CLUSTERING OF SPATIAL DATA BY THE EM ALGORITHM

C. AMBROISE , M. DANG AND G. GOVAERT

Université de Technologie de Compiègne
URA CNRS 817
BP 529 F-60205 Compiègne cedex - France

Abstract.

A clustering algorithm for spatial data is presented. It seeks a fuzzy partition which is optimal according to a criterion interpretable as a penalized likelihood. We propose to penalize the energy function exhibited by Hathaway (1986) with a term taking into account spatial contiguity constraints. The structure of the EM algorithm may be used to maximize the proposed criterion. The Maximization step is then unchanged and the Expectation step becomes iterative. The efficiency of the new clustering algorithm has been tested with biological images and compared with other clustering techniques. [1]

1. Introduction

When classical clustering techniques are used for partitioning spatial data, the resulting classes will often be geographically very mixed. To avoid this phenomenon, the spatial information of the data has to be taken into account. Very different solutions to this problem have been proposed in the literature.

A natural approach consists in using the geographical coordinates of the individuals, more or less heavily weighted, as an additional pairs of variates - see Berry (1966) or Jain and Farrokhnia (1991).

[1] This paper is published in the proceedings of geoENV96 conference, Lisbon, Portugal, November 20-22, 1996. It can be referenced as: Ambroise, C., Dang, M. and Govaert, G. (1997). Clustering of spatial data by the EM algorithm. In *geoENV I - Geostatistics for Environmental Applications*, A. Soares, J. Gómez-Hernandez and R. Froidevaux (eds), 493–504. Quantitative Geology and Geostatistics, vol. 9. Kluwer Academic Publisher.

Another approach groups individuals which are both similar and contiguous (Legendre 1987, Openshaw 1977). This implies the definition of a neighborhood concept. Defining neighborhood relationships is equivalent to building a graph where each element is represented by a node and each neighborhood relationship is an edge. Clustering with spatial contiguity constraints can be described as the succession of two steps:

1. The definition of a neighborhood graph. This can be done with standard algorithms such as a Delaunay triangulation (Green and Sibson 1977) or a Gabriel graph (Gabriel and Sokal 1969).
2. Running of a clustering algorithm while respecting the constraints. Many classical clustering algorithms may be modified to take into account the constraints which are summarized by the graph. In Lebart (1978) a classical hierarchical clustering algorithm is adapted.

These procedures produce classes made of adjacent sites. They may separate into different classes individuals which are very similar, if they are geographically far apart.

Oliver and Webster (1989) propose to run clustering algorithms based on a modified dissimilarity matrix. This modified matrix is a combination of the matrix of geographical distances and the dissimilarity matrix computed from the non geographical variables. This kind of procedure seems to work well but has no statistical justification.

Other spatially constrained clustering methods and approaches have been developed in the field of unsupervised image segmentation (Ripley 1988). The specificity of these methods is that they deal with pixels and their regular grid structure. An image may well be considered as a regular lattice of pixels. In this case, the computation of the neighborhood graph is immediate. The most common choices for the neighborhood graph are the 4 neighbor graph (horizontal and vertical adjacencies only) and the 8 neighbor (including diagonals).

There are numerous unsupervised image segmentation algorithms. In this paper, we consider only statistically based methods. In this framework, the Bayesian approach proposes solutions which may be separated into two families (Masson and Pieczinsky 1993):

1. The local methods make assumptions about the pixel or about small groups of adjacent pixels called "context" (Masson and Pieczinsky 1993).
2. The global methods make assumptions about the whole image and generally use a Markov random field model (Geman and Geman 1984, Besag 1974).

Notice that the classes obtained with these methods differ from those obtained with hard contiguity constrained algorithms: a class does not necessarily form a single patch on the image. Pixels of the same class may be

present in different parts of the image. Thus the statistical models used in unsupervised segmentation algorithms take the contiguity constraints into account but do not impose "one region classes".

