Image segmentation and discriminant analysis for the identification of land cover units in ecology

Article *in* IEEE Transactions on Geoscience and Remote Sensing · October 1997

DOI: 10.1109/36.628781 · Source: IEEE Xplore

1 author:

Agustin Lobo
Spanish National Research Council
85 PUBLICATIONS **1,096** CITATIONS

Image Segmentation and Discriminant Analysis for the Identification of Land Cover Units in Ecology

Agustín Lobo

Abstract—**The textured nature of most natural land cover units as represented in remotely sensed imagery causes limited results of per-pixel classifications. The segmentation algorithm, Iterative Mutually Optimum Region Merging (IMORM), is presented and used to partition images into elements that are thereafter classified by Linear Canonical Discriminant Analysis and a Maximum Likelihood allocation rule. This per-segment approach results in much higher accuracy than the conventional per-pixel approach. Furthermore, separability matrices indicate that many land cover categories cannot be correctly defined by per-pixel statistics.**

I. Introduction

ONE OF THE greatest difficulties in thematic mapping from digital remotely sensed imagery arises from the fact that land-cover categories defined by the user are not smooth but present a textured appearance. For a given object, the variance of its pixel values is not only due to the random variation of the sensor's response, but also to an intrinsic characteristic of the object itself: objects are not characterized by a uniform reflectance value but by a distribution that has a certain spatial autocorrelation. Typically, the importance of considering texture is more evident for processing images of natural areas. In fact, the term "object" in natural scenes tends to be misunderstood, and we will refer to "land cover elements" in this paper instead. Objects are normally understood within artificial scenes, where they tend to be smoother, closer to simple geometric figures, and defined by sharper edges. These properties are so rarely found in natural landscapes that any cluster of pixels with these characteristics immediately captures our attention, often as suspect of being due to human intervention.

The importance of texture appears to be obvious to most authors in remote sensing. But textural analysis is not yet common, and most thematic maps are obtained through multivariate classification on a "per-pixel" basis. Typically, results of per-pixel classifications need to be postprocessed to be presentable. Nevertheless, even if sometimes sophisticated [1]–[3], postprocessing "does not incorporate true spatial characteristics of the class" [4]. One important reason for the relative underuse of textural information is that, while pixel locations are unequivocally defined in the image, the problem of determining the zones in which to measure texture is not trivial. An alternative approach to the one based on per-pixel classification, was present early in the literature on processing techniques for multispectral imagery [5], [6]. This approach consists in the partition of the image into homogeneous elements that are thereafter classified. Although this approach yielded higher classification accuracy [6], it did not become common practice.

In this paper, we present and evaluate the results of a new image segmentation algorithm [Iterative Mutually Optimum Region Merging (IMORM)] to define elements (segments) that are subsequently discriminated through Linear Canonical Discriminant Analysis (LCDA). In this way, segments rather than pixels become the new elements for the discriminant analysis. Although this protocol is generally applicable to land cover mapping, it was developed for ecological applications. Often, classification of remotely sensed imagery for ecological applications requires legends of land cover categories that cannot be discriminated using conventional per-pixel statistics because of low statistical separability (see Section III). Conventional per-pixel classifications often result in low accuracy unless the final categories are very general. While legends with broad categories may be useful for general inventories [7], the lack of more detailed legends has seriously hindered the use of classified remotely sensed imagery in ecology, where subtle distinctions among categories are frequently necessary. In the case of tropical landscapes this problem increases [8]. Also, most ecological applications need to take advantage of the specific spatial distribution of the different categories, for example, spatial templates for population dynamics models or patterns of landscape change over time. While conventional methods may provide enough accuracy for simple first-order statistics, spatial accuracies are often inappropriate. Yet, spatial accuracy is often important in ecology, where the dynamics of transitional zones (ecotones) are a critical issue because of their role as early detectors of environmental change and their high biodiversity.

The importance of texture to discriminate land cover categories from remotely sensed imagery has been recognized by several authors from long ago and many methods for its analysis have been developed. Marceau [9], [10] gives a summary and distinguishes between methods based on structural algorithms (in which textures are defined by rules of spatial interrelationships of primitive units) and methods based on a statistical approach. In the statistical approach, texture is modeled from the distribution of pixel values over a local area. Methods based on a fractal model of the semivariogram are powerful [11], but computationally too intensive.

Manuscript received November 13, 1995; revised September 3, 1996. This work was supported by the Generalitat de Catalunya, Spain, under Grant BPOST90-4530, the Joint Program Ministry of Education and Science, Spain/Fulbright Committee under Grant FU91-50413540 to A. Lobo, and by grants from the Andrew W. Mellon Foundation and the National Aeronautics and Space Administration under Grant NAGW-3124 to S. Levin, Princeton University, Princeton, NJ.

The author is with the Institut de Ciències de la Terra (CSIC), Martí Franqués, 08028 Barcelona, Spain (e-mail: alobo@ija.csic.es).

Publisher Item Idèntifier S 0196-2892(97)04106-5.

Then the list is screened and only those candidate pairs (j, k) that

i) are mutually optimum in terms of $d_{j,k}$ (see below), and
ii) fulfill, for a user-defined threshold (τ), the condition:

$$\delta_{j,k} < \tau$$

(where $\delta_{j,k}$ is obtained from $d_{j,k}$ as explained below) are selected for merging.

In this paper, a candidate pair of adjacent facets (j, k) is mutually optimum if j is the most similar adjacent facet of k and k is the most similar adjacent facet of j. Woodcock and Harward [23] also use a double criterion of mutual optimality and a threshold, but the actual distances $d_{j,k}$ and $\delta_{j,k}$, and the threshold τ are different. Also, the use of a different metrics for selecting candidate pairs and for the final selection of merging pairs is unique to IMORM.

The selected pairs are given a common label (arbitrarily the lowest between j and k), the table of statistics and the list of adjacencies are actualized, and a new iteration is started.

The process stops when no more mergings are possible, either because there are no more mutually optimum pairs or because $\delta_{j,k} \geq \tau$.

The distance $d_{i,j,k}$ between facets j and k is computed in the PC band i, as the normalized difference between two means of samples of normal distributions [27]

$$d_{i,j,k} = \frac{|m_{i,j} - m_{i,k}|}{\sqrt{\frac{s_{i,j}^2}{n_j - 1} + \frac{s_{i,k}^2}{n_k - 1}}} \quad (1)$$

where $m_{i,j}$ and $m_{i,k}$ stand for the means of facets j and k in the ith PC, the s for the standard deviations and n for the number of pixels.

The denominator of (1) estimates the joint standard deviation of both facets if

$$n_j \geq 30$$

and

$$n_k \geq 30 \quad (2)$$

where 30 is used as a cutoff value between large and small samples [27]. If (2) is not true, then the estimate of the joint standard deviation is more complex [27] and (1) is substituted by

$$d_{i,j,k} = \frac{|m_{i,j} - m_{i,k}|}{\sqrt{\frac{(n_j - 1) \cdot s_{i,j}^2 + (n_k - 1) \cdot s_{i,k}^2}{n_j + n_k - 2} \cdot \frac{n_j + n_k}{n_j \cdot n_k}}}. \quad (3)$$

The multivariate $d_{j,k}$ is simply the sum of $d_{i,j,k}$ over all PC bands

$$d_{j,k} = \sum_{i=1}^{p} d_{i,j,k}. \quad (4)$$

The value $\delta_{j,k}$ to be compared with the threshold τ is

$$\delta_{j,k} = \max_{i=1}^{p} (d_{i,j,k}). \quad (5)$$

As the values $d_{i,j,k}$ are distributed according to a Student's t distribution with $(n_i + n_j - 2)$ degrees of freedom, the user just sets the level of confidence. The threshold τ itself is calculated from the Student's t distribution. Equations (1) and (2) are tests on whether the null hypothesis that *both samples come from the same distribution* can be rejected at a given confidence level. Hence, this threshold criterion would prevent any mutually optimum pair to merge if there is any PC for which the null hypothesis can be rejected. The reason for such a conservative merging criterion is that while further merging of any two segments is an easy task, splitting wrongly merged segments is much more cumbersome.

B. Discriminant Analysis

After the segmentation process, a spatially unconstrained, nonhierarchical clustering is performed by applying LCDA. LCDA implies the input of user knowledge in the process, a knowledge that includes many sources of information other than the particular scene being processed. This knowledge takes the form of a data matrix of training samples, that is, image elements that have been identified. LCDA processes the data matrix of training samples to find a linear transformation that enhances discrimination of the different categories or groups. Discriminant canonical analysis is also a feature reduction technique because it yields n components at most, with n being

$$n^2 \min (\text{number of variables} - 1, \text{number of individuals})$$

and often the number of significant canonical components is even smaller. The actual value of n is given by the Bartlett significance test [28], [29] consecutively applied with the $n-k$ (k ranging from 0 to $n-1$) eigenvalues of the within-groups covariance matrix that remain after acceptance of the first k. The test is compared against the χ^2 value with $(n-k)$ $(g-k-1)$ degrees of freedom for a chosen significance level, where g is the number of groups. Summaries of canonical discriminant analysis for remote sensing can be found in [30] and more general presentations in [29], [31]. Despite its obvious interest, applications in the remote sensing literature are not numerous [32]–[38].

