LAND COVER CLASSIFICATION FROM COARSE RESOLUTION TIME SERIES

A. Robin [1,2], S. Le Hégarat-Mascle [1], L. Moisan [2], H. Poilvé [3]

[1] CETP (CNRS), 10-12 av. de l'Europe, 78140 Vélizy, France - mascle@cetp.ipsl.fr

[2] MAP5 (CNRS), 45 rue des Saints-Pères, 75270 Paris cedex 06, France - { robin, moisan }@math-info.univ-paris5.fr

[3] EADS ASTRIUM, 31, avenue des Cosmonautes, 31402 Toulouse Cedex 4, France - poilve@astrium.eads.net

Abstract - Land cover classification requires both temporal and spatial information. Indeed, vegetation temporal evolution is necessary to discriminate the different land cover types. This information can be derived from coarse resolution sensors such as MERIS ($300 \times 300m^2$ pixel size), or SPOT/VGT ($1km^2$ pixel size), whereas high resolution images, such as SPOT4/HRV ones ($20 \times 20m^2$ pixel size), contain the required spatial information. In this paper, a new method is proposed to perform an efficient land cover classification using these two kinds of remote sensing data. This method is based on Bayesian theory and on the linear mixture model permitting, through a simulated annealing algorithm, to perform a high resolution classification from a coarse resolution time series.

Keywords: multiscale classification, land cover, coarse spatial resolution, mixture model, multitemporal series

1. INTRODUCTION

Nowadays, agricultural surveys, forest or environmental monitoring, natural resources management use widely remote sensing data. Land cover classification provides essential information for studies on geosphere-biosphere-atmosphere interaction, analysis of global change and weather forecasting. Indeed, vegetation characteristics have an important impact on surface processes involved in the water or energy exchanges. The advent of remote sensing science provides a powerful tool for land cover monitoring and information extraction. Satellite sensors providing more and more data and information, automatic tools are worth for operational approaches in both environmental researches and management activities.

For vegetation monitoring (and in particular agricultural applications), the time evolution is one of the most discriminating criteria. Hence, land cover classification requires high temporal frequency information. Nowadays, mainly two kinds of imaging sensors are used for these applications: sensors with a high spatial resolution (e.g. SPOT 4, 1 pixel for $20m \times 20m$) but a monthly temporal acquisition frequency, and sensors with a medium or coarse spatial resolution (e.g. MERIS, 1 pixel for $300m \times 300m$, or SPOT-VGT, 1 pixel for $1km^2$) but daily or so temporal acquisition frequency. These sensors are multispectral and hence provide useful information for vegetation studies. One of the key challenge for automatic land cover classification is the combination of information from different resolutions to have both a high discrimination between land cover types and accurate spatial information.

In the remote sensing litterature, numbers of methods have been reported for image fusion. For example, during these last two decades, research studies have been devoted to the problem of reconstructing a HR image from multiple undersampled frames [1]. A possible approach, rather time consuming, consists in deriving multitemporal high spatial resolution (HR) series from coarse resolution (CR) one with a data fusion method (*e.g.* [2]), and then performing classification from this reconstructed HR series. Here, we directly solve the HR/CR classification problem.

In this paper, we propose a Bayesian approach for both supervised and unsupervised algorithm based on the linear mixture model.

Section 2 presents the classification method: having formulated the problem and stated the coarse resolution model, the classification model is deduced and the simulated annealing algorithm used for unsupervised as for supervised classification is described. As an illustration of the performance of the approach, some results obtained processing SPOT data are commented in Section 3 before Section 4 gathers the main lines of the paper as a conclusion.

2. CLASSIFICATION METHOD

The approach we propose here is based on the assumption that the scene geometry is stationary during the considered period. This is realistic for a time period such as one agricultural year. Spatial information can then be extracted at any time of the period *e.g.* processing a segmentation of one HR image. Classification will then be obtained by labellizing all segments according to time series information. We assume CR signal can be expressed according to the HR segmentation using the linear mixture model. In a Bayesian context, assuming the image signal follows a Gaussian law conditionnally to classes, we define an energy corresponding to the maximum *a posteriori* probability of a labellisation knowing the observation. Optimization process necessary to obtain the solution is performed using the simulated annealing technique.

