Unsupervised Feature Selection Using Geometrical Measures in Prototype Space for Hyperspectral Imagery

Mohsen Ghamary Asl, Mohammad Reza Mobasheri, and Barat Mojaradi

Abstract—Feature/band selection is a common technique to overcome the “curse of dimensionality” posed by the high dimensionality of hyperspectral imagery. When the image is characterized by unknown phenomena, an unsupervised approach can be utilized to select the most distinctive and informative bands. The efficiency of an unsupervised feature selection (FS) depends on the criteria to be optimized and the space (e.g., feature space, pixel space, spectral space, etc.) in which the data are represented. Moreover, the determination of the initial feature and the determination of the optimal feature size (the optimal number of distinct bands to be selected) are other challenges faced in unsupervised approaches. In this paper, we propose two unsupervised FS methods by representing bands in the prototype space (PS). The first method proposes a way for selecting the initial feature based on the orthogonal distance from the PS diagonal and determines the optimal feature size by employing the HySime algorithm in the PS. The second method uses two criteria defined by the tangent of the angles between the band vectors in the PS in order to select the initial feature and to describe the band correlations. Meanwhile, the determination of the optimal feature size is embedded in this method. The experimental results on real and synthetic data sets show that our methods are more reliable and can yield a better result in terms of class separability and Friedman test than other widely used techniques.

Index Terms—Hyperspectral imagery (HSI), optimal feature size, prototype space (PS), unsupervised feature selection (FS).

I. INTRODUCTION

HYPER SPECTRAL sensors collect data simultaneously in dozens or hundreds of narrow and adjacent spectral bands for each pixel. These data are commonly used in applications such as classification [1], unmixing [2], anomaly detection [3], target detection [4], and background characterization [5]. From a practical point of view, due to the curse of dimensionality, utilizing all of the bands in an algorithm does not necessarily lead to an improvement in the results. In other words, in the context of hyperspectral data analysis, the determination of the effective bands is vital for the concise representation of phenomena. Therefore, it is of paramount importance to select independent and informative bands that enable us to present objects/classes in an appropriate and detectable manner. In particular, the unsupervised feature selection (FS) is of great value where there is no prior knowledge of phenomena present in the scene. In addition, unsupervised FS could improve some unsupervised tasks such as anomaly detection, data storage, and transmission. Hence, it is ideal to use algorithms that have the ability to desirably reduce the data dimensionality in an unsupervised manner.

Unsupervised dimensionality reduction (DR) methods do not require any prior knowledge or training data. Moreover, they are not directly aimed at optimizing the accuracy in a given classification task [6]. In the relevant literature, different FS/extraction (FS/FE) methods are categorized as two major approaches for unsupervised DR [7]–[9]. Typically, the FE methods transform data into a low-dimensional space by using several criteria that usually change the physical meaning of the original data [10]. Principal component analysis (PCA) [11], independent component analysis (ICA) [12], unsupervised linear feature extraction [13], maximum noise fraction (MNF) [14], and wavelet transform [15] methods are examples of this category. An unsupervised feature extraction algorithm, nonnegative matrix factorization, was used to represent the hyperspectral imagery (HSI) for the subsequent object-oriented classification [16]. A series of feature extraction algorithms was reviewed and combined with object-based segmentation for classification of HSI in [17].

 Manuscript received November 8, 2012; revised March 3, 2013, April 25, 2013, and July 7, 2013; accepted July 18, 2013. (Corresponding author: B. Mojaradi.) M. Ghamary Asl and M. R. Mobasheri are with the Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran 19967-15433, Iran (e-mail: mghamary@dena.kntu.ac.ir; Mobasheri@kntu.ac.ir).
 B. Mojaradi is with the Department of Geomatics Engineering, School of Civil Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran (e-mail: Mojaradi@iust.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2275831
In general, all feature/band selection methods preserve the physical meaning of HSI. Unsupervised FS methods optimizing predefined criteria, regardless of their search strategy stage, are categorized under filter approaches. The filter approach operates independently of any classification algorithm, so redundant and irrelevant features are omitted before the classification process begins. The information-theory-based FS methods, which use criteria like correlation coefficient [18], [19], entropy [20], mutual information [21], linear prediction error (LPE) [22], etc., are examples of the filter approach. Furthermore, supervised measures such as Mahalanobis distance, Bhattacharyya distance, etc. [18], [19], can also be estimated through a clustered image in an unsupervised manner. These algorithms employ search strategies such as sequential forward/backward selection (SFS/SBS) and sequential floating forward/backward selection (SFFS/SFBS) [23] which yield suboptimal results and suffer from high computational burden.

Furthermore, PCA, ICA, MNF, and wavelet-transform-based unsupervised FS methods were proposed in [24]–[26]. There are also some other methods presented by researchers, which make use of computational intelligence (e.g., fuzzy logic, genetic algorithm, and clonal selection) to select informative bands [27]–[30]. Another attempt has been made via band clustering in [31]. Some unsupervised FS algorithms, based on band similarity measurement, were also proposed in [32]. An unsupervised FS based on the similarity measure was also proposed in [33] to reduce the high-dimensional structural and textural features. In addition to these methods, there are other FS methods that operate in the spectral space and make use of the geometrical properties and absorption features of pixel spectra. This means that bands that are located in absorption regions of the spectrum of a pixel or a cluster/class mean are selected as suitable or diagnostic features. One of the most famous algorithms that use these methods for FS is Tetradecorder [34]. In this direction, a new method called classification by diagnosing all absorption features [35] has been proposed. This method uses the absorption and geometrical characteristics of pixels and substances in order to find bands of spectra as the most suitable features for classification.

