Contextual Clustering for Analysis of Functional MRI Data

Eero Salli*, *Student Member, IEEE*, Hannu J. Aronen, Sauli Savolainen, Antti Korvenoja, and Ari Visa, *Senior Member, IEEE*

***Abstract*—We present a contextual clustering procedure for statistical parametric maps (SPM) calculated from time varying three-dimensional images. The algorithm can be used for the detection of neural activations from functional magnetic resonance images (fMRI). An important characteristic of SPM is that the intensity distribution of background (nonactive area) is known whereas the distributions of activation areas are not. The developed contextual clustering algorithm divides an SPM into background and activation areas so that the probability of detecting false activations by chance is controlled, i.e., hypothesis testing is performed. Unlike the much used voxel-by-voxel testing, neighborhood information is utilized, an important difference. This is achieved by using a Markov random field prior and iterated conditional modes (ICM) algorithm. However, unlike in the conventional use of ICM algorithm, the classification is based only on the distribution of background. The results from our simulations and human fMRI experiments using visual stimulation demonstrate that a better sensitivity is achieved with a given specificity in comparison to the voxel-by-voxel thresholding technique. The algorithm is computationally efficient and can be used to detect and delineate objects from a noisy background in other applications.**

***Index Terms*—Clustering, functional magnetic resonance imaging (fMRI), hypothesis testing, segmentation, statistical parametric map.**

Manuscript received April 28, 1999; revised March 9, 2001. This work was supported by the Jenny and Antti Wihuri Foundation, by the Foundation of Technology in Finland, by the National Technology Agency, by the Radiological Society of Finland, by the Academy of Finland, by the Sigrid Jusélius Foundation, by the Paavo Nurmi Foundation, by the Cancer Organizations of Finland, by the Clinical Research Institute Helsinki University Central Hospital, and by the Helsinki University Central Hospital EVO under Grant TYH-0313, Grant TYH-8102, and Grant TYH-9102. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was V. Johnson. *Asterisk indicates corresponding author.*

*E. Salli is with the Laboratory of Biomedical Engineering, Helsinki University of Technology, P.O.B. 2200, FIN-02015 HUT, Espoo, Finland, and the Department of Radiology, Helsinki University Central Hospital, FIN-00029 HUS, Helsinki, Finland (e-mail: eero.salli@hut.fi).

H. J. Aronen is with the Department of Radiology, Helsinki University Central Hospital, FIN-00029 HUS, Helsinki, Finland, the Laboratory of Biomedical Engineering, Helsinki University of Technology, FIN-02015 HUT, Espoo, Finland, and the Department of Clinical Radiology, Kuopio University, FIN-70211 Kuopio, Finland.

S. Savolainen is with the Departments of Radiology and Laboratory Medicine, Helsinki University Central Hospital, FIN-00029 HUS, Helsinki, Finland.

A. Korvenoja is with the Department of Radiology and with the BioMag Laboratory, Helsinki University Central Hospital, FIN-00029 HUS, Helsinki, Finland.

A. Visa is with the Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland.

Publisher Item Identifier S 0278-0062(01)04404-4.

I. Introduction

CURRENT tomographic technologies in medical imaging enable noninvasive studies of brain function by measuring hemodynamic changes related to changes in neuronal activity. The signal changes observed in functional magnetic resonance imaging (fMRI) are mostly based on blood oxygenation level dependent (BOLD) contrast and are usually close to the noise level [1], [2]. Consequently, statistical methods and signal averaging are frequently used to distinguish signals from noise in the data. In most fMRI setups, images are acquired during alternating task (stimulus) and control (rest) conditions.

