BIASED MANIFOLD EMBEDDING FOR PERSON-INDEPENDENT HEAD POSE ESTIMATION

Vineeth Nallure Balasubramanian and Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing (CUbiC), School of Computing and Informatics
Arizona State University, Tempe, AZ 85281, USA

Keywords: Manifold learning, Non-linear dimensionality reduction, Face modeling and analysis, Head pose estimation, Regression analysis.

Abstract: Head pose estimation is an integral component of face recognition systems and human computer interfaces. To determine the head pose, face images with varying pose angles can be considered to lie on a smooth low-dimensional manifold in high-dimensional feature space. In this paper, we propose a novel supervised approach to manifold-based non-linear dimensionality reduction for head pose estimation. The Biased Manifold Embedding method is pivoted on the ideology of using the pose angle information of the face images to compute a biased geodesic distance matrix, before determining the low-dimensional embedding. A Generalized Regression Neural Network (GRNN) is used to learn the non-linear mapping, and linear multi-variate regression is finally applied on the low-dimensional space to obtain the pose angle. We tested this approach on face images of 24 individuals with pose angles varying from -90° to +90° with a granularity of 2°. The results showed significant reduction in the error of pose angle estimation, and robustness to variations in feature spaces, dimensionality of embedding and other parameters.

1 INTRODUCTION

As human-centered computing applications grow each day, human face analysis has grown in its importance as a problem studied by several research communities. The estimation of head pose angle from face images is a significant sub-problem in this respect in several applications like 3D face modeling, gaze direction detection, driver monitoring safety systems, etc. Further, realistic solutions to the problem of face recognition have to be able to handle significant head pose variations, thereby leading to the gain in importance of the automatic estimation of the orientation of the head relative to the camera-centered co-ordinate system. While coarse head pose estimation has been successful to a large extent (Brown and Tian, 2002), accurate person-independent pose estimation, which is very crucial for applications like 3D face modeling, is still in the works.

Current literature (Fu and Huang, 2006) (Raytchev et al., 2004) (Wenzel and Schiffmann, 2005) separates the existing methods for head pose estimation into distinct categories:

- Shape-based geometric analysis, where head pose is discerned from geometric information like the configuration of facial landmarks.
- Model-based methods, where non-linear parametric models are derived before using a classifier like a neural network (Eg. Active Appearance Models (AAMs)).
- Appearance-based methods, where the pose estimation problem is viewed as a pattern classification problem on image feature spaces.
- Template matching approaches, which are largely based on nearest neighbor classification against texture templates/signatures.
- Dimensionality reduction based approaches, where linear/non-linear embedding of the face images is used for pose estimation.

To overcome data redundancy and obtain compact representations of face images, earlier work (Chen et al., 2003) (Raytchev et al., 2004) (Fu and Huang,

Nallure Balasubramanian V. and Panchanathan S. (2007).
BIASED MANIFOLD EMBEDDING FOR PERSON-INDEPENDENT HEAD POSE ESTIMATION.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 76-82

2006) suggests to consider the high-dimensional face image data as a set of geometrically related points lying on a smooth manifold in the high-dimensional feature space.

Different poses of the head, although captured in high-dimensional image feature spaces, can be visualized as data points lying on a low-dimensional manifold in the high-dimensional space. Raytchev et al (Raytchev et al., 2004) stated that the dimension of this manifold is equivalent 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 in image feature space. This conceptualization resulted in a host of dimensionality reduction techniques that are based on the relative geometry of the data points in high-dimensional space. This is the idea that underlies the family of non-linear dimensionality reduction techniques under the umbrella of manifold learning, like Isomap, Locally Linear Embedding (LLE), Laplacian Eigenmaps, Local Tangent Space Alignment (LTSA), etc, which have become popular in recent times.

