Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 283540, 15 pages
doi:10.1155/2008/283540

Research Article

Person-Independent Head Pose Estimation Using Biased Manifold Embedding

Vineeth Nallure Balasubramanian, Sreekar Krishna, and Sethuraman Panchanathan

Center for Cognitive Ubiquitous Computing, Arizona State University, Tempe, AZ 85281, USA

Correspondence should be addressed to Vineeth Nallure Balasubramanian, vineeth.nb@asu.edu

Received 2 June 2007; Revised 16 September 2007; Accepted 12 November 2007

Recommended by Konstantinos N. Plataniotis

Head pose estimation has been an integral problem in the study of face recognition systems and human-computer interfaces, as part of biometric applications. A fine estimate of the head pose angle is necessary and useful for several face analysis applications. To determine the head pose, face images with varying pose angles can be considered to be lying on a smooth low-dimensional manifold in high-dimensional image feature space. However, when there are face images of multiple individuals with varying pose angles, manifold learning techniques often do not give accurate results. In this work, we propose a framework for a supervised form of manifold learning called Biased Manifold Embedding to obtain improved performance in head pose angle estimation. This framework goes beyond pose estimation, and can be applied to all regression applications. This framework, although formulated for a regression scenario, unifies other supervised approaches to manifold learning that have been proposed so far. Detailed studies of the proposed method are carried out on the FacePix database, which contains 181 face images each of 30 individuals with pose angle variations at a granularity of 1°. Since biometric applications in the real world may not contain this level of granularity in training data, an analysis of the methodology is performed on sparsely sampled data to validate its effectiveness. We obtained up to 2° average pose angle estimation error in the results from our experiments, which matched the best results obtained for head pose estimation using related approaches.

1. INTRODUCTION AND MOTIVATION

Head pose estimation has been studied as an integral part of biometrics and surveillance systems for many years, with its applications to 3D face modeling, gaze direction detection, and pose-invariant person identification from face images. With the growing need for robust applications, face-based biometric systems require the ability to handle significant head pose variations. In addition to being a component of face recognition systems, it is important to determine the head pose angle from a face image, independent of the identity of the individual, especially in applications of 3D face recognition. While coarse pose angle estimation from face images has been reasonably successful in recent years [1], accurate person-independent head pose estimation from face images is a more difficult problem, and continues to elicit effective solutions.

There have been many approaches adopted to solve the pose estimation problem in recent years. A broad subjective classification of these techniques with pointers to sample work [2–5] is summarized in Table 1. As Table 1 points out, shape-based geometric and appearance-based methods have been the most popular approaches for many years. However, recent work has established that face images with varying poses can be assumed to lie on a smooth low-dimensional manifold, and this has opened up efforts to approach the problem from the perspectives of non-linear dimensionality reduction.

The computation of low-dimensional representations of high-dimensional observations like images is a problem that is common across various fields of science and engineering. Techniques like principal component analysis (PCA) are categorized as linear dimensionality reduction techniques, and are often applied to obtain the low-dimensional representation. Other dimensionality reduction techniques like multidimensional scaling (MDS) use the dissimilarities (generally Euclidean distances) between data points in the high-dimensional space to capture the relationships between

Table 1: Classification of methods for pose estimation.

Shape-based geometric methods	[6] [7] [5] [8] [9]
Model-based methods	[10] [11] [12] [1]
Appearance-based methods	[13] [14] [15] [16] [17] [18]
Template matching methods	[19] [20]
Dimensionality-reduction-based approaches	[4] [21] [22] [23] [24] [3] [2]

them. In recent years, a new group of non-linear approaches to dimensionality reduction have emerged, which assume that data points are embedded on a low-dimensional manifold in the ambient high-dimensional space. These have been grouped under the term "manifold learning," and some of the most often used manifold learning techniques in the last few years include Isomap [25], Locally Linear Embedding (LLE) [26], Laplacian eigenmaps [27], Local Tangent Space Alignment [28]. The interested reader can refer to [29] for a review of dimensionality reduction techniques.

In this work, different poses of the head, although captured in high-dimensional image feature spaces, are visualized as data points on a low-dimensional manifold embedded in the high-dimensional space [2, 4]. The dimensionality of the manifold is said to be equal to the number of degrees of freedom in the movement during data capture. For example, images of the human face with different angles of pose rotation (yaw, tilt and roll) can intrinsically be conceptualized as a 3D manifold embedded in image feature space.

In this work, we consider face images with pose angle views ranging from -90° to $+90^\circ$ from the FacePix database (detailed in Section 4.1), with only yaw variations. Figure 1 shows the 2-dimensional embeddings of face images with varying pose angles from FacePix database obtained with three different manifold learning techniques—Isomap, Locally Linear Embedding (LLE), and Laplacian eigenmaps. On close observation, one can notice that the face images are ordered by the pose angle. In all of the embeddings, the frontal view appears in the center of the trajectory, while views from the right and left profiles flank the frontal view, ordered by increasing pose angles. This ability to arrange face images by pose angle (which is the only changing parameter) during the process of dimensionality reduction explains the reason for the increased interest in applying manifold learning techniques to the problem of head pose estimation.

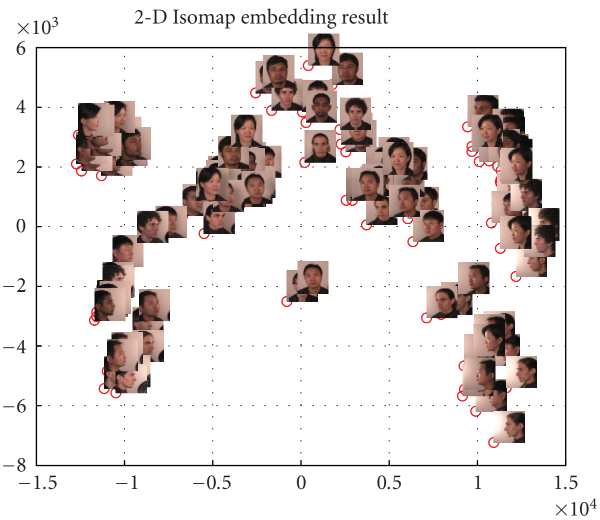
While face images of a single individual with varying poses lie on a manifold, the introduction of multiple individuals in the dataset of face images has the potential to make the manifold topologically unstable (see [2]). Figure 1 illustrates this point to an extent. Although the face images form an ordering by pose angle in the embeddings, face images from different individuals tend to form a clutter. While coarse pose angle estimation may work to a certain acceptable degree of error with these embeddings, accurate pose angle estimation requires more than what is available with these embeddings.

To obtain low-dimensional embeddings of face images ordered by pose angle independent of the number of individuals, we propose a supervised framework to manifold learning. The intuition behind this approach is that while image feature vectors may sometimes not abide by the intrinsic geometry underlying the objects of interest (in this case, faces), pose label information from the training data can help align face images on the manifold better, since the manifold is characterized by the degrees of freedom expressed by the head pose angle.

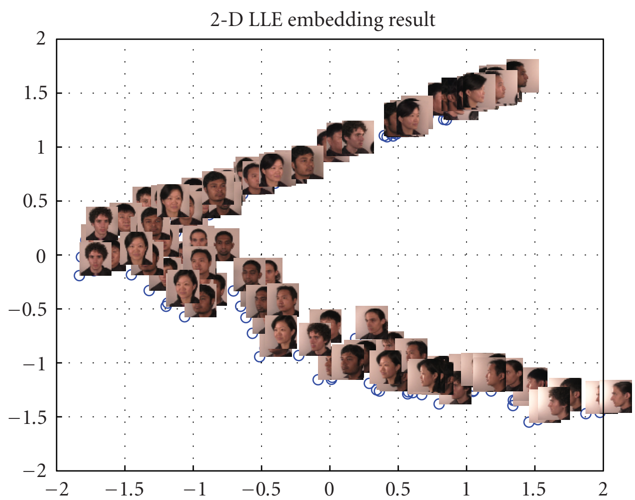
A more detailed analysis of the motivations for this work is captured in Figure 2. Fifty random face images were picked from the FacePix database. For each of these images, the local neighborhood based on the Euclidean distance was studied. The identity and the pose angle of k (=10) nearest neighbors was noted down. The average values of these readings are presented in Figure 2. It is evident from this figure that for most images, the nearest neighbors are dominated by other face images of the same person, rather than other face images with the same pose angle. Since manifold learning techniques are dependent on the choice of the local neighborhood of a data point for the final embedding, it is likely that this observation would distort the alignment of the manifold enough to make fine pose angle estimation difficult.

