A Bayesian Approach to Classification of Multiresolution Remote Sensing Data

Geir Storvik, Roger Fjørtoft, *Member, IEEE*, and Anne H. Schistad Solberg, *Member, IEEE*

***Abstract*—Several earth observation satellites acquire image bands with different spatial resolutions, e.g., a panchromatic band with high resolution and spectral bands with lower resolution. Likewise, we often face the problem of different resolutions when performing joint analysis of images acquired by different satellites. This paper presents models and methods for classification of multiresolution images. The approach is based on the concept of a *reference* resolution, corresponding to the highest resolution in the dataset. Prior knowledge about the spatial characteristics of the classes is specified through a Markov random field model at the reference resolution. Data at coarser scales are modeled as mixed pixels by relating the observations to the classes at the reference resolution. A Bayesian framework for classification based on this multiscale model is proposed. The classification is realized by an iterative conditional modes (ICM) algorithm. The parameter estimation can be based both on a training set and on pixels with unknown class. A computationally efficient scheme based on a combination of the ICM and the expectation–maximization algorithm is proposed. Results obtained on simulated and real satellite images are presented.**

***Index Terms*—Bayesian modeling, classification, expectation–maximization (EM) algorithm, iterative conditional mode (ICM), multiresolution data.**

I. Introduction

The rapidly growing number of earth observation satellites provide a much better coverage in space, time, and the electromagnetic spectrum than in the past. Analysis of compound datasets therefore steadily gains importance. One of the key challenges in multisensor image fusion is how to combine images with different resolution to obtain more precise results.

For all multisensor image analysis tasks, image coregistration is an important prerequisite. For multiresolution datasets, it seems necessary to require subpixel accuracy with respect to the finest resolution. However, we consider multiresolution image registration to be outside the scope of this study.

Many approaches to multisensor image analysis have been proposed, but few of them treat the multiresolution aspect explicitly. The common approach is to resample the different images to be fused to a common pixel resolution. In this paper, we present a multiscale model, which allows preserving the details found in the highest resolution image, while exploiting the spectral information at lower resolution.

The remote sensing literature contains many examples of multiresolution data visualization, e.g., by merging a high-resolution panchromatic band with lower resolution multispectral bands (e.g., see [1]). The purpose is visualization, and the multispectral image is overlaid on the panchromatic image using different colors. Wavelet models are common for this kind of application. However, little work has been done on true multiscale classification models.

In other domains of science, much work on combining data sources at different resolutions exists, e.g., in epidemiology [2], in estimation of hydraulic conductivity for characterizing groundwater flow [3], in estimation of environmental components [4]. These approaches are mainly for situations where the aim is to estimate an underlying *continuous* variable.

For classification problems, Puyou-Lascassies [5] and Zhukov *et al.* [6] considered unmixing of low-resolution data by using class information obtained from classification of high-resolution data. The unmixing is performed through several sequential steps, but no formal model for the complete dataset is derived. Price [1] proposed unmixing by relating the correlation between low-resolution data and high-resolution data resampled to low-resolution, to correlation between high-resolution data and low-resolution data resampled to high resolution. The possibility of mixed pixels was not taken into account.

In [7], separate classifications were performed based on data from each resolution. The resulting resolution-dependent probabilities were averaged over the resolutions.

Multiresolution tree models are sometimes used for multiscale analysis (e.g., see [8]). Such models yield a multiscale representation through a quad tree, in which each pixel at a given resolution is decomposed into four child pixels at higher resolution, which are correlated. This gives a model where the correlation between neighbor pixels depend on the pixel locations in an arbitrary (i.e., not problem-related) way.

The multiscale model presented in this document is based on the concept of a *reference resolution* and is developed in a Bayesian framework [9]. We let the reference resolution correspond to the highest resolution present in the dataset. For each pixel of the input image at the reference resolution, we assume that there is an underlying discrete class. The observed pixel values are modeled conditionally on the classes. The properties of the class label image are described through an *a priori* model. Markov random fields have been selected for this purpose. Data at coarser resolutions are modeled as mixed pixels,

Manuscript received March 15, 2004; revised October 17, 2004. This work was supported by the Norwegian Research Council in the framework of the EOtools Project under Grant 143569/431.

G. Storvik is with the Department of Mathematics, University of Oslo, 0314 Oslo, Norway and also with the Norwegian Computing Center, 0314 Oslo, Norway (e-mail: geirs@math.uio.no).

R. Fjørtoft was with the Norwegian Computing Center, 0314 Oslo, Norway. He is now with the French Space Agency CNES, DCT/SI/AR, 31401 Toulouse, France (e-mail: roger.fjortoft@cnes.fr).

A. H. Schistad Solberg is with the Department of Computer Science, University of Oslo, 0314 Oslo, Norway and also with the Norwegian Computing Center, 0314 Oslo, Norway (e-mail: anne@ifi.uio.no).

Digital Object Identifier 10.1109/TGRS.2004.841395

0196-2892/$20.00 © 2005 IEEE

i.e., the observations are allowed to include contributions from several distinct classes. In this way it is, for example, possible to exploit spectrally richer images at lower resolution to obtain more accurate classification results at the reference level, without smoothing the results as much as if we simply oversampled the low-resolution data to the reference resolution prior to the analysis. Another advantage of using mixed-pixel models is that we simultaneously obtain an "unmixing" of the low-resolution data to the reference resolution, which can also be used for other purposes. We have so far used multivariate Normal distributions to describe the data given the class. This distribution is perfectly suited for optical data, and it can be used as an approximation for multilook synthetic aperture radar (SAR) data. In principle, non-Normal distributions can be used as well, but the computational complexity will increase considerably.

The classification is based on maximization of the *a posteriori* probability (MAP) using the iterative conditional modes (ICM) algorithm [10]. The starting point is a maximum-likelihood (ML) classification based on the data at the reference level.

Parameter estimation for such models quickly becomes computationally demanding. In this paper, we have therefore considered a simple strategy based on an approach proposed by Besag [10], which is an EM-type algorithm switching between classification of unknown class memberships given the current estimated values of the parameters, and estimation of parameters given class memberships defined by the classification.

The results of single- and multiscale classification obtained on simulated and real satellite images are presented and discussed.

II. Model

We will consider a Bayesian framework where a prior spatial model is assumed for the class memberships on the reference resolution (Section II-A). Models for all observed images are constructed by relating them to the class members at the reference resolution (Section II-B).

A. Model for the Class Image

Define a *reference resolution* for which we have the following.

- Each pixel contains only one class.
- All observations are at this or a coarser resolution.