We develop in this paper a new statistical method for spatial clustering which is based on the EM algorithm (Dempster *et al.* 1977) and takes into account the spatial constraints without requiring a partition made of "one region classes". Next section describes the principle of the proposed algorithm. Section 3 relates this method to statistical image segmentation techniques based on Markov Random Fields. Section 4 illustrates the method's performance on the segmentation of a biological gray-level image. The concluding section evocates aspects of this approach that are currently under investigation or seem to be worth of further study.

2. Introducing Spatial Constraints in the EM Algorithm

2.1. MIXTURE ESTIMATION BY THE EM ALGORITHM

In cluster analysis based on Gaussian mixture models (Celeux and Govaert 1995), data are $\mathbb{R}^d$-valued vectors $\mathbf{x}_1, ...\mathbf{x}_n$ assumed to be an identically independently distributed (i.i.d.) sample from a mixture of K normal distributions:

$$f(\mathbf{x}_i|\Phi) = \sum_{k=1}^{K} p_k f_k(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where the p_k are the mixing proportions (for $k = 1, \ldots, K, 0 < p_k < 1$, and $\sum_k p_k = 1$), and $f_k(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ denotes the density of a Gaussian distribution with mean vector $\boldsymbol{\mu}_k$ and variance matrix $\boldsymbol{\Sigma}_k$. This model also assumes that the unobserved vector of labels, $\mathbf{z} = (z_1, \ldots, z_n)$, is an i.i.d. sample of the multinomial distribution:

$$\text{for} \quad 1 \leq i \leq n, 1 \leq k \leq K, \quad P(Z_i = k) = p_k.$$

The EM algorithm is often used to estimate the unknown parameters of the mixture (Dempster *et al.* 1977). It produces a set of parameters that maximizes locally the log-likelihood of the sample, defined as

$$L(\Phi) = \sum_{i=1}^{n} \log f(\mathbf{x}_i).$$

The principle of the EM algorithm consists in building a sequence of estimates Φ^0, Φ^1, ..., Φ^m, over which the log-likelihood monotonically increases (for all m, $L(\Phi^{m+1}) \geq L(\Phi^m)$). The basic idea in EM is to choose Φ^{m+1} from Φ^m that globally maximizes the expectation defined as

$$Q(\Phi|\Phi^m) \quad \triangleq \quad \sum_{\mathbf{z}} P(\mathbf{z}|\mathbf{x}; \Phi^m) \log P(\mathbf{x}, \mathbf{z}; \Phi)$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} \log(p_k f_k(\mathbf{x}_i)) P(Z_i = k|\mathbf{x}_i;\Phi^m) \quad \left(\begin{array}{c}\text{for}\\ \text{mixture}\\ \text{models}\end{array}\right) \quad (2)$$

Thus, starting from an arbitrary value Φ^0, iteration $(m+1)$ of the EM algorithm can be divided in 2 steps:

— **E-step** (*Expectation*): computation of the components of $Q(\Phi|\Phi^m)$ that do not depend on Φ;
— **M-step** (*Maximization*): search $\Phi^{m+1} = \arg\max_\Phi Q(\Phi|\Phi^m)$.

In the case of mixture models, the E-step computes

$$\left[t_{ik}^{m+1} \triangleq P(Z_i = k|\mathbf{x}_i;\Phi^m) = \frac{p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)}{f(\mathbf{x}_i|\Phi^m)} \right]_{\substack{1\le i\le n\\ 1\le k\le K}} . \quad (3)$$

For a gaussian mixture, the M-step yields, for $k = 1, \ldots, K$

$$\boldsymbol{\mu}_k^{m+1} = \frac{1}{n_k}\sum_{i=1}^{n} t_{ik}^{m+1}\mathbf{x}_i \quad (4)$$

$$\boldsymbol{\Sigma}_k^{m+1} = \frac{1}{n_k}\sum_{k=1}^{K}\sum_{i=1}^{n} t_{ik}^{m+1}(\mathbf{x}_i - \boldsymbol{\mu}_k^{m+1})(\mathbf{x}_i - \boldsymbol{\mu}_k^{m+1})^t \quad (5)$$

$$p_k^{m+1} = \frac{n_k}{n} \quad (6)$$

where $n_k = \sum_{i=1}^{n} t_{ik}^{m+1}$.

