Ant colony optimization for image regularization based on a non-stationary Markov modeling

Sylvie Le Hégarat-Mascle, Abdelaziz Kallel, Xavier Descombes

▶ **To cite this version:**

Sylvie Le Hégarat-Mascle, Abdelaziz Kallel, Xavier Descombes. Ant colony optimization for image regularization based on a non-stationary Markov modeling. IEEE Transactions on Image Processing, 2007, 16 (3), pp.865-878. 10.1109/TIP.2007.891150 . hal-00159093

HAL Id: hal-00159093
https://hal.science/hal-00159093v1

Submitted on 18 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Ant Colony Optimization for Image Regularization Based on a Nonstationary Markov Modeling

Sylvie Le Hégarat-Mascle, Abdelaziz Kallel, and Xavier Descombes

***Abstract*—Ant colony optimization (ACO) has been proposed as a promising tool for regularization in image classification. The algorithm is applied here in a different way than the classical transposition of the graph color affectation problem. The ants collect information through the image, from one pixel to the others. The choice of the path is a function of the pixel label, favoring paths within the same image segment. We show that this corresponds to an automatic adaptation of the neighborhood to the segment form, and that it outperforms the fixed-form neighborhood used in classical Markov random field regularization techniques. The performance of this new approach is illustrated on a simulated image and on actual remote sensing images.**

***Index Terms*—Ant colony, classification, image model, Markov random field (MRF).**

I. Introduction

THE development of algorithms for global classification, such as maximum *a posteriori* (MAP) under Markov random field (MRF) modeling has been an area of active study, in particular, since Geman and Geman [1]. For a given neighborhood system with clique potential functions, the equivalence between MRF and Gibbs field is established, and a global energy term is defined that should be minimized. Apart from local minimization processes, such as the iterative conditional mode (ICM), global minimization is generally achieved using the simulated annealing (SA) technique. In this modeling, the neighborhood form is considered to be independent of its location in the image. This may lead to classification errors for pixels located at the segment borders or extremities, since their neighborhoods contain some pixels belonging to other segments or classes. To overcome this problem, one approach consists of defining a line process [2]. Due to the induced complexity, in the framework of image restoration, some specific potentials to preserve edges can be used as an alternative [3], [4]. Another solution could be to relax the assumption of neighbourhood stationarity. The lack of an efficient heuristic is probably the main cause of the absence of published work (to our knowledge) dealing with nonstationary neighborhoods.

S. Le Hégarat-Mascle and A. Kallel are with the CETP/IPSL, F-78140 Vélizy, France (e-mail: sylvie.mascle@cetp.ipsl.fr).

X. Descombes is with the Ariana, Joint Research Group CNRS/INRIA/UNSA, INRIA, 06902 Sophia Antipolis Cedex, France (e-mail: descombes@sophia.inria.fr).

Ant colony optimization (ACO) is currently a popular algorithm [5]. The classical application is to routing in telecommunication networks, on which a number of papers have been published [6]–[11]. ACO has been applied to various problems with good results [12]–[17]. Just as the techniques of simulated annealing and genetic algorithms imitate computing strategies arising from natural physical or biological phenomena, ACO derives from the behavior of social insects. The success achieved using such methods is due, in large part, to the introduction of randomness in the search procedure, permitting to escape from local minima and, thus, achieving a more globally favorable solution. In ACO, the information gathered by simple autonomous mobile agents is shared and exploited to solve the problem.

Ant-based systems have also been used for clustering problems, modeling real ant abilities to sort objects. For instance, in [18], the clustering is based on the definition of an "odor" associated to each sample (represented by an ant) and the mutual recognition of ants sharing a similar "odor" to construct a colonial odor used to discriminate between nest mates and intruders. In other approaches, when a specialized ant meets a given object it collects it with a probability that is the higher the sparser are the objects in this region, and, after moving, it brings in the object with a probability that is the higher the denser are the objects in this region. Besides ACO, ant colony systems (ACS) have been proposed. In [19], it is used for edge detection as follows: The ants evolve on the image itself by reinforcing pheromone levels around pixels with different gray levels or presenting gestalt features. The ant system evolves during generations moving to adjacent pixels at each generation and having more favorable survival and reproduction conditions depending on the local features of the image (for edge detection, the most contrasted the most favorable a region is). In [20], the notion of visibility determined using the maximum variation of gray level of the image intensity is used to drive the ant displacement.

In the context of global image classification under the classical MRF assumption, Ouadfel and Batouche [21] have shown that ACO produces equivalent or better results than other stochastic optimization methods like SA and genetic algorithm. In [21], the authors make an analogy between pixel labeling in the global classification problem and the assignment problem, as it is done in the graph coloring problem. Each ant constructs its own solution, and a trace of the best solution is kept (though the pheromone deposit technique) during the iterative construction of the final solution. Such a use of ACO can be seen as an extension of the SA technique. At each iteration the new solution is constructed taking into account not only the very last one, like in [1], but a combination of several previous solutions (their number depending on the pheromone evaporation

rate). Moreover, at each iteration, not only is one new solution constructed, but several (as many as ants) to produce the one which will be "memorized" (generally the best solution). In this sense, it shares some ideas with the particle filtering [22] and the Bayesian bootstrap filtering [23] techniques.

In this study, we propose to use the ACO by exploiting its ability of self-organization, according to an optimization approach which is similar to the traveling salesman problem, and the routing problem in telecommunication systems. It allows a more flexible image modeling in particular concerning the neighborhood form, and it offers a meta-heuristic to solve it. We consider a MRF modeling with a nonstationary neighborhood. We propose an ACO scheme which jointly estimates the regularized classification map and the optimal nonstationary neighborhood.

Section II deals with the image modeling. In Section III, we present the proposed method based on ACO meta-heuristics. Section IV shows some results, firstly obtained on simulated images, and secondly, obtained on an actual remote sensing image acquired by SPOT/HRVIR satellite sensor. Finally, Section V gathers our conclusions.

II. Image Model and Classification Problem

Let us introduce the following notations. N_l is the dimension in lines of the image and N_c its dimension in columns. The total number of pixels is then $N_l \times N_c$. Ω is the set of the pixel locations, and $|\Omega|$ is its cardinal: $|\Omega| = N_l \times N_c$. The image X is defined as a random field taking values in $\Re^{|\Omega|}$, and the label image L is a random field which takes values in $\Lambda^{|\Omega|}$, where $\Lambda = \{1, \ldots, c\}$ and c is the class number

$$X : \{x_s, s \in \Omega\} \rightarrow \Re^{|\Omega|}$$
$$L : \{l_s, s \in \Omega\} \rightarrow \Lambda^{|\Omega|}.$$

A. Case of Classical Stationary Neighborhood

The problem of image classification is to determine the realization of L, knowing those of X. According to the MAP criterion derived from the Bayes theory, the optimum estimate maximizes $p(X/L).p(L)$. Assuming the distribution law of the pixel values conditionally to their class, i.e., an independent Gaussian in our case, $p(X/L)$ is written as follows:

$$p(X/L) = \prod_{s \in \Omega} \frac{1}{\sqrt{2\pi}.\sigma_{l_s}} .exp\left\{ -\frac{(x_s - \mu_{l_s})^2}{2.\sigma_{l_s}^2} \right\} \quad (1)$$

$$p(X/L) = \left(\frac{1}{\sqrt{2\pi}} \right)^{|\Omega|} .exp\left\{ -\frac{1}{2} \sum_{s \in \Omega} \left[\left(\frac{x_s - \mu_{l_s}}{\sigma_{l_s}} \right)^2 + ln\left(\sigma_{l_s}^2\right) \right] \right\} \quad (2)$$

where μ_{l_s} and $\sigma_{l_s}^2$ are respectively the mean and the variance of the class of pixel s, labeled l_s.

The prior model $p(L)$ is defined assuming that a pixel and its neighbors have a high probability to share the same label. We consider a MRF modeling [24]. Defining a neighborhood system such that, if a pixel s has a pixel s' within its neighborhood $N(s)$, then s is a neighbor of s'

$$s' \in N(s) \Leftrightarrow s \in N(s'). \quad (3)$$

Then, according to the Hammerley–Clifford theorem [25], $p(L)$ follows a Gibbs distribution

$$p(L) = \frac{1}{Z}.\exp\left\{ -\sum_{\gamma \in \Gamma} \mathsf{V}_\gamma(l_s, s \in \gamma) \right\} \quad (4)$$

where Γ is the set of the image cliques γ that describe the interactions between pixels, Z is a normalization constant, and V_γ is the γ potential. Finally, the MAP criterion leads to the minimization of the global energy

$$E = \frac{1}{2}.\sum_{s \in \Omega} \left[\left(\frac{x_s - \mu_{l_s}}{\sigma_{l_s}} \right)^2 + ln\left(\sigma_{l_s}^2\right) \right] + \sum_{\gamma \in \Gamma} \mathsf{V}_\gamma(l_s, s \in \gamma). \quad (5)$$

The first two terms within the brackets are called the "data attachment" energy. The minimization of (5) is performed using the fact that the global energy difference between two label image configurations only differing by one pixel label, either $l_s^{(1)}$ or $l_s^{(2)}$, only depends on this pixel s and its neighborhood

$$\begin{aligned} \Delta E = & \left[\frac{1}{2} \left(\frac{x_s - \mu_{l_s^{(1)}}}{\sigma_{l_s^{(1)}}} \right)^2 + ln\left(\sigma_{l_s^{(1)}}\right) \right. \\ & \left. + \sum_{\gamma \in \Gamma, s \in \gamma} \mathsf{V}_\gamma\left(l_s^{(1)}, l_r, r \in \gamma\right) \right] \\ & - \left[\frac{1}{2} \left(\frac{x_s - \mu_{l_s^{(2)}}}{\sigma_{l_s^{(2)}}} \right)^2 + ln\left(\sigma_{l_s^{(2)}}\right) \right. \\ & \left. + \sum_{\gamma \in \Gamma, s \in \gamma} \mathsf{V}_\gamma\left(l_s^{(2)}, l_r, r \in \gamma\right) \right]. \end{aligned} \quad (6)$$

Generally, the pixel neighborhood is defined in the same manner at each pixel location s, i.e., it is assumed that the neighborhood geometry is stationary. For example, in the case of 4-connectivity, the neighborhood of a pixel located in (i, j) where i is the line number and j the column one, is constituted by the pixels whose locations belong to $V_4(i, j) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$, and for a 8-connectivity, the neighborhood contains the pixels located within ± 1 in line and ± 1 in column. Such an assumption of neighborhood stationarity may not be true in many cases.

