

# Neighborhood Dependent Approximation by Nonlinear Embedding for Face Recognition

Ann Theja Alex, Vijayan K. Asari, and Alex Mathew

Computer Vision and Wide Area Surveillance Laboratory,  
Department of Electrical and Computer Engineering,  
University of Dayton, Dayton, Ohio  
{alex1, vijayan.asari, mathewa3}@notes.udayton.edu

**Abstract.** Variations in pose, illumination and expression in faces make face recognition a difficult problem. Several researchers have shown that faces of the same individual, despite all these variations, lie on a complex manifold in a higher dimensional space. Several methods have been proposed to exploit this fact to build better recognition systems, but have not succeeded to a satisfactory extent. We propose a new method to model this higher dimensional manifold with available data, and use a reconstruction technique to approximate unavailable data points. The proposed method is tested on Sheffield (previously UMIST) database, Extended Yale Face database B and AT&T (previously ORL) database of faces. Our method outperforms other manifold based methods such as Nearest Manifold and other methods such as PCA, LDA Modular PCA, Generalized 2D PCA and super-resolution method for face recognition using nonlinear mappings on coherent features.

**Keywords:** Face Recognition, Manifold Learning, Nonlinear Embedding.

## 1 Introduction

Face recognition is a challenging problem in computer vision. Variations in face images make this a difficult task. Pose, illumination, expression, occlusion etc are different factors that influence recognition accuracy. Different approaches such as PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), ICA (Independent Component Analysis), Modular PCA and several manifold based approaches have evolved over time to deal with this issue. None of the methods offers a complete solution. Many a times, the recognition of frontal face achieves a better recognition rate, but as the pose varies, it becomes difficult to effectively identify faces.

The main objective of the current research is to improve the recognition accuracy in varying facial poses, illumination, expression variations and occlusions. Though PCA is a very popular method, it does not offer high accuracy with the variations in pose and illumination. Modular PCA approach is better than PCA with pose and illumination invariance [1]. However, Modular PCA is a

linear approach. This affects testing accuracy and can introduce false alarms. Many studies have proved that faces of a particular person tend to lie on a non-linear manifold in a higher dimensional space [2]. But methods like PCA and LDA use only Euclidean distance and hence fail to see the nonlinear structure [3]. An alternative is to use Locally Linear Embedding (LLE) [4], which allows data to be handled as lying on a linear neighborhood. This is a piecewise linear approach. The rationale behind this method is that every curve can be represented as a connection of linear segments. Generalizing this concept, a higher dimensional manifold can be represented as a combination of hyper-planes. The linear approach of LLE is not enough to yield good results, as the data points have non-linear relationships between them. Although many other methods such as Nearest Manifold, Locally Linear Embedded Eigenspace analysis, Discriminative manifolds, and 2D PCA evolved over time [5-8], recognizing faces under variations is still an open problem.

The proposed algorithm, Neighborhood Dependent Approximation by Non-linear Embedding (NDANE), aims at addressing this issue and offers a novel method for an effective representation of a manifold. The method is developed from the Hopfield network, LLE and non-linear attractor theories [9].

The Hopfield network is a recurrent network invented by John Hopfield in early 1980s. In Hopfield network, the output of one node is computed as the linear combination of the other inputs. Hopfield suggested that any input can be best approximated as the weighted combination of other related inputs. The Hopfield network supports only linear relationships.

In [4], Locally Linear Embedding is introduced as a dimensionality reduction technique. The approach provides a novel method for nonlinear dimensionality reduction. The method is considered one of the best for nonlinear dimensionality reduction, which is not effectively handled by component analysis methods. In LLE, the inputs are considered as data points in higher dimensional space. All the data points with the same characteristics form a geometric structure called the manifold. As all data points in the manifold belong to a particular class, any data point in the manifold can be redefined as a reconstruction of other data points in the same manifold. In the proposed method, the same observation is used for effective reconstruction of the data points.

In [9], a nonlinear attractor was proposed for learning the relationships in a manifold. This is a variation of the Hopfield network in which a higher dimensional relationship is defined between an input and its neighbors (represented by the other inputs). It is proposed that a higher dimensional polynomial defines the relationship of each image with its neighborhood. This, together with LLE resulted in the development of our method, in which a manifold is represented by a nonlinear approximation. As the manifolds are complex, a nonlinear approach can yield a better representation than a piecewise linear approach.