2.2. THE EM ALGORITHM AS A FUZZY CLUSTERING METHOD

As Hathaway (1986) highlighted it, the EM algorithm in the case of mixture models is formally equivalent to an alternate optimization of function

$$\mathrm{D}(\mathbf{c},\Phi) \triangleq \sum_{k=1}^{K}\sum_{i=1}^{n} c_{ik}\log(p_k f_k(\mathbf{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)) - \sum_{k=1}^{K}\sum_{i=1}^{n} c_{ik}\log(c_{ik}) \quad (7)$$

where $\mathbf{c} = (c_{ik})_{i=1,n\ \ k=1,K}$ defines a fuzzy classification, c_{ik} representing the grade of membership of $\mathbf{x}_i$ to class k ($0 \le c_{ik} \le 1$, $\sum_{k=1}^{K} c_{ik} = 1$, $\sum_{i=1}^{n} c_{ik} > 0$, $1 \le i \le n, 1 \le k \le K$).

To see this, consider the following grouped coordinate ascent on function (7): starting from a given set of initial parameters, Φ^0, criterion $\mathrm{D}(\mathbf{c},\Phi)$ is alternatively optimized over the possible values of the classification matrix $\mathbf{c}$ with fixed mixture parameters, then over the possible values of mixture

parameters Φ with a fixed classification matrix. Thus, iteration $(m+1)$ of this alternate optimization algorithm consists of two steps:

"E"-step: The classification matrix is updated in order to maximize criterion

$$\mathbf{c}^{m+1} = \arg\max_{\mathbf{c}} \mathrm{D}(\mathbf{c}, \Phi^m).$$

Let us write the Lagrangian of $\mathrm{D}(\mathbf{c}, \Phi^m)$ that takes into account the constraints $\sum_{k=1}^{K} c_{ik} = 1, \forall i$:

$$\mathcal{D}(\mathbf{c}) = \mathrm{D}(\mathbf{c}, \Phi^m) + \sum_{i=1}^{n} \lambda_i \left(\left(\sum_{k=1}^{K} c_{ik} \right) - 1 \right), \tag{8}$$

where the λ_i are the Lagrange coefficients corresponding to the constraints.

The necessary conditions of optimality yield:

$$\begin{cases} \frac{\partial \mathcal{D}}{\partial c_{ik}} = \log\left(p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)\right) - 1 - \log c_{ik} + \lambda_i = 0 \\ \sum_{k=1}^{K} c_{ik} = 1 \end{cases}$$

which can also be written as:

$$\begin{cases} c_{ik} = \exp\left\{-1 + \lambda_i + \log\left(p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)\right)\right\} \\ \sum_{k=1}^{K} \exp\left\{-1 + \lambda_i + \log\left(p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)\right)\right\} = 1. \end{cases}$$

The following values are obtained for the c_{ik}^{m+1}:

$$c_{ik}^{m+1} = \frac{p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)}{f(\mathbf{x}_i|\Phi^m)}. \tag{9}$$

which is clearly identical to EM's E-step for mixture models (3).

"M"-step: The parameters are reestimated according to

$$\Phi^{m+1} = \arg\max_{\Phi} \mathrm{D}(\mathbf{c}^{m+1}, \Phi).$$

However, combining (2) with (3), and (7) with (9), it can be seen that

$$\mathrm{D}(\mathbf{c}^{m+1}, \Phi) = Q(\Phi|\Phi^m) - \underbrace{\sum_{k=1}^{K} \sum_{i=1}^{n} c_{ik}^{m+1} \log(c_{ik}^{m+1})}_{\text{independent of } \Phi}$$

so that the same Φ^{m+1} is obtained as in the M-step of the EM algorithm.