Assume that the class image is described by $\mathbf{z} = (z_1, \ldots, z_n)$ where z_i defines the class in pixel i at the reference resolution. Consider the model

$$p(\mathbf{z}|\boldsymbol{\theta}_1) = \frac{1}{C(\boldsymbol{\theta}_1)} \exp\left\{\sum_i \alpha_{z_i} - \beta \sum_{i \sim i'} \phi_{i,i'}(z_i, z_{i'})\right\} \quad (1)$$

where $\phi_{i,i'}(z_i, z_j)$ is a potential function modeling spatial interaction, $\boldsymbol{\theta}_1 = (\alpha_1, \ldots, \alpha_K, \beta)$, K is the number of classes, and $C(\boldsymbol{\theta}_1)$ is a normalization constant. We have here used the notation $i \sim i'$ to denote that pixels i and i' are neighbors. This is an ordinary Markov random field (MRF) model widely used for this purpose [10], [11]. We have in our experiments (Section V) used a variety of the Potts model on a four-neighborhood, where ϕ is equal to the number of neighboring pixels that are attributed to other classes than the central pixel, minus the number of neighboring pixels that belong to the same class as the central pixel, i.e., $\phi_{i,i'} = I(z_i \neq z_{i'}) - I(z_i = z_{i'})$.

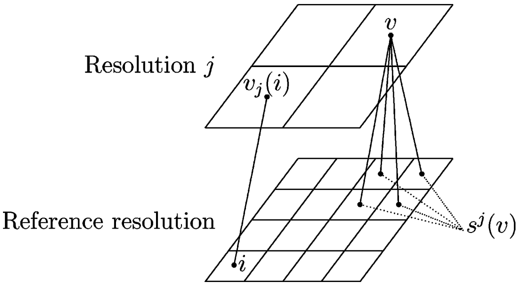


Fig. 1. Illustration of the definition of $s^j(v)$ and $v_j(i)$. For a pixel v at resolution j, lines are drawn down to each of the pixels at the reference resolution which are contained in $s^j(v)$. In this case, $m_j = 4$. For a pixel i at the reference resolution, a line is drawn up to pixel $v_j(i)$ at resolution j.

B. Model for Data

Assume that the observations $\mathbf{y} = (\mathbf{y}^1, \ldots, \mathbf{y}^p)$ are available where $\mathbf{y}^j$ is a (possibly multispectral) image at resolution j.

We will assume that data at different resolutions are conditionally independent, i.e.,

$$p(\mathbf{y}|\mathbf{z}) = \prod_{j=1}^{p} p(\mathbf{y}^j|\mathbf{z}).$$

This is a reasonable assumption for sensors with different spectral properties. A similar assumption is often used for fusion of multitemporal data, e.g., see [12].

In this paper, we will assume that the pixel dimensions at lower resolutions are entire multiples of the pixel dimensions at the reference resolution (level 1) and that the images are perfectly overlapping. We introduce the following notation:

$s^j(v)$ set of pixels at the reference resolution that are contained in pixel v at resolution j;
m^j number of pixels in $s^j(v)$;
$v_j(i)$ pixel at resolution j containing pixel i at the reference resolution.

These definitions are illustrated in Fig. 1.

Data at coarser resolutions are modeled as mixed pixels, i.e., we introduce hidden variables $\tilde{\mathbf{y}}_i^j$ at the reference level that sum up to the observed pixel values $\mathbf{y}_v^j$ at the coarser resolution level j

$$\mathbf{y}_v^j = \frac{1}{m^j} \sum_{i \in s^j(v)} \tilde{\mathbf{y}}_i^j. \quad (2)$$

We can consider $\tilde{\mathbf{y}}_i^j$ as the observation that *would* be obtained if a sensor existed that had the radiometric properties of the sensor at level j, but acquired images at the reference resolution (level 1). A similar assumption was made in, for example, [1] and [6].

We further assume that all $\tilde{\mathbf{y}}_i^j$ are conditionally independent and that

$$\tilde{\mathbf{y}}_i^j|\mathbf{z} \sim N\left(\boldsymbol{\mu}_{z_i}^j, \boldsymbol{\Sigma}_{z_i}^j\right) \quad (3)$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the Normal distribution with expectation vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

In cases where pixels at lower resolutions are not multiples of pixels at the reference resolution, (2) could be generalized to

$$\mathbf{y}_v^j = \frac{\sum_{i\in s^j(v)} a_i^j(v)\tilde{\mathbf{y}}_i^j}{\sum_{i\in s^j(v)} a_i^j(v)} \tag{4}$$

where $a_i^j(v)$ is the size of pixel i at the reference resolution contained in v. Although the framework discussed in this paper will work also in this more general setting, only the simpler situation will be considered.

In the following, we will use

$$\boldsymbol{\theta}_2 = \left\{\left(\boldsymbol{\mu}_k^j, \boldsymbol{\Sigma}_k^j\right),\ j=1,\ldots,p,\ k=1,\ldots,K\right\}$$

i.e., the set of parameters involved in the likelihood functions.

III. Classification

Assuming that the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are known, classification is based on the posterior distribution

$$p(\mathbf{z}|\mathbf{y};\boldsymbol{\theta}) \propto p(\mathbf{z}|\boldsymbol{\theta}_1)\prod_{j=1}^{p} p(\mathbf{y}^j|\mathbf{z};\boldsymbol{\theta}_2). \tag{5}$$

The maximum *a posteriori* estimate is obtained by maximizing (5) with respect to the complete vector $\mathbf{z}$. Explicit MAP solutions have been obtained in [13] for problems with all observations given at the same resolution. Although such algorithms (based on integer programming and Lagrangian-based methods) would be possible also to develop for the problem at hand, we will here consider a simpler version based on the ICM algorithm [10].

The ICM algorithm consists of sequential optimization of the components of $\mathbf{z}$ by maximizing $p(z_i|z_{i'}, i' \neq i, \mathbf{y})$. This quantity will only depend on observations from pixels at all resolutions containing pixel i at the reference resolution. Recall that $v_j(i)$ is defined as the pixel at resolution j containing pixel i at the reference resolution. Using (1) and (5), we have

$$p(z_i|z_{i'}, i'\neq i, \mathbf{y}) \propto \exp\left\{\alpha_{z_i} - \beta\sum_{i':i'\sim i}\phi_{i,i'}(z_i, z_{i'})\right\}\prod_{j=1}^{p} p\left(\mathbf{y}_{v_j(i)}^j|\mathbf{z}\right). \tag{6}$$

The ICM steps consist in calculating the quantities above for $z_i = 1,\ldots,K$ and choosing the class with highest probability. From (2) and (3), we get

$$\mathbf{y}_v^j|\mathbf{z} \sim N\left(\frac{1}{m_j}\sum_{i\in s^j(v)}\boldsymbol{\mu}_{z_i}^j, \frac{1}{m_j^2}\sum_{i\in s^j(v)}\boldsymbol{\Sigma}_{z_i}^j\right) \tag{7}$$

making (6) easy to compute. Also, for the general case (4), the distribution for $\mathbf{y}_v^j|\mathbf{z}$ can be explicitly written down, but then gets somewhat more complex.

IV. Parameter Estimation

The model described in Section II contains two sets of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as defined in Sections II-A and II-B. These parameters must either be estimated or specified by other means [14]. In particular, assuming that a training set $\mathcal{T}$ consisting of pixels with known classes is available, i.e., $\mathcal{T} = \{i; z_i \text{ known}\}$, the parameters in $\boldsymbol{\theta}_2$ can easily be estimated by maximum-likelihood methods. The parameters in the prior models can be given some reasonable values (e.g., all $\alpha_k = 0$ and $\beta \in [1,2]$ for the Potts model).

