EM-based image segmentation using Potts models with external field

Florence Forbes [*] Gilles Celeux Nathalie Peyrard
Inria Rhône-Alpes

April 23, 2002

[*]Correspondence should be sent to Florence Forbes: Fax: (00 33) 4 76 61 54 77, Florence.Forbes@inrialpes.fr

Abstract: Image segmentation using Markov random fields involves parameter estimation in hidden Markov models for which the EM algorithm is widely used. In practice, a simple Markov model is often used to account for the spatial dependencies between pixels, namely the isotropic homogeneous Potts model with no external field. It has the advantage to involve only one interaction parameter and leads in a lot of cases to good results. The absence of an external field parameter implies that all colors have the same weight. In this paper, we investigate the use of additional parameters (external field) that would play the role of the weight terms in mixture distributions and would allow more flexibility for the segmentations. To deal with the difficulties that arise due to the dependence structure in the models, we use a class of EM-like algorithms based on the Mean Field approximation principle and presented in some previous work. We illustrate with numerical experiments the advantages of introducing an external field parameter in the basic Potts model.

Key Words: Image segmentation, EM algorithm, Mean field approximation, Potts model, External field estimation.

1 Introduction

Problems involving incomplete data, where part of the data is missing or unobservable, are common in image analysis. The aim may be to recover an original image which is hidden and has to be estimated from a noisy or blurred version. More generally, the observations may represent measurements, *e.g.* multidimensional variables recorded for each pixel of an image while the hidden data could consist of an unknown class assignment to be estimated from the observations for each pixel. This case is usually referred to as image segmentation. A simple model that allows classification of the pixels is the independent mixture model which assumes that the observations are independent realizations of a mixture distribution. For images, this independence assumption has some limitations and, in this paper, we focus on Markov model-based image segmentation. This involves hidden Markov random fields and have the advantage to take into account spatial information. Let S denote the set of pixels with a neighborhood system defined on it and $N(i)$ be the set of pixels (usually the closest 4 or 8) considered as neighbors of pixel i. A commonly used model for the hidden data is the K-color Potts model: each pixel $i \in S$ takes one of K states, which can represent K different class assignments. Each of them is represented by a binary vector z_i of length K with one component being 1, all others being 0. In its simplest version (homogeneous, isotropic Potts model with no external field) the model involves a single parameter β that accounts for the strength of the inter-pixel interaction and is usually referred to as the interaction parameter,

$$P_G(\mathbf{z} \mid \beta) \;\; = \;\; W(\beta)^{-1} \exp(\sum_{i \in S} \frac{1}{2}\beta \sum_{j \in N(i)} z_i^t z_j) \; , \tag{1}$$

where $\mathbf{z} = (z_i, i \in S)$. In practice, parameter β often does a reasonable job in accounting for spatial dependencies between pixels and leads to satisfying results in a lot of cases ([3], [9], [4], [11], [13], [8], [2], [5]). However, unlike independent mixture models which allow different class or color proportions, this basic version models the different colors in the image as having the same probability. To go further this constraint, a natural generalization is to use the Potts model with external field which involves an additional K-dimensional parameter $\mathbf{V} = (v_1, \ldots, v_K)^t$,

$$P_G(\mathbf{z} \mid \beta, \mathbf{V}) \;\; = \;\; W(\beta, \mathbf{V})^{-1} \exp(\sum_{i \in S} z_i^t \mathbf{V} + \frac{1}{2}\beta \sum_{j \in N(i)} z_i^t z_j) \; . \tag{2}$$

$\mathbf{V}$ acts as weights for the different values of z_i and is often referred to as an external field parameter ([6]). This results in more flexibility but induces additional parameter estimation problems so that this model is much less used in image segmentation problems than the simpler isotropic potts model. Our aim is to investigate whether it leads to a significant gain and a better adequacy especially when the color proportions are very unbalanced in the images to be recovered.

As regards parameter estimation, the EM algorithm is a commonly used algorithm in hidden data problems. In particular, it has been widely used as a classification tool using independent mixture models. For such models, the independence assumption leads to an easy implementation of the algorithm (*cf.* [10]). For Hidden Markov random fields, due to the dependence structure, approximations are required to make the EM algorithm tractable. In this paper, we use approximations based on the mean field principle ([13]). This allows to take the Markovian structure into account while preserving the good features of EM ([5]).

In Section 2, we specify the complete parametric models for the observed and unobserved data. In Section 3, we describe the parameter estimation algorithm we considered for these models and more specifically point out the additional work required to estimate the extra external field parameters. Experiments comparing both models, with and without external field, are reported in Section 4 and a discussion section ends the paper.

