Spatio-Temporal fMRI Analysis Using Markov Random Fields

Xavier Descombes,* Frithjof Kruggel, and D. Yves von Cramon

***Abstract*—Functional magnetic resonance images (fMRI's) provide high-resolution datasets which allow researchers to obtain accurate delineation and sensitive detection of activation areas involved in cognitive processes. To preserve the resolution of this noninvasive technique, refined methods are required in the analysis of the data. In this paper, we first discuss the widely used methods based on a statistical parameter map (SPM) analysis exposing the different shortcomings of this approach when considering high-resolution data. First, the often used Gaussian filtering results in a blurring effect and in delocalization of the activated area. Secondly, the SPM approach only considers false alarms due to noise but not rejections of activated voxels. We propose to embed the fMRI analysis problem into a Bayesian framework consisting of two steps: i) data restoration and ii) data analysis. We, therefore, propose two Markov random fields (MRF's) to solve these two problems. Results on three protocols (visual, motor and word recognition) are shown for two SPM-approaches and compared with the proposed MRF-approach.**

***Index Terms*—Bayesian framework, functional magnetic resonance imaging (fMRI) analysis, Markov random fields (MRF's), signal analysis, signal restoration.**

I. Introduction

RECENT developments in medical imaging technologies allow researchers to study more elaborate cognitive tasks. During the last years, the fMRI, being a noninvasive technique with high spatial and temporal resolution has provided a new interesting option to study cognitive phenomena [1]. A widely used technique in fMRI consists of measuring the blood oxygen level dependent (BOLD) contrast [2]. This measure is related to the difference of concentration between oxygenated and deoxygenated hemoglobin. The change in the hemoglobin composition is related to neuronal activity and produces a very localized slow variation of the fMRI signal. It is possible to repeat a simple simulation over and over again (for instance a visual stimulus) to improve the statistics. However, with cognitive tasks, this becomes problematic (time constraints, learning adaptation of the subject, etc.). Therefore, corresponding to enormous improvements in technology refined techniques from statistics and image processing are required to develop sensitive and robust methods to detect and characterize functional activity. The hemodynamic function is the transfer response of the system under study. The coupling of the neuronal activity to the vascular system can be modeled by convolutions introducing blurring both in space and time domains [3]. Therefore, although the time scale of neuronal activity is milliseconds, the hemodynamic responses have time constants of a few seconds. The purpose of our work is not only to detect a signal correlated with a reference boxcar function but also to get an accurate delineation of the corresponding areas and to characterize and classify the different hemodynamic functions for each activated voxel.

A general and widely used approach has been originally developed for positron emission tomography (PET) images and is currently applied to fMRI [4], [5]. This approach basically entails three steps. A Gaussian filtering is first computed in the space domain in order to increase the signal-to-noise ratio (SNR). This filter decreases the level of noise but also spoils the underlying signal. Therefore, this step is avoided when the noise level is not too high. A statistical parametric map (SPM) is then derived using different statistical tests (t-test, Kolmogorov–Smirnov test, etc.) which have been recently compared in [6]. These SPM are then converted into a z-map modeled by a Gaussian random field (GRF) with unit variance. The clusters above a given threshold are then analyzed using the theory of excursion sets [4], the Euler's characteristics [7] or Monte Carlo simulations [8] (see [9] for a complete review). For a given threshold, a p-value reflecting the significance of the detected activation is then assigned to each cluster essentially depending on its size. A recent survey of those techniques can be found in [10].

However, this approach, although quite general, can be embedded in a more general framework if we consider the huge amount of methods developed in image and signal processing. More elaborate techniques should improve the results obtained with a Gaussian filtering and thresholding. Therefore, some new approaches have appeared in the literature since a couple of years. The famous MUltiple SIgnal Classification (MUSIC) algorithm developed in signal processing is applied to fMRI analysis in [11]. Using a hierarchical description of the z-map, a multiscale detection is proposed in [12]. Some regression based on techniques can be used for matching the signal to a reference waveform leading to parameter maps which characterize the activation [13].

In this paper, we propose a new approach based on Markov random fields (MRF's). MRF's have been introduced in image processing by Besag in 1974 [14]. Since then, they have widely been used, especially for image segmentation, classification

Manuscript received May 28, 1997; revised September 30, 1998. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was X. Hu. *Asterisk indicates corresponding author.*

*X. Descombes was with Max Planck Institute of Cognitive Neuroscience, 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 and D. Y. von Cramon are with Max Planck Institute of Cognitive Neuroscience, 22-26 Inselstrasse, 04103 Leipzig, Germany

Publisher Item Identifier S 0278-0062(98)09423-3.

and image restoration [15], [16], [17]. The main advantage of MRF's lies in their ability to take contextual information into account. A pixel will be classified with respect to its grey level value but also with respect to the value of its neighbors. The dependency between neighboring pixels is modeled by local interactions. All these interactions lead to a global energy which is minimized using an optimization algorithm such as simulated annealing (SA). Derived algorithms are particularly robust to noise due to the contextual dependency. A first approach based on MRF's has been proposed to analyze statistical maps from PET studies in [18]. In this approach a SPM is segmented into three classes (nonactivated voxels, positively activated voxels and negatively activated voxels), leading to an "intelligent smoothing of the pixel labeling." We propose to integrate MRF's earlier in the procedure to take advantage of the contextual information at each step. We first propose to restore a fMRI time series using a spatio-temporal MRF. This restoration increases the SNR as for a Gaussian filtering but does not introduce artifacts (blurring, loss of fine structures, etc.). From the restored signal we then compute the hemodynamic function. Some parameters characterizing the hemodynamic function are extracted for each time series and analyzed by a second MRF leading to an activation map and three parameter maps.

The paper is organized as follows. In section two, we analyze the SPM-based approach. We highlight the different points which can be improved using MRF's. We add basic definitions relative to MRF's and present the most widely used models for image segmentation and image restoration in Section III. In Sections IV and V, we explain the two MRF's for respectively restoring the fMRI data and analyzing the hemodynamic function. We then present results on fMRI for respectively a visual, a motor, and a word recognition protocols. Finally we conclude on future direction of MRF's applied to fMRI analysis.

II. The SPM-Based Approach

Although several approaches have been proposed to analyze functional images, the SPM framework is the most widely used. The usual approach to analyze temporal series in functional images (PET or fMRI) can be divided in the following three main steps:

- Gaussian filtering (optional);
- computation of a statistical map converted into a z-map;
- thresholding of the z-map using spatial extent and/or heights of local maxima.

When considering the different methods developed in image processing, we can remark that each of these three steps can be embedded in a more general framework, which can be stated as follows:

- signal restoration;
- signal analysis;
- decision theory.

Selecting tools from these three huge domains of investigation we can define more powerful algorithms which improve the results of each step using more general frameworks.

A. Signal Restoration

The aim of Gaussian filtering as a preprocessing step is to increase the signal-to-noise ratio (SNR). Usually applied when averaging over several subjects, filtering can also be useful in case of very noisy data which we generally find with fMRI single subject data. However, if such a filter considerably reduces the level of noise, it also may spoil the signal and results in a loss of resolution. Indeed, the Gaussian filter is a low pass filter and as such is not adapted to discontinuities. Several artifacts can result on images after filtering [see Fig. 1(c) and (d)], as follows:

- loss of fine objects;
- blurring of edges;
- fusion of neighboring objects;
- displacement of objects.

To reduce these artifacts, size and standard deviation of the filter are taken as small as possible. The restoration step is also sometimes suppressed when the noise is not too severe. However, noise reduction can be very useful for the analysis step and should be performed if possible.

Noise reduction is a particular case of signal restoration which can be interpreted as an inverse problem. Such a problem arises as soon as a signal issued from a given sensor is analyzed. Therefore, it has been largely studied in image processing, signal processing, astronomy or nuclear physics. A first class of methods consists in defining a filter which preserve the discontinuities while reducing noise. The σ-filter proposed by Lee [19] is one of the most efficient of these methods [20] [see Fig. 1(e)].

Another approach is to reach a compromise between the data and some *a priori* knowledge on the underlying image. Here, several methods have been proposed to find a regularized solution to the problem under consideration. The main idea of these methods is to search a solution which is as close as possible to the data and has regularity properties. Some *a priori* knowledge (about smoothness, homogeneity, amount of edges, etc.) is injected to the solution in order to model these regularity properties. Among these methods, MRF's are very popular because of their ability to manage regularity properties. Moreover, statistical tools to estimate parameters or to optimize the model (reach the solution) are well defined and efficient. Using MRF's, *a priori* properties are modeled by interactions between neighboring pixels. Using well-adapted interactions, we can reduce noise without blurring effect [see Fig. 1(f)].

In this paper, we propose to restore the fMRI signal using an edge preserving MRF. Moreover, the restoration will be achieved in the spatio-temporal space to consider contextual information both in the space domain and in the time domain.

