Journal of Information Hiding and Multimedia Signal Processing ©2015 ISSN 2073-4212
Ubiquitous International Volume **6**, Number **3**, May **2015**

Kernel-Based Nonparametric Fisher Classifier for Hyperspectral Remote Sensing Imagery

Xiaodan Xie and Bohu Li

College of Automation Science and Electrical Engineering
Beihang University, Beijing, China
xiexiaodanbeijing@126.com

Xudong Chai

Beijing Simulation Center, Beijing, China

Received November, 2014; revised December, 2014

Abstract. *Hyperspectral Imagery Sensing (HIS) is widely gained tremendous popularity in many research areas such as remotely sensed satellite imaging and aerial reconnaissance. HIS is an important technique with the measurement, analysis, and interpretation of spectra acquired sensing scene an airborne or satellite sensor. The development of sensor technology brought the developing of collecting image data using hyperspectral instruments with hundreds of contiguous spectral channels. The preprocessing of hyperspectral sensing data is a feasible way through machine learning technology. Among these machine learning methods, kernel learning is a feasible nonlinear feature extraction on hyperspectral sensing data. The nonlinear problems are solved with kernel function, and system performances such as recognition accuracy, prediction accuracy are largely increased. In this paper, we present a novel Kernel-Based Nonparametric Fisher Classifier (KNFC) for hyperspectral remote sensing imagery. Firstly, we have a comprehensive theoretical analysis on Nonparametric Discriminant Analysis. And NDA has its limitations on extracting the nonlinear features owing to the high nonlinear and complex distribution of the hyperspectral imagery data. In order to improve the limitation of NDA on hyperspectral remote sensing imagery, we introduce the kernel trick to NDA to develop Kernel-Based Nonparametric Fisher Classifier to enhance its ability on hyperspectral Imagery Sensing data. The feasibility of the KNFC is testified on the ORL and YALE databases, and then some experiments are implemented on two real data sets including Indian Pines and Washington, D.C. Mall, with various spectral and spatial resolutions reflecting different environments of remote sensing.*

Keywords: Hyperspectral imagery sensing, Kernel learning, Nonparametric discriminant analysis, Kernel-Based nonparametric fisher classifier

1. **Introduction.** The development of sensor technology brought the developing of collecting image data using hyperspectral instruments with hundreds of contiguous spectral channels. As Hyperspectral Imagery Sensing (HIS) is an important technique with the measurement, analysis, and interpretation of spectra acquired sensing scene an airborne or satellite sensor, HIS is widely gained tremendous popularity in many research areas such as remotely sensed satellite imaging and aerial reconnaissance, Dimensionality reduction based preprocessing of hyperspectral sensing data is a feasible way through machine learning technology. Dimensionality reduction methods are the popular face recognition methods [1, 2]. Linear Discriminant Analysis (LDA) as the dimensionality reduction method is widely used in many areas, such as face recognition [5], image retrieval [7],

object recognition, handwriting digit recognition, and texture classification, and so on. However, most LDA-based methods should satisfy two preconditions, i.e., the unimodal distribution of samples and the different scatter of class means of samples, but it is difficult to satisfy these preconditions in the practical applications. LDA often encounters the so-called ill-posed problems when applied to a small samples size problem like face recognition. Furthermore, unless a posteriori probability function is normal (Gaussian), the extracted features are suboptimal in the Bayes sense, although they are optimal with regard to the Fisher criterion. To address such problems of LDA, Fukunaga and Mantock presented a nonparametric discriminant analysis (NDA) method [8].In nonparametric discriminant analysis the between-class scatter is of nonparametric nature. This scatter matrix is generally full rank, thus loosening the bound on extracted feature dimensionality. Also, the nonparametric structure of this matrix inherently leads to extracted features that preserve relevant structures for classification. Recently, Li et al. proposed Nonparametric Discriminant Analysis (NDA) reported an excellent recognition performance for face recognition [9]. However, NDA has its limitations on extracting the nonlinear features of face images for recognition because the distribution of images, such as face images, under a perceivable variation in viewpoint, illumination or facial expression is highly nonlinear and complex, and the linear techniques cannot provide reliable and robust solutions to those face recognition problems with complex face variations. On current kernel learning methods, the performance of many linear learning methods is improved because the data distribution in the nonlinear feature space is easy to classification owing to kernel mapping. The geometrical structure of the data in the kernel mapping space, which is totally determined by the kernel function, has significant impact on the performance of these kernel learning methods. The discriminative ability of the data in the feature space could be even worse if an inappropriate kernel is used. In the previous work, researchers optimized the parameters of kernel function to improve KDA [11, 12, 13], but these methods only choosing the optimal parameter of kernel from a set of discrete values which are created in advance. The geometry structure of data distribution in the kernel space is not be changed only through the changing the parameters of kernel. Xiong proposed a data-depend kernel for kernel optimization [14], and Amari presented support vector machine classifier through modifying the kernel function [15]. Some algorithms using the kernel trick are developed in recent years, such as kernel discriminant analysis (KDA) [17] and support vector machine(SVM) [18]. Researchers have developed a series of KDA algorithms [19, 20, 21].

