INT. J. REMOTE SENSING, 1996, VOL. 17, NO. 16, 3093–3106

The integration of spatial context information in an experimental knowledge-based system and the supervised relaxation algorithm—two successful approaches to improving SPOT-XS classification

C. C. KONTOES

41 Katsimpiri Str, GR-155 61, Holargos, Athens, Greece

and D. ROKOS

National Technical University of Athens, Laboratory for Remote Sensing, Heroon Polytechniou 9, 157 80, Zografos, Athens, Greece

(*Received 28 October 1993; in final form 3 January 1996*)

Abstract. This paper describes two different methods which integrate contextual information in a classification process. This process aims to refine the map products given by the application of a common parametric classification algorithm. The first method is the well known Supervised Relaxation Algorithm, and makes use of the first classification, with additional contextual information. The contextual information is derived either from texture features or from other map products introducing additional information on the existing land use classes. The second method is a knowledge-based system, which makes use of image and geographical context rules. The probability figures, derived from the image classifier and the rule base are combined by the use of the Dempster–Shafer reasoning scheme. Experiments using satellite data from the Loir et Cher region (Central France), together with the appropriate ground truth data, have shown that both methods return improved classification products in terms of thematic and statistical accuracy, compared to using a parametric image classifier alone.

1. Introduction

The requirements, placed in the frame of various operational projects at the European Union level, such as Monitoring Agriculture by Remote Sensing, CORINE, Greek Plan, etc, to produce reliable agricultural statistics and monitor land cover, make the use of high resolution satellite data indispensable. This is the main data source for repeated assessment of quantitative and qualitative information regarding land use on a large scale. The application of low level image analysis techniques (Argialas and Harlow 1990), very often return noisy and unreliable map products which are not of operational value. There are many difficulties in producing acceptable map products, the foremost being, the adequacy of the training sample and the performance of the clustering algorithm. The classification algorithms, either make ideal approximations of the reality (e.g., Gaussian density functions—Morrison 1976, Duda and Hart 1973, and equal or near equal variance covariance matrices, Fisher's linear model application, Marks and Dunn 1974, Gilbert 1969), or require intense efforts in ground truth works (non-parametric algorithms Skidmore and Turner 1988, Shettigara 1991). Thus, the application of such algorithms, may not always be realized, because of time and cost restrictions. Another major reason why inaccuracies occur is that the classification algorithm forces the matching of spectral classes to natural and cultivated crop classes. However, the classification in natural classes, when applied by the photo-interpreter, relies not only on radiometric properties but also on the shape and spatial relations of neighbouring classified pixels/segments and the geographical and biochemical parameters of the study area. There has been much discussion regarding the so-called 'machine assisted photo-interpretation system', which applies logical reasoning, takes account of contextual relations, assesses similarities and disparities among the various objects and occasionally accounts for bio-geophysical parameters using either absolute or subjective measurements (Merchant 1984, Rokos 1988).

It is generally believed, that to achieve the best performance of such a system one should:

(*a*) Exploit the satellite data together with ancillary qualitative and quantitative information.
(*b*) Decide about the most appropriate method to analyse and exploit multi-source information.
(*c*) Explore the relevant human scientific, heuristic and common sense knowledge and
(*d*) Introduce adequate reasoning models into the decision making process (Rokos 1988).

Moreover, the integrated ancillary information may relate to:

(1) *Geographical context*. In order to assist the refinement of pixel/segment classifications on the basis of background geographical parameters (e.g. soil type and texture, subsurface drainage, land suitability, surface slope and elevation, temperature measurements, precipitation level, proximity to road and facilities networks, other environmental parameters).
(2) *Image context*. Aiming to refine pixel/segment classifications on the basis of the class labels assigned to neighbouring pixels/segments on the map layer.

A number of different approaches to the analysis of multiple source data may be reported. Particularly the works of Cibula and Nyquist 1987, Strahler *et al.* 1978, de Jong and Riezebos 1991, Jones *et al.* 1988, Civco 1989, Skidmore 1989 and Tong *et al.* 1987, where the use of satellite data in conjunction with environmental and geographical data return accurate map products of user specified or project depended classes.

