Hyperspectral Image Classification Using a General NFLE Transformation with Kernelization and Fuzzification

Ying-Nong Chen
National Central University
Department of Computer Science and Information Engineering
Taiwan, Taoyuan City
yingnong1218@gmail.com

Yu-Chen Wang
National Central University
Department of Computer Science and Information Engineering
Taiwan, Taoyuan City
m09502062@chu.edu.tw

Chin-Chuan Han
National United University
Department of Computer Science and Information Engineering
Taiwan, Miaoli City
cchan@csie.ncu.edu.tw

Kuo-Chin Fan
National Central University
Department of Computer Science and Information Engineering
Taiwan, Taoyuan City
kcfan@csie.ncu.edu.tw

ABSTRACT

Nearest feature line (NFL) embedding (NFLE) is an eigenspace transformation algorithm based on the NFL strategy. Based on this strategy, the NFLE algorithm generates a low dimensional space in which the local structures of samples in the original high dimensional space are preserved. Though NFLE has successfully demonstrated its discriminative capability, the non-linear manifold structure cannot be structured more efficiently by linear scatters using the linear NFLE method. To address this, a general NFLE transformation, called fuzzy/kernel NFLE, is proposed for feature extraction in which kernelization and fuzzification are simultaneously considered. In the proposed scheme, samples are projected into a kernel space and assigned larger weights based on that of their neighbors according to their neighbors. In that way, not only is the non-linear manifold structure preserved, but also are the discriminative powers of classifiers increased. The proposed method is compared with various state-of-the-art methods to evaluate the performance by several benchmark data sets. From the experimental results, the proposed FKNFLE outperformed the other, more conventional, methods.

Keywords

Hyperspectral image classification, manifold learning, nearest feature line embedding, kernelization, fuzzification

1. INTRODUCTION

Dimensionality reduction (DR) in hyperspectral image (HSI) classification is a critical issue during data analysis because most multispectral, hyperspectral, and ultraspectral images generate high-dimensional spectral images with abundant spectral bands and data. However, it is challenging to classify these spectral data because vast amount of samples have to be collected for training beforehand. Besides, the spectral properties of land covers are too similar to clearly separate them out. Hence, an effective DR is an essential step to extract the salient features for classification. Recently, a number of DR methods have been proposed that can be classified into three categories: linear analysis, manifold learning, and kernelization. Those using linear analysis try to model the linear variation of samples and find a transformation to maximize the global scatter matrix, e.g. principal component analysis (PCA), linear discriminant analysis (LDA), and discriminant common vectors (DCV). Sample scatters are represented in the global Euclidean structure in these methods. They work well for DR or classification if samples are linearly separated or are distributed in a Gaussian function. However, when samples are distributed in a manifold structure, the local structure of a sample in a high dimensional space is not apparent when using global measurement. In addition, the classification performance in the case of linear analysis methods would deteriorate when the decision boundaries are predominantly nonlinear. Manifold learning methods are proposed to reveal the local structure of samples. He et al. propose the locality preserving projection (LPP) method to preserve the local structure of training samples for face recognition. Since LPP presents sample scatter using the relationship between neighbors, the local manifold structure is preserved and the performance is more effective than in the case of the linear analysis methods. Similar to LPP, Tu et al. propose a Laplacian eigenmap (LE) method for land cover classification using polarimetric synthetic aperture radar data. The LE algorithm reduces the dimensions of features from a high-dimensional polarimetric manifold space to an intrinsic low-dimensional manifold space. Wang and He investigated the LPP for DR in HSI classification. Kim et al. utilized the

