Towards Robust Analysis of Satellite Images Using Map Information - Application to Urban Area Detection

Shan Yu, Marc Berthod, Gérard Giraudon

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Towards Robust Analysis of Satellite Images Using Map Information — Application to Urban Area Detection

Shan Yu, Marc Berthod, Gérard Giraudon

N° 3293

November 3, 1997

THÈME 3

ISSN 0249-6399

Towards Robust Analysis of Satellite Images Using Map Information — Application to Urban Area Detection

Shan Yu, Marc Berthod, Gérard Giraudon

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet PASTIS

Rapport de recherche n° 3293 — November 3, 1997 — 32 pages

Abstract: With the rapid development in remote sensing, digital image processing becomes an important tool for quantitative and statistical analysis of remotely sensed images. These images contain most often complex natural scenes. Robust interpretation of such images requires the use of different sources of information about the scenes under consideration. This paper presents an integrated approach to the robust analysis of remotely sensed images by using multi-spectral SPOT image data, as well as map knowledge and contextual information. Several techniques are proposed for the effective use of map information for urban area detection in SPOT images. The first one is concerned with the modeling of SPOT images and map information using Markov random fields, which in turn permits application of various existing energy minimization algorithms for solving image analysis problems. The second one is on a new iterative optimization algorithm which automatically adjusts the optimal values of the parametres of our image model using a feedback control. The third one is about parameter estimation in the Markov random fields which takes into account both the map knowledge and the contextual information. The last one is concerned with the fusion of (intermediate) analysis results by using again the map knowledge in the estimation of the reliability of these results.

Key-words: satellite image analysis, map knowledge, urban area detection, image modeling by Markov random fields, contextual information, data fusion, optimization algorithms

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 93 65 77 77 - International : +33 4 93 65 77 77 — Fax : 04 93 65 77 65 - International : +33 4 93 65 77 65
à partir du 01/01/1998
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Vers une analyse robuste d'images satellitaires en utilisant des connaissances cartographiques — Application à la détection de zones urbaines

Résumé :

Avec le development rapide dans le domaine de la télédétection, les techniques de traitement d'images deviennent les outils importants pour l'analyse quantitative et statistique des images de télédétection. Ces images contiennent très souvent des scènes natuerelles complexes ; leur interpétation robuste nécessite l'utilisation de différentes sources d'information de la scène. Cet article présente une méthode intégrée destinée à l'analyse robuste des images de télédétection en utilisant des données multi-spectrales du SPOT, des connaissances cartographiques, et des informations contextuelles. Plusieurs techniques sont proposées en vue de l'utilisation effective des informations cartographiques pour la détection des zones urbaines dans des images SPOT. La première est sur la modélisation des images SPOT et des informations cartographiques par le formalisme de champs de Markov, qui permet l'application de divers algorithmes de minimization d'energie pour résoudre les problème d'analyse d'images. La deuxième s'agit d'un nouvel algorithme d'optimisation itératif avec l'adaptation automatique des valeurs optimales des paramètres du modèle de champs de Markov par un contrôle rétro-actif. La troisième concerne l'estimation des paramètres dans le modèle de champs de Markov en tenant compte à la fois les connaissances cartographiques et les informations contextuelles. La dernière est la fusion des résultats d'analyse intermédiaires où les connaissance cartographiques sont de nouveau utilisées pour estimer la fiabilité de chacun des résultats intermédiaires.

Mots-clés : analyse d'images de télédétection, connaissances cartographiques, détection de zones urbaines, modélisation d'images par champs de Markov, informations contextuelles, fusion de données, méthodes d'optimisation

1 Introduction

Remotely sensed images have proved to be of great interest for earth resource assessment and environment monitoring. Automatic or semi-automatic analysis of these images becomes an important task in computer vision. However, facing the complexity of such images, many classical image analysis techniques become inefficient. Relatively high rates of miss-interpretation of these images have been reported in various work.

Important efforts have been made to increase the reliability of analysis results of remotely sensed images. One of the most promising approaches is to exploit available information on the scene under consideration. A direct way to do so is to make efficient use of data obtained through different sensors or in different time periods, see [6, 11, 24] and references therein.

In addition to remotely sensed image data, other sources of information, in particular, cartographic information, turn out to be very useful as well. These include different kinds of maps (geographic maps, topographic maps, hydrographic maps, etc.), and some computer data bases like the Geographic Information System (GIS). Such data often contain information not available in multi-spectral images. Combination of multi-spectral image data and cartographic data may certainly improve the reliability in interpretating the scene. Some early work in this direction can be found in [13, 16, 17, 25].

In this paper, we present an integrated approach for the robust analysis of urban areas in remotely sensed images by using map knowledge as well as multi-spectral data and contextual information. The principal technical contribution of the work consists of two aspects. On one hand, we design robust image analysis algorithms by using these available sources of information. On the other hand, we enlarge the set of image analysis algorithms applied to the same image analysis task and we then merge the intermediate results in order to obtain a more reliable final result.

In designing robust image analysis algorithms we propose in particular a new optimization technique using a feedback control mechanism. This approach is based on the iterative update of potentials in the energy function of the Markov random fields (MRFs) in accordance with the map knowledge and the previous analysis result.

Applying different analysis algorithms to the same image analysis task and merging the results provide an attractive way to increase the reliability in remotely sensed image analysis. This can be considered as another fusion approach. The basic idea behind this technique is that different algorithms may yield different results which are possibly redundant and complementary. A judicious fusion of these results may produce a more reliable one.

In addition to these two main aspects, we also make use of image contextual information. Indeed, while information from different sources are helpful in satellite image analysis, contextual information contained in image itself is also an important source to be exploited. It can be used at pixel level or at object level. When used at pixel level, contextual information is usually taken into account by considering neighborhood information. For example, when classifying a satellite image, the class assigned to a site depends not only on the spectral feature of the site itself, but also on the spectral feature of its neighbors [12, 31]. When used at object level, contextual information can aid either in detecting less obvious

objects, or in obtaining a coherent interpretation of the whole scene. For example, in [11], the task of detection of bridge is made easier by considering previously detected rivers.

The paper is organized as follows. In the next section, we provide a general presentation of the work. In Section 3, we describe how to use Markov random fields (MRFs) to model image data, and to formulate our urban area detection problem as an energy minimization problem using Bayes formula. We also describe how to estimate the model parameters by taking into account the heterogeneity of image data. Existing optimization algorithms can be directly applied to solve this energy minimization problem. Experimental results of the performances of some recently developed optimization algorithms and comparison of the effect of model parameters are provided in the Appendix. In Section 4 we propose to integrate map knowledge in the MRF modeling using adapted Bayes formulation, and provide a new iterative optimization technique based on a feedback control mechanism. In Section 5, we present the fusion technique for improving the robustness of image analysis result using map knowledge. Finally, in Section 6, we conclude our presentation with remarks on the future work.

2 General Presentation of Our Approach

As we mentioned previously, due to the complexity of satellite images, many classical image analysis algorithms become inefficient. Two strategies are taken to tackle this problem. On one hand, we design robust image analysis algorithms by integrating auxiliary information. On the other hand, we enlarge the set of image analysis algorithms applied to the same task and then merge the intermediate results in order to obtain a more reliable final result.

The general scheme of our approach is composed of three steps, as illustrated in Figure 1. The first one is data preprocessing. In the second one, several algorithms are applied in parallel to the same problem. Last, the results of these algorithms are evaluated and fusioned to produce the final analysis result.

2.1 Data Preprocessing and Representation of Map Knowledge

Among different kinds of available auxiliary information, we are particularly interested in using map knowledge. Usually the geographic zone of the scene of a satellite image is known. Auxiliary information about the scene is therefore available, either from different kinds of maps or from some digital terrain data bases, like the Geographic Information System (GIS). For the work presented in this paper, we use information obtained from corresponding geographic maps.

Note that maps usually provide delayed information only. Ground changes may occur after the production of the map. Thus, when used for image analysis, map information should be considered as imprecise, uncertain, and out of date, so that it only provides a rough model of the scene.

In order to effectively combine satellite image data and the corresponding geographic map knowledge, some preprocessings are necessary. We consider a digitized map as an

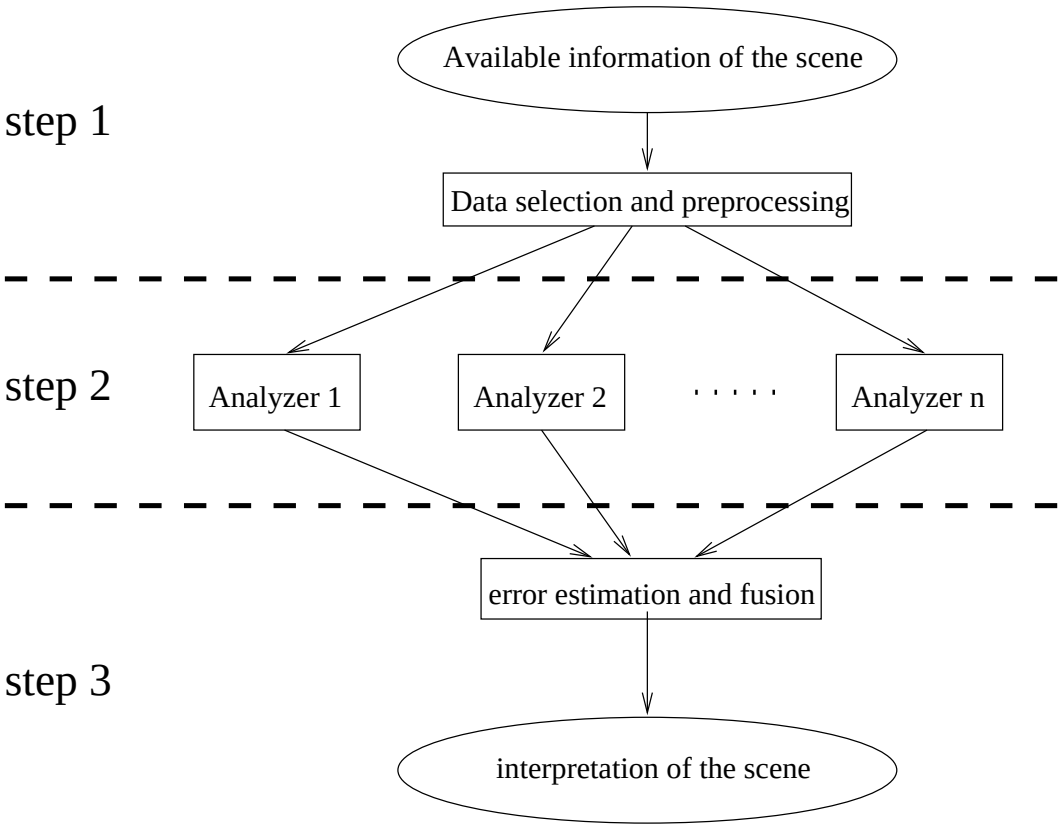


Figure 1: Scheme of the scene interpretation system.

image resulted from a special sensor, and thus treat it in the same way as other digital images. The registration of a satellite image and the corresponding digitized map allows us to have a multi-sensor information at pixel level, and this simplifies the integration of the two sources of information. We use a set of corresponding control points in both the satellite image and the map to accomplish the registration task. Techniques for image registration can be found in, e.g., [4, 19, 22].