B. Statistical Map Analysis Based on Thresholding

The signal being restored, the time series is analyzed by performing statistical tests. Several tests are proposed (t-test, KS-test, correlation, etc) in the literature [5], [21], [10] and have been compared in [6]. In fact, these tests quantify a "distance" between two distributions, the first one corresponding to the baseline and the other to a stimulation state. The time

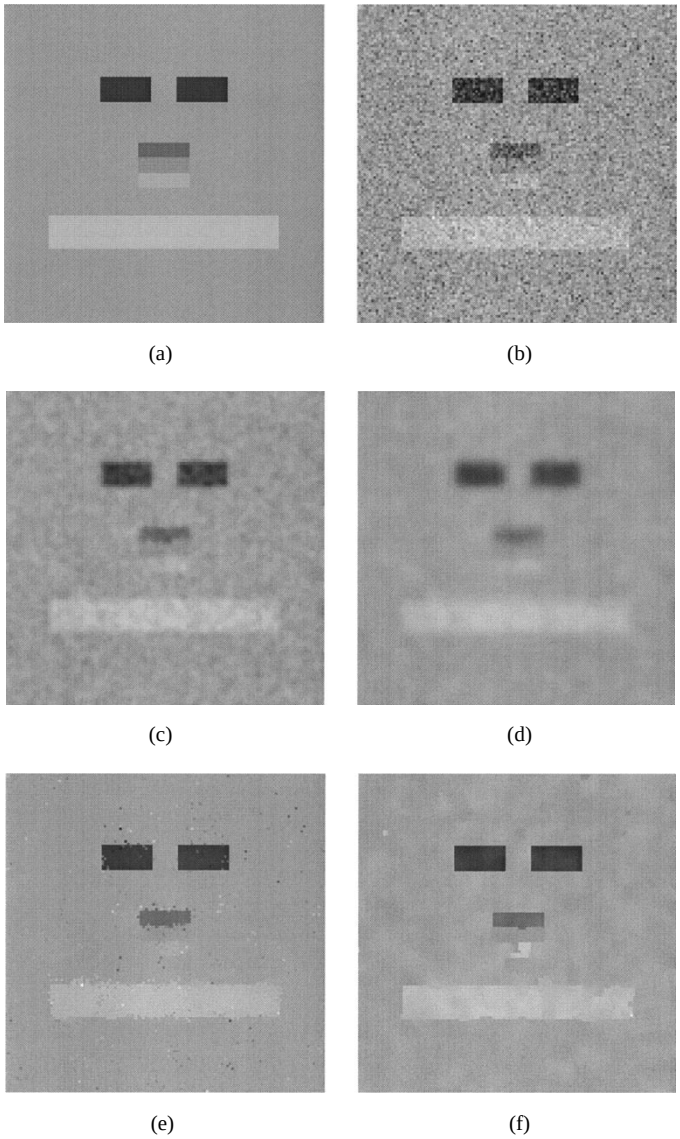

Fig. 1. Example of image restoration on synthetic data. (a) Original data, (b) noisy data Gaussian noise ($\sigma = 20$), (c) restoration with a Gaussian filter ($\sigma = 1$), (d) restoration with a Gaussian filter ($\sigma = 2$), (e) restoration with a sequence of σ-filters, and (f) restoration with a MRF.

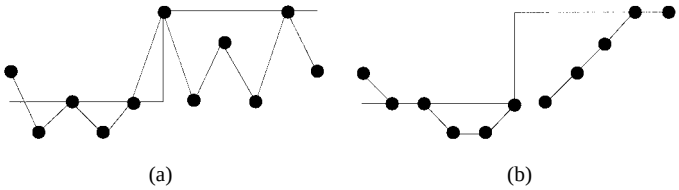

Fig. 2. Two signals with the same t-test value. (a) Noisy and (b) match the stimulus.

dimension is lost here as the only information incorporated is the binary variable "stimulated" or "not stimulated" (baseline). Fig. 2 displays two synthetic time courses. For these two time courses we obtain the same value for a t-test and the same value for the correlation. Therefore, classic fMRI studies classify these two time courses either as noise or as signal with the same significance. However, the signal plotted on Fig. 2(b) seems more significant than on Fig. 2(a) as it is less noisy and better matches the reference boxcar. Indeed, the classic approaches consider two kinds of time samples (activated and baseline periods) but the time dimension is not taken into account in the statistical test during these periods. A very high (with respect to low) value of a single time sample (outlier of the noise distribution) during a stimulated (with respect to baseline) period can considerably increase the result of the statistical test leading to a false alarm. The reverse is also true and leads to false negative responses. One can argue that once the detection of activated areas is performed, the time dimension can be recovered by studying the underlying signal. However, this implies that activated areas have been accurately detected. The detection itself can be improved by using the time information during the analysis process. A statistical test based on correlation can be derived for any kind of waveforms using for example several time delays. However, in that case, the detection depends on the quality of the waveforms choice.

The final detection is usually performed by thresholding the statistical map converted into a z-map. A z-map is modeled by a zero mean unit variance GRF of which the smoothness parameter can be estimated from the z-map using different approximations [7], [22]. Clusters above a given threshold are then analyzed with respect to their size, their maximum value or both of them [23]. To assess the significance of a cluster, the probability that this event (a cluster of this size above the chosen threshold) appears with such a GRF in the search volume is given and referred as the p-value. Derived from a frequencist point of view, this p-value depends on the search volume. The knowledge on the search volume (visual cortex for example) can be incorporate in a prior using a Bayesian approach. Second, the significance is referring to errors of type I (false alarms) only without any considerations of the probability of the cluster to be a signal. In practice, the z-map is thresholded and the connected components with a size greater than a given number of voxels and/or with a high maximum value are considered to be activated. Using a Bayesian framework allows us to compare the probability of a given cluster to be assigned as the noise with its probability to represent an underlying signal. Finally, the clusters are usually only defined by their size or sometimes by their size and maximum value. Using a MRF approach, clusters are defined by the integral of the signal, the length of edges and the contrast with respect to the background.

Consider Fig. 3(a) as an activation map. In Fig. 3(b), we add Gaussian noise so that it can be referred to the z-map. Fig. 3(c) and (d) shows the detection obtained using a threshold and then removing small connected components. Details and fine structures have not been detected. Moreover, as we only consider errors of type I, some holes spoil the activated areas. We have tried to apply a σ-filter before thresholding but the result on Fig. 3(e) shows that lots of details have been lost (see the fine structures on the bottom right quarter and the birds at the top left). Using an adapted MRF (the so-called Chien-model described in [24] and [25]), we can preserve the details. Fig. 3(f) shows that we have detected the birds and most of the lines without having false alarms.

III. Background on Markov Random Fields

In this section, we derive basic definitions concerning MRF's. For a complete description we refer to [26] for a physical approach and to [27] for a mathematical approach. The collective work [17] presents different applications of MRF's in image processing.

An MRF is a discrete stochastic process whose global properties are controlled by means of local properties. They are defined by local conditional probabilities. In the last decade, their use has proliferated in the engineering sciences [14], especially in image processing [15].

A. MRF's and Gibbs Fields

Let us denote by $S \subset Z^{\nu}$ the *set of sites* (lattice) and by Λ the *set of states* (grey levels). $\Omega = \Omega(S) = \Lambda^S$ denotes the set of the applications from S in Λ. Ω is called the *universe*. Consider the σ-algebra $\tau = \mathcal{P}(\Omega)$ (set of the subsets of Ω), and P, a probability on τ. An element of the universe is called an *event* or a *configuration* in Λ^S. It is denoted by Σ_S and the state of the site s by σ_s.

Definition 1: A random field (Σ, P) is said to be *Markovian* if and only if

$$\forall s \in S, \quad \forall \lambda \in \Lambda$$

$$P(\sigma_s = \lambda \mid \sigma_t,\ \forall t \in S - \{s\}) = P(\sigma_s = \lambda \mid \sigma_t,\ \forall t \in \mathcal{N}_s)$$

where $\mathcal{N}_s$ is a finite subset of $S - \{s\}$ and is called the *neighborhood* of s.

This means that if we know the values of the neighbors of pixel s, the law of σ_s is completely defined and does not depend on the values outside the neighborhood of s.

Consider a subset $C \subset \mathcal{P}(s)$ and a family $V = \{V_c, c \in C\}$, where V_c is an application from $\Omega(c) = \Lambda^c$ to R.

Definition 2: A random field (Σ, P) which satisfies the following property:

$$P(\Sigma) = \frac{1}{Z}\exp -\frac{1}{T}U(\Sigma) = \frac{1}{Z}\exp \frac{-\sum_{c\in C} V_c(\sigma_s, s \in c)}{T}$$

is called a *Gibbs field*.

Any element c of C is called a *clique*, V_c is the associated potential. T is a scale parameter referred as *temperature*, whereas $U(\Sigma)$ is the energy of the configuration Σ.

Adding the positivity constraint which states that no configuration has a probability equal to zero, the Hammersley–Clifford theorem [28] establishes the equivalence between MRF's and Gibbs Fields.