locally linear embedding (LLE) method to reduce the dimensionality of HSIs. Li et al. propose the local Fisher discriminant analysis (LFDA) method which integrates the properties of LDA and LPP to reduce the dimensionality of HSI data. Luo et al. propose a discriminative and supervised neighborhood preserving embedding (NPE) method for feature extraction in HSI classification. These manifold learning methods all preserve the local structure of samples and improve on the performance of conventional linear analysis methods. However, the applicability of linear manifold learning is limited to noises. Generally, the discriminative salient features of training samples are extracted using certain evaluation processes. An appropriate kernel function could improve the performance for the given method[13]. The kernelization approaches have been proposed for improving the performance of HSI classification. Boots and Gordon introduced a kernelization method to alleviate the limitation of manifold learning. Scholkopf et al. propose a kernel PCA (KPCA) method for nonlinear DR. KPCA generates a high-dimensional Hilbert space to extract the non-linear structure that is missed by PCA. Furthermore, Lin et al. propose a general framework for multiple kernel learning during DR. They unify the multiple kernel representation, and the multiple feature representations of data are consequently revealed in a low dimension. On the other hand, a composite kernel scheme, a linear combination of multiple kernels, extracts both spectral and spatial data. Chen et al. present a sparse representation of kernels for HSI classification. A query sample is represented via all training samples in an induced kernel space. Moreover, pixels within a local neighborhood are also represented by the combination of training samples. In the previous works, the nearest feature line (NFL) strategy was embedded into the linear transformation for dimension reduction on face recognition and HSI classification. However, the nonlinear and non-Euclidean structures are not efficiently extracted using the linear transformation. Fuzzification and kernelization are two efficient tools for enhancement in nonlinear spaces. The fuzzy methodology is further adopted in previous work. In this study, a general NFLE transformation, called fuzzy-kernel NFLE, is extended for feature extraction in which kernelization and fuzzification are simultaneously considered. In addition, more experimental analysis was conducted in this study. Three benchmark data sets were evaluated in this work. The proposed method was compared with state-of-the-art algorithms for performance evaluation.

2. Related Works

In this study, three approaches, nearest feature line embedding (NFLE), kernelization , and fuzzy *k* nearest neighbor (FKNN)[20], were considered to reduce the feature dimensions for HSI classification. Before the proposed methods, brief reviews of NFLE and kernelization methods are presented in the following: Given N *d*-dimensional training samples $X=[x_1,x_2...x_N]\in R^{d\times N}$ consisting of N_C land-cover classes $C_1,C_2,\ldots,C_{N_C}$. The new samples in a low-dimensional space were obtained by the linear projection $y_i=w^T x_i$, where w is a found linear projection matrix for DR. NFLE is a linear transformation for DR. The sample scatters are represented in a Laplacian matrix form by using the point-to-line strategy which originated from the nearest linear combination (NLC) approach. The objective function is defined and minimized as follows:

$$\begin{aligned} O &= \sum_i \left(\sum_{i\neq m\neq n} \| y_i - L_{m,n}(y_i) \|^2 l_{m,n}(y_i) \right) \\ &= \sum_i \left\| y_i - \sum_j M_{i,j} y_i \right\|^2 \\ &= tr\left(Y(I-M)^T (I-M) Y \right) \\ &= tr\left(w^T X (D-W) w^T X \right) \\ &= tr\left(w^T X L X^T w \right). \end{aligned} \quad (1)$$

Here, point $L_{m,n}(y_i)$ is a projection point on line $L_{m,n}$ for point y_i, and weight $l_{m,n}(\mathbf{y}_i)$ (being 1 or 0) represents the connectivity relationship from point y_i to a feature line $L_{m,n}$ that passes through two points y_m and y_n. The projection point $L_{m,n}(y_i)$ is represented as a linear combination of points y_m and y_n : $L_{m,n}(y_i)=y_m+t_{m,n}(y_n-y_m)$ in which $i\neq m\neq n$, and $t_{m,n}=(y_i-y_m)^T(y_m-y_n)/(y_m-y_n)^T(y_m-y_n)$. Using simple algebra operations, the discriminant vector from point y_i to the projection point $L_{m,n}(y_i)$ can be represented as $y_i-\sum_j M_{i,j}y_j$, in which two values in the *i*th row in matrix M are set as $M_{i,m}=t_{n,m}$, $M_{i,n}=t_{m,n}$, and $t_{n,m}+t_{m,n}=1$, when weight $l_{m,n}(\mathbf{y}_i)=1$. The other values in the *i*th row are set as zero, if