Note that, very often only a subset of information contained in the digitized map is of interest for a specific task of scene analysis. This observation allows us to simplify the representation of the map knowledge by merely extracting information which is directly relevant to our analysis purpose. Here since we are interested in detecting urban area, we can thus consider a scene to be composed of two major classes: urban area and non-urban area. Information corresponding to this a priori requirement is extracted from the digitized map and represented in the form of a binary image (one grey level per class), called map image (see for example Fig. 4-(b)). In the current work, the production of this map image is accomplished by the operator. Automatic extraction of information from digitized maps remains a hard task for the time being. Nevertheless, such information can easily be obtained if a digital terrain data base of the scene is available.

As we shall see later on, this map image will be used in almost all steps of our analysis task: in automatic computation of the model parameters, in iterative improvement of urban area detection using a feedback control scheme, and in fusion of the intermediate analysis results issued from different algorithms.

2.2 Urban Area Detection

The second step of our approach consists in applying several image analysis algorithms to the same scene. Each algorithm will give a result indicating the urban area and the non-urban area in the image. Rather than giving a list of all the algorithms implemented, we shall concentrated on the approach that we proposed recently using Markov random field (MRF) modeling and Bayes formula.

We consider the problem of urban area detection as a scene labeling problem, where each pixel in the image is assigned a label indicating which class (urban area or non-urban area) it belongs to. Note that such a problem is also referred to as a classification problem or a segmentation problem in the literature. We shall use the MRF theory together with Bayes formula to formulate our objective function in terms of the maximum a posteriori (MAP) criterion for optimality (cf. Section 3). The original problem of urban area detection then becomes a problem of minimizing the objective function. Various optimization algorithms can thus be used to find the optimal or suboptimal solutions. For self-containedness of the paper, we provide in the Appendix a brief presentation of some recently proposed algorithms and their performance comparisons.

In order to make more efficient use of information extracted from geographic maps, the SPOT image and the map image are further integrated together by means of adapted Bayes formula in the framework of MRFs (cf. Section 4). Though the derived objective function still has the same, the labeling process is different here. We then devise a new technique

based on a closed-loop control scheme to adjust the behavior of the labeling process (see Figure 2). Any optimization algorithm of MRFs (such as those presented in Appendix) can be used in the algorithm as the kernel to get the intermediately labeling result. After each iteration, the parameters of the MRF model are modified according to the error rate of the current labeling result compared to the a priori knowledge about the growth ratio of the urban area.

2.3 Fusion of Analysis Results

The last step of the approach is the fusion of results issued from different analysis algorithms applied to the same scene (cf. Section 5). For a specific scene analysis task, several algorithms may be available and may yield different results. It is thus necessary to evaluate and compare the performance of each algorithm so as to assess the reliability of each result. With the reliability evaluation at hand, we can either retain the best result, or get an improved one by fusion of these results, see detailed discussions in Section 5.

3 Markov Random Field Modeling Using Bayes Formula for Urban Area Detection

In most vision problems, available information stems from two different sources: observations on image sites for a given occurrence of the problem, and a priori knowledge about the restrictions imposed on the simultaneous labeling of connected neighboring units. This second source of information, typically referred to as the "world model", reflects statistical dependencies between the labels of neighboring sites. Markov random field (MRF) theory provides a convenient and consistent way to model such context-dependent information. The MRFs-Gibbs equivalence, established by Hammersley and Clifford and further developed by Besag [2], gives an explicit formula for the joint distributions of MRFs. This enables us to model vision problems by a mathematically tractable means in the Bayesian framework. According to Bayes theory, when both the a priori distribution and the likelihood function of a pattern are known, the best that can be estimated from these sources of knowledge is the Bayes labeling. The maximum a posteriori (MAP) probability, as a statistical criteria for optimality, has been widely used in MRF modeling of vision problems.

3.1 Image Modeling

In the framework of MRFs, we are given a set of sites $\mathcal{S} = \{s_1, s_2, \cdots, s_N\}$. The sites in $\mathcal{S}$ are related one to another through a neighborhood system: $\mathcal{N} = \{\mathcal{N}_i | \forall i \in \mathcal{S}\}$. Each site has a set of possible labels $\Lambda_i \subseteq \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, $i = 1, 2, \cdots, N$. In most image applications, $\Lambda_i = \Lambda$, so that each site can take any label from Λ. An MRF on these sites is defined by the graph $G = (\mathcal{S}, \mathcal{N})$ and the so-called clique potentials. A clique c of G is a subset of sites in $\mathcal{S}$ where all sites are neighbors to each other. Let $\mathcal{C}$ be the set of all

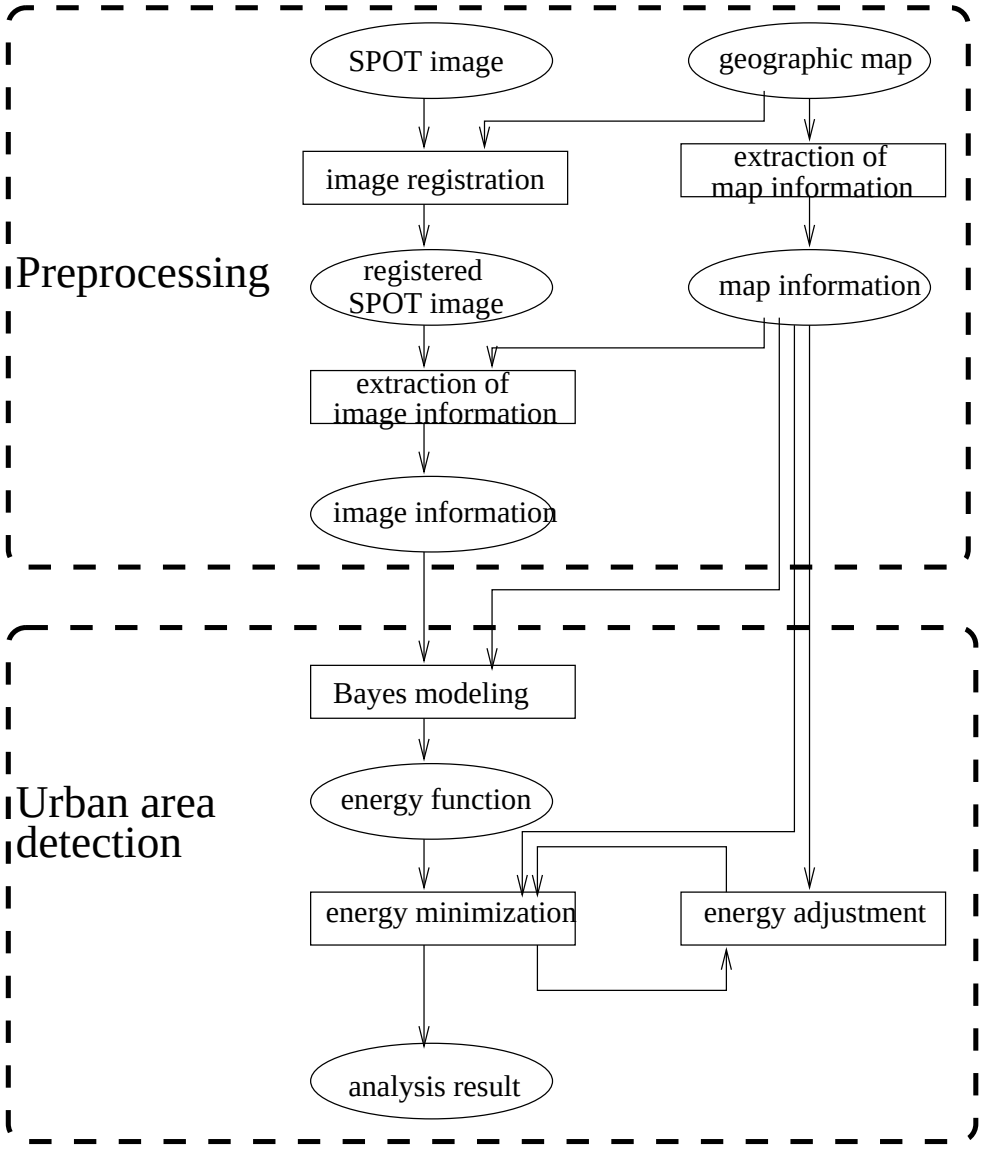


Figure 2: Scheme of urban area detection using map knowledge.

cliques of G, and $\mathcal{C}_i = \{c | s_i \in c\}$ the set of cliques containing site s_i. The number of sites in a clique c is referred to as its degree: $\deg(c)$.

A global discrete labeling L assigns one label $L_i \in \Lambda$ to each site s_i in $\mathcal{S}$. The projection of L to the sites of a given clique c is denoted by L_c. The definition of the MRF is completed by the knowledge of the clique potentials V_{cL_c} for every c in $\mathcal{C}$ and every L in $\mathcal{L}$, where $\mathcal{L}$ is the set of all possible M^N discrete labelings (i.e., the configuration space).

Let $Y = (y_1 \dots y_N)^t$ represent the observed image data, where y_i is the grey level or any other scalar or vector quantities on site s_i. Let $P(L)$ be the probability of having label L, and $P(Y)$ the probability distribution of Y. Bayes formula implies that the conditional probability of having label L given Y is

$$P(L|Y) = \frac{P(Y|L)P(L)}{P(Y)}, \tag{1}$$

where $P(Y|L)$ is the conditional probability distribution of the observations Y given the labeling L. It is usually assumed that this probability at each site is independent of its neighbors:

$$P(Y|L) = \prod_{i=1}^{N} P(y_i|L) = \prod_{i=1}^{N} P(y_i|L_i). \tag{2}$$

The a priori probability of the labeling $P(L)$ defines an MRF. According to Hammersley-Clifford theory [2], $P(L)$ can be expressed as a Gibbs distribution:

$$P(L) = \frac{1}{Z} \exp\left(\sum_{c \in \mathcal{C}} (-V_{cL_c})\right), \tag{3}$$

where Z is the normalizing constant, V_{cL_c} is the potential function whose value depends on c and L_c.

