# A Markov Random Field Model for Classification of Multisource Satellite Imagery

*Anne H. Schistad Solberg*[1], *Torfinn Taxt*[2] *and Anil K. Jain*[3]

[1] Norwegian Computing Center, P.O.Box 114 Blindern, 0314 Oslo, Norway
Phone +47 22 85 25 00 Fax +47 22 69 76 60

[2] University of Bergen, 5000 Bergen, Norway

[3] Michigan State University, East Lansing, Michigan 48824

**Abstract**

In this paper, a general model for multisource classification of remotely sensed data based on Markov Random Fields (MRF) is proposed. A specific model for fusion of optical images, synthetic aperture radar (SAR) images, and GIS (Geographic Information Systems) ground cover data is presented in detail and tested. The MRF model exploits spatial class dependencies (spatial context) between neighboring pixels in an image, and temporal class dependencies between different images of the same scene. By including the temporal aspect of the data, the proposed model is suited for detection of class changes between the acquisition dates of different images. The performance of the proposed model is investigated by fusing Landsat TM images, multitemporal ERS-1 SAR images, and GIS ground-cover maps for land-use classification, and on agricultural crop classification based on Landsat TM images, multipolarization SAR images, and GIS crop field border maps. The MRF model performs significantly better than a simpler reference fusion model it is compared to.

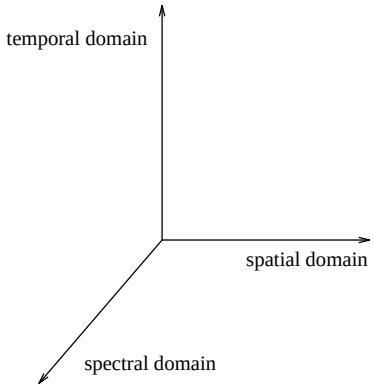


Figure 1: *Contextual information defined in three dimensions.*

# 1 Introduction

For a particular remote sensing application, image data might be available from several different sources. An increasing number of satellite-based sensors are collecting data for geographical, oceanic, geophysical, meteorological, and surveillance studies. In addition to digital satellite imagery, spatial data from Geographic Information Systems (GIS) might also provide information about a particular phenomenon. For many applications, the information provided by individual sensors is incomplete, inconsistent, or imprecise. Additional sources may provide complementary data, and the merging of multisource data can create a more consistent interpretation of the scene, in which the associated uncertainty is decreased.

Consider the task of classifying a satellite image. In addition to the sensed image, prior information about the sensed area in the form of map data or other satellite images acquired during previous surveys might be available to the analyst. Although the prior data are likely to be only partly accurate due to changes in, e.g., area usage, they may still provide additional or complementary information about the scene. To utilize the prior data, the temporal aspect of the data and the reliability of the different sensors or sources need to be considered. Furthermore, the multisource data set need to be co-registered.

In the interpretation of a scene, contextual information is important. Basically, the effect of context in an image labelling problem is that when a pixel is considered in isolation, it may provide incomplete information about the desired characteristics. By considering the pixel in context with other measurements, more complete information might be derived. Context can be defined along three different dimensions, the spectral, the spatial, and the temporal dimension (see Figure **??**). The spectral dimension refers to different bands of the electromagnetic spectrum. These bands may be acquired by a single multispectral sensor, or by two or more different sensors. Modelling of the spectral context or spectral correlation of a multispectral image frequently improves the separation between various ground cover classes compared to a single-band image analysis. Fusion of data from different sensors allows coverage of a broader range of the spectrum. Spatial context is defined by the correlations between spatially adjacent pixels in a spatial neighborhood. The concept of texture is based on the relationship between the measured intensities or pixel values of spatially neighboring pixels [**?**]. In a pixel labelling problem, spatial context refers to the correlations between the class labels of neighboring pixels. Temporal context between multiple images of the same area defines the third contextual dimension. By using multisource data, a better coverage of the temporal dimension is achievable, because one is no longer limited by the repeat cycles of a single satellite.

Markov Random Fields provide a methodological framework which allows the images from different sensors and map data to be merged in a consistent way. Spatial and temporal contextual information is also readily modelled within this framework. In this paper, we present a new general model for multisource classification of remotely sensed data based on Markov Random Field (MRF) modelling.

The performance of the proposed model is investigated by fusing Landsat TM images, multitemporal ERS-1 SAR images, and GIS ground-cover maps for land-use classification, and on agricultural crop classification based on Landsat TM images, multipolarization SAR images, and GIS field border maps. The performance is compared to a simpler fusion model [?] which performs fusion on a pixel-by-pixel level by merging the posterior probabilities from sensor-specific classifiers. The MRF model proposed here performed significantly better than the reference model.

The remainder of the paper is organized as follows: Section 2 reviews previous approaches to data fusion for classification tasks. In section 3, basic elements in a multisource classification model are discussed. The modelling of the multisensor classifier is described in section 4, while section 5 presents the experimental results for the MRF model and the reference model. Finally, a discussion of various aspects of the MRF model and experimental results is provided in section 6.

## 2 Related work

The literature on data fusion in computer vision and machine intelligence domains is substantial. For an extensive review of data fusion, we refer the reader to Abidi and Gonzalez [?]. Multisensor fusion can take place at either the *signal, pixel, feature,* or *symbol* level of representation. Signal-based fusion refers to the combination of signals from different sensors to create a new signal with a better signal-to-noise ratio than the original signals [?]. Pixel-based fusion consists of merging information from different images on a pixel-by-pixel basis to improve the performance of image processing tasks such as segmentation [?]. Feature-based fusion consists of merging features extracted from different signals or images [?]. Symbol-level or decision-level fusion consists of merging information at a higher level of abstraction, e.g., when fusing information from sensors which are very dissimilar or refer to different regions of the environment [?].

For data fusion within remote sensing applications, the main approaches include statistical methods [?, ?], Dempster-Shafer theory [?, ?], and neural networks [?, ?]. To a certain extent, the performances of these three approaches have been compared on the same data set [?, ?]. In these comparisons statistical methods performed the best.

Among the statistical approaches to data fusion for remote sensing applications, three main methods are reported in the literature: The augmented vector approach, stratification, and extended statistical fusion. In the simple augmented vector approach, data from different sources are concatenated as if they were measurements from one single sensor and the fused data vector is then classified as an ordinary single-source measurement [?]. Stratification is often used to incorporate ancillary GIS data in the interpretation process. The GIS data is stratified into categories and then a spectral model for each of these categories is used [?]. The methods based on extended statistical fusion [?, ?] were derived by extending the concepts used for classification of multispectral images involving only one data source. Each data source is considered independently, and the classification results are fused using weighted linear combinations.

Contextual information from neighboring pixels normally improves the classification results compared

to a pixel-by-pixel classification [?]. For a general overview of contextual methods for pattern recognition, the reader is referred to [?, ?, ?]. In general, there are two main types of contextual information [?]: interpixel class dependency and interpixel feature correlation. Interpixel class dependency refers to the correlation among the class labels in a neighborhood. Interpixel feature correlation refers to the relationship between the feature vectors of spatially adjacent pixels in an image. Utilization of interpixel feature correlation requires a model which describes the correlation between adjacent pixel values. However, this results in a model with a large computational complexity [?, ?]. Methods based on interpixel class dependency can be divided into two categories: 1) classification rules based on stochastic models of the behavior of the classes in the scene and the behavior of the conditional distributions of feature vectors given the underlying classes [?, ?]; and 2) simultaneous classification of a group of pixels in the scene, for example by constructing the maximum a posteriori probability estimates of the class labels [?].

Markov Random Field models are commonly used to model spatial and/or temporal context for image classification problems [?, ?, ?]. Geman and Geman [?] used two coupled Markov Random Fields describing an intensity process and a line process (to model discontinuities) for image segmentation. Approaches that are essentially the same as coupled Random Field methods have also been used for data fusion [?, ?].

The methods reviewed above assume that no changes with respect to the pattern classes have occurred between the acquisition of different images. A few studies have been concerned with classification in a time-varying environment. Swain [?] developed a cascade classifier based on multiple single-source observations of a pattern. Jeon [?] presented a spatio-temporal classifier utilizing both the temporal and the spatial context of the image data, without treating the multisensor aspect explicitly. Middelkoop et al. [?] presented a knowledge-based classifier which incorporated land-cover data from the preceding years. Recursive estimation of class labels has been used to model context for character recognition purposes [?, ?]. Raviv [?] originally utilized recursive labelling to model the dependencies between the characters in a word. The recursive modelling of the class labels for a time serie of images presented in this paper is similar to Raviv's approach. In our earlier work [?, ?], we presented a multisensor fusion model incorporating the concepts of time and class changes. This model performed fusion on a pixel-by-pixel level by merging the posterior probabilities from sensor-specific classifiers, taking into account the reliability factors and the likelihood of class changes. The fusion model based on Markov Random Fields described in this paper allows a stronger coupling between the data sources. Fusion can be performed on an intermediate level for pixels in a local neighborhood, while retaining a close link to the original image data.

