

# Use of Ant Colony Optimization for Finding Neighborhoods in Image Non-stationary Markov Random Field Classification

Sylvie Le Hégarat-Masclé<sup>1</sup>, Abdelaziz Kallel<sup>2</sup>, and Xavier Descombes<sup>3</sup>

<sup>1</sup> IEF/AXIS, Université de Paris-Sud 91405, Orsay Cedex, France

<sup>2</sup> CETP/IPSL, 10, 12 avenue de l'Europe 78140, Vélizy, France

<sup>3</sup> CNRS/INRIA/UNSA, INRIA, 06902 Sophia Antipolis, Cedex, France

**Abstract.** In global classifications using Markov Random Field (MRF) modeling, the neighborhood form is generally considered as independent of its location in the image. Such an approach may lead to classification errors for pixels located at the segment borders. The solution proposed here consists in relaxing the assumption of fixed-form neighborhood. Here we propose to use the Ant Colony Optimization (ACO) and to exploit its ability of self-organization. Modeling upon the behavior of social insects for computing strategies, the ACO 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 a same image segment. We show that this corresponds to an automatic adaptation of the neighborhood to the segment form. Performance of this new approach is illustrated on a simulated image and on actual remote sensing images SPOT4/HRV.

## 1 Introduction

Classification processes were among the first attempts to interpret image quantitatively. Global approaches, such as Maximum A Posteriori (MAP), have been possible using Markov Random Fields (MRF) modeling [1]. For a given neighborhood system with clique potential functions, a global energy term is defined that should be minimized. Now, if the neighborhood form is considered as stationary within the image (that is generally the case), classification errors may occur on pixels having neighbors belonging to different classes. To overcome this problem, alternative approaches have been proposed, such as a line process [2] or specific potentials [3,4]. The solution proposed here is to relax the assumption of neighborhood stationary. Ant Colony Optimization [5] (ACO) that belongs to evolutionary computation [6, 7, 8, 9] algorithms has been successfully applied to routing in telecommunication networks (e.g. [10, 11]), quadratic assignment problem [12], graph coloring problem [13], traveling salesman problem [14]. The information gathered by simple autonomous mobile agents, called 'ants' (ACO derives from the behavior of social insects), is shared and exploited to solve a problem. Generally, 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. At each iteration, to produce the solution which will be memorized, several new solutions (as many as ants) are constructed: some of them taking into account a combination of several previous solutions (their number depending on the pheromone evaporation rate), and some of them being more or less randomly constructed. Like for SA or genetic algorithms, the introduction of randomness in the search procedure permits to escape from local minima. In this study, we propose to use the ACO by exploiting its ability of self-organization. Considering a MRF modeling with a non-stationary neighborhood, the ACO scheme jointly estimates the regularized classification map and the optimal non-stationary neighborhood. Section 2 deals with the image modeling; in section 3, the proposed method based on ACO heuristics is presented; section 4 shows some results, firstly on a simulated image, and secondly, on a SPOT/HRV image; finally, section 5 gathers our conclusions.

## 2 Classification Problem Assuming Non-stationary Neighborhood

The problem of image classification is to determine the realization of  $L$ , the label image, knowing those of  $X$ , the ‘radiometric’ image.  $\Omega$  being the set of the pixel locations (image lattice),  $X$  is a random field that takes values in  $\mathbb{R}^{|\Omega|}$ , and  $L$  is a random field that takes values in  $\Lambda^{|\Omega|}$ , where  $\Lambda = \{1, \dots, c\}$  and  $c$  is the class number. The cardinal of  $\Omega$  is  $|\Omega| = N_l \times N_c$ ,  $N_l$  being the image dimension in lines and  $N_c$  its dimension in columns. According to the *Maximum A Posteriori* criterion, the optimum solution maximizes  $p(X/L) \cdot p(L)$ . Assuming that the distribution law of the pixel values conditionally to their class  $l_s$  is an independent Gaussian of mean  $\mu_{l_s}$  and standard deviation  $\sigma_{l_s}$ ,

