Classification and Feature Extraction for Remote Sensing Images From Urban Areas Based on Morphological Transformations

Jon Atli Benediktsson, *Senior Member, IEEE*, Martino Pesaresi, and Kolbeinn Arnason

***Abstract*—Classification of panchromatic high-resolution data from urban areas using morphological and neural approaches is investigated. The proposed approach is based on three steps. First, the composition of geodesic opening and closing operations of different sizes is used in order to build a differential morphological profile that records image structural information. Although, the original panchromatic image only has one data channel, the use of the composition operations will give many additional channels, which may contain redundancies. Therefore, feature extraction or feature selection is applied in the second step. Both discriminant analysis feature extraction and decision boundary feature extraction are investigated in the second step along with a simple feature selection based on picking the largest indexes of the differential morphological profiles. Third, a neural network is used to classify the features from the second step. The proposed approach is applied in experiments on high-resolution Indian Remote Sensing 1C (IRS-1C) and IKONOS remote sensing data from urban areas. In experiments, the proposed method performs well in terms of classification accuracies. It is seen that relatively few features are needed to achieve the same classification accuracies as in the original feature space.**

***Index Terms*—Classification, mathematical morphology, feature extraction, feature selection, high-resolution imagery.**

I. Introduction

IT IS WELL KNOWN that there are two fundamentally different strategies for image segmentation, i.e., edge detection and region growing. Even though the standard approach to segmentation based on mathematical morphology [1] is dependent on edge detection, it is possible to consider a different morphological approach to the segmentation problem. In this paper, we define image features by their morphological intrinsic characteristics, instead of using their boundary [2], [3]. The method is focused on image structural information, and this structural information is collected by applying morphological operators with a multiscale approach and by looking at the residues between the multiscale morphologically transformed image and the original one.

Manuscript received October 18, 2002; revised March 21, 2003. This work was supported in part by the Icelandic Research Council and the Research Fund of the University of Iceland.

J. A. Benediktsson is with the Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

M. Pesaresi is with the Digital Mapping Sector, Environmental Technologies and Services Area, INFORM Srl, 35129 Padova, Italy (e-mail: PesaresiM@Informsrl.it).

K. Arnason is with the National Land Survey of Iceland, 300 Akranes, Iceland (e-mail: kolbeinn@lmi.is).

Digital Object Identifier 10.1109/TGRS.2003.814625

The fundamental operators in mathematical morphology are *erosion* and *dilation* [1]. 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 dual but noninvertible, in general. 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) [2]. Here, the chosen approach for the opening and closing calculation uses a non-Euclidean metric known as filtering by reconstruction [4]. 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 nominally no shape noise, since the shape of the structuring element used in the filtering are adaptive with respect to the structures present in the image itself [5].

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 we know exactly the size of the structures that we want to detect. 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 can be 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 hy-

pothetical spatial domains, and to use the best response of the structures in the image for the classification process.

The idea to use a composition of opening transforms for a morphological segmentation of satellite data was originally proposed for the detection of different urban structures [6], [7]. In the experiments in [6] and [7], segmentation labels were obtained after the arithmetic addition of a series of openings with an increasing SE. The method is only applicable to Boolean maps (binary or two-gray-level images), and it does not use geodesic metric. More recently, Pesaresi and Kanellopoulos [8] used a composition of geodesic opening and closing operations of different sizes in order to build a *morphological profile*. Then, they used a neural network approach for the classification of features. In [3], Pesaresi and Benediktsson proposed a method, which is based on *differential morphological profiles* (DMPs), which is both more general and more robust than the methods in [6]–[8]. Here, the method in [3] will be applied in classification of urban remote sensing data.

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 set which is more effective in discrimination of different urban features, a lot of redundancy will be evident in the feature set. Therefore, it is of interest to see if feature extraction or feature selection can help in finding the most important features in the feature space and if similar classification accuracies can be achieved with a reduced feature set as in the original feature space. Here, we will investigate the use of two feature extraction methods, discriminant analysis feature extraction and decision boundary feature extraction for neural networks, and a simple approach based on sorting the indexes of morphological profiles. After the feature extraction or selection, a conjugate gradient neural network with one hidden layer will be used to classify the data.

The paper is organized as follows. In Section 2, differential morphological profiles will be reviewed. In Section 3, feature extraction and feature selection will be discussed. Experimental results will be given in Section 4 and conclusions drawn in Section 5.

II. Differential Morphological Profiles

Here, we will review the concepts of the *morphological profile* and of the *derivative of the morphological profile* (DMP) [3]. Both of these concepts are used to create a feature vector from a single image, I. Both are based on the repeated use of the opening and closing operators, which are commonly used in mathematical morphology [1].

Let γ_λ^* be a morphological *opening operator by reconstruction* using structuring element SE $=\lambda$ and $\Pi\gamma(x)$ be the *opening profile* at the pixel x of the image I. $\Pi\gamma(x)$ is defined as a vector

$$\Pi\gamma(x) = \{\Pi\gamma_\lambda : \Pi\gamma_\lambda = \gamma_\lambda^*(x), \forall\lambda \in [0, n]\}. \quad (1)$$

Also, let φ_λ^* be a morphological *closing operator by reconstruction* using structuring element SE $=\lambda$. Then, the *closing profile* $\Pi\varphi(x)$ at pixel x of the image I is defined as the vector

$$\Pi\varphi(x) = \{\Pi\varphi_\lambda : \Pi\varphi_\lambda = \varphi_\lambda^*(x), \forall\lambda \in [0, n]\}. \quad (2)$$

In the above, $\Pi\gamma_0(x) = \Pi\varphi_0(x) = I(x)$ for $\lambda = 0$ by the definition of opening and closing by reconstruction [3]. Given (1) and (2), the opening profile can also be defined as a granulometry [1] 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.

