ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Image and Vision Computing xxx (2007) xxx–xxx

image and vision COMPUTING

www.elsevier.com/locate/imavis

Subspace manifold learning with sample weights

Nathan Mekuz *, Christian Bauckhage, John K. Tsotsos

Department of Computer Science and Engineering, Center for Vision Research, York University, CSE 3031, 4700 Keele Street, Toronto, Ont., Canada M3J 1P3

Received 10 November 2005; received in revised form 16 August 2006; accepted 20 October 2006

Abstract

Subspace manifold learning represents a popular class of techniques in statistical image analysis and object recognition. Recent research in the field has focused on nonlinear representations; locally linear embedding (LLE) is one such technique that has recently gained popularity. We present and apply a generalization of LLE that introduces sample weights. We demonstrate the application of the technique to face recognition, where a model exists to describe each face's probability of occurrence. These probabilities are used as weights in the learning of the low-dimensional face manifold. Results of face recognition using this approach are compared against standard nonweighted LLE and PCA. A significant improvement in recognition rates is realized using weighted LLE on a data set where face occurrences follow the modeled distribution.

Keywords: Subspace learning; Nonlinear dimensionality reduction; Locally linear embedding; Face recognition

1. Introduction

The problems of dimensionality reduction and subspace learning are active research topics in machine learning and statistical image analysis [1]. In this context, the goal has often been related to mitigating the effects of the curse of dimensionality [2], compression (e.g., [3]) or the uncovering of latent variables (e.g., blind source separation [4], factor analysis [5]). Specialized dimensionality reduction techniques have also been developed for visualization of high-dimensional data (e.g., [6]).

Subspace learning techniques have also been successfully applied in machine vision, especially in the context of face recognition where they have gained considerable popularity [7,8]. The strong interest in face recognition has been motivated by applications ranging from authentication and security to expression recognition and user interface design. While humans are highly adept at recognizing faces, this task remains a significant challenge for machines. Typically, a face recognition system is trained offline with a set of labeled images prior to being presented with a novel image to recognize. The challenge is to maximize the amount of relevant detail learned from the training data with minimum sensitivity to transformations such as pose and illumination.

Face images are especially suitable for subspace learning: faces are mostly symmetrical, contain many textured and smooth surfaces and have a fairly constant appearance, resulting in strong correlation. Faces normally appear upright, and in many applications a frontal view is available. For these reasons, most face recognition systems have used image-based representations, where faces are represented with characteristic images, and dimensionality reduction techniques are applied for increased storage and comparison efficiency. These techniques effectively learn a subspace of the space spanned by the original input, on which the face images in the training data (approximately) lie. By extension, novel face images are expected to lie close to this manifold.

Some face recognition systems are holistic and use a single model to represent the entire face (e.g., [9]) while others

* Corresponding author. Tel.: +1 416 736 2100x33972.
E-mail address: mekuz@cs.yorku.ca (N. Mekuz).

0262-8856/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.10.007

use geometrical configurations of local features (recognition by parts, e.g., [10]). Our empirical evaluation takes the former approach, but subspace learning also plays a role in parts-based systems, where it may be applied to each part.

1.1. Relationship to previous research

Face recognition has been an extremely fertile research area, and has resulted in many specialized techniques. A complete review is beyond the scope of this paper, but Zhao et al. [8] survey many of the popular techniques in the field. Here, we limit our discussion to the main subspace-based approaches.

Kirby and Sirovich [3] first proposed a low-dimensional representation for face images, based on principal component analysis (PCA, [11]). PCA is a linear transformation into a lower-dimensional coordinate system that preserves maximum variance in the data, thus minimizing mean-square error, computed as a subset of the Karhunen–Loeve rotation [12]. Turk and Pentland later applied this technique to face recognition and detection [13], and introduced the notion of *face space* – the space spanned by the transformation, and the *faceness* of novel images was measured by their distance to face space. The database consisted of low-dimensional PCA projections of the training data, and face recognition was performed by applying a nearest-neighbor classifier in the reduced subspace.