Here, after IMORM, a multivariate data matrix ($\mathbf{X}$) is built with the segments as elements (rows) and the segment's image statistics as variables (columns). In addition to the segment label, the variables are the per-segment means in the spectral bands (or in transformations of them, like the Tasseled Cap or the PC's) and the per-segment coefficients of variation (CV) in PC-1 and PC-2.

Some segments are selected interactively as training fields and their type of land cover is identified. This is done by overlaying the vectorized segmented image over a color composite of the original and applying knowledge acquired in the field and/or with higher-resolution imagery. The vectors of statistics of these segments are extracted from $\mathbf{X}$ and a training field data matrix ($\mathbf{D}$) is built, adding the land cover label as a new column.

The goal is to find a transformation $\mathbf{U}$ such that the classes are optimally separated. LCDA postulates that such a transformation $\mathbf{U}$ is the one in which the ratio of the variance between classes versus the variance within classes is maximized. Discriminant canonical analysis requires the computation of a within-class covariance matrix ($\mathbf{W}$) and a

Then the list is screened and only those candidate pairs (j, k) that

i) are mutually optimum in terms of $d_{j,k}$ (see below), and
ii) fulfill, for a user-defined threshold (τ), the condition:

$$\delta_{j,k} < \tau$$

(where $\delta_{j,k}$ is obtained from $d_{j,k}$ as explained below) are selected for merging.

In this paper, a candidate pair of adjacent facets (j, k) is mutually optimum if j is the most similar adjacent facet of k and k is the most similar adjacent facet of j. Woodcock and Harward [23] also use a double criterion of mutual optimality and a threshold, but the actual distances $d_{j,k}$ and $\delta_{j,k}$, and the threshold τ are different. Also, the use of a different metrics for selecting candidate pairs and for the final selection of merging pairs is unique to IMORM.

The selected pairs are given a common label (arbitrarily the lowest between j and k), the table of statistics and the list of adjacencies are actualized, and a new iteration is started.

The process stops when no more mergings are possible, either because there are no more mutually optimum pairs or because $\delta_{j,k} \geq \tau$.

The distance $d_{i,j,k}$ between facets j and k is computed in the PC band i, as the normalized difference between two means of samples of normal distributions [27]

$$d_{i,j,k} = \frac{|m_{i,j} - m_{i,k}|}{\sqrt{\frac{s_{i,j}^2}{n_j - 1} + \frac{s_{i,k}^2}{n_k - 1}}} \quad (1)$$

where $m_{i,j}$ and $m_{i,k}$ stand for the means of facets j and k in the ith PC, the s for the standard deviations and n for the number of pixels.

The denominator of (1) estimates the joint standard deviation of both facets if

$$n_j \geq 30$$

and

$$n_k \geq 30 \quad (2)$$

where 30 is used as a cutoff value between large and small samples [27]. If (2) is not true, then the estimate of the joint standard deviation is more complex [27] and (1) is substituted by

$$d_{i,j,k} = \frac{|m_{i,j} - m_{i,k}|}{\sqrt{\frac{(n_j - 1)\cdot s_{i,j}^2 + (n_k - 1)\cdot s_{i,k}^2}{n_j + n_k - 2} \cdot \frac{n_j + n_k}{n_j \cdot n_k}}}. \quad (3)$$

The multivariate $d_{j,k}$ is simply the sum of $d_{i,j,k}$ over all PC bands

$$d_{j,k} = \sum_{i=1}^{p} d_{i,j,k}. \quad (4)$$

The value $\delta_{j,k}$ to be compared with the threshold τ is

$$\delta_{j,k} = \max_{i=1}^{p} (d_{i,j,k}). \quad (5)$$

As the values $d_{i,j,k}$ are distributed according to a Student's t distribution with $(n_i + n_j - 2)$ degrees of freedom, the user just sets the level of confidence. The threshold τ itself is calculated from the Student's t distribution. Equations (1) and (2) are tests on whether the null hypothesis that *both samples come from the same distribution* can be rejected at a given confidence level. Hence, this threshold criterion would prevent any mutually optimum pair to merge if there is any PC for which the null hypothesis can be rejected. The reason for such a conservative merging criterion is that while further merging of any two segments is an easy task, splitting wrongly merged segments is much more cumbersome.

B. Discriminant Analysis

After the segmentation process, a spatially unconstrained, nonhierarchical clustering is performed by applying LCDA. LCDA implies the input of user knowledge in the process, a knowledge that includes many sources of information other than the particular scene being processed. This knowledge takes the form of a data matrix of training samples, that is, image elements that have been identified. LCDA processes the data matrix of training samples to find a linear transformation that enhances discrimination of the different categories or groups. Discriminant canonical analysis is also a feature reduction technique because it yields n components at most, with n being

$$n^2 \min (\text{number of variables} - 1, \text{number of individuals})$$

and often the number of significant canonical components is even smaller. The actual value of n is given by the Bartlett significance test [28], [29] consecutively applied with the $n-k$ (k ranging from 0 to $n-1$) eigenvalues of the within-groups covariance matrix that remain after acceptance of the first k. The test is compared against the χ^2 value with $(n-k)$ $(g-k-1)$ degrees of freedom for a chosen significance level, where g is the number of groups. Summaries of canonical discriminant analysis for remote sensing can be found in [30] and more general presentations in [29], [31]. Despite its obvious interest, applications in the remote sensing literature are not numerous [32]–[38].

Here, after IMORM, a multivariate data matrix ($\mathbf{X}$) is built with the segments as elements (rows) and the segment's image statistics as variables (columns). In addition to the segment label, the variables are the per-segment means in the spectral bands (or in transformations of them, like the Tasseled Cap or the PC's) and the per-segment coefficients of variation (CV) in PC-1 and PC-2.

Some segments are selected interactively as training fields and their type of land cover is identified. This is done by overlaying the vectorized segmented image over a color composite of the original and applying knowledge acquired in the field and/or with higher-resolution imagery. The vectors of statistics of these segments are extracted from $\mathbf{X}$ and a training field data matrix ($\mathbf{D}$) is built, adding the land cover label as a new column.

The goal is to find a transformation $\mathbf{U}$ such that the classes are optimally separated. LCDA postulates that such a transformation $\mathbf{U}$ is the one in which the ratio of the variance between classes versus the variance within classes is maximized. Discriminant canonical analysis requires the computation of a within-class covariance matrix ($\mathbf{W}$) and a

TABLE I
RELATIVE INCREMENT OF SEPARABILITIES (%) PRODUCED IN THE AGRICULTURAL SCENE WHEN PER-FIELD STATISTICS ARE USED INSTEAD OF PER-PIXEL STATISTICS. "P.TERRAIN" STANDS FOR PLOUGHED TERRAIN

	Rice	Alfalfa	Sunflw.	Maize	P. terrain	Nat.veg.	Pines	Water
Rice	-	13.800	23.000	28.700	14.000	9.660	11.100	4.840
Alfalfa		-	13.500	27.400	2.980	0.560	0.420	0.000
Sunflw.			-	185.000	195.000	94.200	5.210	0.000
Maize				-	78.100	27.800	6.720	0.000
P. terrain					-	213.000	9.150	0.000
Nat.veg.						-	12.300	0.000
Pines							-	0.000
Water								-

TABLE II
CONFUSION MATRIX USING TEST FIELDS DATA FOR THE CLASSIFICATION OF THE AGRICULTURAL SCENE DONE BY ML AFTER LCDA (SEGMENTS AS ELEMENTS). $\kappa = 0.888(\pm 0.0002)$. ACCURACY (%): 91.03. VALUES IN PIXELS

	Observed categories							
Mapped categories	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	15	0	0	0	0	0	0	0
Class 2	0	142	0	0	75	0	0	0
Class 3	0	0	325	0	22	0	0	0
Class 4	0	0	0	1113	0	95	0	0
Class 5	0	0	27	0	612	0	0	0
Class 6	0	0	0	0	0	335	0	0
Class 7	0	0	0	0	0	90	137	0
Class 8	0	0	0	0	0	0	0	457
Producer's accuracy	1.000	0.654	0.937	0.921	0.958	1.000	0.604	1.000
User's accuracy	1.000	1.000	0.923	1.000	0.863	0.644	1.000	1.000

between-classes covariance matrix ($\mathbf{B}$), both of them from $\mathbf{D}$. $\mathbf{U}$ is the matrix of eigenvectors obtained from the eigenanalysis of $\mathbf{W}^{-1}\mathbf{B}$.

