A Markov Pixon Information Approach for Low-Level Image Description

Xavier Descombes and Frithjof Kruggel

Abstract—The problem of extracting information from an image which corresponds to early stage processing in vision is addressed. We propose a new approach (the MPI approach) which simultaneously provides a restored image, a segmented image and a map which reflects the local scale for representing the information. Embedded in a Bayesian framework, this approach is based on an information prior, a pixon model and two Markovian priors. This model based approach is oriented to detect and analyze small parabolic patches in a noisy environment. The number of clusters and their parameters are not required for the segmentation process. The MPI approach is applied to the analysis of Statistical Parametric Maps obtained from fMRI experiments.

Index Terms—Information, Pixon, Markov Random Fields, image restoration, fMRI analysis.

1 Introduction

A CRUCIAL step before interpreting a scene from an image consists in extracting the underlying information. This step is referred to as the early stage in vision and usually denoted as low-level processing. The main areas in this domain concern image segmentation, image restoration and primitive extraction. The aim of this process is to obtain a synthetic description of the image. The first level of information concerns the detection and the localization of the objects. The second level is a description of the shape of these objects. To achieve this description we have some noisy and/or blurred data and some contextual a priori information. These issues refer to well-known areas of image processing. The extraction of objects is first performed by reducing the number of gray levels to simplify the description of the scene and is referred to as classification or clustering. Adding a priori knowledge leads to segmentation. The object shapes description refers to the problem of image restoration.

In this paper, we propose a new approach to solve these three problems simultaneously. We consider a stochastic model embedded in a Bayesian framework. We describe the scene from data (original image) by three maps referred to as the restored image, the segmented image and the pixon map. The pixon concept has been introduced by Piña and Puetter in [1] to restore astronomy images, and further been refined in [2], [3]. The term *pixon* refers to pixel information. Piña and Puetter model the restored image by the local convolution of a pseudo-image, whose entropy is maximized, by the pixon map. The size of the pixels is given by the resolution of the data. However, to locally describe an image the required resolution is not spatially homogeneous. The background and regions do not require a fine resolution whereas the description of edges and fine objects requires to use all the information given by the data resolution. The pixons are geometric features of varying size. The size of a pixon defines locally the scale of the underlying information. Herein, we define the pseudo-image as the segmentation of the scene. In the classical pixon approach, the aim of the entropy term is to regularize the pseudo-image. In our context, the entropy is used to reduce the number of gray levels in the description (i.e. to define the classes of the segmentation). Therefore, we minimize the entropy of the segmented image histogram. The regularization is performed by adding a Markovian prior. The segmented image provides a rough analysis of the scene using a piecewise constant description. By convolving the segmented image with the pixon map, we obtain a fine description of edges and fine structures. Herein, we consider the restoration of small parabolic patches in a noisy environment. Therefore, the pixon basis consists of a set of parabolas. We propose to estimate the pixon map and the segmented image by modeling them with Markov random fields (MRFs).

The construction of the model involves three steps. We first address the classification problem to obtain the segmented image. However, we do not assume to know the number of classes in the segmented image. This is a major advantage with respect to other algorithms reported in the literature. We introduce an information prior on the segmented image (also referred to as the entropy prior) to perform the classification by extracting a reduced number of gray levels (the labels) describing the scene. The most significant gray levels are selected by the information prior and represent the different classes. The pixels are classified by minimizing a distance with respect to these classes in the same procedure. The information concept was introduced by Shannon in 1948 [4] to quantify the information in a

- *X. Descombes was with Max Planck Institute of Cognitive Neuroscience, Image Processing Group, 22-26 Inselstrasse, 04103, Leipzig, Germany. He is now with INRIA, 2004, route des Lucioles BP 93, 06902 Sophia Antipolis Cedex, France.*
  E-mail: xavier.descombes @sophia.inria.fr.
- *F. Kruggel is with Max Planck Institute of Cognitive Neuroscience, Image Processing Group, 22-26 Inselstrasse, 04103, Leipzig, Germany.*
  E-mail: kruggel @cns.mpg.de.

Manuscript received 5 Sept. 1997; revised 15 Sept. 1998. Recommended for acceptance by K. Mardia.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107575.

string of symbols. The methods which maximize this information are referred to as maximum entropy (ME) methods. This concept has been used in image processing [5] either for restoration tasks [6] or for automatic thresholding [7]. For automatic thresholding the information function, defined as the sum of the two cluster entropies, is maximized. In a restoration framework, the entropy of the image is maximized resulting in a regularized solution (smooth image). In that context, the aim of entropy is to get rid of the noise. In our context, we use the entropy as a prior for the segmented image histogram. We ignore the number of classes so that the state space (set of possible values for the pixels) of the segmented image is the whole gray-level set. To get a segmentation, we have to reduce this set. The histogram of a segmented image is composed of a few peaks corresponding to the different classes. Therefore, the entropy of a segmented image histogram is low. In the proposed approach, we minimize the entropy associated with the segmented image. The corresponding model, which only provides a classification, is referred to as the I model (Information model).

The segmentation allows us to extract the objects but gives a crude description of these objects. To refine this description, we introduce the pixon description and simultaneously estimate the segmentation and the pixon map using an iterative algorithm based on a simulated annealing scheme. The restored image is then defined deterministically from the segmented image and the pixon map by local convolutions.

In the final step, we propose to improve the restoration as well as the segmentation using contextual knowledge. We consider contextual information both on the segmented image and on the pixon map to obtain regularized maps and model this information by Markov Random Fields. This leads to Markov Pixon Information (MPI) model. The previous optimization algorithm is extended to this more general model.

MRFs introduced in the engineering sciences by Besag [8] are widely used as a prior in image segmentation and image restoration because of their ability to add regularizing properties using contextual information [9], [10], [11]. The application of MRFs allows us to yield homogeneous regions when used as a segmentation prior or to smooth regions when used as a restoration prior by modeling local interactions between neighboring pixels. We first consider a Potts model (also known as the Multi Level Logistic model (MLL)) as a prior for the segmented image [9]. This model is widely used for segmentation tasks [12], [13], [14]. In the pixon map, we expect some fine structures delineating the edges of objects. Therefore we select the Chien model as a fine structures preserving model [15], [16]. This completes a low-level description of the image which is referred to as the Markov Pixon Information (MPI) model.

This paper is organized as follows. In Section 2, we introduce successively the information prior and the pixon based description leading to the PI model. MRFs are introduced in Section 3, and the two Markov priors defining the MPI model are derived. Section 4 is devoted to the application which has motivated this work. We consider the description of statistical maps obtained from functional Magnetic Resonance Images (fMRIs). We obtain a restoration of the statistical map and an analysis of the activated areas. Finally, we draw conclusions from this approach in Section 5.

2 The Pixon-Based Approach

In this section, we introduce the PI model. Denote by X the data and, respectively, by R, S, and K the restored image, the segmented image, and the pixon map. The random variable associated with the pixons is their size. For brevity, let us refer to the map of the pixon sizes as the pixon map. Now we address the problem of modeling and maximizing the following conditional probability:

$$P(R, S, K \mid X). \tag{1}$$

Using the Bayes rule and the definition of the conditional probability, we obtain the following result:

$$P(R, S, K \mid X) \propto P(X \mid R, S, K)P(R, S, K) \tag{2}$$

$$\propto P(X \mid R, S, K)P(R \mid S, K)P(S, K). \tag{3}$$

The construction of the proposed description is divided in three steps. In the first step (I model), we do not consider the pixon map, so the restored image is equal to the segmented image. In that case, (3) is reduced to the following:

$$P(S \mid X) \propto P(X \mid S)P(S), \tag{4}$$

where the first term in the right-hand side is referred to as the data attachment term (or goodness of fit (GOF) term), whereas the last term is referred to as the prior on the segmented image. Thus, we have the classical segmentation scheme interpreted as an inverse problem.