The analysis of the image series is frequently based on the computation of a statistical parametric map, and statistical inferences derived from it. For example, a voxel-by-voxel computation of the difference of means of intensities between control and task states normalized by the estimated standard error generates a statistical map that follows the t distribution in the nonactive area, i.e., in the background. Correlation analysis [3]–[5], subspace modeling [6], Fourier, and wavelet transform methods [3], [7], [8], pseudogeneralized least squares analysis using sinusoidal regression [9] and nonparametric Kolmogorov–Smirnov test [2] are examples of other approaches used to create statistical maps. The general linear model [10] is a general framework that includes the simple t statistic and most other parametric tests. Significant active areas are found by thresholding the maps. Methods that assess statistical significance levels based on the spatial extent of the activation cluster after intensity thresholding have been developed to improve sensitivity [11], [12]. Because the spatial extent must exceed a threshold, these methods are occasionally known as dual-parameter thresholding (DPT) techniques [13]. It should be noted that the fMRI time series may be temporally correlated and that the general linear model has been extended to deal with temporal correlations [14], [15]. Several preprocessing steps such as motion correction and temporal filtering are frequently performed before the data analysis. In particular, spatial filtering is frequently used to increase signal-to-noise ratio and validity of inferences based on the theory of Gaussian fields [13], [15], [16].

In addition to the above mentioned inferential data analysis approaches, several methods that emphasize the exploratory nature of the problem have been proposed. These methods include independent spatial component analysis [17], principal component analysis [18], [19], and clustering of the time series [20]–[24].

A single activation region typically consists of numerous voxels. Hence, it may be useful to utilize contextual informa-

tion [25]. In texture segmentation, pixel features are acquired from a pixel neighborhood (see, e.g., [26]). A simple example of a method that utilizes classification information from a voxel's neighborhood is the median filtering of the thresholded image [27]. In addition, the DPT techniques and spatial filtering can be seen as methods that incorporate contextual information into the analysis. An interesting option is to use the intensity value of a pixel and classification information from the pixel neighborhood in the same stage of the classification [25], [28], [29]. This approach can be realized in the context of Markov random fields (MRF) [28]–[31]. MRFs regularize a classification by defining interactions between neighboring pixels. This idea has previously been proposed in the functional brain-imaging context by Holmes [32]. An MRF model is also used for the spatio-temporal analysis of fMRI data in [33] and [34].

In this paper, we present an iterative contextual analysis method based on MRFs. The first step in the method is to compute a statistical parametric map using well-known results from general linear models. Then, the statistical parametric map is clustered, i.e., segmented into nonactive and active regions. The contextual clustering algorithm is based on Besag's iterated conditional modes (ICM) algorithm [29]. However, in our approach the ICM algorithm is modified for performing hypothesis testing by defining an artificial activation class. The null hypothesis is that a voxel is nonactive. If the null hypothesis is rejected, the voxel is considered active. Simulations are used to find false positive (i.e., false activation) rates for varying parameter values of the algorithm. This allows classification so that the probability of false positives is controlled.

In Section II, the contextual clustering rule is derived. In Section III, the method is tested using simulated and measured fMRI data to demonstrate the usefulness of the approach. The method is discussed and conclusions drawn in Section IV.

II. Algorithm

A. Computation of a Statistical Parametric Map

In considering the contextual clustering algorithm, a statistical parametric map can be computed using any existing method as long as the resulting distribution in nonactive voxels can be transformed into a standard Gaussian distribution. For the purposes of this study, an unpaired Student's t statistics with pooled standard error is adequate [35]. Let $I = \{i_1, i_2, \ldots, i_n\}$ be the set of sites, or voxels, of the statistical parametric map. In voxel $i \in I$, the test statistics can be defined as shown in (1) at the bottom of the page, where

- $\overline{X}_{iA}$ and $\overline{X}_{iB}$ means of the observed intensity values in the corresponding voxels of the original time-varying image;
- $\hat{\delta}^2_{iA}$ and $\hat{\delta}^2_{iB}$ estimated variances;
- n_A and n_B the sizes of the task state (subindex A) and control state sets (subindex B).

It is assumed that the signal level rises during activation and, therefore, the expectation of r_{it} is negative on active regions.

