

# Bayesian Morphology: Fast Unsupervised Bayesian Image Analysis

Florence Forbes, Adrian E. Raftery

► **To cite this version:**

Florence Forbes, Adrian E. Raftery. Bayesian Morphology: Fast Unsupervised Bayesian Image Analysis. RR-3374, INRIA. 1998. inria-00073315

**HAL Id: inria-00073315**

**<https://hal.inria.fr/inria-00073315>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***Bayesian Morphology: Fast Unsupervised Bayesian  
Image Analysis***

Florence Forbes , Adrian E. Raftery

**No 3374**

Mars 1998

\_\_\_\_\_ THÈME 4 \_\_\_\_\_



***R*** *apport  
de recherche*



# Bayesian Morphology: Fast Unsupervised Bayesian Image Analysis

Florence Forbes , Adrian E. Raftery

Thème 4 — Simulation et optimisation  
de systèmes complexes  
Projet IS2

Rapport de recherche n° 3374 — Mars 1998 — 48 pages

**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 very 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 image 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.

(Résumé : *tsvp*)

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](mailto: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](mailto:raftery@stat.washington.edu); WEB: [www.stat.washington.edu/raftery](http://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 ÉTRANGÈRES, 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, JOHN WALLACE FOR HELPFUL DISCUSSIONS, TO DR. H. THOMAS ROBERTSON, UW DIVISION OF PULMONARY AND CRITICAL CARE FOR SHARING THE DATA IN FIGURE 12 (A) AND TO XAVIER DESCOMBES, INRIA SOPHIA-ANTIPOLIS FOR PROVIDING FIGURE 6.

## Approches morphologique et bayésienne en analyse d'image: restaurations rapides non supervisées.

**Résumé :** Nous considérons les problèmes de segmentation et de classification ainsi que celui de la restauration d'image lorsque l'image non dégradée est constituée d'un petit nombre de couleurs non ordonnées. Nous nous intéressons en particulier aux questions de performance et de rapidité. Celles-ci n'ont cessé de prendre de l'importance, notamment dans l'analyse d'images de grande taille ou de grande dimension, dans la gestion de grandes bases de données d'images, dans l'analyse en temps réel ou quasi temps réel, et dans l'analyse en ligne de séquences d'images vidéo.

L'analyse bayésienne d'image (Geman et Geman 1984) procure des solutions élégantes à ces problèmes mais est généralement coûteuse en calculs. De plus, les solutions fournies peuvent être sensibles aux propriétés globales non réalistes des modèles adoptés. L'algorithme ICM (Besag 1986) est une alternative plus rapide et qui repose sur des propriétés *locales* des modèles sous-jacents. En particulier, l'estimation des paramètres peut être faite de manière itérative, selon un critère de pseudo-vraisemblance. La morphologie mathématique (Matheron 1975), reconnue pour ses bonnes performances, est également une solution plus rapide mais qui manque de justifications statistiques solides. Le choix des opérateurs à appliquer se fait, par exemple, bien souvent de manière *ad hoc*.

Nous proposons de faire une sorte de synthèse de ces deux types d'approches, *la morphologie bayésienne*, dans le but de combiner les performances de la morphologie mathématique et les bien-fondés statistiques d'ICM. L'idée est que lorsque l'image initiale est discrète (ou lorsque nous disposons d'une première segmentation), nous pouvons adopter des modèles d'image et de bruit pour lesquels (1) l'algorithme ICM est équivalent à une forme de morphologie mathématique, (2) l'algorithme ne dépend pas des valeurs précises des paramètres des modèles. De plus, dans ce cadre, l'estimation par maximum de vraisemblance devient envisageable. Pour des images en niveaux de gris ou pour des images multispectrales, nous nous ramenons au cas précédent en calculant une première segmentation en une image discrète, à l'aide de l'algorithme EM pour les mélanges. L'algorithme qui en résulte est beaucoup plus rapide que l'algorithme ICM traditionnel et ceci d'autant plus que la dimension et la taille des images augmentent. Ses performances sont bonnes sur des exemples simulés et réels.

**Mots-clé :** Restauration bayésienne d'image; Algorithme ICM; Morphologie mathématique; Modèle de Potts; Pseudo-vraisemblance.

## Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Background: Bayesian Image Analysis, ICM and Mathematical Morphology</b>	<b>7</b>
2.1	Bayesian and Morphological Approaches . . . . .	7
2.2	Bayesian Image Restoration and the ICM Algorithm . . . . .	8
2.3	Binary Mathematical Morphology . . . . .	9
<b>3</b>	<b>Bayesian Morphology</b>	<b>11</b>
3.1	Binary Images and Insensitivity Conditions in ICM . . . . .	11
3.2	Unordered-Color Images . . . . .	14
3.3	Grey-scale Images and Marginal Mixture EM Segmentation . . . . .	16
3.4	Multispectral Images and Model-Based Clustering . . . . .	18
<b>4</b>	<b>Online Parameter Estimation</b>	<b>19</b>
4.1	Unsupervised Restoration . . . . .	19
4.2	Estimation Criteria . . . . .	19
4.3	Maximum Likelihood Criterion . . . . .	20
4.4	Maximum Pseudo-Likelihood Criterion . . . . .	22
<b>5</b>	<b>Experiments</b>	<b>23</b>
5.1	Restoring Unordered-Color Images . . . . .	23
5.2	Segmenting Grey-Scale Images . . . . .	32
5.3	Example: PET Image of a Dog Lung . . . . .	33
5.4	Example: Precipitation Climatology . . . . .	33
<b>6</b>	<b>Discussion</b>	<b>44</b>

## List of Tables

1	Results for Mickey image with 15% channel noise . . . . .	26
2	Results for Matisse image with 15% channel noise . . . . .	27
3	Results for images with 30% channel noise . . . . .	29
4	Results for four-color image with 15% channel noise . . . . .	29
5	Results for four-color image with 30% channel noise . . . . .	30
6	Results for grey-level Matisse image with light noise . . . . .	36
7	Results for grey-level Matisse image with heavy noise . . . . .	36
8	Results for dog lung PET image . . . . .	36
9	Results for Legates/MSU precipitation 12-band image . . . . .	36

## List of Figures

1	Insensitivity conditions . . . . .	15
2	Two true binary images: Mickey and Matisse . . . . .	25
3	Degraded Mickey image and unsupervised restorations . . . . .	26
4	Degraded Matisse image and unsupervised restorations . . . . .	27
5	Images with 30% channel noise and different restorations . . . . .	28
6	Four-color test image . . . . .	29
7	Four-color image degraded with 15% noise and unsupervised restorations . . . . .	30
8	Four-color image degraded with 30% noise and unsupervised restorations . . . . .	31
9	Pixel intensity histograms with density EM estimates . . . . .	32
10	Degraded grey-level image with light noise and unsupervised restorations . . . . .	35
11	Degraded grey-level image with heavy noise and unsupervised restorations . . . . .	37
12	A grey-level PET scan image of a dog lung, and unsupervised segmentations. . . . .	38
13	Monthly average precipitation (in mm) for the month of January . . . . .	39
14	Marginal mixture EM of the Legates/MSU precipitation 12-band image . . . . .	40
15	Bayesian morphology segmentation of the Legates/MSU precipitation 12-band image . . . . .	41
16	ICM segmentation of the Legates/MSU precipitation 12-band image . . . . .	42
17	Time series (average) for each class in Figure 15 . . . . .	43

# 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. The problem considered here could also be a first step in methods for solving more specialized object recognition problems, such as deformable templates (Grenander, Chow, and Keenan 1991) or the EP algorithm (Banfield and Raftery 1992).

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 was the first time that a fully statistically grounded method was proposed, but it is very computationally expensive, and the results it produces may be sensitive to unrealistic global properties of the underlying models, as pointed out by Besag (1986).

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 off line from training data if these are available, or estimated iteratively using maximum pseudo-likelihood. Exact maximum likelihood is typically intractable for this model.

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 (Titterton, Smith, and Makov 1985). We call this *marginal mixture EM segmentation*. With multispectral images, the distributions are multivariate, and there has been much progress on estimation methods that combine agglomerative hierarchical clustering methods based on maximum classification likelihood (Banfield and Raftery 1993) with the EM algorithm (Dasgupta and Raftery 1998; Byers and Raftery 1998). When the image is very large, it may be necessary to take a sample of the pixels first, as in Banfield and Raftery (1993). The use of minimum spanning trees (Posse 1998) has the potential to make this approach feasible even with very large images, without subsampling the pixels.

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. In Section 3.1, we focus on binary images. The isotropic Ising model with region-dependent channel transmission noise is considered and the corresponding insensitivity conditions for parameter estimation are derived. For more general unordered-color images, an explicit formulation of ICM and the insensitivity conditions are given in Section 3.2 for the Potts model with channel transmission noise. In Sections 3.3 and 3.4 we outline the extension of Bayesian morphology to grey-scale and multispectral images via marginal mixture EM segmentation. Online parameter estimation, or more precisely adaptive online specification of the algorithm, is described in Section 4. We examine in turn likelihood and pseudo-likelihood criteria for parameter estimation. A number of experiments are presented in Section 5. Connections to other work and possible extensions are discussed in Section 6.

## 2 Background: Bayesian Image Analysis, ICM and Mathematical Morphology

### 2.1 Bayesian and Morphological Approaches

Mathematical morphology (see Serra (1982) for a general reference) and Bayesian image analysis (Geman and Geman 1984; Besag, York, and Mollié 1991) are two of the most successful approaches in image analysis. Morphology is a collection of image transforms based on the set-theoretical operations union and intersection, and can be used as a computationally efficient tool-box for image processing, performing for instance non-linear filtering of noisy images. In practice, it involves selecting cleverly chosen sequences of morphological transformations but it is not generally based on statistical considerations.

Morphology and Bayesian methods arise from different philosophies and are formulated very differently. A question of interest is to investigate the possible relationships between these two approaches. The idea is that this may yield a combination of the power and elegance of the Bayesian approach with the speed and performance of morphology. On one hand, most morphological operators can be implemented in such a way that they require little computation while the Bayesian approach may be more computationally demanding. On the other hand, providing mathematical morphology with some statistical analysis could lead to better insight into how to implement it in practice. With this in mind, we considered the ICM algorithm for some simple models (see Sections 3.1 and 3.2). For binary images, the ICM algorithm can be viewed as a succession of operators that generalize some of the most primitive morphological operators. This formulation reveals some interesting features of ICM that we shall discuss in more detail in Sections 3.1 and 3.2. We first present the ICM algorithm in a Bayesian context and review some basic definitions of binary mathematical morphology.