Experiments have shown, that using multiple source data, together with satellite data, results in significant improvements in the accuracy of the classification. However, the fact that each pixel is examined independently of the surrounding classification labels, and other regional characteristics results in label inconsistencies within the pixel neighbourhood. In the present study the single date SPOT-XS data are classified by two different classification techniques. Both techniques make use of information derived from the image and geographical context of the pixel. The two techniques are the Supervised Relaxation Algorithm (Richards *et al.* 1981) and a knowledge-based system, which has been developed especially for the needs of the study. A detailed description of these two approaches follows in this paper.

2. Study area. Image and geographical context information

The study area, was decided to be one of the regions of interest in the MARS project of the European Union, because satellite data and ground observation reports would be easily accessed. The study area is the Loir et Cher Department, which is one of the major agricultural areas in central France. It represents a big variety in crop types, crop production, field size, field orientation, agricultural systems and background soil types. A SPOT-XS satellite scene captured on 1 May 1986 was provided for this area. Ground truth reports were also available, based on 65 sampled segments which were evenly distributed in the study area, representing 0·5 per cent of the total surface of the region.

The first classification map was obtained following the procedure illustrated in figure 1. Special attention has been given to the reduction of the variance of the sampling data, by minimizing the human type errors in crop reporting with the application of the appropriate clustering to the training data. The accuracy of the classification has been measured in two areas which differ significantly, regarding the land cover classes. The first is the Cour Cheverny area, which is extended NE of the city of Blois. This area is covered mainly by natural vegetation and water surfaces. The second area, is the Herbault region which is one of the main agricultural areas of the Loir et Cher Department. The classification accuracy achieved for the Cour Cheverny and Herbault areas is 71·3 and 64·5 per cent respectively. Figure 2 illustrates the classification scheme which has been used for the Herbault region.

A slightly different classification scheme, has been used for the Cour-Cheverny region, since the land use occupation comprises mainly of forested and natural vegetation classes rather than agricultural ones. The spectral classes which appear at the lower level of the classification tree introduce the classes of the detailed classification map. The classes of the upper layer, that is the super-classes, introduce a rough segmentation of the satellite image, in generalized land use classes. The identification and creation of the super-classes, may be based on either texture (Haralick *et al.* 1973) or spectral properties of the crop types. Texture measurements introduce groups of crops, that is super-classes, with similar texture characteristics (e.g., row crops, homogeneous fully covered canopies, speckled forested areas, etc.). Apart from the texture properties, the statistical merging of classes with close spectral characteristics also defines groups of crops (super-classes) with similar cultivating practices and crop calendars. Experiments have shown, that both texture and spectral properties, can be used for image classification (Weszka *et al.* 1976, Peddle and Franklin 1991). For the purposes of the present study, they have both been used, for the segmentation of the SPOT-XS image in land use classes and super-classes.

The super-class layer introduces additional information regarding the class assignment to the pixel itself, as well as its context. The super-class assignment to the pixel and its context, results in the reduction of the hypothesis space, considered by the algorithm in the classification stage. Its size changes from the total number of crop classes to the size of a sub-group. This reduced size is defined by the number of the corresponding singleton classes, which make up the super-class specified in the classification scheme of figure 2.

Geographical information of the pixel context is another type of additional input, which has been used for this study. This information originated from maps which have been digitized, pre-processed and converted into co-registered raster form, in order to be integrated into the classification methodology. To test the importance of the geographical ancillary information, two GIS data sets were prepared. A soil type map, since soils have a strong influence on land use suitability and crop cultivation. The soil type data were derived from the *Esquisse des Caracteres Hydriques des Sols de la Region Centre from Chambre Regional d' Agriculture du*