Recently, several endmember-selection-based methods, which search for distinctive spectral signatures, were applied to unsupervised FS. These methods are employed in the pixel space [36] where each image band is represented as a vector. The dimensionality of such a space is equal to the number of image pixels. For instance, the geometrical FS (G-FS) [36] and linear prediction (LP) [10] FS methods operate based on N-FINDR [37] and orthogonal subspace projection (OSP) [38] techniques in this space, respectively. The LP method has also been developed based on unsupervised fully constrained least squares linear unmixing (UFCLSLU) method [39] in the pixel space to select distinct features. Since the LP method uses the LPE criterion to select the optimal features, its results are identical to those of OSP [10]. Moreover, a parallel processing technique has been used for unsupervised FS in [40] to alleviate the computational burden of LP/OSP and to speed up the performance without affecting the FS result. A similar attempt was also made in [41] for the LP and G-FS FS algorithms. Due to the large number of bands and the presence of highly correlated pixels in the pixel space, there are practical considerations in LP and G-FS which must be taken into account, i.e., 1) the number of pixels involved in the FS process and 2) the number of features to be selected [10]. Therefore, the aforementioned methods may lead to suboptimal results.

To overcome the high computational cost, G-FS and LP use a representative subset of pixels or a random percentage of all pixels. Recently, an N-FINDR-based pixel selection strategy was also introduced in [41]. In general, the success of each unsupervised FS method is dependent upon several issues: 1) a space having a suitable representation and interpretation of the data; 2) an appropriate criterion; 3) a proper selection of the initial feature to select the rest of the features; and 4) the optimal feature size (the optimal number of distinct bands to be selected) in an unsupervised manner.

This paper aims at selecting distinct and informative features in the prototype space (PS) [42] via a geometrical interpretation of the distinct bands. The PS is constructed based on the spectra of endmembers or cluster centers as the representatives of the phenomena present in the image. Therefore, the dimensionality of this space is equal to the number of distinct phenomena in the image scene [i.e., virtual dimensionality (VD)]. In the PS, bands are represented as points/vectors. In this paper, we propose two unsupervised FS methods operating in such a PS. The innovations of the first proposed method called PFS, which stands for PS-based FS, lie in the use of the PS and the application of a different approach in selecting the initial feature by the OSP operator in the PS. This initial feature is the most informative band with the maximum orthogonal distance from the PS diagonal. The second proposed method called MTD, which stands for maximum tangent discrimination, uses the tangent of the angles between two band vectors in the PS to describe band correlations. This method initiates by choosing the most distinctive feature. It then continues by selecting the optimal features using special tangent-based criteria. Furthermore, these methods are able to determine the optimal feature size.

This paper is organized as follows. In Section I, the problems of hyperspectral data analysis, the need to develop DR algorithms, a brief description of the most common methods, and the objectives of this paper are presented. In the next section, the state-of-the-art methods of unsupervised FS that operate in the pixel space are presented. In Section III, the PS and the proposed methods are explained in detail. In Section IV, the results of the methods and experiments are analyzed, and finally, in Section V, the conclusion is given.

II. RECENT UNSUPERVISED FS METHODS

In the following, detailed descriptions of LP and G-FS as pixel-space-based FS methods are presented.

A. G-FS Method

This method is based on the concept of an automatic endmember extraction algorithm called N-FINDR [37], which is a geometric algorithm. The main assumption in N-FINDR is that the spectrum of each pixel is a linear mixture of the spectra of endmembers present in the image. On the other
hand, for a $B$-dimensional data, the maximum volume of a simplex—a Euclidean geometric spatial element having the minimum number of boundary points—is obtained by $B + 1$ pure pixels [36]. The G-FS method constructs a $P$-dimensional pixel space by band vectors $b_j$

$$ b_j = [b_j^1, b_j^2, \ldots, b_j^B]^T, \quad j = 1, 2, \ldots, B. \tag{1} $$

The theory of G-FS expresses that the points near to or at the ends (vertices) of the simplex are most likely independent features. Now, if a simplex with maximum volume is constructed in the pixel space, the bands with the least amount of correlation can be determined to be optimal features. For situations with high dimensionality, the $n$ extreme features ($n < B$) can be determined by the geometrical algorithm in such a way that a convex simplex with $n$ apexes is formed while other points lie inside it as far as possible.

Despite the high performance of G-FS in selecting suitable features, some notes need to be mentioned. Since the axes in the pixel space are pixel brightness, they are likely to be highly correlated, especially for the pixels of an object. Consequently, any processing in this space is prone to errors resulting from correlations of neighboring pixels. The next point to take into account is the random selection of initial features, a mechanism borrowed from N-FINDR. The factor of randomness in selecting initial samples causes the G-FS algorithm to yield new and sometimes utterly different results each time it is iterated.

B. LP Method

The LP method is based on the similarity and correlation between bands. The concept of LP was originally used in UFCLSLU to extract pure pixels [10]. Based on this concept, a pixel with the maximum reconstruction error in the linear mixture of pure pixels is a distinct pixel that can be regarded as a pixel with the maximum reconstruction error in the linear combination of the selected features. However, this criterion does not always satisfy the classification purpose, as these features do not necessarily contain enough information required for class separability in the image.

III. PROPOSED ALGORITHMS

Evidently, algorithms like G-FS, which rely on a large number of pixels, may be prone to bad pixels, noisy data, or correlated pixels. To remedy these problems, it is better to use algorithms that make use of class spectra in the original feature space. In this paper, we propose two FS methods in the PS. The use of the PS provides our methods with the capability of avoiding the shortcomings of the pixel space. In the first method, PFS, the concept of OSP is used in the PS to describe the informativeness and independence of the bands. The second method, MTD, uses two criteria defined by the tangent of the angles between band vectors in the PS to describe band correlations.

A. PS

In this space, the representatives of the phenomena/classes define the axes. In addition, such a space supplies information concerning band characteristics, unlike the feature space that provides information about the pixels of the phenomena. As it is known, the dimensionality of the pixel space is equal to the number of entire/selected image pixels [36]. However, the dimensionality of the PS is equal to the number of known classes or clusters estimated by subspace identification methods [42]. On the other hand, there is usually little correlation between different objects in the image, as opposed to the high correlation between pixels. Hence, the axes in the PS possess very high independence compared with the axes in the pixel space. Another characteristic of the PS is that, the farther away a band is located from the space diagonal (e.g., band $k$ or $s$ illustrated in Fig. 1), the more information it contains about the separability of the classes present in the image. The reason is that, the closer the spectral response values of objects in a band, the less useful the information that band contains for discriminating objects. Moreover, due to the proximity of spectral values (of different objects in a given band), the band would be close to the diagonal in the PS (e.g., band $g$ or $s$ demonstrated in Fig. 1). Evidently, the reverse is also true. This is illustrated in Fig. 1.