When the training set is small, utilizing observations from pixels with unknown class labels is advantageous. A maximum-likelihood estimate for the complete parameter vector $\boldsymbol{\theta}$ would then be given by the maximization of

$$\begin{aligned} L(\boldsymbol{\theta}) &= p(\mathbf{y}, \mathcal{T}|\boldsymbol{\theta}) \\ &= \sum_{\mathbf{z}\cap\mathcal{T}} p(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}_2)p(\mathbf{z}|\boldsymbol{\theta}_1) \end{aligned}$$

where $\mathbf{z}\cap\mathcal{T}$ means that $\mathbf{z}$ must satisfy the constraints given by the training set $\mathcal{T}$. In practice, numerical methods, such as the EM algorithm, must be applied for this task.

Here, we consider a simpler approach, mainly because the presence of data at several resolutions makes the computational challenges even harder. We follow [10] by applying the following procedure.

Step 1) Obtain an initial guess $\hat{\mathbf{z}}$ of the true scene $\mathbf{z}$ with initial guesses for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.

Step 2) Estimate $\boldsymbol{\theta}_1$ by maximizing the *pseudolikelihood*

$$\tilde{p}(\mathbf{z}|\boldsymbol{\theta}_1) = \prod_i p(z_i|z_{\partial i};\boldsymbol{\theta}_1) \tag{8}$$

on the current $\hat{\mathbf{z}}$ to obtain a new $\hat{\boldsymbol{\theta}}_1$. Here, ∂i is a neighborhood of pixel i, typically the four (or eight) nearest neighbors.

Step 3) Estimate $\boldsymbol{\theta}_2$ by the value that maximizes $p(\mathbf{y}|\hat{\mathbf{z}}, \boldsymbol{\theta}_2)$.

Step 4) Carry out a *single* cycle of ICM based on the current $\hat{\mathbf{z}}$, $\hat{\boldsymbol{\theta}}_1$, and $\hat{\boldsymbol{\theta}}_2$.

Step 5) Return to Step 2) for fixed number of cycles or until approximative convergence.

This procedure can be considered as an algorithm for optimizing

$$p(\mathbf{z}, \mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = p(\mathbf{z}|\boldsymbol{\theta}_1)p(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \tag{9}$$

with respect to both $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ *and* $\mathbf{z}$. This is in contrast to ordinary maximum-likelihood estimation where $\mathbf{z}$ is marginalized out. Basing the estimation on maximization of (9) *can* give biased estimates. For large amounts of data, this bias is hopefully small.

A further approximation is introduced in Step 2) by estimating $\boldsymbol{\theta}_1$ based on the pseudolikelihood for $\boldsymbol{\theta}_1$ given $\mathbf{z}$ rather than the actual likelihood function. This approximation is introduced because of the difficulty in maximizing $p(\mathbf{z}|\boldsymbol{\theta}_1)$ with respect to $\boldsymbol{\theta}_1$.

Some remarks on the different steps are as follows.

1) We follow the general suggestion from [10] by ignoring spatial dependence for the initial guess of $\hat{\mathbf{z}}$. This makes an initial guess for $\boldsymbol{\theta}_1$ unnecessary. An alternative is to use some reasonable initial values for $\boldsymbol{\theta}_1$. Initial guesses for $\boldsymbol{\theta}_2$ can be based on training data.

2) The pseudolikelihood is maximized through Newton's method.

3) This part consists of ordinary maximum-likelihood estimation conditioned on a known class configuration. A complicating factor is, however, that data at coarser resolution will be a mixture of contributions from different classes. This requires specifically tuned algorithms. We propose to use the EM algorithm for this purpose, treating the contributions from different classes in the mixture observation as missing. The details are given in the following subsection.
4) This is equivalent to classification based on the ICM algorithm as described in Section III.

A. *Estimation of* $\boldsymbol{\theta}_2$

Let us assume that $\widehat{\mathbf{z}}$ is known. We want to maximize the likelihood $p(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}_2)$ with respect to $\boldsymbol{\theta}_2$. Now

$$p(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}_2) = \prod_{j=1}^{p}\prod_{v} p\left(\mathbf{y}_v^j|\mathbf{z};\boldsymbol{\theta}_2\right)$$

and from (7) $p(\mathbf{y}_v^j|\mathbf{z};\boldsymbol{\theta}_2)$ is known. If all data were given at the reference resolution, this step would correspond to ordinary maximum-likelihood estimation, and analytical expressions for the optimal value of $\boldsymbol{\theta}_2$ are available. In the general case with data at different resolutions, this is no longer the case, and optimization will have to be performed through a numerical optimizer. Gradient-based optimizers (such as Newton's method) can be applied, but we will consider an alternative approach where the variables $\tilde{\mathbf{y}}_i^j$ defined in (2) are considered as missing data. Estimation of $\boldsymbol{\theta}_2$ can then be performed through the EM algorithm. Define $\tilde{\mathbf{y}} = \{\tilde{\mathbf{y}}_i^j\}$ to be the complete data corresponding to observed data at resolution j. Note that $\mathbf{y}$ is a (deterministic) function of $\tilde{\mathbf{y}}$. The EM algorithm at iteration $s+1$ is defined through the following:

E-step: Compute

$$Q\left(\boldsymbol{\theta}_2,\boldsymbol{\theta}_2^s\right) = E^{p(\tilde{\mathbf{y}}|\mathbf{y},\mathbf{z};\boldsymbol{\theta}_2^s)}\log\left[p(\tilde{\mathbf{y}}|\mathbf{z};\boldsymbol{\theta}_2)\right].$$

M-step: Maximize $Q(\boldsymbol{\theta}_2,\boldsymbol{\theta}_2^s)$ with respect to $\boldsymbol{\theta}_2$ to obtain $\boldsymbol{\theta}_2^{s+1}$.

Here, $E^{p(\mathbf{y})}g(\mathbf{y})$ means the expectation of $g(\mathbf{y})$ with respect to the distribution $p(\mathbf{y})$.

