Nearest Intra-Class Space Classifier for Face Recognition

Wei Liu[1], Yunhong Wang[1], Stan Z. Li[2] and Tieniu Tan[1]
[1]National Lab of Pattern Recognition(NLPR)
Institute of Automation, Chinese Academy of Sciences, P.O.Box 2728, Beijing, P.R.China
{wliu,wangyh,tnt}@nlpr.ia.ac.cn
[2]Microsoft Research Asia, Beijing, P.R.China
szli@microsoft.com

Abstract

In this paper, we propose a novel classification method, called nearest intra-class space (NICS), for face recognition. In our method, the distribution of face patterns of each person is represented by the intra-class space to capture all intra-class variations. Then, a regular principal subspace is derived from each intra-class space using principal component analysis. The classification is based on the nearest weighted distance, combining distance-from-subspace and distance-in-subspace,between the query face and each intra-class subspace. Experimental results show that the NICS classifier outperforms other classifiers in terms of recognition performance.

1. Introduction

In general, appropriate facial representation and effective classification rules are two central issues in most face recognition systems. In this paper, we will mainly explore various nonparametric classification rules to design a robust classifier.

Up to now, a lot of pattern classification methods have been presented. One of the most popular classifiers among them is the nearest neighbor (NN) classifier [1]. Although NN is a very simple and convenient method, the representational capacity of face database is limited by the available prototypes in each class, which restrict the performance of NN.

To extend the capacity of covering more variations for a face class, Li *et al*. presented the nearest feature line (NFL) classifier in literature [2]. Following the work of NFL, Chien *et al*. [3] presented the nearest feature plane (NFP) for face classification. Both methods improve the performance of the NN method by expanding the representational capacity of available prototypes.

The nearest feature space (NFS) classifier [3] shows priority than NFL and NFP by Chien *et al*.'s conclusions. In contrast to NFL and NFP, NFS creates much more virtual prototype feature points, of which a substantial part is redundant and unreasonable, even may incur outliers.

In this paper, we incorporate the advantage of virtual samples with subspace analysis, which has been developed in recent decades starting from the work of E. Oja [4]. Firstly, we construct the intra-class space in which virtual samples are generated according to the learned principal variations. Subsequently, we propose a nearest intra-class space (NICS) classifier which is numerically stable and achieves best performance.

The rest of the paper is organized as follows. In Section 2, nearest feature classifiers (NFL, NFP, NFS) plus some problems associated with them are summarized. Next, we address the idea of intra-class space and derive the NICS classifier. Experimental results are reported in Section 4.

2. Nearest Feature Classifiers

Nearest feature classifiers, have been presented for robust face recognition in presence of varying viewpoints, illumination, expressions, etc. The common merit of these methods is that they all provide an infinite number of virtual prototype feature points of each class.

2.1. Nearest Feature Line and Plane (NFL and NFP)

Let x denote the query and $\{x_i^c | 1 \leq c \leq C, 1 \leq i \leq N_c\}$ represent all prototypes, $P_{i,j}^c$ is the projection point of the query x onto the feature line (FL) $\overline{x_i^c x_j^c}$. The FL distance between the query and FL is determined by $d(x, \overline{x_i^c x_j^c}) = \|x - P_{i,j}^c\|$, where $\|\cdot\|$ represents the Euclidean distance. The NFL distance [2] is the first rank distance which gives the best matched class c^*. The num-

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

ber of projection and distance calculations using NFL is $N_{NFL} = \sum_{c=1}^{C} N_c \cdot (N_c - 1)/2$.

Two potential problems in NFL can be summarized as follows: (a) The method becomes computational intractable when there are too many prototypes in each class; (b) NFL may fail when the prototypes are far away from the query point but the FL distance is very small. Problem (b) can be demonstrated from Fig.1, where we can see that p is much closer to x_1 and x_2 than to x_3 and x_4. However, the FL distance to $L(x_3, x_4)$ is smaller than the distance to $L(x_1, x_2)$, which may lead to wrong classification.

By extending the geometrical concept of line to plane, it is easy to construct NFP [3]. The plane $\widehat{x_i^c x_j^c x_k^c}$ passing through three random feature points (x_i^c, x_j^c and x_k^c) is called feature plane (FP) of x in the class c. The projection $P_{i,j,k}^c$ onto $\widehat{x_i^c x_j^c x_k^c}$ is determined, and $d(x, \widehat{x_i^c x_j^c x_k^c}) = \|x - P_{i,j,k}^c\|$ is the FP distance. The NFP searches the best matched class c^* according to the nearest FP distance. NFP also suffers from problem (a) and (b), takes greater computational cost than NFL. The number of projection and distance calculations using NFP increases to $N_{NFP} = \sum_{c=1}^{C} N_c \cdot (N_c - 1) \cdot (N_c - 2)/6$.

