Bayesian Morphology: Fast Unsupervised Bayesian Image Analysis

Florence Forbes
Inria Rhône-Alpes

Adrian E. Raftery
University of Washington [1]

September 10, 1998

[1]Florence Forbes is Researcher, Projet IS2, Inria Rhône-Alpes, Zirst, 655 av. de l'Europe, 38330 Montbonnot Saint-Martin, France. Email: Florence.Forbes@inrialpes.fr. Adrian E. Raftery is Professor of Statistics and Sociology, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322. Email: raftery@stat.washington.edu; Web: www.stat.washington.edu/raftery. This work was done while the first author was visiting the University of Washington, and was supported by a Lavoisier grant from the Ministère des Affaires Etrangères, Paris and Office of Naval Research grant no. N00014-96-1-0192. The authors are grateful to Julian Besag, Simon Byers, Chris Fraley, Todd Mitchell, Christian Posse, Derek Stanford and John Wallace for helpful discussions, and to Dr. H. Thomas Robertson, UW Division of Pulmonary and Critical Care for sharing the data in Figure 3 (a).

Abstract

We consider the problems of image segmentation and classification, and image restoration when the true image is made up of a small number of (unordered) colors. Our emphasis is on both performance and speed; speed has become increasingly important for analyzing large images and multispectral images with many bands, processing large image databases, real-time or near real-time image analysis, and the online analysis of video.

Bayesian image analysis (Geman and Geman 1984) provides an elegant solution to these problems, but it is computationally expensive, and the solutions it provides may be sensitive to unrealistic global properties of the models on which it is based. The ICM algorithm (Besag 1986) is faster and based on the *local* properties of the models underlying Bayesian image analysis; parameter estimation is performed iteratively via pseudo-likelihood. Mathematical morphology (Matheron 1975) is faster again and is widely considered to perform well, but lacks a statistical basis; method selection (analogous to parameter estimation) is done in a rather *ad hoc* manner.

We propose *Bayesian morphology*, a synthesis of these methods that attempts to combine the speed of mathematical morphology with the principled statistical basis of ICM. The key observation is that when the original image is discrete (or if an initial segmentation has been carried out), then, assuming a Potts model for the true scene and channel transmission noise, (1) the ICM algorithm is equivalent to a form of mathematical morphology; and (2) the segmentation is insensitive to the precise values of the model parameters. Unlike in standard Bayesian images analysis and ICM, it is feasible to do maximum likelihood estimation of the parameters in this setting. For grey-level or multispectral images, we propose an initial segmentation based on the EM algorithm for a mixture model of the marginal distribution of the pixels. The resulting algorithm is much faster than ICM, with gains that increase for more bands and larger images, and has good performance in experiments and for real examples.

Key Words : Bayesian image restoration; Iterated Conditional Modes; Mathematical Morphology; Potts model; Pseudo-likelihood.

Contents

List of Tables

List of Figures

1 Introduction

We consider the problem of image segmentation or classification, or of image restoration when the true scene is made up of a small number of unordered colors. Here we consider only the generic version of this problem, where precise knowledge of the kinds of objects sought is not available; only a general sort of local smoothness similar to that expected by the human eye is assumed.

We consider both performance and speed as criteria for assessing methods. Speed has become increasingly important as the demands on image processing systems have increased over the past 15 years. It is particularly important for analyzing large images, for processing large databases or archives of images, for analyzing multispectral images with many bands, for real-time, near real-time or interactive image analysis, and for the analysis of video, either online or off line.

Bayesian image analysis was proposed as a solution to this problem by Geman and Geman (1984), who introduced the Gibbs sampler for finding the posterior mode of the entire true scene, based on a Markov random field probability model. This is a fully statistically grounded method but it is very computationally expensive. Besag (1986) proposed the iterated conditional modes (ICM) algorithm, which is based on the same model as the methods of Geman and Geman (1984), but uses only the local properties of their model and is faster. The parameters of the model are typically assigned by the user, estimated offline from training data if these are available, or estimated iteratively using maximum pseudo-likelihood. Exact maximum likelihood is typically intractable.

Mathematical morphology was introduced to solve the same general class of problems by Matheron (1975); see also Serra (1982). It consists of applying sequences of morphological operators such as erosions, dilations, openings and closings to the image in an order determined by the user. It is very fast and has been much used; it has been widely reported to perform well in practice. However, it does not have a statistical basis and so its inferential foundation is uncertain. One consequence of this is that the specification of the sequence of operations to be used in any given application tends to be somewhat *ad hoc*, being based on trial and error on the part of the user.

Although Bayesian image analysis and mathematical morphology are aimed at closely related problems, they have largely developed in isolation from one another. Here we introduce an approach called *Bayesian morphology*, which aims to combine the power, elegance and firm statistical foundation of Bayesian image analysis with the speed of mathematical morphology, while retaining the good performance of both. Our key observation comes from considering the situation where the original image to be analyzed consists of the same colors as the true scene, but is a degraded version of it. This can also arise where the original image is grey-scale or multispectral, but an initial, lower-quality segmentation has been performed. Then, if restoration is based on a Potts model for the true scene and the assumption of channel transmission noise, (1) ICM is equivalent to a form of mathematical morphology; and (2) the restoration is the same for a range of values of the model parameters.

Property (1) allows us to perform an ICM restoration using the very fast computational tools of mathematical morphology, while property (2) allows us to incorporate parameter estimation at a low computational cost. It also allows us to do maximum likelihood estimation of the parameters, which is not feasible in standard Bayesian images analysis and ICM. When the original image is grey-scale or multispectral, we propose initializing the method by positing a finite mixture model for the marginal distribution of (possibly multivariate) pixel intensities, and estimating this using the EM algorithm. We call this *marginal mixture EM segmentation*.

The resulting method is much faster than ICM, and in our experiments with synthetically degraded and real images its performance was comparable to that of ICM. The method can be viewed as a special case of either ICM or mathematical morphology. As a morphological method, it has the advantage that the sequence of operators is chosen automatically and in a statistically principled manner via the parameter estimation in ICM, rather than in the usual *ad hoc* manner. The structuring element (morphology) is essentially equivalent to the neighborhood (ICM), which in turn corresponds to the probability model for the true scene. Thus the choice of structuring element can be reduced to a standard problem of statistical model selection, and solved, for example, using Bayesian model selection (Kass and Raftery 1995; Ji and Seymour 1996).

A simple "ancestor" of our approach is to first classify each pixel according to the maximum likelihood classifier, post-process this using the median filter or simple majority vote local smoother, and iterate until convergence. We refer to this method as "blind restoration". Besag (1986) points out that this has the clear disadvantage of losing track of the records themselves, a criticism that also applies to relaxation labeling methods (Rosenfeld, Hummel, and Zucker 1976; Hummel and Zucker 1983). This is correct, and blind restoration tended to perform relatively poorly in our own experiments, as we report below. However, when simple majority vote is replaced by discrete-image ICM, and naive maximum likelihood is replaced by marginal mixture EM segmentation, this disadvantage seems no longer to have serious practical consequences, and may often be more than offset by the increase in speed.

In Section 2, we review the Bayesian, ICM and morphological approaches. In Section 3 we introduce Bayesian morphology. Online parameter estimation is described in Section 4, using likelihood and pseudo-likelihood criteria. A number of experiments are presented in Section 5, including extensions of Bayesian morphology to grey-scale and multispectral images via marginal mixture EM segmentation. Connections to other work and possible extensions are discussed in Section 6.

2 Background: Bayesian Image Analysis, ICM and Mathematical Morphology

2.1 Bayesian Image Restoration and the ICM Algorithm

Bayesian image analysis is based on probability models. It includes a variety of tasks, including image restoration. The true but unknown scene, $x = \{x_i,\ i \in S\}$, where S is a set of pixels, is interpreted as a particular realization of a random vector X. The observed image y is interpreted as a realization of a random vector Y that can be seen as a degraded version of X. The vector Y depends on X through a known conditional probability density function $\mathcal{L}(y|x)$ which incorporates the image formation model and the noise model.

To restore X is to propose an estimator $\hat{X} = \hat{X}(Y)$ of X on the basis of Y. Bayesian image restoration methods are based on the following principles. The true image x is supposed to be a realization of a random field with distribution $\mathbb{P}(x)$. Then the restored image $\hat{x}$ is based on the posterior density of x, namely $\mathbb{P}(x|y) \propto \mathcal{L}(y|x)\mathbb{P}(x)$. A standard restoration criterion consists of maximizing this density, leading to the maximum a posteriori (MAP) estimate of x .

