Appearance-based Face Recognition Using A Supervised Manifold Learning Framework

Bogdan Raducanu
Computer Vision Center
Bellaterra, Barcelona, Spain
bogdan@cvc.uab.es

Fadi Dornaika[1,2]
[1] Dept. of Computer Science and Artificial Intelligence
Univ. of the Basque Country UPV/EHU, San Sebastian, Spain
[2] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Abstract

Many natural image sets, depicting objects whose appearance is changing due to motion, pose or light variations, can be considered samples of a low-dimension nonlinear manifold embedded in the high-dimensional observation space (the space of all possible images). The main contribution of our work is represented by a Supervised Laplacian Eigemaps (S-LE) algorithm, which exploits the class label information for mapping the original data in the embedded space. Our proposed approach benefits from two important properties: i) it is discriminative, and ii) it adaptively selects the neighbors of a sample without using any predefined neighborhood size. Experiments were conducted on four face databases and the results demonstrate that the proposed algorithm significantly outperforms many linear and non-linear embedding techniques. Although we've focused on the face recognition problem, the proposed approach could also be extended to other category of objects characterized by large variance in their appearance.

1. Introduction

Most recognition systems using linear dimensionality reduction methods are bound to ignore the intrinsic structure of the manifold, and this is the bottleneck for achieving highly accurate recognition rates. The reason of failure is twofold: (i) they don't exploit the local topology, performing a holistic analysis instead; and (ii) they often suffer of the 'Small Sample Size' problem when visual data are considered. During the last few years, a large number of approaches have been proposed for constructing and computing an embedded subspace by finding an explicit or non-explicit mapping that projects the original data to a new space of lower dimensionality [9, 13]. These methods can be grouped in two families: linear and non-linear approaches.

The classical linear embedding methods (e.g., PCA, LDA, Maximum Margin Criterion (MMC)[8]) and Locally LDA [7] are demonstrated to be computationally efficient and suitable for practical applications, such as pattern classification and visual recognition. [11] proposed a linear discriminant method called Average Neighborhood Margin Maximization (ANMM). It associates to every sample a margin that is set to the difference between the average distance to heterogenous neighbors and the average distance to the homogeneous neighbors. The linear transform is then derived by maximizing the sum of the margins in the embedded space.

The non-linear methods such as Locally Linear Embedding (LLE)[10], Laplacian Eigenmaps [1], and Isomap focus on preserving the local structure of data. LLE formulates the manifold learning problem as a neighborhood-preserving embedding, which learns the global structure by exploiting the local symmetries of linear reconstructions. Isomap extends the classical Multidimensional Scaling (MDS) [2] by computing the pairwise distances in the geodesic space of the manifold. Essentially, Isomap attempts to preserve geodesic distances when data are embedded in the new low dimensional space. Based on the spectral decomposition of the graph Laplacian, Laplacian Eigenmaps actually try to find Laplacian eigenfunction on the manifold. Maximum Variance Unfolding (MVU) [12] is a global algorithm for nonlinear dimensionality reduction, in which all the data pairs, nearby and far, are considered. MVU attempts to 'unfold' a dataset by pulling the input patterns as far apart as possible subject to the constraints that distances and angles between neighboring points are strictly preserved.

The non-linear embedding methods have been successfully applied to some standard data sets and generated satisfying results in dimensionality reduction and manifold visualization. However, most of these approaches does not take into account the discriminant information that is usually available for many real world problems. Therefore, the application of these methods can be very satisfactory in terms of dimensionality reduction and visualization but

can be fair for classification tasks.

In [4], the authors propose a supervised version of Isomap. This version replaces pairwise Euclidean distances by a dissimilarity function that increases if the pair is heterogeneous and decreases otherwise. Since this algorithm is inherited from Isomap it suffers from the same disadvantage in the sense that outlier samples can give rise to an unwanted embedded space. In [6], the authors exploit label information to improve Laplacian Eigenmaps. The proposed method affects the computation of the affinity matrix entries in the sense that an homogeneous pair of neighbors will have large value and heterogeneous pairs of neighbors will have a small value. Although, the authors show some performance improvement, the proposed method has two drawbacks. First, there is no guarantee that the heterogenous samples will be pushed away from each other. Second, the method has at least three parameters to be tuned.