The *derivative of the opening profile* $\Delta\gamma(x)$ is defined as the vector

$$\Delta\gamma(x) = \{\Delta\gamma_\lambda : \Delta\gamma_\lambda = |\Pi\gamma_\lambda - \Pi\gamma_{\lambda-1}|, \forall\lambda \in [1, n]\}. \quad (3)$$

By duality, the *derivative of the closing profile* $\Delta\varphi(x)$ is the vector

$$\Delta\varphi(x) = \{\Delta\varphi_\lambda : \Delta\varphi_\lambda = |\Pi\varphi_\lambda - \Pi\varphi_{\lambda-1}|, \forall\lambda \in [1, n]\}. \quad (4)$$

Generally, the *derivative of the morphological profile* $\Delta(x)$ or the DMP can be written as the vector

$$\Delta(x) = \left\{\Delta_c : \left\langle \begin{array}{l} \Delta_c = \Delta\varphi_{\lambda=n-c+1}, \forall c \in [1, n] \\ \Delta_c = \Delta\gamma_{\lambda=c-n}, \forall c \in [n+1, 2n] \end{array} \right.\right\} \quad (5)$$

with n equal to the total number of iterations, $c = 1, \ldots, 2n$, and $|n - c|$ = the size of the morphological transform. Let us take a closer look at (5). Near the central position of the DMP vector, c, we have the response for the derivative calculated using small SEs, while at the beginning (position $c = 1$) and at the end (position $c = 2n$), we record the response for the greatest SEs in the closing and opening profiles, respectively. Therefore, if we observe a "centered" DMP, we can argue that small structures are present in the image. On the other hand, an unbalanced DMP (either on the left or right side) indicates the presence of larger structures. These larger structures are either darker [high response in the closing profile $\Delta\varphi(x)$] or lighter [high response in the opening profile $\Delta\gamma(x)$] than the surroundings. Generally speaking, the signal response recorded in the DMP gives information about the size and the type of the structures in the image by observing the placement of the area of the DMP histogram. The size $\mathrm{s} = |n - c|$ is the distance form the center of the DMP, while the type of a structure (darker or lighter than the sourrounding ones) can be argued by observing on which side it is placed in the histogram of the DMP vector values. While the above-mentioned approaches do not require a particular metric for the morphological transforms, the DMP approach requires the use of granulometry and antigranulometry made by opening and closing by reconstruction, using a geodesic metric [1].

Examples of differential morphological profiles for four information classes are given in Fig. 1 for a panchromatic IKONOS image from Reykjavik, Iceland. From Fig. 1, it can be seen that strikingly different characteristics are achieved for the different classes. Therefore, the use of the differential profiles should help in discrimination. However, a lot of redundancy can be seen in the profiles, i.e., the information from all the

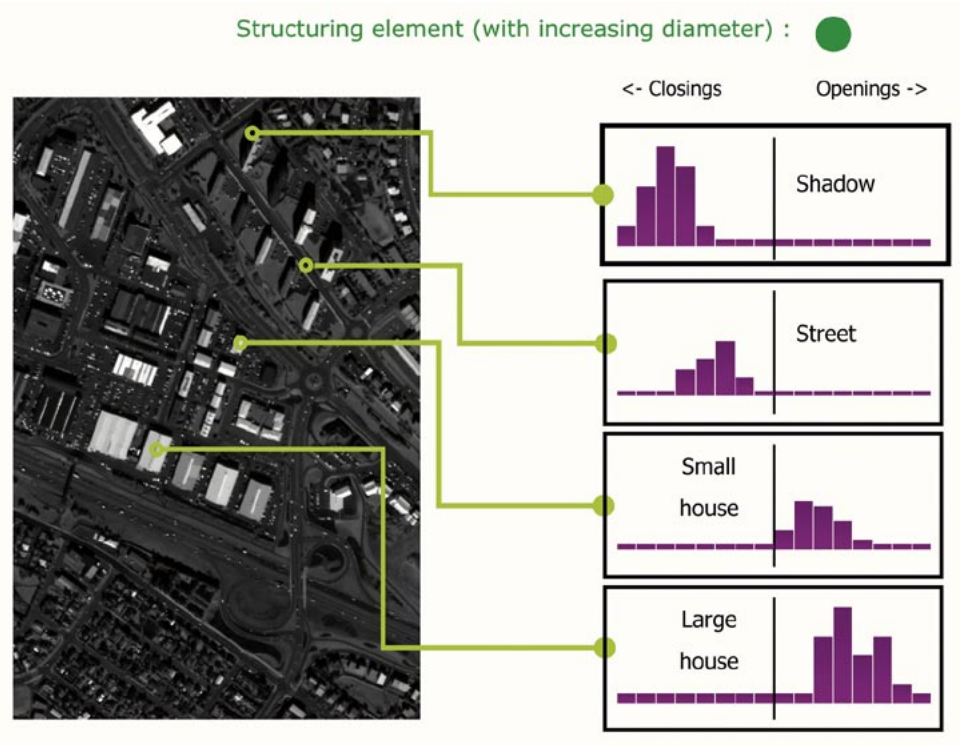

Fig. 1. Differential profiles for four information classes (Shadow, Street, Small house, and Large house) from an IKONOS image from Reykjavik, Iceland. A circular morphological structuring element was used with an increasing diameter. The original image is a 975×639 pixel subsample from an IKONOS frame with full 1-m spatial resolution. The image was acquired on August 9, 2001, when the sun elevation was 42°. The relatively low sun elevation has the effect that shadows of houses are obvious in the image.

channels may not be necessary. Thus, feature extraction or feature selection from the differential profiles is of interest.

III. Feature Extraction and Feature Selection

Feature extraction can be viewed as finding a set of vectors that represent 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 the classifiers. The development of feature extraction methods has been one of the most important problems in the field of pattern analysis and has been studied extensively [9]. Feature extraction methods can be both unsupervised and supervised, and also linear and nonlinear. Here, we will concentrate on linear feature extraction. In linear feature extraction, the number of input dimensions corresponds to the number of eigenvectors selected [9]. The transformed data are determined by

$$Y = P^T X \tag{6}$$

where P 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 and applied successfully in classification of remote sensing data [10]. The best known feature extraction approach for representation is the principal component analysis. Although principal component analysis is optimal in the mean square sense for representation, it is not appropriate for classification [9]. Therefore, other feature extraction approaches have been explored when classification is the main objective. One such method is discriminant analysis feature extraction (DAFE), which is a well-known feature extraction method to enhance separability [9], [10]. However, DAFE has the weakness that it is not directly related to the probability of error in classification. The decision boundary feature extraction (DBFE) [10], [11] proposed by Lee and Landgrebe overcomes many of the problems with the DAFE. Lee and Landgrebe have extended DBFE for neural networks [12], but few feature extraction methods have been proposed for neural networks. DAFE and DBFE will be discussed next, followed by a discussion of a simple feature selection approach, which can be used for the problem at hand.