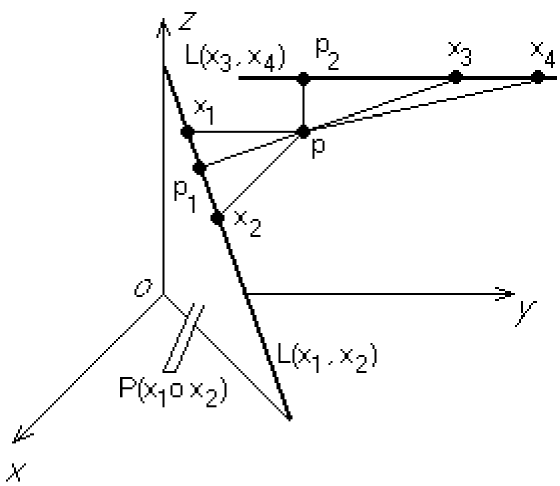

Figure 1. Drawbacks of NFL and NFS

2.2. Nearest Feature Space (NFS)

The NFS classifier [3] is presented to detect the most likely identity of query image by finding the nearest distance to feature spaces (called FS distance). All prototypes per class span a feature space, which is represented by the span

$$S^c = span\{x_1^c, x_2^c, ..., x_{N_c}^c\} \tag{1}$$

a matrix $Z_c = [x_1^c, x_2^c, ..., x_{N_c}^c]$ is built to determine the projection P^c of the query x onto the feature space. We obtain

$$P^c = Z_c(Z_c^T Z_c)^{-1} Z_c^T x \tag{2}$$

then classify the query by finding the nearest feature space among all classes

$$c^* = \arg \min_{1 \le c \le C} d(x, S^c) = \arg \min_{1 \le c \le C} \|x - P^c\| \tag{3}$$

No matter how many prototypes are collected, the number of projection and distance calculations always equals to $N_{NFS} = C$. Hence NFS is very efficient for face recognition.

From the geometrical viewpoint, the feature space is a glob space based on the origin. The FS distance is not a good measure. For example, the feature space spanned by two prototypes x_1 and x_2 in Fig.1 is the plane $P(x_1 O x_2)$. Besides virtual prototypes provided by $L(x_1, x_2)$, much more virtual prototypes on $P(x_1 O x_2)$ are created, which are redundant and unreasonable. Hence the FS distance maybe not suitable for similarity measure, and leads to wrong classification.

Besides above concern, we can find that calculating projection vectors P^c onto each feature space involves matrix inversion($Z_c^T Z_c$), which is close to singular when prototypes per class are numerically near enough. So NFS suffers form numerical instability.

3. Nearest Intra-Class Space (NICS)

The key issue of nearest feature classifiers is that how and where to generate virtual prototype feature points, we think virtual prototypes should be generated in a linear patch. Moreover, we construct the patch as the intra-class space that is demonstrated more reasonable and robust than the FL, FP and FS.

3.1. Intra-Class Space

To efficiently capture the characteristics of the difference between training samples, Moghaddam *et al.* [6] introduced the intra-personal space (IPS). The IPS is constructed by collecting all the difference images between any two image pairs pertaining to the same individual over all persons.

Motivated by IPS, we propose the intra-class space (ICS), which is constructed by collecting all the difference vectors (images) between any two prototype pairs in single class (single person). It is evident that ICS is just a subset of IPS, and it accounts for all intra-class variations in one class.

As addressed in Section 2.2, small differences between prototypes are against the robustness of feature space. Therefore, we produce the principal subspace from the ICS by applying PCA [5]. Our goal is to find a set of orthogonal basis vectors (provided by PCA) capturing the directions of maximum variance in prototype points, and to remove the directions associated with little variance such as eigenvectors corresponding to eigenvalues smaller than a threshold (e.g. 10^{-10}).

Let's denote the ICS by Δ_c. Its construction proceeds as follows: from the training set $\{x_i^c | 1 \le c \le C, 1 \le i \le N_c\}$, we can compute the difference vector $\delta_{i,j}^c = x_i^c - x_j^c$. Now, we have constructed $\Delta_c = \{\delta_{i,j}^c | 1 \le i, j \le N_c\}$. With the availability of the training sample for each ICS Δ_c, we can learn a principal subspace on it, denoted as U_c.

However, the construction of Δ_c will take much computational cost when there are too many prototypes in each class. We try to get subspace U_c without complex space construction. Wang *et al.* [7] proved that the eigenspace of PCA characterizes the difference between any two face images by showing that the covariance matrix for $\{\overrightarrow{x_i}\}$ equals that of $\{(\overrightarrow{x_i} - \overrightarrow{x_j})\}$ after removing the scale. Utilizing the theorem, we conclude that U_c is also the eigenspace of $\{x_i^c | 1 \le i \le N_c\}$ which is easier to compute.