2.1. Problem formulation

Classical Bayesian approach for HR classification assumes that the data image $u : \Omega \to \mathbb{R}$ is a realisation of a random field $\tilde{u}$ on domain Ω, that comes from a realisation $l : \Omega \to L$ of a random field $\tilde{l}$, where $L = \{1, ..., c, ...|L|\}$ is the set of all possible labels (related to the land cover types). Then, the classification problem consists in the estimation of the non observed realisation l of the field $\tilde{l}$ from the observed noisy data u, realisation of $\tilde{u}$. Assuming $\tilde{l}$ is a Markov Random Field, the prior probability of l can be expressed and the probability of u conditionally to l comes from the knowledge of the class statistics (*e.g.* Gaussian distribution). Therefore the *a posteriori* distribution $\mathbb{P}(\tilde{l} = l \mid \tilde{u} = u)$ can be expressed and the classification solution obtained using an optimization technique [3].

Whereas high resolution pixel interactions led to spatial regularity in the HR model, pixel spatial interactions cannot reasonnably be considered in the case of CR data such as SPOT-VGT or MERIS (such pixels represent a surface from $300m \times 300m$ to $1km \times 1km$). Moreover, in the case of agricultural areas, most spatial information is the geometry of the cadaster that can be considered as constant during the agricultural year. Consequently, assuming at least one HR image is available during the agricultural year, we propose to decompose the multi-resolution classification problem into two successive problems: the segmentation of the scene and its labellisation, which aims at giving to each segment a label corresponding to its type of vegetation (*e.g.* 1 for 'corn', 2 for 'forest',etc.). Since the segmentation problem we consider is rather classical (segmentation of one HR image), we refer to the Mumford and Shah variational method [4]. In the following, we focus on the labellization of each segment of the HR segmentation.

2.2. Coarse resolution model

Let Ω' be the CR domain and $v = (v^0, ..., v^{|T|})$ a CR time series. For all date t in the set of available dates T, assume the CR image $v^t : \Omega' \to \mathbb{R}$ is a realisation of a random field $\tilde{v}^t$. Assuming a CR image corresponds to the average of a *pseudo*- HR image, the *linear mixture model* yields that the esperance of the observation measurement performed over a mixed pixel is the weighted average of the observation measurement that could have been performed over *pure pixel* representing each land cover type. This, so called, *linear mixture model* has been widely used for extracting information from remotely sensed images containing mainly mixed pixels, as well for mixed pixel's proportion estimation [5] as for class feature estimation [6]. It has been validated for reflectance measurements, and is verified by definition for fractional cover measurement. The measurement observed in a CR pixel y can then be expressed as

$$v^t(y) = \frac{1}{N} \sum_{c \in L} \sum_{\substack{x \subset y \\ l_x = c}} u^t(x), \tag{1}$$

where x is a HR pixel ($x \subset y$ is the set of all HR pixels contained in the CR pixel y), l_x denotes the label of the HR pixel x and N is the number of HR pixels contained in a CR pixel (resolution ratio). Assume HR image u^t is the realisation of a Gaussian random field $\tilde{u}$ of mean μ_c^t and variance $(\sigma_c^t)^2$. A mixture of Gaussian laws being Gaussian, CR pixel values follow a Gaussian law of mean and variance depending on the class mixture within the CR pixel. More precisely, the value $v^t(y)$ observed in a pixel $y \in \Omega'$ at a date t is a realisation of the random variable $\tilde{v}^t(y) \sim \mathcal{N}(\mu_y^t, (\sigma_y^t)^2)$ where μ_y^t represents the pixel y's mean and $(\sigma_y^t)^2$ its variance. Let us model these parameters according to their class mixture. For each pixel y of the CR domain Ω', let us denote by $\alpha_c(y)$ the relative area in the CR pixel y occupied by the label c. That is, writing $N_c(y)$ the number of HR pixels labelled c within the CR pixel y, the proportion of class c within y is $\alpha_c(y) = N_c(y)/N$ and, by definition, $\sum_{c \in L} \alpha_c(y) = 1$. In this study, as labels are unknown, let us relate segment and class mixture. For each CR pixel y, let us denote by $\beta_k(y)$ the relative area of a segment k within y. The proportion of class c in a pixel y is equal to the sum of the proportions of all segments k labelled by c within y, *i.e*