In nonactive voxels, it is reasonable to assume that intensity values acquired during the task state and control state derive from the same Gaussian distribution. Hence, the t values follow the t distribution with $n_A + n_B - 2$ DOF. If $n_A + n_B - 2$ is large, the standard normal distribution approximation may be reasonable. Otherwise, t distributed values can be readily transformed to Gaussian distributed values. By this we mean a process whereby new values r_i are computed from the original values r_{it} using

$$r_i = q_{\mathrm{norm}}[p_t(r_{it}, df)], \tag{2}$$

where $q_{\mathrm{norm}}(x)$ is the normal inverse distribution function at x and $p_t(x, df)$ is the cumulative distribution function for t distribution with df DOF at x. We will refer to a statistical parametric map computed using (1) a t map and a map transformed using (2) a z map. The transformation from the raw data values to the t values is illustrated in Fig. 1. It should be noted that, after the transformation (2), only the background region (area outside activations) follows a Gaussian distribution. However, this is adequate because only the background distribution is needed to control false positive rates.

B. Contextual Clustering

Following the generation of a z map, the goal is to recognize the active regions from the background. The statistical properties of activation classes are unknown but a statistical characterization of the background class is available. A hypothesis testing approach is used to deal with this kind of classification problem. In the present case, the null hypothesis is that the voxel under consideration is from the background class.

In a standard one-tailed test of a null hypothesis, the null hypothesis would be rejected, and a voxel would be classified as active, if the statistical value of a voxel were significantly smaller than zero. The method presented here is different because the spatial information from the voxel neighborhood is additionally used.

In order to develop a contextual testing method, a hypothetical activation class, or rejection class, is defined. The class determines a critical region corresponding to an "unlikely event". Voxels whose characteristics are "closer" to the rejection class than the background class violate the null hypothesis and are put into the rejection class and considered active.

The background class is known to follow standard normal $(N(0,1))$ distribution from (1) and (2). Here, $N(\mu, \sigma^2)$ defines a Gaussian distribution with mean μ and variance σ^2. We must set certain requirements for the rejection class. At this stage,

$$r_{it} = \frac{\overline{X}_{iB} - \overline{X}_{iA}}{\sqrt{(((n_A - 1)\hat{\delta}^2_{iA} + (n_B - 1)\hat{\delta}^2_{iB})/(n_A + n_B - 2))((1/n_A) + (1/n_B))}}, \tag{1}$$

