Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles

Mathieu Fauvel, Student Member, IEEE, Jón Atli Benediktsson, Fellow, IEEE, Jocelyn Chanussot, Senior Member, IEEE, and Johannes R. Sveinsson, Senior Member, IEEE

Abstract—A method is proposed for the classification of urban hyperspectral data with high spatial resolution. The approach is an extension of previous approaches and uses both the spatial and spectral information for classification. One previous approach is based on using several principal components (PCs) from the hyperspectral data and building several morphological profiles (MPs). These profiles can be used all together in one extended MP. A shortcoming of that approach is that it was primarily designed for classification of urban structures and it does not fully utilize the spectral information in the data. Similarly, the commonly used pixelwise classification of hyperspectral data is solely based on the spectral content and lacks information on the structure of the features in the image. The proposed method overcomes these problems and is based on the fusion of the morphological information and the original hyperspectral data, i.e., the two vectors of attributes are concatenated into one feature vector. After a reduction of the dimensionality, the final classification is achieved by using a support vector machine classifier. The proposed approach is tested in experiments on ROSIS data from urban areas. Significant improvements are achieved in terms of accuracies when compared to results obtained for approaches based on the use of MPs based on PCs only and conventional spectral classification. For instance, with one data set, the overall accuracy is increased from 79% to 83% without any feature reduction and to 87% with feature reduction. The proposed approach also shows excellent results with a limited training set.

Index Terms—Data fusion, extended morphological profile (EMP), feature extraction (FE), high spatial resolution, hyperspectral data, support vector machines (SVMs).

I. INTRODUCTION

IN CLASSIFICATION of remote sensing data from urban areas, the identification of relatively small objects, e.g., houses and narrow streets, is important. Therefore, high spatial resolution of the imagery is necessary for accurate classification. The most commonly available remote sensing data of high spatial resolution from urban areas are single-band panchromatic data. However, using only one high-resolution panchromatic data channel is usually not sufficient for accurate classification of structural information. To overcome that problem, Pesaresi and Benediktsson [1] proposed the use of morphological transformations to build a morphological profile (MP). In [2], the method in [1] was extended for hyperspectral data with high spatial resolution. The approach in [2] is based on using several principal components (PCs) from the hyperspectral data. From each of the PCs, an MP is built. These profiles are used all together in one extended MP (EMP), which is then classified by a neural network. The method in [2] has been shown to perform well in terms of accuracies when compared to more conventional classification approaches. However, a shortcoming of the approach is that it is primarily designed for classification of urban structures and it does not fully utilize the spectral information in the multispectral or hyperspectral data.

However, this type of data contains a lot of information about the spectral properties and the land cover of the data. A finer definition of the classes is possible, and more classes can be considered. Based on the spectral signatures of the classes, many advanced pixel-based classifiers have been proposed, including advanced statistical classifiers [3] and distribution-free approaches such as neural networks and support vector machines (SVMs) [4]. The latter one has shown remarkable abilities to deal with remote multispectral data, particularly with hyperspectral data. However, if the spatial content of the image is not used, the resulting thematic map sometimes looks noisy (salt and pepper classification noise). Approaches involving Markov random field (MRF) and Monte Carlo optimization have been proposed in [5] and [6]. These approaches use the contextual information. The main shortcoming of such algorithms is the computing time, which can be high even for small data sets. Regarding the high dimensionality of recently acquired data, both in the spectral and in the spatial domain, computationally light algorithm is of interest. In this sense, the MP has been proposed as an alternative way to use spatial information [1], [7]. Relative to the MRF-based classifiers, the MP and its extension to a multiband image, the EMP, have the possibility to use geometrical contextual information (shape, size, etc.) and perform well on many kinds of data (panchromatic, multispectral, and hyperspectral data). However, as stated above, a shortcoming of this approach is that it does not fully utilize the spectral information in the data, and consequently, several approaches based on the MP/EMP have been proposed to fully exploit the spatial and the spectral information [8]–[10].

Each data set has its own properties, defining its ability to deal with different natures of classes. Table I sums up the

Manuscript received November 11, 2007; revised February 14, 2008. Current version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008. This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne version published October 30, 2008.
properties of spectral and morphological/spatial data. The first main consideration is the complementary characteristics of the data. It has a consequence in the discrimination ability of such a feature, as will be seen in the experiments. The fusion of two types of information should clearly results in an increase of the classification in terms of global accuracy. The use of spectral information can be critical for classification of nonstructured information in urban areas, e.g., vegetation and soil classes, whereas the use of spatial information can be useful for classification of structured objects, e.g., road and building classes.

The second consideration is the possible redundancy of each feature set. Hence, feature extraction (FE) algorithms could be of an interest.

To include both types of information, an extension to the approach in [2] is proposed in this paper. The proposed method is based on the data fusion of the morphological information and the original data. First, an EMF is created based on the PCs from the hyperspectral data. Second, FE is applied on the original hyperspectral data and the EMF. Finally, the extracted feature vectors from both the original data and the EMF are concatenated into a stacked vector and classified. The proposed approach differ from the approaches in [12]–[14], where the authors had extracted spatial information and used composite kernel to include both types of information. Here, FE algorithms are used to select informative feature from the spectral and spatial domains.

For the multisource classification, SVMs are used rather than a neural network, which was used in our previous experiment with MP/EMP. The superiority of SVMs, implementing structural risk minimization, over the neural classifiers, implementing empirical risk minimization, has been discussed in [4, Ch.9.6 and 12] and in [15], [16]. SVMs aim to discriminate two classes by fitting an optimal separating hyperplane to the training data within a multidimensional feature space by using only the closest training samples. Thus, the approach only considers samples close to the class boundary and works well with small training set, even when high-dimensional data sets are classified. SVMs have already been applied for multisource classification in [17] where several output coding methods were investigated.