Now

$$\log\left[p(\tilde{\mathbf{y}}|\mathbf{z};\boldsymbol{\theta}_2)\right] = \sum_{j=1}^{p}\sum_{k=1}^{K}\log\left[p\left(\tilde{\mathbf{y}}^j(k)|\mathbf{z};\boldsymbol{\mu}_k^j,\boldsymbol{\Sigma}_k^j\right)\right]$$

where $\tilde{\mathbf{y}}^j(k)$ is the set of complete data corresponding to pixels where the class is k, i.e., $\tilde{\mathbf{y}}^j(k) = \{\tilde{\mathbf{y}}_i^j; z_i = k\}$. This shows that

$$Q\left(\boldsymbol{\theta}_2,\boldsymbol{\theta}_2^s\right) = \sum_{j=1}^{p}\sum_{k=1}^{K} Q_k^j\left(\boldsymbol{\mu}_k^j,\boldsymbol{\Sigma}_k^j,\boldsymbol{\theta}_2^s\right)$$

where

$$Q_k^j\left(\boldsymbol{\mu}_k^j,\boldsymbol{\Sigma}_k^j,\boldsymbol{\theta}_2^s\right) = E^{p(\tilde{\mathbf{y}}|\mathbf{y},\mathbf{z};\boldsymbol{\theta}_2^s)}\log\left[p\left(\tilde{\mathbf{y}}^j(k)|\mathbf{z};\boldsymbol{\mu}_k^j,\boldsymbol{\Sigma}_k^j\right)\right].$$

Since $(\boldsymbol{\mu}_k^j, \boldsymbol{\Sigma}_k^j)$ is only contained in $Q_k^j(\boldsymbol{\theta}_2,\boldsymbol{\theta}_2^s)$, maximization can be performed on each source and each class separately. Furthermore

$$\begin{aligned}\log\left[p\left(\tilde{\mathbf{y}}^j(k)|\mathbf{z};\boldsymbol{\mu}_k^j,\boldsymbol{\Sigma}_k^j\right)\right] &= \sum_{i:z_i=k}\log\left[p\left(\tilde{\mathbf{y}}_i^j|z_i;\boldsymbol{\mu}_k^j,\boldsymbol{\Sigma}_k^j\right)\right]\\ &= C - \frac{n_k}{2}\log\left(\left|\boldsymbol{\Sigma}_k^j\right|\right)\\ &\quad - \frac{1}{2}\sum_{i:z_i=k}\left(\tilde{\mathbf{y}}_i^j-\boldsymbol{\mu}_k^j\right)^T\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\\ &\quad \times\left(\tilde{\mathbf{y}}_i^j-\boldsymbol{\mu}_k^j\right)\end{aligned}$$

where C is a constant not depending on the parameters in $\boldsymbol{\theta}_2$ or on the observations, while n_k is the number of pixels from class k. Now, $\tilde{\mathbf{y}}_i^j|\mathbf{y},\mathbf{z} \sim N(\boldsymbol{\eta}_i^j,\mathbf{S}_i^j)$ where a procedure for calculating $\boldsymbol{\eta}_i^j$ and $\mathbf{S}_i^j$ is given in the Appendix. Then, after some algebraic manipulation, we get

$$\begin{aligned}&\mathsf{E}^{p(\tilde{\mathbf{y}}|\mathbf{y},\mathbf{z};\boldsymbol{\theta}_2^s)}\left[\left(\tilde{\mathbf{y}}_i^j-\boldsymbol{\mu}_k^j\right)^T\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\left(\tilde{\mathbf{y}}_i^j-\boldsymbol{\mu}_k^j\right)\right]\\ &\quad = \mathsf{tr}\left(\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\mathbf{S}_i^j\right) + \left(\boldsymbol{\eta}_i^j-\boldsymbol{\mu}_k^j\right)^T\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\left(\boldsymbol{\eta}_i^j-\boldsymbol{\mu}_k^j\right)\end{aligned}$$

where $\mathsf{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$. Furthermore [15, p. 97]

$$\begin{aligned}Q_k^j\left(\boldsymbol{\theta}_2,\boldsymbol{\theta}_2^s\right) = C_j^k &- \frac{n_k}{2}\log\left(\left|\boldsymbol{\Sigma}_k^j\right|\right)\\ &- \frac{1}{2}\mathsf{tr}\left(\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\overline{\mathbf{S}}_k^j\right) - \frac{n_k}{2}\mathsf{tr}\left(\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\mathbf{V}_k^j\right)\\ &- \frac{n_k}{2}\left(\overline{\boldsymbol{\eta}}_k^j-\boldsymbol{\mu}_k^j\right)^T\left[\boldsymbol{\Sigma}_k^j\right]^{-1}\left(\overline{\boldsymbol{\eta}}_k^j-\boldsymbol{\mu}_k^j\right)\end{aligned}$$

where $\overline{\boldsymbol{\eta}}_k^j = (1/n_k)\sum_{i:z_i=k}\boldsymbol{\eta}_i^j$, $\overline{\mathbf{S}}_k^j = (1/n_k)\sum_{i:z_i=k}\mathbf{S}_i^j$ and $\mathbf{V}_k^j = (1/n_k)\sum_{i:z_i=k}(\boldsymbol{\eta}_i^j-\overline{\boldsymbol{\eta}}_k^j)(\boldsymbol{\eta}_i^j-\overline{\boldsymbol{\eta}}_k^j)^T$ From this, we see that the optimal values for $\boldsymbol{\mu}_k^j$ and $\boldsymbol{\Sigma}_k^j$ are

$$\widehat{\boldsymbol{\mu}}_k^j = \overline{\boldsymbol{\eta}}_k^j \quad \widehat{\boldsymbol{\Sigma}_k^j} = \overline{\mathbf{S}}_k^j + \mathbf{V}_k^j.$$

Since this EM algorithm is part of an outer maximization algorithm, only a small number of iterations should be sufficient, at least at an early stage of the outer algorithm. Note that for sources that are observed at the reference resolution, $\tilde{\mathbf{y}}_j = \mathbf{y}_j$ resulting in $\boldsymbol{\eta}_i^j = \mathbf{y}_i^j$ and $\mathbf{S}_i^j = \mathbf{0}$.

V. Results

A. *Simulated Images*

The proposed multiscale classification scheme has been tested on a simulated dataset, where the class properties are derived from real Systeme Pour l'Observation de la Terre (SPOT) MultiSpectral mode (XS) and Landsat Thematic Mapper (TM) images. The reason why we use simulated images is that the new method mainly can be expected to bring improvements in the presence of fine details, e.g., fine structures such as roads and transitions corresponding to region boundaries. Sufficiently detailed ground truth is rarely available for such zones, as the ground truth samples generally are placed well inside homogeneous regions.