According to Fig. 1, bands $s$ and $g$ are landed near the PS diagonal and do not contain helpful information to discriminate cluster 1 and cluster 2. In contrast, bands $t$ and $k$ are informative ones that lie farther away from the diagonal in the PS. In addition, some of the highly correlated bands (e.g., band sets $P$ and $Q$) fall in clusters $P$ and $Q$ in the PS.

B. PFS

According to the aforementioned band characteristics in the PS, the correlation of the bands present in the PS can be taken into account by a measure. Owing to the OSP measure, we
Fig. 1. Band representation in the PS. (a) Spectral space which shows the mean spectra of two clusters. (b) PS constructed by the spectra, which represents some bands (e.g., t, s, k, ... ) and band sets P and Q.

can describe such a correlation between bands. The proposed method called PFS tries to choose the most informative band as the first feature such that its OSP is the largest from the PS diagonal. Then, the other independent features are selected using the OSP measure. The implementation stages of the PFS method are as follows.

1) Subspace identification: the VD of the image is determined using a conventional subspace identification method such as HySime [2], [43].

2) Image clustering: data are clustered into phenomena that may exist in the imaging scene. Since a pixel of a given class may be mixed with its background (i.e., surrounding classes), the minimum number of clusters is probably equal to twice the VD of the image. Our experiments empirically showed that a good upper bound for an optimum number of clusters was twice the VD. In this regard, at this stage, the image is clustered into $L = 2 \times \text{VD}$ clusters using a conventional clustering algorithm, such as K-means.

3) PS construction: using mean vectors of the $L$ spectral clusters in the original feature space, the PS is constructed (Fig. 1).

4) Initial FS: the informative band, which has the longest orthogonal distance from the PS diagonal, is selected as the first feature ($f_1$). To do so, the orthogonal subspace projector $P_\perp^d$ is applied to all band vectors in the PS, where $d_{(B \times 1)} = [1, 1, \ldots, 1]^T$ is defined as the PS diagonal

$$P_\perp^d = I - d(d^T d)^{-1} d^T$$  

$$q_j = P_\perp^d \times b_j$$  

$$f_1 = \arg \max_{b_j} \{\|q_j\|\}, \quad j = 1, 2, \ldots, B$$  

$$U = \{f_1\}$$  

where $q_j$ is the OSP value of the $j$th band ($b_j$) and $U$, which adopts $f_1$ as the first member, indicates the set of features to be selected.

5) FS: the OSP operator $P_\perp^U$ is applied to the rest of the bands to select $f_k$ (the $k$th proper feature) and then continues to select $L$ features

$$P_\perp^U = I - U(U^T U)^{-1} U^T$$  

$$q_j = P_\perp^U \times b_j$$  

$$f_k = \arg \max_{b_j} \{\|q_j\|\}, \quad j = 1, 2, \ldots, B$$  

$$U = \{f_1, f_2, \ldots, f_k\}.$$  

In general, determining the optimal feature size in an unsupervised FS method is a challenge. The optimal feature size can be determined by applying the HySime method in the PS.

C. MTD

FS through a geometrical constraint in the PS is the goal of this paper. Evidently, two coincident vectors, which are dependent/correlated, gain more independence or less correlation as the angle between them approaches 90°. Therefore, here the angle between two band vectors in the PS is used to describe the correlation of the bands. The implementation stages of the MTD method are as follows.

1)–3) The first three stages of the PFS method, including subspace identification, image clustering, and PS construction,
are implemented exactly in the same way and order for the MTD method.

4) The initial FS: the most distinct band, which has the biggest sum of angles from other bands, is selected as the first feature \((f_1)\). In this regard, for a given band, the summation of the tangents of the angles with other bands in the PS can be applied as the criterion \(C_1 = \{C_1^1, C_1^2, \ldots, C_1^B\}\) to determine the most distinct band. For the \(j\)th band, \(C_1^j\) is defined as follows:

\[
C_1^j = \sum_{i=1}^{B} t_{b_i b_j}, \quad j = 1, 2, \ldots, B
\]

where \(t_{b_i b_j} = \tan(\alpha_{ij}) = \cos^{-1}(b_i b_j / |b_i| |b_j|)\). \(b_i\) and \(b_j\) are the \(i\)th and \(j\)th band vectors in the PS. In order to select the most distinct band \(f_1\), the criterion \(C_1^j\) is computed for each band \(j\). Hence, band \(j\) which optimizes the criterion, is selected as \(f_1\) (i.e., the first feature). Hereafter, \(f_i\) refers to the \(i\)th selected feature.

\[
f_1 = \arg\max_{b_j} \{C_1^j\}, \quad j = 1, 2, \ldots, B.
\]

5) FS: the second feature \(f_2\) is selected such that the tangent of the angle between \(f_1\) and \(f_2\) gets maximum. To select the \(k\)th feature, the criterion \(C_2 = \{C_2^1, C_2^2, \ldots, C_2^B\}\) that explains the products of the elements \(t_{f_k b_j}\) \((i = 1, 2, \ldots, k - 1)\) for each band \(j\) is defined as

\[
C_2^j = \prod_{i=1}^{k-1} t_{f_i b_j}, \quad j = 1, 2, \ldots, B.
\]

For the \(i\)th feature in the selected feature set \(U = \{f_1, f_2, \ldots, f_k\}\), a set of elements \(t_{f_i b_j}\) can be calculated. Therefore, the tangent matrix \(T\) with dimension \((k - 1) \times B\) is generated, as in (14), to determine the \(k\)th optimal feature

\[
T_{(k-1)\times B} = \begin{bmatrix} t_{f_1 b_1} & t_{f_1 b_2} & \cdots & t_{f_1 b_B} \\ t_{f_2 b_1} & t_{f_2 b_2} & \cdots & t_{f_2 b_B} \\ \vdots & \vdots & \ddots & \vdots \\ t_{f_k b_1} & t_{f_k b_2} & \cdots & t_{f_k b_B} \end{bmatrix}
\]