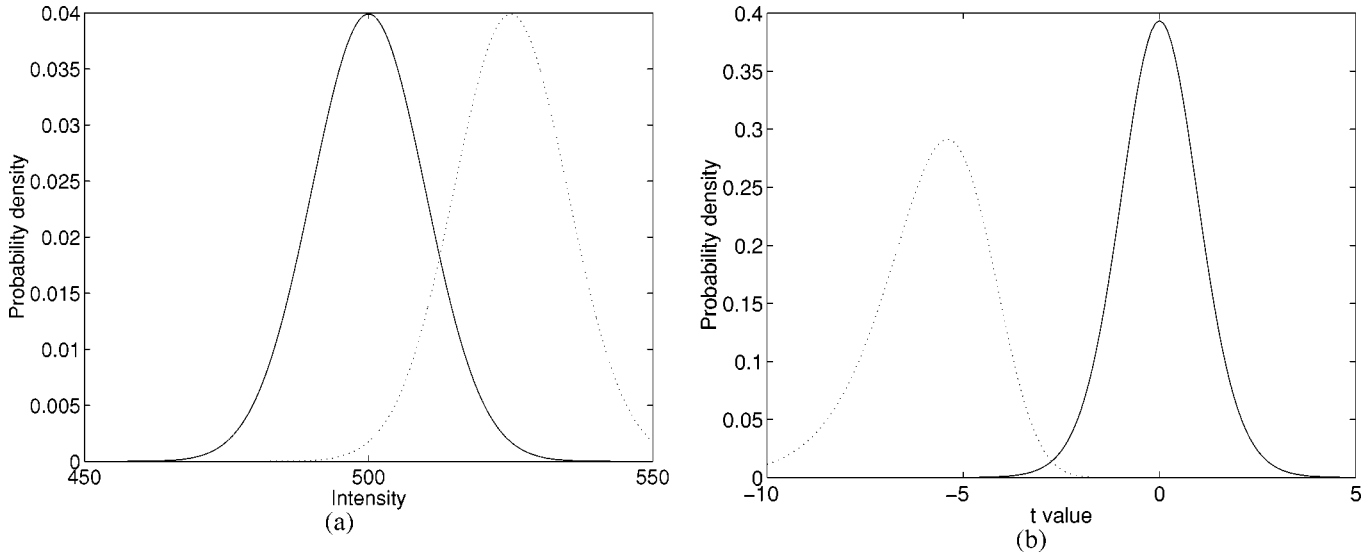

Fig. 1. (a) Probability density functions of fictional Gaussian control state data centered at value 500 (solid line) and task state data centered at value 525 (dotted line). $\delta_A = \delta_B = 100$. (b) Probability density functions of the background voxels (solid line) and of the activation voxels (dotted line) after transformation to t statistics when the degrees of freedom (DOF) is 18. Background voxels follow t distribution and activation voxels noncentral t distribution. "Gaussianization" can be used to transform the t distributed values so that the background voxels follow Gaussian distribution. The difference between the means of the background and activation distributions after the "Gaussianization" is defined to be the separation value $s_\circ$.

suppose that there is no contextual support toward either accepting or rejecting the null hypothesis. Then, in a conventional hypothesis testing, a null hypothesis is rejected at voxel i if r_i is smaller than a user specified threshold a. This is equivalent to the maximum likelihood classification when the background class follows $N(0,1)$ and reject class follows $N(2a,1)$. It is assumed, therefore, that the rejection class follows a $N(2a,1)$ distribution, which guarantees that the rejection probability will be a monotone increasing function of a. The definition of the rejection class is required to formulate the testing problem as a classification problem. The detection problem is now reduced to a two class-segmentation problem that can be solved using standard methods. The ICM algorithm by Besag will be adopted [29]. The derivation of the ICM is now briefly reviewed and after that, it is adapted to the case under discussion, in which only the background distribution is known.

A neighborhood sphere of order q of voxel i is defined as

$$N_i^q = \{j \in I : 0 \leq (i_x - j_x)^2 + (i_y - j_y)^2 + (i_z - j_z)^2 \leq q\}, \tag{3}$$

where i_x, i_y, i_z and j_x, j_y, j_z define the x, y, and z coordinates of the voxel sites i and j. Voxels whose center point is inside the neighborhood sphere, excluding voxel i itself, define a qth order neighborhood. In three dimensions, the first-order neighborhood consists of the six nearest neighbors of each voxel, the second-order neighborhood consists of the eighteen nearest neighbors, and third-order neighborhood consists of the 26 nearest neighbors. In this paper, the third-order neighborhood is used, although other neighborhoods are also possible.

Certain definitions and the concept of the MRF [29]–[31] are presented now. Let $\mathbf{K}$ be the random variable representing the true classification of the image, and $\mathbf{k}$ its realization. Let K_i be the random variable representing the true classification at voxel i, and k_i its realization. Let $m = 0$ represent the background class and $m = 1$ the activation class. Correspondingly, let $k_i = 0$ represent the case that the voxel i belongs to the background and $k_i = 1$ the case that the voxel i is active. Let $f(r_i|k_i)$ be the conditional class density that the voxel i has value r_i, given that the true class is k_i. Let $P(k_i|\mathbf{k}_{I\setminus i})$ be the conditional prior probability (i.e., the probability before the measurement information r_i is considered) of the class k_i at voxel i when the classification in all other voxel locations $\mathbf{k}_{I\setminus i}$ is given. Let $\mathbf{k}_{\partial i}$ represent the classification in the neighborhood of voxel i and $P(k_i|\mathbf{k}_{\partial i})$ be the conditional prior probability of the class k_i at voxel i. Random variable $\mathbf{K}$ is a (locally dependent) MRF if

$$P(K_i = k_i|\mathbf{k}_{I\setminus i}) = P(K_i = k_i|\mathbf{k}_{\partial i}). \tag{4}$$

Hence, the fundamental idea behind the MRF in activation detection is that the prior probability distribution for activation at a voxel is specified conditionally on the activation pattern in the neighborhood of that voxel.