The main contribution of our work is represented by a Supervised Laplacian Eigemaps(S-LE) algorithm, which exploits the class label information for mapping the original data in the embedded space. The use of labels allows us to split graph Laplacian associated to the data in two components: within-class graph and between-class graph. Our proposed approach benefits from two important properties: i) it is discriminative - by implicitly encoding the large margin concept (pushing close homogeneous samples to each other and pulling close heterogeneous sample far apart), and ii) selects adaptively the neighbors around a sample, by exploiting the statistical significance of the data.

The combination between locality preserving property (inherited from the classical LE[1]) and the discriminative property (due to the large margin concept) represents a clear advantage for S-LE, compared with other non-linear embedding techniques, because it finds a mapping which maximizes the distance between data samples from different classes at each local area. In other words, it maps the points in an embedded space where close data with similar labels fall close to each other and where the data from different classes fall far apart.

The automatic and adaptive selection of the neighbors represents also an added value to our algorithm. It is well known that a sensitive matter affecting non-linear embedding techniques is represented by the adequate choice for neighborhood size.

The paper is structured as follows. In section 2 we give a brief review of the classical LE algorithm. Section 3 introduces the newly proposed version of supervised LE. In section 4 we report the experimental results obtained on 4 public face data sets. Finally, section 5 contains our conclusions and the guidelines for future work.

[1]By classical Laplacian Eigenmaps we refer to the algorithm introduced in [1].

2. Review of Laplacian Eigenmaps

Laplacian Eigenmaps is a recent non-dimensionality reduction techniques that aims to preserve the local structure of data [1]. Using the notion of the Laplacian of the graph, this non-supervised algorithm computes a low-dimensional representation of the data set by optimally preserving local neighborhood information in a certain sense. We assume that we have a set of N samples $\{\mathbf{y}_i\}_{i=1}^N \subset \mathbb{R}^D$. Let's define a neighborhood graph on these data, such as a K-nearest-neighbor or ϵ-ball graph, or a full mesh, and weigh each edge $\mathbf{y}_i \sim \mathbf{y}_j$ by a symmetric affinity function $W_{ij} = K(\mathbf{y}_i; \mathbf{y}_j)$, typically Gaussian:

$$W_{ij} = \exp(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\beta}) \tag{1}$$

where β is usually set to the average of squared distances between all pairs.

We seek latent points $\{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^L$ that minimizes $\frac{1}{2}\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 W_{ij}$, which discourages placing far apart latent points that correspond to similar observed points. If $\mathbf{W} \equiv W_{ij}$ denotes the symmetric affinity matrix and $\mathbf{D}$ is the diagonal weight matrix, whose entries are column (or row, since $\mathbf{W}$ is symmetric) sums of $\mathbf{W}$, then the Laplacian matrix is given $\mathbf{L} = \mathbf{D} - \mathbf{W}$. It can be shown that the objective function can also be written as:

$$\frac{1}{2}\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 W_{ij} = tr(\mathbf{Z}^T \mathbf{L}\, \mathbf{Z}) \tag{2}$$

where $\mathbf{Z} = [\mathbf{x}_1^T; \ldots; \mathbf{x}_N^T]$ is the $N \times L$ embedding matrix. The i^{th} row of the matrix $\mathbf{Z}$ provides the vector $\mathbf{x}_i$—the embedding coordinates of the sample $\mathbf{y}_i$.

The embedding matrix $\mathbf{Z}$ is the solution of the optimization problem:

$$\min_{\mathbf{Z}} tr(\mathbf{Z}^T \mathbf{L}\, \mathbf{Z}) \;\; s.t. \;\; \mathbf{Z}^T \mathbf{D}\, \mathbf{Z} = \mathbf{I}, \;\; \mathbf{Z}^T \mathbf{L}\, \mathbf{e} = \mathbf{0} \tag{3}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{e} = (1, \ldots, 1)^T$. The first constraint eliminates the trivial solution $\mathbf{Z} = \mathbf{0}$ (by setting an arbitrary scale) and the second constraint eliminates the trivial solution $\mathbf{e}$ (all samples are mapped to the same point). Standard methods show that the embedding matrix is provided by the matrix of eigenvectors corresponding to the smallest eigenvalues of the following generalized eigenvector problem:

$$\mathbf{L}\, \mathbf{z} = \lambda\, \mathbf{D}\, \mathbf{z} \tag{4}$$

3. Supervised Laplacian Eigenmaps

While the LE may give good results for non-linear dimensionality reduction, it has not been widely used and

