Nearest Neighbor Convex Hull Classification Method for Face Recognition

Xiaofei Zhou^{1,*} and Yong Shi^{1,2}

¹ Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences,
Beijing 100190, China

² College of Information Science and Technology University of Nebraska at Omaha,
Omaha, NE 68182, USA

zhouxf@gucas.ac.cn,
yshi@gucas.ac.cn, yshi@unomaha.edu

Abstract. In this paper, nearest neighbor convex hull (NNCH) classification approach is used for face recognition. In NNCH classifier, a convex hull of training samples of a class is taken as the distribution estimation of the class, and Euclidean distance from a test sample to the convex hull (the distance is called convex hull distance) is taken as the similarity measure for classification. Experiments on face data show that the nearest neighbor convex hull approach can lead to better results than those of 1-nearest neighbor (1-NN) classifier and SVM classifiers.

Keywords: classification, SVM, convex, nearest neighbor convex hull, face recognition.

1 Introduction

As one of most important biometrics technologies, face recognition has become an active research in pattern recognition and data mining area. Similar to many pattern recognition problems, solving classification problem is a key for face recognition. During the past 20 years many classification methods have been successfully applied in face recognition, such as 1-NN [1,2], Neural Networks [3], SVM [4], HMM [5] etc. As a non-parametric pattern recognition approach, the single nearest neighbor (1-NN) classifier is often used for classification. 1-NN is a most intuitive approach based on the nearest neighbor rule, which decides a query to the class including the nearest prototype to it. However, the performance of 1-NN is limited by the available prototypes in each class, which depends on how prototypes are chosen to account for possible sample variations and also how many prototypes are available. Practically no matter how representative the prototypes may be, there are always un-prototyped viewings, because only a finite, often small, number of prototypes are available as compared to all possibilities [6]. To adapt for more prototypes changes than the original prototypes, many classification approaches have been presented by using the linear combinations

---

* Corresponding Author.

of the prototypes to expand the representational capacity of them, such as NLC [6], NFL [7], NFP and NFS [8],>NNL and NNP [9] etc. Some classifiers also utilize convex hull of prototypes to represent training set. In [10], k-local convex distance is used as measure, and k-local convex hulls of each class are used to represent the class. Different from the above method, Jiang et al. [11] and Zhou et al. [12] present a new classification idea called nearest neighbor convex hull (NNCH) classification, which utilizes convex hull of all prototypes per class to represent each class. NNCH method [13] mentioned by Nalbantov et al. has similar ideas with NNCH. The approaches mentioned above are all based on the nearest neighbor rule and provide an infinite number of prototypical points to represent the query. In this paper, we apply the NNCH method [11,12] for face recognition. Like 1-NN, NNCH is a non-parameters method, which uses convex hull of prototypes to represent each class, and Euclidean distance between the query and the convex hull is used as the measure for nearest neighbor classification. The experiments on Yale face database and FERET face database show good performance.

The rest of the paper is organized as follows: Section 2 introduces foundations of NNCH, Section 3 describes NNCH method, Section 4 presents experimental results on face recognition.

2 Foundations of NNCH

Definition 1 [Convex Set]. Let $G \subseteq R^d$. Say that G is convex set if, for each x_1 and x_2 in G , the line segment between x_1 and x_2 is contained in G : $\forall x_1, x_2 \in G$ and $\lambda \in [0,1]$, $\lambda x_1 + (1-\lambda)x_2 \in G$.

That is, G is convex if, for all x_1 and x_2 in G and $\lambda \in [0,1]$, the point $\lambda x_1 + (1-\lambda)x_2$ lies in G .

Given a set S with n elements $\{x_1, x_2, \dots, x_n\}$, where $x_i = (x_{i1}, \dots, x_{id})^T$ is d dimensional vector in a feature space. We use $\text{co}(S)$ to denote the convex hull of S , $\text{co}(S)$ is the smallest convex set containing set S :

Definition 2 [Convex Hull]. The convex hull of a set $S \subset R^d$ is the smallest convex set containing set S : $\text{co}(S) = \{ \sum_{i=1}^n \alpha_i x_i \mid x_i \in S, \alpha_i \geq 0, \sum_{i=1}^n \alpha_i = 1, i=1,2,\dots,n \}$.