The second step introduces the pixon map. Equation (3) is written as follows:

$$P(S, K \mid X) \propto P(X \mid S, K)P(S). \tag{5}$$

2.1 The Information as a Prior for Image Classification

The most commonly used definition of information was introduced by Shannon in 1948 [4]. This quantity specifies the average information contained in a string of symbols transmitted over a communication line. The information per symbol set is given by:

$$H\left(A = \{y_0, \ldots, y_n\}\right) = -\sum_{i=1}^{n} p_i \log_2 p_i, \tag{6}$$

where n is the total number of symbols and p_i is the probability of occurrence of the ith symbol ($p_i = P(y_k = i)$). The information is also referred to as Shannon entropy by analogy with statistical physics.

In information theory, several interesting properties of the function H have been demonstrated. For our purpose, we mention a result obtained by Khinchin [17]. Dealing with information measures Khinchin showed that the defined H is the only function which satisfies the three following properties:

1) H is maximized for a uniform probability law (all the p_is are equal) and minimized for Dirac distributions (all the p_is but one are equal to 0).

2) Adding a new state j such that $p_j = 0$ does not affect the value of H.
3) $H(A, B) = H(A) + E_A(H(B/A))$, where E refers to as expectation, which is reduced to $H(A, B) = H(A) + H(B)$ if A and B are independent.

Because of these properties, this information function has been widely used in a Bayesian context for image restoration. Referred to as maximum entropy methods, the derived algorithms use this information measure as a prior. In this paper, the entropy is used as the prior on the histogram image and is minimized. Consider that the samples $y_0, ..., y_S$ represent the pixels of the image, the set $\{1, ..., n\}$ being the gray-level set. The entropy (information) function of the histogram is then defined as:

$$H\left(A = \{y_0, \ldots, y_S\}\right) = -\sum_{i=1}^{n} p(y_u = i)\log_2 p(y_u = i) = -\sum_{i=1}^{n} \frac{N_i}{S}\log_2\left(\frac{N_i}{S}\right) \quad (7)$$

where N_i is the number of pixels having the gray value i.

The entropy of the histogram is then defined as in (6), where the p_is represent the histogram values. The three properties found by Khinchin can be formulated as follows:

1) The histogram entropy is minimized for a Dirac distribution corresponding to one single class in the segmented image.
2) Adding gray levels which do not occur in the image does not change H.
3) The information given by n-independent sensors is equal to the sum of their individual information.

The first property will tend to reduce the number of gray levels defining the histogram. The second property states that the information criterion only depends on the pixel values in the image and not on the gray-level set of the image representation. Finally, the third property is useful in case of multisensor data, as it allows us to check the independency between sensors or to quantify the redundancy. In image restoration, the entropy of the image itself is considered. To smooth the image, the image entropy is maximized. The histogram of the segmented image is defined by the sum of Dirac distributions corresponding to each cluster. This histogram is represented by intervals with value 0 separated by peaks. So we want to obtain a histogram sharper than the histogram of the initial image. The first of the three mentioned properties indicates to minimize the entropy as we expect a strongly nonuniform law for the segmented image.

In this paper, we first consider the classification of data from a single sensor X in a Bayesian framework. Let L denote the lattice and S the segmented image. We assume a Gaussian model for the noise and consider that the pixels in X are independent conditionally on S so that the GOF is defined as follows:

$$P(X|S) = \prod_{l\in L}\frac{1}{\sqrt{2\pi\sigma^2}}\exp -\frac{\|x_l - s_l\|^2}{2\sigma^2}. \quad (8)$$

Note that the state space of the segmented image is the whole gray-level set. The actual reduced number of classes is obtained from the information prior. Therefore the standard deviation in the GOF term does not depend on classes but refers to the noise in the whole image.

The energy associated with the information prior is written as follows:

$$H = -\alpha_I\sum_{i=1}^{n} p_i \ln p_i = -\alpha_I\sum_{i=1}^{n}\frac{N_i}{N}\ln\frac{N_i}{N}, \quad (9)$$

where N_i is the number of pixel in state i (with gray level i), and N is the total number of pixels in the segmented image. In the I approach, the minimized energy is the sum of the information prior and the GOF term:

$$E_I = \frac{1}{2\sigma^2}\sum_{l\in L}(s_l - x_l)^2 - \alpha_I\sum_{i=1}^{n}\frac{N_i}{N}\ln\frac{N_i}{N}, \quad (10)$$

where σ is the standard deviation of the data.

2.2 The PI Model

The minimum entropy provides a first description of the image by reducing the number of gray levels and classifying the pixels. However, this approach has two main limitations. First, the information function does not take the spatial correlations in the image into account. The contribution of a given pixel to the prior depends only on its gray-level value but not on the relation with its neighbors. Second, the information entity is the pixel itself and is homogeneous on the whole image. Most images do not exhibit a uniform spatial resolution. Parts of an image such as the background contain very little information; parts containing detailed objects require a high resolution. To obtain a multiscale description of the image in order to represent a spatially varying resolution, Piña and Puetter have introduced the concept of pixon [1]. This approach has been successfully developed in astronomy to restore telescope images [2], [3], [18].

A first approach consists in defining "hard" pixons. The value of a hard pixon represents the mean brightness of a given region, while the size of the region is spatially dependent inducing a multiscale description. Such a pixon represents a degree of freedom in describing the image in this particular region. The degrees of freedom used to describe the image is then reduced in homogeneous regions. However, in this approach the accuracy of the description may actually be reduced. To achieve a multiscale description without loss of accuracy, Piña and Puetter have introduced fuzzy pixons [1]. Here, the image is modeled by the local convolutions of a pseudo-image and a pixon map. A maximum entropy prior models the pseudo-image. In our approach, the pseudo-image is replaced by the segmented image, and we use a minimum entropy prior on the histogram as stated above. The pixon model is written as follows:

$$R(r_l, l \in L) = (K \otimes S)(r_l, l \in L), \quad (11)$$

which is interpreted as follows:

$$\forall l \in L, (K \otimes S)(r_l) = \int_{V_l} K_l(l, l')s_{l'}dl', \quad (12)$$

where $K_l(l, l') = K_l(l' - l)$ is the pixon in l and V_l is called the base of K_l.

There are a wide range of choices for the pixon shape which have not yet been investigated. Following [3], we consider radially symmetric pixons defined by truncated paraboloids:

$$K_l(l'-l) = K(l'-l, d_l) = \begin{cases} \frac{1}{C_l}\left(1 - \frac{\|l'-l\|^2}{d_l^2}\right) & \text{if } \|l'-l\| \le d_l \\ 0 & \text{if } \|l'-l\| > d_l \end{cases} \quad (13)$$

where C_l is a normalizing factor such that $\|K_l\| = 1$.

Therefore, a pixon is completely defined by its size d which corresponds to the random variables of the pixon map. In practice, as we deal with discrete data, the different admissible sizes belong to the set $\{0, 1, ..., d_{\max}\}$. The pixon basis contains $d_{\max} + 1$ shapes.

To describe an image, we estimate three maps (images):

1) the restored image R,
2) the segmented image S,
3) the pixon map K.

The pixon map reflects the local resolution of features in the image. For instance, the values of d_l tends to be low on edges, whereas homogeneous regions are represented by a higher pixon size. Second, a spatially varying resolution of noise, which can spoil the segmentation, is embedded into the pixon map. As for the I model, the information prior is introduced in the segmented image model to reduce the number of gray levels in this map.

First, we estimate the segmented image and the pixon map. The restored image is then obtained using the local convolution described in (11). We embed our estimation problem into a Bayesian scheme. In this paper, we consider the data to be noisy but not blurred, so we have the following model:

$$X = R + \eta = K \otimes S + \eta, \quad (14)$$

where η represents an additive Gaussian noise.

In the PI model, no prior model is given on the pixon-map. This implicitly induces a uniform law for $P(K)$ (i.e., on each pixel the a priori distribution of the pixon size is uniform on $\{0, 1, ..., d_{\max}\}$). $P(S)$ is defined as for the I model.

