Multi-resolution Segmentation and Shape Analysis for Remote Sensing Image Classification

Selim Aksoy and H. Gökhan Akçay
Bilkent University
Department of Computer Engineering
Bilkent, 06800, Ankara, Turkey
{saksoy,akcay}@cs.bilkent.edu.tr

Abstract— **We present an approach for classification of remotely sensed imagery using spatial information extracted from multi-resolution approximations. The wavelet transform is used to obtain multiple representations of an image at different resolutions to capture different details inherently found in different structures. Then, pixels at each resolution are grouped into contiguous regions using clustering and mathematical morphology-based segmentation algorithms. The resulting regions are modeled using the statistical summaries of their spectral, textural and shape properties. These models are used to cluster the regions, and the cluster memberships assigned to each region in multiple resolution levels are used to classify the corresponding pixels into land cover/land use categories. Final classification is done using decision tree classifiers. Experiments with two ground truth data sets show the effectiveness of the proposed approach over traditional techniques that do not make strong use of region-based spatial information.**

I. Introduction

Remote sensing image analysis has been an important research area for the last four decades. There is also an extensive literature on classification of remotely sensed imagery using parametric or nonparametric statistical or structural techniques [1]. Advances in satellite technology and computing power have enabled the study of multi-modal, multi-spectral, multi-resolution and multi-temporal data sets for applications such as urban land use monitoring and management, GIS and mapping, environmental change, site suitability, agricultural and ecological studies.

The usual choice for the level of processing image data has been pixel-based analysis in both academic and commercial remote sensing image analysis systems. However, a recent study [2] that investigated classification accuracies reported in the last 15 years showed that there has not been any significant improvement in the performance of classification methodologies over this period. We believe that the reason behind this problem is the fact that there is a large semantic gap between the low-level features used for classification and the high-level expectations and scenarios required by the users.

Remote sensing experts use spatial information to interpret the land cover because pixels alone do not give much information about image content. Image segmentation techniques [3] automatically group neighboring pixels into contiguous regions based on similarity criteria on pixels' properties. Even though image segmentation has been heavily studied in image processing and computer vision fields, and despite the early efforts [4] that use spatial information for classification of remotely sensed imagery, segmentation algorithms have only recently started receiving emphasis in remote sensing image analysis. Examples of image segmentation in the remote sensing literature include region growing [5] and Markov random field models [6] for segmentation of natural scenes, hierarchical segmentation for image mining [7], region growing for object level change detection [8], and boundary delineation of agricultural fields [9].

We model spatial information by segmenting images into spatially contiguous regions and classifying these regions according to the statistics of their pixel properties and shape features. To develop segmentation algorithms that group pixels into regions, first, we perform multi-resolution analysis using wavelets [10], [11] to model image content in different levels. These levels are used to capture different details inherently found in different structures. Then, each image obtained after the multi-resolution analysis is segmented using clustering-based and mathematical morphology-based algorithms. These algorithms are developed to produce oversegmented regions, especially at higher resolution levels, to capture the details of small structures.

Resulting regions are modeled using the statistical summaries of their spectral and textural properties along with shape features that are computed from region polygon boundaries. Then, these attributes are used as features to cluster the regions. Finally, the cluster memberships assigned to each region in multiple levels of the resolution hierarchy are used to classify the corresponding pixels into land cover/land use categories defined by the user. Final classification is done using decision tree classifiers.

The rest of the paper is organized as follows. Multi-resolution analysis based on wavelets is presented in Section II. Segmentation of images is described in Section III. Feature data used for modeling pixels and regions are described in Section IV. Classification of pixels and regions are discussed in Section V. Experiments are presented in Section VI and the approach is summarized in Section VII.

This work was supported by the TUBITAK CAREER Grant 104E074.

0-7803-8977-8/05/$20.00 ©2005 IEEE.