Having stated the motivation behind this work, the broad objectives of this work are to contribute to pattern recognition in biometrics by establishing a supervised form of manifold learning as a solution to accurate person-independent head pose angle estimation. These objectives are validated with experiments to show that the proposed supervised framework, called the Biased Manifold Embedding, provides superior results for accurate pose angle estimation over traditional linear (principal component analysis, e.g.) or non-linear (regular manifold learning techniques) dimensionality reduction techniques, which are often used in face analysis applications.

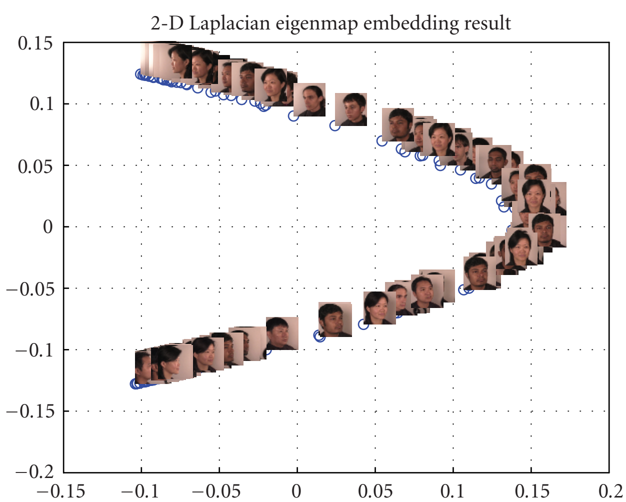
The contributions of this work lie in the proposition, validation and analysis of the Biased Manifold Embedding (BME) framework as a supervised approach to manifold-based dimensionality reduction with application to head pose estimation. This framework, although primarily formulated for a regression scenario, unifies other supervised approaches to manifold learning that have been proposed



(a) Embedding with the Isomap algorithm

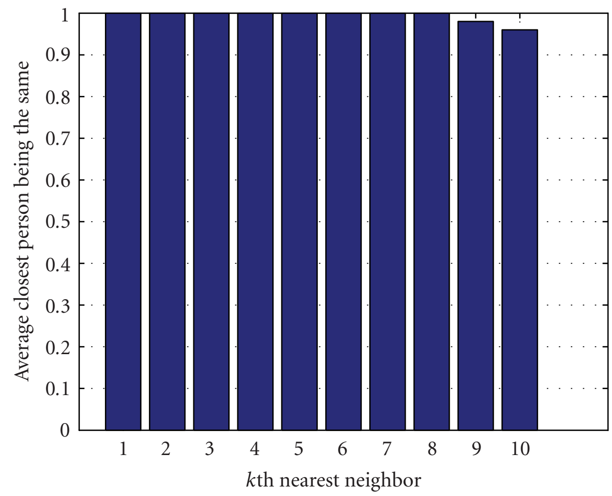


(b) Embedding with the LLE algorithm

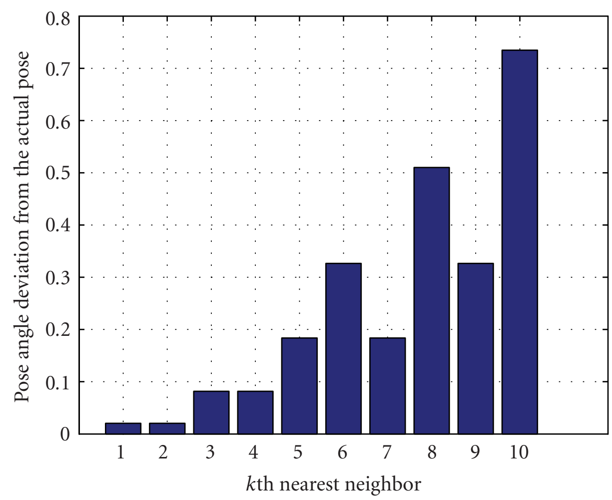


(c) Embedding with the Laplacian eigenmap algorithm

FIGURE 1: Embedding of face images with varying poses onto 2 dimensions.



(a) Analysis of the identity of the nearest neighbors. A 0.9 value for average closest person being the same indicates that 9 out of 10 images had the person himself/herself as the corresponding kth neighbor by Euclidean distance



(b) Analysis of the pose angle of the nearest neighbors

FIGURE 2: Analysis of the k (= 10) nearest neighbors (by Euclidean distance) of a face image in high-dimensional feature space. It is evident and intuitive that face images in the high-dimensional image feature space tend to have the face images of the same person as the closest neighbors. Since manifold learning methods are dependent on local neighborhoods for the entire construction; this could affect fine estimation of head pose angle. The more the number of individuals is, the worse the clutter becomes.

so far. The application of the framework to the problem of head pose estimation has been studied using images from the FacePix database, which contains face images with a granularity of 1° variations in pose angle. Both global and local approaches to manifold learning have been considered in the experimentation. Since it is difficult to obtain this level of granularity of pose angle in training data with biometric applications in the real world, the proposed framework has been evaluated with sparsely sampled data from the FacePix database. Considering that manifold learning methods are

Figure 3: The data capture setup for FacePix.

known to fail with sparsely sampled data [29, 30], these experiments also serve to evaluate the effectiveness of the proposed supervised framework for such data.

While this framework was proposed in our recent work [2] with initial results, the framework has been enhanced to provide a unified view of other supervised approaches to manifold learning in this work. A detailed analysis of the motivations, modification of the framework to unify other supervised approaches to manifold learning, the evaluation of the framework on sparse data samples, and comparison to other related approaches are novel contributions of this work.

A review of related work on manifold learning, head pose estimation, and other supervised approaches to manifold learning is presented in Section 2. Section 3 details the mathematical formulation of the Biased Manifold Embedding framework from a regression perspective, and extends it to classification problems. This section also discusses how the proposed framework unifies other supervised approaches to manifold learning. An overview of the FacePix database, details of the experimentation and the hypotheses tested for, and the corresponding results are presented in Section 4. Discussions and conclusions with pointers to future work follow in Sections 5 and 6.

2. RELATED WORK

A classification of different approaches to head pose estimation was presented in Section 1. In this section, we discuss approaches to pose estimation using manifold learning, that are related to the proposed framework, and review their performance and limitations. In addition, we also survey existing supervised approaches to manifold learning. So far, to the best of the authors' knowledge, these supervised techniques have not been applied to the head pose estimation problem, and hence, we limit our discussions to the main ideas in these formulations.

2.1. Manifold learning and pose estimation

Since the advent of manifold learning techniques less than a decade ago, a reasonable amount of work has been done using manifold-based dimensionality reduction techniques for head pose estimation. Chen et al. [22] considered multiview face images as lying on a manifold in high-dimensional feature space. They compared the effectiveness of kernel discriminant analysis against support vector machines in learning the manifold gradient direction in the high-dimensional feature space. The images in this work were synthesized from a 3D scan. Also, the application was restricted to a binary classifier with a small range of head pose angles between -10° and $+10^\circ$.

Raytchev et al. [4] studied the effectiveness of Isomap for head pose estimation against other view representation approaches like the Linear Subspace model and Locality Preserving Projections (LPP). While their experiments showed that Isomap performed better than the other two approaches, the face images used in their experiments were sampled at pose angle increments of 15°. In the discussion, the authors indicate that this dataset is insufficient to provide for experiments with accurate pose estimation. The least pose angle estimation error in all their experiments was 10.7°, which is rather high.

Hu et al. [24] developed a unified embedding approach for person-independent pose estimation from image sequences, where the embedding obtained from Isomap for a single individual was parametrically modeled as an ellipse. The ellipses for different individuals were subsequently normalized through scale, translation and rotation based transformations to obtain a unified embedding. A Radial Basis Function interpolation system was then used to obtain the head pose angle. The authors obtained good results with the datasets, but their approach relied on temporal continuity and local linearity of the face images, and hence was intended for image/video sequences.

In more recent work, Fu and Huang [3] presented an appearance-based strategy for head pose estimation using a supervised form of Graph Embedding, which internally used the idea of Locally Linear Embedding (LLE). They obtained a linearization of manifold learning techniques to treat out-of-sample data points. They assumed a supervised approach to local neighborhood-based embedding and obtained low pose estimation errors; however, their perspective of supervised learning differs from how it is addressed in this work.

