Fuzzy Image Classification and Combinatorial Optimization Strategies for Exploiting Structural Knowledge

Hirotaka Suzuki [1], Pascal Matsakis [2] and Jacky Desachy [3]

[1] Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse (IRIT)
118 route de Narbonne, 31062 Toulouse Cedex, France (e-mail: Hirotaka.Suzuki@irit.fr)

[2] University of Missouri-Columbia, Department of Computer Engineering and Computer Science
201 Engineering Building West, Columbia, MO 65211, USA (e-mail: pmatsakis@cecs.missouri.edu)

[3] Université des Antilles et de la Guyane, UFR Sciences Exactes et Naturelles
Campus de Fouillole, 97159 Pointe à Pitre, Guadeloupe, France (e-mail: Jacky.Desachy@univ-ag.fr)

***Abstract*-The present work describes a new approach to integrating structural knowledge into the image classification process. First, a fuzzy classifier produces a fuzzy partition of the image. Then, the defuzzified (crisp) partition is tried to be improved. According to the membership degrees in the fuzzy partition, the system selects a set of pixels and associates a set of candidate classes with each of them. The initial crisp partition is improved by reassigning each selected pixel to one of the classes it may belong to. This is performed by a combinatorial optimization strategy. The aim is to maximize the adequacy between the regions defined by the crisp partition and the available structural knowledge. First experiments on synthetic data as well as on simple real data show the applicability of our approach.**

I. Introduction

In remote sensing image classification, integration of additional data into conventional spectral analysis-based approaches has been investigated with the aim of achieving higher accuracy. The multisource classification model relies on pixel-by-pixel classification techniques that combine spectral information with various forms of data related to individual pixels (e.g., multisensor data, multitemporal data, ancillary data like digital elevation models, geological and topographic maps, symbolic data such as model-based knowledge represented by if-then rules, etc.) [1][2][3]. In particular, data fusion methods relying on Dempster-Shafer's theory and fuzzy set theory have been successfully employed for combining data with uncertain and incomplete information from different sources [4][5][6]. However, as expressed by the contextual classification model, the pixels should not be classified independently from their neighbors [7]. Relaxation methods are representative of this concept [8]. Nevertheless, human photo interpreters also implicitly use *structural* knowledge in the manual classification process. They not only consider contextual information but also information about the shape of and the spatial relations between the image regions. This type of knowledge has not been utilized in former systems.

The present work is part of our ongoing investigation about the integration of structural knowledge into the image classification process [9][10]. Essentially, we are faced with three fundamental problems: (1) How to represent the expert knowledge? (2) How to measure the adequacy between an image region and the knowledge that is supposed to concern it? (3) How to exploit such a measurement in the classification process? The first two problems are discussed briefly in Section II. Our main interest here is in the third problem. In Section III, we describe a Region Modification Approach by Simulated Annealing. The approach is validated in Section IV with some experimental results.

II. Knowledge Representation

Expert knowledge contains uncertain, incomplete and vague information. Consequently, the use of a fuzzy inference system appears to be justified. Fuzzy production rules for image classification are typically of the following form (although the premise and consequent are sometimes reversed) [3][4][7]:

If (class c),
then (V_1 is A_1) and ... and (V_i is A_i) and ... and (V_N is A_N).

The consequent term characterizes the environmental context of the class. It is composed of elementary propositions such as "V_i is A_i," where V_i denotes a variable and A_i a fuzzy set. These propositions are connected by logical ANDs (ORs are also admitted). When the V_i's are related to pixels (i.e., when each pixel can be considered independently of the others), the production rule represents *pixel-related knowledge*. It involves pixel-related features (e.g., spectral data, altitude) [4].

