Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **6**, November **2016**

Directional Derivative and Feature Line Based Subspace Learning Algorithm for Classification

Jeng-Shyang Pan[1]

[1]Fuzhou University of International Studies and Trade
Fujian Province, China
jspan@cc.kuas.edu.tw

Lijun Yan[2,3]*, Xingming Liu[2], Keke Wu[2]

[2]Shenzhen Institute of Information Technology
Shenzhen, China
[3]Fujian Provincial Key Lab of Big Data Mining and Applications
Fujian University of Technology, China

Received September, 2016; revised September, 2016

Abstract. *A novel subspace learning algorithm based on nearest feature line and directional derivative gradient is proposed in this paper. The proposed algorithm combines neighborhood discriminant nearest feature line analysis and directional derivative gradient to extract the local discriminant features of the samples. Directional derivative gradient gives some directional features of samples. Feature line can extract some discriminant feature of samples. A discriminant power criterion based on nearest feature line is used to find the most discriminant direction in this paper. The proposed algorithm aims to extract some discriminant and directional features of samples. Some experiments are implemented to evaluate the proposed algorithm and the experimental results demonstrate the effectiveness of the proposed algorithm.*

Keywords: Directional derivative gradient; Feature extraction; Nearest feature line.

1. **Introduction.** Over the past 20 years, biometric and related technologies have become very popular in person authentication, computer vision and machine learning [1, 2, 3]. Many researches on biometric were based on image classification, so a lot of image feature extraction algorithms were proposed. Principal Component Analysis (PCA) [4], Linear Discriminant Analysis (LDA)[5] are some of most popular approaches. However, PCA projects the original samples to a low dimensional space, which is spanned by the eigenvectors associated with the largest eigenvalues of the covariance matrix of all samples. PCA is the optimal representation of the input samples in the sense of minimizing the mean squared error. However, PCA is an unsupervised algorithm, which may lead to a lower recognition accuracy. Linear subspace analysis (LDA) finds a transformation matrix U that linearly maps high-dimensional sample $x \in R^n$ to low-dimension data y by $y = U^T x \in R^m$, where $n > m$. LDA can calculate an optimal discriminant projection by maximizing the ratio of the trace of the between-class scatter matrix to the trace of the within-class scatter matrix. LDA takes consideration of the labels of the input samples and improves the classification ability. However, LDA suffers from the small sample size (SSS) problem. Some algorithms using the kernel trick are developed in recent years [6], such as kernel principal component analysis (KPCA)[7, 8], kernel discriminant analysis

(KDA) [9] and Locality Preserving Projection[10] used in many areas[11]. Researchers have developed a series of KDA and related algorithms[12, 13, 14].

An alternative way to handle the above problem is to extract features from the face image matrix directly. In resent years, several methods based on matrix are proposed, such as Two-Dimensional Principal Component Analysis (2DPCA) [15] and Two-Dimensional Linear Discriminant Analysis (2DLDA) [16]. 2DPCA and 2DLDA can extract the features in a straightforward manner based on the image matrix projection. And these algorithms, not only greatly reduce the computational complexity, but also enhance the recognition effectiveness. Many works based on matrix were presented in these years [17].

The above algorithms are based on Euclidean Distance. Nearest feature line (NFL) [18] is a classifier, proposed by Li in 1998, firstly. In particular, it performs better when only limited samples are available for training. The basic idea underlying the NFL approach is to use all the possible lines through every pair of feature vectors in the training set to encode the feature space in terms of the ensemble characteristics and the geometric relationship. As a simple yet effective algorithm, the NFL has shown its good performance in face recognition, audio classification, image classification, and retrieval. The NFL takes advantage of both the ensemble and the geometric features of samples for pattern classification. Some improved algorithms were proposed in the recent years [19, 20].

While NFL has achieved reasonable performance in data classification, most existing NFL-based algorithms just use the NFL metric for classification and not in the learning phase. While classification can be enhanced by NFL to a certain extent, the learning ability of existing subspace learning methods remains to be poor when the number of training samples is limited. To address this issue, a number of enhanced subspace learning algorithms based on the NFL metric have been proposed, recently. For example, Zheng et al. proposed a Nearest Neighbour Line Nonparametric Discriminant Analysis (NFL-NDA) [21] algorithm, Pang et al. presented a Nearest Feature Line-based Space (NFLS)[22] method, Lu et al. proposed the Uncorrelated Discriminant Nearest Feature Line Analysis (UDNFLA) [23], Yan et al. proposed the Neighborhood Discriminant Nearest Feature Line Analysis (NDNFLA) [24] and some improve algorithms based on NFL [25].

