

Conformal Embedding Analysis with Local Graph Modeling on the Unit Hypersphere

Yun Fu, Ming Liu, and Thomas S. Huang
Beckman Institute for Advanced Science and Technologies
University of Illinois at Urbana-Champaign
{yunfu2, mingliu1, huang}@ifp.uiuc.edu

Abstract

We present the Conformal Embedding Analysis (CEA) for feature extraction and dimensionality reduction. Incorporating both conformal mapping and discriminating analysis, CEA projects the high-dimensional data onto the unit hypersphere and preserves intrinsic neighbor relations with local graph modeling. Through the embedding, resulting data pairs from the same class keep the original angle and distance information on the hypersphere, whereas neighboring points of different class are kept apart to boost discriminating power. The subspace learned by CEA is gray-level variation tolerable since the cosine-angle metric and the normalization processing enhance the robustness of the conformal feature extraction. We demonstrate the effectiveness of the proposed method with comprehensive comparisons on visual classification experiments¹.

1. Introduction

Subspace learning and feature extraction are essential techniques to deal with many practical problems in computer vision and pattern recognition [2]. A common way to address these objectives is to seek a generalized projection from the input space to the desired low-dimensional space, with labeling information when available, such that the intrinsic data relations are revealed. Traditional techniques, such as Principal Components Analysis (PCA) [8] and Linear Discriminant Analysis (LDA) [9], and their kernelized variations [23], measure the Euclidean distance between data points and obtain the global projection via an optimization formulation assuming the Gaussian distribution in data space [16]. In general, these approaches are often based upon the assumption that the training data are drawn from the same underlying distribution as the testing

data. Unfortunately, due to the limitation in data collecting and large variations of the sensory inputs, it is usually hard to guarantee that the training data have these desired characteristics with statistically sufficient.

Without ignoring the nice linear properties of the traditional methods, recent studies reveal that local features and intrinsic geometric structures [17, 11] in the input space take on more discriminating power for classification. Such techniques assume that the high-dimensional data can be considered as a set of geometrically related points lying on a smooth low-dimensional manifold [7]. Actually, the object space is usually a sub-manifold of very low dimensionality embedded in the ambient space. The representative nonlinear manifold learning algorithms, such as Locally Linear Embedding (LLE) [10], Isomap [12], Laplacian Eigenmaps (LE) [1], and Semidefinite Programming (SDE) [13], aim at discovering the geometric properties of the data space. Particularly, embedding learning is closely related to preserving the local topological structure of neighborhood connections. Because it remains a difficult issue to map new data to the learned manifold space [28], these methods cannot be easily extended for generalized classification problems.

Recent researches also suggest to use the linear approximation form for the nonlinear methods in the style of Graph Embedding (GE) [14], such as Locality Preserving Projections (LPP) [5], Locally Embedded Analysis (LEA) [3], and Neighborhood Preserving Embedding (NPE) [6]. The basic idea is to observe and model the local manifold structure registered on an affinity graph. However, these algorithms mainly focus on preserving data localities and similarities in the manifold space so that the discriminating power can not be guaranteed sufficiently high. As a result, the projected data points of different classes may still overlap after embedding. To overcome this limitation, it has been shown that the discriminating power can be boosted by Fisher criterion [2] and Kernel trick [23]. Local Discriminant Embedding (LDE) [19] or Locality Sensitive Discriminant Analysis (LSDA) [27] and the 2D/kernel variants achieve good face recognition accuracy by integrating the information of

¹This research was funded in Part by the U.S. Government VACE program, and in part by the NSF Grant CCF 04-26627. The views and conclusions are those of the authors, not of the US Government or its Agencies.

neighbor and class relations between data points. Neighborhood Discriminant Projection (NDP) [26] models the data distribution using both LPP and NPE in accordance with Fisher criterion. Kernel PCA and Kernel LDA show generally good performance for face recognition [23]. Other ways, such as Orthogonal LPP (OLPP) [18], geodesic distance based LDA [25], and extended Isomap [24], can additionally enhance the discriminating capability as well.

In addition to the above perspective, we also note that, in practice, it is often the case that Euclidean metric is incapable of capturing the intrinsic similarities between image data points, especially when the gray-level variation is large, e.g. image underexposed and overexposed cases. Some recent researches have suggested to estimate local angles and distances for the locally isometric embedding [13]. The basic idea is to support the existence of a rotation, reflection and/or translation for the local mapping between data points and their neighbors. Moreover, for distance metric selection, many classifiers rely on the distances between the data in the input space.