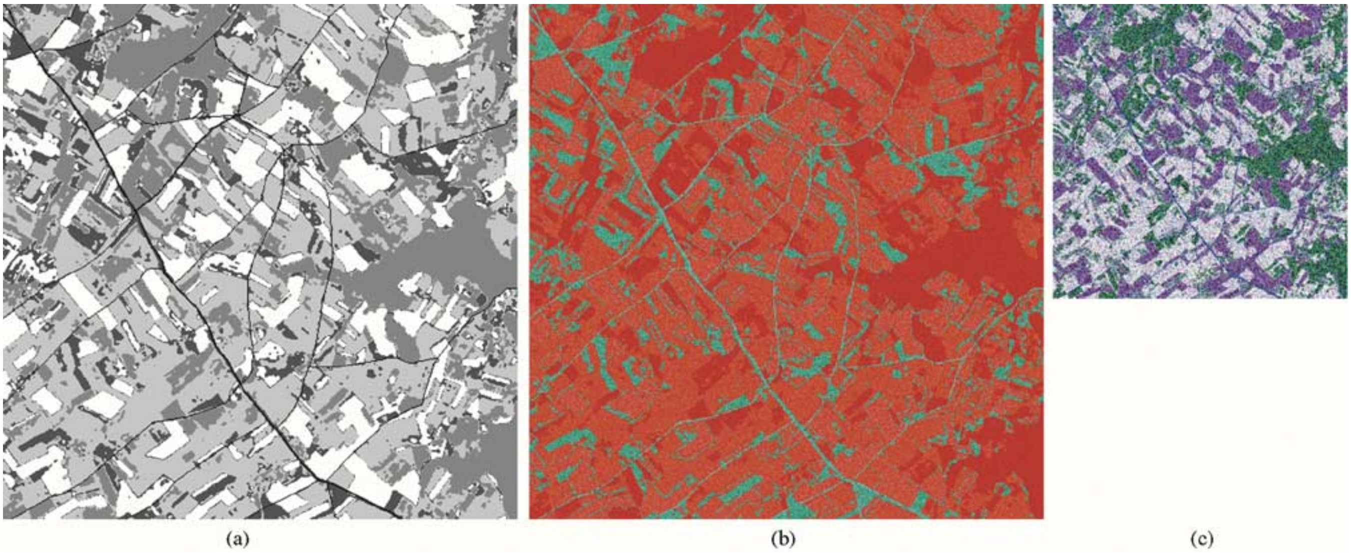


Fig. 2. Simulated images for multiscale classification. (a) 512 × 512 class label image with five classes. (b) Three-band 512 × 512 simulated image with class characteristics derived from SPOT XS data. (c) Six-band 256 × 256 simulated image with class characteristics derived from Landsat TM data (three bands shown).

The class label image that we used is shown in Fig. 2(a). It represents an agricultural scene and is derived from a class image used in a previous study [16], by manually adding roads, so that the number of classes becomes 5. It should be noted that the class image not is a realization of a Markov random field.

Covariance matrices and mean vectors for roads (class 1), urban areas (class 2), forests (class 3) and two kinds of agricultural crops (class 4 and class 5) were estimated from a SPOT XS image of Bourges, France, and the associated ground truth. The three-band image shown in Fig. 2(b) was obtained by simulating multivariate Normal noise according to the estimated class properties, and by attributing the signal vectors to the pixels of the ideal class image.

Likewise, a six-band image was simulated from the class properties of a Landsat TM image of Kjeller, Norway, and the label image shown in Fig. 2(a). However, the estimated covariance matrices were first multiplied by a factor of 4, then 2 × 2 pixel blocks of the simulated image were averaged, and the entire image was down-sampled by 2 × 2. Hence, we obtained the image shown in Fig. 2(c), which has 2 × 2 times lower resolution than the initial class image, and contains mixed pixels at region boundaries, in agreement with our multiscale model. The class properties of the simulated image are very close to those of the original Landsat TM image.

The simulated images look somewhat noisier than real SPOT XS and Landsat TM images. This is due to the fact that slow spatial variations in the class properties of real images contribute to the overall class variance. In a simulated image generated as described above, the total class variance can be observed between neighboring pixels.

A 100 × 100 pixel square in the upper left corner of the ideal class image shown in Fig. 2(a), containing samples of all five classes, was defined as training set, and the rest of the ideal class image constituted the test set.

A Potts model was assumed for the class image, while the simulated images were assumed to follow the data model described in Section II-B.

To make the results as representative as possible, we decided to separate the estimation and classification phases, by first running the ICM loop until convergence with the training set fixed to estimate the radiometric and spatial parameters, and then, by running the ICM until convergence without fixing the test set and without parameter estimation, in order to obtain the final classification. For practical applications, it would be sufficient to run the ICM until convergence only once, with fixed training set, estimating the parameters and simultaneously iterating toward the final classification.

The result of the multiscale classification scheme is shown in Fig. 3(a), and the confusion matrix (rows = reference, columns = result) is given in Table I. The convergence criterion was reached after seven iterations with the training set fixed, and the computing time on a Linux personal computer with a Pentium III 733-MHz processor was 8 min 33 s. For the additional ICM loop without parameter estimation and fixed training set (which can be omitted in practical applications), convergence was reached after six iterations and the processing time was 3 min 42 s. The overall probability of correct classification, computed on the test set, was 97.8%.

To assess the improvement brought by the new multiscale approach, we compared it with the corresponding single-scale analysis of the same dataset. First we resampled the image with Landsat TM properties to the reference resolution by replacing each pixel by a block of 2 × 2 identical pixels, and concatenated the resulting image bands with those of the image with SPOT XS properties, so that we got a 512 × 512 with nine bands. ICM loops for parameter estimation and classification were carried out until convergence. The only difference from the multiscale algorithm is that the computations related specifically to the multiscale model are let out. The result is shown in Fig. 3(b) and the confusion matrix is given in Table II. Convergence was reached after 13 iterations and 1 min 42 s for the combined estimation and classification ICM loop, and after four iterations and 16 s for the final (optional) classification loop.

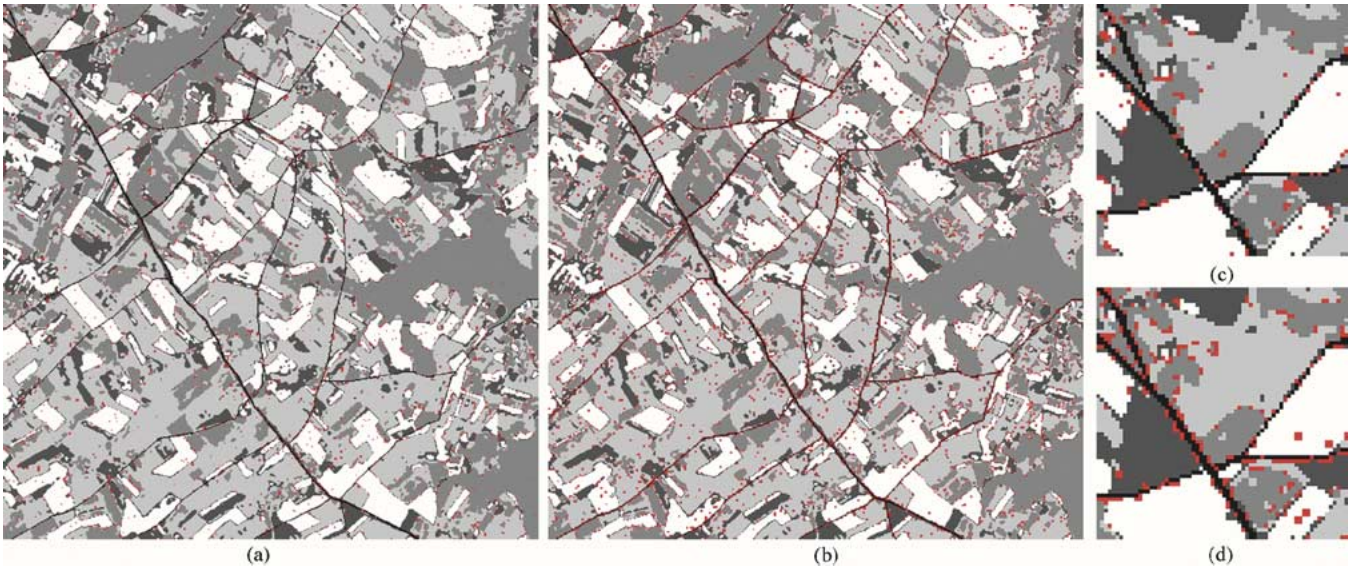


Fig. 3. Classification result of (a) the multiscale classification scheme and (b) the corresponding single-scale classification scheme applied after resampling of the second image to the reference resolution. Correctly classified pixels are shown in gray levels, and misclassified pixels are shown in red. The upper left corners are magnified to show some representative details of (c) the multiscale result and (d) the single-scale result.