A. DAFE

As stated above, DAFE is a method, which is intended to enhance separability. A within-class scatter matrix Σ_W and a between-class scatter matrix Σ_B are defined [9], [10]. The criterion used for optimization of separability may be defined as

$$J = \mathrm{tr}\left(\Sigma_W^{-1}\Sigma_B\right) \tag{7}$$

where $\mathrm{tr}(\,)$ denotes the trace of a matrix. New feature vectors are selected to maximize the criterion. The necessary transformation from X to Y in (6) is found by taking the eigenvalue-eigenvector decomposition of the matrix $(\Sigma_W^{-1}\Sigma_B)$. The transformation matrix then becomes the set of normalized eigenvectors of $(\Sigma_W^{-1}\Sigma_B)$ corresponding to the eigenvalues in a decreasing

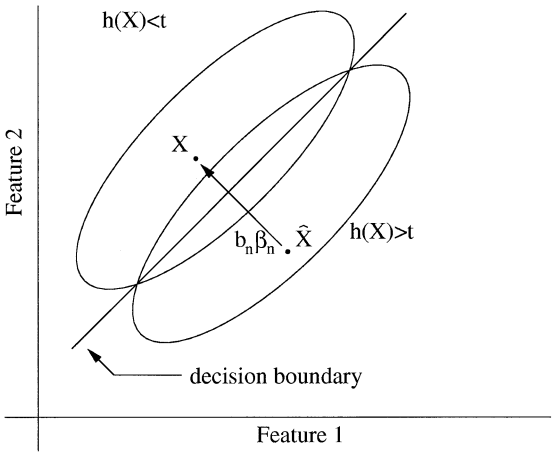

Fig. 2. Example of a discriminantly informative feature, β_n. Here, X denotes a pattern, b_n a constant, $h(X)$ a likelihood ratio, and t the decision boundary. As X is moved along the direction of β_n, the decision changes [11].

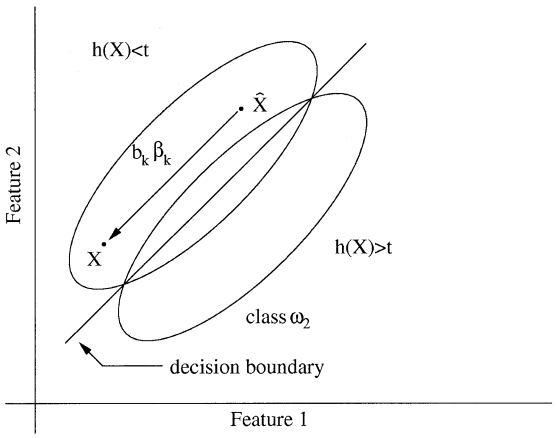

Fig. 3. Example of a discriminantly redundant feature, β_k. Here, X denotes a pattern, b_k a constant, $h(X)$ a likelihood ratio, and t the decision boundary. As X is moved along the direction of β_k, the classification result will be the same [11].

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 [9], [10].

B. DBFE

Lee and Landgrebe [11] showed that discriminantly informative features and discriminantly redundant features can be extracted from the decision boundary itself. They also showed that discriminating informative feature vectors have a component, which is normal to the decision boundary at least at one point on the decision boundary. Further, discriminating redundant feature vectors are orthogonal to a vector normal to the decision boundary at every point on the decision boundary. In [11], *a decision boundary feature matrix* (DBFM) was defined in order to extract *discriminatly informative features* (see Fig. 2) and *discriminatly redundant features* (see Fig. 3) from the decision boundary. It can be shown that the rank of the DBFM is the smallest dimension where the same classification accuracy can be obtained as in the original feature space. Also, the eigenvectors of the DBFM, corresponding to nonzero eigenvalues, are the necessary feature vectors to achieve the same classification accuracy as in the original feature space [11]. As stated above, Lee and Landgrebe have extended their approach for neural networks [12] as demonstrated by the procedure below.

1) DBFE Procedure for Neural Networks (Two-Pattern Class Case) [12]:

Step 1) Train the neural network using all features.

Step 2) For each training sample correctly classified as class ω_1, find the nearest sample correctly classified as class ω_2. Repeat the same procedure for the samples classified as class ω_2.

Step 3) The line connecting a pair of samples found in Step 2 must pass through the decision boundary, since the pair of samples are correctly classified differently. By moving along this line, find the point on the decision boundary or near the decision boundary within a threshold.

Step 4) At each point found in Step 3), estimate the normal vector $\mathbf{N}_i$.

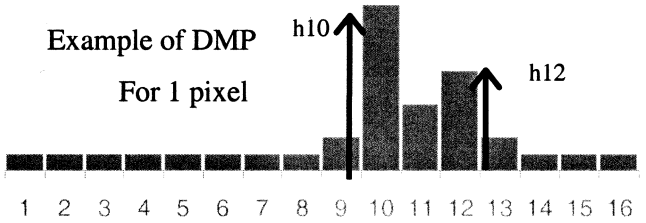

Fig. 4. Use of the indexes of the DMP for feature selection. Here, the heights h10 and h12 are greater than the heights for other indexes. For the pixel in this example, the input vector of the neural network would be three dimensional, i.e., the original gray level value of the pixel and the vector (10, 12).

Step 5) Estimate the decision boundary feature matrix using the normal vectors found in Step 4) by

$$\Sigma_{\mathrm{DBFM}} = 1/L \sum \mathbf{N}_i \mathbf{N}_i^{\mathrm{T}} \tag{8}$$

where L is the number of correctly classified samples and T is the transpose operator.

Step 6) Select the eigenvectors of the decision boundary feature matrix as new feature vectors according to the magnitude of the corresponding eigenvalues.

If there are more than two classes, the procedure can be repeated for each pair of classes after the network is trained for all the classes. Then the total decision boundary feature matrix can be calculated by averaging the decision boundary feature matrices of each pair of classes.