In the last few years of the application of manifold learning techniques, there have been limitations that have been identified [29, 30]. While all these techniques capture the geometry of the data points in the high-dimensional space, the disadvantage of this family of techniques is the lack of a projection matrix to embed out-of-sample data points after the training phase. This makes the method more suited for data visualization, rather than classification/regression problems. However, the advantage of these techniques to capture the relative geometry of data points enthuses researchers to adopt this methodology to solve problems like head pose estimation, where the data is known to possess geometric relationships in a high-dimensional space.

These techniques are known to depend on a dense sampling of the data in the high-dimensional space. Also, Ge et al. [31] noted that these techniques do not remove correlation in high-dimensional spaces from their low-dimensional representations. The few applications of these techniques

Figure 4: Sample face images with varying pose and illumination from the FacePix database.

to pose estimation have not exposed the limitations yet—however, from a statistical perspective, these generic limitations intrinsically emphasise the requirement for the training data to be distributed densely across the surface of the manifold. In real-world applications like pose estimation, it is highly possible that the training data images may not meet this requirement. This brings forth the need to develop techniques that can work well with training data on sparsely sampled manifolds too.

2.2. Supervised manifold learning

In the last few years, there have been efforts to formulate supervised approaches to manifold learning. However, none of these approaches have explicitly been used for head pose estimation. In this section, we review the main ideas behind their formulations, and discuss the major novelties in our work, when compared to the existing approaches.

Ridder et al. [32] came up with one of the earliest supervised frameworks for manifold learning. Their framework was centered around the idea of defining a new distance metric for Locally Linear Embedding, which increased inter-class distances and decreased intra-class distances. This modified distance metric was used to compute the dissimilarity matrix, before computing the adjacency graph which is used in the dimensionality reduction process. Vlassis et al. [33] formulated a supervised approach that was intended towards identifying the intrinsic dimensionality of given data using statistical methods, and using the computed dimensionality for further analysis.

Li and Guo [34] proposed a supervised Isomap algorithm, where a separate geodesic distance matrix is constructed for the training data from each class. Subsequently, these class-specific geodesic distance matrices are merged into a discriminative global distance matrix, which is used for the multidimensionality scaling step. Vlachos et al. [35] proposed the WeightedIso method, where the Euclidean distance between data samples is scaled with a constant factor $\lambda(<1)$ if the class labels of the samples are the same. Geng et al. [36] extended the work from Vlachos et al. towards visualization applications, and proposed the S-isomap (supervised isomap), where the dissimilarity between two points is defined differently from the regular geodesic distance. The dissimilarity is defined in terms of an exponential factor of the Euclidean distance, such that the intraclass distance never exceeds 1, and the interclass distance never falls below $1 - \alpha$, where α is a parameter that can be tuned based on the application.

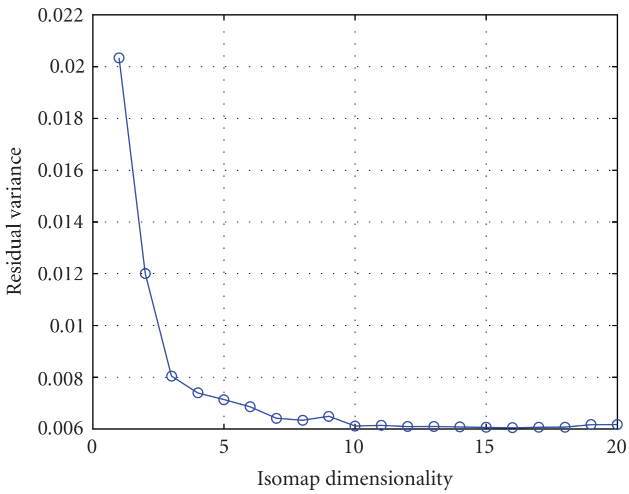
Zhao et al. [37] proposed a supervised LLE (SLLE) algorithm in the space of face images preprocessed using Independent Component Analysis. Their SLLE algorithm constructs these neighborhood graphs with a strict constraint imposed: only those points in the same cluster as the point under consideration can be its neighbors. In other words, the primary focus of the proposed SLLE is restricted to reveal and preserve the neighborhood in a cluster scope.

The approaches to supervised manifold learning discussed above primarily consider the problem from a classification/clustering perspective. In our work, we view the class labels (pose labels) as possessing a distance metric by themselves, that is, we approach the problem from a regression perspective. However, we also illustrate how it can be applied to classification problems. In addition, we show how the proposed framework unifies the existing approaches. The mathematical formulation of the proposed framework is discussed in the next section.

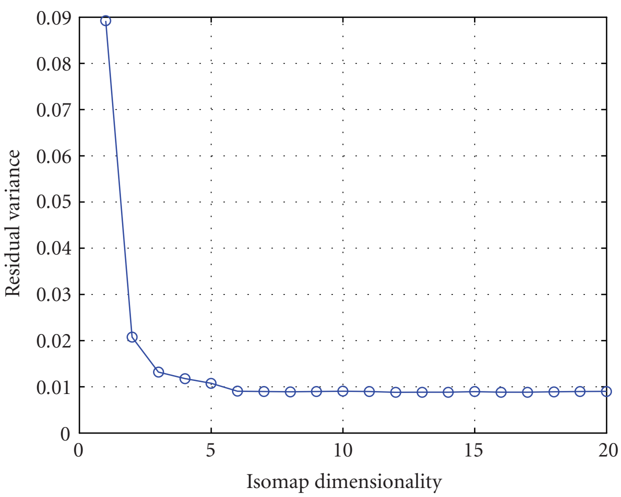
3. BIASED MANIFOLD EMBEDDING: THE MATHEMATICAL FORMULATION

In this section, we discuss the mathematical formulation of the Biased Manifold Embedding approach as applied in the head pose estimation problem. In addition, we then illustrate how this framework unifies other existing supervised approaches to manifold learning.

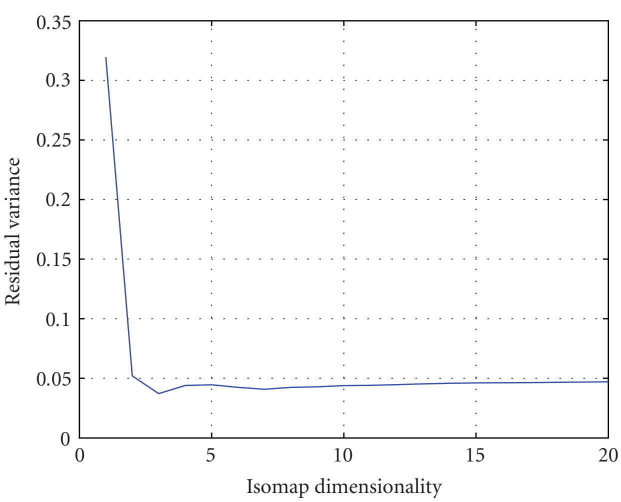
Manifold learning methods, as illustrated in Section 1, align face images with varying poses by an ordering of the pose angle in the low-dimensional embeddings. However, the choice of image feature vectors, presence of image noise and the introduction of the face images of different individuals in the training data can distort the geometry of the manifold. To ensure the alignment, we propose the Biased Manifold Embedding framework, so that face images whose pose angles are closer to each other are maintained nearer to each other in the low-dimensional embedding, and images with farther pose angles are placed farther, irrespective of the



(a) Face images with 5° pose angle intervals



(b) Face images with 2° pose angle intervals



(c) Face images with 1° pose angle intervals

FIGURE 5: Plots of the residual variances computed after embedding face images of 5 individuals using Isomap.

(a) Gray scale image

(b) Laplacian of Gaussian (LoG) tranformed image

FIGURE 6: Image feature spaces used for the experiments.

identity of the individual. In the proposed framework, the distances between data points in the high-dimensional feature space are biased with distances between the pose angles of corresponding images (and hence, the name). Since a distance metric can easily be defined on the pose angle values, the problem of finding closeness of pose angles is straightforward.

We would like to modify the dissimilarity/distance matrix between the set of all training data points with a factor of the pose angle dissimilarities between the points. We define the modified biased distance between a pair of data points to be of the fundamental form:

$$\widetilde{D}(i,j) = \lambda_1 \times D(i,j) + \lambda_2 \times f(P(i,j)) \times g(D(i,j)), \quad (1)$$

where $D(i,j)$ is the Euclidean distance between two data points x_i and x_j, $\widetilde{D}(i,j)$ is the modified biased distance, $P(i,j)$ is the pose distance between x_i and x_j, f is any function of the pose distance, g is any function of the original distance between the data samples, and λ_1 and λ_2 are constants. While we defined this formulation after empirical evaluations of several formulations for the dissimilarity matrix, we found that this formulation, in fact, unifies other existing supervised approaches to manifold learning that modify the dissimilarity matrix.