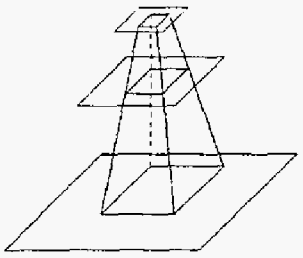
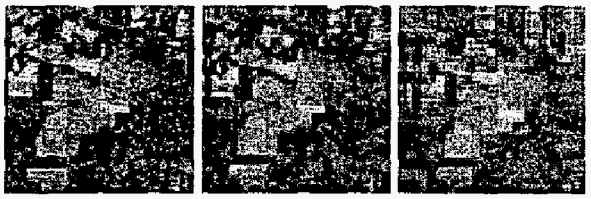
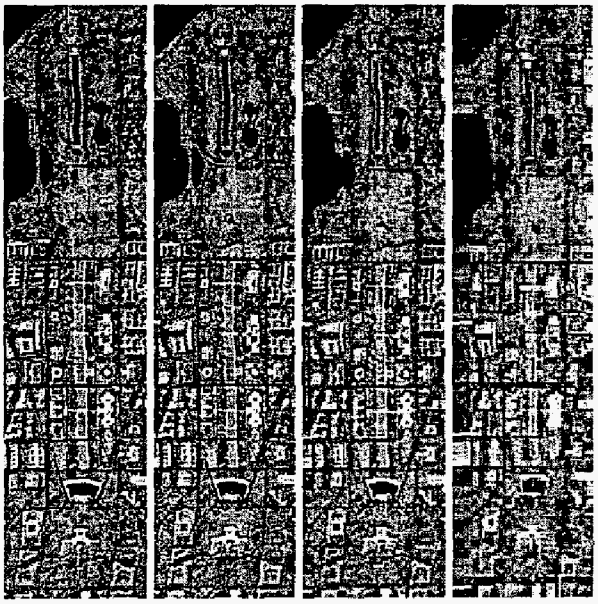


Fig. 1. Wavelet pyramid representation for multiple resolutions. The bottom image is at the original resolution. Images that get smaller towards top are results of successive wavelet decompositions.



(a) 145×145 pixels (original) (b) 76×76 pixels (c) 41×41 pixels

Fig. 2. Wavelet decomposition of the Indian Pine data set. The bands 50, 27 and 17 were used to generate the false color images and these images were resized to show the details.



(a) $1,280 \times 307$ (orig.) (b) 643×157 pixels (c) 325×82 pixels (d) 166×44 pixels

Fig. 3. Wavelet decomposition of the DC Mall data set. The bands 63, 52 and 36 were used to generate the false color images and these images were resized to show the details.

II. Multi-resolution Analysis

Different physical structures that we want to recognize in images may generally have very different sizes. In order to cope with this variability, either the features used must be designed to be size invariant or the image must be processed at different resolutions, since different resolutions characterize different structures in the image. As the resolution gets coarser from that of the original image, larger structures that provide the general image context can be represented without being convoluted with the details. It is therefore natural to analyze first the image content at a coarse resolution and then gradually increase the resolution [10]. This process is also similar to the strategy used by the human vision system [11].

In this work, we use the wavelet transform [10], [11] to obtain multiple representations of an image at different resolutions. The wavelet transform provides a hierarchical framework for interpreting the image. At each level of the hierarchy, the image is passed through a low-pass filter that provides a smooth approximation, and a band-pass filter that captures the details. After the filtering, the corresponding images are subsampled by two and the resolution is reduced by half. This procedure can be repeated for further decomposition using a filter bank. The low-pass filtered versions can be used as the representations that best approximate the original image at multiple resolutions.

The resulting low-pass filtered smooth approximations can be represented as a pyramid as shown in Fig. 1. Let 2^j ($j \leq 0$) represent the resolution of an image where $j = 0$ is the resolution of the original image. A single pixel at resolution 2^j covers a block of 2^{-j} pixels in the original image. Processing of a spatial neighborhood of size K at this resolution is equivalent to processing over a neighborhood of size $2^{-j}K$ in the original image. Therefore, the coarse information is analyzed over large neighborhoods whereas the detail information is analyzed over small neighborhoods [11].