In this paper, we present a novel Kernel-Based Nonparametric Fisher Classifier (KNFC) for hyperspectral remote sensing imagery. Firstly, we have a comprehensive theoretical analysis on Nonparametric Discriminant Analysis. And NDA has its limitations on extracting the nonlinear features owing to the high nonlinear and complex distribution of the hyperspectral imagery data. In order to improve the limitation of NDA on hyperspectral remote sensing imagery, we introduce the kernel trick to NDA to develop Kernel-Based Nonparametric Fisher Classifier to enhance its ability on hyperspectral Imagery Sensing data. The feasibility of the KNFC is testified on the ORL and YALE databases, and then some experiments are implemented on two real data sets including Indian Pines and Washington, D.C. Mall, with various spectral and spatial resolutions reflecting different environments of remote sensing.

2. **Kernel-Based Nonparametric Fisher Classifier.** In this section, we present a novel Kernel-Based Nonparametric Fisher Classifier (KNFC). Firstly, we have a comprehensive theoretical analysis on Nonparametric Discriminant Analysis. And we introduce

the kernel trick to NDA to develop Kernel-Based Nonparametric Fisher Classifier to enhance its ability on the nonlinear feature extraction method.

Firstly, we analyze Nonparametric Discriminant Analysis as follows. Supposed that

$$S_W = \sum_{i=1}^{c}\sum_{k=1}^{k_1}\sum_{l=1}^{n_i}\Big(x_i^l - N(x_i^l, i, k)\Big)\Big(x_i^l - N(x_i^l, i, k)\Big)^T \tag{1}$$

$$S_B = \sum_{i=1}^{c}\sum_{j=1, j\neq i}^{c}\sum_{k=1}^{k_2}\sum_{l=1}^{n_i} w(i,j,k,l)\Big(x_i^l - N(x_i^l, j, k)\Big)\Big(x_i^l - N(x_i^l, j, k)\Big)^T \tag{2}$$

where $w(i,j,k,l)$ is defined as

$$w(i,j,k,l) = \frac{min\{d^\alpha(x_i^l, N(x_i^l, i, k)), d^\alpha(x_i^l, N(x_i^l, j, k))\}}{d^\alpha(x_i^l, N(x_i^l, i, k)) + d^\alpha(x_i^l, N(x_i^l, j, k))} \tag{3}$$

where $d(v_1, v_2)$ is the Euclidean distance between two vector v_1 and v_2, and α is the a parameter ranging from zero to infinity which control the changing of the weight with respect to the ratio of the distance. $N(x_i^l, j, k)$ is the kth nearest neighbor from class j to the vector x_i^l which from lth sample of ith class.

Algorithm Procedure:

Step 1. Calculate kth nearest neighbor vector $N(x_i^l, j, k)$ from class j to the vector x_i^l which from lth sample of ith class.

Step 2. Calculate the Euclidean distance $d(v_1, v_2)$ between two vector v_1 and v_2, and the weight parameter $w(i,j,k,l)$.

Step 3. Calculate within-scatter matrix S_W.

Step 4. Calculate the matrix S_B.