Turk and Pentland's technique has been very influential but PCA's linear model is suboptimal for image data. Murase and Nayar [14] present an extension of Turk and Pentland's technique that represents continuous appearance variations of objects using spline interpolation. The test image is first projected onto a low-dimensional space, in order to identify the object. Once the object is identified, the image is projected onto a new subspace, defined specifically for that object. The resulting subspaces are nonlinear and appear as manifolds in high-dimensional space.

Other linear dimensionality reduction techniques that have been successfully applied to face recognition include independent component analysis (ICA, [15]) and linear discriminant analysis (LDA, [16]). ICA seeks components that are statistically independent (rather than de-correlated). It is argued to provide a more localized decomposition than PCA. LDA is a linear transformation related to the Fisher linear discriminant (FLD, [17]), that seeks to maximize class separability in the projection, based on known class labels [12].

As a least-squares technique, PCA suffers from sensitivity to outliers. Several approaches have been proposed to increase the robustness of PCA. De la Torre and Black propose using M-estimators [18]. Skočaj et al. [19] use a generalized version of PCA that introduces image and pixel weights, and the effect of outliers in the data is reduced by reducing their weights. The weights control the learning of the subspace (the training phase) and are also used in the recognition classifier. Weighted subspace learning is also a central element of our technique, although the motivation for introducing the weights is different.

The space spanned by images of an object under different variations is highly nonlinear, and the application of linear dimensionality reduction techniques to image data results in suboptimal object recognition performance [20–23]. Several of the recently proposed subspace learning techniques model the subspace manifold as a connected patchwork of locally linear surfaces. These models are especially well suited to manifolds where the intrinsic dimensionality varies in different areas of the manifold or where the locally intrinsically low-dimensional patches have a globally varying orientation. Also, in cases where the manifold is discontinuous, these techniques make it possible to model clusters separately. One popular technique from this category is locally linear embedding (LLE, [21]), which is particularly appealing due to its simplicity and the existence of a closed-form solution (discounting the computation of the eigenvectors). LLE computes the local structure around each input point based on its neighbors. LLE has been applied to face recognition, and the resultant face manifold has been shown to provide better classification opportunities than the face space produced by PCA. This suggests that for the purpose of face recognition, the local structure of the manifold is a better discriminant than the global Euclidean structure.

1.2. Weighted manifold learning

In addition to controlling the effect of outliers, weighted input data can also be used to tune the learning of the manifold, so that data samples can be considered according to their reliability or significance. This can be useful in a face recognition system that is trained with a large number of appearances for each face, where a likelihood function can be defined for the occurrence of the different appearances in a novel image. In nonweighted subspace learning algorithms, the user needs to balance the need for a sufficient number of training samples with the risk that including uncommon and infrequent appearances in the training data may adversely affect the representation of the common appearances. Our approach extends a popular nonlinear manifold learning technique, namely LLE, to work with sample weights. If the probability of occurrence for different face appearances is known (or can be estimated), then our technique eliminates this dilemma, and effectively models the subspace based on the available weights.

1.3. Overview

We propose a locally linear dimensionality reduction technique based on locally linear embedding (LLE), where the input data is labeled with weights, which are used to bias the transformation to model certain parts of the input more effectively than others. This paper consists of four sections. The first section has motivated the need for sample weights in dimensionality reduction and for local

Please cite this article in press as: N. Mekuz et al., Subspace manifold learning with sample weights, Image Vis. Comput. (2007), doi:10.1016/j.imavis.2006.10.007

learning of the manifold. Section 2 describes our technical approach and the algorithm used. In Section 3 we present empirical results of our algorithm applied to face recognition. Section 4 summarizes our work. Finally, Section 5 suggests directions for future research.

2. Sample weights

Our technique extends LLE to allow for weighted samples in the training phase where the weights bias the transformation to favor certain input points over others. If input data represents observations from an unknown manifold, then weights may be used to represent the reliability of the observations.