1.1. Our Approach

Considering the foregoing discussions, we are motivated to develop a new subspace learning framework, called Conformal Embedding Analysis (CEA), by integrating both conformal embedding nature and discriminating criterion. In particular, CEA can map the high-dimensional data onto the unit hypersphere [20, 21] with a local graph modeling so that the low dimensional representation of the data is provided with robust discriminating power for pattern classification. It is worthwhile to highlight some aspects of CEA algorithm as follows.

1) CEA is a supervised subspace learning method, which is provided with good discriminating power by incorporating the labeling information of neighborhood and class for each *same-class* sub-manifold and *diff-class* sub-manifold. This property conforms to the Fisher Criterion so as to be more reliable for multi-class classification problems.

2) Since the high-dimensional data are modelled and embedded with a local affinity graph in the conformal mapping manner, CEA is based on a different geometric intuition from traditional methods, and as a result, new properties are provided for efficient subspace learning and feature representation.

3) The subspace learned by CEA is gray-level variation tolerable since the cosine-angle metric and the normalization processing enhance the robustness of the conformal feature extraction.

In the following sections, we first present the generic formulation of CEA algorithm in section 2, and then justify the theory in section 3. A kernel version of CEA is formulated in section 4. We demonstrate the effectiveness of CEA in section 5 with experiments and conclude the paper at last.

2. Conformal Embedding Analysis

Assume that the given image samples are denoted as the vector-represented data set $\{\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, n\}$ with high dimensionality and the corresponding class labels $\{l_i \in \{1, \dots, n_c\}, i = 1, \dots, n\}$, where D is equal to the number of image pixels and each datum \mathbf{x}_i belongs to a class l_i . The basic objective of CEA is to find a low dimensional representation $\{\mathbf{z}_i \in \mathbb{R}^d, i = 1, \dots, n\}$ of the original data with discriminant property through a transformation function $\mathcal{F}(\cdot)$, where d denotes the reduced low dimension. For mathematical convenience, we write the original data matrix as $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n] \in \mathbb{R}^{D \times n}$ and the embedded data matrix as $\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2 \dots \mathbf{z}_n] \in \mathbb{R}^{d \times n}$, where $d \ll D$ and $\mathbf{Z} = \mathcal{F}(\mathbf{X})$. The CEA algorithm is stated by the following steps.

- *Preprocessing.* Preprocess each of the original image data $\{\mathbf{x}_i\}_{i=1}^n$ by some particular image processing algorithms, e.g. histogram equalization, and form the matrix \mathbf{Y} for $\{\mathbf{y}_i\}_{i=1}^n$ on each column, which is one-to-one corresponded to $\{\mathbf{x}_i\}_{i=1}^n$.
- *Normalization.* Scale each column vector of \mathbf{Y} to be norm-one to obtain $\{\tilde{\mathbf{y}}_i\}_{i=1}^n$, where $\tilde{\mathbf{y}}_i = \frac{\mathbf{y}_i}{\|\mathbf{y}_i\|}$, and form the normalized $D \times n$ matrix $\tilde{\mathbf{Y}}$. Note that $\tilde{\mathbf{y}}_i$ and \mathbf{x}_i is one-to-one corresponded for $i = 1, \dots, n$.
- *Constructing Conformal Graph.* Define two conformal affinity graphs \mathcal{G}_s and \mathcal{G}_d both with n nodes. The i -th node corresponds to the data \mathbf{x}_i . For \mathcal{G}_s , we only consider each pair of data \mathbf{x}_i and \mathbf{x}_j from the *same* class with $l_i = l_j$. An edge is constructed between nodes i and j if $\tilde{\mathbf{y}}_j$ is among k_s largest conformal neighbors of $\tilde{\mathbf{y}}_i$ and vice versa. For \mathcal{G}_d , we only consider each pair of data \mathbf{x}_i and \mathbf{x}_j from the *different* class with $l_i \neq l_j$. An edge is constructed between nodes i and j if $\tilde{\mathbf{y}}_j$ is among k_d largest conformal neighbors of $\tilde{\mathbf{y}}_i$ and vice versa. Note that k_s and k_d can be different and chosen with empirical values.
- *Choosing Conformal Weights.* Define the $n \times n$ conformal affinity matrix \mathbf{W}_s of graph \mathcal{G}_s and \mathbf{W}_d of graph \mathcal{G}_d . If node i and j are connected, the weight of the edge between \mathbf{x}_i and \mathbf{x}_j is set by $w_{ij} = w_1$. Otherwise, $w_{ij} = w_2$ if node i and j are not connected. The parameter w_1 and w_2 are defined in the next section.
- *Computing Embedded Subspace.* Define the $D \times d$ projection matrix $\mathbf{P} = [\mathbf{p}_1 \dots \mathbf{p}_d]$. Find the subspace bases $\{\mathbf{p}_i\}_{i=1}^d$ as the eigenvectors of