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

The prior model  $p(L)$  is defined assuming a MRF modeling. Neighborhood system is such that, if pixel  $s'$  belongs to pixel  $s$  neighborhood  $N(s)$ , then  $s$  is a neighbor of  $s'$ :

$$s' \in N(s) \Leftrightarrow s \in N(s'). \quad (2)$$

According to the Hammerley-Clifford theorem,  $p(L)$  follows a Gibbs distribution:

$$p(L) = \frac{1}{Z} \cdot \exp \left\{ - \sum_{\gamma \in \Gamma} \mathcal{V}_\gamma(l_s, s \in \gamma) \right\}, \quad (3)$$

where  $\Gamma$  is the set of the image cliques  $\gamma$  that describe the interactions between pixels,  $Z$  is a normalization constant, and  $\mathcal{V}_\gamma$  is the  $\gamma$  potential. Generally, the clique potentials are defined such that a pixel and its neighbors have a high probability to share the same label. Finally, the MAP criterion leads to the minimization of

$$E = \frac{1}{2} \sum_{s \in \Omega} \left[ \left( \frac{x_s - \mu_{l_s}}{\sigma_{l_s}} \right)^2 + \log(\sigma_{l_s}^2) \right] + \sum_{\gamma \in \Gamma} \mathcal{V}_\gamma(l_s, s \in \gamma). \quad (4)$$

The minimization of (4) is performed using the fact that the global energy difference between two label image configurations only differing by one pixel label only depends on this pixel  $s$  and its neighborhood. Under the assumption of pixel neighborhood having the same geometry at each pixel location  $s$ , this calculation can be performed handling a reasonable number of terms. However, the diversity of the areas or object shapes makes questionable the stationarity assumption for neighborhood form. Therefore, here we propose to relax this assumption and to adopt an approach where the neighborhood form is automatically adjusted. At each location  $s$ , the neighborhood is constructed based on three criteria:

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

Now, for the classification problem, (5) is still valid except that the cliques are now defined over non stationary 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

$$E = \sum_{s \in \Omega} \left[ \frac{(x_s - \mu_{l_s})^2}{2\sigma_{l_s}^2} + \log(\sigma_{l_s}) + \beta \frac{|\{r \in N(s); l_r \neq l_s\}|}{2} \right], \quad (5)$$

where  $|\{r \in N(s); l_r \neq l_s\}|$  is the number of neighbors having a different label than  $s$ , and  $\beta$  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. 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. We showed [15] that it corresponds to assume the existence of another random field  $H$  that corresponds to the definition of the non-stationary neighborhood in every pixel, and that the couple  $(L, H)$  is Markovian.

At each pixel location  $s$ , the optimal neighborhood is researched with the constraint of its cardinality (third criterion for neighborhood construction), and such that the symmetry property (2) is satisfied: once a pixel has constructed its neighborhood ( $h_s$ ), the neighborhoods of some other pixels (those belonging to  $h_s$ ) are already partially constructed, and subsequent neighborhood constructions should take into account already achieved neighborhood constructions.

Then, one must use a metaheuristic to find the ‘optimal’ neighborhood configuration defined (i) knowing the image observation, i.e.  $X$  realization, and (ii) assuming the image label, i.e.  $L$  realization.

### 3 Application of the ACO Meta-heuristic

Ant Colony Optimization mimics to the way social insects are able to solve some optimization problems. The ant problem is to find the shortest path between their nest and food. 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.

More conceptually, the problem of the ant colony is the following: given a function to minimize, different solutions are examined (randomly in a small percentage of cases), and each of them is memorized (thanks to the pheromone deposit) depending on its quality. In our case, we use ACO because of the analogy that can be done between the research of an optimal path by ants and the research of an optimal set of connected pixels from a given ‘originate’ pixel, and the required interaction (obtained using pheromone deposit) between local neighborhood solutions for (2) ascertaining. We now explain the way ACO is used for neighborhood construction. Assuming that the order of neighbor selection is without importance, in 8-connectivity, during the neighborhood construction, the following neighbor can be selected among any of the pixels located in a range of  $[1, +1]$  lines and column of an already selected neighbor.

According to the third neighborhood criterion, the cardinal of the neighborhood of pixel  $s$ ,  $|N(s)|$ , is assumed to be constant. In the following, it is noted  $N_n$ . Denoting  $\delta(.,.)$  the Kroenecker function:  $\delta(i, j) = 1$  if  $i = j$ ,  $\delta(i, j) = 0$  otherwise, and  $h_s = N(s)$  the  $s$  neighborhood  $|\{r \in N(s); l_r \neq l_s\}|$  writes  $\sum_{r \in h_s} 1 - \delta(l_s, l_r)$ . Then, using (2), (5) is:

$$E = \sum_{s \in \Omega} \left\{ \sum_{r \in h_s} \left[ \frac{\beta}{2} (1 - \delta(l_s, l_r)) + \frac{1}{N_n} \left( \frac{(x_r - \mu_r)^2}{2\sigma_{l_r}^2} + \log(\sigma_{l_r}) \right) \right] \right\}. \quad (6)$$

Now, according to the Markovian property of  $(L, H)$ , we are able to compare the energy of two configurations only differing by the label  $l_s$  of pixel  $s$  and the neighborhood realizations  $h_t$  of the pixels  $t$  included in  $W_{N_n}(s)$ , the  $(2N_n + 1) \times (2N_n + 1)$  sized neighborhood around  $s$ . 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. The algorithm is as follows [15].