Since this scheme yields the same calculations as EM applied to mixtures, the latter can be viewed as an alternate optimization method of criterion $\mathrm{D}(\mathbf{c}, \Phi)$. The criteria themselves are related by $L(\Phi^m) = \mathrm{D}(\mathbf{c}^{m+1}, \Phi^m)$. This interpretation makes clearer the relationships between EM and other clustering techniques, which often optimize criteria very similar to $\mathrm{D}(\mathbf{c}, \Phi)$ (Celeux and Govaert 1995).

2.3. PENALIZATION OF HATHAWAY'S CRITERION

We propose to regularize Hathaway's criterion (7) with a term which takes into account the spatial information relative to the data. Let $\mathbf{V}$ be the "neighborhood matrix":

$$v_{ij} = \begin{cases} \alpha > 0 & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors} \\ 0 & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are not neighbors.} \end{cases}$$

To compute the matrix $\mathbf{V}$, one can use the graph structure described in the introduction, or a function of the spatial distance between the elements of the data set.

The regularizing term we propose is the following:

$$\mathrm{G}(\mathbf{c}) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ik}.c_{jk}.v_{ij}. \tag{10}$$

The more the classes contain adjacent elements, the greater this term is. Notice that $\mathrm{G}(\mathbf{c})$ is only function of the classification matrix. The new criterion we consider is then:

$$\mathrm{U}(\mathbf{c}, \Phi) = \mathrm{D}(\mathbf{c}, \Phi) + \beta.\mathrm{G}(\mathbf{c}) \tag{11}$$

where the coefficient $\beta \geq 0$ gives more or less weight to the spatial homogeneity term relatively to $\mathrm{D}(\mathbf{c}, \Phi)$.

A modified version of the EM algorithm can be used to maximize this criterion. We named this version *Neighborhood EM algorithm* (NEM). Since the new regularizing term does not contain any parameter of the mixture, the Maximization step remains unchanged. The Estimation step (maximization of the criterion relatively to the classification matrix with fixed mixture parameters) changes:

1. Initialization: a neighborhood matrix $\mathbf{V}$ is computed according to the spatial relationship; arbitrary initial values are chosen for the parameters of the mixtures $\Phi^{(0)}$ as well as for the classification matrix, $\mathbf{c}^{(0)}$, (n lines, K columns).
2. At each iteration the followings steps are realized until convergence:

(a) **E-step:**

$$\mathbf{c}^{m+1} = \arg\max_{\mathbf{c}} U(\mathbf{c}, \Phi^m).$$

The necessary conditions of optimality take the following form:

$$\begin{cases} \frac{\partial U}{\partial c_{ik}} = \log\left(p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)\right) + 1 - \log c_{ik} + \lambda_i + \beta \sum_{j=1}^n c_{jk} v_{ij} \\ \sum_{k=1}^K c_{ik} = 1 \end{cases}$$

which may be written as

$$\begin{cases} c_{ik} = \exp\left\{\log(p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)) + 1 + \lambda_i + \beta \sum_{j=1}^n c_{jk} v_{ij}\right\} \\ \sum_{k=1}^K \exp\left\{\log(p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m)) + 1 + \lambda_i + \beta \sum_{j=1}^n c_{jk} v_{ij}\right\} = 1. \end{cases}$$

Finally we get the following equation:

$$c_{ik}^{m+1} = \frac{p_k^m f_k(\mathbf{x}_i|\boldsymbol{\mu}_k^m, \boldsymbol{\Sigma}_k^m) \cdot \exp\{\beta \sum_{j=1}^n c_{jk}^{m+1} v_{ij}\}}{\sum_{\ell=1}^K p_\ell^m f_\ell(\mathbf{x}_i|\boldsymbol{\mu}_\ell^m, \boldsymbol{\Sigma}_\ell^m) \cdot \exp\{\beta \sum_{j=1}^n c_{j\ell}^{m+1} v_{ij}\}} \qquad (12)$$

which suggests an iterative computing algorithm of the form $\mathbf{c} = g(\tilde{\mathbf{c}})$, where $\tilde{\mathbf{c}}$ is the old classification matrix. From a practical point of view, a few iterations produce a reasonable new classification matrix $\mathbf{c}$, which can be used for the next *M-step*.

