Taylor & Francis
Taylor & Francis Group

A new kernel method for hyperspectral image feature extraction

Bin Zhao[a,b], Lianru Gao[a], Wenzhi Liao[c] and Bing Zhang[a,b]

[a]Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China; [b]University of Chinese Academy of Sciences, Ringgold Standard Institution, Beijing, China; [c]Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium

ABSTRACT

Hyperspectral image provides abundant spectral information for remote discrimination of subtle differences in ground covers. However, the increasing spectral dimensions, as well as the information redundancy, make the analysis and interpretation of hyperspectral images a challenge. Feature extraction is a very important step for hyperspectral image processing. Feature extraction methods aim at reducing the dimension of data, while preserving as much information as possible. Particularly, nonlinear feature extraction methods (e.g. kernel minimum noise fraction (KMNF) transformation) have been reported to benefit many applications of hyperspectral remote sensing, due to their good preservation of high-order structures of the original data. However, conventional KMNF or its extensions have some limitations on noise fraction estimation during the feature extraction, and this leads to poor performances for post-applications. This paper proposes a novel nonlinear feature extraction method for hyperspectral images. Instead of estimating noise fraction by the nearest neighborhood information (within a sliding window), the proposed method explores the use of image segmentation. The approach benefits both noise fraction estimation and information preservation, and enables a significant improvement for classification. Experimental results on two real hyperspectral images demonstrate the efficiency of the proposed method. Compared to conventional KMNF, the improvements of the method on two hyperspectral image classification are 8 and 11%. This nonlinear feature extraction method can be also applied to other disciplines where high-dimensional data analysis is required.

1. Introduction

Hyperspectral images can synchronously obtain spatial and spectral information of earth targets and have extremely high spectral resolution which makes it possible to precisely identify the target classes (Plaza et al. 2009; Li et al. 2012). However, hyperspectral data have relatively more bands. The number of bands can reach up to hundreds, which produce challenges for analysis and application of hyperspectral data. On the one hand, the large data size and high calculation complexity of hyperspectral images generate high requirements for hardware equipment. On the other hand, the high-dimensional features of hyperspectral data are unconducive to the precise classification of hyperspectral images. The higher the number of hyperspectral data bands is, the more the samples are required to achieve the desired classification accuracy. Yet, it is difficult to obtain labeled samples, which usually demands significant human and material resources (Hughes 1968).

In this study, dimensionality reduction is considered to be an effective method to solve these problems (Jia, Kuo and Crawford 2013; Benediktsson, Palmason and

Sveinsson 2005). Hyperspectral dimensionality reduction aims to simplify and optimize image features. It could effectively express high-dimensional data information, greatly reduce data size, and enable rapid and precise extraction of target information. Therefore, dimensionality reduction is important to the effective use of hyperspectral images. The common methods of dimensionality reduction are divided into two major categories: band selection and feature extraction. According to relevant evaluation indicators, band selection chooses band subsets from original data. The criteria of band selection are mainly based on information theory and spectral variance, such as methods based on information entropy proposed by Bajcsy and Groves (2004), mutual information to evaluate the information content of different bands proposed by Guo et al. (2006), and information covariance proposed by Ball et al. (2007). Feature extraction refers to the transformation of original data to the optimized subspace using mathematical transformation and thereby generating novel spectral data features. Traditional feature extractions include the linear and nonlinear feature extraction methods.

CONTACT Lianru Gao gaolr@radi.ac.cn

Principal component analysis (PCA) (Roger 1996) and minimum noise fraction (MNF) (Green et al. 1988) are the two commonly used linear dimensionality reduction methods. PCA uses information content as evaluation index of feature extraction and sorts the components by descending order of image information content after transformation. However, some studies found that, when noise distribution is uneven in each band, PCA transformation could not guarantee that the image components were sorted in accordance with quality (Green et al. 1988). Noise fraction (NF) is used as evaluation index of MNF. After MNF transformation, the components are sorted in accordance with image quality and are not affected by the noise distribution (Lee, Woodyatt and Berman 1990). However, MNF is unable to achieve reliable performances in real applications (Gao et al. 2013b). Some work has reported that the fundamental reason constraining the transformation of MNF was inaccuracy in the calculation of NF (Greco 2006; Liu et al. 2009; Gao et al. 2011; Zhao, Gao and Zhang, 2016; Gao et al. 2017). The information content for a particular hyperspectral image remains unchanged; the calculation accuracy of NF mainly depends on the noise estimation results. The original MNF transformation uses spatial neighborhood information to estimate noise. However, hyperspectral images usually have low spatial resolution with severely mixed pixels. There are relatively large errors related to the sole use of spatial information to calculate NF. Furthermore, the estimation results are unstable. Therefore, Gao et al. (2011; Gao, Du, et al., 2013; Gao, Zhang, et al., 2013; 2012) introduced the optimized minimum noise fraction (OMNF) method. This method adopts spectral and spatial decorrelation (SSDC) method to estimate noise by considering the high correlation between bands. This measure improves the results of the noise estimation and increases the accuracies of NF. Ultimately, the performance of original MNF transformation is improved. However, the above-mentioned methods can only extract linear features of hyperspectral data and are unable to find the nonlinear features.

Currently, the widely used nonlinear feature extraction methods are primarily kernel-based. The advantages of these methods mainly are that nonlinear features can be extracted effectively. Based on kernel function, these algorithms can transform the original data into a higher dimensional feature space. Thus, linear inseparable data in the original space can be separated in high-dimensional feature space. Nielsen (2011), Gómez-Chova and Nielsen (2011), Nielsen and Vestergaard (2012) proposed the kernel-based minimum noise fraction (KMNF) method based on this principle. This method has achieved great results in variation detection of remote sensing images with high spatial resolution, while failing to obtain the desired effect for hyperspectral images in practical application. From the theoretical analyses of KMNF, similar to MNF, KMNF also adopts spatial neighborhood information to estimate noise; thus, NF is also inaccurate. To overcome this problem, the optimized kernel minimum noise fraction (OKMNF) method was proposed (Zhao, Gao and Zhang 2016). This method also uses SSDC to estimate noise; thus, the performance of KMNF was improved greatly by introducing SSDC to estimate noise. However, in practical applications, it was found that the performance of OKMNF is still unstable for hyperspectral images with complex surface features. The main reason is: there is certain inadequacy in using SSDC to evaluate noise in hyperspectral images. In SSDC, the minimum region of multiple linear regression (MLR) is determined by empirical value, such as 6×6 pixels and 8×8 pixels (Gao et al. 2013a). This would involve other surface features inevitably to minimum regions, such as borders, when hyperspectral images have complex earth objects. Hence, the consistency of classes cannot be ensured within all divided areas, and this would produce errors while estimating noise and decrease the performance of feature extraction.