For each pixel i, let $S\backslash\{i\}$ denote all other pixels in S. One of the most popular modeling assumptions is to consider the image x as being a realization of a Markov random field. This means that for all pixels i in S, $\mathbb{P}(x)$ satisfies $\mathbb{P}(x_i|x_{S\backslash\{i\}}) = \mathbb{P}(x_i|x_{N(i)})$, i.e., the conditional distribution of $\mathbb{P}(x)$ depends only on the values of pixels in a subset $N(i)$ of $S\backslash\{i\}$, called the *neighborhood* of pixel i. Another usual assumption is that, given $X = x$, the Y_i are conditionally independent and have the same conditional density function $f(y_i|x_i)$ that depends only on x_i. Thus $\mathcal{L}(y|x)$ can be written as the product $\mathcal{L}(y|x) = \prod_{i \in S} f(y_i|x_i)$.

Finding the MAP estimate under these assumptions can require heavy computation. A less computationally demanding method that provides a fast approximation to the MAP is the Iterated Conditional Modes (ICM) algorithm (Besag 1986). The ICM algorithm is iterative. Given a current estimate $\hat{x}$ of the image, a new one is computed by visiting each pixel in turn. When at pixel i, the current value there is replaced by the value that maximizes the conditional density

$$\mathbb{P}(x_i|\hat{x}_{S\setminus\{i\}}, y), \tag{1}$$

given all other current pixel values $\hat{x}_{S\setminus\{i\}}$ and the fixed observation y. This choice is motivated by the following equality that holds for any pixel i,

$$\mathbb{P}(x|y) = \mathbb{P}(x_i|x_{S\setminus\{i\}}, y)\ \mathbb{P}(x_{S\setminus\{i\}}|y)\ .$$

When pixels are updated sequentially, choosing values that maximize the conditional probability $\mathbb{P}(x_i|x_{S\setminus\{i\}}, y)$ increases the posterior density $\mathbb{P}(x|y)$ and ensures the convergence to a local maximum of $\mathbb{P}(x|y)$.

Under the previous modeling assumptions, maximizing the conditional density (1) is equivalent to maximizing

$$f(y_i|x_i)\ \mathbb{P}(x_i|\hat{x}_{N(i)}), \tag{2}$$

since only the dependence on x_i is relevant for the maximization. For unordered-color images, we shall focus, in Sections 3.1 and 3.2, on models simple enough that this maximization can be done explicitly. We will show that for binary images the algorithm can be formulated using a morphological terminology.

2.2 Binary Mathematical Morphology

Mathematical morphology was first introduced as a tool for investigating geometric structure in binary images. In this context, binary images are usually viewed as subsets of a two-dimensional space, usually the two-dimensional discrete plane $\mathbb{Z}^2$ or some finite subset S of it (the pixels), or equivalently as mappings from this set of pixels to $\{0, 1\}$. Geometric information can be extracted from a binary image by probing it with a small shape known as the structuring element. This element is a subset of $\mathbb{Z}^2$. To make it more similar to neighborhood structures in a Markov random field setting, it is assumed to contain the origin $(0, 0)$ and to be symmetric. Note that this also simplifies some of the following definitions. Probably the most commonly used example of such an element is the 3×3 square consisting of nine pixels.

Let I be an image and B a symmetric structuring element. For each pixel $i \in S$, let $B(i)$ be the translation of B by i, $B(i) = \{i + j, j \in B\}$. The neighborhood $N(i)$ of a pixel i is thus the translation by i of a symmetric structuring element N. A typical N would be the set of eight pixels in a 3×3 square when the center of the square is not included.

The most primitive morphological operators are erosion and dilation. They are respectively defined as follows,

$$\begin{aligned} I \ominus B &= \{i \in S,\ B(i) \subseteq I\} , \\ I \oplus B &= \{i \in S,\ B(i) \cap I \neq \emptyset\} . \end{aligned}$$

A pixel i belongs to I eroded by B if $B(i)$ is totally contained within I, while for pixel i to be in the dilation of I by B, it is enough that one pixel of $B(i)$ belongs to B .

Erosion and dilation are dual notions. The so-called duality principle plays an important role in mathematical morphology. The dual, ψ^*, of a morphological operator ψ is defined by $\psi^*(I) = (\psi(I^c))^c \quad \forall I \subset \mathbb{Z}^2$, where I^c represents the complement of I . The erosion of the background of an image is equivalent to the dual dilation of its foreground. When an operator is equal to its dual, it is said to be self-dual. Self-dual operators treat the background and foreground of an image identically, and may therefore be desirable when no a priori information is available on what is foreground and what is background.

When the intention is to remove noise from an image, another desirable property is idempotence. An operator ψ is said to be idempotent when $\psi(\psi(I)) = \psi(I) \quad \forall I \subset \mathbb{Z}^2$.

Erosion and dilation are not idempotent but they can be combined into two idempotent operators called opening and closing. An opening is an erosion followed by a dilation,

$$I \circ B = (I \ominus B) \oplus B .$$

Its effect is roughly to delete small isolated parts and remove thin filaments of an image. A closing is a dilation followed by an erosion,

$$I \bullet B = (I \oplus B) \ominus B .$$

A closing is an opening of the image background. Opening and closing are dual notions in the sense that $I \bullet B = (I^c \circ B)^c$.

Openings and closings are the morphological operators which are commonly used to clean noisy images. They are based on erosions and dilations. The latter can be further generalized. They are particular cases of weighted rank operators as defined in Heijmans

(1994). Let B be a structuring element with m points and denote by I_j the value (0 or 1) at pixel j when the image I is associated with a mapping from S to $\{0,1\}$, i.e.

$$I_j = \begin{cases} 1 & \text{if } j \in I \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = \{w_1, w_2, \ldots, w_m\}$ be the weights w_i associated with each pixel i in B and r be a threshold or rank for $w_1, \ldots, w_m, r \in \mathbb{Z}$. The weighted rank operator $\rho_{B,W,r}$ is defined by

$$\rho_{B,W,r}(I) \quad = \quad \{i \in S, \sum_{j \in B(i)} w_j \; I_j \geq r\} \, . \tag{3}$$

If the weights are positive, such an operator is increasing (with respect to set inclusion). The negative of $\rho_{B,W,r}$ is the weighted rank operator $\rho_{B,W,r'}$ with $r' = \sum_{i=1}^{m} w_i + 1 - r$. It follows that $\rho_{B,W,r}$ is self-dual if and only if $2r = \sum_{i=1}^{m} w_i + 1$. In that case, $\rho_{B,W,r}$ is called a weighted median operator.

If $w_i = 1$ for all i and $1 \leq r \leq m$ then $\rho_{B,W,r}$ is called a rank operator and is simply denoted by $\rho_{B,r}$. For any two operators ψ_1 and ψ_2, let us denote by $\psi_1 \leq \psi_2$ the situation where $\psi_1(I) \subset \psi_2(I)$ for every $I \subset \mathbb{Z}^2$. Then we have $\rho_{B,m} \leq \rho_{B,m-1} \leq \ldots \leq \rho_{B,1}$. The operators $\rho_{B,m}$ and $\rho_{B,1}$ are erosion and dilation by B, respectively. More generally, rank operators can be decomposed as finite unions of erosions or intersections of dilations. For instance, we have $\rho_{B,r}(I) = \bigcup_{B_0 \in \mathcal{V}_0} (I \ominus B_0)$, where $\mathcal{V}_0 = \{B_0 \subset B \; , \; |B_0| = r\}$. Thus the rank operators are morphological operators in the sense of Matheron (1975) and Serra (1982). If m is odd and $r = (m+1)/2$ then $\rho_{B,r}$ is self-dual and is usually referred to as the median operator. It has interesting cleaning capabilities and is sometimes combined with openings and closings to filter noisy images.

These rank operators are the ones that appear in the formulation (10) of ICM below. They appear for different structuring elements depending on the neighborhood structure chosen in the Bayesian analysis.

