optimal $\mathbf{p}^*$, maximizing J_1, is the top eigenvector of the eigenvalue problem

$$\tilde{\mathbf{M}}_b\mathbf{p} = \lambda\tilde{\mathbf{M}}_w\mathbf{p} \tag{11}$$

Adding both sides of Eq. (11) by $\lambda\tilde{\mathbf{M}}_b\mathbf{p}$, we can find

$$\tilde{\mathbf{M}}_b\mathbf{p} = \frac{\lambda}{1+\lambda}\tilde{\mathbf{M}}_t\mathbf{p} = \mu\tilde{\mathbf{M}}_t\mathbf{p} \tag{12}$$

Since both $\tilde{\mathbf{M}}_w$ and $\tilde{\mathbf{M}}_t$ are nonsingular, we can find that Eq. (11) and Eq. (12) share the same eigenvector with different eigenvalues. Therefore, the optimal discriminant transformation, maximizing the criterion J_1, is $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b$. □

Since $\tilde{\mathbf{M}}_w$ is singular in the case of the SSS problem, we can't directly computed the eigenvector $\mathbf{v}$ in the following generalized eigenvalue problem

$$\tilde{\mathbf{M}}_b\mathbf{v} = \lambda\tilde{\mathbf{M}}_w\mathbf{v} \tag{13}$$

Theorem 4. *When* $\tilde{\mathbf{M}}_w$ *is singular,* $\lambda_i = \bar{\lambda}_i^b/\bar{\lambda}_i^w$ *and* $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{u}}_i^b$ *are the* i*th eigenvalue and the corresponding eigenvector in Eq. (13).*

Proof. Let $rank(\tilde{\mathbf{M}}_w) = \ell_1^w < \ell_1^t$, $\ell_2^w = \ell_1^t - \ell_1^w$, $rank(\tilde{\mathbf{M}}_b) = \ell_1^b \le \ell_1^t$ and $\ell_2^b = \ell_1^t - \ell_1^b$. According to **Theorem** 1, $\bar{\mathbf{\Sigma}}_w$ and $\bar{\mathbf{\Sigma}}_b$ can be rewritten as

$$\bar{\mathbf{\Sigma}}_w = diag(\bar{\lambda}_1^w, \bar{\lambda}_2^w, \cdots, \bar{\lambda}_{\ell_1^t}^w) = \begin{bmatrix} \mathbf{0}_w & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_w \end{bmatrix}$$

$$\bar{\mathbf{\Sigma}}_b = \mathbf{I} - \bar{\mathbf{\Sigma}}_w = diag(\bar{\lambda}_1^b, \bar{\lambda}_2^b, \cdots, \bar{\lambda}_{\ell_1^t}^b) = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_b \end{bmatrix}$$

where $\mathbf{0}_w \in \mathbb{R}^{\ell_2^w \times \ell_2^w}$ and $\mathbf{0}_b \in \mathbb{R}^{\ell_2^b \times \ell_2^b}$ are zero matrices, $\mathbf{I}_w \in \mathbb{R}^{\ell_2^b \times \ell_2^b}$ and $\mathbf{I}_b \in \mathbb{R}^{\ell_2^w \times \ell_2^w}$ are identity matrices, $\mathbf{D}_w = diag\left(\bar{\lambda}_{\ell_2^w+1}^w, \bar{\lambda}_{\ell_2^w+2}^w, \cdots, \bar{\lambda}_{\ell_1^b}^w\right)$ and $\mathbf{D}_b = diag\left(\bar{\lambda}_{\ell_2^w+1}^b, \bar{\lambda}_{\ell_2^w+2}^b, \cdots, \bar{\lambda}_{\ell_1^b}^b\right)$ are diagonal matrices, $0 < \bar{\lambda}_{\ell_2^w+1}^w \leq \bar{\lambda}_{\ell_2^w+2}^w \leq \cdots \leq \bar{\lambda}_{\ell_1^b}^w < 1$ and $1 > \bar{\lambda}_{\ell_2^w+1}^b \geq \bar{\lambda}_{\ell_2^w+2}^b \geq \cdots \geq \bar{\lambda}_{\ell_1^b}^b > 0$.