$$\mathbf{A} = (\tilde{\mathbf{Y}}(\mathbf{D}_s - \mathbf{W}_s)\tilde{\mathbf{Y}}^T)^{-1}\tilde{\mathbf{Y}}(\mathbf{D}_d - \mathbf{W}_d)\tilde{\mathbf{Y}}^T, \quad (1)$$

where $\mathbf{D}[i, i] = \sum_j w_{ij}$, that correspond to the d largest eigenvalues of \mathbf{A} . The conformal embedding $\mathcal{F}(\mathbf{x}_i)$ of \mathbf{x}_i is as follows.

$$\mathbf{x}_i \longrightarrow \mathbf{y}_i \longrightarrow \tilde{\mathbf{y}}_i \longrightarrow \mathbf{z}_i = \mathbf{P}^T \tilde{\mathbf{y}}_i = \mathcal{F}(\mathbf{x}_i). \quad (2)$$

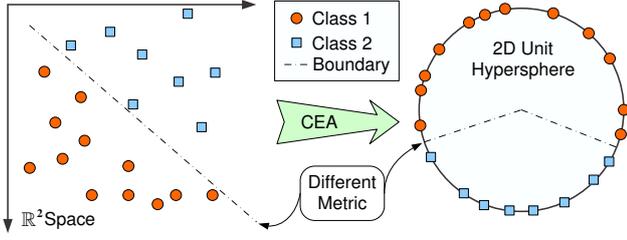


Figure 1. CEA perspective with data normalization.

In the matrix form, we have $\mathbf{Z} = \mathcal{F}(\mathbf{X}) = \mathbf{P}^T \tilde{\mathbf{Y}}$.

- **Largest Conformal Affinity Classification.** For a testing data point $\mathbf{x}_t \in \mathbb{R}^D$, we can project it into the learned subspace by $\mathbf{z}_t = \mathcal{F}(\mathbf{x}_t)$. The classification problem is defined to predict $l_t := l_{i^*}$, satisfying

$$\mathbf{z}_{i^*} = \mathcal{F}(\mathbf{x}_{i^*}) = \arg \max_{\mathbf{z}_i} \mathbf{z}_t^T \mathbf{z}_i \quad (3)$$

3. Theoretical Justifications of CEA

The above transformation from $\{\mathbf{x}_i\}_{i=1}^n$ to $\{\tilde{\mathbf{y}}_i\}_{i=1}^n$ can encode the orientation information of vectors \mathbf{x}_i in the normalized unit vectors $\tilde{\mathbf{y}}_i$. It projects the original data points onto a high-dimensional unit hypersphere for gray-level variation tolerable feature representation. Figure 1 illustrates the CEA perspective on a toy data set. Note that the distance metric is changed from Euclidean to cosine-angle.

Unlike PCA and LDA, our CEA algorithm reflects the local angle relations between neighboring data based on the metric of cosine value. On the other hand, we consider the Fisher Criterion for discriminant embedding, which is not the similar learning manner in LPP and LEA. To obtain an embedding for discriminative subspace learning, we have three aspects of objective. 1) Preserve the same-class conformal affinity while keeping away the diff-class conformal affinity after the embedding. 2) If the two original high-dimensional data are close (large conformal affinity), then the embedded low-dimensional points are close as well. 3) The embedded sub-manifold can better reflect the class relations with respect to the labeling information.

3.1. Conformal Affinity Weights

After preprocessing and normalization, the relationship of any pair of transformed data can be effectively modelled by the cosine-angle value of the two vectors on the unit hypersphere. Since $\tilde{\mathbf{y}}_i$ and $\tilde{\mathbf{y}}_j$ are both unit vectors, we have $\tilde{\mathbf{y}}_i \cdot \tilde{\mathbf{y}}_j = \|\tilde{\mathbf{y}}_i\| \cdot \|\tilde{\mathbf{y}}_j\| \cdot \cos \theta_{ij} = \cos \theta_{ij}$, where θ_{ij} is the angle between the two vectors. In general, the conformal affinity distance can be defined by $\text{dist}(\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_j) = 1 - \theta_{ij}$. Hence we have $\tilde{\mathbf{y}}_i \cdot \tilde{\mathbf{y}}_j \in [-1, 1]$ and $\text{dist}(\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_j) \in [0, 2]$. We define 3 optional modes of w_{ij} configuration to form the matrices \mathbf{W}_s and \mathbf{W}_d in the formulation.