In general, the function f could be picked from the family of reciprocal functions ($f \in \mathcal{F}_R$) based on an application. In this work, we set $\lambda_1 = 0$ and $\lambda_2 = 1$ in (1), function g as the constant function (= 1), and the function f as

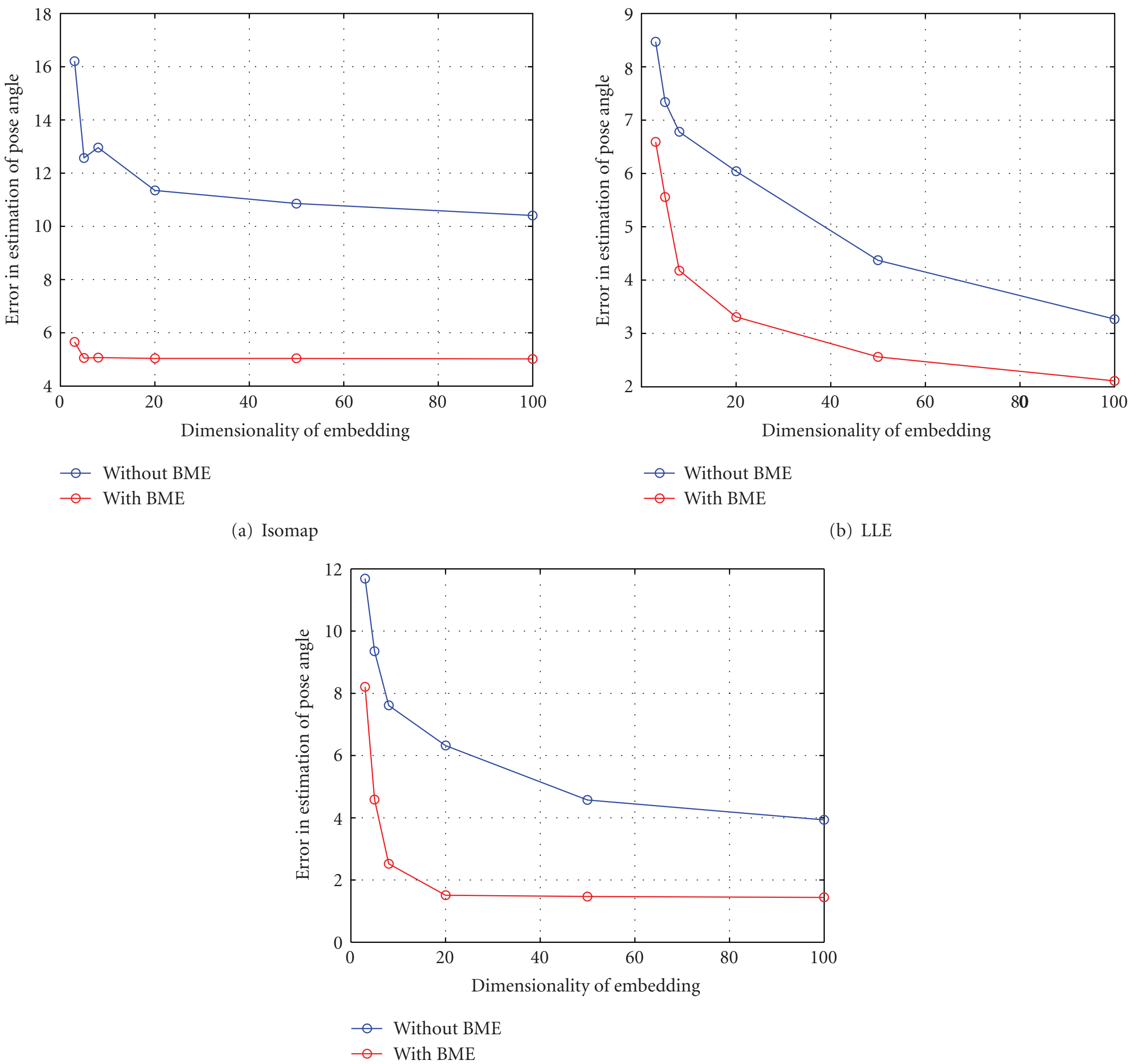
$$f(P(i,j)) = \frac{1}{\max_{m,n} P(m,n) - P(i,j)}. \quad (2)$$

This function could be replaced by an inverse exponential or quadratic function of the pose distance, for example. To ensure that the biased distance values are well-separated for different pose distances, we multiply this quantity by a function of the pose distance:

$$\widetilde{D}(i,j) = \frac{\alpha(P(i,j))}{\max_{m,n} P(m,n) - P(i,j)} * D(i,j), \quad (3)$$

where the function α is directly proportional to the pose distance, $P(i,j)$, and is defined in our work as

$$\alpha(P(i,j)) = \beta * |P(i,j)|, \quad (4)$$



(a) Isomap

(b) LLE

(c) Laplacian eigenmap

FIGURE 7: Pose estimation results of the BME framework against the traditional manifold learning technique with the gray scale pixel feature space. The red line indicates the results with the BME framework.

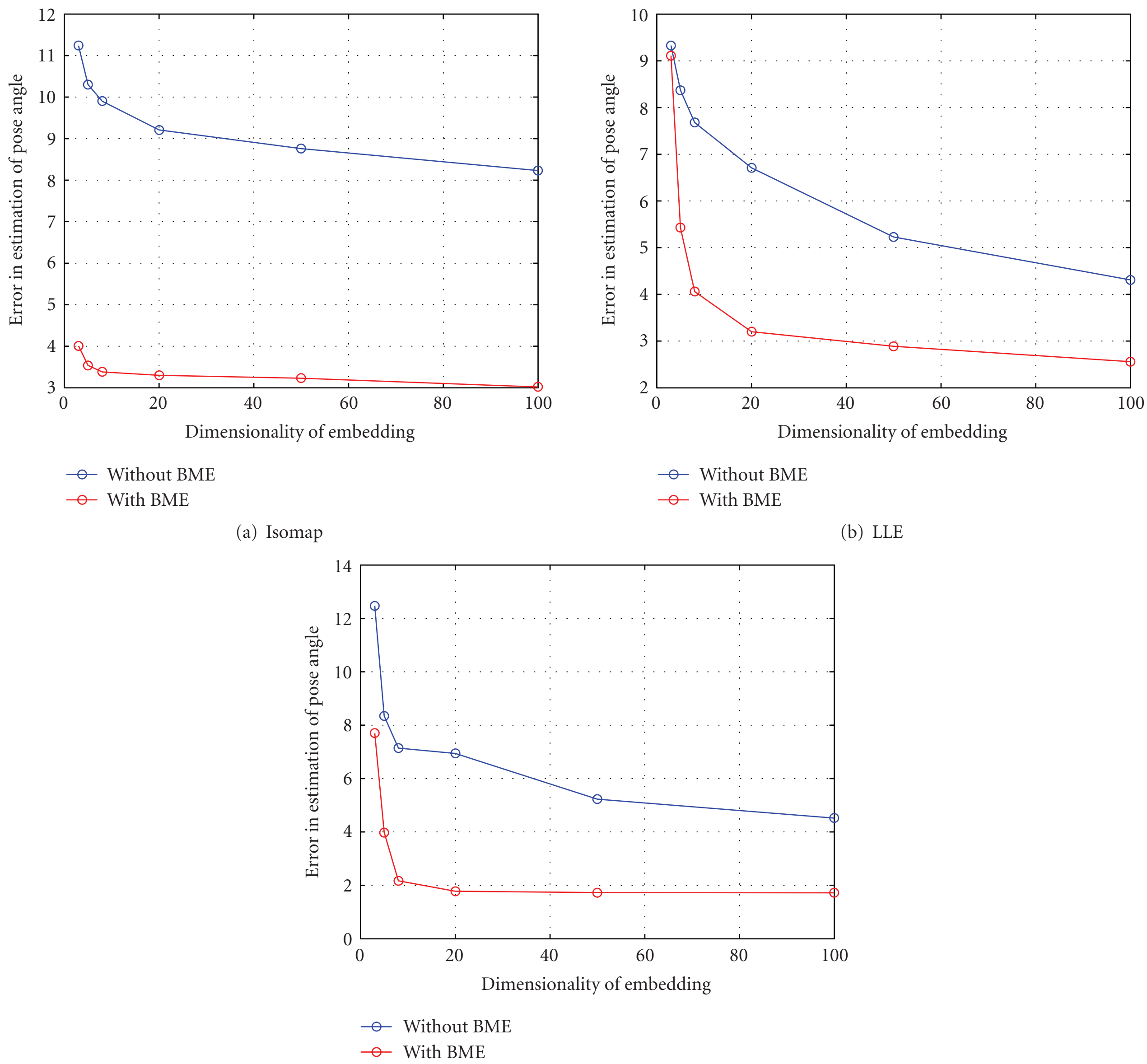
where β is a constant of proportionality and allows parametric variation for performance tuning. In our current work, we used the pose distance as the one-dimensional distance, that is, $P(i,j) = |P_i - P_j|$, where P_k is the pose angle of x_k.