Assigning weights to training samples can come in useful in various scenarios. For example, in a face recognition application, higher weights may be assigned to areas of the face considered more important for recognition, such as the eyes. Lower weights can be used to reduce the effect of outliers (e.g., using weighted PCA, [19]). Weights may also be applied to images in the training set, for example assigning a higher weight to more recent images of an individual, or images that are considered more likely to be encountered in the test set. The latter application requires a model of the probabilities of occurrence of different appearances. In this paper, we present an extension of LLE – a promising nonlinear dimensionality reduction technique especially effective for face recognition [21] – that allows weighted training data.

2.1. The use of weights

Generalized PCA, and specifically weighted PCA is a common multivariate statistical technique [11]. Skočaj et al. use weighted PCA in their object recognition system [19]. Weights are applied to individual pixels (*spatial weights*) and images (*temporal weights*). Spatial weights are used to describe the reliability (e.g., missing pixels, outliers) or importance of different parts of the image, while temporal weights are used to bias the learning of the subspace in favor of more recent images of each subject. The technique works by maximizing the weighted variance in the low-dimensional space (face space) in order to achieve lower reconstruction error for certain target images or pixels. Image weights may also be used to tune a user authentication system to achieve better recognition rates for individuals that require more reliable recognition performance.

There are also other situations where image weights may be useful. Face images are subject to variation due to scale, orientation (i.e., rotation with respect to the camera's optical axis) and pose (relative to the camera), facial expression, occlusion, and lighting conditions, and the presence or absence of features such as beards, glasses, or articles of clothing such as scarfs and hats. To account for these variations, face recognition systems are typically trained with multiple images of each subject. Nonweighted techniques present a dilemma for users. On the one hand, it is desirable to train the system with as many appearance variations as possible to facilitate good recognition under different conditions. On the other hand, the inclusion of training samples which represent extreme deviations from the norm and which are not very likely to occur in test data may taint the representation of the higher-likelihood samples, and adversely affect their recognition.

Skočaj et al. propose weighted PCA algorithms for processing weighted images and images with weighted pixels, both as a batch process, and as an EM-based adaptive online algorithm [19].

The batch technique incorporates the weights both in the computation of the sub-space and in the classification of the test image. Given image weights $\{\eta_i\}_{i=1}^N$, the subspace is computed by transforming the input vectors $\{\vec{x}_i\}_{i=1}^N$ lying in data space $\mathbb{R}^M$, after subtracting the mean, as follows:

$$\vec{x}'_i = \sqrt{\eta_i}(\vec{x}_i - \bar{\vec{x}}), \quad i = 1, \ldots, N \tag{1}$$

where $\vec{x}'_i$ are the transformed images from which the covariance matrix and its eigenvectors are computed. Similarly, spatial weights $\{\pi_j\}_{j=1}^M$ are introduced as follows:

$$x'_{ij} = \sqrt{\pi_j}(x_{ij} - \bar{x}_j, \quad i = 1, \ldots, N, j = 1, \ldots, M \tag{2}$$

where $\vec{x}'_i$ are the images adjusted for spatial weights. Combining the two, a set of images modified to reflect both image and spatial weights, $\{\vec{x}'_i\}_{j=1}^N$ is computed as follows:

$$x'_{ij} = \sqrt{\eta_i \pi_j}(x_{ij} - \bar{x}_j, \quad i = 1, \ldots, N, j = 1, \ldots, M \tag{3}$$

These modified images are used as an input to nonweighted PCA to obtain a weight-adjusted subspace.

Our weighted learning technique incorporates sample weights into the learning process. These weights can be set to the test images' probability of occurrence to bias the learning process in favor of certain face appearances at the expense of others. For instance, we may assume that of all variations caused by in-plane rotation, upright faces are most likely. Our empirical testing used a Gaussian distribution model to represent the probability of occurrence of face images transformed by in-plane rotation, where the mean was centered around the upright position.