Then, the \(k\)th feature is selected using the following argument:

\[
f_k = \arg\max_{b_j} \{C_2^j\}, \quad j = 1, 2, \ldots, B.
\]

After selecting the \(k\)th feature, matrix \(T\) is updated by adding a row containing the tangents of the angles between the \(k\)th feature and all \(B\) bands. In this regard, the algorithm continues until the tangent of the angle between \(f_k\) and \(f_1\) becomes equal to the minimum value of the first row of \(T\) (i.e., the algorithm returns to a status similar to the one where \(f_1\) was selected). On the other hand, MTD explores the PS to evaluate the independence of bands compared to the selected feature set. Then, the algorithm stops when it finds a band having the minimum angle with the first selected feature \(f_1\). It is noteworthy that, through this procedure, the algorithm achieves the optimal feature size. A sample of the FS order of the MTD method is illustrated in Fig. 2. In this figure, the angle between band vectors corresponding to features 1 and 17 is the smallest one causing the MTD method to stop. However, feature 17 is not selected as the proper feature.

IV. EXPERIMENTS AND RESULTS

A. Area of Study

The data sets used in the experiments were the well-known AVIRIS and Hyperion data. The AVIRIS data set was collected in June 1992 over the Indian Pine Test Site with the size of 145 × 145 pixels. The spatial resolution is about 20 m, and the radiometric resolution is 10 bits in 220 spectral bands. Bands 104–108, 150–163, and 220 are noisy and water vapor absorption bands that do not contain useful information for the class separability purpose. Hence, these bands were removed, and 200 bands were retained for the experiments [1]. This scene contains two-thirds agriculture and one-third forest or other natural perennial vegetation. The data are now publicly available on the Internet [44]. The crops were all in the early spring stage, with background residual ranging from “clean till” to “no till,” and were reported to exist in 16 classes. In our experiments, we consider a real situation in which most of the similar classes are included in the evaluations. Hence, 12 classes with an adequate number of labeled samples were selected for experiments. Table I shows the characteristics of the selected classes.

A subset of the Hyperion data set, with a size of 450 × 256 pixels, located in the Okavango Delta, Botswana (available on the Internet [45]), was also used in the experiments. The data set contains 14 different land-cover types consisting of seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the delta. In the subset, eight classes...
TABLE I
SAMPLES OF THE DATA SETS USED IN THE EXPERIMENTS

<table>
<thead>
<tr>
<th></th>
<th>Indian Pine</th>
<th></th>
<th>Botswana</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of Class</td>
<td># Samples</td>
<td></td>
<td>Name of Class</td>
<td># Samples</td>
</tr>
<tr>
<td>Wheat</td>
<td>212</td>
<td></td>
<td>Hippo Grass</td>
<td>101</td>
</tr>
<tr>
<td>Corn</td>
<td>234</td>
<td></td>
<td>Water</td>
<td>210</td>
</tr>
<tr>
<td>Bldg-Grass-Tree-Drives</td>
<td>380</td>
<td></td>
<td>Acacia Shrublands</td>
<td>248</td>
</tr>
<tr>
<td>Hay-windrowed</td>
<td>489</td>
<td></td>
<td>Firescar</td>
<td>259</td>
</tr>
<tr>
<td>Grass/Pasture</td>
<td>497</td>
<td></td>
<td>Mixed Mopane</td>
<td>268</td>
</tr>
<tr>
<td>Soybean-clean</td>
<td>614</td>
<td></td>
<td>Acacia Grasslands</td>
<td>305</td>
</tr>
<tr>
<td>Grass/Trees</td>
<td>747</td>
<td></td>
<td>Acacia Woodlands</td>
<td>314</td>
</tr>
<tr>
<td>Corn-min</td>
<td>834</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soybeans-notill</td>
<td>968</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woods</td>
<td>1294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn-notill</td>
<td>1434</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soybeans-min</td>
<td>2468</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10171</td>
<td></td>
<td>Total</td>
<td>1822</td>
</tr>
</tbody>
</table>

Fig. 3. Utilized data sets. (a) Hyperion Botswana scene along with different land-cover types of the Okavango Delta. (b) AVIRIS Indian Pine scene. (c) Different land-cover types present in Indian Pine.

with an adequate number of labeled samples were selected (Table I). The data from this region were collected by the NASA Earth Observing 1 (EO-1) satellite for the calibration/validation purpose of the mission in 2001. The Hyperion sensor on EO-1 acquires data at a 30-m spatial resolution over a 7.7-km strip in 242 bands, covering the 400–2500-nm portion of the spectrum in 10-nm windows. Uncalibrated and noisy bands that contain water absorption features had been previously removed, resulting in 145 bands [46].

Some sample bands of the data sets used in the experiments along with their ground truths are shown in Fig. 3. It is noteworthy that the experiments were conducted using the original and noise-whitened data sets of Indian Pine and Botswana.

C. Preprocessing

As mentioned before, the dimensionality of the PS can be obtained by the HySime method. Hence, the VDs of the AVIRIS and Hyperion data sets were determined to be 16 and 20, respectively, and were used as a reference for the dimensionality of the PS. In all of the following experiments, the AVIRIS and Hyperion data sets were clustered into 32 and 40 (i.e., $2 \times VD$) clusters, respectively, by the $K$-means algorithm [19]. Since the number of bands to be selected cannot be determined by the LP, G-FS, and PFS methods, they were set to 32 and 40 for the AVIRIS and Hyperion data sets, respectively. In contrast to these methods, the MTD algorithm is able to determine the number of bands to be selected in a straightforward manner without any preprocessing. Meanwhile, since the PS provides the capability of avoiding the shortcomings of the pixel space, the LP and G-FS methods were also conducted in the PS.