$j \neq m \neq n$. The mean squared distance in Eq. (1) for all training points to their NFLs is next obtained as $tr\left(w^T XLX^T w\right)$, in which $L = D - W$, and matrix D is a matrix of the column sums of the similarity matrix W. From the consequences of Yan *et al.* [22], matrix W is defined as $W_{i,j} = \left(M + M^T - M^T M\right)_{i,j}$ when $i \neq j$, and zero otherwise; $\sum_j M_{i,j} = 1$. Matrix L in Eq. (1) is represented as a Laplacian matrix. For more details, refer to [18, 19]. Consider the class labels in supervised classification, two parameters K_1 and K_2 are manually determined in calculating the within-class scatter S_w and the between-class scatter S_b, respectively:

$$\mathbf{S}_w = \sum_{k=1}^{N_C}\left(\sum_{x_i \in C_k}\sum_{L_{m,n} \in F_{K_1}(x_i, C_k)} \left(x_i - L_{m,n}(x_i)\right)\left(x_i - L_{m,n}(x_i)\right)^T\right) \quad (2)$$

$$\mathbf{S}_b = \sum_{k=1}^{N_C}\left(\sum_{x_i \in C_k}\sum_{l=1, l\neq k}^{N_C}\sum_{L_{m,n} \in F_{K_2}(x_i, C_l)} \left(x_i - L_{m,n}(x_i)\right)\left(x_i - L_{m,n}(x_i)\right)\right) \quad (3)$$

.

$F_{K_1}\left(x_i, C_k\right)$ indicates the set of K_1 NFLs within the same class, C_k, of point x_i, i.e. $l_{m,n}(y_i) = 1$, and $F_{K_2}\left(x_i, C_l\right)$ is a set of K_2 NFLs belonging to the different classes of point x_i. The Fisher criterion $tr\left(\mathrm{S}_b / \mathrm{S}_w\right)$ is then maximized to find the projection matrix w, which is composed of the eigenvectors with the corresponding largest eigenvalues. A new sample in the low-dimensional space can be obtained by the linear projection $y = w^T x$, and the NN (one-NN) matching rule is applied for template matching. In kernel LDA, consider the nonlinear mapping function from a space X to a Hilbert space H, $\phi : x \in X \rightarrow \phi(x) \in H$, the within-class and between-class scatter in space H are calculated as

$$\mathbf{S}_w^{\phi} = \sum_{k=1}^{N_C}\left(\sum_{x_i \in C_k}\left(\phi(x_i) - \bar{\phi}_k\right)\left(\phi(x_i) - \bar{\phi}_k\right)^T\right), \quad (4)$$

and

$$\mathbf{S}_b^{\phi} = \sum_{k=1}^{N_C}\left(\bar{\phi}_k - \bar{\phi}\right)\left(\bar{\phi}_k - \bar{\phi}\right)^T . \quad (5)$$

Here, $\bar{\phi}_k = \frac{1}{n_k}\sum_{i=1}^{n_k}\phi(x_i)$ and $\bar{\phi} = \frac{1}{N}\sum_{i=1}^{N}\phi(x_i)$ represent the class mean and the population mean in space H, respectively. To generalize LDA to the nonlinear case, the dot product trick is exclusively used. The expression of dot product on the Hilbert space H is given by the following kernel function: $k(x_i, x_j) = k_{i,j} = \phi^T(x_i)\phi(x_j)$. Let the symmetric matrix K of N by N be a matrix composed of dot product in feature space H, i.e. $K(x_i, x_j) = \left\langle \phi(x_i) \cdot \phi(x_j) \right\rangle = \left(k_{i,j}\right)$ and, $i, j = 1, 2, ..., N$. The kernel operator K makes the possibility: the construction of the linear separating function in space H is equivalent to that of the nonlinear separating function in space X. Kernel LDA also maximizes the between-class scatter and minimizes the within-class scatter, i.e. $\max\left(S_b^{\phi} / S_w^{\phi}\right)$. This maximization is equivalent to the following eigenvector resolution: $\lambda S_w^{\phi} w = S_b^{\phi} w$. There is a set of coefficients α for $w = \sum_{i=1}^{N}\alpha_i \phi(x_i)$ such that the largest eigenvalue gives the maximum of the scatter quotient $\lambda = w^T S_b^{\phi} w / w^T S_w^{\phi} w$.

3. Fuzzy Kernel Nearest Feature Line Embedding

According to the aforementioned surveys, a training DR scheme effectively extracts the discriminant features from the non-Euclidean and non-linear space. To this end, fuzzy kernel nearest feature line embedding (FKNFLE) is proposed for HSI classification. The idea of FKNFLE is to incorporate the fuzziness and kernelization into the manifold learning method. The kernel function not only generates a non-linear feature space for well discriminant analysis, but also increases the robustness to noise during the training phase. Manifold learning methods preserve the local structure of samples in the Hilbert space. On the other hand, the fuzzy *k*-nearest neighbor (FKNN) method extracts the non-Euclidean structures of training samples for enhancing discriminative capability.