Step 5. Calculate the projection matrix W.

NDA is essentially a linear feature extraction algorithm. Although it is reported good performance on face recognition, however, there still is the space to improve its recognition performance through enhancing its ability to extracting the nonlinear facial features owing to the variation in illumination. In order to solve this limitation, we introduce the kernel trick develop a novel feature method called Nonparametric Kernel Discriminant Analysis (NKDA) as follows. Kernel methods have been widely used in the previous works [3, 4, 6]. The main idea is based on a conceptual transformation from the input space into a nonlinear high-dimensional feature space. Supposed that M training samples $\{x_1, x_1, \cdots, x_M\}$ with L class labels take values in an N-dimensional space R^N, the data in R^N are mapped into a feature space F via the following nonlinear mapping:

$$\Phi : \mathbb{R}^N \longrightarrow F, x \longmapsto \Phi(x) \tag{4}$$

We have theoretical analysis on NDA in the feature space F to develop NKDA. Supposed that

$$S_W^\Phi = \sum_{i=1}^{c}\sum_{k=1}^{k_1}\sum_{l=1}^{n_i}\Big(\Phi(x_i^l) - N(\Phi(x_i^l), i, k)\Big)\Big(\Phi(x_i^l) - N(\Phi(x_i^l), i, k)\Big)^T \tag{5}$$

$$S_B^\Phi = \sum_{i=1}^{c}\sum_{j=1, j\neq i}^{c}\sum_{k=1}^{k_2}\sum_{l=1}^{n_i} w^\Phi(i,j,k,l)\Big(\Phi(x_i^l) - N(\Phi(x_i^l), j, k)\Big)\Big(\Phi(x_i^l) - N(\Phi(x_i^l), j, k)\Big)^T \tag{6}$$

where $w^{\Phi}(i,j,k,l)$ is defined as

$$w^{\Phi}(i,j,k,l) = \frac{min\{d^{\alpha}(\Phi(x_i^l), N(\Phi(x_i^l), i, k)), d^{\alpha}(\Phi(x_i^l), N(\Phi(x_i^l), j, k))\}}{d^{\alpha}(\Phi(x_i^l), N(\Phi(x_i^l), i, k)) + d^{\alpha}(\Phi(x_i^l), N(\Phi(x_i^l), j, k))} \tag{7}$$

where $d(\Phi(v_1), \Phi(v_2))$ is the Euclidean distance between two vector v_1 and v_2 in the kernel space, and α is the the a parameter ranging from zero to infinity which control the changing of the weight with respect to the ratio of the distance. $N(\Phi(x_i^l), i, k)$ is the kth nearest neighbor from class j to the vector x_i^l which from lth sample of ith class with the similarity measure with Euclidean distance in the kernel space. In order to the clear description of NKDA, we rewrite the equation as

$$S_W^{\Phi} = \sum_{i=1}^{c}\sum_{k=1}^{k_1}\sum_{l=1}^{n_i}\Big(\Phi(x_i^l) - \Phi(y_i^k)\Big)\Big(\Phi(x_i^l) - \Phi(y_i^k)\Big)^T \tag{8}$$

$$S_B^{\Phi} = \sum_{i=1}^{c}\sum_{j=1, j\neq i}^{c}\sum_{k=1}^{k_2}\sum_{l=1}^{n_i} w^{\Phi}(i,j,k,l)\Big(\Phi(x_i^l) - \Phi(y_i^k)\Big)\Big(\Phi(x_i^l) - \Phi(y_i^k)\Big)^T \tag{9}$$

$$w^{\Phi}(i,j,k,l) = \frac{min\{d^{\alpha}(\Phi(x_i^l) - \Phi(y_i^k)), d^{\alpha}(\Phi(x_i^l) - \Phi(y_i^k))\}}{d^{\alpha}(\Phi(x_i^l) - \Phi(y_i^k)) + d^{\alpha}(\Phi(x_i^l) - \Phi(y_i^k))} \tag{10}$$

where $\Phi(y_i^k) = N(\Phi(x_i^l), i, k)$ and $\Phi(y_j^k) = N(\Phi(x_i^l), j, k)$. //It is easy to obtain

$$d^{\alpha}(\Phi(x_i^l), \Phi(y_i^k)) = \parallel \Phi(x_i^l) - \Phi(y_i^k) \parallel^{\alpha} = \Big(k(x_i^l, x_i^l) - 2k(x_i^l, y_i^k) + k(y_i^k, y_i^k)\Big)^{\frac{\alpha}{2}} \tag{11}$$