1) **Balanced Soft Weights.** If node i and j are connected, the weight of the edge between \mathbf{x}_i and \mathbf{x}_j is set by $w_{ij} = \exp(\frac{-\text{dist}(\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_j)}{t}) = \exp(\frac{\tilde{\mathbf{y}}_i \cdot \tilde{\mathbf{y}}_j - 1}{t})$. Otherwise, $w_{ij} = 0$ if node i and j are not connected.

2) **Unbalanced Soft Weights.** If node i and j are connected, for \mathbf{W}_s , the weight is set by $w_{ij}^{(s)} = \exp(\frac{\tilde{\mathbf{y}}_i \cdot \tilde{\mathbf{y}}_j - 1}{t_s})$, while for \mathbf{W}_d , the weight is set by $w_{ij}^{(d)} = \exp(\frac{\tilde{\mathbf{y}}_i \cdot \tilde{\mathbf{y}}_j - 1}{t_d})$, where t_s and t_d can be different. Otherwise, $w_{ij} = 0$ if node i and j are not connected.

3) **Balanced Rigid Weights.** If node i and j are connected, the weight of the edge between them is set by $w_{ij} = 1$. Otherwise, $w_{ij} = 0$ if they are not connected.

The \mathbf{W}_s and \mathbf{W}_d are symmetric and nonnegative matrices which are defined over all data points to model the local conformal affinity structure of the manifold. The basic criteria is to penalize the distance measure between data via the weights for the embedding.

3.2. Conformal Discriminant Embedding

The objective function of CEA can be written as,

$$\arg \max_{\mathbf{p}} \varepsilon(\mathbf{p}) = \sum_{i,j=1}^n (1 - \mathbf{z}_i \cdot \mathbf{z}_j) \cdot w_{ij}^{(d)} \quad (4)$$

$$\text{subject to } \sum_{i,j=1}^n (1 - \mathbf{z}_i \cdot \mathbf{z}_j) \cdot w_{ij}^{(s)} = \delta, \text{ and } \|\mathbf{z}_i\| = 1.$$

where δ is a constant number, such as 1. This formulation essentially couples the neighbor and class information through the elements of \mathbf{W}_s and \mathbf{W}_d . To solve the optimization problem, the Eq. 4 can be derived in matrix style following some algebraic steps.

$$\varepsilon(\mathbf{p}) = \sum_{i,j=1}^n (1 - \mathbf{z}_i^T \mathbf{z}_j) \cdot w_{ij}^{(d)} = \sum_{i,j=1}^n (\mathbf{z}_i^T \mathbf{z}_i - \mathbf{z}_i^T \mathbf{z}_j) \cdot w_{ij}^{(d)} \quad (5)$$

Substituting \mathbf{z}_i and \mathbf{z}_j with $\mathbf{P}^T \tilde{\mathbf{y}}_i$ and $\mathbf{P}^T \tilde{\mathbf{y}}_j$ respectively according to Eq. 2, we have

$$\begin{aligned} \varepsilon(\mathbf{P}) &= \sum_{i,j=1}^n \text{trace}\{\mathbf{P}^T \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T \mathbf{P} - \mathbf{P}^T \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_j^T \mathbf{P}\} \cdot w_{ij}^{(d)} \\ &= \text{trace}\{\mathbf{P}^T (\sum_{i,j=1}^n \tilde{\mathbf{y}}_i w_{ij}^{(d)} \tilde{\mathbf{y}}_i^T - \sum_{i,j=1}^n \tilde{\mathbf{y}}_i w_{ij}^{(d)} \tilde{\mathbf{y}}_j^T) \mathbf{P}\} \\ &= \text{trace}\{\mathbf{P}^T \tilde{\mathbf{Y}} (\mathbf{D}_d - \mathbf{W}_d) \tilde{\mathbf{Y}}^T \mathbf{P}\} \end{aligned} \quad (6)$$

Thus, the objective function and constraint in Eq. 4 are equivalent to be reformulated as

$$\arg \max_{\mathbf{P}} \varepsilon(\mathbf{P}) = \text{trace}\{\mathbf{P}^T \tilde{\mathbf{Y}} (\mathbf{D}_d - \mathbf{W}_d) \tilde{\mathbf{Y}}^T \mathbf{P}\} \quad (7)$$

$$\text{subject to } \text{trace}\{\mathbf{P}^T \tilde{\mathbf{Y}} (\mathbf{D}_s - \mathbf{W}_s) \tilde{\mathbf{Y}}^T \mathbf{P}\} = \delta.$$