2 Hidden Markov random fields

A typical example of pixels set S in image analysis is the two dimensional lattice with a second order neighborhood system. For each site, the neighbors are the eight sites surrounding it. Image segmentation involves observed data and unobserved data to be recovered. In this paper, the unobserved data is modeled as a discrete Markov random field, $\mathbf{Z} = \{Z_i,\ i \in S\}$, defined on S, each Z_i taking values in a finite set $\{e_1, \ldots, e_K\}$ where e_k is the K-dimensional vector with kth composant set to 1 and all others to 0. A commonly used distribution for $\mathbf{Z}$ is the K-color Potts model defined in (2). When $\mathbf{V}$ is zero, no color is favored, *i.e.* at a given site i, if no information on the neighboring sites is available, then all colors appear with the same probability at site i. When β is zero, there is no interaction between pixels and the variables are independent. When β is zero, it comes from (2), for all i in S and all

k in $\{1, \ldots, K\}$,

$$P_G(Z_i = e_k) \;=\; \frac{\exp(v_k)}{\sum\limits_{l=1}^{K} \exp(v_l)},$$

from which we can see that $\mathbf{V} = (v_1, \ldots, v_K)$ acts as weights for the different possible values of z_i. Note that $\mathbf{V}$ is defined up to a constant c, *i.e.* a value $\mathbf{V}'$ defined as $\mathbf{V}' = (v_1 + c, \ldots, v_K + c)^t$ would lead to the same Potts model since $P_G(\mathbf{z} \mid \mathbf{V}') = P_G(\mathbf{z} \mid \mathbf{V})$. In particular, we can choose $\mathbf{V}$ such that $\sum\limits_{k=1}^{K} \exp(v_k) = 1$.

In hidden Markov models, the observations $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$, according to a density f which is assumed to be of the following type (θ is a parameter and the f_i's are given),

$$\begin{aligned} f(\mathbf{y} \mid \mathbf{z},\ \theta) \;&=\; \prod_{i \in S} f_i(y_i \mid z_i,\ \theta) \\ &=\; \exp\{\sum_{i \in S} \log f_i(y_i \mid z_i,\ \theta)\}, \end{aligned} \tag{3}$$

assuming that all the $f_i(y_i \mid z_i,\ \theta)$ are positive. This makes the model similar to an independent mixture model (*cf.* [10]). An independent mixture model could be seen as a hidden Markov model where the hidden field $\mathbf{Z}$ is one of independent identically distributed variables, which is the case for a Potts model with $\beta = 0$. For the f_i's, Gaussian distributions are considered in most applications. If z_i is in class k, f_i is the Gaussian distribution with parameters μ_k and σ_k, μ_k and σ_k being scalar mean and standard deviation, in the univariate situation, and vector means and covariance matrices in the multivariate case. The vector parameter to be estimated is then $(\beta, \mathbf{V}, \theta)$ with $\theta = \{(\mu_k, \sigma_k), k = 1, \ldots, K\}$.

In image segmentation problems, the question of interest is generally to recover the unknown image $\mathbf{z}$, interpreted as a classification into a finite number K of labels. This classification usually requires values for the vector parameter $\Psi = (\theta, \beta, \mathbf{V})$. If unknown, an estimation of Ψ can be obtained via the EM algorithm that we describe below.

3 Parameter estimation using an EM-based procedure

Assuming Ψ unknown, our aim is to get the maximun likelihood estimate of this parameter knowing the observations $\mathbf{y}$. The log-likelihood of the model is

$$L(\Psi) \;=\; \log P_G(\mathbf{y} \mid \Psi) \;=\; \log \sum_{\mathbf{z}} P_G(\mathbf{y}, \mathbf{z} \mid \Psi).$$

The EM algorithm [7] is an iterative algorithm aiming at maximizing this log-likelihood by maximizing at iteration q,

$$Q(\Psi \mid \Psi^{(q)}) \;=\; \mathbb{E}_{\Psi^{(q)}}[\log P_G(\mathbf{y}, \mathbf{Z} \mid \Psi) \mid \mathbf{Y} = \mathbf{y}] \;,$$

the expectation of the complete log-likelihood knowing the observation $\mathbf{y}$ and a current estimate $\Psi^{(q)}$. Using Bayes' rule, Q can be further written as follows

$$Q(\Psi \mid \Psi^{(q)}) \;=\; \mathbb{E}_{\Psi^{(q)}}[\log P_G(\mathbf{y} \mid \mathbf{Z}, \theta) \mid \mathbf{Y} = \mathbf{y}] + \mathbb{E}_{\Psi^{(q)}}[\log P_G(\mathbf{Z} \mid \beta, \mathbf{V}) \mid \mathbf{Y} = \mathbf{y}].$$

The first term does not depend on β and $\mathbf{V}$ while the last one does not involve θ. Therefore we will write,

$$\begin{aligned} Q(\theta \mid \Psi^{(q)}) &= \mathbb{E}_{\Psi^{(q)}}[\log P_G(\mathbf{y} \mid \mathbf{Z}, \theta) \mid \mathbf{Y} = \mathbf{y}] \\ &= \sum_{i \in S} \sum_{z_i} P_G(z_i \mid \mathbf{y}, \Psi^{(q)}) \;\log f_i(y_i \mid z_i, \theta) \end{aligned} \tag{4}$$

$$\begin{aligned} Q(\beta, \mathbf{V} \mid \Psi^{(q)}) &= \mathbb{E}_{\Psi^{(q)}}[\log P_G(\mathbf{Z} \mid \beta, \mathbf{V}) \mid \mathbf{Y} = \mathbf{y}] \\ &= -\log W(\beta, \mathbf{V}) \\ &\quad - \sum_{i \in S} \sum_{\mathbf{z}_{i \cup N(i)}} z_i^t(\mathbf{V} + \tfrac{1}{2}\beta \sum_{j \in N(i)} z_j)\; P_G(\mathbf{z}_{i \cup N(i)} \mid \mathbf{y}, \Psi^{(q)}), \end{aligned} \tag{5}$$