(b) **M-step:**

$$\begin{aligned} \Phi^{m+1} &= \arg\max_{\Phi} U(\mathbf{c}^{m+1}, \Phi) \\ &= \arg\max_{\Phi} D(\mathbf{c}^{m+1}, \Phi). \end{aligned}$$

Thus, to compute the parameters of the mixture, one can use the same formulae as in the M-step of the EM algorithm.

Empirically the NEM algorithm converges with all tested problems, but no proof of local convergence is available yet.

At each iteration the spatial information modifies the partition according to the importance of the β coefficient. The Estimation step smoothes the map of the labels.

The preceding algorithm solves two different tasks which are highly dependent: it estimates the parameters of the mixture and finds a partition. The two tasks are performed at the same time.

2.4. HARD PARTITION

The NEM algorithm may be easily adapted to seek a hard partition ($c_{ik} \in \{0, 1\}$):

- The easiest way is to consider the fuzzy partition obtained at the convergence, and to assign each individual i to the most probable class according to the a posteriori probabilities, i.e. the class k that maximizes c_{ik}.
- As in the CEM algorithm (Celeux and Govaert 1992), it is possible to add an intermediate classification step between the E and the M step.

3. A Bayesian Interpretation

Notice that it is possible to have a Bayesian interpretation of the NEM algorithm. Maximizing the criterion $U(\mathbf{c}, \Phi)$ is equivalent to maximizing

$$\exp\{U(\mathbf{c}, \Phi)\} = \exp\{D(\mathbf{c}, \Phi)\} \cdot \exp\{\beta \cdot G(\mathbf{c})\}. \tag{13}$$

We may write

$$\exp\{\beta \cdot G(\mathbf{c})\} \propto P(\mathbf{c}) \tag{14}$$

where $P(\mathbf{c})$ is a Gibbs distribution with energy function $-\beta \cdot G(\mathbf{c})$. On the other hand, the expression $\exp\{D(\mathbf{c}, \Phi)\}$ may be interpreted as $P_\Phi(\mathbf{x}|\mathbf{c})$, the conditional density of the sample $\mathbf{x}$, knowing the classification matrix $\mathbf{c}$, with parameters Φ characterizing the Gaussian mixture. Thus,

$$\begin{aligned} \exp\{U(\mathbf{c}, \Phi)\} &\propto P(\mathbf{c}) \cdot P_\Phi(\mathbf{x}|\mathbf{c}) \\ &\propto P_\Phi(\mathbf{c}|\mathbf{x}) \end{aligned}$$

where $P_\Phi(\mathbf{c}|\mathbf{x})$ is the posterior distribution of the classification matrix $\mathbf{c}$.

Maximizing $U(\mathbf{c}, \Phi)$ is equivalent to finding the matrix $\mathbf{c}$ that maximizes the posterior distribution $P_\Phi(\mathbf{c}|\mathbf{x})$, which is a Gibbs distribution with energy function $-U$. Thus, the NEM algorithm may be interpreted as an algorithm that searches the maximum a posteriori (MAP) estimate of the classification matrix $\mathbf{c}$.