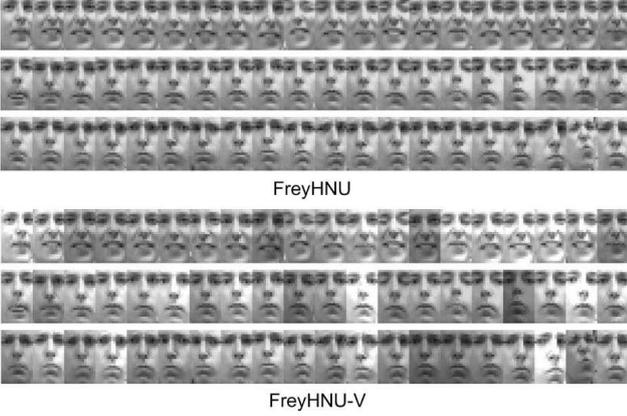


Figure 2. Sample face images from FreyHNU and FreyHNU-V databases. Each row represents a specific expressional state.

The column vectors $\{\mathbf{p}_i\}_{i=1}^d$ of projection matrix \mathbf{P} can be obtained by solving the eigenvalue problem

$$(\tilde{\mathbf{Y}}(\mathbf{D}_d - \mathbf{W}_d)\tilde{\mathbf{Y}}^T)\mathbf{p} = \lambda(\tilde{\mathbf{Y}}(\mathbf{D}_s - \mathbf{W}_s)\tilde{\mathbf{Y}}^T)\mathbf{p}. \quad (8)$$

Obviously, $\tilde{\mathbf{Y}}(\mathbf{D}_d - \mathbf{W}_d)\tilde{\mathbf{Y}}^T$ and $\tilde{\mathbf{Y}}(\mathbf{D}_s - \mathbf{W}_s)\tilde{\mathbf{Y}}^T$ are both positive semidefinite and symmetric matrices. The vectors $\{\mathbf{p}_i\}_{i=1}^d$ are the generalized eigenvectors of matrix \mathbf{A} in Eq. 1 corresponding to the d largest eigenvalues by solving the eigen decomposition equation $\mathbf{A}\mathbf{P} = \lambda\mathbf{P}$ with Singular Value Decomposition (SVD). Once the projection matrix \mathbf{P} is calculated from training, any testing datum \mathbf{x}_t can be first preprocessed to be \mathbf{y}_t . Then the normalized datum $\tilde{\mathbf{y}}_t = \mathbf{y}_t / \|\mathbf{y}_t\|$ can be transformed into $\mathbf{z}_t = \mathbf{P}^T \tilde{\mathbf{y}}_t$. It is straightforward to implement the largest conformal affinity classification through the low-dimensional data representation in the learned subspace.

4. Kernel CEA

To enhance the classification power of the CEA algorithm, we perform the kernel analysis on the preprocessed and normalized data to explore a new algorithm—Kernel CEA. The basic idea is to map the preprocessed and normalized data vectors $\{\tilde{\mathbf{y}}_i\}_{i=1}^n$ from the original input space, \mathbb{R}^D , to a higher or even infinite dimensional feature space, \mathbb{R}^F , by a nonlinear mapping function: $\Phi: \mathbb{R}^D \rightarrow \mathbb{R}^F, F > D$. Denote the $n \times n$ kernel matrix \mathbf{K} in the feature space \mathbb{R}^F by $\mathbf{K}[i, j] = \mathbf{k}(\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_j) = \Phi(\tilde{\mathbf{y}}_i) \cdot \Phi(\tilde{\mathbf{y}}_j)$. The optimization problem of kernel CEA becomes

$$\arg \max_{\mathbf{Q}} \psi(\mathbf{Q}) = \mathbf{Q}^T \mathbf{K}(\mathbf{D}_d - \mathbf{W}_d) \mathbf{K} \mathbf{Q} \quad (9)$$

$$\text{subject to } \mathbf{Q}^T \mathbf{K}(\mathbf{D}_s - \mathbf{W}_s) \mathbf{K} \mathbf{Q} = \delta.$$

where δ is a constant number, such as 1. The column vectors $\{\mathbf{q}_i\}_{i=1}^d$ of projection matrix $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_d]$ can be

Method	FreyHNU [Dim.]	FreyHNU-V [Dim.]
Error Rate (%)		
K-means	15.93 [336]	63.35 [336]
K-W+N+K-means	9.84 [10]	15.06 [100]
CEA+K-means	8.65 [3]	12.29 [100]

Table 1. Face clustering performance on Frey’s faces.

obtained by solving the eigenvalue problem

$$(\mathbf{K}(\mathbf{D}_d - \mathbf{W}_d)\mathbf{K})\mathbf{q} = \lambda(\mathbf{K}(\mathbf{D}_s - \mathbf{W}_s)\mathbf{K})\mathbf{q}. \quad (10)$$

Considering a test datum $\tilde{\mathbf{y}}_t$ whose projection is $\Phi(\tilde{\mathbf{y}}_t)$, the low-dimensional projection can be formed by $\mathbf{z}_t = \sum_{j=1}^n q_j \mathbf{k}(\tilde{\mathbf{y}}_j, \tilde{\mathbf{y}}_t)$, where q_j is the entry of column vector \mathbf{q}_i . We can choose multi-dimensional subspaces for the kernel embedding.