## 3 Basic elements in a multisource classification model

In this section, we will briefly discuss different elements in the multisource classification model. The particular modelling of each module for the task of land-use classification will be discussed in section ??. A schematic view of the model is given in Figure ??. To specify the model, each of the modules and their interaction must be defined. The basic elements are the following:

- *Input data.* The input images are assumed to be geocoded and co-registered. If the satellite flight parameters are available, automatic transformation of each image to a common map projection is possible. Without these parameters, the transformation parameters must be computed from manually selected control points. A transformation will alter the image statistics. If the images contain textured regions or significant noise with known characteristics, then a transformation

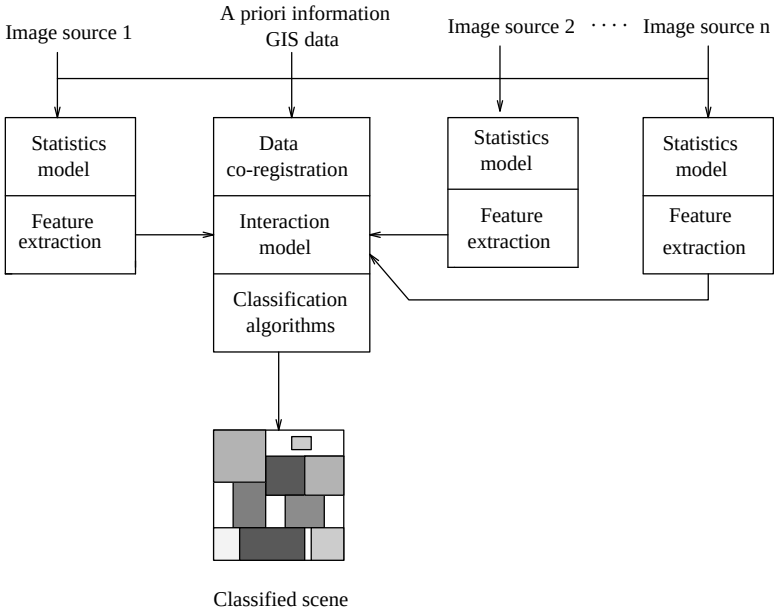


Figure 2: *A schematic diagram of a multisource classification model.*

may be undesirable. If texture is present for only one of the sensors, fusion can be performed in the geometry of this sensor and the classified scene can then be transformed to the desired projection.

- *An image statistic model.* A sensor-specific image statistic model should contain information about the underlying noise characteristics of the sensor and the intrinsic variability (class noise) of the ground cover classes.

- *Ancillary information.* The backscatter signature for a ground cover measured with a specific sensor normally changes with time and season. Factors such as soil moisture, temperature, wind conditions, and snow cover typically affect the signature. The signature change is most significant for microwave imaging.

- *A model for spatial context.* The interpixel class dependency or interpixel feature correlation is modelled for the pixels in a local spatial neighborhood.

- *A model for temporal context.* Temporal context can be modelled for interpixel class dependency or interpixel feature correlation in the same way as spatial context. The ground cover class of individual pixels, in most cases, remains constant over small time intervals. Information about the likelihood of changes with respect to the classes over time for a certain pixel location may provide useful knowledge about the class dependencies. Accurate modelling of temporal interpixel feature correlation for multisensor images is difficult. This task requires modelling of the temporal changes in spectral signature for each of the sensors due to different weather conditions. A distinction between signature variations due to seasonal changes within the same ground cover class and variations due to an actual change between two classes must also be made. In this paper, we will limit our attention to class dependency constraints.

- *An interaction model.* An essential part of the fusion model is the interaction between the sensor-specific modules, e.g. if fusion is performed at the pixel, feature, or object level.

- *A classification algorithm.* A classification rule must be defined to find a pixel labelling for the scene which is reasonable according to the data and the prior model. This is done by specifying a criterion function or loss function for the scene. An algorithm for optimization of the criterion function is also needed.

- *Parameter estimation procedures.* The parameters of different modules must be specified or estimated. For a complete multisource classification system, parameter estimation procedures should be provided. However, standard techniques for parameter estimation might not be applicable in a complex fusion model.

A methodological framework which allows us to merge all these elements together in a common model is Markov Random Fields. MRFs are frequently used for modelling spatial and/or temporal context [**?**, **?**]. A powerful property of these models is that the prior information and the observed data from different sensors can be easily combined through the use of suitable *energy functions*. The interaction between the different modules in the fusion model is defined in terms of a sum of sensor-specific energy functions.

# 4 Multisensor classification modelling

In this section, we present the general multisource classification model and the specific model for fusion of Landsat TM images, ERS-1 SAR images, and GIS ground cover data. In Table **??**, a list of notation used throughout this paper is given. In the general multisource or sensor fusion case, we have a set of measurements or images from $n$ sensors (Figure **??**). Let the $M \times N$ image from

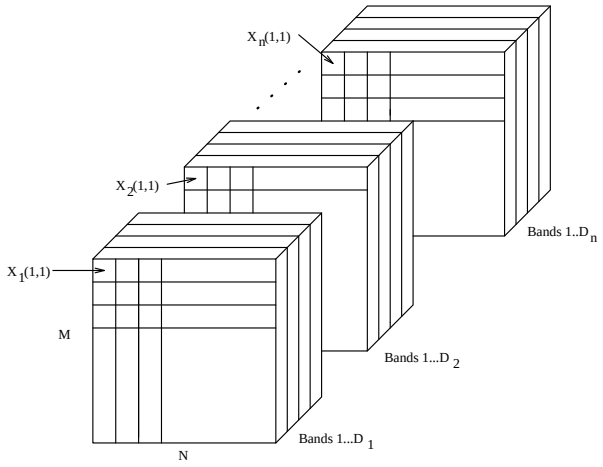


Figure 3: *Image data.*

sensor $s$ consist of $MN$ pixels or feature vectors $X_s(1,1), ..., X_s(M,N)$, $s = 1, 2, ..., n$, where $X_s(i,j) = (x_s(i,j,1), ..., x_s(i,j,D_s))$, and $D_s$ is the number of features or spectral bands for sensor $s$. The scene consists of $K$ true classes, $\omega_1, ..., \omega_K$, with prior probabilities $P(\omega_1), ..., P(\omega_K)$. We assume that the multisource images have been registered. The class of pixel $(i,j)$ is denoted by $\mathcal{C}(i,j)$. Let $X_s$ denote the set of pixels for the whole scene, that is, $X_s = \{X_s(i,j); 1 \leq i \leq M, 1 \leq j \leq N\}$, and $\mathcal{C} = \{\mathcal{C}(i,j), 1 \leq i \leq M, 1 \leq j \leq N\}$ the corresponding set of labels for the whole scene; $\mathcal{C}(i,j) \in \{\omega_1, \omega_2, ..., \omega_k\}$. Let $P(X_1, ..., X_n \mid \mathcal{C})$ denote the conditional probability density of feature vectors $X_1, ..., X_n$ given the scene labels $\mathcal{C}$. The task is to assign each pixel to the class that maximizes the posterior probabilities $P(\mathcal{C} \mid X_1, ..., X_n)$. In a Bayesian formulation, the relationship between the measurements and the prior information is given by

$$P(\mathcal{C} \mid X_1, ..., X_n) = \frac{P(X_1, ..., X_n \mid \mathcal{C}) P(\mathcal{C})}{P(X_1, ..., X_n)}, \tag{1}$$

where $P(\mathcal{C})$ is the prior model for the class labels, and $P(\mathcal{C} \mid X_1, ..., X_n)$ is the posterior probability.

| Symbol | Explanation |
|---|---|
| $M, N$ | Image dimensions |
| $X_s^t$ | Image from sensor $s$ at time $t$ |
| $n$ | Number of sensors |
| $D_s$ | Dimension of feature vector from sensor $s$ |
| $K$ | Number of classes |
| $P(\omega_k)$ | Prior probability for class $\omega_k$ |
| $\mathcal{C}^t(i,j)$ | Class label of pixel $(i,j)$ at time $t$ |
| $\mathcal{C}^t$ | Set of class labels for the scene at time $t$ |
| $l_g(i,j)$ | Line process for pixel $(i,j)$ for GIS field border map |
| $l_c(i,j)$ | Line process for pixel $(i,j)$ derived from the class labels $\mathcal{C}^t$ |
| $\alpha_s$ | Reliability factor for source $s$ |
| $\mathbf{X}_s^t$ | $\{X_s^r; r \leq t\}$, multitemporal image set |
| $\mathbf{C}^t$ | $\{\mathcal{C}^r; r \leq t\}$ |
| $\beta_{sp}$ | Parameter to control the spatial energy |
| $\beta_{temp}$ | Parameter to control the temporal energy |
| $\beta_g$ | Parameter to control the energy associated with the map |
| $Y_s, \theta_{kl}, \sigma^2, \mu_y$ | Parameters of the SAR texture model |
| $U(\cdot)$ | Energy functions |
| $V(\cdot)$ | Potential functions |

Table 1: *Legend of notation*

Below, we divide the fusion model into two parts. The fixed time module describes the fusion of data collected simultaneously, or within a short time interval where no class changes have occurred. The second part, the multitemporal module, describes fusion of images captured at different dates where changes in the classes are allowed.

### 4.1 Fusion of images sensed at the same time

The task here is to create a consensus interpretation of the scene by assigning each pixel to the class that maximizes the posterior probability $P(\mathcal{C} \mid X_1, ..., X_n)$, taking into account the reliability factor of each sensor. To simplify the modelling we use the following conditional independence assumption.

**Assumption 1:** $P(X_1, ..., X_n \mid \mathcal{C}) \equiv P(X_1 \mid \mathcal{C}) \cdots P(X_n \mid \mathcal{C})$.

This assumption means that we consider the measurements from the different sensors to be conditionally independent. This may not be valid for all applications, particularly for related sensors onboard the same space platform. However, for the task of fusing data from optical and microwave sensors on different satellites, this assumption seems more reasonable. Without it, the form of the joint conditional probability density, $P(X_1, ..., X_n \mid \mathcal{C})$ takes a very complex form which is difficult to estimate.

To combine data from sensors with different reliability or discrimination ability, we assign weights to sensor measurements according to the reliability factor of the sensor [**?**]. Thus, we seek to maximize the likelihood function

$$\mathcal{L}(X_1, ..., X_n \mid \mathcal{C}) = P(X_1 \mid \mathcal{C})^{\alpha_1} \cdots P(X_n \mid \mathcal{C})^{\alpha_n} P(\mathcal{C}), \tag{2}$$

where $\alpha_s$, $0 \leq \alpha_s \leq 1$, is defined as the reliability factor for source $s$. If source $s$ is totally unreliable, then $\alpha_s$ should be zero, so $P(X_s \mid \mathcal{C})^{\alpha_s} = 1$, and the source will not contribute to the likelihood function

$\mathcal{L}(X_1, ..., X_n, | \mathcal{C})$. For a highly reliable source, the reliability factor should be close to 1. We will use the overall classification accuracy for each single-source classifier as the value of its reliability factor.