Once $\mathbf{U}$ has been found, it is applied not only to $\mathbf{D}$ but to the entire $\mathbf{X}$ as well.

$$\mathbf{E} = \mathbf{UD}$$
$$\mathbf{Y} = \mathbf{UX}$$
$$\mathbf{Z} = \mathbf{UC}$$

$\mathbf{U}$ is also applied to $\mathbf{C}$, the matrix of centroids, to produce the transformed coordinates. Elements (segments in this case) in $\mathbf{E}$ and $\mathbf{Y}$ are then allocated to their closest class, normally according to a maximum likelihood (ML) rule. Allocation of the elements of the transformed training field data-matrix $\mathbf{E}$ is used to test the adequacy of the training fields and several iterations between the LCDA of $\mathbf{E}$ and field and/or image inspection may be needed until an adequate $\mathbf{D}$ matrix is obtained. A test-field data matrix analogous to $\mathbf{D}$ but not used to find $\mathbf{U}$ is also transformed. Allocation of the elements of this transformed test-field data matrix is used to evaluate the overall accuracy of the process. Finally, allocation of the elements of $\mathbf{Y}$ yields a land cover map by labeling each pixel with the category code to which its segment has been allocated.

III. EXAMPLES

The first example of the above procedure is the processing of a Landsat TM subscene (from path: 199, row: 31, date: 12 July 1993) of an irrigated area in a semi-arid landscape in the Ebro Valley (Spain). The superior results of per-field statistics versus per-pixel statistics, also found by other authors in other agricultural scenes (i.e., [39]–[41]), are shown by the increase of separabilities as measured by the Jeffries–Matusita distance [29], [30] (Table I).

Nevertheless, the extensive use of per-field statistics would imply manual digitizing of the fields or the use of GIS databases that are seldom available and up-to-date. A study was conducted by Lobo *et al.* [37] to explore the use of IMORM to substitute per-field statistics by per-segment statistics.

After segmentation, training and test segments for six categories were defined in the image and LCDA was applied with success (Table II), suggesting that IMORM can significantly increase the accuracy of conventional methods based on per-pixel statistics while avoiding the digitizing needed for per-field statistics. Nevertheless, it must be pointed out that the use of image segmentation requires a modification of the current field surveying, which at the moment produces only per-field descriptions. As a consequence, the study reported in [37] could not formally establish the identity in the ground of classes that had been interpreted from the imagery.

The second example is digitized aerial color infrared photograph from the Jasper Ridge Grassland, California. A diapositive of this photograph was digitized with a color scanner with a final resolution of 13.5 cm on the ground per pixel side. After segmentation, visual and field inspection determined that the landscape could be partitioned into four land-cover

TABLE III
SEPARABILITIES ACCORDING TO THE JEFFRIES–MATUSITA DISTANCE BETWEEN THE GRASSLAND COVER CATEGORIES USING PER-PIXEL STATISTICS. THE POSSIBLE RANGE IS 0–2

	Bunch grasses	Annuals (dense cover)	Annuals (sparse cover)	Bare Soil
Bunch Grasses	0.000	1.448	1.654	1.863
Annuals (dense cover)	1.448	0.000	1.012	1.642
Annuals (sparse cover)	1.654	1.012	0.000	1.266
Bare Soil	1.863	1.642	1.266	0.000

TABLE IV
SEPARABILITIES ACCORDING TO THE JEFFRIES–MATUSITA DISTANCE BETWEEN THE GRASSLAND COVER CATEGORIES USING PER-SEGMENT STATISTICS. THE POSSIBLE RANGE IS 0–2

	Bunch grasses	Annuals (dense cover)	Annuals (sparse cover)	Bare Soil
Bunch Grasses	0.000	1.979	1.994	2.000
Annuals (dense cover)	1.979	0.000	1.862	2.000
Annuals (sparse cover)	1.994	1.862	0.000	1.997
Bare Soil	2.000	2.000	1.997	0.000

categories: bunch grasses, dense cover of annuals, sparse cover of annuals, and bare soil [38]. Training-field data matrices were built both in a per-segment and a per-pixel basis and used to run the LCDA. Low values of separability (Table III) reveal that the above categories cannot be defined from per-pixel statistics, in contrast to results produced by per-segment statistics (Table IV).

Three different classifications were performed:

i) per-pixel allocation using per-pixel statistics;
ii) per-pixel allocation using per-segment statistics. This would be close to common supervised classifications, where per-pixel allocations are performed by maximum likelihood according to training fields interactively drawn from the display; and
iii) per-segment allocation using per-segment statistics, following the method described in Section II.

The Kappa index of agreement (κ) for the test fields was maximum for iii) and minimum for ii) (Table V).

The third example is a subscene of a Landsat TM image (path: 233, row: 70, quadrant 3 from 28 September 1993) over the Chimanes Forest in the Bolivian Amazon (Fig. 1). After georectification by polynomial fitting using field-collected GPS coordinates, the image was also submitted to IMORM. A tentative legend of six land cover classes in the forested area was produced by image inspection. Up to now, analysis of field data allow for a reliable identification of two out of the six classes—"Non-flooded alluvial plains forest" (NFA) and "Low land seasonally flooded forest" (LSF)—and for a reasonable guess on the identity of two more classes: "Palm Forest" (PF) and "Swamp Forest" (SWF) [42]. Again, separabilities from per-pixel statistics are very poor compared to the ones resulting from per-segment statistics (Tables VI and VII).

TABLE V
ACCURACY RESULTS OF THE THREE CLASSIFICATIONS OF THE GRASSLAND IMAGE

Classification method	κ	s.d.	Accuracy
(i)	0.444	0.050	62.0 %.
(ii)	-0.045	0.109	17.5 %
(iii)	0.770	0.037	84.5 %

Accuracy of per-segment classification by maximum likelihood after LCDA was very high (Table VIII), although, in this case, test fields were not visited but selected from the image at the same time as the training fields. Preliminary analysis of our field data suggests that, at least for NFA and LSF, the classification also matches the field description [42].

A study of textural features of four types of tropical forests in the Chimanes scene was conducted by comparing the angular second moment (ASM), contrast (C), and second-order entropy (H') calculated from one GLCM for each cover type as proposed by Haralick *et al.* [13]:

$$\mathrm{ASM} = \sum p(i, j)^2$$

$$C = \sum (i - j)^2 \cdot p(i, j)^2$$

$$H' = -\sum p(i, j) \cdot \log_2 p(i, j)$$

where $p(i, j)$ is the probability of gray levels i and j to co-occur in the 3 × 3 pixel windows and recorded in the GLCM. All three features are thus second-order statistics. GLCM's are computed from 3 × 3 pixel moving windows using the program *i.texture* contributed by McCauley to the GRASS package [43], [44].

The three second-order textural features separate the four forest types in a similar way, with SF and PF being the smoothest and NFA being the coarsest (Fig. 2). This ordination agrees both with visual inspection and with what could be expected from the structural characteristics of these forests, although further analysis of the field data [42] are needed. A very important point is that the same pattern is found regardless of the particular image transformation used: the two first Tasseled Cap components [45] and the two first PC's. This coincidence reinforces the hypothesis that second-order textural features are related to structural characteristics of the landscape elements rather than to their spectral characteristics. Fig. 3 presents values of two textural features computed from first-order statistics, the coefficient of variation (CV) and the first-order entropy (H'). These values do not show a consistent relationship with visual perception and widely vary with the particular transformation used. These results suggest that the GLCM features should be included as variables for the LCDA of segments, although the small size of some segments could represent a problem.

IV. DISCUSSION

The well known procedure of supervised classification is one of the most attractive for users of remotely sensed imagery in the natural sciences because it combines user's knowledge (introduced by locating training fields) with machine performance and statistical rigor (which essentially extends user's

