Discriminant Feature Extraction Based on Center Distance

Hui Yan, Wankou Yang, Jian Yang, Jingyu Yang
Nanjing University of Science & Technique, China, Nanjing, 210094
betty@njust.edu.cn

Abstract

In this paper, a novel discriminant feature extraction algorithm employing center-based distance is proposed for face recognition. This new method, which is a supervised linear dimensionality reduction and feature extraction approach, computes the center-based distance between each training sample-pairs in the same class and the distance between each training sample-pair belonging to different classes. Then the high-dimensional data are embedded into a low-dimensional space, preserving the within-class geometric structure on a submanifold via maximum variance projection. Many experiments on ORL and Yale face database indicate that this method is highly effective.

1. Introduction

Face recognition has received extensive attention in recent years. Over the past 20 years, numerous algorithms have been proposed, among which the most well known ones are Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2, 3]. However, both assume an approximate Gaussian distribution of each class in the data space, and both may fail to work when the data are distributed in a nonlinear way. Recent studies reveal that face images, parameterized by some continuous variables, usually belong to a submanifold of intrinsically low dimension [4, 5-11]. In this respect, a number of nonlinear embedding methods have emerged. The representative ones include Isomap [6], local linear embedding (LLE) [5, 7], Laplacian eigenmaps (LE) [8], local tangent space alignment (LTSA) [12], etc. These algorithms differ in the representations of local geometries that they attempt to preserve. While these manifold learning methods generate satisfying results in dimensionality reduction and manifold visualization, these approaches does not define how to project new data into the embedded space, so they are not suitable for recognition problems in practical application. As a result, linear approximation approaches to nonlinear embedding [13] have been sought, such as LPP [14] and LEA [15], which are sufficiently effective to deal with practical problems in machine learning and computer vision.

Despite the differences of the aforementioned approaches, they can all be unified within the graph-embedding (GE) framework and its extensions [16]. Furthermore, GE framework offers a general platform for developing new dimensionality reduction algorithms.

To fulfill the task of classification and preserve within-class property, we propose a novel algorithm for linear dimensionality reduction and feature extraction within the framework of GE. The new algorithm exhibits several attractive properties:

(a) *The preservation of within-class structure*: The new algorithm preserves center-based distances, which can capture more extract information than Euclid distance between any two data points with the same class.

(b) *The discriminating power in classification*: On the assumption that the data distributed on different submanifolds belong to different classes, we want to map the inputs from different manifolds as far as possible, i.e. maximize the pairwise distances belonging to different classes in the low dimensional subspace. Besides, within-class intrinsic geometric structures in the input space may produce additional discriminating power for classification.

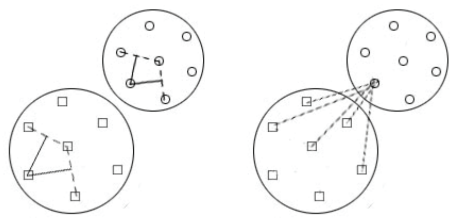
(c) *The employment of the graph embedding (GE) framework*: According to the GE framework, we define an intrinsic graph that describes within-class geometry, and a penalty graph that describes the separability of between-class data, as illustrated in Fig.1.

In Fig.1, the new algorithm defines a corresponding intrinsic graph that describes the center-based distance between each training sample-pairs in the same class, with constraints from the penalty graph that characterizes a statistical or geometric property of the distance between each training sample-pair belonging to different classes.

(d) *The applicability in recognition problems*: The new algorithm is a linear embedding algorithm. Given a new high-dimensional data point, it automatically finds the corresponding low-dimensional point on manifold via the linear projection.

The rest of this study is organized as follows. In

Sec.2, we review the center-based nearest neighbor classifier. The proposed algorithm is described in detail in Sec.3. In Sec.4, many experiments with face image data are carried out to evaluate our new algorithm. Finally, the conclusions are summarized in Sec.5.