B. Preprocessing

As mentioned before, the dimensionality of the PS can be obtained by the HySime method. Hence, the VDs of the AVIRIS and Hyperion data sets were determined to be 16 and 20, respectively, and were used as a reference for the dimensionality of the PS. In all of the following experiments, the AVIRIS and Hyperion data sets were clustered into 32 and 40 (i.e., $2 \times VD$) clusters, respectively, by the $K$-means algorithm [19]. Since the number of bands to be selected cannot be determined by the LP, G-FS, and PFS methods, they were set to 32 and 40 for the AVIRIS and Hyperion data sets, respectively. In contrast to these methods, the MTD algorithm is able to determine the number of bands to be selected in a straightforward manner without any preprocessing. Meanwhile, since the PS provides the capability of avoiding the shortcomings of the pixel space, the LP and G-FS methods were also conducted in the PS.

The PFS and MTD methods along with the LP and G-FS methods were tested on the original data and were compared with each other. The proposed methods depend on image clustering in order to initialize. Therefore, to assess the effect of clustering on the reliability of the DR algorithms, both data sets were clustered ten times by the $K$-means clustering method. For each clustered image, the proposed FS methods were carried out. To evaluate the performance of all of the FS methods, 100 and 50 training samples were randomly chosen for each class in Indian Pine and Botswana data sets (Table I), respectively. The remaining samples were used as test ones. The overall accuracy (OA) and kappa coefficient (KC) of the quadratic discriminant classifier (QDC), linear discriminant classifier (LDC), $k$ nearest neighbor ($k$NN), and support vector machine (SVM) classifiers were used as performance measures. Moreover, to assess the OA and KC measures, the training samples were randomly chosen 30 times for each clustered image (i.e., 300 times for all 10 clusters). Meanwhile, the Indian Pine and Botswana data sets were classified only by $k$NN and SVM classifiers in full dimensionality (i.e., without FS) using the same training samples as in reduced dimension. The obtained results are given in Table II.

The learning curves that determine the maximum results (max), the results that correspond to optimal feature size for a given classifier in terms of OA, are shown in Fig. 4. The highest average results of OA and KC obtained by each classifier along with their standard deviations (SDs) and the optimal feature sizes for both data sets are included in Table II. As can be seen in the Indian Pine data set, PFS and MTD provide superior results using the QDC classifier. They, respectively, obtained 72.38% and 75.20% OAs with 1.80 and 1.58 SDs using 19 and 16 features. Put accurately, the proposed methods have better performance compared with LP and G-FS which achieved lower OAs and higher SDs. It is noteworthy that MTD selected 16 features for discriminating the 12 classes so that a conventional classifier (i.e., QDC) can show a better performance with a limited number of training samples (i.e., 100 samples for each class). As shown in Table II, QDC obtains the best results for the Indian Pine data set, and LDC does so for the Botswana data set. For the Indian Pine data set in full dimensionality, $k$NN demonstrates its superiority over the SVM classifier as the obtained OA is 69.73%. Meanwhile, the comparison between SVMs and QDC shows that the former
TABLE II
AVERAGE RESULTS OF FULL DIMENSIONALITY AND THE MAXIMUM RESULTS OF THE AVERAGE LEARNING CURVES ALONG WITH THEIR CORRESPONDING SDs AND OPTIMAL FEATURE SIZES (NUMBER OF FEATURES FOR EACH RESULT), OBTAINED BY EACH CLASSIFIER FOR THE ORIGINAL DATA SETS

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classifier</th>
<th>OA SD</th>
<th>KC SD</th>
<th># F</th>
<th>OA SD</th>
<th>KC SD</th>
<th># F</th>
<th>OA SD</th>
<th>KC SD</th>
<th># F</th>
<th>OA SD</th>
<th>KC SD</th>
<th># F</th>
<th>Full Dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pine</td>
<td>QDC</td>
<td>70.81</td>
<td>66.19</td>
<td>8</td>
<td>69.94</td>
<td>65.32</td>
<td>14</td>
<td>72.38</td>
<td>67.66</td>
<td>19</td>
<td>75.20</td>
<td>70.29</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANN</td>
<td>65.89</td>
<td>61.59</td>
<td>7</td>
<td>61.17</td>
<td>57.21</td>
<td>32</td>
<td>66.81</td>
<td>62.45</td>
<td>14</td>
<td>67.56</td>
<td>63.16</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LDC</td>
<td>63.55</td>
<td>59.41</td>
<td>27</td>
<td>64.00</td>
<td>59.83</td>
<td>31</td>
<td>67.29</td>
<td>62.91</td>
<td>18</td>
<td>65.47</td>
<td>61.20</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SVMs</td>
<td>64.35</td>
<td>60.14</td>
<td>30</td>
<td>64.77</td>
<td>60.42</td>
<td>32</td>
<td>68.30</td>
<td>63.94</td>
<td>23</td>
<td>68.82</td>
<td>64.35</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>63.08</td>
<td>53.21</td>
<td>30</td>
<td>64.53</td>
<td>63.69</td>
<td>32</td>
<td>2.98</td>
<td>3.14</td>
<td>2.84</td>
<td>2.95</td>
<td>2.95</td>
<td>3.04</td>
<td></td>
</tr>
</tbody>
</table>

OA: Overall Accuracy (%) ; KC: Kappa Coefficient (%) ; SD: Standard Deviation ; # F: Number of Features

has a significantly smaller OA and a larger SD. Moreover, for the Botswana data set and again in full dimensionality, the comparison between SVMs and LDC shows that the former has a little smaller OA and a larger SD.

