Classification of Hyperspectral Data From Urban Areas Based on Extended Morphological Profiles

Jón Atli Benediktsson, *Fellow, IEEE*, Jón Aevar Palmason, *Student Member, IEEE*, and Johannes R. Sveinsson, *Senior Member, IEEE*

***Abstract*—Classification of hyperspectral data with high spatial resolution from urban areas is investigated. A method based on mathematical morphology for preprocessing of the hyperspectral data is proposed. In this approach, opening and closing morphological transforms are used in order to isolate bright (opening) and dark (closing) structures in images, where bright/dark means brighter/darker than the surrounding features in the images. A morphological profile is constructed based on the repeated use of openings and closings with a structuring element of increasing size, starting with one original image. In order to apply the morphological approach to hyperspectral data, principal components of the hyperspectral imagery are computed. The most significant principal components are used as base images for an extended morphological profile, i.e., a profile based on more than one original image. In experiments, two hyperspectral urban datasets are classified. The proposed method is used as a preprocessing method for a neural network classifier and compared to more conventional classification methods with different types of statistical computations and feature extraction.**

***Index Terms*—Hyperspectral remote sensing data, morphological profiles, neural networks, principal components.**

I. Introduction

HIGH-RESOLUTION hyperspectral remote sensing data from urban areas have recently become available. Such data provide both detailed structural and spectral information about urban scenes. Therefore, the data should be useful for information and classification. However, the classification of hyperspectral urban data is a challenging problem for two main reasons. First, the hyperspectral data contain a lot of information about the spectral properties of the land cover in the data, but no spatial information is inherent in the spectral data. Second, the individual images from an urban scene contain spatial information but very limited information about the spectral nature of the data. Consequently, a joint spectral/spatial classifier is needed for classification of urban hyperspectral data, but few such approaches have been proposed. In this paper, we propose such an approach, i.e., a morphological method that is based on making use of both the spectral and spatial information for classification.

Manuscript received April 7, 2004; revised September 3, 2004. This work was supported in part by the Icelandic Research Council and in part by the Research Fund of the University of Iceland. An early version of this paper was presented at the IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data—A Workshop Honoring Prof. David A. Landgrebe, Oct. 27–28, NASA Goddard Visitor Center, Washington, DC, USA, 2003.

The authors are with the Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is; jaep@hi.is; sveinsso@hi.is).

Digital Object Identifier 10.1109/TGRS.2004.842478

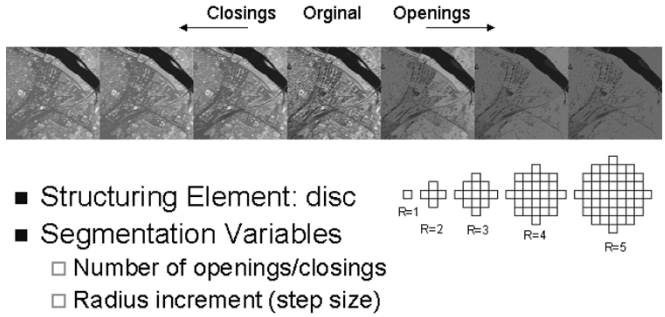


Fig. 1. Morphological profile based on a circular structuring element, three openings, and three closings. In the shown profile, circular structural elements with $R = 2, 4$, and 6 were used.

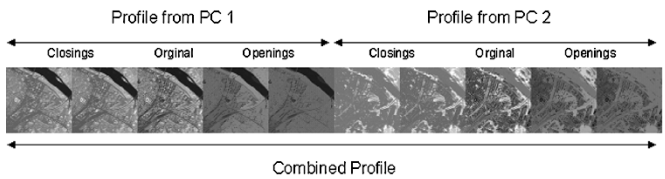


Fig. 2. Combined morphological profile for two PCs with two openings and two closings for each PC. In the combined profile, circular structural elements with $R = 2$ and 4 were used.

Classification of urban hyperspectral data has been discussed in some recent papers where the primary focus has often been on either advanced pixel classifiers [1], [2] or feature extraction [3]. In addition, approaches based on spectral signature recognition in urban areas have recently been proposed [4], [5]. Most such approaches do not consider the spatial content of the data, the reason being that the resolution of hyperspectral data usually has been low. Other papers on analysis of urban data have addressed spectral mixture analysis for hyperspectral data [6] or the identification of spectral signatures in mixed pixels [7]. Such approaches are important in urban areas, since the spatial scale of urban objects usually is on the order of a few meters.

The proposed approach is based on mathematical morphology [8], [9], [19] and has recently been used successfully in classification of panchromatic urban data. Pesaresi and Benediktsson [10] used a composition of geodesic opening and closing operations of different sizes in order to build a morphological profile and a neural network approach for the classification of features. A potential problem with approaches based on morphological profiles is that these methods create a large feature set from one original image by applying a series of opening and closing transforms. Although the use of morphological profiles should help in creating an image feature

Fig. 3. Pavia data. Three-channel color composite of the area used for classification.

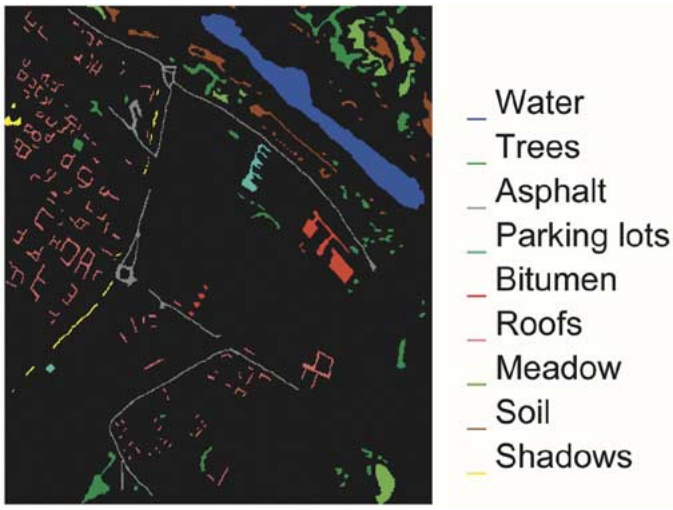


Fig. 4. Pavia data. The available reference data.