Let $\mathbf{r}$ represent the vector of all r_is. It is assumed that the values of the statistical parametric map are conditionally independent so that the conditional density of $\mathbf{r}$ is

$$l(\mathbf{r}|\mathbf{k}) = \prod_{i=1}^{n} f(r_i|k_i). \tag{5}$$

Generally, in the maximum *a posteriori* probability based segmentation of the image $\mathbf{r}$, the segmentation $\mathbf{k}$ is selected so that the *a posteriori* probability $P(\mathbf{k}|\mathbf{r})$ is globally maximized. This problem can be solved by applying Bayes' rule and maximizing

$$P(\mathbf{k}|\mathbf{r}) \propto l(\mathbf{r}|\mathbf{k})p(\mathbf{k}) \tag{6}$$

where $p(\mathbf{k})$ represents prior information that is in our case a Markovian prior.

The ICM algorithm can be used to find a local maximum of (6) through the following steps: 1) Initial classification is chosen so that $f(r_i|k_i)$ is maximized at each voxel i separately; 2) the

classification is updated at each voxel so that the new class has maximum conditional probability, given the value r_i and current classification in the neighborhood of voxel i, that is $\hat{\mathbf{k}}_{\partial i}$. The probability to be maximized at voxel i is

$$P(k_i|\mathbf{r}, \hat{\mathbf{k}}_{I\setminus i}) \propto f(r_i|k_i)p_i(k_i|\hat{\mathbf{k}}_{\partial i}). \tag{7}$$

Step 2) is repeated until the classification does not change or begins to oscillate between states.

The Hammersley-Clifford theorem [30] states that an MRF can be written as a Gibbs field, provided that no segmentation has a probability equal to zero, as follows:

$$P(\mathbf{k}) = \frac{1}{Z}\exp[-U(\mathbf{k})] = \frac{1}{Z}\exp\left[-\sum_{c\in C} V_c(\mathbf{k})\right] \tag{8}$$

where

U energy function;
C set of cliques;
V_c potential associated with the clique c,
Z being the normalization constant.

A clique is a set of voxels that are neighbors of one another. We restrict our attention to the potentials between two neighboring voxels, i.e., to the doubletons. One-site cliques, i.e., the singletons, can be ignored by assigning the value zero to their potential functions. Let the potential be zero when both voxels of a clique belong to the same class, and β otherwise. Then, consider two realizations which differ only at voxel i. From (8), the conditional prior probability of class m at voxel i is derived as

$$p_i(m|\mathbf{k}_{\partial i}) \propto \exp\left[\beta u_i(m)\right] \tag{9}$$

where $u_i(m)$ is the number of neighbors of the voxel i having class m. The class densities are defined to be Gaussian, i.e.,

$$f(r_i|m) = \frac{1}{\sqrt{2\pi\sigma_m^2}}\exp\left[\frac{-(r_i-\mu_m)^2}{2\sigma_m^2}\right]. \tag{10}$$

By using (9) and (10), and $\sigma^2 = \sigma_m^2$, we find that (7) is maximized when the class m for voxel i is chosen so that

$$\frac{1}{2\sigma^2}(r_i-\mu_m)^2 - \beta\hat{u}_i(m) \tag{11}$$

is minimized. Here $\hat{u}_i(m)$ is the number of neighborhood voxels currently classified to class m. Positive β encourages neighbors to be of the same class.

Besag [29] applied the ICM algorithm exclusively for two-dimensional images. However, it extends readily to three-dimensional (3-D) images by simply defining $u_i(m)$ in three dimensions. Three-dimensional variations of the ICM algorithm have been used earlier, for example in the segmentation of structural MR brain images in [36], [37]. By minimizing (11) in three dimensions using a third-order neighborhood and using $\mu_0 = 0$, $\mu_1 = 2a$, and $\sigma^2 = 1$, the following contextual classification rule is derived:

$$r_i + \frac{\beta}{a}[\hat{u}_i(1) - 13]\begin{cases} < a \Rightarrow \hat{k}_i = 1 \\ \geq a \Rightarrow \hat{k}_i = 0. \end{cases} \tag{12}$$

When $\beta = 0$, (12) yields the context-free thresholding rule that is also used to initialize the classification. Later in this paper, we will refer to the value of the normal cumulative distribution function at a as a nominal alpha (α_n).