To overcome the above limitations, we introduce image segmentation to feature extraction algorithm. Image segmentation divides image into several subregions which do not intersect each other. The features are homogeneous within each subregion, while there are significant differences from different subregions. Traditional image segmentation methods can be divided into four categories: threshold-based image segmentation, region-based image segmentation, edge-based image segmentation, and theory-based image segmentation (Huang 1998; Kanungo et al. 2002). The key to threshold-based image segmentation is to determine a suitable threshold to divide images. Common threshold processing techniques include the adaptive threshold method, global threshold method, and optimal threshold method. The region-based image segmentation searches region directly. This method has two different search modes, including region growing and region splitting with merging. The edge-based image segmentation attempts to divide images through detecting marginal information of different regions. However, this method is extremely sensitive to noise and thus only suitable for low-noise images. Theory-based image segmentation combines theories and methods based on the development of various disciplines in new theories. In recent years, common theory-based image segmentation methods are primarily based on the mathematical morphology, fuzzy theory, gene coding, wavelet transform, machine learning, and clustering analysis. These image segmentation methods based on different specific theories have their own applications. In particular, image segmentation based on clustering analysis, by considering not only spatial information but also spectral information, is more suitable for the processing of hyperspectral images.
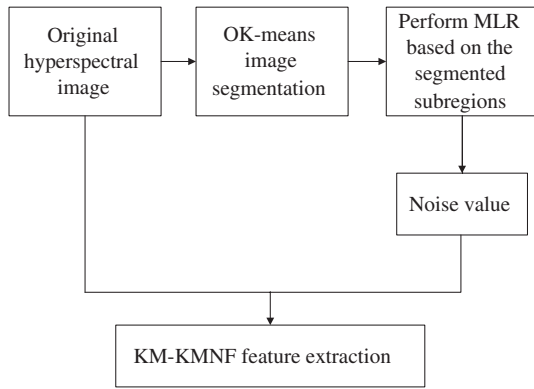
Figure 1. Flowchart of the KM-KMNF algorithm.

This paper proposes a new kernel method (KM-KMNF) for hyperspectral image feature extraction. Figure 1 shows the flowchart of the proposed KM-KMNF method.

The method exploits the homogeneous region generated by image segmentation as the minimum region for multiple linear regression. By integrating the spatial information of the homogeneous region with spectral decorrelation, KM-KMNF improves noise estimation for feature extraction, enabling better performances on feature extraction and the post-applications (e.g. classification).

The remainder of this paper is organized as follows: Section 2 introduces the KM-KMNF algorithm. Section 3 evaluates the classification performance of KM-KMNF in comparison with another established dimensionality reduction methods by using two real hyperspectral data sets. Finally, Section 4 concludes the paper.

2. KM-KMNF algorithm

2.1. Optimized K-means clustering method

Conventional K-means clustering is a rather effective unsupervised clustering method. With a certain similarity as a measure criterion, this method divides hyperspectral images into different subregions. The characteristics are homogeneous within the same subregion, while there are significant differences for different subregions. For a given hyperspectral image X containing n pixels and b bands, the algorithmic flow of K-means clustering is as follows: at the beginning, k pixels are randomly chosen as initial cluster centers from hyperspectral image. In addition, other pixels are assigned to the most similar clusters by computing their similarity with all cluster centers. The next step is to calculate the cluster centers of new clusters and to repeat the process until the standard measure function begins to converge. After K-means clustering, each cluster is a subregion of image segmentation. However, as showed in Figure 2(c) and Figure 4(c), pixels within the same subregions are distributed discretely. Thus, not all pixels within the same subregion can form a connected region. Moreover, homogeneity of the internal features cannot be guaranteed in the same

subregion. Therefore, the segmentation results of the conventional K-means clustering cannot meet the algorithm requirements. To solve these problems, we optimize the K-means (OK-means) clustering algorithm. Firstly, in the segment of hyperspectral images, to ensure the homogeneous features within the subregions, the number of subregions k should be far higher than that of normal segmentation. Then, to ensure the connection within each subregion, the search area of pixels, in which the similarity of pixel with its adjacent cluster center be measured, should be limited. Finally, pixels are assigned to the most similar clusters. By conducting OK-means image segmentation, all pixels within the same subregion can form a connected region, and homogeneity of the internal features can be guaranteed in each subregion. Therefore, OK-means image segmentation algorithm is more suitable for the processing of hyperspectral images. The segmentation results of the OK-means algorithm are shown in Figure 2(d) and Figure 4(d).

The flow of algorithm is as follows:

Step 1: Initialize K cluster centers. For a hyperspectral image X with n pixels and b bands, the ideal subregion size is $\sqrt{n/k} \times \sqrt{n/k}$. With $\sqrt{n/k}$ as step length, K pixels are selected as initial cluster centers ($c_1, c_2, ..., c_k$).

Step 2: Assign pixel x_i to the most similar cluster. According to the formula $d_{xc} = \sqrt{(x_{i,1} - c_{v,1})^2 + (x_{i,2} - c_{v,2})^2 + \cdots + (x_{i,b} - c_{v,b})^2}$, the similarity between pixel x_i and cluster center c_v located in the spatial neighborhood is calculated, and this pixel is assigned to the cluster with the greatest similarity. In other words, this pixel is labeled with the cluster that is most similar to it. Where the cluster center located in the spatial neighborhood of x_i is where the cluster center is covered in the area of $2\sqrt{n/k} \times 2\sqrt{n/k}$ with x_i as the center.