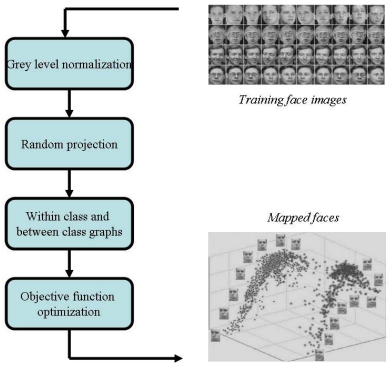

Figure 1. Supervised Laplacian Eigenmaps embedding for the face recognition problem.

assessed for classification tasks [5]. Indeed, many experiments show that the recognition rate in the embedded space is highly depending on the choice of the neighborhood size in the reconstructed graph [13]. Choosing the ideal size in advance can be a very difficult task. Moreover, the introduced mapping by LE does not exploit the discriminant information given by the labels of data. In this section, we present our Supervised LE algorithm which has two interesting properties: i) the obtained embedding respects both discriminant and geometrical structure in data, ii) there is no use of a predefined fixed neighborhood size since the neighbors of every sample are adaptively selected according to the local density and similarity between data samples.

3.1. Two graphs and adaptive neighborhood sets

In order to discover both geometrical and discriminant structure of the data manifold, we split the global graph in two components: the within-class graph G_w and between-class graph G_b. Let $l(\mathbf{y}_i)$ be the class label of $\mathbf{y}_i$. For each data point $\mathbf{y}_i$, we compute two subsets, $N_b(\mathbf{y}_i)$ and $N_w(\mathbf{y}_i)$. $N_w(\mathbf{y}_i)$ contains the neighbors sharing the same label with $\mathbf{y}_i$, while $N_b(\mathbf{y}_i)$ contains the neighbors having different labels. We stress the fact that unlike the classical LE that relies on a fixed neighborhood size, our algorithm selects the sets $N_b(\mathbf{y}_i)$ and $N_w(\mathbf{y}_i)$ according to the local sample point $\mathbf{y}_i$ and its similarities with the rest of samples. To this end, each set is defined for each sample point $\mathbf{y}_i$ and is computed in two consecutive steps. First, the average similarity of the sample $\mathbf{y}_i$ is computed by the total of all similarities with the rest of the data set (Eq. (5)). Second, the sets $N_w(\mathbf{y}_i)$ and $N_b(\mathbf{y}_i)$ are computed using Eqs. (6) and (7), respectively.

$$AS(\mathbf{y}_i) = \frac{1}{N}\sum_{k=1}^{N} \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_k\|^2}{\beta}\right) \tag{5}$$

$$N_w(\mathbf{y}_i) = \{\mathbf{y}_j \mid l(\mathbf{y}_j) = l(\mathbf{y}_i), \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\beta}\right) > AS(\mathbf{y}_i)\} \tag{6}$$

$$N_b(\mathbf{y}_i) = \{\mathbf{y}_j \mid l(\mathbf{y}_j) \neq l(\mathbf{y}_i), \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\beta}\right) > AS(\mathbf{y}_i)\} \tag{7}$$

Equation (6) means that the set of within-class neighbors of the sample $\mathbf{y}_i$, $N_w(\mathbf{y}_i)$, is all data samples that have the same label of $\mathbf{y}_i$ and that have a similarity higher then the average similarity associated with $\mathbf{y}_i$. There is a similar interpretation for the set of between-class neighbors $N_b(\mathbf{y}_i)$. From Equations (6) and (7) it is clear that the neighborhood size is not the same for every data sample. This mechanism adapts the set of neighbors according to the local density and similarity between data samples in the original space.

3.2. Two affinity matrices

Let $\mathbf{W}_w$ and $\mathbf{W}_b$ be the weight matrices of G_w and G_b, respectively. These matrices are defined as:

$$W_{w,ij} = \begin{array}{l} \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\beta}\right) if \mathbf{y}_j \in N_w(\mathbf{y}_i) \; or \; \mathbf{y}_i \in N_w(\mathbf{y}_j) \\ 0, otherwise \end{array}$$

$$W_{b,ij} = \begin{array}{l} \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\beta}\right) if \mathbf{y}_j \in N_b(\mathbf{y}_i) \; or \; \mathbf{y}_i \in N_b(\mathbf{y}_j) \\ 0, otherwise \end{array}$$

3.3. Optimal mapping

Each data sample $\mathbf{y}_i$ is mapped into a vector $\mathbf{x}_i$. The aim is to compute the embedded coordinates $\mathbf{x}_i$ for each data sample. The objective functions are:

$$\min \frac{1}{2}\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 W_{w,ij} \tag{8}$$

$$\max \frac{1}{2}\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2 W_{b,ij} \tag{9}$$

