

# Mutual Neighborhood Based Discriminant Projection for Face Recognition

Ben Niu<sup>1</sup>, Simon Chi-Keung Shiu<sup>1</sup>, and Sankar Pal<sup>2</sup>

<sup>1</sup>Department of Computing, Hong Kong Polytechnic University, Hong Kong, China  
{csckshiu, csniuben}@comp.polyu.edu.hk

<sup>2</sup>Indian Statistical Institute, Kolkata, India  
sankar@isical.ac.in

**Abstract.** Linear Discriminant Analysis is optimal under the assumption that the covariance matrices of the conditional densities are normal and all identical. However, this doesn't hold for many real world applications, such as Facial Image Recognition, in which data are typically under-sampled and non-Gaussian. To address this deficiency the Non-Parametric Discriminant method has been developed, but it requires model selection to be carried out for selecting the free control parameters, making it not easy for use in practice. We proposed a method, Mutual Neighborhood based Discriminant Projection, to overcome this problem. MNDP identifies the samples that contribute most to the Bayesian errors and highlights them for optimization. It is more convenient for use than NDA and avoids the singularity problem of LDA. On facial image datasets MNDP is shown to outperform Eigenfaces and Fisherfaces under various experimental conditions.

**Keywords:**  $k$ -nearest neighbors, mutual neighborhood, discriminant projection, face recognition.

## 1 Introduction

Linear Discriminant Analysis [1] is optimal for feature extraction assuming that the conditional densities are normal and all identical. However, for face recognition problems, datasets are often under-sampled and non-Gaussian. The Fisher-Raleigh criterion that defines LDA may lose its effectiveness for feature extraction and classification.

To address this problem, Fukunaga has developed the Nonparametric Discriminant Analysis method (NDA) [2] based on the idea of  $k$ -nearest neighborhood construction [3],[4]. Samples are assigned weights for training to increase the discriminant power of the extracted features. However, the weight function of NDA contains free parameters that have to be configured through model selections. They affect the weight function exponentially and range from 0 to the infinity. Users have to be careful in choosing them for maximal performance, making it not easy for use as LDA in practice.

We proposed an alternative approach, Mutual Neighborhood based Discriminant Projection (MNDP) to address this problem. The idea is that a majority of the Bayesian errors in classification arise from the samples that stay very close but belong to different

classes. To improve classification, we should find out and highlight them for discriminant feature extraction. This can be achieved by incorporating the idea of mutual  $k$  nearest neighborhoods. That is, if 2 samples in the data space are within the  $k$  nearest neighborhood of each other then we pair up and pick out them for optimization to reduce the recognition errors. Differing from NDA, MNDA doesn't require the extra costs for choosing the weight parameters in model selection. Also, we can learn more projection vectors in NDA than in LDA to improve the recognition rate. It doesn't have the singularity problem of LDA. We have done experiments to evaluate MNDA on the AR databases. Our results indicate that MNDA outperforms Eigenfaces and Fisherfaces significantly and consistently for face recognition.

## 2 Related Works

For 2 class problem, NDA is formulated by Eq. (1)

$$W_{opt} = \arg \max_W \frac{|W^T \tilde{S}_B W|}{|W^T S_W W|}, \tag{1}$$

where  $\tilde{S}_B = \frac{1}{N_1} \sum_{u=1}^{N_1} w_u (x_u^1 - m_2(x_u^1))(x_u^1 - m_2(x_u^1))^T + \frac{1}{N_2} \sum_{v=1}^{N_2} w_v (x_v^2 - m_1(x_v^2))(x_v^2 - m_1(x_v^2))^T$ .