The convex hull of a set S is simply the set of all linear combinations of elements of S in which the coefficients of elements of S are nonnegative and sum to 1. Such constrained linear combinations are known as convex combinations.

In this paper the distance measure based on convex hull is used for classification, i.e., the distance from a query to a convex hull of training samples of a class is taken as the similarity measure. Virtually, the distance problem is a projection problem. Given a nonempty set $G \subset R^d$ and a vector y , the projection problem is the problem of determining the point $\hat{x} \in G$ that is the closest to y among all $x \in G$ (with respect to the Euclidean distance). Formally, the problem is given by

$$\min_{x \in G} \|y - x\|^2 \quad (1)$$

When the set G is a closed and convex set, the solution exists and it is unique, as seen in the following theorem.

Theorem 1 [Projection onto Closed Convex Set]. Let H is a Hilbert space, and $G \subset H$ be a nonempty closed convex set and $y \in H$.

- (Existence and uniqueness) There is a unique $\hat{x} \in G$ such that $\hat{x} = \arg \min_{x \in G} \|y - x\|$. $\hat{x}$ is called the projection of y onto G and is denoted by $p_G(y)$.
- (Characterization) A point $\hat{x} \in S$ is the projection $p_G(y)$ if $\langle y - \hat{x}, x - \hat{x} \rangle \leq 0$ for all $x \in S$.

The theorem guarantees the existence and uniqueness of the projection of a vector on a closed convex set. That is, for a nonempty closed convex set $G \subset H$ and arbitrary $y \in H$, $d(y, G) = \min_{x \in G} \|y - x\|$, where $d(y, G)$ can be computed, and there is a unique $\hat{x} \in G$ such that $\|y - \hat{x}\| = \min_{x \in G} \|y - x\|$. The projection $\hat{x}$ is the convex combination that is nearest to the query y . The distance between the query and nearest point in the convex hull is used as the similarity of the query and the class set. The unique solution y to the projection problem is referred to as the projection of $\hat{x}$ on $co(S)$.

3 Nearest Neighbor Convex Hull Classifier

Inspired by both nearest neighbor rule and the geometric interpretation of SVM, Jiang et al. [11] and Zhou et al. [12] presented a new classification method: nearest neighbor convex hull (NNCH) method. They use convex hull to extend the representation of prototypes and adopt nearest neighbor rule to realize classification.

The idea of NNCH is to expand representational capacity of prototypes of each class by convex combinations. This virtually provides an infinite number of prototypical points, and thus can account for more prototypical changes than the original prototypes. In the calculation of distance between a query vector and a class, the query is projected to the convex hull spanned by the prototypes of this class. The projection point is the convex combination that is nearest to the query. The distance between the query and convex hull is used as the basis for classification. Based on such distances, the conventional 1-NN classification, which compares each prototype individually, is extended to the nearest neighbor convex hull classification, which compares the convex hulls of each class: the query is classified to the class of the nearest convex hull.

The idea using convex hull of all samples in a class to represent the class is also derived from the geometric interpretation of SVM. SVM is a robust methodology [14, 15] for classification. Intuitively, given a set of points belonging to two classes, SVM finds the optimal hyperplane that separates binary class data without errors, while maximizing the distance from either class to the hyperplane. In geometry, this is equivalent to separating the convex hulls of all samples in each class with maximal margin [16, 17]. So the

convex hull of all samples in a class can represent the class space of samples well. In NNCH, such convex hulls are considered to extend the representation of prototypes. But unlike maximal margin rule of SVM, NNCH uses the nearest neighbor rule to classify. In fact, NNCH can be considered as an approach separating the convex hull by nearest neighbor rule. In reference [10], k -local convex distances are used for classification, which emphasis the local linear construction of sample distribution and use local convex hulls to represent the class. But in NNCH, general convex hull of a class is presented to represent sample distribution, which is inspired by the convex hull in SVM.

3.1 Convex Hull Distance

In NNCH classifier, a convex hull of training samples of a class is taken as distribution estimation of the class, and Euclidean distance from a test sample to the convex hull (the distance is called convex hull distance) is taken as the similarity measure for classification.