B. Case of Nonstationary Neighborhood

In this section, we consider a MRF with a nonstationary neighborhood form, optimal with respect to the classification map in the sense that two pixels having different labels are unlikely to be neighbors. Let us begin with a simple example of an image with 4×5 pixels and two labels. Fig. 1 shows the nonstationary neighborhood system optimal with respect to a binary labeling (triangles and circles represent the label). Note

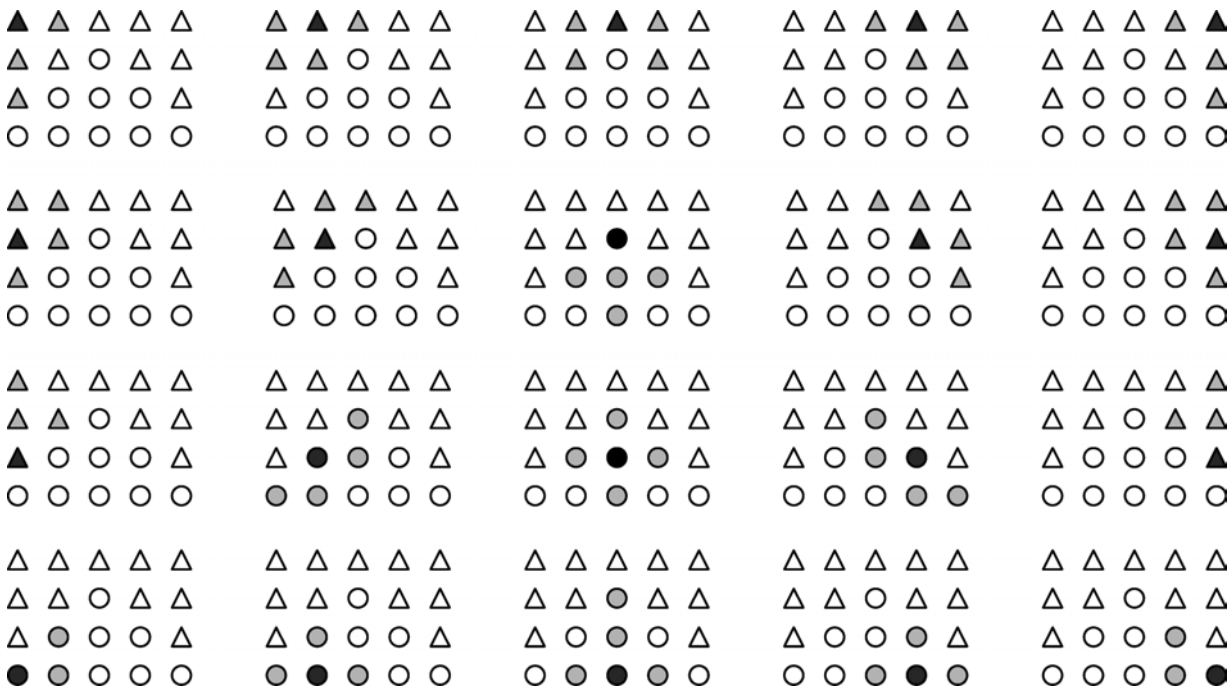

Fig. 1. Neighborhoods of each of the pixels on the simple example of an image with 4×5 pixels and two classes. The class of a pixel is represented either by a circle or a triangle. The neighborhood of pixel s is represented with s in black and its neighbors in gray.

that reciprocity property (3) is verified. More generally, the nonstationarity of the neighborhood is motivated by the diversity of the areas or object shapes: more or less large, elongated, thin, with some peaks, etc. In this study, we relax the neighborhood stationarity assumption and propose an algorithm where the neighborhood form is automatically adjusted. The used criteria to construct the neighborhood at each location pixel are:

- the neighbor pixels are connected;
- the neighbor pixels have a high probability to share the same label;
- each pixel has the same number of neighbors.

The last assumption allows us to have for each pixel of the image the same weighting between the prior and the "data attach" terms.

Now, for the classification problem, (5) is still valid except that the cliques are now defined over nonstationary neighborhoods. In the simplest case, only cliques of cardinal 2 are considered. Assuming the Potts model for the potentials, for a given neighborhood system, the function to minimize is given by

$$E = \sum_{s \in \Omega} \left[\frac{(x_s - \mu_{l_s})^2}{2.\sigma^2_{l_s}} + ln(\sigma_{l_s}) + \beta \times \frac{|\{r \in N(s); l_r \neq l_s\}|}{2} \right] \tag{7}$$

where $|\{r \in N(s); l_r \neq l_s\}|$ is the number of neighbors having a different label than s, and β is a positive parameter weighting the relative importance of the "data attach" term and the neighborhood one. The factor 1/2 is due to the fact that, for cliques of order 2, their potentials are counted 2 times when the sum is done over the pixels rather than over the image clique set.

Relaxing the neighborhood stationarity assumption, the s optimal neighborhood now depends on the label l_s. Therefore, we cannot directly obtain an expression of the energy difference between two label image configurations as simple as (6). Indeed, now the cliques involving s are not the same ones when $l_s = l_s^{(1)}$ and when $l_s = l_s^{(2)}$ since they depend on the neighborhood geometry which varies with l_s. Moreover, due to the constraint of constant neighborhood cardinal, when s neighborhood is changed, the neighborhoods of some other pixels are also changed: The pixel s previous neighbors, that have lost one neighbor (namely s), have to find another pixel neighbor replacing it, and the new s neighbors, that have gain one neighbor (s), have to get rid of another pixel neighbor, and so on.

To reprocess one convenient mathematical framework, we assume the existence of another random field H that corresponds to the definition of the nonstationary neighborhood in every pixel. Such an approach is similar to the line process one [1], or, more recently, and in a more general form, to the triplet Markov field approach introduced by Pieczynski [26], [27]. In our case, the values of the restriction of H to a pixel s, denoted H_s, are the sets of s "active" neighbor pixels. Its cardinal is $|N(s)| = N_n$, except at the image borders where it is equal to $1 + |N(s)|/2$, and at the image corners where it is equal to $1 + |N(s)|/4$. Each H_s value verifies the connection condition (first criterion to construct the pixel neighborhood), and, when s varies, H_s values are not independent: They are linked by the reciprocity constraint.

We assume that the couple (L, H) is a Markov field. Actually, H is a Markov field since, due to the reciprocity of the connex nonstationary-form neighborhood

$$P(H_s = h_s | H_{-s} = h_{-s}) = P(H_s = h_s | H_t = h_t, t \in W_{N_n}(s)) \tag{8}$$

where H_s is H in s, h_s one realization of H_s, H_{-s} is H except in s, h_{-s} one realization of H_{-s}, and $W_{N_n}(s)$ is a neighborhood of size $2N_n + 1$ lines and $2N_n + 1$ columns around s

(fixed-square neighborhood), and N_n is the size of the nonstationary neighborhood (e.g., 4 or 8). We remember that $|N(s)|$, is assumed to be constant (equal to N_n), i.e., only the assumption on the neighborhood shape is relaxed, but not those on its cardinality. Then, assuming L is a Markov field (as usually), the couple (L, H) is also a Markov field over the W_{N_n} neighborhood system. We define the cliques set C over W_{N_n} neighborhood system as in (9), shown at the bottom of the page, where $\beta > 0$, $\zeta > 0$, and $\zeta \gg \beta$.

Note that with such a definition, all active neighborhood solutions even those not verifying the reciprocity condition are possible. Then, classically

$$\begin{aligned} &P(H_s = h_s, L_s = l_s | H_{-s} = h_{-s}, L_{-s} = l_{-s}) \\ &= \frac{exp\left\{-\sum_{c\in C} V_c(h_t, l_t, \forall t \in c)\right\}}{\sum_{h_s=\omega, l_s=i} exp\left\{-\sum_{c\in C} V_c(h_t, l_t, \forall t \in c)\right\}} \\ &= \frac{exp\left\{-\sum_{c\in C: s\in c} V_c(h_s, l_s, h_t, l_t, \forall t \in c, t \neq s)\right\}}{\sum_{h_s=\omega, l_s=i} exp\left\{-\sum_{c\in C: s\in c} V_c(\omega, i, h_t, l_t, \forall t \in c, t \neq s)\right\}} \end{aligned} \quad (10)$$

after simplification of $exp\{-\sum_{c\in C: s\notin c} V_c(\omega, i, h_t, l_t, \forall t \in c, t \neq s)\}$ that was present at the numerator and at the denominator of the fraction. In the following, we denote $U_s(h_s, l_s)$ the local energy "attached to" pixel s : $U_s(h_s, l_s) = \sum_{c\in C: s\in c} V_c(h_s, l_s, h_t, l_t, \forall t \in c, t \neq s) = \sum_{t\in W_{N_n}(s)} V_c(h_s, l_s, h_t, l_t)$. Then

$$\begin{aligned} &\frac{P\left(H_s = h_s^{(1)}, L_s = l_s^{(1)} | H_{-s} = h_{-s}, L_{-s} = l_{-s}\right)}{P\left(H_s = h_s^{(2)}, L_s = l_s^{(2)} | H_{-s} = h_{-s}, L_{-s} = l_{-s}\right)} \\ &\quad = exp\left\{-\left[U_s\left(h_s^{(1)}, l_s^{(1)}\right) - U_s\left(h_s^{(2)}, l_s^{(2)}\right)\right]\right\}. \end{aligned} \quad (11)$$

Now, comparing two configurations only differing by h_s, if, for one of these two configurations, the reciprocity condition is not verified for the active neighborhood (h), the probability of this latter will be very low once $\zeta \gg \beta$. Here, because of the size of the solution space, simulated annealing, that investigates any solution even very improbable, fails, and a classical heuristic such as the ICM would also fail, due to the high number of local configurations. Therefore, alternatively, we propose to define a meta-heuristic that focuses on the solutions verifying the h reciprocity and that converges to a favorable local minimum.

