

A High-Reliability, High-Resolution Method for Land Cover Classification into Forest and Non-forest

Roger Trias-Sanz^{1,2} and Didier Boldo¹

¹ Institut Géographique National,
2/4, avenue Pasteur, 94165 Saint-Mandé, France
Roger.Trias.Sanz@ieee.org

² SIP-CRIP5, Université René Descartes,
45, rue des Saints-Pères, 75006 Paris, France

Abstract. We present several methods for per-region land-cover classification based on distances on probability distributions and whole-region probabilities. We present results on using this method for locating forest areas in high-resolution aerial images with very high reliability, achieving more than 95% accuracy, using raw radiometric channels as well as derived color and texture features. Region boundaries are obtained from a multi-scale hierarchical segmentation or from a registration of cadastral maps.

1 Introduction

Several research projects are underway at the French national mapping agency (IGN) which attempt to partially or fully automate the production of fine-scale topographical maps. As part of this research effort, the goal of the Salade project is to produce a high-reliability land cover classification of rural areas in France, using high-resolution (50cm) orthorectified digital imagery with color and near-infrared channels, and cadastral maps.

The goal is to be able to fully automate interpretation of easy areas, and to use human interpreters only in complicated regions. Thus, considered classification algorithms must also produce a confidence map which indicates the algorithm's confidence in the produced classification, for each region. The classification must be very accurate where the algorithm gives high confidence values, but may be incorrect in some areas as long as the algorithm shows its lack of confidence for these areas. As an intermediate result in this research, we present in this paper several methods for classifying pixels into forest and non-forest in high-resolution aerial images, with very high reliability.

Typical classification systems classify each pixel separately using classifiers such as k -nearest neighbors, neural networks, or maximum a posteriori (MAP) classification. The raw radiometry vector for each pixel is used as the input for the classifiers. The class map—the output, which gives the class for each image pixel—is usually noisy. In some cases, the class map is geometrically regularized using ad hoc methods, median filtering, or probabilistic relaxation [1, 2] in order to reduce noise. Other algorithms incorporate contextual constraints in the classification process itself by means of Markov Random

Fields [3, 4], and are notoriously slow. If an a priori segmentation of the image into homogeneous regions is available, per-region classification algorithms exist that give the majority class for all pixels in a region [5, 6].

The algorithms presented here are of the per-region kind. Regions are obtained either from the registration of geometrically imprecise cadastral maps onto the images [7, 8], or from a fine image segmentation if cadastre data is not available. They are assumed to be homogeneous, that is to say, they contain only one land cover class. The probability distribution for the input vectors within each region is compared to the distributions for each terrain class created in a training phase, either using a distance on probability distributions or by a probabilistic approach. That distance or probability is itself used as a measure of the algorithm's confidence in the classification.

By using per-region classification we avoid the noise problem of per-pixel classifiers. Availability of prior region data—in the form of registered cadastral maps—means that we do not introduce any artificial regularity as may be the case with MRF-based classifiers or probabilistic relaxation regularization.

In the remainder of this article we present the training procedure (section 2.1) and the classification algorithms (section 2.2). In section 3 we show a thorough evaluation of each algorithm on two real test areas.

2 Algorithm

Classification involves three steps. First, a partition of the image to be classified is obtained. The cadastre, which partitions terrain into cadastre regions, can be used, but it should be registered to the image (see [7] or [8]) because edges of cadastre regions are geometrically imprecise and do not exactly correspond to actual edges in a land cover class map. Alternatively, a partition can be obtained by simply running a segmentation algorithm configured to give a segmentation with small regions; there is a risk of obtaining biased results in the classification confidence measures if the segmentation uses the same image channels as the classification, which is why using the cadastre is a better option. If a cadastre region is not homogeneous, but contains more than one land cover class, confidence will be very low for whatever single class is given for the region, and the problem may be detected.