Step 3: Correct cluster center. The centers of pixels belonging to the same cluster labels are regarded as the new cluster centers.

Step 4: Calculate distance D. This distance is the difference between pixels and their cluster centers. The algorithm will be terminated if D converges. Otherwise, it is required to return to Step 2.

2.2. KM-KMNF algorithm

Remote sensing images are inevitably degraded by certain types of noise due to the influence of sensor instrument errors and other environmental factors. The hyperspectral image X, which contains n pixels and b bands, is composed of two parts signal and noise (Green et al. 1988; Landgrebe and Malaret 1986),

$$x(p) = x_s(p) + x_N(p) \tag{1}$$

where $x(p)$ is the pixel vector at position p, $x_s(p)$ and $x_N(p)$ are the signal and noise contained in $x(p)$, respectively. For optical images, the signal and noise are commonly
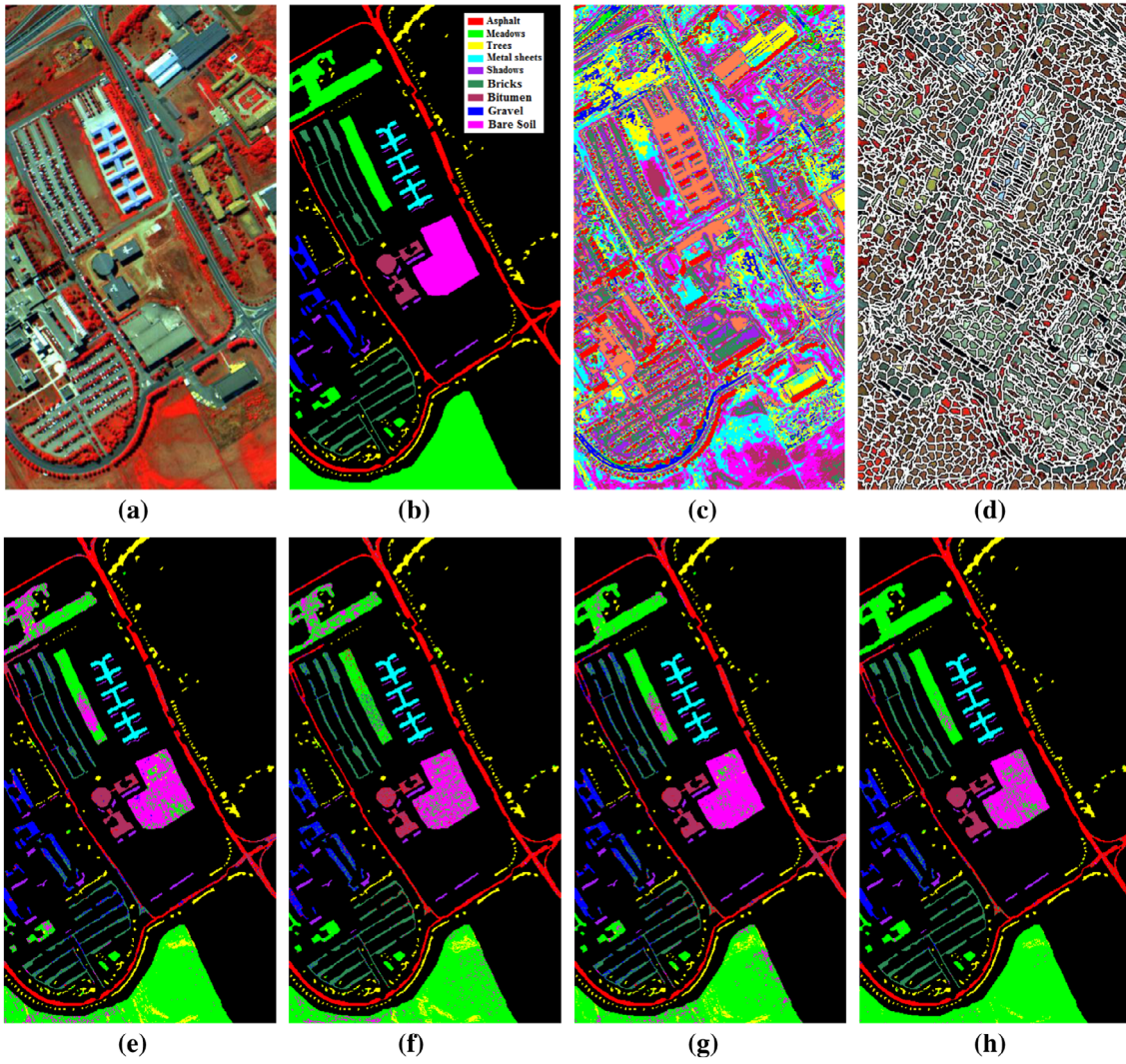
Figure 2. Experiment results on Pavia University data (a) Original Pavia University image. (b) Ground reference map containing nine land-cover classes. (c) Original K-means clustering result (same-colored region belongs to the same subregion). (d) The result of the image segmentation using the OK-means clustering method (each closed region is a subregion). The results of the ML classification after using different dimensionality reduction methods (number of features = 8), (e) OMNF method, (f) KMNF method, (g) OKMNF method, and (h) KM-KMNF method.

considered independent of each other (Landgrebe and Malaret 1986; Green et al. 1988). Thus, the covariance matrix S of image X can be expressed as the sum of the noise covariance matrix S_N and the signal covariance matrix S_S:

$$S = S_N + S_S \qquad (2)$$

Using $\tilde{x}_k$ to represent the mean value of all pixels in the k-th band of the image, the mean matrix X_{mean} with n rows b columns can be obtained.

$$X_{mean} = \begin{bmatrix} \tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_b \\ \tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_b \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_b \end{bmatrix} \qquad (3)$$

Then, the centralization matrix Z of X can be expressed as follows:

$$Z = X - X_{mean} \qquad (4)$$

The covariance matrix of image X can be expressed as

$$S = Z^T Z / (n - 1) \qquad (5)$$

Similarly, Let us consider $\tilde{x}_{Nk}$ as the average of the noise in k-th band, the mean noise matrix X_{mean} with n rows b columns can be obtained.