Now, the problem is to construct the neighborhood with the constraint that each pixel, except for on the image border, has the same number of neighbors, and such that the symmetry property (3) is satisfied. For example, in 8-connectivity, eight neighbors have to be selected, each of them been located in a range of $[-1, +1]$ lines and the column of another neighbor. The order of selection of the neighbors is without importance. Therefore, during the neighborhood construction, the following neighbor can be selected among any of the pixels located in a range of $[-1, +1]$ lines and the column of an already selected neighbor. As illustrated in Fig. 2, for the choice of the first neighbor, eight locations are possible, for the second neighbor, the number of possible locations ranges between ten and 12, depending on the first neighbor position, for the third neighbor, the possible location number ranges between 12 and 16 depending on the relative positions of the previous neighbors, and so on. Table I gives the number of possible configurations of connected neighborhoods with cardinal N_n. Besides, when a pixel changes its label, generally it will also completely change its neighborhood (e.g., consider a pixel s at the border between two regions having different labels. Depending on its own label l_s, s will take preferably its neighborhood in one or the other of the two regions). Therefore, to base neighborhood estimation on the current one is not adequate. Moreover once a pixel has constructed its "active" neighborhood (h_s), the "active" neighborhoods of some other pixels (those belonging to h_s) are already partially constructed, due to the symmetry property (3): Subsequent neighborhood constructions should take into account already achieved neighborhood constructions. Therefore, if the pixel "active" neighborhoods are constructed sequentially, the obtained result depends

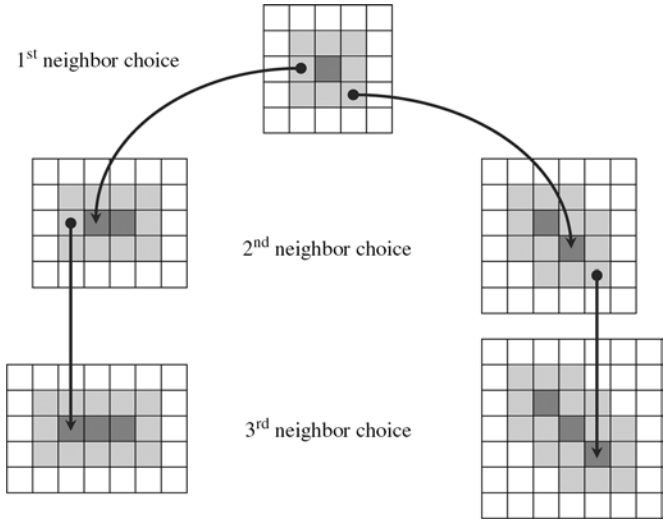

Fig. 2. Possible locations for the choice of the first, second, and third neighbor. Dark gray cells indicate the already selected neighbors, and light gray cells indicate the possible neighbors.

$$\forall (s,t) \in \Omega^2, V_c(h_s, h_t, l_s, l_t) = \begin{cases} 0, & \text{if } (s \notin h_t \text{ and } t \notin h_s) \forall (l_s, l_t) \\ -\beta/2, & \text{if } (s \in h_t \text{ and } t \in h_s \text{ and } l_s = l_t) \\ +\beta/2, & \text{if } (s \in h_t \text{ and } t \in h_s \text{ and } l_s \neq l_t) \\ \zeta, & \text{if } (s \in h_t \text{ and } t \notin h_s) \text{ or } (s \notin h_t \text{ and } t \in h_s) \end{cases} \quad (9)$$

TABLE I
NUMBER OF POSSIBLE CONFIGURATIONS OF CONNEX NEIGHBORHOODS WITH CARDINAL N_n

N_n	2	3	4	5	6	7	8
4-connexity	18	76	315	1296	5320	21800	89190
8-connexity	60	440	3190	22992	165144	1183528	/

on the order of selection of the pixels and provides no guaranty of optimality. Since the number of W_{N_n} neighborhood solutions is too large (even restricted to those satisfying the reciprocity condition) to allow a practical comparison of each of them, a meta-heuristic has to be defined to find the "optimal" W_{N_n} neighborhood configuration: i) knowing the image observation, i.e., X realization, and ii) assuming the image label, i.e., L realization.

Among the meta-heuristics, we focus on the ACO. This choice is motivated by the two following analogies: the first one between the research of an optimal path (between nest and food) by ants and the research of an optimal set of connected pixels from a given "originate" pixel, and the second one between the environment modification through pheromone deposit and the symmetry property (3). Indeed, in ACO the environment is used as a medium to register any already achieved (partial or complete) solution and take it into account in the research of some other solutions. In our case, it allows us to model the fact that the construction of a pixel "active" neighborhood should be guided by the previous "active" neighborhood constructions (either of this pixel or of some other pixels). Moreover, as we will see in next section, it allows introducing a competition between W_{N_n} neighborhood constructions: The pixel "active" neighborhoods are researched in parallel, and the "better" ones are the first ones modifying the environment. There is no mathematical proof of the optimality of the obtained solution, but the performance of the ACO approach to a large number of various problems.

III. ACO METHOD

ACO is modeled on the problem-solving ability of social insects such as ants. The nature-inspired basis of the technique is as follows. While searching for food, ants deposit trails of a chemical substance called pheromones to which other ants are attracted. As shorter paths to food will be traversed more quickly, they have a better chance of being sought out and reinforced by other ants before the volatile pheromones evaporate. Other ants then follow suit, and so on. Thus, using pheromones and random search procedures, the colony is able to find the shortest paths to food rapidly.

More conceptually, the problem of the ant colony is the following: Given a function to minimize (the delay between the ant's nest and food), different solutions are examined (ant exploration), and each of them is memorized (thanks to the pheromone deposit) depending on its quality, and guides the research of the next solutions until convergence.

A. Application to Neighborhood Construction

In this study, we use the ants to find some solutions over the over the W_{N_n} neighborhood system satisfying the symmetry property (3), i.e., without cliques having potential value equal to ζ. Considering a pixel s, we aim at comparing two configurations only differing by the potential values of the cliques included in $W_{N_n}(s)$, the $(2N_n+1)\times(2N_n+1)$ sized neighborhood around s. Indeed, it allows us to use a similar computation as given by (11), simply replacing the "local energy attached to pixel s," by a "local energy attached to $W_{N_n}(s)$." Practically, this means that, for any pixel t of $W_{N_n}(s)$ having n_{out} "active" neighbor(s) outside of $W_{N_n}(s)(n_{out}\in[0,N_n])$, these n_{out} neighbors are fixed, and only the other $(N_n - n_{out} = n_{in})$ "active" neighbors are "free" and can be changed with other possible neighbors belonging to $W_{N_n}(s)$. This is the choice of these "free" neighbors that will be optimized by the ants.

Denoting $\delta(.,.)$ the Kroenecker function: $\delta(i,j)=1$ if $i=j$, $\delta(i,j)=0$ otherwise, $|\{r\in N(s); l_r = l_s\}|$ writes $\sum_{r\in h_s}\delta(l_s,l_r)$. Then, using (3) (i.e., considering only h configurations verifying the reciprocity condition), (7) can be written in a more convenient way to guide the construction of the neighborhoods

$$E=\sum_{s\in\Omega}\left[\sum_{r\in h_s}\left[\frac{\beta}{2}\times(1-\delta(l_r,l_s)) + \frac{1}{N_n}\left\{\frac{(x_r-\mu_{l_r})^2}{2.\sigma_{l_r}^2}+ln(\sigma_{l_r})\right\}\right]\right] \quad (12)$$

and local energy relative to $W_{N_n}(s)$ is similarly computed

$$E\left(W_{N_n}(s)\right)=\sum_{t\in W_{N_n}(s)}\left[\sum_{r\in h_t\cap W_{N_n}(s)}\left[\frac{\beta}{2}\times(1-\delta(l_r,l_t)) + \frac{1}{N_n}\left\{\frac{(x_r-\mu_{l_r})^2}{2.\sigma_{l_r}^2}+ln(\sigma_{l_r})\right\}\right]\right]. \quad (13)$$

In (13), $h_t\cap W_{N_n}(s)$ is the set of t free neighbors. For each pixel s, the active neighborhoods h_t are constructed so that they minimize the sum within the brackets and satisfy the symmetry property. To achieve this construction, ants are used as follows.