In unsupervised FS methods, it is not possible to attain the learning curves. Consequently, the optimal feature size cannot be obtained. To tackle this deficiency, the VD of the data in the feature space \( \text{VD}_{fs} \) or \( 2 \times \text{VD}_{fs} \) can be used as a reference for the optimal feature size. It is considerable that the \( \text{VD}_{fs} \) is equivalent to the VD mentioned previously. In contrast, in the PFS method, the optimal feature size is equal to the VD in the PS \( \text{VD}_{ps} \). Furthermore, MTD inherently provides us with the optimal feature size to stop the algorithm (i.e., the number of bands corresponding to the last result). To show the efficiency of the proposed unsupervised FS methods, the OAs of LP, G-FS, PFS, and MTD are given in Table III. As can be seen, the number of selected features by MTD is close to \( \text{VD}_{fs} \) determined by HySime for both data sets. Furthermore, in the original Indian Pine data set, the last OA of MTD achieved by QDC is significantly better than the OAs of LP and G-FS in \( \text{VD}_{fs} \) and \( 2 \times \text{VD}_{fs} \) features. On the other hand, in the Botswana data set, the results of MTD are comparable with those of LP and G-FS. However, it is evident that PFS is stronger than LP and G-FS in \( \text{VD}_{fs} \) and \( 2 \times \text{VD}_{fs} \) dimensions in both original and noise-whitened data sets of Indian Pine and Botswana. Moreover, through the comparison of learning curves, it is observed in Table III that, in \( \text{VD}_{ps} \) features, PFS achieves higher accuracies than the maximum accuracies of LP and G-FS.

In this paper, the Friedman test [47], [48] is considered as an appropriate nonparametric statistical test for validating the FS methods based on multiple classifiers. The classifiers are used as judges (i.e., blocks in the test) to rank the FS methods (i.e., treatments in the test) independently and to evaluate whether the difference among the FS methods is statistically significant.

We performed the Friedman tests on the FS methods for all classification accuracies using the first-five, first-ten, etc., and also \( \text{VD}_{fs} \) and \( 2 \times \text{VD}_{fs} \) features. The \( p \)-values and mean ranks provided by the Friedman tests are given in Table IV. It is considerable that, for feature sizes greater than the optimal feature size achieved by MTD, this method does not participate in the Friedman test. For both data sets, the optimal feature size which refers to the last result achieved by MTD is 16 in this experiment (Fig. 4 and Table III). As shown in Table IV, MTD achieves remarkably better results in the Indian Pine data set. However, in most test cases, PFS adopts higher mean ranks, while MTD, LP, and G-FS receive lower mean ranks, respectively. For each data set, the \( p \)-values estimated by the Friedman test are (approximately) equal to zero. This states that the null hypothesis \( (H_0) \) is rejected for high confidence levels (e.g., 0.95 or 0.99), which means that the difference in the behavior among the FS methods is statistically significant.

The comparison of Fig. 4 and Table IV shows a good matching. This means that, for a given number of features, an FS method with higher accuracies obtained by all/most of the classification methods takes a higher mean rank in the Friedman test. For example, according to Table IV, for the Indian Pine Data set, using 16 features (i.e., the \( \text{VD}_{fs} \) of the data), the methods MTD, PFS, G-FS, and LP take the highest to lowest mean ranks, respectively. This ranking is confirmed by the accuracies demonstrated in Fig. 4(a)–(d) and Table III for each FS method, i.e., the MTD method achieves the highest OAs, and LP achieves the lowest OAs in all/most cases (i.e., classification methods).

D. Experiment on Noise-Whitened Data

The conditions in which remotely sensed images are acquired (e.g., atmospheric and environmental conditions) inevitably make them noisy. The noise component varies from band to band. High levels of noise could even make a band more distinct than adjacent bands. Noise whitening could help us have more realistic bands [10]. A noise-whitening process can
be conducted via noise estimation, but it is not a simple task to estimate the noise of an image [10]. Nevertheless, it might be estimated via a multiple regression theory based approach [49], [50], which outperforms the shift difference method [46]. Therefore, after removing the bad bands, noise whitening was applied to the rest of the bands. To evaluate the effect of noise whitening on hyperspectral data, the same experiments were again conducted using the noise-whitened data set through the previous experimental setup. It should be noted that the noise-whitened data were used only for FS processes, while the classifications were carried out on original data. The results are given in Fig. 5 and Table V, showing the average of results for 300 classifications performed by QDC, LDC, \(k\)NN, and SVM classifiers. These tests and the classification results point out the acceptable effect of noise whitening on FS. Comparing the results between full and reduced dimensionalities shows that the OA obtained by SVMs is 9.19% less than the OA obtained by QDC in the Indian Pine data set. The OA achieved by SVMs in full dimensionality is almost close to the one obtained by LDC in reduced space for both data sets. Although this achievement by SVMs is remarkable, it is noteworthy that the SDs of SVMs are bigger than those of other classifiers in all data sets. The QDC and LDC yield proper results for the Indian Pine and Botswana data sets, respectively. Once more, the PFS and MTD algorithms achieve better performance in contrast to LP and G-FS in the Indian Pine data set. The PFS also demonstrates an acceptable performance in the Botswana data set. Furthermore, all of the FS methods profit from noise whitening. In particular, the proposed PFS and MTD methods achieved higher accuracies in the Indian Pine data set. The former gained 0.66% improvement with 14 features, and the latter gained 3.37% improvement with 15 features. In this data set, noise whitening improved the accuracies of the LP and G-FS methods up to 1.94% and 2.57% compared with the first experiment. In addition, the optimal feature sizes increased from 8 to 13 and 14 to 26, respectively. The experimental results show slight improvements in the Botswana data set. Nevertheless, the OAs and SDs of all methods show remarkable improvements in both data sets.
**TABLE III**  
COMPARISON BETWEEN OAS OF THE FS METHODS IN THE OPTIMAL FEATURE SIZES DETERMINED BY THE SUPERVISED AND UNSUPERVISED APPROACHES

<table>
<thead>
<tr>
<th>Method</th>
<th>Indian Pine (Classifier: QDC)</th>
<th>Botswana (Classifier: LDC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dataset</td>
<td># F</td>
</tr>
<tr>
<td>LP</td>
<td>OD</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>NWD</td>
<td>16</td>
</tr>
<tr>
<td>VD_5</td>
<td>OD</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>NWD</td>
<td>16</td>
</tr>
<tr>
<td>2VD_5</td>
<td>OD</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>NWD</td>
<td>16</td>
</tr>
<tr>
<td>MTD</td>
<td>OD</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>NWD</td>
<td>19</td>
</tr>
<tr>
<td>PFS</td>
<td>OD</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>NWD</td>
<td>16</td>
</tr>
</tbody>
</table>