In this paper, the proposed approach has been compared to statistical classification methods and SVM classification. Experiments were conducted on two different high-resolution remote sensing data sets from urban areas. The effectiveness of the proposed methodology with a limited training set has also been assessed.

This paper is organized as follows. Section II reviews the use of morphological transformations for processing of hyper-spectral imagery in urban areas. In Section III, the considered supervised FE approaches are introduced. SVMs are discussed in Section IV. The applied data fusion schemes are discussed in Section V. Experimental results obtained on two ROSIS data sets from urban areas are presented in Section VI. Finally, conclusions are drawn in Section VII.

II. EMP

Mathematical morphology is a theory aiming to analyze the spatial relationship between pixels. For a remote sensing application, several morphological operators are available for extracting geometrical information. An overview of operators can be found in [18]. In the following section, some basic notions of mathematical morphology are reviewed. Then, the concept of the morphological profile and its extension to multi-valued data are detailed.

A. Mathematical Morphology

The two fundamental operators in mathematical morphology are erosion and dilatation [19]. These operators are applied to an image with a set of known shapes, called the structuring elements (SEs). To erode an image consists of finding where the SE fits the objects in the image. The dilation, which is dual to the erosion, shows where the SE hits the objects.

Opening and closing are combinations of erosion and dilation. These operators are removed from the original image structures of size less than the SE. However, they also modify structures which are still present in the image after the opening/closing. Thus, they can introduce fake objects in the image. To avoid this problem, geodesic morphology and reconstruction should be used [19]. Opening and closing by reconstructions are connected operators that satisfy the following assertion: If the structure of the image cannot contain the SE, then it is totally removed; else, it is totally preserved. For a given SE, geodesic opening or geodesic closing allows one to know the size or shape of some objects present in the image: The objects that are smaller than the SE are deleted, whereas the others (that are bigger than the SE) are preserved. To determine the shape or size of all elements present in an image, it is necessary to use a range of different SE sizes. This concept is called granulometry.

Granulometries are typically used for the analysis of the size distribution of the structures in the images. Classical granulometry by opening is built by successive opening operations with an SE of an increasing size. By doing so, the image is progressively simplified. By using connected operators, like opening by reconstruction, no shape noise is introduced.

MPs are defined by using the granulometry. An MP is composed of the opening profile (OP) and the closing profile (CP). The OP at the pixel x of the image I is defined as an n-dimensional vector

$$\text{OP}_i(x) = \gamma_R^{(i)}(x) \quad \forall i \in [0, n]$$

where $\gamma_R^{(i)}$ is the opening by reconstruction with an SE of a size $i$ and $n$ is the total number of openings. In addition, the CP at the pixel x of image I is defined as an n-dimensional vector

$$\text{CP}_i(x) = \phi_R^{(i)}(x) \quad \forall i \in [0, n]$$

where $\phi_R^{(i)}$ is the closing by reconstruction with an SE of a size $i$. Clearly, we have $\text{CP}_0(x) = \text{OP}_0(x) = I(x)$. By collating
the OP and the CP, the MP of an image \( I \) is defined as a \( 2n + 1 \)-dimensional vector

\[
\text{MP}(x) = \{\text{CP}_n(x), \ldots, I(x), \ldots, \text{OP}_n(x)\}
\]  

(3)

Example of MP is shown in Fig. 1. Thus, from a single image results a multiband image, whose dimension corresponds to the number of transformations, and spatial information is now contained in the MP for each pixel. However, an MP is built with only one band. Therefore, the spectral information is lost. One approach to deal with this problem is to extract several images that contain some parts of the spectral information and then build the MP on each of the individual images. This approach, namely, the EMP, is discussed in the following.

B. EMP

In order to apply this approach to hyperspectral data, characteristic images need to be extracted. In [11], it was suggested to use several PCs of the hyperspectral data for such a purpose. Hence, the MP is applied on the first PCs, corresponding to a certain amount of the cumulative variance, and a stacked vector is built with the MP on each PC. This yields to the EMP. Following the previous notation, the EMP is an \( m(2n + 1) \)-dimensional vector:

\[
\text{MP}_{\text{ext}}(x) = \{\text{MP}_{\text{PC1}}(x), \ldots, \text{MP}_{\text{PCm}}(x)\}
\]  

(4)

where \( m \) is the number of retaining PCs. An example of EMP is shown in Fig. 2.

As with multispectral data, the MP/EMP may include some redundancy. Classical feature reduction algorithm can be applied, as detailed in the following section.

III. SUPERVISED FEATURE EXTRACTION (FE)

FE 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 of interest. 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 FE, the number of input dimensions corresponds to the number of selected eigenvectors [3]. The transformed data are determined by

\[
\tau = \Phi^T x
\]  

(5)

where \( \Phi \) is the transformation matrix composed of the eigenvectors of the feature matrix, \( x \) is the data in the input space, and \( \tau \) is the transformed data in the feature space. We have in general \( \dim(x) \geq \dim(\tau) \). Several statistical extraction approaches have been proposed for remote sensing data [3], including decision boundary FE (DBFE) and nonparametric weighted FE (NWFE).

A. DBFE

It was shown in [20] that both discriminantly informative features and redundant features can be extracted from the decision boundary between two classes. The features are extracted from the decision boundary feature matrix (DBFM). The eigenvectors of the DBFM corresponding to nonzero eigenvalues are the necessary feature vectors to achieve the same classification accuracy as in the original space. The efficiency of the DBFE is related to the training set and can be computationally intensive.