Let $\mathbf{V} = \mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{U}}_b = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{\ell_1^t}]$. According to **Theorem** 1, $\mathbf{V}^T\tilde{\mathbf{M}}_w\mathbf{V} = \bar{\mathbf{\Sigma}}_w$ and $\mathbf{V}^T\tilde{\mathbf{M}}_b\mathbf{V} = \bar{\mathbf{\Sigma}}_b$. Then, $\tilde{\mathbf{M}}_w\mathbf{V} = (\mathbf{V}^{-1})^T\bar{\mathbf{\Sigma}}_w$ and $\tilde{\mathbf{M}}_b\mathbf{V} = (\mathbf{V}^{-1})^T\bar{\mathbf{\Sigma}}_b$. Let $\mathbf{V}' = (\mathbf{V}^{-1})^T = [\mathbf{v}'_1, \mathbf{v}'_2, \cdots, \mathbf{v}'_{\ell_1^t}]$, we can find, for $i = 1, \cdots, \ell_1^t$

$$\tilde{\mathbf{M}}_w\mathbf{v}_i = \bar{\lambda}_i^w\mathbf{v}'_i \tag{14}$$

$$\tilde{\mathbf{M}}_b\mathbf{v}_i = \bar{\lambda}_i^b\mathbf{v}'_i \tag{15}$$

$$\bar{\lambda}_i^w\tilde{\mathbf{M}}_b\mathbf{v}_i = \bar{\lambda}_i^b\tilde{\mathbf{M}}_w\mathbf{v}_i \tag{16}$$

Then, $\tilde{\mathbf{M}}_b\mathbf{v}_i = (\bar{\lambda}_i^b/\bar{\lambda}_i^w)\tilde{\mathbf{M}}_w\mathbf{v}_i$. Thus, $\lambda_i = \bar{\lambda}_i^b/\bar{\lambda}_i^w$ and $\mathbf{\Sigma}_t^{-1/2}\bar{\mathbf{u}}_i^b$ are the ith eigenvalue and the corresponding eigenvector in Eq. (13). □

Since $\tilde{\mathbf{M}}_w$ is symmetric positive semi-definite and $\bar{\lambda}_1^w = \cdots = \bar{\lambda}_{\ell_2^w}^w = 0$, $\mathbf{v}_i$ lies in the null space of $\tilde{\mathbf{M}}_w$ for $i = 1, \cdots, \ell_2^w$ according to Eq. (14). Due to $0 < \bar{\lambda}_{\ell_2^w+1}^w \leq \cdots \leq \bar{\lambda}_{\ell_1^b}^w < 1$, $\mathbf{v}_i$ doesn't lie in the null space of $\tilde{\mathbf{M}}_w$ for $i = \ell_2^w + 1, \cdots, \ell_1^b$. According to the definition of λ_i, we can find that $\lambda_i = +\infty$ for $i = 1, \cdots, \ell_2^w$, $0 < \lambda_i < +\infty$ for $i = \ell_2^w + 1, \cdots, \ell_1^b$ and $\lambda_{\ell_2^w+1} \geq \cdots \geq \lambda_{\ell_1^b} > 0$. Due to $\mathrm{J}_1(\mathbf{v}_i) = \lambda_i$, the robust discriminant transformation vectors can be first extracted from the null space of $\tilde{\mathbf{M}}_w$, and then from the range space of $\tilde{\mathbf{M}}_w$.