$$X_{Nmean} = \begin{bmatrix} \tilde{x}_{N1} & \tilde{x}_{N2} & \cdots & \tilde{x}_{Nb} \\ \tilde{x}_{N1} & \tilde{x}_{N2} & \cdots & \tilde{x}_{Nb} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_{N1} & \tilde{x}_{N2} & \cdots & \tilde{x}_{Nb} \end{bmatrix} \qquad (6)$$

The centralization matrix Z_N of the noise matrix X_N can be expressed as:

$$Z_N = X_N - X_{mean} \qquad (7)$$

The covariance matrix S_N of the noise matrix X_N can be expressed as:

$$S_N = Z_N^T Z_N / (n - 1) \qquad (8)$$

The NF is defined as the ratio of the noise covariance matrix to the total covariance matrix of image $\boldsymbol{X}$. Therefore, for the linear combination $\boldsymbol{a}^T\boldsymbol{Z}(p)$,

$$\text{NF} = \boldsymbol{a}^T\boldsymbol{S}_N\boldsymbol{a}/\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a} = \boldsymbol{a}^T\boldsymbol{Z}_N^T\boldsymbol{Z}_N\boldsymbol{a}/\boldsymbol{a}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{a} \quad (9)$$

where $\boldsymbol{a}$ is the eigenvector of NF. In NF, it is significant that the noise is estimated reliably. The original KMNF method primarily uses the spatial neighborhood 3×3 information of a hyperspectral image to estimate noise $\boldsymbol{Z}_N$, as shown below:

$$\begin{aligned}n_{i,j,k} &= z_{i,j,k} - \hat{z}_{i,j,k} \\ &= z_{i,j,k} - (-z_{i-1,j-1,k} + 2z_{i,j-1,k} - z_{i+1,j-1,k} + 2z_{i-1,j,k} \\ &\quad + 5z_{i,j,k} + 2z_{i+1,j,k} - z_{i-1,j+1,k} + 2z_{i,j+1,k} - z_{i+1,j+1,k})/9\end{aligned}$$

$$(10)$$

where $z_{i,j,k}$ is the pixel value in the i-th row, j-th column, and k-th band of hyperspectral image Z, $\hat{z}_{i,j,k}$ is the estimated value of this pixel, and $n_{i,j,k}$ represents the estimated noise value of $z_{i,j,k}$. Many works have reported that the estimated results have a relatively large error and are unstable when the noise is estimated by only using the spatial neighborhood information alone (Gao et al. 2013a; 2013b). In comparison, the results are more accurate using the SSDC method in OKMNF, as shown in Equation (11):

$$\begin{aligned}n_{i,j,k} &= z_{i,j,k} - \hat{z}_{i,j,k} \\ &= z_{i,j,k} - (a + bz_{i,j,k-1} + cz_{i,j,k+1} + dz_{p,k})\end{aligned} \quad (11)$$

where the parameters a, b, c, and d are the coefficients of the MLR, and $z_{p,k}$ represents the pixel value of the adjacent spatial positions at the same band. To estimate noise, we calculate the inversion parameters of the MLR, which is equivalent to identify minimum region of the MLR. In OKMNF, a region covering 6×6 pixels is set as the minimum region to obtain MLR parameters by virtue of the empirical value. However, this segmentation method would inevitably separate other classes to the same minimum region, such as boundary, this makes it difficult to guarantee homogeneous category within the region. In comparison, in the KM-KMNF algorithm, subregion X_{sub} generated by OK-means clustering is used as the minimum region to estimate noise, ensuring the homogeneity of the pixel features in the minimum region. Meanwhile, within each minimum region, the noise estimated by spectral decorrelation (which only considers the spectral information of hyperspectral image and the correction made by spatial neighborhood pixel $z_{p,k}$ to $z_{i,j,k}$ in Equation (11)) is eliminated. Thus, the impact of heterogeneous information of the image spatial neighborhood on the noise estimation results can be avoided. The equation for the spectral decorrelation excluding the spatial neighborhood information is as follows:

$$\begin{aligned}n_{i,j,k} &= z_{i,j,k} - \hat{z}_{i,j,k} \\ &= z_{i,j,k} - (a + bz_{i,j,k-1} + cz_{i,j,k+1})\end{aligned} \quad (12)$$

The adopted MLR model can be expressed as:

$$X_{sub} = B\mu + \varepsilon \quad (13)$$

$$\boldsymbol{X}_{\text{sub}} = \begin{bmatrix} x_{1,2,k} \\ x_{1,3,k} \\ \vdots \\ x_{w,h,k} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & x_{1,2,k-1} & x_{1,2,k+1} \\ 1 & x_{1,3,k-1} & x_{1,3,k+1} \\ \vdots & \vdots & \vdots \\ 1 & x_{w,h,k-1} & x_{w,h,k+1} \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$(14)$$

where $\boldsymbol{B}$ represents the spectral neighborhood matrix, $\boldsymbol{\mu}$ is the coefficient matrix, and $\boldsymbol{\varepsilon}$ is the residual value. Then, $\boldsymbol{\mu}$ could be estimated by:

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{X}_{\text{sub}} \quad (15)$$

The signal value could be estimated through:

$$\hat{\boldsymbol{X}}_{\text{sub}} = \boldsymbol{B}\hat{\boldsymbol{\mu}} \quad (16)$$

The noise value can be obtained from the difference between the true and estimated values:

$$\boldsymbol{N}_{\text{sub}} = \boldsymbol{X}_{\text{sub}} - \hat{\boldsymbol{X}}_{\text{sub}} \quad (17)$$

Then, estimating noise in all subregions was performed by image segmentation. The noise estimation result N of the whole hyperspectral image can be obtained accurately. To acquire the kernel minimum noise fraction (KNF), the noise estimation result is inserted into the NF calculation Equation (9), and NF is performed with dual transformation $\boldsymbol{a} \propto \boldsymbol{Z}^T\boldsymbol{b}$ and kernel transformation (Nielsen and Vestergaard 2012).

$$\text{KNF} = \boldsymbol{b}^T\kappa_N\kappa_N^T\boldsymbol{b}/\boldsymbol{b}^T\kappa^2\boldsymbol{b} \quad (18)$$

where κ is the Gaussian radial basis function (RBF).