$P(X \mid K, S)$ represents the GOF term and is given by:

$$P(X|K,S) = \prod_{l \in L} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \left(\frac{\left\| x_l - \int_{V_l} K_l(l,l')s(l')dl' \right\|^2}{2\sigma^2} \right) \quad (15)$$

$$= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^{|L|}} \exp - \left(\sum_{l \in L} \frac{\left\| x_l - \int_{V_l} K_l(l,l')s(l')dl' \right\|^2}{2\sigma^2} \right) \quad (16)$$

We use the information given in (9) as a prior for the pseudo-image. The solution we search for consists in the configuration (K, S), which minimizes the following energy:

$$E_{PI} = \frac{1}{2\sigma^2} \sum_{l \in L} \left\| x_l - \int_{V_l} K_l(l,l')s(l')dl' \right\|^2 - \alpha_I \sum_{i=1}^{n} \frac{N_i}{N} \ln \frac{N_i}{N}, \quad (17)$$

where N_i is the number of pixel in state i (with gray level i), and N is the total number of pixels in the segmented image.

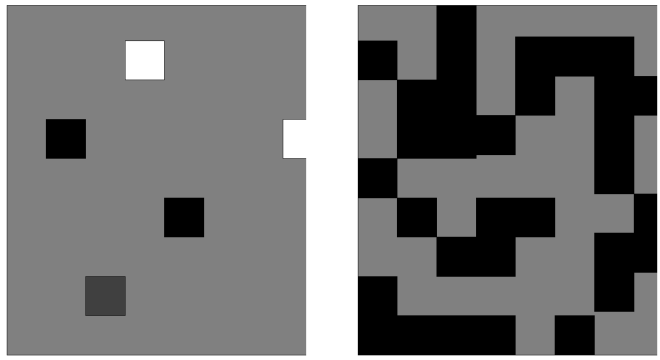
2.3 Optimization Schemes

To find the global minimum of the energy E_{PI}, we have to reach the Maximum A Posteriori (MAP) criterion for the conditional probability $P(K, S \mid X)$. We have to estimate two images. Moreover, the information prior is a global constraint, and the pixon map introduces some interactions into the GOF term. This leads to a complex energy function containing local minima. Therefore the MAP criterion requires a sophisticated optimization algorithm. To overcome this problem, suboptimal algorithms are proposed in [2] and [3]. A first approximation is made by introducing the following iterative scheme:

1) Optimize $P(K^{(n)}, S \mid X)$ with respect to S, where $K^{(n)}$ is the current estimate of K.
2) Optimize $P(K, S^{(n)} \mid X)$ with respect to K, where $S^{(n)}$ is the current estimate of S.
3) If the algorithm has not converged, increment n and go to step 1, otherwise stop.

A second approximation is made for each of the two first steps. The MAP criterion is not reached, but suboptimal optimization algorithms are used. In [2], the entropy prior is not taken into account in step 1. The optimization then leads to the Maximum Likelihood estimators. It is reported in [3] that the entropy does not significantly affect the results in practice when using astronomy data. Therefore, even when this term is considered, a conjugate gradient converging towards a local minimum is used. In our approach, we consider noisy data and the histogram entropy is minimized. The entropy prior is of great importance in selecting relevant gray levels to describe the image without noise. Therefore we propose to use a stochastic algorithm to reach the MAP criterion in step 1. We have implemented a simulated annealing algorithm whose convergence to the MAP criterion has been proved in [9]. In [2] and [3], the second step is performed by selecting at each point the pixon which minimizes the energy. The best pixon is calculated for each point, and then the pixon map is updated. In this approach, the results tend to be overresolved: Pixons tend to have small sizes in the first iterations. Thus, a lower bound for the pixon size is imposed during the first iterations, which decreases successively.

We have implemented two algorithms to perform the restoration. The first one follows the previous scheme except that simulated annealing is introduced in the first step to reach the MAP criterion for the segmented image optimization. A second algorithm optimizes $P(K, S \mid X)$ with respect to both K and S by sampling from the product space $\mathcal{K} \times S$. This algorithm is based on a simulated annealing scheme, which theoretically reaches the global energy minimum and leads to the MAP criterion of the complete model.



(a) (b)

Fig. 1. Avoided configurations using MRFs. (a) Noisy configuration. (b) Two spatially mixed classes.

3 Pixon Map and Segmented Image Regularization Using Markov Models

Using the PI model, we achieve a multiscale description of the image by using an adaptive resolution in this description. However, we did not consider noise reduction in the segmented image and in the pixon map. In this section, we consider an additional prior to take the contextual information in images into account. The main idea is to obtain regularized configurations for both the segmented image and the pixon map. This kind of prior knowledge is widely used for image segmentation and is well modeled by Markov Random Fields [10], [12], [19]. In this section, we derive two MRFs which are introduced in the PI model as priors for both the segmented image and the pixon map.

For simplicity, we assume that S and K are mutually independent. Equation (3) is written as follows:

$$P(S, K \mid X) \propto P(X \mid S, K)P(S)P(K). \tag{18}$$

In this section, we define the two Markovian priors and derive an optimization scheme for the general MPI model.

3.1 Basic Definitions

Markov Random Fields have been introduced in image processing by Besag in 1974 [8]. In 1984, Geman and Geman showed an application in binary image restoration while proving the convergence of the simulated annealing used to optimize the model [9]. An MRF is a probabilistic model defined by local conditional probabilities.

Consider a set of sites $L = \{l\}$ (the lattice of the image) and a state space Λ (set of gray values). An image is represented by an element, denoted $\Sigma_L = \{\sigma_l, l \in L\}$, of the configuration space $\Omega = \Lambda^L$. Consider a random field P defined on Ω. P is said to be a Markov Random Field if:

$$\forall l \in L, \forall \sigma_l \in \Lambda, \qquad P(\sigma_l, l \in L) > 0, \tag{19}$$

$$\forall l \in L, \forall \sigma_l, \sigma_m \in \Lambda,$$
$$P(\sigma_l \mid \sigma_m, m \in L - \{l\}) = P(\sigma_l \mid \sigma_m, m \in \mathcal{N}_l), \tag{20}$$

where $\mathcal{N}_l$ is a finite subset of L called the neighborhood of l, and, $\{\mathcal{N}_l, l \in L\}$ is the neighborhood system satisfying the two following properties:

$$l \notin \mathcal{N}_l, \tag{21}$$

$$l \in \mathcal{N}_m \Rightarrow m \in \mathcal{N}_l. \tag{22}$$

The Hammersley-Clifford theorem establishes that an MRF can be written as a Gibbs Field:

$$\forall \Sigma_L \in \Omega, P(\Sigma_L) = \frac{1}{Z}\exp - U(\Sigma_L) = \frac{1}{Z}\exp - \left(\sum_{c \in C} V_c(\sigma_l, l \in c)\right) \tag{23}$$

where U is the energy function, C is the set of cliques and V_c is the potential associated with the clique c, Z being the normalization constant also named the partition function in statistical physics. Therefore, to define a Markovian prior, we have to define a set of cliques and their associated potential.

3.2 The Segmented Image Prior

The aim of the Markovian prior for the segmented image is to get a realization composed of homogeneous regions. The information prior reduces the complexity in the gray-level set. Indeed, the compromise between the GOF and the entropy term leads to a description of the segmented image with a reduced number of gray levels, which can be interpreted as a rough description (or a classification) of the image. In the PI model, no constraint about spatial homogeneity is imposed. Therefore we can obtain some noisy configuration or strongly spatially mixed classes (see Fig. 1). To avoid such configurations, we propose to add a Markovian prior in the segmented image model. We apply the Potts model, which is widely used in image segmentation [10], [13], [19].

We consider the pairwise cliques defined by the eight connectivity:

$$C = \{\{l = (i, j), m = (i + 1, j)\}, \{l = (i, j), m = (i, j + 1)\}, \{l = (i, j), m = (i - 1, j + 1)\}, \{l = (i, j), m = (i + 1, j + 1)\}\}. \tag{24}$$

The Potts potential associated with a clique $c = \{l, m\}$ is written as follows:

$$V_c(s(l), s(m)) = \beta\delta_{s(l) \neq s(m)}, \tag{25}$$

where δ_A is equal to 1 if condition A is satisfied and 0 otherwise.