Substituting Eqs. (2) and (3) into Eq.(1), and dropping off $P(Y)$ and Z which do not depend on the labeling L (the solution to the optimization problem remains unchanged), the a posteriori distribution of the labeling $P(L|Y)$ can be rewritten as:

$$P(L|Y) \propto \prod_{i=1}^{N} P(y_i|L_i) \ \exp\left(\sum_{c \in \mathcal{C}} (-V_{cL_c})\right), \tag{4}$$

where $\propto$ stands for "proportional to".

The first term on the right hand side of Eq. (4) is the label-to-data fitting term. A commonly used model is that the observed value y_i (assumed to be a vector in general sense) is drawn from a Gaussian distribution, which gives:

$$P(y_i|L_i = l_i) = \frac{1}{\sqrt{2\pi|\Sigma_{l_i}|}} \exp\left(-\frac{1}{2}(y_i - \vec{\mu}_{l_i})^T \Sigma_{l_i}^{-1} (y_i - \vec{\mu}_{l_i})\right), \tag{5}$$

where $\vec{\mu}_{l_i}$ and Σ_{l_i} are respectively the mean vector and the covariance matrix of the feature vector Y for class l_i, $(y_i - \vec{\mu}_{l_i})^T$ is the transpose of the vector $(y_i - \vec{\mu}_{l_i})$.

Combining Eqs. (4) and (5), the a posteriori energy to be minimized can be expressed as:

$$\begin{aligned} U(L|Y) &= U(Y|L) + U(L) \\ &= \sum_{s \in \mathcal{S}} \left(\log |\Sigma_{l_s}| + (y_s - \vec{\mu}_{l_s})^T \Sigma_{l_s}^{-1} (y_s - \vec{\mu}_{l_s}) \right) + \sum_{c \in \mathcal{C}} V_{cL_c}. \end{aligned} \tag{6}$$

where $U(Y|L)$ is a disparity measure between the labels and the observed data. High disparity values, corresponding to high energy terms, are discouraged. The a priori energy $U(L) = \sum_{c \in \mathcal{C}} V_{cL_c}$ imposes constraints on the labels of neighboring sites in the MRF. Constraints on labels of two neighboring sites are the lowest order constraints to convey contextual information. They are widely used because of their simple form and low computational cost. In this paper, we shall only consider energies of cliques of orders 1 and 2. The first term in the right-hand-side of (6) shall be considered as energies of cliques of order 1, the second term as energies of cliques of order 2.

Concerning energies of cliques of order 2, for sake of mathematical and computational convenience, most MRF vision models are assumed to be homogeneous and isotropic, so that V_{cL_c} is independent of the relative position of the clique c in $\mathcal{S}$, and independent of the orientation of c. Under these assumptions, the Ising model, or more generally, the Potts model are frequently used for image segmentation problems:

$$V_{cL_c} = \beta \delta(L_i - L_j), \tag{7}$$

when $c = (i, j)$, β is a constant coefficient, where $\delta(x) = 0$ if $x = 0$ and $\delta(x) = 1$ otherwise.

In our experimentation, we find that in many cases, the above model does not provide satisfactory solutions to the problem of satellite image classification. Indeed, satellite images, especially those containing urban scenes, often exhibit heterogeneous surface coverage. The MRF defined on them do not usually satisfy the criteria of homogeneity and isotropicity. Thus, we adopted an inhomogeneous and anisotropic MRF model, and define the second order clique potentials as the compatibility coefficient

$$V_{cL_c} = \beta_{(i,j)}(L_i, L_j) \tag{8}$$

which depends on the clique $c = (i, j)$, both on its position in the image and on its orientation.

As a result, the original problem of urban area detection now becomes that of the minimization of the following energy function:

$$\begin{aligned} U(L|Y) &= U(Y|L) + U(L) \\ &= \sum_{s \in \mathcal{S}} \left(\log |\Sigma_{l_s}| + (y_s - \vec{\mu}_{l_s})^T \Sigma_{l_s}^{-1} (y_s - \vec{\mu}_{l_s}) \right) + \sum_{(i,j) \in \mathcal{C}} \beta_{(i,j)}(L_i, L_j). \end{aligned} \tag{9}$$

3.2 Estimation of Model Parameters

Parameters involved in the above energy function (Eq. (9)) are concerned with the energies of cliques of orders 1 and 2. Energies of cliques of order 1 are related to individual sites. These parameters are often obtained by supervised or unsupervised learning techniques. Energies of cliques of order 2, however, are related to mutual constraints of the sites. When the MRFs are supposed to be homogeneous and isotropic, the estimation of these parameters is usually carried out by optimizing a statistical criterion using techniques such as maximum likelihood, coding, pseudo-likelihood, etc. In our inhomogeneous and anisotropic MRF model, they are estimated in a local adaptive window centered at each clique, see below.

3.2.1 Estimation of Parameters of Cliques of Order 1

The image is to be classified into two classes: urban area and non-urban area. Under the assumption that the observed data are drawn from a Gaussian distribution, it is natural to use the mean value and the standard deviation of each class to be the model parameters. The map image allows us to automatically select the training zones and thus to compute the model parameters for each class. It also helps to choose the optimal set of texture descriptors for the image. Indeed, high resolution urban scene satellite images are often very rich in textures. It is therefore more appropriate to use texture measures rather than grey level values of the image to label different classes. Texture analysis has been extensively studied in the literature. In the present work, we do not propose new texture descriptors. Rather, we study how to select from the existing texture descriptors, those which are the most effective for the image being analyzed. To this end, we proceed in the following way [31]:

- Firstly, we restrict our selection within a limited number of texture descriptors having good performance according to results presented in the literature and confirmed by our experimentations. Such texture descriptors include those based on co-occurrence matrix [10], on local histogram measures [15], on some gradient measures [9], and Laws' texture detectors [14].
- Secondly, we combine these texture descriptors into several sets such that the descriptors of the same set produce texture measures which are as uncorrelated as possible (this can be checked by means of correlation matrix), and that the whole set is capable to discriminate certain types of textures. Each set contains a small number of texture operators, typically between 3 to 6 descriptors.
- Lastly, in order to determine which set of texture descriptors has the best capability to discriminate different regions for the image to be analyzed, we apply a simple Bayes classifier to the image with each of the sets. The set yielding the best (preliminary) classification result (compared to the map image) is considered to be the most appropriate one for the image under consideration. The corresponding texture measures are then used as input image data for our urban area detection algorithms.

3.2.2 Estimation of Parameters of Cliques of Order 2

For parameters related to cliques of order 2, as we mentioned previously, several estimation techniques exist for homogeneous and isotropic MRFs. For inhomogeneous and anisotropic MRFs, however, there does not seem to be any technique proposed in the literature for model parameter estimation.

For our inhomogeneous and anisotropic MRF model, the potential parameter $\beta_{(i,j)}(\lambda, \lambda')$ depends not only on the position of the clique c in $\mathcal{S}$, but also on the orientation of c. We interprate it as a compatibility coefficient for site i taking label λ and site j taking label λ' [28]. If we have a statistical model for the problem domain, the value of this compatibility coefficient can be determined using that model. Unfortunately, in most cases, we do not posses such a model. Therefore, we estimate its value by taking statistics over small local regions around clique c.

To estimate the value of $\beta_{(i,j)}(\lambda, \lambda')$, we consider the image to be a graph $G = (V, E)$ whose vertices in V represent image sites, whose (weighted) arcs in E represent the constraints on the label assignment of neighboring sites. We partition the set of arcs E into D subsets $E_1, E_2, \cdots, E_D$, with $\cup_{d=1}^{D} E_d = E$ and $E_{d_1} \cap E_{d_2} = \emptyset$, $d_1 \neq d_2$. E can be partitioned into 4-connection or 8-connection subsets depending on the definition of the neighborhood system. For the case of the 8-connection neighborhood system, E_1 ($E_2, E_3, E_4, E_5, E_6, E_7, E_8$, respectively) contains all arcs (i, j) such that j is the North-West (North, North-East, West, East, South-West, South, and South-East, respectively) neighbor of i.

Let $W_n(i)$ be the window of size $n \times n$ centered at site i. Define $E_{d,n}(i)$ as the projection of E_d over $W_n(i)$:

$$E_{d,n}(i) = \{(h, k) \mid h \in W_n(i),\ (h, k) \in E_d\}.$$

Let $P_i(\lambda)$ be the initial estimate of the probability that label $\lambda \in \Lambda$ is assigned to site $i \in \mathcal{S}$. Denote by $P_{i,j}(\lambda, \lambda')$ the joint probability that label $\lambda \in \Lambda$ is assigned to site i and label $\lambda' \in \Lambda$ is assigned to site j. For any given local window size $n \times n$, the estimation of the joint probability is computed as:

$$P_{i,j}(\lambda, \lambda') = \frac{1}{|E_{d,n}(i)|} \sum_{(h,k) \in E_{d,n}(i)} P_h(\lambda) \cdot P_k(\lambda'), \qquad (i, j) \in E_d, \quad 1 \leq d \leq D, \tag{10}$$

where $|E_{d,n}(i)|$ is the cardinality at $E_{d,n}(i)$. The potential parameters are obtained accordingly, cf. [20],

$$\beta_{i,j}(\lambda, \lambda') = \frac{P_{i,j}(\lambda, \lambda')}{P_i(\lambda) \cdot P_j(\lambda')} \tag{11}$$

Figure 3 illustrates the computation of $\beta_{ij}(\lambda, \lambda')$. E is defined by the 8-connection neighborhood system on a lattice. Sites i and j form a second order clique in the North-West direction. The local window (represented by shaded sites) of size 3×3 centered at site i indicates the region over which the statistic estimation of $\beta_{ij}(\lambda, \lambda')$ is carried out. The arrows represent the pairs of neighbors considered in the estimation of $\beta_{ij}(\lambda, \lambda')$.