In summary, the biased distance between a pair of points can be given by

$$\widetilde{D}(i,j) = \begin{cases} \dfrac{\alpha(P(i,j))}{\max_{m,n} P(m,n) - P(i,j)} * D(i,j), & P(i,j) \neq 0, \\ 0, & P(i,j) = 0. \end{cases} \quad (5)$$

This biased distance matrix is used for Isomap, LLE and Laplacian eigenmaps to obtain a pose-ordered low-dimensional embedding. In case of Isomap, the geodesic distances are computed using this biased distance matrix. The LLE and Laplacian eigenmaps algorithms are modified to use these distance values to determine the neighborhood of each data point. Since the proposed approach does not alter the algorithms in any other way other than the computation of the biased dissimilarity matrix, it can easily be extended to other manifold-based dimensionality reduction techniques which rely on the dissimilarity matrix.

In the proposed framework, the function $P(i,j)$ is defined in a straightforward manner for regression problems. Further, the same framework can also be extended to classification problems, where there is an inherent ordering in the class labels. An example of an application with such



(a) Isomap

(b) LLE

(c) Laplacian eigenmap

Figure 8: Pose estimation results of the BME framework against the traditional manifold learning technique with the Laplacian of Gaussian (LoG) feature space. The red line indicates the results with the BME framework.

a problem is head pose classification. Sample class labels could be "looking to the right," "looking straight ahead," "looking to the left," "looking to the far left," and so on. The ordering in these class labels can be used to define a distance metric. For example, if the class labels are indexed by an ordering $k = 1, 2, \ldots, n$ (where n is the number of class labels), a simple expression for $P(i, j)$ is

$$P(i, j) = \gamma \times \mathrm{dist}(|i - j|), \tag{6}$$

where i and j are the indices of the corresponding class labels of the training data samples. The dist function could just be the identity function, or could be modified depending on the application.

3.1. A unified view of other supervised approaches

In the next few paragraphs, we discuss briefly how the existing supervised approaches to manifold learning are special cases of the Biased Manifold Embedding framework. Although this discussion is not directly relevant to the pose estimation problem, this shows the broader appeal of this idea.

Ridder et al. [32] proposed a supervised LLE approach, where the distances between the samples are artificially increased if the samples belonged to different classes. If the samples are from the same class, the distances are left unchanged. The modified distances are given by

$$\Delta' = \Delta + \alpha \times \max(\Delta)\Lambda, \qquad \alpha \in [0, 1]. \tag{7}$$

Going back to (1), we arrive at the formulation of Ridder et al. by choosing $\lambda_1 = 1$, $\lambda_2 = \alpha \times \max(\Delta)$, function $g(D(i,j)) = 1$ for all i, j, and function $f(P(i,j)) = \Lambda$.

Li and Guo [34] proposed the SE-Isomap (Supervised Isomap with Explicit Mapping), where the geodesic distance matrix is constructed differently for intra-class samples, and is retained as is for inter-class data samples. The final distance matrix, called the discriminative global distance matrix G, is of the form

$$G = \begin{bmatrix} \rho_1 G_{11} & G_{12} \\ G_{21} & \rho_2 G_{22} \end{bmatrix}. \tag{8}$$

Clearly, this representation very closely resembles the choice of parameters we have chosen in our pose estimation work. In (1), the formulation of Li and Guo would simply mean choosing $\lambda_1 = 0$, $\lambda_2 = 1$, function $f(P(i,j)) = 1$, and function $g(D(i,j))$ can be defined as

$$g(D(i,j)) = \begin{cases} D(i,j), & P(i) \neq P(j), \\ \rho_i \times D(i,j), & P(i) = P(j). \end{cases} \tag{9}$$

The work of Vlachos et al. [35]—the WeightedIso method—is exactly the same in principle as Li and Guo. For data samples belonging to the same class, the distance is scaled by a factor $1/\alpha$, where $\alpha > 1$; else, the distance is left undisturbed. This can be exactly formulated as discussed above for Li and Guo. The work of Geng et al. [36] is based on the WeightedIso method, and the authors extended the WeightedIso method with a different dissimilarity matrix (which would just mean a different definition for $D(i,j)$ in the proposed BME framework), and parameters to control the distance values.

Zhao et al. [37] formulated the S-LLE (supervised LLE) method, where the distance between points that belonged to different classes was set to infinity, that is, the neighbors of a particular data point had to belong to the same class as the point. Again, this would be rather straight-forward in the BME framework, where the function $g(D(i,j))$ can be defined as

$$g(D(i,j)) = \begin{cases} \infty, & P(i) \neq P(j), \\ D(i,j), & P(i) = P(j). \end{cases} \tag{10}$$

Having formulated the Biased Manifold Embedding framework, we discuss the experiments performed and the results obtained in the next section.

4. BIASED MANIFOLD EMBEDDING FOR HEAD POSE ESTIMATION: EXPERIMENTATION AND RESULTS

4.1. The FacePix database

In this work, we have used the FacePix database [38] built at the Center for Cognitive Ubiquitous Computing (CUbiC) for our experiments and evaluation. Earlier work on face analysis have used databases such as FERET, XM2VTS, the CMU PIE Database, AT & T, Oulu Physics Database, Yale Face Database, Yale B Database, and MIT Database for evaluating the performance of algorithms. Some of these databases provide face images with a wide variety of pose angles and illumination angles. However, none of them use a precisely calibrated mechanism for acquiring pose and illumination angles. To achieve a precise measure of recognition robustness, FacePix was compiled to contain face images with pose and illumination angles annotated in 1 degree increments. Figure 3 shows the apparatus that is used for capturing the face images. A video camera and a spot light are mounted on separate annular rings which rotate independently around a subject seated in the center. Angle markings on the rings are captured simultaneously with the face image in a video sequence, from which the required frames are extracted.

The FacePix database consists of three sets of face images: one set with pose angle variations, and two sets with illumination angle variations. Each of these sets are composed of a set of 181 face images (representing angles from -90° to $+90^\circ$ at 1 degree increments) of 30 different subjects, with a total of 5430 images. All the face images (elements) are 128 pixels wide and 128 pixels high. These images are normalized, such that the eyes are centered on the 57th row of pixels from the top, and the mouth is centered on the 87th row of pixels. The pose angle images appear to rotate such that the eyes, nose, and mouth features remain centered in each image. Also, although the images are down sampled, they are scaled as much horizontally as vertically, thus maintaining their original aspect ratios. Figure 4 provides two examples extracted from the database, showing pose angles and illumination angles ranging from -90° to $+90^\circ$ in steps of 10°. For earlier work using images from this database, please refer [38]. There is ongoing work on making this database publicly available.

4.2. Finding the intrinsic dimensionality of the face images

An important component of manifold learning applications is the computation of the intrinsic dimensionality of the dataset provided. Similar to how linear dimensionality reduction techniques like PCA use the measure of captured variance to arrive at the number of dimensions, manifold learning techniques are dependent on knowing the intrinsic dimensionality of the manifold embedded in the high-dimensional feature space.

We performed a preliminary analysis of the dataset to extract its intrinsic dimensionality, similar to what was performed in [25]. Isomap was used to perform nonlinear dimensionality reduction on a set of face images from 5 individuals. Different pose intervals of the face images were selected to vary the density of the data used for embedding. The residual variances after computation of the embedding are plotted in Figure 5. The subfigures illustrate that most of the residual variance is captured in one dimension of the embedding. This goes to prove that there is only one dominant dimension in the dataset. As the pose intervals used for the embedding becomes lesser, that is, the density of the data becomes higher, this observation is even more clearly noted. The data captured in the FacePix database have pose variations only along one degree of freedom (the yaw), and this result corroborates the fact that these face images could

TABLE 2: Results of head pose estimation using principal component analysis and manifold learning techniques for dimensionality reduction, in the gray scale pixel feature space.