The pixels  $t$  emit ants that gather the information about neighbor label, along some paths of connected pixels including  $t$ . During their paths, neighborhood constructions, the ants select the following neighbor from ‘routing indicators’.

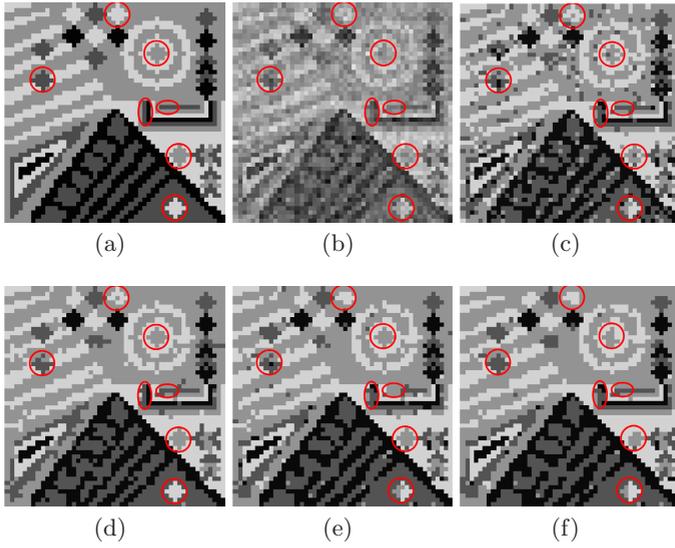
These latter are based on previously deposited pheromones (either by ants emitted by  $t$  or by ants emitted by other pixels having chosen  $t$  as neighbor) and the ‘energy hop’  $E_{t \rightarrow r}$  defined as:

$$E_{t \rightarrow r} = \beta(1 - \delta(l_t, l_r)) + \frac{1}{N_n} \left( \frac{(x_t - \mu_{l_t})^2}{2\sigma_{l_t}^2} + \log(\sigma_{l_t}) \frac{(x_r - \mu_{l_r})^2}{2\sigma_{l_r}^2} + \log(\sigma_{l_r}) \right). \quad (7)$$

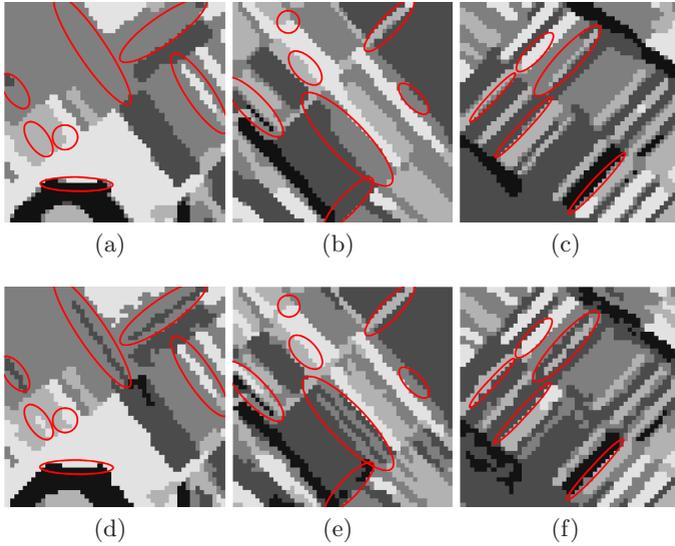
(7) shows that the following neighbor  $r$  is chosen considering its label  $l_r$  (relative to the emitting pixel label  $l_t$ ) and also its ‘data attach’ energy, which gives a hint about the  $l_r$  confidence. According to the ACO procedure, the next neighbor is chosen either randomly according to the probability of random exploration, or minimizing a function of  $E_{t \rightarrow r}$  and pheromone deposition. Practically, to simulate the pheromone deposition we define, for each pixel  $s$  a ‘neighborhood’ matrix of size  $(2N_n + 1) \times (2N_n + 1)$  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  $[N_s](r)$  the value of the neighborhood matrix of  $s$  in  $r$  ( $r \in W_{N_n}(s)$ ,  $[N_s](r) = [N_r](s)$ ), we define the cost of the choice of  $r$  as  $t$  following neighbor as  $c_{t \rightarrow r} = \beta(1 - [N_t](r)) + E_{t \rightarrow r}$ .