B. The Bayesian Framework

The usual framework for segmentation using MRF's is the Bayes formalism. Let $X(1), \cdots, X(n)$ be the data, representing different images and Y be the segmented image (for example a binary image represented activated and nonactivated areas). The aim of segmentation is to maximize the conditional probability $P(Y \mid X(1), \cdots, X(n))$. The known quantities are the *a priori* model which is represented by $P(Y)$ and models the expected result and the data attachment (goodness of fit) which is represented by the conditional probability $P(X(1), \cdots, X(n) \mid Y)$. We, thus, use the Bayes rule which states that

$$P(Y \mid X(1), \cdots, X(n)) = \frac{P(X(1), \cdots, X(n) \mid Y)P(Y)}{P(X(1), \cdots, X(n))}. \tag{1}$$

As $X(1), \cdots, X(n)$ are known, to segment the image means to maximize the product in (1) numerator.

In our context, Y is supposed to be a MRF so that the prior $P(Y)$ can be written as a Gibbs field

$$P(Y) = \frac{1}{Z}\exp -\sum_{c\in C} V_c(y_s, s \in c). \tag{2}$$

To compute the data attachment model $P(X(1), \cdots, X(n) \mid Y)$, we consider that the conditional probabilities $P(x_s(1), \cdots, x_s(n) \mid y_s)$ are mutually independent and strictly positive. In that way, we have

$$\begin{aligned} &P(X(1), \cdots, X(n) \mid Y) \\ &\quad = \prod_{s\in S} P(x_s(1), \cdots, x_s(n) \mid y_s) \\ &\quad = \exp -\sum_{c=\{s\}\subset S} -\ln P(x_s(1), \cdots, x_s(n) \mid y_s). \end{aligned} \tag{3}$$

We then have

$$\begin{aligned} &P(Y \mid X(1), \cdots, X(n)) \\ &\quad \propto \frac{1}{Z}\exp -\left[\sum_{c\in C} V_c(y_s, s \in c) \right. \\ &\qquad \left. - \sum_{c=\{s\}} \ln P(x_s(1), \cdots, x_s(n) \mid y_s)\right]. \end{aligned} \tag{4}$$

To minimize the energy we use a SA algorithm. This algorithm is iterative and considers conditional probabilities on each site. Using the Markov property, these conditional probabilities only depend on the neighborhood. The main idea of SA algorithm is to accept increasing of the energy function in order to escape from the local minima. The number and height of these jumps of energy are controlled by a scaling parameter referred as temperature: T. This temperature slowly decreases during iterations and convergence is obtained when T tends to 0. The SA algorithm converges toward the global minimum of the energy function and does not depend on the initialization [15]. It can be described as follows.

1) Initialize a random configuration $Y = (y_s)$, set $T = T_0$.
2) For each site s:
   a) Choose a random value *new* different from the current value *cur*.
   b) Compute the local energies $U(y_s = \text{cur} \mid \mathcal{N}_s)$ and $U(y_s = \text{new} \mid \mathcal{N}_s)$.
   c) If $U(y_s = \text{new} \mid \mathcal{N}_s) < U(y_s = \text{cur} \mid \mathcal{N}_s)$ set y_s to *new*, otherwise set y_s to *new* with probability $\exp -[\frac{\Delta U}{T}]$ where $\Delta U = U(y_s = \text{new} \mid \mathcal{N}_s) - U(y_s = \text{cur} \mid \mathcal{N}_s)$.

3) If the stopping criterion is not reached decrease T and go to Step 2).

To adhere to the theoretical properties of convergence we should consider a logarithmic decrease of the temperature. However, to obtain a faster algorithm we have considered a linear decrease of rate 0.95. For the kind of energy functions we use, this decrease rate is slow enough to achieve a good solution.

In the next subsections, we recall the most popular MRF's in image processing, for segmentation and restoration tasks.

C. The Most Used MRF's in Image Processing

The binary Ising and m-ary Potts models are the most commonly used for image segmentation. The energy associated with the Ising model is written as follows:

$$U(\Sigma_S) = -\sum_{c=\{s,\,s'\}\in C} \beta_{s,s'}\delta_{y_s=y_{s'}} - h\sum_{s\in S}\delta_{y_s=x_s} \qquad (5)$$

where $\beta_{s,s'}\delta_{y_s=y_{s'}}$ represents the interaction between sites s and s' which tend to result in homogeneous regions and $h\delta_{y_s=x_s}$ is the data driven term (X being a noisy binary configuration).

The SA algorithm leads to a compromise between both terms by minimizing the energy. Adding *a priori* spatial information improves the segmentation obtained using a simple maximum-likelihood (ML) approach (which is a threshold in the binary case) by regularizing the solution. The Ising model penalizes neighbors which are in a different state (have a different value). Some more sophisticated models have also been proposed [24]. The improvements obtained with respect to thresholding are shown on Fig. 3.

In the restoration problem we consider that we have some data X representing a noisy version of Y. The restoration process consists in recovering Y from X. In a probabilistic framework, we are searching for the configuration Y which maximizes the conditional probability $P(Y \mid X)$. Suppose we have a model of the noise, namely we know $P(X \mid Y)$. Using Bayes law, the restoration process then consists in maximizing the product $P(X \mid Y)P(Y)$ where $P(Y)$ represents the prior.

Consider an additive independent Gaussian noise of mean zero and standard deviation σ. The data attachment term $P(X \mid Y)$ is then written as follows:

$$P(X \mid Y) = \prod_{s\in S} P(x_s \mid y_s) = \prod_{s\in S}\frac{1}{\sqrt{2\pi\sigma^2}}\exp -\frac{(x_s-y_s)^2}{2\sigma^2}$$
$$= \exp -\sum_{s\in S}\left(\frac{(x_s-y_s)^2}{2\sigma^2} + \frac{1}{2}\ln\left(2\pi\sigma^2\right)\right). \qquad (6)$$

A first prior can be obtained using Gaussian interactions. Consider two neighbors s and s'. The potential associated with clique $c = \{s, s'\}$ is the written as follows:

$$V_c(y_s, y_{s'}) = \beta(y_s - y_{s'})^2. \qquad (7)$$

Such a potential is shown on Fig. 4(a). When the difference between two neighbors increases, the potential value increases as a quadratic function. The *a priori* effect is then to smooth the data as the neighbors tend to have the same value.

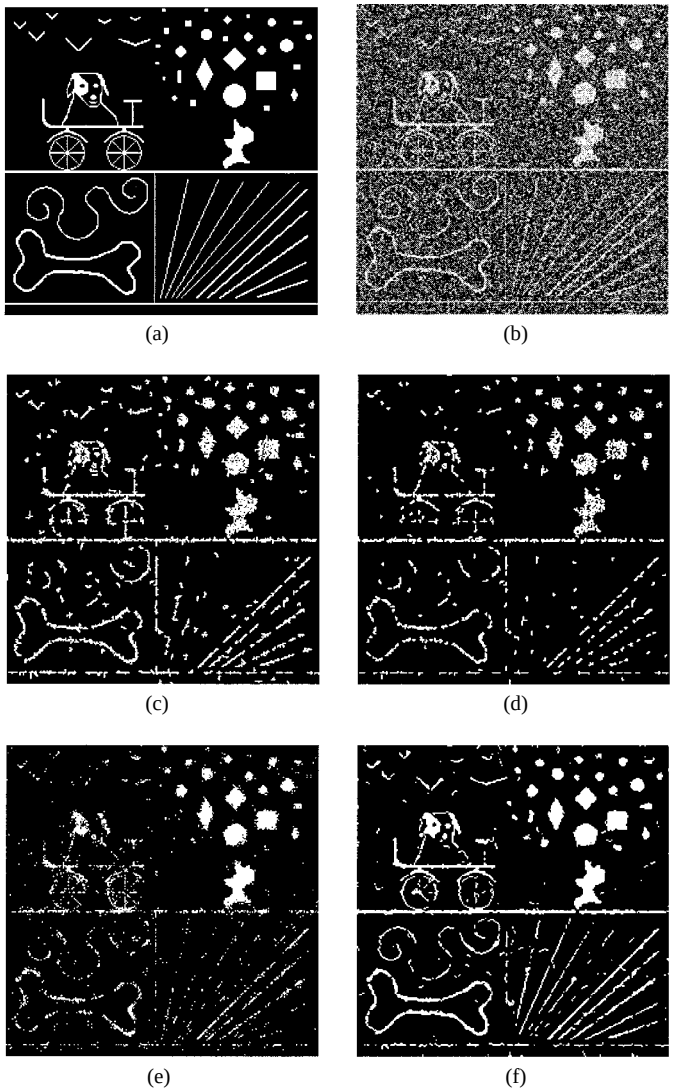

Fig. 3. Example of image segmentation on synthetic data. (a) Original data, (b) noisy data: Gaussian noise ($\sigma = 50$), (c) segmentation with a threshold ($\mathrm{th} = 128$) and deleting connected components smaller than five voxels, (d) segmentation with a threshold ($\mathrm{th} = 148$) and deleting connected components smaller than three voxels, (e) segmentation with a sequence of σ-filters followed by a threshold ($\mathrm{th} = 128$), and (f) segmentation with a fine structures preserving MRF [24].