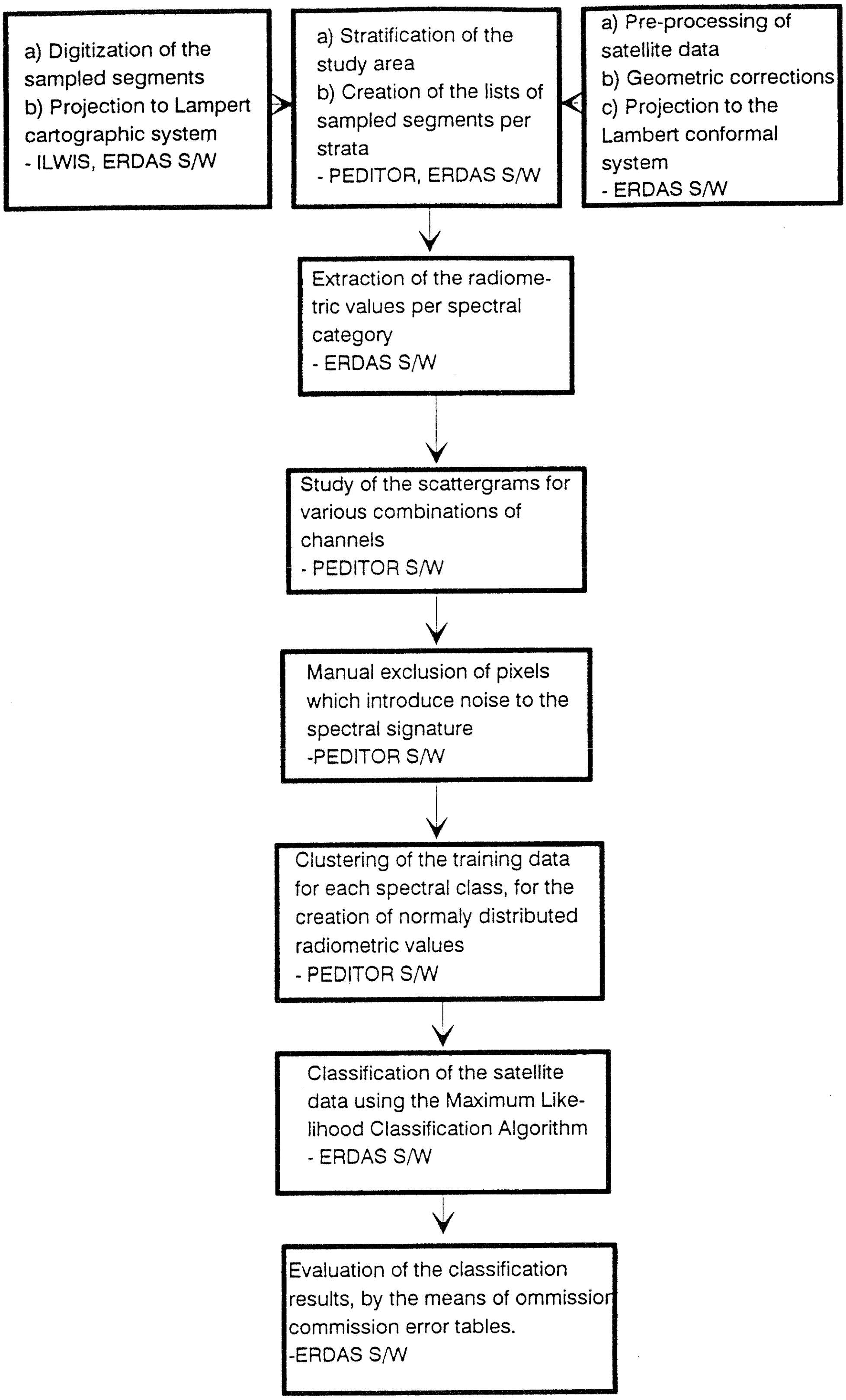

Figure 1. Works realized in order to obtain the first classification map.

Centre (1:250 000), which is the only relevant cartographic document for the whole region. The second data set used, describes the proximity to the main road network, since land use and crop cultivation is influenced by the ease of access to the transportation networks. The road network was digitized from the topographic map of the region, in the scale of 1:25 000.