NFLE has successfully been applied in HSI classification. Noise variations and high degree non-linear data distributions limit the performance of manifold learning. A kernel trick is used to alleviate this problem. The details of FKNFLE are introduced in the following: Let $\phi : x \in X \rightarrow \phi(x) \in H$ be a nonlinear mapping from a low dimensional space to a high-dimensional Hilbert space H. The mean squared distance for all training points to their NFLs in the Hilbert space is written as follows:

$$\sum_i \left\| \phi(y_i) - L_{m,n}(\phi(y_i)) \right\|^{(2)}$$
$$= \sum_i \left\| \phi(y_i) - \sum_j M_{i,j}\phi(y_j) \right\|^2$$
$$= tr\left(\phi^T(Y)(I-M)^T(I-M)\phi(Y)\right)$$
$$= tr\left(\phi^T(Y)(D-W)\phi(Y)\right)$$
$$= tr\left(w^T\phi(X)L\phi^T(X)w\right). \quad (6)$$

Then, the object function in Eq. (6) is minimized and expressed as a Laplacian matrix. The eigenvector problem of kernel NFLE in the Hilbert space is expressed as:

$$\left[\phi(X)L\phi^T(X)\right]w = \lambda\left[\phi(X)D\phi^T(X)\right]w. \quad (7)$$

To extend NFLE to its kernel version, the implicit feature vector, $\phi(x)$, does not need to be obtained explicitly. The dot product expression of two samples is exclusively applied in the Hilbert space with a kernel function as follows: $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. The eigenvectors of Eq. (7) are represented by the linear combinations of $\phi(x_1)$, $\phi(x_2)$, $\cdots$, $\phi(x_N)$. The coefficient α_i is $w = \sum_{i=1}^{N} \alpha_i \phi(x_i) = \phi(X)\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T \in R^N$. Then, the eigenvector problem is as follows:

$$KLK\alpha = \lambda KDK\alpha. \quad (8)$$

Let the coefficient vectors, $\alpha^1, \alpha^2, \ldots, \alpha^N$, be the solutions of Eq. (8) in a column format. Given a testing point, z, the projections onto the eigenvectors, w^k, are obtained as follows:

$$\left(w^k \cdot \phi(z)\right) = \sum_{i=1}^{N} \alpha_i^k \langle \phi(z), \phi(x_i) \rangle = \sum_{i=1}^{N} \alpha_i^k K(z, x_i), \quad (9)$$

where α_i^k is the i^{th} element of the coefficient vector, α^k. The kernel function RBF (radial basis function) is used in this study. Thus, the within-class and between-class scatters in a kernel space are defined as follows:

$$\mathbf{s}_w^{\phi} = \sum_{k=1}^{N_C} \left(\sum_{\phi(x_i) \in C_k} \sum_{L_{m,n} \in F_{K_1}(\phi(x_i), C_k)} \left(\phi(x_i) - L_{m,n}(\phi(x_i))\right)\left(\phi(x_i) - L_{m,n}(\phi(x_i))\right)^T \right) \quad (10)$$

$$\mathbf{s}_b^{\phi} = \sum_{k=1}^{N_C} \left(\sum_{\phi(x_i) \in C_k} \sum_{l=1, l \neq k}^{N_C} \sum_{L_{m,n} \in F_{K_2}(\phi(x_i), C_l)} \left(\phi(x_i) - L_{m,n}(\phi(x_i))\right)\left(\phi(x_i) - L_{m,n}(\phi(x_i))\right)^T \right). \quad (11)$$

The kernelized manifold learning could preserve the non-linear local structure in a Hilbert space. The distances in the NFLE approach are calculated by the Euclidean distance-based measurement. On the other hand, the non-Euclidean structure of training samples can be further extracted by fuzzification. The FKNN algorithm[20] enhances the discriminant power among samples by assigning the higher membership grades to the samples whose neighbors are with the same class. By doing so, the non-Euclidean structures are extracted, and the discriminative power of samples can be enhanced. The idea of FKNFLE using the fuzzification trick is described in the following.