The pixels t emit ants. The ants gather the information about neighbor label, along some paths of connected pixels including t, the emitting pixel. During their paths, or neighborhood solution constructions, the ants select the following neighbor from "routing indicators." These latter are based on previously deposed pheromones (either by ants emitted by t or by ants emitted

by other pixels having chosen t as neighbor) and the "energy hop" $E_{t\to r}$ defined as

$$E_{t\to r} = \left[\beta \times (1-\delta(l_r,l_t)) + \frac{1}{N_n}\left\{\frac{(x_t-\mu_{l_t})^2}{2.\sigma_{l_t}^2} + ln\,(\sigma_{l_t}) + \frac{(x_r-\mu_{l_r})^2}{2.\sigma_{l_r}^2} + ln\,(\sigma_{l_r})\right\}\right]. \quad (14)$$

Equation (14) shows that the following neighbor r is chosen considering its label l_r (relative to the emitting pixel label l_t) and also its "data attach" energy, which gives a hint about the l_r confidence. According to the ACO procedure, the next neighbor is chosen either randomly according to the probability of random exploration (necessary once stepping over a "bad" neighbor is necessary to reach "good" neighbors.), or minimizing a function of $E_{t\to r}$ and pheromone deposition. Practically, to simulate the pheromone deposition we define, for each pixel s a "neighborhood" matrix of size $(2N_n+1)(2N_n+1)$ representing all the possible neighbor pixel locations from s. The "neighborhood" matrix values are real, with the matrix norm equal to 1, i.e., this matrix is somewhat a fuzzy representation of the neighbor feature of the pixels around s. Denoting $[\mathbf{N}_\mathrm{s}](r)$ the value of the neighborhood matrix of s in $r(r\in W_{N_n}(s), [\mathbf{N}_\mathrm{s}](r) = [\mathbf{N}_\mathrm{r}](s))$, we define the cost of the choice of r as t following neighbor as $c_{t\to r} = \beta(1-[\mathbf{N}_t](r)) + E_{t\to r}$. Note that this weighting between pheromone deposition and energy hop terms is empirical.

Arriving at a selected neighbor pixel, an ant waits a time proportional to the cost $c_{t\to r}$, before selecting the next neighbor. Each ant has to find a number of neighbors equal to the number of "free active neighbors" of its emitting pixel. When an ant achieves this number, it stops and returns.

On its return, the ant deposits pheromones on the visited neighbor pixels. Practically, for each pixel r visited by an ant emitted by pixel t, $[\mathbf{N}_\mathrm{t}](r)$ is increased by the quantity of pheromone deposit q. Due to the reciprocity condition (3), pheromones are also deposited on the "neighborhood" matrices of the r selected neighbors, on the pixel corresponding to t in these "neighborhood" matrices, $[\mathbf{N}_\mathrm{r}](t)$.

Due to the waiting time proportional to the cost $c_{t\to r}$, pixels on "good" paths are visited frequently by ants, thus increasing neighborhood matrix values for pixels contained in those paths and diminishing those of the other pixels. Ants on poorer paths arrive later, and have only short-lived effect on the estimation of the neighborhood matrices.

B. Global Classification Algorithm

The ant algorithm works in the following way.

1) Initializations: data image (X realization) and class features loading, blind classification, stop criterion definition, initialization of the *a priori* parameters: pheromone deposit quantity q, probability of random exploration p_e, experience duration T_e, and initialization of the "neighborhood matrix" consisting of a set of possible neighbors (located in a range of $\pm N_n$ in lines and $\pm N_n$ in columns) with an isotropic neighborhood.
2) As long as the stop criterion is not verified.
   2.1. For each pixel s:
      2.1.1. for each pixel t of $W_{N_n}(s)$, computation of the number $n_{in}(t)$ of active neighbors located inside $W_{N_n}(s) : \forall t \in W_{N_n}(s), n_{in}(t) = |h_t^{(1)} \cap W_{N_n}(s)|$, and setting $\forall t \in W_{N_n}(s), h_{[s]}^{(1)}(t) = \{h_t^{(1)} \cap W_{N_n}(s)\}$
      2.1.2. Computation of the current $W_{N_n}(s)$ local energy term according to the current neighborhood matrices of t pixels included in $W_{N_n}(s)$ and their current labels

$$E_1 = \sum_{t\in W_{N_n}(s)}\left[\sum_{r\in h_{[s]}^{(1)}(t)}\left[\frac{\beta}{2}\times\left(1-\delta\left(l_r^{(1)},l_t^{(1)}\right)\right) + \frac{1}{N_n}\left\{\frac{\left(x_r-\mu_{l_r^{(1)}}\right)^2}{2.\sigma^2_{l_r^{(1)}}} + ln\left(\sigma_{l_r^{(1)}}\right)\right\}\right]\right]. \quad (15)$$

      2.1.3. For each new label to test $l_s^{(2)}$.
      2.1.4. Optimization of the $W_{N_n}(s)$ active neighborhood: $W_{N_n}(s)$ pixels t emit simultaneously several ants. Each ant stores the label of its emitting pixel $l_t^{(2)}(\forall t \neq s, l_t^{(2)} = l_t^{(1)})$.
         2.1.4.1. Each ant advances to subsequent pixels r either minimizing the cost function $c_{t\to r}$, or with a small "exploration probability" p_e, by picking a next neighbor pixel randomly. On the way, at each visited pixel, it waits a time proportional to the cost value corresponding to the chosen neighbor.
         2.1.4.2. After $n_{in}(t)$ hops, an ant emitted by pixel t has constructed an active neighborhood solution (or its restriction to $W_{N_n}(s)$) denoted $h_{[s]}^{(2)}(t)$. Then, it follows the identical path in the reverse direction. At each pixel r on the return path, i.e., $r \in h_{[s]}^{(2)}(t)$, the neighborhood matrix value for the ant source pixel t is updated, as are the neighborhood matrix values for the ant-visited pixels r

$$\forall r \in h_{[s]}^{(2)}(t), \quad \begin{cases} [\mathbf{N}_t](r) = \frac{1}{K_t}\times([\mathbf{N}_t](r)+q) \\ [\mathbf{N}_r](t) = \frac{1}{K_r}\times([\mathbf{N}_r](t)+q) \\ \forall r' \in [\mathbf{N}_r]\cap W_{N_n}(s), \\ [\mathbf{N}_r](r') = \frac{1}{K_r}\times[\mathbf{N}_r](r') \end{cases}$$

$$\forall r' \in [\mathbf{N}_t]\cap W_{N_n}(s), \quad [\mathbf{N}_t](r') = \frac{1}{K_t}\times[\mathbf{N}_t](r') \quad (16)$$

         where $[\mathbf{N}_\mathrm{s}](r)$ its r^{th} element of s neighborhood matrix $[\mathbf{N}_\mathrm{s}]$, and K_s is a normalization factor such that the sum of $[\mathbf{N}_\mathrm{s}]$ elements be equal to 1. q is an *a priori* parameter representing the quantity of pheromone deposit.
         2.1.4.3. If the "experience duration" T_e is not expired, each time an ant is back to its source pixel t, t generates a new ant, and go to step 2.1.4.1.

2.1.5. When the experience has finished, computation of the $W_{N_n}(s)$ local energy corresponding to the tested label $l_s^{(2)}$ and estimated pixel active neighborhood

$$E_2 = \sum_{t \in W_{N_n}(s)} \left[\sum_{r \in h^{(2)}_{[s]}(t)} \left[\frac{\beta}{2} \times \left(1 - \delta\left(l_r^{(2)}, l_t^{(2)}\right)\right) + \frac{1}{N_n} \left\{ \frac{\left(x_r - \mu_{l_r^{(2)}}\right)^2}{2.\sigma^2_{l_r^{(2)}}} + ln\left(\sigma_{l_r^{(2)}}\right) \right\} \right] \right]. \quad (17)$$

For notation convenience, we subscript (2) all labels in (17) but note that $\forall t \neq s, l_t^{(2)} = l_t^{(1)}$.

2.1.6. Decision to change the s label from $l_s^{(1)}$ to $l_s^{(2)}$ if $\Delta E = E_2 - E_1$ is negative.

2.2. New estimation of the parameter "experience duration," equal to five times the average duration of an ant path.

In step 2.1.3, the tested labels $l_s^{(2)}$ can be chosen either randomly or systematically ranging from 1 to c (the number of classes).

With this algorithm, the convergence is ensured by the fact that the global energy is decreased at each step. Indeed, for a configuration change restricted to s label and active neighbors r of t pixels, such that t and r pixels belong to $W_{N_n}(s)$, the global energy difference induced is equal to $W_{N_n}(s)$ local energy difference. Therefore, accepting only configuration changes corresponding to $W_{N_n}(s)$ energy decreases ensures the global energy decrease, just as for the iterative conditional modes (ICM) [28]. However, as for the ICM, the achieved result is only a local minimum depending on the initialization. To overcome this problem and obtain a global minimum, step 2.1.5 could be stochastic and some changes may be accepted even if ΔE be positive, just as for the classical simulated annealing application to MAP classification. Now, on our images the algorithm is already rather long; therefore, we prefer to keep step 2.1.5 determinist.

There are two kinds of algorithm parameters: the image model parameters and the ACO algorithm ones.

The image model parameters are:

- the class features $\mu_i, \sigma_i, i \in \{1, \ldots, c\}$;
- the neighborhood size N_n;
- the neighborhood weight β.

The ACO algorithm parameters are:

- the exploration probability p_e;
- the quantity of pheromone deposit q;
- the experience duration T_e.

The class features are assumed to be known: They have been estimated previously either from learning areas, model simulations, or a clustering algorithm. Different values of neighborhood size have been tested, ranging from to 4 to 12. We find the algorithm is not very sensitive to the exact value; in our case, good results were obtained for N_n about equal to 8 (which is the same value as in classical MRF considering the neighborhood surrounding the pixels). Finally, β value has been estimated empirically, just as for the classical MAP classification.

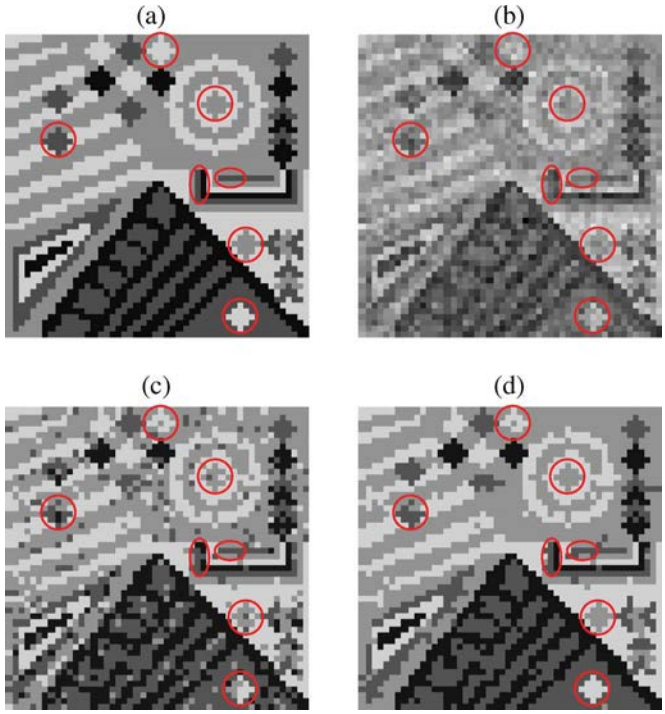

Fig. 3. Simulated data (noise $\sigma = 40$): (a) label image, (b) data image with Gaussian white noise, and obtained classifications, (c) blind result, and (d) ACO result.