Fisher criterion is defined by

$$J(V) = \frac{V^T S_B^{\Phi} V}{V^T S_W^{\Phi} V} \tag{12}$$

Where V is the discriminant vector, and S_B^{Φ} and S_W^{Φ} are he between classes scatter matrix and the total population scatter matrix respectively. According to the Mercer kernel function theory, any solution V belongs to the span of all training pattern in R^N. Hence there exist coefficients $c_p (p = 1, 2, \cdots, M)$ such that

$$V = \sum_{p-1}^{M} c_p \Phi(x_p) = \Psi_{\alpha} \tag{13}$$

where $\Psi = [\Phi(x_1), \Phi(x_1), \cdots, \Phi(x_M)]$ and $\alpha = [c_1, c_1, \cdots, c_M]^T$.

$$J(\alpha) = \frac{\alpha^T B \alpha}{\alpha^T W \alpha} \tag{14}$$

where $B = \sum\limits_{i=1}^{c}\sum\limits_{j=1,j\neq i}^{c}\sum\limits_{k=1}^{k_2}\sum\limits_{l=1}^{n_i} w^{\Phi}(i,j,k,l)B(i,j,k,l)$, $W = \sum\limits_{i=1}^{c}\sum\limits_{k=1}^{k_2}\sum\limits_{l=1}^{n_i} W(i,k,l)$, $B(i,j,k,l) = K_1(i,j,k,l)^T K_1(i,j,k,l)$, $W(i,k,l) = K_2(i,k,l)^T K_2(i,k,l)$, $K_1(i,j,k,l) = [k(x_1, x_(i)^l), \cdots, k(x_M, x_(i)^l)] - [k(x_1, y_(j)^k), \cdots, k(x_M, y_(j)^k)]$ and $K_2(i,k,l) = [k(x_1, x_i^l), \cdots, k(x_M, x_i^l)] - [k(x_1, Y_i^k), \cdots, k(x_M, y_i^k)]$, where k_1 and k_2 are the total number of k nearest neighbors.

The projection matrix $V = [\alpha_1, \alpha_2, \cdots, \alpha_d]$ is easy to be obtained by the eigenvectors of $W^{-1}B$ corresponding to the d largest eigenvalue. Calculate the eigenvectors of $W^{-1}B$ is the same to simulataneous diagonalization of W and B. Firstly, the eigenvector matrix Φ and the corresponding eigenvalue matrix Θ of W are solved, and then project the class centers onto $\Phi\Theta^{\frac{1}{2}}$, thus B is transformed to $B_K = \Theta^{-\frac{1}{2}}\Phi^T B \Phi \Theta^{-\frac{1}{2}}$. Finally, solve the eigenvector matrix Ψ and the eigenvalue matrix Λ of B_K, the projection matrix V is equal to $V = \Phi\Theta^{-\frac{1}{2}}$.

3. **Experimental Results.** We perform two sets of experiments to testify the feasibility of the proposed Kernel-Based Nonparametric Fisher Classifier on ORL and YALE databases, and then evaluate the performance of hyperspectral remote sensing imagery on two real data sets including Indian Pines and Washington, D.C. Mall with various spectral and spatial resolutions.

In this paper, we emphasize KNFC based classification method on image analysis for breast cancer diagnosis. The framework of the real application system is shown in Figure 1. The practical system collects the image of the patient via the imaging sensor, and the images are preprocessed with the image preprocessing modular. The algorithm is implemented with software not hardware in the practical applications. Many images are saved in the images database. In the system, all algorithms are implemented with the software, while imaging modular is implemented with hardware.