Arriving at a selected neighbor pixel, an ant waits a time proportional to the cost  $c_{t \rightarrow r}$ , 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. Then, it stops and on its return deposits pheromones on the visited neighbor pixels. Practically, for each pixel  $r$  visited by an ant emitted by pixel  $t$ ,  $[N_t](r)$  is increased by the quantity of pheromone deposit  $q$ . Due to (2), pheromones are also deposited on the ‘neighborhood’ matrices of the  $r$  selected neighbors, on the pixel corresponding to  $t$  in these ‘neighborhood’ matrices,  $[N_r](t)$ . Due to the waiting time, pixels on ‘good’ paths are visited frequently by ants, thus increasing neighborhood matrix values for pixels contained in those paths and diminishing other ones.

Then, the classification global algorithm is as follows. First are set the image parameters: neighborhood size  $N_n$  and weight  $\beta$ , and the ACO parameters: pheromone deposit quantity  $q$ , random exploration probability  $p_e$ , and experience duration  $T_e$ . For initialization, a blind classification of  $X$  is performed. As long as the stopping criterion is not verified, the pixels  $s$  are considered successively. For each  $s$ , the local energy  $E_1$  term is computed according to the current label  $l_s^{(1)}$  of  $s$  and the current neighborhoods  $h_t$  of the pixels  $t$  included in  $W_{N_n}(s)$ . For each new label to test  $l_s^{(2)}$ , reconstruct the neighborhoods  $h_t$  of the pixels  $t$  included in  $W_{N_n}(s)$  using ACO:  $W_{N_n}(s)$  pixels  $t$  emit ants, which explore neighborhood solutions and actualize neighborhood matrices as described previously; each time an ant is back to its source pixel  $t$ , if the ‘experience duration’  $T_e$  is not expired,  $t$  generates a new ant. At the end of the ACO experience, the new the local energy  $E_2$  term is computed. Then, the decision to change the  $s$  label from  $l_s^{(1)}$  to  $l_s^{(2)}$  is taken if the local energy has decreased, i.e. if  $\Delta E = E_1 - E_2$  is positive. With this algorithm, the convergence is ensured by



**Fig. 1.** Simulated data (noise  $\sigma = 40$ ): (a) ‘true’ label image, (b) data image, and obtained classifications: (c) blind result, (d) ACO result, (e) 8-connectivity stationary neighborhood result using SA algorithm, (f) line process [1] result



**Fig. 2.** Classification results: isotropic neighborhood (first line), and adaptive neighborhood (second line)-  $50 \times 50$  pixel subparts of an actual SPOT image

the fact that the global energy is decreased at each step, just as for the Iterative Conditional Modes (ICM, [16]).

## 4 Results

We first consider simulated data with Gaussian  $\mathcal{N}(\mu_i, \sigma_i)$  conditional distribution. Fig.1a shows the 4-class label image, and Fig.1b the data image when  $\mu_i \in \{100, 200, 300, 400\}$ , and  $\sigma_i$  equal to 40 for all  $i$ . Classical classification corresponding to (4) minimization with stationary neighborhood, either for null clique potential functions (blind classification, Fig.1c), or for clique potential functions corresponding to 8-connectivity Potts model was first performed. The MAP is obtained using simulated annealing process, testing different  $\beta$  parameter values and keeping the best result (Fig.1e). The blind result is very noisy. With the classical Potts stationary neighborhood, most of the ‘isolated’ errors have been corrected. To correct the packet errors the  $\beta$  parameter must be increased, however doing that some fine structures are lost. To overcome this limitation in performance, one has to change the image model revisiting the assumption of stationary neighborhood. Fig.1f is the result provided by the line process [1], much more sophisticated (and complex) than the isotropic neighborhood model, and *a priori* able to preserve image fine structures while regularizing the configuration. The ACO result is presented on Fig.1d. Among the considered approaches, ACO leads to the best result even if some errors remain due to the high level of noise. Some examples illustrating the interest of the proposed approach are pointed on Fig.1.