However, the most subspace learning based image feature extraction algorithms only extract the statistical features of images and ignore the features of images as two-dimensional signals. In this paper, a novel image feature extraction algorithm, named Directional Discriminant Analysis (DDA), is proposed. Its effectiveness of the proposed method is verified by some experiments on AR face database. The rest of the paper is organized as follows. In section 2 introduces some preliminaries. In section 3, we give the presentation of the proposed method. In section 4, the experiments are implemented to justify the superiority of the proposed algorithm. And conclusions are made in section 5.

2. **Preliminaries.**

2.1. **Directional Derivative Gradient.**

Given a signal f and a direction vector $(\sin\theta, \cos\theta)$, the first directional derivative [26] $f'_\theta(r, c)$ of f in the direction θ can be treated as the component of the gradient ∇f along the direction vector, that is,

$$f'_\theta(r, c) = \frac{\partial f}{\partial r}\sin\theta + \frac{\partial f}{\partial c}\cos\theta \tag{1}$$

Then the points in $I_{\alpha,\beta}$ a digital line with slope α and intercept β. Given an α, the digital lines with different β can cover the plane.

2.2. **Nearest feature line.** Nearest feature line is a classifier [18]. It is first presented by Stan Z. Li and Juwei Lu. Given a training samples set, $X = \{x_n \in R^M : n = 1, 2, \cdots, N\}$, denote the class label of x_i by $l(x_i)$, the training samples sharing the same class label with x_i by $P(i)$, and the training samples with different label with x_i by $R(i)$. NFL generalizes each pair of prototype feature points belonging to the same class: $\{x_m, x_n\}$ by a linear function $L_{m,n}$, which is called the feature line. The line $L_{m,n}$ is expressed by the span $L_{m,n} = sp(x_m, x_n)$. The query x_i is projected onto $L_{m,n}$ as a point $x^i_{m.n}$. This projection can be computed as

$$x^i_{m,n} = x_m + t(x_n - x_m) \tag{2}$$

where $t = [(x_i - x_n)(x_m - x_n)]/[(x_m - x_n)^T(x_m - x_n)]$.

The Euclidean distance of x_i and $x^i_{m,n}$ is termed as FL distance. The less the FL distance is, the bigger probability that x_i belongs to the same class as x_m and x_n is. Fig. 1 shows a sample of FL distance. In Fig. 1, the distance between y_p and the feature line $L_{m,n}$ equals to the distance between y_q and y_p, where y_p is the projection point of y_q to the feature line $L_{m,n}$.

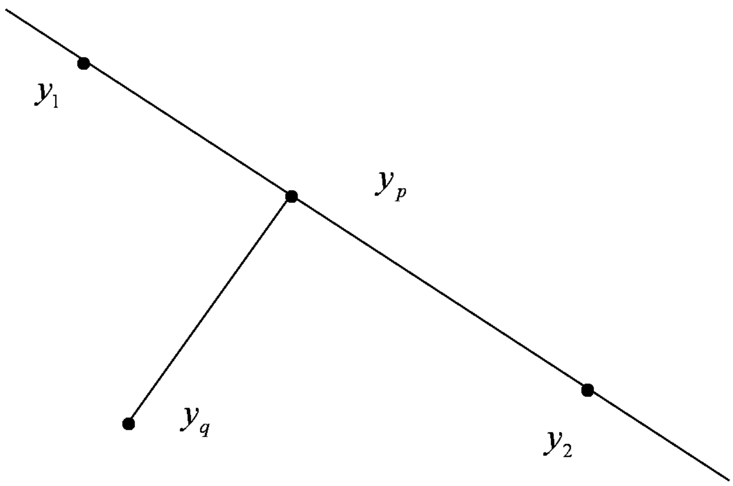

FIGURE 1. Feature Line Distance

2.3. **NDNFLA.** Let's introduce two definitions firstly.