Equation (12) is intuitively appealing. The left part of (12) can be understood as a contextually corrected location on the null distribution. When the majority of the neighboring voxels belong to the background, the location is moved to the right on the distribution; in other words, to a less significant position. When the majority of the neighboring voxels belong to the rejection class, the location is moved to the left on the distribution.

In order to avoid certain unwanted effects, such as the growing of nonconvex regions, the values that the parameter β may have must be restricted. Through the parameter β, the trade-off between segmentation accuracy (error rates near the edges of activations) and weighting of contextual information can be adjusted. For example, if we set

$$\beta = \frac{a^2}{6} \tag{13}$$

we require that at least 19 of the voxels in the neighborhood of a background voxel are classified in the rejection class before the background voxel falls in the rejection class with probability of 0.5 or higher. For example, imagine a voxel with $r_i = 0$ surrounded by 19 active voxels. When condition (13) is used, the left part of (12) reaches the decision level a. According to our experiments, (13) ensures good segmentation accuracy and prevents unwanted filling effects. By substituting (13) into (12), the following rule used iteratively in most of our experiments is ultimately derived:

$$r_i + \frac{a}{6}[\hat{u}_i(1) - 13]\begin{cases} < a \Rightarrow \hat{k}_i = 1 \\ \geq a \Rightarrow \hat{k}_i = 0. \end{cases} \tag{14}$$

The propagation of data through the analysis is illustrated in Fig. 2. There are two important aspects in the implementation of the algorithm. The first is the order in which the voxels are updated. The second is how the border voxels of the image matrix are handled. The results presented in this paper are based on synchronous updating, in which all voxels are updated simultaneously, unless otherwise stated. For clarity, one updating of all voxels will be refered to as a cycle because the term "iteration" is occasionally used to refer to the updating of one voxel in a sequential updating. Regarding behavior on image borders, it is assumed that voxels outside the image volume belong to the background.

Concern could be raised about detection of activation regions with a mean of less than $2a$, and/or variance different from one and/or type of distribution that is non-Gaussian. The decision rule (12) provides a partial answer to this problem. Consider an activation region whose median value is smaller than a. During the initialization, the majority of the voxels are classified as active. Thereafter, inside the activation region, the majority of the neighboring voxels belong to the activation class and the use of (12) classifies more voxels to the activation class.

In hypothesis testing, the two central questions are: 1) What is the probability of classifying a truly nonactive voxel to the activation class with given a (probability of Type I error); and

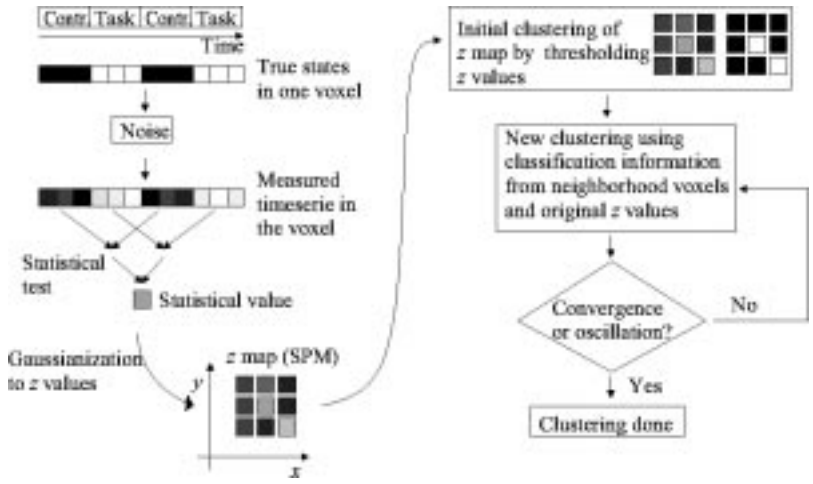

Fig. 2. Computation of a z map and contextual clustering of it.

2) what fraction of the activation voxels are identified with given α (power of the test)? These issues are discussed in Section III.