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp[-\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2/(2\sigma^2)] \quad (19)$$

where σ is the proportional parameter, which can be obtained by calculating the mean distance between the observed value x_i and x_j (Nielsen 2011). To sort the image components according to the quality order after dimensionality reduction, KNF needs to be minimized, thus by solving the transformation matrix for feature extraction. Solving the minimized KNF equals to solve the symmetric generalized eigenvalues that can be solved using the maximized Rayleigh entropy (Nielsen 2011) which does not elaborate the detailed solution process. It should be noted that as far as KNF is concerned, the kernel noise value is calculated in the kernel space. In

other words, the real X of hyperspectral image and the estimated $\hat{X}$ of the MLR are converted to the kernel space at the same time. Subsequently, the noise value is obtained in the kernel space instead of the original space (Nielsen and Vestergaard 2012). Finally, the feature extraction results can be expressed as:

$$Y = \kappa\boldsymbol{b} \qquad (20)$$

3. Experiment and analysis

We compare the KM-KMNF with OMNF, KMNF, and OKMNF on two real hyperspectral remote sensing images. The features extracted by each method are used as input of the maximum likelihood (ML) classification and support vector machine (SVM) to validate the performances. The SVM classifier with RBF kernels in MATLAB SVM Toolbox, LIBSVM (Chang and Lin 2011; Liao et al. 2012), is applied in our experiments. The fivefold cross-validation is used to find the best parameters in SVM (Liao et al. 2012). Each experiment runs ten times, and the average value of these ten experiments is reported for comparison.

3.1. Experiment on Pavia University data

The Italian hyperspectral data of Pavia University were collected using the reflective optical system imaging spectrometer in 2001. The data contain 610×340 pixels and 103 spectral bands. In addition, the spectrum ranges from 0.43 μm to 0.86 μm, and the data have a spatial resolution of 1.3 m. The image contains nine different types of surface classes (Figure 2(a) and (b)). About 50% of the samples are randomly selected as the training samples. The remaining 50% of the samples are used as testing samples. The sample categories and quantities are shown in Table 1. The classification accuracy of the different feature extraction methods is shown in Table 2 and Figure 3. The results of classification are shown in Figure 2(e–h).

In Table 2, the classification accuracy of the four different feature extraction methods increases with the growing number of features. In particular, KM-KMNF produces a higher accuracy than the other three methods. Compared with KMNF and OKMNF, the accuracy of KM-KMNF could be improved by 10.95 and 4.69%, respectively, which indicates that incorporating the spatial information through image segmentation to the KM-KMNF algorithm as the minimum region of the MLR benefits post-applications. The accuracies of KM-KMNF and OKMNF are much higher than that of KMNF. In the meantime, the KM-KMNF performs better than OKMNF. The results show that for the noise estimation in KMNF algorithm, the spectral information is more important than the spatial dimension information of hyperspectral image. Moreover, Figure 2(e–h) shows that the classification results after KM-KMNF transformation are more consistent with the actual distribution of the surface features than those of other methods.

Table 1. Pavia University samples used for training and testing.

Classes	Training	Testing
Asphalt	3315	3316
Meadows	9324	9325
Gravel	1049	1050
Trees	1532	1532
Metal sheets	672	673
Bare soil	2514	2515
Bitumen	665	665
Bricks	1841	1841
Shadows	473	474
Total	21,385	21,391

Table 2. Overall accuracy of the ML classification on Pavia University image using different dimensionality reduction methods.

| Number of features	KMNF, %	OMNF, %	OKMNF, %	KM-KMNF, %
3	67.63	73.70	74.13	78.23
4	70.99	78.53	77.71	80.45
5	73.85	79.66	79.24	84.25
6	78.28	81.01	83.03	85.59
7	81.45	82.79	86.11	88.06
8	83.06	84.05	87.24	88.94
9	83.37	85.96	89.31	90.60
10	84.04	87.96	90.02	91.42
11	85.31	90.71	91.10	92.56
12	86.15	91.54	91.46	93.74
13	86.85	91.67	92.39	94.26
14	87.79	92.01	92.44	94.81
15	88.61	92.24	92.46	94.45
16	88.88	92.21	92.37	94.52
17	88.97	92.61	92.56	94.17
18	88.58	92.15	92.19	94.20
19	88.59	92.45	92.57	93.49
20	88.25	92.13	92.55	93.52
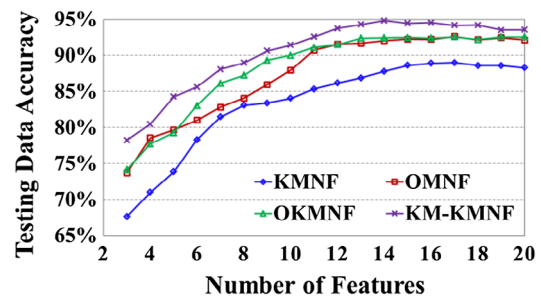
Figure 3. Comparison of the accuracies of the ML classification on Pavia University image using different number of features.