The prior on the segmented image, including the entropy term, is written as follows:

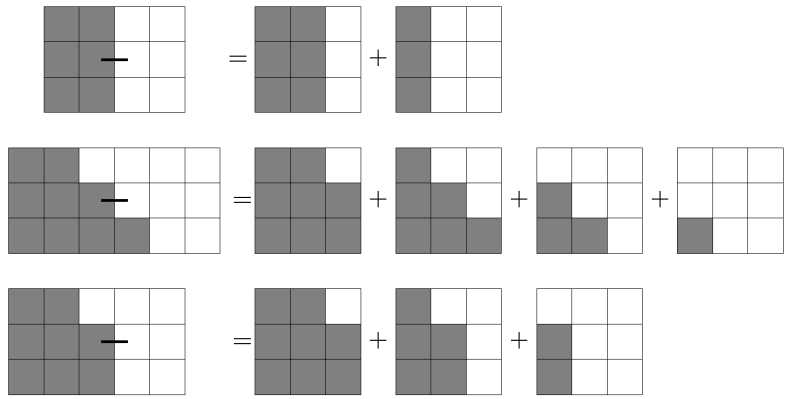


Fig. 2. Equations associated with the edge constraints for the Chien model (each sum of potentials is equal to e).

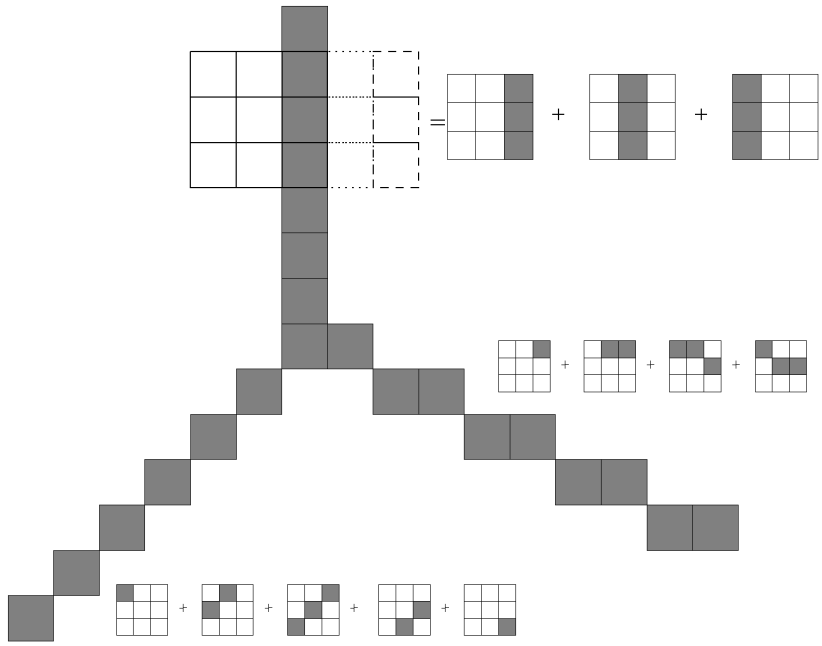


Fig. 3. Equations associated with the line constraints for the Chien model (each sum of potentials is equal to *l*).

$$P(S) = \frac{1}{Z}\exp-\left(-\alpha_I \sum_{i=1}^{n} \frac{N_i}{N} \ln \frac{N_i}{N} + \beta \sum_{c=\{l,m\}\in C} \delta_{s(l)\neq s(m)}\right), \quad (26)$$

where Z is the partition function (normalization constant).

3.3 The Pixon Map Prior

The pixon map offers a multiscale description of the image which corresponds to the local spatial resolution of the image. We expect to get spatial homogeneity from the map representing the pixon size. However, we want the pixon map to represent fine structures. The boundary of objects (fine structures) will correspond to low values for the pixon size whereas the interior of homogeneous objects will tend to higher values. We propose to apply the Chien model as a prior for the pixon map. This model, which has been proposed in [15], has the property to preserve fine structures and lines in the resulting configurations. Herein we only briefly define this model, further details can be found in [16] and [20].

The Chien model has been originally defined as a binary model. The set of cliques is composed by the 3 × 3 pixel square inducing a neighborhood of 5 × 5 for each pixel. For each clique, we have 2^9 = 512 different binary configurations. These configurations are ordered into classes considering that two configurations, which can be obtained from each other by a rotation or/and by exchanging black and white pixels belong to the same class. We obtained 51 classes denoted $C(i)$, $i \in \{1, ..., 51\}$, which are represented in Fig. 4. Setting the potential of the uniform configuration to 0, we have at this step 50 parameters, denoted $V_c(i)$. To reduce the number of parameters, we consider several local configurations. We first define a penalty equal to e for each unit of edge for boundaries in any direction. Writing the local energy of an edge, we obtain linear equations depending on the different potentials involved in an edge description (see Fig. 2; for instance, we have, for a vertical boundary: $V_c(13) + V_c(13) = e$). We also define a penalty equal to l for each unit of line (see

i	Rep.	$V_c^{(b)}(i)$	i	Rep.	$V_c^{(b)}(i)$	i	Rep.	$V_c^{(b)}(i)$
1		0	18		n	35		$0.71e$
2		n	19		$0.63e$	36		n
3		$0.35e$	20		n	37		n
4		n	21		n	38		n
5		$0.28e$	22		n	39		n
6		n	23		$0.56l - 0.32e$	40		n
7		n	24		n	41		n
8		n	25		n	42		n
9		$0.71e$	26		$l - e$	43		n
10		n	27		n	44		n
11		$0.47l - 0.23e$	28		$0.47l - 0.23e$	45		n
12		n	29		$0.48e$	46		n
13		$0.5e$	30		n	47		n
14		$0.35e$	31		n	48		n
15		n	32		n	49		n
16		$0.56l - 0.32e$	33		n	50		n
17		n	34		n	51		n

Fig. 4. Binary configurations and their associated potential in the Chien model.

Fig. 3). We solve the derived system of linear equations and obtain a value for potentials depending on e and l. The remaining potentials correspond to configurations referred to as noise and are set to n. The value of the potentials are given in Fig. 4. Therefore, the Chien model is characterized by three parameters:

1) e denotes the elementary cost of edges,
2) l denotes the elementary cost of lines,
3) n denotes the cost of noise.

The energy associated with the binary Chien model is written as follows:

$$U^{(b)}(K) = \sum_{c \in C_K} V_c^{(b)}(i), \tag{27}$$

where

$$C_K = \{\{(i-1, j-1), (i-1, i), (i-1, j+1), (i, j-1), (i, j), (i, j+1), (i+1, j-1), (i+1, j), \quad (i+1, j+1)\}, l = (i, j) \in L\}$$

and $V_c^{(b)}(i)$ is defined accordingly to Fig. 4.

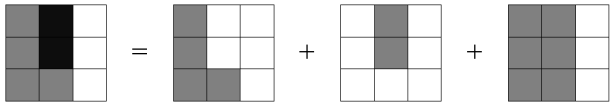


Fig. 5. Construction of an *m*-ary model from the binary Chien model.

From the binary model, an *m*-ary model is derived. Consider an *m*-ary configuration *i* on a clique *c*. For each state *d* (possible value for the d_s), denote by i_d the binary configuration defined by setting all the pixels in *c* equal to *d* to 1 and the other pixels to 0. The *m*-ary potential associated with the *m*-ary configuration *i* is then written as the sum of the corresponding potentials in the binary model:

$$V_c(i) = \sum_d V_c^{(b)}(i_d). \tag{28}$$

An example of this construction is shown in Fig. 5. For this configuration containing three labels, the potential is defined as follows:

$$V = V^{(b)}(29) + V^{(b)}(10) + V^{(b)}(13) = 0.98e + n. \tag{29}$$

3.4 The MPI Model and Optimization

In the MPI model (Markov Pixon Information), we introduce three priors in a pixon-based approach. The information prior is introduced for the segmented image as described for the PI model. We propose to add two Markov priors both for the segmented image and the pixon map. We also use a Gaussian model for the GOF. To summarize, the global energy is written as follows:

$$E_{MPI} = E_{GOF}(X, K, S) + E_I(S) + E_M^{(1)}(S) + E_M^{(2)}(K), \tag{30}$$

where:

$$E_{GOF}(X, K, S) = \frac{1}{2\sigma^2} \sum_{l \in L} \left\| x_l - \int_{V_l} K_l(l, m) s(m) dm \right\|^2 \tag{31}$$