TABLE I
PAVIA DATA. INFORMATION CLASSES AND SAMPLES

Class	Name	Training Samples	Test Samples
1	Shadows	202	4083
2	Roofs	205	2302
3	Parking lots	206	1136
4	Asphalt	205	1301
5	Trees	204	1630
6	Meadows	201	132
7	Soil	315	2041
8	Bitumen	202	491
9	Water	119	159
Total		1859	13275

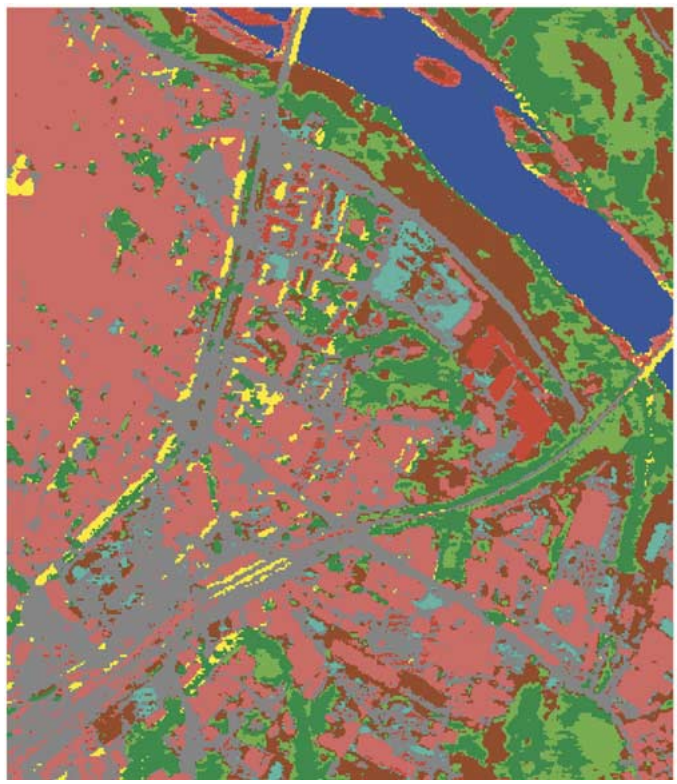


Fig. 5. Pavia data. Classification map obtained by maximum-likelihood classification of seven DAFE features based on original statistics. The overall accuracy of test data was 97.2% in this case.

set that is more effective in discrimination of different urban features, a lot of redundancy will be evident in the feature set. Therefore, feature extraction can be used in finding the most important features in the feature space. In [11], Benediktsson *et al.* extended the method in [10] and successfully included decision boundary feature extraction to reduce the redundancy in the morphological profile.

In both [10] and [11], the approaches were applied on panchromatic remote sensing data of high spatial resolution. It is of interest to extend the approaches in [10] and [11] for classification of hyperspectral urban data. Recently, Plaza *et al.* [26] proposed an extension of the approach. Their approach looks at both the spectral and spatial information in the imagery and is based on a *morphological endmember selection*, which integrates spectral/spatial information from the hyperspectral data. That means for hyperspectral data, that a characteristic image needs to be extracted from the data. Another much simpler approach was suggested in [12], i.e., using only the first principal component (PC) of hyperspectral image data to build a morphological profile.

In this paper, the proposed extended morphological profile method is compared to statistical classification methods based on different statistical computations and feature extraction methods that have previously been applied in classification of

TABLE II
PAVIA DATA. OVERALL TEST ACCURACIES IN PERCENTAGE FOR STATISTICAL APPROACHES WITH DIFFERENT TYPES OF STATISTICS COMPUTATIONS AND FEATURE EXTRACTION METHODS

Statistics	Feature Extraction	No. of Features	Overall Accuracy (%)
Original	No	40	95.9
LOOC	No	40	96.7
Enhanced	No	40	92.8
Original	DAFE	7	97.2
Original	DBFE	20	96.6
Original	NWFE	5	95.8

TABLE III
PAVIA DATA. CLASS-SPECIFIC TEST ACCURACIES IN PERCENTAGE FOR STATISTICAL APPROACHES WITH DIFFERENT TYPES OF STATISTICS COMPUTATIONS AND FEATURE EXTRACTION METHODS: OS, LOOC, AND ES

Class	Name	OS	LOOC	ES	DAFE with OS	DBFE with OS	NWFE with OS
1	Shadows	97.4	97.3	95.9	100.0	99.9	99.7
2	Roofs	98.2	97.2	88.6	96.6	95.2	98.2
3	Parking lots	94.5	93.9	88.5	96.3	95.1	98.1
4	Asphalt	97.1	97.7	95.5	93.3	97.2	96.2
5	Trees	91.4	94.4	80.1	94.5	87.2	90.7
6	Meadow	84.8	78.0	99.2	72.7	93.9	85.6
7	Soil	100.0	99.9	100.0	99.7	99.0	100.0
8	Bitumen	95.7	96.5	92.5	99.2	91.0	96.3
9	Water	39.0	76.1	78.6	83.0	58.5	44.7
	AVE	88.5	92.4	91.0	92.8	90.8	89.9
	OA	95.9	96.7	92.8	97.2	95.8	96.6

hyperspectral data from urban areas. The paper is organized as follows. In Section II, the proposed mathematical morphology approach to classification of hyperspectral data from urban areas is discussed. Feature extraction is discussed in Section III and statistical parameter estimation in Section IV. Experimental results are given in Section V and conclusions drawn in Section VI.

II. MORPHOLOGICAL PROFILES FOR HYPERSPECTRAL DATA

Recent theoretical advances in mathematical morphology, such as definitions of *leveling* and *morphological spectrum*, form a theoretical framework, which is used for the formal definition of the morphological profiles [8], [14]. Standard morphological segmentation approaches are based on an edge-detection phase (watershed line extraction on a gradient image), but here we consider a morphological segmentation method, which avoids the gradient calculation, and can be applied either to single-scale or multiscale image processing problems.

Watershed line detection [15] is the main tool of mathematical morphology used for image segmentation. Watershed segmentation was introduced in image analysis by Beucher and Lantuéjoul [16] and defined mathematically by both Meyer [17] and Najman and Schmitt [18]. However, except for a few simple cases where the target object is brighter than the background or vice versa, watershed segmentation cannot be applied directly.