In prior work in this domain, (Raytchev et al., 2004) and (Hu et al., 2005) employed a straightforward approach to learn the non-linear mapping onto the low-dimensional space through manifold learning, and estimated the pose angle using a pose parameter map. In the work carried out so far, the pose information of the given face images is ignored while computing the embedding. In this light, we propose a novel improvement to traditional manifold learning techniques, called the Biased Manifold Embedding approach, which provides a semantic bias to the manifold-based embedding process, using pose information from the given face image data. While the proposed Biased Manifold Embedding method is illustrated using Isomap in this paper, it can easily be extended to other manifold learning techniques with minor adaptations. As broader impact, the work proposed here is a significant contribution to a supervised approach to manifold-based non-linear dimensionality reduction techniques across all regression problems.

We discuss the background with a brief description of the Isomap algorithm, followed by related work and an insight into the significance of our work in Section 2. Section 3 details the mathematical formulation of the proposed Biased Manifold Embedding method. The experimental setup and the methodology of our experiments are briefed in Section 4. The results of the experiments are discussed in Section 5. We then discuss the advantages and limitations of the approach in the concluding section in Section 6, and provide future directions to this work.

2 BACKGROUND

2.1 Non-linear Dimensionality Reduction Using Isomap

Finding low-dimensional representations of high-dimensional data is a common problem in science and engineering. High-dimensional observations are prevalent in all fields: images, spectral data, instrument readings, etc. Techniques like Principal Component Analysis (PCA) are recognized as linear dimensionality reduction techniques, because of the linear projection matrix obtained from the eigen vectors of the covariance matrix. Techniques like Multi-Dimensional Scaling (MDS) are grouped under non-linear dimensionality reduction techniques. However, MDS uses the L2 (Euclidean) distance between data points in the high-dimensional space to capture their similarities. If the data points were to lie on a manifold in the high-dimensional space, Euclidean distances do not capture the geometric relationship between the data points. In such cases, it is beneficial to consider the geodesic (along the surface on which the data points lie) distances between the data points to obtain a more truthful representation of the data.

To capture the global geometry of the data points, Tanenbaum et al (Tanenbaum et al., 2000) proposed Isomap to compute an isometric low-dimensional embedding of a given set of high-dimensional data points (See Algorithm 1).

While Isomap captures the global geometry of the data points in the high-dimensional space, the disadvantage of this family of manifold learning 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 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. Figure 1 shows the visualization results of using Isomap to embed face images onto 2 dimensions. Faces of 10 individuals with 11 pose angles (-75° to +75° in increments of 15) were used to perform this embedding. The feature space considered here was the space

learning techniques to treat out-of-sample data points.

There has been recent work by (Ridder et al., 2003) and (Yu and Tian, 2006) to obtain a supervised approach to manifold learning techniques. However, their approaches are strictly oriented towards classification problems, and do not exploit the label information as possible for regression problems like head pose estimation.

2.3 Proposed Approach

While manifold learning techniques like Isomap capture the global geometrical relationship between data points in the high-dimensional image feature space, they do not use the pose label information of the training data samples. Unlike class labels in classification problems, pose information can be viewed as an ordered single-dimensional label with an established distance metric. This can provide valuable input to the embedding process.

In this work, we propose a biased manifold-based embedding for head pose estimation. We use the given pose information to bias the non-linear embedding to obtain accurate pose angle estimation. The significance of our contribution is realized in the fact that the proposed Biased Manifold Embedding method, although validated in this work with Isomap, can be extended to other manifold learning techniques with minor modifications, and in general, can be applied to all regression problems that use manifold learning methods. In addition, while most current approaches use face images sampled with pose angles at increments of 10-15° (Raytchev et al., 2004), we use the FacePix database (Little et al., 2005) that includes images of faces taken at a wide range of precisely measured pose angles with a readily available granularity of 1°. This reinforces the validity of our experiments with the proposed approach.

3 BIASED MANIFOLD EMBEDDING

In the Biased Manifold Embedding method, we propose to use the pose angle information of the training data samples to obtain a more meaningful embedding with a view to solve the problem of pose estimation. The fundamental idea of our approach is that face images with nearer pose angles must be nearer to each other in the low-dimensional embedding, and images with farther pose angles are placed farther, irrespective of the identity of the individual. We achieve this with a modification to the computation of the geodesic distance matrix. Since a distance metric can easily be defined on the pose angle values, the problem of finding closeness of pose angles is straight-forward.