B. NWFE

To overcome the limitations of the DBFE, Kuo and Landgrebe [21] proposed the NWFE. NWFE is based on the discriminant analysis FE by focusing on samples near the eventual decision boundary. The main ideas of the NWFE are as follows: 1) putting different weights on every sample to compute the local means and 2) defining nonparametric between-class and within-class scatter matrices [3].

Many experiments have shown the effectiveness of these approaches for the classification of hyperspectral data [3]. They are usually applied on the spectral data, but they are successfully applied to the EMP [11].

IV. CLASSIFICATION BY SVM

So far, in our previous approach [2], [7], [11], [22], the classification was done with either a statistical classifier (Gaussian maximum likelihood (ML)), a neural network, or a fuzzy classifier. Here, it is proposed to use the SVMs. Early work in classification of remotely sensed images by SVM showed excellent results [17], [23], [24]. In [15], several SVM-based classifiers were compared to other classical classifiers such as a K-nearest neighbor classifier and a neural network classifier, and the SVM using the kernel method outperformed the other classifiers in terms of accuracy. Multiclass SVM performances were also positively compared with a discriminant analysis classifier, a decision tree classifier, and a feedforward neural network classifier with a limited training set [25]. SVMs show good results in the situation of limited training set in [26]. Semisupervised SVMs were also investigated for multispectral data classification [27], [28].

SVM is surely among the most used kernel learning algorithms. It performs robust nonlinear classification of samples using the kernel trick. The idea is to find a separating hyperplane in some feature space induced by the kernel function while all the computations are done in the original space [4]. A good introduction to SVM for pattern recognition can be found in [29]. Given a training set \( S = \{(x^1, y_1), \ldots, (x^\ell, y_\ell)\} \in \mathbb{R}^n \{-1; 1\} \), the decision function is found by solving the convex optimization problem

\[
\max_{\alpha} g(\alpha) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j k(x^i, x^j)
\]

subject to \( 0 \leq \alpha_i \leq C \) and \( \sum_{i=1}^{\ell} \alpha_i y_i = 0 \)  

(6)
where $\alpha$’s are the Lagrange coefficients, $C$ is a constant that is used to penalize the training errors, and $k$ is the kernel function. To be an acceptable kernel, $k$ should be a positive semidefinite function [30]. One classical effective kernel is the Gaussian kernel

$$k(x^i, x^j) = \exp \left( -\frac{\|x^i - x^j\|^2}{2\sigma^2} \right) \tag{7}$$

where the norm is the Euclidean norm and $\sigma \in \mathbb{R}^+$ tunes the flexibility of the kernel. A short comparison of kernels for remotely sensed image classification can be found in [26].

When the optimal solution of (6) is found, i.e., the $\alpha_i$, the classification of a sample $x$ is achieved by looking to which side of the hyperplane it belongs

$$y = \text{sgn} \left( \sum_{i=1}^{l} \alpha_i y_i k(x^i, x) + b \right). \tag{8}$$

To deal with multiclass classification problem, the pairwise approach was used in our experiments [31]. More advanced multiclass approaches applied to remote sensing data can be found in [15]. For the particular case of one-class classification, a dedicated methodology is proposed in [32].

The SVMs are mainly a nonparametric method, yet some parameters need to be tuned before the optimization. In the Gaussian kernel case, there are two parameters: $C$, which is the penalty term, and $\sigma$, which is the width of the exponential. It is usually done by a cross-validation step, where several values are tested. In our experiments, $C$ was fixed to 200 and $\sigma^2 \in \{0.5, 1, 2, 4\}$ was selected by using a fivefold cross validation. The SVM optimization problem was solved by using library LIBSVM [33].

V. DATA FUSION

The proposed method is based on the data fusion of the morphological information and the original data. In a previous work [34], it was proposed to fuse the classification results of two SVM classifiers, each one working with either the spectral or the EMP data. It consisted in an appropriate adaptive fusion scheme based on the output’s characteristics of the SVM. The results in terms of accuracy were increased, but it needed two training of SVM, that could be time-consuming.

Here, it is proposed to use a multisource strategy to fuse spectral and spatial information. First, an EMP is created based on applying the PCA on the hyperspectral data. Second, FE is applied on both the EMP and the original hyperspectral data. Finally, the extracted feature vectors are concatenated into one stacked vector and classified by the SVM.

In the morphological processing, we usually retain PCs corresponding to 99% of the cumulative variance. This is done in order to reduce the redundancy in the data but keep most of the variation. The EMP is built by using the $m$ PCs that correspond to the 99% variance. Each MP is composed of $n$ geodesic openings, $n$ geodesic closing, and the corresponding PC. The SE is a disk with initial radius of $r$ pixels. The size increment is $s$. Hence, each MP has $2n + 1$ features, and the EMP has $m(2n + 1)$ features. Noting $x_{\varphi}$ as the features associated to the spectral bands and $x_{\omega}$ as the features associated to the EMP, the corresponding extracted features from the FE algorithm are as follows:

$$x_{\varphi} = \Phi^T_{\varphi} x_{\varphi} \tag{9}$$

$$x_{\omega} = \Phi^T_{\omega} x_{\omega}. \tag{10}$$

The stack vector is finally $y = [x_{\varphi}, x_{\omega}]^T$.

Fig. 3 shows the data fusion scheme. Note that in this paper, only morphological information is extracted, but it is possible to extract other types of spatial information with other processing and include them in the stacked vector.