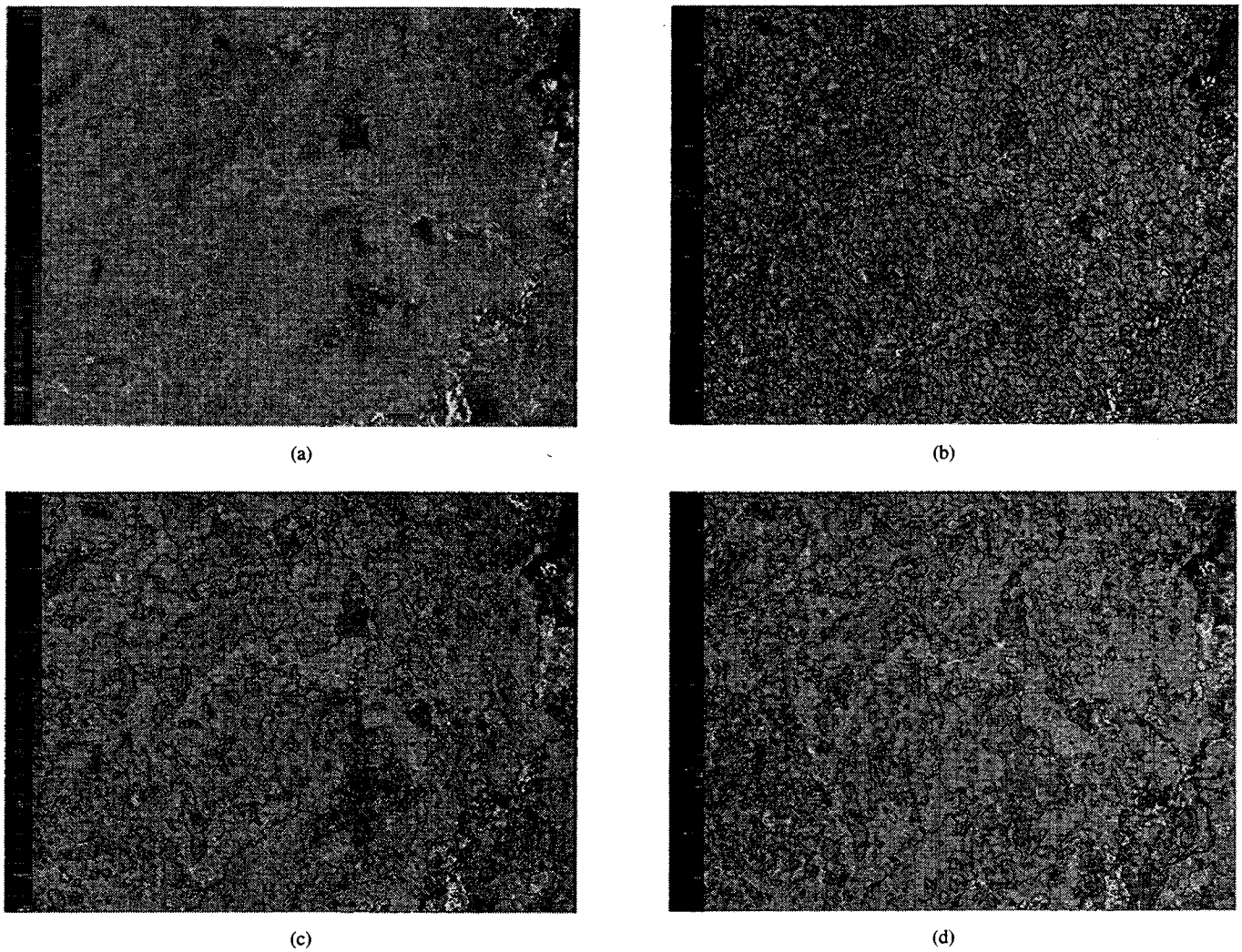


Fig. 1. Segmentation and classification by LCDA of the georectified Landsat TM scene of the tropical forests of Chimanes, Bolivia. The area covered is 16.380 × 11.940 km^2 with a spatial resolution of 30 × 30 m^2. (a) The Tasseled Cap color composite of the area. (b) Contours of the IMORM segmentation overlaid on (a). IMORM was run from an EPS of PC-1. (c) Contours of the classification of segments (b) by LCDA. An error in the PF class is indicated. Symbols: Swamp Forest (SF); Palm Forest (PF); Non-flooded Alluvial-plains Forest (NFA); and Low Seasonally Flooded Forest (LSF). (d) Analogous to (c) but IMORM was run from a combination of the EPS of PC-1 and edges detected on PC-2.

TABLE VI

SEPARABILITIES ACCORDING TO THE JEFFRIES–MATUSITA DISTANCE BETWEEN THE TROPICAL LAND COVER CATEGORIES USING PER-PIXEL STATISTICS. THE POSSIBLE RANGE IS 0 TO 2. SWF: "SWAMP FOREST"; PF: "PALM FOREST"; NFA: "NON-FLOODED ALLUVIAL PLAINS FOREST"; LSF: "LOW LAND SEASONALLY FLOODED FOREST"; U1 AND LSF2 ARE STILL UNKNOWN LAND COVER CATEGORIES

	SWF	PF	NFA	U1	LSF	LSF2
SWF	0.00	2.00	2.00	1.99	1.90	1.27
PF	2.00	0.00	1.17	1.98	1.62	1.95
NFA	2.00	1.17	0.00	1.74	1.18	1.89
U1	1.99	1.98	1.74	0.00	1.77	1.80
LSF	1.90	1.62	1.18	1.77	0.00	1.10
LSF2	1.27	1.95	1.89	1.80	1.10	0.00

knowledge to the entire image under well known hypothesis on the underlying distributions). Supervised classification discloses a major contradiction: while training fields are always outlined by the user as solid areas, allocations are done on a per-pixel basis instead of on a per-area basis. As a consequence, supervised classification, as normally practiced, seldom produces satisfactory results and probably the most common classification procedure relies on an initial unsupervised clustering that is subsequently reduced by regrouping the spectral classes into land-cover categories according to the user's criteria. As we mentioned in Section I, this procedure results in low accuracy unless the final categories are very general.

Two important reasons for the relative failure of supervised classification techniques as normally practiced are that the original multispectral space is seldom transformed and the reliance on per-pixel statistics. Use of LCDA and related techniques solve the first problem and classification of segments the second.

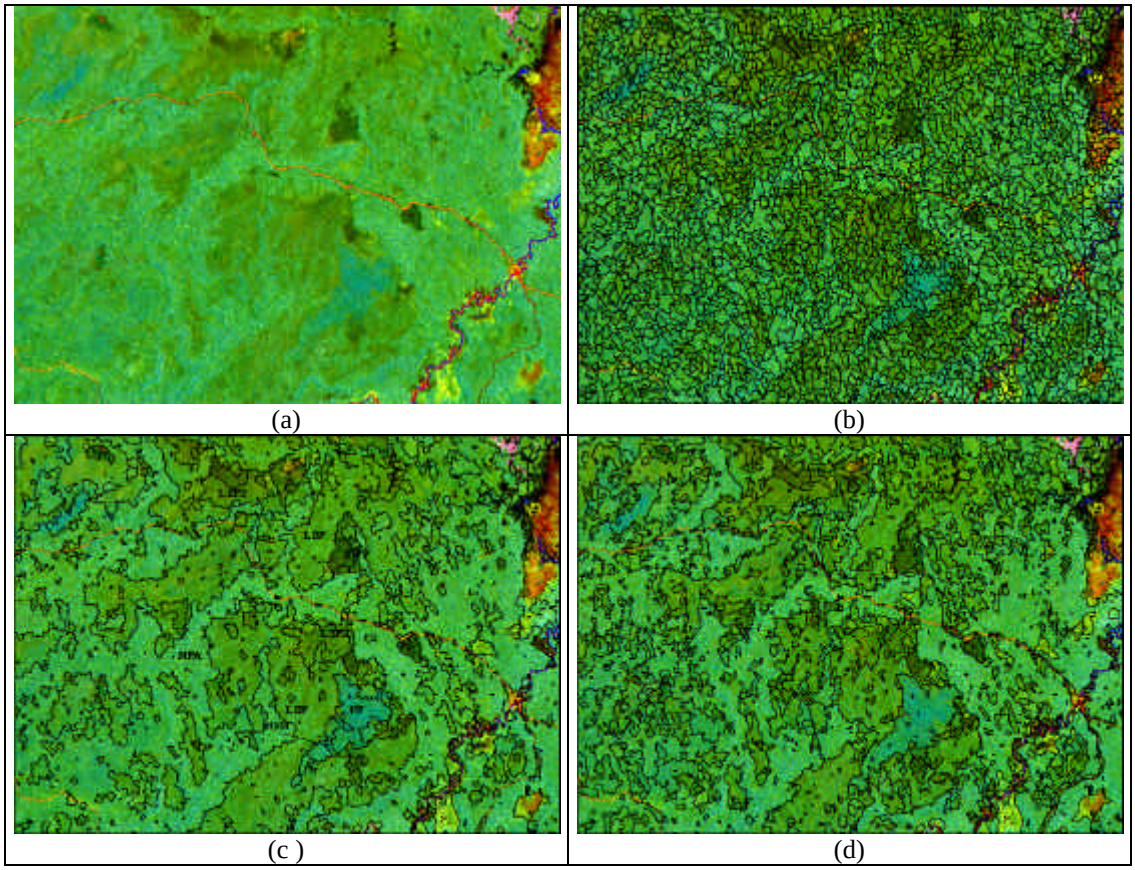
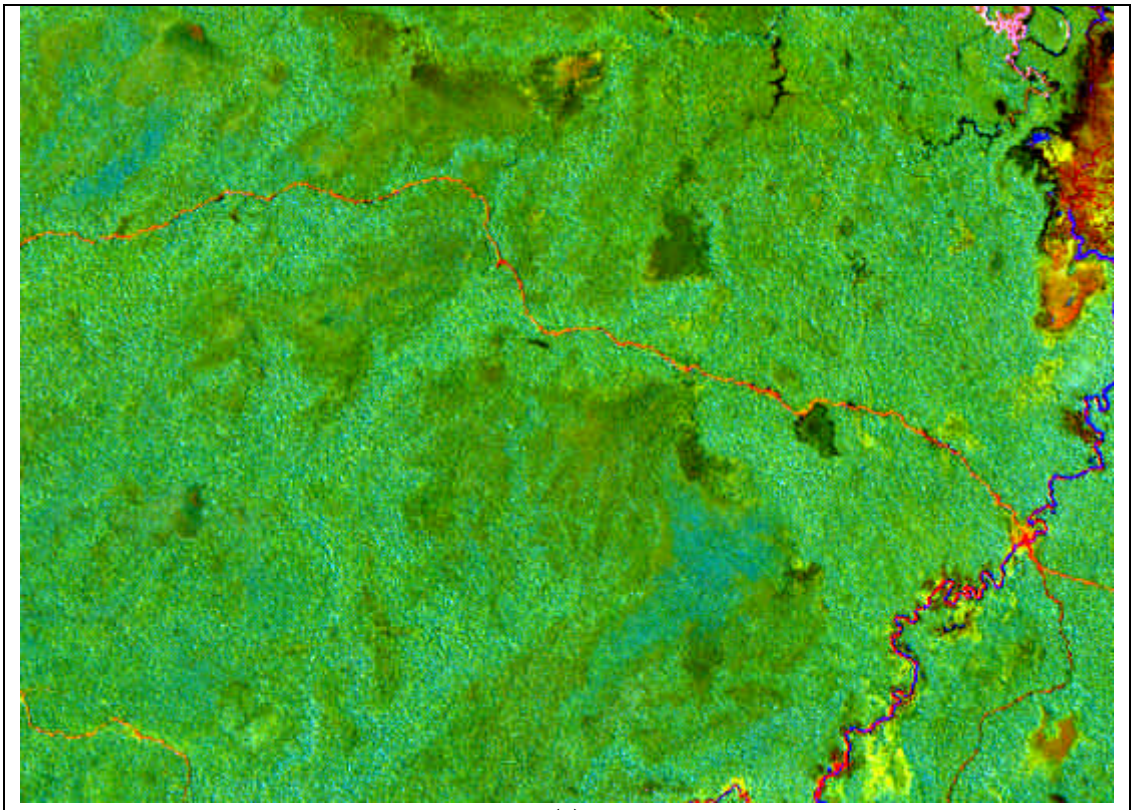
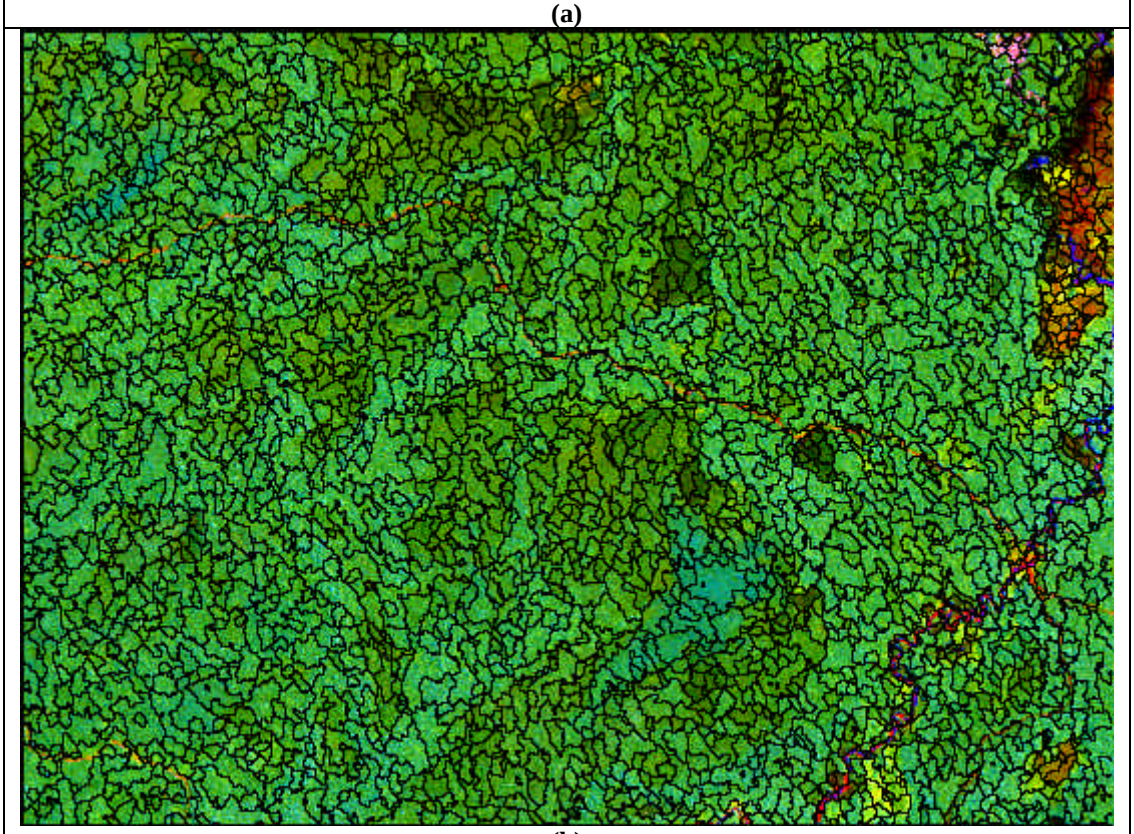