$$\alpha_c(y) = \sum_{\substack{k \in S \\ l_k = c}} \beta_k(y) \ , \tag{2}$$

where S denotes the number of segments contained in the segmentation and l_k represents the label of the segment k.

For all date t and pixel y, the mean

$$\mu_y^t = \sum_{c \in L} \alpha_c(y) \mu_c^t = \sum_{c \in L} \sum_{\substack{k \in S \\ l_k = c}} \beta_k(y) \mu_c^t. \tag{3}$$

As the measurement observed in a CR pixel (1) can also be written

$$v^t(y) = \sum_{c \in L} \alpha_c(y) \sum_{\substack{x \subset y \\ l_x = c}} \frac{u^t(x)}{N_c(y)}, \tag{4}$$

the variance is computed using the variance expression for Gaussian mixture and hence, for each date t and pixel y, it writes

$$\begin{aligned} (\sigma_y^t)^2 &= \sum_{c \in L} \alpha_c^2(y) N_c(y) (\frac{1}{N_c(y)})^2 (\sigma_c^t)^2 && (5) \\ &= \frac{1}{N} \sum_{c \in L} \alpha_c(y) (\sigma_c^t)^2 && (6) \\ &= \frac{1}{N} \sum_{c \in L} \sum_{\substack{k \in S \\ l_k = c}} \beta_k(y) (\sigma_c^t)^2. && (7) \end{aligned}$$

Now the coarse resolution model is set up, let us define the *optimal* labellisation corresponding to an observation v.

2.3. labellisation model

Consider $\bar{l} = (\bar{l}_1, ..., \bar{l}_{|S|}) \in L^{|S|}$ the labels'vector of the segments indexed by $k = (1, ..., |S|)$. We are looking for the *optimal* labellisation of the random field $\tilde{l}$ knowing the coarse resolution observation v. According to Gaussian hypothesis and notations from section 2.2, the probability of observing a series v in the CR pixel y, conditionally to the labellisation l,

$$\mathbb{P}(v(y) \mid \tilde{l} = l) = \prod_{t \in T} \frac{1}{\sigma_y^t \sqrt{2\pi}} \exp\left(-\frac{(v^t(y) - \mu_y^t)^2}{2{\sigma_y^t}^2}\right), \tag{8}$$

with μ_y^t and σ_y^t respectively given by (3) and (5), assuming conditionnal independance of $\tilde{v}^t|\tilde{l}$. The *a posteriori* distribution $\mathbb{P}(\tilde{l} = l \mid \tilde{v}^t = v) = (\mathbb{P}(v^t|\tilde{l} = l)\mathbb{P}(l))/\mathbb{P}(v^t)$. Notice that $\mathbb{P}(v^t)$ is a constant for the considered problem, and the prior $\mathbb{P}(l)$ is also considered as constant (equiprobability of l) in absence of more information and spatial interaction modelling. Therefore, for a fixed date t, the maximum *a posteriori* is the labellisation vector $\bar{l}$ solution of the problem

$$\min_{\{l \in L^{|S|}\}} \sum_{y \in \Omega'} \left(\frac{\left(v^t(y) - \mu_y^t\right)^2}{(\sigma_y^t)^2} + \ln{(\sigma_y^t)^2} \right). \tag{9}$$