Now, consider the following expert knowledge: "class 1 appears principally in the shape of little circular regions." At this point the pixels can no longer be classified independently from the others. This type of knowledge is called *structural* knowledge. In the example above, there are two elemental knowledge terms, "circular region" and "little region," which implicitly involve specific features (e.g., aspect ratio, area). In the present paper, we assume that we know how to relate an appropriate set of variables to all of the elemental structural knowledge, and further, how to measure these variables. A multisource fuzzy inference system can be used to represent structural knowledge, and to measure its adequacy with image objects [2][7]. For example, in the framework of hierarchical fuzzy production rules, the structural knowledge SK_c about the class c is represented as follows:

Upper level rule:
If (class c),
then (ESK_1) and ... and (ESK_j) and ... and (ESK_N).
Lower level rule:
If (ESK_j),
then (V_1^j is A_1^j) and ... and (V_i^j is A_i^j) and ... and ($V_{M_j}^j$ is $A_{M_j}^j$).

The consequent term of the upper level rule is composed of elemental knowledge terms ESK_j connected by logical ANDs (ORs are also admitted). In the lower level rule, each ESK_j is represented by measurable variables V_i^j and fuzzy sets A_i^j. Note that the V_i^j's are not concerned with pixels but with regions. Consider a region R that may be assigned to class c. The membership degree in A_i^j of the value V_i^j obtained at R

corresponds to the degree of truth $\mu_{A_i^j}(R)$ of the proposition "V_i^j is A_i^j." The logical combination of the $\mu_{A_i^j}(R)$'s gives $q^{ESK_j}(R)$, which is the degree of adequacy between the region R and the elemental structural knowledge ESK_j. Finally, at the upper level, the logical combination of the $q^{ESK_j}(R)$ gives $q^{SK_c}(R)$, the degree of adequacy between R and SK_c. Consider, for instance, the knowledge SK_2: "If (class 2), then (ESK_1) and (ESK_2)" where the lower level rules are "If (ESK_1), then (V_1^1 is A_1^1) or (V_2^1 is A_2^1)" and "If (ESK_2), then (V_1^2 is A_1^2)." We get: $q^{SK_2}(R) = min\{max\{\ \mu_{A_1^1}(R), \mu_{A_2^1}(R)\ \}, \mu_{A_1^2}(R)\ \}$. In the following, we assume we are able to evaluate any adequacy degree.

III. Knowledge Integration

Let X^{init} be a crisp partition of the studied image. X^{init} is composed of crisp regions. Consider any region R. It has been assigned to some class. If structural knowledge about this class is available, then a degree of adequacy can be computed and associated with R (Section II). We define a *global adequacy degree* $Q(X^{init})$—the degree of adequacy between X^{init} and all the available structural knowledge—as follows: $Q(X^{init}) = F(q_1, ..., q_r, ..., q_n)$, where F denotes a combination operator, q_r the adequacy degree associated with the r-th region, and n the total number of regions on which knowledge is available. For our experiments (Section IV), we used the classical arithmetic mean. However, other operators can be considered (e.g., *min*).

The aim of the Region Modification Approach (RMA) presented here is to improve classification accuracy by exploiting the ability to evaluate global adequacy degrees. More precisely, the RMA strives to increase the global adequacy degree of X^{init} by reassigning some pixels. The problem corresponds to a combinatorial optimization problem. It aims at finding an optimal reassignment of pixels, i.e., a partition X that maximizes $Q(X)$. Three heuristic iterative methods are well known as efficient methods for answering optimization problems and have been successfully applied to clustering and image classification [11][12][13]: Simulated Annealing (SA), Tabu Search, and Genetic Algorithm. In a previous study [9], we tested the RMA on a simple image, and compared the use of the three heuristic methods. SA seemed to be the most appropriate. Therefore, in this paper, we focus on the RMA by SA. Here is the algorithm:

Initialization: Select a value for *maxLoop1* (maximum number of iterations for **loop 1**), *maxLoop2* (maximum number of iterations for **loop 2**), and T_0 (initial temperature). Select a cooling schedule $\alpha(T)$ (in Section IV we use the classical $\alpha(T)=T_0/t$). Initialize the iteration counter t and the current partition X^{now} (i.e., $t \leftarrow 1$ and $X^{now} \leftarrow X^{init}$).