Given a set $S \subset \mathbf{R}^d$, $S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$, the distance function between a query $\mathbf{x}$ and the convex hull of S can be written in detail:

$$\begin{aligned}
 d^2(\mathbf{x}, co(S)) &= \min_{\boldsymbol{\eta} \in co(S)} \|\mathbf{x} - \boldsymbol{\eta}\|^2 \\
 &= \min_{\boldsymbol{\alpha}} \left\| \mathbf{x} - \sum_{i=1}^k \alpha_i \mathbf{x}_i \right\|^2 \\
 &= \min_{\boldsymbol{\alpha}} [(\mathbf{x} \cdot \mathbf{x}) - 2 \sum_{i=1}^k \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j)] \\
 \text{s.t. } \sum_{i=1}^k \alpha_i &= 1, \alpha_i \geq 0, i=1, 2, \dots, k
 \end{aligned} \tag{2}$$

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$, $\mathbf{e} = (1, 1, \dots, 1)_{1 \times k}^T$, $\mathbf{0} = (0, 0, \dots, 0)_{1 \times k}^T$, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\alpha}$, $\mathbf{e}^T \boldsymbol{\alpha} = 1, \boldsymbol{\alpha} \geq \mathbf{0}$. Equation (2) can be written in matrix:

$$\begin{aligned}
 d^2(\mathbf{x}, co(S)) &= \min_{\mathbf{y} \in co(S)} \|\mathbf{x} - \boldsymbol{\eta}\|^2 \\
 &= \min_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{X}\boldsymbol{\alpha}\|^2 \\
 &= \min_{\boldsymbol{\alpha}} (\mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\alpha}) \\
 \text{s.t. } \mathbf{e}^T \boldsymbol{\alpha} &= 1, \boldsymbol{\alpha} \geq \mathbf{0}
 \end{aligned} \tag{3}$$

As $\mathbf{x}^T \mathbf{x}$ is constant, which can be discard from the optimal problem. The optimal problem we need to solve is

$$\begin{aligned}
 \min_{\boldsymbol{\alpha}} \quad & 2\mathbf{x}^T \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\alpha} \\
 \text{s.t. } \quad & \mathbf{e}^T \boldsymbol{\alpha} = 1, \boldsymbol{\alpha} \geq \mathbf{0}
 \end{aligned} \tag{4}$$

Equation (4) is a convex quadratic optimization. In this paper, we use the MATLAB optimal tools to realize the solution.

Supposed $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_k^*)^T$ is the optimum solution for (4), projection $\mathbf{y}$ is the convex combination of elements in S , with the coefficient $\alpha_1^*, \alpha_2^*, \dots, \alpha_k^*$:

$$\mathbf{y} = \sum_{i=1}^k \alpha_i^* \mathbf{x}_i \quad (5)$$

Then we can compute $d^2(\mathbf{x}, co(\mathbf{S}))$ by $\mathbf{y}$:

$$d^2(\mathbf{x}, co(\mathbf{S})) = \|\mathbf{x} - \mathbf{y}\|^2 = \left\| \mathbf{x} - \sum_{i=1}^k \alpha_i^* \mathbf{x}_i \right\|^2. \quad (6)$$

3.2 Nearest Neighbor Convex Hull Algorithm

We assume to have a multi-labeled training set $\mathbf{S} = \{(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_n, c_n)\}$, where $\mathbf{x}_i \in \mathbf{R}^d$ is a sample and $c_i \in \{1, \dots, l\}$ is its corresponding class or label, i.e. there are l categories training sets: $\mathbf{S}_1 = \{\mathbf{x}_i | c_i = 1\}$, $\mathbf{S}_2 = \{\mathbf{x}_i | c_i = 2\}$, ..., $\mathbf{S}_l = \{\mathbf{x}_i | c_i = l\}$. The l convex hulls from different category training sets are:

$$\begin{aligned} co(\mathbf{S}_1) &= \left\{ \sum_{i, c_i=1} \alpha_i \mathbf{x}_i \mid \sum_{i, c_i=1} \alpha_i = 1, \alpha_i \geq 0 \right\}, \\ &\dots, \\ co(\mathbf{S}_l) &= \left\{ \sum_{i, c_i=l} \alpha_i \mathbf{x}_i \mid \sum_{i, c_i=l} \alpha_i = 1, \alpha_i \geq 0 \right\}. \end{aligned}$$