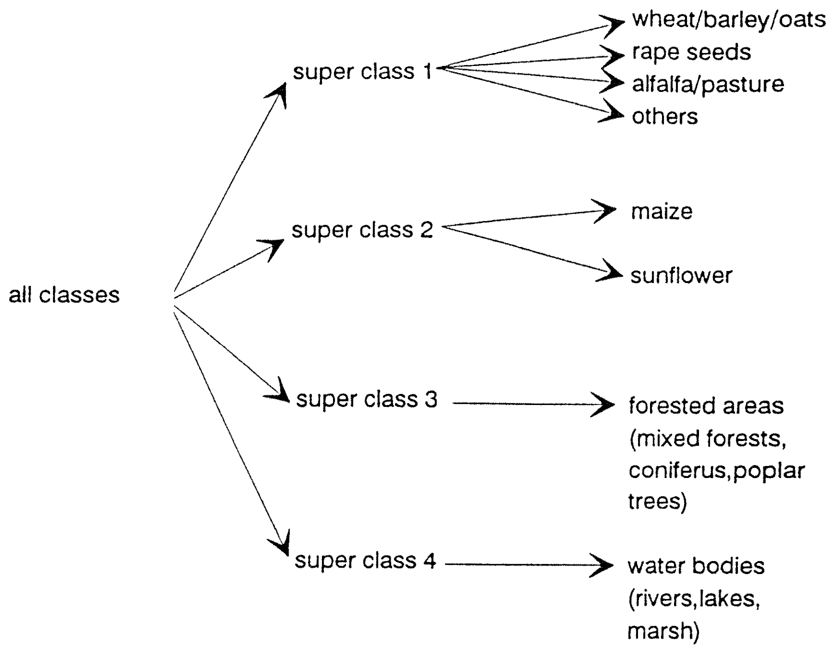

Figure 2. Classification scheme used for the Herbault region in Loir et Cher.

3. Experimenting with the Supervised Relaxation Algorithm

3.1. *Model description*

The use of the Supervised Relaxation Algorithm in the implementation of the post-classification process on the SPOT-XS satellite data with the integration of ancillary contextual information, is illustrated in the present section. The Supervised Relaxation Algorithm, which has been suggested by Richards *et al.* 1981, is an iterative and heuristic approach that extracts contextual information in a scene, by using multi-type data characteristics. The core of the algorithm, is the probability updating rule, introduced by Zucker and Mohammed 1978. Its iterative application, results in a better consistency among pixel labels, with respect to their neighbourhoods and the auxiliary data sources. It has been shown, that after a certain number of iterations, the algorithm converges to a probability vector, which is close to the one of the initial classification. At the same time the label consistency increases and the labelling uncertainty decreases (Zucker and Mohammed 1978, Faugeras and Pratt 1981). The Supervised Relaxation Algorithm is given by the following formula:

$$P_i^{(k+1)}(\lambda_r) = \frac{P_i^{(k)}(\lambda_r)R_i^{(k)}(\lambda_r)}{\sum_{\lambda_s=1}^{m} P_i^{(k)}(\lambda_s)R_i^{(k)}(\lambda_s)} \tag{1}$$

where, the spatial factor R is given as:

$$R_i^{(k)}(\lambda_r) = T_i(\lambda_r) \sum_{j=1}^{J} c_{ij} \sum_{\lambda_s=1}^{m} P_{ij}(\lambda_r \mid \lambda_s) P_j^{(k)}(\lambda_s) \tag{2}$$

and the factor T as:

$$T_i(\lambda_r) = 1 + \beta(mP_i^{(0)}(\lambda_r) - 1) \tag{3}$$

Also with $P_i^{(k+1)}(\lambda_r)$ it is denoted the probability that to the pixel i at the $k+1$ iteration is assigned the label λ_r, $P_{ij}(\lambda_r \mid \lambda_s)$ denotes the conditional probability that to the pixel i is assigned the label λ_r, given that the neighbourhood pixel j belongs to the class λ_s, P_j (λ_s) is the probability that to the pixel j is given the label λ_s,

$R_i^{(k)}(\lambda_r)$ is the so-called spatial factor which expresses the influence of the pixel neighbourhood in the classification of the pixel at the k iteration, $T_i(\lambda_r)$ is an operator based on the probability vector of the initial classification (iteration 0) and supervises the output of the algorithm at each iteration and, the factor β is the so-called 'degree of supervision'. This factor is defined heuristically by the user in the range $[0, 1]$ and expresses his confidence to the ancillary information.