where for any subset A of S, $\mathbf{z}_A$ stands for $\{z_i, i \in A\}$. There are two difficulties in evaluating Q in this case. Both the partition function $W(\beta, \mathbf{V})$ and the conditional probabilities, $P_G(z_i \mid \mathbf{y}, \Psi^{(q)})$ and $P_G(\mathbf{z}_{i \cup N(i)} \mid \mathbf{y}, \Psi^{(q)})$, cannot be computed exactly. We used approximations based on the mean field principle ([6]). The mean field approach consists of neglecting fluctuations from the mean in the environment of each pixel. More generally, we talk about mean field-like approximations when the value at site i does not depend on the values at other sites which are all set to constants (not necessarily the means) independently of the value at site i.

The general form of the algorithms we propose consists of repeating the following steps,

(1) Create, from the observations $\mathbf{y}$ and some current parameter estimates $\Psi^{(q-1)}$, a configuration $\tilde{\mathbf{z}}^{(q)}$. For each site i, set the neighbors to $\tilde{\mathbf{z}}^{(q)}_{N(i)}$ and replace the marginal distribution $P_G(\mathbf{z} \mid \beta, \mathbf{V})$ by

$$P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{z} \mid \beta, \mathbf{V}) \quad = \quad \prod_{i \in S} P_G(z_i \mid \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta, \mathbf{V}). \tag{6}$$

(2) Apply the EM algorithm for the model defined by (3) and (6), with starting values $\theta^{(q-1)}$, $\beta^{(q-1)}$ and $\mathbf{V}^{(q-1)}$, to get updated estimates $\theta^{(q)}$, $\beta^{(q)}$ and $\mathbf{V}^{(q)}$. The joint distribution $P_G(\mathbf{y}, \mathbf{z} \mid \Psi)$ is thus replaced by

$$P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{y}, \mathbf{z} \mid \Psi) \quad = \quad \prod_{i \in S} \{f_i(y_i \mid z_i, \theta) \; P_G(z_i \mid \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta)\}. \tag{7}$$

Because of step **(1)**, the two problems encountered when considering the EM algorithm with the exact joint distribution disappear. The computation of the normalizing constant in $P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{z} \mid \beta, \mathbf{V})$ becomes easy, and the E step reduces to the computation, in (4) and (5), of conditional probabilities corresponding to the approximation of the conditional distribution $P_G(\mathbf{z} \mid \mathbf{y}, \Psi^{(q)})$. This approximation derives naturally from the approximation (6) of $P_G(\mathbf{z} \mid \beta)$ as

$$P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{z} \mid \mathbf{y}, \Psi^{(q)}) \quad = \quad \frac{f(\mathbf{y} \mid \mathbf{z}, \theta^{(q)}) \; P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{z} \mid \beta^{(q)}, \mathbf{V}^{(q)})}{P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{y} \mid \Psi^{(q)})}, \tag{8}$$

which can be further simplified, using (3) and (6), into

$$P_{\tilde{\mathbf{z}}^{(q)}}(\mathbf{z} \mid \mathbf{y}, \Psi^{(q)}) \quad = \quad \prod_{i \in S} \{ \frac{f_i(y_i \mid z_i, \theta^{(q)}) \; P_G(z_i \mid \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta^{(q)}, \mathbf{V}^{(q)})}{\sum_{z_i} f_i(y_i \mid z_i, \theta^{(q)}) \; P_G(z_i \mid \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta^{(q)}, \mathbf{V}^{(q)})} \}$$

$$= \prod_{i\in S} P_G(z_i \mid y_i, \tilde{\mathbf{z}}^{(q)}_{N(i)}, \Psi^{(q)}) \tag{9}$$

$$= \prod_{i\in S} P_{\tilde{\mathbf{z}}^{(q)}}(z_i \mid y_i, \Psi^{(q)}).$$

It follows the corresponding approximations of $Q(\theta \mid \Psi^{(q)})$ and $Q(\beta, \mathbf{V} \mid \Psi^{(q)})$, defined in (4) and (5),

$$Q(\theta \mid \Psi^{(q)}) \approx \sum_{i\in S}\sum_{z_i} P_{\tilde{\mathbf{z}}^{(q)}}(z_i \mid y_i, \Psi^{(q)}) \ \log f_i(y_i \mid z_i, \theta), \tag{10}$$

and

$$Q(\beta, \mathbf{V} \mid \Psi^{(q)}) \approx \sum_{i\in S}\sum_{z_i} P_{\tilde{\mathbf{z}}^{(q)}}(z_i \mid y_i, \Psi^{(q)}) \ \log P_{\tilde{\mathbf{z}}^{(q)}}(z_i \mid \beta, \mathbf{V}). \tag{11}$$

The following M step in EM becomes tractable. As noted by [12], the likelihood takes the form of a likelihood from independent observations from finite mixture of the same component densities but the sets of mixing weights vary for each site i depending on the choice of $\tilde{\mathbf{z}}^{(q)}$.