For an arbitrary query $(\mathbf{x}, c)$, $\mathbf{x} \in \mathbf{R}^d$, c is class label which is unknown, we need to give the c value. We respectively compute the square distance between $\mathbf{x}$ and each class convex hull:

$$\begin{aligned} d^2(\mathbf{x}, co(\mathbf{S}_1)) &= \min_{\boldsymbol{\eta} \in co(\mathbf{S}_1)} \|\mathbf{x} - \boldsymbol{\eta}_1\|^2, \\ &\dots, \\ d^2(\mathbf{x}, co(\mathbf{S}_l)) &= \min_{\boldsymbol{\eta} \in co(\mathbf{S}_l)} \|\mathbf{x} - \boldsymbol{\eta}_l\|^2. \end{aligned}$$

We take $d^2(\mathbf{x}, co(\mathbf{S}_j))$ as the similarity of $\mathbf{x}$ and the j th class, and classify $\mathbf{x}$ to the class of the nearest neighbor convex hull:

$$c = \underset{j}{\operatorname{argmin}} d^2(\mathbf{x}, co(\mathbf{S}_j)), j = 1, 2, \dots, l. \quad (7)$$

Compared with 1-NN, NNCH presents infinite samples convex combined by training samples, whereas 1-NN has the limited training set. Compared with SVM, NNCH can be directly used for multi-class problems. For SVM, it usually needs to decompose multi-class problems into many two-class problems.

4 Experiments

We apply NNCH classifier on face recognition. The comparison experiments of NNCH with 1-NN and SVM are conducted on two face databases, ORL and FERET face

databases. For SVM, three kernels, linear kernel $k = (\mathbf{x} \cdot \mathbf{y})$, quadratic kernel $k = (1 + (\mathbf{x} \cdot \mathbf{y}))^2$ and gaussian kernel $k = \exp (-0.5(\|\mathbf{x} - \mathbf{y}\|/\sigma)^2)$ are chosen, and a bottom-up binary tree-structured approach [4] is used to extend SVM to deal with multi-class problems. All of our experiments are carried out under Matlab 7.0 platform.

The first experiment is performed on ORL face database (<http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>). ORL database contains 40 distinct persons. Each person has ten different images labeled with the number 1,2, ... ,10. The original face images were all sized 92×112 pixels with 256-level gray scale. Fig.1 shows ten face images of one subject. In our experiments, all images are transformed to JPEG format, and downsampled to 16×16 by bicubic interpolation. We use the first two images of each individual and their mirrors as training set. The remaining 320 images are as test set.

Fig. 1. Ten face images of one person in ORL face database

Table 1 gives the comparisons of different methods on face recognition accuracy. The penalty parameter of SVMs adopt $C = \infty$, and for gaussian kernel SVM, $\sigma = 3$.

Table 1. Results of experiments on ORL face database

Classification methods	Recognition accuracy
1-NN	86.56%
SVM(Linear kernel)	85.63%
SVM(quad kernel)	83.00%
SVM(gaussian kernel)	86.88 %
NNCH	87.81%

The second experiment is performed on a subset of FERET face database (http://www.itl.nist.gov/iad/humanid/feret/feret_master.html). The subset contains 200 distinct persons. Each person has seven different images labeled with “ba, bd, be, bf, bg, bj, bk”. The original face gray images are cropped to 80×80 pixels. Fig.2 shows the cropped face images from one subject.

Fig. 2. Seven face images for one person in FERET face database

In our experiments, the images are also downsampled to 40×40 by bicubic interpolation, and the illumination normalization technique is done for compensating illumination variations of all faces. The first four images (labeled with “ba, bd, be, bf”) of each person and their mirrors are used for training, and the rest 600 images are used for testing.

Table 2 gives the comparisons of different methods on face recognition accuracy. The penalty parameter for SVMs adopt $C=\infty$, and for gaussian kernel SVM, $\sigma=50$.