Dimension of embedding	Error in pose estimation			
	PCA	Isomap	LLE	Laplacian eigenmap
10	11.37°	12.61°	6.60°	7.72°
20	9.90°	11.35°	6.04°	6.32°
40	9.39°	10.98°	4.91°	5.08°
50	8.76°	10.86°	4.37°	4.57°
75	7.83°	10.67°	3.86°	4.17°
100	7.27°	10.41°	3.27°	3.93°

TABLE 3: Results of head pose estimation using principal component analysis and manifold learning techniques for dimensionality reduction, in the LoG feature space.

Dimension of embedding	Error in pose estimation			
	PCA	Isomap	LLE	Laplacian eigenmap
10	9.80°	9.79°	7.41°	7.10°
20	8.86°	9.21°	6.71°	6.94°
40	8.54°	8.94°	5.80°	5.91°
50	8.03°	8.76°	5.23°	5.23°
75	7.92°	8.47°	4.83°	4.89°
100	7.78°	8.23°	4.31°	4.52°

be visualized as lying on a low-dimensional (ideally, one-dimensional) manifold in the feature space.

4.3. Experimentation setup

The setup of the experiments conducted in the subsequent sections is described here. All of these experiments were performed with a set of 2184 face images, consisting of 24 individuals with pose angles varying from -90° to $+90^\circ$ in increments of 2°. The images were subsampled to 32×32 resolution, and two different feature spaces of the images were considered for the experiments. The results presented here include the grayscale pixel intensity feature space and the Laplacian of Gaussian (LoG) transformed image feature space (see Figure 6). The LoG transform, which captures the edge map of the face images, was used since pose variations in face images can be considered a result of geometric transformation, and texture information can be considered redundant. The images were subsequently rasterized and normalized.

Unlike linear dimensionality reduction methods like Principal Component Analysis, manifold learning techniques lack a well-defined approach to handle out-of-sample extension data points. Different methods have been proposed [39, 40] to capture the mapping from the high-dimensional feature space to the low-dimensional embedding. We adopted the generalized regression neural network (GRNN) with radial basis functions to learn the nonlinear mapping. GRNNs are known to be a one-pass "learning" system and are known to work well with sparsely sampled data. This approach has been adopted by earlier researchers [37]. The parameters involved in training the network are minimal (only the spread of the radial basis function), thereby facilitating better evaluation of the proposed framework. Once the low-dimensional embedding was obtained, linear multivariate regression was used to obtain the pose angle of the test image. To ensure generalization of the framework, 8-fold cross-validation was used in these experiments. In this validation model, 1911 face images (91 images each of 21 individuals) were used for the training phase in each fold, while all the remaining images were used in the testing phase. The parameters, that is, the number of neighbors used and the dimensionality of embedding, were chosen empirically.

4.4. Using manifold learning over linear dimensionality reduction for pose estimation

Traditional approaches to pose estimation that rely on dimensionality reduction use linear techniques (PCA, to be specific). However, with the assumption that face images with varying poses lie on a manifold, nonlinear dimensionality reduction would be expected to perform better. We performed experiments to compare the performance of manifold learning techniques with principal component analysis. The results of head pose estimation comparing PCA against manifold learning techniques with the experimentation setup described in the previous subsection are tabulated in Tables 2 and 3. While these results have been noted as obtained, our empirical observations indicated that the number of significant digits could be considered up to one decimal place.

As the results illustrate, while Isomap and PCA perform very similarly, both the local approaches, that is, Locally Linear Embedding and Laplacian eigenmaps, show 3-4° improvement in pose angle estimation over PCA, consistently.

Table 4: Summary of head pose estimation results from related approaches in recent years.

Reference	Method	Best result in pose angle estimation: error/accuracy	Notes
[22]	Fisher manifold learning	About 3°	Face images only in $[-10^\circ, 10^\circ]$ interval
[18]	Kernel PCA + support vector machines	97%	Face images only in 10° intervals (this was framed as a classification problem of identifying the pose angle as one of these intervals)
[4]	Isomap	About 11°	Face images sampled at 15° increments
[4]	LPP	About 15°	Face images sampled at 15° increments
[3]	LEA	About 2°	Best results so far
Current work	BME using Laplacian eigenmap	About 2°	Results similar to [3]
Current work	BME using Isomap, LLE	About 3°	—

Table 5: Results from experiments performed with sparsely sampled training dataset for each of the manifold learning techniques with and without the BME framework on the gray scale pixel feature space. The error in the head pose angle estimation is noted.

Number of training images	Error using isomap		Error using LLE		Error using Laplacian eigenmap	
	without BME	with BME	without BME	with BME	without BME	with BME
570	12.13°	3.26°	5.95°	5.88°	10.27°	3.84°
475	11.70°	6.01°	6.58°	6.95°	9.47°	3.71°
380	8.19°	7.61°	6.47°	6.72°	9.59°	4.72°
285	8.39°	8.75°	6.36°	6.71°	9.12°	5.61°
190	8.75°	8.58°	6.77°	7.03°	10.05°	7.76°
95	11.27°	9.22°	9.43°	8.45°	15.44°	14.54°

4.5. *Supervised manifold learning for person-independent pose estimation: Experiments with Biased Manifold Embedding*

While manifold learning techniques demonstrate reasonably good results for pose estimation over linear dimensionality reduction techniques, we hypothesize that the supervised approach to manifold learning performs better for accurate results with person-independent pose estimation. In our next set of experiments, we evaluate this hypothesis. The error in the pose angle estimation process is used as the criterion for the evaluation.

The proposed BME framework was applied to face images from the FacePix database, and the performance was compared against the performance of regular manifold learning techniques. These experiments were performed against global (Isomap) and local (Locally Linear Embedding and Laplacian eigenmaps) approaches to manifold learning. The error in the estimated pose angle (against the ground truth from the FacePix database) was used to evaluate the performance.

The results of these experiments are presented in Figures 7 and 8. The blue line indicates the performance of the manifold learning techniques, while the red line stands for the performance from the Biased Manifold Embedding approach. As evident, the error significantly drops with the proposed approach. All of the approaches perform better with the LoG feature space, as compared to using plain gray scale pixel intensities. This corroborates the intuitive assumption that the head pose estimation problem is one of geometry of face images, and the texture of the images can be considered redundant. However, we believe that it would be worthwhile to perform a more exhaustive analysis with other feature spaces as part of our future work. Also, it is clear from the error values obtained that the BME framework substantially improves the head pose estimation performance, when compared to other manifold learning techniques or principal component analysis.

It can also be observed that the results obtained from the local approaches, that is, Locally Linear Embedding and Laplacian eigenmaps, far outperform the global approach, viz. Isomap. Considering that isomap is known to falter when

Table 6: Results from experiments performed with sparsely sampled training dataset with and without the BME framework on the LoG feature space.

Number of training images	Error using Isomap		Error using LLE		Error using Laplacian eigenmap	
	without BME	with BME	without BME	with BME	without BME	with BME
570	10.63°	3.19°	8.76°	7.99°	9.01°	3.57°
475	12.08°	3.73°	8.08°	7.63°	8.56°	3.99°
380	11.34°	6.40°	8.16°	8.48°	8.47°	5.00°
285	13.96°	6.66°	8.14°	8.49°	9.30°	6.69°
190	15.46°	6.96°	8.72°	8.68°	12.27°	8.84°
95	11.93°	8.59°	8.77°	8.77°	30.17°	15.79°

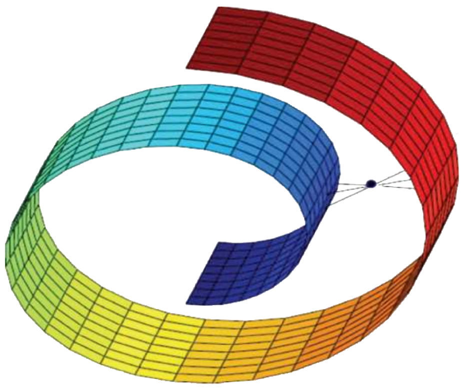


Figure 9: Example of topological instabilities that affect Isomap's performance. An outlier could short-circuit the geometry of the manifold and destroy its geometrical structure. In such a case, global approaches like Isomap fail to find an appropriate low-dimensional embedding.

there is topological instability [41]; the relatively low performance with both the feature spaces suggests that the manifold of face images constructed from the FacePix database may be topologically unstable. In reality, this would mean that there are face images which short-circuit the manifold in a way that the computation of geodesic distances is affected (see Figure 9). There have been recent approaches to overcome the topological instability by removing critical outliers in a preprocessing step [40].