2.2. Locally linear manifold learning

Nonlinear manifold learning techniques (e.g., [20–23]) have been developed for a variety of applications. Roweis and Saul [21] demonstrate several synthetic visualization examples where three-dimensional data (such as the Swiss roll example depicted in Fig. 1) is reduced to two dimensions. Variance in the depicted examples does not align along any specific linear axis, and consequently PCA fails to produce a meaningful projection. Empirical evidence shows that the space of face images is indeed highly nonlinear. Other research (e.g., [18]) points out and illustrates PCA's extreme sensitivity to outliers. Outliers present an even more severe problem in techniques which de-correlate

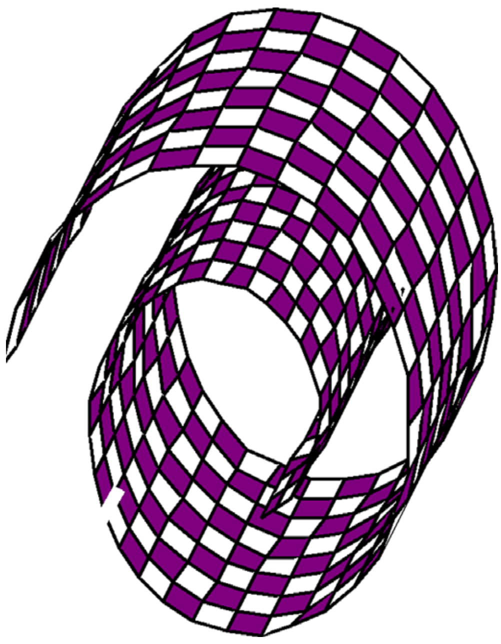

Fig. 1. The Swiss roll manifold depicted here is highly nonlinear. In cases such as this, techniques that seek dimensionality reduction (such as PCA) will fail to discover the manifold's shape.

higher order statistics such as ICA or general projection pursuit.

Locally linear embedding (LLE) is a locally linear data reduction technique which appears to be especially effective for face recognition. This is confirmed by independent studies (e.g., [24]). LLE computes dimensionality reduction that preserves the local neighborhood structure of the input data in the low-dimensional transformation. The transformation models the subspace manifold as a connected patchwork of locally linear surfaces. The local nature of the learning results in a robust representation, since outliers only affect the learning in their immediate neighborhood. On the other hand, LLE requires that the manifold be adequately sampled.

If the manifold is sufficiently sampled, then in small neighborhoods, points lie on nearly linear patches of the manifold. This is commonly justified using Taylor's theorem [25] which states that any differentiable function is linear at the limit in a small area around a point. LLE works by identifying local neighborhood distance relationships, and finding a mapping into a lower dimensionality that preserves them as much as possible. The neighborhood configuration is expressed as the set of coefficients for each point, that best reconstructs it as a linear combination of its neighbors.

Two formulations to define a data point's neighborhood have been proposed. The simpler formulation defines the neighborhood as the k nearest neighbors in the Euclidean sense. The other formulation defines the neighborhood of a point as all data points that fall within distance t from it. The advantage of the former approach is that the entire neighborhood configuration may be represented as a sparse matrix, with a number of nonzero elements that is known ahead of time. Our empirical testing follows Roweis [21] and uses the k nearest neighbors approach, but the technique may be applied to the ε-radius ball formulation easily.

2.3. Extending LLE with weights

Our goal is to extend LLE with sample weights similarly to the weighted PCA technique described above. The non-weighted LLE computation performs the following steps:

(a) Define each data point's neighborhood, either by identifying the k nearest Euclidean neighbors, or by finding all data points that are within distance ε from it. An efficient implementation using kd-trees or ball trees can perform this step in $\mathcal{O}(N \log N)$ time, but even with a brute force $\mathcal{O}(MN^2)$ implementation, this step is relatively inexpensive.

(b) For each data point $\vec{x}_i$, compute a set of coefficients $\vec{\xi}_i$ that best reconstruct it from its neighbors. Together these sets can be combined into a sparse $N \times N$ matrix. This step can be performed in $\mathcal{O}(MNk^3)$ time, where M is the input dimensionality and k is the (amortized) neighborhood size.