The wavelet decompositions of the images used in the experiments in this paper are shown in Figs. 2 and 3. The first image consists of a 145×145 section of an AVIRIS data set with 220 spectral bands recorded over a mixed agriculture/forestry landscape in the Indian Pine Test Site [1]. The second image consists of an airborne hyperspectral data flightline over the Washington DC Mall area and has $1,280 \times 307$ pixels with 191 spectral bands [1]. For hyperspectral images like these, we apply the wavelet decomposition independently to each spectral band to form new images at lower resolutions with the same number of bands.

III. Image Segmentation

Different wavelet levels capture different details inherently found in different structures. Image segmentation is used to group pixels that belong to the same structure with the goal of delineating each individual structure as an individual region. We have experimented with several segmentation algorithms from the computer vision literature. Algorithms that are based on graph clustering [12], mode seeking [13] and classification [14] have been reported to be successful in moderately sized color images with relatively homogeneous

structures. However, we could not apply these techniques successfully to our data sets because the huge amount of data in hyperspectral images made processing infeasible due to both memory and computational requirements, and the detailed structure in high resolution remotely sensed imagery restricted the use of sampling that has been often used to reduce the computational requirements of these techniques.

The segmentation approach we have used in this work consists of clustering and mathematical morphology. First, the k-means algorithm [15] is used to cluster the spectral data. After this unsupervised clustering step, each pixel is assigned the label of the cluster that it belongs in the spectral feature space. Since the k-means algorithm uses only spectral information and ignores spatial correlations, the resulting segmentation may contain isolated pixels with labels different from those of their neighbors. We use an iterative split-and-merge algorithm [16] to convert this intermediate step to contiguous regions as follows:

1) Merge pixels with identical labels to find the initial set of regions and mark these regions as foreground,
2) Mark regions with areas smaller than a threshold as background using connected components analysis [3],
3) Use region growing to iteratively assign background pixels to the foreground regions by placing a window at each background pixel and assigning it to the label that occurs the most in its neighborhood.

This procedure corresponds to a spatial smoothing of the clustering results. We further process the resulting regions using mathematical morphology operators [3] to automatically divide large regions into more compact sub-regions as follows [16]:

1) Find individual regions using connected components analysis for each label,
2) For all regions, compute the erosion transform [3] and repeat:
   a) Threshold erosion transform at steps of 3 pixels in every iteration,
   b) Find connected components of the thresholded image,
   c) Select sub-regions that have an area smaller than a threshold,
   d) Dilate these sub-regions to restore the effects of erosion,
   e) Mark these sub-regions in the output image by masking the dilation using the original image,

   until no more sub-regions are found,
3) Merge the residues of previous iterations to their smallest neighbors.

The parameters for the algorithms were empirically chosen to produce oversegmented regions, especially at higher resolution levels, to capture the details of small structures. For example, the value of k for clustering was set to powers of 2 between 2 and 16, and the minimum area threshold for merging was set to multiples of 2 linearly between 4 and 10 pixels for the wavelet levels $j = 0, -1, -2, -3$. The neighborhood size for growing was fixed as 3×3 for all levels. Fig. 4 shows examples of segmentation results for different resolutions.

(a) Indian Pine segmentation (b) DC Mall segmentation

Fig. 4. Segmentation examples for the Indian Pine and the DC Mall data sets. For both sets, the first image shows segmentation at the original resolution and the second image shows segmentation at the second wavelet level ($j = -2$). Region boundaries are marked as white.