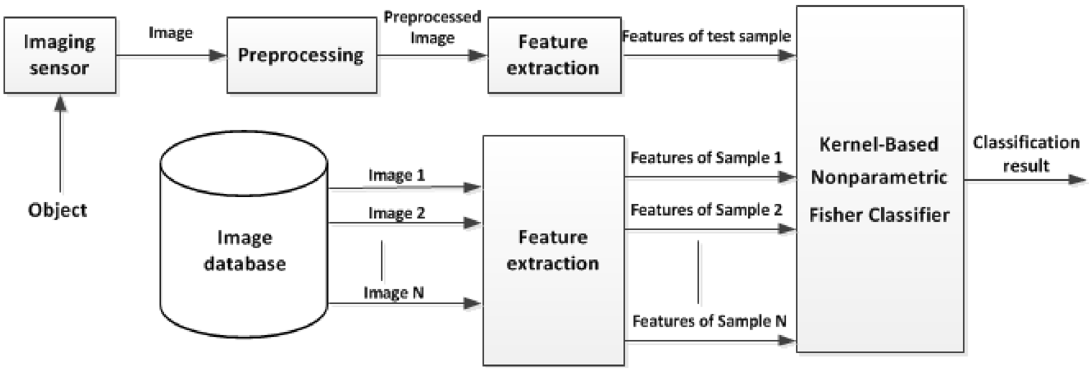


Figure 1. The framework of a real system example.

Firstly, we implement our algorithm in the two face databases, ORL face database [11], developed at the Olivetti Research Laboratory, Cambridge, U.K., is composed of 400 grayscale images with 10 images for each of 40 individuals. To reduce computation complexity, we resize the original ORL face images sized pixels with a 256 gray scale to pixels. We randomly select 5 images from each subject, 200 images in total for training, and the rest images are used to test the performance. Yale face database [16] was constructed at the Yale Center for Computational Vision and Control. It contains 165 grayscale images of 15 individuals in this database. These images are taken under different lighting condition (left-light, center-light, right-light), and different facial expression (normal, happy, sad, sleepy, surprised, and wink), and with/without glasses. We run each set of experiments for 10 times, and the averaged results are used to evaluate the performance of the proposed algorithm. The experiments are implemented on a Pentium 3.0 GHz computer with 512MB RAM and programmed in the MATLAB platform. The procedural parameters are chosen with cross-validation method. The experiments are implemented to testify the feasibility of improving the NDA with kernel trick under the same conditions. We choose three sets of parameters, i.e.,$k_1 = 2$,$k_2 = 2$,$k_1 = 2$, $k_2 = 3$ and $k_1 = 3$,$k_1 = 3$, for NDA and NKDA under the different number of features for face recognition. The results are shown in Table 1, Table 2 and Table 3. We compare their recognition performance on the same feature dimensionality from 5 to 40.

As shown in Table 1, 2 and 3, under the same conditions, KNFC outperforms NDA on the recognition performance, which demonstrates that the kernel method is feasible to improve NDA. It is worth to emphasize that the recognition accuracy of KNFC is higher about 20% than NDA with 5 features. KNFC achieves the higher recognition accuracy under the same feature dimension.

Table 1. Performance on ORL database($k_1 = 2, k_2 = 2$)(%)

Feature dimension	5	10	15	20	25	30	35	40
KNFC	70.70	87.80	91.40	93.35	93.60	94.25	94.35	94.40
Linear Method(NDA)	52.05	78.75	85.65	88.80	90.20	91.80	92.55	93.15

Table 2. Performance on ORL database($k_1 = 2, k_2 = 3$)(%)

Feature dimension	5	10	15	20	25	30	35	40
KNFC	70.85	87.50	91.25	93.30	93.45	94.10	94.35	94.45
Linear Method(NDA)	52.00	78.90	85.60	88.90	90.15	91.80	92.55	93.25