When transformation vector $\tilde{\mathbf{v}}$ lies in the null space of $\tilde{\mathbf{M}}_w$ not in the null space of $\tilde{\mathbf{M}}_b$, $\mathrm{J}_1(\tilde{\mathbf{v}}) = +\infty$, which means that the null space of $\tilde{\mathbf{M}}_w$ is an effective discriminant space with respect to the criterion J_1. Assume that both $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ lie in the null space of $\tilde{\mathbf{M}}_w$, $\tilde{\mathbf{v}}_1^T\tilde{\mathbf{M}}_b\tilde{\mathbf{v}}_1 > 0$ and $\tilde{\mathbf{v}}_2^T\tilde{\mathbf{M}}_b\tilde{\mathbf{v}}_2 > 0$; however, $\mathrm{J}_1(\tilde{\mathbf{v}}_1) = \mathrm{J}_1(\tilde{\mathbf{v}}_2) = +\infty$. That means every vector, lying in the null space of $\tilde{\mathbf{M}}_w$ not in the null space of $\tilde{\mathbf{M}}_b$, can make J_1 infinite. In this case, the criterion J_1 can't characterize the separability of the inter-class graph. Therefore, we should replace J_1. In this paper, we extract the robust discriminant transformation vectors from the null space of $\tilde{\mathbf{M}}_w$ based on the following criterion:

$$\mathrm{J}_3(\mathbf{p}) = \frac{\mathbf{p}^T\tilde{\mathbf{M}}_b\mathbf{p}}{\mathbf{p}^T\mathbf{p}} \tag{17}$$

where $\mathbf{p}^T\mathbf{p}$ normalizes $\mathbf{p}^T\tilde{\mathbf{M}}_b\mathbf{p}$ so that the optimal vector can't be selected randomly. Therefore, the optimal discriminant transformation vector $\mathbf{p}^*$ is

$$\mathbf{p}^* = \arg \max_{\substack{\mathbf{p} \\ \mathbf{p}^T\tilde{\mathbf{M}}_w\mathbf{p}=0}} \mathrm{J}_3(\mathbf{p}) \tag{18}$$

Let $\tilde{\lambda}_i^w$ and $\tilde{\mathbf{u}}_i^w$ be the ith eigenvalue and the corresponding eigenvector of $\tilde{\mathbf{M}}_w$, $0 \leq \tilde{\lambda}_1^w \leq \cdots \leq \tilde{\lambda}_{\ell_1^t}^w$. We define $\tilde{\mathbf{U}}_w^a = [\tilde{\mathbf{u}}_1^w, \cdots, \tilde{\mathbf{u}}_{\ell_2^w}^w]$ and $\tilde{\mathbf{U}}_w^b = [\tilde{\mathbf{u}}_{\ell_2^w+1}^w, \cdots, \tilde{\mathbf{u}}_{\ell_1^t}^w]$.

Since the vector $\mathbf{p}$ in Eq. (18) lies in the null space of $\tilde{\mathbf{M}}_w$, let $\mathbf{p} = \tilde{\mathbf{U}}_w^a \mathbf{z}$. Therefore, the objective function in Eq. (18) can be changed to

$$\mathbf{z}^* = \arg\max_{\mathbf{z}} \frac{\mathbf{z}^T (\tilde{\mathbf{U}}_w^a)^T \tilde{\mathbf{M}}_b \tilde{\mathbf{U}}_w^a \mathbf{z}}{\mathbf{z}^T \mathbf{z}} \tag{19}$$

Thus, the optimal $\mathbf{z}^*$ is the top eigenvector of the eigenvalue problem $(\tilde{\mathbf{U}}_w^a)^T \tilde{\mathbf{M}}_b \tilde{\mathbf{U}}_w^a$. Let the column vectors $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{\ell_2^w}$ be the leading eigenvectors of the eigenvalue problem $(\tilde{\mathbf{U}}_w^a)^T \tilde{\mathbf{M}}_b \tilde{\mathbf{U}}_w^a$. Thus, the optimal discriminant transformation of J_3 in Eq. (17) is $\mathbf{P}_1^* = \tilde{\mathbf{U}}_w^a [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{\ell_2^w}]$.