Consider N samples in the reduced space $Y = [y_1, y_2 ..., y_N]$ and their corresponding fuzzy membership grades, $\pi(y_i)$, for each sample, y_i. The objective function is re-defined as follows:

$$O = \sum_i \pi(y_i) \left(\sum_{i \neq m \neq n} \left\| y_i - L_{m,n}(y_i) \right\|^2 l_{m,n}(y_i) \right)$$
$$= \sum_i \pi(y_i) \left\| y_i - \sum_j M_{i,j} y_j \right\|^2$$
$$= tr\left(Y^T(FEI - FEM)^T(FEI - FEM)Y\right)$$
$$= tr\left(Y^T(FED - FEW)Y\right)$$
$$= tr\left(Y^T(D_{fuzzy} - W_{fuzzy})Y\right)$$
$$= tr\left(w^T X L_{fuzzy} X^T w\right) \quad (10)$$

Here, each sample is assigned a fuzzy grade, $\pi(y_i)$. Element $M_{i,j}$ denotes the connectivity relationship between point y_i and line $L_{m,n}$ which is the same as that in Eq. (1). Two non-zero terms, $M_{i,n} = t_{m,n}$ and $M_{i,m} = t_{n,m}$, are set, and $\sum_j M_{i,j} = 1$. Using simple algebra operations, the objective function with fuzzification is represented in a Laplacian matrix form in which the fuzzy terms, $\pi(y_i)$, constitute the column vector, F, with size $N \times 1$, and E is a row vector of all those with size $1 \times N$.

Similarly, given N samples $\phi(\mathrm{X}) = \{\phi(x_1), \phi(x_2), \ldots, \phi(x_N)\}$ in a Hilbert space, the membership grade of a specified sample, $\phi(x_i)$, and its K_3 neighbors is designed in the following equation for computing the within-class scatter:

$$\pi(x_i) = \begin{cases} 0.51 + (0.49 * (q_i / K_3)), & \text{if } q_i \geq \theta_{within}; \\ 0.49 * (q_i / K_3) & \text{otherwise.} \end{cases} \quad (11)$$

Here, value q_i is the number of samples whose labels are the same label of $\phi(x_i)$ among K_3 nearest neighbors, and θ_{within} is a manual threshold. If $q_i = K_3$, then $\pi(x_i)$ returns to 1, i.e. all neighbors are in the same class. Adding the fuzzy term $\pi(x_i)$, the within-class scatter matrix becomes:

$$\mathbf{S}_w^{\phi F} = \sum_{k=1}^{N_C}\left(\sum_{\phi(x_i)\in C_k} \pi(x_i) * \sum_{L_{m,n}\in F_{K_1}(\phi(x_i),C_k)} (\phi(x_i) - L_{m,n}(\phi(x_i)))(\phi(x_i) - L_{m,n}(\phi(x_i)))^T \right) \quad (14)$$

Similarly, a fuzzy term $\lambda(x_i)$ is also adopted to evaluate the membership grade of $\phi(x_i)$ and its neighbors during the computation of between-class scatter as follows:

$$\lambda(x_i) = \begin{cases} 0.51 + (0.49 * (p_i / K_4)) & \text{if } p_i \geq \theta_{between}; \\ 0.49 * (p_i / K_4) & \text{otherwise.} \end{cases} \quad (12)$$

Here, value p_i is the number of samples with labels different from $\phi(x_i)$ among K_4 nearest neighbors, and $\theta_{between}$ is a given threshold. If $p_i = K_4$, term $\lambda(x_i)$ is returned to 1. That means all neighbors have labels different from $\phi(x_i)$. The fuzzy term $\lambda(x_i)$ is added into the between-class scatter matrix to generate a new one as:

$$\mathbf{S}_b^{\phi F} = \sum_{k=1}^{N_C}\left(\sum_{\phi(x_i)\in C_k} \lambda(x_i) * \sum_{l=1, l\neq k}^{N_C} \sum_{L_{m,n}\in F_{K_2}(\phi(x_i),C_l)} (\phi(x_i) - L_{m,n}(\phi(x_i)))(\phi(x_i) - L_{m,n}(\phi(x_i)))^T \right) \quad (16)$$

Hence, kernelization and fuzzification are simultaneously integrated into the NFLE transformation for feature extraction. In this paper, a general format for the NFLE learning method using kernelization and fuzzification is proposed to be used for DR. The advantages of the proposed method are threefold: the kernelization strategy generates a non-linear feature space for the discriminant analysis and increases the robustness to noise for manifold learning; the kernelized manifold learning preserves the local manifold structure in a Hilbert space as well as the locality of the manifold structure in the reduced low dimensional space; non-Euclidean structures are extracted for improving discriminative abilities using the FKNN strategy.