## 2 Theoretical Background

The proposed supervised learning algorithm is based on the fact that facial images are data points in a higher dimensional space, whose dimension is determined by the number of pixels in the image. The data points that correspond to the face of a particular person are observed to lie on a single nonlinear manifold. As the images in a manifold belong to the same person, there is a relationship between these images. A relationship can be modeled as proposed by Hopfield and as established by Seow and Asari [9]. In this paper, the same idea is extended so that it can be used to represent closely related points in terms of the neighborhood points. In other words, each image can be represented as the relationship of a few images in the manifold to which it belongs. Since manifolds can have a broad set of images of the same person, we do not consider all the images in the manifold to reconstruct an image. The nearest neighbors to each image are identified using Euclidean distance as the distance measure. Thus we construct small neighborhoods within the same manifold. Each image can be represented as the relationship of the neighborhood images. This relationship is defined as a nonlinear relationship, as it is observed that a nonlinear relationship offers a better representation of the manifold than any piece wise linear approach.

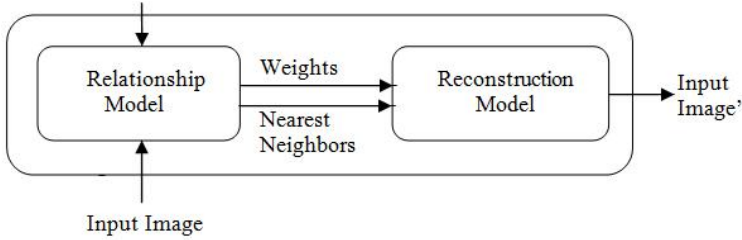
### 2.1 Architecture

A simple schematic diagram of the proposed algorithm is provided in this section (Fig. 1 and Fig. 2). We have a pipeline approach with two pipelines for handling training and testing phases.

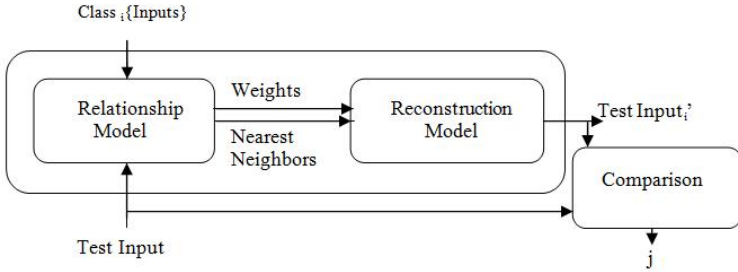
The algorithm is supervised and hence the training input will already be divided into classes. In the training phase, there are two separate stages. The first stage identifies the  $k$  nearest neighbors of an input image from the class to which the input image belongs. The nearest neighbors are the neighbors with the minimum Euclidean distances. The set of the nearest neighbors is also called the Proximity set. This step also models a higher dimensional polynomial relationship to represent the data point under consideration, using the nearest neighbors identified. The coefficients of the terms in the polynomial are called the weights. This stage is called the Relationship Modeling stage. The polynomial relationship that is used in this stage is described in more detail in section 2.2. The second stage is the Reconstruction stage, where the actual reconstruction of the data points is done using the neighbors and the weights. As the outcome of this stage, we obtain a reconstructed image. The process is repeated for all images in the training set. A set of reconstructed inputs are generated at the end of this stage.

In the Relationship Modeling stage of the testing phase, for each test image, we find the  $k$  nearest neighbors from each class. These neighbors are identified from the reconstructed training set inputs. Thus corresponding to each test input image, there are as many neighborhood sets as there are classes. In the Reconstruction stage, the test images are reconstructed based on the neighbors from all classes. The best reconstruction of the test image comes from the class

to which the test image should belong. This concept is used for classification. The testing procedure is described in section 2.3. The geometric foundation of the method is described in section 2.4.



**Fig. 1.** The schematic represents the stages in training. The inputs are reconstructed as the relationship of the nearest neighbors.