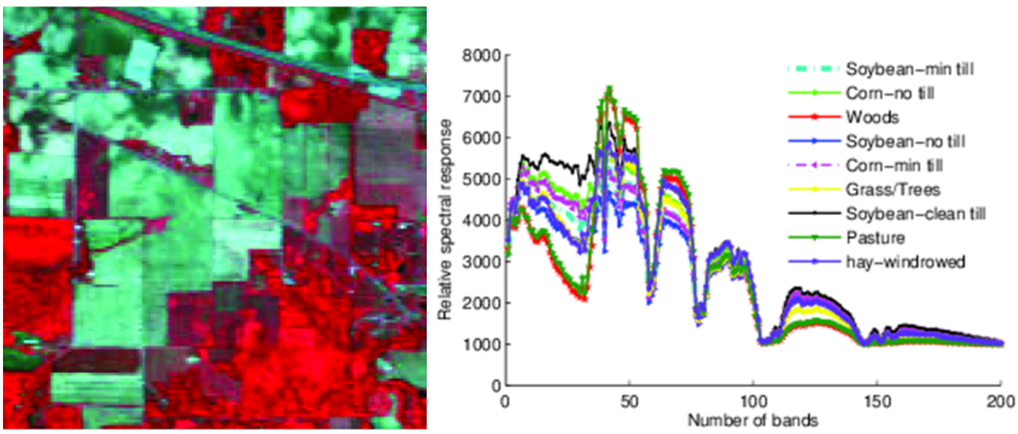
Table 3. Performance on ORL database($k_1 = 3, k_2 = 3$)(%)

Feature dimension	5	10	15	20	25	30	35	40
KNFC	70.85	86.85	91.70	93.35	93.95	94.45	94.75	94.90
Linear Method(NDA)	53.10	78.95	86.30	89.10	90.15	92.00	92.80	92.95

On the YALE database, the experimental results are performance on YALE databases including NDA and KNFC methods. We select the dimensionality through cross validation method under the different parameter setting, as shown in Table 4.

Table 4. Performance on YALE database(%)

Parameter settings	$k_1 = 3, k_2 = 3$	$k_1 = 3, k_2 = 3$	$k_1 = 3, k_2 = 3$
KNFC	89.23	92.12	92.56
Linear Method(NDA)	87.82	89.23	90.11



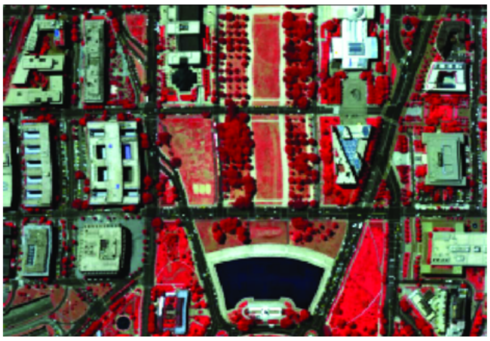
(a) Three band false color composite (b) Spectral signatures

Figure 2. Indian Pines data

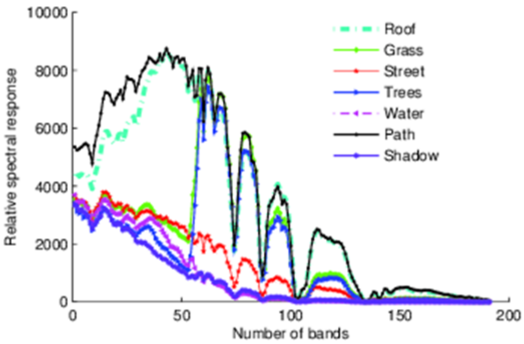
Secondly, we implement the experiments on Indian Pines and Washington, D.C. Mall databases. 1) Indian Pines data: the first test set to be used was the well-known Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image scene, which was captured over the agricultural region of Northwestern Indiana in June 1992, with spectral resolution of 224 bands covering the $0.4 - 2.5\mu m$ range and spatial resolution of 20m per pixel. After removing the noisy and water-vapor absorption bands, 200 bands reserved for experiments.

Although the whole scene consists of 145×145 pixels with 16 classes of interest, ranging the size from 20 to 2468 pixels, only 9 classes with high number of samples are selected. Some examples are shown in Figure 2.