TABLE IV
PAVIA DATA. EIGENVALUES OF THE PRINCIPAL COMPONENTS

	Value	$\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}$
λ_1	2.27×10^5	78.2 %
λ_2	5.22×10^4	18.0 %
λ_3	8.39×10^3	2.9 %
λ_4	1.16×10^3	0.4%
$\lambda_5+\ldots+\lambda_{40}$		$<0.5\%$

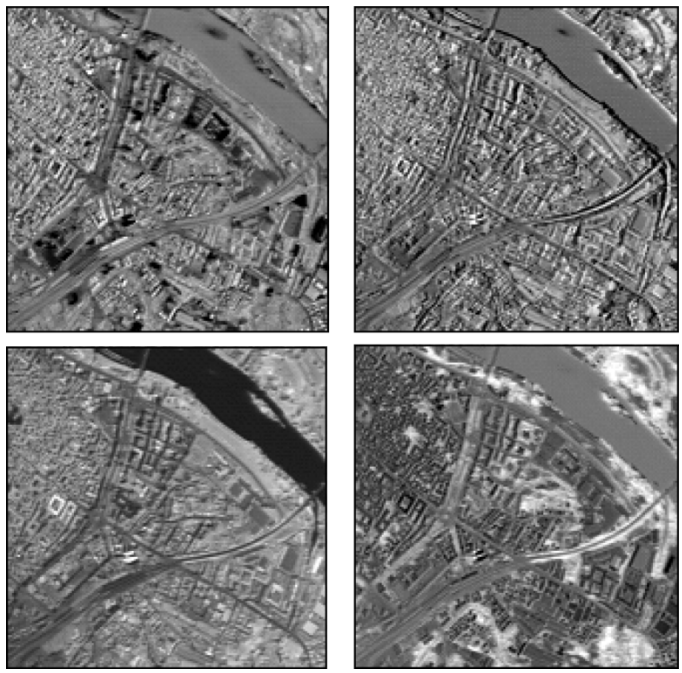


Fig. 6. Pavia data. The first four PCs: (lower left) first component, (lower right) second component, (upper left) third component, and (upper right) fourth component.

TABLE V
PAVIA DATA. OVERALL TEST ACCURACIES IN PERCENTAGE FOR NEURAL NETWORK CLASSIFICATION AFTER MORPHOLOGICAL PROCESSING OF THE FIRST AND SECOND PRINCIPAL COMPONENTS

# openings = # closings	# of PCs	Feature Extraction	# of Features	Overall Accuracy(%)
0	1	No	1	48.8
4	1	No	9	72.6
8	2	No	18	93.9
8	2	DAFE	8	79.4
8	2	DBFE	13	83.4
8	2	NWFE	14	89.4

Generally, the method is applied to images that have been transformed by a gradient-like operator based on the measure of the local slope of the gray-level function. Watershed extraction generally involves the thinning of the gradient image with a homotopic transformation, and the detection of basins as regions and crest lines as boundaries of these regions. For these reasons, the watershed approach generally leads to finding the features of

TABLE VI
PAVIA DATA. CLASS-SPECIFIC TEST ACCURACIES IN PERCENTAGE FOR NEURAL NETWORK CLASSIFICATION OF MORPHOLOGICAL PROFILES

Index	Name	PC1	Morphological Profile Based on PC1	Morphological Profile Based on PC1 and PC2	Morphological Profile Based on PC1 and PC2 with DAFE	Morphological Profile Based on PC1 and PC2 with DBFE	Morphological Profile Based on PC1 and PC2 with NWFE
1	Shadows	100.0	100.0	100.0	100.0	100.0	100.0
2	Roofs	0.0	25.9	93.4	82.2	70.0	82.5
3	Parking lots	0.0	94.5	92.8	0.0	84.6	89.0
4	Asphalt	0.0	2.5	88.8	55.6	77.4	68.8
5	Trees	0.0	82.2	95.8	91.8	79.4	94.6
6	Meadow	0.0	94.0	100.0	86.4	89.4	97.7
7	Soil	94.0	89.8	79.0	86.5	74.8	86.4
8	Bitumen	96.1	85.7	81.5	73.3	74.5	82.3
9	Water	0.0	84.3	68.6	70.4	66.7	87.4
	AVE	32.2	62.8	89.1	62.2	79.6	87.6
	OA	48.8	72.6	93.9	79.5	83.4	89.4

an image based on an edge-detection strategy. This characteristic of watershed segmentation is responsible for the difficulty in applying this technique in cases where the sensor resolution approaches the size of the objects in the scene.

The fundamental operators in mathematical morphology are erosion and dilation. When mathematical morphology is used in image processing, these operators are applied to an image with a set of a known shape, called a structuring element (SE). The application of the erosion operator to an image gives an output image, which shows where the SE fits the objects in the image. On the other hand, the application of the dilation operator to an image gives an output image, which shows where the SE hits the objects in the image. The erosion and dilation operators are in general dual but noninvertible. All other morphological operators can be expressed in terms of erosion and dilation.

Two commonly used morphological operators are opening and closing. The idea behind opening is to dilate an eroded image in order to recover as much as possible of the eroded image. In contrast, the idea behind closing is to erode a dilated image in order to recover the initial shape of image structures that have been dilated. The filtering properties of the opening and closing operators are based on the fact that not all structures from the original image will be recovered when these operators are applied. It is a common practice to use the opening and closing transforms in order to isolate bright (opening) and dark (closing) structures in images, where bright/dark means brighter/darker than the surrounding features in the images.

In order to isolate features with a thinner support than a given SE, a widely used technique is to take the residuals of the opening, closing, and original images, by a morphological transformation called *top-hat* and *inverse top-hat* (or bot-hat) [8]. The chosen approach for the opening and closing calculation should use a non-Euclidean metric known as filtering by reconstruction [20], [21]. The reason for using the reconstruction approach is that this family of morphological filters has proven to have a better shape preservation than classical morphological filters. In fact, reconstruction filters introduce less shape noise, since an interaction between the shape of the structures present in the image and the shape of the structuring element is used in the filtering.

Some structures may have a high response for a given SE size, and a lower response for other SE sizes, depending on the interaction between the SE size and the size of the structure. Sometimes, the size of the structures that we want to detect is known exactly. However, that is often not possible, and then a single-SE-size approach appears to be too simplistic. For these reasons, in exploratory or more complex cases, it is often a good idea to use a multiscale approach based on a range of different SE sizes. This can allow us to explore a range of different hypothetical spatial domains and to use the best response of the structures in the image for the classification process (see Fig. 1).

Given the above-proposed notion of morphological characteristic, it is straightforward to extend the same concept to multiscale processing, by introducing the concepts of *morphological profile* and of the *derivative of the morphological profile* (DMP).