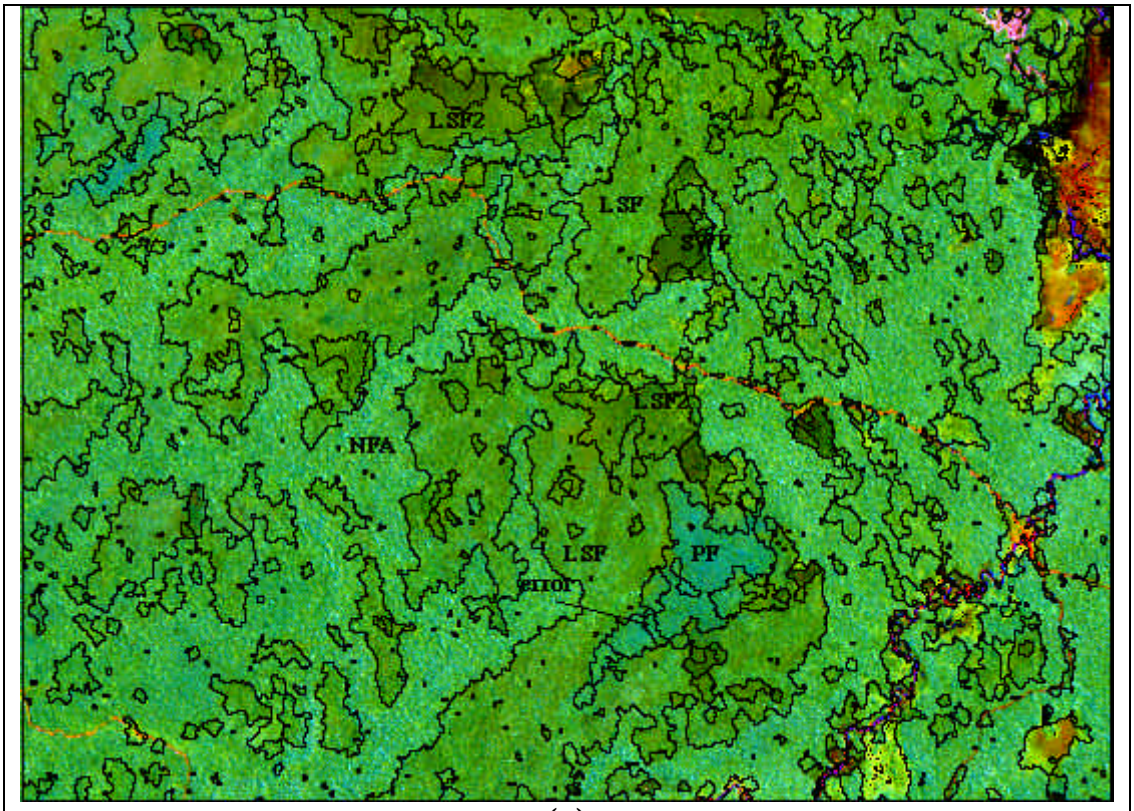
Figure 1. (A. Lobo, 1997. IEE TGARS 35 (5): 1141)



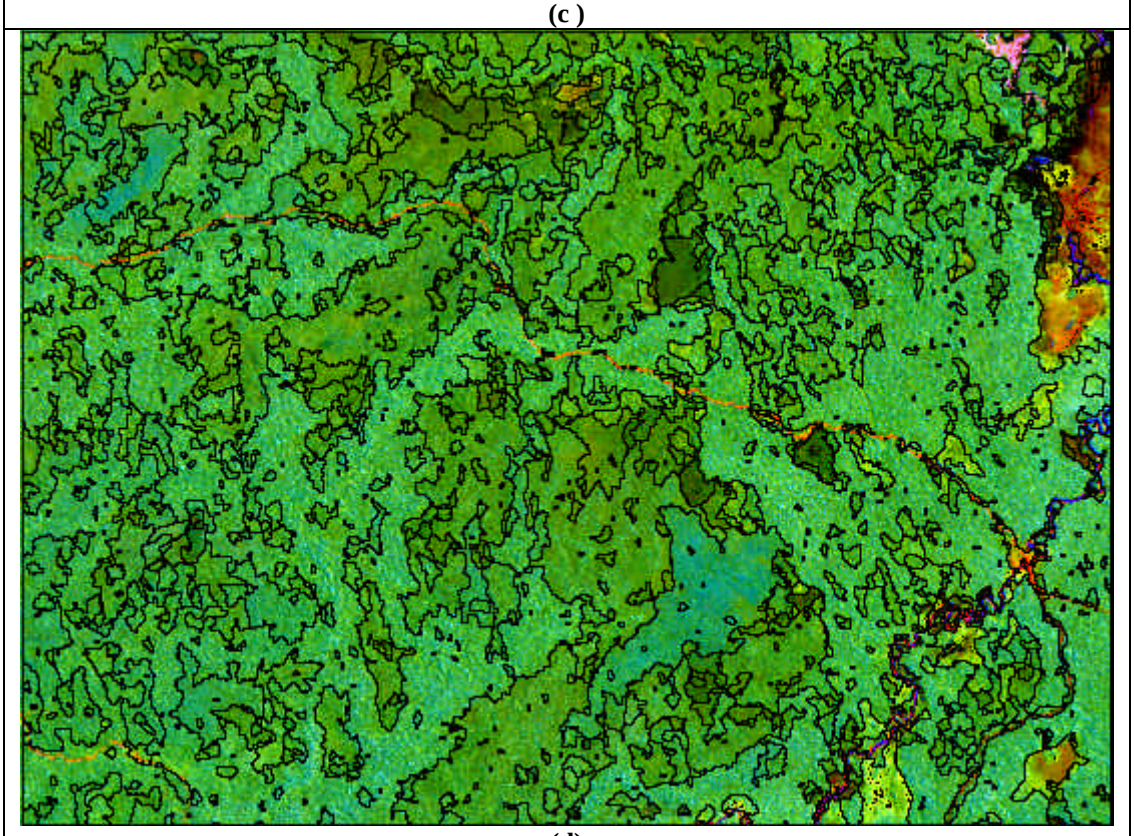
(a)