**Fig. 2.** The schematic represents the stages in testing. The test inputs are reconstructed as the relationship of the nearest neighbors from each class ‘i’, to get TestInput<sub>i</sub>. The best representation is given by the class to which the test image should ideally belong. The best representation is identified and is represented as ‘j’.

### 2.2 Training Algorithm

In the training process, the inputs are first grouped into different classes based on the labels, such that there is one class corresponding to each person. Each class is identified as a manifold and the aim of the training phase is to best approximate a manifold. If the images in a class have a high variance, instead of considering each class as a manifold, the class has to be partitioned into sub manifolds before processing. This partitioning can be repeated several times before the variances within a sub-manifold are sufficiently low. This can result in a tree like structure of repeated partitioning of the manifolds. However, in this paper, our focus is on the basic framework. Once the classes are identified, the  $k$  nearest neighbors of each training inputs in their respective classes are found. This forms a proximity set. Each training input is represented as the weighted sum of elements in the proximity set as given by equation (1).

$$X_i = \sum_{j=1}^k \sum_{m=1}^d W_{mji} X_j^m + L. \quad (1)$$

where  $k$  is the number of neighbors and  $d$ , the degree of the polynomial.

The degree  $d$  can be any positive integer, depending on the nonlinear relationship. If  $d=2$ , the resulting polynomial relationship is a second degree equation. We have used  $d=2$  in our experiments.  $L$  is a constant term. Since this representation is an approximation, we need to estimate the error in such a representation. The deviation of a representation from the actual data point can be characterized as a squared error given by equation (2).

$$E = (X_i - \sum_{j=1}^k \sum_{m=1}^d W_{mji} X_j^m + L)^2. \quad (2)$$

The weights corresponding to each term are found by minimizing the mean square error given by equation (2). The original training images are replaced with the form given in equation (1). Intuitively, this process incorporates the characteristics of its neighborhood into an image point.

### 2.3 Testing Algorithm

The first step in the testing algorithm is to find the  $k$  nearest neighbors in each class. This gives as many nearest neighbor sets as there are classes. The weights corresponding to each representation are found by minimizing  $E$  in equation (2). Each test image is represented as the weighted sum of elements in its proximity set. This gives as many representations as there are classes. Redefining each of the test image as the weighted sum of the neighborhood images gives as many representations as there are classes. The Euclidean distance,  $D_{class}$ , of each representation to the test image is computed. The class corresponding to the representation whose Euclidean distance is minimum, is taken as the class to which the test data belongs. This is given by equation (3).

$$j = \min[D_{class}]. \quad (3)$$

Here,  $j$  is the class which gives the minimum distance. The best reconstruction can only be provided by the set of neighborhood images in the class to which the image can belong. This is because the relationship in equation (1) can model only related images. If the test image belongs to a class, then the images in that particular class are only related to the test image and images from other classes have little or no relationship. So the reconstruction produced by the right class is a better approximation than the ones produced by other classes. Thus the reconstruction that results in minimum Euclidean distance is the one that is produced by the class to which the test image should belong.

## 2.4 Why NDANE Works?

Fig.3 illustrates two manifolds, M1 and M2, and a data point T. The geometric foundation of the method can be understood by analyzing such a topology. As there are only a few points on the manifold, the manifold is not fully defined. The distance  $d_1$  between the data point T and M1 is less than  $d_2$ , the distance between the data point and the closest data point on M2. In a Euclidean distance based algorithm, the distances  $d_3$  (distance between T and the closest data point on M1) and  $d_2$  are taken into consideration. This gives an incorrect classification of T as belonging to M2, although the actual distance of T to M1 ( $d_1$ ) is smaller than all other distances. Interpolating the manifold with available data points produces the point P on M1, whose distance to T is less than that to any point on M2. In our method, we complete the manifold by modeling it as a hyper-surface in image space. This completion process allows the calculation of the actual distance of the point to the closest manifold.