The exploration probability, the quantity of pheromone deposit and the experience duration should allow the emergence of a stable ant solution. For this purpose, they will not be chosen independently: The lower the exploration probability, the higher the pheromone quantity, the lower the experience duration. In our case, the experience duration is deduced from the mean waiting time (of the ants in visited pixels) $\bar{t}_{\text{wait}}$ as: $T_e = 5.N_n.\bar{t}_{\text{wait}}$, where the factor 5 has been chosen empirically, and is not very sensitive. The pheromone deposit quantity is 0.4, and the exploration probability 0.04. These empirical choices have been done among several tested values. However, the obtained results were not very sensitive provided that the decrease of p_e and q are sufficiently slow, and T_e sufficiently large.

IV. Results

A. Method Validation

First of all, we validate the proposed algorithm on simulated data. Fig. 3(a) shows the label image. There are four classes, centered on $\{100, 200, 300, 400\}$, respectively, and having standard deviation equal to 40 (variance 1600). The pixel value distribution, for a given class, is Gaussian $\mathbf{N}(\mu_i, \sigma_i)$. Fig. 3(b) shows the data image in the case of a Gaussian white noise.

Fig. 3(c) shows the blind classification result. Fig. 3(d) represents the result of the ACO method. The parameters are the following: $|N(s)| = 8$, $\beta = 1.0$, and initial values of $(p_e, q, T)_e)$ are (4%,40%,100). For comparison (in the following), the β value has been chosen equal to those empirically optimized for the stationary neighborhood MAP. We note that the blind result is very noisy. Analyzing the ACO result versus this blind

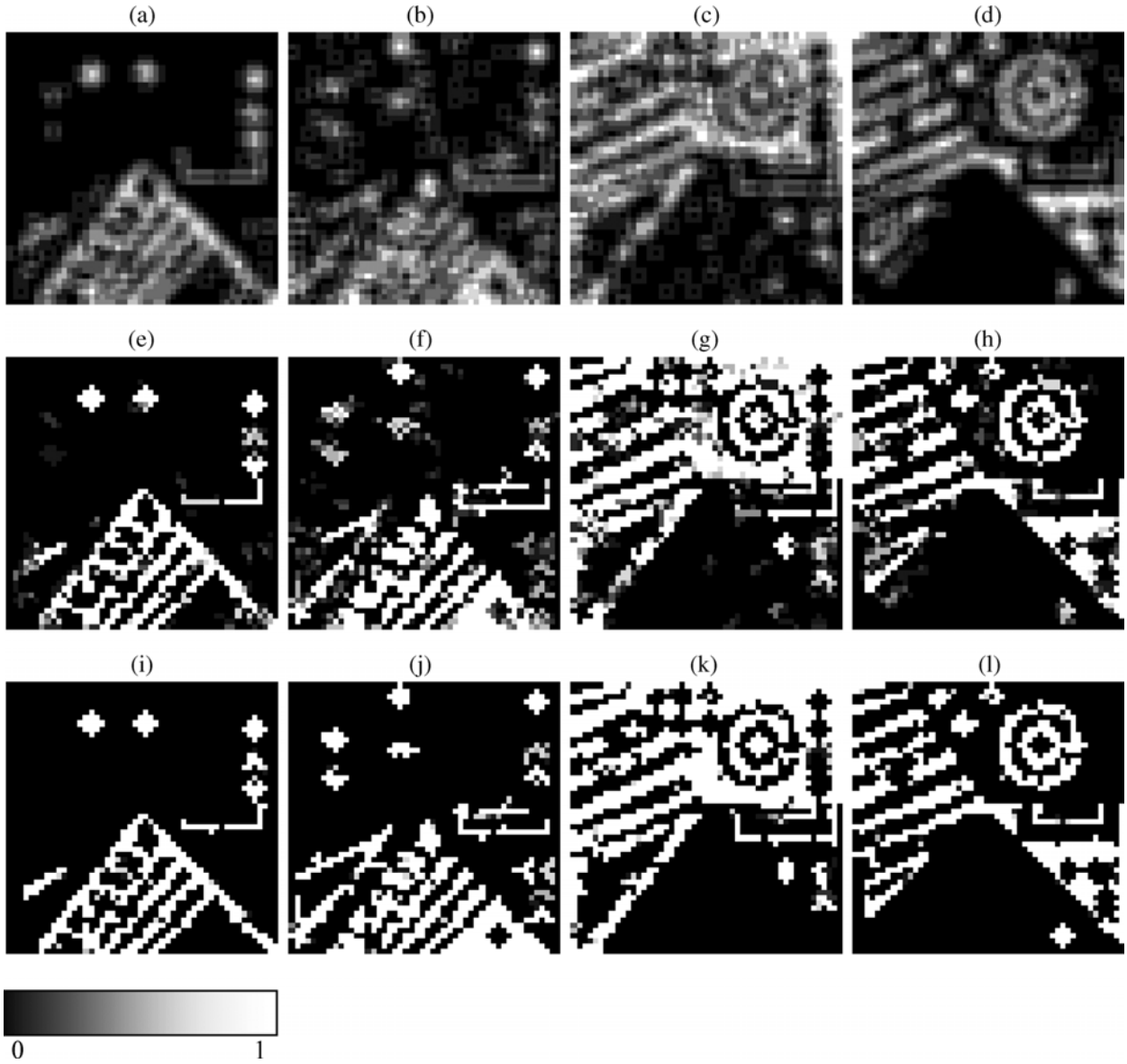

Fig. 4. Normalized neighborhood potential functions of the four classes (one by column) computed over the auto-adaptive neighborhood: first row corresponds to the classical fixed-form neighborhood; second row is after ACO first iteration; and last row shows the result after ACO convergence.

classification (i.e., spatial regularization versus data attach), we note that regularization seems to have occurred at the expected places. Before further analysis and comparison with alternative classification, we check whether the algorithm succeed in the construction of the adaptive neighborhoods.

Fig. 4 shows the normalized neighborhood potential, $|\{r \in N(s); l_r = l_s\}|/|N(s)|$ corresponding to the four classes. We check that the neighborhoods are constructed to be label homogeneous. Before convergence, the potentials are estimated over the fixed-form 8-connectivity square neighborhood (initialization). We see that many pixels have mixed-class neighborhood. After convergence, even at the class border most of the potential values are very close to 1 (white color) or 0 (black color), indicating a homogeneous neighborhood. Indeed, the effect of having a nonstationary auto-adaptive neighborhood is that the pixels having a different label are not included in the neighborhood. Only in some very few cases, the ants fail to construct a completely uniformly-labeled neighborhood, probably because of the relatively high level of noise in the data image. However, we note that the neighborhood function value of the pixel label is always much closer to 1 than the value obtained by classical 8-connectivity.

We now aim at comparing the ACO result with the results obtained using alternative approaches. Fig. 5(a) and (b) show the ICM one and the MAP obtained by simulated annealing, both using 8-connectivity isotropic neighborhood and the Potts model (different values of β have been tested, only the best result is presented). Fig. 5(c) is the result provided by the line process as described in [1]. Fig. 5(d) is the result provided by the Chien model [4], [29]. This model, defined on a 5×5 neighborhood, depends on three parameters which control the local energy of edges, lines, and noise. These two approaches (line process and Chien model) have been selected, for comparison with the proposed method, as much more sophisticated (and

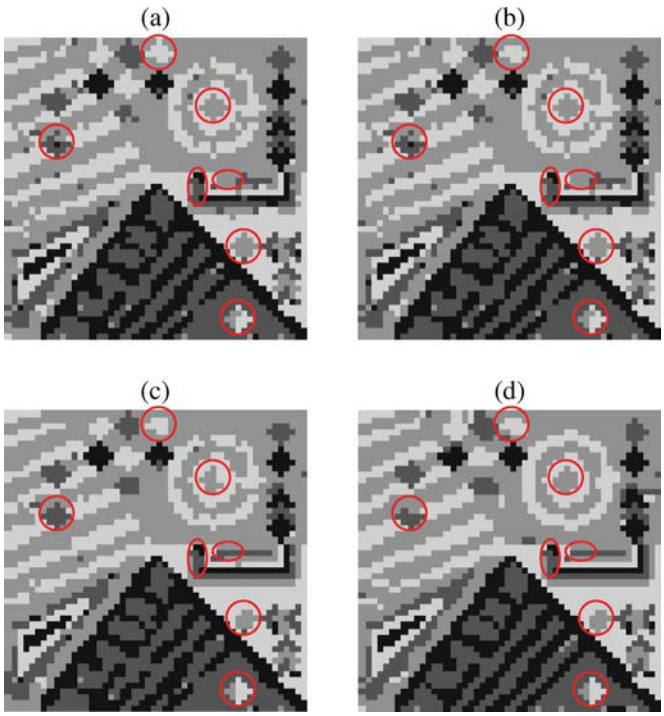

Fig. 5. Simulated data (noise $\sigma = 40$), alternative classification results: (a), (b) best result with 8-connectivity and Potts model respectively using ICM and simulated annealing algorithms, (c) result using line process [1], (d) result using high-order cliques (Chien model, [4]).

complex) than the isotropic neighborhood model. They are *a priori* able to preserve image fine structures while regularizing the configuration. The ACO result was presented in Fig. 3(d).