VI. EXPERIMENTS

A. Data Set

Airborne data from the ROSIS-3 (Reflective Optics System Imaging Spectrometer) optical sensor are used for the experiments. The flight over the city of Pavia, Italy, was operated by the Deutsches Zentrum fur Luft-und Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, and managed and sponsored by the European Union. According to specifications, the number of bands of the ROSIS-3 sensor is 115 with a spectral coverage ranging from 0.43 to 0.86 $\mu$m. The data have been atmospherically corrected.
but not geometrically corrected. The spatial resolution is 1.3 m per pixel. Two data sets were used in the experiment.

1) University area: The first test set is around the Engineering School at the University of Pavia. It is 610 × 340 pixels. Some channels (12) have been removed due to noise. The remaining 103 spectral dimensions are processed. Nine classes of interest are considered, i.e., trees, asphalt, bitumen, gravel, metal sheets, shadow, bricks, meadows, and soil.

2) Pavia center: The second test set is the center of Pavia. The Pavia center image was originally 1096 × 1096 pixels. A 381-pixel-wide black stripe in the left part of image was removed, resulting in a “two parts” image. This “two parts” image is 1096 × 715 pixels. Some channels (13) have been removed due to noise. The remaining 102 spectral dimensions are processed. Nine classes of interest are considered, i.e., water, trees, meadows, bricks, soil, asphalt, bitumen, tiles, and shadows.

Available training and testing set for each data set are given in Tables II and III, and Fig. 4 shows false color images for both data sets.

The classification accuracy was assessed with the following:

1) an overall accuracy (OA) which is the number of well-classified samples divided by the number of test’s samples;
2) an average accuracy which represents the average of class classification accuracy;
3) a kappa coefficient of agreement (κ) which is the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone [35].

These criteria were used to compare classification results and were computed by using the confusion matrix. Furthermore, the statistical significance of differences was computed by using McNemar’s test, which is based upon the standardized normal test statistic [36]

\[ Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \]  

(11)

where \( f_{12} \) indicates the number of samples classified correctly by classifier 1 and incorrectly by classifier 2. The difference in accuracy between classifiers 1 and 2 is said to be statistically significant if \( |Z| > 1.96 \). The sign of \( Z \) indicates whether classifier 1 is more accurate than classifier 2 \( (Z > 0) \) or vice versa \( (Z < 0) \). This test assumes related testing samples and thus is adapted to our situation since the training and testing set were the same for each experiment.

The FEs were done with MultiSpec [3], whereas the morphological operations were done with the image processing toolbox of Matlab. The SVM classification was done by using the LIBSVM through its Matlab interface [33]. From previous experiments on the same data set, the Gaussian kernel provides the best results and was used for the experiments [26]. The range of each feature, be it spectral or morphological, was stretched between zero and one.

To obtain a baseline result for comparison, the classification was also done by using the Gaussian ML classifier on the hyperspectral data using MultiSpec.\(^1\) FE was done by using

\(^1\)For the ML, the kappa coefficient was not accessible.
TABLE V
UNIVERSITY AREA. SUMMARY OF THE GLOBAL AND THE CLASS-SPECIFIC TEST ACCURACIES IN PERCENTAGE FOR THE CLASSIFICATION. THE NUMBERS OF FEATURES FROM THE SPECTRAL DATA AND THE MORPHOLOGICAL DATA, RESPECTIVELY, ARE GIVEN IN BRACKETS

<table>
<thead>
<tr>
<th>Features</th>
<th>ML (NWFE)</th>
<th>Spectral</th>
<th>EM</th>
<th>Spec.</th>
<th>EMP</th>
<th>DBFE 95%</th>
<th>NWFE 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>80.10</td>
<td>79.48</td>
<td>79.14</td>
<td>83.53</td>
<td>87.97</td>
<td>87.59</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>87.00</td>
<td>88.14</td>
<td>84.30</td>
<td>89.39</td>
<td>89.43</td>
<td>88.93</td>
<td></td>
</tr>
<tr>
<td>κ</td>
<td>-</td>
<td>74.47</td>
<td>73.25</td>
<td>79.13</td>
<td>84.40</td>
<td>83.89</td>
<td></td>
</tr>
<tr>
<td>Class 1</td>
<td>76.00</td>
<td>84.36</td>
<td>94.50</td>
<td>95.33</td>
<td>90.92</td>
<td>86.80</td>
<td></td>
</tr>
<tr>
<td>Class 2</td>
<td>73.90</td>
<td>66.20</td>
<td>72.82</td>
<td>73.46</td>
<td>85.91</td>
<td>86.95</td>
<td></td>
</tr>
<tr>
<td>Class 3</td>
<td>70.80</td>
<td>71.99</td>
<td>53.22</td>
<td>65.89</td>
<td>57.88</td>
<td>63.26</td>
<td></td>
</tr>
<tr>
<td>Class 4</td>
<td>96.70</td>
<td>98.01</td>
<td>98.89</td>
<td>99.18</td>
<td>99.22</td>
<td>98.53</td>
<td></td>
</tr>
<tr>
<td>Class 6</td>
<td>87.60</td>
<td>93.12</td>
<td>58.11</td>
<td>84.15</td>
<td>85.32</td>
<td>82.62</td>
<td></td>
</tr>
<tr>
<td>Class 7</td>
<td>92.00</td>
<td>91.20</td>
<td>96.09</td>
<td>97.22</td>
<td>95.19</td>
<td>96.61</td>
<td></td>
</tr>
<tr>
<td>Class 8</td>
<td>87.30</td>
<td>92.26</td>
<td>95.27</td>
<td>96.12</td>
<td>95.84</td>
<td>95.38</td>
<td></td>
</tr>
<tr>
<td>Class 9</td>
<td>99.10</td>
<td>96.62</td>
<td>91.24</td>
<td>93.66</td>
<td>95.14</td>
<td>90.60</td>
<td></td>
</tr>
</tbody>
</table>