C. Feature Selection Based on the Sorting of the Indexes of the DMP

Using feature extraction methods can be computationally intensive, since they are based on estimating parameters in full feature space. Therefore, a simple feature selection approach is proposed here for comparison. This approach is based on the idea to characterize each pixel with the index corresponding to the size and kind of morphological transformation where we observe the maximum values of the DMP, since these "maximum indexes" should reflect the most important information about morphological characteristics of the structures in the image (see Fig. 4). These indexes can then be used in conjunction with

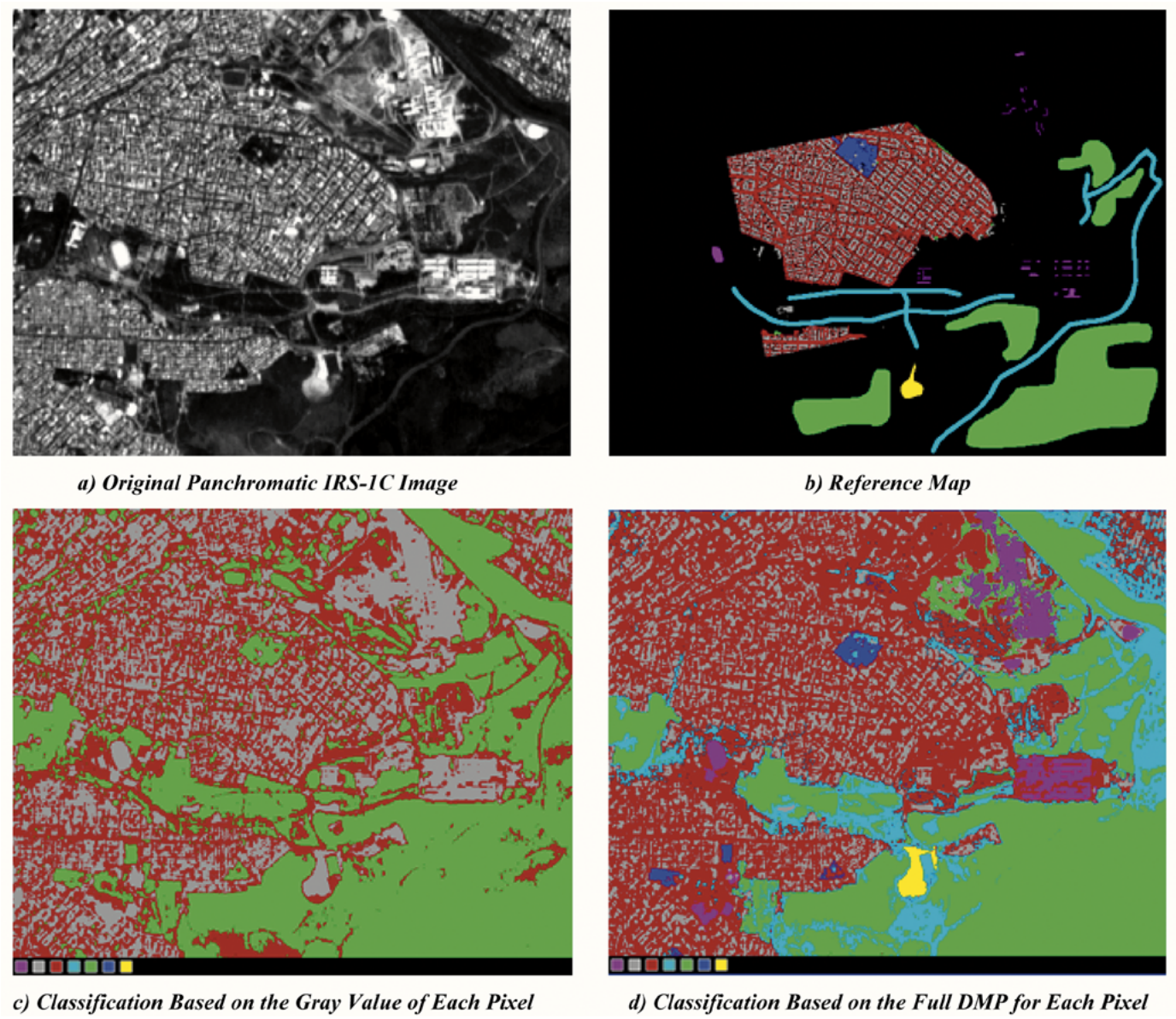

Fig. 5. (a) (upper left) Original Panchromatic IRS-1C Image, (b) (upper right) reference map, (c) (lower left) classification based on the gray value of each pixel (one feature used), (d) (lower right) classification based on the full DMP for each pixel (16 features used).

the original gray level image. We can continue sorting the indexes according to the value of the corresponding DMP (in a descending order) and to extract only the first n indexes. The goal is to find stepwise the "n," which gives good classification accuracies without much computation load. However, it should be kept in mind that this simple feature selection approach should not in general be expected to outperform feature extraction methods, which are based on the linear combination of the features. The goal here is to see if such a heuristic approach can be used to pick out the most important feature with very little effort.

IV. Experimental Results

The proposed approaches where applied on two high-resolution images: 1) Indian Remote Sensing 1C (IRS-1C) panchromatic image from Athens, Greece and 2) IKONOS panchromatic image from Reykjavik, Iceland. In both experiments, a 17-dimensional morphological profile was created (eight closings, eight openings along with the original image) using a circular morphological structuring element with an increasing diameter. A conjugate gradient neural network was used for classification. The two feature extraction approaches, DAFE and DBFE, were applied along with the feature selection based on the sorting of the indexes. The classification accuracies for the different feature sets were compared to accuracies achieved for the full differential morphological profile. In each case, the number of hidden neurons in a neural network classifier was selected based on empirical experiments. The general rule was to select the number of hidden neurons as twice the number of input features. However, in all cases several different implementations were investigated and the one that gave the highest overall accuracies was reported.