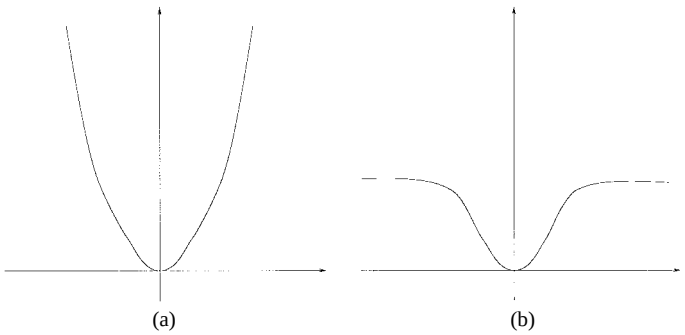

Fig. 4. Potentials for image restoration. (a) Gaussian potential and (b) Φ-model.

However, the energy associated with neighbors on both sides of an edge can be too high and the resulting restoration may result in a blurring effect despite of the data. To overcome

this effect, one approach consists in defining a line process on the dual lattice [29]. Between two neighboring pixels, a binary variable is introduced and inhibits the interaction if an edge is detected. This leads to two coupled MRF's for which the derived optimization algorithms and parameter estimation methods become complex. Therefore, some potentials have been defined to preserve edges during the restoration process. In this paper, we consider Φ-functions [30], [31] [see Fig. 4(b)]

$$V_c(y_s, y_{s'}) = \frac{-\beta}{1 + \frac{(y_s - y_{s'})^2}{\delta^2}}. \tag{8}$$

IV. An Edge Preserving Restoration Process

In this section, we derive a spatio-temporal model to restore fMRI time series. Activation areas can represent fine structures following the cerebral cortex (typically two or three voxels wide). Therefore, we pay particular attention to preserve edges in the spatial domain. Moreover, this restoration is only a preprocessing step. The aim of this preprocessing step is to increase the SNR without spoiling the signal.

A. A Spatio-Temporal MRF

We consider a time series of 2-D slices. The spatial domain is then reduced to a lattice on the plane. However, the extension to a time series of volumes is straightforward but requires more computation time. In the experiments we have managed, this approximation is justified by the gap between slices. So, we have developed a three-dimensional (3-D) restoration model where the third dimension represents time.

1) Data Attachment: The datasets have been corrected for motion-induced artifacts. The different slices have been registered onto a reference slice using an affine transformation. This affine transformation is estimated by maximizing the image cross correlation [32] with the Simplex algorithm. The slow-time dependent fluctuations have been removed using a highpass filter to correct baseline nonstationarities. We then subtract the mean of the time series and normalize the signal on each voxel to remove the fluctuations due to the anatomical structures.

For the restored data $y(s,t)$ we expect to get a value near the mean of the different $x(s,t)$. If we use Gaussian potentials for the data attachment, the energy will dramatically increase if the pixel value is set too far from the data. The first-order potentials are written as follows:

$$\forall (s,t) \in S \times \{0, \cdots, n\} V^{at}(s,t) = \alpha \big(y(s,t) - x(s,t)\big)^2. \tag{9}$$

where α is a strictly positive parameter.

2) Spatial a Priori Model: In the space domain we consider pairwise interactions induced by the 8-connectivity. As we do not want to blur the activated areas, we use edge-preserving potentials defined by a Φ-model on the pairwise cliques induced by the 8-connectivity

$$\forall s \in S, \quad \forall \epsilon \in \{(1,0),(0,1),(1,-1),(1,1)\} \quad \forall t \in \{0, \cdots, n\}$$

$$V^{\mathrm{sp}}(s,t;\epsilon) = \frac{-\beta_{\mathrm{sp}}}{1 + \delta^2 / \big(y(s,t) - y(s+\epsilon, t)\big)^2}. \tag{10}$$

where β_{sp} is a strictly positive parameter.

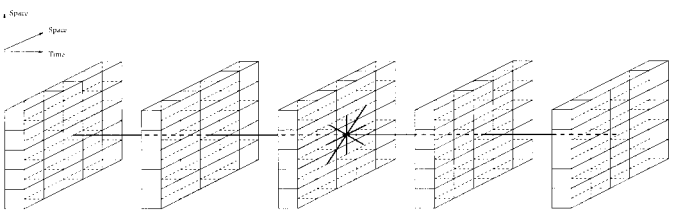

Fig. 5. The 12 neighbors of the restoration model.

3) Temporal a Priori Model: Using pairwise interaction for regularization tends to decrease the range of the first derivative. Even in case of an edge preserving potential, we smooth the data by controlling the first derivative. To study the time course of the signal we want to fit as closely as possible the original first derivatives as the hemodynamic function is defined by the first derivative of the signal. However, we want to regularize this derivative in order to regularize the time course of the hemodynamic function. We define the temporal prior to minimize the variation of the derivative, i.e., to minimize the second-order derivative $2y(s,t) - y(s,t-1) - y(s,t+1)$. So, we consider cliques of three elements to get a potential depending on the second derivative. We use Φ-potentials on these cliques which are written as follows:

$$\forall s \in S, \ \forall t \in \{1, \cdots, n-1\}, \ V^{ti}(s,t) = \frac{-\beta_{ti}}{1 + \delta^2 / \big(2y(s,t) - y(s,t-1) - y(s,t+1)\big)^2} \tag{11}$$

where β_{ti} is a strictly positive parameter.

4) Neighborhood and Energy Function: The induced neighborhood of the spatio-temporal MRF is a 12 connectivity (see Fig. 5), in which each voxel has eight spatial neighbors belonging to its own slice and four temporal neighbors located at the same place but representing previous and next instants.

The global energy function of the MRF is written as follows:

$$\begin{aligned} U = \sum_{s \in S} \sum_{t=1}^{n-1} V^{ti}(s,t) + \sum_{s \in S} \sum_{t=0}^{n} \big(& V^{\mathrm{sp}}(s,t;(1,0)) \\ & + V^{\mathrm{sp}}\big(s,t;(0,1)\big) + V^{\mathrm{sp}}\big(s,t;(1,1)\big) \\ & + V^{\mathrm{sp}}\big(s,t;(1,-1)\big)\big) + \sum_{s \in S} \sum_{t=0}^{n} V^{at}(s,t) \end{aligned} \tag{12}$$

Consider a given voxel (s,t) outside the boundaries. The local energy on (s,t) given its neighbors which is used in the SA algorithm is written as follows:

$$\begin{aligned} U_L(s,t \mid \mathcal{N}_{(s,t)}) = & \sum_{\epsilon_t \in \{-1,0,1\}} V^{ti}(s, t+\epsilon_t) \\ & + \sum_{\epsilon \in \{-1,0,1\} \times \{-1,0,1\}, \epsilon \neq (0,0)} V^{\mathrm{sp}}\big(s,t;(\epsilon_x, \epsilon_y)\big) \\ & + V^{at}(s,t). \end{aligned} \tag{13}$$

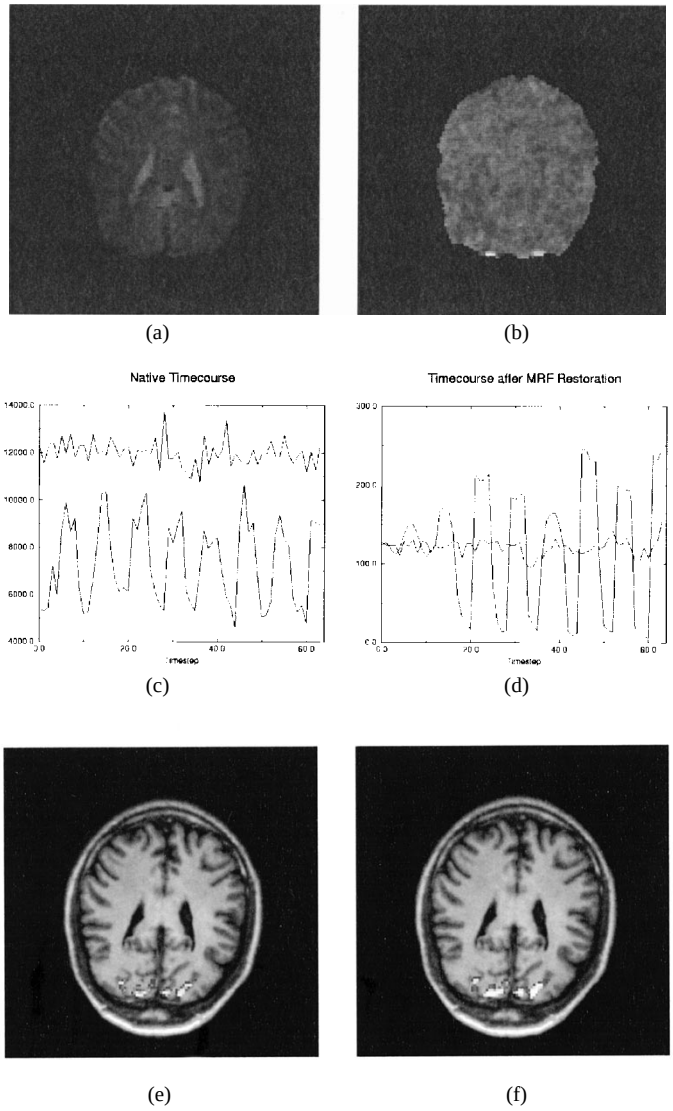

Fig. 6. Example of the restoration process of fMRI. (a) Original data (thirty-second slice), (b) restored data (thirty-second slice), (c) timecourse of the signal for an activated and a closely located nonactivated voxel, (d) timecourse at the same locations after restoration, (e) Z-map overlay of original data, and (f) Z-map overlay of restored data.