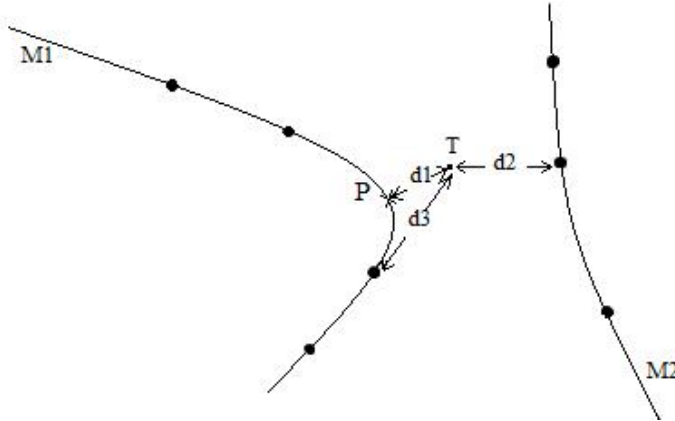


Fig. 3. Embedding a test data point on a manifold

## 3 Experiments

The proposed algorithm was tested using the pre-cropped images in the Extended Yale Face database B [10][11], the AT&T (previously ORL) database [12] and Sheffield (previously UMIST) database [13].

### 3.1 Experiments on Extended Yale Face Database B

The Extended Yale Face Database B has 16128 images of 28 human subjects under 9 poses and 64 illumination conditions [10][11]. Our experiments were conducted on a subset of the database. Images of 5 different people were used for testing and training. 10 images of each person were randomly chosen for

training. Another set of 10 images were chosen for testing. The training and testing images do not overlap. In this experiment  $d = 2$  and  $k=5$  in equation (1). This is a five class classification problem. We achieved a 100% recognition in this subset of the Extended Yale Face Database.

### 3.2 Experiments on AT&T (Previously ORL) Database of Faces

The AT&T database has a collection of 400 images of different people in various illumination conditions, pose and expression variations. The database is divided into 8 subsets. Each subset included 5 classes corresponding to the 5 individuals in the database. Out of the 10 available images per person, 5 were chosen for training and rest 5 for testing. In this experiment  $d=2$  and  $k=2$  in equation (1). The experimental results are given in Fig.4.

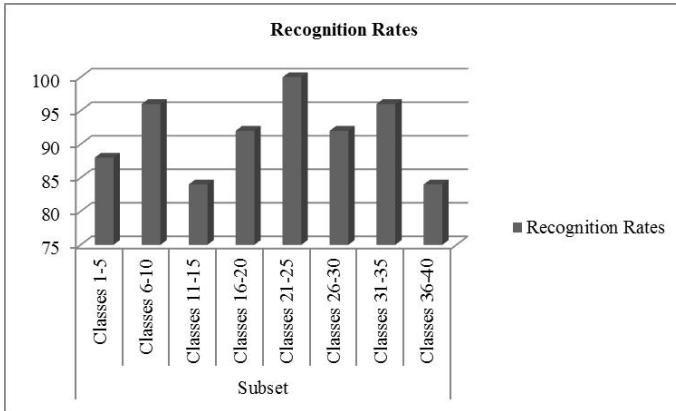


Fig. 4. Recognition rates on each subset of the AT&T database

### 3.3 Experiments on Sheffield (Previously UMIST) Database

The Sheffield (previously UMIST) database consists of 575 images of 20 individuals. There are images from different races, gender and poses. A sample data set is provided in Fig.5. It shows the reconstructed images which we obtained on reconstructing each image as relationship of its  $k$  nearest neighbors. In this experiment,  $d = 2$  and  $k=5$  in equation (1). A sample test set and the corresponding reconstructions are shown in Fig.6. It can be seen that the test image reconstructions are almost the same as the original images. In this database, number of images per person differs, so we have chosen half of the available number for training and the other half for testing. We have trained the system with 290 images and have tested the system with the remaining 285 images.

In all the three databases on which we conducted the experiments, our algorithm yielded a better result than other methods (Table 1 and Fig.7). We have achieved 95.44% recognition for the complete set of UMIST database. These results are better than the best results in [6], [7] and [14].