4. Experimental Results

In this section, the experimental results are discussed to demonstrate the effectiveness of the proposed method for HSI classification. Three HSI benchmarks are given for evaluation. The first data set, Indian Pines Site (IPS) image, was generated from AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) which was captured by the Jet Propulsion Laboratory and NASA/Ames in 1992. The IPS image was captured from 6 miles in the western area of the *Northwest Tippecanoe County* (NTC). A false color IR image of dataset IPS is shown in Fig. 2(a). Although dataset IPS contained 220 bands and 16 land-cover classes, only ten classes were used in the experiments: *Corn-no till*(1428), *Corn-min till*(830), *Grass/Pasture*(478), *Grass/Trees*(730), *Hay-windrowed*(483), *Soybeans-no till*(972), *Soybeans-min till*(2455), *Soybeans-clean till*(593), *Woods*(1265), and *Bldg-Grass-Tree-Drives*(386). The numbers in parentheses are the collected pixel numbers in dataset IPS. The ground truths of 9,620 pixels were manually labeled for training and testing. Nine hundred training samples of ten classes were randomly chosen from 9,620 pixels, and the remaining samples were used for testing. The other two HSI data sets adopted in the experiments were obtained from the Reflective Optics System Imaging Spectrometer (ROSIS) instrument covering the City of Pavia, Italy. Two scenes, the university area and the Pavia city centre containing 103 and 102 data bands both with a spectral coverage from 0.43 to 0.86 um and a spatial resolution of 1.3m. The image sizes of these two areas were 610x340 and 1096x715 pixels, respectively. Figs. 2(b) and 2(c) show the false color IR image of these two data sets. Nine land-cover classes were available in each data set, and the samples in each data set were separated into two subsets, i.e. one training and one testing set. Given the Pavia University data set, ninety training samples per class were randomly collected for training, and the 8,046 remaining samples were tested for performance evaluation. Similarly, the numbers of training and testing samples used for the Pavia Centre data set were 810 and 9,529, respectively. The proposed methods, NFLE[18,19], KNFLE, FNFLE[24], and FKNFLE, were compared with two state-of-the-art algorithms, i.e. nearest regularized subspace (NRS) [23] and NRS-LFDA [23]. The parameter configurations for both algorithms NRS[1] and NRS-LFDA can be referred to in [23]. The gallery samples were randomly chosen for training the transformation matrix, and the query samples were matched with the gallery samples using the NN matching rule. Each algorithm was run 30 times to obtain the average rates. To obtain the appropriate reduced dimensions of FKNFLE, the available

[1] The source codes are available from the web site https://github.com/eric-tramel/NRSClassifier

training samples were chosen to evaluate the overall accuracy versus the reduced dimensions in the benchmark datasets. The proposed method was compared with various classification methods on computational time. All methods were implemented by MATALB codes on a personal computer with an i7 2.93-GHz CPU and 12.0 GB RAM. The comparisons of various algorithms on computational time were tabulated in Table I for the IPS, Pavia University, and Pavia City Centre datasets. Considering the training time, the proposed FKNFLE algorithm was generally faster than NRS and NRS-LFDA; 2 times and 15 times, respectively. Due to the fuzzification process, algorithms FKNFLE and FNFLE were slower than KNFLE and NFLE; 13 times and 15 times, respectively.

From Tables II to IV, the producer's accuracy, overall accuracy, kappa coefficients, and user's accuracy defined by the error matrices (or confusion matrices) [25] were calculated for performance evaluation. They are briefly defined in the following.

The user's accuracy and the producer's accuracy are two widely used measures for class accuracy. The user's accuracy is defined as the ratio of the number of correctly classified pixels in each class by the total pixel number classified in the same class. The user's accuracy is a measure of commission error, whereas the producer's accuracy measures the errors of omission and indicates the probability that certain samples of a given class on the ground are actually classified as such. The kappa coefficient, also called the kappa statistic, is defined to be a measure of the difference between the actual agreement and the changed agreement.

Table I: The training and testing time of various algorithms for the benchmark datasets (seconds).

Datasets	IPS		Pavia University		Pavia City Centre	
Algorithms	Training	Testing	Training	Testing	Training	Testing
	900	8720	810	8046	810	9529
NFLE-NN	10	18	9	16	9	20
KNFLE-NN	12	18	11	16	11	20
FNFLE-NN	155	18	140	16	140	20
FKNFLE-NN	156	18	141	16	141	20
NRS	326	326	294	300	294	351
LFDA-NRS	2331	327	2098	301	2098	352

Table II: The classification error matrix for data set IPS (in percentage).