4.6. Comparison with related pose estimation work

In comparing related approaches to pose estimation which have different experimental design criteria, the results are summarized below in Table 4. The results obtained from the BME framework match the best results so far obtained by [3], considering face images with pose angle intervals of 1°. The best results are obtained when BME is used with Laplacian eigenmap. When LLE or Isomap is used, the error goes marginally higher and hovers about 3°.

4.7. Experimentation with sparsely sampled data

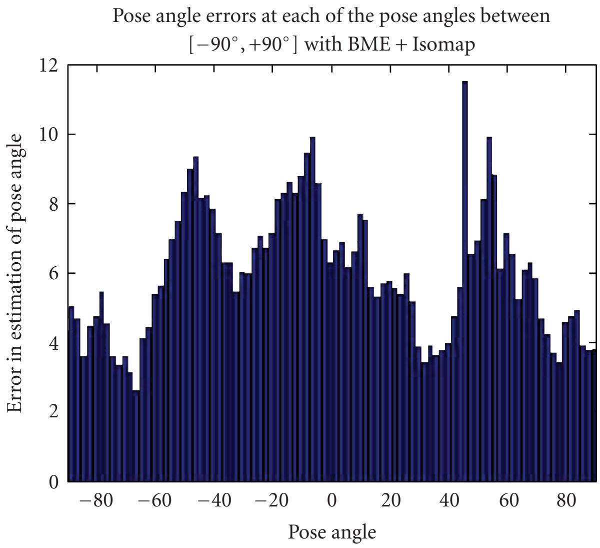
Manifold learning techniques have been known to perform poorly on sparsely sampled datasets [29]. Hence, in our next set of experiments, we propose that the BME framework, through supervised manifold learning, performs reasonably well even on sparse samples, and evaluate this hypothesis.

In these experiments, we sampled the available set of face images sparsely (by pose angle) and used this sparse sample of the face images dataset for training, before testing with the entire dataset. In these experiments, face images of all the 30 individuals in the FacePix database were used. The set of training images included face images in pose angle intervals of 10°, that is, only 19 out of the total 181 images for each individual were used in the training phase. Subsequently, the number of training images (total number of images is 5430) was progressively reduced in steps to observe the performance. These experiments were carried out for Isomap, LLE and Laplacian eigenmaps for both the feature spaces. To maintain uniformity of results and to aid comparison, all these trials embedded the face images onto a 8-dimensional space, and 50 neighbors were used for constructing the embedding (as in the earlier section). The results are presented in Tables 5 and 6. Note the results obtained with BME and without BME for Isomap and Laplacian eigenmap in both these tables. The results show significant reduction in error. However, the results for LLE do not reflect this observation.

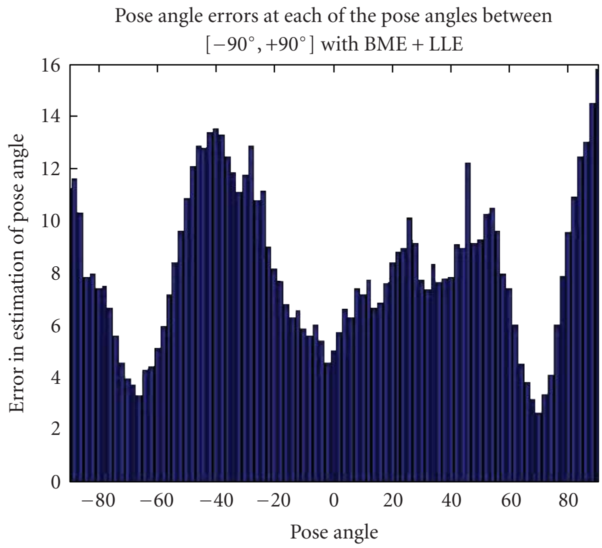
The results validate our hypothesis that the BME framework performs better even with sparsely sampled datasets. With Isomap and Laplacian eigenmap, the application of the BME framework improves the performance of pose estimation substantially. However, we note that Locally Linear Embedding performed as well even without the Biased Manifold Embedding framework. This suggests that in tasks of unsupervised learning (like clustering), where there are no class labels to supervise the learning process, Locally Linear Embedding may be a good technique to apply for sparsely sampled datasets.

5. DISCUSSION

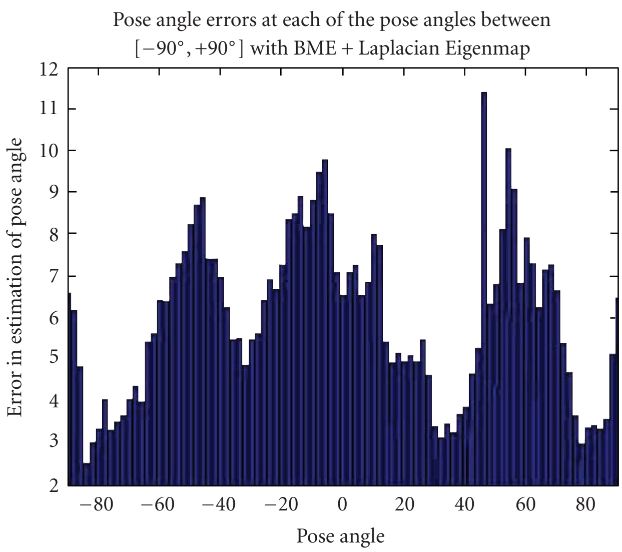
The results from the previous section show the merit of the proposed supervised framework for manifold learning as effective for head pose estimation. As mentioned before, using the pose information to supervise the manifold learning process may be looked at as obtaining a better estimate of the geometry of the manifold, based on the exact parameters/degrees of freedom (in our case, the pose angles) that define the intrinsic dimensionality of the manifold. This in



(a) Biased manifold embedding with Isomap



(b) Biased manifold embedding with LLE



(c) Biased manifold embedding with Laplacian eigenmap

Figure 10: Analysis of the average error in pose estimation for each of the views between $[-90^\circ, +90^\circ]$.

turn improves the performance of the head pose estimation methodology.

As an integral focus for biometric systems that require person-independent head pose estimation, our observations from the experiments indicate that local approaches to manifold learning (Locally Linear Embedding and Laplacian eigenmaps) provide the best results for head pose estimation with a dataset like FacePix. As mentioned before, the relatively low performance of Isomap could be attributed to a possible instability in the topology of the manifold, which could be caused by some outlier face images. A deeper study of the detection of the presence of such an instability, and the kind of face images that may cause this instability, is certainly warranted, and will be considered in our future work.

For a better understanding of the results, we analyzed how the errors in the pose estimation process were spread out on the interval $[-90^\circ, +90^\circ]$. Figure 10 shows the head pose estimation error in each of the views in this pose angle interval. While we expected to see a better performance at the frontal view, this was not very evident in any of the three approaches. We also hoped to identify particular regions of pose angle views of face images where the framework consistently performs relatively poor. However, these plots do not provide any coherent information on identifying such views of face images.

The analysis of the performance of the techniques on sparsely sampled set of face images reveals that while Isomap and Laplacian eigenmaps provide increased performance when there is an increase in the number of training images, Locally Linear Embedding provides consistent results and may be the choice when the dataset is sparsely sampled, and the number of available samples is less. Another observation from these results showed that even if the training data is sparsely sampled in terms of the pose angles, populating the dataset with more samples of face images of other individuals helps compensate for the lack of face images in the intermediate pose angle regions to a reasonable extent.

It is also important to note that while the Biased Manifold Embedding framework holds promise, the technique works better as the number of face images available for training is increased, and as the spectrum of training images becomes more representative of the test face images. Further, since we have used generalized regression neural networks (GRNNs) in this work, GRNNs [42] are also known to perform better with more training samples. However, as the training sample set gets larger, the memory requirements for a GRNN for computation become heavier, and this may be a cause for concern.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed an approach to person-independent head pose estimation based on a novel framework called the Biased Manifold Embedding for supervised manifold learning. Under the credible assumption that face images with varying pose angles lie on a low-dimensional manifold, nonlinear dimensionality reduction based on manifold learning techniques possesses strong potential for face analysis in biometric applications. We compared the proposed framework with regularly used approaches like principal component analysis and other manifold learning techniques, and we found the results to be reasonably good for head pose estimation. While the framework was primarily intended for regression problems, we have also shown how this framework unifies earlier approaches to supervised manifold learning. The results that we obtained from pose estimation using the FacePix database match the best results obtained so far and demonstrate the suitability of this approach for similar applications.