Considering time series, the mean vector writes $\mu_y = (\mu_y^t)_{t \in T}$ for all pixel y and, assuming all dates are independant, the covariance matrix Σ_y is diagonal with diagonal values $(\ (\sigma_y^t)^2\ )_{t \in T}$. Hence, (9) becomes

$$\min_{\{l \in L^{|S|}\}} \sum_{y \in \Omega'} \left({}^t(v(y) - \mu_y) \Sigma_y^{-1} (v(y) - \mu_y) + \ln\left(\det(\Sigma_y)\right)\right). \tag{10}$$

From this energy, we propose, on the one hand, a supervised algorithm leading to a solution of (10) and, on the other hand, an unsupervised algorithm solving a simplyfied problem. Indeed, if class features (mean and variance) are known *a priori*, a solution of (10) can be derived using a global optimization process. In practice, class features are generally unknown: they depend on the acquisition date, local vegetation growing process, the atmospheric conditions, etc. Therefore, unsupervised classification method can be more appropriate. However, if class means can be estimated during the minimization process (as explained in the following section), class variances estimation is much more sensitive and time-consuming. Hence, according to some robustness criteria, the unsupervised approach we propose assumes equal class variances rather than inaccurate estimated ones. This assumption leads, from (6), to constant variances for all pixel y and (10) simplifies as

$$\min_{\{l \in L^{|S|}\}} \sum_{y \in \Omega'} \sum_{t \in T} \left(v^t(y) - \mu_y^t \right)^2 . \tag{11}$$

An algorithm of simulated annealing type is described in the next section as an optimisation process for both the supervised and unsupervised method.

2.4. Algorithm

Because of the size of the solution space, a systematic search of the minimum is impossible. As no heuristic seems justified for this problem, we chose a simulated annealing algorithm. It has been widely used for different optimization problems [7]. As far as the unsupervised approach is concerned, the algorithm takes as inputs a HR segmentation, a CR time series of the same scene and the number of labels required. It returns the labellisation solution of (11) and the class features. For sake of simplicity, denote E_l the global energy to minimize in (11). It stands for the energy corresponding to a labellization l. Each step of the algorithm changes randomly one segment label in the labellisation and test whether it makes the energy decreasing or not. Denote $E_{l_{prev}}$ the energy correponding to the previous labellization. The algorithm is the following.

Compute proportions $(\beta_k(y))_{y,k}$ for all pixel y.
Initialize randomly the labellisation.
Estimate label's mean (linear regression).
Initialize temperature T_0 to the graph's diameter.
While a label is not rejected $n_r \times |S|$ times successively, do
  for $i = 0$ to $|S|$
    select randomly a segment k and a label c for this segment,
    re-estimate label's means,
    compute $\Delta E = E_l - E_{l_{prev}}$
    if $\Delta E \leq 0$ accept the label change,
    else reject it with a probability $\exp(-\Delta E / T)$.
  $T = (q)^n T_0$ where n is the number of done iterations.

Theroretically, the temperature descent must be logarithmic to lead to convergence but empirically, setting $n_r = 400$ and $q = 0.999$ provided good results in our experiments.
The supervised algorithm takes as inputs a HR segmentation, a CR time series of the same scene and the class features. It returns the labellization solution of (10). The algorithm is the same as the unsupervised one except for the label's mean estimation and re-estimation (since mean and variance are *a priori* known) and E_l now refers to the complete energy to minimize in (10). As the class feature estimation step is time-consuming, this version is more rapid than the unsupervised one. Let us now analyse the results we obtain in the case of simulated and actual data.

3. RESULTS

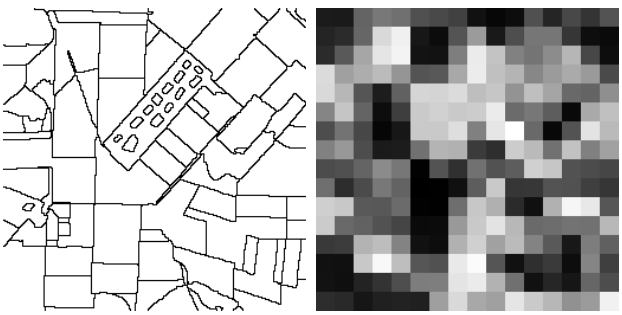
First, both methods have been validated using simulated data. Then, their performance on actual data have been checked.