$$E_I(S) = -\alpha_I \sum_{i=1}^{n} \frac{N_i}{N} \ln \frac{N_i}{N} \tag{32}$$

$$E_M^{(1)}(S) = \beta \sum_{c=\{l,m\} \in C} \delta_{s(l) \neq s(m)} \tag{33}$$

$$E_M^{(2)}(K) = \sum_{c \in C_K} \sum_l V_c^{(b)}(i_l). \tag{34}$$

In the MPI model, we consider some spatial interactions both on the pixon map and on the segmented image. In the first algorithm derived for the PI model, the optimization of the pixon-map consists in selecting independently the best pixon size on each pixel. This leads to poor results, as we have introduced some interactions in the pixon map. Therefore, we have to use simulated annealing for both the segmented image and the pixon map. We use the second version of the algorithm which directly samples the product space $S \times \mathcal{K}$. If we include the Markovian priors during the first iterations, small clusters are favored on the segmented image and on the pixon map which represent very deep local minima. The Metropolis dynamic used in the proposed simulated scheme is based on single pixel changes. So the information prior can not suppress a whole cluster within this dynamic in practice and does not reduce the number of gray levels in the segmented image. Indeed, the information is a global constraint as it is defined by the configuration on the whole image whereas Markov random fields are defined by the sum of local constraints involving a few number of pixels. Therefore, the dynamic associated with the information prior is much slower than the one associated with the Markovian priors. To solve this problem, we first compute some iterations of the simulated annealing optimizing the PI model. The Markovian priors involved in the MPI model are integrated during the optimization when the information prior has already reduced the number of values in the segmented image. In practice, we compute 4,000 iterations of the simulated annealing. During the first 3,000 iterations, the Markovian priors are not taken into account in the energy function. The cooling schedule is given by a geometric law: $T(n + 1) = 0.95\ T(n)$, with the initial temperature $T(0) = 1{,}000$. The program requires two minutes on a standard workstation for a 128×128 parametric map.

4 Analysis of Statistical Maps

The application which has motivated this work is the analysis of statistical parametric maps from fMRI studies. The fMRI data consist in time series of 2D slices (or 3D volumes). In this paper, we have considered 2D slices for simplicity of notations (and computation time considerations). The different slices in the time domain correspond to different acquisitions of the same slice in the human brain during two or more stimulus conditions. In categorical experiments, the protocol consists in several periods of *A* acquisitions during a baseline time (no stimulation) followed by *B* acquisitions under a stimulus condition. The aim of fMRI analysis is to compare the distributions of the gray level in a given voxel and to detect voxels for which the distributions during baseline and stimulated periods differ significantly. These voxels are referred to as activated voxels.

4.1 fMRI Analysis

The most widely used approach to analyze fMRI consists in three steps [21], [22], [23], [24], [25]:

1) optional signal restoration (filtering, ...),
2) computation of a statistical map,
3) thresholding, using correction for multiple comparison.

A recent review of fMRI analysis can be found in [26]. Consider an fMRI time series denoted by $F = (f(l, t))$, where $l \in L$ refers to the spatial coordinates and $t \in T$ refers to the time coordinate. We also consider herein that initial data have been filtered or restored if necessary. Denote T_A the subset of *T* corresponding to condition *A* shifted by the time corresponding the hemodynamic response delay and T_B the subset of *T* corresponding to condition *B* with the same shift. A statistical parametric map (SPM) corresponds to an image for which the value in a given voxel is given by a statistical test comparing the voxel distribution during T_A and the voxel distribution during T_B:

$$\forall l \in L,\ SPM(l) = test(d_A(l), d_B(l)), \tag{35}$$

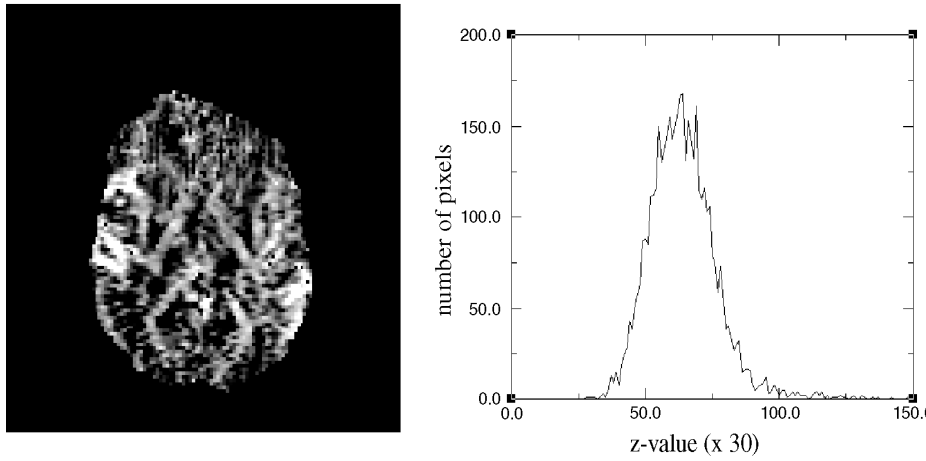


Fig. 6. Statistical Parametric Map (t-test) and its histogram.

where:

$$d_A(l) = \{f(l, t), t \in T_A\} \text{ and } d_B(l) = \{f(l, t), t \in T_B\}. \quad (36)$$

Commonly used statistical tests are the parametric t-test and Pearson correlation or the nonparametric Kolmogorov-Smirnov test. An example of an SPM obtained with a t-test is shown in Fig. 6. To detect activated voxels, thresholding is usually performed. The threshold is defined using a p-value, which takes into account the spatial extent of the detected areas, to discard false alarms [22], [27]. However, misdetections (false-negative alarms) are not considered in these approaches. Extended but flat areas (lower than the chosen threshold) can also represent an underlying activation. Moreover, defining a threshold seems quite arbitrary when considering the histogram in Fig. 6. Therefore, alternative approaches are studied to analyze fMRI [28]. Herein we propose to describe the SPM using the MPI approach. First, the restored image will remove noise (both creating false alarms and misdetections) and increase the signal to noise ratio. Second, the segmented image leads to a description of areas containing information (representing a signal different from the background) and areas referred to as background. Therefore, we avoid the threshold in the SPM analysis but analyze the information in the SPM using well defined modes in the restored image.

4.2 Results Using the MPI Approach

In this subsection, we study the results obtained, respectively, with the I model, the PI model, and the MPI model. Fig. 7a shows a SPM obtained using a t-test on a Flash MRI sequence from an experiment studying language. Before computing the t-test, data sets are preprocessed to correct baseline fluctuations and movement artifacts [29]. Fig. 7b, Fig. 7c, and Fig. 7d represent the restored images obtained with, respectively, the I model, the PI model, and the MPI model. Fig. 7e, Fig. 7f, and Fig. 7g are the corresponding pixon maps, and Fig. 7h, Fig. 7i, and Fig. 7j the segmented images. Note that for the I model, the pixon map is not defined (see Fig. 7e), so that the restored and segmented images are identical (Fig. 7b and Fig. 7h). The gray-level scale represents the intensity for the restored images. For the pixon maps and the segmented images, the gray levels only define the labels (respectively, the pixon sizes and the classes). The description obtained with the I model is reduced to a classification of the image. This classification is automatic as it does not require the number of classes nor any initial value for these clusters. The result is then interpreted as an automatic multilevel thresholding. The PI model allows us to reconstruct the image in accordance with the classification. Comparing Fig. 7b and Fig. 7c, we can notice that the I-approach description is quite crude, as each object has a constant value and hard boundaries. We can also remark by comparing the segmented images from the I model (Fig. 7h) and from the PI model (Fig. 7i) that the two classes represented in gray, detected in the I approach, originate from two different resolutions of noise rather than from real information in the original image. These two classes are found in the pixon map (Fig. 7f) but not in Fig. 7i. However, the pixon map and the segmented image obtained with the PI model are quite noisy, resulting in a suboptimal restored image. The two classes representing the background are still present in Fig. 7c. Using MRFs allows us to regularize both the pixon map (see Fig. 7g) and the segmented image (see Fig. 7j). The areas containing some information (activated areas) are detected on the segmented image, whereas the different resolutions of the noise are represented on the pixon map. The minimum size for pixons (black pixels) is obtained either for fine objects (activated areas, in our case) or for local outliers in the noise.