Maximizing (10) is straightforward since in our case a closed-form expression is available for $\theta^{(q+1)}$. Maximizing (11) can be done numerically setting its gradient to zero which corresponds to find $(\beta, \mathbf{V})$ that solve the following equations:

$\forall k = 1 \ldots K$,

$$\sum_{i\in S} P_G(Z_i = e_k | \mathbf{Z}_{N(i)} = \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta, \mathbf{V}) = \sum_{i\in S} P_G(Z_i = e_k | \mathbf{Y} = y, \mathbf{Z}_{N(i)} = \tilde{\mathbf{z}}^{(q)}_{N(i)}, \Psi^{(q)}) \tag{12}$$

and

$$\sum_{i\in S} I\!E[Z_i^t \sum_{j\in N(i)} \tilde{z}_j^{(q)} | \mathbf{Z}_{N(i)} = \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta, \mathbf{V}] = \sum_{i\in S} I\!E[Z_i^t \sum_{j\in N(i)} \tilde{z}_j^{(q)} | \mathbf{Y} = y, \mathbf{Z}_{N(i)} = \tilde{\mathbf{z}}^{(q)}_{N(i)}, \Psi^{(q)}] \tag{13}$$

To be more specific, let $q^{(q)}_{1,k}(\beta, \mathbf{V})$ and $p^{(q)}_{1,k}(\Psi)$ denote

$$q^{(q)}_{1,k}(\beta, \mathbf{V}) = P_G(Z_i = e_k \mid \mathbf{Z}_{N(i)} = \tilde{\mathbf{z}}^{(q)}_{N(i)}, \beta, \mathbf{V}) = \frac{\exp(e_k^t \mathbf{V} + \beta e_k^t \sum_{j\in N(i)} \tilde{\mathbf{z}}_i^{(q)})}{\sum_{l=1}^{K} \exp(e_l^t \mathbf{V} + \beta e_l^t \sum_{j\in N(i)} \tilde{\mathbf{z}}_i^{(q)})}$$

and

$$p_{1,k}^{(q)}(\Psi) \quad = \quad f_i(y_i \mid e_k, \theta)\; q_{1,k}^{(q)}(\beta, \mathbf{V}).$$

Then the problem is to find β and $\mathbf{V}$ that satisfy

$$\forall k = 1 \ldots K, \quad \sum_{i \in S} q_{1,k}^{(q)}(\beta, \mathbf{V}) - p_{1,k}^{(q)}(\Psi^{(q)} \quad = \quad 0$$

and

$$\sum_{i \in S} \sum_{k=1}^{K} e_k^t \sum_{j \in N(i)} \tilde{z}_j^{(q)}(q_{1,k}^{(q)}(\beta, \mathbf{V}) - p_{1,k}^{(q)}(\Psi^{(q)}) \quad = \quad 0.$$

Note that the first equation is due to the introduction of parameter $\mathbf{V}$ in addition to β (equation (13)). There is no extra theoretical difficulty in adding equation (12) but in practice, it adds K additional equations to the one-dimensional problem (13).

4 Experiments

Because of the introduction of extra parameters and computational cost, the Potts model with external field has not received great attention. However, The additional flexibility and information in the complete Potts model (2), which takes into account the external field parameter $\mathbf{V}$ and does not assume a priori equal probabilities for the different colors, can be beneficial. We performed numerical experiments to compare the performance of the complete Potts model (2) and the simplified Potts model (1) as regards segmentation purpose. We first considered experiments on synthetic noise-corrupted images. The degraded images were obtained by adding some zero mean Gaussian noise with standard deviation σ to the K-color original images. For each image, the colors were all degraded using the same value of σ. We then used a model with second order neighborhood, (*i.e.* the eight closest neighbors for each pixel), which is in general more realistic. For comparison purpose, the algorithms were always initialized using the same initial classification. We used EM for independent mixtures to produce a reasonable first guess. As regards the original images, we considered cases with different color proportions. The idea is that non equal color proportions in a realization is likely to measure non equal color probabilities in the model. Figures 1 (a) and 2 (a) are synthetic images showing one color in small proportion, typically less than 10% while in Figures 3 (a) and 4 (a) the color proportions are all greater than 20%. In addition, in Figure 4 (a) proportions are almost equal (see Table 2). The restorations obtained using the simplest model with no external field (1) and the complete model (2) are shown in Figures 1 to 4 (c) and (d) respectively. The corresponding parameter estimations and error rates are given in Tables 1 and 2. In addition to global error rates, *i.e.* proportions of misclassified pixels, color dependent error rates are computed considering misclassified pixels for each color separately. We also reported the color proportions in the final restorations.

As one could expect, the gain in using the complete model is greater when color proportions are unbalanced and especially when one of them is small (Figures 1 and 2). We stopped the algorithms after 500 iterations but we noticed that greater numbers of iterations were clearly in favor of the complete model in these cases: the complete model tends to give better results while the model with no external field leads to degraded estimations where the color in smaller proportion tends to disappear. For a color in small proportion, it seems that the complete model allows better estimations of the parameters associated to this color resulting then in a better restoration. This is particularly striking in the 4-color triangle image (Figure 2 (c)) where one color tends to disapear probably due to a bad estimation of its mean and variance (Table 1). For the 3-color triangle image, the color proportions are close to equal and the two models lead to equivalent results.