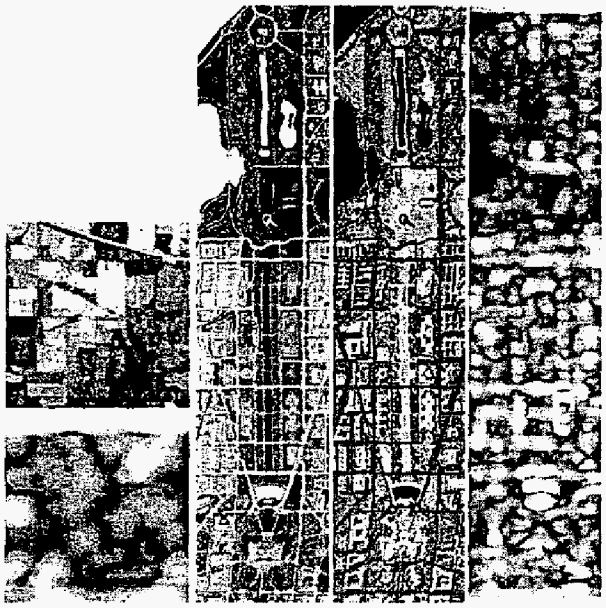
IV. Feature Extraction

We use multiple feature representations for both pixels and segmented regions. These features correspond to spectral, textural and shape properties, and are described below.

A. Pixel Level Data

The DC Mall and Indian Pine images consist of 191 and 220 spectral bands, respectively. To simplify computations and to avoid the curse of dimensionality, we use the 9-band subset that came with the original data for the Indian Pine data set. For the DC Mall data set, we apply Fisher's linear discriminant analysis (LDA) [15] that finds a projection to a new set of bases that best separates the data in a least-squares sense. The resulting number of bands for this set is 6 (corresponding to 7 classes as described in Section VI). These values are used as spectral features.

We also apply principal components analysis (PCA) [15] to images at each resolution and then extract Gabor texture features [17] by filtering the first principal component image at each resolution with Gabor kernels at different scales and orientations. We used kernels rotated by $n\pi/4, n = 0, \ldots, 3$, at three scales resulting in feature vectors of length 12. Examples for pixel features are shown in Fig. 5.



(a) Indian Pine features (b) DC Mall features

Fig. 5. Pixel feature examples for the Indian Pine and the DC Mall data sets. In (a), the first PCA band and the corresponding Gabor image for 0 degree orientation at the third scale are shown from top to bottom. In (b), the first LDA band, the first PCA band and the corresponding Gabor image for 0 degree orientation at the third scale are shown from left to right. Histogram equalization was applied to all images for better visualization.

B. Region Level Data

Regions are modeled using the statistical summaries of their spectral and textural properties along with shape features that are computed from region polygon boundaries. The statistical summary for a region is computed as the means and standard deviations of features of the pixels in that region. The shape properties of a region correspond to its area, orientation of the region's major axis with respect to the x axis, eccentricity (ratio of the distance between the foci to the length of the major axis; e.g., a circle is an ellipse with zero eccentricity), Euler number (1 minus the number of holes in the region), solidity (ratio of the area to the convex area), extent (ratio of the area to the area of the bounding box), spatial variances along the x and y axes, and spatial variances along the region's principal (major and minor) axes [3], resulting in a feature vector of length 10.

V. Image Classification

Image classification is usually done by using pixel features as input to classifiers such as minimum distance, maximum likelihood, neural networks or decision trees. However, large within-class variations and small between-class variations of these features at the pixel level and the lack of spatial information limit the accuracy of these classifiers.

In this work, we perform classification using region level information. First, the region features at each resolution are clustered using the k-means algorithm. This process assigns a cluster label to each region for each feature used. In particular, for each region at each resolution, we obtain three labels from

- clustering of the statistics of the original 9 bands,
- clustering of the statistics of the 12 Gabor bands,
- clustering of the 10 shape features,

respectively, for the Indian Pine data set, and three labels from

- clustering of the statistics of the 6 LDA bands,
- clustering of the statistics of the 12 Gabor bands,
- clustering of the 10 shape features,

respectively, for the DC Mall data set.

These region level labels can be converted to pixel level features by collecting the labels of the regions at multiple resolutions corresponding to each pixel. We use two successive wavelet levels as the multi-resolution approximation for both data sets. Therefore, a pixel at the original resolution is assigned a new feature vector of length 9 containing three values from the corresponding region at the original resolution, three values from the corresponding region at the first wavelet level, and three values from the corresponding region at the second wavelet level. Region correspondences are found using the dependencies between different resolutions as shown in the wavelet pyramid in Fig. 1.