Next, probability models are estimated for each terrain class. This involves manually defining, in a training image, a set of polygons for each terrain class, constructing a histogram for the radiometries—or other input features such as texture or derived color channels—of pixels in a class, selecting a probability model for each class, and estimating the parameters for these models, using the radiometry histograms. Generalization occurs at this point, by preferring a worse-fitting but simpler model over a more complex, better-fitting one. This is discussed in section 2.1.

Finally, each region in the image to be classified is processed in turn. The histogram of the region is constructed, and compared to the probability models for each terrain class. The most probable class, or that with the probability model closest to the region's histogram, is selected for that region. This is discussed in section 2.2.

2.1 Estimation

Let $X = (x_1, \dots, x_N)$ be the sequence of values of all N training pixels of a certain land cover class, in arbitrary order. For a d -channel image, $x_i \in D \subset \mathbb{R}^d$. Let x_{ij} denote the j -th component of x_i . Let $\bar{\mu}$ be the sample mean of X , $\bar{\mathbf{S}}$ its sample covariance matrix, \bar{s}_{nm} be the elements of $\bar{\mathbf{S}}$, and $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ the set of terrain classes in our problem. In this step the parameters for several probability models are estimated by from the sample X . In addition, a prior probability for class ω_i , $p(\omega_i)$, is calculated as the ratio of training pixels of that class to the total number of training pixels —region size is assumed not to depend on terrain class. The model which best balances complexity and quality of fit to the data is selected for each terrain class.

Model Estimation. In the implementation used for the evaluation of section 3, we provide, in addition to a standard d -dimensional Gaussian model, the following probability models. Specific applications may require the definition of additional models.

Laplacian Random Variable. A 1-dimensional Laplacian random variable V of parameters λ and m has probability density function $v(x) = \frac{1}{2\lambda}e^{|x-m|/\lambda}$. The higher-dimensional version is not easily found in the literature; we can define one as follows: Let $A = (A_1, \dots, A_d)$ be a vector of independent 1-dimensional Laplacian random variables of mean zero ($m = 0$) and $\lambda = 1$ (and therefore variance = 2). The random variable $W = \mathbf{M} \cdot A + \mu$, where \mathbf{M} is an invertible $d \times d$ matrix and $\mu \in \mathbb{R}^d$, is a d -dimensional Laplacian random variable. Its probability density function $w(x)$ is

$$w(x) = \frac{1}{2^d |\det \mathbf{M}|} \cdot e^{-\|\mathbf{M}^{-1}(x-\mu)\|_1}, \tag{1}$$

where $\|\cdot\|_1$ is the L_1 norm, $\|(x_1, x_2, \dots, x_n)\|_1 = \sum_{i=1}^n |x_i|$. Its mean is $E(W) = \mu$ and its covariance matrix is $\text{cov } W = 2 \cdot \mathbf{M} \cdot \mathbf{M}^T$, where \mathbf{M}^T denotes matrix transposition. Estimates $\hat{\mu}$ and $\hat{\mathbf{M}}$ for its parameters μ and \mathbf{M} can be calculated from the sample’s mean $\bar{\mu}$ and covariance $\bar{\mathbf{S}}$ —using a Cholesky decomposition— with

$$\hat{\mu} = \bar{\mu}, \quad 2 \cdot \hat{\mathbf{M}} \cdot \hat{\mathbf{M}}^T = \bar{\mathbf{S}}. \tag{2}$$

Rectangular Uniform Random Variable. A d -dimensional random variable V with constant probability over a rectangular support $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d] \subset \mathbb{R}^d$, and zero elsewhere. Its a and b parameters are related to its mean and covariance as $E(V_j) = (b_j - a_j)/2$ and $E((V_j - E(V_j))^2) = ((b_j - a_j)^2)/12$, and we can therefore give estimates \hat{a} and \hat{b} for parameters a and b from the sample X as

$$\hat{a} = \bar{\mu} - w/2, \quad \hat{b} = \bar{\mu} + w/2, \quad \text{where } w_i = \sqrt{12 \bar{s}_{ii}}, \quad w = (w_1, \dots, w_d). \tag{3}$$