TABLE I
INFORMATION CLASSES, TRAINING, AND TEST SAMPLES FOR IRS-1C PANCHROMATIC DATA

Class N	Information Class	Number of Training Samples	Number of Test Samples
1	Large Buildings	180	547
2	Small Buildings	4138	12454
3	Roads inside Urban Area	4736	14243
4	Roads outside Urban Area	1683	5061
5	Open Spaces outside Urban Area	6851	20580
6	Open Spaces inside Urban Area	305	894
7	Wastelands	120	373
Total		18013	54152

TABLE II
TEST ACCURACIES IN PERCENTAGE FOR ORIGINAL IRS-1C IMAGE, THE ENTIRE DIFFERENTIAL PROFILE, AND ADDED FEATURES BASED ON THE SORTING APPROACH

Class No.	Original Gray Value	Original Gray Value + Index of 1st Max	Original Gray Value + Indexes of 1st & 2nd Max	Entire Differential Profile
1	0.0	0.0	82.3	87.6
2	84.8	67.7	67.4	67.5
3	46.4	70.2	74.4	69.6
4	0.0	19.6	28.9	53.4
5	98.3	96.8	98.3	98.0
6	0.0	0.0	0.0	15.8
7	0.0	0.0	0.0	70.0
Average Accuracy	32.8	36.3	50.2	66.0
Overall Accuracy	69.4	72.6	76.0	77.7

A. Experiment With IRS-1C Data From Athens, Greece

The satellite image in experiment 1 is taken from an agricultural area connoted by scattered settlement (northeast of Athens, Greece) and was recorded by the IRS-1C panchromatic sensor, which has a ground spatial resolution of 5 m. The image is a subsample of 800×800 pixels (4 km $\times$ 4 km) from an original scene of about 15 $\times$ 15 pixels (75 km $\times$ 75 km). Fig. 5(a) shows the subsampled image with a min-max histogram stretching. The copresence of objects and regions of different sizes is obvious in the scene. A reference image with six information classes is available [see Fig. 5(b)]. From the available training data, training samples and test samples were selected (see Table I). Approximately, one third of the samples were used for training and the rest for testing the approaches.

The data were at first classified using the original IRS-1C image and the results compared to the classification of the entire DMP; see Fig. 5(c) and (d). The difference between Fig. 5(c) and (d) is striking. In the figures, there are two colors for the open spaces; green for the open space outside urban areas and dark blue for the open space inside urban areas. In Fig. 5(c), the neural network cannot distinguish between the two classes. This stresses the fact that the morphological operations on the original image give information about the size of each structure. In the same way, big buildings (magenta) and small houses (gray) are successfully classified only in Fig. 5(d). We also notice the yellow class (wastelands), which is very successfully classified in Fig. 5(d).

Table II shows the classification accuracies for the test samples when the original IRS-1C image was used, along with feature sets based on the sorting of indexes method, and the entire differential morphological profile. In Table II, overall accuracy represents the accuracy for all samples whereas average accuracy represents the average classification accuracy for the samples. The visual classification results in Fig. 5(c) and (d) are reflected in Table II where it is seen that using the original IRS-1C image alone, only gave 69.4% overall accuracy for test data. By adding more features, both the overall and average accuracies improved. Also, when the original IRS-1C was used as the only input, all large buildings (class 1) were classified as small buildings (class 2). The reason is that there is no information about the size of the structures in the original image but the use of morphological transformations helped in providing that. It is also interesting that the two maximum indexes were needed to classify large buildings with any accuracy at all.

The best overall and average accuracies in Table II were achieved for the entire differential profile as was expected. In Table II, the entire differential profile was the only feature set that could be used to discriminate wastelands (class 7) and open spaces inside urban (class 6) from the other classes. However, in terms of overall accuracies, using the feature set consisting of the original IRS-1C image and the two maximum indexes (76%) gave similar results to the case when the entire differential profile (77%) was used.

The DAFE method and the DBFE for neural networks approach were applied on the DMP data. The cumulative eigenvalues for these approaches are shown in Fig. 6. According to Fig. 6 four transformed features contain almost all the variance, for both approaches. Here, it must be kept in mind that the criteria for the approaches are different.

Table III shows the achieved accuracies when up to six channels were used for the DAFE but six channels represent the full transformed feature set based on the DAFE. According to Table III it is evident that the DAFE did not perform according to expectations. The overall classification accuracies for the first three feature sets [DA(1), DA(2), and DA(3)] were lower than for the original gray value in Table II. With four and five features, similar accuracies were achieved as for the sorting feature selection with two and three features. However, when six DAFE features were used, similar overall accuracies and higher average accuracies were achieved as compared to the classification of the entire differential profile in Table II. The main reason for the poor performance of the DAFE, especially when few features were used, is due to the relatively small variation between adjacent differential morphological features. Because of the low variation, the matrix $\Sigma_W^{-1}\Sigma_B$ was close to being singular for many of the classes. These singularity problems make the DAFE results unreliable.

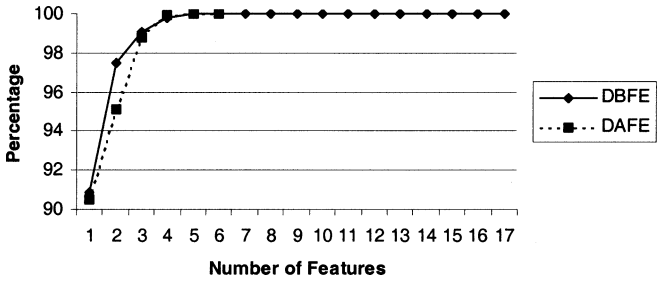

Fig. 6. IRS-1C data. The cumulative eigenvalues in percentage for the DBFE and DAFE approaches.

TABLE III
IRS-1C DATA. TEST ACCURACIES IN PERCENTAGE FOR THE MOST IMPORTANT DAFE CHANNELS. THE NUMBER IN PARENTHESES REPRESENTS THE NUMBER OF DISCRIMINANT ANALYSIS FEATURES USED IN CLASSIFICATION

Class No.	DA(1)	DA (2)	DA (3)	DA (4)	DA (5)	DA (6)
1	0.0	20.7	88.1	88.1	88.1	73.1
2	0.0	26.8	29.6	39.4	63.3	62.9
3	0.0	78.8	76.4	70.4	73.9	79.7
4	0.0	0.0	0.0	47.3	9.4	22.0
5	100.0	98.5	99.2	97.4	99.3	99.0
6	0.0	76.6	79.6	77.6	82.0	81.9
7	0.0	97.6	98.1	97.9	97.5	97.9
Average Accuracy	14.3	57.0	77.2	75.8	73.4	73.8
Overall Accuracy	38.0	66.3	67.5	71.7	75.4	77.7

Table IV shows the achieved accuracies when up to four channels were used for the DBFE. It is interesting to note that using three and four DBFE channels outperformed even the classification of the entire differential profile in Table II in terms of overall and average accuracies. Using the feature set based on DBFE with four features gave higher accuracies than the classification of the entire differential profile for classes 6 and 7. However, higher accuracies were achieved for class 4 (roads outside urban) using the feature set based on the entire differential profile. In contrast to the DAFE, the DBFE had no singularity problems.