We then compared our algorithms on a real image of a cell (Figure 5, [1]). Surprisingly, although one class of interest (the center of the cell) is in a much smaller proportion, the algorithms lead to equivalent results. The initial classification used is shown in Figure 5 (b) and is already satisfying in terms of color proportions (see Table 3). However, differences

Ji image				
	β	**V**	μ	σ
True values	-	-	1.00 2.00	0.70 0.70
no external field	0.53	0.00 0.00	0.82 2.00	0.63 0.69
complete model	0.56	0.20 0.00	1.02 2.01	0.71 0.69
4-color triangle image				
True values	-	-	1.00 2.00 3.00 4.00	0.50 0.50 0.50 0.50
no external field	1.78	0.00 0.00 0.00 0.00	1.00 1.45 2.21 4.01	0.49 0.00 0.64 0.50
complete model	2.02	0.19 0.26 0.11 -0.08	1.00 2.00 2.97 4.01	0.49 0.50 0.48 0.50
Kin image				
True values	-	-	1.00 2.00	0.90 0.90 0.90 0.90
no external field	0.56	0.00 0.00	0.99 2.00	0.89 0.89
complete model	0.59	0.00 0.00	1.00 1.99	0.89 0.89
3-color triangle image				
True values	-	-	0.00 1.00 2.00	0.50 0.50 0.50
no external field	1.54	0.00 0.00 0.00	0.00 1.00 2.00	0.50 0.50 0.50
complete model	1.68	-0.02 0.02 -0.04	0.00 1.00 2.00	0.50 0.50 0.50

Table 1: Degraded K-color images: parameter estimation using Potts models with or without external field.

Ji image			
	proportions	error rate	error rates by color
True values	5 % 95 %	-	-
no external field	3% 97%	3 %	50% 0%
complete model	5% 95%	3 %	34% 1%
4-color triangle image			
True values	37% 34% 9% 20%	-	-
no external field	37% 0% 43% 20%	34%	1% 100% 1% 0%
complete model	38% 33% 9% 20%	1%	0% 1% 0% 0%
Kin image			
True values	21% 79%	-	-
no external field	20% 80%	6%	17% 3%
complete model	20% 80%	5%	13% 3%
3-color triangle image			
True values	37% 34% 29%	-	-
no external field	38% 33% 29%	0.6%	0% 1% 0%
complete model	37% 34% 29%	0.6%	1% 1% 0%

Table 2: Degraded K-color images: color frequencies and error rates for restorations using Potts models with or without external field.

Cell image					
	β	**V**	μ	σ	proportions
no external field	1.28	0.00 0.00 0.00 0.00	-3.40 -0.59 0.41 0.78	1.23 0.31 0 .16 0.14	5% 31% 41% 23%
complete model	1.28	0.02 0.10 -0.07 -0.02	-3.41 -0.58 0.41 0.77	1.22 0.32 0.16 0.15	5% 31% 41 23%
Independent mixture model	-	-	-3.54 -0.58 0.43 0.71	1.18 0.35 0.18 0.19	4% 32% 38% 26%

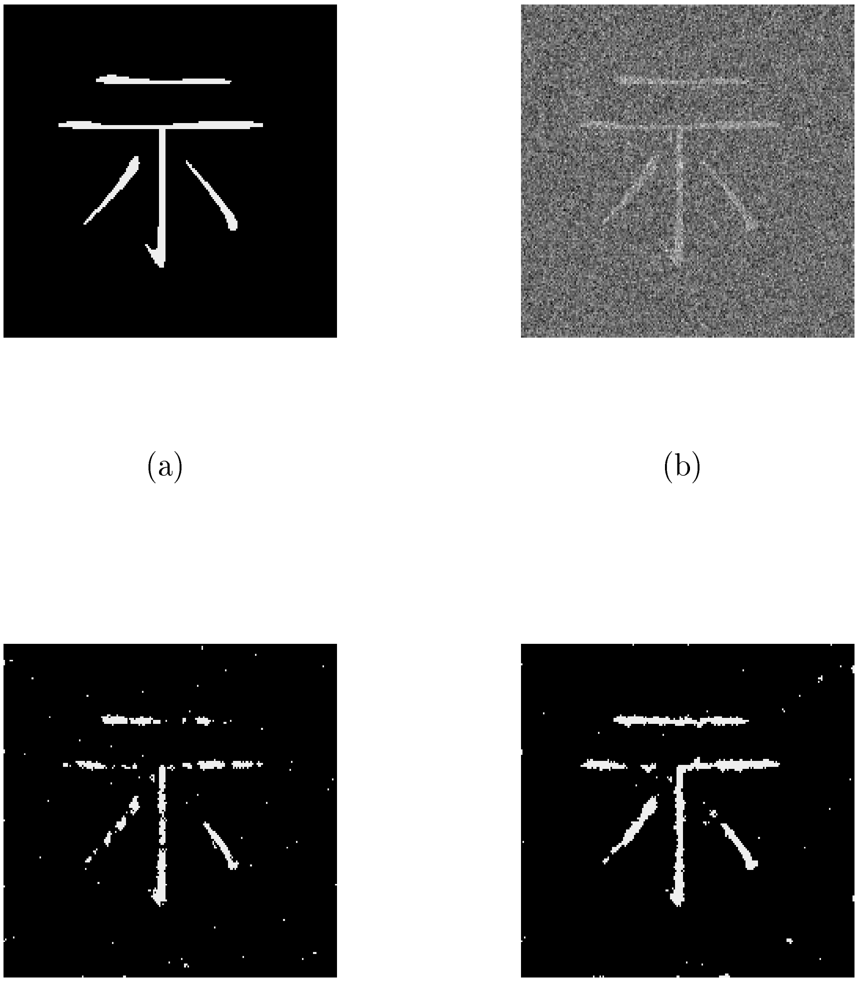
Table 3: Cell image: parameter estimation for 4-color segmentations using Potts models with or without external field and EM for independent mixtures.