2) D.C. Mall data: the second test set was acquired by the airborne hyperspectral digital imagery collection experiment (HYDICE) sensor over a Mall in Washington D.C. on August 23, 1995. The whole urban image size is 1280 307 pixels with the spatial resolution of 1.5m by pixel and 210 spectral bands in the 0.4-2.4 region. Several undesirable bands influenced by the atmospheric absorption are discarded, leaving 191 bands for experiments. From the original image, we crop a [870-1080] [1-307] subset with a size of 211 307, which composed of 7 classes of land-covers (i.e. roof, grass, street, trees water, path and shadow).Some examples are shown in Figure 3.



(a) Three band false color composite



(b) Spectral signatures

FIGURE 3. D.C. Mall data

(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 4. The former 6 principle components of the Indian Pines data.(a)-(f) are the first to sixth principle components, respectively.

On the hyperspectral image databases, all the experimental results are the mean accuracies over 10 repetitions, which helps to compare different methods in a fair and reasonable way. As to the feature reduction, the PCA results of both data sets are shown in Figure 4 and Figure 5, respectively, which contain the former 6 (for the Indian Pines data) or 4 (for the D.C. Mall data) principle components. The experiments are shown in Table 5 and Table 6. On the same conditions, the proposed kernel-based classifier performs better than the linear classifier.

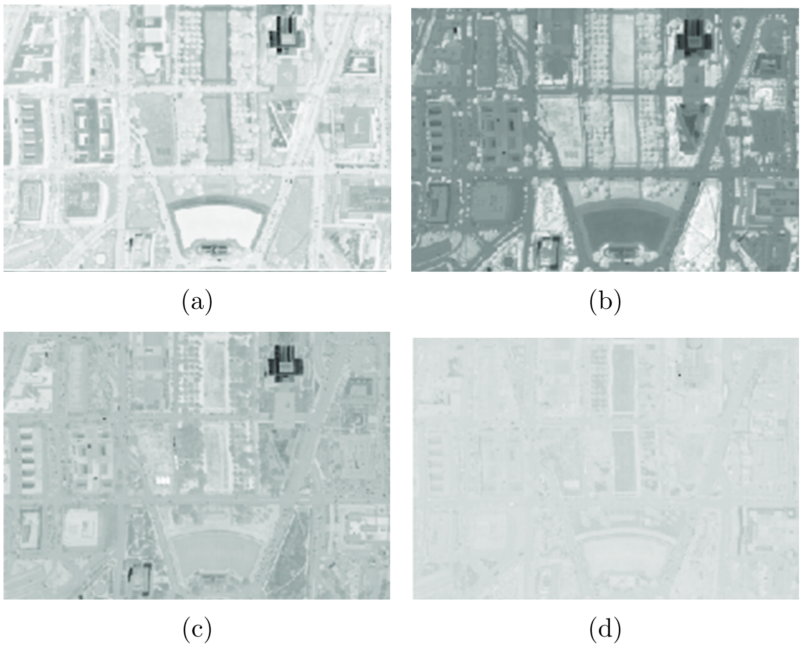


FIGURE 5. The former 4 principle components of the D.C. Mall data. (a)-(d) are the first to fourth principle components, respectively.

TABLE 5. Results of different methods for the Indian Pines data(%)

Classes	1	2	3	4	5	6	7	8	9
NDA	74.78	70.12	96.23	73.34	77.23	94.12	76.23	86.12	96.23
KNFC	77.21	72.34	98.23	75.54	79.34	96.04	78.23	88.34	98.45

TABLE 6. Results of different methods for the D.C.Mall data(%)

Classes	roof	grass	street	trees	water	path	shadow
NDA	80.12	92.34	91.23	92.24	96.23	96.12	91.23
KNFC	82.34	94.78	93.56	94.78	98.87	98.43	93.87

4. **Conclusions.** In this paper, we present a novel Kernel-Based Nonparametric Fisher Classifier (KNFC) for hyperspectral remote sensing imagery. A comprehensive theoretical analysis on Nonparametric Discriminant Analysis (NDA) is implemented, and NDA has its limitations on extracting the nonlinear features owing to the high nonlinear and complex distribution of the hyperspectral imagery data. the kernel trick to NDA to develop Kernel-Based Nonparametric Fisher Classifier to enhance its ability on hyperspectral Imagery Sensing data. KNFC also has its ability on image recognition, video retrieval and processing. Besides the good performance on recognition accuracy, the time consuming is one problem endured by NKDA, so how to improve the efficiency of NKDA is the future work.