III. Experiments

A. Simulated Data

Simulation studies were carried out using z maps of 32 $\times$ 32 $\times$ 32 voxels in size. The z maps were generated using the following procedure. First, a 64 $\times$ 64 $\times$ 64 background image with mean zero and variance one was generated using the Gaussian pseudorandom number generator of Matlab 5.3 (The MathWorks, Inc., Natick, MA). In order to take spatial correlations into account, this image was filtered and sampled to the size of 32 $\times$ 32 $\times$ 32 image using the method that will be described and used in Section III-C. Activation was modeled using a discrete sphere centered at (15,15,15), radius 6.5, inside which a smaller empty sphere centered at (17,15,15), radius 3.5 (Fig. 3) was located. At the voxel locations defined by this sphere, intensities were replaced by values drawn from an activation distribution (Gaussian or uniform) having a mean $-s_{\mathrm{o}}$. Various values were given to s_{o}. The number of activation voxels was 986 while the remaining 31 782 voxels were assigned to the background.

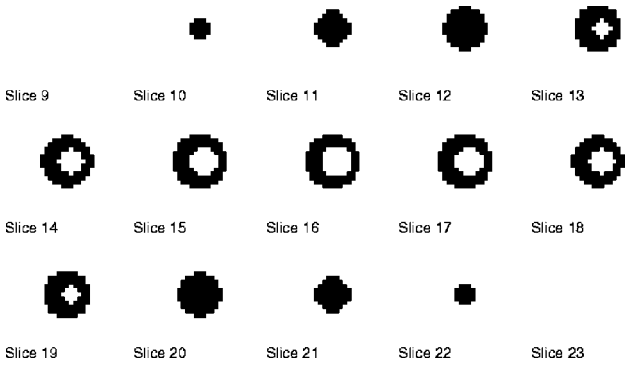

Fig. 3. Fifteen spatial levels of the simulated activation image 32 $\times$ 32 $\times$ s 32 voxels of size. All other levels are empty. The activation is a discrete sphere centered at (15,15,15), radius 6.5 inside which is a smaller empty sphere centered at (17,15,15), radius 3.5.

B. Measured Data

The magnetic resonance imaging of a volunteer was performed with a 1.5-T Siemens Magnetom Vision system (Siemens, Erlangen, Germany) using a gradient-echo echo-planar (EPI) sequence (TE 76 ms, TR 2.4 s, flip angle 90°, field of view 256 $\times$ 256 mm, matrix size 64 $\times$ 64, 16 slices, slice thickness 3 mm, gap 1 mm), and a standard head coil. A checkerboard visual stimulus flashing at 8 Hz rate (task condition, 24 s) was alternated with a visual fixation marker on a gray background (control condition, 24 s). In total, 110 samples (3-D volumes) were acquired. The brain was segmented from the EPI slices to enable identification of voxels belonging to the brain volume. The brain was assumed to remain at a fixed location during the scanning, so it was considered sufficient to segment only one volume. The sample number 60 was selected. Segmentation was achieved by thresholding the volume using an intensity threshold of 120. The result of the segmentation is used as a visualization aid in Fig. 9. Finally, a simple linear detrending (baseline correction) was conducted for each voxel. The first three samples of each control state and the first sample of each task state were not included in statistical analysis to account for the delay in the hemodynamic response.

C. Estimation of False Positive Probabilities

An estimate was made of the probabilities of classifying the voxel to the activation class under the null hypothesis, i.e., voxel-wise false positive rates (fpr), and the family-wise fprs, i.e., the probabilities of falsely detecting an activation in a whole volume. It is assumed that the intensity values of z maps follow standard normal distribution when no activations exist. This is the usual assumption made, which is normally well satisfied (see, e.g., [13]), although deviations from normality are possible [38].