Iteration loop 1

> **Iteration loop 2**
>
> 1. From X^{now}, generate a candidate partition X^{cand}.
> 2. Calculate $\delta = Q(X^{now}) - Q(X^{cand})$. If $\delta<0$, replace X^{now} by X^{cand}. Otherwise, select a random number $r \in [0,1]$, and if $r<exp(\delta/T)$, then replace X^{now} by X^{cand}.
>
> Repeat **loop 2** *maxLoop2* times.
>
> Update the counter ($t \leftarrow t+1$) and the temperature ($T \leftarrow \alpha(T)$).

Repeat **loop 1** until t exceeds *maxLoop1* or $Q(X^{now})$ reaches a predefined value. X^{now} is the final partition.

It is assumed that the crisp partition X^{init} is obtained by defuzzifying the partition $\tilde{X}$ issued from some fuzzy pre-classification. From a crisp partition X^{now} (current solution, equal to X^{init} at the first iteration), an alternative crisp partition X^{cand} is generated by reassigning a randomly chosen pixel, $\bar{P}$, to a randomly chosen class, $\bar{C}$. For any pixel p and any class c, let $\mu_c(p)$ be the membership value of p to class c in $\tilde{X}$, let $X^{init}(p)$ be the class of p in X^{init} and $X^{now}(p)$ the class of p in X^{now}. We have: $\mu_{X^{init}(p)}(p) \geq \mu_c(p)$. It is obvious that a pixel p satisfying $\mu_{X^{init}(p)}(p) >> \mu_c(p)$ for any $c \neq X^{init}(p)$ would not constitute a judicious choice for $\bar{P}$. Similarly, a class c such that $\mu_{X^{init}(\bar{P})}(\bar{P}) >> \mu_c(\bar{P})$ would not constitute a good choice for $\bar{C}$. Let $\gamma(p)$ be the value $min_{c \neq X^{init}(p)}\{\mu_{X^{init}(p)}(p) - \mu_c(p)\}$. "Random" choices can be restrained so that the lower $\gamma(p)$, the higher the probability of setting $\bar{P}$ to p, and then, the higher $\mu_c(\bar{P})$, the higher the probability of setting $\bar{C}$ to c. Therefore, the crisp partition X^{cand} is generated as follows (with σ being a predefined threshold between 0 and 1):

[Step 1] Select $\bar{P}$ by a weighted random choice: the probability of choosing a given pixel p_0 is set to $\gamma^*(p_0)/\Sigma_p \gamma^*(p)$, where $\gamma^*(p)$ equals $1-\gamma(p)$ if $\gamma(p) \leq \sigma$, and equals 0 otherwise.

[Step 2] Select $\bar{C}$ by a weighted random choice: the probability of choosing a given class c_0 is set to $\mu^*_{c_0}(\bar{P})/\Sigma_c \mu^*_c(\bar{P})$, where $\mu^*_c(\bar{P})$ equals $\mu_c(\bar{P})$ if $\mu_{X^{init}(\bar{P})}(\bar{P}) - \mu_c(\bar{P}) \leq \sigma$ and equals 0 otherwise.

[Step 3] If $\bar{C}$ is equal to $X^{now}(\bar{P})$, return to Step 1. Otherwise, modify X^{now} by reassigning $\bar{P}$ to class $\bar{C}$, and call X^{cand} the partition that results from this modification.

Note that the fuzzy partition $\tilde{X}$ can be either probabilistic, as in [14], or possibilistic, as in [15]. In the second case, $\Sigma_c \mu^*_c(\bar{P})$ is not necessarily equal to 1, even if $\sigma=1$. Finally, if the pre-classification process is crisp, its output X^{init} can be fuzzified and used to produce $\tilde{X}$. Let p be a pixel, and let R be a region assigned to class c in X^{init}. If p belongs to R, we can give $\mu_c(p)$ a value directly proportional to the distance of p from the boundary of R. Otherwise, we can give $\mu_c(p)$ a value inversely proportional to the distance of p from all regions assigned to class c. Then, the RMA would tend to round the regions which are expected to be round, to elongate those which are expected to be elongated, etc.