TABLE VI
UNIVERSITY AREA. STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION ACCURACIES

<table>
<thead>
<tr>
<th></th>
<th>EMP/Spectral</th>
<th>Spectral/Spec. EMP</th>
<th>Spectral/DBFE 95%</th>
<th>Spectral/NWFE 80%</th>
<th>Spec. EMP/DBFE 95%</th>
<th>Spec. EMP/NWFE 80%</th>
<th>DBFE 95%/NWFE 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-1.36</td>
<td>-9.55</td>
<td>-37.98</td>
<td>-36.61</td>
<td>-25.19</td>
<td>-21.79</td>
<td>2.32</td>
</tr>
</tbody>
</table>

TABLE VII
UNIVERSITY AREA. GLOBAL ACCURACIES IN PERCENTAGE WITH DIFFERENT FE METHODS. THE NUMBERS OF FEATURES FROM THE SPECTRAL DATA AND THE MORPHOLOGICAL DATA, RESPECTIVELY, ARE GIVEN IN BRACKETS

<table>
<thead>
<tr>
<th>Feature extraction</th>
<th>Cum. Var.</th>
<th>Features</th>
<th>OA</th>
<th>AA</th>
<th>κ</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBFE</td>
<td>99</td>
<td>60 (45,15)</td>
<td>84.77</td>
<td>89.98</td>
<td>80.43</td>
</tr>
<tr>
<td>95</td>
<td>27 (27,10)</td>
<td>87.97</td>
<td>88.94</td>
<td>84.46</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>28 (20,8)</td>
<td>86.49</td>
<td>88.94</td>
<td>82.56</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>19 (14,5)</td>
<td>82.95</td>
<td>87.51</td>
<td>77.27</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>14 (10,4)</td>
<td>76.93</td>
<td>84.63</td>
<td>71.26</td>
<td></td>
</tr>
<tr>
<td>NWFE</td>
<td>99</td>
<td>62 (42,20)</td>
<td>84.15</td>
<td>88.89</td>
<td>79.61</td>
</tr>
<tr>
<td>95</td>
<td>26 (16,12)</td>
<td>82.90</td>
<td>87.25</td>
<td>77.98</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>18 (10,8)</td>
<td>82.64</td>
<td>86.77</td>
<td>77.65</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>13 (7,6)</td>
<td>87.59</td>
<td>88.93</td>
<td>83.85</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>10 (5,5)</td>
<td>79.57</td>
<td>87.39</td>
<td>74.45</td>
<td></td>
</tr>
</tbody>
</table>

TABLE VIII
UNIVERSITY AREA. PROCESSING TIME IN SECONDS AS FUNCTION OF DIMENSIONALITY AND NUMBER OF SUPPORT VECTORS

<table>
<thead>
<tr>
<th></th>
<th>Spectral</th>
<th>EMP</th>
<th>Spec. EMP</th>
<th>DBFE 95%</th>
<th>NWFE 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>103</td>
<td>27</td>
<td>130</td>
<td>37</td>
<td>13</td>
</tr>
<tr>
<td>Training</td>
<td>2074</td>
<td>850</td>
<td>3257</td>
<td>1184</td>
<td>859</td>
</tr>
<tr>
<td>Number of SVs</td>
<td>1085</td>
<td>405</td>
<td>529</td>
<td>727</td>
<td>572</td>
</tr>
<tr>
<td>Classification</td>
<td>76</td>
<td>9</td>
<td>41</td>
<td>18</td>
<td>9</td>
</tr>
</tbody>
</table>

and the EMP is not statistically significant (see Table VI). Note that it is consistent with the characteristics of the scene: The university area is a mix between man-made structures and natural materials. Therefore, the morphological information is not as useful as it could be in a very dense urban area. When a careful analysis is done on the class-specific accuracies, we can see from Table V that each approach performed well for complementary classes, e.g., the spectral approach performed better for classes 3, 6, and 9, whereas the EMP approach performed better for classes 1, 2, 7, and 8. After the data fusion, we have to look at these classes and see if the best information was used, i.e., if the classification accuracy increased for these classes.

The experiment was then performed with the concatenated vector. The vector was made of the 103 spectral bands and the 27 features of the EMP. The vector was directly used as an input to the SVM. The classification results are reported in Table V. As can be seen from the table, the global accuracies increased. The κ value in percentage is 79.13% against 74.47% for the spectral approach and 73.25% for the EMP, and the differences are statistically significant (see Table VI). Regarding the class-specific accuracies, the results in terms of accuracies have increased for classes 1, 7, and 8 when compared to both individual approaches. In fact, all the classes are more accurately classified than the worst respective cases for the individual approaches.

In the last experiment, feature reduction was applied on the morphological data and the original data before the concatenation. Then, the stacked vector was classified by the SVM. Table VII gives the summary of the test accuracies for several

B. University Area Data Set

PCs were computed from the hyperspectral data. The results for the eigenvalues are shown in Table IV. The left column gives the component number, the center column the eigenvalues in percentage of the total amount of variance, and the right column the cumulative amount of variance. From the table, three PCs were necessary to retain 99% of the variance criterion. EMPs were built according to the scheme presented in Section V: A circular SE with a step size increment of two was used. Four openings and closings were computed for each PC, resulting in an EMP of dimension 27.