The choice of the size of window $W_n(i)$ depends on the characteristics of the image under consideration. A common idea is that image sites in a small region are most likely to have

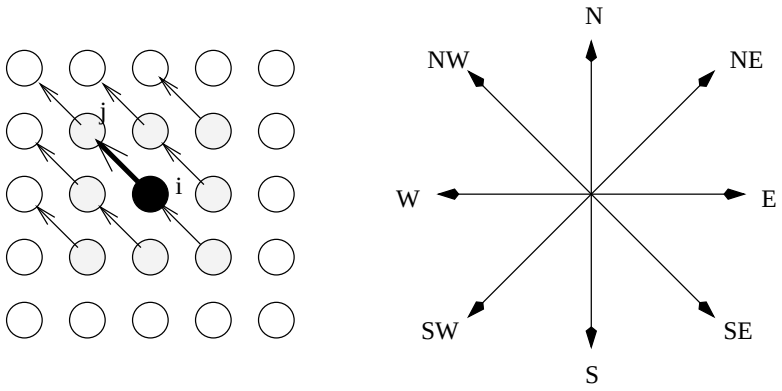


Figure 3: A window of size 3×3 centered at site i for the estimation of $\beta_{ij}(\lambda, \lambda')$ of clique $c = (i, j)$ in the North-West direction.

the same features. Thus, the window size n represents, in certain sense, the inverse of degree of likelihood between image sites. On the other hand, we should not take too small regions due to statistical biases. For example, the Gaussian white noise in the images can influence significantly feature measures of individual sites. A simple way to decrease the effects of additive noise is to use a low-pass filter on the image. In order to get an effective smoothing, the window size should be large enough. Hence, the size of region $W_n(i)$ represents, in this regard, the degree of smoothing. There is a tradeoff between the likelihood of the sites in a region and the effect of noise elimination in the region. Our experience shows that usually regions of size $n = 5, 7, 9$ and 11 provide very good estimates of the parameter $\beta_{ij}(\lambda_i, \lambda_j)$.

When the precision of the labeling is important, we can further determine adaptively the size of window $W_n(i)$ for each site i. This can be performed as follows: for every site i and for different size n, where $1 \leq n \leq N'$, and N' is a fixed value, we compute the variance of the quantity $P_h(\lambda_i) \cdot P_k(\lambda_j)$, where $(h, k) \in E_{d,n}(i)$, $1 \leq d \leq D$. We choose the size of window $W_n(i)$ to be the one that yields the smallest variance of the quantity $P_h(\lambda_i) \cdot P_k(\lambda_j)$.

The above method of estimating potential parameters for cliques of order 2 can easily be extended to the computation of potential parameters for cliques of orders higher than 2: $\gamma_{i_1,i_2,\cdots,i_m}(\lambda_1, \lambda_2, \cdots, \lambda_m)$. It can further be generalized to take into account the distance of the sites in a region:

$$P_{i,j}(\lambda, \lambda') = K_{i,n} \sum_{(h,k) \in E_{d,n}(i)} \alpha_{i,h} P_h(\lambda) \cdot P_k(\lambda'), \qquad (i, j) \in E_d, \quad 1 \leq d \leq D, \qquad (12)$$

where $K_{i,n}$ is the normalization constant, $\alpha_{h,i}$ is a coefficient depending on the distance $d(i, h)$ between site i and site h. For example, $\alpha_{i,h}$ can be defined as $1/(1 + d(i, h))$.

This estimation method (Eq. (10)) can also be used in homogeneous and isotropic MRFs by taking the average of $P_{(i,j)}(\lambda_i, \lambda_j)$ in all the directions with the whole image taken as the window.

3.3 Energy Minimization by Optimization Algorithms

Once the urban area detection problem is formulated as an energy minimization problem, it can be solved by an optimization algorithm. Quite a lot literature exists on optimization algorithms in this framework, see [1] and references therein.

In the Appendix we have included a brief description of some recently proposed algorithms: Deterministic Pseudo-Annealing (DPA), Game Strategy Approach (GSA), and Modified Metropolis Dynamics (MMD). We have also included our investigations on the effectiveness of these algorithms and the impact of model parameter estimation techniques.

4 Integration of Map Knowledge in Bayes Modeling and in the Labeling Process

To make more efficient use of map information, we integrate it directly in the Bayes modeling of the image by MRFs and in the labeling process so as to increase the robustness of the analysis result.

4.1 Integrating Map Knowledge in Bayes Modeling

We partition the set of image sites into two subsets: $\mathcal{S} = \mathcal{S}' \cup \mathcal{S}''$, $\mathcal{S}' \cap \mathcal{S}'' = \emptyset$, where $\mathcal{S}'$ is the subset of sites corresponding to the urban area indicated by the map image. Let L' and L'' be the projection of labeling L on $\mathcal{S}'$ and $\mathcal{S}''$, respectively. Let $\boldsymbol{\lambda}$ be the vector of labels of sites in $\mathcal{S}'$ such that all components are equal to the label "urban area".

According to Bayes formula, we obtain the a posteriori probability of the labeling L, given a realization of the observation Y on all the sites in $\mathcal{S}$, and the a priori knowledge of labels on the sites in $\mathcal{S}'$:

$$P(L|Y,\boldsymbol{\lambda}) = \frac{P(Y,\boldsymbol{\lambda}|L)\cdot P(L)}{P(Y,\boldsymbol{\lambda})} = \frac{P(Y|L)\cdot P(\boldsymbol{\lambda}|L)\cdot P(L)}{P(Y,\boldsymbol{\lambda})}. \tag{13}$$

We assume that the urban region presented in the satellite image is a superset of that represented in the corresponding geographic map. In other words, we assume that urban areas are growing as time goes. With this assumption, $P(\boldsymbol{\lambda}|L)$ is either 0 or 1. Thus, the a posteriori probability $P(L|Y,\boldsymbol{\lambda})$ can be maximized only when $P(\boldsymbol{\lambda}|L) = 1$, which implies that $L' = \boldsymbol{\lambda}$. It then follows from Eq. (13) that

$$P(L|Y,\boldsymbol{\lambda}) = P(L''|Y,L' = \boldsymbol{\lambda}) = \frac{P(Y|L'')\cdot P(L'')}{P(Y,L' = \boldsymbol{\lambda})} \tag{14}$$

Analogous to the derivations in Section 3, we obtain that the maximization of $P(L|Y,\boldsymbol{\lambda})$ is equivalent to the minimization of the following energy function:

$$U(L|Y,\boldsymbol{\lambda}) = \sum_{s\in\mathcal{S}''} \left(\log|\Sigma_{l_s}| + (y_s - \vec{\mu}_{l_s})^T \Sigma_{l_s}^{-1}(y_s - \vec{\mu}_{l_s})\right) + \sum_{(i,j)\in\mathcal{C}''} \beta_{(i,j)}(L_i, L_j), \tag{15}$$

where $\mathcal{C}''$ is the set of cliques such that one of the vertices is in $\mathcal{S}''$.

One observes that Eq. (15) differs from Eq. (9) only in the set of sites (i.e. $\mathcal{S}''$ or $\mathcal{S}$) to be included in the evaluation of the energy function. Thus, when map information is integrated in the modeling stage, one only needs to set $L' = \boldsymbol{\lambda}$ and consider L'' on $\mathcal{S}''$.

4.2 Integrating Map Knowledge in the Labeling Process through a Feedback Control Scheme

The above energy function can be minimized by any of the optimization algorithms such as those of section A, which results in a labeling (i.e. urban area and non-urban area) of the scene. In order for the detected urban area to be closer to the expected one, we devise an an iterative process which adjusts the potentials in the energy function according to a feedback control scheme, see figure 2. In each iteration, the same optimization algorithm is used as the kernel (in our case, the GSA algorithm has been used) to carry out the labeling process [29].

Assume first that the growth ratio of the urban area, noted as τ, after the map was produced is available. This ratio can be obtained, for example, from estimating population growth or from town planning projects [23]. Let α be the growth ratio of the urban area computed according to the current labeling result compared to the map information.

Intuitively, if $\alpha < \tau$, the detected urban area is smaller than the expected one. Therefore, we increase the probability of the "urban area" label and decrease the probabilities of the other labels. Otherwise, if $\alpha > \tau$, the detected urban area is larger than the expected one. Therefore, we decrease the probability of the "urban area" label and increase the probabilities of the other labels.

Observe that for any site, the more neighbors belong to urban area, the more likely this site belongs also to urban area. Hence, the modification of these probabilities is such that the increase and decrease are proportional or inversely proportional to the number of neighbors having the "urban area" label.

Let λ_0 be the "urban area" label, d the degree of the graph G, and v_i the number of neighbors of site s_i having label λ_0. The adjustment procedure is performed in the following way (γ is an adaptation parameter, $0 < \gamma < 1$):
If $\alpha < \tau$, then,

$$\begin{aligned} P_i(\lambda_0) &:= P_i(\lambda_0) + (1 - P_i(\lambda_0)) \cdot \gamma \cdot \frac{v_i}{d}; \\ P_i(\lambda) &:= P_i(\lambda) - P_i(\lambda) \cdot \gamma \cdot \frac{v_i}{d}, \qquad \lambda \neq \lambda_0. \end{aligned}$$

If $\alpha > \tau$, then

$$\begin{aligned} P_i(\lambda_0) &:= P_i(\lambda_0) - P_i(\lambda_0) \cdot \gamma \cdot \frac{1}{1 + v_i}; \\ P_i(\lambda) &:= P_i(\lambda) + \left(\frac{1}{|\Lambda| - 1} - P_i(\lambda) \right) \cdot \gamma \cdot \frac{1}{1 + v_i}, \qquad \lambda \neq \lambda_0. \end{aligned}$$

The adjustment of initial probabilities are then taken into account in the computation of energy potentials of cliques of order 2 using Eq. (10).

Note that even if the growth ratio of urban area is unknown, one can set $\tau = 0$ and still apply the iterative scheme for urban area detection. In most of all our experiments, the growth ratio α of urban area computed at the first iterations is larger than the expected one τ. Our experiences indicate that after several iterations the computed ratio α becomes close to τ, and that the modifications of labeling results of successive iterations are then very small and almost stagnant. Therefore, in the absence of information on τ (we set $\tau = 0$ in this case), one can stop the iteration when this phenomenon is observed.

4.3 Experimental Results

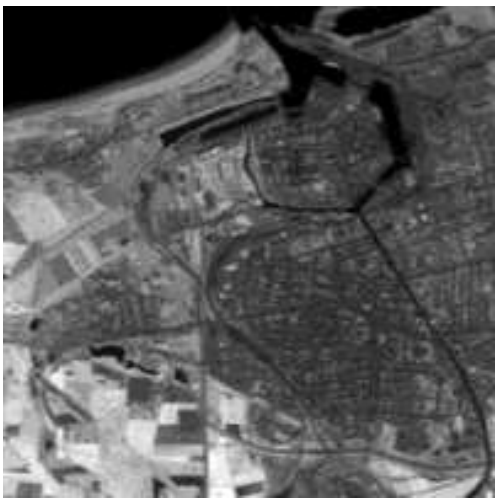
Figure 4 illustrates an experimental example of urban area detection. Figure 4-(a) is a portion of the SPOT XS3 image over the region Calais in France. Since we are interested in finding urban areas alone, we use just two classes to label different regions of the image: λ_1 = "urban area", λ_2 = "non-urban area". The map information corresponding to our interest is shown in Figure 4-(b), where the white color represents urban areas, the dark color represents non-urban areas. It is assumed that such information is not perfect. As we can see, the small urban areas at the top left corner in Figure 4-(a) are not present in image 4-(b).