(c) Compute low-dimensional vectors $\{\vec{y}\}_{i=1}^{N}$ in data space $\mathbb{R}^m$ which best reproduce the coefficients obtained in the previous step, in a new low-dimensional space of the desired target dimensionality. The computational complexity of this step is $\mathcal{O}(mN^2)$, where m is the target dimensionality.

Our research extends LLE to work with weighted samples, where the weights control the degree to which each input point affects the construction of the low-dimensional embedding, or the learned manifold. If the input points represent observations, then the weights can be used to represent their reliability. Since the local shape of the manifold at each point is learned from its neighbors, the neighbors' weights have to be included in the computation in a way that increases the influence of neighbors with high weights. This requires modifying the first two steps of the algorithm.

In its first step, the algorithm selects representative points for each input point's neighborhood (nearest Euclidean neighbors). Next, a set of coefficients is computed for each data point that reconstructs it optimally from its neighbors. Given neighbors {*neighbor*$_1$,…,*neighbor*$_k$} at point $\vec{x}_p$, in order to compute coefficients at $\vec{x}_p$, the local covariance matrix C of the neighbors is computed, about $\vec{x}_p$, where the elements c_{ij} of C are defined as follows:

$$c_{i,j} = (\vec{x}_p - \vec{x}_i)^{\mathrm{T}} (\vec{x}_p - \vec{x}_j) \quad (4)$$

for $i, j\ \varepsilon \in \{neighborl_1, \ldots, neighbor_k\}$. This results in a symmetric semipositive definite Gram matrix which defines the difference vectors $\vec{x}_i - \vec{x}_j$) up to isometry. Next, optimal coefficients $\{\xi_i\}_{i=1}^k$ that minimize $\vec{x}_p$'s reconstruction error from its neighbors can be easily computed using Lagrange multipliers, or equivalently by solving,

$$C\vec{\xi} = [1, 1, \ldots, 1]^{\mathrm{T}} \tag{5}$$

for $\vec{\xi}$ and normalizing so that $\sum_{i=1}^k \xi_i = 1$. This is repeated for each input point $\vec{x}_p$.

Therefore, the computation of Euclidean distances affects both the choice of points used to represent each input point's neighborhood and the formation of the local covariance matrices in the second step. With weighted samples, weighted distances need to be computed which combine Euclidean distances with the points' weights. Given weights $w_i \geqslant 0$ for data points $\vec{x}_i$, respectively, the adjusted distance d'_{ij} between points $\vec{x}_i$ and $\vec{x}_j$ is given by,

$$d'^2_{ij} = \frac{d^2_{ij}}{w_i w_j} = \frac{(\vec{x}_p - \vec{x}_i)^{\mathrm{T}}(\vec{x}_p - \vec{x}_j)}{w_i w_j} \tag{6}$$

The adjusted distances d'_{ij} are used both for selecting neighbors at each input point in the first step of the algorithm and for the computation of the Gram matrix in the second step. Point $\vec{x}_i$'s neighbors in the weighted scheme are therefore the k' points with the smallest d'_{ij} or alternatively all points for which $d'_{ij} \leqslant \varepsilon'$, where k' and ε' are adjusted parameters. Any nonnegative values may be used for the weights. Note that the transformation effected by the weights is linear and therefore only their relative values are important. In other words, multiplying all weights by a constant has no effect on the algorithm's output.

Similar to the computation of the distances, the computation of the local covariance matrix C′ for $\vec{x}_p$ is modified as follows,

$$c'_{i,j} = \sqrt{w_i w_j}(\vec{x}_p - \vec{x}_i)^{\mathrm{T}}(\vec{x}_p - \vec{x}_j) \tag{7}$$

where $i, j \in \{neighbor_1, \ldots\ neighbor_{\mathrm{k}}\}$. Now replacing C with C′ in equation [5] after normalization, yields weight-adjusted reconstruction coefficients $\vec{\xi}'$ that incorporate sample weights, formally,

$$C'\vec{\xi}' = [1,\ 1,\ \ldots,\ 1]^{\mathrm{T}} \tag{8}$$

Once the weight-adjusted coefficients $\vec{\xi}'$ have been computed, the final step computes the embedding based on these adjusted coefficients, and therefore no changes are required in the final step of the algorithm.