We now consider actual data acquired by the SPOT4/HRVIR sensor, having pixel size equal to  $20 \times 20m^2$  and from whose measurements a ‘vegetation index’ can be derived for the study of vegetation areas (e.g. agricultural areas). Fig.2 shows the results of classification considering five main classes of vegetation densities (corresponding to different growing stages), and pointed areas where the great complexity of the landscape, and the thinness some fields illustrates the interest of adaptive neighborhood approach (relative to isotropic one) both allowing better preservation of fine structures (areas 1, 2, 4, 5, 11, 13, 14, 16) and a removal of blind classification errors (3, 12) even in the presence of mixed pixels (7, 9, 16, 17, 18, 19).

## 5 Conclusion

In this study, we present a method to estimate non-stationary neighborhood shape in the framework of MRF Bayesian classification. It uses the meta-heuristic ACO, based on the behavior of social insects (ants) and their ability to find optimal solution thanks to the deposit of pheromones for their communications. Applied, as we propose, to the construction of pixel neighborhoods in global classification problems, it yields a performance superior to that of classical fixed form neighborhoods. The advantage of having a neighborhood shape, which automatically adapts to the image segment, clearly appears in the case of images

containing fine elements. The proposed method shows a stable performance relative to image parameter  $\beta$ , and is relatively robust to the exact fitting of the ACO parameters that can be calibrated to the proposed default values.

## References

1. Geman, S., Geman, D.: Stochastic relaxation, gibbs distribution and bayesian restoration of images. *IEEE Trans. on PAMI* 6, 721–741 (1984)
2. Geman, D., Geman, S., Graffigne, C., Dong, P.: Boundary detection by constrained optimization. *IEEE Trans. on PAMI* 12, 609–628 (1990)
3. Geman, S., Reynolds, G.: Constrained restoration and recovery of discontinuities. *IEEE Trans. on PAMI* 14, 367–383 (1992)
4. Descombes, X., Kruggel, F., von Cramon, Y.: Spatio-temporal fmri analysis using markov random fields. *IEEE Trans. on Medical Imaging* 17(6), 1028–1039 (1998)
5. Dorigo, M., Maniezzo, V., Colorni, A.: The ant system: optimization by a colony of cooperating agents. *IEEE Trans. on Systems, Man, and Cyber.* 26, 29–41 (1996)
6. Yao, X.: *Evolutionary Computation: Theory and Applications*. World Scientific, Singapore (1999)
7. Tan, K., Lim, M., Yao, X., L., P., W. (eds.): *Recent Advances in simulated Evolution And Learning*. World Scientific, Singapore (2004)
8. Rodríguez-Vázquez, K., Fonseca, C.M., Fleming, P.J.: Identifying the structure of nonlinear dynamic systems using multiobjective genetic programming. *IEEE Trans. on Sys., Man, and Cyber.* 34(4), 531–545 (2004)
9. Yeun, Y.S., Ruy, W.S., Yang, Y.S., Kim, N.J.: Implementing linear models in genetic programming. *IEEE Trans. Evol. Comp.* 8(6), 542–566 (2004)
10. di Caro, G., Dorigo, M.: Antnet: distributed stigmeric control for communications networks. *J. of Artificial Intelligence Research* 9, 317–365 (1998)
11. Sigel, E., Denby, B., Le Hégarat-Masclé, S.: Application of ant colony optimization to adaptive routing in a leo telecommunications satellite network. *Ann. of Telecommunications* 57, 520–539 (2002)
12. Colorni, A., Dorigo, M., Maffioli, F., Maniezzo, V., Righini, G., Trubian, M.: Heuristics from nature for hard combinatorial problems. *Int. Trans. in Operat. Research* 3, 1–21 (1996)
13. Costa, D., Hertz, A.: Ants can colour graphs. *J. of the Operat. Resea. Soc.* 48, 295–305 (1997)
14. Dorigo, M., Gambardella, L.: Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Trans. on Evol. Comp.* 1, 53–66 (1997)
15. Le Hégarat-Masclé, S., Kallel, A., Descombes, X.: Ant colony optimization for image regularization based on a non-stationary markov modeling. *IEEE Trans. on Image Processing* 16(3), 865–878 (2007)
16. Besag, J.: On the statistical analysis of dirty pictures. *J. of the Royal Statistical Society, Series B* 3(48), 259–302 (1986)