Classes	Reference Data										User's Accuracy
	1	2	3	4	5	6	7	8	9	10	
1	79.20	3.43	0.28	0.35	0	5.46	9.73	1.54	0	0	79.20
2	5.90	81.81	0	0.12	0	1.33	6.39	4.34	0	0.12	81.81
3	0	0	97.49	1.46	0.21	0.42	0	0.21	0.42	0.84	97.49
4	0	0	0.27	96.30	0	0	0	0	0	3.42	96.30
5	0	0	0.42	0	99.58	0	0	0	0	0	99.58
6	5.14	0.21	0.10	0.41	0	88.89	4.42	0.72	0	0.10	88.89
7	10.59	5.58	0.29	0.33	0.04	9.78	69.98	3.30	0	0.12	69.98
8	1.35	4.05	1.52	0.34	0	1.69	1.85	88.53	0	0.67	88.53
9	0	0	3.32	0.16	0	0	0	0	90.83	5.69	90.83
10	0	0	3.89	5.70	0	0	0	0.26	10.88	79.27	79.27
Producer's Accuracy	77.51	86.04	90.62	91.57	99.75	82.63	75.76	89.51	88.94	87.85	

Kappa Coefficient: **0.821** Overall Accuracy: **83.34%**

Table III: The classification error matrix for data set Pavia University (in percentage).

Classes	Reference Data									User's Accuracy
	1	2	3	4	5	6	7	8	9	
1	90.18	3.15	0	0	0	3.24	1.35	1.26	0.81	90.18
2	2.31	92.50	0	2.31	0	1.85	0	1.01	0	92.50
3	0	0	90.07	2.38	1.58	0.99	2.97	0.99	0.99	90.07
4	0	1.23	2.84	90.24	1.42	1.42	1.51	1.32	0	90.24
5	0.63	1.13	0.75	1.26	91.91	0.63	1.64	0.88	1.13	91.91
6	1.10	1.19	1.38	1.56	1.19	92.54	0.55	0.46	0	92.54
7	0	1.12	0.51	0.61	2.24	0	93.25	1.22	1.02	93.25
8	0.47	1.42	0.95	1.42	2.38	1.90	0	90.76	0.66	90.76
9	1.14	0	2.15	2.01	0	2.29	0	2.15	90.22	90.22
Producer's Accuracy	94.10	90.92	91.30	88.65	91.25	88.25	92.08	90.71	95.14	

Kappa Coefficient: **0.910** Overall Accuracy: **91.31%**

Table IV: The classification error matrix for data set Pavia City Centre (in percentage).

Classes	Reference Data									User's Accuracy
	1	2	3	4	5	6	7	8	9	
1	98.61	0.17	0.51	0.34	0.34	0	0	0	0	98.61
2	1.04	97.47	0.43	0	0	0.34	0.17	0.52	0	97.47
3	0.59	0.82	96.23	0.69	0.99	0	0	0	0.69	96.23
4	0	0.56	0.66	96.68	0.37	0.47	0.66	0.56	0	96.68
5	0	0	0.43	0.34	97.73	0.26	0.34	0.34	0.52	97.73
6	0.35	0.26	0.61	0	0	98.15	0	0.26	0.35	98.15
7	0.35	0.26	0	0.35	0	0.44	98.23	0.35	0	98.23
8	0	0	0.37	0.30	0.37	0.52	0.45	97.43	0.52	97.43
9	0.39	0.59	0.79	0.29	0.29	0	0	0	97.60	97.60
Producer's Accuracy	97.32	97.34	96.20	97.67	97.64	97.97	98.38	97.96	97.91	

Kappa Coefficient: **0.971** Overall Accuracy: **97.59%**

5. SECTIONS

In this paper, a general NFLE transformation, FKNFLE, for HSI classification is proposed. Kernelization and fuzzification were both considered in NFLE in extracting non-linear and non-Euclidean structures. In addition, the locality of the manifold structure of samples was preserved. High-dimensional HSI data were reduced to low-dimensional features by the proposed FKNFLE transformation. Two state-of-the-art algorithms, NRS and NRS-LFDA, were compared with the proposed FKNFLE. Three land-cover benchmarks, IPS, Pavia University, and Pavia City Centre, were tested for performance evaluation. The experimental results demonstrated that FKNFLE outperformed the other algorithms.