When the extracted transformation vector lies in the range of $\tilde{\mathbf{M}}_w$, the objective function is

$$\max_{\substack{\mathbf{p} \\ \mathbf{p}^T \tilde{\mathbf{M}}_w \mathbf{p} \neq 0}} \frac{\mathbf{p}^T \tilde{\mathbf{M}}_b \mathbf{p}}{\mathbf{p}^T \tilde{\mathbf{M}}_w \mathbf{p}} \tag{20}$$

Since the vector $\mathbf{p}$ in Eq. (20) lies in the range space of $\tilde{\mathbf{M}}_w$, let $\mathbf{p} = \tilde{\mathbf{U}}_w^b \mathbf{z}'$. Therefore, Eq. (20) can be rewritten as:

$$\max_{\substack{\mathbf{z}' \\ \mathbf{z}'^T (\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_w \tilde{\mathbf{U}}_w^b \mathbf{z}' \neq 0}} \frac{\mathbf{z}'^T (\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_b \tilde{\mathbf{U}}_w^b \mathbf{z}'}{\mathbf{z}'^T (\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_w \tilde{\mathbf{U}}_w^b \mathbf{z}'} \tag{21}$$

Due to $(\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_w \tilde{\mathbf{U}}_w^b = diag(\tilde{\lambda}_{\ell_2^w+1}^w, \cdots, \tilde{\lambda}_{\ell_1^t}^w)$ where $0 < \tilde{\lambda}_{\ell_2^w+1}^w \leq \cdots \leq \tilde{\lambda}_{\ell_1^t}^w$, $(\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_w \tilde{\mathbf{U}}_w^b$ is nonsingular. Therefore, the optimal $\mathbf{z}'^*$ is the top eigenvector of $((\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_w \tilde{\mathbf{U}}_w^b)^{-1} (\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_b \tilde{\mathbf{U}}_w^b$. Let the column vectors $\mathbf{z}'_1, \mathbf{z}'_2, \cdots, \mathbf{z}'_{\ell_1^w}$ be the leading eigenvectors of the eigenvalue problem $((\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_w \tilde{\mathbf{U}}_w^b)^{-1} (\tilde{\mathbf{U}}_w^b)^T \tilde{\mathbf{M}}_b \tilde{\mathbf{U}}_w^b$. Thus, the optimal discriminant transformation of the objective function in Eq. (20) is $\mathbf{P}_2^* = \tilde{\mathbf{U}}_w^b [\mathbf{z}'_1, \mathbf{z}'_2, \cdots, \mathbf{z}'_{\ell_1^w}]$.

Based on the analysis of Eq. (16), we have known that the robust discriminant transformation vectors can be first extracted from the null space of $\tilde{\mathbf{M}}_w$, and then from the range space of $\tilde{\mathbf{M}}_w$. Therefore, the robust discriminant transformation of Supervised Graph Embedding is $\mathbf{Q} = \mathbf{U}_t [\mathbf{P}_1^*, \mathbf{P}_2^*]$ when $\tilde{\mathbf{M}}_w$ is singular.

4 Experiments

In order to validate GSSGE, we apply GSSGE for computing the discriminant transformation of Local Discriminant Embedding (LDE) [2]. Extensive experiments on FERET [10] are performed to demonstrate the effectiveness and robustness of GSSGE. Since ULDA/GSVD [6], Exact Algorithm [7], LDA/FKT [8] and LDA/GSVD [9] are essentially equivalent, we only compare ULDA/GSVD with other algorithms without loss of generality. Therefore, the system performance of GSSGE is compared to the ones of classical LDA [1], ULDA/GSVD, LDE [2] and LSDA [3]. Since the dimension of the facial image is often very high,

Fig. 1. Twenty facial images of ten individuals in the FERET database

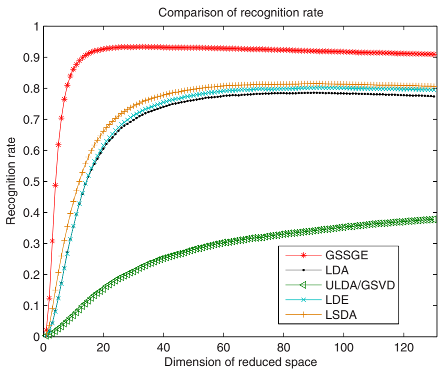

Fig. 2. Recognition rate versus dimension m of reduced space

which can result in the SSS problem, the experimental results can demonstrate the robustness of each method for dealing with the SSS problem.