Raw Non-parametric Model. We can also estimate a “non-parametric” random variable whose probability density function follows the distribution histogram of the sample X . Let’s partition the data domain D into equally-shaped histogram bins, $\{D_i\}_i$, with $\cup_i D_i = D$ and $\forall i \neq j. D_i \cap D_j = \emptyset$. Let’s construct a relative frequency histogram h from X by counting the number of values in X that belong to the i -th histogram bin,

for each i : $h_i = \text{card}\{j : x_j \in D_i\}/N$. We have $\sum_i h_i = 1$. We can define a random variable V with values in D with probability density function $v(x)$

$$v(x) = h_i \quad \text{with } i \text{ such that } x \in D_i. \tag{4}$$

In this implementation, several such models are estimated for different partitions of D .

Kernel Density Estimation. With raw non-parametric estimation, even if the underlying density is known to be smooth, a large sample size is necessary in order to obtain an estimation which has both a high number of histogram bins and a high count in each histogram bin. *Kernel density estimation* [9, 10] can be used to overcome this problem, by convolving a histogram h_i with many bins by a suitable d -dimensional kernel—in this implementation, Gaussian kernels of different variances.

Model Selection. Finally, one probability model must be selected from those estimated in the previous step. We cannot simply select the model that best fits the sample, because this will tend to select models with a high number of degrees of freedom, and cause *overfitting*. To avoid this we will select, for each land cover class, the estimated model with the lowest value of the *Bayes Information Criterion*, or BIC [11], a measure which balances the fit to the data and a measure of the model complexity, defined as

$$BIC = -2 \ln p(X | M) + p \ln N, \tag{5}$$

where $p(X | M)$ is the probability of the observations given the model, p is the number of parameters in the model (the complexity), and N is the number of observations.

2.2 Classification

Let $X = (x_1, \dots, x_N)$ be the sequence of values of all N pixels in a certain region to be classified, taken in arbitrary order. Using the estimation procedures in the last step, for each class ω_i we have an estimated random variable V_{ω_i} with probability density function v_{ω_i} . The goal of this step is to assign a class to the region X , $\kappa(X) \in \Omega$, and to obtain a measure of confidence for that classification, $c(X)$, which should be higher for more confident decisions. In this we assume that the pixel values in a region depend only on the class of that region but not on the class of other regions.

Classification by Distribution Distance. Let's partition the data domain D into equally-shaped histogram bins, $\{D_i\}_i$, with $\cup_i D_i = D$ and $\forall i \neq j. D_i \cap D_j = \emptyset$. Let's construct a relative histogram h from X by counting the number of values in X that belong to the i -th histogram bin, for each i : $h_i = \text{card}\{j : x_j \in D_i\}/N$. In our implementation, d -dimensional square bins, each of side 8, were chosen.

Distribution distance-based classification algorithms depend on the definition of $d(X, V)$, a dissimilarity measure between a distribution X and a random variable V (not necessarily a distance in the mathematical sense). For a given dissimilarity, the class for region X is the one closest to it, and we propose to use as confidence measure the dissimilarity measure itself:

$$\kappa(X) = \operatorname{argmin}_{\omega \in \Omega} d(X, V_\omega), \quad c(X) = - \min_{\omega \in \Omega} d(X, V_\omega). \quad (6)$$