REFERENCES

[1] Y. Xie, L. Setia and H. Burkhardt, Face Image Retrieval Based on Concentric Circular Fourier-Zernike Descriptors, International Journal of Innovative Computing, *Information and Control*, vol. 4, no. 6, pp. 1433-1444, 2008.

[2] H. Ryu, V. Dinh and M. Kim, Real-time Multi-view Face Tracking Combining Learning Based Classifier and Template Matching, *ICIC Express Letters*, vol. 1, no. 2, pp. 185-189, 2007.

[3] J. B. Li, J. S. Pan and S. C. Chu, Kernel Class-wise Locality Preserving Projectio, *Information Sciences*, vol. 178, no. 7, pp. 1825-1835, 2008.

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

[5] J. B. Li and J. S. Pan, A novel pose and illumination robust face recognition with a single training image per person algorithm, *Chinese Optics Letters*, vol. 6, no. 4, pp. 255-257, 2008.

[6] J. B. Li, S. C. Chu, J. S. Pan and J. H. Ho, Adaptive Data-Dependent Matrix Norm Based Gaussian Kernel for Facial Feature Extraction, International Journal on Innovative Computing, *Information and Control*, vol. 3, no. 5, pp. 1263-1272, 2007.

[7] C. Zhou, X. Wei, Q. Zhang and B. Xiao, Image Reconstruction for Face Recognition Based on Fast ICA, *International Journal of Innovative Computing, Information and Control*, vol. 4, no. 7, pp. 1723-1732, 2008.

[8] K. Fukunaga andJ. Mantock, Nonparametric discriminant analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 6, pp. 671-678, 1983.

[9] Z. Li, Dahua Lin and Xiaoou Tang. Nonparametric Discriminant Analysis for Face Recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 755-761, 2009.

[10] F. Samaria, A., Parameterisation of a Stochastic Model for Human Face Identification, *Proc. of 2nd IEEE Workshop on Applications of Computer Vision*, pp. 138-142, 2008.

[11] J. Huang, P. C. Yuen, W. S. Chen and J. H. Lai, Kernel Subspace LDA with Optimized Kernel Parameters on Face Recognition, *Proc. of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[12] J. S. Pan, J. B. Li and Z. M. Lu, Adaptive quasiconformal kernel discriminant analysis,*Neurocomputing* , vol. 71, no. 13-15, pp. 2754-2760, 2008.

[13] W. S. Chen, P. C. Yuen, J. Huang and D. Q. Dai, Kernel Machine-Based One-Parameter Regularized Fisher Discriminant Method for Face Recognition, *IEEE Trans. Systems, Man and Cybernetics-Part B: Cybernetics*, vol. 35, no. 4, pp. 658-669, 2005.

[14] H. Xiong, M. N. Swamy and M. O. Ahmad, Optimizing the kernel in the empirical feature space,*IEEE Trans. on Neural Networks*, vol. 16, no. 2, pp. 460-474, 2005.

[15] S. Amari and S. Wu, Improving support vector machine classifiers by modifying kernel functions, *Neural Networks*, vol. 12, no. 6, pp. 783-789, 1999.

[16] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,*IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[17] B. Scholkopf, A. Smola and K.R. Muller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.

[18] D. C. Tao, X. O. Tang, X. L. Li and X.D. Wu, Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1088-1099, 2006.

[19] Y. Liang, C. Li, W. Gong and Y. Pan, Uncorrelated linear discriminant analysis based on weighted pairwise Fisher criterion, *Pattern Recognition*, vol. 40 , no. 12, pp. 3606-3615, 2007.

[20] Y. J. Zheng, J. Yang, J. Y. Yang and X. J. Wu, A reformative kernel Fisher discriminant algorithm and its application to face recognition, *Neurocomputing*, vol. 69, no.13-15, pp. 1806-1810, 2006.

[21] B. Ma, H. Y. Qu and H. S. Wong, Kernel clustering-based discriminant analysis, *Pattern Recognition*, vol. 40, no. 1, pp. 324-327, 2007.