Fig. 5. Sample training set on left and the reconstructed images on right



Fig. 6. Sample test set on left and the reconstructed images on right

Table 1. Comparison of Recognition rates (Other methods Vs Proposed method)

Algorithm	%recognition on Sheffield database
PCA [7]	86.87
KPCA [7]	87.64
LDA [7]	90.87
2DPCA [7]	92.9
DCV [7]	91.51
B2DPCA [7]	93.38
K2DPCA [7]	94.77
MLA+NN [6]	94.29
Super Resolution Method [14]	93
<b>Proposed Method</b>	<b>95.44</b>

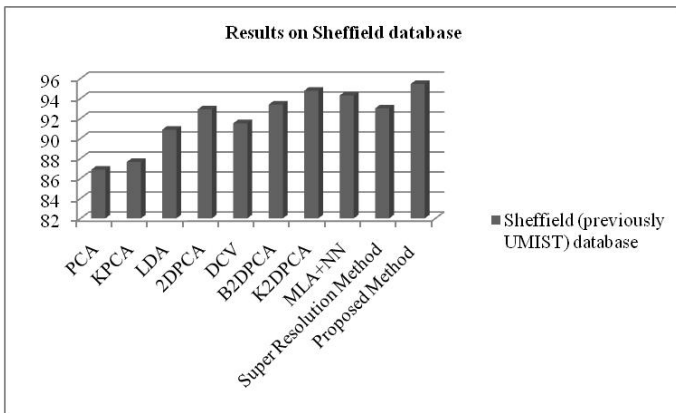


Fig. 7. Recognition rate comparison (Other methods Vs Proposed method)



## 4 Conclusion

We have proposed a robust mathematical model for representing complex manifolds formed by face images and a method to map test images to these manifolds. Our method outperforms conventional methods like PCA, Modular PCA, LDA and the methods in [6], [7] and [14]. All the experiments in this paper were conducted on the Sheffield (previously UMIST) database, AT&T (previously ORL) database and the Extended Yale Face Database B. The results obtained on these databases substantiate the mathematical foundation of our concept. Further experiments on larger databases such as FRGC database and FERET database are progressing. Research is also in progress to find a method to parameterize the value of  $k$  used in equation (1) depending on the data set variance and the data availability.

**Acknowledgments.** The authors acknowledge the creators of cropped Sheffield (previously UMIST) Face Database, Extended Yale Face Database B and AT&T ORL database for making these databases available for face recognition research.

## References

1. Gottumukkal, R., Asari, V.: An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters* 25(4), 429–436 (2004)
2. Park, S.W., Savvides, M.: An extension of multifactor analysis for face recognition based on submanifold learning. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2645–2652 (2010)
3. Wu, Y., Chan, K.L., Wang, L.: Face recognition based on discriminative manifold learning. In: *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004*, vol. 4, pp. 171–174 (2004)
4. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500), 2323–2326 (2000)
5. Yang, Q., Tang, X.: Recent Advances in Subspace Analysis for Face Recognition. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) *SINOBIOMETRICS 2004*. LNCS, vol. 3338, pp. 275–287. Springer, Heidelberg (2004)
6. Zhang, J., Stan, Z.L., Wang, J.: Nearest Manifold Approach for Face Recognition. In: *6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 223–228 (2004)
7. Kong, H., Wang, L., Teoh, E.K., Wang, J., Venkateswarlu, R.: Generalized 2D principal component analysis for face image representation and recognition. *Neural Networks* 18(5/6), 585–594 (2005)
8. Fu, Y., Huang, T.S.: Locally Linear Embedded Eigenspace Analysis. In: *IFP-TR, UIUC*, vol. 2005, pp. 2–5 (2005)
9. Seow, M., Asari, V.: Towards representation of a perceptual color manifold using associative memory for color constancy. *Neural Networks, The Official Journal of the International Neural Network Society, European Neural Network Society & Japanese Neural Network Society* 22, 91–99 (2009)
10. Lee, K.C., Ho, J., Kriegman, D.: Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence* 27(5), 684–698 (2005)

11. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 643–660 (2001)
12. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, vol. 5(7), pp. 138–142 (1994)
13. Graham, D., Allinson, N.M.: Characterizing Virtual Eigensignatures for General Purpose Face Recognition. In: *Face Recognition: From Theory to Applications*. NATO ASI Series F, Computer and Systems Sciences, vol. 163, pp. 446–456 (1998)
14. Huang, H., He, H.: Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features. *IEEE Transactions on Neural Networks* 22(1), 121–130 (2011)