The overall classification accuracies for test data as a function of the number of features for the different feature extraction methods and the simple feature selection approach are shown in Fig. 7. From Fig. 7, it is clear in all cases that classification of the feature set based on the decision boundary was more accurate than classification of the feature sets based on DAFE and on sorting the indexes in the DMP. This result is in accordance with Tables II–IV. It is interesting to see in Fig. 7 that the overall accuracies for the DBFE and the simple sorting feature selection did not improve much after four features had been included in the feature sets. On the other hand, the accuracies fluctuated a lot.

B. Experiment With IKONOS Data From Reykjavik, Iceland

An IKONOS panchromatic image (see Fig. 1) from Reykjavik, Iceland was used in the second experiment. The image is

TABLE IV
IRS-1C DATA. TEST ACCURACIES IN PERCENTAGE FOR THE MOST IMPORTANT DBFE CHANNELS ACCORDING TO THE DBFM. THE NUMBER IN PARENTHESES REPRESENTS THE NUMBER OF DECISION BOUNDARY FEATURES USED IN CLASSIFICATION

Class No.	DB (1)	DB (2)	DB (3)	DB (4)
1	0.0	88.1	39.5	87.6
2	72.8	76.0	62.6	66.4
3	66.1	56.7	81.1	77.6
4	0.0	51.5	26.3	33.4
5	98.9	97.6	98.8	98.8
6	0.0	65.5	75.4	79.0
7	0.0	12.6	81.5	94.1
Average Accuracy	32.8	36.3	50.2	66.0
Overall Accuracy	71.5	76.6	77.9	79.2

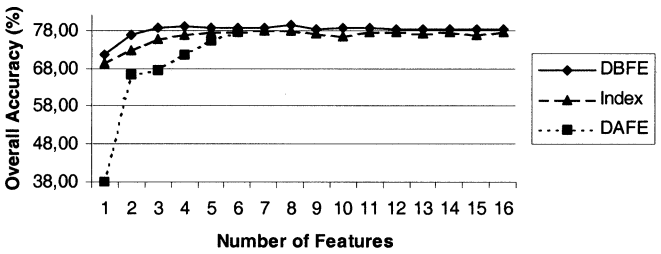

Fig. 7. IRS-1C data. Overall classification accuracies for test data as a function of the nmber of features for the DBFE and DAFE approaches and the approach based on sorting the maximum indexes in the DMP (index).

TABLE V
INFORMATION CLASSES, TRAINING, AND TEST SAMPLES FOR IKONOS PANCHROMATIC DATA

Class No.	Information Class	Number of Training Samples	Number of Test Samples
1	Large Buildings	2810	2806
2	Small Buildings	1673	1663
3	Streets	763	763
4	Open Areas	1398	1399
5	Residential Lawns	644	642
6	Shadows	3766	3770
Total		11054	11043

a 975×639 pixel subsample from an IKONOS frame with full 1-m spatial resolution. The image was acquired on August 9, 2001, when the sun elevation was 42°. The relatively low sun elevation has the effect that shadows of houses are obvious in the image. Therefore, shadows are defined as a separate class. The test area is in the middle part of Reykjavik. It comprises residential, commercial, and open areas.

In this experiment, we concentrated on six information classes (see Table V). These classes were: large buildings,

TABLE VI
TEST ACCURACIES IN PERCENTAGE FOR ORIGINAL IKONOS IMAGE, THE ENTIRE DIFFERENTIAL PROFILE, AND ADDED FEATURES BASED ON THE SORTING APPROACH

Class No.	Original Gray Value	Original Gray Value + Index of 1st Max	Original Gray Value + Indexes of 1st & 2nd Max	Entire Differential Profile
1	81.8	86.4	80.7	94.7
2	0.0	18.4	76.7	85.4
3	94.8	93.2	94.4	95.1
4	58.4	58.3	58.1	96.6
5	93.5	84.9	87.7	94.8
6	90.3	92.3	81.5	99.1
Average Accuracy	69.8	72.3	79.9	94.1
Overall Accuracy	70.9	75.0	78.8	95.1

small buildings, streets, open areas, residential lawns, and shadows. The large buildings are mainly confined to the commercial area in the upper and middle part of the image. They appear in different shades of gray, from bright to very dark. Most of these houses are flat-topped but some of them have an uprise resulting in facets of different slopes and different gray values in the IKONOS image. The small houses are mostly apartment houses in the lower part and the upper right corner of the image. Most of the apartment houses are single but some are semidetached. As with the large houses, the individual residential houses appear in different gray values, from dark gray to white. They are situated on private lots with trees, bushes and grass lawns. There are actually two types of streets in the imagery; broad streets and narrow streets—but they are considered as a single class. Broad streets are streets with two lanes in each direction. The main broad street in the test area separates the commercial area from the residential area in the lower part of the IKONOS image. Narrow streets have a single lane in each direction or they are one-way streets. Narrow streets are both in the residential area and in the commercial area. Open areas are rather homogeneous grass grown areas, completely void of other types of vegetation such as trees. They are scattered throughout the image but the most prominent open areas are adjacent to the broad street across the lower part of the image. The residential lots or lawns are small areas connected with the individual apartment houses. They are covered with heterogeneous vegetation, i.e., trees, grass, and bushes.

The information classes and training and test samples are listed in Table V. As can be seen from Table V, approximately half of the labeled samples where used for training. The other samples were used to test the neural network classifiers.