$N_1$  and  $N_2$  are the number of samples in class 1 and class 2.  $x_u^1$  and  $x_v^2$  are 2 samples from class 1 and class 2.  $m_2(x_u^1)$  and  $m_1(x_v^2)$  are the mean vector of the  $k$  nearest neighbors of  $x_u^1$  in class 2 and  $x_v^2$  in class 1.  $w_u$  is the weight assigned to  $x_u^1$ ,

$$w_u = \frac{\min\{d^\alpha(x_u, x_{kNN}^1), d^\alpha(x_u, x_{kNN}^2)\}}{d^\alpha(x_u, x_{kNN}^1) + d^\alpha(x_u, x_{kNN}^2)}. \tag{2}$$

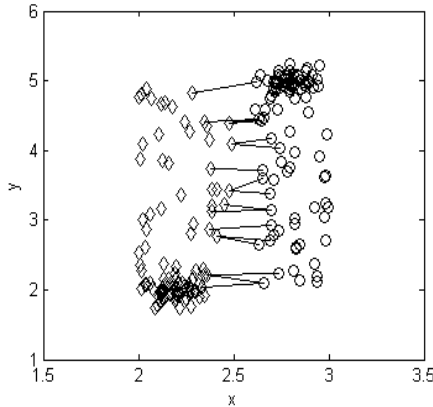
In Eq. 2,  $x_{kNN}^1$  and  $x_{kNN}^2$  are the  $k$ -th nearest neighbor of  $x_u$  in class 1 and class 2,  $d(\cdot, \cdot)$  is a distance measure,  $d(x_u, x_{kNN}^1)$  and  $d(x_u, x_{kNN}^2)$  are the radii of the  $kNN$  neighborhoods,  $\alpha$  is a free control parameter that ranges from zero to the infinity. The function has the property that near the classification boundary the weight takes large values and drops off to zero as going faraway. The parameters,  $\alpha$ , determine how fast  $w_u$  falls to zero. They range from 0 to the infinity and affect the weight function exponentially. Users have to be careful in choosing these parameters for performance, making it not easy for use as LDA in practice.

### 3 Concept of MNDP

Let  $X$  denote the space of observations,  $X \subseteq \mathbb{R}^n$ .  $X_1, X_2 \subset X, X_1 \cap X_2 = \emptyset$  are 2 classes of training samples. The between-class mutual  $k$ -nearest neighborhood is defined as the set of the sample pairs with mutual neighbor relations,

$$P = \{(x_1^i, x_2^j) \mid x_1^i \in X_1, x_2^j \in X_2, x_1^i \in \text{kNN}(x_2^j), x_2^j \in \text{kNN}(x_1^i)\}, \quad (3)$$

$x_1^i$  and  $x_2^j$  is the  $i$ -th and the  $j$ -th sample in classes  $X_1$  and  $X_2$ , respectively.  $\text{kNN}(x_1^i)$  is the set of the  $k$ -nearest neighbors of  $x_1^i$  in  $X_2$ .  $\text{kNN}(x_2^j)$  is the set of the  $k$ -nearest neighbors of  $x_2^j$  in  $X_1$ . By building the mutual  $k$ -nearest neighborhoods, we can readily identify the pairs of samples who are the mutual neighbors, as shown in Fig. 1.



**Fig. 1.** Mutual neighbors identified for 2 data classes (represented as circles and diamonds).

### 4 Formulation of MNDP

The objective function of MNDP is defined by,

$$W_{\text{opt}} = \arg \max_W \frac{|W^T S_N W|}{|W^T S_F W|} \quad (4)$$

$S_N$  is the between-class mutual kNN scatter matrix,  $S_N = \frac{1}{M} \sum_{i=1}^M (x_1^i - x_2^i)(x_1^i - x_2^i)^T$ ,

$M$  is the number of the sample pairs in the mutual kNN neighborhood,  $(x_1^i, x_2^i) \in P$ .  $S_F$

is the within-class kFN scatter matrix,  $S_F = \frac{1}{L} \sum_{j=1}^L (x_1^j - x_2^j)(x_1^j - x_2^j)^T$ ,  $L$  is the number

of the sample pairs in the kFN neighborhoods,  $(x_1^j, x_2^j) \in Q$ . The MNDP projections,  $W_{opt}$ , are obtained by solving the eigenvectors for the matrix  $S_F^{-1}S_N$ .

We demonstrated the effectiveness of MNDP over LDA using simulated data, as shown in Fig. 2. LDA and MNDP both work well for normal distributions (Left). But LDA fails to learn the optimal projection when the densities are not typically Gaussian (Right). However, MNDP can work well under both conditions.