3 Bayesian Morphology

In this section, we first specify the ICM algorithm for some simple models. We show in Section 3.1 that for binary images, the ICM estimate of the true pixel (update rule (2)) is equivalent to the application of a mathematical morphology rank operator. When the noise and model parameters are known, ICM can be seen as a succession of rank operators whose ranks depend on the known parameters (see (10) below). When the noise and model parameters are not known, a natural approach is to consider the unsupervised version of ICM which consists of alternating between parameter estimation and restoration (see Section 4). One advantage of such a mathematical morphology formulation of the restoration step (update rule (10) below) is to reveal the existence of insensitivity conditions for the parameters. This means that the final restoration is not sensitive to the precise values of the noise and model parameters. This is a key observation that we use to reduce the complexity of the estimation step in traditional unsupervised ICM and to save a lot of computation time (see Theorem 2 below). We call the resulting algorithm and its extension to unordered-color images (Section 3.2), Bayesian morphology. When performed on unordered-color images, it is equivalent to ICM in the sense that the final restoration or classification is the same. In this case, it differs from ICM essentially in the way the parameter estimation step is carried out. According to the insensitivity conditions, point estimates need not be computed.

3.1 Binary Images and Insensitivity Conditions in ICM

In the case of binary images, $x_i \in \{0,1\}$ and $y_i \in \{0,1\}$ for all pixels i . A commonly used prior distribution for the true image X is the attractive isotropic Ising model (in fact, the two-color Potts model),

$$\mathbb{P}(x) \quad = \quad Z(\beta)^{-1} \exp(\beta v(x)) \; , \tag{4}$$

where

$$v(x) \quad = \quad \sum_{i \sim j} \delta(x_i, x_j) \tag{5}$$

is the number of pairs of neighboring pixels having the same color in x. In the above sum, $i \sim j$ denotes the statement that the pixels i and j are neighbors and $\delta(x_i, x_j)$ refers to the Kronecker delta function, equal to 1 if x_i and x_j are the same, and to 0 otherwise. The quantity $Z(\beta)$ is the normalizing constant, or partition function,

$$Z(\beta) \quad = \quad \sum_{x} \exp(\beta v(x)) \; . \tag{6}$$

This function is usually difficult to compute because of the intractably large number of terms in the summation. The conditional distributions of $\mathbb{P}(x)$ have the simple form

$$\mathbb{P}(x_i | x_{S\setminus\{i\}}) \quad \propto \quad \exp(\beta u_i(x_i)) \; , \tag{7}$$

where

$$u_i(x_i) \quad = \quad \sum_{j \in N(i)} \delta(x_i, x_j) \tag{8}$$

is the number of neighbors of pixel i having color x_i. The model depends on a parameter β which is taken to be positive, reflecting the assumption that neighboring pixels tend to be of the same color.

The true images are then assumed to be degraded by a region-dependent flip-flop, or channel transmission noise characterized by the two parameters p_{01} and p_{10}, where $p_{01} = I\!P(y_i = 1|x_i = 0)$ and $p_{10} = I\!P(y_i = 0|x_i = 1)$. We shall assume, without loss of generality, that p_{01} and p_{10} belong to the open interval $(0, 1/2)$. The y_i are assumed to be conditionally independent and the likelihood is $\mathcal{L}(x|y) = \prod_{i \in S} f(y_i|x_i)$, where $f(y_i|1) = (1 - p_{10})^{y_i}(p_{10})^{1-y_i}$ and $f(y_i|0) = (p_{01})^{y_i}(1 - p_{01})^{1-y_i}$.

For these noise and image models, the update rule (2) in ICM can be written down explicitly. In what follows, $\lceil\ \rceil$ and $\lfloor\ \rfloor$ denote the floor and the ceiling functions, respectively.

Theorem 1 *For an Ising model with region-dependent flip-flop noise, the current ICM estimate of the true image at pixel i is updated by changing $\hat{x}_i$ to x_i^* according to the rule*

$$x_i^* \quad = \quad \begin{cases} 1 & \textit{if} \qquad\qquad u_i(1) - u_i(0) \geq 2w_1 \\ 0 & \textit{if} \qquad\qquad u_i(1) - u_i(0) < -2w_0 \\ y_i & \textit{if} \ -2w_0 \leq u_i(1) - u_i(0) < 2w_1 \ , \end{cases} \tag{9}$$

where w_0 and w_1 are positive integers that depend on the noise and model parameters p_{01}, p_{10}, β, namely

$$w_0 = \left\lfloor \frac{1}{2\beta} \log\left(\frac{1 - p_{10}}{p_{01}}\right) \right\rfloor \ ,$$
$$w_1 = \left\lceil \frac{1}{2\beta} \log\left(\frac{1 - p_{01}}{p_{10}}\right) \right\rceil \ .$$

Note that $u_i(1) + u_i(0) = |N(i)| = m$, so that $u_i(1) - u_i(0) = 2u_i(1) - |N(i)|$ is even for every symmetric neighborhood $N(i)$, where $|N(i)|$ is the number of elements in $N(i)$. For binary images, we have $u_i(1) = \sum_{j \in N(i)} x_j$, and an equivalent formulation of (9) can be given in terms of the rank operators defined in (3):

$$x_i^* \quad = \quad \begin{cases} [\rho_{N,r_0}(\hat{x})]_i & \text{if } y_i = 0 \\ [\rho_{N,r_1}(\hat{x})]_i & \text{if } y_i = 1, \end{cases} \tag{10}$$

where $r_0 = \frac{m}{2} + w_1$ and $r_1 = \frac{m}{2} - w_0$ and where $[z]_i$ is the value of the image z at pixel i. A shorter way to write (10) is $x_i^* = [\rho_{N,r_i}(\hat{x})]_i$, where $r_i = \frac{m}{2} + w_1 - y_i(w_0 + w_1)$.

The update rule (9) depends on the parameters p_{01}, p_{10} and β only through the values of w_0 and w_1. When these integers are known, each pixel can be updated easily by applying the rule (9) or, equivalently, the rank operators ρ_{N,r_0} or ρ_{N,r_1}. In the general case, p_{01}, p_{10} and β are unknown and so are w_0 and w_1. However, we have the following result.

Theorem 2 *If there exist integers k_0 and k_1 such that p_{01}, p_{10}, β satisfy*

$$\frac{1}{2(k_0+1)} \log\left(\frac{1-p_{10}}{p_{01}}\right) < \beta \leq \frac{1}{2k_0} \log\left(\frac{1-p_{10}}{p_{01}}\right) , \tag{11}$$

$$\frac{1}{2k_1} \log\left(\frac{1-p_{01}}{p_{10}}\right) \leq \beta < \frac{1}{2(k_1-1)} \log\left(\frac{1-p_{01}}{p_{10}}\right), \tag{12}$$

then for all these values of p_{01}, p_{10} and β, w_0 and w_1 are equal to k_0 and k_1 respectively. These conditions can be equivalently formulated in terms of p_{01} and $p_{11} = 1 - p_{10}$,

$$p_{01}\, a^{k_0} \leq p_{11} < p_{01}\, a^{k_0+1} \tag{13}$$

$$p_{01}\, \frac{1}{a^{k_1-1}} + 1 - \frac{1}{a^{k_1-1}} < p_{11} \leq p_{01}\, \frac{1}{a^{k_1}} + 1 - \frac{1}{a^{k_1}} , \tag{14}$$

where $a = \exp(2\beta)$.

It follows that the update rule (9) does not depend on the exact values of p_{01}, p_{10} and β, but rather on how they are related to each other. We refer to equations (11) and (12) (or (13) and (14)) as the *insensitivity conditions.* Figure 1 shows the sets of values of p_{01} and p_{11} within which w_0 and w_1 are constant. Estimating w_0 and w_1 then consists of finding k_0 and k_1 such that the insensitivity conditions hold. We can restrict our search to values of k_0 in $\{0, 1, \ldots, |N(i)|/2 - 1\}$ and k_1 in $\{1, \ldots, |N(i)|/2\}$. Indeed, if $w_1 \geq |N(i)|/2 + 1$, then $y_i = 0$ automatically implies $x_i^* = 0$. If $w_0 \geq |N(i)|/2$, then $y_i = 1$ implies $x_i^* = 1$. If both previous conditions hold, then $x_i^* = y_i$. A particular case occurs when $w_1 = 1 + w_0$. For these values ρ_{N,r_0} and ρ_{N,r_1} are dual operators. Note that this is satisfied when $p_{01} = p_{10}$, i.e. when foreground and background noise are treated identically. Another particular case is $w_1 = 1 - w_0$, which actually corresponds to $w_0 = 0$ and $w_1 = 1$. It occurs when

$$\beta > \frac{1}{2} \log\left(\frac{1-p_{10}}{p_{01}}\right) , \tag{15}$$

$$\beta \geq \frac{1}{2} \log\left(\frac{1-p_{01}}{p_{10}}\right), \tag{16}$$

leading to the theorem below.