IV. Experimental Results

A. One-Dimensional Synthetic Data

The fuzzy partition shown in Fig. 1(a) classifies 161 ordered elements into 7 classes. The crisp partition X^{init} associated with this fuzzy partition is shown in Fig. 1(b). Fig. 1(c) presents the product of manual classification by human expert. It is the control partition. We consider the following structural knowledge: "classes 1, 2 and 7 appear as little regions; classes 4 and 6 appear as medium size regions; classes 3 and 5 appear as big regions; for each $c \neq 1$, a region of class c appears next to and on the right-hand side of a region of class $c-1$." This synthetic set of data and knowledge is inspired by the geomorphological classification of an atoll [17]. The data represents the pixels of a cross section of the atoll rim. Along the cross section, from the ocean to the lagoon, the geomorphological classes are not situated in random order. The class *reef front* appears first, then the class *outer reef flat*, and the class *coral reef conglomerate*, etc. To represent this knowledge, we chose a hierarchical set of fuzzy production rules (Section II), where two-level rules

are associated with each class. In the upper level we have two elementary propositions. One concerns the size of regions and the other concerns the spatial relations between regions. For instance, this is SK_2: "If (class 2), then (ESK_1: "little region") and (ESK_2: "the neighboring region on the right-hand side is of class 1")." In the lower level, the variable *size* (i.e., number of pixels) is attached to the first elemental knowledge ESK_1. For any region R, the variable *size* is used to evaluate $q^{ESK_1}(R)$, the degree of adequacy between ESK_1 and R (Fig. 1(d)). The value $q^{ESK_2}(R)$ is obtained in a simpler way. Let c be the class to which R is assigned: $q^{ESK_2}(R)=1$ if R appears next to a region of class $c-1$ (on the right-hand side), and $q^{ESK_2}(R)=0$ otherwise. Finally, the adequacy degree $q^{SK_2}(R)$ is $min\{q^{ESK_1}(R), q^{ESK_2}(R)\}$. At this stage, for any partition X, it is possible to evaluate $Q(X)$, the degree of adequacy between X and the whole available structural knowledge (Section III). For the control partition (Fig. 1(c)), we obtain $Q=0.96$. For X^{init} (Fig. 1(b)), we get $Q=0.32$. The percentage of well classified pixels is 82.6%.

The RMA by SA has been tested 200 times: 100 times with $\sigma=0.03$, $maxLoop2=10$, $T_0=0.001$, and 100 times with $\sigma=0.1$, $maxLoop2=500$, $T_0=0.01$. With the first threshold, 66 pixels are candidate for reassignment (out of 161), and each one can be reassigned to 2 candidate classes (out of 7). In other words, there are $2^{66}=7.38\times10^{19}$ possible solutions in the search space. With each iteration, the global adequacy degree increases (Table 1). Some tests output the control partition after 50 iterations only: Q is 0.96, and the percentage of well classified pixels is 100%. They also output other partitions with $Q=0.96$. After 400 iterations, every test gives $Q=0.96$, and on average, the percentage of well classified pixels is 98.9% (+16.3%, in 776.4 ms of CPU time on a P300 MMX laptop). Table 2 shows the results obtained with the second set of parameters. When $\sigma=0.1$, 126 pixels are candidate for reassignment: 86 pixels can be reassigned to 2 candidate classes each, and 40 pixels to 3 candidate classes each. This time, there are 9.41×10^{44} possible solutions in the search space. The convergence is much slower. After 10 iterations (computation of 5000 partitions since $maxLoop2=500$), some tests output partitions for which $Q=1$. After 100 iterations, 72 tests out of 100 produce partitions for which Q is higher than 0.96. However, on average, the percentage of well classified pixel is only 84.6% (+2.0% only, in 8367.4 ms of CPU time). Note that the system always tries to find a partition that maximizes Q, even if the degree of adequacy attached to the control partition is not maximum (which is the case here). As a result, even if the system succeeds in reaching a global maximum ($Q=1$), the output may not correspond to the control partition, and the percentage of well classified pixels may not be 100%.