The mathematical formulation of the Biased Manifold Embedding method is given below. We would like the ideal modified geodesic distance between a pair of data points to be of the form:

$$\tilde{D}(i,j) = f(P(i,j)) \otimes D(i,j)$$

where $D(i,j)$ (= d_M in Algorithm 1) is the geodesic distance between two data points x_i and x_j, $\tilde{D}(i,j)$ is the modified biased geodesic distance, $P(i,j)$ is the pose distance between x_i and x_j, f is any function of the pose distance, and $\otimes$ is a binary operator. If $\otimes$ was chosen as the multiplication operation, the function f would be chosen as inversely proportional to the pose distance, $P(i,j)$. In a more general perspective, the function f could be picked from the family of reciprocal functions ($f \in \mathcal{F}_R$) based on the needs of an application. In this work, we choose the function as:

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

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

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

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

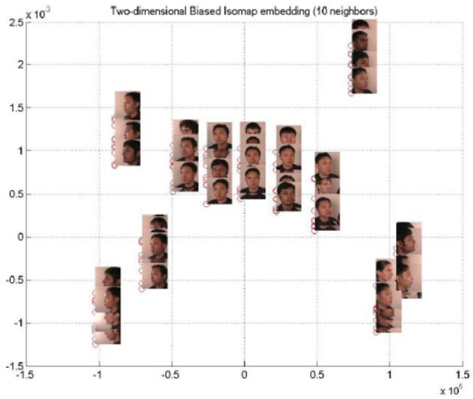
$$\alpha(P(i,j)) = \beta * |P(i,j)|$$

where β is a constant of proportionality, and allows parametric variation for performance tuning. In our work, we have used the pose distance as the one-dimensional distance i.e. $P(i,j) = |Pi - Pj|$, where P_k is the pose angle of x_k. In summary, the biased geodesic distance between a pair of points can be given by:

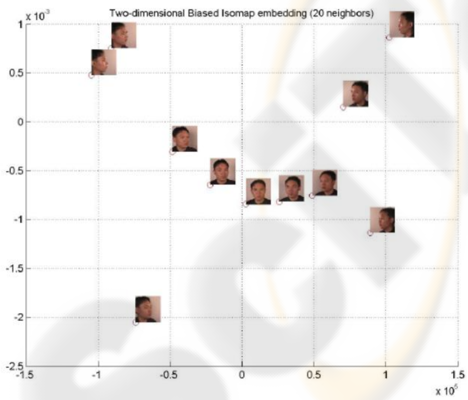
$$\tilde{D}(i,j) = \begin{cases} \frac{\alpha(P(i,j))}{max_{m,n}P(m,n)-P(i,j)} * D(i,j) & P(i,j) \neq 0, \\ 0 & P(i,j) = 0. \end{cases} \quad (1)$$

Classical MDS is applied on this biased geodesic distance matrix to obtain the embedding. The proposed modification impacts only the computation of the geodesic distance matrix, and hence, can easily be extended to other manifold-based dimensionality reduction techniques that use the geodesic distance.

Figure 2 shows the results of using Biased Isomap to embed the same facial images used in Figure 1 onto 2 dimensions. The embedded images establish the tendency of the method to elicit person-independent representations of the pose angles of the given facial images. As expected from the formulation of the method (see Figure 2), the face images of all individuals with the same pose angle have merged onto the same data point in 2 dimensions. This renders an embedding that is more conducive to determine the pose angle from the face images.



(a) Biased Isomap embedding with 10 neighbors



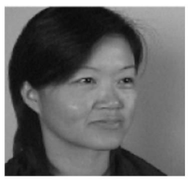
(b) Biased Isomap embedding with 20 neighbors

Figure 2: Biased Isomap Embedding of face images with varying poses onto 2 dimensions. Note in 2(b) that all the face images with the same pose angle have merged onto the same 2D point.