## 2.2 Bayesian Image Restoration and the ICM Algorithm

Bayesian image analysis is based on probability models. It includes a variety of tasks of which one of the most straightforward is 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  $x$ .

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  $\mathbb{P}(x)$  satisfies, for all pixels  $i$  in  $S$ ,

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

i.e., the conditional distribution of  $\mathbb{P}(x)$  depends only on the values of pixels in a subset  $N(i)$  of  $S \setminus \{i\}$ , 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

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

When pixels are updated sequentially, choosing values that maximize the conditional probability  $\mathbb{P}(x_i|x_{S\setminus\{i\}}, y)$  increases the posterior distribution 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

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

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.

### 2.3 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  $3 \times 3$  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  $i$ ,

$$B(i) = \{i + j, j \in B\}. \quad (3)$$

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  $3 \times 3$  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  $I$ .

Erosion and dilation are dual notions. The so-called duality principle plays an important role in mathematical morphology. The dual,  $\psi^*$ , of a morphological operator  $\psi$  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  $\psi$  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  $W = \{w_1, w_2, \dots, w_m\}$  be the weights  $w_i$  associated with each pixel  $i$  in  $B$  and  $r$  be a threshold or rank for  $w_1, \dots, w_m, r \in \mathbb{Z}$ . The weighted rank operator  $\rho_{B,W,r}$  is defined by

$$\rho_{B,W,r}(I) = \{i \in S, \sum_{j \in B(i)} w_j I_j \geq r\} . \quad (4)$$

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  $\psi_1$  and  $\psi_2$ , 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

$$\rho_{B,m} \leq \rho_{B,m-1} \leq \dots \leq \rho_{B,1}.$$

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 (11) of ICM. They appear for different structuring elements depending on the neighborhood structure chosen in the Bayesian analysis. One interest of such a formulation is that it shows how the problematic model and noise parameter estimation can be reduced to the easier estimation of the ranks. More details are given in the next section.

### 3 Bayesian Morphology

#### 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,

$$\mathbb{P}(x) = Z(\beta)^{-1} \exp(\beta v(x)), \quad (5)$$

where

$$v(x) = \sum_{i \sim j} \delta(x_i, x_j) \quad (6)$$

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,

$$Z(\beta) = \sum_x \exp(\beta v(x)). \quad (7)$$

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

$$\mathbb{P}(x_i | x_{S \setminus \{i\}}) \propto \exp(\beta u_i(x_i)), \quad (8)$$

where

$$u_i(x_i) = \sum_{j \in N(i)} \delta(x_i, x_j) \quad (9)$$

is the number of neighbors of pixel  $i$  having color  $x_i$ . The model depends on a parameter  $\beta$  which is taken to be positive, reflecting the assumption that neighboring pixels tend to be of the same color. Note that in the original Ising model the state space is  $\{-1, 1\}$ . The model above is equivalent by identifying the state space to  $\{0, 1\}$  and dividing  $\beta$  by 2. This is more natural in our image setting and we shall still refer to it as the Ising model.

The true images are then assumed to be degraded by a region-dependent flip-flap, or channel transmission noise characterized by the two parameters  $p_{01}$  and  $p_{10}$ , where

$$p_{01} = \mathbb{P}(y_i = 1 | x_i = 0) ,$$

$$p_{10} = \mathbb{P}(y_i = 0 | x_i = 1) .$$

We shall assume, without loss of generality, that  $p_{01}$  and  $p_{10}$  belong to the open interval  $(0, 0.5)$ . 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

$$f(y_i | 1) = (1 - p_{10})^{y_i} (p_{10})^{1 - y_i} ,$$

$$f(y_i | 0) = (p_{01})^{y_i} (1 - p_{01})^{1 - y_i} .$$

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

For an Ising model with region-dependent flip-flap 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

$$x_i^* = \begin{cases} 1 & \text{if } u_i(1) - u_i(0) \geq 2w_1 \\ 0 & \text{if } u_i(1) - u_i(0) < -2w_0 \\ y_i & \text{if } -2w_0 \leq u_i(1) - u_i(0) < 2w_1 , \end{cases} \quad (10)$$

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

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

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

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,  $u_i(1)$  can be written

$$u_i(1) = \sum_{j \in N(i)} x_j,$$

and an equivalent formulation of (10) can be given in terms of the rank operators defined in (4):

$$x_i^* = \begin{cases} [\rho_{N,r_0}(\hat{x})]_i & \text{if } y_i = 0 \\ [\rho_{N,r_1}(\hat{x})]_i & \text{if } y_i = 1, \end{cases} \quad (11)$$

where

$$\begin{aligned} r_0 &= \frac{m}{2} + w_1 \\ r_1 &= \frac{m}{2} - w_0, \end{aligned}$$

and where  $[z]_i$  is the value of the image  $z$  at pixel  $i$ . A shorter way to write (11) 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 (10) depends on the parameters  $p_{01}$ ,  $p_{10}$  and  $\beta$  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 (10) or, equivalently, the rank operators  $\rho_{N,r_0}$  or  $\rho_{N,r_1}$ . Unfortunately, in the general case,  $p_{01}$ ,  $p_{10}$  and  $\beta$  are unknown and so are  $w_0$  and  $w_1$ . However, we have the following result.

‡ If there exist integers  $k_0$  and  $k_1$  such that  $p_{01}, p_{10}, \beta$  satisfy

$$\frac{1}{2(k_0 + 1)} \log \left( \frac{1 - p_{10}}{p_{01}} \right) < \beta \leq \frac{1}{2k_0} \log \left( \frac{1 - p_{10}}{p_{01}} \right), \quad (12)$$

$$\frac{1}{2k_1} \log \left( \frac{1 - p_{01}}{p_{10}} \right) \leq \beta < \frac{1}{2(k_1 - 1)} \log \left( \frac{1 - p_{01}}{p_{10}} \right), \quad (13)$$

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

$$p_{01} a^{k_0} \leq p_{11} < p_{01} a^{k_0+1} \quad (14)$$

$$p_{01} \frac{1}{a^{k_1-1}} + 1 - \frac{1}{a^{k_1-1}} < p_{11} \leq p_{01} \frac{1}{a^{k_1}} + 1 - \frac{1}{a^{k_1}}, \quad (15)$$

where  $a = \exp(2\beta)$ . It follows that the update rule (10) does not depend on the exact values of  $p_{01}, p_{10}$  and  $\beta$ , but rather on how they are related to each other. We refer to equations (12) and (13) (or (14) and (15)) 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  $w_1$  consists then in finding  $k_0$  and  $k_1$  such that the insensitivity conditions hold.



We can restrict our search to values of  $k_0$  in  $\{0, 1, \dots, |N(i)|/2 - 1\}$  and  $k_1$  in  $\{1, \dots, |N(i)|/2\}$ . 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  $\rho_{N,r_0}$  and  $\rho_{N,r_1}$  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

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

$$\beta \geq \frac{1}{2} \log \left( \frac{1 - p_{01}}{p_{10}} \right), \quad (17)$$

leading to the theorem below. ¶ If  $p_{01}, p_{10}, \beta$  satisfy conditions (16) and (17), then the update rule (10) does not involve any parameters. It becomes

$$x_i^* = \begin{cases} 1 & \text{if } u_i(1) > u_i(0) \\ 0 & \text{if } u_i(1) < u_i(0) \\ y_i & \text{if } u_i(1) = u_i(0). \end{cases} \quad (18)$$

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 (18) 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 restorations consists of estimating  $w_0$  and  $w_1$ . We call the resulting approach *Bayesian morphology*. We will show, in Section 4, that this turns out to be much easier than computing point estimates for  $p_{01}, p_{10}$  and  $\beta$ . 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 a color) indicating the pixel's class. It may appear less natural in image restoration where grey-level images are usually considered. However, it is the most straightforward generalization of the isotropic Ising model to more than two unordered colors and we will show that insensitivity conditions also exist for this model allowing us to derive fast unsupervised restorations for more general images.

The standard nearest-neighbor Potts model is defined by the joint distribution

$$\mathbb{P}(x) = Z(\beta)^{-1} \exp(\beta v(x)), \quad (19)$$

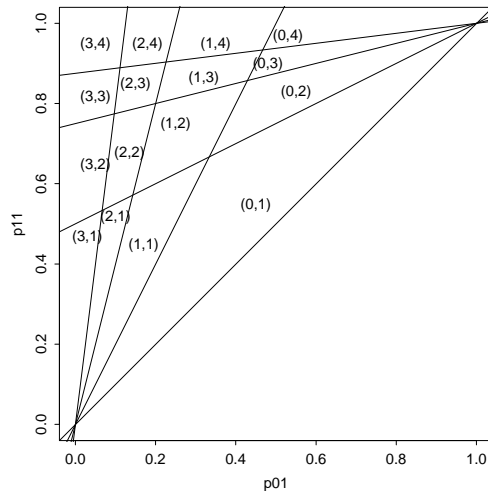


Figure 1: Insensitivity Conditions in Terms of  $p_{01}$  and  $p_{11}$  for an Eight-Pixel Neighborhood with  $\beta = 0.35$ . The lines delimit regions where  $w_0$  and  $w_1$  are constant. Their respective values are shown in parentheses.

where the expressions (6), (7), (8), (9) 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 (Higdon 1994, Appendix A; Morris, Descombes, and Zerubia 1996).

To derive insensitivity conditions as in Section 3.1, an important point 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,

$$f(y_i|x_i) = \begin{cases} 1-p & \text{if } y_i = x_i \\ \frac{p}{C-1} & \text{otherwise.} \end{cases} \quad (20)$$

This corresponds to a 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 (20) 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 (10). 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. 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  $i$

$$x_i^* = \begin{cases} c & \text{if } u_i(c) - u_i(y_i) \geq w \\ y_i & \text{otherwise,} \end{cases} \quad (21)$$

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 (21) depends on  $\beta$  and  $p$  only through the value of  $w$ . This leads to the following insensitivity conditions. ¶ For the Potts model with channel noise, if there exists an integer  $k$  such that  $\beta$  and  $p$  satisfy

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

or, equivalently,

$$\frac{1}{k} \log \left( \frac{(1-p)(C-1)}{p} \right) \leq \beta < \frac{1}{k-1} \log \left( \frac{(1-p)(C-1)}{p} \right), \quad (23)$$

or in terms of  $p$

$$\frac{1}{1 + (C-1)^{-1} \exp(\beta k)} \leq p < \frac{1}{1 + (C-1)^{-1} \exp(\beta(k-1))} , \quad (24)$$

then  $w$  is equal to  $k$  and the ICM update rule (21) does not depend on the exact values of  $\beta$  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.