This Bayesian interpretation assumes that there are two random fields: $\mathbf{X} = \{\mathbf{X}_s, s \in S\}$, which is the observed random field and $\mathbf{C} = \{\mathbf{C}_s, s \in S\}$, which is the random field corresponding to the classification matrix (S is the set of sites). The $\mathbf{X}_s$ take their values in $\mathbb{R}^d$ and the $\mathbf{C}_s = (C_{s1}, \cdots, C_{sK})$ in a subset of $[0,1]^K$. Both fields $\mathbf{X}$ and $\mathbf{C}$ follow Gibbs Distributions.

This approach may be used in image segmentation and provides an alternative to the existing unsupervised fuzzy segmentation algorithms (Kent and Mardia 1991, Caillol *et al.* 1993).

In the case of a hard classification matrix, i.e. the $C_s = (C_{s1}, \cdots, C_{sK})$ are random variables which take their values in $\{0,1\}^K$, and with the following neighborhood matrix:

$$v_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors} \\ 0 & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are not neighbors,} \end{cases}$$

the a priori distribution of the classification matrix is a Markov Random Field model very frequently used in image segmentation (Strauss 1977, Besag 1986):

$$\begin{aligned} P(\mathbf{C} = \mathbf{c}) &= \frac{1}{Z} \exp\{\beta G(\mathbf{c})\} \\ &= \frac{1}{Z} \exp\{\beta \sum_{i<j} v_{ij} (\mathbf{c}_i \cdot \mathbf{c}_j)\} \\ &= \frac{1}{Z} \exp\{\beta (\#\text{neighbors belonging to the same class})\} \end{aligned}$$

where Z is a normalizing constant.

4. An Application to Biological Images

A sample of living cells is laid on a nutritive substance. After a few days new living cells appear and form a thin but visible layer around the original sample. Biologists are interested in determining the surface of the new layer. We have a collection of images of size 512 by 512 [2] of such experimental results and present here the analysis of a representative image.

We aim to distinguish three different kinds of patterns, in order to determine automatically the size of the area covered by the new cells. A "good" segmentation from the biologist point of view should separate the image in three different areas representing:

- the original sample;
- the nutritive substance;
- the new cells.

We have tested three different algorithms:

- the classical EM algorithm, which does not take into account the spatial information;
- the Gibbsian EM algorithm (Chalmond 1989); this unsupervised segmentation method applies the principle of EM to a Markov random field model for the labels $\mathbf{Z}$, and an independant distribution of data $\mathbf{X}$ conditionally on $\mathbf{Z}$; here, the computation of $Q(\Phi|\Phi^m)$ in the E-step requires to simulate the posterior distribution of $\mathbf{Z}$ conditionnally on $\mathbf{x}$; this simulation is achieved by a Gibbs sampler (Geman and Geman 1984);

[2] We are indebted to Eugenio Grapa, from the Université de Technologie de Compiègne, Laboratoire de biologie cellulaire expérimentale, for the biological pictures