Figure 4-(c) shows the urban areas detected by the method presented in Section 3 with GSA as the optimization algorithm. In this method, map information is only used to select optimal texture feature descriptors of the input image. There is no integration of map knowledge in the image model nor in the labeling process. Figure 4-(d) provides the urban areas detected by the method presented in this section, with GSA as the kernel optimization algorithm.

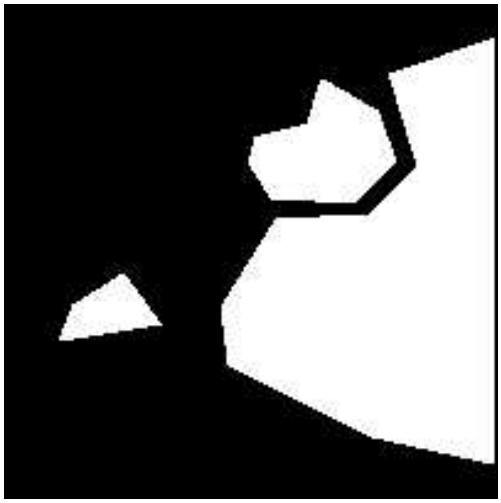
Comparing Figures 4-(c) and (d), one observes that the method presented in this section improves the quality of scene labeling algorithms. The urban areas detected by this last method is more complete and closer to the reality. As far as this particular image is concerned, the undetected urban region at the center of Figure 4-(c) is mainly due to the fact that in this region there is a museum and some monuments surrounded by green area. Hence the texture features of this region resemble more those of the countryside rather than those of the urban area in the image. This region is detected as urban area in Figure 4-(d) owing to the enforced utilization of the map knowledge in the above method.

5 Improving Image Analysis Quality through Fusion

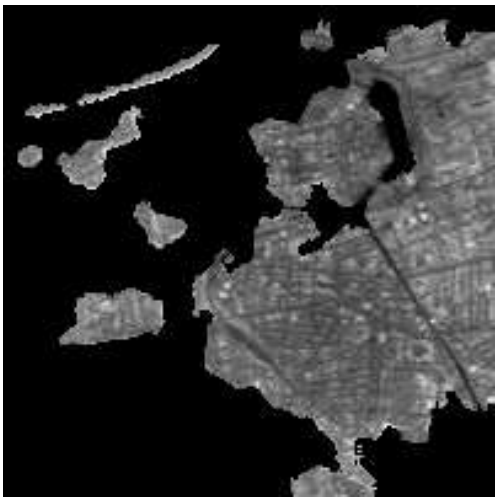
We now present the third step of our work on urban area detection (cf. Figure 1). The work presented in the previous section has been focused on designing robust image analysis algorithms. In particular, we use information extracted from geographic maps to improve analysis qualities of SPOT images. However, satellite images are often very complex. One



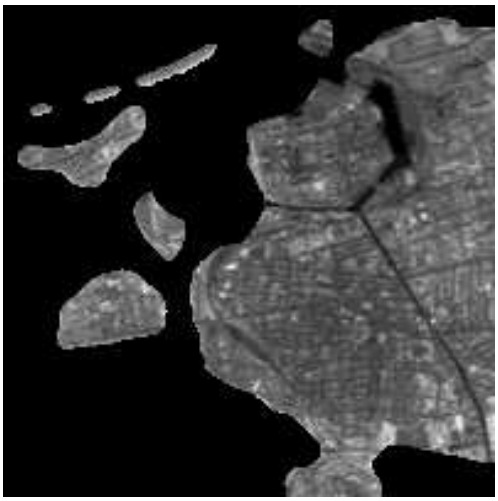
(a)



(b)



(c)



(d)

Figure 4: (a): A portion of the SPOT XS3 image over the region Calais in France. (b): Map information about the urban area obtained from the corresponding geographic map. The white region represents urban area, the dark region represents non-urban area. (c): Urban area detected by the simple GSA. (d): Urban area detected by the iterative GSA.

single algorithm may not succeed in analyzing the images with sufficient accuracy and reliability.

For a specific scene analysis task, several algorithms may be available and may yield different but possibly partially redundant and complementary results. In what follows, we investigate the possibility of the fusion of these results, in accordance with their performance evaluated using map knowledge, so as to increase the reliability of the final result. The work is performed in two steps, as in [27] First, we evaluate the quality of each (intermediate) result with the help of map knowledge. Then, based on this evaluation, we consider the problem of fusion of these results.

5.1 Quality Evaluation

To evaluate the quality of an image analysis result, we first compute a local error measure $E(s)$ for each site s, then sum it up over all the image sites to get a global error measure E_{global} for the result. The global error measure determines whether a result is good enough to be used in the fusion process. For those retained for the fusion process, the local error measure determines how each pixel in different results should be taken into account in the fusion process: a pixel with a smaller local error measure has a higher reliability, and thus has more influence in the fusion process.

The ideal way to assess the quality of an image analysis result would be to use the ground truth of the scene. Unfortunately, it is not always available. In this work, we use information extracted from geographic maps instead. Note that although map information, being simplified and labeled, is easy to use, it provides only a rough model of the scene. It is therefore imprecise and uncertain. This factor should be taken into account when used for evaluating the quality of image analysis results.

For any $s \in \mathcal{S}$, let $Map(s)$ denote the label of site s according to the map information, and $Img(s)$ the label of site s given by the image analysis result. Suppose that $Map(s) = \lambda_i$, $Img(s) = \lambda_j$. The change of label from λ_i (in the map) to λ_j (in the image analysis result) may incur an "error" that we call a risk:

$$x_{\lambda_i,\lambda_j} \begin{cases} = 0 & \text{if } i = j, \\ > 0 & \text{if } i \neq j. \end{cases}$$

The value of x_{λ_i,λ_j} depends on the specific labels λ_i and λ_j. Generally speaking, the risk of changing from a label of a natural object (e.g. the forest) to a label of a constructed object (e.g. a road) is smaller than that of changing from constructions to natural scenes. For example, the risk of a site of label "countryside" according to the map changing to label "road" in the image analysis result is smaller than that of a site of label "town" in the map becoming label "forest" in the image analysis result. In our work we use information from both the map and the image analysis result, as well as some common sense knowledge about the world model, to assign a value to the risk.

An image site cannot be analyzed without being related to its neighbors. Let s' be a site in the neighborhood $\mathcal{V}_s$ of site s, and $d(s, s')$ the distance between s and s'. We compute

the following two measures:

$$a(s) = \sum_{s' \in \mathcal{V}_s} \frac{\mathbf{1}_{\{Map(s')=\lambda_j\}}}{d(s,s')} \tag{16}$$

$$b(s) = \sum_{s' \in \mathcal{V}_s} \frac{\mathbf{1}_{\{Img(s')=\lambda_j\}}}{d(s,s')}, \tag{17}$$

where $\mathbf{1}_{\{\bullet\}}$ is the indicator function. Measure $a(s)$ is the number of sites in $\mathcal{V}_s$ which are labeled as λ_j in the map, inversely weighted by their distance to site s. It represents the support that site s gets from its neighbors in the map for taking λ_j as its label. Measure $b(s)$ represents the support that site s gets from the image analysis result for this decision. The size of the neighborhood depends on the resolution of the image. Higher the resolution is, larger the size could be.

The local error measure $E(s)$ of site s is then defined as:

$$E(s) = \frac{x_{\lambda_i,\lambda_j}}{r \cdot a(s) + b(s) + 1}, \tag{18}$$

where r is some positive coefficient. As both image analysis result and the map contain uncertainty and imprecision, we use the coefficient r to reflect the degree of confidence we have in the map with respect to the resulting image.

The global error measure E_{global} is simply the sum of the local error measures over all the image sites.

5.2 Fusion Process

Once we obtain the global error measure of an image analysis result, we can decide whether this result is good enough to be used in the fusion process. To do this, we simply take K best results among all available ones. Another way to de this is to use a threshold h: only those for which $G_{global} < h$ are retained for the fusion process. However it is nontrivial to determine the value of h.

Though a result selected to be used in the fusion process has a relatively high quality, all sites in this result do not have the same reliability for the fusion. Thus we compute a confidence coefficient for each site of the result based on the local error measure of this site:

$$C(s) = 1 - \frac{E(s)}{\sum_{\lambda_i,\lambda_j \in \Lambda} x_{\lambda_i,\lambda_j}}, \tag{19}$$

where the denominator is for the normalization of $E(s)$ so that $0 \leq C(s) \leq 1$.

The decision of the final label for a site depends on the accumulation of the confidence coefficient of the same site through all the retained results:

$$\lambda(s) = \arg\max_{\lambda_i \in \Lambda} \sum_{k=1}^{K} C_k(s) \mathbf{1}_{\{Img_k(s)=\lambda_i\}}, \tag{20}$$

where $Img_k(s)$ is the k^{th} input image to the fusion process, and $C_k(s)$ the confidence coefficient of site s in this result.

5.3 Experimental Results

Figure 5 illustrates the experimental result of application of the fusion process presented in this section on the same scene as in Figure 4. Figures 5-(a), (b) illustrate the urban areas of the scene detected by the two methods presented in Sections 3 and 4 respectively, where GSA is used as the optimization algorithm. Figures 5-(c) is obtained from another algorithm using wavelet transform and region growing techniques. Other algorithms such as those based on neural networks [26] can also be used. Figure 5-(d) is the map image. Figure 5-(e) is obtained by the fusion of the above three (intermediate) results and the map information. Values of different parameters used for this application are as follows: $K = 3$, $r = 2$, $x_{\lambda_1,\lambda_2} = 20$, $x_{\lambda_2,\lambda_1} = 1$. As for the size of the neighborhood, the second order neighborhood system is used, i.e., the 8 pixels in the 3×3 window centered at site s have been considered in equations (16) and (17). The improvement by fusion process is clear from this experimentation.