### 3.3 Grey-scale Images and Marginal Mixture EM Segmentation

In the previous cases we considered, the observed and unobserved images were of the same type. This is natural when the observations are assumed to be a noisy version of the unknown scene that we want to restore. In standard segmentation (or classification) problems, the goal is to assign to each pixel of the observed image a label indicating the pixel's class. This can be seen as a restoration problem where the image to be restored consists of unordered colors (classes) but the observations are usually ordered (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. It seems that there is no gain in using them rather than estimating the model parameters directly. 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 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  $f(y_i|x_i)$  (e.g. in the case of grey-scale Gaussian noise, the noise mean and variance for each class). Alternatively, *representative* training data must be available, consisting of pixels whose class is known. These conditions will often not be met, and requiring them seems undesirably restrictive for unsupervised restoration. Here we describe a simple alternative, *marginal mixture EM segmentation*, which yields a reasonable initial classification without requiring either specialized user knowledge or training data.

Suppose that there are  $C$  colors ( $c = 1, \dots, C$ ) in the true image, and let  $g_c(y_i) = f(y_i|x_i = c)$ ,  $c = 1, \dots, C$ , where the distribution  $g_c(\cdot)$  may have a parametric form specified by parameters  $\theta_c$  (e.g. the mean and variance if  $g_c(\cdot)$  is normal). Then if one forms the *marginal* distribution of grey-scale pixel intensities, one would expect it to be of the form of a finite mixture of  $C$  distributions:

$$g(y) = \sum_{c=1}^C \pi_c g_c(y), \quad (25)$$

where  $\pi_c \geq 0$  and  $\sum_{c=1}^C \pi_c = 1$ . Maximum likelihood estimators of the parameters  $\pi = (\pi_1, \dots, \pi_C)$  and  $\theta = (\theta_1, \dots, \theta_C)$  can be found using the EM algorithm (Dempster, Laird, and Rubin 1977; McLachlan and Basford 1988).

In its standard form, with each pixel considered as a separate observation, the EM algorithm may take a long time with images of a standard size (e.g.  $512 \times 512 = 262,144$  pixels). However, a tremendous saving can be achieved by recognizing that the number of observed distinct grey-scale values is much smaller (often 256, as for eight-bit images), grouping the pixels accordingly, and expressing the expected log-likelihood at the E step as a sum over the distinct grey-scale values rather than over the pixels. Then the required computer time is reduced roughly from the order of the number of pixels to the order of the number of distinct grey-scale values (e.g. from 262,144 to 256 for an eight-bit  $512 \times 512$  image, a reduction of about three orders of magnitude).

The final step is to calculate the posterior probability for each pixel of being of each color, given the estimated parameter values. This is done using Bayes's theorem (McLachlan and Basford 1988). Then each pixel is classified to the most likely color.

A byproduct of this is that it could be used to assess the number of distinct colors in the true image if this is unknown, at least roughly. This is done by equating the hypothesis that there are  $C$  colors with the corresponding finite mixture model (25) with  $C$  components, and comparing these models using approximate Bayes factors (Kass and Raftery 1995; Roeder and Wasserman 1997), which can be done very efficiently if the BIC approximation is used. Of course, this approach ignores spatial dependence and so may be expected to overestimate the true number of colors, but it would provide a first solution to a problem that does arise in practice but has not been much considered in the literature. Research on modifications of this that take account of the spatial dependence would be useful.

### 3.4 Multispectral Images and Model-Based Clustering

For multispectral images, the same approach as for grey-scale images can be used, in principle. The distributions  $g_c(\cdot)$  then become multivariate distributions (typically multivariate normal), the finite mixture model (25) becomes a mixture of multivariate rather than univariate distributions, and marginal mixture EM segmentation can be used, as before.

There is an additional difficulty, however, namely the problem of initializing the EM algorithm. For grey-scale images, the mixture model is univariate, and it is easy to obtain reasonable initial values of the model parameters, for example by ranking the pixels by intensity and then dividing them into equal-sized chunks. The results are usually fairly insensitive to the precise way in which this is done. For multispectral images, however, the mixture model is multivariate, and specifying reasonable initial values of the parameters is not so easy. In addition, the results can be sensitive to the way it is done, as the EM algorithm is guaranteed only to converge to a local maximum of the likelihood.

We have found that agglomerative hierarchical *model-based clustering* (Banfield and Raftery 1993; Celeux and Govaert 1995) provides a sensible way of doing this. This provides a clustering of the pixels according to their intensities (ignoring spatial information), where the clustering criterion is the classification likelihood of the mixture model (i.e. the likelihood given the classification). 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). Indeed, this clustering could be used to provide an original image for Bayesian morphology, avoiding the EM algorithm altogether, although this seems likely to perform less well (Dasgupta and Raftery 1998).

For standard-sized images, such as  $256 \times 256$  or  $512 \times 512$ , 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 (25) 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.

## 4 Online Parameter Estimation

### 4.1 Unsupervised Restoration

When the model and noise parameters are unknown, a natural approach to unsupervised restoration is to estimate the parameters first and then to estimate the true scene given the parameter estimates. This is usually done iteratively by iterating between parameter estimation and restoration, following the general procedure proposed in Besag (1986). At each cycle of iteration, new values of the prior parameters are computed from the current restored image, which is in turn updated using the new parameter values. Therefore, the most frequent approach (see for instance Descombes, Morris, and Zerubia 1996) is to study estimation procedures for the parameters independently, and then to include them in a restoration or segmentation algorithm.

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 that we shall specify in Section 4.2 . We shall consider two such criteria, maximum likelihood in Section 4.3 and maximum pseudo-likelihood in Section 4.4 .

### 4.2 Estimation Criteria

An estimation criterion is a function that measures the quality of a set of parameters. We shall write  $\beta$  to denote an element of this set 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 usually not known)

and  $f(\beta)$  is computable for any given value of  $\beta$ . Typically, if  $F$  is a convex function 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  $\beta$ , we shall refer to it 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 (23), 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  $\beta$ , a natural estimator of  $w$  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

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

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 the next subsections, we consider in detail the cases corresponding to maximum likelihood and maximum pseudo-likelihood estimation.

### 4.3 Maximum Likelihood Criterion

Consider the standard Potts model (19). The maximum likelihood estimator of  $\beta$  can be obtained by maximizing the log-likelihood defined by

$$\log(\mathbb{P}(x)) = \beta v(x) - \log(Z(\beta)).$$

The derivative of the log-likelihood is given by

$$\frac{d \log(\mathbb{P}(x))}{d\beta} = v(x) - \mathbb{E}_{\beta}[v(X)],$$

where

$$\mathbb{E}_{\beta}[v(X)] = \sum_x v(x) \mathbb{P}(x),$$

which depends on  $\beta$  through  $\mathbb{P}(x)$ . Therefore the maximum likelihood estimator  $\hat{\beta}$  of  $\beta$  satisfies

$$\mathbb{E}_{\hat{\beta}}[v(X)] = v(x). \tag{26}$$

In addition,  $\mathbb{E}_\beta[v(X)]$ , viewed as a function of  $\beta$ , is increasing since it is easy to check that

$$\frac{d\mathbb{E}_\beta[v(X)]}{d\beta} = \text{Var}_\beta(v(X)). \tag{27}$$

Therefore  $\mathbb{E}_\beta[v(X)]$  is a monotone function of  $\beta$  for which the value at  $\hat{\beta}$  is known and equal to  $v(\hat{x})$ . Considering the model in Section 3.2, assume that  $\hat{x}$  and  $\hat{p}$  are the current estimates of the true scene and the noise parameter. In order to compute a new restoration following ICM, we do not need to estimate  $\beta$  itself, but only the integer  $w$ . According to the insensitivity condition (23), this consists of finding the integer  $k$  that satisfies

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

If a likelihood criterion is adopted for the estimation of  $\beta$ , the corresponding value of  $w$  can be determined as follows. First, the function  $\mathbb{E}_\beta[v(X)]$  must be evaluated at  $\beta = \log \left( \frac{(1 - \hat{p})(C - 1)}{\hat{p}} \right)$  and compared with  $\mathbb{E}_{\hat{\beta}}[v(X)]$  which is equal to  $v(\hat{x})$  and can be easily extracted from  $\hat{x}$ . If  $\mathbb{E}_\beta[v(X)]$  at  $\beta = \log \left( \frac{(1 - \hat{p})(C - 1)}{\hat{p}} \right)$  is lower than  $v(\hat{x})$  then, since  $\mathbb{E}_\beta[v(X)]$  is an increasing function of  $\beta$ ,

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

In this case,  $w = 1$ . Otherwise, we need to evaluate  $\mathbb{E}_\beta[v(X)]$  at  $\beta = \frac{1}{2} \log \left( \frac{(1 - \hat{p})(C - 1)}{\hat{p}} \right)$  and compare it with  $v(\hat{x})$ , and so on.

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  $\beta$  and forming empirical averages. The required samples can be generated using Markov chain Monte Carlo. This can be computationally demanding, but can be greatly accelerated for the Potts model using the Swendsen-Wang algorithm (Swendsen and Wang 1987). It considers clusters instead of pixels and converges faster than single site updating algorithms such as the Gibbs sampler. In addition, the Swendsen-Wang algorithm explores the state space more freely so that fewer iterations are required to obtain accurate estimates.

Even when few samples and evaluations are required, online sampling takes time and may slow down the restoration procedure considerably. Approaches that may be faster have been considered by Geyer and Thompson (1992) and Descombes, Morris, and Zerubia (1996). A simple sample reuse scheme that can be implemented adaptively is considered by Descombes, Morris, and Zerubia (1996). They also suggest using a data base where empirical averages are stored for a certain number of values of the parameters; this would allow one to suppress online sampling completely. Geyer (1991) uses samples for a set of values of  $\beta$  to construct a mixture distribution which has appreciable probability in all regions of interest. For small size images (such as  $32 \times 32$ ), it is possible to consider a small set of values of  $\beta$ . The main problem seems to



be the extension of the method to real-size images (at least  $256 \times 256$ ), for which the required number of values of  $\beta$  becomes prohibitively large.