TABLE I
CONFUSION MATRIX FOR THE MULTISCALE RESULT IN FIG. 3

$p(\hat{z}\|z)$	$\hat{z}=1$	$\hat{z}=2$	$\hat{z}=3$	$\hat{z}=4$	$\hat{z}=5$
$z=1$	0.893	0.107	0.000	0.000	0.000
$z=2$	0.015	0.985	0.000	0.000	0.000
$z=3$	0.000	0.000	0.993	0.006	0.001
$z=4$	0.000	0.000	0.007	0.966	0.027
$z=5$	0.000	0.000	0.008	0.006	0.986

TABLE II
CONFUSION MATRIX FOR THE SINGLE-SCALE RESULT IN FIG. 3

$p(\hat{z}\|z)$	$\hat{z}=1$	$\hat{z}=2$	$\hat{z}=3$	$\hat{z}=4$	$\hat{z}=5$
$z=1$	0.944	0.056	0.000	0.000	0.000
$z=2$	0.022	0.978	0.000	0.000	0.000
$z=3$	0.010	0.000	0.957	0.031	0.002
$z=4$	0.010	0.000	0.005	0.940	0.044
$z=5$	0.008	0.000	0.002	0.037	0.952

TABLE III
PERFORMANCE ON SIMULATED DATASET

Method	Data	Accuracy	Time
Multi-scale ICM	XS + TM	97.8 %	8 m 33 s
Single-scale ICM	XS + TM	95.2 %	1 m 42 s
Single-scale ICM	XS only	90.9 %	45 s
Single-scale ICM	TM only	78.3 %	54 s

The overall probability of correct classification was here 95.2%. For comparison, we reached 90.9% using only the three-band image with SPOT XS properties, and 78.3% when using only the resampled six-band image with Landsat TM properties.

The most important performance measures are summarized in Table III. The computing time excludes the optional second ICM loop, that can be omitted in practice. We see that there are only 9.1% of misclassified pixels when applying the single-scale ICM scheme to the image with SPOT XS properties only. The same method applied to a concatenation of this image and a resampled version of the image with Landsat TM properties further reduces the proportion of misclassified pixels by 4.3%, and another 2.6% are eliminated with the new multiscale method, so that only 2.2% of misclassified pixels remain. However, the last improvement is obtained at the cost of five times higher computing time.

As can be seen from Fig. 3, the multiscale approach, as could be expected, mainly improves the classification accuracy near fine structures and boundaries between regions. Careful study of the confusion matrices in Tables I and II reveals that the multiscale scheme for this dataset does not improve the number of correctly classified pixels of class 1 (roads), as the spectral characteristics of this class are quite distinct, but it significantly reduces the confusion with this class, i.e., that it performs better in the neighborhood of these fine structures. It also performs significantly better at the boundaries between larger regions.

It can be argued that plain 2 × 2 up-sampling of the low-resolution image may cause a block effect in the result of the single-scale classification of the concatenated image bands, and that higher order interpolation could yield a smoother result. We have therefore tested the impact of using cubic interpolation. Single-scale classification of the nine image bands in this case yields 96.7% of correctly classified pixels, which is a 1.5% improvement with respect to the previous single-scale result, but still 1.1% poorer than the multiscale classification. Visual inspection reveals that the cubic interpolation brings little or no improvement near fine structures.

B. Real Images

Comparison of the single-scale and the multiscale estimation and classification methods has also been carried out on an extract of a SPOT 5 image of Paris, as shown in Fig. 4. It consists of a 1600 × 1600 panchromatic band with 2.5-m pixel size and four 400 × 400 spectral bands with 10-m pixel size.

Independent training and test sets for six classes were created based on visual inspection and are shown together in Fig. 4(a)