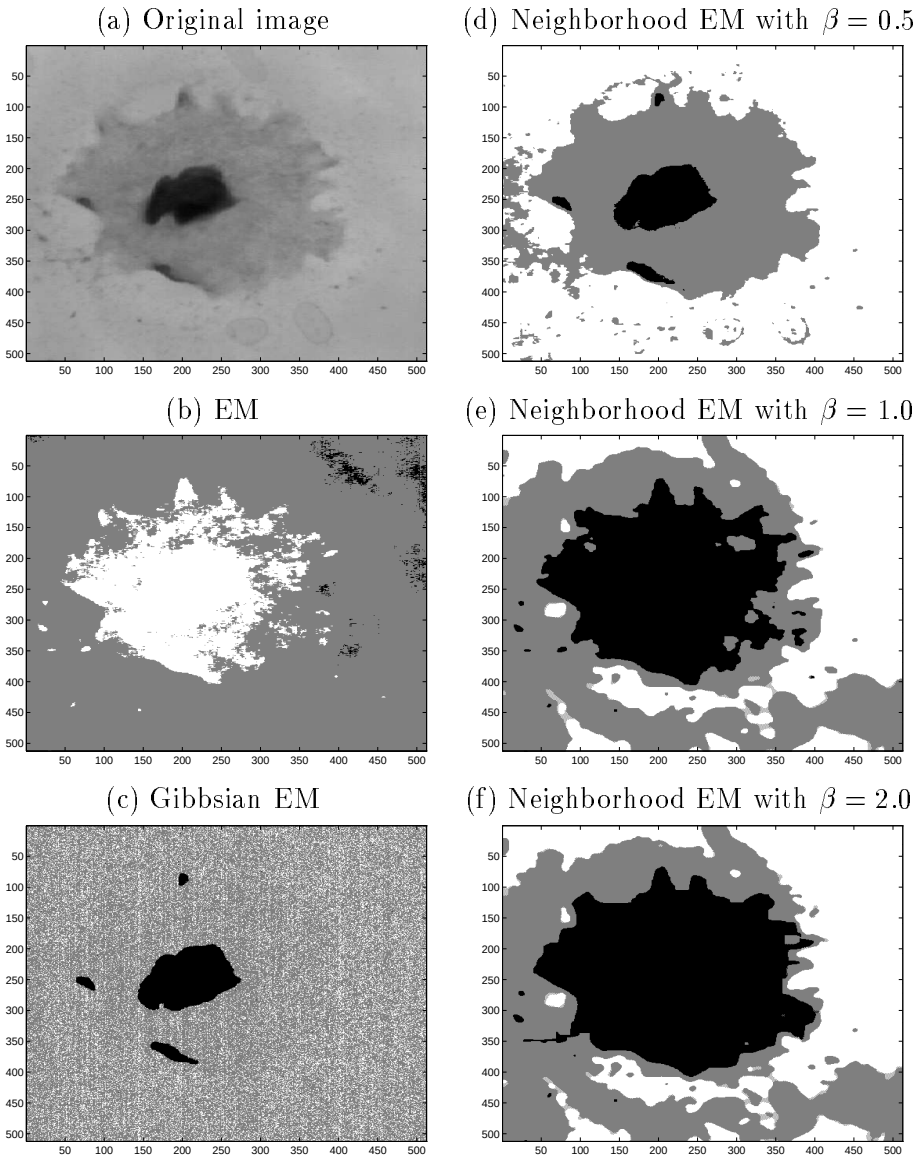

Figure 1. Segmentation of a biological image by different algorithms

– the proposed version of the EM algorithm, which maximizes the previously defined penalized likelihood.

In these experiments, the algorithms were initialized by a "histogram" method, often used in gray level image segmentation. This method consists in clustering the pixels by thresholding the histogram of the gray levels.

The thresholds are computed in order to have K classes of equal size. The clustering obtained in this way serves as an initial partition to start the algorithms.

On the original image (Figure 1(a)), the human vision distinguishes clearly the three classes. In fact, each class is far from having an uniform gray level: some pixels representing the new cells have exactly the same color as the nutritive substance pixels. While the human eyes automatically make the necessary adjustments, the unsupervised segmentation of this kind of image is more difficult than one could think at first.

The EM algorithm isolates the nutritive substance, but tends to empty the class corresponding to the original sample (Figure 1(b)) .

The Gibbsian EM algorithm does not make any distinction between the new cells and the nutritive substance (Figure 1(c)).

The NEM algorithm was tested with different values for the parameter β (Figure 1(d)-(f)). When the parameter is very small, the same results are obtained as with the EM algorithm. This seems to be logical, because if we set β to zero, NEM becomes the classical EM algorithm. When $\beta = 0.5$, the result is very satisfying and allows the automatic computation of the surface covered by the new cells. If we try to run NEM with greater values of the parameter, the spatial information becomes preponderant and the segmentation does not make a lot of sense.

5. Discussion

The preceding example showed the practical efficiency of the proposed algorithm on this particular problem. We tested the NEM algorithm on other segmentation tasks and got regularly good results once the β parameter was tuned.