B. Restoration Results

We have performed restoration on the data sets analyzed in the next section. Fig. 6 shows one example obtained with a visual stimulation protocol. Fig. 6(a) shows a native slice. Fig. 6(b) is obtained after registration, normalization and restoration. Fig. 6(c) and (d) shows the time series of an activated and a nonactivated voxels respectively before and after the restoration. Note that the normalization changes the scale of the signal. The ratio between the activated voxel and the noise of the nonactivated voxel has been increased by the restoration process. Finally, Fig. 6(e) and (f) shows the result of a thresholded t-test before and after the restoration. The increasing of the SNR obtained by the restoration results in higher z-scores and a higher sensitivity.

V. Signal Analysis and Activation Detection

In this section, we analyze the restored data in order to obtain a classification between activated and nonactivated areas. We also extract and regularize parameters characterizing the hemodynamic function at each activated voxel. This step corresponds to the two last steps in the SPM-approach (statistical test and thresholding using spatial extend).

A. Signal Model and Parameter Maps

We consider a time series with two periodical conditions of period T referred as baseline and stimulated state. The analysis consists in detecting and characterizing any signal which appear during the stimulated periods. We make the assumption that at the end of a baseline period, the past is forgotten, namely the signal does not depend on the previous stimulated periods.

Let h_s be the hemodynamic function of the voxel s. The signal is modeled by the following convolution:

$$y_s(t) = (h_S * x)(t) + \eta_s(t) = \int_R x(\tau) h_s(t-\tau)\, d\tau + \eta_s(t) \tag{14}$$

where $\eta_s(t)$ is a Gaussian noise, $x(t)$ represents the stimulus and is given by

$$\begin{aligned} &x(t) = 1, \quad \text{if } t \in]0, T] \\ &x(t) = 0, \quad \text{if } t = 0 \quad \text{or} \quad t \in [T+1, 2T[. \end{aligned} \tag{15}$$

We make the assumption that the support of the hemodynamic function is included in the interval $[0, T]$, namely h is null outside this interval. Then we have the following equation:

$$y_s(t) = \int_0^T h_s(t-\tau)\, d\tau + \eta_s(t) \tag{16}$$

and:

$$\begin{aligned} &\frac{dy_s(t)}{dt} = h_s(t) + \frac{d\eta_s(t)}{dt}, \quad \text{if } t \in [1, T], \\ &\frac{dy_s(t)}{dt} = -h_s(t-T) + \frac{d\eta_s(t)}{dt}, \quad \text{if } t \in]T, 2T[. \end{aligned} \tag{17}$$

From the restored data we can compute the hemodynamic function at each voxel. In fact, we have two noisy estimates of $h_s(t)$ for each interval $[2nT, 2(n+1)T \quad]$, one at time t and one at time $T+t$. We compute the hemodynamic function as the average of the two estimates:

$$\begin{aligned} h_s(t) &= \frac{1}{2}\left(\frac{dy_s(t)}{dt} - \frac{dy_s(t+T)}{dt}\right) \\ &= \frac{1}{2}\big(y_s(t+1) - y_s(t) + y_s(t+T+1) - y_s(t+T)\big). \end{aligned} \tag{18}$$

We define several parameters characterizing the hemodynamic function and regularize the associated maps. First we define a map which indicates the presence of activation. This map is driven by the norm of the hemodynamic function. If this norm is lower than a threshold n_1, we consider a very low probability of any activation, if it is higher than n_2 we tend to detect activation, between both the contextual information has a higher weight in the decision. We consider three labels on the activation map. The label 0 corresponds to nonactivated voxels. The label -1 refers to detected signals

which do not match the convolution equation (16). In practice it corresponds to signals for which the maximum of the absolute value of the hemodynamic function corresponds to a negative value. This occurs if the signal has significantly decreases during the stimulated period. It can be due to a response shorter than the stimulus or to an outliar of the noise distribution. When analyzing the results, we cannot give any cognitive or physiological interpretation of these areas. They are, therefore, considered as false alarms. Finally, the label 1 corresponds to activation matching the model of (16). We also define three parameter maps to characterize the different kinds of activation. These parameters are i) the norm of the hemodynamic function, ii) the maximum of the absolute value of the hemodynamic function, and iii) the time when this maximum occurs. So we have the four following 2-D maps:

$\forall s \in S, \quad a(s) \in \{0, -1, 1\}$: activation

$\forall s \in S, \quad n(s) \in R^+$ (or grey levels set): norm

$\forall s \in S, \quad m(s) \in R^+$ (or grey levels set): maximum

$\forall s \in S, \quad i(s) \in \{1, \cdots, T\}$: time when the max occurs

B. A Two Level MRF

In this subsection we describe a MRF to regularize the four maps. The structure of the site space has two levels. The first one consists in the activation map whereas the lower level is defined by the three parameter maps (norm, maximum and time).

1) Data Attachment: From the restored volume we can compute data for each parameter map. In fact, we compute the hemodynamic function by taking the derivative of the signal and estimating the norm, the maximum of the absolute value and the time when this maximum occurs. These estimates represent the data and are denoted $d_n(s)$, $d_m(s)$ and $d_i(s)$. We define first-order potentials stating that the result should be close to the data. Therefore, we consider Gaussian potentials for the norm and the maximum maps. The time parameter depends on the time resolution of the fMRI experiment and typically does not exceed eight different values. We use a potential more adapted to this kind of state space

$$\forall s \in S, \quad V^n(s) = \alpha_m \big(n(s) - d_n(s)\big)^2 \tag{19}$$

$$V^m(s) = \alpha_m \big(m(s) - d_m(s)\big)^2 \tag{20}$$

$$V^i(s) = -\alpha_i \cos\Big(\big(i(s) - d_i(s)\big)\frac{\pi}{2T}\Big). \tag{21}$$

As stated previously the activation map can take three values (-1, 0, $+1$). The data attachment term of the activation map depends on the norm estimate and the sign of the maximum of the absolute value of the hemodynamic function. A given voxel $a(s)$ is attracted to the value zero, if the estimate norm is lower than a threshold n_1 and to -1 or $+1$ is the estimate norm is greater than n_2. Between these two values the data attachment is lower and the prior will take a higher weight to take the decision. These values define the sensitivity and the specificity of the result and can be related to the threshold used in the SPM-approach. Taking low values leads to more activation areas than for higher values. If we increase the interval between n_1 and n_2 we increase the influence of the prior model. A very localized high response will be detected as well as a smaller response on a wider area. We do not have access to a p-value for this procedure. However, some Monte Carlo simulations could provide this information. In practice, n_1 and n_2 are set by defining the percentage of activation which is considered as significant.

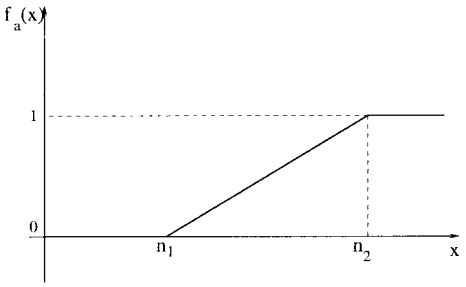

Fig. 7. Function defining the first-order potential of the activation map (see text for details).

The derived potential is written as follows:

$$\forall s \in S, \quad V^a(s) = \alpha_a\big((a(s)=0) - \big(d_m(s)<0\big)\big(a(s)=-1\big) - \big(d_m(s)>0\big)\big(a(s)=1\big)\big) f_a\big(d_n(s)\big) \tag{22}$$

where $f_a(x)$ is defined on Fig. 7 and where $(x = y)$ equal to one, if $x = y$ and to zero, otherwise.

2) The Prior Model: We now define interactions within the activation map. The activation map can be modeled by an image consisting of homogeneous areas: the background and several clusters, the activated areas. We define a prior widely used in image segmentation. Activation areas are expected to be small and finely structured. To avoid over regularization, the parameter maps are used to inhibit interactions between the activated areas and the background. Using the 8-connectivity we consider a Potts model of pairwise interactions. In this way, our model favors activated areas that are homogeneous regions. The associated potential is written as follows:

$$\forall s \in S, \quad \forall u \in \{\{-1,0,1\} \times \{-1,0,1\}\} - \{(0,0)\},$$

$$V^{\mathrm{aa}}(s,u) = -\beta_{\mathrm{aa}}\big(a(s) = a(s+u)\big). \tag{23}$$

We now have to define interactions on the parameter maps. We first address the value zero in the parameter maps to nonactivated voxels. Other values represent the variation of the parameter in activated areas. We regularize the parameter maps inside activated areas of the same class or inside nonactivated areas but do not consider interactions between activated and nonactivated voxels or between activated voxels belonging to different classes (-1 and 1).