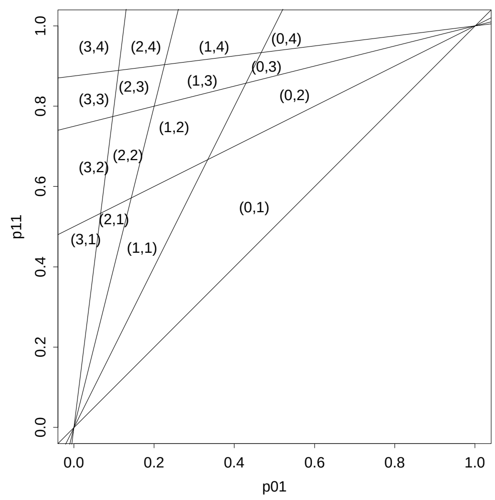


Figure 1: Insensitivity Conditions in Terms of p_{01} and p_{11} for an Eight-Pixel Neighborhood with $\beta = 0.35$. The lines delimit regions where w_0 and w_1 are constant. Their respective values are shown in parentheses. For example, in the largest triangle region, notation $(0, 1)$ means that for all values of p_{01} and p_{11} lying in this region, $w_0 = 0$ and $w_1 = 1$.

Theorem 3 *If p_{01}, p_{10}, β satisfy conditions (15) and (16), then the update rule (9) does not involve any parameters. It becomes*

$$x_i^* = \begin{cases} 1 & if \quad u_i(1) > u_i(0) \\ 0 & if \quad u_i(1) < u_i(0) \\ y_i & if \quad u_i(1) = u_i(0) \, . \end{cases} \tag{17}$$

This corresponds to the blind restoration proposed in Zhang, Shirazi, and Noda (1996) . Other blind restorations can be derived by assuming different values for w_0 and w_1. They may be more appropriate depending on the true parameter values (see Section 5). However, the update rule (17) consists of applying the majority rule and may be viewed as a reasonable first choice of blind restoration for many images. Note that this choice corresponds to the largest region in Figure 1.

Our alternative to arbitrary blind restoration consists of estimating w_0 and w_1. We will show, in Section 4, that this turns out to be much easier than computing point estimates for p_{01}, p_{10} and β. But first we consider similar developments for color images.

3.2 Unordered-Color Images

A widely used model for images with C unordered exchangeable colors is the Potts model defined below. Such a model is adapted to image segmentation or classification, where the goal is to assign to each pixel of the observed image a label (or color) indicating the pixel's class. The standard nearest-neighbor Potts model is defined by the joint distribution

$$\mathbb{P}(x) \quad = \quad Z(\beta)^{-1} \exp(\beta v(x)) \,, \tag{18}$$

where the expressions (5), (6), (7), (8) remain valid, keeping in mind that they may refer to more than two colors. Note that, in the infinite size limit, this model has a phase transition at $\beta_{crit} = \log(1 + \sqrt{C})$ when there are four neighbors (Georgii 1988).

To derive insensitivity conditions as in Section 3.1, the key step is to formulate the update rule in ICM so that the discrete nature of the images can be exploited in a simple way. For that reason, we restrict attention to the following independent noise model,

$$f(y_i|x_i) \quad = \quad \begin{cases} 1-p & \text{if } y_i = x_i \\ \frac{p}{C-1} & \text{otherwise.} \end{cases} \tag{19}$$

This corresponds to transmission with channel noise. A good transmission is assumed to be the most probable, which means that $C(1-p) > 1$. More general models are possible but they lead to more complex update rules. For instance, a region-dependent degradation where p in (19) would be replaced by values p_{x_i} that may depend on x_i could be considered. This would better generalize Section 3.1 but formulating it properly would require the estimation of $(C-1)(3C-2)/2$ integers instead of the two, w_0 and w_1, in (9). Adopting the simple noise model above leads to a formulation of ICM update rule very similar to that in the binary image case. Let $\hat{x}$ be the current restoration as in Section 3.1.

Theorem 4 *For a Potts model with channel noise, ICM works as follows. When at pixel i, the current value $\hat{x}_i$ is changed to x_i^* according to the procedure below:*

- *find the color which is taken most often by the neighbors of pixel i in $\hat{x}$, i.e. find c in $\{\hat{x}_j, j \in N(i)\}$ such that $u_i(c)$ is maximized.*
- *Then, update pixel i*

$$x_i^* \quad = \quad \begin{cases} c & \textit{if } u_i(c) - u_i(y_i) \geq w \\ y_i & \textit{otherwise,} \end{cases} \tag{20}$$

where w is a positive integer that depends on the parameters,

$$w = \left\lceil \frac{1}{\beta} \log\left(\frac{(1-p)(C-1)}{p}\right) \right\rceil \,.$$

The update rule (20) depends on β and p only through the value of w. This leads to the following insensitivity conditions.

Theorem 5 *For the Potts model with channel noise, if there exists an integer k such that β and p satisfy*

$$k-1 < \frac{1}{\beta}\log\left(\frac{(1-p)(C-1)}{p}\right) \leq k, \tag{21}$$

or, equivalently,

$$\frac{1}{k}\log\left(\frac{(1-p)(C-1)}{p}\right) \leq \beta < \frac{1}{k-1}\log\left(\frac{(1-p)(C-1)}{p}\right), \tag{22}$$

or in terms of p

$$\frac{1}{1+(C-1)^{-1}\exp(\beta k)} \leq p < \frac{1}{1+(C-1)^{-1}\exp(\beta(k-1))}, \tag{23}$$

then w is equal to k and the ICM update rule (20) does not depend on the exact values of β and p .

Estimating w then reduces to finding k such that one of the equivalent expressions above is satisfied. In Section 4, we propose a way to do so, avoiding the need for point estimation.

4 Online Parameter Estimation

When the model and noise parameters are unknown, a natural approach to unsupervised restoration is first to estimate the parameters and then to estimate the true unknown scene x (see Section 2.1) given the parameter estimates. This is usually done iteratively by iterating between parameter estimation and restoration, following the general procedure proposed in Besag (1986). In most cases, estimating the noise parameters does not present a problem (maximum likelihood estimators are usually available in closed form) and the main difficulty arises in the estimation of the parameters of the model for the true scene (or "prior" model). Maximum likelihood estimation is usually intractable, except maybe in some cases (see Goutsias 1991), because of the practical impossibility of computing the partition function. In practice, estimation is based on approximations such as maximum pseudo-likelihood (Besag 1975). If the Potts model is "true", this may lead to inaccurate estimators, especially when the dependency is high (Geyer 1991), but for actual images it may yield better restorations, given the often unrealistic *global* behavior of the Potts model (Besag 1986).

In our setting, however, the amount of computation that maximum likelihood estimation requires appears to be reasonable. The idea is to take advantage of the insensitivity conditions given in Sections 3.1 and 3.2 . They allow one to find the restoration based on the maximum likelihood estimate without ever actually finding the maximum likelihood estimator explicitly, a fact that leads to great computational savings.

Our algorithms are based on estimation criteria. An estimation criterion is a function that measures the quality of a parameter. We shall write β to denote such a parameter and $\hat{\beta}$ the value at which such a function is maximized, assuming that such a maximum exists. More generally, $\hat{\beta}$ may be an estimator of a parameter of interest, and we shall call an estimation criterion any monotone function f such that $f(\hat{\beta})$ is known (although $\hat{\beta}$ is not necessarily known) and $f(\beta)$ is computable for any given value of β. Typically, if F is a concave function with maximum at $\hat{\beta}$, its derivative $\frac{dF}{d\beta}$ is decreasing and equal to zero at $\hat{\beta}$. If this derivative is computable for given values of β, $f = \frac{dF}{d\beta}$ or $f = -\frac{dF}{d\beta}$ can be referred to as an estimation criterion.