5. Experiments

For the experiments, we first design a facial image clustering problem for the insight of the gray-level variation tolerable property of CEA. The face recognition experiments on benchmark database ORL are later presented to demonstrate the discriminating property of CEA.

5.1. Image Clustering

We use Brendan Frey’s 1,965 gray-scale face images [10, 4] taken from sequential frames of a video for face clustering test. The images, in a resolution of 28×20 , show variations in face expression and view rotation. All the images are cropped to the resolution of 24×14 with a rectangular mask. A professional manually labels each image with respect to the expressional state and partition the data into 4 data sets—happy, neutral, unhappy, and others—with 618, 587, 634, and 126 images for each [4]. We choose the happy, neutral and unhappy data to obtain a specific database (FreyHNU) for our experiments, which has in total 1,839 images and 336 dimension (pixels) for each. To test the CEA property—gray-level variation tolerable, we also generate another database FreyHNU-V with synthetic gray-level variation. For each image from FreyHNU, we scale the gray-level of each pixel with a randomly selected constant multiplier in [0.8, 1.2]. Figure 2 illustrates some sample face images from the two databases.

We compare the K-means (on original data), Kernel-Whitening[15]+normalization+K-means, and CEA+K-means on the two database for 3-class face clustering. Each centroid is the mean of the points in that cluster, after normalizing those points to unit Euclidean length. Table 1 summarizes the clustering error rates, reduced dimension and performance comparison. As can be seen, CEA+K-means outperforms the other methods on the two databases. When the gray-level has large variations,



Figure 3. Sample face images from ORL databases.

CEA+K-means exhibits robust and reliable performance than the other methods, which infers its gray-level variation tolerable property.

5.2. Face Recognition

In this section, we evaluate the performance of proposed CEA for face recognition. Since CEA is a linear subspace learning method², we compare it with PCA [8], LDA [9], LPP [5], and LSDA [27], four most popular linear methods in face recognition. Here, PCA is unsupervised while the other four are in supervised manner. We choose the benchmark databases ORL to test these algorithms. We use nearest neighbor classifier for classification. The baseline recognition is performed in the original image space without any dimensionality reduction.

The ORL [22] database contains in total 400 images of 40 subjects with 10 gray-scale face images for each. The images show all frontal and slight tilt/rotation of the face up to 20 degrees. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open or closed eyes, smiling or not smiling) and w/o glasses. Figure 3 illustrates 20 sample images of two individuals in the database. The images are manually aligned, cropped and resized to 32×32 , with 256 gray levels per pixel. The baseline feature of each image is represented by a 1,024-dimensional column vector.

For face recognition, we perform 7 different database partitions for cross-validation evaluation. The 7 training sets are formed by images of each individual with indexes 1-3, 4-6, 7-9, 1-4, 5-8, 1-5, and 6-10. The rest images of each case form the testing sets. All the comparison methods are performed on the original image vectors without prior dimensionality reduction. Table 2 summarizes the recognition accuracy of baseline, PCA, LDA, LPP, LSDA, CEA along with the subspace dimension corresponding to the best performance. It can be seen that CEA consistently outperforms the other five methods in all the 7 cross-validation cases with highest accuracy of 91.07%, 89.29%, 92.14%, 95.83%, 95.42%, 96.50%, and 98.00%, respectively. The dimensions of the 7 CEA subspaces corresponding to the best results are 33, 34, 67, 43, 43, 49, and 46 respectively.

²Actually the transformation function $F(\cdot)$ is nonlinear because of the normalization procedure. But, after the preprocessing and normalization, we still try to learn a linear subspace for data representation.

Since the LDA-like methods have already achieved high recognition accuracies (86% ~ 96%) here, the recognition improvement (1% ~ 2%) by CEA is sufficiently significant. Moreover, the latest work LSDA [27] was reported outperforming MFA [14] and LDE [19], hence the comparison of CEA with LSDA is sufficient to validate the superiority of DSA over these algorithms. Figure 4 shows the plots of error rate versus dimension of subspace for the seven cases. It shows that CEA provides a significant improvement for face recognition accuracy with comparative low dimensional feature representations.