between the algorithms appear when starting from a less satisfying classification: an equal proportion classification was produced by sorting the pixels values and dividing them in 4 intervals with the same number of pixels. The results are shown in Figures 5 (e) and (f). The algorithm using the complete model seems to deal better with this unfavorable initialization probably due to the extra flexibility coming from the additional parameters.

This also illustrates the difficulty in the choice of initializations. One possibility is to use random initialization or more or less arbitrary initial values or restoration. However, we made various experiments, not reported here, that showed that starting values were of significant effect and could critically affect the quality of the subsequent segmentations. The initialization choice can then be seen as an issue introducing an extra random effect in the algorithm. One way to avoid this could be to use a systematic procedure. Using EM for independent mixtures was enough to illustrate some differences in the models performance but other initialization would worth studying.

5 Discussion

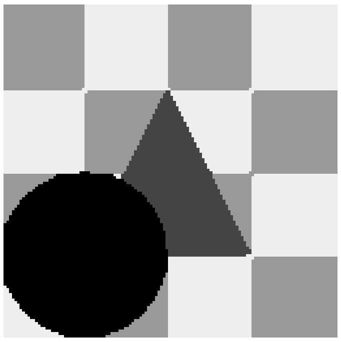
The basic Potts model (no external field), although performing well in a lot of cases, may not always capture the image characteristics well enough. A natural generalization is to use Potts models with potentials on singletons (site-dependent external field). This induces additional parameter estimations but should result in more flexibility and better adequacy especially when the colors proportions are very unbalanced in the images to be recovered and when the

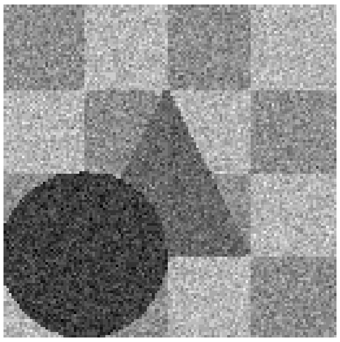


(a) (b)

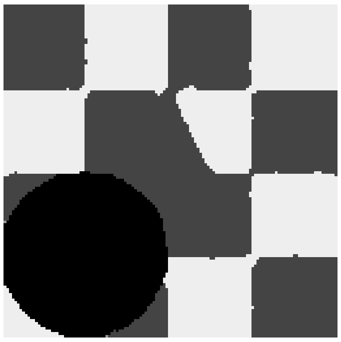
(c) (d)

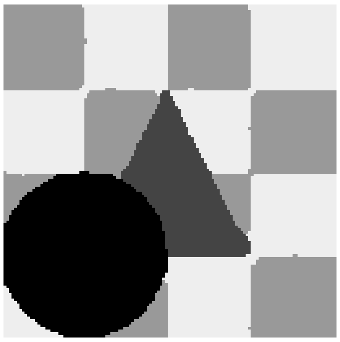
Figure 1: 2-color Ji image with very unbalanced color proportions, degraded with Gaussian noise: (a) 2-color image, (b) degraded image ($\sigma = 0.7$), (c) 2-color segmentation using a Potts model with no external field, (d) 2-color segmentation using a Potts model with a class dependent external field





(a) (b)





(c) (d)

Figure 2: 4-color image with unbalanced color proportions, degraded with Gaussian noise: (a) 4-color image, (b) degraded image ($\sigma = 0.5$), (c) 4-color segmentation using a Potts model with no external field, (d) 4-color segmentation using a Potts model with a class dependent external field