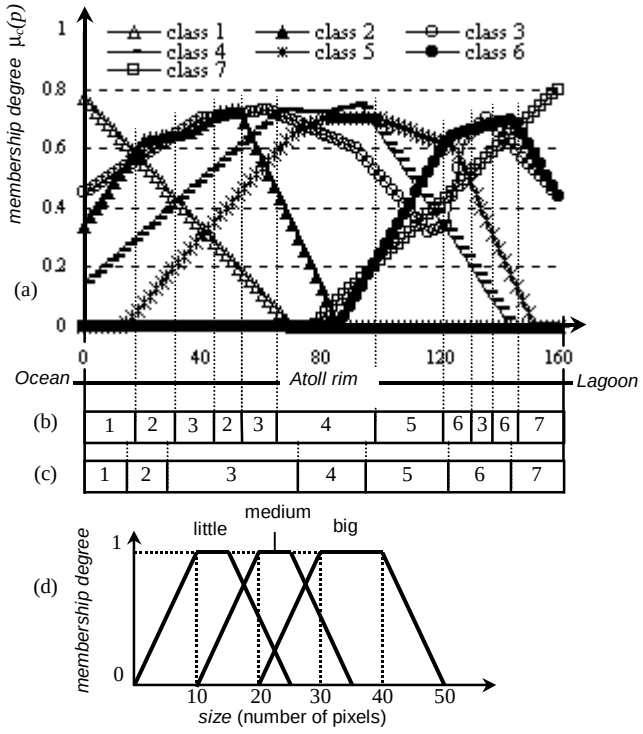

Fig. 1. Synthetic data. (a) Fuzzy partition (pre-classification output). (b) Corresponding crisp partition. (c) Control partition issued from manual classification. (d) Linguistic values for *size*.

B. Two-Dimensional Real Data

At this time, we are preparing a series of tests on LANDSAT 7 multispectral images of atolls. To classify this type of image, the human photo interpreter implicitly utilizes structural knowledge. In this paper, and for our first experiments on 2D real data, a very simple image is considered. Its manual classification does not require specific expertise. It is a very small RGB image (115×54) that represents a knife handle with three rivets, as shown in Fig. 2(a). There are four classes: *background*, *shadow*, *handle* and *rivet*. The defuzzification of the fuzzy partition generated by some supervised classifier produced the crisp partition shown in Fig. 2(b). This pre-classification stage was based on the spectral analysis of individual pixels, and will not be described here. In the crisp partition, 33% of the *rivet* pixels (51 pixels out of 156) are misclassified. We can clearly see anomalies on the right and left rivets. With the aim of improving this pre-classification result, the following structural knowledge is considered: "a rivet appears as a little circular region." To represent this knowledge, we utilize the linguistic variable *area* (for "little region," see Fig. 2(f)), and the variables *aspect ratio* and *density* (for "circular region," see Fig. 2(d)(e)). The degree of adequacy between the initial crisp partition and the knowledge is $Q=0.23$.