6. Conclusion

We have presented the CEA algorithm that projects the high-dimensional data onto the unit hypersphere and learns the embedding with a local graph modeling subject to discriminating criteria and conformal mapping nature. The feature learned by CEA is gray-level variation tolerable benefiting from the cosine-angle metric and the normalization processing. The effectiveness of our method is demonstrated by extensive simulation and comparison with image clustering and face recognition experiments. For future research, we will investigate the tensorization formulation and its solution for CEA.

References

- [1] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, no. 6, pp. 1373-1396, 2003. 1
- [2] R.O. Duda, E.H. Peter, and G.S. David, *Pattern Classification* (2nd ed.), Wiley Interscience, 2000. 1
- [3] Y. Fu and T.S. Huang, *Locally Linear Embedded Eigenspace Analysis*, www.ifp.uiuc.edu/~yunfu2/papers/LEA-Yun05.pdf, IFP-TR, UIUC, Jan. 1, 2005. 1
- [4] Y. Fu and T.S. Huang, "Unsupervised Locally Embedded Clustering For Automatic High-Dimensional Data Labeling," *IEEE Conf. on ICASSP'07*, pp. 1057-1060, 2007. 4
- [5] X.F. He, S.C. Yan, Y.X. Hu, P. Niyogi, and H.-J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Trans. on PAMI*, vol. 27, no. 3, pp. 328-340, 2005. 1, 5
- [6] X.F. He, D. Cai, S.C. Yan, and H.J. Zhang, "Neighborhood Preserving Embedding," *IEEE Conf. on ICCV'05*, vol. 2, pp. 1208-1213, 2005. 1
- [7] J. Ham, D.D. Lee, S. Mika, and B. Schölkopf, "A Kernel View of the Dimensionality Reduction of Manifolds," *ICML'04*, 2004. 1
- [8] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," *IEEE Conf. on CVPR*, pp. 586-591, 1991. 1, 5
- [9] P.N. Belhumeur, J.P. Hefanpha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711-720, 1997. 1, 5
- [10] S.T. Roweis and L.K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323-2326, 2000. 1, 4

Method	1 ~ 3 (%)	4 ~ 6 (%)	7 ~ 9 (%)	1 ~ 4 (%)	5 ~ 8 (%)	1 ~ 5 (%)	6 ~ 10 (%)
Baseline	76.79 [1024]	76.79 [1024]	76.43 [1024]	85.83 [1024]	80.00 [1024]	87.50 [1024]	84.00 [1024]
PCA	76.79 [63]	76.79 [117]	76.43 [50]	85.83 [103]	80.00 [64]	87.50 [49]	84.00 [133]
LDA	86.07 [39]	87.14 [19]	90.36 [33]	90.83 [38]	90.00 [38]	94.00 [36]	95.00 [34]
LPP	86.43 [41]	87.50 [53]	88.57 [83]	91.25 [79]	92.92 [36]	93.00 [107]	96.50 [76]
LSDA	89.29 [57]	88.57 [40]	91.43 [40]	95.00 [40]	94.58 [39]	95.00 [49]	96.00 [40]
CEA	91.07 [33]	89.29 [34]	92.14 [67]	95.83 [43]	95.42 [43]	96.50 [49]	98.00 [46]

Table 2. Face recognition accuracy comparison on the ORL database. The number inside [] is the reduced dimension for the best result.