Let the matrix $\mathbf{Z}$ denotes $[\mathbf{x}_1^T; \ldots; \mathbf{x}_N^T]$, it can be shown that the above objective functions can be written as:

$$\min tr(\mathbf{Z}^T \mathbf{L}_w \mathbf{Z}) \tag{10}$$

$$\max tr(\mathbf{Z}^T \mathbf{L}_b \mathbf{Z}) \tag{11}$$

where $\mathbf{L}_w = \mathbf{D}_w - \mathbf{W}_w$ and $\mathbf{L}_b = \mathbf{D}_b - \mathbf{W}_b$.

Using the scale constraint $\mathbf{Z}^T \mathbf{D}_w \mathbf{Z} = \mathbf{I}$ and the equation $\mathbf{L}_w = \mathbf{D}_w - \mathbf{W}_w$, the above two objective functions can be combined into one single objective function:

$$arg\max_{\mathbf{Z}}\{\gamma\, tr(\mathbf{Z}^T\,\mathbf{L}_b\,\mathbf{Z})+(1-\gamma)\, tr(\mathbf{Z}^T\,\mathbf{W}_w\,\mathbf{Z})\} s.t.\, \mathbf{Z}^T\,\mathbf{D}_w\,\mathbf{Z}=\mathbf{I}$$

where $\gamma \in [0,1]$ is a balance parameter. Let the matrix $\mathbf{B}$ denotes $\gamma\,\mathbf{L}_b + (1-\gamma)\,\mathbf{W}_w$, the problem becomes:

$$arg\max_{\mathbf{Z}} tr(\mathbf{Z}^T\,\mathbf{B}\,\mathbf{Z}) \quad s.t.\;\; \mathbf{Z}^T\,\mathbf{D}_w\,\mathbf{Z}=\mathbf{I}$$

This gives the following generalized eigenvalue problem having the following form:

$$\mathbf{B}\,\mathbf{z}=\lambda\,\mathbf{D}_w\,\mathbf{z} \tag{12}$$

Let the column vectors $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_L$ be the generalized eigenvectors of (12) according to their eigenvalue: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. Then, the $N \times L$ embedding matrix $\mathbf{Z} = [\mathbf{x}_1^T; \ldots; \mathbf{x}_N^T]$ will be given by concatenating the obtained eigenvectors $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_L]$.

Figure 2. Some samples in Extended Yale data set.

Figure 3. Some samples in PF01 data set.

Figure 4. Some samples in PIE data set.

4. Experimental results

In this section, we report the experimental results obtained from the application of our proposed algorithm to the problem of face recognition.

Face recognition is one of the most studied problems and a large literature has been devoted to this issue [14]. Face recognition represents an intuitive and non-intrusive method of recognizing people and this is why it became one of the three identification methods used in e-passports and a biometric of choice for many other security applications. Facial image data are often complex to understand and difficult to process due to their high variability in appearance. For this reason, it is mandatory to discover a meaningful low dimensional structure hidden in high dimensional observation data space [3]. Therefore, appearance-based face recognition was usually preceded by a given transform-based dimensionality reduction technique.

4.1. Face data sets

In this study, four face data sets are considered:

1. The UMIST face data set[2]. The UMIST data set contains 575 gray images of 20 different people. The images depict variations in head pose.

2. The Extended Yale Face Database B[3]. It contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions. In our study, a subset of 1800 images has been used. Figure 2 shows some face samples in the Extended Yale Face Database B.

3. The PF01 face data set[4]. It contains the true-color face images of 103 people, 53 men and 50 women, representing 17 various images (1 normal face, 4 illumination variations, 8 pose variations, 4 expression variations) per person. All of the people in the database are Asians. There are three kinds of systematic variations, such as illumination, pose, and expression variations in the database. Some samples are shown in Figure 7.

4. The PIE face data set[5] contains 41,368 images of 68 people. Each person is imaged under 13 different poses, 43 different illumination conditions, and with 4 different expressions. In our study, we used a subset of the original data set, considering 29 images per person. Some samples are shown in Figure 4.

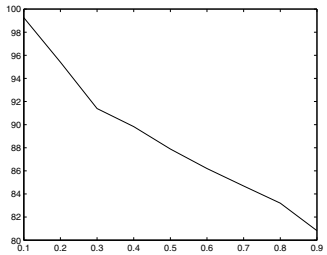

Figure 5. Average recognition rate as function of the blending parameter γ using a part of the UMIST data set.