OD: Original Data; NWD: Noise-Whitened Data  
# F: Number of Features; OA: Overall Accuracy (%)

**TABLE IV**  
p-VALUES AND MEAN RANKS OF THE FOUR FS METHODS ACHIEVED BY THE FREIDMAN TEST, USING THE ORIGINAL DATA, BASED ON THE FIRST-FIVE, FIRST-TEN, ETC., AND ALSO VD_{fs} AND 2VD_{fs} FEATURES

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Indian Pine (VD_{fs}=10)</th>
<th>Botswana (VD_{fs}=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td># F</td>
<td>p-value</td>
<td>Friedman</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>603.409</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>856.756</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>332.249</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>321.934</td>
</tr>
<tr>
<td>25</td>
<td>9.50E-08</td>
<td>223.868</td>
</tr>
<tr>
<td>30</td>
<td>2.02E-08</td>
<td>275.095</td>
</tr>
<tr>
<td>40</td>
<td>1.85E-27</td>
<td>376.540</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>662.309</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>610.815</td>
</tr>
<tr>
<td>15</td>
<td>1.89E-27</td>
<td>607.239</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>421.622</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>425.803</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>463.423</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>369.482</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>386.687</td>
</tr>
</tbody>
</table>

# F: Number of Features

Evidently, the noise-whitening task shows tendency to increase the OA in all methods. Taking note of Tables II and V and subplots of learning curves (Figs. 6 and 7), it can be illustrated that PFS and MTD, while showing a better performance than G-FS and LP, demonstrate less reaction to noise whitening. It is noteworthy that, in the original and noise-whitened data sets of Indian Pine, the MTD results exceed the maximum results of LP and G-FS attained by learning curves.

Moreover, the Friedman test was performed on the noise-whitened data sets. The p-values achieved by the tests are (approximately) equal to zero, which again (as in Section IV-C) demonstrates that the difference among the FS methods is statistically significant. The p-values and mean ranks provided by the Friedman tests are given in Table VI. As shown, in the noise-whitened data sets, PFS again takes higher ranks (i.e., owns the best performance) and keeps its superiority in most of the feature spaces.

Similar to the comparison made in the previous experiment (i.e., Section IV-C) between Fig. 4 and Table IV, in this experiment, Fig. 5 and Table VI have a good matching.

### E. Experiment on the Synthetic Data

In the previous experiments, unsupervised FS methods were conducted on real images. In this experiment, we aim at evaluating the FS methods in a controlled environment. Therefore, the unsupervised FS performance is compared with the supervised FS in a synthetic image with respect to the signal-to-noise ratio (SNR) and the number of distinct features. A synthetic image consisting of $100 \times 100$ mixed pixels was simulated by the code provided in [51]. In order to generate the image, nine spectra were selected from the USGS spectral library according to the cuprite mineral materials in NV, available in [52]. Then, the spectra were mixed based on the Dirichlet distribution. Fig. 8 shows the spectra of the endmembers having 224 spectral bands. After the synthetic image was generated, it was corrupted by a Gaussian additive noise in order to reach SNRs of $20:1$ and $30:1$.

The VDs (i.e., VD_{fs}) of the synthetic images in both SNRs were determined to be 9 by the HySime method. Moreover, the optimal number of features significant to the problem in the clean original synthetic image was obtained by a search strategy (i.e., SFS) which optimized the OAs of LDC. The results obtained by the SFS method are given in Fig. 9. As shown, LDC obtained 99.02% OA using 8 features. This synthetic image with a known number of features and classes is used to evaluate the proposed unsupervised FS methods.

The results of the unsupervised FS methods conducted on the synthetic data sets are given in Table VII. In general, the OAs were improved by increasing the SNR in the data sets for all FS methods. Moreover, in all FS methods except MTD, the optimal feature sizes decrease by increasing SNR. As shown, the number of optimal features obtained by PFS is equal to $2VD_{fs}$ (i.e., 18 features), with OAs of 86.90% and 87.62% in low- and high-SNR data sets, respectively. However, the number of optimal features obtained by PFS is more than the number of known distinct features. Meanwhile, the selected feature numbers provided by PFS are not necessarily the same as the known distinct features in the data sets. Moreover, in the high-SNR data set, the OA obtained by PFS using 18 (i.e., $2VD_{fs}$) features in an unsupervised manner is 11.40% less than the OA obtained by SFS using 8 features. On the other hand, SVM and $k$NN classifiers demonstrate almost the same performance in dealing with synthetic image in full dimensionality. The maximum OAs obtained by both classifiers in full dimensionality are smaller than the one obtained by LDC in reduced space.

In this part of this paper, the effect of FS as a preprocessing step is assessed on both computational cost decrease and accuracy improvement of the classifiers. The obtained results in the reduced and full dimensionality spaces are compared together and are given in Table VIII. As shown, the computational cost for the $k$NN classifier is almost the same as that of other FS
methods. In addition, in the synthetic data set, its accuracy is comparable with the best result by PFS. Nevertheless, in the Indian Pine data set, $k$NN in full dimensionality achieved a poor result in comparison with QDC in reduced space. As shown, the computational cost for the SVM classifier conducting on both data sets in full dimensionality is higher than that of the FS methods. Therefore, the FS can reduce the computational cost. On the other hand, the OA provided by QDC in reduced space improved remarkably by 10.14% in the Indian Pine data set. Although in some cases SVMs show better performance in dealing with high dimension space, it suffers from large values of SDs. In contrast to SVMs, some simple classifiers like QDC with small SDs demonstrate a better performance in reduced space.

V. CONCLUSION

In this paper, two effective unsupervised FS methods called PFS and MTD have been proposed based on the PS. The PS provides them with the capability of avoiding the shortcomings of the pixel space used by conventional methods. PFS proposes a way for selecting the initial feature based on the orthogonal distance from the PS diagonal. It attempts to choose the most informative band as the first feature, i.e., the one being farthest away from the space diagonal. Then, the other distinct features are selected by the OSP measure. MTD uses two criteria defined by the tangent of the angles between band vectors in the PS to describe band correlations. For the PFS method, the optimal feature size is estimated via the HySime method in the PS, while in MTD, it is determined as an embedded process. These methods were evaluated using two real data sets and a synthetic one and were compared with the LP and G-FS methods.