It is not clear how to select good values for k or ε in weighted or non-weighted LLE, other than empirically. The neighborhood size effectively controls how local the features preserved by the LLE transformation will be. Whatever values are chosen, the introduction of weights changes the calculation of distances and therefore different values need to be used for k' and ε'. These values will likely be higher than in the nonweighted case, since the weighting creates unequal dependence of each point on its different neighbors, and greater sensitivity to noise.

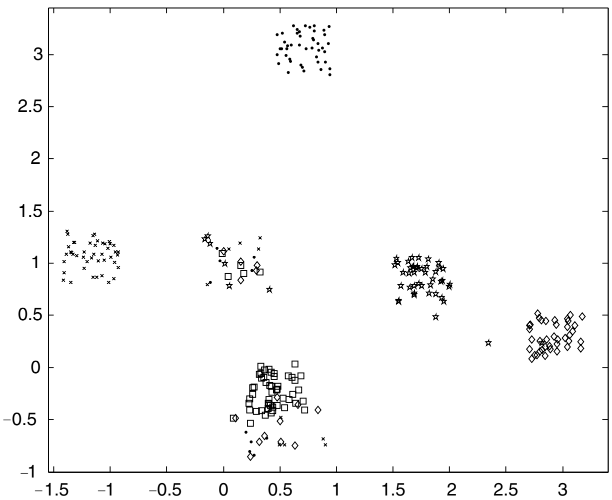

Fig. 2. The first two dimensions of the weighted LLE projection of face images, using $k = 7$, $d = 15$. Each marker type represents the face image of a different subject.

3. Empirical evaluation

We have designed a series of empirical tests to analyze the effectiveness of our method. We used a standard database of face images and added images of faces rotated in-plane by small angles. We chose in-plane rotation as a mutator since it can be easily generated synthetically and since faces normally appear more or less upright. The face images in the resulting database were assigned normally-distributed probabilities of occurrence with respect to the rotation angle, with a mean at the upright position, as depicted in Fig. 3. We used m-fold cross validation, meaning one data set was used for both training and testing. While the modeled probabilities were input as weights in the training phase, the test images appeared at probabilities that followed the modeled distribution. The goal of our evaluation was to assess the effectiveness of our method against other subspace learning techniques on this scenario.

Fig. 2 depicts a two-dimensional plot of the projection created by weighted LLE given the weighted face images we used in our experiments. Different markers in the plot represent face images of different individuals. Clusters are visible in the plot suggesting that the resulting projection should be conducive to good classification.

3.1. Methodology

In our tests we used the Yale face database [16], which consists of 165 images of 15 subjects recorded under different lighting conditions and depicting a variety of facial expressions. Each face was rotated by $\theta_1 \in \{-8°, -4°, 0°, 4°, 8°\}$, creating a total of 825 face images. We cropped and aligned the images to a final size of 80×80, and to further reduce the cost of computations, we preprocessed the resulting images with PCA, retaining 100 principal compo-

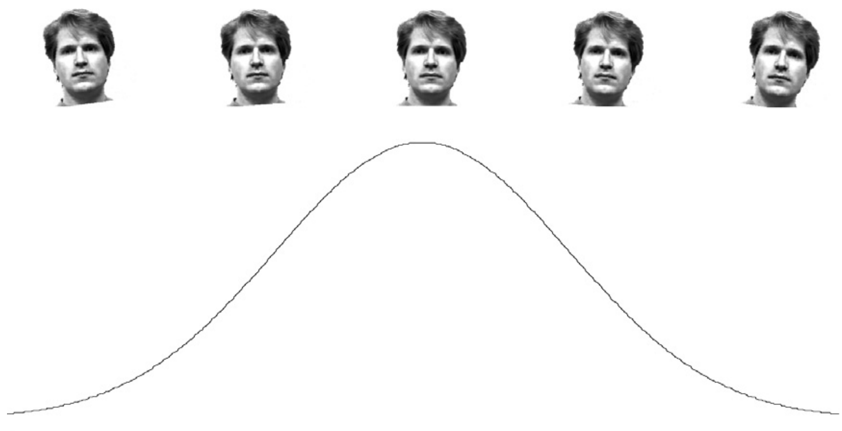

Fig. 3. A Gaussian probability distribution model for the probability of occurrence of faces mutated using in-plane rotation.

nents, capturing well over 99 percent of all variance present.