We first define pairwise interactions between the activation map and the different parameter maps. These interactions penalize configurations for which a given voxel is different from zero on a given parameter map and is equal to zero on the activation map (nonactivated voxel). This term avoids inconsistencies between the different maps. The associated potentials are written as follows:

$$\forall s \in S, \quad V^{an}(s) = \alpha_{an}\big(a(s)=0\big)\big(n(s)\neq 0\big) \tag{24}$$

$$V^{\mathrm{am}}(s) = \alpha_{\mathrm{am}}\big(a(s)=0\big)\big(m(s)\neq 0\big) \tag{25}$$

$$V^{ai}(s) = \alpha_{ai}\big(a(s)=0\big)\big(i(s)\neq 0\big). \tag{26}$$

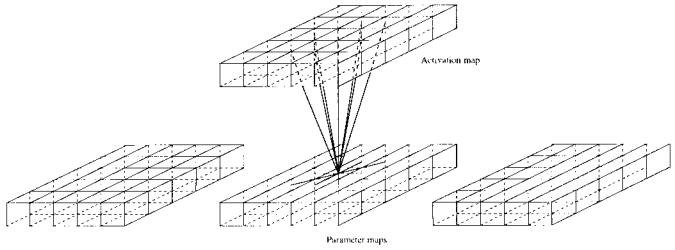

Fig. 8. Neighborhood of a parameter map site.

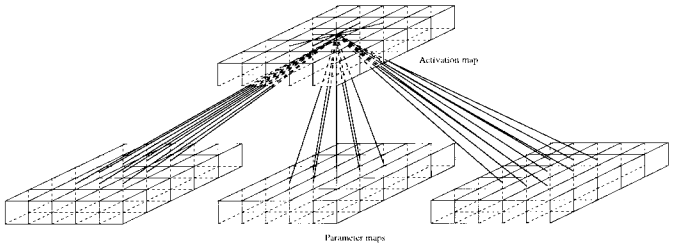

Fig. 9. Neighborhood of an activation map site.

Then we define regularizing constraints inside each parameter map. However, to avoid interactions between activated and nonactivated voxels or between activated voxels of different classes, these constraints also depend on the activation map. We perform a smoothing inside each cluster but this smoothing is not spoiled by adjacent nonactivated voxels. The norm and the maximum parameters lie in the same scale defined by the grey levels. The potentials associated with these two parameters are of the same kind and are chosen Gaussian to get a strong smoothing inside a given activated area. We can consider such a smoothing because we have inhibited the interactions between the activation areas and the background. As for the data attachment term, the potential associated with the time parameter is adapted to a state space containing few values

$$\forall s \in S, \quad \forall u \in \{-1,0,1\} \times \{-1,0,1\} - \{(0,0)\}$$

$$V^{\mathrm{na}}(s, s+u) = \beta_{\mathrm{na}}\big(n(s) - n(s+u)\big)\big(a(s) = a(s+u) \neq 0\big) \tag{27}$$

$$V^{ma}(s, s+u) = \beta_{ma}\big(m(s) - m(s+u)\big)\big(a(s) = a(s+u) \neq 0\big) \tag{28}$$

$$V^{ia}(s, s+u) = -\beta_{ia}\cos\Big((i(s) - i(s+u))\frac{\pi}{2T}\Big)\big(a(s) = a(s+u) \neq 0\big).$$

The induced neighborhood depends on the map. Fig. 8 shows the neighborhood of a voxel belonging to a parameter map, and Fig. 9 shows the neighborhood of a voxel belonging to the activation map.

VI. Results on fMRI

We have validated these new methods on several data sets and for several protocols. The data have been obtained with a 3.0 Tesla Medspec 30/100 scanner from Bruker Medizintechnik GmbH (Ettlingen, Germany). Experiments have been performed at the Max Planck Institute of Cognitive Neuroscience in Leipzig. The functional data sets contain 64 two-dimensional (2-D) scans of a gradient-echo FLASH sequence with TR $= 80.5$ ms, TE $= 40$ ms and FOV $= 25$ cm [33], [34]. The original matrices contain 128×64 voxels and have been zero-filled before the Fourier transform to yield a spatial resolution of 1.93 mm $\times$ 1.93 mm and a slice thickness of 5 mm with 2 mm gap. The overlays are shown on 2-D high resolution 512×512 anatomical slices acquired with a T_1-weighted MDEFT sequence [35].

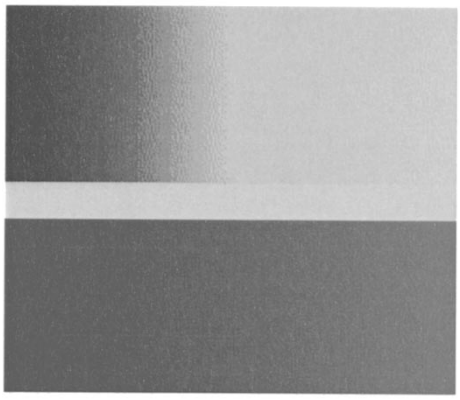

Fig. 10. Coding colors for the two activation classes (green: label -1, orange label 1). Scale: min (red, green): 2.0, max (white, blue): 6.0 (z-value for t-test and regression analysis). Scale: min (red, green): 2%, max (white, blue): 4% (percentage for activation for the MRF analysis).

For each dataset we have 64 time samples with 5-s intervals. Each period contains four samples under simulation (20 s) and four baseline samples (20 s). Eight such periods make up a time series. These experiments have been performed by healthy volunteers from 20 to 30 yr. old.

The look up tables corresponding to the two activation classes are shown on Fig. 10. The green scale corresponds to the label (-1), i.e., to signals which do not match the convolution model, whereas the red-yellow-white scale is used to code the activation matching the time course of the stimulation protocol. For the SPM approaches the red-yellow-white scale represents the z-value from 2.0 (red) to 6.0 (white). The MRF approach results show the norm of the activation which has been scaled to fit the same look-up table.

For the three experiments we have compared the proposed method with the SPM-approach. We first have used a t-test on original data or on data filtered with a Gaussian filter of variance one when no significant activation was detected on original data. The activation is obtained by thresholding the derived z-map using spatial extend. We have also compared the results with a regression method based on a Gaussian model of the hemodynamic function taking into account a correction for lag and dispersion [36]. The signal is modeled by the convolution of a Gaussian (the hemodynamic function) with the stimulation function plus an independent additive Gaussian noise. The mean of the hemodynamic function (referred as to the lag), the standard deviation (referred as to the dispersion) and the amplitude are estimated using a least square procedure. This method has been applied on original data.

A. Applicability of the Methods

Herein we only consider periodical ON-OFF experimental designs, for which the three compared procedures can be applied, and address the quality of the results. However, the application range of the procedures is of importance for future designs. The t-test compares two distributions and makes no assumptions about the time course of the signal. It is not applicable if more than two conditions exist in the experimental design. The regression analysis may be combined with hemodynamic modeling to improve the specificity and sensitivity of fMRI signal detection. In addition, it can take several experimental conditions into account. The modeling procedure, however, requires a certain number of timesteps in a period of the experimental design (a trial) to allow a successful adaption of the model function. In the proposed approach we make the assumption of a periodical ON-OFF sequence. However, this assumption is only required for the definition of the hemodynamic response in (14), as we consider the convolution with a step function. The Markovian approach can be used with more complex experimental designs. We currently investigate the restoration of the signal for different experimental designs using MRF's.

B. Visual Protocol

Subjects were asked to fixate a central point on an 8-Hz alternating checker board pattern projected by a LCD system. Three axial slices have been considered, the central slice corresponding to the visual cortex. Fig. 11 shows the results on two slices which cut the visual cortex approximately at 40° from the AC-PC plane. There is a cross validation of the three methods as the results are consistent between the different methods. Using a t-test, the filtering step is not necessary to detect activation. However, to obtain a good detection we have used a low value for the threshold $(t = 2.0)$ resulting in false alarms in the precentral sulcus [see Fig. 11(a)]. This activation is considered as false positive because there is no neurophysiological reason for a precentral activation in such a simple visual task. Indeed, none of the more specific methods shows this activation. We could have used Gaussian filtering to increase the threshold, however, this leads to blurred activated areas. Using a regression analysis allows us to increase the threshold $(t = 3.5)$ without detecting false alarms. In the proposed methods we get a higher activation [see Fig. 11(f)]. Note that "activated" voxels in extra cerebral compartments are classified as false alarms, as for the sinus rectus on Fig. 11(f).