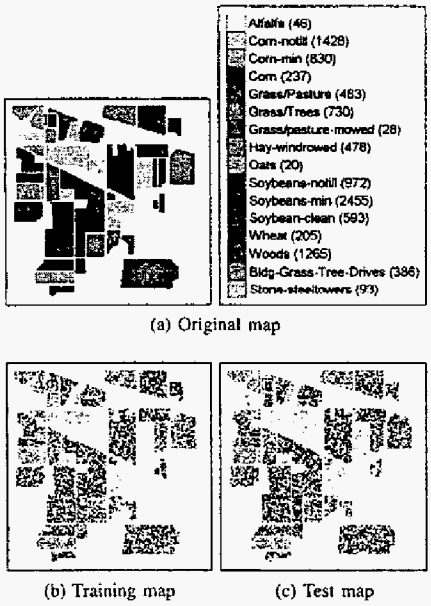
In the next section, we evaluate the performance of these new features for classifying pixels into land cover/land use categories defined by the user. Classification is done using a binary decision tree classifier with the gini impurity criterion [15], and its performance is compared to that of a traditional maximum likelihood classifier with the multivariate Gaussian with full covariance matrix assumption for each class.

VI. Experiments

The proposed algorithms were evaluated using the Indian Pine and DC Mall data sets. Multi-resolution analysis, image segmentation and feature extraction were applied to both images as described in the previous sections. Finally, region features were extracted and classifiers were trained using the corresponding pixel features.

The 16 land cover classes that were used for the Indian Pine data set include alfalfa, corn-notill, corn-min, corn, grass/pasture, grass/trees, grass/pasture-mowed, hay-windrowed, oats, soybeans-notill, soybeans-min, soybean-clean, wheat, woods, building-grass-tree-drives, and stone-steeltowers. A thematic map with ground truth labels for 10,249 pixels was supplied with the original data [1]. The ground truth was divided into half as independent training (5,128 pixels) and test (5,121 pixels) sets. Both training and testing were done using the labels at the original resolution. Details are given in Fig. 6.

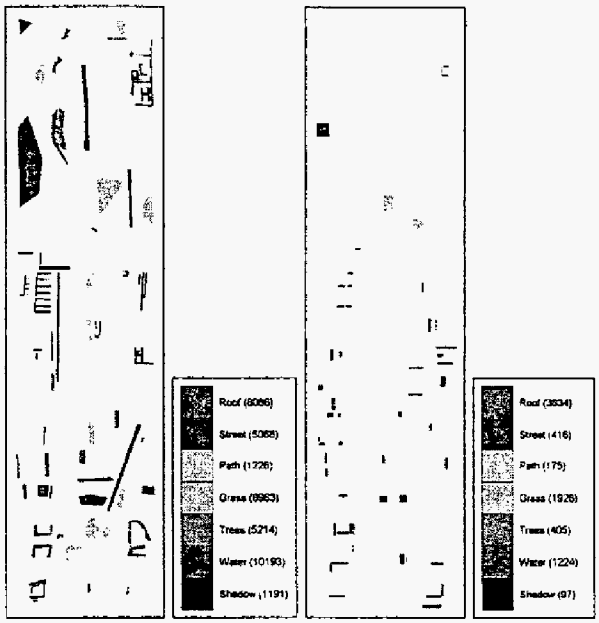
The 7 land cover classes that were used for the DC Mall data set include roof, street, path, grass, trees, water, and shadow. A thematic map with ground truth labels for 8,079 pixels was supplied with the original data [1]. We used this ground truth for testing and separately labeled 37,941 pixels for training. Both training and testing were done using the labels at the original resolution. Details are given in Fig. 7.



(a) Original map

(b) Training map (c) Test map

Fig. 6. Training and testing ground truth maps for the Indian Pine data set. The number of pixels for each class are shown in parenthesis in the legend.