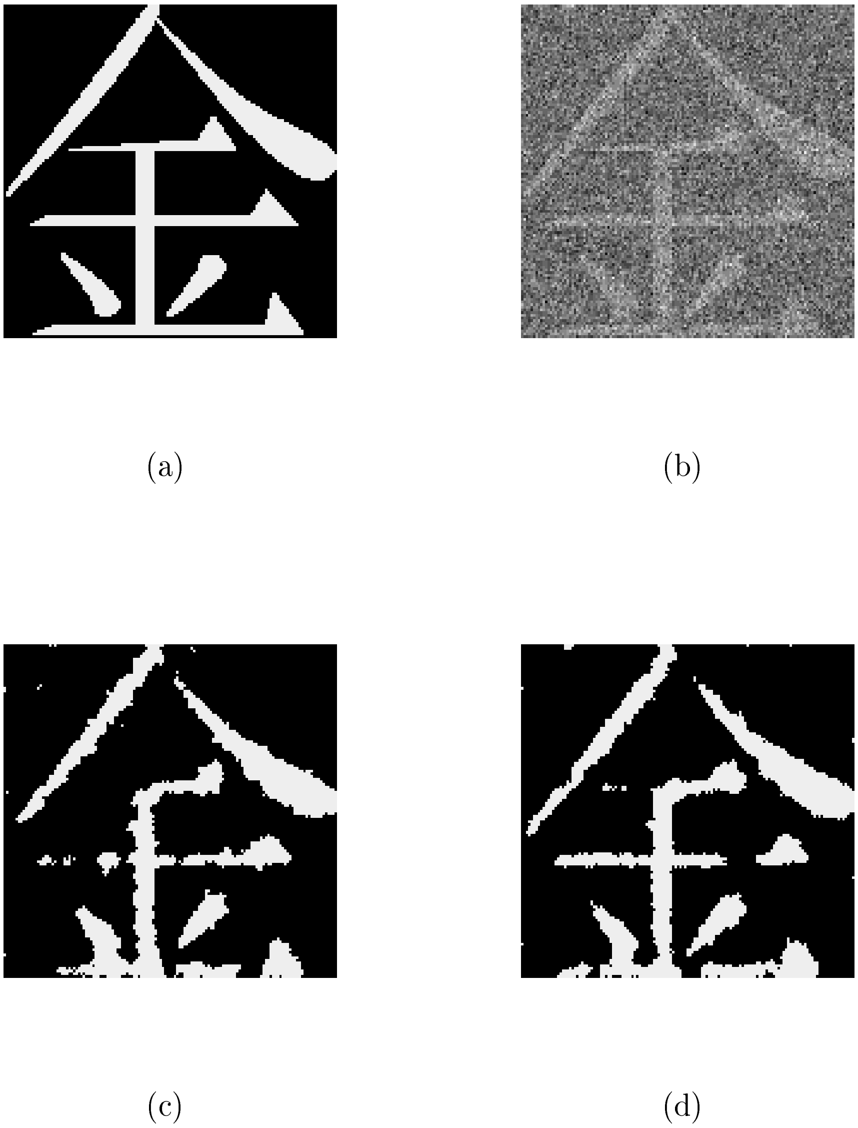
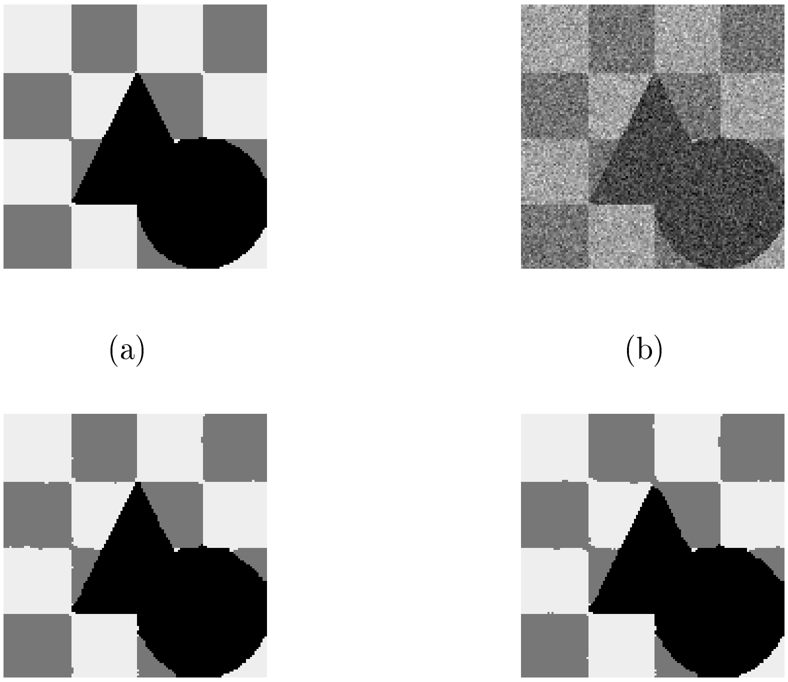


Figure 3: 2-color Kin image with less unbalanced color proportions, degraded with Gaussian nois e: (a) 2-color image, (b) degraded image (σ = 0.9), (c) 2-color segmentation using a Potts model with no external field, (d) 2-color segmentation using a Potts model with a class dependent external field



(a) (b) (c) (d)

Figure 4: 3-color image with close to equal color proportions, degraded with Gaussian noise: (a) 3-color image, (b) degraded image ($\sigma = 0.5$), (c) 3-color segmentation using a Potts model with no external field, (d) 3-color segmentation using a Potts model with a class dependent external field