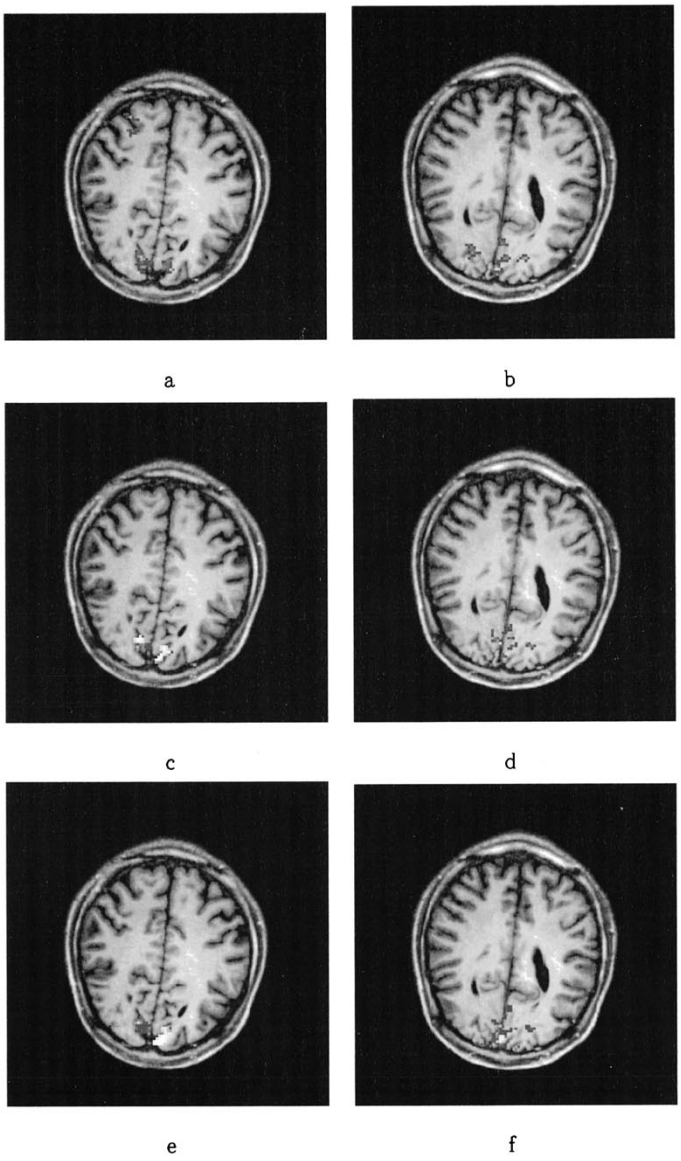

Fig. 11. Visual experiment.t-test on original data $(a+b)$, regression analysis on original data $(c + d)$, MRF analysis on restored data $(e + f)$.

C. Motor Protocol

A finger tapping task was performed by the subjects with a rapid repetitive opposition movement of the first two fingers. Fig. 12 shows results on two axial slices 36 with respect to 43 mm above the AC-PC plane. In this motor task, activation is expected in the motor cortex, the ipsilateral sensory cortex and the SMA. Only the MRF method detects these three areas. Gaussian filtering leads to activated areas badly located and spreading into the white matter [see Fig. 12(b)]. Using MRF allows us to increase the sensitivity with respect to the regression analysis as most of the false alarms detected are classified in green. We get a better activation in the motor cortex and also detect some activation in the supplementary motor area (SMA) [see Fig. 12(e)].

D. Word Protocol

A random series of words and pronounceable nonwords was heard by the subjects during the stimulation period. Their task was to count the nonwords during the stimulation. Two sagittal slices through the center of the temporal lobe on the left side are shown on Fig. 13. For this experiment, the z-map are not too noisy and Gaussian filtering can be avoided. The underlying signal modeling of the regression analysis and the Markovian approach allows us to increase the sensitivity with respect to the t-test. Lower activation along the sulcus temporalis inferior (STI) is detected with the regression analysis [see Fig. 13(c) and (d)].

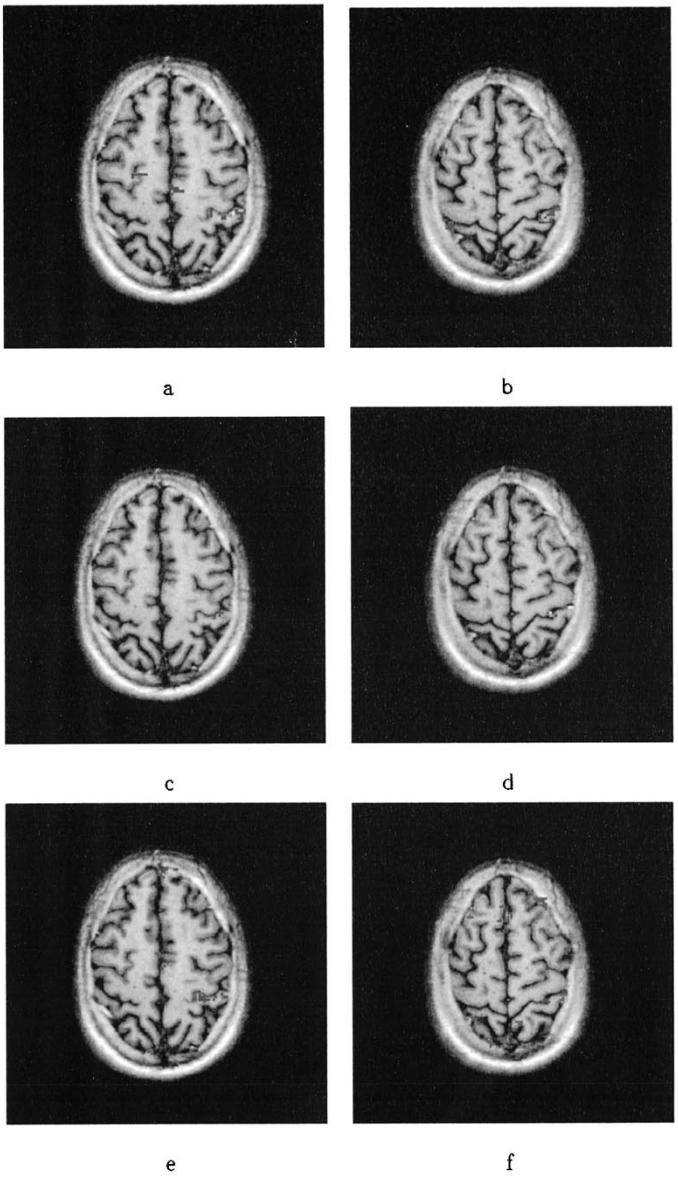

Fig. 12. Finger tapping experiment. t-test on filtered data $(a+b)$, regression analysis on original data $(c+d)$, MRF analysis on restored data $(e+f)$.

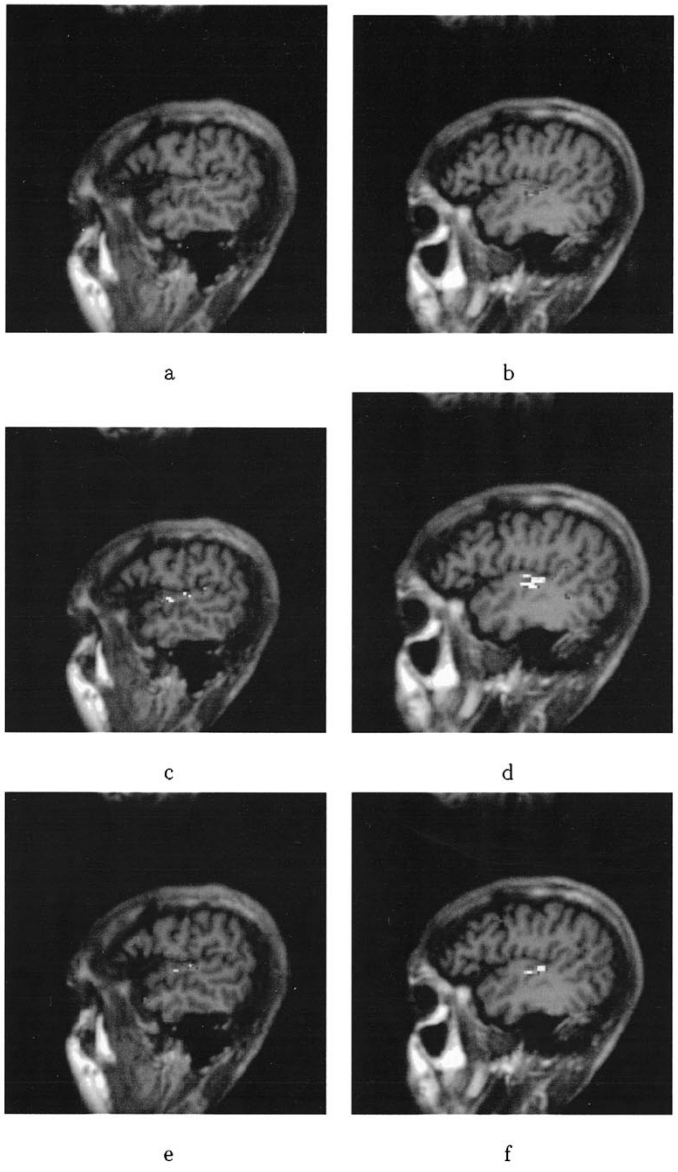

Fig. 13. Word recognition experiment. t-test on original data $(a+b)$, regression analysis on original data $(c+d)$, MRF analysis on restored data $(e+f)$.