3.1. Validation on simulated data

Simulations are made such that the two main image model assumptions are verifyied: linear mixture model for CR data and Gaussian distribution conditionally to classes. More precisely, from a HR classification of a 256×256 subpart of a SPOT/HRV image, actual class features are estimated. Then, these parameters are used to simulate HR images randomly according to Gaussian laws conditionally to classes. Then, CR simulated images result from the spatial average of HR simulated images at the coarse resolution considered. The quality of the obtained labellisation (from CR data) is measured by comparison to the labellisation obtained from HR data. This latter hence stands for the reference.
Obtained results of the supervised approach reach a performance upper than 99.5% of good classification for simulated CR data with a resolution ratio $|\Omega'|/|\Omega| = 16 \times 16$. Two main types of errors have been observed. On the one hand, tiny segments may be misclassified if their occupation rate within the CR pixel is to weak to have an effective contribution in the energy. On the other hand and unusually, connex segments can be misclassified because of simulated annealing failure to reach the global minimum. In this case, the algorithm is stuck in a local minimum, very close to the global minimum. For example, we encountered a case where three connex segments are misclassified because labellisation error correction for only one of them increases the energy and only exact labellisation simultaneously of all of them lets the energy decrease. We may improve the algorithm changing the temperature descent law for a logarithmic one but as this kind of errors is unusual we decided, so far, to keep the proposed temperature descent law as a compromise between time processing and result performance.
The unsupervised algorithm has been validated on the same simulated data. Notice that, theoretically, the unsupervised approach (minimization (11)) does not necessarily lead exactly to the same minimum with CR data than with HR data. However, in practice they are very close and results obtained are compared to the HR reference. From simulated CR data, the unsupervised approach reaches 99% of good classification (comparing to HR classification), that is almost as good as the supervised approach. A few more little segments are misclassified, as the energy considered only takes into account segment's means. Indeed, class variances have been supposed equal whereas one class has been simulated with a much smaller variance.

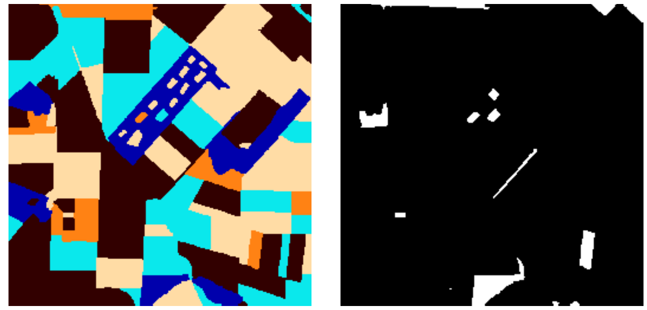
3.2. Application to actual data

The proposed approach has then been applied to a subpart of an actual SPOT/HRV time series of 8 images provided by the CNES agency in the framework of the European ADAM project. This time series has been pre-processed with an algorithm derived from [8], leading to land cover fraction series. This parameter is linear and competitive for land cover discrimination. The Mumford and Shah segmentation method has been processed on one HR image Fig. 1a. CR MERIS data have been simulated from the HR time series (see Fig. 1b), by averaging of a factor 16 in each direction (line and column). Fig. 2 shows the unsupervised result obtained with 5 classes. Fig. 2c presents, for each class, the mean estimation from CR data and from the reference HR data. We notice, in a satisfying way, that these estimated class means are very close, showing unsupervised means estimation from CR data is rather accurate. CR resulting labelisation is compared to the result obtained with HR series: the classification from CR series Fig. 2a is up to 97% identical to the one obtained from HR series. The error map presented on Fig. 2b shows that additional errors may be larger than the size of the CR pixel. This happens essentially when a segment covers several CR pixels with small proportions on each CR intersected pixel and a mean very close to its neighbours one. The energy of the error labellisation is hence very close to the reference one, but still minimum. We also note that, in some very few cases, whole CR pixels are misclassified (see the triangular error segment at the bottom of the map 2b). In particular, this may happen if this pixel's value belongs to the distribution queue. It seems that the use of larger multitemporal series could overcome such problems.