(b)



(c)



(d)

Figure 1. (A. Lobo, 1997. IEE TGARS 35 (5): 1141)

TABLE VII
SEPARABILITIES ACCORDING TO THE JEFFRIES–MATUSITA DISTANCE BETWEEN THE TROPICAL LAND COVER CATEGORIES USING PER-SEGMENT STATISTICS. THE POSSIBLE RANGE IS 0 TO 2

	SWF	PF	NFA	U1	LSF	LSF2
SWF	0.00	2.00	2.00	2.00	2.00	2.00
PF	2.00	0.00	1.99	2.00	1.99	2.00
NFA	2.00	1.99	0.00	1.99	2.00	2.00
U1	2.00	2.00	1.99	0.00	1.96	1.98
LSF	2.00	1.99	2.00	1.96	0.00	1.90
LSF2	2.00	2.00	2.00	1.98	1.90	0.00

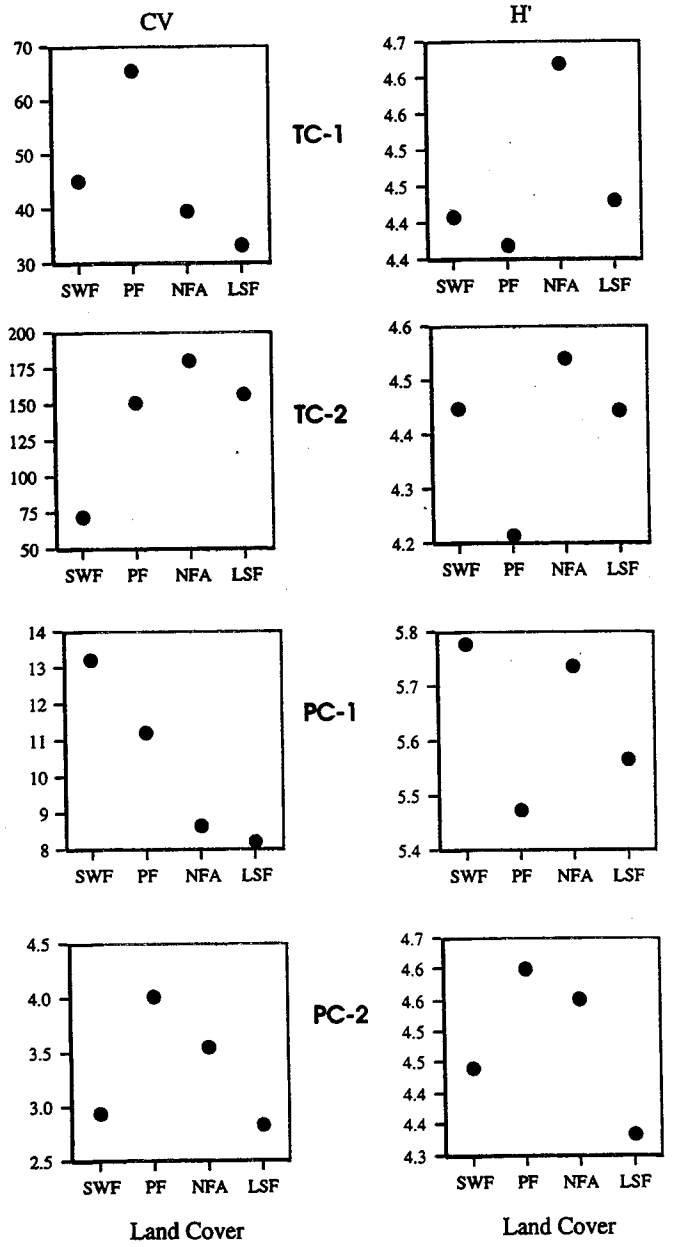


Fig. 2. Textural features computed from the grey-level co-occurrence matrices in four tropical forest types: Swamp forest (SF); Palm forest (PF); Non-flooded Alluvial-plains forest (NFA); and Low Seasonally Flooded Forest (LSF).

Unlike individual pixel classification, errors occur as patches and not as scattered individual pixels in segment classification. This makes errors visually more obvious and easier to identify and study. Most errors detected in our work to date are due to undersegmentation, and of those, the undersegmentation comes from the EPS processing. Undersegmented facets cannot be corrected by IMORM or the LCDA. Even more, undersegmented facets tend to have a high variance and tend to attract correct neighboring facets in IMORM, which yields larger erroneous segments. Afterwards, those wrong segments concentrate most of the wrong identifications by LCDA. One example of these errors is shown in Fig. 1(c). An undersegmented facet produced by EPS at the transition between LSF and PF gets even larger after IMORM and is finally identified as LSF by the LCDA. As a result, about one third of the segment area is incorrectly identified and, even worse, the transition edge between both types of forest is also incorrect.

The reason for the undersegmentation in the EPS results is that, unlike IMORM, EPS is not using all the multispectral information but only PC-1. Although PC-1 contains an important part of the total variance (72.42%, 74.86%, and 72.73%, respectively, in the three examples presented in the previous section), it does not contain all. In the example of Fig. 1, the transition between both types of forest was much more obvious in PC-2 than in PC-1. An EPS conducted on PC-2 for the undersegmented facet further partitions it, allowing IMORM to correctly integrate and LCDA to correctly identify the new segments. Such a finer EPS partition of the formerly incorrect facet also induced a better segmentation in the neighboring area.

Nevertheless, EPS cannot be conducted on PC-2 and then be combined with the result from PC-1 into one single LFI because this produces a large oversegmentation in most of the image. It has to be borne in mind that EPS applied to PC-1 produces a LFI that allows IMORM and LCDA to perform very well in most of the image, as seen in the previous section. A way to introduce in the process those major discontinuities that can be better detected in PC-2 is to combine the LFI produced by the EPS of PC-1 with edges obtained by the zero-crossings of the Laplacian transformation of the Gaussian-smoothed PC-2. Such a combination produces a finer LFI for IMORM. Fig. 1(d) shows the result of IMORM and LCDA run with such a finer LFI. Note that the erroneous partition of the PF patch indicated in Fig. 1(c) is no longer present in Fig. 1(d).

One problem with the use of segment classification versus individual pixel classification is that poor estimation of covariance matrices can occur for land-cover categories with a low relative cover, because this low presence results in a low number of training fields. This poor estimation of the covariance matrix may result, for those categories, on better