First, the classification with SVM was done by using the spectral information and the EMP. The best ML accuracy was obtained by using eight features extracted with the NWFE, following Landgrebe’s recommendations in [3]. The results are reported in Table V. Regarding the global accuracies, both SVM approaches perform equally well, for instance, the difference between the classification using the spectral information and the EMP is not statistically significant (see Table VI).

In the last experiment, feature reduction was applied on the morphological data and the original data before the concatenation. Then, the stacked vector was classified by the SVM. Table VIII gives the summary of the test accuracies for several
values of the variance criterion for the DBFE and NWFE. Best results were obtained with 95% and 80% variance criterion for the DBFE and NWFE, respectively. By using 95% of the variance criterion with DBFE, the hyperspectral data were reduced to 27 features and the EMP to 10 features. With NWFE and 80%, seven features were extracted from the hyperspectral data and six from the EMP. Again, as can be seen in Table VI, differences between the classification accuracies are statistically significant.

Considering the class-specific accuracies, the DBFE approach improved the classification for class 2, whereas class 3 was less accurately classified than with the concatenated full hyperspectral data and EMP. However, the DBFE outperformed the individual classifications of the spatial or spectral information. On this data set, the classification of the DBFE-feature-extracted data gave the best classification results. Similar comments can be made for the accuracies obtained with classifications of the NWFE. Still, the number of features needed to achieve the same accuracy is significantly lower for the NWFE approach than for the DBFE. Since the SVM is linearly related to the dimensionality of the data, lower dimensional data reduced the training time and increased the speed of the classification.

To assess this increase, comparison of the processing time (training and classification process) for the different approaches was made. Table VIII gives the summary of the results which are clearly different according to the features used. The training time could depend on several factors:

1) the dimension of the data;
2) the size of the training set;
3) the number of parameters for the kernel.

For our given problem, items 2) and 3) are the same. Reducing the size of the data is beneficial for the processing time, since data with lower dimensionality (EMP and NWFE) have the shortest processing time. For the best case (NWFE), the gain is about 73%.

### TABLE IX

<table>
<thead>
<tr>
<th>Value</th>
<th>Cumulative Val.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>82.94</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>14.82</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>1.70</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.19</td>
</tr>
</tbody>
</table>

For the classification processing time, two factors have an influence: the dimension of the data and the number of support vectors [nonzero $\alpha_i$ in (8)]. Thus, approaches with low dimensionality and few support vectors perform the classification task of the whole image faster (EMP and NWFE). Nevertheless, the classification processing is really fast by comparison to the training time, in all the cases.

Classification maps for the different approaches are shown in Fig. 5.

**C. Pavia Center Data Set**

For the second test, the scene is a very dense urban area in the center of the city of Pavia. Because of that, morphological information should be useful for the discrimination. PCs were computed from the hyperspectral data. The results for the eigenvalues are shown in Table IX. From the table, three PCs were necessary to retain 99% of the variance criterion. The EMP was built according to the scheme presented in Section V: A circular SE with a step size increment of two was used. Four openings and closings were computed for each PC, resulting in an EMP of dimension 27.

SVM classification was applied to the original hyperspectral data and the EMP. The best ML accuracy was obtained by using 29 features extracted with the DBFE. The results are reported in Table X. From the table, it can be seen that SVM classifier achieved excellent global accuracies. In these experiments, the morphological approach performs better than the spectral-based approach. Table XI shows the statistical significance.
TABLE X
Pavia Center, Summary of the Global and the Class-Specific Test Accuracies in Percentage for SVM Classification. The Numbers of the Features from the Spectral Data and the Morphological Data, Respectively, Are Given in Brackets

<table>
<thead>
<tr>
<th>Features</th>
<th>ML (DBFE)</th>
<th>Spectral EMP</th>
<th>Spec. EMP</th>
<th>DBFE 99%</th>
<th>NWFE 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>94.50</td>
<td>97.67</td>
<td>98.69</td>
<td>99.69</td>
<td>98.65</td>
</tr>
<tr>
<td>102</td>
<td>94.00</td>
<td>95.60</td>
<td>97.69</td>
<td>98.07</td>
<td>97.30</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>96.71</td>
<td>98.15</td>
<td>98.15</td>
<td>98.10</td>
</tr>
<tr>
<td>Class 1</td>
<td>91.50</td>
<td>98.35</td>
<td>99.08</td>
<td>98.66</td>
<td>99.17</td>
</tr>
<tr>
<td>Class 2</td>
<td>92.00</td>
<td>91.23</td>
<td>91.62</td>
<td>93.52</td>
<td>90.00</td>
</tr>
<tr>
<td>Class 3</td>
<td>97.70</td>
<td>96.76</td>
<td>96.18</td>
<td>95.95</td>
<td>96.54</td>
</tr>
<tr>
<td>Class 4</td>
<td>86.90</td>
<td>88.45</td>
<td>98.40</td>
<td>98.77</td>
<td>98.92</td>
</tr>
<tr>
<td>Class 5</td>
<td>95.60</td>
<td>96.97</td>
<td>98.81</td>
<td>99.42</td>
<td>99.27</td>
</tr>
<tr>
<td>Class 6</td>
<td>94.40</td>
<td>96.32</td>
<td>97.98</td>
<td>98.36</td>
<td>98.45</td>
</tr>
<tr>
<td>Class 7</td>
<td>96.40</td>
<td>96.01</td>
<td>97.89</td>
<td>98.22</td>
<td>97.91</td>
</tr>
<tr>
<td>Class 8</td>
<td>99.30</td>
<td>99.40</td>
<td>99.74</td>
<td>99.79</td>
<td>99.81</td>
</tr>
<tr>
<td>Class 9</td>
<td>92.30</td>
<td>99.93</td>
<td>99.44</td>
<td>99.93</td>
<td>98.60</td>
</tr>
</tbody>
</table>