A. Definition

Let $\Pi\gamma(x)$ be the *opening profile* at the point x of the image I defined as a vector

$$\Pi\gamma(x) = \{\Pi\gamma_\lambda\colon \Pi{\gamma^*}_\lambda(x), \qquad \forall\lambda \in [0,\ldots,n]\} \quad (1)$$

and let $\Pi\phi(x)$ be the *closing profile* at the point x of the image I defined as a vector

$$\Pi\phi(x) = \{\Pi\phi_\lambda\colon \Pi{\phi^*}_\lambda(x), \qquad \forall\lambda \in [0,\ldots,n]\} \quad (2)$$

with ${\gamma_0}^*(x) = {\phi_0}^*(x) = I(x) \Rightarrow \Pi\gamma_0(x) = \Pi\phi_0(x) = I(x)$ for $\lambda = 0$ by definition of opening and closing by reconstruction. Given (1) and (2), the opening profile can also be defined as a granulometry made with opening by reconstruction, while the closing profile can be defined as antigranulometry made with closing by dual reconstruction. The derivative of the morphological profile is defined as a vector where the measure of the slope of the opening–closing profile is stored for every step of an increasing SE series.

When the morphological profile approach is applied to hyperspectral data, a characteristic image needs to be extracted from the data. As stated above, it was suggested in [12] to use the first PC of the hyperspectral data for such a purpose. Although that approach seems reasonable because principal component analysis is optimal for data representation in the mean square sense, it should not be forgotten that with only one PC, the hyperspectral data are reduced from potentially several hundred data channels into one single data channel. In addition, although the first

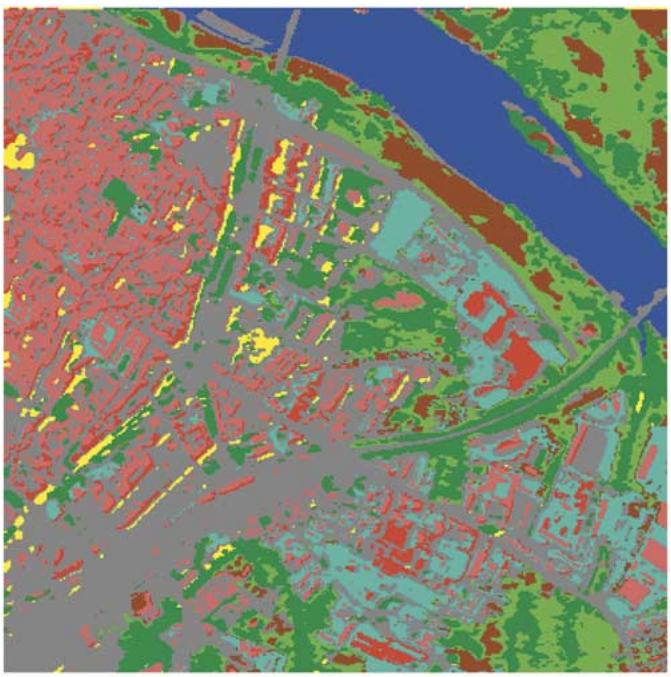


Fig. 7. Pavia data. A classification map obtained by applying a neural network classifier on an extended morphological profile with two PCs, four openings, and four closings. The neural network classifier was applied to 18 features. The overall accuracy of the test data was 93.9% in this case.

PC may represent most of the variation in the image, some important information may be contained in the other PCs. Therefore, we propose an extension to the approach in [12] and build an extended morphological profile from several different PCs. For example, we could decide to use the PCs that account for around 90% of the total variation in the image. If two PCs contain 90% of the variation, the extended profile would be a double morphological profile where the first profile would be based on the first PC and the second profile on the second PC (see Fig. 2). The extended profile would then be considered a single stacked vector to be used in neural network classification. Obviously, the computations will be more extensive because of this approach. On the other hand, better information should be extracted from the hyperspectral data than for the simple approach proposed in [12]. Also, some redundancies should be observed for the extended morphological profile. Therefore, feature extraction may be important. Feature extraction is the topic of the next section.

III. Feature Extraction

Feature extraction can be viewed as finding a set of vectors that represents an observation while reducing the dimensionality. In pattern recognition, it is desirable to extract features that are focused on discriminating between classes. Although a reduction in dimensionality is desirable, the error increment due to the reduction in dimension has to be without sacrificing the discriminative power of classifiers. In linear feature extraction, the number of input dimensions corresponds to the number of selected eigenvectors [27]. The transformed data are determined by $Y = \Phi^T X$, where Φ is the transformation matrix composed of the eigenvectors of the feature matrix, X is the data in the original feature space, and Y is the transformed data in the new feature space. Several feature extraction approaches have been proposed for remote sensing data [27].

Fig. 8. Washington, DC data. Three-channel color composite of the area used for classification.

A. Discriminant Analysis Feature Extraction (DAFE)

Discriminant analysis is a method that is intended to enhance separability. A within-class scatter matrix Σ_W and a between-class scatter matrix Σ_B are defined in [22] and [27]. The criterion used for optimization of separability may be defined as

$$J = \mathrm{tr}\left(\Sigma_W^{-1}\Sigma_B\right) \tag{3}$$

where $\mathrm{tr}(\,)$ denotes the trace of a matrix. New feature vectors are selected to maximize the criterion.

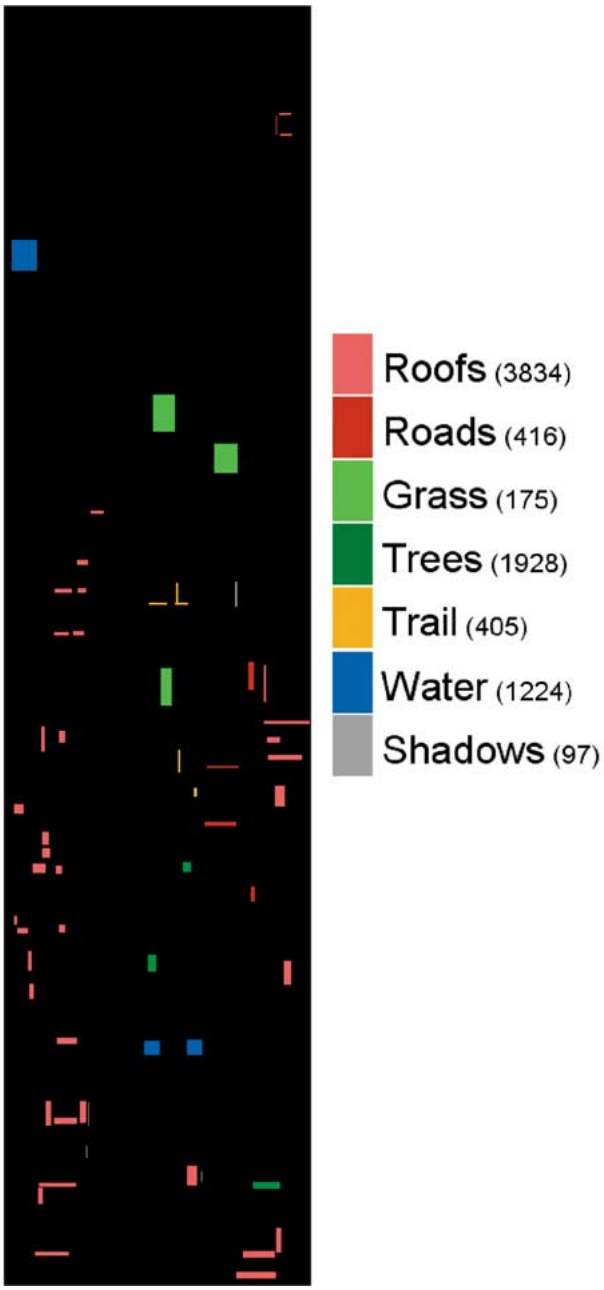


Fig. 9. Washington, DC data. The available reference data. The number of available reference samples for each class is given in parantheses.