[2] *http://www.shef.ac.uk/eee/research/vie/research/face.html*

[3] *http://vision.ucsd.edu/ ∼ leekc/ExtYaleDatabase/ExtYaleB.html*

[4] *http://nova.postech.ac.kr/special/imdb/imdb.html*

[5] *http://www.ri.cmu.edu/projects/project_418.html*

4.2. Data preparation

Figure 1 illustrates the main steps of the application of S-LE to the problem of face recognition. The initial face data set is projected on the embedded face subspace using the S-LE algorithm, whose steps have been summarized by a 4 block-diagram. A face image is recognized using the nearest neighbor (NN) classifier applied in this low dimensional space. The parameters of the proposed algorithm are: i) the heat Kernel parameter β, and ii) the parameter γ that balances the impact of within class and between class graphs. In our experiments, the heat Kernel parameter β is set to the average of squared distances between all pairs. The parameter γ is estimated by cross-validation that is carried out on a part of the data set. Figure 5 illustrates the obtained average recognition rate as a function of the parameter γ when a part of the UMIST data set is used as a validation set. Similar behaviors were obtained with other face data sets.

4.3. Evaluation methodology

We have compared our method with six different manifold learning methods, namely PCA, LDA, ANMM, KPCA, Isomap, and LE. For methods relying on neighborhood graphs (ANMM, Isomap, and LE), several trials have been performed in order to choose the optimal neighborhood size. The final values correspond to those giving the best recognition rate.

For each face data set and for every method, we conducted three groups of experiments for which the percentage of training samples was set to 30%, 50%, and 70% of the whole data set, respectively. The remaining data was used for testing. The partition of the data set was done randomly. The best (average) performance obtained by these algorithms, based on ten random spits, are shown in Table 1. The number appearing in parenthesis corresponds to the optimal dimensionality of the embedded subspace (at which the maximum recognition rate has been reported). We can observe that: i) the S-LE outperforms all other methods on all four face data sets, ii) for some databases and some cases LDA method gave better results than the ANMM method, and iii) the performances of S-LE and LDA are comparable for the third group of experiments for which the training/test percentage was set to 70%-30%, However, the difference in performance becomes more significant for the first and second groups of experiments for which the training/test percentage was set to 30%-70% and 50%-50%, respectively, where S-LE clearly outperforms LDA.

For a given embedding method, the recognition rate was computed for several dimensions belonging to $[1, L_{max}]$. For most of the tested methods L_{max} is equal to the number of samples used (except for LDA, where the maximum dimension is equal to the number of classes minus one). Its rate was reported in Table 1. Figures 6 and 7 illustrates the average recognition rate associated with Extended Yale and PF01 data sets, respectively. The maximum dimension depicted in the plots was set to a fraction of L_{max}, in order to guarantee meaningful results. Moreover, we can observe that after a given dimension the recognition rate associated with the three methods PCA, KPCA, and Isomap becomes stable. However, the recognition rate associated with LE and S-LE methods decreases if the number of used eigenvectors becomes large—a general trend associated with many non-linear methods. This means that the last eigenvectors do not have any discriminant information, lacking completely of statistical significance.

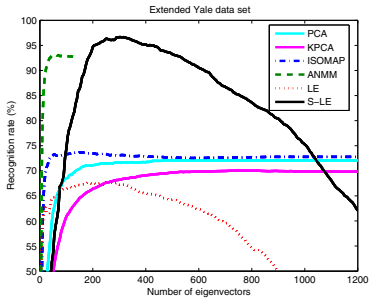

Figure 6. Average recognition rate as function of the number of eigenvectors obtained with Extended YALE data set. The training/test percentage was set to 30%-70%

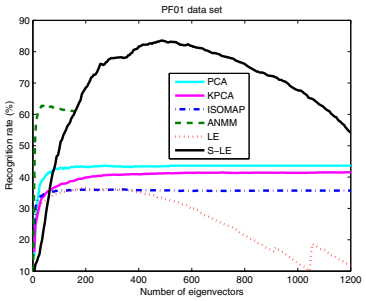

Figure 7. Average recognition rate as function of the number of eigenvectors obtained with PF01 data set. The training/test percentage was set to 50%-50%

5. Conclusions and Future Work

We proposed a novel supervised non-linear dimensionality reduction technique, namely Supervised Laplacian Eigenmaps (S-LE). S-LE learns a mapping by exploiting the class label information. Our algorithm benefits from two important properties: i) it is discriminative, and ii) it selects adaptively the neighbors around a sample. For validation purposes, we applied our method to the face recognition problem. The proposed algorithm was compared with several linear and non-linear embedding techniques. In the future, we will try to find for a given classification task the optimal set of eigenvectors using a feature selection paradigm.