One such measure is the Kullback-Leibler divergence, also called the relative entropy. The Kullback-Leibler divergence d between a sample X and a random variable V of probability density function v is

$$d(X, V) = kl(X, V) = 2 \sum_i (h_i \log h_i - h_i \log p(V \in D_i)), \quad (7)$$

where, of course, $p(V \in D_i) = \int_{z \in D_i} v(z) dz$. Another measure is the χ^2 statistic, commonly used in the χ^2 test [12],

$$d(X, V) = \chi^2(X, V) = N \sum_i \frac{(h_i - p(V \in D_i))^2}{p(V \in D_i)}, \quad (8)$$

omitting from the sum any term with $h_i = p(V \in D_i) = 0$. However, the χ^2 statistic depends linearly on N , the number of pixels in the region. While this does not affect the choice of class for the region, it may affect the confidence measures. If we refuse to classify regions where the confidence measure falls under a certain threshold, we might keep an incorrect classification for a small region while rejecting a correct classification for a large region. The problem is avoided by using this confidence measure instead:

$$c(X) = - \min_{\omega \in \Omega} \frac{1}{N} \chi^2(X, V_\omega). \quad (9)$$

Classification by Per-region Probability. Let us further assume that, in a region, the value for a pixel is independent from the values of other pixels—as in standard per-pixel classification; existing dependencies can be captured through the use of texture features. We can calculate the probability of all pixels belonging to a certain class, and choose the class giving maximum probability. From a maximum a posteriori approach (MAP), this gives

$$\begin{aligned} \kappa_{\text{MAP}}(X) &= \operatorname{argmax}_{\omega \in \Omega} p(\omega|X) = \operatorname{argmax}_{\omega \in \Omega} \frac{p(X|\omega) p(\omega)}{p(X)} = \operatorname{argmax}_{\omega \in \Omega} p(X|\omega) p(\omega) \\ &= \operatorname{argmax}_{\omega \in \Omega} p(\omega) \prod_{n=1}^N p(x_n|\omega) = \operatorname{argmax}_{\omega \in \Omega} p(\omega) \prod_{n=1}^N v_\omega(x_n). \end{aligned} \quad (10)$$

There is open debate on whether MAP or maximum likelihood (ML) approaches should be used. Briefly, MAP gives less probability to classes found rarely in the training data. ML gives equal chances to all classes, which is not optimal for risk minimization but may be desired in some cases. From an ML approach we obtain the classification

$$\kappa_{\text{ML}}(X) = \operatorname{argmax}_{\omega \in \Omega} p(X|\omega) = \operatorname{argmax}_{\omega \in \Omega} \prod_{n=1}^N p(V_\omega = x_n) = \operatorname{argmax}_{\omega \in \Omega} \prod_{n=1}^N v_\omega(x_n). \quad (11)$$