6. REFERENCES

M. Turk and A.P. Pentland, “Face recognition using eigenfaces,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1991, pp. 586-591.

P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using

class specific linear projection," *IEEE Trans. Pattern Ana Anal. Mach. Intell*., vol. 19, no. 7, pp. 711-720, Jul. 1997.

H. Cevikalp, M. Neamtu, M. Wikes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Ana Anal. Mach. Intell*., vol. 27, no. 1, pp. 4-13, Jan. 2005.

S. Prasad and L. Mann Bruce, "Information fusion in kernel-induced spaces for robust subpixel hyperspectral ATR," *IEEE Trans. Geosci. Remote Sens. Lett*., vol. 6, no. 3, pp. 572-576, Jul. 2009.

X. He, S. Yan, Y. Ho, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Ana Anal. Mach. Intell*., vol. 27, no. 3, pp. 328-340, Mar. 2005.

S. T. Tu, J. Y. Chen, W. Yang, and H. Sun, "Laplacian eigenmaps-based polarimetric dimensionality reduction for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 170-179, Jan. 2011.

Z. Wang and B. He, "Locality preserving projections algorithm for hyperspectral image dimensionality reduction," in *Proc. 19th Int. Conf. Geoinf.*, Jun. 24-26, 2011, pp. 1-4.

D. H. Kim, and L. H. Finkel, "Hyperspectral image processing using locally linear embedding," in *Proc. 1st Int, IEEE EMBS Conf. Neural Eng*., Italy, Mar. 20-22, 2003, pp. 316-319.

W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geosci. Remote Sens. Letters*, vol. 8, no. 5, pp. 894-898, Sep. 2011.

W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

R. B. Luo, W. Z. Liao, and Y. G. Pi, "Discriminative supervised neighborhood preserving embedding feature extraction for hyperspectral-image classification," *Telkomnika*, vol. 10, no. 5, pp. 1051–1056, 2012.

B. Boots, and G. J. Gordon, "Two-manifold problems with applications to nonlinear system Identification," *Proc. 29th International Conference on Machine Learning*., United Kindom, Jun. 26- Jul. 1, 2012.

F. Odone, A. Barla, and A. Verri, "Building kernels from binary strings for image matching," *IEEE Trans. Image Processing*., vol. 14, no. 2, pp.169-180, Feb. 2005.

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, Jul. 1998.

Y. Y. Lin, T. L. Liu, and C. S. Fuh, "Multiple kernel learning for dimensionality reduction," *IEEE Trans. Pattern Analysis. Mach Intell*., vol. 33, no. 6, pp.1147-1160, Jun. 2011.

J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens*., vol. 51, no. 9, pp.4816-4829, Sep. 2013.

Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens*., vol. 51, no. 1, pp.217-231, Jan. 2013.

Y. N. Chen, C. C. Han, C. T. Wang, and K. C. Fan, "Face recognition using nearest feature space embedding," *IEEE Trans. Pattern Analysis. Mach Intell*., vol. 33, no. 6, pp.1073-1086, Jun. 2011.

Y. L. Chang, J. N. Liu, C. C. Han, and Y. N. Chen, "Hyperspectral image classification using nearest feature line embedding approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 278–287, Jan. 2014.

J. M. Keller, M. R. Gray, and J. A. Givens, Jr., "A fuzzy k-nearest neighbor algorithm," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 15, no. 4, pp.580-585, July/August. 1985.

S. Z. Li, "Face recognition based on nearest linear combinations," *Proc. Computer Vision and Pattern Recognition*, pp. 839-844, 1998.

S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a framework for dimensionality reduction," *IEEE Trans. Pattern Analysis. Mach Intell*., vol. 29, no. 1, pp.40-51, Jun. 2007.

W. Li, E. W. Tramel, S. *Prasad*, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014.

Y. N. Chen, C. C. Han, and K. C. Fan, "Use fuzzy nearest feature line embedding for hyperspectral image classification," *Proc. 4th International Conf. Earth Observations and Societal Impacts*, Taiwan, Jun. 22-24, 2014.

T. M. Lillesand and R. W. Kiefer, *Remote Sensing and Image Interpretation*. New York: Wiley, 2000.