Table 2. Results of experiments on FERET face database

Classification methods	Recognition accuracy
1-NN	75.00%
SVM(linear kernel)	81.50%
SVM(quadratic kernel)	83.00%
SVM(gaussian kernel)	83.67%
NNCH	86.50%

From the results of above experiments, it is clear that as a whole the face classification performance of the NNCH is competitive with that of 1-NN and SVMs. In Table 1 and Table 2, NNCH on face recognition all can obtain the highest recognition accuracy among the methods. In Table 1, the recognition rate of NNCH (87.81%) is higher than 1-NN (86.56%) and three kernel SVMs (85.63%, 83.00% and 86.88 %); In Table2, the recognition rate of NNCH is 86.50%, which is much higher 11.5% than 1-NN, also higher 5% than linear SVM, and higher almost 3% than nonlinear SVMs.

5 Conclusions

This paper introduces a novel pattern classification method called nearest neighbor convex hull (NNCH) approach for face recognition. In NNCH, a convex combination of vectors belonging to a class is used to define a measure of distance from a query vector to the class. The measure is defined as the Euclidean distance from the query to the nearest convex combination. Experiments on face data show that NNCH leads to better results than those of 1-nearest neighbor (1-NN) classifier and SVM classifiers.

Acknowledgments. This work was partially supported by the National Grand Fundamental Research 973 Program of China under Grant No.2004CB720103, by the National Nature Science Foundation of China under Grant No.70531040, No.70621001, No.10601064 and No.70501030 and by a research grant from BHP Billion Co., Australia.

References

1. Jin, Z., Yang, J.Y., Hu, Z.S., Lou, Z.: Face Recognition based on Uncorrelated Discriminant Transformation. *Pattern Recognition* 34(7), 1405–1416 (2001)
2. Yang, J.Y., Zhang, David, Yang, J.Y., et al.: Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE PAMI* 26, 1131–1137 (2004)
3. Ranganath, S., Arun, K.: Face Recognition Using Transform Features and Neural Networks. *Pattern Recognition* 30(10), 1615–1622 (1997)
4. Guo, G., Li, S.Z., Chan, K.L.: Face Recognition by Support Vector Machines. In: *Proc. fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 196–201 (2000)
5. Samaria, F., Young, S.: HMM based Architecture for Face Recognition. *Image and Computer Vision* 12, 537–543 (1994)
6. Li, S.Z.: Face Recognition based on Nearest Linear Combinations. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 839–844 (1998)
7. Li, S.Z.: Performance Evaluation of the Nearest Feature Line Method in Image Classification and Retrieval. *IEEE Trans. On PAMI* 22(11), 1335–1339 (2000)
8. Chen, J.T., Wu, C.C.: Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition. *IEEE Trans. On PAMI* 24(12), 1644–1649 (2002)
9. Zheng, W.M., Zou, C., Zhao, L.: Face Recognition Using Two Novel Nearest Neighbor Classifiers. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. v-725–728 (2004)
10. Vincent, P., Bengio, Y.: K-local Hyperplane and Convex Distance Nearest Neighbor Algorithms. In: Dietterich, T.G., Becker, G.Z. (eds.) *Advances in Neural Information Processing Systems* 14, pp. 985–992. MIT Press, Cambridge (2002)
11. Jiang, W.H., Zhou, X.F., Yang, J.Y.: P-norm Nearest Neighbor Convex Hull Classification Algorithms (in Chinese). *Journal of Harbin Institute of Technology* 38, 982–984 (2006); In: *Proceedings of the 7th Chinese Symposium on Intelligent Robot*(2006)
12. Zhou, X.F., Jiang, W.H., Yang, J.Y.: Kernel Nearest Neighbor Convex Hull Classification Algorithm (in Chinese). *Journal of Image and Graphics*, 1209–1213 (2007)
13. Nalbantov, G.I., Groenen, P.J.F., Bioch, J.C.: Nearest Convex Hull Classification, Report (2007), <http://repub.eur.nl/publications/index/728919268/NCH10.pdf>
14. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
15. Deng, N.Y., Tian, Y.J.: *A New Approach in Data Mine-Support Vector Machine*. Science Press, Beijing (2004)
16. Bennett, K.P., Bredensteiner, E.J.: Duality and Geometry in SVM Classifiers. In: Langley, P. (ed.) *Proceedings of the 17th International Conference on Machine Learning*, pp. 57–64. Morgan Kaufmann, San Francisco (2000)
17. Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K.: A Fast Iterative Nearest Point Algorithm for Support Vector Machine Classifier Design. *IEEE Transactions on Neural Networks* 11(1), 124–136 (2000)