#### 4.4 Maximum Pseudo-Likelihood Criterion

Maximum likelihood is usually intractable in the standard Markov random field cases when point estimation of the parameters is required; maximum pseudo-likelihood is a natural alternative. The term ‘‘pseudo-likelihood’’ was introduced by Besag (1975) to refer to the product of conditional probabilities

$$\prod_{i \in S} P(x_i | x_{S \setminus \{i\}}). \quad (28)$$

For the Potts model,

$$p_i(x_i) = Z_i(\beta)^{-1} \exp(\beta u_i(x_i)),$$

where  $p_i(x_i) = P(x_i | x_{S \setminus \{i\}})$  and

$$Z_i(\beta) = \sum_{c \in \{1, \dots, C\}} \exp(\beta u_i(c)).$$

The pseudo-likelihood function (28) is then maximized by maximizing the log-pseudo-likelihood

$$F(\beta) = \sum_{i \in S} (\beta u_i(x_i) - \log(Z_i(\beta))).$$

The first and second derivatives can be easily calculated and expressed in a form similar to that in the maximum likelihood case,

$$\begin{aligned} \frac{dF(\beta)}{d\beta} &= \sum_{i \in S} (u_i(x_i) - \mathbb{E}_{p_i}[u_i(X_i)]) \\ &= \sum_{i \in S} u_i(x_i) - \sum_{i \in S} \mathbb{E}_{p_i}[u_i(X_i)], \end{aligned} \quad (29)$$

where

$$\mathbb{E}_{p_i}[u_i(X_i)] = \sum_{c \in \{1, \dots, C\}} u_i(c) p_i(c).$$

Similar equations to those for maximum likelihood estimation ((26) and (27)) hold. The first sum in (29) does not depend on  $\beta$  and the second sum depends on  $\beta$  through the  $p_i$ 's. The maximum pseudo-likelihood estimator of  $\beta$  satisfies

$$\sum_{i \in S} \mathbb{E}_{p_i}[u_i(X_i)] = \sum_{i \in S} u_i(x_i).$$

Given a current restoration  $\hat{x}$ , the quantity  $\sum_{i \in S} \mathbb{E}_{p_i}[u_i(X_i)]$  can be computed exactly for all given values of  $\beta$  and is an increasing function of  $\beta$  since

$$\frac{d \sum_{i \in S} \mathbb{E}_{p_i}[u_i(X_i)]}{d\beta} = \sum_{i \in S} \text{Var}_{p_i}[u_i(X_i)],$$

where  $\text{Var}_{p_i}[u_i(X_i)]$  is the variance of  $u_i(X_i)$  under  $p_i$ . The quantity  $\sum_{i \in S} u_i(\hat{x}_i)$  is also easily calculated from  $\hat{x}$ . An advantage over maximum likelihood estimation is therefore 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 different steps of the Bayesian morphology algorithm are specified below for the model of Section 3.2. We show application of this algorithm to two  $200 \times 200$  binary images and to a  $128 \times 128$  four-color image degraded by channel transmission noise.

The unsupervised restoration algorithm is as follows. Given a current estimate  $\hat{x}$  of the image we want to restore, the noise parameter  $p$  is first estimated from  $y$  and  $\hat{x}$  using a likelihood criterion. The maximum likelihood estimate of  $p$  is given by

$$\hat{p} = \operatorname{argmax}_p \mathbb{P}(y|\hat{x}) = 1 - \frac{\sum_{i \in S} \delta(\hat{x}_i, y_i)}{|S|},$$

where  $|S|$  is the number of pixels.

In standard unsupervised procedures, the following step consists of estimating the model parameter  $\beta$  using  $y$ ,  $\hat{x}$  and  $\hat{p}$ . In our algorithm,  $\beta$  is not directly estimated. Given an estimation criterion (see Section 4.2), we look for the integer  $k$  that satisfies

$$f\left(\frac{1}{k} \log\left(\frac{(1-p)(C-1)}{p}\right)\right) \leq b < f\left(\frac{1}{k-1} \log\left(\frac{(1-p)(C-1)}{p}\right)\right), \quad (30)$$

where  $f$  and  $b$  depend on the estimation criterion chosen. The constant  $b$  is equal to  $v(\hat{x})$  for maximum likelihood and to  $\sum_{i \in S} u_i(\hat{x}_i)$  ( $= 2v(\hat{x})$ ) for maximum pseudo-likelihood.

A new restoration is then computed from the current one,  $\hat{x}$ , using  $y$  and  $\hat{w} = k$  according to the update rule (21), and the whole procedure is iterated until convergence.

The true binary images that we used are shown in Figure 2. Figures 3 and 4 show these images degraded with 15% channel noise, and different restorations: Bayesian morphology using the likelihood and pseudo-likelihood criteria, and blind restoration using the simple majority vote (18), as in Zhang, Shirazi, and Noda (1996). Figure 5 presents restorations when the images are degraded with 30% channel noise.

We adopted a two-color Potts model and an eight-pixel neighborhood (the nearest eight neighbors). The choice of the neighborhood may substantially affect the restoration, especially when the noise increases. For images such as those in Figure 2, an eight-pixel neighborhood provides better restorations than a four-pixel neighborhood (the diagonally adjacent pixels are ignored). The problem of choosing the most appropriate neighborhood for the image at hand requires more consideration. In particular, it is related to the problem of choosing a structuring element in mathematical morphology.

The performance of our Bayesian morphology algorithm is the same as that of an unsupervised version of ICM. As a numerical indicator, we simply use the error rate or the proportion of misclassified pixels. Note that this may not always coincide with visual criteria. In Figure 5, for instance, restoration (b) may be preferred to (c) since it keeps more details although it does not remove enough noise and the error rate is higher (similarly (e) may be preferred to (f)). More sophisticated indicators that take into account visual features are given in Baddeley (1992a) and Baddeley (1992b). In Tables 1, 2 and 3, 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 (30). 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 (if a data base is not available) 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 and 0.45 for images degraded with 15% and 30% channel noise, respectively. Starting with a large initial estimate of the noise parameter ( $\hat{p}_0 = 0.45$ ) corresponds to choosing 1 as an initial value of  $w$ , i.e. an initial restoration  $\hat{x}_0$  which is the image of  $y$  after applying a median operator, or one iteration of a simple majority rule smoother.

Note that for a two-color Potts model, the term  $(u_i(c) - u_i(y_i))$  in the update rule (21) 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 unnecessary, the distinction is interesting when comparing maximum likelihood and maximum pseudo-likelihood estimators of  $\beta$  since a smaller value of  $w$  corresponds to a larger value of  $\beta$ . In Tables 1 and 2, 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 images (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 higher estimate of  $w$  is consistent with the empirical observation by Geyer and Thompson (1992) that maximum pseudo-likelihood estimators tend to overestimate  $\beta$ . It is not clear whether maximum likelihood estimation should be preferred, since more accurate estimation of  $\beta$  may not always result in better restorations, especially if the model used for the true scene does not capture its global characteristics well.

For the binary images of Figure 2 with 15% channel noise, it is clear from Figures 3 and 4 that blind restoration performs badly. Too much detail is lost and the error rate is higher. For 30% channel noise (Figure 5), blind restoration and pseudo-likelihood restoration ((c) and (f)) are equivalent and lead to a smaller error rate than maximum likelihood restoration. However, maximum likelihood restorations ((b) and (e)) seem to keep more details and may still be preferred.

The same conclusions do not necessarily hold for images consisting of more homogeneous regions such as that of Figure 6. For 15% channel noise (Figure 7), pseudo-likelihood estimation seems to be a good

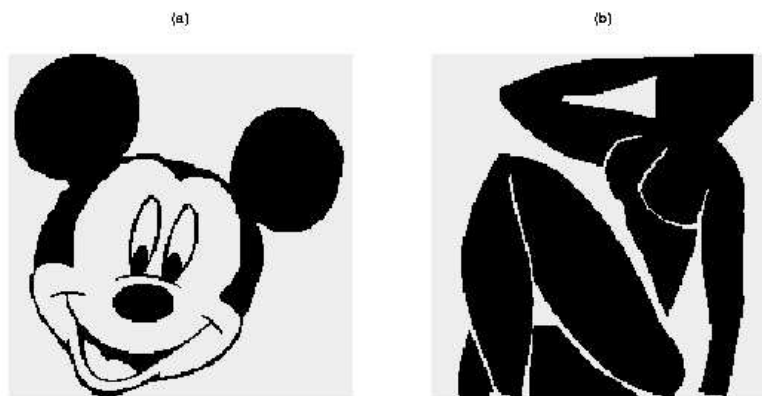


Figure 2: Two  $200 \times 200$  True Binary Images: (a) Mickey; (b) Matisse.

compromise between likelihood estimation which does not remove enough noise near the borders between regions and blind restoration which smooths the edges too much. When the noise increases (Figure 8), blind restoration (d) is visually better mainly because the details kept in the other restorations are almost exclusively noise. The results are summarized in Tables 4 and 5.

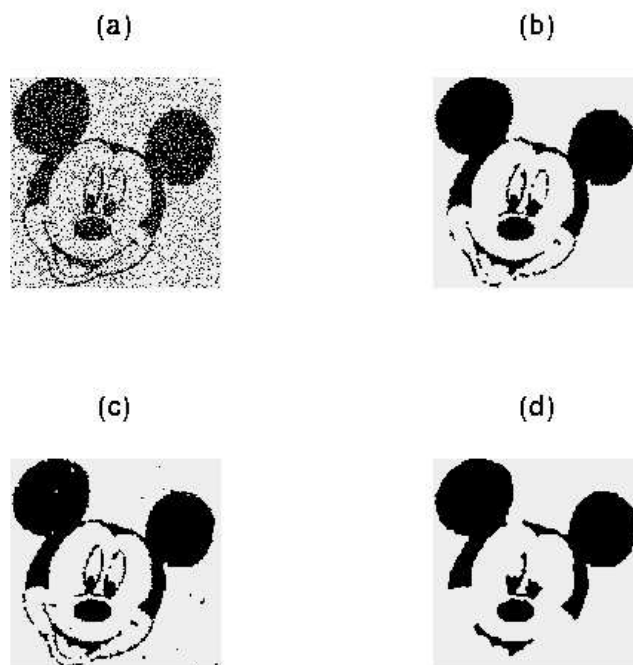


Figure 3: Degraded Mickey Image and Different Unsupervised Restorations: (a) Original image degraded with 15% channel noise. (b) Bayesian morphology based on a likelihood criterion. (c) Bayesian morphology based on a pseudo-likelihood criterion. In (b) and (c), the initial estimate of the noise parameter  $\hat{p}_0$  was 0.35. (d) Blind restoration using simple majority rule.