As future work, we wish to extend this work to experiment on other datasets like the USF database [3], which have similar granularity of pose angle in the face image database. We hope that this would provide more inputs on the generalization of this framework. We plan to implement this as part of a wearable platform to perform real-time pose classification from a live video stream, to study its applicability in real-world scenarios. We also hope to study the potential detection of the existence of topological instabilities that may affect the performance of global manifold learning approaches like Isomap, and come up with solutions to circumvent such issues in pose estimation and other face analysis applications. Further, as manifold learning techniques continue to be applied in pose estimation and similar applications, it becomes imperative to carry out an exhaustive study to identify the kind of image feature spaces that are most amenable to manifold-based assumptions and analysis.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation NSF-ITR Grant no. IIS-0326544.

REFERENCES

[1] L. M. Brown and Y.-L. Tian, "Comparative study of coarse head pose estimation," in *Proceedings of the IEEE Workshop on Motion and Video Computing*, pp. 125–130, Orlando, Fla, USA, December 2002.

[2] V. N. Balasubramanian, J. Ye, and S. Panchanathan, "Biased manifold embedding: a framework for person-independent head pose estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, Minneapolis, Minn, USA, June 2007.

[3] Y. Fu and T. S. Huang, "Graph embedded analysis for head pose estimation," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (AFGR '06)*, vol. 2006, pp. 3–8, Southampton, UK, April 2006.

[4] B. Raytchev, I. Yoda, and K. Sakaue, "Head pose estimation by nonlinear manifold learning," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 4, pp. 462–466, Cambridge, UK, August 2004.

[5] M. T. Wenzel and W. H. Schiffmann, "Head pose estimation of partially occluded faces," in *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision (CRV '05)*, pp. 353–360, Victoria, Canada, May 2005.

[6] J. Heinzmann and A. Zelinsky, "3D facial pose and gaze point estimation using a robust real-time tracking paradigm," in *Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition (AFGR '98)*, pp. 142–147, Nara, Japan, April 1998.

[7] M. Xu and T. Akatsuka, "Detecting head pose from stereo image sequence for active face recognition," in *Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition (AFGR '98)*, pp. 82–87, Nara, Japan, April 1998.

[8] K. N. Choi, P. L. Worthington, and E. R. Hancock, "Estimating facial pose using shape-from-shading," *Pattern Recognition Letters*, vol. 23, no. 5, pp. 533–548, 2002.

[9] Y. Hu, L. Chen, Y. Zhou, and H. Zhang, "Estimating face pose by facial asymmetry and geometry," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (AFGR '04)*, pp. 651–656, Seoul, Korea, May 2004.

[10] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[11] H. Rowley, S. Baluja, and T. Kanade, "Neural network based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[12] S. Gundimada and V. Asari, "An improved SNoW based classification technique for head-pose estimation and face detection," in *Proceedings of the 34th Applied Imagery Pattern Recognition Workshop (AIPR '05)*, pp. 94–99, Washington, DC, USA, October 2005.

[13] Y. Wei, L. Fradet, and T. Tan, "Head pose estimation using gabor eigenspace modeling," in *Proceedings of International Conference on Image Processing (ICIP '02)*, vol. 1, pp. 281–284, Rochester, NY, USA, September 2002.

[14] P. Fitzpatrick, "Head pose estimation without manual initialization," Tech. Rep., AI Lab, MIT, Cambridge, UK, 2000.

[15] B. Tordoff, W. W. Mayol, T. D. Campos, and D. Murray, "Head pose estimation for wearable robot control," in *Proceedings of the 13th British Machine Vision Conference (BMVC '02)*, pp. 807–816, Cardiff, UK, September 2002.

[16] S. O. Ba and J.-M. Odobez, "A probabilistic framework for joint head tracking and pose estimation," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 4, pp. 264–267, Cambridge, UK, August 2004.

[17] S. O. Ba and J.-M. Odobez, "Evaluation of multiple cue head pose estimation algorithms in natural environments," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '05)*, pp. 1330–1333, Amsterdam, The Netherlands, July 2005.

[18] S. Z. Li, Q. D. Fu, L. Gu, B. Scholkopf, Y. Cheng, and H. Zhang, "Kernel machine based learning for multi-view face detection and pose estimation," in *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV '01)*, vol. 2, pp. 674–679, Vancouver, BC, Canada, July 2001.

[19] M. Bichsel and A. Pentland, "Automatic interpretation of human head movements," Tech. Rep. 186, Vision and Modeling Group, MIT Media Laboratory, 1993.

[20] S. J. McKenna and S. Gong, "Real-time face pose estimation," *Real-Time Imaging*, vol. 4, pp. 333–347, 1998.

[21] S. Srinivasan and K. L. Boyer, "Head pose estimation using view based eigenspaces," in *roceedings of the 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 4, pp. 302–304, Quebec City, Canada, August 2002.

[22] L. Chen, L. Zhang, Y. Hu, M. Li, and H. Zhang, "Head pose estimation using fisher manifold learning," in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Face and Gestures (AMFG '03)*, pp. 203–207, Nice, France, October 2003.

[23] Y. Zhu and K. Fujimura, "Head pose estimation for driver monitoring," in *Proceedings of IEEE Intelligent Vehicles Symposium (IVS '04)*, pp. 501–506, Parma, Italy, June 2004.

[24] N. Hu, W. Huang, and S. Ranganath, "Head pose estimation by non-linear embedding and mapping," in *Proceedings of the International Conference on Image Processing (ICIP '05)*, vol. 2, pp. 342–345, Genova, Italy, September 2005.

[25] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[26] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[27] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[28] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM Journal of Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.

[29] L. V. D. Maaten, E. O. Postma, and H. V. D. Herik, "Dimensionality reduction: a comparative review," Tech. Rep., University Maastricht, Amsterdam, The Netherlands, 2007.

[30] M.-C. Yeh, I.-H. Lee, G. Wu, Y. Wu, and E. Y. Chang, "Manifold learning, a promised land or work in progress?" in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '05)*, pp. 1154–1157, Amsterdam, The Netherlands, July 2005.

[31] X. Ge, J. Yang, T. Zhang, H. Wang, and C. Du, "Three-dimensional face pose estimation based on novel non-linear discriminant representation," *Optical Engineering*, vol. 45, no. 9, Article ID 090503, 3 pages, 2006.

[32] D. de Ridder, O. Kouropteva, O. Okun, M. Pietikäinen, and R. P. W. Duin, "Supervised locally linear embedding," in *Proceedings of the International Conference on Artificial Neural Networks and Neural Information Processing*, vol. 2714, pp. 333–341, Istanbul, Turkey, June 2003.

[33] N. Vlassis, Y. Motomura, and B. Kröse, "Supervised dimension reduction of intrinsically low-dimensional data," *Neural Computation*, vol. 14, no. 1, pp. 191–215, 2002.

[34] C.-G. Li and J. Guo, "Supervised isomap with explicit mapping," in *Proceedings of the 1st IEEE International Conference on Innovative Computing, Information and Control (ICICIC '06)*, Beijing, China, August 2006.

[35] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas, "Non-linear dimensionality reduction techniques for classification and visualization," in *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD '02)*, pp. 645–651, Edmonton, Alberta, Canada, July 2002.

[36] X. Geng, D.-C. Zhan, and Z.-H. Zhou, "Supervised nonlinear dimensionality reduction for visualization and classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 6, pp. 1098–1107, 2005.

[37] Q. Zhao, D. Zhang, and H. Lu, "Supervised LLE in ICA space for facial expression recognition," in *Proceedings of International Conference on Neural Networks and Brain (ICNNB '05)*, vol. 3, pp. 1970–1975, Beijing, China, October 2005.

[38] G. Little, S. Krishna, J. Black, and S. Panchanathan, "A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 2, pp. 89–92, Philadelphia, Pa, USA, March 2005.

[39] Y. Bengio, J. F. Paiement, P. Vincent, and O. Delalleau, "Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering," in *Proceedings of the 18th Annual Conference on Neural Information Processing Systems (NIPS '04)*, Vancouver, BC, Canada, December 2004.

[40] H. Choi and S. Choi, "Robust kernel isomap," *Pattern Recognition*, vol. 40, no. 3, pp. 853–862, 2007.

[41] M. Balasubramanian and E. L. Schwartz, "The isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, p. 7, 2002.

[42] D. F. Specht, "A generalized regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, 1991.