(a)Intrinsic Graph (b) Penalty Graph

Fig.1 Intrinsic Graph (a) and Penalty Graph (b)

2. Center-based nearest neighbor classifier

Center-based nearest neighbor classifier [17] considers a kind of line for classification. Let x_i^c be a training sample of class c, let o^c be the center of class c, $o^c = \frac{1}{N_c}\sum_{i=1}^{N_c} x_i^c$ .For an unknown sample x, define center-based line(CL) $\overline{x_i^c o^c}$, which is the straight line passing through x_i^c and o^c , as illustrated in Fig.2. This CL is used to capture the information to achieve better classification performance. Define distance from x to CL as

$$d(x, \overline{x_i^c o^c}) = \| x - p^{c,i} \| \tag{1}$$

where $p^{c,i}$ is the projection of x onto the CL, $\| \bullet \|$ is the Euclidean norm. The projection point $p^{c,i}$ is calculated according to the equation

$$p^{c,i} = x_i^c + \mu(o^c - x_i^c) \tag{2}$$

where $\mu \in \mathrm{R}$, which is called position parameter and formalized as

$$\mu = \frac{(x - x_i^c)^{\mathrm{T}}(o^c - x_i^c)}{(o^c - x_i^c)^{\mathrm{T}}(o^c - x_i^c)} \tag{3}$$

where T is the transpose operator.

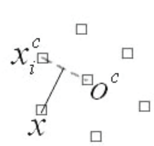


Fig.2 Distance from x to CL

3. Discriminant feature extraction based on center-based distance

Suppose we have the original data set $X = \{x_i : x_i \in R^N\}_{i=1}^n$ and the label set w_k $(k = 1,2,....t)$. The goal of linear feature extraction is to learn a projection matrix A, which projects x_i to $y_i = A^{\mathrm{T}} x_i$, where $y_i \in R^d$ is the projected data with $d << N$, so that in the projected space the data different classes can be effectively discriminated.

3.1. Within-class structure

Be different from the definition in LPP [14], in this paper, the within-class relation matrix W is given as below:

$$W_{ij} = \begin{cases} \exp(-\frac{d^2(x_i, \overline{x_j \mathrm{o}^c})}{\sigma}), (x_i \in w_c \Lambda x_j \in w_c) \\ 0 \qquad , \qquad \text{otherwise} \end{cases}$$

We call $d^2(x_i, \overline{x_j \mathrm{o}^c})$ the center-based distance from x_i to x_j , which has been defined in Sec.2. This distance has more capacity of representation for sample classes than the original samples and thus can capture more information.

Then the within-class of outputs can be rewritten as:

$$\sum_{i,j} ||y_i - y_j||^2 W_{ij} \tag{6}$$

(6) reduces to

$$\sum_{i,j} (y_i^{\mathrm{T}} y_i - y_i^{\mathrm{T}} y_j + y_j^{\mathrm{T}} y_j - y_j^{\mathrm{T}} y_i) W_{ij}$$

$$= \mathrm{tr} \sum_{i,j} (y_i W_{ij} y_i^{\mathrm{T}} - y_i W_{ij} y_j^{\mathrm{T}} + y_j W_{ij} y_j^{\mathrm{T}} - y_j W_{ij} y_i^{\mathrm{T}})$$

$$= \mathrm{tr}(\mathrm{Y}(\mathrm{D}+\mathrm{S})\mathrm{Y}^{\mathrm{T}} - \mathrm{Y}(2\mathrm{W})\mathrm{Y}^{\mathrm{T}})$$

$$= \mathrm{tr}(\mathrm{Y}(\mathrm{D}+\mathrm{S}-2\mathrm{W})\mathrm{Y}^{\mathrm{T}})$$

where D and S are diagonal matrices, i.e. $\mathrm{D}_{ii} = \sum_j W_{ij}$, $\mathrm{S}_{jj} = \sum_i W_{ij}$.

The new objective function is constructed to

$$\sum_{i,j} ||y_i - y_j||^2 W_{ij} = \sum_{i,j} ||x_i - x_j||^2 W_{ij} \tag{7}$$

3.2. Dissimilarities among submanifolds

We can use the sum of the squared distance to measure the dissimilarities among submanifolds. We can construct a label matrix η to mark the label information of each point, where η shows the label information as follows:

$$\eta_{ij} = \begin{cases} 0, \text{if } x_i \text{and} x_j \text{ have the same class label} \\ 1, \text{otherwise} \end{cases}$$

Then the dissimilarities among submanifolds can be defined as the following equation:

$$J_D = \sum_{i,j}^{n} \eta_{ij} (y_i - y_j)(y_i - y_j)^{\mathrm{T}}$$
$$= 2\sum_{i} y_i \eta_{ii} y_i^{\mathrm{T}} - 2\sum_{i,j} y_i \eta_{ij} y_j^{\mathrm{T}}$$
$$= 2tr\{Y(Q-\eta)Y^{\mathrm{T}}\} \tag{8}$$

where Q is a diagonal matrix, i.e. $Q_{ii} = \sum_j \eta_{ij}$.