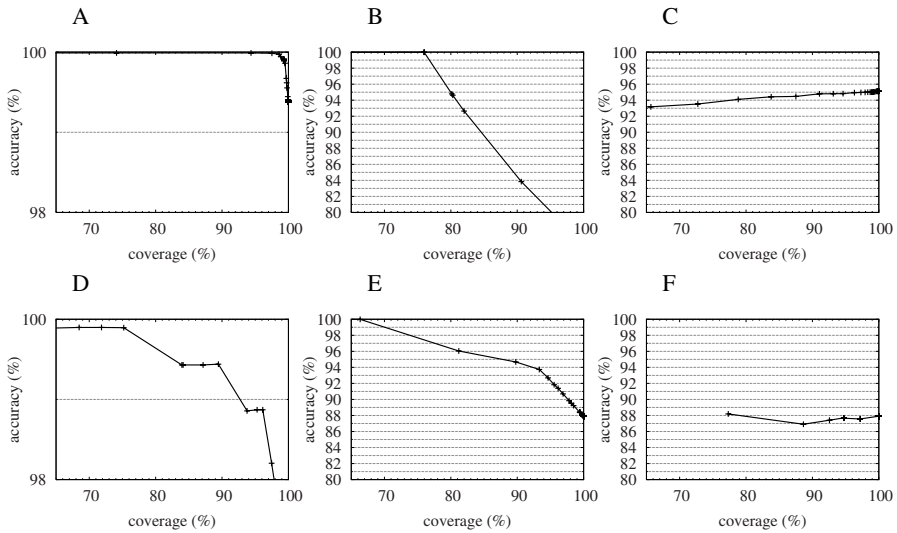


Fig. 1. Several graphs of classification accuracy as a function of coverage showing excellent (A, D), correct (B, E), and undesired behaviors (C, F). A: Saint-Leger, segmentation regions, $MAP^{1/N}$, “forest” class. B: Toulouse, cadastre regions, $MAP^{1/N}$, “forest” class. C: Toulouse, segmentation, $MAP^{1/N}$, global accuracy. D: Saint-Leger, cadastre, KL , global accuracy. E: Toulouse, segmentation, $MAP^{1/N}$, “forest” class. F: Toulouse, segmentation, $MAP^{1/N}$, “forest” class, using $c(X) = \max_{\omega \in \Omega} p(\omega) \prod_{n=1}^N v_{\omega}(x_n)$ instead

We propose to use as confidence measures

$$c_{MAP}(X) = \max_{\omega \in \Omega} p(\omega) \left(\prod_{n=1}^N v_{\omega}(x_n) \right)^{\frac{1}{N}}, \quad c_{ML}(X) = \max_{\omega \in \Omega} \left(\prod_{n=1}^N v_{\omega}(x_n) \right)^{\frac{1}{N}}. \quad (12)$$

where, as with equations 8 and 9, the $1/N$ exponent makes the measure independent of region size (compare Fig. 1 E, with the $1/N$, with Fig. 1 F, without).

3 Evaluation

We have implemented the classification algorithms described in section 2.2, and the estimation methods described in section 2.1. Using these, we present results for two sites. In the first, *Saint-Leger*, we use aerial images at 80cm resolution, resampled to 50cm, with red, green, and blue channels. In the second, *Toulouse*, we use digital aerial images at 20cm resolution, also resampled to 50cm, with red, green, blue, and near-infrared channels (the images before resampling were not available to us). Training and evaluation polygons are defined on both sites. Classification is not only performed on the raw radiometric channels but also on derived color and textural channels—the latter because few channels are available but spatial resolution is very high: Following suggestions in [13], for Saint-Leger we use the raw green channel, the second Karhunen-Loève transformed channel (from [14], $k_2 = (red - blue)/2$), and the local proportion

of pixels whose gradient module is over a threshold [15]. For Toulouse we use the green channel, the same gradient-based measure, and the normalized difference vegetation index (NDVI).

Using the training polygons, probability models are obtained, for each site, for terrain classes “field”, “forest”, “road”, “bare soil”, “orchard”, “quarry”, “uncultivated field”, “other vegetation”, “water”, “sand”, “vineyard”, and “shrubs”. For each class, several models are computed as described in section 2.1 and the best fit—according to the Bayes Information Criterion—is selected. Regions are obtained either from a fine segmentation of the images [16] or by registering geometrically imprecise cadastral maps onto the images [7, 8]. Per-region classification is then performed using the methods described in section 2.2. Finally, all classes except the “forest” class are merged into a single “non-forest” class. This is done after classification, and not at the modeling stage, because non-forest classes have quite distinct signatures which would be more difficult to model jointly.

For baseline comparison, we have also implemented a per-pixel MAP classifier, a per-pixel ML classifier, and a per-region classifier based on majority vote [5] from per-pixel MAP:

$$\kappa_{\text{MAP}}(\{x\}) = \operatorname{argmax}_{\omega \in \Omega} p(\omega)v_{\omega}(x), \tag{per-pixel MAP}, \tag{13}$$

$$\kappa_{\text{ML}}(\{x\}) = \operatorname{argmax}_{\omega \in \Omega} v_{\omega}(x), \tag{per-pixel ML}, \tag{14}$$

$$\kappa_{\text{maj-MAP}}(X) = \operatorname{argmax}_{\omega \in \Omega} \operatorname{card}\{x_i \in X : \kappa_{\text{MAP}}(x_i) = \omega\}, \tag{majority MAP}. \tag{15}$$