(a) Training map (b) Test map

Fig. 7. Training and testing ground truth maps for the DC Mall data set. The number of pixels for each class are shown in parenthesis in the legend.

Confusion matrices for the cases where all region-based features were used with the decision tree classifier are shown in Tables I and II for the Indian Pine and DC Mall data sets, respectively. Confusion matrices for the maximum likelihood classifier are not given due to page limitations but the results are summarized in Table III. Classification maps are shown in Fig. 8.

The results show that the proposed approach performed significantly better than the traditional maximum likelihood classifier with Gaussian density assumption for the Indian Pine data set and gave comparable results for the DC Mall data set. Using texture features in addition to the spectral ones improved the performance of both approaches. In addition, using multi-resolution approximation and spatial information with region features and shape properties improved the results for the proposed approach further but the maximum likelihood classifier could not avoid producing groups of misclassified pixels due to the lack of spatial information.

VII. Summary

We have presented an approach for classification of remotely sensed imagery using multi-resolution and spatial techniques. Wavelet decomposition was used to model image content in different levels. Then, each resolution level was independently segmented into contiguous regions using clustering and mathematical morphology-based algorithms. The resulting regions were modeled using the statistical summaries of their spectral and textural features and shape properties. Then, these models were used to cluster the regions, and the cluster labels assigned to each region in multiple levels of the resolution hierarchy were used to classify the corresponding pixels with a decision tree classifier. We investigated the performance of multi-resolution analysis and usefulness of different region features in classification. Experiments with two data sets showed the effectiveness of the proposed approach over the traditional maximum likelihood classifier because of the use of spatial information extracted from multi-resolution approximations.

Acknowledgment

We would like to thank Prof. David A. Landgrebe and Mr. Larry L. Biehl from the Laboratory for Applications of Remote Sensing, Purdue University, Indiana, U.S.A., for the Indian Pine and DC Mall data sets.

References

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons, Inc., 2003.

[2] G. G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 433–440, March 2005.

[3] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Addison-Wesley, 1992.

[4] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Transactions on Geoscience Electronics*, vol. GE-14, no. 1, pp. 19–26, January 1976.

[5] C. Evans, R. Jones, I. Svalbe, and M. Berman, "Segmenting multispectral Landsat TM images into field units," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 5, pp. 1054–1064, May 2002.

[6] A. Sarkar, M. K. Biswas, B. Kartikeyan, V. Kumar, K. L. Majumder, and D. K. Pal, "A MRF model-based segmentation approach to classification for multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 5, pp. 1102–1113, May 2002.

TABLE I

CONFUSION MATRIX WHEN ALL REGION FEATURES WERE USED WITH THE DECISION TREE CLASSIFIER FOR THE INDIAN PINE DATA SET. CLASSES WERE LISTED IN FIG. 6.

True labels	Assigned labels																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	22	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	23
2	0	669	1	0	0	0	0	0	0	13	19	10	0	2	0	0	714
3	0	0	397	2	0	1	0	0	1	5	2	0	7	0	0	0	415
4	0	0	2	116	0	0	0	0	0	0	0	0	0	0	0	0	118
5	0	1	5	0	234	0	1	0	0	0	0	0	0	0	0	0	241
6	0	0	0	0	0	361	0	0	0	0	2	0	0	2	0	0	365
7	0	0	0	0	6	0	8	0	0	0	0	0	0	0	0	0	14
8	0	0	0	0	0	0	0	239	0	0	0	0	0	0	0	0	239
9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	10
10	0	11	2	0	0	0	0	0	0	423	25	25	0	0	0	0	486
11	5	29	0	0	0	5	0	0	0	5	1,181	0	0	2	0	0	1,227
12	0	0	1	1	0	0	0	0	0	2	0	290	0	0	0	2	296
13	0	0	0	0	1	0	0	0	0	0	0	0	101	0	0	0	102
14	0	0	0	0	3	1	0	0	0	0	2	0	0	626	0	0	632
15	0	0	0	0	1	0	0	0	0	0	0	1	0	0	191	0	193
16	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	44	46
Total	27	710	408	119	245	368	9	239	11	448	1,232	328	108	632	191	46	5,121