The algorithm was also applied to cluster a data set consisting of socio-economical indices on a set of adjacent counties. The first results are encouraging. Further experiments should test the performance of the algorithm on other spatial data which have an irregular lattice structure.

The main difficulty in applying the NEM algorithm consists in deciding on the value of β. In this problem we validate the value using our knowledge of what the "good" solution should look like. It would be easier to have an automatic estimation of this parameter. To this purpose, the estimation methods that were developed in the framework of image segmentation could be a source of inspiration (Chalmond 1989, Younes 1989, Pieczynski 1994). However, since most of these methods were specifically designed for a regular grid structure of the sites, they would need to be adapted to work on irregularly distributed spatial data.

Theoretical results on the convergence properties of NEM have been recently established. They will be published in a forthcoming paper.

References

Berry, B.J.L. (1966). Essay on commodity flows and the spatial structure of the indian economy. Research paper 111. University of Chicago, departement of geography.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society* **35**, 192–236.

Besag, J.E. (1986). Spatial analysis of dirty pictures. *Journal of the Royal Statistical Society* **48**, 259–302.

Caillol, H., A. Hillion and W. Pieczinsky (1993). Fuzzy random fields and unsupervised image segmentation. *IEEE transactions on geoscience and remote sensing* **31**(4), 801–810.

Celeux, G. and G. Govaert (1992). A classification EM algorithm for clustering and two stochastic versions. *Computational statistics and data analysis* **14**, 315–332.

Celeux, G. and G. Govaert (1995). Gaussian parsimonious clustering models. *Pattern Recognition* **28**, 781–793.

Chalmond, B. (1989). An iterative gibbsian technique for reconstruction of m-ary images. *Pattern Recognition* **22**(6), 747–761.

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

Gabriel, K.R. and R.R. Sokal (1969). A new statistical approach to geographic variation analysis. *Systematic Zoology* **18**, 259–278.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Green, P.J. and R. Sibson (1977). Computing dirichlet tesselations in the plane. *The Computer Journal* **21**, 168–173.

Hathaway, R.J. (1986). Another interpretation of the EM algorithm for mixture distributions. *Journal of Statistics & Probability Letters* **4**, 53–56.

Jain, A.K. and F. Farrokhnia (1991). Unsupervised texture segmentation using gabor filters. *Pattern Recognition* **24**(12), 1167–1186.

Kent, J.T. and K.V. Mardia (1991). Spatial classification using fuzzy membership. *IEER Trans. Patt. Anal. Machine Intell.* **10**(5), 135–143.

Lebart, L. (1978). Programme d'agrégation avec contraintes. *Cahier de l'analyse des données* **3**, 275–287.

Legendre, P. (1987). Constrained clustering. In: *Developments in Numerical Ecology* (P. Legendre and L. Legendre, Eds.). Vol. 14 of *NATO ASI Series G: Ecological Sciences*. Springer-Verlag. Heidelberg. pp. 289–307.

Masson, P. and W. Pieczinsky (1993). SEM algorithm and unsupervised statistical segmentation of satellite images. *IEEE transactions on geoscience and remote sensing* **31**(3), 618–633.

Oliver, M.A. and R. Webster (1989). A geostatistical basis for spatial weighting in multivariate classification. *Mathematical Geology* **21**, 15–35.

Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region-building, partitionning and spatial modelling. *Transactions of the institute of british geographers* **2**, 459–472.

Pieczynski, W. (1994). Champs de markov cachés et estimation conditionnelle itérative. *Traitement du Signal* **11**(2), 141–153.

Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press. Cambridge.

Strauss, D.J. (1977). Clustering on coloured lattices. *Journal of appl. probab.* **14**, 135–143.

Younes, L. (1989). Parametric inference for imperfectly observed gibbsian fields. *Probability Theory and Related Fields* **82**, 625–645.