Consider now the histograms of the restored images. Again, the data are preprocessed to correct for movement artifacts and baseline fluctuations. Fig. 8 represents the histograms (red curves) of the restored images obtained with the I model, the PI model, and the MPI model. These histograms are composed of well-separated modes describing the content of the image (compare with the histogram of the original image in black). We have selected the modes representing the highest values which correspond to the activated pixels in our application. In the I approach, the modes are reduced to intervals of width 1. All the information is then contained in the three images corresponding to peaks 1, 2, and 3. The PI model brings a more detailed information in the restored map but pixels in a given mode are very similar to those obtained with the I

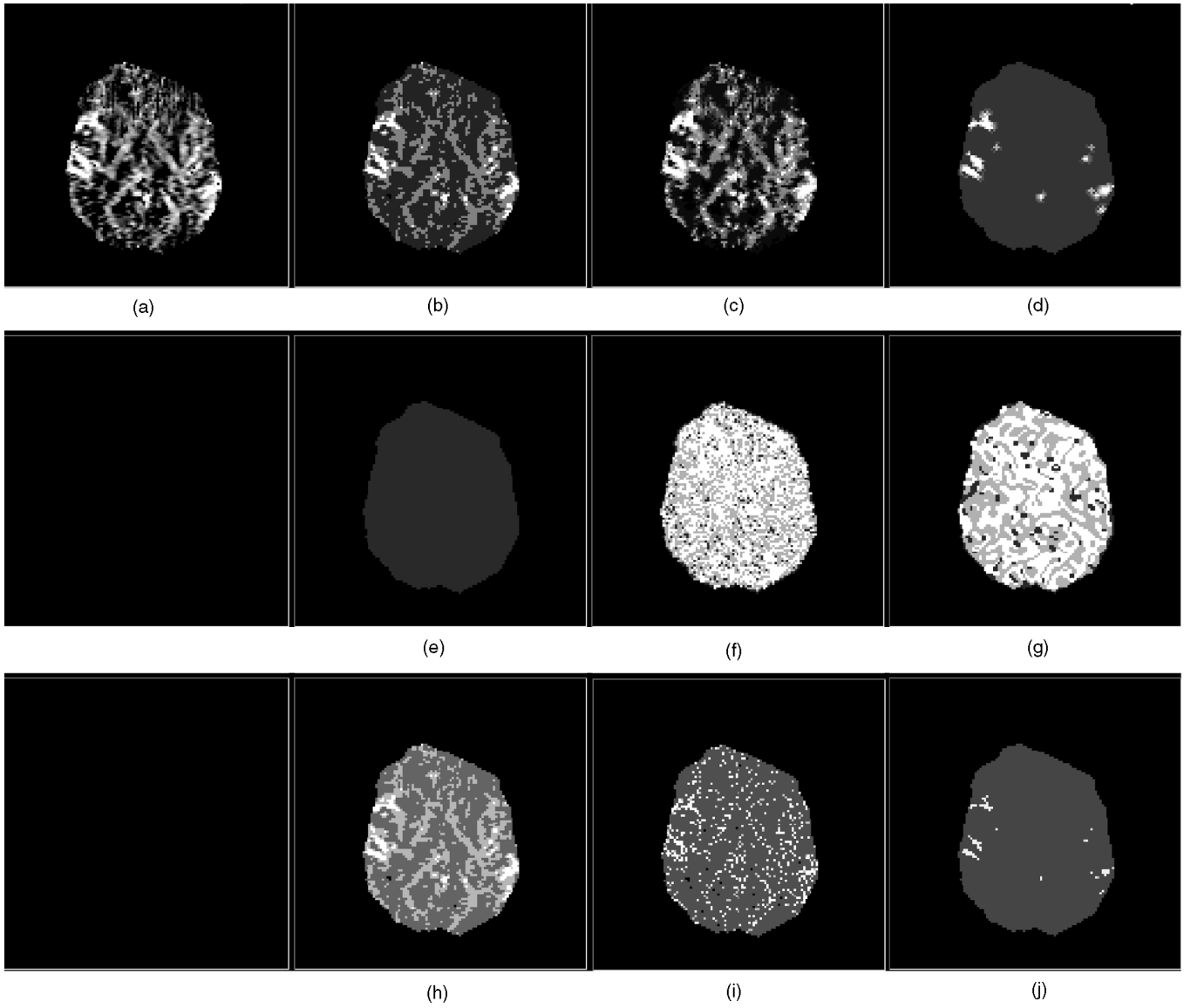


Fig. 7. (a) Initial SPM. (b)-(d) Restored images. (e)-(g) Pixon map and (h)-(j) segmented images, using the I model (b), (e), and (h); the PI model (c), (f), and (g); and the MPI model (d), (g), (j).

model. With the MPI model, the different maps corresponding to the modes are more regularized. We obtained a classification of the activated voxels with respect to their intensity. The "continuous" description is detailed on the restored image.

The proposed approach has also been validated with fMRI series obtained by echo planar imaging (EPI), which features a high temporal resolution. An example is shown in Fig. 9. For a conventional analysis, row a contains results from SPM obtained by a t-test, row b the SPM after thresholding (th = 2.5) and assessment of significance by using the spatial extent [22], overlaid onto an anatomical scan. For comparison, row c contains the MPI-restored image, row d the same SPM, thresholded by th = 1.0, and in row e, overlaid onto an anatomical scan. While both methods show essentially the same information, it is interesting to find that in the left side of the image (which corresponds to the left brain hemisphere), there is a general, low-level activation spread over the temporal lobe and the insula. Because the threshold can be lowered with MPI restoration, low-level activation is much better retained: Activation in the basal ganglia (thalamus, putamen, and caudatum) as well as in the anterior insula is much more clearly segmented. These findings are well explicable in the context of the experiments: Subjects had to perform a grammatical judgment on aurally presented sentences. It is clear that in this language-related task, the left hemisphere is generally involved, whereas in the right hemisphere only activations in the primary auditory cortices (and superficial veins) are found. The thalamic activation is especially clearly depicted in the left hemisphere, which is not apparent in comparison with the thresholded SPM. In retrospect, one may find the thalamic activation in the unrestored SPM but would hardly judge this as significant. This finding was replicated in 12 other subjects performing the same task. Thus, the MPI approach is able to recover more information from the SPM.