We shall use such functions to find w_0 and w_1 if the model is that of Section 3.1, and to find w if the model is that of Section 3.2. The two cases are similar and we focus on the model of Section 3.2. The value of w is the integer k that satisfies (22), which for any increasing function f is equivalent to

$$f\left(\frac{1}{k}\log\left(\frac{(1-p)(C-1)}{p}\right)\right) \leq \; f(\beta) \; < f\left(\frac{1}{k-1}\log\left(\frac{(1-p)(C-1)}{p}\right)\right) .$$

Therefore, if $\hat{\beta}$ is an estimator of β, a natural estimator of w, $\hat{w}$, satisfies

$$\frac{1}{\hat{w}}\log\left(\frac{(1-p)(C-1)}{p}\right) \leq \; \hat{\beta} \; < \frac{1}{\hat{w}-1}\log\left(\frac{(1-p)(C-1)}{p}\right).$$

Equivalently, if f is such that $f(\hat{\beta})$ is equal to a known constant b, $\hat{w}$ satisfies

$$f\left(\frac{1}{\hat{w}}\log\left(\frac{(1-p)(C-1)}{p}\right)\right) \leq \; b \; < f\left(\frac{1}{\hat{w}-1}\log\left(\frac{(1-p)(C-1)}{p}\right)\right). \tag{24}$$

If $f\left(\frac{1}{k}\log\left(\frac{(1-p)(C-1)}{p}\right)\right)$ is known for different integer values k, finding $\hat{w}$ reduces to comparing real values, which is easier than computing $\hat{\beta}$. In Forbes and Raftery (1997), we considered in detail the cases corresponding to maximum likelihood and maximum pseudo-likelihood estimation for the standard Potts model (18). For maximum likelihood estimation, we showed that an appropriate function f and the corresponding value b are
$f(\beta) = \mathbb{E}_\beta[v(X)] = \sum_u v(u)\mathbb{P}(u)$, which depends on β through $\mathbb{P}$, and

$b = \mathbb{E}_{\hat{\beta}}[v(X)] = v(x)$, where $\hat{\beta}$ is the maximum likelihood estimator of β.

The exact computation of the above expectations is intractable. However, estimates can be obtained by sampling the distribution $\mathbb{P}(x)$ for the corresponding values of β and forming empirical averages. The required samples can be generated using Markov chain Monte Carlo. This can be computationally demanding, but we were able to accelerate it greatly for the Potts model using the Swendsen-Wang algorithm (Swendsen and Wang 1987). Our algorithm requires only a small number of samples and evaluations. Other approaches that try to save online samplings include Geyer and Thompson (1992), Geyer (1991), and Descombes, Morris, and Zerubia (1996).

Pseudo-likelihood estimation appears as a natural alternative to maximum likelihood, which is usually intractable in the standard Markov random field cases when point estimation of the parameters is required. The term "pseudo-likelihood" was introduced by Besag (1975) to refer to the product of conditional probabilities

$$\prod_{i \in S} \mathbb{P}(x_i \mid x_{S \setminus \{i\}}). \tag{25}$$

For the Potts model, $\mathbb{P}(x_i \mid x_{S \setminus \{i\}}) = p_i(x_i) = Z_i(\beta)^{-1} \exp(\beta u_i(x_i))$, where $Z_i(\beta) = \sum_{c \in \{1,\ldots,C\}} \exp(\beta u_i(c))$. In this case, an appropriate function f is $f(\beta) = \sum_{i \in S} \mathbb{E}_{p_i}[u_i(X_i)]$, and the corresponding constant b is $b = \sum_{i \in S} u_i(x_i) = 2v(x)$. An advantage over maximum likelihood estimation is the possibility of computing $\sum_{i \in S} \mathbb{E}_{p_i}[u_i(X_i)]$ exactly.

5 Experiments

5.1 Restoring Unordered-Color Images

We first consider pure restorations of unordered-color images. The steps of the Bayesian morphology algorithm are not listed here but more details and other images can be found in Forbes and Raftery (1997). Figure 2 shows a binary image degraded with 15% channel noise, and different restorations: Bayesian morphology using the likelihood and pseudo-likelihood criteria (see Section 4), and blind restoration using the simple majority vote (17), as in Zhang, Shirazi, and Noda (1996).

The performance of our Bayesian morphology algorithm is the same as that of an unsupervised version of ICM. In Table 1, we give the final estimates of p and w, the number of iterations before convergence and the number of calls to the function f used in (24).

The number of times the estimation criterion (function f) is evaluated is relevant to the speed of the algorithm, especially for maximum likelihood, since such evaluations require online sampling and may considerably slow down the restoration procedure. To initialize our procedure, we set the initial value of the noise parameter p to 0.35.

Note that for a two-color Potts model, the term $(u_i(c) - u_i(y_i))$ in the update rule (20) is even. Therefore, an odd value of w and the even integer following it actually lead to the same updating. Thus, in this case the number of calls to f can be divided by 2. Although computationally unnecessarily, the distinction is interesting when comparing maximum likelihood and maximum pseudo-likelihood estimators of β since a smaller value of w corresponds to a larger value of β. In Table 1, the pseudo-likelihood criterion results in an estimate of w equal to 3, instead of 4 when a likelihood criterion is used. For binary images, those values of w correspond to the same updating of the current restoration, so that for the noisy image (a), likelihood and pseudo-likelihood criteria lead to similar restorations ((b) and (c)). The restorations are different, however, because the first updating is different in each procedure. The fact that a pseudo-likelihood criterion leads to a lower estimate of w is consistent with the empirical observation by Geyer and Thompson (1992) that maximum pseudo-likelihood estimators tend to overestimate β. As illustrated in Figure 2, blind restoration performs badly. Too much detail is lost and the error rate is higher.

The choice of estimation criterion may be important in this setting. Theoretical studies and empirical comparison of the maximum likelihood and maximum pseudo-likelihood estimators for Markov random fields have been carried out in various references. Both the maximum likelihood and maximum pseudo-likelihood estimators have been shown to be consistent (see Comets and Gidas (1992) and Gidas (1993)). This does not necessarily say very much about their properties when used on finite lattices. In Tjelmeland (1996), simulation experiments with maximum likelihood and maximum pseudo-likelihood estimators for binary Markov random fields on moderately-sized hexagonal lattices (2,883 to 30,603 pixels) were presented. The results indicate that both estimators give good results for the Ising model and for models with weak interactions between neighboring pixels. For models with strong interactions, the maximum pseudo-likelihood estimator produces unsatisfactory estimates, whereas maximum likelihood estimator still gives reliable results. However, as observed in Tjelmeland (1996), estimated parameter values substantially different from the true values do not necessarily mean that the visual appearance of the corresponding realizations of the model differs from that of the true model. This is an important point in image analysis where modeling per se may not be the primary goal.

In our setting, the parameter estimation step cannot be considered independently of the restoration step. The main advantage in using insensitivity conditions is to avoid the problematic computation of pointwise estimates. Therefore, our comparison of the estimation criteria must be based on the quality of the final restoration or segmentation. How to evaluate this quality depends on the task at hand. Also, when considering a visual criterion for assessing the quality of an estimator, the actual performance may be masked by the choice of the prior model. On one hand, more accurate parameter estimation (via maximum likelihood) allows more reliable comparisons of different priors. In particular, the Potts model and, more generally, models that involve only pairwise interactions (Descombes, Morris, and Zerubia (1996), Tjelmeland and Besag (1998)) can be shown not to capture the global characteristics of images well. On the other hand, these models may perform surprisingly well in tasks such as restoration although they may perform poorly for other tasks (see Tjelmeland and Besag (1998)). Descombes, Morris, and Zerubia (1996) show cases where the restored images using the Potts model as a prior, may still be noisy. However, to obtain better restorations, a common trick is to over-regularize the solution by increasing β . In that sense, it is not surprising that a pseudo-likelihood criterion may produce good results. Ideally, we recommend the use of accurate parameter estimation and more realistic priors such as the higher order Markov random fields described in Descombes, Mangin, Pechersky, and Sigelle (1995) and Tjelmeland and Besag (1998). But this is much more demanding in terms of computation and time so that in practice the Potts model may still be prefered by users.

5.2 Segmenting Grey-Scale and Multidimensional Images

In the previous cases, the observed and unobserved images were of the same type. Often, however, the image to be restored consists of unordered colors (classes) but the observations are usually ordered real values (measurements). The ICM algorithm can be used for such restorations, assuming a different noise model (Besag, York, and Mollié 1991), but our use of the insensitivity conditions cannot be extended efficiently for this model. We believe however, that they can still be used to obtain acceptable segmentations rapidly.