Recognition rates were measured using the leave one out strategy (m-fold cross validation). The system was trained with all images but one, and classification was attempted on the remaining image. This process was repeated for all images in the database, and error (misclassification) rates were recorded for four subspace learning algorithms:

- PCA on the original upright faces only.
- Nonweighted LLE ($k = 3$) on the original upright faces only.
- Nonweighted LLE ($k = 7$) on all 825 images.
- Weighted LLE ($k = 7$) on all images where the weights w_i were set to the images' probability of occurrence using a Gaussian model.

3.2. Results

We compared the recognition error rates achieved by the above four methods, at various dimensionality settings. The results are summarized in Fig. 4, which plots the error rate as a function of the target dimensionality of the embedding space.

The results are consistent across all target dimensionalities. The error rates realized by weighted LLE are consistently lower than those of the other three algorithms.

The error rates achieved by PCA are consistently the highest. Interestingly, this observation contradicts other tests [21] that find a crossover point around $d = 18$ above which PCA actually achieves better recognition rates than LLE.

3.3. Discussion

The results support our hypothesis that in cases where training samples have an unequal but known likelihood of occurring in test data, weighted LLE successfully learns a

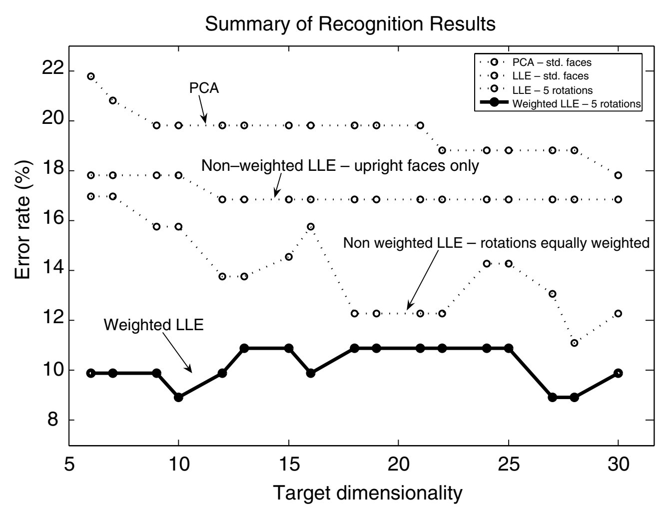

Fig. 4. Recognition errors of four dimensionality reduction configurations using the Yale faces [16] database.

subspace manifold which takes these probabilities into account, and this manifold results in superior classification ability. Weighted LLE achieved superior recognition rates against the other three techniques at all dimensionalities, with improvement ranging from 2.5% to 7%. However, while the other three techniques show better recognition results at higher target dimensionalities, weighted LLE seems to have a constant error rate of around 10 percent across all dimensionalities, although we did not test at dimensionalities higher than 30. Also, it may be possible to improve these results with further tuning of the parameters.

4. Summary

We have presented a novel approach to weighted manifold learning, by extending the locally linear embedding (LLE) algorithm with sample weights. The weights influence the computation of the low dimensional embedding by biasing the modeling of the neighborhoods in favor of data points with higher weights. This technique may be used where input observations have associated measures of reliability, significance or probability of occurring in a test data point.

We tested our technique on face recognition using an extended database of face images with unequal frequencies of occurrence and compared recognition results against nonweighted LLE as well as PCA. Our algorithm produced embeddings which emphasized the more frequently occurring faces, and realized significantly superior recognition rates against the other techniques, in all target dimensionalities.