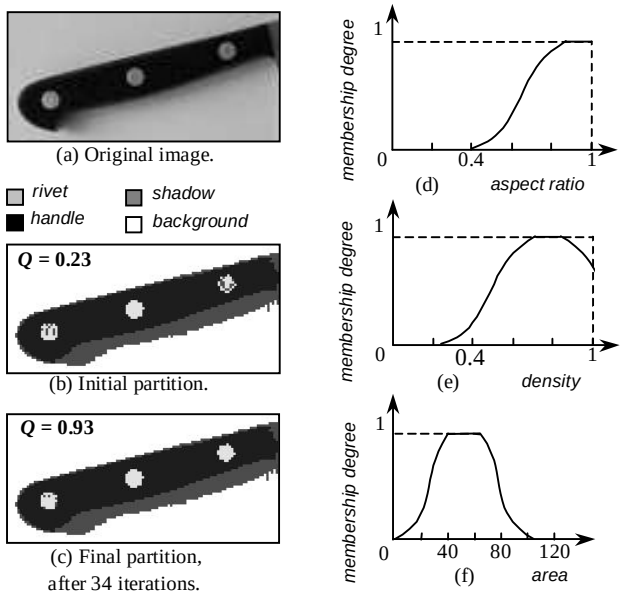

Fig. 2. Real data.

TABLE 1. RESULTS OF 100 TESTS WITH PARAMETERS σ=0.03, *maxLoop2*=10, T_0=0.001.

Iteration	# of partitions	Time (ms.)			Global adequacy degree, *Q*			Percentage of well classified pixels		
		Min	Average	Max	Min	Average	Max	Min	Average	Max
0	0	0	0	0	0.32	0.32	0.32	82.6	82.6	82.6
50	500	50	102.3	220	0.39	0.55	0.96	80.1	90.9	100
100	1000	160	205.9	330	0.43	0.71	0.96	81.4	93.7	100
200	2000	330	396.4	500	0.52	0.91	0.96	84.5	97.7	100
300	3000	540	587.6	720	0.63	0.94	0.96	90.1	98.6	100
400	4000	710	776.4	880	0.96	0.96	0.96	98.1	98.9	100

TABLE 2. RESULTS OF 100 TESTS WITH PARAMETERS σ=0.1, *maxLoop2*=500, T_0=0.01.

Iteration	# of partitions	Time (ms.)			Global adequacy degree, *Q*			Percentage of well classified pixels		
		Min	Average	Max	Min	Average	Max	Min	Average	Max
0	0	0	0	0	0.32	0.32	0.32	82.6	82.6	82.6
1	500	50	139.7	220	0.32	0.34	0.54	68.3	81.7	88.8
10	5000	870	1235.3	1490	0.36	0.65	1.0	50.3	73.0	96.8
50	25000	4180	4650.7	5000	0.64	0.94	1.0	59.6	83.7	96.9
100	50000	7850	8367.4	9390	0.73	0.96	1.0	69.6	84.6	96.9
200	100000	15210	15860.3	18340	0.73	0.96	1.0	69.6	84.6	96.9

The RMA by SA has been tested 100 times, with σ=0.05. With this threshold, 75 pixels are candidate for reassignment. The 100 partitions that have been generated are all partitions with *Q*=0.93. However, they are all different, and the percentage of misclassified *rivet* pixels varies from 19% (30 pixels out of 156) to 22% (34 pixels out of 156). Fig. 2(c) shows one output. The percentage of misclassified *rivet* pixels has been reduced from 33 % to 21 %. Despite the high adequacy degree, the rivets obtained are not perfectly circular (especially the left rivet that contains patches misclassified as *shadow*). It is obvious that the quality of the results produced by RMA not only depends on the quality of the initial fuzzy partition, but also on the quality of the structural knowledge, and on the quality of its representation. Note that the average time for partitions with *Q*=0.93 to be generated was 38.1 seconds of CPU time on a P300 MMX laptop.

V. CONCLUSION

In this paper, we have introduced a novel approach that aims at improving the automatic classification of remote sensing images by exploiting expert structural knowledge. It is based on the computation of a fuzzy partition, and the use of a combinatorial optimization strategy. We have presented first experiments on a 1D synthetic image, as well as on a RGB real image of simple and strongly structured objects. The results are encouraging. The crisp partitions issued from the pre-classification stage were coherently modified by appropriately reassigning initially misclassified pixels. However, many factors can affect the results (e.g., the quality of the fuzzy partition, the quality of the knowledge, the way to represent that knowledge and to evaluate its degree of adequacy with image regions), and much work still has to be done. Also, the computational time is a practical issue that cannot be ignored. Remote sensing images often contain millions of pixels. For instance, how to choose the threshold that is used to select candidate classes? The lower the threshold, the lower the computational time, but the less significant the potential improvement over the initial crisp partition. Which value constitutes a good compromise? We intend to introduce a dynamic threshold controlled with a decreasing function of the number of iterations. We also intend to integrate contextual information into the pixel reassignment process. At the moment, only one pixel differentiates a candidate partition from its parent partition. Many pixels could be simultaneously reassigned, especially neighbors with similar membership values.