Evaluation results are summarized in table 1. This shows the ratio of correctly classified pixels at 100% coverage, —accepting all classifications regardless of their confidence value— for the Saint-Leger and Toulouse sites using the presented classification algorithms, and using either an image oversegmentation or registered cadastre data as input regions to the per-region classifiers. The Saint-Leger test site has an area of 5.7 km². The Toulouse test site has an area of 37.6 km². However, since only 7.0 km² of cadastre data is available for the Toulouse site, results for that site using registered cadastre edges as regions are less significant; results under “Toulouse cadastre” in table 1 correspond to the

Table 1. Classification accuracies —correctly classified pixels in the evaluation set— at full coverage

Method:	Test site:	Saint-Leger		Toulouse	
	Regions	segmentation	cadastre	segmentation	cadastre
KL (eq. 7)		99.4%	97.0%	93.5%	85.7%
χ^2 (eq. 8)		99.4%	97.6%	87.6%	63.3%
χ^2/N (eq. 9)		99.4%	97.6%	87.6%	63.1%
MAP ^{1/N} (eqs. 10 and 12)		99.4%	97.0%	95.1%	90.1%
ML ^{1/N} (eqs. 11 and 12)		99.4%	97.0%	95.0%	90.1%
Per-pixel MAP (eq. 13)		98.2%	98.2%	92.8%	92.0%
Per-pixel ML (eq. 14)		97.6%	97.3%	90.0%	86.8%
Majority vote MAP (eq. 15)		99.4%	97.8%	93.1%	84.8%

subset of the Toulouse site where cadastre data is available. Using regions extracted from cadastral maps instead of an over-segmentation of the source image gives slightly worse accuracy, but output regions tend to have higher geometrical quality. Execution time is approximately 40 s/km² on a 2.4GHz Intel Pentium 4 processor for a single algorithm.

For the Saint-Leger site, all algorithms except per-pixel MAP and ML achieve better than 99% pixel-wise accuracy —accuracy is defined as the ratio of correctly classified pixels in the evaluation set over the total number of pixels in the evaluation set. For the much more diverse and complex Toulouse site, the best algorithm, MAP^{1/N}, achieves an accuracy of over 95%; this is slightly better than the per-pixel MAP and ML methods, and has the advantage of not having the dotted noise typical of per-pixel classification. It is also slightly better than a majority vote of a per-pixel MAP classification. For comparison, recent papers —which often use per-pixel classification on images of lower spatial resolution but higher spectral resolution— report accuracies around 85%–90% [17, 18, 19].

In some cases, accuracy increases with decreasing coverage —as a larger fraction of the least-confidently-classified regions is rejected, accuracy in the accepted regions increases. This indicates meaningful confidence values. In that case we can obtain higher accuracies at the cost of decreased coverage. In other cases, however, the calculated confidence measure appears not to be indicative of the quality of classification. In Fig. 1 we show several confidence measure graphs, some showing the desired behavior (increasing accuracy as coverage decreases) and some not. For example, in the Toulouse