VII. Conclusion

In this paper we have proposed a new approach to restore and analyze fMRI data using MRF's in a Bayesian framework and showed that MRF's and more generally tools from image processing can represent a good alternative to the SPM approach. We have applied this approach in two steps. Firstly, we apply a signal restoration algorithm to improve the SNR. We define a 3-D MRF where the third dimension represents time. We use some Φ-functions as potentials in the space domain in order to preserve discontinuities due to edges. After this restoration step, one can apply the SPM approach. The sensitivity of the detection is improved as the SNR is increased but, contrary to low pass filters, the MRF restoration dœs not spoil the signal. One possible extension of this model can be to incorporate a convolution in the associated inverse problem. In that way, we expect to improve the resolution of the initial data in the space domain. The open question concerns the model of the convolution kernel.

Secondly, we analyze the signal to detect and characterize activated voxels. Using MRF's allows us to perform the detection in the same process as for the analysis. We do not make any assumption about the shape of the response of the activated voxels. Statistical test or correlation using the constant-wise function of baseline and activated periods as a reference are only sensitive to a mean activation during the stimulation periods. They lack robustness with respect to outliers. Analyzing the signal after the detection can be valid only if activated voxels are really detected. Moreover, MRF's allow us to detect lower signals as they manage some contextual information. The analysis of the hemodynamic response is achieved using three parameters, the norm, the maximum and the time when this maximum occurs. The

analysis of these three parameters depends of course on the activation detection, but we consider that the activation also depends on these parameters. When the values of these parameters in neighboring voxels are far from each other, the probability of detection is lower as the associated hemodynamic responses are not consistent in the spatial domain. Therefore, we consider a two level MRF modeling interactions between the activation map and the three parameter maps. The detection of an activated area, thus, depends on the norm of the hemodynamic response and some contextual information on this norm but also on the consistency of the hemodynamic function parameters along this area.

Our current research concerns the analysis of more complex experimental protocols than ON-OFF sequences and the classification of the activated voxels with respect to the shape of their hemodynamic function. We work on the definition of several different classes. A first class may characterize the BOLD effect in vessels (veins) and some other classes may represent different types of neuronal responses.

ACKNOWLEDGMENT

The authors wish to thank Dr. D. Norris for providing the data sets.

REFERENCES

[1] J. W. Belliveau, D. N. Kennedy, R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff, M. S. Cohen, J. M. Viver, T. J. Brady, and B. R. Rosen, "Functional mapping of the human visual cortex by magnetic resonance imaging," *Sci.*, vol. 254, pp. 716–719, 1991.

[2] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," in *Proc. Nat. Acad. Sci. USA 87*, 1990, pp. 9868–9872.

[3] A. Villringer and U. Dirnagl, "Coupling of brain activity and cerebral blood flow: Basis of functional neuroimaging," *Cerebrovasc. Brain Metab. Rev.*, vol. 7, no. 3, pp. 240–276, 1995.

[4] 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.

[5] K. J. Friston, J. Ashburner, A. Holmes, J. B. Poline, R. Turner, R. Wise, C. Price, R. Dolan, P. Fletcher, J. Coull, and C. Frith, "Statistical parametric mapping and functional neuroimaging," presented at *Inst. Neurol. Short Course*, Functional Imaging Laboratory, London, U.K., May 17–18, 1996.

[6] J. Xiong, J. H. Gao, J. L. Lancaster, and P. T. Fox, "Assessment and optimization of functional MRI analyses," *Human Brain Mapping*, vol. 4, pp. 153–167, 1996.

[7] K. J. Worsley, A. C. Evans, S. Marret, and P. Neelin, "A three-dimensional statistical analysis for CBF activation studies in Human Brain," *Journal of Cerebral Blood Flow and Metabolism*, vol. 12, pp. 900–918, 1992.

[8] S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, and D. C. Noll, "Improved assessment of significant activation in fMRI: Use of a cluster-size threshold," *Magn. Reson. in Med.*, vol. 33, pp. 636–647, 1995.

[9] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans, "A unified statistical approach for determining significant signals in images of cerebral activation," *Human Brain Mapping*, pp. 58–73, 1996.

[10] N. Lange, "Tutorial in biostatistics, statistical approaches to human brain mapping by fMRI," *Statistics in Med.*, vol. 15, pp. 389–428, 1996.

[11] K. Sekihara and H. Koizumi, "Detecting cortical activities from fMRI time-course data using the MUSIC algorithm with forward and backward co-variance averaging," *Magn. Reson. in Med.*, vol. 35, pp. 807–813, 1996.

[12] J. B. Poline and B. M. Mazoyer, "Analysis of individual brain activation maps using hierarchical description and multiscale detection," *IEEE Trans. Med. Imag.*, vol. 13, pp. 702–710, Aug. 1994.

[13] E. Bullmore, M. Brammer, S. C. R. Williams, S. Rabe-Hesketh, N. Jannot, A. David, J. Mellers, R. Howard, and P. Sham, "Statistical methods of estimation and inference for functional MR image analysis," *Magn. Reson. in Med.*, vol. 35, pp. 261–277, 1996.

[14] J. Besag, "Spatial interaction and statistical analysis of lattice systems," *Acad. Roy. Statistical Soc. Series B*, vol. 36, pp. 721–741, 1974.

[15] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, June 1984.

[16] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 39–55, Jan. 1987. <Author: Is no. 9 correct? (Jan or Sept.)

[17] R. Chellappa and A. Jain, *Markov Random Fields: Theory and Application*. New York: Academic, Collective Work, 1993.

[18] A. Holmes and I. Ford, "A Bayesian approach to significance testing for statistic images from PET," in *Quantification of Brain Function. Tracer Kinetics and Image Analysis in Brain PET*, B. V. K. Uemura, N. Lassen, T. Jones, and I. Kanno. Amsterdam, the Netherlands: Elsevier, 1993.

[19] J. S. Lee, "Digital image smoothing and the sigma filter," *Comput. Vision, Graphics, Image Processing*, vol. 24, pp. 255–269, 1983.

[20] G. Palubinskas, "A comparative study of smoothing filters in magnetic resonance images," in *Proc. IASTED (SIP'96)*, Orlando, FL, 1996, pp. 155–159.

[21] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of fMRI time-series," *Human Brain Mapping*, vol. 2, pp. 49–78, 1994.

[22] K. J. Worsley and K. J. Friston, "Analysis of fMRI time-series revisited again," *NeuroImage*, vol. 2, pp. 173–181, 1995.

[23] J. B. Poline, K. J. Worsley, A. C. Evans, and K. J. Friston, "Combining spatial extend and peak intensity to test for activations in functional imaging," *NeuroImage*, vol. 5, pp. 83–96, 1996.

[24] X. Descombes, J. F. Mangin, E. Pechersky, and M. Sigelle, "Fine structures preserving model for image processing," in *Proc. 9th SCIA 95 Uppsala*, Sweden, 1995, pp. 349–356.

[25] X. Descombes, R. Morris, and J. Zerubia, "Quelques améliorations la segmentation d'images Bayesienne Première partie: Modelisation," *Traitement du Signal*, vol. 14, no. 4, pp. 373–382, 1997.

[26] H. O. Georgii, "Gibbs measures and phase transitions," *De Gruyter—Studies in Mathematics*, vol. 9, 1998.

[27] X. Guyon, *Champs aléatoires sur un réseau : Modelisations, statistique et applications, Techniques Stochastiques*, Editions Masson, 1992.

[28] J. M. Hammersley and P. Clifford, "Markov random fields on finite graphs and lattices," unpublished, 1971.

[29] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 609–628, July 1990.

[30] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," in *Proc. 46th Session of ISI*, 1987, pp. 22–26.

[31] S. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, Mar. 1992.

[32] K. J. Friston, S. Williams, R. Howard, R. S. J. Frackowiak, and R. Turner, "Movement-related effects fMRI timeseries," *Magn. Reson. in Med.*, vol. 35, pp. 346–355, 1996.

[33] A. Haase, J. Frahm, D. Matthaei, W. Hänicke, and K. D. Merboldt, "Flash imaging. Rapid NMR imaging using low flip-angle pulses," *J. Magn. Reson.*, vol. 67, pp. 258–266, 1986.

[34] K. Ugurbil, M. Garwood, J. Ellermann, K. Hendrich, R. Hinke, X. Hu, S. G. Kim, R. Menon, H. Merkle, S. Ogawa, and R. Salmi, "Imaging at high magnetic fields: Initial experiences at 4T," *Magn. Reson. Quart.*, vol. 9, pp. 259–277, 1993.

[35] J. H. Lee, M. Garwood, R. Menon, G. Adriany, P. Andersen, C. L. Truwit, and K. Ugurbil, "High contrast and fast three-dimensional magnetic resonance imaging at high fields," *Magn. Reson. in Medicine*, vol. 34, pp. 308–312, 1995.

[36] J. Rajapakse, F. Kruggel, J. Maisog, and Y. von Cramon, "Modeling hemodynamic response for analysis of function MRI time-series," submitted for publication.