We will use a Markov random field to model the spatial context in $P(\mathcal{C})$. For a more detailed description of Markov random fields in general and the specific model used here, see, e.g., [?, ?, ?]. Markov modelling of the spatial contextual term gives:

$$P(\mathcal{C}(i,j) \mid \mathcal{C}(k,l); \{k,l\} \neq \{i,j\}) = P(\mathcal{C}(i,j) \mid \mathcal{C}(k,l); \{k,l\} \in \mathcal{G}_{ij}) = \frac{1}{Z} e^{-U(\mathcal{C})/T}, \tag{3}$$

where $U$ is called the energy function, $Z$ is a normalizing constant, and $T$ is a constant that stands for "temperature", a term inherited from statistical physics. $\mathcal{G}_{ij}$ is the local neighborhood of pixel $(i,j)$. We will use a second-order neighborhood. Maximizing $P(\mathcal{C}(i,j) \mid \mathcal{C}(k,l); \{k,l\}$ is equivalent to minimizing $U(\mathcal{C})$. $U(\mathcal{C}(i,j))$ is given by

$$U(\mathcal{C}(i,j)) = \sum_{\{k,l\} \in \mathcal{G}_{ij}} \beta I(\mathcal{C}(i,j), \mathcal{C}(k,l)) \tag{4}$$

where

$$\begin{aligned} I(\mathcal{C}(i,j), \mathcal{C}(k,l)) &= -1 \text{ if } \mathcal{C}(i,j) = \mathcal{C}(k,l) \\ &= 0, \text{ if } \mathcal{C}(i,j) \neq \mathcal{C}(k,l), \end{aligned} \tag{5}$$

and $\beta$ is a parameter. This potential function corresponds to counting the number of pixels in the neighborhood assigned to the same class as pixel $(i,j)$. If $\beta = 0$, this reduces to a noncontextual model. As $\beta$ increases, more homogeneous image regions will be favored.

The multisensor classification algorithm consists of maximizing $\mathcal{L}(X_1, ...X_n \mid \mathcal{C})$, which is equivalent to minimizing

$$U(X_1, ...X_n, \mathcal{C}) = \sum_{s=1}^{n} \alpha_s U_{data}(X_s) + U_{sp}(\mathcal{C}), \tag{6}$$

where $U_{sp}$ is given by Eq. (??), and $\alpha_s$, $s = 1, ..., n$ are the sensor-specific reliability factors. For a complete multisensor classifier, specification of the sensor-specific class-conditional energy function $(U_{data}(X_s))$ is needed (see Section ??).

## 4.2 Fusion of multitemporal images

To extend the model to incorporate the temporal aspect, let $X_s^t = \{X_s^t(i,j)\}$ represent the measurements from sensor $s$ at time $t$. Allowing changes with respect to the classes, let $\mathcal{C}^t = \{\mathcal{C}^t(i,j)\}$ represent the class labels at time $t$. Define $\mathbf{X}_s^t = \{X_s^r; r \leq t\}$ and $\mathbf{C}^t = \{\mathcal{C}^r; r \leq t\}$. At each time $t$, we seek a recursive estimate of the labels $\mathbf{C}^t$ based on the measurements, $\mathbf{X}_s^t$, up to that time. Thus, we maximize

$$P(\mathbf{C}^t \mid \mathbf{X}_s^t) = \frac{P(\mathbf{C}^t, \mathbf{X}_s^t)}{P(\mathbf{X}_s^t)}. \tag{7}$$

A recursive scheme allows the images to be analyzed frame by frame, in the order in which they are obtained. Also, processing one frame at a time keeps the computational requirements at a manageable level compared to processing the frames simultaneously.

The following derivation of recursive class label estimates is according to Green and Titterington [?]. To make the model tractable, we assume

**Assumption 2:** $P(X_s^t \mid \mathbf{C}^t, \mathbf{X}_s^{t-1}) \equiv P(X_s^t \mid \mathcal{C}^t)$.

**Assumption 3:** $P(\mathcal{C}^t \mid \mathbf{C}^t) \equiv P(\mathcal{C}^t \mid \mathcal{C}^{t-1})$.

Assumption 2 implies that the noise associated with a particular image frame is independent of the noise or class labels at different times. Assumption 3 corresponds to a first-order Markov assumption for the class labels over time. The two assumptions give

$$P(\mathbf{C}^t, \mathbf{X}_s^t) = (\prod_{m=1}^{t} P(X_s^m \mid \mathcal{C}^m) P(\mathcal{C}^m \mid \mathcal{C}^{m-1})) P(\mathcal{C}^0), \tag{8}$$

where $\mathcal{C}^0$ is the initial labelling of the scene. Then, a recursive estimate of $\mathcal{C}^t$, given the previous labels and current measurements, is:

$$P(\mathcal{C}^t \mid \mathbf{C}^{t-1}, \mathbf{X}_s^t) \propto P(\mathcal{C}^t \mid \mathcal{C}^{t-1}) P(X_s^t \mid \mathcal{C}^t). \tag{9}$$

The true labels $\mathcal{C}^{t-1}$ are unknown, so the classification algorithm will be based on the estimated labels $\hat{\mathcal{C}}^{t-1}$

$$P(\mathcal{C}^t \mid \hat{\mathcal{C}}^{t-1}, \mathbf{X}_s^t) \propto P(\mathcal{C}^t \mid \hat{\mathcal{C}}^{t-1}) P(X_s^t \mid \mathcal{C}^t). \tag{10}$$

Using the estimated labels $\hat{\mathcal{C}}^{t-1}$ in the classification algorithm is a suboptimal solution which ignores the uncertainty in the estimates $\hat{\mathcal{C}}^{t-1}$. An alternative would be to base the classification on

$$P(\mathcal{C}^t \mid \mathbf{X}_s^t) = \sum_{\mathcal{C}^{t-1}} P(\mathcal{C}^t, \mathcal{C}^{t-1} \mid \mathbf{X}_s^t), \tag{11}$$

but with computational aspects in mind we chose the simpler approximation in Eq. (**??**).

### 4.2.1 Specification of the temporal contextual term

We assume a Markov dependence of the class labels at time $t$ to specify the temporal contextual term $P(\mathcal{C}^t \mid \mathcal{C}^{t-1})$:

**Assumption 4:** $P(\mathcal{C}^t(i,j) \mid \mathcal{C}^{t-1}, \mathcal{C}^t(k,l); \{k,l\} \in \mathcal{G}_{ij}^t) \equiv P(\mathcal{C}^t(i,j) \mid \mathcal{C}^{t-1}(q,r), \mathcal{C}^t(k,l),$
$\{k,l\} \in \mathcal{G}_{ij}^t, \{q,r\} \in \mathcal{G}_{ij}^{t-1})$,

where $\mathcal{G}_{ij}^{t-1}$ includes pixel $(i,j)$ and its second-order neighbors, all in frame $t-1$ (see Figure **??**).

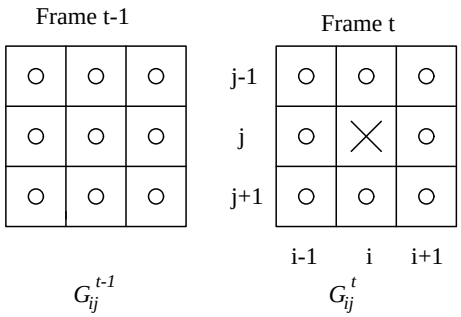


Figure 4: Spatio-temporal neighborhood of pixel $(i,j)$.

Rewriting this expression in terms of energy functions and Gibbs random fields gives:

$$(12) \qquad P(\mathcal{C}^t(i,j) \mid \mathcal{C}^{t-1}(q,r), \mathcal{C}^t(k,l); \{k,l\} \in \mathcal{G}_{ij}, \{q,r\} \in \mathcal{G}_{ij}^{t-1}) = \frac{1}{Z} e^{U(\mathcal{C}^t(i,j), \mathcal{C}^{t-1}(q,r), \mathcal{C}^t(k,l))/T}.$$

The energy function $U(\cdot)$ of Eq. (**??**) is still difficult to specify because it involves both spatial and temporal contextual information. For simplicity, we assume that the spatial and time-dependent aspects of $\mathcal{C}^t$ are separable.

**Assumption 5:** $U(\mathcal{C}^t(i,j), \mathcal{C}^{t-1}(q,r), \mathcal{C}^t(k,l)) \equiv$
$U_{sp}(\mathcal{C}^t(i,j), \mathcal{C}^t(k,l)) + U_{temp}(\mathcal{C}^t(i,j), \mathcal{C}^{t-1}(q,r))$.

Let the spatial context be

$$(13) \qquad U_{sp}(\mathcal{C}^t(i,j), \mathcal{C}^t(k,l)) = \beta_{sp} \sum_{\{k,l\} \in \mathcal{G}_{ij}} I(\mathcal{C}^t(i,j), \mathcal{C}^t(k,l))$$

and the temporal context be

$$(14) \qquad U_{temp}(\mathcal{C}^t(i,j), \mathcal{C}^{t-1}(q,r)) = -\beta_{temp} \sum_{\{q,r\} \in \mathcal{G}_{ij}^{t-1}} Pr(\mathcal{C}^t(i,j) \mid \mathcal{C}^{t-1}(q,r)).$$

The indicator function $I(k,l)$ is given by Eq. (**??**) and $Pr(\mathcal{C}^t(i,j) \mid \mathcal{C}^{t-1}(q,r))$ is the transition probability for a change from class $\mathcal{C}^{t-1}(q,r)$ to class $\mathcal{C}^t(i,j)$ between time $t-1$ and time $t$. $\beta_{sp}$ and $\beta_{temp}$ are user-specified parameters. Thus, the multitemporal fusion algorithm is based on minimizing

$$(15) \qquad U(X_1, ..., X_n, \mathcal{C}^t, \mathcal{C}^{t-1}) = \sum_{s=1}^{n} \alpha_s U_{data}(X_s, \mathcal{C}^t) + U_{sp}(\mathcal{C}^t) + U_{temp}(\mathcal{C}^t, \mathcal{C}^{t-1}),$$

where $U_{sp}$ and $U_{temp}$ are given by Eqs. **??** and **??**, respectively, and $U_{data}$ is the sensor-specific image statistic function, which will be defined in Section **??**.