Fig. 10. Washington, DC data. Classification map obtained by maximum-likelihood classification of five NWFE features based on original statistics. The overall accuracy of test data was 99.8% in this case.

The necessary transformation from X to Y is found by taking the eigenvalue–eigenvector decomposition of the matrix $\Sigma_W^{-1}\Sigma_B$ and then taking the transformation matrix as the normalized eigenvectors corresponding to the eigenvalues in a decreasing order. For this method, the maximum rank of Σ_B is $K-1$ for a K-class problem, which indicates that at maximum $K-1$ features can be extracted by this approach [22].

B. Decision Boundary Feature Extraction (DBFE)

Lee and Landgrebe [30] showed that both discriminately informative features and discriminately redundant features can be extracted from the decision boundary itself. They also showed that discriminately informative feature vectors have a component that is normal to the decision boundary at least at one point

TABLE VII
WASHINGTON, DC DATA. INFORMATION CLASSES AND SAMPLES

Class	Name	Training Samples	Test Samples
1	Roofs	40	3794
2	Road	40	376
3	Grass	40	135
4	Trees	40	1888
5	Trail	40	365
6	Water	40	1184
7	Shadows	40	57
Total		280	7779

on the boundary. Further, discriminately redundant feature vectors are orthogonal to a vector normal to the decision boundary at every point on the boundary. In [30], a decision boundary feature matrix (DBFM) was defined to extract discriminately informative features and discriminately redundant features from the decision boundary. The rank of DBFM is not dependent on the number of classes.

C. Nonparametric Weighted Feature Extraction (NWFE)

Although the DBFE overcomes some of the problems with the DAFE, it does have its limitations, i.e., it can be computationally intensive, and because it is based on the training samples, it does not perform well for small numbers of training samples [22]. To assume the advantages of both the DBFE and DAFE and mitigate their limitations, Kuo and Landgrebe [3] proposed nonparametric weighted feature extraction. In particular, NWFE improves DAFE by focusing on samples near the eventual decision boundary location, and different weights are put on every sample to compute the "local means" and defining new nonparametric between-class and within-class scatter matrices to get more features. The NWFE can give as many features as in the original feature space and has been shown to perform well in experiments [22].

IV. Statistics Estimation in Multispectral Analysis

Supervised statistical classification relies heavily on the available training samples, which are used to estimate the mean vectors and the covariance matrices for the individual classes [27]. If the mean vectors and the covariance matrices are only based on the available training data (called *original statistics*), problems may be observed in hyperspectral classification because of the *Hughes phenomenon* [22]. There are several approaches that have been proposed to overcome the Hughes phenomenon in hyperspectral classification. Such methods include *enhanced statistics* and leave one out (LOO) statistics, which are discussed next.

A. Statistical Enhancement of Remotely Sensed Data

A well-trained classifier must successfully model the distribution of the entire dataset, but the modeling must be done in such a way that the different classes of interest are as distinct from one another as possible. Shahshahani and Landgrebe [29] accomplished the modeling by an iterative calculation based on both the training samples and a systematic sampling of all the pixels in the scene. Their method is based on the EM algorithm and is called *enhanced statistics*. In the method, the statistics are adjusted or *enhanced* so that, while still being defined by the training samples, the collection of class conditional statistics better fit the entire dataset. This amounts to a hybrid supervised/unsupervised training scheme.

B. LOO Statistics

With limited training samples, discriminant functions that do not contain many parameters may provide improved classification results even though they result in simpler decision boundaries. Hoffbeck and Landgrebe [28] proposed a covariance estimator that provides this advantage. This covariance estimator examines mixtures of the sample covariance, diagonal sample covariance, common covariance, and diagonal common covariance [28] This optimization criterion is referred to as the leave one out method, and the obtained covariance is referred to as leave one out covariance (LOOC).

TABLE VIII
Washington, DC Data. Overall Test Accuracies in Percentage for Statistical Approaches With Different Types of Statistics Computations and Feature Extraction Methods

Statistics	Feature Extraction	No. of Features	Overall Accuracy (%)
Original	No	38	97.3
LOOC	No	38	99.5
Enhanced	No	38	86.6
Original	DAFE	6	96.4
Original	DBFE	9	97.3
Original	NWFE	5	99.8

V. Experimental Results

Two hyperspectral datasets from urban areas were used in classification. One dataset is from the city of Pavia, Italy. The other dataset is from Washington, DC. In both cases, the proposed morphological preprocessing method was applied along with a backpropagation neural network classifier with one hidden layer. The number of neurons in the hidden layer is computed as the geometric average of the number of inputs and outputs. The number of openings (and closings) varied for the experiments, and different radius increments were also tested for the SE. The results are compared to statistical Gaussian ML classification based on different types of statistics computations and feature extraction. The statistics methods applied are original statistics, LOOC statistics and enhanced statistics [13]. The considered feature extraction approaches are DAFE, DBFE [30], and NWFE [22]. A feature set based on the 99% variance was selected for the DBFE and NWFE methods, but a feature set based on 95% variance was selected for the DAFE approach. MultiSpec [23] was used for the statistical classification but Matlab for the morphological preprocessing and neural network classification. The results from the experiments are discussed below.