In the first experiment, the optimal features were selected for the two original data sets. According to the results obtained by learning curves in both data sets, the PFS method performed better than the LP and G-FS methods in terms of the OA and KC measures. Moreover, compared to LP and G-FS, MTD demonstrated a much better performance in the Indian Pine data set and an acceptable performance in the Botswana data set.
set. To evaluate the efficiency of the unsupervised FS methods, the OAs and the optimal feature sizes, which were obtained in an unsupervised fashion by learning curves, were compared together on both data sets. The comparisons showed that the PFS method achieved better results in both supervised and unsupervised fashions. In particular, the OAs of PFS obtained with \( VD_{ps} \) features were better than the maximum OAs of LP and G-FS achieved by learning curves. The optimal feature sizes achieved by MTD were close to the number of classes reported in both data sets. The OA of MTD, corresponding to its optimal feature size in an unsupervised fashion, was much better than the OAs of LP and G-FS in the Indian Pine data set, where \( VD_{ps} \) or \( 2VD_{ps} \) features were used. Similar comparison results were attained in the Botswana data set for both LP and G-FS methods, where LP demonstrated a better result with \( 2VD_{ps} \) bands. In order to evaluate the efficiency of the FS methods, the case of applying the full dimensional data was also considered in the experiments. Both real data sets were classified only by \( k \)NN and SVM classifiers in full dimensionality. In the Indian Pine data set, the comparison between SVMs in full dimensionality and QDC in reduced space showed that the former had a significantly smaller OA and a larger SD. Almost similar results were attained for the Botswana data set. In addition to the comparisons mentioned previously, the Friedman tests were conducted on both data sets, which showed that the difference among the FS methods were statistically significant.

The second experiment was conducted by using the noise-whitened data sets. The Friedman test on these data sets once again demonstrated similar results as in the original data sets. Once more, the PFS algorithm reached higher accuracies in
Fig. 7. Average learning curves achieved by the LDC classifier for the original and noise-whitened Botswana data sets. (a) LP. (b) G-FS. (c) PFS. (d) MTD.

TABLE VI
*p*-VALUES AND MEAN RANKS OF THE FOUR FS METHODS ACHIEVED BY FRIEDMAN TEST, USING THE NOISE-WHITENED DATA, BASED ON THE FIRST-FIVE, FIRST-TEN, ETC., AND ALSO BASED ON VD_{fs} AND 2VD_{fs} FEATURES

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Mean Ranks of the Four FS Methods</th>
<th>Friedman p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botswana (FD=20)</td>
<td>LP G-FS PFS MTD</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>716.89 199.770 559.368 925.963</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>779.158 150.595 787.010 685.237</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>532.184 168.451 749.641 951.724</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>444.439 189.354 717.707 -</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>546.914 226.547 758.039 -</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>359.695 381.738 610.068 -</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>44.06e-278 270.097 433.211 648.193</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>684.447 282.997 969.305 501.252</td>
<td></td>
</tr>
</tbody>
</table>

In the third experiment, a synthetic image with a known number of features and classes was used to evaluate the proposed unsupervised FS methods in two scenarios: SNR = 20 and SNR = 30. The experimental results showed that the OA obtained by PFS using 18 (i.e., 2VD) features in an unsupervised manner was 11.40% less than the OA obtained by SFS using 8 features (i.e., the optimal feature size determined in a supervised fashion). Furthermore, for the SVM classifier in the Indian Pine and synthetic data sets in full dimensionality, the computational cost was higher than that of conducting FS and classification processes.

In sum, PFS demonstrated better results in an unsupervised fashion in comparison to the situation in which learning curves were used for the LP and G-FS methods. Meanwhile, the OAs comparison with other methods in both data sets. Furthermore, PFS and MTD profited from noise whitening as they achieved higher accuracies with less number of features in the Indian Pine data set. In this data set, noise whitening improved the accuracy of the LP and G-FS methods as the optimal feature size increased. Nevertheless, an improvement in the OAs of all methods and a decrease in their SDs were remarkable in the data set. However, the experimental results did not show significant improvements in the Botswana data set.
of MTD obtained in an unsupervised manner were almost the same as the maximum OAs obtained by the learning curves.

### REFERENCES


Mohsen Ghahary Asl received the B.S. degree in geodesy and geomatics engineering and the M.S. degree in remote sensing from K.N. Toosi University of Technology, Tehran, Iran, in 2004 and 2006, respectively, where he is currently working toward the Ph.D. degree in remote sensing, conducting research on unsupervised dimensionality reduction of hyperspectral imagery.

His research interests lie in the areas of remote sensing (i.e., multispectral and hyperspectral data analysis and applications), image processing, and pattern recognition.

Mohsen Ghahary Asl received the B.S. degree in physics from Tehran University, Tehran, Iran, in 1975, the M.S. degree from Toledo University, Toledo, OH, USA, in 1980, and the Ph.D. degree in remote sensing from James Cook University of North Queensland, Townsville, Qld., Australia, in 1995.

He is currently a Researcher with K.N. Toosi University of Technology, Tehran. He has more than 150 publications. His research interests include modeling and applied and environmental remote sensing. His current interests are spectroradiometry of vegetation and information extraction from hyperspectral data.

Barat Mojaradi received the B.S. degree in geodesy and geomatics engineering from Tabriz University, Tabriz, Iran, in 1998 and the M.S. and Ph.D. degrees in remote sensing from K.N. Toosi University of Technology, Tehran, Iran, in 2000 and 2008, respectively.

He is currently an Assistant Professor with the Iran University of Science and Technology, Tehran. He was a visiting Ph.D. student with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands, from February to August 2007. He is a reviewer of several national and international journals. His special research interests are the development of image processing algorithms for multispectral and hyperspectral optical data and the development of models for land surface processes.

Dr. Mojaradi was the recipient of ISPRS06 Innovative Solving Problem award in The Netherlands.