5. Future work

The selection of neighbors at each point has a significant effect on the transformation (both in the weighted and nonweighted case). To date, little research has been done on techniques for selecting these neighborhood graphs, and values for k or ε are often chosen empirically. The introduction of weights further complicates this issue. With weights, the neighborhood size is no longer dictated only by the locality of the features to be preserved by the transformation. Care must also be taken to ensure that each neighborhood sufficiently represents the local shape of the manifold when weights are included.

Also, neighborhood formulations that rely solely on Euclidean relationships do not enforce selection of points in general position. In extreme cases, the selection of collinear points may result in singularity conditions. Traditional approaches such as Delaunay triangulations are computationally impractical in high dimensionality and tend to select neighbors which are geodesically far apart.

Finally, LLE does not generalize as well as PCA. Projecting a novel data point requires a re-computation of the embedding to include the new point.

Acknowledgment

The authors gratefully acknowledge the help of Konstantinos Derpanis in reviewing this paper.

References

[1] E.J. Keogh, K. Chakrabarti, M.J. Pazzani, S. Mehrotra, Dimensionality reduction for fast similarity search in large time series databases, Knowledge and Information Systems 3 (3) (2001) 263–286.

[2] R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, NJ, USA, 1961.

[3] M. Kirby, L. Sirovich, Low-dimensional procedure for the characterization of human faces, Optical Society of America 4 (3) (1987) 519–524.

[4] C. Jutten, J. Herault, Blind separation of sources, Signal Processing 24 (1) (1991) 1–10.

[5] L.L. Thurstone, Multiple factor analysis, Psychological Review 38 (1931) 406–427.

[6] C.M. Bishop, M.E. Tipping, A hierarchical latent variable model for data visualization, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3) (1998) 281–293.

[7] M.-H. Yang, D. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (1) (2002) 34–58.

[8] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (4) (2003) 399–458.

[9] W. Zhao, R. Chellappa, A. Krishnaswamy, Discriminant analysis of principal components for face recognition, in: Third International Conference on Face & Gesture Recognition, IEEE Computer Society, Washington, DC, USA, 1998, p. 336.

[10] S.Z. Li, X. Hou, H. Zhang, Q. Cheng, Learning spatially localized, parts-based representation., in: IEEE Conference on Computer Vision and Pattern Recognition (1), 2001, pp. 207–212.

[11] I. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.

[12] K. Fukunaga, Introduction to statistical pattern recognition, Second ed., Academic Press Professional Inc., San Diego, CA, USA, 1990.

[13] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.

[14] H. Murase, S. Nayar, Visual learning and recognition of 3-d objects from appearance, International Journal of Computer Vision 14 (1) (1995) 5–24.

[15] B. Draper, K. Baek, M. Bartlett, J. Beveridge, Recognizing faces with PCA and ICA, Computer Vision and Image Understanding 92 (2) (2003) 115–137.

[16] P.N. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[17] R.A. Fisher, The use of multiple measures in taxonomic problems, Annual Eugenics 7 (1936) 179–188.

[18] F. De la Torre, M. Black, Robust principal component analysis for computer vision, in: International Conference on Computer Vision, 2001, pp. 362–369.

[19] D. Skočaj, H. Bischof, A. Leonardis, A robust PCA algorithm for building representations from panoramic imagesEuropean Conference on Computer Vision, vol. IV, Springer, Verlag, 2002, pp. 761–775.

[20] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.

[21] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[22] A. Shashua, A. Levin, S. Avidan, Manifold pursuit: A new approach to appearance based recognition., in: International Conference on Pattern Recognition, vol. 3, 2002, pp. 590–594.

[23] J. Tenenbaum, V. de Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.

[24] M.-H. Yang, Face Recognition using Extended Isomap, in: IEEE International Conference on Image Processing, 2002, pp. 117–120.

[25] B. Taylor, Methodus incrementorum directa et inversa, London (1715).

Please cite this article in press as: N. Mekuz et al., Subspace manifold learning with sample weights, Image Vis. Comput. (2007), doi:10.1016/j.imavis.2006.10.007