A. Results for Data From Pavia, Italy

The data in this experiment are very fine resolution hyperspectral data, part of the records of four flight lines over the urban area of Pavia, in northern Italy [13]. The flight was done in the framework of the HySens project, managed by Deutsches Zentrum fuer Luft- und Raumfahrt (the German Aerospace Center) and sponsored by the European Union within the transnational access to major research infrastructures (Contract no. HPRI-CT-1999-00075). The urban area was imaged by means of the Digital Airborne Imaging Spectrometer (DAIS). The flight altitude was chosen as the lowest available for the airplane, which resulted in a spatial resolution of 2.6 m. The lines were chosen so that the images are partially overlapping,

TABLE IX
WASHINGTON, DC DATA. CLASS-SPECIFIC TEST ACCURACIES IN PERCENTAGE FOR STATISTICAL APPROACHES WITH DIFFERENT TYPES OF STATISTICS COMPUTATIONS AND FEATURE EXTRACTION METHODS: OS, LOOC, AND ES

Class	Name	OS	LOOC	ES	DAFE	DBFE	NWFE
1	Roofs	99.7	99.3	73.4	98.6	99.3	99.9
2	Road	75.3	98.7	97.9	95.5	82.7	98.1
3	Grass	93.3	100.0	100.0	99.3	99.3	100.0
4	Trees	99.9	100.0	99.9	98.0	100.0	100.0
5	Trail	79.2	99.2	98.6	90.7	79.5	99.2
6	Water	99.2	99.9	99.3	89.4	99.7	99.9
7	Shadow	64.9	100.0	77.2	82.5	84.2	96.5
	AVE	87.4	99.6	92.3	93.4	92.1	99.1
	OA	97.3	99.5	86.6	96.4	97.7	99.8

Fig. 11. Washington, DC data. The first four principal components: (left) first component to the (right) fourth component.

TABLE X
WASHINGTON, DC DATA. EIGENVALUES OF PRINCIPAL COMPONENTS

	Value	$\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}$
λ_1	4.51×10^7	66.0%
λ_2	2.15×10^7	31.5%
λ_3	1.19×10^6	1.7 %
$\lambda_4+\ldots+\lambda_{39}$		$< 0.9\%$

TABLE XI
WASHINGTON, DC DATA. OVERALL TEST ACCURACIES IN PERCENTAGE FOR NEURAL NETWORK CLASSIFICATION AFTER MORPHOLOGICAL PROCESSING OF THE FIRST AND SECOND PRINCIPAL COMPONENTS

# openings = # closings	# of PCs	Feature Extraction	# of Features	Overall Accuracy(%)
0	1	No	1	49.4
6	1	No	13	87.2
6	2	No	26	98.9
6	2	DAFE	5	92.9
6	2	DBFE	20	96.8
6	2	NWFE	14	98.5

which allows studying the effect of the directional reflectance of urban materials on mapping accuracy [24]. In the experiments, it was decided to take into account only bands of reflective energy and thus skip the seven thermal infrared bands. Also, the middle infrared bands above 1.958-μm wavelengths were skipped because of heavy noise in those bands. Therefore, only the first 40 (out of 80) spectral bands of DAIS (visible and near-infrared bands, from 0.496–1.756 μm) are considered in the experiments. The test site is the middle part of the DAIS 7915 image swath 1 over center of the city of Pavia, Italy (see Fig. 3). Its size is 400×400 pixels. The test site comprises a number of different land cover classes, from a dense residential area on one side of the river to open areas and meadows on the other side. In detail, the land cover classes that can be differentiated are: water, trees, asphalt, parking lot cover, bitumen, brick roofs, meadows, bare soil, and shadows. Ground truth information provided by the University of Pavia is available for all these classes (see Table I and Fig. 4).

The test accuracies for the applied statistical approaches are shown in Tables II and III. The Gaussian ML classifier was applied in all cases. Table II summarizes the results in terms of the number of features used. The highest overall classification accuracy was achieved when DAFE was used (seven data channels). The classification map for this best case is shown in Fig. 5. The performance of the classification of the DAFE and NWFE features is noteworthy. In classification of small feature sets (seven and five features, respectively), similar or higher overall test accuracies are achieved as for classification of the full feature space (40 data channels). The classification of 20 DBFE also gives similar overall accuracies for 20 features. It is interesting to note that enhanced statistics do not work very well in the full feature space. The main reason for this is that some noise was observed in the data although most of the noisy channels had been removed.

PCs were computed from the hyperspectral data. The results for the eigenvalues are shown in Table IV, and the first four principal component images are shown in Fig. 6. From Fig. 6, it is clear that the individual principal component images provide different information. In Table IV, the left column gives the component number, the center column the eigenvalues for the components, and the right column the percentage for the eigenvalue of the total eigenvalue sum. By looking at the right column, it is seen that 96.2% of the total eigenvalue sum is contained in the first two components, whereas the first component had 78.2% of the total. Therefore, it can be expected that the first two PCs are needed in the morphological preprocessing of the hyperspectral data (the upper two images in Fig. 6).

For the morphological processing, a circular SE with a step size increment of 2 gave the best results. Four openings and four

TABLE XII
WASHINGTON, DC DATA. CLASS-SPECIFIC TEST ACCURACIES IN PERCENTAGE FOR NEURAL NETWORK CLASSIFICATION OF MORPHOLOGICAL PROFILES

Index	PC1	Morphological Profile Based on PC1	Morphological Profile Based on PC1 and PC2	Morphological Profile Based on PC1 and PC2 with DBFE	Morphological Profile Based on PC1 and PC2 with NWFE
1	75.0	96.3	98.0	93.7	97.1
2	99.5	98.9	99.2	100.0	100.0
3	100.0	100.0	100.0	100.0	100.0
4	99.5	100.0	100.0	100.0	100.0
5	90.1	98.9	99.5	99.5	98.6
6	99.9	99.9	99.9	100.0	100.0
7	96.5	93.0	96.5	100.0	100.0
AVE	94.4	98.1	99.0	99.0	99.4
OA	87.2	98.9	98.9	96.8	98.5

closings were computed for each PC. The test accuracies for the neural network classification of the morphological profiles are shown in Tables V and VI. From Table V, it is seen that the best overall accuracy was obtained when an extended morphological profile based on the first two principal components was used with no feature extraction, resulting in an overall test accuracy of 93.9%. Also, the use of two PCs in an extended profile resulted in far higher accuracies than when only one PC was used for classification, regardless of the feature extraction method used. With feature extraction, the best performance was seen with the NWFE. When no morphological processing was done on a single PC, the accuracy was poor as expected, i.e., only about 56%.

The class-specific accuracies for the morphological approach are shown in Table VI. There it can be seen that the classification of the morphological profile based on two PCs gave higher class-specific accuracies for most classes than the classification approach based on the use of one PC. In particular, the method that uses two PCs and no feature extraction outperformed the one-PC profile significantly in terms average classification accuracy. The classified map for the morphological profile based on two PCs without feature extraction is shown in Fig. 7. Details of the given urban structures are evident in Fig. 7. For both Figs. 5 and 7, it must be kept in mind that they are based on very limited training data. More training data would be needed to classify the whole area satisfactorily.