4 EXPERIMENTAL SETUP AND METHODOLOGY

The proposed Biased Isomap Embedding approach was compared against the traditional Isomap method for non-linear dimensionality reduction in the head pose angle estimation process. We used the FacePix face database (Little et al., 2005) (see Figure 3) built at the Center for Cognitive Ubiquitous Computing (CUbiC), which has face images with precisely measured pose variation. In this work, we consider a set of 2184 face images, consisting of 24 individuals with pose angles varying from -90° to +90° in increments of 2°. The images were subsampled to 32 x 32 resolution, and 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 4). The LoG transform was used since pose variation in face images is a result of geometric transformation, and texture information may not be really useful for the pose estimation problem. This was also reflected in preliminary experiments conducted with Gabor filters and Fourier-Mellin transformed images. The images were subsequently rasterized and normalized.

Figure 3: The data capture setup for FacePix.



(a) Grayscale image

(b) Laplacian of Gaussian (LoG) tranformed image

Figure 4: Image feature spaces used for the experiments.

Non-linear dimensionality reduction techniques like manifold learning do not provide a projection matrix to handle test data points. While different approaches have been used by earlier researchers to capture the mapping from the high-dimensional feature space to the low-dimensional embedding, we adopted a Generalized Regression Neural Network (GRNN) with Radial Basis Functions to learn the non-linear mapping. This approach has been adopted earlier by Zhao et al (Zhao et al., 2005). Additionally, the parameters involved in training the network (just the spread of the Radial Basis Function) are minimal, thereby facilitating better evaluation of the proposed method. Once the low-dimensional embedding was obtained, linear multi-variate regression was used to obtain the pose angle of the test image.

The proposed Biased Isomap Embedding method was compared with the traditional Isomap approach using resubstitution and 8-fold cross-validation models. In the resubstitution model, 100 data points were randomly chosen from the training sample for the testing phase. The error in estimation of the pose angle was used as the metric for performance evaluation. In the 8-fold cross-validation model, face images of 3 individuals were used for the testing phase in each fold, while all the remaining images were used in the training phase. In addition to these experiments, the variation in accuracy of the proposed method with the embedding dimension and the number of neighbors for the embedding was studied.

5 RESULTS AND DISCUSSION

The results for the resubstitution model are presented in Table 1. The improved performance of the Biased Isomap Embedding method for head pose estimation is unanimously reflected in the significant reduction in error values across the image feature spaces. However, validation using the resubstitution model is preliminary since test samples are picked from the training sample set itself. For more robust validation, we implemented 8-fold cross-validation over the images from 24 individuals. The results of these experiments are shown in Table 2. The results with the cross-validation model corroborate our claim of the performance gain. Both of these experiments were carried out with an embedding dimension of 8, with a choice of 50 neighbors for the embedding. The pose angle estimate error is consistently under 4°, which is a substantial improvement over earlier work (Raytchev et al., 2004).

In addition, the performance of the Biased Manifold Embedding was analyzed with varying dimensions of embedding, and choice of the number of neighbors used for embedding. Table 3 captures the results for different embedding dimensions with the number of neighbors fixed at 50. Table 4 captures the results for varying number of neighbors for the embedding with the embedding dimension fixed at 8. Grayscale pixel intensities of the face images were used for these independent experiments.

Table 1: Results using the resubstitution model.

Feature Space	Error using traditional Isomap	Error using Biased Isomap
Grayscale	11.39	1.98
Laplacian of Gaussian	8.80	2.31

Table 2: Results using the 8-fold cross-validation model.

Feature Space	Error using traditional Isomap	Error using Biased Isomap
Grayscale	10.55	3.68
Laplacian of Gaussian	9.10	3.38

Table 3: Analysis of performance with varying dimensions of embedding.

Dimension of Embedding	Error using traditional Isomap	Error using Biased Isomap
100	10.41	5.02
50	10.86	5.04
20	11.35	5.04
8	12.96	5.07
5	12.57	5.05
3	16.21	5.66

As evident from the results, the significant reduction in the error of estimation of pose angle substantiates the effectivness of the proposed approach. In addition, as the results in Tables 2, 3 and 4 illustrate, the Biased Manifold Embedding method is robust to variations in feature spaces, dimensions of embedding and choice of number of neighbors. While the traditional Isomap embedding has fluctuating results for these parameters, the range of error values obtained for the Biased Manifold Embedding method across these parameter changes suggests the high stability of the method, thanks to the biasing of the embedding.