## 4.3 Inclusion of GIS data

Different types of GIS data can be included in the model. We now describe how ground cover maps or crop field border maps can be modelled in the same general framework.

### 4.3.1 Ground cover map

The multisensor classifier can be extended to incorporate GIS ground cover data. Assume that $\mathcal{C}^g = \{\mathcal{C}^g(i,j); 1 \le i \le M, 1 \le j \le N\}$ represents a ground cover map of the area, co-registered with the image data, with ground cover types corresponding to the set of desired classes $\omega_1, ..., \omega_K$. The classification of the scene at time $t$ is to be based on all current and previous images and previous ground cover classifications,

$$(16) \qquad P(\mathcal{C}^t \mid \hat{\mathcal{C}}^{t-1}, \mathbf{X}_s^t, \mathcal{C}^g) \propto P(\mathcal{C}^t \mid \hat{\mathcal{C}}^{t-1}, \mathcal{C}^g) P(X_s^t \mid \mathcal{C}^t).$$

In accordance with $P(\mathcal{C}^t \mid \hat{\mathcal{C}}^{t-1})$, we assume a Markov dependency of the class labels and rewrite in terms of energy functions:

$$(17) \qquad P(\mathcal{C}^t(i,j) \mid \mathcal{C}^t(k,l), \mathcal{C}^{t-1}(q,r), \mathcal{C}^g(q,r)) = \frac{1}{Z} e^{U(\mathcal{C}^t(i,j), \mathcal{C}^t(k,l), \mathcal{C}^{t-1}(q,r) \mathcal{C}^g(q,r))/T}$$

Again, we assume that the spatial and time-dependent aspects of $\mathcal{C}^t$ are separable, and that the temporal variations in the image are independent of the temporal variations concerning the map data:

$$U(\mathcal{C}^t(i,j),\mathcal{C}^t(k,l),\mathcal{C}^{t-1}(q,r),\mathcal{C}^g(q,r)) = U_{sp}(\mathcal{C}^t(i,j),\mathcal{C}^t(k,l)) + U_{temp}(\mathcal{C}^t(i,j),\mathcal{C}^{t-1}(q,r)) + U_{\mathrm{GIS}}(\mathcal{C}^t(i,j),\mathcal{C}^g(q,r)). \quad (18)$$

$U_{sp}(\cdot)$ and $U_{temp}(\cdot)$ are given by Eqs. (??) and (??), respectively, and the temporal context from the map is

$$U_{\mathrm{GIS}}(\mathcal{C}^t(i,j),\mathcal{C}^g(q,r)) = -\beta_g \sum_{\mathcal{G}^g_{ij}} Pr(\mathcal{C}^t(i,j) \mid \mathcal{C}^g(q,r)), \quad (19)$$

where $Pr(\mathcal{C}^t(i,j) \mid \mathcal{C}^g(q,r))$ is the transition probability for a change from class $\mathcal{C}^g(q,r)$ to class $\mathcal{C}^t(i,j)$ between the creation of the map and the current time $t$, and $\beta_g$ is a parameter. The multitemporal fusion including GIS data is based on minimizing

$$U(X_1,...,X_n,\mathcal{C}^t,\mathcal{C}^{t-1},\mathcal{C}^g) = \sum_{s=1}^{n} \alpha_s U_{data}(X_s,\mathcal{C}^t) + U_{sp}(\mathcal{C}^t) + U_{temp}(\mathcal{C}^t,\mathcal{C}^{t-1}) + U_{\mathrm{GIS}}(\mathcal{C}^t,\mathcal{C}^g). \quad (20)$$

### 4.3.2 Crop ownership data

If crop ownership data or field border maps are available, field-by-field classification is possible. However, field-by-field classification is based on the assumption that each field contains only one crop type, and that the whole field is harvested or processed at the time of image acquisition. This assumption might not be valid, so these types of hard constraints are not always desirable. Instead, softer constraints can be used, in which pixels within the same field are encouraged to be assigned to the same class, and penalized if they are assigned to different classes. This can readily be accomplished in a MRF model by the use of a line process. Geman and Geman [?] introduced the line process to smooth an image except at discontinuities in grey levels. Let the field border map be represented by a line process $L_g = \{l_g(i,j), 1 \le i \le M, 1 \le j \le N\}$ (see Figure ??). An edge site in the lattice indicates whether the two pixels it separates are in the same field ($l_g(i,j) = 0$), or if a field border is present between them ($l_g(i,j) = 1$). Also attach a line process $L_c = \{l_c(i,j), 1 \le i \le M, 1 \le j \le N\}$ to the class labels of the scene by letting $l_c(i,j) = 0$ if the two pixels it separates are assigned to the same class, and $l_c(i,j) = 1$ if they are assigned to different classes. Note that the labels $l_c(i,j)$ are not estimated directly; they are derived from the class labels $\mathcal{C}^t$. We introduce a coupling term between the two line processes:

$$W[l_c(i,j), l_g(i,j)] = \begin{cases} 0 & \text{if } l_c(i,j) = l_g(i,j) \\ 1 & \text{otherwise.} \end{cases} \quad (21)$$

The coupling term $W[l_c(i,j), l_g(i,j)]$ can be interpreted as an additional penalty term which increases the energy function if an edge site has different values in the map and the classified scene, e.g., if two pixels in the same field are assigned to different classes. This results in the following energy function to be minimized:

$$U(X_1,...,X_n,\mathcal{C}^t,\mathcal{C}^{t-1},L_g) = \sum_{s=1}^{n} \alpha_s U_{data}(X_s,\mathcal{C}) + U_{sp}(\mathcal{C}^t) + U_{temp}(\mathcal{C}^t,\mathcal{C}^{t-1}) + U_{l_g}(L_c,L_g), \quad (22)$$

where

$$U_{l_g}(l_c(i,j), l_g(i,j)) = -\beta_g \sum_{\mathcal{G}^g_{ij}} W(l_c(i,j), l_g(i,j)). \quad (23)$$

The summation in Eq. (??) is over all edge sites in the neighborhood.

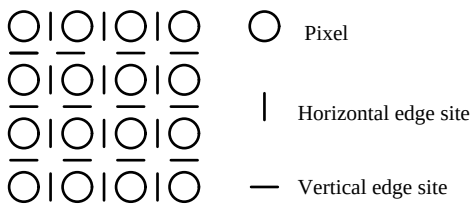


Figure 5: Pixel and line (edge) sites

### 4.4 Classification algorithms

The classification of a scene involves assigning class labels to all the image pixels. One would like to obtain a pixel labelling that is reasonable according to the data and the prior model. A popular criterion is to find the labelling that maximizes the *posterior* distribution (MAP). This is equivalent to a minimum-risk solution with a 0-1 loss function, and to a minimum-energy solution. The MAP estimate is the mode of the posterior distribution for $\mathcal{C}$, given $X_1, ..., X_n$. Finding the MAP estimate requires minimization of a function of $M \times N$ variables. Simulated annealing can be used to find an iterative solution to this problem [**?**]. The computational demands of the simulated annealing algorithm are well known.

A computationally feasible alternative to the simulated annealing algorithm is Besag's *Iterated Conditional Modes (ICM)* algorithm [**?**], which converges to a local minimum of the energy function. The basic idea behind the ICM algorithm is the following relationship between the class labels for pixel $(i, j)$, the current estimate of the labels for the neighboring pixels, and the sensed image $X$:

$$P(\mathcal{C}(i,j) \mid X, \mathcal{C}(k,l); \{k,l\} \in \mathcal{G}_{ij}) \propto P(X(i,j) \mid \mathcal{C}(i,j)) P(\mathcal{C}(i,j) \mid \mathcal{C}(k,l))$$

Another alternative optimization criterion is *Maximizing of the Posterior Marginals (MPM)* [**?**], which is based on minimizing a loss function describing segmentation error. The MPM approach is similar to simulated annealing without any cooling schedule. A more detailed discussion of different algorithms can be found in [**?**]. Dubes and Jain [**?**] reported only small differences in the segmentation performance of the different algorithms. However, for segmentation of complex SAR images, Rignot and Chellappa [**?**] found that MAP estimation and MPM estimation performed significantly better than the ICM algorithm.

In a preliminary experiment using a limited test data set, we compared the error rates for MAP estimation based on simulated annealing and the ICM algorithm. No significant difference in error rates were found. Due to the high computational complexity of the simulated annealing procedure and the similar error rates observed in the preliminary experiment, only the ICM algorithm was used in the full scale experiments reported here.

### 4.5 Image statistic models

To specify the particular fusion model used in our experiments, the image statistic models for Landsat TM, and ERS-1 SAR, and airborne MAESTRO multipolarization SAR images must be defined.

### 4.5.1 Landsat TM image statistic

The six non-thermal bands of Landsat TM images were used for this study. The pixel resolution of the geocoded image in UTM projection was $30m \times 30m$. The TM image was not severely degraded by noise, and was well modelled by a multivariate normal distribution (based on inspection of the univariate histograms). With computational aspects in mind, we will only include interpixel feature correlation. Thus, we assume that $P(X_s \mid \mathcal{C}) = \prod_{(i,j)} P(X_s(i,j) \mid \mathcal{C}(i,j))$. $P(X_s(i,j) \mid \mathcal{C}^t(i,j))$ can be rewritten in terms of energy functions based on the multivariate normal distribution

$$U_{data}(X_s(i,j),\mathcal{C}^t(i,j)) = \frac{D_s}{2} ln \mid 2\pi\Sigma_k \mid + \frac{1}{2}(X_s(i,j) - \mu_k)^T \Sigma_k^{-1}(X_s(i,j) - \mu_k). \tag{24}$$

$\Sigma_k$ and $\mu_k$ are the class-conditional covariance matrix and mean vector for class $k$, respectively, and $D_s$, $s = 1, ..., n$ is the number of spectral bands for source $s$.