B. Results for Data From Washington, DC

In the second experiment, analysis was done on a Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne hyperspectral data flightline over the Washington, DC Mall (see Fig. 8). Two hundred and ten bands were collected in the 0.4–2.4-μm region of the visible and infrared spectrum. The water absorption bands were then deleted, resulting in 189 channels. In order to reduce the feature set further, every fifth data channel was selected from the 189 channels. The analyzed feature set contained 38 data channels. The dataset contains 1280 scan lines with 307 pixels in each scan line. The Washington, DC dataset has been studied extensively by Kuo and Landgrebe [3], [22], [25] and Dundar and Landgrebe [1]. The dataset is available in the student CD-ROM of [22]. The MultiSpec project file provided on the student CD-ROM was used to select training samples in our experiments.

There are seven information classes in the Washington, DC data. The information classes and training and test samples are listed in Table VII and shown in Fig. 9. As can be seen from Table VII, only 40 samples per class were used here to train the classifier. All other samples were used to test the neural classifiers. The low number of training samples was expected to have a significant effect on the classification accuracies in the experiments.

The test accuracies for the applied statistical approaches are shown in Tables VIII and IX. The Gaussian ML classifier was applied in all cases. Table VIII summarizes the results in terms of the number of features used. As in Experiment 1, the performance of the ML classifier when applied to the NWFE feature set is noteworthy. Here, the NWFE outperforms all other feature extraction methods in terms of overall test accuracies when the ML classifier is applied. The classification map for the ML classification of the NWFE feature set is shown in Fig. 10. In a similar way as in Experiment 1, enhanced statistics do not show a good performance in the full feature space. Here, the problem is mostly related to the individual class specific accuracy of the largest class (roofs).

As in the experiment on the Pavia data, PCs were computed from the hyperspectral data. The first four principal component images are shown in Fig. 11. Again, striking visual differences are seen between the individual PCs. The eigenvalues for the individual components are given in Table X. By looking at the right column of Table X, it is seen that 97.5% of the total eigenvalue sum is contained in the first two components but only 66.0% in the first component. Therefore, we expect the first two PCs to be needed in the morphological preprocessing of the hyperspectral data.

In the morphological processing of the Washington, DC data, a circular SE with a step size increment of 2 gave the best results. Six openings and six closings were computed for each PC. The test accuracies for the neural network classification of the morphological profiles are shown in Tables XI and XII. From

Fig. 12. Washington, DC data. A classification map obtained by applying a neural network classifier on an extended morphological profile with two PCs, four openings, four closings, and no feature extraction. The neural network classifier was applied to 26 input features. The overall accuracy of the test data was 98.9% in this case.

Fig. 13. Washington, DC data. A classification map for obtained by applying a neural network classifier on an extended morphological profile with two PCs, four openings, four closings, and nonparametric weighted feature extraction. The neural network classifier was applied to 14 NWFE features. The overall accuracy of the test data was 98.5% in this case.

Table XI, it can be seen that the best overall and average accuracies were obtained when the first two principal components were used in an extended morphological profile with no feature extraction, giving an overall test accuracy of 98.5%. The use of 14 NWFE features also gave a similar overall accuracy. These classification results can be considered outstanding, especially in light of the fact that only 40 training samples were used for each class. Also, the use of two PCs in an extended profile resulted in higher accuracies than the use of only one PC, regardless of whether feature extraction war used or not. When no morphological processing was done on a single PC, the overall accuracy was only 49.4%, as expected.

The class-specific accuracies for the morphological approach are shown in Table XII. It is clear from the table that the use of two PCs resulted in increased average accuracies over the method based on one PC. Also, the use of DBFE with the morphological profile outperformed the DAFE, but both feature extraction methods gave good accuracies. The classified map for two PCs and no feature extraction is shown in Fig. 12, and the comparable result with NWFE feature extraction is given in Fig. 13.

VI. Conclusion

Classification of hyperspectral data from urban areas has been discussed. A method was proposed, which is based on using several principal components from the hyperspectral data, building a morphological profile for each of the PCs and using them all together in one extended morphological profile. The extended morphological profiles were then classified with a neural network, with and without feature extraction. In experiments on two datasets, the proposed approach was compared to a Gaussian ML classifier with several different feature extraction and statistics estimation methods. The proposed approach performed well in terms of accuracies and was comparable in accuracies to the Gaussian ML classifier, especially when decision boundary feature extraction was applied on the extended morphological profile.

In this paper, principal component analysis was chosen for dimension reduction because it is both simple and gives optimal representation of data in the mean squared sense. Several different dimension reduction techniques can be used to reduce the hyperspectral data. Such approaches include, for example, independent component analysis [31] and projection pursuit [22]. A topic of future research is to investigate the use of other dimensionality reduction techniques to create characteristic images for the morphological profiles.

The excellent performance of the proposed morphological approach in experiments is interesting because it mostly uses spatial information instead of the rich spectral information available in the hyperspectral data. In our current research, we are working toward making more use of the spectral information by fusing spectral and spatial classification results. Such fusion could be important for pixels which should be classified based on spectral rather than spatial information, e.g., vegetation inside urban areas.

Acknowledgment

The authors would like to thank P. Gamba (University of Pavia, Italy) for providing the reference data for the Pavia dataset. The Washington, DC data were obtained from the student CD-ROM which accompanies D.A. Landgrebe's book, *Signal Theory Methods in Multispectral Remote Sensing* [22].

References

[1] M. M. Dundar and D. A. Landgrebe, "Toward an optimal supervised classifier for the analysis of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 271–277, Jan. 2004.

[2] ——, "A model-based mixture-supervised classification approach in hyperspectral data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 12, pp. 2692–2699, Dec. 2002.

[3] B.-C. Kuo and D. A. Landgrebe, "A robust classification procedure based on mixture classifiers and nonparametric weighted feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2486–2494, Nov. 2002.

[4] L. Fiumi, "Evaluation of MIVIS hyperspectral data for mapping covering materials," in *Proc. IEEE/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Areas*, Rome, Italy, Nov. 2001, pp. 324–327.

[5] S. Roessner, K. Segl, U. Heiden, and H. Kaufmann, "Automated differentiation of urban surfaces based on airborne hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1525–1532, Jul. 2001.

[6] S. Bhaskar and B. Datt, "Sub-pixel analysis of urban surface materials. A case study of perth, W. Australia," in *Proc. IGARSS*, vol. 4, Honolulu, HI, Jul. 2000, pp. 1535–1537.