A natural idea, when dealing with more general images (possibly multidimensional data), is to first derive an initial classification using simple and fast procedures such as the maximum likelihood classifier ($\hat{x}_i$ is chosen to maximize $f(y_i|x_i)$ at each pixel i separately). These procedures do not take spatial information into account and may result in an unacceptably large number of misclassified pixels. Therefore a better classification can be computed by

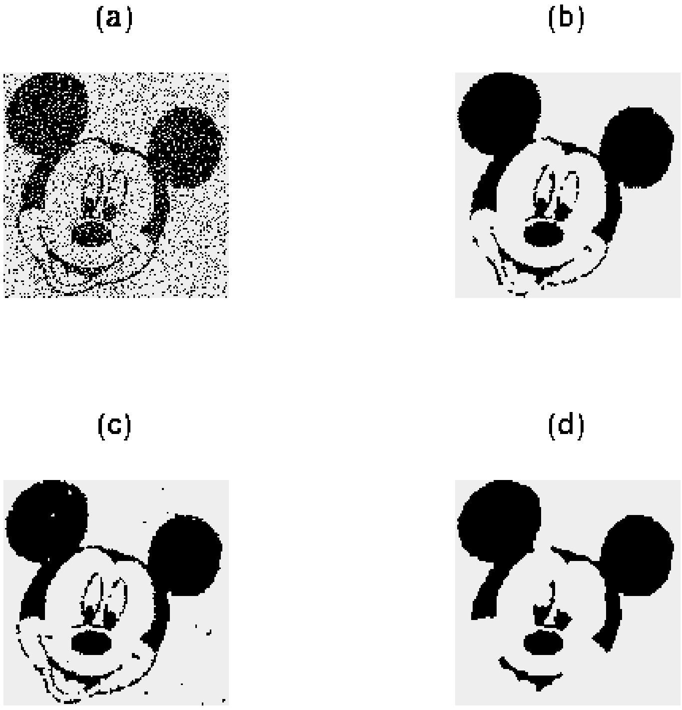


Figure 2: Degraded Mickey Image and Different Unsupervised Restorations: (a) Original image degraded with 15% channel noise. (b) Bayesian morphology based on a likelihood criterion. (c) Bayesian morphology based on a pseudo-likelihood criterion. In (b) and (c), the initial estimate of the noise parameter $\hat{p}_0$ was 0.35. (d) Blind restoration using simple majority rule.

Table 1: Results for Figure 2 (Mickey image with 15% channel noise). f, k and p are defined in the text.

	image	iterations	calls to f	final k	final p	error rate
(a)	15% channel noise					14.9%
(b)	Likelihood criterion ($\hat{p}_0 = 0.35$)	7	30	4	14.4%	1.8%
(c)	Pseudo-likelihood criterion ($\hat{p}_0 = 0.35$)	9	30	3	13.9%	1.8%
(d)	Blind restoration ($w = 1$)	20		1	16.4%	3.2%

assuming the initial classification to be a noisy version of the final one according to the model in Section 3.2 . The performance of such an alternative depends on the quality of the initial restoration. It cannot be expected to perform quite as well as algorithms that do not lose track of the original (continuous, grey-level, etc.) data, but in our experiments it performs similarly and it can be much faster.

The maximum likelihood classifier requires either that the user know in advance or be prepared to specify the parameters that define $f(y_i|x_i)$ Alternatively, *representative* training data must be available, consisting of pixels whose class is known. These conditions will often not be met. Here we used a simple alternative, *marginal mixture EM segmentation*, which yields a reasonable initial classification without requiring either specialized user knowledge or training data. The *marginal* distribution of grey-scale pixel intensities is assumed to be of the form of a finite mixture of distributions. More details of this procedure can be found in Forbes and Raftery (1997).

For multispectral images, the same approach as for grey-scale images can be used, in principle. There is an additional difficulty, however, namely the important problem of initializing the EM algorithm. We have found that agglomerative hierarchical *model-based clustering* (Banfield and Raftery 1993; Celeux and Govaert 1995) provides a sensible way of doing this; see Dasgupta and Raftery (1998). This can be done using the `mclust` software, which is part of S-PLUS, and can also be obtained at `http://lib.stat.cmu.edu/general/mclust` (Fortran version), or `http://lib.stat.cmu.edu/S/mclust` (S version).

For standard-sized images, such as 256×256 or 512×512, this clustering can be computationally expensive, but clustering on the basis of a subsample of pixels is fast and often works well; see Banfield and Raftery (1993). The resulting parameter estimates for the mixture model can be used to initialize the EM algorithm, or to provide an initial classification directly via discriminant analysis (Banfield and Raftery 1993; Bensmail and Celeux 1996).

Another way of getting a fast first classification via marginal mixture EM segmentation without first subsampling the pixels is described in Posse (1998). It consists of combining a Minimum Spanning Tree algorithm with a classification procedure so that a large number of points (pixels) can be handled.

5.3 Example: Grey-level Image of a Dog Lung

Figure 3(a) is a grey-level image of a dog lung. The aim here was to distinguish the lung from the rest of the image in order to measure the heterogeneity of the tissue in the region of interest. Only pixels in this delimited area will then be considered to compute a heterogeneity measure, such as a coefficient of variation. The EM algorithm was used to compute the three-component segmentation (b) which is a good initial segmentation apart from some misclassified pixels. The final segmentations using standard ICM and Bayesian morphology are shown in (c) and (d).

Although perhaps visually more pleasing, the standard ICM segmentation (c) is slightly too smooth for the specific purpose of the study and the Bayesian morphology segmentation (d) may be preferred. Note that this observation is consistent with the corresponding estimates $\hat{\beta}$ of β. In the Bayesian morphology case, only bounds are available for $\hat{\beta}$ ($1.05 \leq \hat{\beta} < 1.40$). For ICM, the estimate is $\beta = 1.44$, and this is higher than the higher bound, confirming that the subsequent restoration may be smoother. The Bayesian morphology segmentation represents the non-lung area inside the lung as an "island", while the ICM segmentation connects it to the area outside the lung; the researchers preferred the Bayesian morphology solution in this respect. Also, standard ICM produces a more jagged restoration of the outside circle than Bayesian morphology, and again in this respect our external knowledge indicates the Bayesian morphology solution to be more faithful to reality. In addition, Bayesian morphology was 3.5 times faster than ICM.

5.4 Example: Precipitation Climatology

Input data are data for a global precipitation climatology that has been produced at the Joint Institute for the Study of the Atmosphere and Ocean; they are available on the Web at `tao.atmos.washington.edu/legates_msu`. The spatial resolution of this climatology is 2.5 degrees in each of latitude and longitude, which leads to a set of twelve 144×72 maps representing stations or points (pixels) at which monthly average precipitation (in mm) has been recorded or extrapolated, for each individual calendar month. Figure 4 shows such a map for the month of January.

Possible goals of classifications of these data into a small number of components include building climatic regionalizations to show climatic variability, or defining local forecast zones consisting of groups of stations, each of which would be considered a single locale for forecasting purposes, thus reducing the total number of stations. More background can be found in Fovell (1997) where similar data have been considered for the purpose of delineating climate zones of the conterminous United States. In that study, standard clustering techniques that do not take spatial location and dependence into account were used. They have the disadvantage of producing small separate entities which are not climatologically meaningful. Using ICM or Bayesian morphology to classify these data has the advantage of producing more spatially cohesive regionalizations. Small isolated regions are removed by automatically reassigning the stations (pixels) located in them, reducing the spatial fragmentation of the classifications.

We computed six-component classifications following a suggestion in Fovell (1997). As these data are far from normally distributed, a non-linear transform was first applied: the power 0.25 of each record was taken. A fast first segmentation of the 144×72 12-band image was then obtained using the technique described in Posse (1998). Figure 5 shows this classification after applying Bayesian morphology with an initial value of $\hat{w}$ set to 1. The Bayesian morphology segmentation ($1.15 \leq \hat{\beta} < 1.53$) appears too fragmented in some regions while the ICM segmentation ($\hat{\beta} = 1.77$) appears too smooth. In the latter, significant regions are removed near the islands of Madagascar and New Zealand, for instance. Another advantage of Bayesian morphology is that for this 12-band image, it was 13 times faster than ICM.