TABLE XI
Pavia Center, Statistical Significance of Differences in Classification Accuracies

<table>
<thead>
<tr>
<th>EMP/Spectral</th>
<th>EMP/Spec. EMP</th>
<th>EMP/DBFE 99%</th>
<th>EMP/NWFE 99%</th>
<th>Spec. EMP/DBFE 95%</th>
<th>Spec. EMP/NWFE 80%</th>
<th>DBFE 95%/NWFE 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.74</td>
<td>-0.06</td>
<td>1.44</td>
<td>-8.14</td>
<td>1.42</td>
<td>-7.75</td>
<td>-9.44</td>
</tr>
</tbody>
</table>

TABLE XII
Pavia Center, Global Accuracies in Percentage With Different FE Methods. The Numbers of Features from the Spectral Data and the Morphological Data, Respectively, Are in Brackets

<table>
<thead>
<tr>
<th>Feature extraction</th>
<th>Cum. Var.</th>
<th>Features</th>
<th>OAA</th>
<th>AA</th>
<th>κ</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBFE</td>
<td>99</td>
<td>66 (51,15)</td>
<td>98.65</td>
<td>97.30</td>
<td>98.10</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>41 (31.10)</td>
<td>98.37</td>
<td>96.86</td>
<td>97.70</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>31 (23.8)</td>
<td>98.08</td>
<td>96.71</td>
<td>97.29</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>22 (17.5)</td>
<td>98.53</td>
<td>97.22</td>
<td>97.42</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>17 (13.4)</td>
<td>98.53</td>
<td>97.31</td>
<td>97.42</td>
</tr>
<tr>
<td>NWFE</td>
<td>99</td>
<td>66 (44.20)</td>
<td>98.87</td>
<td>97.95</td>
<td>98.41</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>31 (19.12)</td>
<td>98.58</td>
<td>96.66</td>
<td>97.90</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>21 (12.9)</td>
<td>98.41</td>
<td>97.28</td>
<td>97.71</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>14 (8.6)</td>
<td>98.24</td>
<td>96.63</td>
<td>97.52</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>10 (6.4)</td>
<td>98.39</td>
<td>96.39</td>
<td>97.73</td>
</tr>
</tbody>
</table>

of differences between the classification accuracies for the different approaches. This is consistent with the characteristics of the picture: It is a very dense urban area, and morphological processing provides discriminative information. In terms of accuracies, the main improvement in the classification is achieved for class 4. The other classes are classified equally accurately. The data fusion should thus improve the classification of class 4 while preserving very good results for the other classes.

Next, the experiment was performed by using the concatenated vector. The vector was made of the 102 spectral bands and the 27 features of the EMP. This vector was used as an input for the SVM without any additional processing. The classification results are reported in Table X. The differences of classification accuracies between the EMP and the concatenated vector are not statistically significant, since the McNemar’s test is almost equal to zero (see Table XI). Thus, both EMP and concatenated vector perform equally well.

As in the previous experiment, feature reduction was applied both on the morphological data and on the original data before the concatenation. Then, the stacked vector was classified by the SVM. Table XII gives the summary of the test accuracies for several values for the variance criterion for the DBFE and NWFE. The best results are obtained with 99% variance criterion for both DBFE and NWFE. By using 99% of the variance of the DBFE, the hyperspectral data are reduced to 51 features, and the EMP is reduced to 15 features. With the NWFE and 99% of the variance criterion, 44 features were extracted from the hyperspectral data and 20 from the EMP. The results are given in Table X.

For this experiment, the DBFE does not help for the classification since the Z-test is not significant. On the other hand, similar classification accuracy is reached with far less features, nearly half the size of the previous feature set, thus decreasing the total training and classification time. The NWFE leads to a significant increase of the classification accuracies, \( |Z| = 7.75 \), by comparison to the best results obtained with the concatenation vector, which is contrary to the previous experiment. Classification maps for the different approaches are shown in Fig. 6. Visually, the thematic map produced with the classification of the NWFE features seems less noisy than the one obtained with the classification of the DBFE features. This is particularly true in the top-left corner which corresponds to a very dense urban area.

Regarding the computing time, the results for the training and the classification are reported in Table XIII. As expected, using FE methods reduces the processing time for both the training and the classification.

Classification maps for the different approaches are shown in Fig. 6.

D. Small Training Set Experiment: University Area

To assess the effectiveness of the proposed methodology for a limited training set, we have randomly extracted a few
training samples from the training set. For this experiment, we used 20 samples for each class, which represents less than 5% of the original training set. We have used the same EMP but had some problems with the DBFE; the covariance matrix was noninvertible (the NWFE does not suffer from this problem). In order to overcome this shortcoming and to apply the DBFE anyway, we use the leave out on covariance to estimate the covariance matrix and perform a statistical enhancement with unlabeled samples; both algorithms were implemented in the MultiSpec software [37], [38]. We have repeated the training sample selection and the classification process five times, and the mean classification results are reported in this paper.

As with the previous experiments, we perform the classification using the spectral or the morphological feature with SVM. The ML produced very poor results, simply close to random classification and hence not reported. The global accuracies are reported in Table XIV. Statistical significance of differences is reported in Table XV.

First of all, the test results are lower than those in Tables V and VII, due to the limited training set. For instance, with the concatenated feature vector, the OA and the $\kappa$ are, respectively, 83.53% and 79.13% for the original training set, whereas using a limited training set, the OA and the $\kappa$ are, respectively, 75.35% and 68.66%. Nevertheless, with a very small training set, the results are still good.