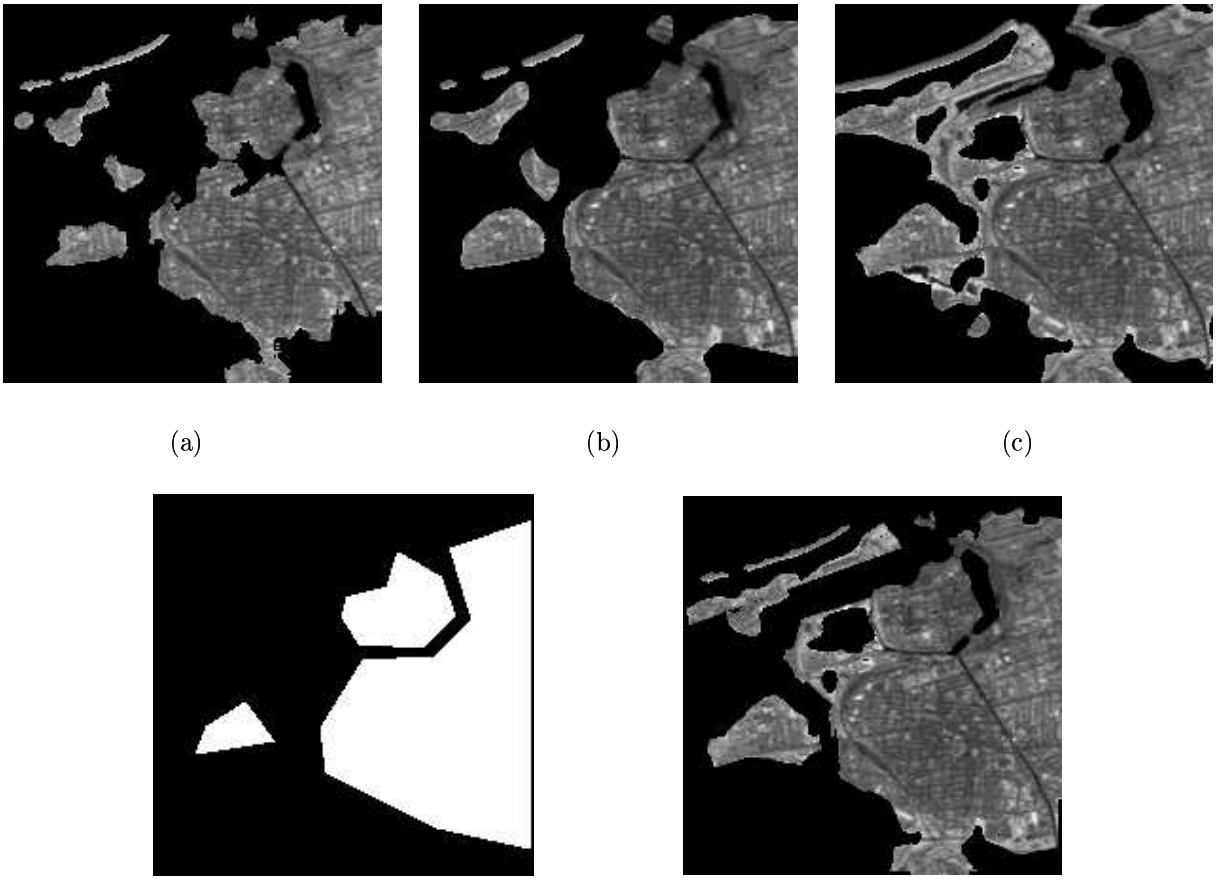
6 Conclusions

In this paper, we have presented an ongoing work on the interpretation of remotely sensed images by integrating multi-spectral satellite data, geographic map knowledge and image contextual information. We first provided an overview of the approach. Then we described a work on the utilization of map information to improve the effectiveness and robustness of urban area detection in SPOT images. Finally we presented a method to merge image analysis results issued from different algorithms performing the same task on a satellite image so as to increase the reliability of the final result.

Our experiences have shown that information extracted from geographic maps has great potential in guiding remotely sensed image analysis and in the design of robust algorithms for cartographic object detection. Though map information usually provides only a rough model of the scene, such a model is very useful in the learning of knowledge about the scene and in the optimization of parameters of image analysis algorithms.

Our work also confirms that context information plays an important role in the task of scene interpretation. When used at pixel level, context information provides neighborhood information around a pixel, and helps to increase the reliability of each detected object. When used at object level, it indicates the relative relations among different cartographic objects, which is useful for obtaining a coherent interpretation of the whole scene.

In order to make more efficient use of available information about the scene, techniques of data fusion are necessary at different levels of the analysis procedure. At input data level, they permit to integrate different sources of information to obtain most useful information and its appropriate representations about the scene. At object level, we can merge results issued from different analysis algorithms to reach a more reliable interpretation.



(a) (b) (c)

(d) (e)

Figure 5: (a), (b), (c): Images of urban area dectected by different methods. (d): Map information about the urban area obtained from the corresponding geographic map. The white region represents urban area, the dark region represents non-urban area. (e): Urban area obtained by fusion of the above three images and the map information show in figure 4-(b), using the method presented in Section5.

Until now, our main efforts have been mainly focused on the use of map information to improve the results of urban area detection. The detection of other cartographic objects with the help of map information will also be investigated. Furthermore, information from GIS will be used whenever it is available for the scene.

The ultimate goal of the work is to build an automatic and robust scene interpretation system. In order for the results to be reliable, the system should contain a set of robust image analysis tools, and should make full use of available information about the scene under analysis. Many problems remain to be solved for the scene interpretation to be effective and efficient. What kind of input data should be used for detecting a specific object? How to select the most effective image analysis tools for a specific task? What should the sequence of objects to be detected? What kind of a priori knowledge could be used to guide an analysis task? How to manage different kinds of information? How to model the imprecision and the uncertainty of the information? The control of the whole interpretation process, the management of different information and the selection of image analysis tools would definitely require a knowledge-based system, such as MESSIE, a multi-specialist architecture developed in our laboratory for scene interpretation tasks, cf. [5, 7].

A Energy Minimization by Optimization Algorithms

In this appendix, we present experimental results obtained by applying three recently developed optimization algorithms for the minimization of the objective function (9) established in Section 3. We also illustrate the impact of the ways model parameters are estimated.

A.1 Brief Descriptions of the Algorithms

The optimization algorithms we implement and compare here are among the most effective algorithms for MRFs (cf. [1]), and are concerned with Deterministic Pseudo-Annealing (DPA), Game Strategy Approach (GSA), and Modified Metropolis Dynamics (MMD). Details about these three algorithms and comparisons with other well-known optimization algorithms in the literature can be found in [1]. Below we briefly present the basic ideas of these three algorithms.

Game Strategy Approach (GSA). This algorithm is based on the game theory [30]. We define a game by considering image sites as players, and labels as the strategies of the players. The set of all possible plays corresponds to the set of all possible configurations of the labeling. The payoff function of each player depends on the strategy of the player and also on those of its neighbors. The total payoff of the team takes into account our objective function $f(L)$. With this game view point, we let the players choose their strategies at discrete time epochs $k = 0, 1, 2, \cdots$. At any time $k \geq 1$, each player has one-step delayed information on the strategies of its neighbors, and each player chooses independently its new strategy to maximize its expected payoff. The label updating scheme is randomized with a

probability α $(0 < \alpha < 1)$, which not only guarantees the convergence of the algorithm but also makes the final labeling result less dependent on the initialization. It is shown in [30] that the set of solutions of the game (i.e. Nash equilibria) is identical to the set of local maxima of our objective function $f(L)$.

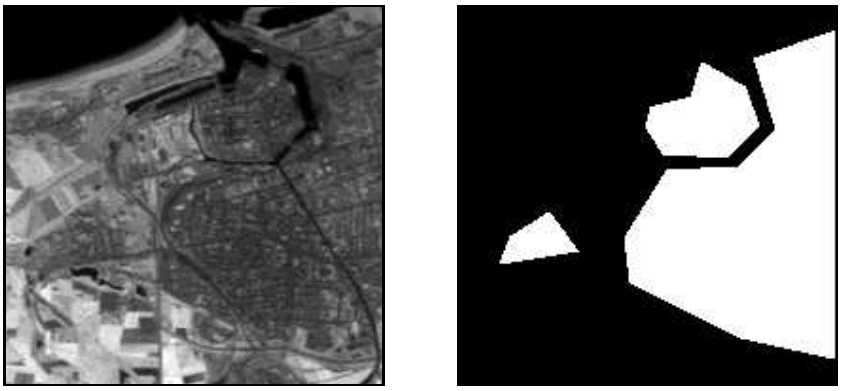
Deterministic Pseudo Annealing (DPA). Algorithm DPA is a variation of annealing which shares some common flavors with Mean-Field Annealing [8, 32], as well as Graduated Non-Convexity [3, 21]. The basic idea of DPA is to introduce weighted labelings, which assign a weighted combination of labels to any site, and then to build a merit function of all the weighted labels. This function, a polynomial with non-negative coefficients, is an extension to a compact domain of $\mathcal{R}^{\mathcal{N}}$ of an application defined on the finite (but very large) set of labelings; its only extrema under suitable constraints correspond to discrete labelings. The algorithm of DPA consists of changing temporarily the constraints, and thus the subset on which f is maximized, so as to convexify this function, find its unique global maximum, and then track down the solution until the original constraints are restored, thus obtaining a usually good discrete labeling.

Modified Metropolis Dynamics (MMD). The algorithm of MMD is a modified version of Metropolis Dynamics [18]: the choice of the new label state is made randomly using a uniform distribution; the rule to accept a new state is deterministic. In other words, the difference between the original Metropolis Dynamics and MMD is on the value of the variable ξ which is used in the rule to accept a new state. In the original method, ξ is chosen randomly at each iteration; in MMD, ξ is a constant, chosen at the beginning of the algorithm. At high temperature, the behavior of MMD is similar to the stochastic techniques. When the temperature goes down, it becomes deterministic.

A.2 Performance Comparison and the Impact of Parameter Estimation Techniques

Here we show the experimental results obtained by applying the above three algorithms to minimize the energy function (9) for urban area detection. We also provide their performance comparison with respect to different parameter estimation techniques.

Figure 6-(a) is a portion of the SPOT XS3 image over the region Calais in France. The scene is quite representative of a complex urban environment. Figure 6-(b), the map image, is used for automatic estimation of parameters of the first order cliques as described in Section 3.2.1. When we take an heterogeneous and anisotropic MRF model, i.e., parameters of the second order cliques are estimated as described in Section 3.2.2, cf. Eqs. (10) and (11), we obtain results shown in Figures 7-(a), (b), (c), issued from the algorithms of GSA, DPA and MMD, respectively. The white regions represent the detected urban areas, and the dark regions represent the non-urban areas. We can see that the three algorithms have quite similar performances. Figures 7-(d), (e), (f) illustrate the detected urban scenes from these labeling results.