3.3. Objective function

J_D can be maximized to separate different submanifolds further. The latter will be approached by keeping the center-based distance between within-class samples. So the new algorithm can be expressed as follows:

$$\max \sum_{i,j}^{n} \eta_{ij} ||y_i - y_j||^2$$
$$\text{s.t.} \sum_{i,j} ||y_i - y_j||^2 W_{ij} = \sum_{i,j} ||x_i - x_j||^2 W_{ij} \tag{9}$$

(9) reduces to

$$\max \ tr\{A^{\mathrm{T}} X (Q-\eta) X^{\mathrm{T}} A\} \tag{10}$$
$$\text{s.t.} tr(A^{\mathrm{T}} X (\mathrm{D}+\mathrm{S}-2\mathrm{W}) X^{\mathrm{T}} A) = tr(X(\mathrm{D}+\mathrm{S}-2\mathrm{W})X^{\mathrm{T}})$$

This constrained optimization problem can be figured out by enforcing Lagrange multiplier. A function J(A) can be linearly constructed by the objective function and the constraint:

$$J(A) = \max \ tr\{A^{\mathrm{T}} X (Q-\eta) X^{\mathrm{T}} A -$$
$$\lambda (A^{\mathrm{T}} X (\mathrm{D}+\mathrm{S}-2\mathrm{W}) X^{\mathrm{T}} A - X(\mathrm{D}+\mathrm{S}-2\mathrm{W})X^{\mathrm{T}})\}$$

The optimal transformation matrix A can be obtained from

$$\frac{\partial J(A)}{\partial A} = 2X(Q-\eta)X^{\mathrm{T}} A - 2\lambda X(\mathrm{D}+\mathrm{S}-2\mathrm{W})X^{\mathrm{T}} A = 0$$
$$X(Q-\eta)X^{\mathrm{T}} A = \lambda X(\mathrm{D}+\mathrm{S}-2\mathrm{W})X^{\mathrm{T}} A \tag{11}$$

From Eq.(11), it is shown that A is composed of the eigenvectors associated with the d top eigenvalues by solving the corresponding generalized eigenequation.

4. Experiments and applications

In this section, we design experiments to evaluate the performance of the new approach, in comparison with the performance of PCA, LDA, and SLPP (Supervised LPP) [18], on two well-known face image databases (ORL and Yale) for face recognition. The K-nearest neighbor classifier (K=1) is employed for simplicity.

4.1. Experiments on the ORL Database

The ORL database contains images from 40 individuals, each providing 10 different images. In the experiments, images are cropped based on the centers of eyes and the cropped images are normalized to the 40×40 pixel arrays with 256 gray levels per pixel. p(=3,5) images per person are randomly selected for training and the rest are used for testing. For each given p, we average the realizations over 20 random splits. Fig.3 illustrates the optimized recognition rates versus the variation of dimensions. It shows that the new method outperforms the others when the dimension is over 62 and 40 respectively, when 3 or 5 images per person are selected for training.

The best average results and the standard deviations with the corresponding reduced dimensions, listed in Tab.1, demonstrate that our proposed method consistently outperforms better than PCA, LDE, SLPP.

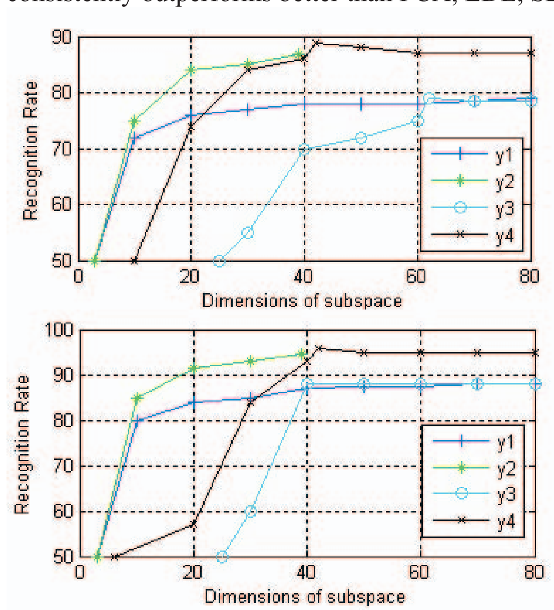


Fig.3 Recognition rate versus dimensionality reduction on the ORL database: (up) 3 train and (down) 5 train.