Note that, depending on the goal of the analysis, the data may be pre-processed differently. If the goal is to construct local forecast zones, it may be judged prudent to standardize the records to eliminate level (mean) and seasonality (variance) distinctions. In the present application, we were more concerned with the construction of climatic zones for which the level and seasonality components are useful information. Time series can be computed by averaging the data for the members in each class of Figure 5 (see Figure 17 in Forbes and Raftery (1997)). In this classification, class 1 is represented in white while class 6 is in black. Class 1 and 2 correspond to dry regions with different seasonalities. Class 6 corresponds to wet areas with light seasonal variations. Class 3 includes regions with high variability and dry-summer climates, while class 5 includes continental interiors which generally exhibit a wet-summer/dry-winter cycle. Class 4 is characterized by moderate rainfall all year round, with a slight peak in late summer and fall. This includes much of the industrialized world: most of Europe, eastern North America, eastern Australia and, arguably, Japan, as well as large parts of southern Asia and both major oceans. This seems to be a rather heterogeneous class, suggesting that more than six classes may be needed.

6 Discussion

We have introduced the Bayesian morphology approach to image segmentation and to the restoration of unordered-color images. For grey-level and multispectral images, this consists of first segmenting the image using marginal mixture EM segmentation, and then applying

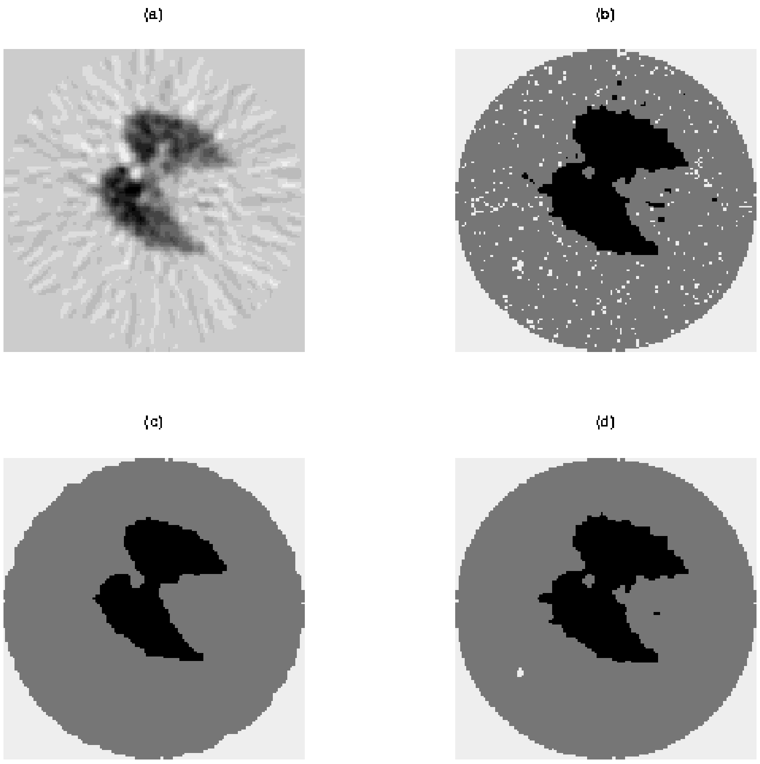
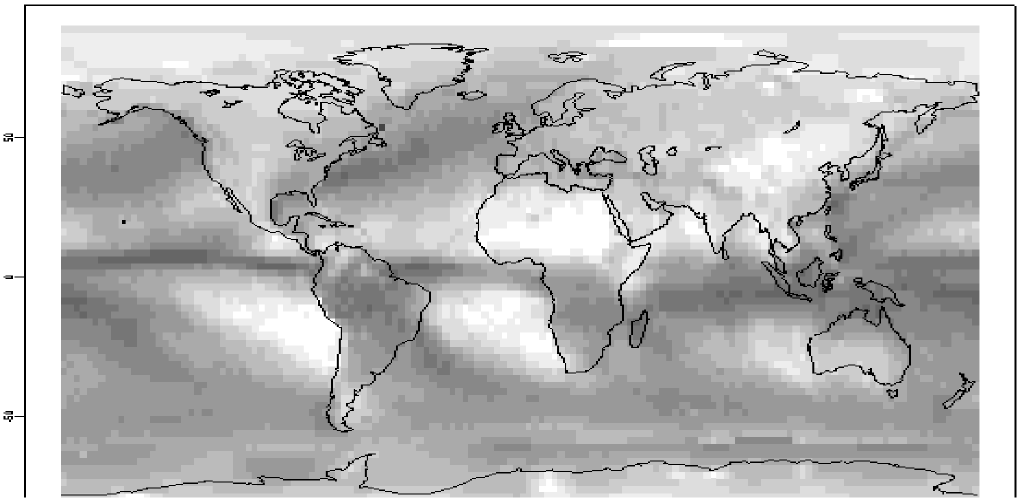


Figure 3: A Grey-Level Image of a Dog Lung, and Different Unsupervised Segmentations: (a) Original grey-level image. (b) Initial three-component segmentation using the marginal mixture model EM method. (c) Standard ICM segmentation, starting from (b). (d) Bayesian morphology segmentation, modeling the initial segmentation (b) as being degraded by channel noise.



ial

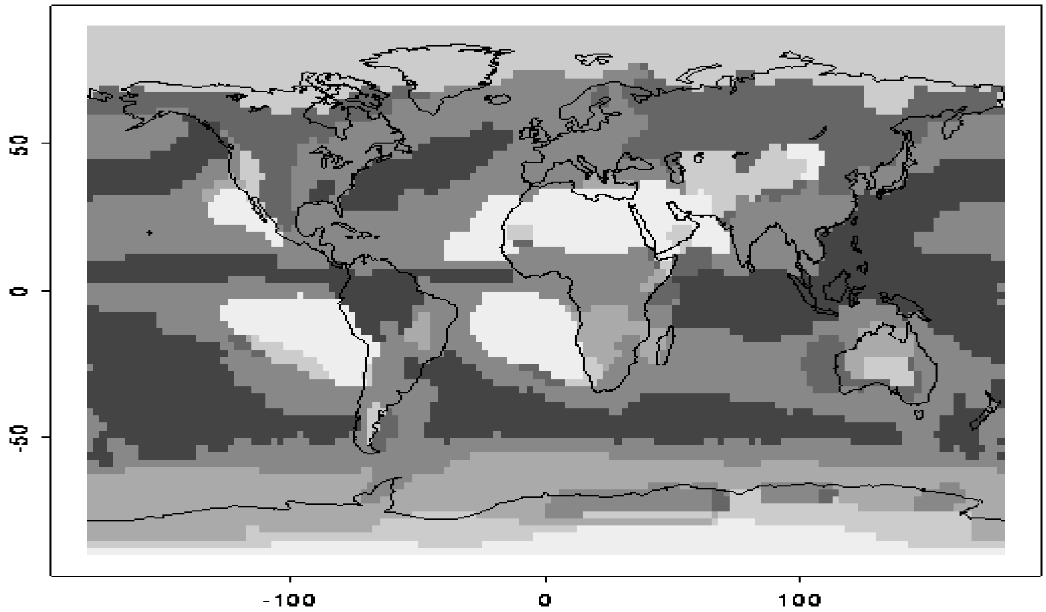


Figure 5: Bayesian morphology segmentation of the Legates/MSU precipitation 12-band image, modeling the initial Marginal Mixture EM segmentation as being degraded by channel noise. The initial value of $\hat{w}$ was set to 1.

discrete-valued ICM to the resulting segmentation. We have shown that (1) discrete-valued ICM is equivalent to a form of mathematical morphology; and (2) we can find the restoration corresponding to optimal parameter estimates without actually having to find these estimates. These two results lead to considerable computational savings.

Bayesian morphology performs similarly to standard ICM in our experiments, but is much faster. In our grey-level image experiments, it was three to eight times faster, and the savings were greater for the images that required the most CPU time. For our experiments with multispectral images the savings were even greater. For one 12-band image, Bayesian morphology was over 13 times faster than ICM. This opens the possibility of applying Bayesian image analysis in contexts where it has not been feasible previously because great speed was required, such as real-time and interactive image analysis and real-time video processing.