Table 4: Analysis of performance with varying number of neighbors for embedding.

Number of Neighbors	Error using traditional Isomap	Error using Biased Isomap
30	11.56	5.10
50	12.96	5.06
100	13.83	5.03
200	12.59	5.06
500	14.36	5.07

6 CONCLUSION

We have proposed the Biased Manifold Embedding method, a novel supervised approach to manifold learning techniques for regression problems. The proposed method was validated for accurate person-independent head pose estimation. The use of pose information in the manifold embedding process improved the performance of the pose estimation process significantly. The pose angle estimates obtained using this method are accurate, and can be relied upon with an error margin of 3-4°. Our experiments also demonstrated that the method is robust to variations in feature spaces, dimensionality of embedding and the choice of the number of neighbors for the embedding. The proposed method can easily be extended from the current Isomap implementation to cover the envelop of other manifold learning techniques, and can be developed as a framework for biased manifold learning to cater to all regression problems at large.

6.1 Limitations and Future Work

As mentioned earlier, a significant drawback of manifold learning techniques is the lack of a projection matrix to treat new data points. While we used the GRNN to learn the non-linear mapping in this work, there have been other approaches adopted by various researchers. Bengio et al (Bengio et al., 2004) proposed a mathematical formulation focussed to overcome this problem. We plan to use these approaches to support the validity of our approach. Besides, we intend to extend the Biased Manifold Embedding implementation to LLE and Laplacian Eigenmaps to establish it as a framework for non-linear dimensionality reduction in regression applications. On a lesser significant note, another limitation of the current approach is that the number of neighbors chosen to obtain the embeddding has to be more than the number of individuals in the face images. This is because different individuals with the same pose angle are assigned a zero distance value in the biased geodesic distance matrix. We plan to modify our algorithm to overcome this limitation. In addition, the function of pose distance used to bias the geodesic distance matrix can be varied to study the applicability of different reciprocal functions for pose estimation.

REFERENCES

Bengio, Y., Paiement, J. F., Vincent, P., and Delalleau, O. (2004). Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering.

Brown, L. and Tian, Y.-L. (2002). Comparative study of coarse head pose estimation.

Chen, L., Zhang, L., Hu, Y. X., Li, M. J., and Zhang, H. J. (2003). Head pose estimation using fisher manifold learning. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG03)*, pages 203–207, Nice, France.

Fu, Y. and Huang, T. S. (2006). Graph embedded analysis for head pose estimation. In *7th International Conference on Automatic Face and Gesture Recognition*, Southampton, UK.

Hu, N., Huang, W., and Ranganath, S. (2005). Head pose estimation by non-linear embedding and mapping. In *IEEE International Conference on Image Processing*, page 342345, Genova.

Little, G., Krishna, S., Black, J., and Panchanathan, S. (2005). A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose and illumination angle. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, page 8992, Philadelphia, USA.

Raytchev, B., Yoda, I., and Sakaue, K. (2004). Head pose estimation by nonlinear manifold learning. In *17th International Conference on Pattern Recognition (ICPR04)*, Cambridge, UK.

Ridder, D. d., Kouropteva, O., Okun, O., Pietikainen, M., and Duin, R. P. (2003). Supervised locally linear embedding. In *International Conference on Artificial Neural Networks and Neural Information Processing*, 333341.

Tanenbaum, J. B., Silva, V. d., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):23192323.

Wenzel, M. T. and Schiffmann, W. H. (2005). Head pose estimation of partially occluded faces. In *Second Canadian Conference on Computer and Robot Vision (CRV05)*, page 353360, Victoria, Canada.

Yu, J. and Tian, Q. (2006). Learning image manifolds by semantic subspace projection. In *ACM Multimedia*, Santa Barbara, CA, USA.

Zhao, Q., Zhang, D., and Lu, H. (2005). Supervised lle in ica space for facial expression recognition. In *International Conference on Neural Networks and Brain (ICNNB05)*, volume 3, page 19701975, Beijing, China.