TABLE II

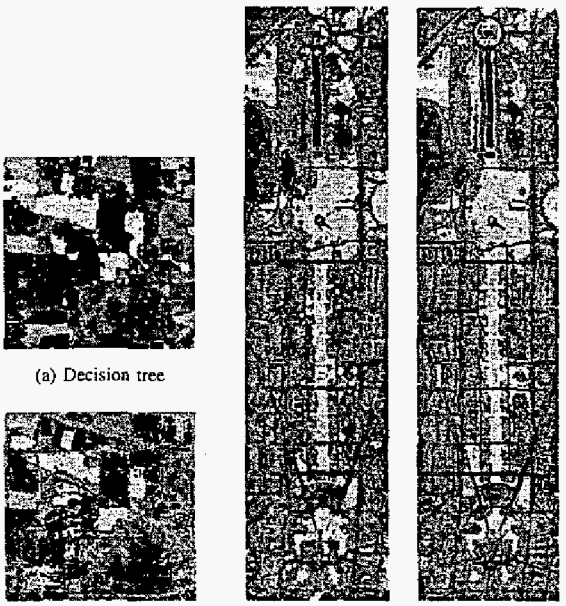
CONFUSION MATRIX WHEN ALL REGION FEATURES WERE USED WITH THE DECISION TREE CLASSIFIER FOR THE DC MALL DATA SET. CLASSES WERE LISTED IN FIG. 7.

True labels	Assigned labels							
	1	2	3	4	5	6	7	Total
1	3,651	159	0	0	21	0	3	3,834
2	21	394	0	0	1	0	0	416
3	0	0	175	0	0	0	0	175
4	0	0	0	1,928	0	0	0	1,928
5	6	0	0	6	393	0	0	405
6	1	0	0	0	2	1,221	0	1,224
7	1	10	0	0	11	0	75	97
Totals	3,680	563	175	1,934	428	1,221	78	8,079

TABLE III

SUMMARY OF CLASSIFICATION ERROR USING THE REGION FEATURES WITH THE DECISION TREE CLASSIFIER AND THE PIXEL FEATURES WITH THE MAXIMUM LIKELIHOOD GAUSSIAN CLASSIFIER.

Datasets	Indian Pine		DC Mall	
	Decision tree	Gaussian	Decision tree	Gaussian
Training	0.0190	0.1153	0.0035	0.0282
Testing	0.0560	0.2261	0.0514	0.0324



(a) Decision tree (b) Gaussian (c) Decision tree (d) Gaussian

Fig. 8. Final classification maps with the decision tree and maximum likelihood Gaussian classifiers for the Indian Pine and DC Mall data sets. Class color codes were listed in Figs. 6 and in Fig. 7, respectively.

[7] J. C. Tilton, G. Marchisio, K. Koperski, and M. Datcu, "Image information mining utilizing hierarchical segmentation," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, Toronto, Canada, June 2002, pp. 1029–1031.

[8] G. G. Hazel, "Object-level change detection in spectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 553–561, March 2001.

[9] A. Rydberg and G. Borgefors, "Integrated method for boundary delineation of agricultural fields in multispectral satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2514–2520, November 2001.

[10] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[11] S. Mallat, "Wavelets for a vision," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 604–614, April 1996.

[12] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, August 2000.

[13] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002.

[14] P. Paclik, R. P. W. Duin, G. M. P. van Kempen, and R. Kohlus, "Segmentation of multi-spectral images using the combined classifier approach," *Image and Vision Computing*, vol. 21, no. 6, pp. 473–482, June 2003.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, Inc., 2000.

[16] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, "Learning Bayesian classifiers for scene classification with a visual grammar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 581–589, March 2005.

[17] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, August 1996.