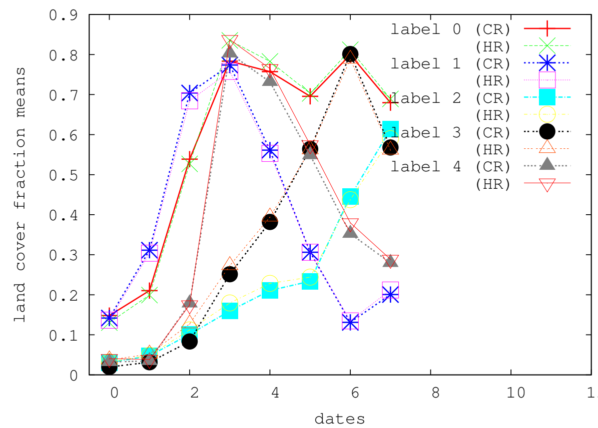


(a) HR Segmentation (b) Input CR image

Fig. 1. Inputs of the algorithm: on the left, the HR segmentation (100 segments) and on the right, a CR image (simulated MERIS) extracted from the time series.



(a) Classification (b) Errors



(c) Estimated means

Fig. 2. Outputs of the unsupervised algorithm: (a) the HR labellisation obtained from a CR time series for 5 labels, (b) the errors map compared to the labellisation obtained from HR time series (about 3% of misclassification) and (c) estimated class means represented as a time function (CR time series date 0 to 7).

4. CONCLUSION

In this paper, we proposed a Bayesian method for high resolution classification from coarse resolution time series. In practice, two algorithms have been described. Both provide a HR classification from CR time series. The first one, supervised, leads to the optimal energy but is quite sensitive to class variance estimation. The second one is totally unsupervised but the class number.

As the obtained results are convincing, future work will deal more performance analysis according to different parameters (resolution data, number of dates, etc.). Moreover, the validity of the model's assumptions will be studied in the case of different types of actual data.

5. REFERENCES

[1] R. Molina, M. Vega, J. Abad, and A.K. Katsaggelos, "Parameter estimation in bayesian high-resolution image reconstruction with multisensors," *IEEE Trans. on Image Processing*, vol. 12, no. 12, pp. 1655–1667, 2003.

[2] G. Flouzat, O. Amram, F. Laporterie, and S. Cherchali, "Multiresolution analysis and reconstruction by a morphological pyramid in the remote sensing of terrestrial surfaces," *Signal Processing*, vol. 81, no. 10, pp. 2171–2185, October 2001.

[3] S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution and bayesian restoration of images," *IEEE Trans. on Pattern An. and Mach. Intell.*, vol. 6, no. 6, pp. 721–741, 1984.

[4] J-M. Morel and S. Solimini, *Variational methods in image segmentation*, Birkhauser, 1995.

[5] H. Cardot, R. Faivre, and M. Goulard, "Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data.," *J. of Applied Stat.*, vol. 30, pp. 1185–1199, 2003.

[6] R. Faivre and A. Fischer, "Predicting crop reflectances using satellite data observing mixed pixels," *J. of Agr. Bio. and Env. Stat.*, vol. 2, pp. 87–107, 1997.

[7] P.J. van Laarhoven and E.H. Aarts, *Simulated Annealing: Theory and Applications*, Springer, 1987.

[8] S. Jacquemoud and F. Baret, "A model of leaf optical properties spectra," *Remote sensing of Environment*, vol. 34, pp. 75–91, 1990.

Acknowledgments

We thank the Centre National d'Etudes Spatiales (CNES) for providing the ADAM data base. A. Robin is funded by a grant from the CNES and Astrium/EADS.