### 4.5.2 ERS-1 SAR image statistic

To model the image statistic of ERS-1 SAR images, we could either use a probability density function or a more complex model describing the speckle statistics and autocorrelation function. For both approaches, a number of different methods have been reported [?, ?, ?, ?, ?]. The ERS-1 images used in this study were amplitude SAR images processed at Tromsø Satellite Station, Norway. For amplitude images, no simple expression of the resulting speckle probability distribution can be written down [?].

Another complicating factor was the need to co-register the SAR images with the TM images and map data. A geometric transformation of the images alters the spatial speckle statistics, so we did not attempt an exact modelling of speckle statistics, but used the more general multiplicative autoregressive random field model (MAR) [?], which incorporates textural information. Previous studies indicate that the use of textural information, in addition to the backscatter values themselves, improves the discrimination ability of the classifier [?]. The parameters of the model are used as texture descriptors. For details of the texture model, see [?, ?]. The parameters were computed from the original SAR image. The multivariate normal distribution is used to model the feature vectors.

### 4.5.3 Image statistic for MAESTRO amplitude images

A 4-look C-band amplitude SAR image with the polarizations HH, HV, and VV from the MAESTRO program [?] was used in this study. Only the amplitude images of different polarizations were available to us, thus statistical models for classification of fully polarimetric data (see, e.g., [?]) could not be applied. Each of the intensities $|HH|^2$, $|HV|^2$, and $|VV|^2$ is known to follow the Gamma distribution [?]. The co-polarized and cross-polarized intensities are normally assumed to be uncorrelated [?], but this is not the case for the co-polarized intensities $|HH|^2$ and $|VV|^2$. The joint distribution of $|HH|^2$ and $|VV|^2$ becomes more complex. The image statistic of a 4-look image matches the theoretical distribution of a 3-look image because the four looks are not completely independent [?]. As the number of looks increases, the Gamma distribution will approach a Gaussian distribution. We compared the classification error rates based on the 3-look Gamma distribution and the Gaussian distribution for each of the three single intensities $|HH\ |^2$, $|HV|^2$, and $|VV\,|^2$. No significant increase in classifier performance was found for the Gamma classifier.

Speckle filtering has been shown to improve the classification accuracy significantly (see, e.g., [?]). We found the effect of preprocessing in terms of speckle reduction to be much larger than the effect

**Algorithm for fixed date multisource classification by ICM method**

**(1)** Initialize $\hat{\mathcal{C}}$ for each pixel by choosing $\mathcal{C}(i,j)$ as the class that minimizes the non-contextual energy function $\sum_{s=1}^{n} \alpha_s U_{data}(X_s(i,j), \mathcal{C}(i,j))$, where $U_{data}(\cdot)$ is given by Eq. (**??**).

**(2)** For all pixels $(i,j)$, update $\hat{\mathcal{C}}(i,j)$ by the class $\mathcal{C}(i,j)$ that minimizes Eq. (**??**).

**(3)** Repeat **(2)** $N_{iter}$ times.

Figure 6: *Algorithm for fixed date classification.*

**Algorithm for multitemporal multisource classification by ICM method**

**(1)** At time $t_0$, initialize $\hat{\mathcal{C}}^0$ by, for each pixel, choosing the class that minimizes $\sum_{s=1}^{n} \alpha_s U_{data}(X_s(i,j), \mathcal{C}(i,j))$, where $U_{data}(\cdot)$ is given by Eq. (**??**).

**(2)** For time $t = t_k$:

**(3a)** No GIS data:
For all pixels $(i,j)$, update $\hat{\mathcal{C}}^t(i,j)$ by the class $\mathcal{C}^t(i,j)$ that minimizes Eq. (**??**).

**(3b)** With GIS ground cover map:
For all pixels $(i,j)$, update $\hat{\mathcal{C}}^t(i,j)$ by the class $\mathcal{C}^t(i,j)$ that minimizes Eq. (**??**).

**(3c)** With GIS field border map:
For all pixels $(i,j)$, update $\hat{\mathcal{C}}^t(i,j)$ by the class $\mathcal{C}^t(i,j)$ that minimizes Eq. (**??**).

**(4)** Repeat (3) $N_{iter}$ times.

**(5)** Repeat (2) - (4) for times $t = t_1, ..., t_{final}$.

Figure 7: *Algorithm for multitemporal classification.*

of choosing a particular distribution function for the backscatter values. Adaptive filters for speckle reduction often smooth homogeneous regions in the image significantly, while retaining high-contrast regions resulting in an image with edges and heterogeneous regions preserved. In a preliminary experiment, the classification accuracies were computed after filtering by the mean filter, the adaptive Lee-filter [**?**], and the adaptive sigma filter. No significant differences were found, although the mean filter performed best. The averaging of $m$ pixels in a 4-look image will result in an image with the same type of distribution as the original, but where the parameter corresponding to the number of looks ($N = 4$) is replaced by $4m$. (The effective number of looks will be slightly smaller because the averaged pixels are not independent.) With a $7 \times 7$ window size, the distribution of the filtered image will be close to the normal distribution. Thus, we chose to model the multipolarization image by a multivariate normal distribution. The degree of texture in the images has not been investigated, but agricultural areas are normally quite homogeneous.

### 4.5.4 The complete multisource classification algorithm

Having specified the sensor-specific image statistic functions for Landsat TM, ERS-1 SAR images, and MAESTRO multipolarization SAR images, the ICM multisource classification algorithms for the fixed date mode and the temporal mode are given in Figures **??** and **??**.

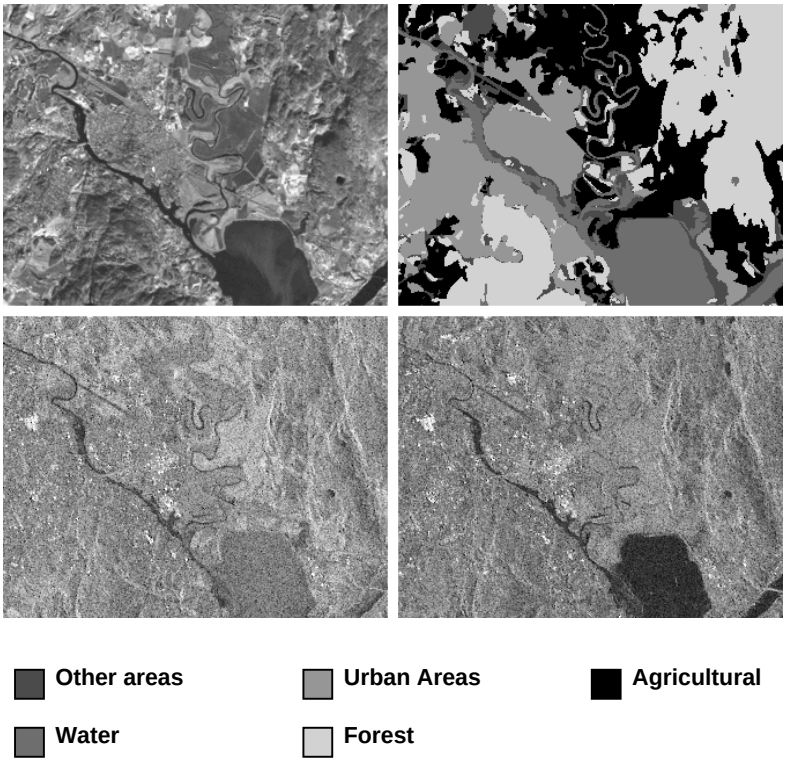


Figure 8: *Original images and ground cover maps. (a) Landsat TM image, Oct. 20, 1991; (b) Ground cover map; (c) ERS-1 SAR image, Oct. 17, 1991; (d) ERS-1 SAR image, Nov. 19, 1991. The signature variations between the two SAR images are due to different weather conditions. (ESA/EURIMAGE/TSS/Satellitbild)*

# 5 Experimental results

To test the performance of the multisource classification model, two data sets were used. The first task consisted of land-use classification based on Landsat TM and ERS-1 SAR images, and the second task was crop classification based on Landsat TM and airborne multipolarization SAR images.

## 5.1 Land-use classification based on Landsat TM and ERS-1 SAR images

A data set consisting of one Landsat TM image from Oct. 20, 1991, and three ERS-1 SAR images from Oct. 17, Oct. 20, and Nov. 19, 1991, of Kjeller, Norway was used (Figure **??**). A flat test site was chosen due to lack of a geocoding facility for ERS-1 SAR images in Norway. The data set was acquired for a project which monitored plowing in agricultural areas in the fall. Runoff from fields tilled in the fall is a major source of pollution. Due to the normally high cloud coverage and the lighting conditions in Norway in the late fall, the chance of getting a cloud-free Landsat TM scene in the short time period after the tilling activities have ended and before the ground is covered by snow is low. Although ERS-1 SAR images alone may not provide enough information to monitor changes in tilling, they may be used in combination with a Landsat TM image acquired on a cloud-free day earlier in the fall to detect the changes. The ability of ERS-1 SAR images to discriminate between different crop classes is dependent on factors like soil moisture and temperature.

A five-class ($K = 5$) classification problem was considered, which included the more common ground

cover classes water, urban areas, forests, and two classes of agricultural areas: plowed and unplowed. Between the acquisitions of different images, a large portion of the agricultural fields had been tilled. Ground control samples from the agricultural areas were available for all the image acquisition dates.

The SAR images were co-registered with the TM image, which was available in UTM projection. Large variations in the SAR signature due to varying temperature and soil moisture content for land areas, and wind conditions for water areas were observed in the SAR images.

To check the effect of including map data in classification, a topographic map in the series M711 was digitized and transformed to the UTM projection. The GIS data originated from two map revisions, each covering only a part of the area. The original maps were revised in 1976 and 1988, respectively. This means that the GIS data for parts of the scene were inaccurate with respect to changes in the area usage after 1976. During this period, some areas changed from forest/agricultural type to urban type.