(a) (b)

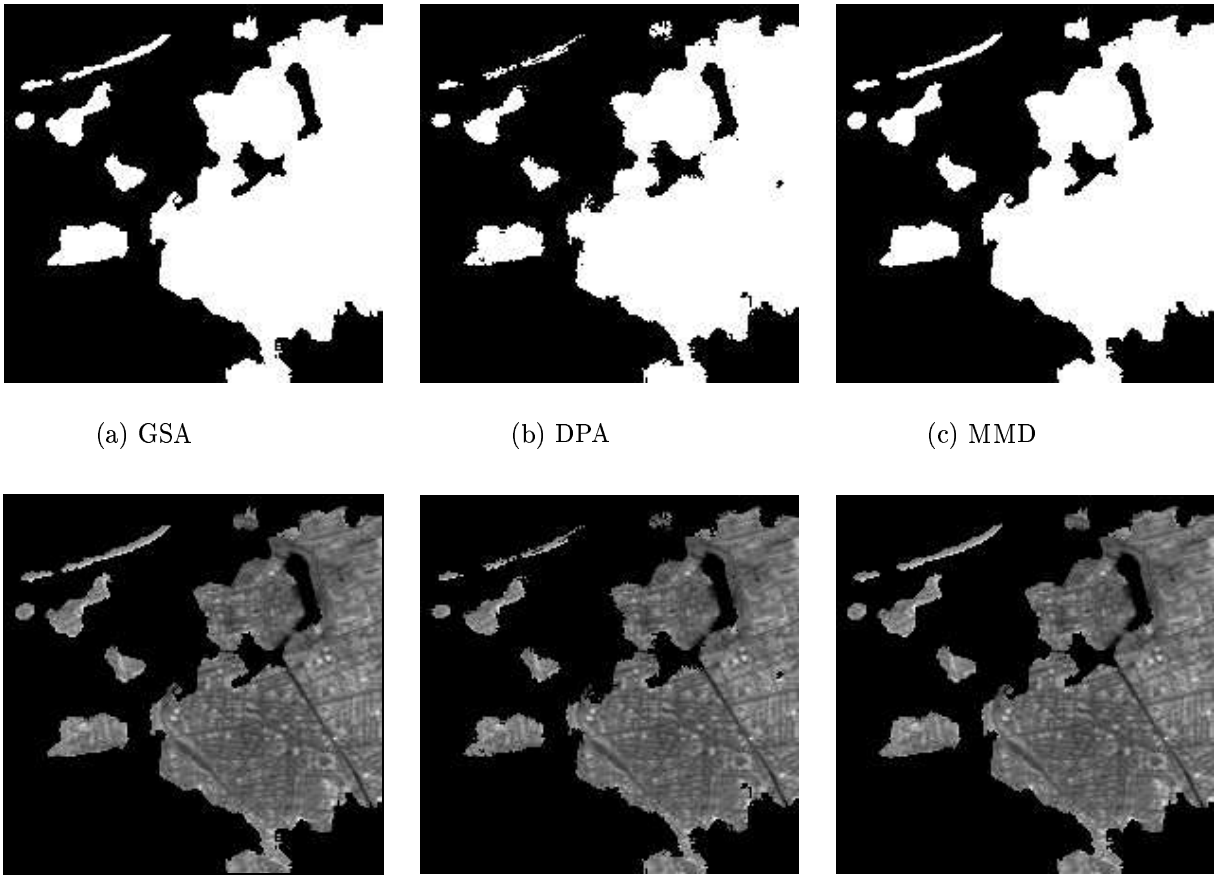
Figure 6: (a): A portion of the SPOT XS3 image over the region Calais in France. (b): The map image obtained from the corresponding geographic map.

For comparison, we also take the homogeneous and isotropic MRF model, i.e., parameters of the second order cliques are obtained using the Potts model, cf. Eq. (7). The labeling results are shown in Figures 8-(a), (b), (c). Figures 8-(d), (e), (f) are the corresponding detected urban scenes from these labeling results.

In order to see clearer the impact of model parameters, we use two synthetic images (cf. Figure 9) so as to compare between the homogeneous and isotropic model and the heterogeneous and anisotropic model. These comparison results are illustrated in Figures 10 and 11. Table 1 shows the error rates of the labeling results. The columns "heterogeneous model" correspond to our estimation method based on small windows. We can see from this table that the error rates with heterogeneous model are reduced by about a half compared to the homogeneous model (Potts model).

	Image of Figure 9-(a)		Image of Figure 9-(b)	
	Potts model	heterogeneous model	Potts model	heterogeneous model
GSA	5.63 %	3.03 %	9.33 %	5.91 %
DPA	5.33 %	2.78 %	10.25 %	6.03 %
MMD	5.41 %	2.68 %	9.23 %	5.78 %

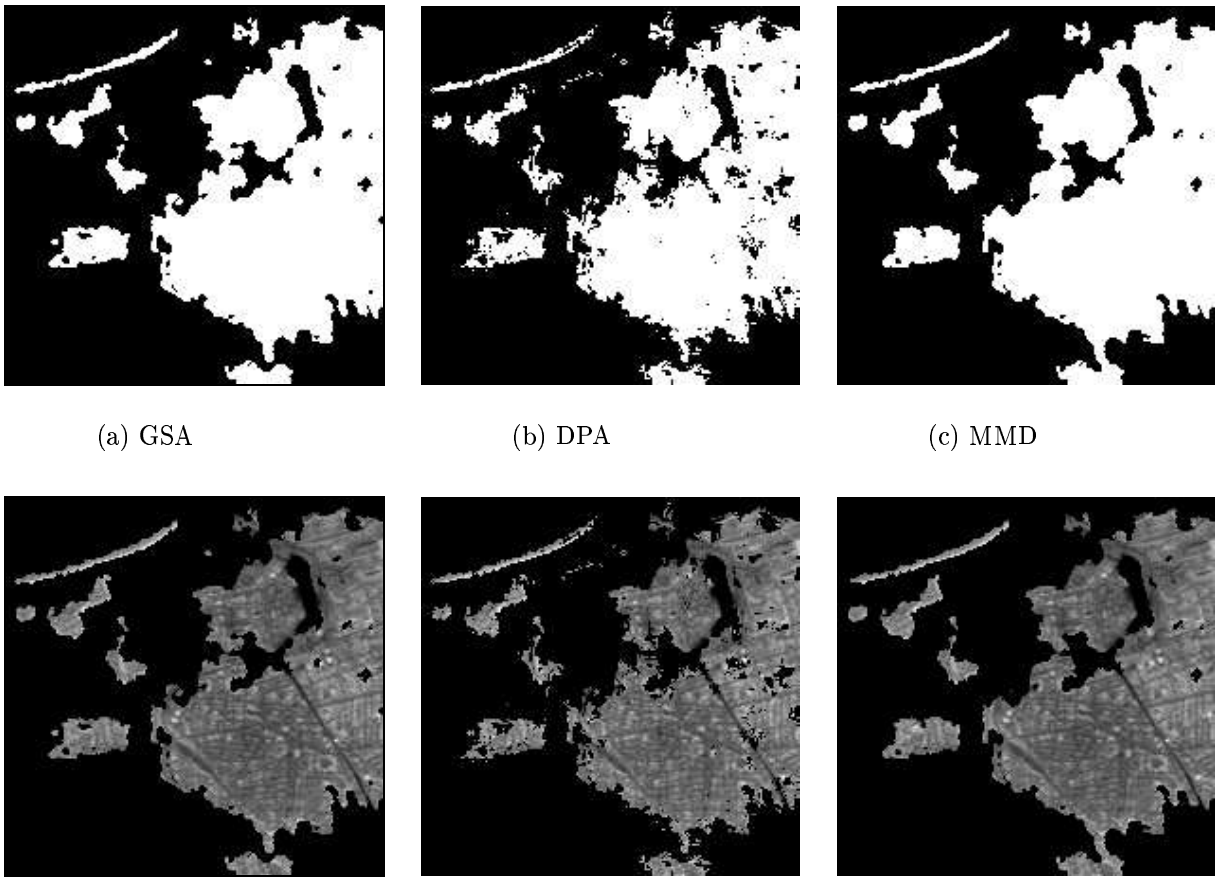
Table 1: Error rates of texture labeling by different algorithms (GSA, DPA, MMD) and different parameter estimations in MRF models.



(a) GSA (b) DPA (c) MMD

(d) GSA (e) DPA (f) MMD

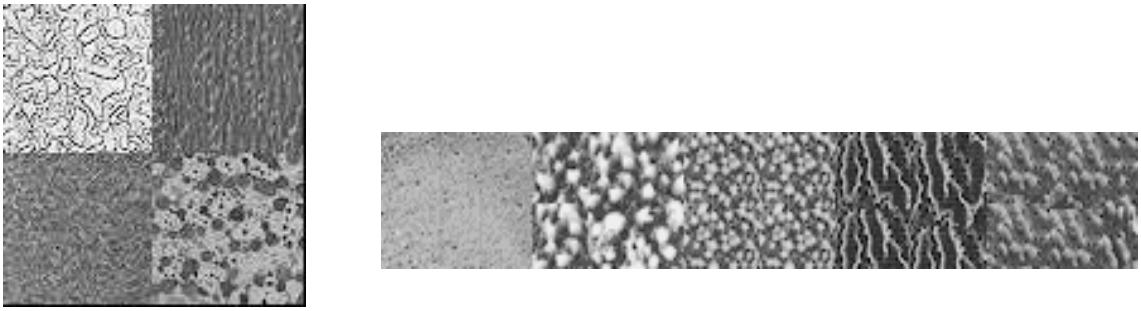
Figure 7: (a), (b), (c): Labeling results obtained by the algorithms of GSA, DPA and MMD, with a heterogeneous and anisotropic MRF model. White regions represent urban areas, Dark regions represent non-urban areas. (d), (e), (f): Detected urban scenes from the above labeling results.



(a) GSA (b) DPA (c) MMD

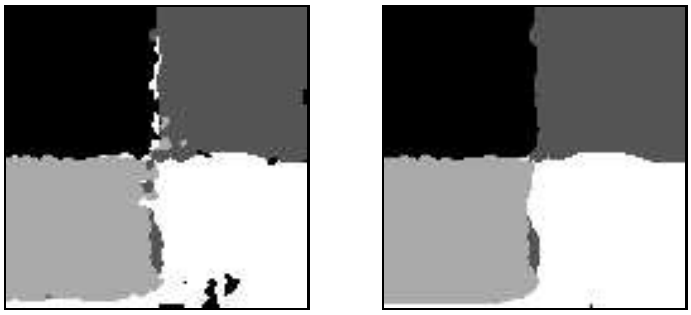
(d) GSA (e) DPA (f) MMD

Figure 8: (a), (b), (c): Labeling results obtained by the algorithms of GSA, DPA and MMD, with a homogeneous and isotropic MRF model. White regions represent urban areas, Dark regions represent non-urban areas. (d), (e), (f): Detected urban scenes from the above labeling results.