From the above given formulas, it becomes clear that, if the user confidence to the initial classification is low ($\beta \rightarrow 0$), or the initial classification map is characterized by maximum entropy ($H_{\max} = (m-1)/m$, with $m =$ total number of classes), the Supervised Relaxation Algorithm, is identical to the Probabilistic Relaxation Algorithm, where initial classification is not considered in the post-classification process.

In the present study, the contribution of the initial classification map to the post-classification process was not expected to be of special importance, since it was dominated by significant missclassification errors. Therefore, it was decided to use the additional ancillary information for the supervision of the algorithm. In order to calculate the probability vector $P_i^{(k+1)}(\lambda_r)$ at each iteration, the following map layers have been used:

(*a*) The classification map which is the output of the Algorithm at iteration k.
(*b*) The probability layer associated with the classification map in (*a*) above.
(*c*) The super-class classification map of the study area.
(*d*) The probability layer associated with the super-class classification.

Spatial dependency among labels is incorporated via the compatibility coefficients $P_{ij}(\lambda_r, \lambda_s)$. Their computation is based on the following formulation:

The conditional probability is given as:

$$P_{ij}(\lambda_r \mid \lambda_s) = \frac{P_{ij}(\lambda_r, \lambda_s)}{\overline{P_j}(\lambda_s)} \tag{4}$$

For each one of the pixels i and j, it is possible to calculate two different likelihood values, denoting their assignment to classes λ_r and λ_s, respectively. The production of the two likelihood values, is based on the use of the two probability layers in (*b*) and (*d*) above. The value $\overline{P_j}(\lambda_s)$ is calculated as the mean value of the likelihoods $P_j(\lambda_s)$, which are present in the interior of a certain geometric window. This window, is user defined and it is much larger than the 4-connected one, examined by the Supervised Relaxation Algorithm. The joint probability $P_{ij}(\lambda_r, \lambda_s)$, is also estimated inside the same geometric window. It is given as a function of the degree of supervision β and the product of the two likelihood values mentioned above. The product of the likelihood values, reflects the influence of the ancillary super-class layer in the post-classification process. The calculation of the parameters $\overline{P_j}(\lambda_s)$ and $P_{ij}(\lambda_r, \lambda_s)$, is realized in various sites inside the study area, since their values are expected to change from place to place on the image plane. Therefore, the geometric window scans the image plane and for each new position new calculations are realized. One can easily conclude the following, in regards to the definition of the compatibility coefficients, as well as the likelihood updating rule of equation (1):

(*a*) If the classification of the pixel, is supported by the super-class classification (that is the given super-class is an ancestor of the final class with respect to the classification hierarchy of figure 2), any increase to the degree of supervision β results in the increase of the pixel classification likelihood.

(*b*) In contrast, if the classification label given to a certain pixel, is not supported by the super-class classification, any increase to the parameter β will result in decrease to the pixel classification likelihood.

3.2. *Model application*

The Supervised Relaxation Algorithm was tested on the two areas Cour-Cheverny and Herbault, of the Loir et Cher region. In general the results were in agreement with what Zucker and Mohammed (1978) and Faugeras and Pratt (1980) have shown. After a certain number of iterations, the algorithm converges to a probability vector, which is close to the one of initial classification and it is not changing further with the iteration number (case of Herbault region). However, in the present study, the experiments have shown, that in some cases and as a function of the degree of supervision, the accuracy level reaches a certain maximum value and then turns to lower values. Examining the output map, at each iteration, it became clear that the classification of some pixels found at the boundaries of classified objects, is subject of further change. This results in object expansion on the classification map and thus in significant changes of the thematic and statistical accuracy of the classification map from iteration to iteration. This problem becomes more apparent in intense agricultural areas, where changes in field boundaries have a negative impact on overall statistical and thematic accuracy.