[7] S. Roessner, K. Segl, U. Heiden, and H. Kaufmann, "Analysis of spectral signatures of urban surfaces for their identification using hyperspectral HyMap data," in *Proc. IEEE/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Areas*, Rome, Italy, Nov. 2001, pp. 173–177.

[8] P. Soille, *Morphological Image Analysis—Principles and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

[9] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2042–2055, Sep. 2002.

[10] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Mar. 2001.

[11] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[12] J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Morphological transformations and feature extraction for urban data with high spectral and spatial resolution," in *Proc. IGARSS*, vol. 1, Toulouse, France, 2003, pp. 470–472.

[13] F. Dell'Acqua, P. Gamba, and A. Ferrari, "Exploiting spectral and spatial information for classifying hyperspectral data in urban areas," in *Proc. IGARSS*, vol. 1, Toulouse, France, 2003, pp. 464–466.

[14] M. Pesaresi and J. A. Benediktsson, "Image segmentation based on the derivative of the morphological profile," in *Mathematical Morphology and Its Applications to Image and Signal Processing*, J. Goustsias, L. Vincent, and D. S. Bloomberg, Eds. Norwell, MA: Kluwer, 2000.

[15] F. Meyer and S. Beucher, "Morphological segmentation," *J. Vis. Commun. Image Represent.*, vol. 11, pp. 21–46, 1990.

[16] S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection," presented at the *Int. Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation*, Rennes, France, Sep. 1979.

[17] F. Meyer, "Integrals, gradients and watershed lines," in *Proc. Mathematical Morphology and Its Applications to Signal Processing*, Barcelona, Spain, May 1993, pp. 70–75.

[18] L. Najman and M. Schmitt, "Definition and some properties of the watershed of a continuous function," in *Proc. Mathematical Morphology and Its Applications to Signal Processing*, Barcelona, Spain, May 1993, pp. 75–81.

[19] J. Serra, *Image Analysis and Mathematical Morphology*. London, U.K.: Academic, 1982.

[20] J. Serra and P. Salembier, "Connected operators and pyramids," in *Proc. SPIE Conf. Non-Linear Algebra and Morphological Image Processing*, vol. 2030, San Diego, CA, 1993, pp. 65–76.

[21] J. Crespo, J. Serra, and R. Schafer, "Theoretical aspects of morphological filters by reconstruction," *Signal Process.*, vol. 47, pp. 201–225, 1995.

[22] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.

[23] C. Lee and D. A. Landgrebe, "Decision boundary feature extraction for neural networks," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 75–83, Jan. 1997.

[24] B.-C. Kuo and D. A. Landgrebe, "Improved statistics estimation and feature extraction for hyperspectral data classification," School Elect. Comput. Eng., Purdue Univ., West Lafayette, IN, Tech. Rep. TR-ECE 01-6, Nov. 2001.

[25] ——, "A covariance estimator for small sample size classification problems and its application to feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 814–819, Apr. 2002.

[26] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A new method for target detection in hyperspectral imagery based on extended morphological profiles," in *Proc. IGARSS*, vol. 6, Toulouse, France, Jul. 21–25, 2003, pp. 3772–3774.

[27] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. London, U.K.: Academic, 1990.

[28] J. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 763–767, Jul. 1996.

[29] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087–1095, Sep. 1994.

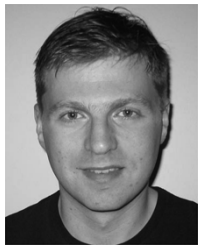
[30] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 388–400, Apr. 1993.

[31] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

Jón Atli Benediktsson (S'86–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1984, 1987, and 1990, respectively.

He is currently a Professor of electrical and computer engineering at the University of Iceland. He has held visiting positions at the School of Computing and Information Systems, Kingston University, Kingston upon Thames, U.K. (1999–2004), the Joint Research Centre of the European Commission, Ispra, Italy (1998), Denmark's Technical University (DTU), Lyngby (1998), and the School of Electrical and Computer Engineering, Purdue University (1995). His research interests are in pattern recognition, neural networks, remote sensing, image processing, and signal processing, and he has published extensively in those fields. He was a Fellow at the Australian Defence Force Academy (ADFA), Canberra, in August of 1997. From 1999 to 2004, he was Chairman of the energy company Metan Ltd. His research interests are in remote sensing, pattern recognition, neural networks, image processing, and signal processing, and he has published extensively in those fields.

Dr. Benediktsson received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, and in 2000, he was granted the IEEE Third Millennium Medal. He is Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGARS) and was an Associate Editor TGARS from 1999 to 2002. He coedited (with Professor David A. Landgrebe) a Special Issue on Data Fusion of TGARS (May 1999). In 2002, he was appointed Vice President of Technical Activities in the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) and (with Paolo Gamba and Graeme G. Wilkinson) a Special Issue on Urban Remote Sensing from Satellite (October 2003). From 1996 to 1999, he was the Chairman of GRSS' Technical Committee on Data Fusion and was elected to the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) for the term 2000 to 2002, and in 2002, he was appointed Vice President of Technical Activities of GRSS. He was the founding Chairman of the IEEE Iceland Section and served as its Chairman from 2000 to 2003. He is the Chairman of the University of Iceland's Science and Research Committee (from 1999), a member of Iceland's Science and Technology Council, and a member of the Nordic Research Policy Council. He was a member of a NATO Advisory Panel of the Physical and Engineering Science and Technology Sub-Programme (2002–2003). He is a member of Societas Scinetiarum Islandica and Tau Beta Pi.



Jón Aevar Palmason (S'03) was born in Akranes, Iceland, in 1976. He received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, in 2000. He is currently pursuing the M.S. degree from the University of Iceland.

Since 2000, he has been with Og Vodafone (formerly Islandssimi), Reykjavik, an Icelandic telecommunication company. His research interests are pattern recognition, remote sensing, and signal processing.

Johannes R. Sveinsson (S'86–M'90–SM'02) received the B.S. degree from the University of Iceland, Reykjavik, and the M.S. (Eng.) and Ph.D. degrees from Queen's University, Kingston, ON, Canada, all in electrical engineering.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, University of Iceland. From 1981 to 1982, he was with the Laboratory of Information Technology and Signal Processing, University of Iceland. He was a Visiting Research Student with the Imperial College of Science and Technology, London, U.K. from 1985 to 1986. At Queen's University, he held teaching and research assistantships and received Queen's Graduate Awards. From November 1991 to 1998, he was with the Engineering Research Institute, University of Iceland, as a Senior Member of the research staff and a Lecturer in the Department of Electrical Engineering, University of Iceland. His current research interests are in systems and signal theory.

Dr. Sveinsson is a member of SIAM.