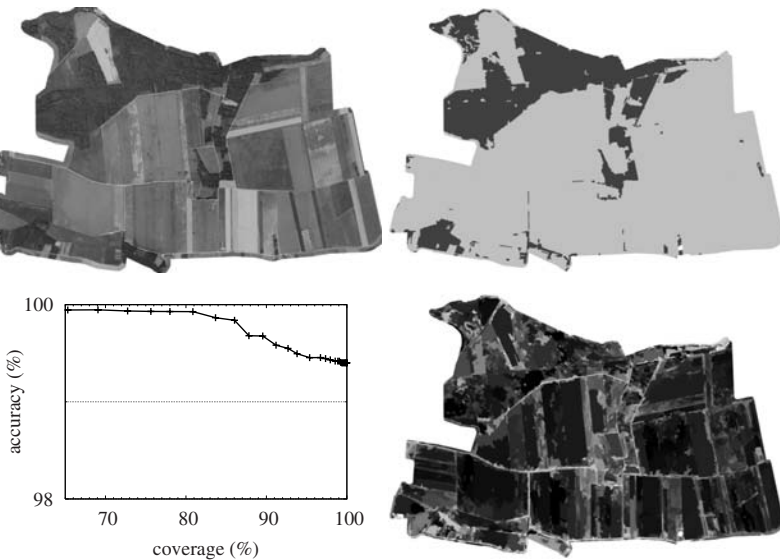


Fig. 2. Saint-Leger site, classified using the χ^2/N method (eq. 9) and segmentation regions. Top left: input image. Top right: class map (white: unclassified; dark gray: forest; light gray: non-forest). Bottom right: confidence values (darker is higher confidence). Bottom left: classification accuracy (% correctly classified pixels) as a function of coverage (% classified pixels, depending on a threshold on the classification confidence value)

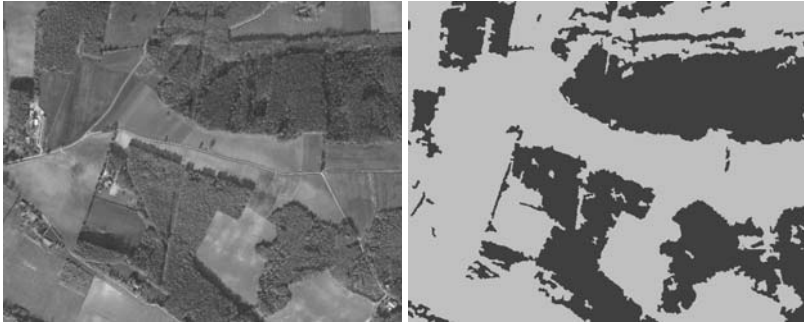


Fig. 3. A portion of the Toulouse site, classified using the $\text{MAP}^{1/N}$ method (eqs. 10 and 12) and segmentation regions. Left: input image. Right: class map (white: unclassified; dark gray: forest; light gray: non-forest). Graphs of accuracy versus coverage are given in Fig. 1 C and E

site, using regions obtained from an over-segmentation, and the $\text{MAP}^{1/N}$, the accuracy of the “forest” class can be increased from 87.9% at 100% coverage to 96.0% at 81% coverage (Fig. 1 E).

Figures 2 and 3 show, for the Saint-Leger site and a portion of the Toulouse site respectively, the source image and the class map using segmentation-derived regions and the χ^2/N and $\text{MAP}^{1/N}$ algorithms. For Saint-Leger, we also show the classification accuracy as a function of coverage, and the confidence map; accuracy indeed increases with decreasing coverage, indicating that the confidence measure is meaningful there. For the Toulouse site, global confidence measures —shown in Fig. 1 C— are not meaningful and cannot be used to obtain higher accuracy at lower coverage; confidence classes for certain classes, however, behave correctly —such as that for the “forest” class shown in Fig. 1 E.

4 Conclusion

We have presented several methods for land-cover classification. To avoid the noise usually associated with per-pixel classification, we chose per-region algorithms. However, in contrast to the more common per-region classification method, which involves a per-pixel classification followed by a majority vote, our methods use distances on probability distributions and whole-region probabilities.

These methods are evaluated for locating forest areas in high-resolution color and color-infrared aerial images with very high reliability, achieving more than 95% accuracy. Because few channels are available, but spatial resolution is high, we use texture features in addition to raw radiometry and derived color channels as classification input. Region boundaries are obtained from a multi-scale hierarchical segmentation or from a registration of cadastral maps —the latter gives lower classification accuracies but better spatial precision. When using segmentation regions, the presented $\text{MAP}^{1/N}$ method outperforms the per-pixel baseline methods also evaluated. We obtain slightly more accurate classifications than other studies, while, thanks to the use of textural features, being able to use higher spatial resolution images —and therefore having much improved spatial precision.