Table 1: Results for Figure 3 (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%

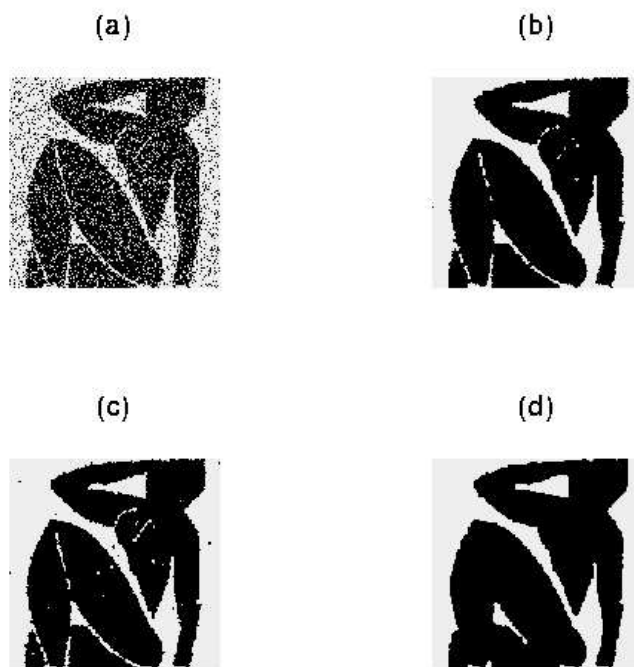


Figure 4: Degraded Matisse Image and Different Unsupervised Restorations: (a) Original image degraded with 15% channel noise. (b) Bayesian morphology based on a likelihood criterion. (c) Bayesian morphology based on a pseudo-likelihood criterion. In (b) and (c), the initial estimate of the noise parameter  $\hat{p}_0$  was 0.35. (d) Blind restoration using simple majority rule.

Table 2: Results for Figure 3 (Matisse image with 15% channel noise).

	image	iterations	calls to f	final k	final p	error rate
(a)	15% channel noise					15.2%
(b)	Likelihood criterion ( $\hat{p}_0 = 0.35$ )	5	22	4	14.5%	1.7%
(c)	Pseudo-likelihood criterion ( $\hat{p}_0 = 0.35$ )	4	15	3	14.1%	1.7%
(d)	Blind restoration ( $w = 1$ )	19		1	15.9%	2.1%

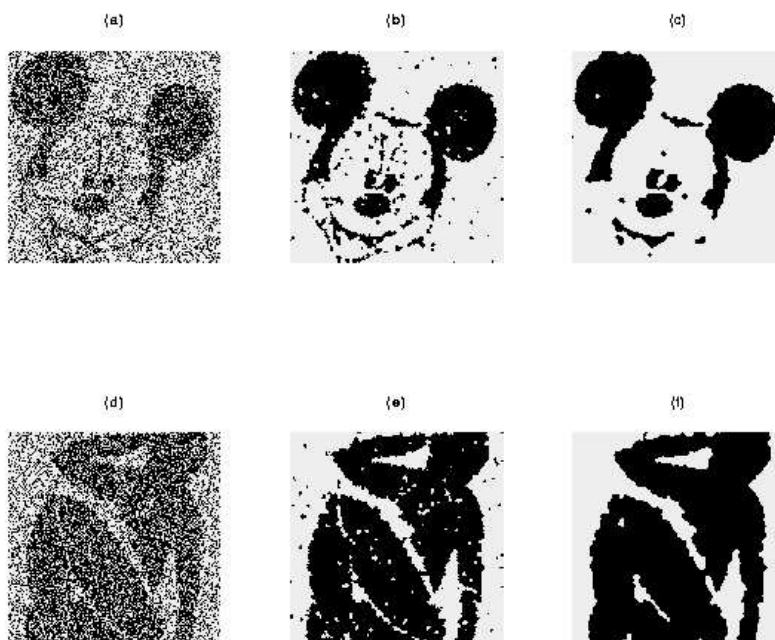


Figure 5: The Mickey and Matisse Images Degraded with 30% Channel Noise ((a) and (d)) and Different Unsupervised Restorations. (b) and (e) are Bayesian morphology restorations based on a likelihood criterion and restorations (c) and (f) are Bayesian morphology restorations based on a pseudo-likelihood criterion. The initial value of the noise parameter was set to 0.45, which leads to an initial value of  $\hat{w}$  equal to 1. Restorations (c) and (f) are the same as those obtained with blind restoration using a simple majority rule.

Table 3: Results for Binary Images Degraded With 30% Channel Noise.

image	iterations	calls to f	final k	final p	error rate
(a) 30% channel noise					29.6%
(b) Likelihood criterion ( $\hat{p}_0 = 0.45$ )	5	16	3	26.2%	6.2%
(c) Pseudo-likelihood criterion ( $\hat{p}_0 = 0.45$ ) and blind restoration ( $w = 1$ )	19	28	1	29.9%	4.8%
(d) 30% channel noise					30.5%
(e) Likelihood criterion ( $\hat{p}_0 = 0.45$ )	6	20	3	27.0%	6.3%
(f) Pseudo-likelihood criterion ( $\hat{p}_0 = 0.45$ ) and blind restoration ( $w = 1$ )	15	20	1	30.2%	4.2%

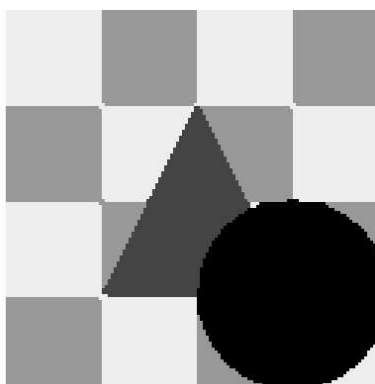


Figure 6: A  $128 \times 128$  Four-Color Test Image.

Table 4: Results for Figure 7 (four-color image with 15% channel noise).

image	iterations	calls to f	final k	final p	error rate
(a) 15% channel noise					15.0%
(b) Likelihood criterion ( $\hat{p}_0 = 0.35$ )	3	19	5	14.2%	0.8%
(c) Pseudo-likelihood criterion ( $\hat{p}_0 = 0.35$ )	4	17	3	14.7%	0.5%
(d) Blind restoration ( $w = 1$ )	4		1	15.2%	0.7%



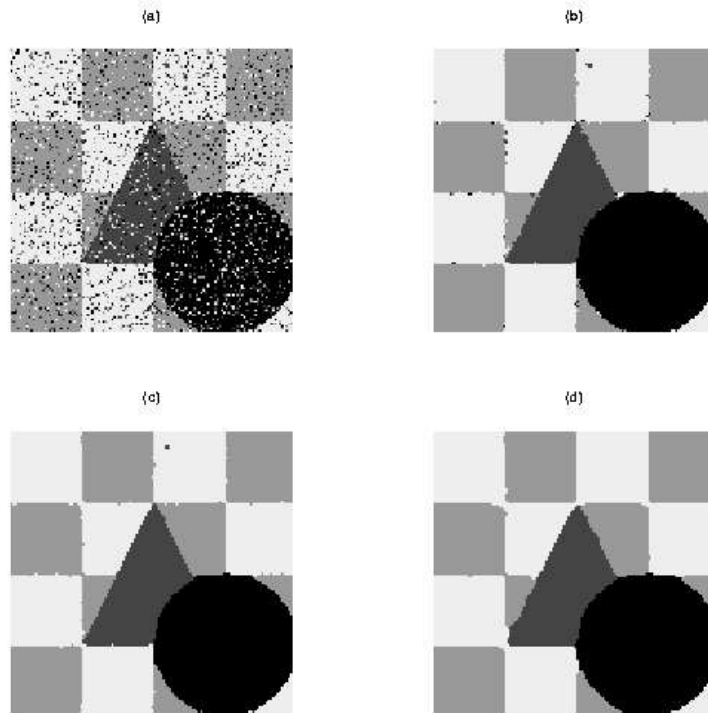


Figure 7: Four-Color Image Degraded with 15% Channel Noise and Different Unsupervised Restorations: (a) Original image degraded with 15% channel noise. (b) Bayesian morphology based on a likelihood criterion. (c) Bayesian morphology based on a pseudo-likelihood criterion. In (b) and (c), the initial estimate of the noise parameter  $\hat{p}_0$  was 0.35. (d) Blind restoration using simple majority rule.

Table 5: Results for Figure 8 (four-color image with 30% channel noise).

image	iterations	calls to f	final k	final p	error rate
(a) 30% channel noise					30.4%
(b) Likelihood criterion ( $\hat{p}_0 = 0.45$ )	5	24	4	28.9%	1.9%
(c) Pseudo-likelihood criterion ( $\hat{p}_0 = 0.45$ )	5	20	3	29.4%	1.5%
(d) Blind restoration ( $w = 1$ )	4		1	30.4%	1.2%

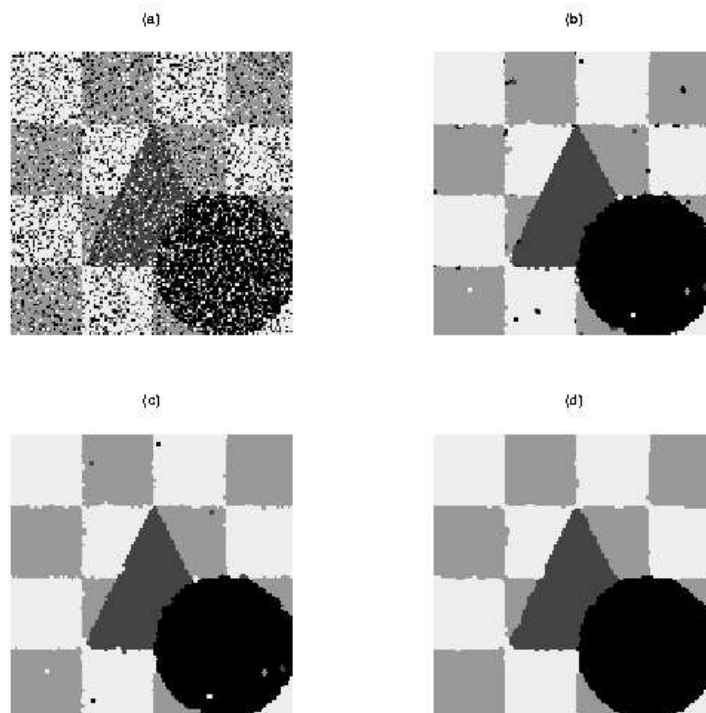


Figure 8: Four-Color Image Degraded with 30% Channel Noise and Different Unsupervised Restorations: (a) Original image degraded with 30% channel noise. (b) Bayesian morphology based on a likelihood criterion. (c) Bayesian morphology based on a pseudo-likelihood criterion. In (b) and (c), the initial estimate of the noise parameter  $\hat{p}_0$  was 0.35. (d) Blind restoration using simple majority rule.