30%-70%	UMIST	Extended Yale	PF01	PIE
PCA	88.08% (45)	72.06% (465)	33.28% (385)	30.76% (370)
LDA	85.35% (10)	79.82% (10)	62.16% (55)	63.38% (65)
ANMM	96.4% (61)	93.0% (51)	54.2% (41)	57.9% (59)
KPCA	89.82% (85)	70.04% (725)	34.91% (940)	39.25% (1030)
Isomap	84.11% (25)	73.69% (125)	31.39% (115)	36.47% (200)
LE	77.88% (40)	67.69% (185)	32.71% (170)	34.97% (330)
S-LE	**99.25%** (15)	**96.65%** (300)	**80.73%** (405)	**89.28%** (70)
50%-50%				
PCA	94.44% (65)	81.73% (395)	43.62% (270)	39.25% (330)
LDA	90.27% (15)	95.94% (25)	80.40% (30)	60.33% (65)
ANMM	99% (23)	95.9% (75)	62.8% (53)	69.7% (59)
KPCA	95.79% (85)	79.40% (820)	41.53% (1180)	50.34% (1190)
Isomap	91.63% (45)	79.23% (165)	36.13% (330)	45.02% (210)
LE	86.52% (40)	74.00% (445)	36.44% (200)	42.09% (385)
S-LE	**99.68%** (15)	**99.92%** (45)	**83.54%** (490)	**93.56%** (85)
70%-30%				
PCA	99.42% (25)	88.15% (540)	48.16% (455)	46.53% (360)
LDA	97.10% (10)	97.55% (25)	89.56% (60)	68.51% (60)
ANMM	98.58% (51)	95.88% (76)	65.40% (55)	76.56% (63)
KPCA	98.49% (155)	83.79% (820)	46.22% (1150)	58.96% (1195)
Isomap	94.68% (20)	81.82% (48)	39.23% (350)	51.26% (180)
LE	90.69% (40)	75.99% (445)	39.10% (340)	46.36% (365)
S-LE	**98.58%** (10)	**99.98%** (40)	**92.28%** (405)	**96.14%** (120)

Table 1. Best recognition accuracy obtained with six face data sets. The training/test percentage was set to 30%-70%, 50%-50%, and 70%-30%.

Acknowledgment. This work was partially supported by the Spanish Government under the project TIN2010-18856.

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. 465, 466

[2] I. Borg and P. Groenen. *Modern Multidimensional Scaling: theory and applications*. Springer-Verlag New York, 2005. 465

[3] B. Draper, K. Baek, M. Bartlett, and J. R. Beveridge. Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, 91:115–137, 2003. 468

[4] X. Geng, D. Zhan, and Z. Zhou. Supervised nonlinear dimensionality reduction for visualization and classification. *IEEE Transactions on systems, man, and cybernetics-part B: cybernetics*, 35:1098–1107, 2005. 466

[5] P. Jia, J. Yin, X. Huang, and D. Hu. Incremental Laplacian Eigenmaps by preserving adjacent information between data points. *Pattern Recognition Letters*, 30(16):1457–1463, 2009. 467

[6] Q. Jiang and M. Jia. Supervised laplacian eigenmaps for machinery fault classification. In *World Congress on Computer Science and Information Engineering*, 2009. 466

[7] T. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3):318–327, 2005. 465

[8] H. Li, T. Jiang, and K. Zhang. Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans. on Neural Networks*, 17(1):157–165, 2006. 465

[9] L. Saul, K. Weinberger, F. Sha, J. Ham, and D. Lee. *Semisupervised Learning*, chapter Spectral methods for dimensionality reduction. MIT Press, Cambridge, MA, 2006. 465

[10] L. K. Saul, S. T. Roweis, and Y. Singer. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003. 465

[11] F. Wang, X. Wang, D. Zhang, C. Zhang, and T. Li. Marginface: A novel face recognition method by average neighborhood margin maximization. *Pattern Recognition*, 42:2863–2875, 2009. 465

[12] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1):77–90, 2006. 465

[13] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extension: a general framework for dimensionality reduction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007. 465, 467

[14] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003. 468