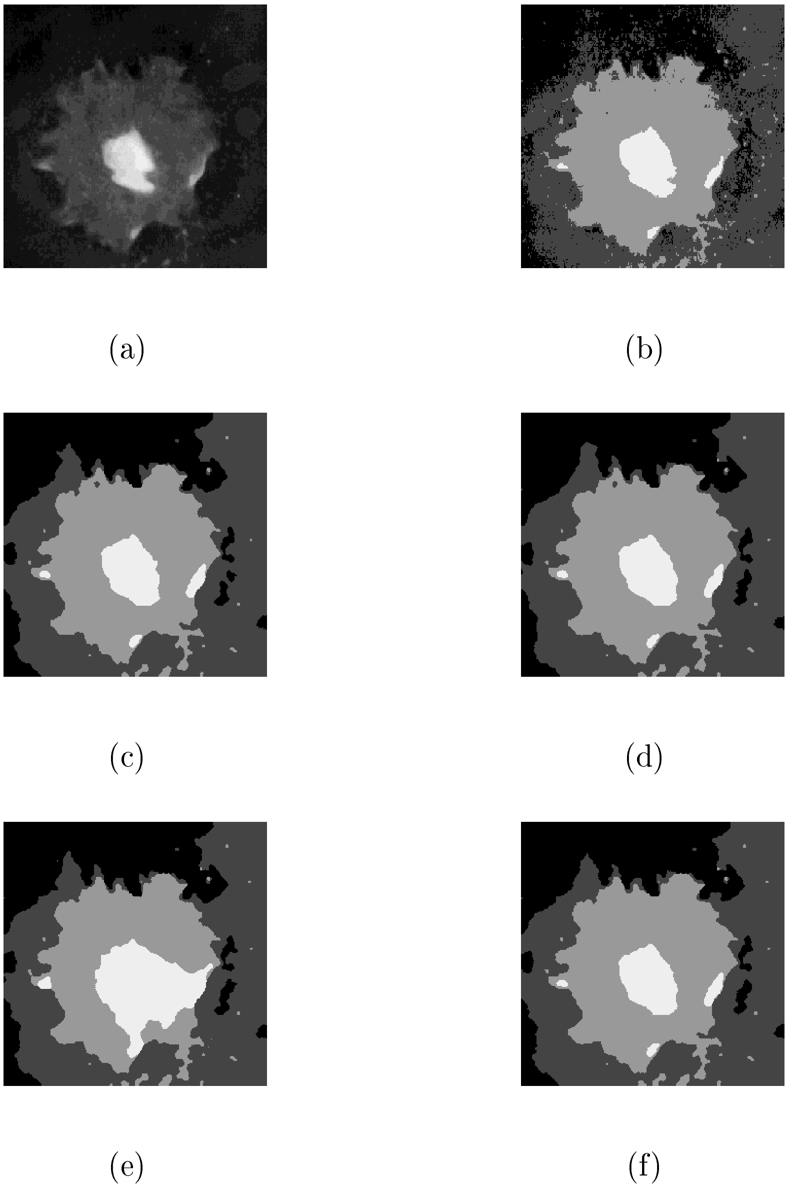


Figure 5: Cell image: (a) cell image, (b) 4-color segmentation using EM for independent mixtures, (c) 4-color segmentation using a Potts model with no external field and segmentation (b) as an initial classification, (d) 4-color segmentation using a Potts model with a class dependent external field and segmentation (b) as an initial classification, (e) 4-color segmentation using a Potts model with no external field and an initial classification in equal class proportions, (f) 4-color segmentation using a Potts model with a class dependent external field and an initial classification in equal class proportions.

quality of the initialization is uncertain. Our experiments tend to confirm this. Starting from satisfying initializations, we obtained better restorations using a complete model in the case of small proportions. This seems to result from a better estimation of the color parameters. The extra flexibility seems also to help dealing with uncertain initializations. In practice, the question is to decide if the complete model is worth using. As regards the computational time, it depends on the number of colors. In our examples, it varies from 2 to 3 times more than for the simpler Potts model but remains very reasonable. In some applications one may specifically look for objects or features representing a class in small proportion. In that case, the use of the complete model may lead to significantly different results and is worthing the extra computational cost especially when the quality of the initialization is uncertain.

Acknowledgements

The authors thank F. Chouly for his help when performing the numerical experiments.

References

[1] C. Ambroise. *Approche probabiliste en classification automatique et contraintes de voisinage*. PhD thesis, Université de Technologie de Compiègne, France, 1996.

[2] S.A. Barker and P. J. W. Rayner. Unsupervised image segmentation using markov random field models. *Pattern Recognition*, 33:587–602, 2000.

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

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

[5] G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *To appear in Pattern Recognition*, 2002.

[6] D. Chandler. *Introduction to Modern Statistical Mechanics.* Oxford University Press, 1987.

[7] A. P. Dempster, N. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, series B*, 39:1–38, 1977.

[8] E. Levitan, M. Chan, and G. T. Herman. Image-modelling gibbs prior. *Graphical Models and Image processing*, 57(2):117–130, 1995.

[9] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82(397):76–89, 1987.

[10] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, 2000.

[11] T. N. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, 40(4):901–914, 1992.

[12] W. Qian and D. M. Titterington. Estimation of parameters in hidden Markov models. *Phil. Trans. R. Soc. Lond. A*, (337):407–428, 1991.

[13] J. Zhang. The Mean Field Theory in EM Procedures for Markov Random Fields. *IEEE Transactions on Signal Processing*, 40(10):2570–2583, 1992.