Assume $U_c = [u_1^c, u_2^c, ..., u_{r_c}^c]$, where r_c denotes the intrinsic dimension of Δ_c and is usually $N_c - 1$. An arbitrary point lying in Δ_c is given by $\sum_j \gamma_j u_j^c$. Then the ICS generates virtual samples $x_\vee^c$ as below:

$$x_\vee^c - x_i^c \in \Delta_c \Longleftrightarrow x_\vee^c - x_i^c = \sum_{j=1}^{r_c} \gamma_j u_j^c = U_c \gamma \qquad (4)$$

where $i = 1, 2, ..., N_c$, and $\gamma \in R^{r_c}$ is the scale vector. By averaging Eq.(4) over i, we derive the unified formula

$$x_\vee^c - \overline{x^c} \in \Delta_c \Longleftrightarrow x_\vee^c - \overline{x^c} = U_c \gamma \qquad (5)$$

where $\overline{x^c}$ is the class mean vector. Eq.(5) addresses ICS from the geometrical perspective. We regard the ICS as a reference coordinate system with $\overline{x^c}$ as its origin, so the coordinates of virtual samples form the scale vector γ. Moreover, we have the following relation

$$\|x_\vee^c - \overline{x^c}\| = \|\gamma\| \qquad (6)$$

3.2. Distance-From-Subspace and Distance-In-Subspace

For any query point x, we define its distance-from-subspace by $d(x, \Delta_c)$ and distance-in-subspace by $d(x|\Delta_c)$. Firstly, we will find the best matched virtual sample for x in class c. Following the viewpoint of the IPS reference coordinate system, the virtual sample is just the projection point p^c of x in the reference coordinate system.

We define distance-from-subspace $d(x, \Delta_c)$ as the Euclidean distance between x and its projection p^c in the ICS Δ_c. Projecting the difference vector $x - \overline{x^c}$ into Δ_c, we get the difference vector $p^c - \overline{x^c}$ and its coordinates $U_c^T(x - \overline{x^c})$.

Due to Pythagorean theorem and Eq.(6), $d(x, \Delta_c)$ is given by

$$\begin{aligned} d(x, \Delta_c) &= \|x - p^c\| = \sqrt{\|x - \overline{x^c}\|^2 - \|p^c - \overline{x^c}\|^2} \\ &= \sqrt{\|x - \overline{x^c}\|^2 - \|U_c^T(x - \overline{x^c})\|^2} \end{aligned} \qquad (7)$$

Subsequently, we define distance-in-subspace $d(x|\Delta_c)$. To overcome the problem (b) described in Section 2.1, we must take the distances between the virtual samples and prototypes into account. We still consider the projected difference vector $p^c - \overline{x^c}$, which pertains to Δ_c and may be linearly decomposed to components in principal directions by Eq.(5). Because of different variance in each direction, we characterize $d(x|\Delta_c)$ as a Mahalanobis distance between p^c and $\overline{x^c}$

$$d(x|\Delta_c) = \sum_{j=1}^{r_c} \frac{y_j^2}{\lambda_j^c} \qquad (8)$$

where $[y_1, y_2, ..., y_{r_c}]^T = U_c^T(x - \overline{x^c})$ and eigenvalue λ_j^c corresponds to eigenvector u_j^c. In fact, the distance-in-subspace $d(x|\Delta_c)$ expresses rationality of the best matched virtual sample of x in Δ_c. The smaller $d(x|\Delta_c)$ is, the more likely the virtual sample p^c is in class c.

3.3. NICS Method

Now, we propose NICS classification rule which balances the two factors: distance-from-subspace and distance-in-subspace. By weighting $d(x, \Delta_c)$ and $d(x|\Delta_c)$, we obtain the ICS distance $d_{\Delta_c}(x)$ and find the nearest ICS distance to classify the query x

$$c^* = \arg\min_{1 \le c \le C} d_{\Delta_c}(x) = \arg\min_{1 \le c \le C} d(x, \Delta_c) + \alpha \cdot d(x|\Delta_c) \qquad (9)$$

where α (which we set to 10) is a regularization parameter that controls the trade-off between distance-from-subspace and distance-in-subspace.

4. Experimental Results

To demonstrate the efficiency of our method, extensive experiments are done on different face data sets. All methods are compared on the same training sets and testing sets, including NN, NFL, NFP, NFS and our method NICS. All experiments are implemented using the MATLAB V5.3 under Pentium IV PC with a clock speed of 1.69GHZ.

4.1. ORL Database

There are 10 different images for each subject in the *ORL* face database composed of 40 distinct subjects. All the subjects are in up-right, frontal position. The size of each face image is 92×112. The first line of Fig.2 shows 6 images of the same subject. The standard eigenface method [5] is first applied to the set of training images to reduce the dimension of facial image. In our experiments, we use 60 eigenfaces for each facial feature.