The reported error rates for the classifier are averages over ten experiments. Each experiment consisted of partitioning the control data randomly into a training set and a test set of approximately the same size. The numbers of training samples and test samples were typically 1300 for each of the classes urban areas, water, and forests, and varied from 200 to 1200 for the two agricultural classes.

The performance of the proposed MRF fusion model was compared to the fusion model of [**?**, **?**]. The simpler reference model performs fusion on a pixel-by-pixel basis by merging the posterior probabilities from sensor-specific classifiers, taking into account the reliability factors and the likelihood of class changes. Table **??** shows the overall classification error rates for the two fusion models. For fusion of images acquired at the same time, the MRF fusion model resulted in an average error rate of $3.1 \pm 0.1\%$ (an average error of 3.1%, with a standard deviation of 0.1), compared to $6.3 \pm 0.3\%$ for the reference model. With the inclusion of GIS data, the corresponding error rates were reduced to $1.4 \pm 0.1\%$ for the MRF model and $2.9 \pm 0.2\%$ for the reference model. For the temporal fusion model, the average error rate for the MRF model was 5.6%, compared to 7.2% for the reference model. The average error rates with GIS data included were 2.7% for the MRF model and 4.4% for the reference model, respectively.

| Images fused | Reference method | MRF method |
|---|---|---|
| TM and SAR Oct. 20, 1991 | $6.3 \pm 0.3$ % | $3.1 \pm 0.1\%$ |
| TM, SAR Oct. 20, 1991 and GIS | $2.9 \pm 0.2\%$ | $1.4 \pm 0.1\%$ |
| TM and SAR Oct. 17, 1991 | $8.1 \pm 0.4\%$ | $6.9 \pm 0.2\%$ |
| TM, SAR Oct. 17, 1991 and GIS | $4.8 \pm 0.3\%$ | $3.1 \pm 0.1\%$ |
| TM and SAR Nov. 19, 1991 | $6.4 \pm 0.2\%$ | $4.3 \pm 0.3\%$ |
| TM, SAR Nov. 19, 1991 and GIS | $3.9 \pm 0.2\%$ | $2.5 \pm 0.2\%$ |

Table 2: Performance of the fusion models. Classification error rates (in %) are averages over 10 experiments.

The classification accuracies for different classes are given in Table **??**. For fusion of the Oct. 20 images, the MRF method improved the classification accuracy for water areas and urban areas compared to the reference method. For the temporal fusion model, the main difference between the MRF fusion method and the reference method was the improved classification accuracy for urban areas and agricultural areas (plowed and unplowed). By including GIS data, the classification accuracy improved for all the classes in general, but the improvements for the agricultural classes were minor. This was to be expected, because the GIS data did not contain any information about the agricultural processing of the fields.

The parameter values used in the MRF fusion model are given in Table **??**. These parameter values

were determined experimentally. For the reliability factors $\alpha_{SAR}$ and $\alpha_{TM}$, the overall single-source classification accuracy was used. The commonly reported value of 1.5 [**?**] for the spatial contextual factor $\beta_{sp}$ performed well for the fixed or constant time fusion model. The temporal fusion model involves additional terms, and the $\beta_{sp}$ value needs to be adjusted for the temporal case to achieve the best performance ($\beta_{sp} = 0.5$ for the temporal model). Adjustments were also necessary for the $\beta_g$ parameter. Transition probabilities of 0.5 between the two agricultural classes for the image data and of 0.1 between the forest/agricultural areas and urban areas for the GIS map were used. For the reference model, the parameter values used are given in [**?**].

The ability of the two fusion models to detect changes was tested by comparing the classification of the agricultural fields for the Oct. 17 image and the Nov. 19 image with the Oct. 20 TM image. The results are shown in Table **??**. On an average, the accuracy in detecting the actual changes in plowing was 75% for the MRF model, compared to 62% for the reference model.

### 5.2 Crop classification based on Landsat TM and multipolarization SAR images

A 4-look C-band SAR image from the MAESTRO survey [**?**] from August 16, 1989, and a Landsat TM image from August 23, 1989 of Feltwell, UK, were used. Only the amplitude image consisting of the three polarizations $|HH|$, $|HV|$, and $|VV|$ were available for our study. Figure **??** shows the SAR image corresponding to $HH$ polarization and the Landsat TM image (band 4). The two images were manually co-registered. The SAR images were calibrated, but they still contained intensity variations in the range direction because of varying incidence angles. A common approach to compensate for the incidence angle effect is to normalize with respect to the total power [**?**], but this eliminates most of the information contained in the power intensity. Discrimination between two objects with similar scattering characteristics, but different backscattering strength will then be difficult. Instead, we chose to partition the image into five subimages corresponding to different ranges, and then to train and classify each subimage separately.

The classification problem considered involved the following agricultural classes: wheat, sugar beet, potatoes, carrots, grass, stubble, and bare soil. Crop maps from August 16 and 22, 1989, were used to train and test the classifier [**?**]. Between the two acquisition dates, a number of the fields had been harvested. The number of training pixels for each class was typically in the range 1000 to 4000, except for potatoes in one of the SAR subimages, where only 85 pixels were used.

Table **??** shows the classification error rates for the single-source classifiers, the MRF fusion model, and the reference model. Reliability factors of $\alpha_s = 0.6$ for SAR and $\alpha_{TM} = 0.8$ for Landsat TM were used, and $\beta_g$ was set to 10 to control the influence of the field border map, while the other parameter values were the same as given in Table **??**. From the changes in the field maps, class transition probabilities were estimated. The overall classification accuracy for the August 16, 1989, SAR image was 59.9%, while it was 70.3 % for the August 23, 1989, Landsat TM image. The reference fusion model improved the classification accuracy slightly to 71.3%, while the multitemporal MRF fusion model resulted in 73.0 % correct classification. By fusing C-band SAR with Landsat TM images, the overall classification accuracy improved by 2.7%, compared to Landsat alone. This relatively small improvement in accuracy can be explained by the suboptimal frequency, with respect to crop type separation, used in C-band SAR. With the inclusion of the crop field border map, the classification accuracy increased significantly to 79.6%.

Figure **??** shows classification maps of a selected subscene for single-source classification based on Landsat TM (upper left), the reference fusion model (upper right), the MRF fusion model (lower left),

and the MRF model with GIS field borders (lower right). The classification maps based on Landsat TM alone and the reference fusion model appear very similar; they both contain some small, noisy regions. For the MRF fusion model, some of the noise has been removed and the classification map appears cleaner. The classification map produced by including the field map is clearly more consistent than the one produced without map data. To a large extent, the pixels within each field were assigned to the same class, and significant details were preserved.

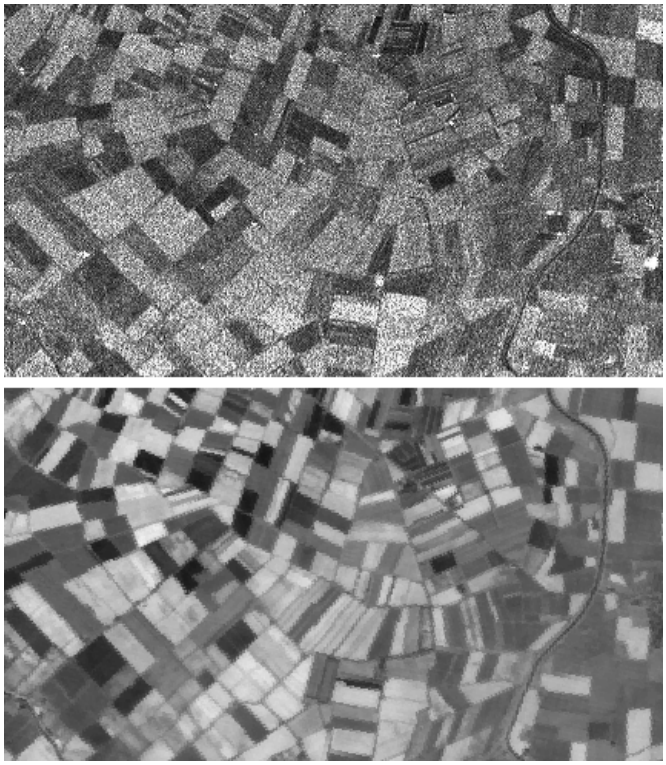

Figure 9: *Images of the Feltwell, UK, scene. MAESTRO SAR image, August 16, 1989, HH polarization (upper part); Landsat TM image, August 23, 1989 (Band 4 is shown) (lower part). (JRC/ESA)*

## 6 Discussion

In this paper, we have presented a Markov Random Field model for multisource classification of remotely sensed data. The method exploits spatial class dependencies between neighboring pixels in an image, and temporal class dependencies between different images. By including the temporal aspect of the data, the model is suitable for detecting class changes between the acquisition of different images. GIS data can also be incorporated into the classification process.

The performance of the proposed model was evaluated by fusing Landsat TM images, ERS-1 SAR images, and GIS ground cover data for land-use classification, and by fusing MAESTRO multipolarization SAR images with Landsat TM images and a GIS crop field border map for crop classification. It was compared to a simpler fusion model [?], and the MRF method resulted in slightly higher (2%) classification accuracy, on average, when the same data were used as input to the two models. When GIS field border data was included in the MRF model, the classification accuracy of the MRF model improved by 8%. For change detection in agricultural areas, 75% of the actual class changes were detected by the MRF model, compared to 62% for the reference model.

The MRF model provides a framework for integrating various types of spatial data. This was demonstrated by including a crop field border map in the classification process. By using line processes to encourage the classifier to assign pixels within the same field to the same class, significant improvements were found both in term of visual interpretation of classification maps and classification accuracy.