In addition, we have proposed several confidence measures that should indicate those classifications the classifier is unsure about; this would allow a trade-off between classification accuracy and coverage —leaving some areas unclassified, to be processed by a human operator, in exchange for higher accuracy in the classified regions. Evaluation shows that these confidence measures behave as expected for some cases but not all: further research is needed on the subject. Additional research into better-fitting probability models and into optimal texture features and color transformations is also desirable. Good results in these areas may make it possible to obtain similar accuracies on classifications with a richer set of land cover classes.

References

1. Rosenfeld, A., Hummel, R., Zucker, S.: Scene labeling by relaxation operations. *IEEE Trans. on Systems, Man, and Cybernetics* **6** (1976) 320–433
2. Wang, J.P.: Stochastic relaxation on partitions with connected components and its application to image segmentation. *IEEE Trans. on PAMI* **20** (1998) 619–636
3. Deng, H., Clausi, D.A.: Unsupervised image segmentation using a simple MRF model with a new implementation scheme. In: *Proc. 17th ICPR, Cambridge, UK, IAPR (2004)*
4. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on PAMI* **6** (1984) 721–741
5. Aplin, P., Atkinson, P.M.: Predicting missing field boundaries to increase per-field classification accuracy. *Photogr. Eng. and Remote Sensing* **70** (2004) 141–149
6. De Wit, A.J.W., Clevers, J.G.P.W.: Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing* **25** (2004) 4091–4112
7. Trias-Sanz, R.: An edge-based method for registering a graph onto an image with application to cadastre registration. In: *Proc. of ACIVS 2004, Brussels, Belgium (2004)* 333–340
8. Trias-Sanz, R., Pierrot-Deseilligny, M.: A region-based method for graph to image registration with application to cadastre data. In: *Proc. ICIP, Singapore, IEEE (2004)*
9. Rosenblatt, M.: Remarks on some nonparametric estimates of a density functions. *Annals of Mathematical Statistics* **27** (1956) 642–669
10. Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, USA (1986)
11. Schwarz, G.: Estimating the dimension of a model. *Annals of Statistics* **6** (1978) 461–464
12. Cramér, H.: *Mathematical Methods of Statistics*. Princeton U. Press, Princeton, USA (1946)
13. Giffon, S.: *Classification grossière par radiométrie et texture en zones agricoles*. Master's thesis, Université Jean Monnet, Saint Étienne, France (2004) (In French).
14. Van de Wouwer, G., Scheunders, P., Livens, S., Van Dyck, D.: Wavelet correlation signatures for color texture characterization. *Pattern Recognition* **32** (1999) 443–451
15. Baillard, C.: *Analyse d'images aériennes stéréo pour la restitution 3-D en milieu urbain*. PhD thesis, École Nat. Sup. des Télécommunications, Paris, France (1997) (In French).
16. Guigues, L., Le Men, H., Cocquerez, J.P.: Scale-sets image analysis. In: *Proc. ICIP, Barcelona, Spain, IEEE (2003)*
17. Thomas, N., Hendrix, C., Congalton, R.G.: A comparison of urban mapping methods using high-resolution digital imagery. *Photogr. Eng. and Remote Sensing* **69** (2003) 963–972
18. Haapanen, R., Ek, A.R., Bauer, M.E., Finley, A.O.: Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment* **89** (2004) 265–271
19. Cablk, M.E., Minor, T.B.: Detecting and discriminating impervious cover with high-resolution IKONOS data using principal component analysis and morphological operators. *International Journal of Remote Sensing* **24** (2003) 4627–4645