Table VI shows the classification accuracies for the test samples when the original IKONOS image was used, along with feature sets based on the sorting of indexes method, and the entire differential morphological profile. As in experiment one; using the original image only (now IKONOS) gave neither high overall nor high average accuracies. Again, by adding more features, both the overall and average accuracies improved. Also, when the original IKONOS image was used as the only input, all

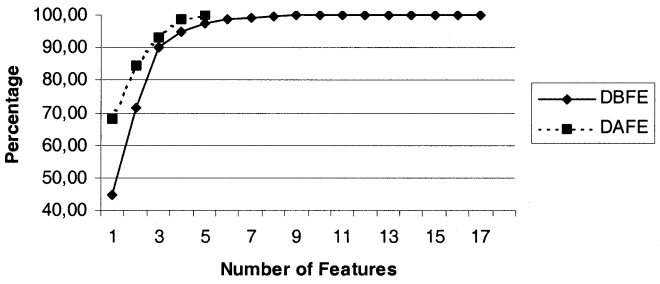

Fig. 8. IKONOS data. The cumulative eigenvalues in percentage for the DBFE and DAFE approaches.

large buildings (class 1) were classified as small buildings (class 2). It is also interesting (similar to experiment 1) that two maximum indexes were needed to classify large buildings with an accuracy that is greater than chance. As in experiment 1, the best overall and average accuracies in Table VI were achieved for the entire differential profile. The difference from experiment one is that much more improvement was achieved. A probable reason for this improvement is that the area being classified here is more uniform than in experiment 1. In this experiment, proportionately more training samples were used when compared to experiment 1. Here, the ratio of training and test samples was approximately 1 whereas in experiment 1, this ratio was 1/3.

The DAFE method and the DBFE for neural networks approach were applied on the DMP data. The cumulative eigenvalues for these approaches are shown in Fig. 8. As can be seen in Fig. 8, four transformed features contain approximately 95% or more of the variance, for both approaches.

Table VII shows the achieved accuracies when up to five channels were used for the DAFE but five channels represent the full transformed feature set based on the DAFE. As can be seen from Table VII, classification based on the DAFE underperformed classification using the simple sorting feature selection in a similar way to the results achieved in experiment 1. However, in this experiment when the full DAFE feature set was used [DA(5)], the overall accuracies were not close to the overall accuracies achieved by the use of the entire differential profile.

TABLE VII
IKONOS DATA. TEST ACCURACIES IN PERCENTAGE FOR THE MOST IMPORTANT DAFE CHANNELS. THE NUMBER IN PARENTHESES REPRESENTS THE NUMBER OF DISCRIMINANT ANALYSIS FEATURES USED IN CLASSIFICATION

Class No.	DA(1)	DA (2)	DA (3)	DA (4)	DA (5)
1	0.0	90.7	92.3	93.0	91.4
2	0.0	0.0	39.4	51.5	71.3
3	0.0	1.6	80.6	0.0	82.4
4	0.0	0.0	58.6	72.3	59.3
5	86.8	86.8	86.8	86.8	86.8
6	100.0	96.5	96.3	94.7	95.9
Average Accuracy	31.1	45.9	75.7	66.4	81.2
Overall Accuracy	39.2	61.1	80.3	77.9	84.7

TABLE VIII
IKONOS DATA. TEST ACCURACIES IN PERCENTAGE FOR THE MOST IMPORTANT DBFE CHANNELS ACCORDING TO THE DBFM. THE NUMBER IN PARENTHESES REPRESENTS THE NUMBER OF DECISION BOUNDARY FEATURES USED IN CLASSIFICATION

Class No.	DB (1)	DB (2)	DB (3)	DB (4)
1	96.8	92.1	91.6	93.9
2	45.6	81.0	83.8	85.2
3	0.0	0.3	90.2	93.6
4	61.5	69.4	89.6	94.5
5	25.9	67.1	86.8	89.7
6	90.9	98.4	99.3	99.6
Average Accuracy	53.5	68.1	90.2	91.1
Overall Accuracy	71.8	81.8	92.4	94.3

As in experiment 1, the main reason for the poor performance of the DAFE is due to the low variation between adjacent differential morphological features. Here, the matrix $\Sigma_W^{-1}\Sigma_B$ was close to being singular for many of the classes. Therefore, the DAFE results should be taken with reservations.

Table VIII shows the achieved accuracies when up to four channels were used for the DBFE. In contrast to experiment 1, using three and four DBFE channels was outperformed in classification when the entire differential profile was used in terms of both overall and average accuracies. However, using the feature sets based on DBFE with three and four features gave outstanding overall and average accuracies in the experiment (more than 90% accuracy). It is interesting that three DBFE features were needed to classify class 3 (streets) with any accuracy at all. On the other hand, class 3 was classified very accurately when only the original image was used (see Table VI).

The overall classification accuracies for test data as a function of the number of features for the DBFE and DAFE approaches and the feature selection method based on sorting the indexes

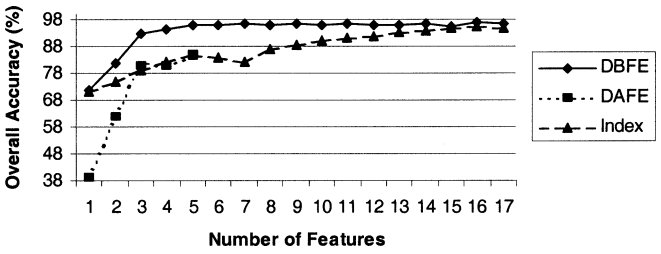

Fig. 9. IKONOS data. Overall classification accuracies for test data as a function of the number of features for the DBFE and DAFE approaches and the approach based on sorting the maximum indexes in the DMP (index).

are shown in Fig. 9. From Fig. 9 it is seen that classification of the feature set based on the decision boundary was in most cases more accurate than classification of the feature set based on sorting indexes in the DMP. The overall test accuracies seem to reach "saturation" when the DBFE feature set with five features was used. However, the overall classification accuracies based on sorting the indexes of the DMP show an increasing trend. The overall accuracies for both methods were similar when 15 or more features are used. The results using the DAFE were comparable to the simple sorting approach when three to five features were used. However, the DAFE results should be taken with reservations because of the singularity problems.

V. CONCLUSION

Classification and feature extraction for urban data have been considered. It is well known that remote sensing images with high-resolution are needed for classification of urban images. However, using only single band high-resolution panchromatic data is not sufficient to classify structural information accurately. To overcome this problem, Pesaresi and Benediktsson [2] proposed the use of morphological transformations to build a differential morphological profile that could be used for classification by neural networks. However, the drawback of their method concerns the necessity of looking at a range of increasing opening and closing by reconstruction operations. Therefore, the resulting differential profile can be high-dimensional. In this paper, methods to preprocess the differential morphological profiles were investigated in order to reduce the computational load when differential morphological profiles are used for classification by neural networks.