With the ICM 8-connectivity isotropic neighborhood, most of the "isolated" errors have been corrected relative to the blind classification [presented in Fig. 3(c)]. Packet errors can not be corrected without damaging the fine structures. The simulated annealing produces an almost equivalent result (slightly better thanks to the global minimization instead of the local one). The fact that the ICM and the simulated annealing produce very close results shows that the remaining errors are not due to the optimization process but rather to the image model, and in particular (but not exclusively) to the assumption of stationary neighborhood. This is confirmed by the examination of the other results, where several errors have been corrected either due to the use of a more complex image model (line process, Chien model), or to the relaxation of the stationary neighborhood assumption (ACO). Among the considered approaches, ACO leads to the best result, even if some errors remain due to the high level of noise in the initial image: The fine structures are more or less preserved and most classification errors are corrected. Several examples illustrating the interest of the proposed approach are pointed out in Figs. 3 and 5. For example, one can see that the fine longitudinal structures in the middle right of the image have been lost by the ICM or MAP classical approaches. Preserving it with classical isotropic neighborhood would require to decrease β parameter, but at the expense of the correction of some blind classification errors. Using some more sophisticated approaches, such as the line process, the Chien model, or the adaptive neighborhood, allows preserving these structures. On the other hand, there are some cases, such as those pointed by the circles in Figs. 3 and 5 (except the bottom right one), where the simple classical image model (with two-order potential functions) performs better than the line process or the Chien model, but slightly worse than the adaptive neighborhood one. Finally, the area pointed by the bottom right circle shows a case where only the auto-adaptive neighborhood is able to correct initial blind classification errors. This can be explained by the fact that our model is less constrained than the line process which requires *a priori* on the structure of edges or the Chien model which constrains the orientations of shapes and lines [30]. The proposed model, while avoiding the over-regularization effect of the Potts model, adapts locally the neighborhood and is, therefore, more flexible than more complex priors. Note also that here we focus on the cases where the more interesting approach is the proposed one, but there are some cases where it is one of the alternative methods (generally the line process or the Chien model).

Quantitatively, the performance of a result is measured by the ratio of well-classified pixels, or by the Kappa function, both defined from the confusion matrix by

$$r_{\mathrm{GC}} = \sum_{i=1}^{c} M_{ii} \Big/ \sum_{i=1}^{c}\sum_{j=1}^{c} M_{ij} \tag{18}$$

$$\begin{cases} \mathrm{Kappa} = (r_{\mathrm{GC}} - b)/(1-b) \\ \text{where } b = \left(\sum_{i=1}^{c} M_{ij}\right) \\ \qquad \times \left(\sum_{j=1}^{c} M_{ij}\right) \Big/ \left(\sum_{i=1}^{c}\sum_{j=1}^{c} M_{ij}\right)^2 . \end{cases} \tag{19}$$

Table II shows the performance of different classification results obtained from the data images with different white noise levels: standard deviations of the classes respectively equal to 20, 40, and 60, whereas the class centers remain the same in the three cases. The performance is compared in terms of the percentage of good classification r_{GC} and the Kappa value obtained from blind classification, fixed square form (8-connectivity) neighborhood MAP approach, line process, Chien model and the ACO auto-adaptive 8-connectivity result. We note that the improvement between auto-adaptive neighborhood (thanks to ACO) and fixed-form neighborhood modeling (two-order clique MAP, optimized by simulated annealing) is until few percents. However, we also note that in the case of high noise level, the auto-adaptive neighborhood ACO is less performing than the Chien model. For further understanding, Fig. 6 presents the simulated data images and classification results respectively obtained using the ACO and the Chien model, in the two cases of $\sigma = 20$ noise and $\sigma = 60$ noise. Now, remember that the auto-adaptive neighborhood approach does not use any *a priori* information about neighborhood (except its cardinal), conversely the some approaches privileging some neighborhood form through the definition of the clique potentials. From Fig. 6(d)–(f), we see that, in the case of important noise, due to the local optimization, the less constraining approach may fail to recover the actual solution, conversely to the case of the lower noise [illustrated in Figs. 3(d), 5(d), and 6(b) and (c)].

TABLE II
COMPARISON OF THE PERFORMANCE OF THE DIFFERENT CLASSIFICATION ALGORITHMS IN THREE CASES OF SIMULATED DATA

Classification	blind	Fixed-form 8 connectivity	Line process	Chien-model	Auto-adaptive form 8 connectivity
$\sigma = 20$	$r_{GC} = 98.28$ Kappa = 97.63	$r_{GC} = 99.40$ Kappa = 99.17	$r_{GC} = 98.72$ Kappa = 98.23	$r_{GC} = 99.04$ Kappa = 98.67	$r_{GC} = 99.76$ Kappa = 99.67
$\sigma = 40$	$r_{GC} = 85.00$ Kappa = 79.40	$r_{GC} = 91.36$ Kappa = 88.06	$r_{GC} = 92.32$ Kappa = 89.38	$r_{GC} = 92.56$ Kappa = 89.74	$r_{GC} = 93.60$ Kappa = 91.16
$\sigma = 60$	$r_{GC} = 67.44$ Kappa = 55.70	$r_{GC} = 77.80$ Kappa = 69.40	$r_{GC} = 80.00$ Kappa = 72.49	$r_{GC} = 82.40$ Kappa = 75.91	$r_{GC} = 80.40$ Kappa = 73.15

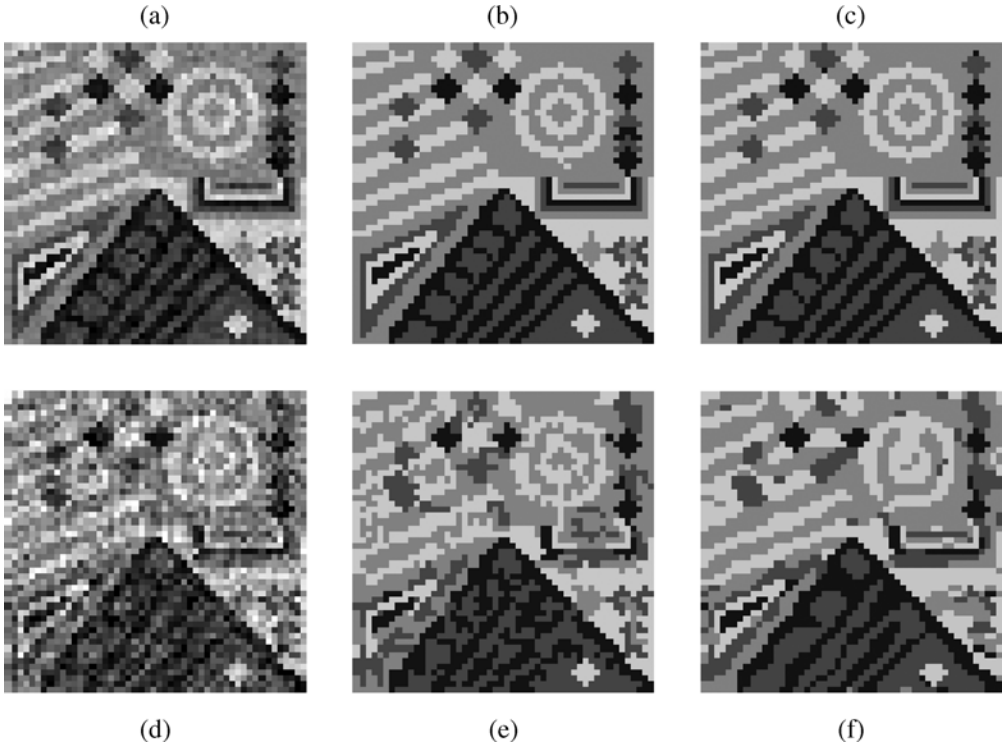

Fig. 6. Simulated data: (a) data image with Gaussian white noise ($\sigma = 20$); (b) ACO classification result; (c) Chien model classification result; (d) data image with Gaussian white noise ($\sigma = 60$); (e) ACO classification result; (f) Chien model classification result.

We point out that the global improvement in the results corresponding to the proposed approach is due to the use of nonstationary neighborhood rather than to the ACO. Indeed, ACO used with a classical fixed-form MRF model would lead to results similar to simulated annealing ones [21]. Here, ACO is used as one way to derive the nonstationary neighborhood solution. In particular, as an alternative to the ACO, an optimization using a Gibbs sampler with simulated annealing was tried to retrieve the couple $(l, h) = \{(l_s, h_s), s \in \Omega\}$ (defined in Section II-B). However, the energy landscape on the solution space is too much irregular to achieve satisfying results: In particular, changing only one pixel neighborhood configuration h_s at each step makes impossible a direct transition between two configurations checking the reciprocity condition (there is at least one intermediate nonreciprocal neighborhood configuration, i.e., with high energy).

B. Application to Actual Data

In this section, data were acquired by the SPOT4/HRVIR sensor, which performs measurements in the three following channels: green, red, and near infrared, and has a pixel size equal to 20×20 m^2. For the study of vegetation areas (e.g., forest, agricultural areas), vegetation indices are estimated from the radiometric measurements: either empirical vegetation indices defined from simple channel combinations [31], such as the normalized difference vegetation index (NDVI), or vegetation indices obtained from physical radiative transfer modeling, requiring the knowledge of acquisition conditions, such as the vegetation fraction cover (FCOV).