3.2. Experiment on Indian Pines data

Indian Pines hyperspectral data are acquired by the Airborne Visible/Infrared Imaging Spectrometer in 1992, which contains 145×145 pixels and 220 bands. The spectrum ranges from 0.4 μm to 2.5 μm, and the spatial resolution is 20 m. Only 200 bands of the image are taken into account in this experiment, whereas the noise bands and the atmospheric vapor absorption bands are excluded. Nine large classes are considered in this experiment (Figure 4(a) and (b)). About 25% percent of the samples are randomly selected as the training samples. The remaining 75% of the samples are used as testing samples. The sample categories and quantities are shown
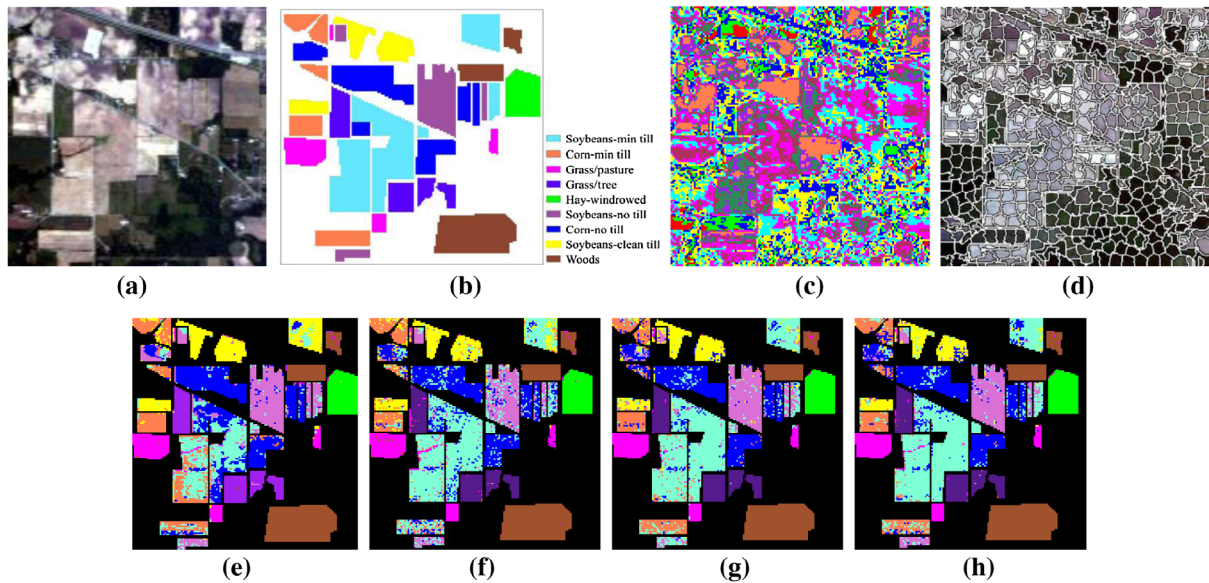
Figure 4. Experiment results on Indian Pines data (a) Original Indian Pines image. (b) Ground reference map containing nine land-cover classes. (c) Original K-means clustering result (same-colored region belongs to the same subregion). (d) The result of image segmentation using the optimal K-means clustering method (each closed region is a subregion). The results of the ML classification after using different dimensionality reduction methods (number of features = 8), (e) OMNF method, (f) KMNF method, (g) OKMNF method, and (h) KM-KMNF method.

Table 3. Samples from Indian Pines used for training and testing.

Classes	Training	Testing
Corn-no till	359	1075
Corn-min till	209	625
Grass/Pasture	124	373
Grass/Trees	187	560
Hay-windrowed	122	367
Soybean-no till	242	726
Soybean-min till	617	1851
Soybean-clean till	154	460
Woods	324	970
Total	2338	7007

in Table 3. The classification accuracy of the different classifiers (ML and SVM) using different feature extraction methods is shown in Table 4, Table 5, and Figure 5. The classification results are shown in Figure 4(e–h).

Tables 4 and 5 show that KM-KMNF contributes a higher accuracy than the other algorithms as the number of features increases, similar to that of the Pavia University data. This further demonstrates that the efficiency of KM-KMNF on feature extraction. Figure 5 also reveals that the accuracies of the three methods. By taking the spatial and spectral information of image into consideration, our KM-KMNF produces higher accuracy than that of the KMNF method which solely considers the spatial information for noise estimation. This confirms that the accurate calculation of the noise is crucial to the performance of the feature extraction for KMNF and MNF-based methods. In Table 4, we also can find that most standard deviations of KM-KMNF are relatively smaller than the others, indicating the stable performance of KM-KMNF. From Figure 3 and Figure 5,

Table 4. Overall accuracies of the ML classification on Indian Pines image using different dimensionality reduction methods.

| Number of features	KMNF(Std)	OMNF(Std)	OKMNF(Std)	KM-KMNF(Std)
3	59.79% (0.012 3)	66.56% (0.004 8)	64.41% (0.005 2)	67.65% (0.004 3)
4	66.62% (0.011 6)	72.83% (0.004 7)	67.53% (0.007 7)	72.75% (0.003 0)
5	69.72% (0.017 2)	74.80% (0.004 4)	76.60% (0.004 1)	77.50% (0.004 6)
6	73.55% (0.008 5)	76.54% (0.004 2)	78.48% (0.002 6)	80.39% (0.003 1)
7	74.77% (0.006 7)	77.81% (0.004 5)	79.94% (0.003 3)	81.98% (0.003 7)
8	76.27% (0.003 8)	77.96% (0.004 2)	82.84% (0.005 3)	84.31% (0.004 0)
9	76.84% (0.001 7)	82.02% (0.004 3)	84.84% (0.006 2)	84.36% (0.007 1)
10	77.90% (0.004 2)	83.92% (0.003 9)	85.48% (0.003 9)	85.54% (0.004 8)
11	78.62% (0.006 5)	84.89% (0.003 7)	85.92% (0.003 7)	86.43% (0.004 5)
12	79.17% (0.017 6)	86.27% (0.003 3)	86.33% (0.003 3)	87.17% (0.004 5)
13	79.79% (0.015 0)	86.68% (0.005 7)	86.61% (0.005 7)	88.00% (0.002 4)
14	80.43% (0.013 7)	87.70% (0.004 0)	86.89% (0.001 6)	89.01% (0.004 6)
15	80.74% (0.002 0)	88.07% (0.002 7)	87.40% (0.001 7)	89.30% (0.003 4)
16	80.89% (0.004 5)	88.54% (0.003 9)	87.75% (0.001 0)	89.70% (0.003 1)
17	80.93% (0.007 3)	88.44% (0.003 4)	87.94% (0.002 2)	89.65% (0.003 4)
18	81.07% (0.006 2)	88.94% (0.004 4)	87.91% (0.001 2)	89.64% (0.003 7)
19	81.20% (0.005 4)	88.88% (0.004 2)	88.17% (0.002 0)	89.61% (0.003 3)
20	81.18% (0.005 0)	88.94% (0.004 2)	88.51% (0.003 3)	89.75% (0.003 2)
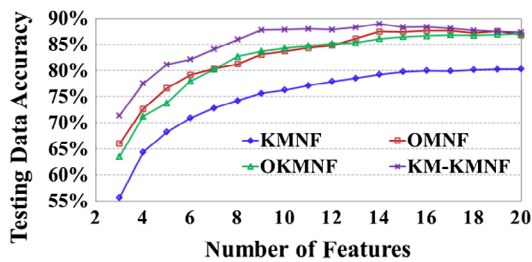
Figure 5. Comparison of the accuracies of the SVM classification on Indian Pines image after using different dimensionality reduction methods.