In the paper, the use of two feature extraction methods and a simple feature selection approach were investigated. The considered approaches were: 1) decision boundary feature extraction for neural networks, 2) discriminant analysis feature extraction, and 3) simple feature selection based on sorting the indexes of the DMP using the value of the discrete derivative. The approaches were applied on two datasets. One dataset consisted of IRS-1C panchromatic data from Athens, Greece and the other of panchromatic IKONOS data from Reykjavik, Iceland. The use of the differential profiles in classification was seen to be valuable. Although the original data were different in many ways (e.g., resolution, sun angle), good overall accuracies were achieved for both datasets.

The DBFE has the advantage that it is based on the classifier used (here a neural network), and it can provide as many features as the input data. In experiments, the classification of

the feature sets based on the DBFE outperformed the other approaches in terms of overall and average accuracies. In comparison, the DAFE had singularity problems and did not give reliable results. The simple sorting feature selection performed relatively well. It is computationally simple and does not rely on statistics in the full feature space. Consequently, the simple method does not explore all aspects of the feature space. Other more advanced feature selection approaches should give higher accuracies and will be investigated in our future research. Examples of such methods are approaches based on forward feature selection or backward feature elimination. On the other hand, the proposed simple feature selection could be used when computational speed is more important than accuracy.

ACKNOWLEDGMENT

The authors appreciate the comments of the anonymous reviewers. Their suggestions have significantly improved the paper. The authors acknowledge the contributions of M. Perrin (Ecole Nationale Superierure d' Ingenieurs Electiciens de Grenoble) and K. Benediktsson (Decode Genetics, Reykjavik) to this work, but M. Perrin worked on this project at the Univeristy of Iceland, Reykjavik, for four months in 2002. The authors thank C. Lee (Yonsei University) and D. A. Landgrebe (Purdue University) for the DBFE code they provided.

REFERENCES

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

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

[3] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 309–320, Feb. 2001.

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

[5] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geosciences and remote sensing," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 2042–2055, Sept. 2002.

[6] M. Pesaresi, "Analisi Numerica dello spazio edificato nella città diffusa,", Venice, Italy, Tech. Rep. IUAV DAEST, 1993.

[7] A. Bianchin and M. Pesaresi, "Outils de morphologie mathématique appliquée aux images satellite pour l'analyze de l'urbanization diffuse," in *Proc. EGIS-MARI 94 Conf.*, Paris, France, Mar. 29–Apr. 1 1994, pp. 2085–2094.

[8] M. Pesaresi and I. Kannellopulos, "Detection of urban features using morphological based segmentation and very high resolution remotely sensed data," in *Machine Vision and Advanced Image Processing in Remote Sensing*, I. Kanellopoulos, G. G. Wilkinson, and T. Moons, Eds. Berlin, Germany: Springer-Verlag, 1999.

[9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

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

[11] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 388–400, Apr . 1993.

[12] ——, "Decision boundary feature extraction for neural networks," *IEEE Trans. Neural Networks*, vol. 8, pp. 75–83, Jan. 1997.

Jon Atli Benediktsson (S'86–M'90–SM'99) received the Ph.D. degree from the School of Electrical Engineering at Purdue University, West Lafayette, IN, in 1990.

He is currently a Professor of electrical and computer engineering at the University of Iceland, Reykjavik. He is also a Visiting Professor at the School of Computing and Information Systems, Kingston University, Kingston upon Thames, U.K. He has held visiting positions at the Joint Research Centre of the European Commission, Ispra, Italy, Denmark's Technical University (DTU), Lyngby, and the School of Electrical and Computer Engineering, Purdue University. His research interests are in pattern recognition, neural networks, remote sensing, image processing, and signal processing, and he has published extensively in those fields.

Dr. Benediktsson was a Fellow at the Australian Defence Force Academy (ADFA), Canberra, in August of 1997. In 1991, he received from Purdue University the Stevan J. Kristof Award 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 2003. In 2002, he was appointed Vice President of Technical Activities in the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS). From 1996 to 1999, he was the Chairman of GRSS' Technical Committee on Data Fusion. He coedited (with Professor David A. Landgrebe) a Special Issue on Data Fusion of TGARS (May 1999). He is the current and founding Chairman of the IEEE Iceland Section.

Martino Pesaresi graduated in Town and Regional Planning from IUA, Venice, Italy, in 1992 with an overall first class mark of 110/110. In 1993, he followed a specialization course on digital photogrammetry at the International Centre for Mechanical Sciences (CISM), Udine, Italy.

Since 1998, he has been a Visiting Professor of spatial statistics at the University of Venice, Venice, Italy. He has pursued research activities involving the use of remotely sensed data in urban analysis with the Centre d'Analyze et de Mathématique Sociales (CAMS), Ecole des Hautes Etudes en Sciences Sociales (EHESS) from 1991 to 1992 and with the Laboratoire d'Informatique Appliquée (LIA), ORSTOM, France, from 1992 to 1993. He was a Research Assistant with the Dipartimento di Analisi Economica e Sociale (DAEST) and the Dipartimento di Urbanistica (DU), both of the University of Venice (IUAV) from 1992 to 1995 and with the Dipartimento di Architettura e Urbanistica (DAU) , University of Pescara, Pescara, Italy, from 1996 to 1997. From 1992 to 1997, he was a Consultant for many public and private companies as well for professionals involved in town and regional planning activities. From 1997 to 2000, he was working with a research contract at the EC Joint Research Centre, Space Applications Institute, EGEO Unit, Advanced Methods Sector. Since February 2001, he has been with the INFORM srl, Padova, Italy, leading the Geographical Information System and Remote Sensing Applications Group. His main interests are social-economic and spatial data analysis, regional analysis, GIS implementation, and satellite image processing for analysis of the urban and natural environment, with special attention to image textural and image morphological segmentation techniques and new structure-based multiscale image fusion approaches.

Kolbeinn Arnason received the B.S. degree in geophysics from the University of Iceland, Reykjavik, Iceland, in 1977. He received the M.S. (Dipl. Geophys.) degree from the Institute for Geophysics and the Ph.D. (Dr.rer.nat.) degree from the Institute for Geological Remote Sensing, both from Ludwig-Maximilians University, Munich, Germany.

He is currently working on a joint remote sensing project between the National Land Survey of Iceland and the Engineering Research Institute, University of Iceland. His research interests are remote sensing applications, image processing, and geographical information science.