ACKNOWLEDGEMENT

The authors wish to thank D. Dubois and H. Prade for their precious advice and comments.

REFERENCES

[1] J. A. Richards, D. A. Landgrebe, and P. H. Swain, "A means for utilizing ancillary information in multispectral classification," *Remote Sensing Environ.*, no. 12, pp. 463-477, 1982.

[2] F. Russo, and G. Ramponi, "Fuzzy methods for multisensor data fusion," *IEEE Trans. Instrumentation and Measurement*, vol. 43, no. 2, pp. 288-294, 1994.

[3] P. Clark, C. Feng, S. Matwin, and K. Fung, "Improving image classification by combining statistical, case-based and model-based prediction methods," *Fundamenta Informatica*, vol. 30, no. 3-4, pp. 227-240, 1996.

[4] E. H. Zahzah, and J. Desachy, "Numeric and symbolic data combination for satellite image interpretation," in *Proc. IGARRS'93*, Tokyo, Japan, 1993, pp. 1704-1706.

[5] L. Tong, J. A. Richards, and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 25, pp. 283-293, May 1987.

[6] I. Bloch, "Information combination operators for data fusion: A comparative review with classification," in *Image and Signal Processing for Remote Sensing, Proc. SPIE*, J. Desachy, Ed, 1994, vol. 2315, pp. 148-159.

[7] E. Binaghi, P. Madella, M.G. Montesano, and A. Rampini, "Fuzzy contextual classification of multisource remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, no. 2, pp. 326-339, 1997.

[8] 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.

[9] H. Suzuki, P. Matsakis, and J. Desachy, "Exploitation de connaissances structurelles en classification d'images : utilisation de méthodes heuristiques d'optimisation combinatoire," in *Proc. ORASIS 2001*, Cahors, France, June 2001, pp. 455-64.

[10] H. Suzuki, P. Matsakis, S. Andréfouët, and J. Desachy, "Satellite image classification using expert structural knowledge: a method based on fuzzy partition computation and simulated annealing," in *Proc. IAMG 2001*, Cancún, Mexico, December 2001, in press.

[11] M. Bao, "Classification of multi-temporal SAR images and INSAR coherence images using adaptive neighborhood model and simulated annealing approach," in *Proc. 20th Asian Conf. for Remote Sensing*, Hong Kong, China, 1999.

[12] K. S. Al-Sultan, "A tabu search approach to the clustering problem," *Pattern Recognition*, vol. 28, no. 9, pp. 1443-51, 1995.

[13] M. Sonka, S. K. Tadikonda, and S. M. Collins, "Knowledge-based interpretation of MR brain images," *IEEE Trans. Med. Imag.*, vol. 15, no. 4, pp. 443-452, 1996.

[14] I. Gath, and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. PAMI*, vol. 11, pp. 773-781, 1989.

[15] R. Krishnapuram, and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, 1993.

[16] C. R. Reeves, *Modern Heuristic Techniques for Combinatorial Problems, Blackwell Scientific Publications*. Blackwell Scientific Publications, 1993.

[17] S. Andréfouët, M. Claereboudt, P. Matsakis, J. Pagès, and P. Dufour, "Typology of atoll rims in Tuamotu Archipelago (French Polynesia) at landscape scale using SPOT HRV images," *Int. J. Remote Sensing*, vol. 22, no. 6, pp. 987-1004, 2001.