as the number of extracted features increases, the performance of each feature extraction method first increases and then decreases. It confirms that feature extraction can improve the classification performance on hyperspectral images, and most information can be preserved even with a few extracted features. With SVM classifier, we have similar findings as that of the ML classifiers. The KM-KMNF produces better results than the others. In particular, KM-KMNF with ML classifier outperforms those of SVM classifier.

We also conduct experiments by increasing the number of training samples, as shown in Table 6. We find that as the number of training samples increases, the classification accuracy first increases and then keeps a relative stable status. To compare the efficiency of feature extraction methods, we take Indian Pines data as an example (by extracting 20 features), and the consumed time of KM-KMNF, OKMNF, OMNF, and KMNF is 21.57 s, 19.99 s, 20.71 s, and 1.50 s, respectively. We

find that KM-KMNF consumes relative more time but with better performances. When extracting less features, KM-KMNF can reduce the consumed time, and get the best performance when using around 14 features.

4. Conclusions

This paper is to propose a new kernel minimum noise fractional transformation for the feature extraction of hyperspectral data. Our method uses spectral-dimension decorrelation for calculation of the NF, and further improves the accuracy of the NF by introducing the spatial information through image segmentation to determine the minimum region of the MLR. Moreover, the conventional K-means clustering method has been improved to make OK-means algorithm more suitable for hyperspectral image segmentation. Two real hyperspectral image data sets are used for the experiments, and the classification accuracy is used as the index to evaluate the performances of feature extraction algorithms. The results show that the accuracy of the MNF based on image segmentation is much higher than that of the original KMNF algorithm. Meanwhile, it is also much higher than that of the OMNF and OKMNF algorithms. Better noise estimation brings better performances on both feature extraction and its post-application to hyperspectral image classification for MNF- and KMNF-based methods.

Notes on contributors

Bin Zhao is a Postgraduate of Electronics and Communication Engineering at the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. His research interests include hyperspectral image feature extraction and classification.

Lianru Gao is currently a Professor with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. He received

Table 5. Overall accuracies of the SVM classification on Indian Pines image using different dimensionality reduction methods.

Number of features	KMNF, %	OMNF, %	OKMNF, %	KM-KMNF, %
3	55.65	66.00	63.52	71.39
4	64.38	72.66	71.26	77.55
5	68.23	76.64	73.78	81.15
6	70.89	79.18	77.99	82.16
7	72.82	80.33	80.21	84.10
8	74.23	81.27	82.72	85.96
9	75.64	83.07	83.70	87.78
10	76.26	83.72	84.32	87.85
11	77.12	84.33	84.78	88.00
12	77.91	84.83	85.15	87.85
13	78.52	86.14	85.30	88.30
14	79.25	87.49	86.10	88.98
15	79.77	87.46	86.47	88.32
16	80.00	87.61	86.66	88.38
17	79.90	87.67	86.85	88.12
18	80.14	87.18	86.74	87.71
19	80.24	87.56	86.92	87.52
20	80.30	86.82	87.06	87.33

Table 6. Overall classification accuracies by using different number of training samples.

Methods	Training samples/class (Pavia University)				Training samples/class (Indian Pines)			
	50	100	150	200	50	100	150	200
KMNF	80.95%	82.82%	82.93%	83.55%	74.72%	75.29%	75.82%	76.44%
KM-KMNF	86.94%	87.72%	88.46%	88.26%	83.11%	84.23%	84.28%	84.20%

the BS degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the PhD degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2007. His research interests include hyperspectral image processing and information extraction. He has published more than 100 papers in China and abroad.

Wenzhi Liao, since 2012, has been working as a Postdoc at Ghent University and as a Postdoctoral Fellow with the Fund for Scientific Research, Flanders-FWO. He received the BS degree in mathematics from Hainan Normal University, Haikou, China, in 2006, the PhD degree in engineering from South China University of Technology, Guangzhou, China, in 2012, and the PhD degree in computer science engineering from Ghent University, Ghent, Belgium, in 2012. His current research interests include pattern recognition, remote sensing, and image processing. In particular, his interests include mathematical morphology, multitask feature learning, multisensor data fusion, and hyperspectral image restoration.

Bing Zhang is a Professor and the Deputy Director of the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. He received the BS degree in geography from Peking University, Beijing, China, and the MS and PhD degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China. His research interests include the development of Mathematical and Physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas, such as geology, hydrology, ecology, and botany. He has developed five software systems in the image processing and applications. He has authored more than 300 publications, including 171 journal papers.

References

Bajcsy, P., and P. Groves. 2004. "Methodology for Hyperspectral Band Selection." *Photogrammetric Engineering & Remote Sensing* 70 (7): 793–802.

Ball, J. E., L. M. Bruce, T. West, and S. Prasad. 2007. "Level Set Hyperspectral Image Segmentation Using Spectral Information Divergence-based Best Band Selection." IEEE International Conference: the Geoscience & Remote Sensing Symposium, Barcelona, Spain, July 23–28. doi: 10.1109/igarss.2007.4423739.

Benediktsson, J. A., J. A. Palmason, and J. R. Sveinsson. 2005. "Classification of Hyperspectral Data from Urban Areas based on Extended Morphological Profiles." *IEEE Transactions on Geoscience & Remote Sensing* 43 (3): 480–491. doi:10.1109/tgrs.2004.842478.