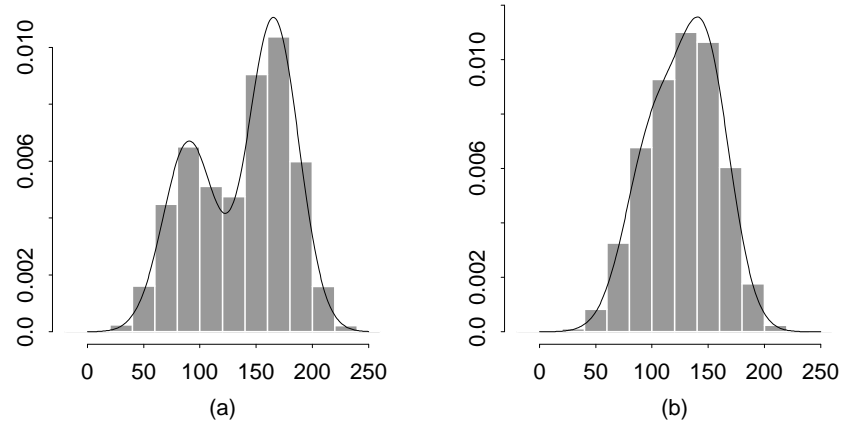


Figure 9: Pixel intensity histograms with overlaid marginal mixture EM density estimates. Matisse image degraded by superposition of Gaussian noise and converted to 256 grey levels. The noise standard deviation is (a) 0.3 ; (b) 0.5 .

## 5.2 Segmenting Grey-Scale Images

We created noisy grey-level images by adding some Gaussian noise to image (b) of Figure 2 and discretizing the result into 256 grey levels. The standard deviation was successively set to 0.3 and 0.5, leading to Figures 10 (a) and 11 (a). The marginal mixture model EM algorithm was then applied to obtain classifications with error rates of 4.5% and 15.25% in Figures 10(b) and 11(b), respectively. Figure 9 shows the respective pixel intensity histograms with the overlaid density estimates.

We used these classifications as initial classifications when applying standard ICM for grey-level images and computed the classifications (c). They were also considered as the original images to which Bayesian morphology was applied. This led to the classifications (d). In the latter case, the noise parameter  $p$  was initialized at 0.15 and 0.45, respectively. Restorations (d) have slightly higher error rates (Tables 6 and 7) but provide a visual impression similar to the restorations (c). They require far fewer evaluations of the estimation criterion  $f$ . Computation times are given when a pseudo-likelihood criterion is used. Note that the gain is more pronounced for larger images and when sampling is considered.

### 5.3 Example: PET Image of a Dog Lung

Figure 12(a) is a grey-level image of a dog lung obtained using the PET (Photon Emission Tomography) technique. One aim of the study is to measure the heterogeneity of the tissue in the region of interest. The classification task appears here as a pre-processing of the data. For image (a), the aim was therefore to isolate the relevant part (the lung) from the background and the artifacts produced by PET technique. Only pixels in this delimited area will then be considered to compute a heterogeneity criterion, 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 visually better, 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. 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. The results are summarized in Table 8 .

### 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](http://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  $144 \times 72$  maps representing stations or points (pixels) at which monthly average precipitation (in mm) has been recorded or extrapolated, for each individual calendar month. Figure 13 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. It has 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.

As regards the interpretation of such classifications, precipitation data alone cannot be expected to represent climate variability and to provide a complete picture of the climatic zones. Climatic regionalizations typically include additional temperature data. In Fovell (1997), temperature and precipitation are supposed to be both necessary and sufficient for the delineation of climatic zones. However, for various reasons, climatic regionalizations that proceed without directly mixing the two disparate data types are recommended. Two independent clustering solutions were formed from temperature and precipitation data and then merged together in some fashion. An important issue is the number of components in each classification. For the precipitation data, we computed six-component classifications following a suggestion in Fovell (1997). As those data are far from being normally distributed a non-linear transform was first applied: the power 0.25 of each record was taken. A fast first segmentation of the  $144 \times 72$  12-band image was then obtained using the technique described in Posse (1998). This classification is shown in Figure 14. Figure 15 shows this classification after applying Bayesian morphology with an initial value of  $\hat{w}$  set to 1. Figure 16 shows the segmentation obtained by applying ICM. The Bayesian morphology segmentation appears too fragmented in some regions while the ICM segmentation 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. The results are summarized in Table 9.

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. Figure 17 shows the time series computed by averaging the data for the members in each class of Figure 15. In this classification, class 1 is represented in white while class 6 is in black. Class 1 (*e.g.* Sahara, eastern South Pacific, eastern South Atlantic) and class 2 (*e.g.* Polar regions, American South-West, and Australian Outback) correspond to dry regions with different seasonalities. Class 6 corresponds to wet areas with light seasonal variations (*e.g.* southern ocean, equatorial Pacific, rain forests of South America). Class 3 includes regions with high variability and dry-summer climates (*e.g.* the Californian coast, the South-East Mediterranean region, Northern Australia) 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.

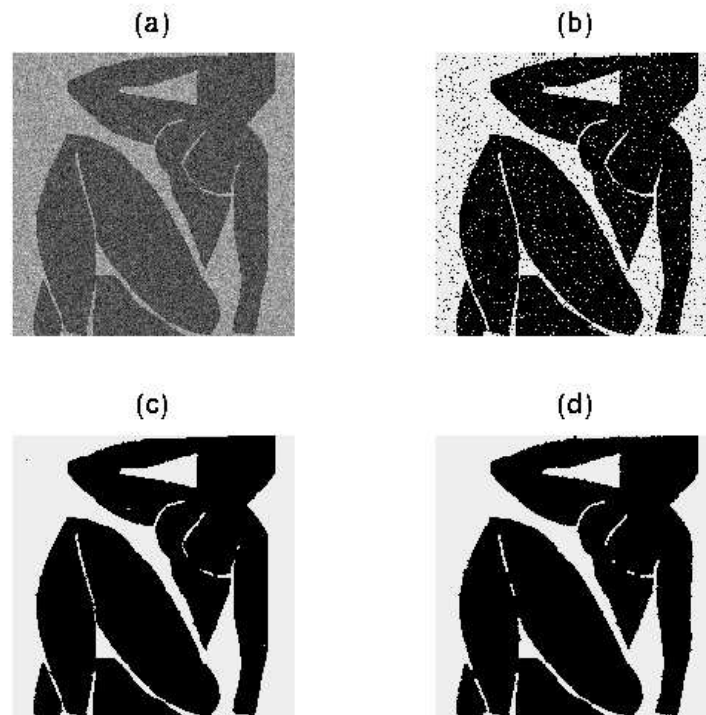


Figure 10: Degraded Grey-Level Image with Light Noise and Unsupervised Restorations. (a) Matisse image degraded by superposition of Gaussian noise with standard deviation  $\sigma = 0.3$ . (b) Initial segmentation using the marginal mixture model EM method. (c) Standard ICM reconstruction. (d) Bayesian morphology reconstruction. This was obtained by modeling the initially segmented image (b) as being degraded with channel noise.

Table 6: Results for Figure 10 (grey-level Matisse image with light noise).

	image	iterations	calls to f	final k	final p	final $\beta$	error	CPU time
(a)	Gaussian noise $\sigma = 0.3$							
(b)	Initial 2-component classification (EM)						4.5%	
(c)	ICM for grey-level images (Gaussian noise)	7	89			1.03	0.3%	12.26 s
(d)	Bayesian morphology ( $\hat{p}_0 = 0.15$ )	3	16	4	4.4%		0.6 %	2.96 s

Table 7: Results for Figure 11 (grey-level Matisse image with heavy noise).

	image	iterations	calls to f	final k	final p	final $\beta$	error	CPU time
(a)	Gaussian noise $\sigma = 0.5$							
(b)	Initial 2-component classification (EM)						15.25%	
(c)	ICM for grey-level images (Gaussian noise)	14	176			0.91	1.2%	23.43 s
(d)	Bayesian morphology ( $\hat{p}_0 = 0.45$ )	4	13	3	14.6%		1.7 %	2.98 s

Table 8: results for Figure 12 (dog lung PET image).

	image	iterations	calls to f	final k	final p	final $\beta$	CPU time
(a)	PET scan of a dog lung 128 $\times$ 128						
(b)	Initial 3-component classification (EM)	50					
(c)	ICM for grey-level images (Gaussian noise)	6	68			1.44	4.22 s
(d)	Bayesian morphology ( $\hat{p}_0 = 0.45$ )	3	14	4	2.9%		1.20 s

Table 9: results for Figures 14, 15, 16 (Legates/MSU precipitation 12-band image).

	image	iterations	calls to f	final k	final p	final $\beta$	CPU time
	Initial 6-component classification (mclust + MST)						
	ICM for grey-level 12-band images (Gaussian noise)	14	160			1.77	46.05 s
	Bayesian morphology (initial $\hat{w} = 1$ )	12	48	4	4.8%		3.53 s

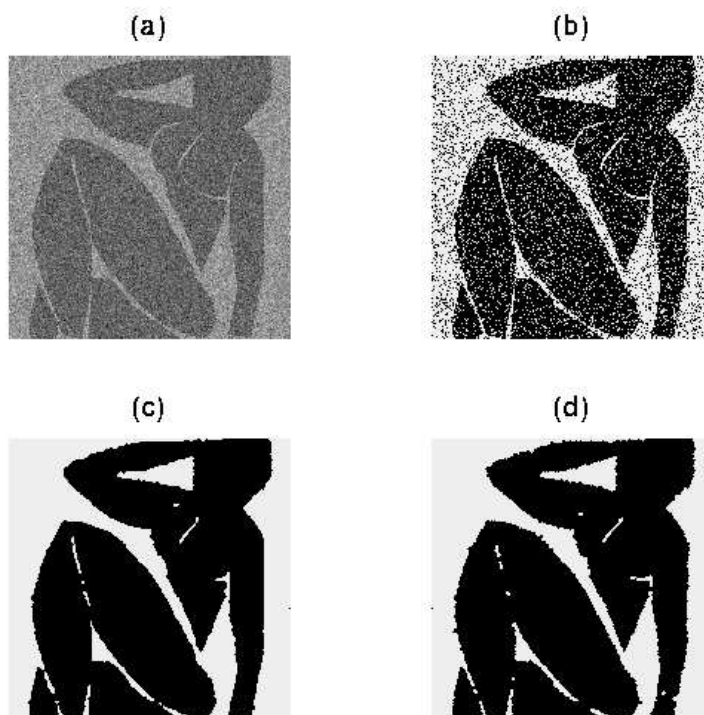


Figure 11: Degraded Grey-Level Image with Heavy Noise and Unsupervised Restorations. (a) Matisse image degraded by superposition of Gaussian noise with standard deviation  $\sigma = 0.5$ . (b) Initial segmentation using the marginal mixture model EM method. (c) Standard ICM reconstruction. (d) Bayesian morphology reconstruction.