The considered data were acquired over an agricultural test site in Romania in June 2001 (ADAM database). From the whole image, a subset of 256×256 pixels was extracted, and

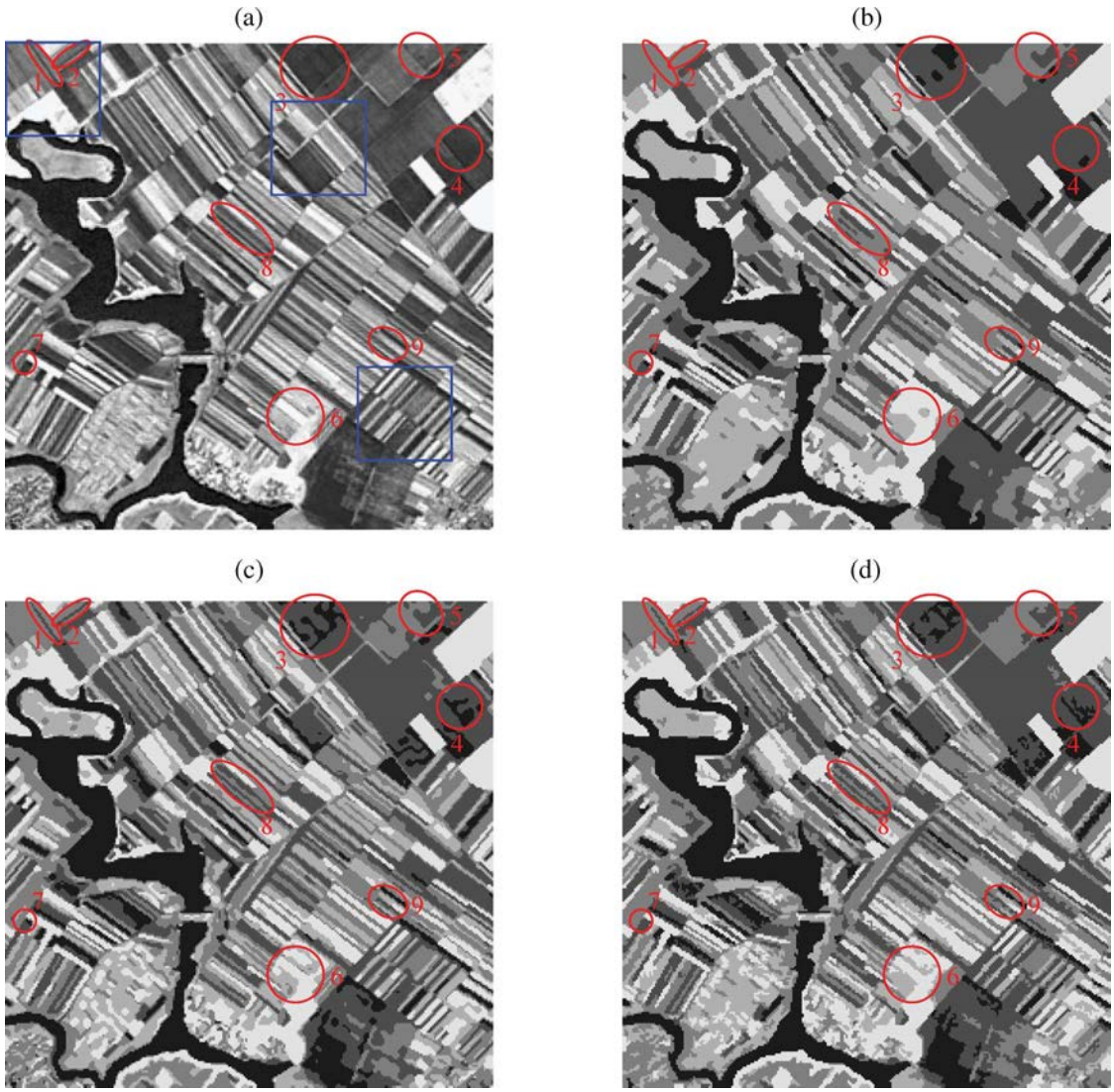

Fig. 7. SPOT4 data and results: (a) vegetation fraction cover image; (b) blind classification result; (c) classification result using the classical isotropic 8-connectivity neighborhood; (d) classification result using the self-organized neighborhood.

the FCOV image was calculated [32]. It is shown in Fig. 7(a). Five main classes of vegetation densities can be distinguished: from null vegetation cover (water or bare soil), small vegetation cover (corresponding to winter cereal fields almost all at stubble state at the end of June), to maximum vegetation cover (corresponding to fully developed crops at the end of June, e.g., beans, or grassland areas). Table III shows the features of these five classes estimated by fuzzy c-means [33], and their color code for classification. We note the great complexity of the landscape, and, in particular, the thinness of most of the fields (excepting winter cereal ones) as a result of both agricultural practices and sensor pixel size. In addition, this image presents two main issues for classification regularization. First, some large fields may present heterogeneous stages of vegetation (e.g., from clear bare soil to stubble), such as those pointed by circled areas 3–6 in Fig. 7(a). These field subparts have generally a geometric form that is based on soil geometry (namely field borders) and soil properties (namely moisture, roughness). For classification, one may expect that they will either be removed by regularization (when their corresponding vegetation stages are not very different from the local background ones), or preserved but retaining their highly structured (i.e., not smoothed) boundaries so that they can still be identified as heterogeneous subareas rather than well delimited fields. Second, due to the pixel size (20×20 m^2), the image

TABLE III
CLASS FEATURES FOR THE SPOT4 VEGETATION FRACTION COVER IMAGE

Class number	center	variance	color
1	0.563	1.0	
2	0.608	1.0	
3	0.0443	1.0	
4	0.172	1.0	
5	0.316	1.0	

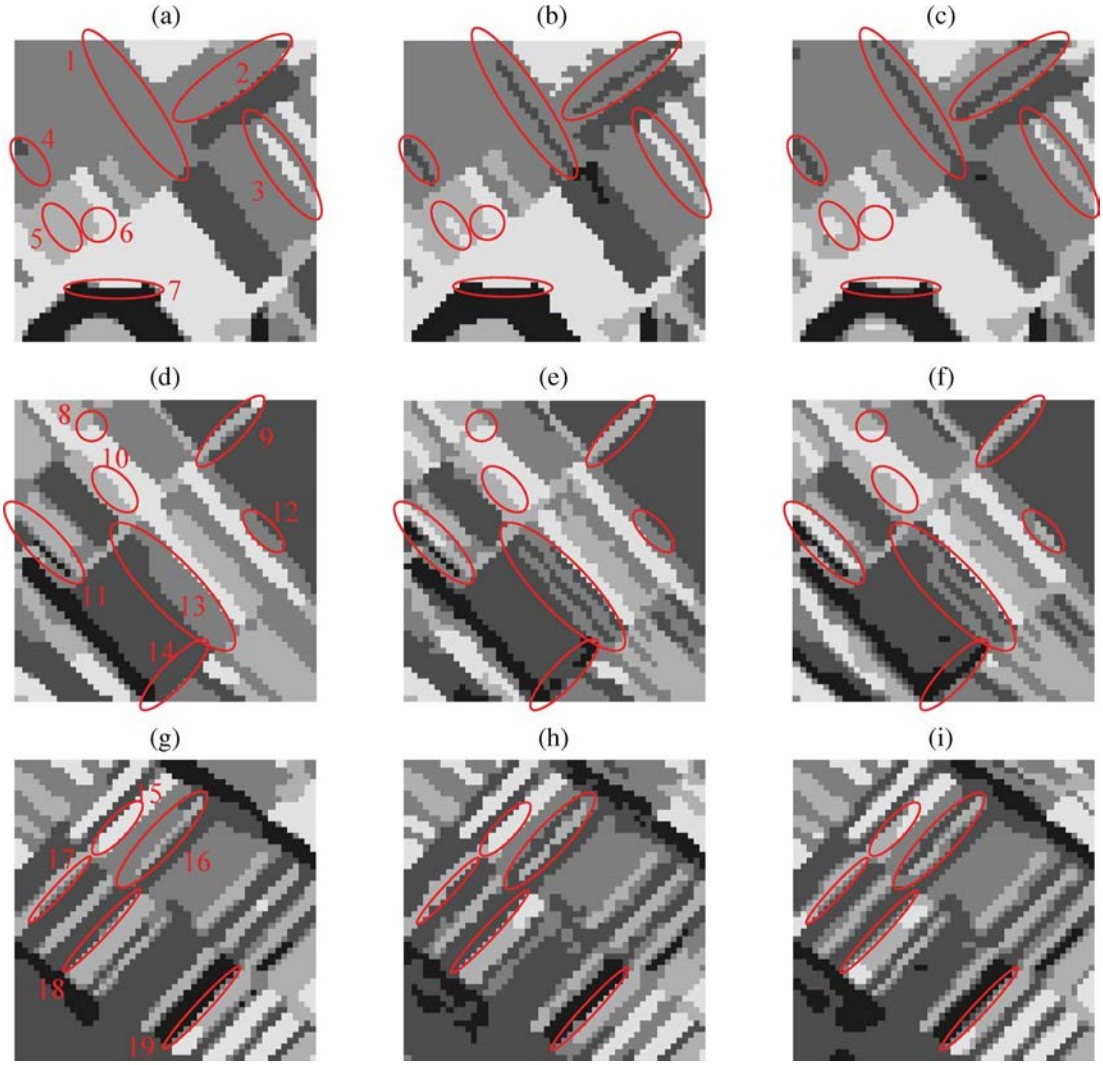

Fig. 8. Comparison of the classification results: respectively, from left to right, isotropic neighborhood (a), (d), (g), adaptive neighborhood (b), (e), (h), and Chien model (c), (f), (i), in the case of three 50 × 50 pixel subparts of Fig. 7 (pointed by blue squares): upper left corner, upper middle right, and bottom right.

contains some mixed pixels, in particular corresponding to field borders. Considering hard classification, one may expect that these mixed pixels be classified either in one of the classes composing the mixture (and not in an intermediary class, e.g., low vegetation density class for a pixel straddling a dense vegetation and a bare soil field). We will now see how the different classification methods manage with these issues.