For FE, NWFE with 99% of the cumulative variance provides the best results: The obtained OA is 85.42%, and the $\kappa$ is 80.87%, which is closed to the best results obtained with the full training set (OA = 97.87% and $\kappa$ = 84.40%, see Table V). The $|Z|$ between the best results with limited training and the best results with full training set is equal to 13.65.

Furthermore, the accuracies are better than those obtained with the full training set with the spectral or morphological information alone. It is also important to note that NWFE outperforms the DBFE without any statistical enhancement.

Considering the processing time, with only 20 samples for each class, the training, as well as the classification of the entire data set, is done in 1 or 2 s.

VII. Conclusion

Classification of hyperspectral data with a fine spatial resolution has been investigated. The contribution of this paper is a methodology to include both spatial and spectral information in the classification process by a data fusion scheme. Experimental results on two ROSIS data sets showed excellent accuracies and improvements compared to those obtained with pixel-based classifiers and the EMP-based classifier.

The use of FE was motivated by the fact that the full stacked vector contains a lot of redundancies because there is a redundancy in the hyperspectral data [3] as well as in the EMP [11], which was confirmed by the experiments. On the other hand, SVMs are known to be robust to dimensionality. Therefore, the use of feature reduction for SVMs could be disputable. However, in the experiments, lower dimensional data decreased the processing time, which can be crucial for some applications, and more importantly, it has been shown that SVMs can suffer from the dimensionality when many features are irrelevant [39]. By construction, the stacked vector may contain many copies of the same information, and an FE step may finally be needed to ensure correct classification on every data set, which is confirmed by the experiments (usefulness of
feature reduction for the classification of remote sensing data with SVMs was also assessed in [40]).

It is clear that FE helps for the classification of hyperspectral data, but it is not clear which one of the FE methods should be used for the fusion of morphological and spectral features. From a theoretical point of view, the NWFE was derived because of some intrinsic problems with the DBFE [3], i.e., “DBFE can involve lengthy calculations, and more significantly, it does not perform as well for small numbers of training samples.” Hence, the NWFE might be more preferable, particularly when a small training set is available. The experiments performed with a limited training set confirmed this.

In conclusion, the proposed fusion method succeeded in taking advantage of the spatial and the spectral information simultaneously. It outperformed previous results [2], [10]. Our current research is oriented to the definition of additional spatial features, such as textural characteristics [41], to be included in the feature vectors.

ACKNOWLEDGMENT

The authors would like to thank Mr. J. A. Palmason for his contributions, the IAPR—TC7 for providing the data, and Prof. P. Gamba and Prof. F. Dell’Acqua of the University of Pavia, Italy, for providing reference data. The authors would also like to thank the reviewers for their many helpful comments.

REFERENCES


[38] B. M. Shahshahani and D. A. Landgrebe, “The effect of unlabeled samples has published extensively in those fields.


Mathieu Fauvel (S’06) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in image and signal processing from the Grenoble Institute of Technology (INPG), Grenoble, France, in 2004, 2004, and 2007, respectively.

He is currently with the Grenoble Images Speech Signals and Automatics Laboratory, Signal and Image Department, INPG. He is also with the Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland. His research interests include remote sensing, data fusion, pattern recognition, multicomponent signal, and image processing.

Dr. Fauvel is a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Jocelyn Chanussot (M’04–SM’04) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (INPG), Grenoble, France, in 1995, and the Ph.D. degree from Savoie University, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l’Arment (DGA-French National Defense Department). Since 1999, he has been with INPG, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of signal and image processing. He is currently conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory, INPG. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and Remote Sensing Recognition. He is the Cochair of the GRS Data Fusion Technical Committee and a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society. He is the founding President of IEEE Geoscience and Remote Sensing French chapter.

Jón Atli Benediktsson (S’84–M’90–SM’99–F’04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively.

He was a Fellow with the Australian Defence Force Academy, Canberra, A.C.T., in August 1997. From 1999 to 2004, he was a Chairman of the Energy Company Metan Ltd. He has held visiting positions with the School of Electrical and Computer Engineering, Purdue University (1995), the Joint Research Centre of the European Commission, Ispra, Italy (1998), Denmark’s Technical University, Lyngby, (1998), the School of Computing and Information Systems, Kingston University, Kingston upon Thames, U.K. (1999–2004), and the Department of Information and Communication Technology, University of Trento, Trento, Italy (2002–present). He is currently the Director of Academic Development and Innovation and a Professor of electrical and computer engineering with the University of Iceland. His research interests include remote sensing, pattern recognition, neural networks, image processing, and signal processing, and he has published extensively in those fields.

Johannes R. Sveinsson (S’86–M’90–SM’02) received the B.S. degree in electrical engineering from the University of Iceland, Reykjavik, and the M.S. (Eng.) and Ph.D. degrees in electrical engineering from Queen’s University, Kingston, ON, Canada. From 1981 to 1982, he was with the Laboratory of Information Technology and Signal Processing, University of Iceland. From 1985 to 1986, he was a Visiting Research Student with the Imperial College of Science and Technology, London, U.K. At Queen’s University, he held teaching and research assistantships and received Queen’s Graduate Awards. From November 1991 to 1998, he was a Senior Member of the Research Staff and a Lecturer with the Engineering Research Institute and the Department of Electrical Engineering, respectively, of the University of Iceland, where he is currently a Professor with the Department of Electrical and Computer Engineering. His research interests include systems and signal theory.