Tab.1 Best recognition rates (percent) with the reduced dimensions (in parentheses) of four methods on ORL

Method	3 Train	5 Train
PCA	79.24 ± 2.40(118)	88.16 ± 2.47(194)
LDA	86.93 ± 1.83(39)	94.61 ± 1.51(39)
SLPP	79.07 ± 2.34(62)	88.12 ± 2.43(40)
The proposed method	88.80 ± 1.52(42)	95.75 ± 1.73(39)

4.2. Experiments on the Yale Database

The Yale face dataset contains 11 grayscale images for each of the 15 individuals. In our experiment, the images were also resized to 40×40. *6* images per person are randomly selected for training and the rest are used for testing. We average the realizations over 20 random splits. Tab.2 shows the average best recognition rates with the reduced dimensions (in parentheses) of four methods. As observed in Fig.2, this figure indicates that the performance of this new method is still better than those of the other three methods, only with less dimensionality.

Several experiments have been carried out on these two face databases. The new method performs better than PCA, LDA, and SLLE. It is shown that our method is more competitive for preserving intrinsic structure from raw face images based on center-based distance.

Why can our unsupervised method outperform the supervised methods LDA and SLPP? In our opinion, the possible reason is that our method is more robust than the supervised methods to outliers. The outlier images may cause errors in the estimate of within-class scatter in LDA and computing *K*-nearest neighbor samples within the same class in SLPP, thus, make their projection inaccurate. In contrast, our method builds the adjacency relationship of data points using k-nearest neighbors and groups the data in a natural way. Most outlier images of different persons are grouped into new different clusters. Furthermore, the other important cause is that the abilities of within-class structure preserving and classification, based on the measurement of center-based distances, have been combined into the properties of our new algorithm.

Tab.2 Comparison of four methods on Yale

Method	Recognition rate	Feature Dimensions
PCA	81.2%	40
LDA	94.3%	14
SLPP	92.1%	30
The proposed method	95.2%	13

5. Conclusion

A novel discriminant feature extraction algorithm is proposed for face recognition. Our method has two prominent characteristics. On the one hand, it is a representative algorithm based on the preservation of the within-class geometric properties. On the other hand, it is applicable in recognition, since it maximizes the distance between any two samples belonging to different classes. And the experiments on face recognition demonstrate the effectiveness of our algorithm.

Acknowledgments

This work is partially supported by the National Science Foundation of China under Grants No. 60332010, No. 60503026, No. 60472060, No. 60473039, and No. 60632050.

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," J. Cogn Neurosci.3(1), 71–86 (1991).

[2] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell. 19(7), 711–720 (1997).

[3] D. L. Swets and J. Weng, "Using discriminant eigen features for image retrieval," IEEE Trans. Pattern Anal. Mach. Intell. 18(8), 831–836 (1996).

[4] G. Shakhnarovich and B. Moghaddam, "Face recognition in subspaces," in Handbook of Face Recognition, Springer-Verlag, Berlin (2004).

[5] L. Saul and S. Roweis, "Think globally, fit locally: Unsupervised learning of nonlinear manifolds," J. Mach. Learn. Res.4, 119–155 (2003).

[6] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science 290, 2319–2323 (2000).

[7] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science 290, 2323–2326 (2000).

[8] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," Proc. Adv. Neural Infor. Process. Syst., 14, Canada (2001).

[9] H. S. Seung and D. D. Lee, "The Manifold ways of perception," Science 290, 2268–2269 (2000).

[10] A. Shashua, A. Levin, and S. Avidan, "Manifold pursuit: a new approach to appearance based recognition," Proc. Intl. Conf. Patt. Recog. (Aug. 2002).

[11] T. Zhang, J. Yang, D. Zhao, and X. Ge, "Linear local tangent space alignment and application to face recognition," Neurocomputing 70, 1547–1553 (2007)..

[12] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," SIAM J. Sci. Comput. (USA) 26(1), 313–338 (2004).

[13] R.O. Duda, E.H. Peter and G.S. David, Pattern Classification (2nd ed.), Wiley Interscience, 2000.

[14] Y. Fu and T. Huang, "Locally Linear Embedded Eigenspace Analysis," IFP-TR, Univ. of Illinois at Urbana-Champaign, Jan.2005.

[15] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face Recognition Using Laplacianfaces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328-340, Mar. 2005.

[16] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 1, pp.40–51, Jan. 2007.

[17] Qingbin Gao, Zhengzhi Wang. Center-based nearest neighbor classifier [J]. Pattern Recognition, vol.40,no.1,pp. 346～349, 2007.

[18] Caifeng Shan, Shaogang Gong, etc.. A comprehensive empirical study on linear subspace methods for facial expression analysis. Computer Vision and Pattern Recognition Workshop, 153~159, June 2006.