The evolution of the classification accuracy, with respect to the iteration number and the degree of supervision β, for the regions Cour-Cheverny and Herbault, is illustrated in figures 3 and 4 respectively. It should be noted, that the iterative aspect of the Supervised Relaxation Algorithm, makes its application very time consuming. A single iteration on a single processor VAX 11/785 machine needs approximately

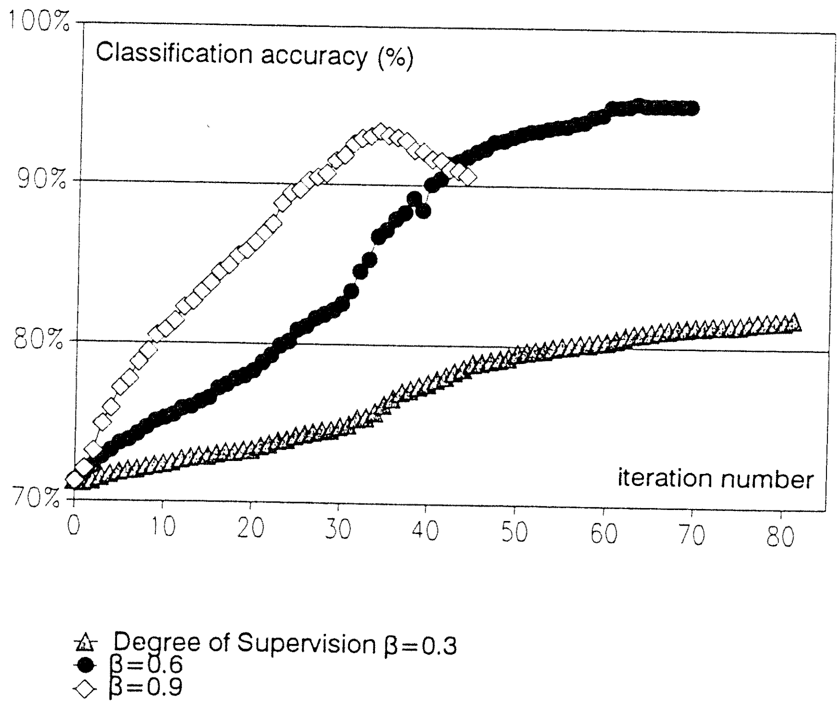

Figure 3. Change of the classification accuracy as a function of the degree of supervision and the iteration number. The study area is around the Cour Cheverny city.

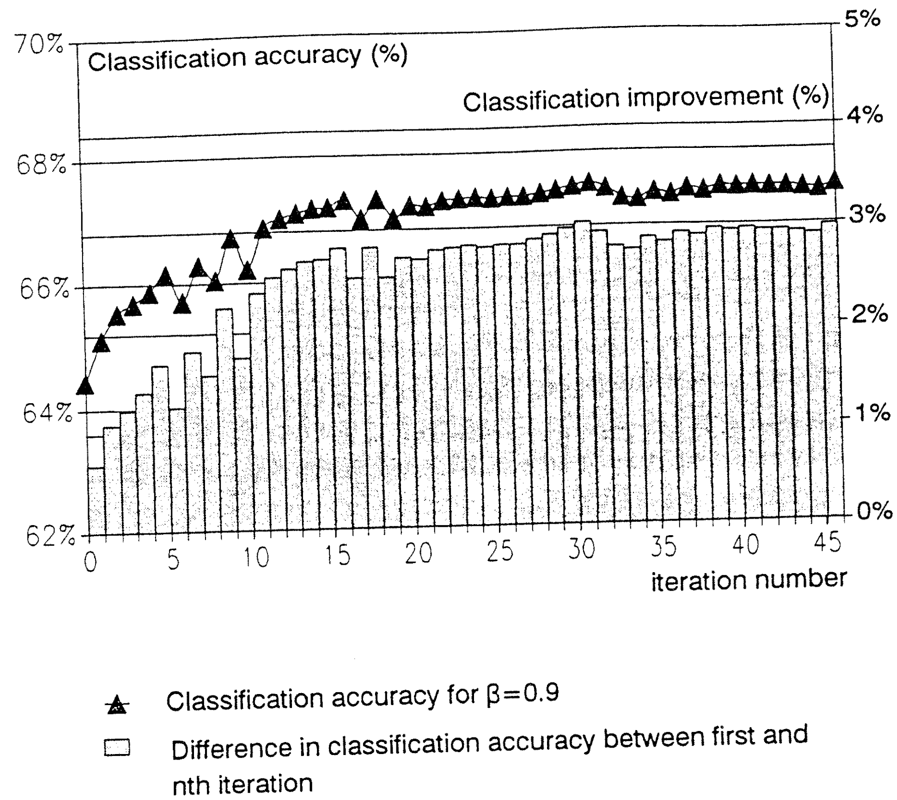

Figure 4. Classification accuracy and classification improvement as a function of the iteration number. The study area is the agricultural region Herbault.