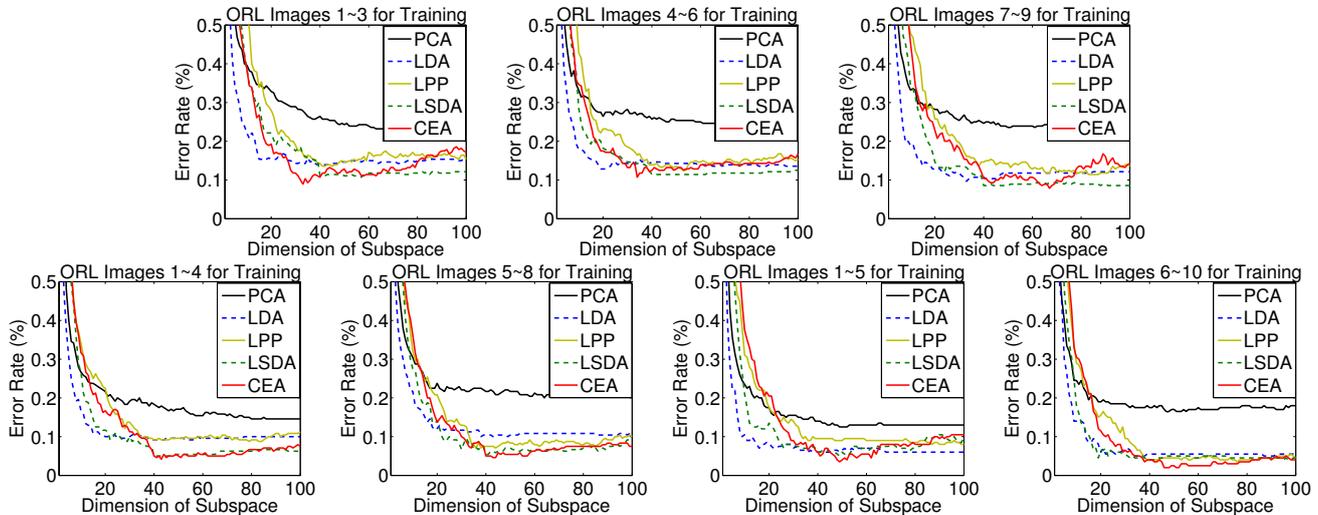


Figure 4. Face recognition error rate vs. dimension of subspace on ORL databases using PCA, LDA, LPP, LSDA, and CEA. The seven cases of individual images with indexes 1 ~ 3, 4 ~ 6, 7 ~ 9, 1 ~ 4, 5 ~ 8, 1 ~ 5, and 6 ~ 10 for training respectively.

- [11] L.K. Saul and S.T. Roweis, "Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds," *J. of Machine Learning Research*, vol. 4, pp. 119-155, 2003. 1
- [12] J.B. Tenenbaum, V.de Silva, and J.C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319-2323, 2000. 1
- [13] K.Q. Weinberger and L.K. Saul, "Unsupervised Learning of Image Manifolds by Semidefinite Programming," *IEEE Conf. on CVPR'04*, vol. 2, pp. 988-995, 2004. 1, 2
- [14] S.C. Yan, D. Xu, B.Y. Zhang, and H.J. Zhang, "Graph Embedding: A General Framework for Dimensionality Reduction," *IEEE Conf. on CVPR'05*, pp. 830-837, 2005. 1, 5
- [15] D.M.J. Tax and P. Juszczak, "Kernel Whitening for One-Class Classification," *Int'l Journal of Pattern Recognition and Artificial Intelligence*, vol.17, no.3, pp. 333-347, 2003. 4
- [16] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 696-710, 1997. 1
- [17] A. Weingessel and K. Hornik, "Local PCA Algorithms," *IEEE Trans. on NN*, vol. 11, no. 6, pp. 1242-1250, 2000. 1
- [18] D. Cai, X.F. He, J.W. Han, and H.-J. Zhang, "Orthogonal Laplacianfaces for Face Recognition," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3608-3614, 2006. 2
- [19] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local Discriminant Embedding and Its Variants," *IEEE Conf. on CVPR'05*, pp. 846-853, 2005. 1, 5
- [20] A.B.A. Graf, A.J. Smola, and S. Borer, "Classification in a Normalized Feature Space Using Support Vector Machines," *IEEE Trans. on NN*, vol. 14, no. 3, pp. 597 - 605, 2003. 2
- [21] A. Banerjee, I.S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the Unit Hypersphere Using Von Mises-Fisher Distributions," *J. of Mach. Learn. Res.*, vol.6, pp.1345-1382,2005. 2
- [22] F. Samaria and A.C. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," *IEEE Workshop on Applications of Computer Vision*, pp. 138-142, 1994. 5
- [23] M.-H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *IEEE Conf. on FG'02*, pp. 215-220, 2002. 1, 2
- [24] M.-H. Yang, "Extended Isomap for Pattern Classification," *AAAI'02*, pp. 224-229, 2002. 2
- [25] S.C. Yan, H.J. Zhang, Y.X. Hu, B. Y. Zhang, and Q.S. Cheng, "Discriminant Analysis on Embedded Manifold," *ECCV'04*, pp. 121-132, 2004. 2
- [26] Q.B. You, N.N. Zheng, S.Y. Du, and Y. Wu, "Neighborhood Discriminant Projection for Face Recognition," *IEEE Conf. on ICPR'06*, pp. 532-535, 2006. 2
- [27] D. Cai, X.F. He, K. Zhou, J.W. Han, and H.J. Bao, "Locality Sensitive Discriminant Analysis," *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'07)*, India, Jan. 2007. 1, 5
- [28] Y. Bengio, J.F. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, "Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering," *NIPS'04*, 2004. 1