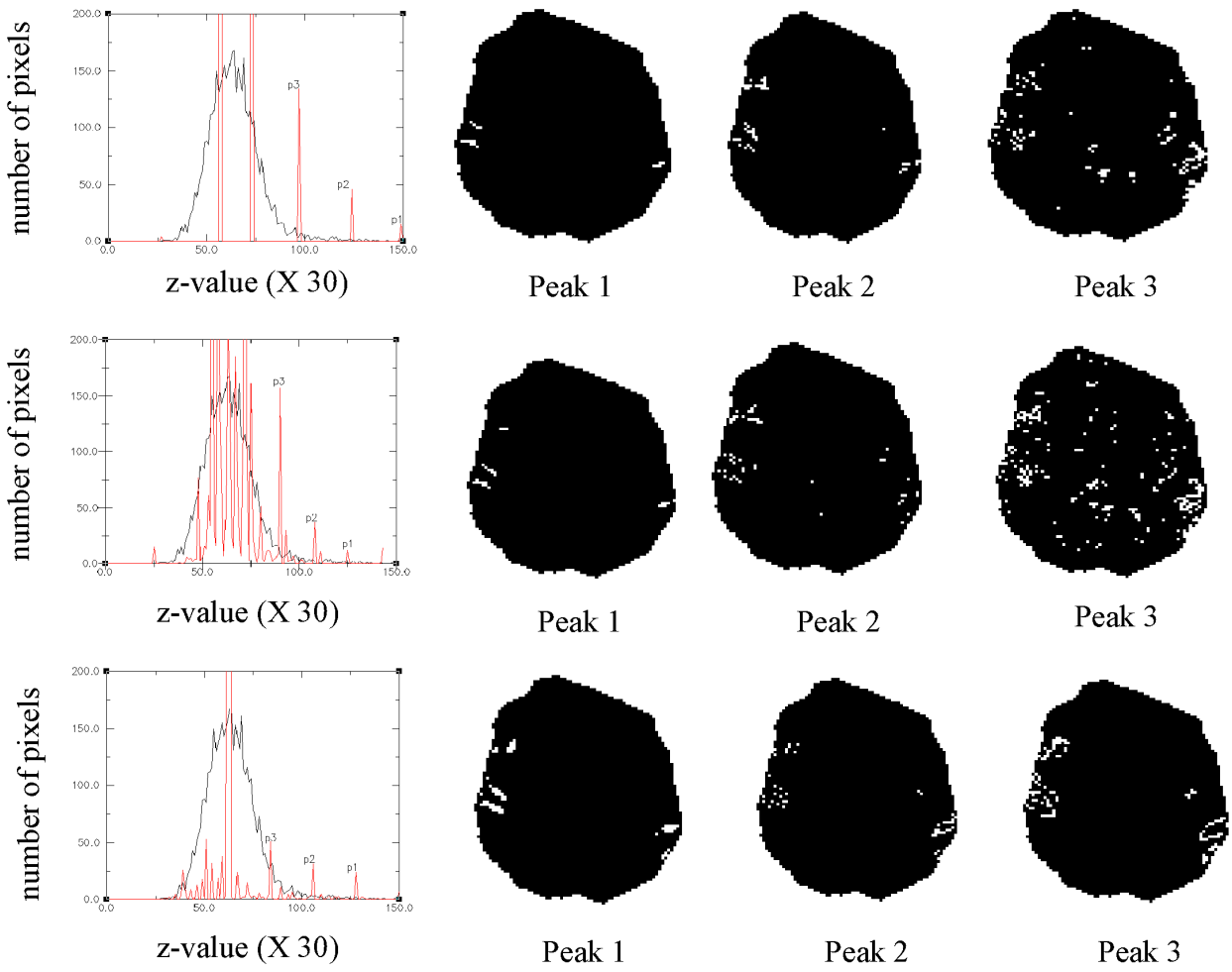


Fig. 8. SPM analysis using the I (top row), PI (middle row), and MPI (bottom row) models: histograms of the restored image (red lines) and original image (black lines), pixels belonging to the peak $p1$, pixels belonging to the peak $p2$, and pixels belonging to the peak $p3$.

5 Conclusion

In this paper, we have addressed the problem of low level description of images. The main goal of our approach is to perform a restoration and a segmentation of an image within the same algorithm which leads to an analysis of the information contained in the image corresponding to early vision. The obtained segmentation allows us to analyze the restored image. The proposed approach is embedded in a Bayesian framework which is a powerful tool to integrate a priori knowledge. The proposed algorithm produces a restored image, a segmented image, and a pixon map, reflecting the resolution of the underlying data. A first prior consists of minimizing the entropy associated with the segmented image histogram. We thus obtain a classification process which does not require the number of clusters. Markovian prior models are introduced to model both the segmented image and the pixon map in order to obtain regularized solutions. This new approach has been validated by analyzing statistical parametric maps obtained from fMRI studies.

The segmentation and the restoration are computed in the same step. The pixon map improves the segmentation by considering the local resolution of the data. This approach is unsupervised with respect to the number of clusters and their parameters. The parameters involved are the weights of the different priors and the parameters associated with the two MRFs. We do not address the parameter estimation problem in this paper. Our experience shows that these parameters are robust for a given application. Indeed, once fixed for a data set, the same values provide satisfactory results for the different experiments. We have validated this approach on two different kinds of data (Flash and EPI fMRI sequences) and obtained results for the same parameter values (σ^2 = 1,000, α_l = 1, β = 0.001, e = 0.0004, l = 0.0008, n = 0.0008). The Markovian parameters can be estimated using the method proposed in [10] for the Potts model and the method proposed in [16] for the Chien model. We are currently studying cross-validation techniques to estimate the hyper-parameters [30]. Finally, an Expectation Maximization (EM) scheme can be considered to get a completely unsupervised algorithm.

One point related to fMRI analysis concerns the significance value of the detected areas. Using spatial extend analysis and correction for multiple comparison, a p-value is usually derived in the usual thresholding approach. The smallest connected components are removed according to

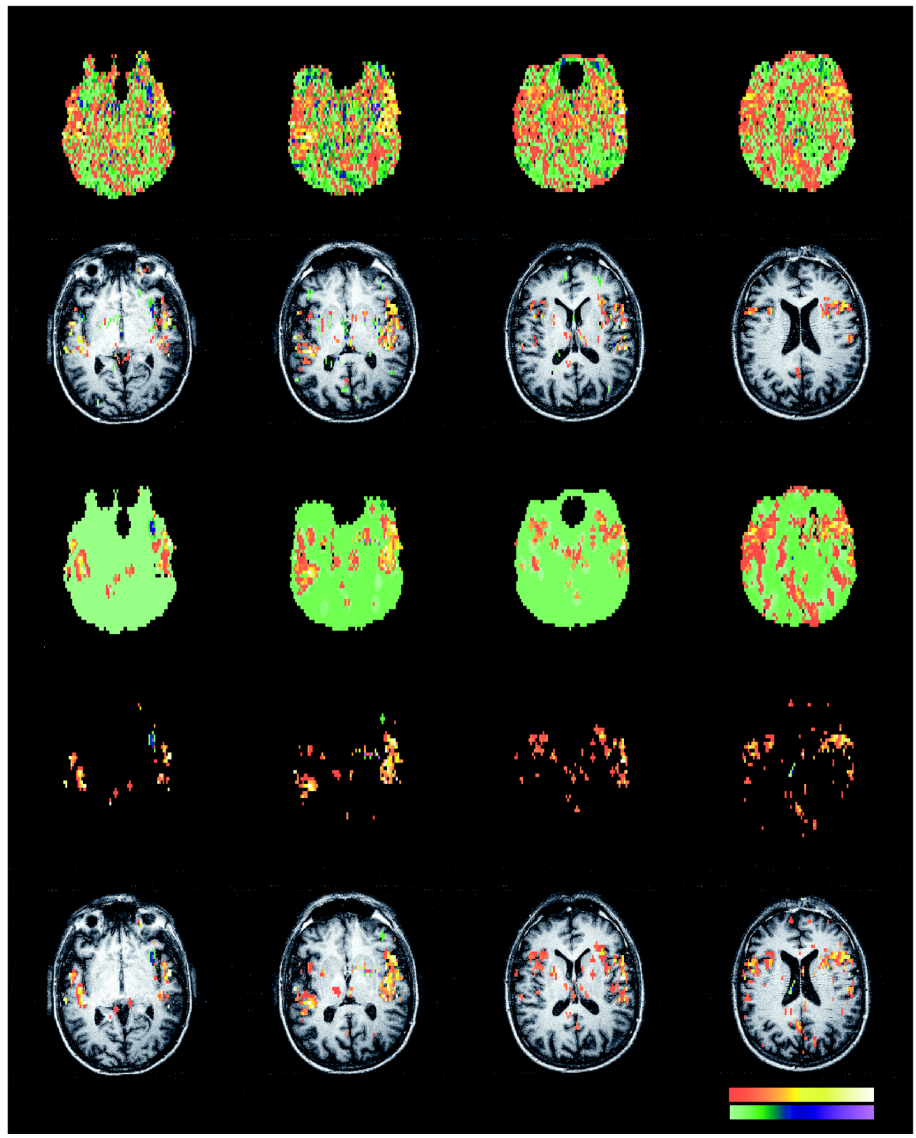


Fig. 9. Comparison of results from an fMRI study in language comprehension. (row a) Original SPM. (row b) Overlay of original SPM, threshold th = 2.5, assessment of significance by spatial extent with $p < 0.05$. (row c) MPI-restored SPM. (row d) MPI-restored SPM, threshold th = 1. (row e) Overlay of MPI-restored SPM, threshold th = 1).

this p-value leading to a false alarm rate lower than 0.05. However, this approach has various shortcomings. First, the formula which defines the p-value makes assumptions which are only asymptotically valid for high thresholds. While this provides reasonable results for positron emission tomography (PET) data, Monte Carlo simulations for fMRI data show considerable deviations from these assumptions. Second, this approach considers false positive alarms but not false negative alarms. Finally, the formula depends on the volume of the studied data set.