Chang, C. C., and C. J. Lin. 2011. "LIBSVM: A Library for Support Vector Machines." *ACM Transactions on Intelligent Systems and Technology* 2 (3): 1–27.

Gao, L., B. Zhang, Z. Chen, and L. Lei. 2011. "Study on the Issue of Noise Estimation in Dimension Reduction of Hyperspectral Images." The 3rd Workshop Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, June 6–9. doi: 10.1109/whispers.2011.6080944.

Gao, L., Q. Du, W. Yang, and B. Zhang. 2012. "A Comparative Study on Noise Estimation for Hyperspectral Imagery." The Workshop on Hyperspectral Image & Signal Processing: Evolution in Remote Sensing. Shanghai, China, June 4–7. doi: 10.1109/whispers.2012.6874262.

Gao, L., Q. Du, B. Zhang, W. Yang, and Y. Wu. 2013. "A Comparative Study on Linear Regression-Based Noise Estimation for Hyperspectral Imagery." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 6 (2): 488–498. doi:10.1109/jstars.2012.2227245.

Gao, L., B. Zhang, X. Sun, S. Li, Q. Du, and C. Wu. 2013. "Optimized Maximum Noise Fraction for Dimensionality Reduction of Chinese HJ-1A Hyperspectral Data." *EURASIP Journal on Advances in Signal Processing* 2013 (1): 1–12. doi:10.1186/1687-6180-2013-65.

Gao, L. R., B. Zhao, X. P. Jia, W. Z. Liao, and B. Zhang. 2017. "Optimized Kernel Minimum Noise Fraction Transformation for Hyperspectral Image Classification." *Remote Sensing* 9 (6): doi:10.3390/rs9060548.

Gómez-Chova, L., and A. A. Nielsen. 2011. "Explicit Signal to Noise Ratio in Reproducing Kernel Hilbert Spaces." The IEEE International Conference: the Geoscience & Remote Sensing Symposium, Vancouver, BC, Canada, July 24–29. doi: 10.1109/igarss.2011.6049993.

Greco, M. 2006. "Analysis of the Classification Accuracy of A New MNF-based Feature Extraction Algorithm." *Proceedings of SPIE* 6365 63650V-1. doi:10.1117/12.689961.

Green, A. A., M. Berman, P. Switzer, and M. D. Craig. 1988. "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal." *IEEE Transactions on Geoscience & Remote Sensing* 26 (1): 65–74. doi:10.1109/36.3001.

Guo, B., S. R. Gunn, R. I. Damper, and J. D. B. Nelson. 2006. "Band Selection for Hyperspectral Image Classification Using Mutual Information." *IEEE Geoscience & Remote Sensing Letters* 3 (4): 522–526. doi:10.1109/lgrs.2006.878240.

Huang, Z. 1998. "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values." *Data Mining & Knowledge Discovery* 2 (3): 283–304. doi:10.1023/a:1009769707641.

Hughes, G. 1968. "On the Mean Accuracy of Statistical Pattern Recognizers." *IEEE Transactions on Information Theory* 14 (1): 55–63. doi:10.1109/tit.1968.1054102.

Jia, X., B. C. Kuo, and M. M. Crawford. 2013. "Feature Mining for Hyperspectral Image Classification." *Proceedings of the IEEE* 101 (3): 676–697. doi:10.1109/jproc.2012.2229082.

Kanungo, T., D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. 2002. "An Efficient k-Means Clustering Algorithm: Analysis and Implementation." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 24 (7): 881–892. doi:10.1109/tpami.2002.1017616.

Landgrebe, D. A. and E. Malaret. 1986. "Noise in Remote-Sensing Systems: The Effect on Classification Error." *IEEE Transactions on Geoscience & Remote Sensing* GE-24(2): 294–300. doi:10.1109/tgrs.1986.289648.

Lee, J. B., A. S. Woodyatt, and M. Berman. 1990. "Enhancement of High Spectral Resolution Remote-Sensing Data by A Noise-adjusted Principal Components Transform." *IEEE Transactions on Geoscience & Remote Sensing* 28 (3): 295–304. doi:10.1109/36.54356.

Li, W., S. Prasad, J. E. Fowler, and L. M. Bruce. 2012. "Locality-Preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis." *IEEE Transactions on Geoscience & Remote Sensing* 50 (4): 1185–1198. doi:10.1109/TGRS.2011.2165957.

Liao, W., A. Pizurica, P. Scheunders, W. Philips, and Y. Pi. 2012. "Semisupervised Local Discriminant Analysis for Feature Extraction in Hyperspectral Images." *IEEE Transactions on Geoscience & Remote Sensing* 51 (1): 184–198. doi:10.1109/tgrs.2012.2200106.

Liu, X., B. Zhang, L. R. Gao, and D. M. Chen. 2009. "A Maximum Noise Fraction Transform with Improved Noise Estimation for Hyperspectral Images." *Science in China* 52 (9): 1578–1587. doi:10.1007/s11432-009-0156-z.

Nielsen, A. A. 2011. "Kernel Maximum Autocorrelation Factor and Minimum Noise Fraction Transformations." *IEEE Transactions on Image Processing* 20 (3): 612–624. doi:10.1109/tip.2010.2076296.

Nielsen, A. A., and J. S. Vestergaard. 2012. "Parameter Optimization in the Regularized Kernel Minimum Noise Fraction Transformation." IEEE International Conference: Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, July 22–27. doi: 10.1109/igarss.2012.6351561.

Plaza, A., J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, and J. Chanussot. 2009. "Recent Advances in Techniques for Hyperspectral Image Processing." *Remote Sensing of Environment* 113 (1): S110–S122. doi:10.1016/j.rse.2007.07.028.

Roger, R. E. 1996. "Principal Components transform with simple, automatic noise adjustment." *International Journal of Remote Sensing* 17 (14): 2719–2727.

Zhao, B., L. Gao, and B. Zhang. 2016. "An Optimized Method of Kernel Minimum Noise Fraction for Dimensionality Reduction of Hyperspectral Imagery." IEEE International Conference: the Geoscience & Remote Sensing Symposium, Beijing, China, July 10–15. doi: 10.1109/igarss.2016.7729003.