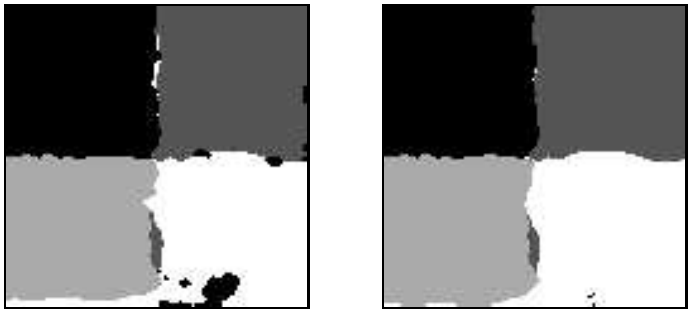
(a) (b)

Figure 9: Two synthetic texture images



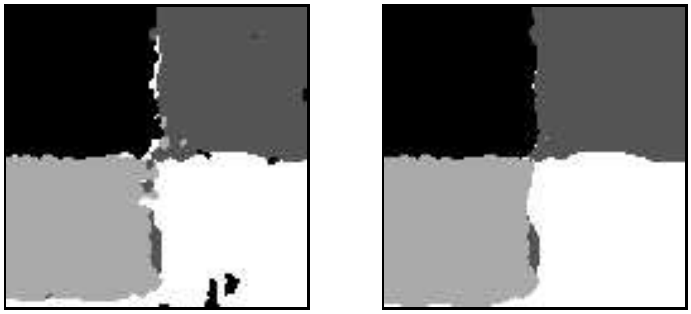
(a) (b)

Texture labeling by GSA. Model parameters estimated by:
(a) Potts model; (b) our method.



(c) (d)

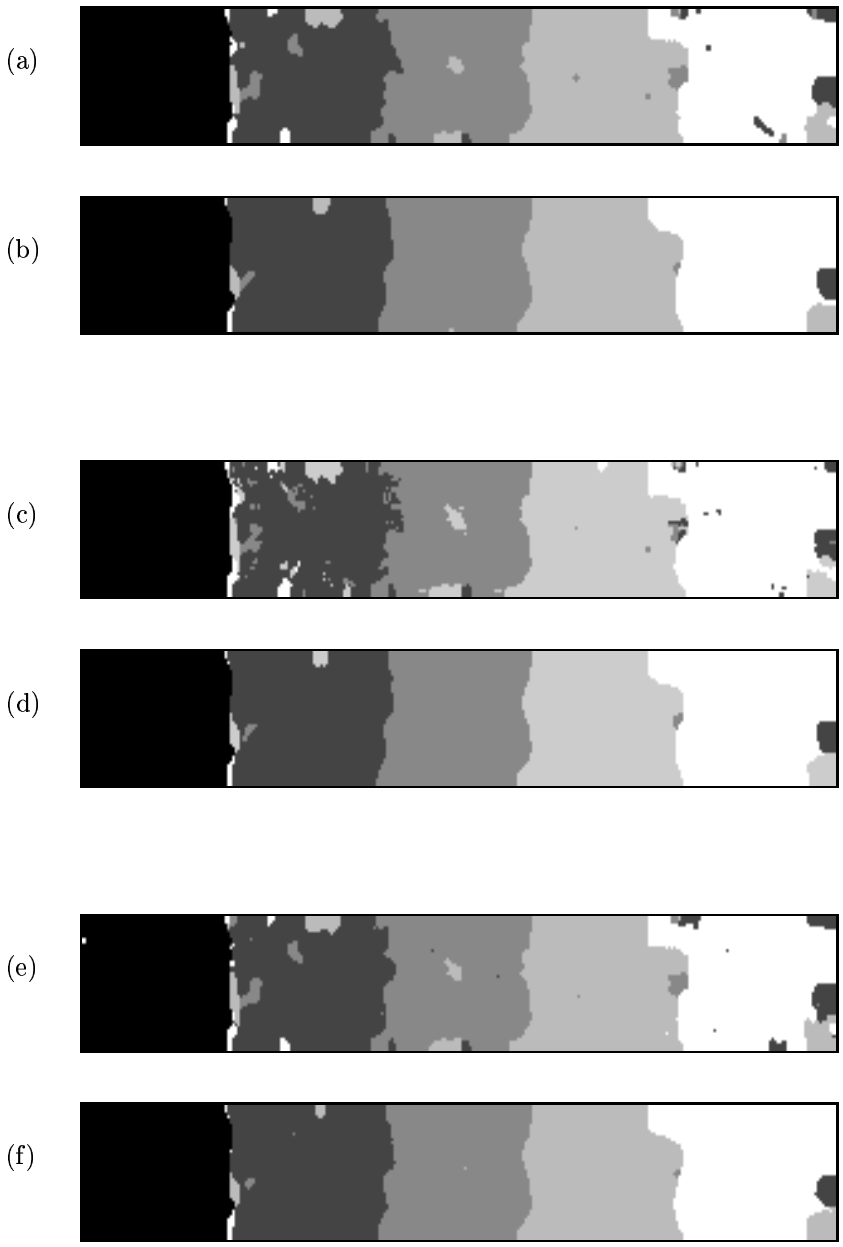
Texture labeling by DPA. Model parameters estimated by:
(c) Potts model; (d) our method.



(e) (f)

Texture labeling by MMD. Model parameters estimated by:
(e) Potts model; (f) our method.

Figure 10: Comparison results of texture labeling of the image in Figure 9-(a)



Texture labeling by GSA.
Model parameters estimated by:
(a) Potts model;
(b) our method.

Texture labeling by DPA.
Model parameters estimated by:
(c) Potts model;
(d) our method.

Texture labeling by MMD.
Model parameters estimated by:
(e) Potts model;
(f) our method.

Figure 11: Comparison results of texture labeling of the image in Figure 9-(b)

References

[1] M. Berthod, Z. Kato, S. Yu, and J. Zerubia. Bayesian image classification using Markov random fields. *Image and Vision Computing*, 14(4):285–295, May 1996.

[2] J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Statis. Soc. B.*, 68:259–302, 1986.

[3] A. Blake and A. Zisserman. Visual reconstruction. *MIT Press, Cambridge - MA*, 1987.

[4] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[5] V. Clément, G. Giraudon, S. Houzelle, and F. Sandakly. Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture. *IEEE Transaction on Geoscience and Remote Sensing*, 31(4), July 1993.

[6] S. J. Ford and D. M. McKeown. Utilization of multispectral imagery for cartographic feature extraction. In *IEEE Image Understanding Workshop DARPA*, pages 805–820, Jan 1992.

[7] P. Garnesson, G. Giraudon, and P. Montesinos. Un systeme multi-spécialistes en vision : Application à l'interprétation en imagerie aérienne. *Traitement du Signal*, 9(5), 1992.

[8] D. Geiger and F. Girosi. Parallel and deterministic algorithms for MRFs : surface reconstruction and integration. In *Proc. ECCV90*, pages 89–98, Antibes, France, 1990.

[9] V. Gertner and H. Andre. Extraction d'éléments texturés dans les images spot. Rapport de stage, essi 3, ESSI 3, 1990.

[10] R. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, 67(5):786–804, May 1979.

[11] S. Houzelle and G. Giraudon. Data fusion using SPOT and SAR images for bridge and urban area extraction. In *Proc. of Int. Geoscience and Remote Sensing Symposium*, Espoo, Finlande, 1991.

[12] B. Jeon and D. A. Landgrebe. Classification with spatio-temporal interpixel class dependency contexts. *IEEE Trans. on Geoscience and Remote Sensing*, 30(4):663–672, July 1992.

[13] B. Kiracofe, D. Houck, and C. Lucas. Automated change detection: where does the analyst fit? In *Proc. of SPIE Image Understanding and the Man-Machine Interface*, volume 758, pages 181–184, 1987.

[14] K. Laws. *Textured Image Segmentation.* Ph.d. thesis, Univ. of Southern California, Los Angeles, USA, 1980.

[15] G. Lowitz. Mapping the local information content of a spatial image. *Pattern Recognition*, 17(5):545–550, 1984.

[16] D. M. McKeown. The role of artificial intelligence in the integration of remotely sensed data with geographic information system. *IEEE Trans. on Geoscience and Remote Sensing*, 25(3):330–338, May 1987.

[17] D. M. McKeown and J. L. Denlinger. Map-guided feature extraction from aerial imagery. In *IEEE Workshop on Computer Vision*, April 1984.

[18] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. of Chem. Physics, Vol. 21, pp 1087-1092*, 1953.

[19] A. C. Morris and A. Stevens. Ground control determination for registration of satellite imagery using digital map data. *Photogrammetric Record*, 12(72):809–822, October 1988.

[20] S. Peleg. A new probabilistic relaxatioh scheme. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-2:362–369, 1980.

[21] A. Rangarajan and R. Chellappa. Generalised Graduated Non-Convexity algorithm for Maximum A Posteriori image estimation. *Proc. ICPR*, pages 127–133, Jun. 1990.

[22] M. Roux. *Recalage d'images multi-sources, application au recalage d'une image SPOT et d'une carte.* Ph.d. thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, Sept. 1992.

[23] G. Sadler and M. Barnsley. Use of population density data to improve classification accuracies in remotely-sensed images of urban areas. In *First European Conference on Geographical Information Systems*, Netherlands, April 1990.

[24] A. H. Schistad Solberg, A. K. Jain, and R. Taxt. Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. *IEEE Trans. on Geoscience and Remote Sensing*, July 1994.

[25] F. Wang. A knowledge-based vision system for detecting land changes at urban fringes. *IEEE Trans. on Geoscience and Remote Sensing*, 31(1):136–145, January 1993.

[26] K. Weigl and S. Yu. A hybrid approach for image classification using neural network and markov random field. In *Proc. of the International Conference on Artificial Neural Networks*, volume 2, pages 321–326, Paris, France, Oct. 1995.

[27] S. Yu. Improving satellite image analysis quality by data fusion. In *Proc. of 1995 International Geoscience And Remote Sensing Sumposium*, volume 3, pages 2164–2166, Firenze, Italy, July 1995.

[28] S. Yu. Efficient use of contextual information in image labeling algorithms. In *Fourth International Conference on Control, Automation, Robotics and Vision*, volume 2, pages 1388–1392, Singapore, Dec. 1996.

[29] S. Yu and M. Berthod. Urban area detection in satellite images using map knowledge by a feedback control technique. In *Proc. of the 12th International Conference on Pattern Recognition*, volume 1, pages 98–102, Jerusalem Israel, Oct 1994.

[30] S. Yu and M. Berthod. A game strategy approach to relaxation labeling. *Computer Vision and Image Understanding*, 61(1):32–37, Jan 1995.

[31] S. Yu, M. Berthod, and G. Giraudon. Satellite image segmentation using textural information, contextural information and map knowledge. In *Proc. of IEEE Systems, Man, and Cybernetics Conference*, volume 3, pages 355–360, Le Touquet, France, October 1993.

[32] J. Zerubia and R. Chellappa. Mean field approximation using compound Gauss-Markov Random field for edge detection and image restoration. *Proc. ICASSP, Albuquerque, USA*, 1990.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399