The proposed MPI model targets the analysis of SPMs obtained from functional cerebral images. However, the MPI principle offers wider perspectives. MPI algorithms can be derived for other kinds of data using different pixon shapes. We have proposed to use symmetric parabolic shapes to recover activation patches in fMRI studies. These shapes are also adapted to restore astronomy data but seem restrictive for more complex images. How to find the best pixon basis for a given application is an open issue.

A second field which can benefit from the MPI approach is the segmentation. For segmentation tasks, the pixon model can be removed. The MI approach provides an unsupervised segmentation algorithm. The information prior allows us to automatically determine the number of clusters. The clustering problem (finding the number of classes and their localization in the histogram) and the regularization problem are addressed within the same model.

ACKNOWLEDGMENTS

The authors wish to thank Prof. von Cramon for support and comment, Dr Norris for providing the datasets, and M. Meyer for designing the experiments.

REFERENCES

[1] R.K. Pina and R.C. Puetter, “Bayesian Image Reconstruction: The Pixon and Optimal Image Modeling,” *P.A.S.P.*, vol. 105, pp. 630–637, 1993.

[2] R.C. Puetter and R.K. Pina, "The Pixon and Bayesian Reconstruction," *SPIE*, vol. 1,946, pp. 405–416, 1993.
[3] T.R. Metcalf, H.S. Hudson, T. Kosugi, R.C. Puetter, and R.K. Pina, "Fractal Pixon Image Reconstruction for Yohkoh's Hard X-Ray Telescope," *Ap. J.*, vol. 466, pp. 585–594, 1996.
[4] C.E. Shannon, "A Mathematical Theory of Communication," *Bell Systems Technical J.*, vol. 27, p. 379, 1948.
[5] S.F. Gull and J. Skilling, "Maximum Entropy Method in Image Processing," *Proc. Inst. Elec. Eng. F*, vol. 131, pp. 646–659, 1984.
[6] B.R. Frieden, "Restoring With Maximum Likelihood and Maximum Entropy," *J. Optical Soc. Amer.*, vol. 62, pp. 511–518, 1972.
[7] A.K.C. Wong and P.K. Sahoo, "A Gray-Level Thresholding Selection Method Based on Maximum Entropy Principle," *IEEE Trans. Systems Man and Cybernetics*, vol. 19, pp. 866–871, 1989.
[8] J. Besag, "Spatial Interaction and Statistical Analysis of Lattice Systems," *Acad. Royal Statistical Soc., Series B*, vol. 36, pp. 721–741, 1974.
[9] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
[10] H. Derin and H. Elliott, "Modelling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39–55, Jan. 1987.
[11] S. Geman and G. Reynolds, "Constrained Restoration and Recovery of Discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[12] S. Lakshmanan and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 799–813, Aug. 1989.
[13] C.S. Won and H. Derin, "Unsupervised Segmentation of Noisy and Textured Images Using Markov Random Fields," *Computer Vision Graphics and Image Processing*, vol. 4, pp. 308–328, 1992.
[14] Z. Liang, J.R. MacFall, and D.P. Harrington, "Parameter Estimation and Tissue Segmentation From Multispectral MR Images," *IEEE Trans. Medical Imaging*, vol. 13, no. 3, pp. 441–449, 1994.
[15] X. Descombes, J.F. Mangin, E. Pechersky, and M. Sigelle, "Fine Structures Preserving Model for Image Processing," *Proc. Ninth SCIA 95*, pp. 349–356, Uppsala, Sweden, 1995.
[16] X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood," *IEEE Trans. Image Processing*, in press.
[17] A.I. Khinchin, *Mathematical Foundations of Information Theory*. Dover, 1957.
[18] R.C. Puetter, "Pixon-Based Multiresolution Image Reconstruction and the Quantification of Picture Information Content," *Int'l J. Image Systems and Technologies*, vol. 6, pp. 314–331, 1995.
[19] B.S. Manjunath and R. Chellappa, "Unsupervised Texture Segmentation Using Markov Random Field Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 5, pp. 478–482, May 1991.
[20] X. Descombes, R. Morris, and J. Zerubia, "Quelques améliorations la segmentation d'images bayesienne. Première partie: Modélisation," *Traitement du Signal*, vol. 14, no. 4, pp. 373–382, 1997.
[21] K.J. Worsley, A.C. Evans, S. Marret, and P. Neelin, "A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain," *J. Cerebral Blood Flow and Metabolism*, vol. 12, pp. 900–918, 1992.
[22] K.J. Friston, K.J. Worsley, R.S.J. Frackowiak, J.C. Mazziotta, and A.C. Evans, "Assessing the Significance of Focal Activations Using Their Spatial Extent," *Human Brain Mapping*, vol. 1, pp. 210–220, 1994.
[23] K.J. Friston, A.P. Holmes, J.B. Poline, P.J. Grasby, S.C.R. Williams, R.S.J. Frackowiak, and R. Turner, "Analysis of fMRI Time-Series Revisited," *NeuroImage*, vol. 2, pp. 45–53, 1995.
[24] K.J. Worsley and K.J. Friston, "Analysis of fMRI Time-Series Revisited Again," *NeuroImage*, vol. 2, pp. 173–181, 1995.
[25] N. Lange, "Tutorial in Biostatistics: Statistical Approaches to Human Brain Mapping by fMRI," *Statistics in Medicine*, vol. 15, pp. 389–428, 1996.
[26] S. Rabe-Hesketh, E.T. Bullmore, and M.J. Brammer, "The Analysis of Functional Magnetic Resonance Images," *Statistical Methods in Medical Research*, vol. 6, pp. 215–237, 1997.
[27] E.T. Bullmore, M.J. Brammer, S.C.R. Williams, S. Rabe-Hesketh, N. Janot, A.S. David, J.D.C. Mellers, R. Howard, and P. Sham, "Statistical Methods of Estimation and Inference for Functional MR Image Analysis," *Magnetic Resonance in Medicine*, vol. 35, pp. 261–277, 1996.
[28] X. Descombes, F. Kruggel, and D.Y. von Cramon, "Spatio-Temporal fMRI Analysis Using Markov Random Fields," *IEEE Trans. Medical Imagery*, vol. 17, no. 6, pp. 1,028-1,039, Dec. 1998.
[29] F. Kruggel, X. Descombes, and Y.D. Von Cramon, "Preprocessing of fMR Datasets," *IEEE Workshop on Biomedical Image Analysis*, Santa Barbara, Calif., 1998. Los Alamitos, Calif.: IEEE CS Press, 1998.
[30] A.M. Thompson, J.C. Brown, J.W. Kay, and D.M. Titterington, "A Study of Methods of Choosing the Smoothing Parameter in Image Restoration by Regularization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 326–339, Apr. 1991.

Xavier Descombes received the bachelor's degree in telecommunications from the Ecole Nationale Superieure des Telecommunications (ENST) Paris, France, in 1989, the master of science in mathematics from the University of Paris VI in 1990, and the PhD degree in signal and image processing from the ENST in 1993. He has been a postdoctoral researcher at ENST in 1994, at the Katolieke Universitat Leuven in 1995, at the Institut National de Recherche en Informatique et Automatique (INRIA) in 1996 and a visiting scientist in the Max Planck Institute of Leipzig in 1997. He is currently a permanent researcher at INRIA. His research interests include Markov Random Fields, stochastic geometry and stochastic modeling in image processing.

Frithjof Kruggel earned his diploma in chemistry from the Ruhr-University (Bochum, Germany) in 1983. He studied medicine in Bochum, Zurich, and Munich. In Munich, he received his MD (with honors) from the Ludwig-Maximilians-University in 1987. He joined the newly founded Max-Planck-Institute for Cognitive Neuroscience (Leipzig, Germany) in 1995, where he has built up and leads the workgroup on signal and image processing. Dr. Kruggel's research interests focus on the development of new algorithms to increase the information yield from neurophysiological measurements in cognitive science.