The improvement in classification accuracy obtained by fusing C-band SAR data with a more recent Landsat TM image was minor. Larger improvements for agricultural studies might be achieved by fusing optical data with SAR images of other frequencies, e.g. P-band SAR. However, we have demonstrated a good potential for fusing a more recent SAR image with prior Landsat TM images and using a multitemporal fusion model allowing class changes. This approach can be very useful for monitoring agricultural areas in northern areas like Norway, where the availability of cloud-free optical images is limited due to high cloud coverage and lighting conditions.

The MRF fusion model is related to the fusion model of [?, ?] and to the multitemporal classifier of [?]. The model is based on many of the concepts which were presented in [?, ?], but MRF modelling allows a stronger coupling between the different source-specific elements in the model. Fusion is performed on an intermediate level for pixels in a local neighborhood, compared to the pixel-by-pixel fusion reported in [?, ?]. The multitemporal classifier of [?] is also based on MRF modelling. It may also be used as a multisensor classifier, but since it does not treat the multisensor aspect, one would not expect it to perform well on data from sensors with different reliability.

The MRF fusion model involves a number of parameters controlling the contribution of different information sources, and spatial and temporal contextual information. In this paper, we have not discussed estimation of these parameters. The parameter values used were determined empirically in a limited experimental study. The sensitivity to changes in the parameter values needs to be further investigated. An automatic parameter estimation method for the fusion model is desirable, but difficult to establish.

To keep the computational complexity of the model at an acceptable level, we selected the ICM classification algorithm. The ICM algorithm is suboptimal and converges only to a local minimum of the energy function [?]. An alternative algorithm which provides a solution that is closer to the optimal solution, but which is computationally more attractive, is a multiresolution algorithm based on the MPM criterion [?].

For efficient use of multisource data in regular remote sensing applications, the problem of automatic image registration must be solved. The establishment of common repositories/data bases for sharing data would increase the interest and potential for multisource applications. A large-scale study of fusion models should also be performed.

Combining the well-founded theoretical basis of Markov Random Field models for classification tasks and the encouraging experimental results in our small-scale study, we conclude that the proposed MRF model is useful for classification of multisource satellite imagery.

**Acknowledgement**
This work was supported by a grant from The Research Council of Norway. We are grateful to the State Pollution Control Authority (SFT) and the Norwegian Space Center for providing the ERS-1 SAR images. The MAESTRO SAR images were kindly made available to us by NORUT-IT and JRC/ESA.

| Images fused | Class | Reference method | MRF method |
|---|---|---|---|
| TM and SAR Oct. 20, 1991 | 1 | $95.2 \pm 1.0\%$ | $99.7 \pm 0.1\%$ |
| | 2 | $88.6 \pm 0.6\%$ | $95.2 \pm 0.4\%$ |
| | 3 | $97.6 \pm 0.4\%$ | $98.4 \pm 0.3\%$ |
| | 4 | $96.3 \pm 0.9\%$ | $96.8 \pm 0.6\%$ |
| | 5 | $90.7 \pm 1.0\%$ | $91.6 \pm 0.9\%$ |
| TM, SAR Oct. 20, 1991 and GIS | 1 | $99.7 \pm 0.1\%$ | $100.0 \pm 0.0\%$ |
| | 2 | $95.5 \pm 0.3\%$ | $99.9 \pm 0.1\%$ |
| | 3 | $98.7 \pm 0.3\%$ | $99.5 \pm 0.1\%$ |
| | 4 | $95.8 \pm 0.8\%$ | $96.0 \pm 0.6\%$ |
| | 5 | $91.2 \pm 0.9\%$ | $93.0 \pm 0.9\%$ |
| TM and SAR Oct. 17, 1991 | 1 | $91.9 \pm 1.0\%$ | $90.0 \pm 0.5\%$ |
| | 2 | $90.5 \pm 0.7\%$ | $94.6 \pm 0.7\%$ |
| | 3 | $98.1 \pm 0.4\%$ | $98.7 \pm 0.2\%$ |
| | 4 | $83.5 \pm 1.2\%$ | $86.9 \pm 0.7\%$ |
| | 5 | $86.2 \pm 1.5\%$ | $89.0 \pm 1.4\%$ |
| TM, SAR Oct. 17, 1991 and GIS | 1 | $96.0 \pm 1.1\%$ | $96.2 \pm 0.2\%$ |
| | 2 | $96.9 \pm 0.9\%$ | $100.0 \pm 0.0\%$ |
| | 3 | $99.1 \pm 0.2\%$ | $99.7 \pm 0.1\%$ |
| | 4 | $83.8 \pm 1.2\%$ | $89.7 \pm 0.6\%$ |
| | 5 | $85.4 \pm 1.5\%$ | $88.0 \pm 1.6\%$ |
| TM and SAR Nov. 19, 1991 | 1 | $98.4 \pm 0.3\%$ | $99.8 \pm 0.0\%$ |
| | 2 | $90.7 \pm 0.7\%$ | $93.9 \pm 0.7\%$ |
| | 3 | $98.0 \pm 0.4\%$ | $99.0 \pm 0.2\%$ |
| | 4 | $75.9 \pm 3.8\%$ | $78.7 \pm 3.7\%$ |
| | 5 | $89.3 \pm 0.3\%$ | $89.9 \pm 0.7\%$ |
| TM, SAR Nov. 19, 1991 and GIS | 1 | $99.9 \pm 0.1\%$ | $99.9 \pm 0.0\%$ |
| | 2 | $97.6 \pm 0.4\%$ | $99.9 \pm 0.1\%$ |
| | 3 | $99.4 \pm 0.2\%$ | $99.6 \pm 0.2\%$ |
| | 4 | $75.9 \pm 3.8\%$ | $78.6 \pm 3.6\%$ |
| | 5 | $90.2 \pm 0.3\%$ | $90.4 \pm 0.7\%$ |

Table 3: Classification accuracy (in %) for different classes (class 1: water; class 2: urban areas; class 3: forest; class 4: agricultural areas, unplowed; class 5: agricultural areas, plowed).

| Fixed time mode | |
|---|---|
| Reliability factors | $\alpha_{SAR} = 0.7$, $\alpha_{TM} = 0.95$ |
| Spatial context | $\beta_{sp} = 1.5$ |
| Temporal context | $\beta_g = 0.9$ |
| Temporal mode | |
| Reliability factors | $\alpha_{SAR} = 0.7$, $\alpha_{TM} = 0.95$ |
| Spatial context | $\beta_{sp} = 0.5$ |
| Temporal context | $\beta_{temp} = 0.3$ $\beta_g = 0.6$ |

Table 4: Parameter values used for the fusion algorithm. Six iterations are used in the ICM algorithm.

| Images fused | Reference method | MRF method |
|---|---|---|
| TM and SAR Oct. 17, 1991 | 50.2% | 66.7% |
| TM and SAR Oct. 19, 1991 | 74.1% | 83.4 % |

Table 5: Classification accuracy for detecting actual changes in agricultural fields.

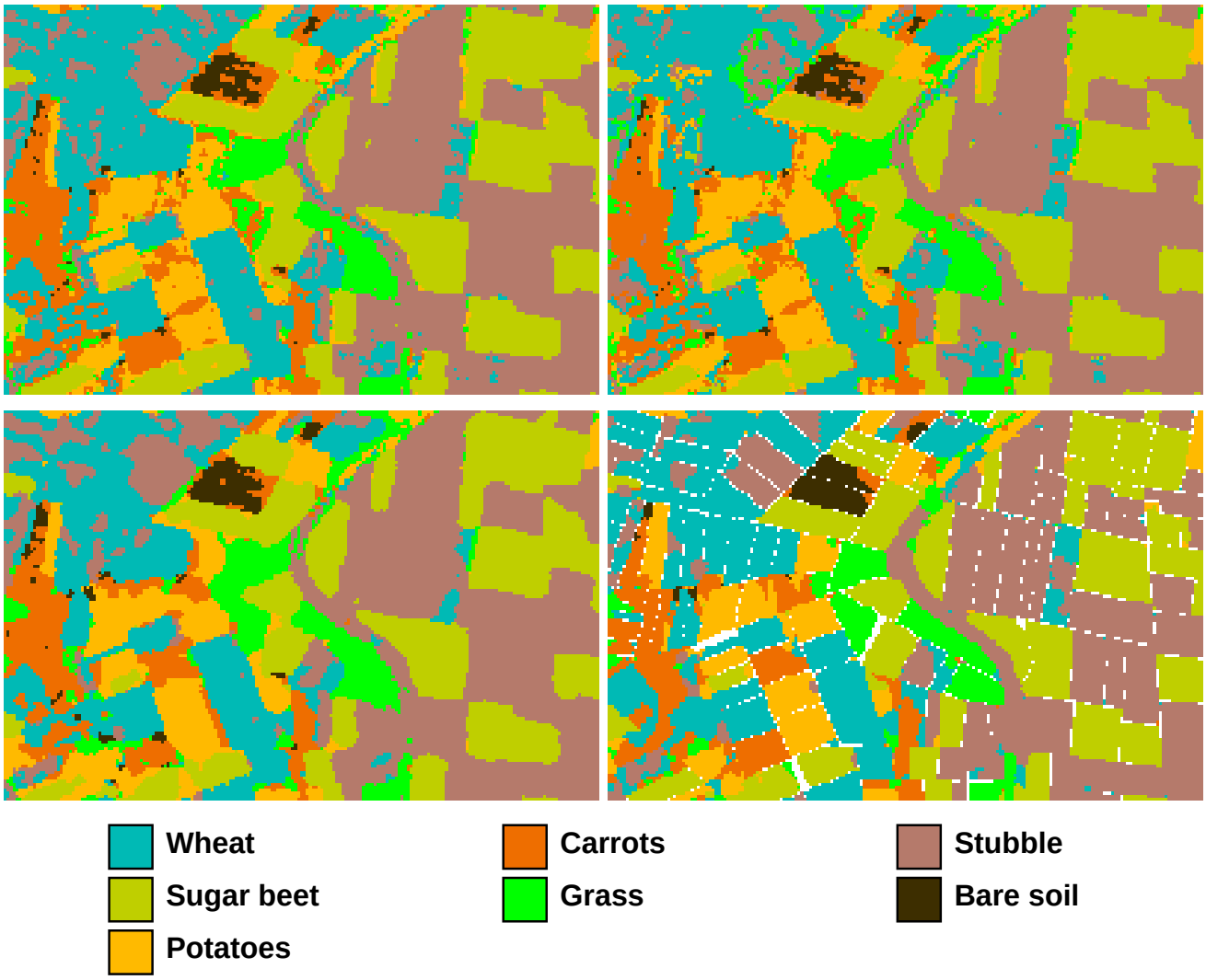


Figure 10: *Classification maps for different methods. (a) Landsat TM alone (upper left) (b) Reference fusion model (upper right) (c) MRF fusion model (lower left) (d) MRF fusion model with GIS field border map. The field borders are indicated in white in (d) (the gaps are due to a resampling of the borders to produce this overlay).*