Definition 2.1. *Homogeneous neighborhoods: For a sample x_i, its k nearest homogeneous neighborhood N_i^o is the set of k most similar data which are in the same class with x_i.*

Definition 2.2. *Heterogeneous neighborhoods: For a sample x_i, its k nearest Heterogeneous neighborhoods N_i^e is the set of k most similar data which are not in the same class with x_i.*

In NDNFLA approach, the optimization problem is as follows:

$$\begin{aligned} \max J(W) = (&\sum_{i=1}^{N} \frac{1}{NC^2_{\left|N_i^e\right|}} \sum_{x_m, x_n \in N_i^e} \left\|W^T x_i - W^T x^i_{m,n}\right\|^2 \\ &- \sum_{i=1}^{N} \frac{1}{NC^2_{\left|N_i^o\right|}} \sum_{x_m, x_n \in N_i^o} \left\|W^T x_i - W^T x^i_{m,n}\right\|^2) \end{aligned} \tag{3}$$

Using matrix computation,

$$\begin{aligned}&\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^e\right|}}\sum_{x_m,x_n\in N_i^e}\left\|W^Tx_i-W^Tx^i_{m,n}\right\|^2\\&=\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^e\right|}}\sum_{x_m,x_n\in N_i^e}\mathrm{tr}[W^T(x_i-x^i_{m,n})(x_i-x^i_{m,n})^TW]\\&=\mathrm{tr}\{W^T\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^e\right|}}\sum_{x_m,x_n\in N_i^e}[(x_i-x^i_{m,n})(x_i-x^i_{m,n})^T]W\}\end{aligned}\tag{4}$$

where tr denotes the trace of a matrix. Similar with the above,

$$\begin{aligned}&\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^o\right|}}\sum_{x_m,x_n\in N_i^o}\left\|W^Tx_i-W^Tx^i_{m,n}\right\|^2\\&=\mathrm{tr}\{W^T\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^o\right|}}\sum_{x_m,x_n\in N_i^o}[(x_i-x^i_{m,n})(x_i-x^i_{m,n})^T]W\}\end{aligned}\tag{5}$$

Then the problem becomes

$$\max J(W)=\mathrm{tr}[W^T(A-B)W]\tag{6}$$

where

$$A=\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^e\right|}}\sum_{x_m,x_n\in N_i^e}[(x_i-x^i_{m,n})(x_i-x^i_{m,n})^T]\tag{7}$$

$$B=\sum_{i=1}^{N}\frac{1}{NC^2_{\left|N_i^o\right|}}\sum_{x_m,x_n\in N_i^o}[(x_i-x^i_{m,n})(x_i-x^i_{m,n})^T]\tag{8}$$

A length constraint $w^Tw=1$ is imposed on the proposed NDNFLA. Then, the optimal projection W of NDNFLA can be obtained by solving the following eigenvalue problem.

$$(A-B)w=\lambda w\tag{9}$$

Let $w_1,w_2,\cdots,w_q$ be the eigenvectors of formula(9) corresponding to the q largest eigenvalues ordered according to $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_q$. An $M\times q$ transformation matrix $W=[w_1,w_2,\cdots,w_q]$ can be obtained to project each sample $M\times 1$ x_i into a feature vector $q\times 1$ y_i as follows:

$$y_i=W^Tx_i,\qquad i=1,2,\cdots,N\tag{10}$$

3. **The proposed algorithm.** In this paper, a Directional Discriminant Analysis (DDA) is proposed to extract the directional features of the images. In DDA approach, the directional features with a most discriminant direction are extracted based on directional derivative gradient of images.

Firstly, discriminant power criterion based on NFL is proposed in this section. Let $X=\{X_1,X_2,\cdots,X_N\}\subset R^{d_1\times d_2}$ denote the prototype sample set. X_i^θ denotes the directional derivative gradient of images X_i with the direction θ. Then transform the matrix X_i^θ to a vector $x_i^\theta\in R^D$, where $D=d_1\times d_2$. Denote $X_\theta=\{x_1^\theta,x_2^\theta,\cdots,x_N^\theta\}\subset R^D$.

Let $l_i(\theta)$ denote the number of FLs in the same class with x_i^θ among its k nearest feature lines in X_θ. Then, let $L(\theta)=\sum_{i=1}^N l_i(\theta)$. At last, let

$$J_{DP}(\theta)=\frac{L(\theta)}{k*N}\tag{11}$$

According to the formula(11), it is clear that the bigger $J_{DP}(\theta)$ is, the more discriminant features are. So the most discriminant direction can be find by maximizing J_{DP}.

The main idea of the proposed feature extraction algorithm, is to extract the local discriminant features from the most discriminant direction. The detailed procedure of proposed method is as follows:

Training stage:

Step 1, using the discriminant power criterion based on NFL, find the most discriminant direction θ_0;

Step 2, perform the directional derivative gradient operator with direction θ_0 to all the prototype samples to get a new prototype samples set;

Step 3, apply NDNFLA to find the optimal transformation matrix W on the new prototype samples set;

Step 4, Extract the features of prototype samples following formula(10).