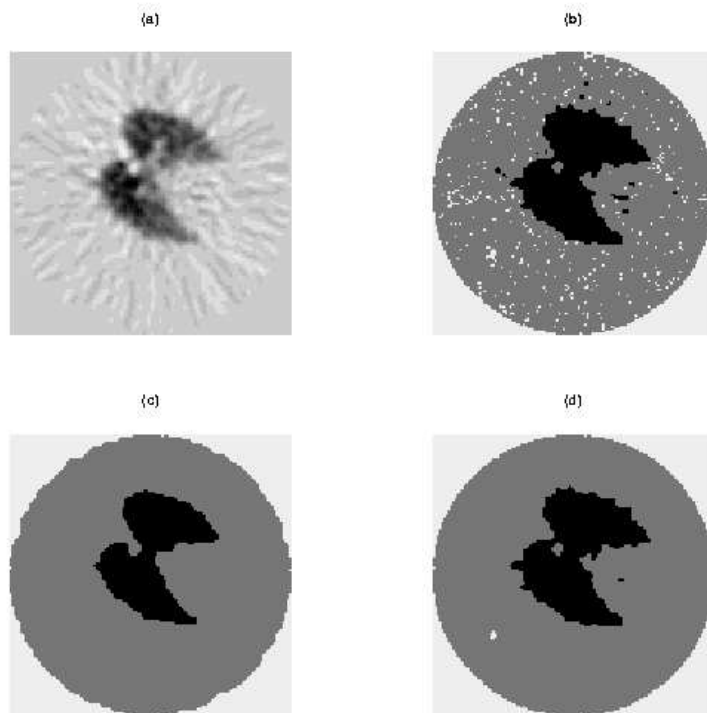


Figure 12: A Grey-Level PET Scan Image of a Dog Lung, and Different Unsupervised Segmentations: (a) Original image derived from a PET scan. (b) Initial three-component segmentation using the marginal mixture model EM method. (c) Standard ICM segmentation, starting from (b). (d) Bayesian morphology segmentation, modeling the initial segmentation (b) as being degraded by channel noise.

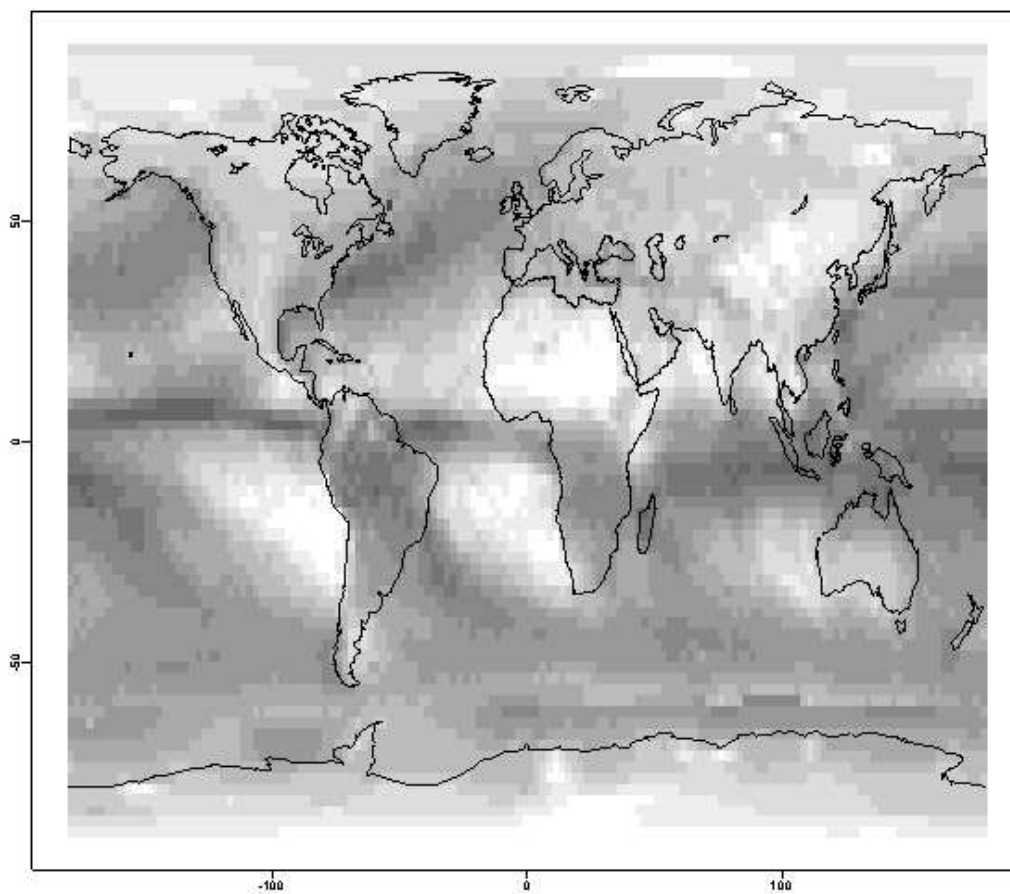


Figure 13: Monthly average precipitation (in mm) for the month of January. The spatial resolution is 2.5 degrees in each of latitude and longitude and results in a  $144 \times 72$  image.

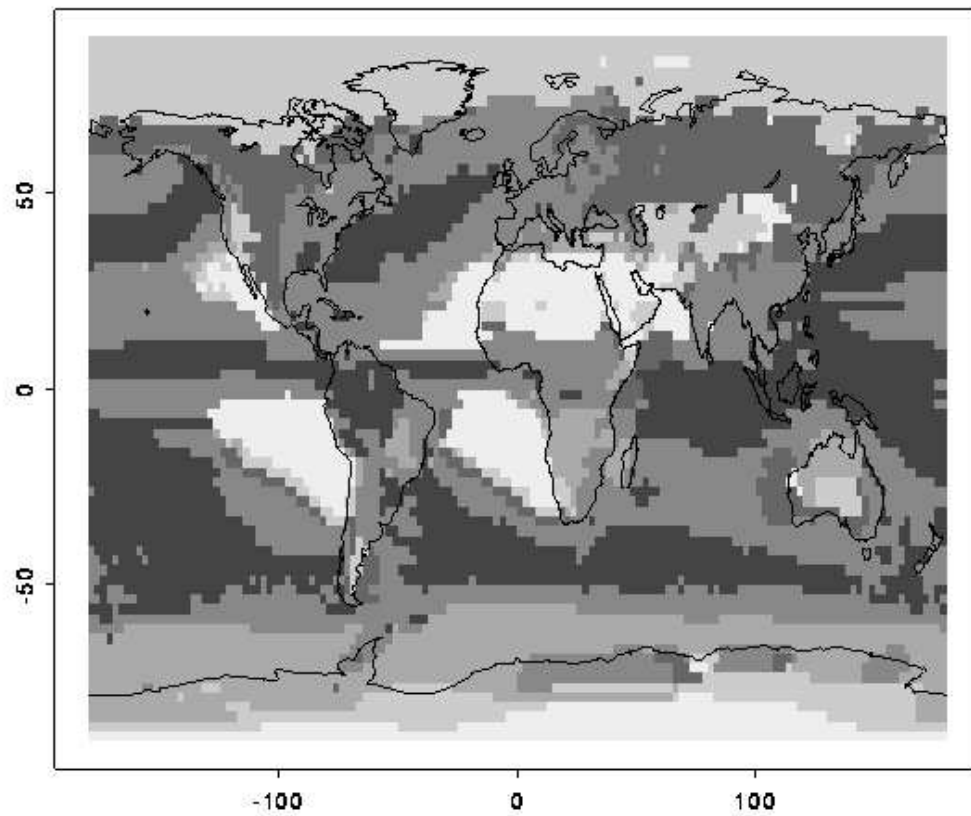


Figure 14: Marginal mixture EM segmentation of the Legates/MSU precipitation 12-band image: initial six-component segmentation using a combination of mclust and minimum spanning tree. This takes no account of spatial location and dependence.