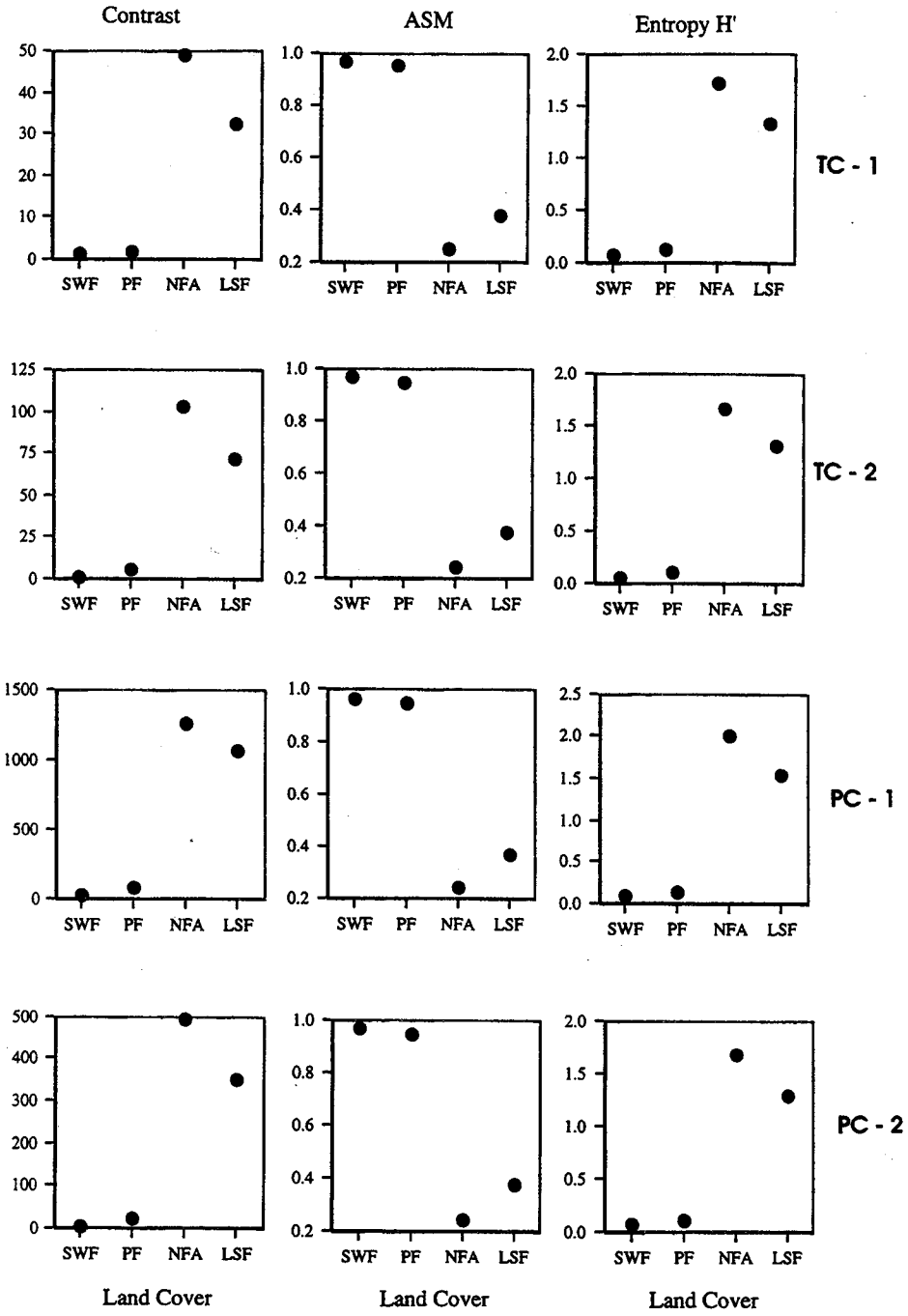


Fig. 3. Textural features (CV: coefficient of variation; H': entropy) computed from the histogram of grey-levels in four tropical forest types: SwampForest (SF); Palm Forest (PF); Non-flooded Alluvial-plains Forest (NFA); and Low Seasonally Flooded Forest (LSF).

accuracies by allocating by minimum Euclidean distance than by maximum likelihood. Increasing the number of training fields is not as easy with segments as it is with individual pixels and, often, it would require either using up virtually all the category's cover for the training fields (undesirable for accuracy testing) or increasing the area of study. The use of nonparametric classifiers and decision-boundary feature extraction (DBF, [46]) instead of LCDA could be particularly indicated for classifying segments, avoiding computation of covariance matrices.

The use of image segmentation as a previous step to classification by LCDA implies the need for a high spatial resolution imagery relative to the size of the land cover units of interest, that is, an H-resolution scene model in the terminology of Strahler *et al.* [47]. In H-resolution models the elements of the scene are larger than the resolution cells (pixels). The use of EPS, which requires a radius to be set by the user, further restricts the minimum size of features that will be discriminated. This is rarely a problem with natural vegetation, but it must be considered if small or thin features (i.e., roads or streams) are considered to be important in the final result. Nevertheless, other methods exist to enhance, detect, and extract these features, which can be added to the result from IMORM.

TABLE VIII
CONFUSION MATRICES (IN PIXELS) FOR TRAINING AND TEST FIELDS IN THE CHIMANES SCENE FOR A PER-SEGMENT CLASSIFICATION BY LCDA. $\kappa = 0.920$. ACCURACY (%) 93.91

Training fields							Test fields						
	Observed categories							Observed categories					
Classified categories	SWF	PF	NFA	U1	LSF	LSF2	Classified categories	SWF	PF	NFA	U1	LSF	LSF2
SWF	219	0	0	0	0	0	SWF	140	0	0	0	0	0
PF	0	347	0	0	0	0	PF	0	407	0	0	0	0
NFA	0	0	914	0	0	0	NFA	0	0	364	0	0	0
U1	0	0	0	171	0	0	U1	0	0	0	31	0	0
LSF	0	0	0	0	1032	0	LSF	0	0	0	0	144	0
LSF2	0	0	0	0	0	483	LSF2	74	0	0	0	0	55

Image segmentation is inadequate for L-resolution scene models, in which the elements of the scene are smaller than pixels and for which techniques such as spectral mixture analysis are appropriate. Nevertheless, both approaches can be undertaken in parallel to finally combine their results to produce a more complete analysis of the scene.

A set of nested hierarchic segmentations, as proposed by Woodcock and Hayward [23], is possible with IMORM by running the program with different values of τ. Nevertheless, in the protocol described here the hierarchical levels of segmentation should correspond to hierarchical levels in the legend and imply different methods and information. Here, IMORM is used for the fine segmentation. Afterwards, segments merge according to the results of LCDA, which implies a different approach and the introduction of user's expertise. The LCDA level produces a land cover legend that is as fine as the spectral and textural characteristics of the segments in one hand, and the user's information on the other, permit. The next step in the hierarchy would imply grouping LCDA categories into broader ones (i.e., NFA and LSF into Alluvial Plains Tropical Forest) according to logical rules. The information stressed at each level is different and the weight of the external (from an image point of view) information is increasingly higher. At the IMORM level the information used is spatial and spectral; at the LCDA level the spectral-textural information is used along with specific models introduced by the user; at the grouping level the identity of the LCDA categories, sometimes complemented with some spatial information (i.e., to include city parks in a urban category) are the input for the logical rules set by the user. This approach, that uses both image and nonimage information, is more adequate to create a hierarchical legend in accordance with the concept of hierarchical organization of landscapes [48]–[50]. Theoretical and practical ecological problems related to changes of scale [51]–[56] will greatly benefit of a correct hierarchical legend.

Finally, likelihood values resulting from maximum likelihood allocation in the canonical space are useful and suggest an ecological reflection. Display of maximum likelihood values for each segment helps to detect areas that are not well represented by the current training-fields set. Also, attributing a vector of likelihood values to each segment agrees with a view of landscape types as a continuum, in which several modes are distinguishable. This approach is analogous to fuzzy classification [57]–[59]. Such a flexible model needs the development of newer accuracy measurements as well as newer techniques for the analysis of land cover change. Image segmentation and LCDA are tools that qualitatively improve land cover maps by enhancing accuracy and enabling finer and more flexible legends.

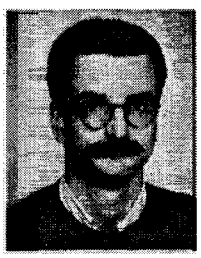
ACKNOWLEDGMENT

The author would like to acknowledge comments from B. Philpott, Cornell University, and two anonymous reviewers, which significantly improved the original manuscript.

REFERENCES

[1] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 420–433, 1976.
[2] F. Gonzalez-Alonso and S. Lopez-Soria, "Using contextual information to improve land use classification of satellite images in central Spain," *Int. J. Remote Sensing*, vol. 12, no. 11, pp. 2227–2235, 1991.
[3] P. H. Swain, S. B. Vardeman, and J. C. Tilton, "Contextual classification of multispectral image data," *Pattern Recognit.*, vol. 13, pp. 429–441, 1981.
[4] J.-H. Lee and W. D. Philpot, "Spectral texture pattern matching: A classifier for digital imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, pp. 545–554, July 1991.
[5] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. GE-14, no. 1, pp. 19–26, 1976.
[6] D. A. Landgrebe, "The development of a spectral classifier for Earth observational data," *Pattern Recognit.*, vol. 12, pp. 165–175, 1980.
[7] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer, "A land use and land cover classification system for use with remote sensor data," *Geological Survey Professional Paper 964*, Geological Survey, 1976.
[8] S. A. Sader, "Spatial characteristics of forest clearing and vegetation regrowth as detected by Landsat Thematic Mapper imagery," *Photogramm. Eng. Remote Sens.*, vol. 61, no. 9, pp. 1145–1151, 1995.
[9] D. Marceau, "A review of image classification procedures with special emphasis on the grey-level co-occurrence matrix method for texture analysis," ISTS-EOL-TR89-007, Earth-Observations Lab., Inst. Space Terrestrial Sci., Dept. Geography, Univ. Waterloo, Waterloo, Ont., Canada, 1989.
[10] D. J. Marceau, P. J. Howarth, J.-M. M. Dubois, and D. J. Gratton, "Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 513–519, July 1990.
[11] G. Ramstein and M. Raffy, "Analysis of the structure of radiometric remotely sensed images," *Int. J. Remote Sensing*, vol. 10, no. 6, pp. 1049–1073, 1989.
[12] A. Rosenfeld and E. Troy, "Visual texture analysis," in *IEEE Symp. Feature Extraction Selection Pattern Recognition*, Argonne, IL, 1970.
[13] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 610–621, 1973.