Fig. 7(b)–(d) shows the classification results. The comparison is only qualitative due to the lack of sufficiently accurate ground truth data. The blind classification result (not shown) is rather "noisy." This effect can be much reduced taking into account neighborhood. Fig. 7(b) shows the classification results using fixed-form 8-connectivity square neighborhood, Fig. 7(c) shows the Chien model result, and Fig. 7(d) shows the self-organized (auto-adaptive) neighborhood classification. In Fig. 7(b), we note the effect of the isotropic form of the neighborhood, namely a removal of the fine structures, either completely (as shown by circled areas 1 and 2) or partially (as shown by circled areas 8 and 9). However, this effect cannot be alleviated (by decreasing the β parameter) without seriously damaging the reduction of blind result noise. Indeed, Fig. 7(b) is the best compromise we obtain between noise regularization and actual complexity preservation. Considering either the Chien model classification result or the adaptive neighborhood one, we note that the fields corresponding to fine structures are restituted. Now, with the Chien model result (like with the isotropic neighborhood one), we note an unrealistic smoothing of the boundaries for the heterogeneous vegetation stage subareas (main issue 1 shown by circled areas 3 to 6). Both approaches have built up some erroneous "blob" structures, that is not the case using the adaptive neighborhood approach. Finally, we note that there are some cases where the blind classification errors are removed only in Fig. 7(d) (ACO result), as shown by circled area 7 or in Fig. 8.

Fig. 8 allows a more precise comparison of the classification results (respectively isotropic neighborhood, adaptive neighborhood and Chien model, from left to right) in the case of three subparts of the previous 256 × 256 considered image. The circled areas focus on some areas illustrating the interest of the adaptive neighborhood:

- better preservation of fine structures relative to the isotropic neighborhood model (circled areas 1, 2, 4, 5, 11, 13, 14, 16);
- better restitution of the actual geometric form of the structures relative to the Chien model (circled areas 6, 8, 10, 13, 14, 15);
- better removal of blind classification errors (3, 12) in particular in the presence of mixed pixels (7, 9, 16, 17, 18, 19).

To be fair, we also admit that there are some areas where the Chien model leads to a more plausible result, in particular if the problem of mixed pixel misclassification is withdrawn because the mixture is between classes contiguous in the feature space. Indeed, for longitudinal fields the used constraint of shapes and lines can be helpful to retrieve the actual field geometry.

V. Conclusion

The ACO applied to the construction of pixel neighborhoods in global classification problems has been shown to result in near optimal neighborhood functions and to yield a performance superior to that of classical fixed-form neighborhoods in the case of simulated images with different noise levels. This optimization scheme allows us to consider a nonstationary MRF with respect to the neighborhood and provides a numerical tool to estimate locally the neighborhood.

The advantage of having a neighborhood shape which automatically adapts to the image segment clearly appears in the case of images containing fine elements, as shown in the simulated example. Classifications using a classical isotropic 8-connectivity neighborhood fail to recover the true label image, regardless of the optimization algorithm: iterative conditional mode, or simulated annealing, since the limits are not due to the optimization but to the image modeling. More sophisticated image models such as the Chien model or the line process improves somewhat the results. Finally, auto-adaptive neighborhood model seems the more flexible and the more suitable to handle image complexity. Similar results have also been obtained on actual data where the superiority of a self-organized neighborhood was demonstrated qualitatively.

Besides its performance, an interest of the method is its robustness to the parameter fitting. In particular, the ACO parameters (exploration probability, quantity of pheromone deposit, and experience duration) can be calibrated to proposed default values. The β parameter can be taken equal to the classical isotropic 8-connectivity neighborhood optimal value. This possibility to use default parameter value or empirical estimation provided by quick algorithms (such as the classical isotropic ICM) is a serious advantage since the proposed algorithm is much slower (due to the additional complexity). Indeed, future work will deal with current approach approximations in order to reduce computation time. One could also investigate recent optimization techniques such as nonparametric belief propagation [34], that has been shown to be efficient for the exploration of very large solution space [such as the neighborhood configurations for all the pixels belonging to $W_{N_n}(s)$], using a Gaussian mixture approximation [35] for the messages.

Acknowledgment

The authors would like to thank CNES for providing the SPOT data and ASTRIUM for providing the vegetation fraction cover data.

References

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[2] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 609–628, Jul. 1990.

[3] S. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[4] X. Descombes, F. Kruggel, and Y. von Cramon, "Spatio-temporal fMRI analysis using Markov random fields," *IEEE Trans. Med. Imag.*, vol. 17, no. 6, pp. 1028–1039, Jun. 1998.

[5] M. Dorigo, V. Maniezzo, and A. Colorni, "The ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 1, pp. 29–41, Feb. 1996.

[6] R. Schoonderwoerd, O. Holland, and J. Bruten, "Ant-like agents for load balancing in telecommunications networks," in *Proc. ACM Agents*, Marina del Rey, CA, 1997, pp. 209–216.

[7] G. di Caro and M. Dorigo, AntNet: "A mobile agents approach to adaptive routing," Tech. Rep. IRIDIA97-12, Univ. Libre de Bruxelles, Bruxelles, Belgium, 1997.

[8] ——, "AntNet: Distributed stigmeric control for communications networks," *J. Artif. Intell. Res.*, vol. 9, pp. 317–365, 1998.

[9] E. Bonabeau, F. Heneaux, S. Guérin, D. Snyers, P. Kuntz, and G. Theraulaz, "routing in telecommunications networks with smart ant-like agents," Working Paper, 98-01-003, Santa Fe Inst., Santa Fe, NM, 1998.

[10] M. Heusse, D. Snyers, S. Guérin, and P. Kuntz, "Adaptive Agent-Driven Routing and Load Balancing in Communication Networks," Tech. Doc. RR-98001-IASC, Ecole Nat. Sup. Télécomm. Bretagne, France, 1998.

[11] E. Sigel, B. Denby, and S. Le Hégarat-Mascle, "Application of ant colony optimization to adaptive routing in a LEO telecommunications satellite network," *Ann. Telecomm.*, vol. 57, pp. 520–539, 2002.

[12] V. Maniezzo, A. Colorni, and M. Dorigo, "The ant system applied to the quadratic assignment problem," Tech. Rep. IRIDIA/94-28, Université Libre de Bruxelles, Bruxelles, Belgium, 1994.

[13] A. Colorni, M. Dorigo, F. Maffioli, V. Maniezzo, G. Righini, and M. Trubian, "Heuristics from nature for hard combinatorial problems," *Int. Trans. Oper. Res.*, vol. 3, pp. 1–21, 1996.

[14] D. Costa and A. Hertz, "Ants can colour graphs," *J. Oper. Res. Soc.*, vol. 48, pp. 295–305, 1997.

[15] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Jan. 1997.

[16] M. Dorigo, G. Di Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," *Artif. Life*, vol. 5, pp. 137–172, 1999.

[17] L. M. Gambardella, E. Taillard, and M. Dorigo, "Ant colonies for the quadratic assignment problem," *J. Oper. Res. Soc.*, vol. 50, pp. 167–176, 1999.

[18] N. Labroche, N. Monmarché, and G. Venturini, "A new clustering algorithm based on the ants chemical recognition system," in *Proc. 15th Eur. Conf. Artificial Intelligence*, Lyon, France, Jul. 2002, pp. 345–349.

[19] C. Fernandes, V. Ramos, and A. Rosa, "Self-regulated artificial ant colonies on digital image habitats," *Int. J. Lateral Comput.*, vol. 2, no. 1, pp. 1–8, 2005.

[20] H. Nezamabadi-Pour, S. Saryazdi, and E. Rashedi, "Edge detection using ant algorithms," *Soft Comput.*, vol. 10, no. 7, pp. 623–628, 2006.

[21] S. Ouadfel and M. Batouche, "MRF-based image segmentation using ant colony system," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 2, pp. 12–24, 2003.

[22] P. Del Moral, "Non linear filtering: Interacting particle solution," *Markov Processes and Related Fields*, vol. 2, no. 4, pp. 555–580, 1996.

[23] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/nongaussian Bayesian state estimation," *Proc. Inst. Elect. Eng. F*, vol. 140, no. 2, pp. 107–113, 1993.
[24] M. Bartlett, *The Statistical Analysis of Spatial Pattern*. New York: Wiley, 1975.
[25] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, vol. 36, pp. 192–236, 1974.
[26] D. Benboudjema and W. Pieczynski, "Unsupervised image segmentation using triplet Markov fields," *Comput. Vis. Image Understand.*, vol. 99, pp. 476–498, 2005.
[27] W. Pieczynski and D. Benboudjema, "Multisensor triplet Markov fields and theory of evidence," *Image Vis. Comput.*, to be published.
[28] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 3, no. 48, pp. 259–302, 1986.
[29] X. Descombes, J. F. Mangin, E. Pechersky, and M. Sigelle, "Fine structure preserving Markov model for image processing," in *Proc. 9th Scand. Conf. Image Analysis*, Uppsala, Sweden, 1995, pp. 156–166.
[30] X. Descombes and E. Pechersky, "Droplet shapes for a class of models in Z2 at zero temperature," *J. Statist. Phys.*, vol. 111, no. 1–2, pp. 129–169, 2003.
[31] G. Rondeaux, "Vegetation monitoring by remote sensing: A review of biophysical indices," *Photo-Interpretation*, vol. 3, pp. 197–216, 1995.
[32] S. Jacquemoud and F. Baret, "PROSPECT: A model of leaf optical properties spectra," *Remote Sens. Environ.*, vol. 34, pp. 75–91, 1990.
[33] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means algorithm," *Comput. Geosci.*, vol. 10, pp. 191–203, 1984.
[34] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, vol. 1, pp. I-605–I-612.
[35] A. T. Ihler, E. B. Sudderth, W. T. Freeman, and A. S. Willsky, "Efficient multiscale sampling from products of Gaussian mixtures," *Neural Inf. Process. Syst.*, 2003.