37 m 44 s of CPU time. The use of parallel processing techniques would eventually reduce a lot the CPU time requirements.

4. Experimenting with the knowledge-based system

4.1. *Handling uncertainty. Use of belief functions*

One of the main drawbacks of the models commonly used in classification approaches, is that pixels are forced to take a single class label, even though the likelihood values committed to competing labels do not differ significantly. Thus, a method for handling uncertainty is required, in order to avoid class assignments which are vaguely supported or even to classify pixels in one of the super-classes, especially when there is not enough evidence for single class assignment. The latter is a very important feature, since a correct super-class is arguably thematically more useful than an incorrect fine class of the lower level of the classification hierarchy of figure 2. It is also desirable to be able to deal with a large body of evidence provided, either by a set of supplementing data-sources (super-class classification, probability layers), or/and by a set of rules encompassing the relevant scientific, heuristic and common-sense knowledge of the analyst. On the basis of the above given reflections, it was decided to use a numerical reasoning scheme which is based on the belief function method of the Dempster–Shafer theory of evidence. The Gordon–Shortliffe (1985) approximation of the theory of evidence has been adopted, for the needs of the present study, for two reasons:

(*a*) To reduce the complexity of the Dempster–Shafer model application.
(*b*) To perform a reasoning model, which accounts for hypothesis spaces like the ones of figure 2 (classification hierarchy where any of the descendant classes has only one linkage with an ancestor class).

Another important feature of the theory of evidence, is that it provides the means for the mathematical expression of the notions *ignorance* and *uncertainty*, which are considered as necessary components in the decision-making process.

The application of the Gordon–Shortliffe approximation, handles uncertainty and manipulates an unlimited amount of confirming or disconfirming (positive and negative) evidence, according to the following scheme:

(*a*) The algorithm makes the synthesis of all confirming evidence values committed to nodes of the classification tree and computes the basic probability assignment functions.
(*b*) It propagates the positive evidence through the classification tree and combines appropriately the evidence by the use of the Dempster–Shafer combination rule. It calculates the final belief values committed to each one of the classes of the classification tree.
(*c*) The algorithm makes the synthesis of disconfirming evidence in the three following cases:

Case 1: Negative evidence, is committed to the ancestor X of a certain class A. According to the theory, when strong disconfirming evidence is committed to a hypothesis X which is an ancestor of class A, it results in the decrease of belief in A.
Case 2: Negative evidence, is committed to the descendant class of a certain parent class. In such a case the final belief value committed to the parent class is also decreased.
Case 3: Negative evidence, is committed to a class X which makes up the only one 'brother' class of a certain class A. This has the effect of increasing the belief in A.

4.2. *Knowledge-based system design*

The knowledge base system separates the knowledge required to solve the pixel classification problem from the problem-solving mechanism. It was previously mentioned, that the problem-solving mechanism utilizes the Gordon–Shortliffe approximation of the Dempster–Shafer reasoning model. The knowledge base comprises a set of 29 rules relating to image context, 23 rules accounting for the geographic context of the pixel and a set of data and facts. Data and facts refer to the classification of the 4-connected window, which scans the classification maps in both levels of the classification tree, plus the GIS maps which introduce geographical parameters, that is the soil map and the road network map.

Data and facts are variables instantiated to integers, reals or Boolean values which provide the knowledge based system with various types of information like:

The 4-connected window location.
The central pixel label.
The corresponding 'ancestor' class label according to the classification tree structure.
The number of pixel labels in the 4-connected which match the classification at the super-class level.
The number of surrounding labels which match the label of the central pixel in the 4-connected window.