[14] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures of terrain classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-4, pp. 269–285, 1976.
[15] L. S. Davis, S. A. Johns, and J. K. Aggarwal, "Texture analysis using generalized co-occurrence matrices," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, pp. 251–259, 1979.
[16] L. S. Davis, M. Clearman, and J. K. Aggarwal, "An empirical evaluation of generalized co-occurrence matrices," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 214–221, 1981.
[17] M. D. Levine, *Vision in Man and Machine.* New York: McGraw-Hill, 1985.
[18] E. Sali and H. Wolfson, "Texture classification in aerial photographs and satellite data," *Int. J. Remote Sensing,* vol. 13, no. 18, pp. 3395–3408, 1992.
[19] S. Hsu, "Texture-tone analysis for automated landuse mapping," *Photogramm. Eng. Remote Sens.*, vol. 44, pp. 1393–1404, 1978.
[20] J. R. Irons and G. W. Petersen, "Texture transforms of remote sensing data," *Remote Sens. Environ.*, vol. 11, pp. 359–370, 1981.
[21] A. M. Nazif and M. D. Levine, "Low level image segmentation: An expert system," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 5, pp. 555–577, 1984.
[22] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comp. Vis., Graph. Image Proc.*, vol. 12, pp. 100–132, 1985.
[23] C. Woodcock and V. J. Harvard, "Nested-hierarchical scene models and image segmentation," *Int. J. Remote Sensing,* vol. 13, no. 16, pp. 3167–3187, 1992.
[24] P. Legendre, "Constrained clustering," in *Developments in Numerical Ecology,* P. Legendre and L. Legendre, Eds. Berlin, Germany: Springer-Verlag, 1987, pp. 289–307.
[25] P. Legendre and M.-J. Fortin, "Spatial pattern and ecological analysis," *Vegetatio,* vol. 80, pp. 107–138, 1989.
[26] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs,* M. Nadler, Ed. New York: Plenum, 1980.
[27] R. R. Sokal and F. J. Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research.* New York: Freeman, 1995.
[28] L. Legendre and P. Legendre, *Numerical Ecology.* Amsterdam, The Netherlands: Elsevier, 1983.
[29] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition.* New York: Wiley, 1992.
[30] J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction,* 2nd ed. Berlin, Germany: Springer-Verlag, 1993.
[31] W. R. Dillon and M. Goldstein, *Multivariate Analysis. Methods and Applications,* W. A. Shewhart and S. S. Wilks, Eds. New York: Wiley, 1984.
[32] B. F. Merembeck and B. J. Turner, "Directed canonical analysis and the performance of classifiers under its associated linear transformation," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-18, pp. 190–196, 1980.
[33] R. J. Hobbs, J. F. Wallace, and N. A. Campbell, "Classification of vegetation in the Western Australian wheatbelt using landsat mss data," *Vegetatio,* vol. 80, no. 2, pp. 91–106, 1989.
[34] C.-C. T. Chen and D. A. Landgrebe, "A spectral feature design system for the *hiris/modis* era," *IEEE Trans. Geosci. Remote Sensing,* vol. 27, pp. 681–686, Nov. 1989.
[35] T. L. Henderson, A. Szilagyi, M. F. Baumgarder, C. C. T. Chen, and D. A. Landgrebe, "Spectral band selection for classification of soil organic matter content," *Soil Sci. Soc. Amer. J.,* vol. 53, no. 6, pp. 1778–1784, 1989.
[36] J. F. Wallace, N. A. Campbell, G. A. Wheaton, and D. J. McFarlane, "Spectral discrimination and mapping of waterlogged cereal crops in Western Australia," *Int. J. Remote Sensing,* vol. 14, no. 14, pp. 2731–2743, 1993.
[37] A. Lobo, O. Chic, and A. Casterad, "Multisensorial classification of Mediterranean crops: Per-pixel versus per-patch statistics and image segmentation," *Int. J. Remote Sensing,* vol. 17, no. 12, p. 2385, 1996.
[38] A. Lobo, K. Moloney, and N. Chiariello, "Fine-scale mapping of a grassland from digitized aerial photography: An approach using image segmentation and discriminant analysis," *Int. J. Remote Sensing*, to be published.
[39] D. A. Landgrebe, "Machine processing of remotely acquired data," in *Remote Sensing of Environment,* L. A. D. S. Simonett, Ed. Reading, MA: Addison-Wesley, 1976, pp. 349–373.
[40] J. Megier, W. Mehl, and R. Ruppelt, "Per-field classification and application to SPOT, simulated SAR and combined SAR-MSS data," in *18th Int. Symp. Rem. Sens. Environ.*, Univ. Michigan, Ann Arbor, 1984, pp. 1011–1018.
[41] M. I. Pedley and P. J. Curran, "Per-field classification: An example using SPOT HRV imagery," *Int. J. Remote Sensing,* vol. 12, pp. 2181–2192, 1991.
[42] A. Lobo and R. E. Gullison, "Mapping the tropical landscapes of Beni (Bolivia) from landsat-TM imagery: Beyond the forest–non forest legend," in *Forest Biodiversity in Europe, Australia and Africa: Research, Monitoring and Modeling,* F. Dallmeier and J. Comiskey, Eds. New York: Parthenon, to be published.
[43] M. Shapiro, J. Westervelt, D. Gerdes, M. Larson, and R. Brownfield, *GRASS 4.1 Users Manual,* U.S. Army Corps Eng., Construct. Eng. Res. Lab., 1993.
[44] J. D. McCauley and B. E. Engel, "Comparison of scene segmentations: SAMP, ECHO, and maximum likelihood," *IEEE Trans. Image Processing,* submitted for publication.
[45] E. P. Crist and R. C. Cicone, "A physically based transformation of Thematic Mapper data: The TM tasseled cap," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-22, pp. 256–263, 1984.
[46] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 4, pp. 388–400, 1993.
[47] A. H. Strahler, C. E. Woodcock, and J. A. Smith, "On the nature of models in remote sensing," *Remote Sens. Environ.*, vol. 20, pp. 121–139, 1986.
[48] D. L. Urban, R. V. O'Neill, and H. H. Shugart, "Landscape ecology," *BioSci.*, vol. 37, no. 2, pp. 119–127, 1987.
[49] R. T. T. Forman and M. Godron, *Landscape Ecology.* New York: Wiley, 1986.
[50] R. V. O'Niell, D. L. DeAngelis, J. B. Waide, and T. F. H. Allen, *A Hierarchical Concept of Ecosystem.* Princeton, NJ: Princeton Univ. Press, 1986.
[51] A. Moody and C. E. Woodcock, "Scale-dependent errors in the estimation of land-cover proportions—Implications for land-cover datasets," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 5, pp. 585–594, 1994.
[52] ——, "The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing," *Landscape Ecology,* vol. 10, no. 6, pp. 363–379, 1995.
[53] M. Turner, V. Dale, and R. Gardner, "Predicting across scales: Theory development and testing," *Landscape Ecology,* vol. 3, no. 3/4, pp. 245–252, 1989.
[54] W. L. Baker, "Spatially heterogeneous multi-scale response of landscapes to fire suppression," *Oikos,* vol. 66, pp. 66–71, 1993.
[55] L. L. Pierce and S. W. Running, "The effects of aggregating subgrid land surface variation on large-scale estimates of net primary production," *Landscape Ecology,* vol. 10, no. 4, pp. 239–253, 1995.
[56] J. Townshend and C. Justice, "Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations," *Int. J. Remote Sensing,* vol. 9, no. 2, pp. 187–236, 1990.
[57] G. M. Foody, "A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: An example from lowland heath," *Photogramm. Eng. Remote Sens.*, vol. 58, pp. 221–225, 1992.
[58] G. M. Foody and D. P. Cox, "Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions," *Int. J. Remote Sensing,* vol. 15, no. 3, pp. 619–631, 1994.
[59] G. M. Foody, "Fully fuzzy supervised image classification," in *Remote Sensing Action, Proc. 21st Annu. Conf. Remote Sensing Soc.,* Southampton, U.K., 1995, Remote Sensing Soc., pp. 1187–1194.



Agustín Lobo received the B.Sc. degree in biology from the Universidad Complutense de Madrid, Spain, the M.Sc. degree from the Universitée d'Aix-Marseille II, France, and the Ph.D. degree from the Universitat de Barcelona, Spain, specializing in the analysis of digital imagery for environmental applications.

From 1991 to 1994, he was with Cornell University, Ithaca, NY, and Princeton University, Princeton, NJ, where he worked on the modeling and analysis of spatial pattern from high resolution digital imagery. He is an ecologist who uses digital imagery as his primary, but not only, source of information. He is currently a Research Associate at the Institut de Ciències de la Terra (CSIC), Barcelona, Spain, where he works on environmental applications of satellite imagery and GIS with emphasis on the dynamics of vegetation cover and its relationship to climate.

Dr. Lobo is a member of the International Users Committee of the VEGETATION instrument. He was awarded a Postdoctoral Fulbright Fellowship from Cornell University and Princeton University.