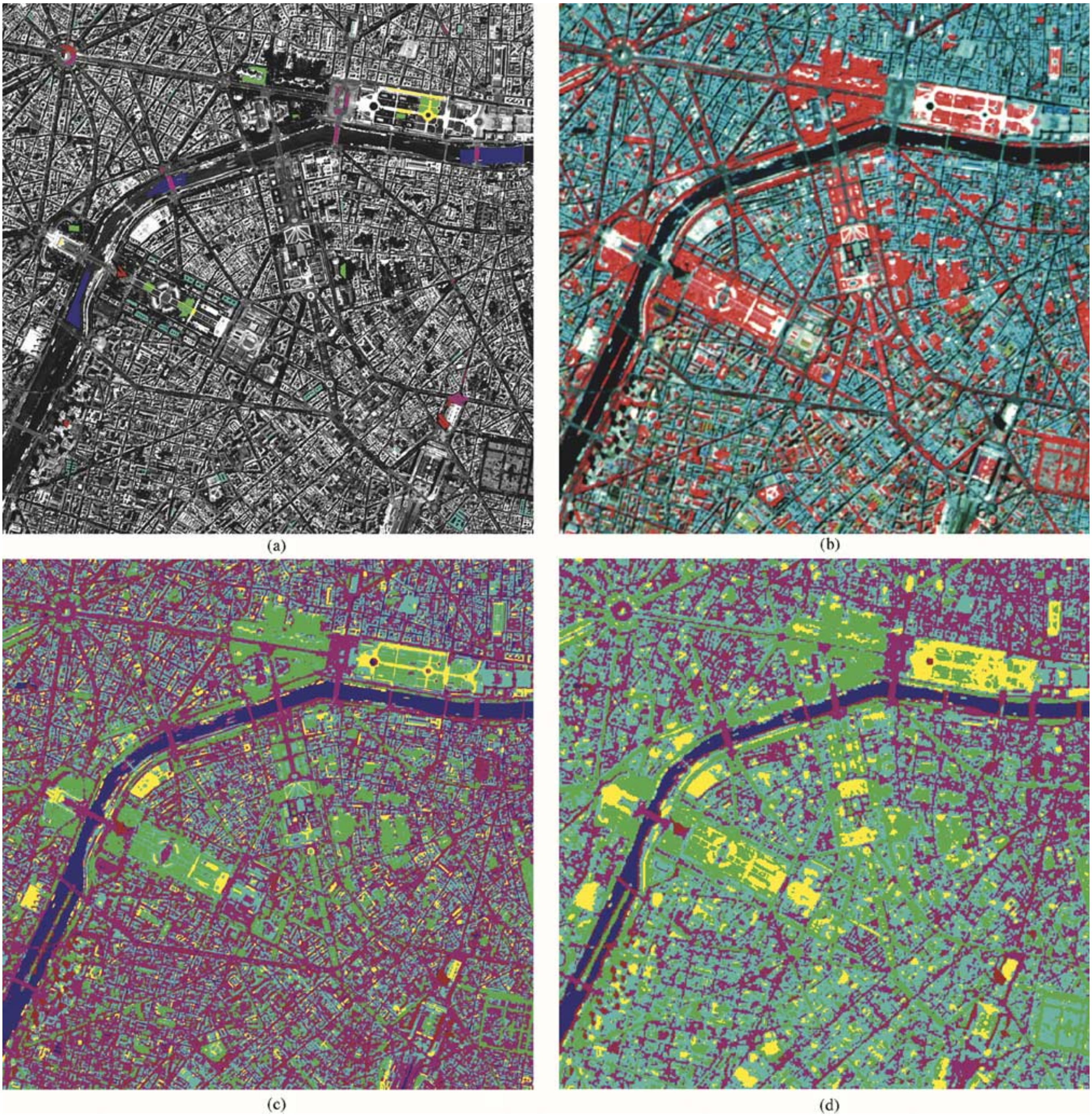


Fig. 4. SPOT 5 image of downtown Paris, France. Copyright CNES 2002—Distribution SPOT Image. (a) 1600 × 1600 extract of the panchromatic band (2.5 m) with training and test sets overlaid. (b) 4 × 4 magnified color composite of the corresponding 400 × 400 extract of spectral bands 3, 2, and 1 (10 m). Classification results obtained with (c) the multiscale scheme and (d) the single-scale scheme after resampling to the reference resolution.

overlaid to the panchromatic band. The classes and the colors representing them are as follows:

- Water (blue);
- Shadow (brown);
- Vegetation (green);
- Building (cyan);
- Road (purple);
- Path (yellow).

Using the same settings as for the simulated dataset presented in Section V-A, we performed iterative parameter estimation and classification with the multiscale and the single-scale schemes. The classified images are shown in Fig. 4(c) and (d), and the confusion matrices are presented in Tables IV and V, respectively.

The overall probability of correct classification in the test set is here 88.7% for the multiscale scheme and 78.0% for the single-scale scheme. Several differences can be noticed visually. First of all, fine structures such as streets and shadows of buildings are better preserved for the multiscale approach, despite the fact that the estimated regularity parameter is the same in both cases ($\beta \approx 2.2$). We also notice that the decision limits between the classes Road (purple) and Building (cyan)

TABLE IV
CONFUSION MATRIX FOR THE MULTISCALE RESULT IN FIG. 4

$p(\hat{z}\|z)$	$\hat{z}=1$	$\hat{z}=2$	$\hat{z}=3$	$\hat{z}=4$	$\hat{z}=5$	$\hat{z}=6$
$z=1$	0.945	0.038	0.000	0.000	0.017	0.000
$z=2$	0.000	0.653	0.000	0.000	0.347	0.000
$z=3$	0.000	0.000	0.912	0.085	0.001	0.002
$z=4$	0.000	0.000	0.000	0.630	0.105	0.264
$z=5$	0.005	0.000	0.004	0.000	0.991	0.000
$z=6$	0.000	0.000	0.000	0.058	0.000	0.942

TABLE V
CONFUSION MATRIX FOR THE SINGLE-SCALE RESULT IN FIG. 4

$p(\hat{z}\|z)$	$\hat{z}=1$	$\hat{z}=2$	$\hat{z}=3$	$\hat{z}=4$	$\hat{z}=5$	$\hat{z}=6$
$z=1$	0.813	0.136	0.019	0.008	0.024	0.000
$z=2$	0.000	0.597	0.064	0.060	0.279	0.000
$z=3$	0.000	0.000	0.691	0.000	0.000	0.309
$z=4$	0.000	0.000	0.000	0.624	0.228	0.147
$z=5$	0.002	0.004	0.044	0.016	0.935	0.000
$z=6$	0.000	0.000	0.000	0.000	0.000	1.000

as well as Building (cyan) and Path (yellow) have become different through the iterative estimation process for the two approaches, so that, for example, the class Building is somewhat underrepresented in the multiscale result and overrepresented in the single-scale result. More generally, the single-scale classification mainly reflects the information in the lower resolution spectral bands, whereas the multiscale method better preserves the structural information in the high-resolution panchromatic band.

Using cubic interpolation instead of plain 4 × 4 up-sampling of the four spectral bands only increases the portion of correctly classified pixels from 78.8% to 79.9%. Also visually the two single-scale results are very similar and considerably poorer than multiscale classification, which here yields an accuracy of 88.7%.

It should be noted that this real dataset deviates from the multiscale model presented in Section II in several ways as follows.

- The 2.5- and 10-m bands are not independent (simultaneous acquisition and spectral overlap).
- There are mixed pixels at the reference resolution.
- There is significant intraclass variation.

Nevertheless, we observe significant improvements with respect to single-scale analysis. The variability of the properties of thematic classes in natural scenes is a well-known problem that typically leads to interclass confusion, and iterative estimation and classification methods are particularly vulnerable.

VI. CONCLUSION

This paper describes a Bayesian approach for integration of multiresolution image data. The approach is based on the concept of a reference resolution. Data at this and lower resolutions are connected to the reference resolution through a fully specified statistical model. Algorithms for parameter estimation and classification based on the multiscale model are proposed, and results and comparisons with single-scale classification are presented for simulated and real satellite images.

The results obtained on a simulated dataset that fits the multiscale model clearly demonstrate that the multiscale scheme produce higher classification accuracy near fine structures and region boundaries, as compared to single-scale analysis where the images with coarser resolution have first been resampled to the reference solution.

A similar comparison was carried out on a SPOT 5 image with a high-resolution panchromatic band and four spectral bands with 4 × 4 times lower resolution. Even though this dataset only approximately fits the multiscale model, a significant improvement of the classification accuracy was observed. Visual inspection confirmed a better preservation of fine structures than for single-scale analysis.

The proposed multiscale estimation and classification method provides a better way of exploiting spectrally rich images at lower resolution together with images at the reference resolution, when the goal is to obtain accurate and spatially detailed classification results at the reference resolution. The method has so far only been implemented for cases where the pixel dimensions at lower resolutions are entire multiples of the pixel size at the reference resolution. Extensions to more general situations are possible, but require modifications of the algorithms and imply a higher implementational complexity. Further, the approach is best suited for cases where there is not a too high difference in resolution between the observed images.