Classification stage:

Step 1, perform the directional derivative gradient operator with direction θ_0 to the query sample;

Step 2, Extract the features of query following formula(10);

Step 3, Classify with NFL.

4. **Experimental results.** To evaluate the performance of the proposed algorithm, some experiments are implemented on AR face database [27] and ORL face database [28]. In the following, we assess the feasibility and performance of the proposed method for face recognition with one sample problem. Comparative performance is carried out against some popular face recognition algorithm such as the PCA, UDNFLA, NFLS, NDNFLA. The following experiments are implemented on a PC with Athlon 2.5GHz CPU and 768MB RAM and programmed in the MATLAB platform. For each database, five image samples are selected randomly for training and the other samples are used for test. This procedure will be performed ten times. The maximum average recognition rate (MARR) will be used to evaluate the performance of different algorithms. In the following experiments, NN is the classifier for classfication.

4.1. **Experimental results on AR face database.** AR face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different illumination conditions, facial expressions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. In the following experiments, only nonoccluded images of 120 people in AR face database are selected. Five images per person are randomly selected for training and the other images are for testing. This system also runs 20 times. Some samples of AR face database are shown in Fig. 2. Table 1 tabulates MARRs of these algorithms on AR face database. Clearly, MARR of the proposed algorithm is higher than other approaches.

4.2. **Experimental results on ORL face database.** The ORL face database contains 400 face images, 10 different face images per person for 40 individuals. Some face images are captured at different times . There are facial expressions (open or closed eyes, smiling or non-smiling) and facial details (glass or no glasses). These face images are taken with a tolerance for some tilting and rotation of the face up to 20 . All face images are gray with 256 levels and size of 112 92. Yale face database contains 165 images of 15 individuals (each person has 11 different images). These images are under variations with following facial expressions or configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised and wink. All images are gray with 256

Figure 2. Some samples of AR face database

Table 1. MARR of different algorithms on AR face database

Algorithms	MARR	Feature dimension
fisherface	0.9481	120
PCA+NN	0.7604	120
PCA+NFL	0.8521	190
UDNFLA	0.9353	120
NFLS	0.9126	190
NDNFLA	0.9690	150
Proposed algorithm	0.9735	130

levels and size of 100 100 pixels. Fig. 3 gives some examples in ORL face database.

Figure 3. Some samples of AR face database

Table 2 tabulates MARRs of these algorithms on AR face database. The MARR of the proposed algorithm is higher than those of PCA+NN, PCA+NFL

5. **Conclusions.** In this paper, a novel image feature extraction algorithm, called Directional Discriminant Analysis is proposed. DDA is based on the directional derivative gradient and nearest feature line. It can find the optimal direction adaptively to extract the most discriminant directions based on NFL. The experimental results on ORL

Table 2. MARR of different algorithms on ORL face database

Algorithms	MARR	Feature dimension
fisherface	0.9361	130
PCA+NN	0.8872	120
PCA+NFL	0.8893	180
UDNFLA	0.9217	110
NFLS	0.9082	120
NDNFLA	0.9438	120
Proposed algorithm	0.9560	140

face database and AR face database show the effectiveness of the proposed algorithm. The MARR of the proposed algorithm is higher that those of PCA, UDNFLA, NFLS, NDNFLA.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (NO. 61371178, NO. 61202456, and NO. 70903016), the Natural Science Foundation of Guangdong (2015A030310511, 2014A030310299) and Project of Shenzhen Science and Technology Innovation Committee (JCYJ20150417094158026, JCYJ20150417094158019).

REFERENCES

[1] J.-B. Li, S.-C. Chu , J.-S. Pan , and L. C. Jain, Multiple Viewpoints Based Overview for Face Recognition. *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, pp. 352–369, 2012

[2] S. Krinidis, and I. Pitas, Statistical Analysis of Human Facial Expressions, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, pp. 241–260, 2010

[3] L. Zhang, L. Zhang, D. Zhang, and. Guo, Phase Congruency Induced Local Features for Finger-Knuckle-Print Recognition, *Pattern Recognition*, vol. 45, 2522–2531, 2012

[4] J. Wu, Z. H. Zhou, Face Recognition with One Training Image Per Person. *Pattern Recognition Letters* 23, pp. 1711–1719, 2002

[5] P. N. Belhumenur, J. P. Hepanha, D. J. Kriegman, Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Analysis Machine Intelligence* 19, pp. 711–720, 1997.