Bayesian morphology may also contribute improvements to the practice of mathematical morphology. This is already very fast and performs well; however it lacks a formal statistical foundation for inference, and the specification of the sequence of morphological operations to be used is done in a rather *ad hoc* manner. Because Bayesian morphology is based on a statistical model, issues such as the choice of a structuring element and the choice of the sequence of morphological operations can be reduced to issues of statistical model selection, and addressed, at least in principle, with Bayesian model selection by computing approximate Bayes factors (Kass and Raftery 1995; Ji and Seymour 1996).

A limitation of our study is that the insensitivity conditions we use seem to be limited to simple image and noise models. It may be interesting to investigate further whether the analysis of more complex image models can be exactly or approximately divided into different steps involving unordered-color image restorations. Performing those restorations faster will then be relevant for a larger class of problems. For grey-level and multispectral images, one possible extension would be to iterate between the EM and ICM steps of the Bayesian morphology algorithm.

Another possible extension is the use of other models for the unobserved image. The Ising and Potts models do not always capture the image characteristics well enough. Other models have been proposed and studied by Descombes, Mangin, Pechersky, and Sigelle (1995), Tjelmeland and Besag (1998), Descombes, Morris, and Zerubia (1996). They are higher order interaction Markov random fields which involve three parameters regulating the presence of noise, edge and line configurations. Each pixel is considered to have a 5×5 neighborhood consisting of the 24 nearest pixels. Such a neighborhood structure induces cliques which are 3×3 squares. For binary Markov random fields, a configuration on a 3×3

clique can be classified using the symmetries (symmetry black-white, rotation) in one of 51 classes. A potential function is associated with each class and constraints are imposed by relation between these functions. They involve three parameters e, l, n. The parameters e and l can be interpreted as the edge and line energies per unit of length. The parameter n allows one to control the amount of noise. It is associated with noise configurations, typically those which are not involved in edge or line features. The resulting distribution has the form

$$I\!P(x) \;\propto\; \exp(-ev_0(x) - lv_1(x) - nv_2(x)) \;,$$

where v_0, v_1, v_2 depend on the 3×3 configurations in x. In particular, $v_2(x)$ is the number of noise configurations in x .

Our study can be easily extended to a one parameter version of these models. We can restrict ourselves to a situation where only the noise parameter n is unknown by setting the edge and line parameters to some constants. The reason for such a restriction is that insensitivity conditions still hold for models with several parameters but it seems that there is no real gain from using them. Note that these models require more computation and, although they are more flexible, users may still prefer the limited but simple Potts model in most cases.

Other estimation criteria such as those that take into account the observed image y may also be worth investigating. In Qian and Titterington (1992), the parameters are estimated by maximizing

$$\prod_{i \in S} I\!P(y_i \mid \hat{x}_{S\setminus\{i\}}) \;.$$

Focusing on β leads to equations similar to those in Section 4 (see Forbes and Raftery (1997)) but the choice of an appropriate function f is not straightforward.

In addition, we believe that the morphological point of view can simplify the implementation of statistical algorithms such as ICM and give better insight into their convergence properties, in particular through the properties of iterations of morphological transformations (see Heijmans 1994). This also includes considering alternatives to the commonly used Markov random fields priors such as, for instance, morphologically constrained Markov random fields (Sivakumar and Goutsias 1997). The latter incorporate geometric properties more clearly and may induce algorithms with clear morphological characteristics. Note that although our study and Sivakumar and Goutsias (1997) both attempt to link Bayesian image analysis and mathematical morphology, the goals and results are quite different. Sivakumar and Goutsias (1997) use ideas from mathematical morphology to try to build more complex and realistic Markov random field models for images. Applications of such models to texture simulation are illustrated in Carstensen (1992). In our study, on the other hand, we have identified equivalences between simple and basic versions of both methodologies, and tried to use them to find ways in which insights from each approach could yield improvements in the other one.

References

Banfield, J. D. and A. E. Raftery (1993). Model-based Gaussian and Non-Gaussian Clustering. *Biometrics 49*, 803–821.

Bensmail, H. and G. Celeux (1996). Regularized Gaussian discriminant analysis through eigenvalue decomposition. *Journal of the American Statistical Association 91*, 1743–1748.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician 24*, 179–195.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, series B 48*, 259–302.

Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics 43*, 1–59.

Carstensen, J. M. (1992). *Description and simulation of visual texture*. Ph. D. thesis, Institute of Mathematical Statistics and Operations Research. Technical University of Denmark, Lyngby.

Celeux, G. and G. Govaert (1995). Gaussian Parsimonious Clustering Models. *Pattern Recognition 28*, 781–793.

Comets, F. and B. Gidas (1992). Parameter estimation for Gibbs distributions from partially observed data. *The Annals of Applied Probability 2*(1), 142–170.

Dasgupta, A. and A. E. Raftery (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association 93*, 294–302.

Descombes, X., J.-F. Mangin, E. Pechersky, and M. Sigelle (1995). Fine structures preserving Markov model for image processing. *Proceedings of the 9th Scandinavian Conference on Image Analysis*, 349–356.

Descombes, X., R. Morris, and J. Zerubia (1996). Estimation of Markov random field prior parameters using Markov chain Monte-Carlo maximum likelihood. Technical Report 3015, INRIA, Sophia Antipolis, France. `<http://www.inria.fr/rapports/sophia/RR-3015.html>`.

Forbes, F. and A. E. Raftery (1997). Bayesian Morphology: Fast Unsupervised Bayesian Image Analysis. Technical Report 325, Department of Statistics, University of Washington, Seattle. `<http://www.stat.washington.edu/tech.reports/tr325.ps>`.

Fovell, R. G. (1997). Consensus clustering of U.S. temperature and precipitation data. *Journal of Climate 10*, 1405–1427.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence 6*, 721–741.

Georgii, H. O. (1988). *Gibbs Measures and Phase Transitions.* de Gruyter, Berlin.

Geyer, C. J. (1991). Reweighting Monte-Carlo mixtures. Technical Report 568, School of Statistics, University of Minnesota.

Geyer, C. J. and E. A. Thompson (1992). Constrained Monte-Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, series B 54*, 657–699.

Gidas, B. (1993). Parameter estimation for Gibbs distributions from fully observed data. In *Chellapa, R. and Jain, A. (eds.) Markov Random Fields - Theory and Applications*, pp. 471–498. Academic Press, Boston.

Goutsias, J. K. (1991). Unilateral approximation of Gibbs random field images. *Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing 53*, 240–257.

Heijmans, H. J. (1994). *Morphological Image Operators.* Academic Press: Boston.

Hummel, R. A. and S. W. Zucker (1983). On the foundations of relaxation labelling processes. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence 5*, 267–287.

Ji, C. and L. Seymour (1996). A consistent model selection procedure for Markov random fields based on penalized pseudolikelihood. *Annals of Applied Probability 6*, 423–443.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–795.

Matheron, G. (1975). *Random Sets and Integral Geometry*. New York: John Wiley.

Posse, C. (1998). Hierarchical Model-based Clustering for Large Data Sets. *Unpublished manuscript, Department of Statistics, University of Minnesota.*

Qian, W. and M. Titterington (1992). Stochastic relaxations and EM algorithms for Markov random fields. *Journal of Statistical Computation and Simulation 40*, 55–69.

Rosenfeld, A., R. A. Hummel, and S. W. Zucker (1976). Scene labelling by relaxation operation. *I.E.E.E. Transactions on Systems, Man and Cybernetics 6*, 420–433.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.

Sivakumar, K. and J. K. Goutsias (1997). Morphologically constrained discrete random sets. In *Advances in Theory and Applications of Random Sets, D. Jeulin (Ed.)*. World Scientific Publishing Company.

Swendsen, R. H. and J. S. Wang (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters 58*, 86–88.

Tjelmeland, H. (1996). *Stochastic models in reservoir characterization and Markov random fields for compact objects*. Ph. D. thesis, Department of Mathematical Sciences. Norwegian University of Science and Technology, Trondheim.

Tjelmeland, H. and J. Besag (1998). Markov random fields with higher order interactions. *Scandinavian Journal of Statistics*, To appear.

Zhang, B., M. N. Shirazi, and H. Noda (1996). Blind restoration of degraded binary Markov random field images. *Graphical models and Image Processing 58*, 90–98.