| | Sources/Fusion model | | | | |
|---|---|---|---|---|---|
| Class | SAR | TM | Reference model | MRF model | MRF model with field borders |
| Wheat | 49.2 | 76.2 | 77.8 | 73.7 | 83.4 |
| Sugar beet | 63.8 | 79.5 | 79.9 | 78.8 | 89.2 |
| Potatoes | 70.4 | 63.4 | 67.0 | 70.5 | 77.2 |
| Carrots | 41.0 | 58.1 | 62.1 | 60.3 | 71.4 |
| Grass | 78.9 | 86.6 | 86.8 | 87.6 | 92.7 |
| Stubble | 79.9 | 61.7 | 61.1 | 68.8 | 66.0 |
| Bare soil | - | 63.6 | 63.6 | 66.7 | 78.8 |
| Total | 59.9 | 70.3 | 71.3 | 73.0 | 79.6 |

Table 6: *Classification accuracy (in %) for different classes for the MAESTRO crop classification. (Bare soil was not present in the SAR image.)*

## References


[1] M. A. Abidi and R. C. Gonzalez. *Data Fusion in Robotics and Machine Intelligence.* Academic Press, Inc., 1992.

[2] N. Ayache and O. Faugeras. Building, registrating, and fusing noisy visual maps. *Int. J. Robot. Res.*, 7:45–64, 1988.

[3] J. A. Benediktsson and P. H. Swain. A method of statistical multisource classification with a mechanism to weight the influence of the data sources. In *IEEE Symp. Geosc. Rem. Sens. (IGARSS)*, pages 517–520, Vancouver, Canada, July 1989.

[4] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy. Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Trans. Geosc. Remote Sensing*, 28:540–552, 1990.

[5] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.

[6] R. Chellappa and A. Jain (editors). *Markov Random Fields: Theory and Application.* Academic Press, 1993.

[7] P. B. Chou and C. M. Brown. Multimodal reconstruction and segmentation with Markov Random Fields and HCF optimization. In *Proceedings of the 1988 DARPA Image Understanding Workshop*, pages 214–221, 1988.

[8] P. N. Churchill. MAESTRO 1 proceedings of the data quality workshop 5-6/7-1990. Technical report, Report Number IRSA/MWT/4.90, Version 1.0, July 1990, CEC-JRC/ESA, 1990.

[9] H. Derin, P. A. Kelly, G. Vezina, and S. G. Labitt. Modeling and segmentation of speckled images using complex data. *IEEE Trans. Geosc. Remote Sensing*, 28:76–87, 1990.

[10] R. C. Dubes and Anil K. Jain. Random field models in image analysis. *J. Applied Statistics*, 16:131–163, 1989.

[11] S. E. Franklin. Ancillary data input to satellite remote sensing of complex terrain phenomena. *Computers and Geosciences*, 15:799–808, 1989.

[12] R. T. Frankot and R. Chellappa. Lognormal random-field models and their applications to radar image synthesis. *IEEE Trans. Geosc. Remote Sensing*, 25:195–206, 1987.

[13] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6:721–741, 1984.

[14] P. J. Green and D. M. Titterington. Recursive methods in image processing. In *46th Session of the International Statistical Institute*, pages 1–17, 1987.

[15] R. M. Haralick. Decision making in context. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-5:417–428, 1984.

[16] J. Haslett. Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context. *Pattern Recognition*, 18:287–296, 1985.

[17] S. J. Henkind and M. C. Harrison. An analysis of four uncertainty calculi. *IEEE Trans. Syst., Man Cybern.*, 18:700–714, 1988.

[18] J. Kittler and J. Föglein. Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2:13–29, 1984.

[19] O. Jakubowicz. Autonomous reconfiguration of sensor systems using neural nets. In *Proc. SPIE: Sensor Fusion*, pages 197–203, Orlando, Florida, April 1988.

[20] B. Jeon and D. Landgrebe. Spatio-temporal contextual classification of remotely sensed multispectral data. In *Proc. of 1990 IEEE Intern. Conf. on Syst., Man, and Cybern.*, pages 342–344, Los Angeles, CA, 1990.

[21] B. Jeon and D. A. Landgrebe. Classification with spatio-temporal interpixel class dependency contexts. *IEEE Trans. Geosc. Rem. Sens.*, 30:663–672, 1992.

[22] E. Holbæk Hanssen, H. Tjelmeland, O. M. Johannessen, T. Olaussen, and R. Karpuz. Speckle reduction and maximum likelihood classification of SAR images from sea ice recorded during MIZEX 87. In *IEEE Symp. Geosc. Rem. Sens. (IGARSS)*, pages 755–758, Vancouver, Canada, July 1989.

[23] N. Khazenie and M. M. Crawford. Spatio-temporal autocorrelated model for contextual classification. *IEEE Trans. Geosc. Remote Sensing*, 28:529–539, 1990.

[24] H. Kim and P. H. Swain. A method for classification of multisource data using interval-valued probabilities and its application to HIRIS data. In *Proc. Workshop on Multisource Data Integration in Remote Sensing*, pages 75–82. NASA Conference Publication 3099, Maryland, June 1990.

[25] D. G. Leckie. Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination. *Photogrammetric Engineering and Remote Sensing*, 56:1237–1246, 1990.

[26] J-S Lee. Speckle analysis and smoothing of SAR images. *Computer Graphics and Image Processing*, 7:24–32, 1981.

[27] J-S Lee, K. Hoppel, S. A. Mango, and A. R. Miller. Intensity and phase statistics of multi-look polarimetric SAR imagery. In *IEEE Symp. Geosc. Rem. Sens. (IGARSS)*, pages 813–816, Tokyo, Japan, August 1993.

[28] T. Lee, J. A. Richards, and P. H. Swain. Probabilistic and evidential approaches for multisource data analysis. *IEEE Trans. Geosc. Remote Sensing*, 25:283–293, 1987.

[29] Q. Lin and J. P. Allebach. Combating speckle in SAR images: Vector filtering and sequential classification based on a multiplicative noise model. *IEEE Trans. Geosc. Remote Sensing*, 28:647–653, 1990.

[30] A. L. Maffett and C. C. Wackerman. The modified Beta density function as a model for synthetic aperture radar clutter statistics. *IEEE Trans. Geosc. Remote Sensing*, 29:277–283, 1991.

[31] J. Marroquin, S. Miter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82:76–89, 1987.

[32] J. Middelkoop and L. L. F. Janssen. Implementation of temporal relationships in knowledge based classification of satellite images. *Photogrammetric Engineering & Remote Sensing*, 57:937–945, 1991.

[33] A. Owen and P. Switzer. A neighbourhood based classifier for Landsat data. Techn. Report, Dep. of statistics, Stanford University, 1982.

[34] J. Raviv. Decision making in markov chains applied to the problem of pattern recognition. *IEEE Trans. Inform. Theory*, 13:536–551, 1967.

[35] J. M. Richardson and K. A. Marsh. Fusion of multisensor data. *Int. J. Robot. Res*, 7:78–96, 1988.

[36] E. Rignot and R. Chellappa. Segmentation of synthetic-aperture-radar complex data. *J. Opt. Soc. Am. A*, 8:1499–1509, 1991.

[37] E. Rignot and R. Chellappa. Segmentation of polarimetric synthetic aperture radar data. *IEEE Trans. Image Processing*, 1:281–299, 1992.

[38] A. H. Schistad and A. K. Jain. Texture analysis in the presence of speckle noise. In *IEEE Symp. Geosc. Rem. Sens. (IGARSS)*, pages 884–886, Houston, Texas, May 1992.

[39] A. H. Schistad Solberg, A. K. Jain, and T. Taxt. Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images. *IEEE Trans. Geosc. Rem. Sens.*, 32:768–778, 1994.

[40] A. H. Schistad Solberg, A. K. Jain, and T. Taxt. Fusion of multitemporal satellite images and GIS data for land-use classification. In *Proceedings of the 8th Scandinavian Conference on Image Analysis*, pages 691–697, Tromsø May 25-28, 1993.

[41] A. Schistad Solberg. A Markov random field model for classification of multisource satellite imagery. Techn. Report, Norwegian Computing Center BILD/09/94, 1994.

[42] P. H. Swain. Bayesian classification in a time-varying environment. *IEEE Trans. Sys. Man and Cyber.*, 8:879–883, 1978.

[43] G. T. Toussaint. The use of context in pattern recognition. *Pattern Recognition*, 10:189–204, 1978.

[44] M. Tuceryan and A. K. Jain. *Texture Analysis, in Handbook of Pattern Recognition and Computer Vision*, pages 235–276. World Scientific Publishing Company, 1993.

[45] F. T. Ulaby and C. Elachi. *Radar Polarimetry for Geoscience Applications.* Artec House Inc., 1990.

[46] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave Remote Sensing, Active and Passive, Vols. I-III.* Artech House Inc., 1981, 1982, 1986.

[47] M. G. Wooding. Crop survey of the feltwell area - august 1989, short report and crop map prepared within the MAESTRO-1 experimental programmme. Technical report, British National Space Centre, January 1990.

[48] W.A. Wright. A Markov random field approach to data fusion and colour segmentation. *Image Vision Comp.*, 7:144–150, 1989.