[6] J.-B. Li, J.-S. Pan, and Z.-M. Lu, Kernel Optimization-Based Discriminant Analysis for Face Recognition, *Neural Computing and Applications* vol. 18, 603–612, 2009

[7] J. Yang , A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence* vol. 27, pp. 230–244, 2005

[8] J.-B. Li, L.-J. Yu, S.-H. Sun, Refined Kernel Principal Component Analysis Based Feature Extraction, *Chinese Journal of Electronics*, vol. 20, 467–470, 2011

[9] J.-S. Pan, J.-B. Li, and Z.-M. Lu, Adaptive Quasiconformal Kernel Discriminant Analysis, *Neurocomputing*, vol. 71, 2754–2760, 2008

[10] J.-B. Li, J.-S. Pan,S.-M. Chen, Kernel Self-Optimized Locality Preserving Discriminant Analysis for Feature Extraction and Recognition, *Neurocomputing*, vol. 74, 3019–3027, 2011

[11] J.-B. Li, J.-S. Pan, and S.-C. Chu, Kernel Class-wise Locality Preserving Projection, *Information Sciences*, vol. 178, 1825–1835, 2008

[12] J.-B. Li, H. Gao, H. and J.-S. Pan,Common Vector Analysis of Gabor Features with Kernel Space Isomorphic Mapping for Face Recognition, *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 10, October 2010.

[13] Y. Xu, D. Zhang, Z. Jin, M. Li, and J. Y. Yang, A Fast Kernel-Based Nonlinear Discriminant Analysis for Multi-Class Problems. *Pattern Recognition*, vol. 39, pp. 1026–1033, 2006

[14] Y. Xu, D. Zhang, Represent and Fuse Bimodal Biometric Images at the Feature Fevel: Complex-Matrix-Based Fusion Scheme, *Opt. Eng.*, vol. 49, pp. 037002, 2010

[15] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, 131–137, 2004

[16] J. Yang, D. Zhang, X. Yong, and J. Y. Yang, Two-dimensional discriminant transform for face recognition, *Pattern Recognition* , vol. 37, 1125–1129, 2005

[17] L. Yan, J.-S. Pan, S.-C Chu, K. K. Muhammad, Adaptively weighted sub-directional two-dimensional linear discriminant analysis for face recognition, *Future Generation Computer Systems*, vol. 28, 232–235, 2012

[18] S. Z. Li, J. Lu, Face Recognition Using the Nearest Feature Line Method, *IEEE Trans. Neural Networks*, vol. 10, PP. 439–443, 1999

[19] Q. Feng, J.-S. Pan, and L. Yan, Restricted Nearest Feature Line with Ellipse for Face Recognition, *Journal of Information Hiding and Multimedia Signal Processing* vol. 3, 297–305, 2012

[20] Q. Feng, J.-S. Pan, and L. Yan, Nearest Feature Centre Classifier for Face Recognition, *Ellectronics Letters*, vol. 48, 1120-1122, 2012

[21] Y. -J. Zheng, J. -Y. Yang, J. Yang, X. -J. Wu, Z. Jin, Nearest Neighbour Line Nonparametric Discriminant Analysis for Feature Extraction, *Electronics Letters*, vol. 42, PP. 679–680, 2006

[22] Y. Yang, Y. Yuan, X. Li, Generalised Nearest Feature Line for Subspace Learning, *Electronics Letters*, vol. 43, pp. 1079–1080, 2007

[23] J. Lu, Y. P. Tan, Uncorrelated Discriminant Nearest Feature Line Analysis for Face Recognition, *IEEE Signal Processing Letter*, vol. 17, pp. 185–188, 2010

[24] L. Yan, Z. Pan, S. C. Chu, J.F. Roddick, Neighborhood Discriminant Nearest Feature Line Analysis for Face Recognition, *Journal of Internet Technology* , vol. 14 344–347, 2013

[25] L. Yan, C. Wang, S.-C. Chu, J.-S. Pan, Discriminant Analysis Based on Nearest Feature Line, In: *Proceedings of Intelligent Science and Intelligent Data Engineering 2013*. LNCS, vol. 7751, pp. 356–363. Springer, Heidelberg, 2013

[26] O. A. Zuniga, R. M. Haralick, Integrated Directional Derivative Gradient Operator, *IEEE Trans. Systems, Man, and Cybernetics*, vol. 17, 508–517, 1987

[27] A. M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report 24, 1998.

[28] Olivetti and Oracle Research Laboratory, ORL face database. Available: http://www.cam-orl.co.uk/facedatabase.html, 1994.