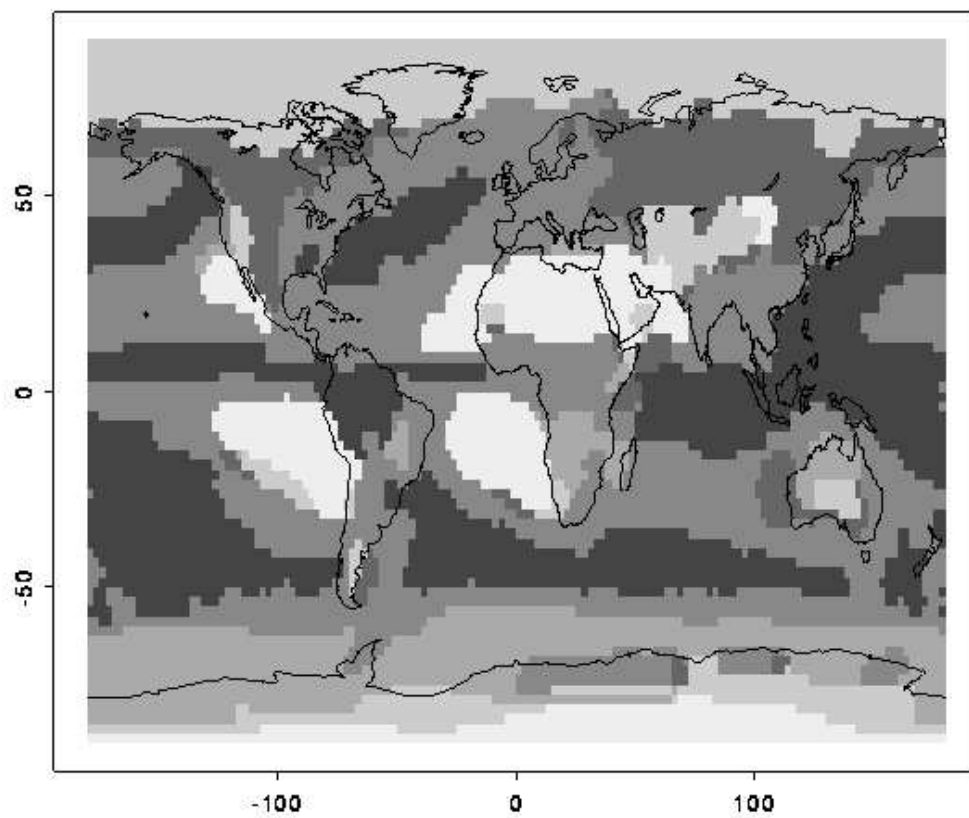


Figure 15: Bayesian morphology segmentation of the Legates/MSU precipitation 12-band image, modeling the initial segmentation, Figure 14 (a), as being degraded by channel noise. The initial value of  $\hat{w}$  was set to 1.

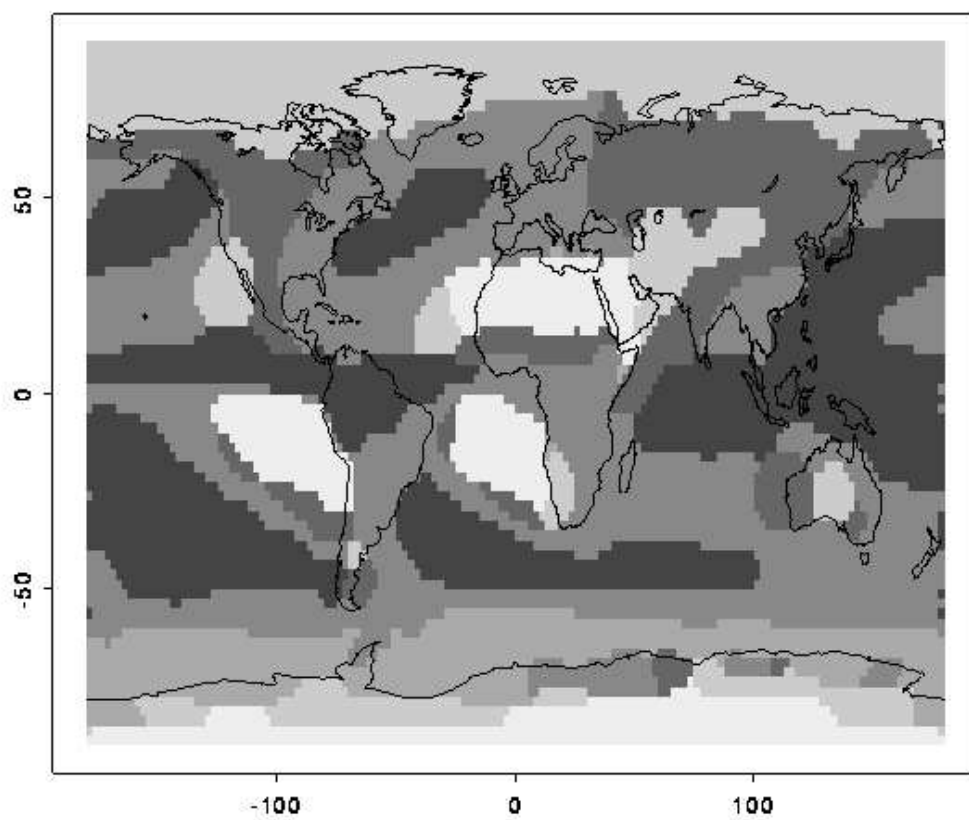


Figure 16: Standard ICM segmentation of the Legates/MSU precipitation 12-band image, starting from Figure 14 (a).

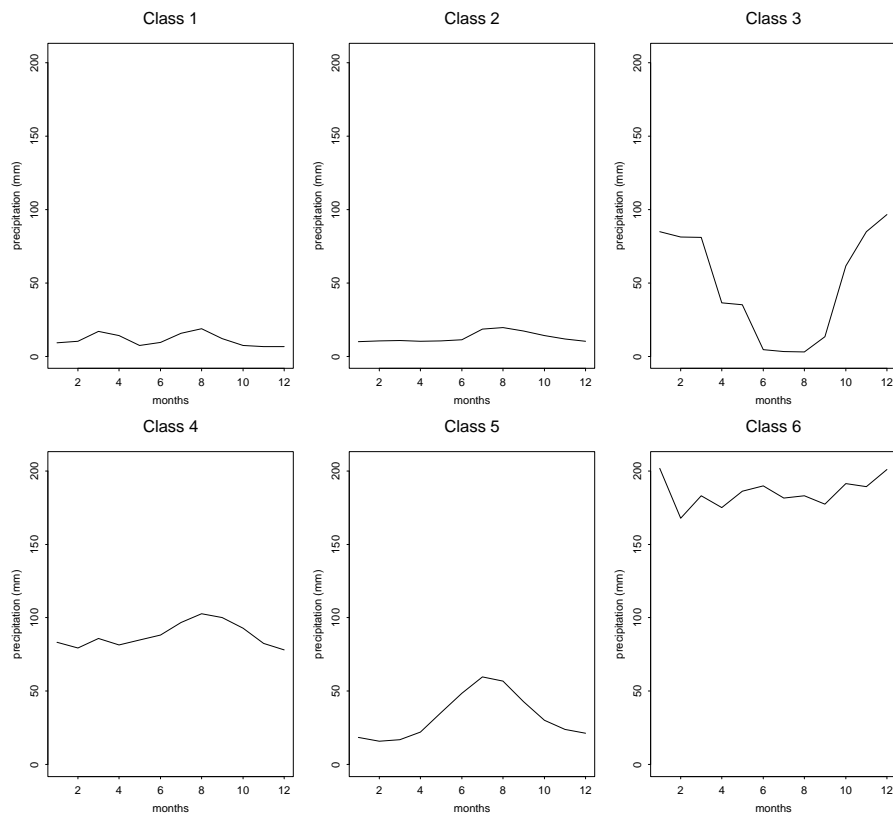


Figure 17: Time series computed by averaging the data for the members in each class of Figure 15.

## 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 discrete-valued ICM to the resulting segmentation. We have shown that (1) the latter 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 greater savings were realized 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 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, real-time video processing, and so on.

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 in 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  $5 \times 5$  neighborhood consisting of the 24 nearest pixels. Such a neighborhood structure induces cliques which are  $3 \times 3$  squares. For binary Markov random fields, a configuration on a  $3 \times 3$  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

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

where  $v_0, v_1, v_2$  depend on the  $3 \times 3$  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 exist for models with several parameters but it seems that there is no real gain in 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 Titterton (1992), the parameters are estimated by maximizing

$$\prod_{i \in S} P(y_i | \hat{x}_{S \setminus \{i\}}) .$$

Focusing on  $\beta$  leads to the following equations. If

$$F(\beta) = \log \left( \prod_{i \in S} P(y_i | \hat{x}_{N(i)}) \right) ,$$

then

$$\frac{dF(\beta)}{d\beta} = \sum_{i \in S} \mathbb{E}_{q_i} [u_i(X_i)] - \mathbb{E}_{p_i} [u_i(X_i)]$$

where

$$\begin{aligned} q_i(x_i) &= p(x_i | y_i, \hat{x}_{N(i)}) \\ &\propto p(y_i | x_i) p_i(x_i) , \end{aligned} \tag{31}$$

and

$$\frac{d^2 F(\beta)}{d\beta^2} = \sum_{i \in S} Var_{q_i} [u_i(X_i)] - Var_{p_i} [u_i(X_i)] .$$

We tried to use formulations similar to those in Section 4.4. The main difference is that both terms in (31) depend on  $\beta$ . However, the choice of an appropriate function  $f$  is not clear.



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

- Baddeley, A. J. (1992a). An error metric for binary images. In *W. Förstner and S. Ruwiedel, editors, Robust Computer Vision: Quality of Vision Algorithms*, pp. 59–78. Wichmann Verlag.
- Baddeley, A. J. (1992b). Errors in binary images and a  $L^p$  version of the Hausdorff metric. *Nieuw Archief voor Wiskunde 10*, 157–183.
- Banfield, J. and A. E. Raftery (1992). Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *Journal of the American Statistical Association 87*, 7–16.
- 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.

- Byers, S. D. and A. E. Raftery (1998). Nearest neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association* 93, to appear.
- 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.
- 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, To appear.
- Dempster, A. P., N. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, series B* 39, 1–38.
- 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.
- 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.
- 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.
- 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.
- Grenander, U., Y. Chow, and D. M. Keenan (1991). Hands: a pattern theoretic study of biological shapes. *Research Notes on Neural Computing* 2.
- Heijmans, H. J. (1994). *Morphological Image Operators*. Academic Press: Boston.
- Higdon, D. (1994). *Spatial Applications of Markov Chain Monte-Carlo for Bayesian Inference*. Ph. D. thesis, University of Washington.

- 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.
- McLachlan, G. J. and K. E. Basford (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.
- Morris, R., X. Descombes, and J. Zerubia (1996). An analysis of some models used in image segmentation. Technical Report 3016, INRIA, Sophia Antipolis, France.
- Posse, C. (1998). Hierarchical Model-based Clustering for Large Data Sets. *In preparation*.
- Qian, W. and M. Titterton (1992). Stochastic relaxations and EM algorithms for Markov random fields. *Journal of Statistical Computation and Simulation* 40, 55–69.
- Roeder, K. and L. A. Wasserman (1997). Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association* 92, 894–902.
- 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.
- Titterton, D. M., A. F. Smith, and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester, U.K.: John Wiley.
- 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.



---

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,  
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY  
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex  
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN  
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex  
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

---

Éditeur  
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399