Fig.3 shows the average error rates (%) as functions of the number of training samples. In each round, $k (k = 2, 3, .., 9)$ images are randomly selected from the database for training and the remaining images of the same subject for testing. For each k, 20 tests are performed with different configuration of training and test set, and the results are averaged. The results plus recognition time at $k = 5$ are listed in Tab.1, which demonstrates that our method outperforms other classifiers.

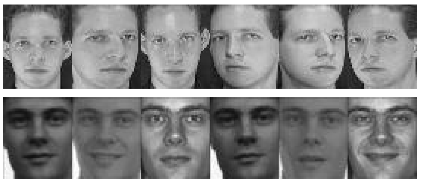

Figure 2. Samples from one subject in ORL and FERET

4.2. FERET Database

We tested our method on the more complex and challenging *FERET* database. We selected 70 subjects from the *FERET* database with 6 up-right, frontal-view images of each subject. The face images have much more variations in lighting, facial expressions and facial details. The second line of Fig.2 shows one subject from the selected data set.

The eye locations are fixed by geometric normalization. The size of face images is normalized to 92×112. Training and test process are similar to those on the *ORL* database. Similar comparisons between those methods are performed. This time 100 eigenfaces are used and k changes between 2 to 5, and the corresponding averaging error rates (%) are plotted in Fig.4. Tab.1 lists the average error rates and recognition time at the case $k = 4$. There are encouraging results, which show that the performance of our method is significantly better than other classifiers.

Table 1. Comparison of Classifiers

Classifier	Error Rate(%)		Recognition Time(ms)	
	ORL	FERET	ORL	FERET
NN	5.55	21.43	3.85	6.63
NFL	4.47	18.50	31.1	40.0
NFP	4.02	18.21	59.5	51.5
NFS	4.67	17.79	4.04	65.8
NICS	3.75	15.43	3.50	8.28

5. Acknowledgements

This work is sponsored by the Natural Sciences Foundation of China under grant No. 60121302 and 60335010. The authors thank Zhouyu Fu for his suggestions. Great appreciations especially to Yilin Dong for her encourage.

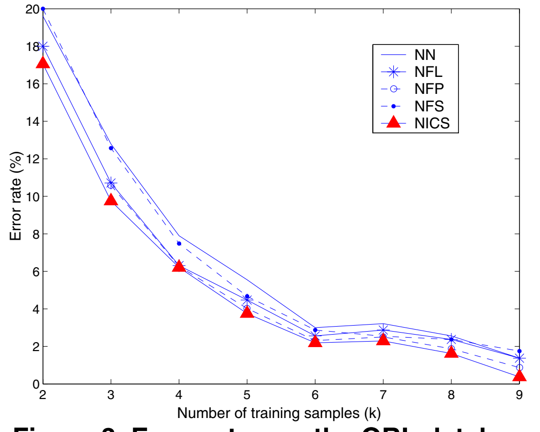

Figure 3. Error rates on the ORL database

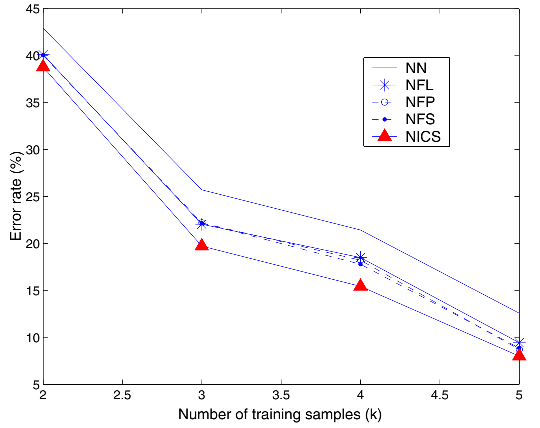

Figure 4. Error rates on the FERET database

References

[1] T.M. Cover and P.E. Hart. "Nearest Neighbor Pattern Classification", *IEEE Transactions on Information Theory*, 13(1): 21-27, 1967

[2] S.Z. Li and J.W. Lu. "Face Recognition Using the Nearest Feature Line Method", *IEEE Transactions on Neural Networks*, 10(2): 439-443, 1999

[3] J.T. Chien and C.C. Wu. "Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12): 1644-1649, 2002

[4] E. Oja. "Subspace methods of Pattern Recognition", *Research Sudies Press Ltd.*, 1983

[5] M. Turk and A. Pentland. "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, 3(1): 71-86, 1991

[6] B. Moghaddam, T. Jebara and A. Pentland. "Bayesian Modeling of Facial Similarity", *Advances in Neural Information Processing System*, 11: 910-916, 1998

[7] X.G. Wang and X.O. Tang. "Unified Subspace Analysis for Face Recognition", *Proc. of Int. Conf. on Computer Vision*, 1: 679-686, 2003