This dataset consists of all the 1195 people from the FERET Fa/Fb data set. There are two face images for each person. We preprocessed these original images by aligning transformation and scaling transformation so that the two eyes were aligned at the same position. Then, the facial areas were cropped into the resulting images. The size of each cropped image is 64×64, with 256 grey levels per pixel. We didn't perform further preprocessing. Fig. 1 shows the twenty facial images of ten individuals from this dataset.

We selected randomly 495 people for training and used the remaining 700 people as testing. For each testing people, one face image is in the gallery and the other is for probe. Thus, this dataset has no overlap between the training set and gallery/probe set, which results in the requirement of generalizable ability from known objects in the training set to unknown objects in the gallery/probe set for each method. Therefore, the result on the dataset from the FERET database is convincing to evaluate the robustness of each method. We performed 50 times to choose randomly the training set. The final result is the average recognition rate over 50 random training sets. Fig. 2 illustrates the plot of recognition rate versus the dimension m of reduced space for LDA, ULDA/GSVD, LDE, LSDA and GSSGE. From the experimental results, we can find that GSSGE is superior to the other methods. However, it is obvious that ULDA/GSVD works badly.

5 Conclusions

In this paper, we first show that supervised graph embedding algorithms share a general criterion (Generalized Rayleigh Quotient). Through thorough perspective to Generalized Rayleigh Quotient, we propose a general solution, called *General Solution for Supervised Graph Embedding (GSSGE)*, to extract the robust discriminant transformation. Experimental results on FERET database demonstrate the effectiveness and robustness of GSSGE. Furthermore, because our proposed GSSGE is a general solution for Supervised Graph Embedding, GSSGE can be used to extract the robust discriminant transformation of Supervised Graph Embedding algorithms, such as LDA [1] and LSDA [3], and so on.

Acknowledgments. This work was supported by the National High-Tech Research and Development Plan of China under Grant No. 20060101Z1059, and the National Basic Research Program of China under Grant No. 2006CB708303.

References

1. Belhumeur, P.N., Hepanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19(7), 711–720 (1997)
2. Chen, H.-T., Chang, H.-W., Liu, T.-L.: Local discriminant embedding and its variants. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 846–853 (2005)
3. Cai, D., He, X., Zhou, K., Han, J., Bao, H.: Locality Sensitive Discriminant Analysis. In: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (2007)
4. Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., Lin, S.: Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. IEEE Trans. Pattern Anal. Machine Intell. 29(1), 40–51 (2007)
5. You, Q., Zheng, N., Du, S., Wu, Y.: Neighborhood Discriminant Projection for Face Recognition. In: Proceedings of the International Conference on Pattern Recognition, vol. 2, pp. 532–535 (2006)
6. Ye, J., Janardan, R., Li, Q., Park, H.: Feature Reduction via Generalized Uncorrelated Linear Discriminant Analysis. IEEE Trans. Knowledge Data Eng. 18(10), 1312–1322 (2006)
7. Ye, J., Janardan, R., Park, C.H., Park, H.: An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Trans. Pattern Anal. Machine Intell. 26(8), 982–994 (2004)
8. Zheng, S., Sim, T.: When Fisher meets Fukunage-Koontz: A New Look at Linear Discriminant. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 323–329 (2006)
9. Howland, P., Park, H.: Generalizing discriminant analysis using the generalized singular value decomposition. IEEE Trans. Pattern Anal. Machine Intell. 26(8), 995–1006 (2004)
10. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Machine Intell. 22(10), 1090–1104 (2000)