APPENDIX
DISTRIBUTION OF $\tilde{\mathbf{y}}$

We will consider the expectation and covariance matrix of $\{\tilde{\mathbf{y}}_i^j\}$, conditioned on

$$\frac{1}{m^j} \sum_{i' \in s_j(v_j(i))} \tilde{\mathbf{y}}_{i'}^j = \mathbf{y}_{v_j(i)}^j.$$

Since $\{\tilde{\mathbf{y}}_i^j\}$ follows a Normal distribution, also the conditional distribution will be Normal. In the following, we will use $v = v_j(i)$ to simplify the notation. Define $\tilde{\mathbf{y}}_v^j$ to be the vector with elements $\{\tilde{\mathbf{y}}_i^j, i \in s^j(v)\}$. Then

$$\mathsf{E}\left[\begin{matrix}\tilde{\mathbf{y}}_v^j \\ \mathbf{y}_v^j\end{matrix}\middle|\,\mathbf{z}\right] = \begin{pmatrix}\boldsymbol{\mu}_1 \\ \overline{\boldsymbol{\mu}}_2\end{pmatrix}$$

where $\boldsymbol{\mu}_1$ is the vector with elements $\{\boldsymbol{\mu}_{z_i}^j, i \in s^j(v)\}$, while

$$\overline{\boldsymbol{\mu}}_2 = \frac{1}{m^j} \sum_{i \in s^j(v)} \boldsymbol{\mu}_{z_i}^j.$$

Furthermore

$$\mathsf{Var}\left[\begin{matrix}\tilde{\mathbf{y}}_i^j \\ \mathbf{y}_i^j\end{matrix}\middle|\,\mathbf{z}\right] = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \overline{\boldsymbol{\Sigma}}_{22}\end{pmatrix}$$

where $\boldsymbol{\Sigma}_{11}$ is a block diagonal matrix containing $\{\boldsymbol{\Sigma}_{z_i}^j, i \in s^j(v)\}$ as block diagonals, $\boldsymbol{\Sigma}_{12}$ is a column vector of matrices with elements $\{(1/m^j)\boldsymbol{\Sigma}_{z_i}^j, i \in s^j(v)\}$ and $\overline{\boldsymbol{\Sigma}}_{22} = (1/(m^j)^2) \sum_{i \in s^j(v)} \boldsymbol{\Sigma}_{z_i}^j$. Then [15, p. 63]

$$\begin{aligned}\boldsymbol{\eta}_v^j &= \mathsf{E}\left[\tilde{\mathbf{y}}_i^j | \mathbf{y}_v^j, \mathbf{z}\right] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\overline{\boldsymbol{\Sigma}}_{22}^{-1}\left(\mathbf{y}_v^j - \overline{\boldsymbol{\mu}}_2\right) \\ \mathbf{S}_v^j &= \mathsf{Var}\left[\tilde{\mathbf{y}}_i^j | \mathbf{y}_v^j, \mathbf{z}\right] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\overline{\boldsymbol{\Sigma}}_{22}^{-1}\boldsymbol{\Sigma}_{21}.\end{aligned}$$

From this, we deduce that $\tilde{\mathbf{y}}_i^j|\mathbf{y} \sim N(\boldsymbol{\eta}_i^j, \mathbf{S}_i^j)$ where $\boldsymbol{\eta}_i^j$ is the (vector) component of $\boldsymbol{\eta}_v^j$ corresponding to pixel i, while $\mathbf{S}_i^j$ is the block diagonal of $\mathbf{S}_v^j$ corresponding to pixel i.

The first time these quantities need to be calculated, some initial values for $\boldsymbol{\mu}_k^j$ and $\boldsymbol{\Sigma}_k^j$ are needed. A possible choice is

$$\boldsymbol{\mu}_k^j = \mathbf{0}, \quad \boldsymbol{\Sigma}_k^j = \boldsymbol{I}$$

in which case

$$\boldsymbol{\eta}_i^j = \mathbf{y}_v^j, \quad \mathbf{S}_i^j = \mathbf{0}$$

in the first iteration of the inner EM algorithm.

REFERENCES

[1] J. C. Price, "Combining multispectral data of differing spatial resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1199–1203, 1999.

[2] N. G. Best, K. Ickstadt, and R. L. Wolpert, "Spatial poisson regression for health and exposure data measured at disparate resolutions," *J. Amer. Stat. Assoc.*, vol. 452, pp. 1076–1088, 2000.

[3] M. M. Daniel and A. S. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, no. 1, pp. 164–180, 1997.

[4] D. Hirst, G. Storvik, and A. R. Syversveen, "A hierarchical modeling approach to combining environmental data at different scales," *J. R. Stat. Soc., C*, vol. 52, no. 3, pp. 377–390, 2003.

[5] P. Puyou-Lascassies, A. Podaire, and M. Gay, "Extracting crop radiometric responses from simulated low and high spatial resolution satellite data using a linear mixing model," *Int. J. Remote Sens.*, vol. 15, no. 18, pp. 3767–3784, 1994.

[6] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhäckel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[7] M. M. Crawford, S. Kumar, M. R. Ricard, J. C. Gibeaut, and A. Neuenshwander, "Fusion of airborne polarimetric and interferometric SAR for classification of coastal enviromnents," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1306–1315, May 1999.

[8] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 41–63, Jan. 1994.

[9] J. Besag, "Toward Bayesian image analysis," *J. Appl. Stat.*, vol. 16, no. 3, pp. 395–407, 1989.

[10] ——, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc., B*, vol. 48, no. 3, pp. 259–302, 1986.

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

[12] A. H. S. Solberg, "Contextual data fusion applied to forest map revision," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1234–1243, May 1999.

[13] G. Storvik and G. Dahl, "Lagrangian based methods for finding MAP solutions for MRF models," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 469–479, Mar. 2000.

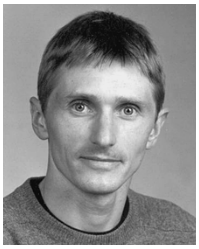
[14] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, no. 2, pp. 131–164, 1989.

[15] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. London, U.K.: Academic, 1979.

[16] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 675–686, Mar. 2003.

Geir Storvik received the M.S and Ph.D. degrees in statistics from the University of Oslo, Oslo, Norway, in 1986 and 1993, respectively.

Since 1993, he has been an Associate Professor with the Department of Mathematics, University of Oslo. His main interest are in statistical computing, Bayesian hierarchical models, spatio-temporal modeling, and image analysis.



Roger Fjørtoft (M'01) received the M.S. degree in electronics from the Norwegian Institute of Technology, Trondheim, in 1993, and the Ph.D. degree in signal processing, image processing, and communications from the Institut National Polytechnique, Toulouse, France, in 1999.

Since 2000, he has been with the Norwegian Computing Center, Oslo, Norway, where he works on automatic image analysis for remote sensing applications. He is currently a Visiting Scientist in the Altimetry and Radar Department, French Space Agency (CNES), Toulouse.

Anne H. Schistad Solberg (S'92–M'96) received the M.S. and Ph.D. degrees in image analysis from the University of Oslo, Oslo, Norway, in 1989 and 1995, respectively.

She is currently an Associate Professor in the Digital Signal Processing and Image Analysis Group, Department of Informatics, University of Oslo. Her research interests include SAR image analysis, oil spill detection, hyperspectal imagery, statistical classification, and data fusion.