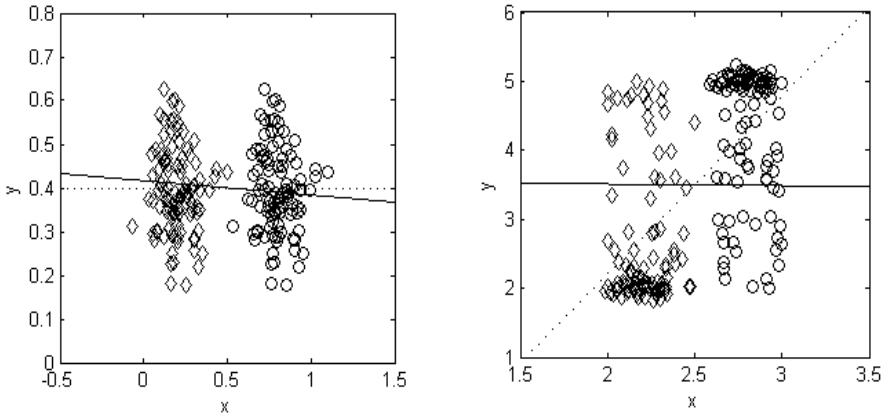


Fig. 2. First component of LDA (dot line) and MNDP (solid line)

## 5 Experimental Results

We evaluated MNDP for Face Recognition and compared it with Eigenfaces and Fisherfaces. The results on AR databases [5] indicate that MNDP achieves better recognition rate under various experiment conditions than others. We found that intriguingly it is most effective when the Cosine distance measure was utilized for classification on AR database. The accuracy rate was improved about 10 percent.

### 5.1 AR Database

We selected the sample images from 126 individuals, manually cropped and normalized them to  $50 \times 40$  pixels. The neighborhood size was set to 3. For each person 7 samples out of the 14, as shown in Fig. 3, are chosen randomly for training, the remainders for testing. The  $L_1$ ,  $L_2$  norms and the Cosine distances are utilized for  $l$ -NN classification. We repeated the experiment 20 times. The recognition rates, their standard deviations, and the number of components employed for classification are shown in Table 1.



Fig. 3. Sample images of one subject from AR database

Table 1. Top average recognition rate (%) on AR database

Method	Distance measures		
	L1	L2	Cosine
Eigenfaces	$92.08 \pm 1.23(80)$	$80.92 \pm 1.23(80)$	$85.48 \pm 1.23(80)$
Fisherfaces	$84.61 \pm 1.32(79)$	$82.74 \pm 1.21(80)$	$87.59 \pm 1.28(55)$
MNDP	$95.14 \pm 1.12(40)$	$90.81 \pm 0.81(41)$	$95.57 \pm 0.81(40)$

## 6 Conclusion

Linear Discriminant Analysis is optimal under the assumption that the covariances of the conditional densities are normal and all identical. This however doesn't hold for many real world applications, such as Face Recognition, in which data are typically under-sampled and non-Gaussian. Fukunaga has developed the Non-Parametric Discriminant method to address this deficiency but it was not quite easy for use as model selections have to be carried out to choose the optimal weight parameters. We proposed an alternative approach, Mutual Neighborhood based Discriminant Projection, to overcome this problem. MNDP highlights the mutually neighboring samples, which are considered to contribute most to the Bayesian errors for discriminant feature analysis. This was achieved by utilizing the idea of mutual  $k$  nearest neighborhood. MNDP is easier for use than NDA and avoids the singularity problem one of the limitations of LDA. On facial image datasets, MNDP achieved better recognition result than Eigenfaces and Fisherfaces under various experiment conditions.

## References

1. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic Press, Boston (1990)
2. Fukunaga, K., Mantock, J.: Nonparametric Discriminant Analysis. IEEE Trans. Pattern Anal. Machine Intell. 5, 671–678 (1983)

3. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood Preserving Embedding. In: Proceedings of the Tenth IEEE International Conference on Computer Vision, pp. 1208–1213 (2005)
4. Hu, D., Feng, G., Zhou, Z.: Rapid and brief communication: Two-dimensional Locality Preserving Projections (2DLPP) with its Application to Palmprint Recognition. *Pattern Recognition* 40(1), 339–342 (2007)
5. Martinez, A., Benavente, R.: The AR Face Database. CVC Technical Report, no. 24 (June 1998)