Although normality is assumed, the spatial independence of the noise is not. We now describe the estimation procedure for the voxel-wise false positive rates presented in fMRI context by Forman *et al.* [12]. Both a 2-D and 3-D correlated noise were

TABLE I
ESTIMATED VOXEL-WISE FALSE POSITIVE RATES (fpr) IN AN IMAGE WITHOUT ACTIVATIONS

α_n	a	N	No filt.	$f_{w,2D}=0.2$	$f_{w,2D}=0.4$	$f_{w,2D}=0.6$	$f_{w,3D}=0.6$
0.05	−1.645	15000	0.00000009	0.00000009	0.00000007	0.00000010	0.00000015
0.07	−1.476	15000	0.0000015	0.0000015	0.0000014	0.0000016	0.0000024
0.09	−1.341	15000	0.000011	0.000011	0.000011	0.000013	0.000020
0.11	−1.227	5000	0.000051	0.000052	0.000053	0.000062	0.000104
0.13	−1.126	5000	0.000183	0.000183	0.000192	0.000226	0.00042
0.17	−0.954	1000	0.00131	0.00131	0.00142	0.00175	0.00360
0.21	−0.806	1000	0.00589	0.00593	0.00675	0.0086	0.0173
0.25	−0.675	400	0.0201	0.0204	0.0243	0.0310	0.0546
0.29	−0.553	400	0.0574	0.0583	0.0706	0.0847	0.1199

N is the number of simulated (random) z maps used in simulations.
$f_{w,n\text{-D}}$ is the standard deviation of n-dimensional Gaussian filter used to create spatial correlations.

TABLE II
ESTIMATED PROBABILITY OF DETECTING AT LEAST ONE FALSE POSITIVE VOXEL IN A z MAP OF 64 × 64 × 16 VOXELS

α_n	No filt.	$f_{w,2D}=0.3$	$f_{w,2D}=0.6$	$f_{w,3D}=0.6$
0.05	0.007	0.007	0.007	0.008
0.06	0.028	0.028	0.028	0.032
0.07	0.09	0.09	0.10	0.11
0.08	0.25	0.25	0.25	0.28
0.09	0.51	0.51	0.52	0.55

$f_{w,n\text{-D}}$ is the standard deviation of n-dimensional Gaussian filter used to create spatial correlations. Each result is based on 30 000 simulate D z maps.

modeled. In the 2-D case, 128 × 128 × 16 image was constructed using a pseudorandom number generator producing normally distributed numbers with mean zero and unit variance. The slices of the image were filtered with two-imensional Gaussian filter of standard deviation $2f_{w,2\text{-D}}$ voxels to approximate spatially stationary and continuously differentiable autocorrelation function of statistical parametric maps. The size of the convolution matrix was 5 × 5 voxels. To decrease the effect of pixelation, the filtered 128 × 128 × 16 image was decimated to 64 × 64 × 16 image by taking the mean value of each four-voxel block. This procedure returns the effective standard deviation of filter to $f_{w,2\text{-D}}$ voxels. The 3-D noise correlation was created in a similar way, but a 128 × 128 × 64 image was filtered with a 5 × 5 × 5 Gaussian filter of standard deviation $2f_{w,3\text{-D}}$ voxels and decimated using eight-voxel blocks. In related fMRI studies, the estimated correlations in planes corresponded to filter widths $f_{w,2\text{-D}} < 0.6$ [12], [38]. This is consistent with our findings and we perform the estimation of false positive rates and other studies using $f_{w,n\text{-D}} \leq 0.6$. Because the correlation in z direction is likely to be at most the same as it is within planes, a 3-D filter $f_{w,3\text{-D}} = 0.6$ is used to represent the worst case.

The filtered images were scaled into normal deviates and fed to the contextual clustering algorithm. The estimated voxel-wise false positive rates with different α_n and filter widths are tabulated in Table I. The corresponding results for obtaining at least one active voxel in the whole volume are tabulated in Table II. The number of significant digits presented for the estimates in these tables was selected so that the difference between minimum and maximum of 95% confidence interval was at most 4 units of the least significant digit for that estimate.

D. Simulations

The simulated z maps were clustered using $\alpha_n = 0.05$, 0.09, 0.13, 0.21, 0.29, 0.37 that correspond to $a \approx -1.64$, −1.34, −1.13, −0.81, −0.55, −0.33, respectively. Parameter β had value $\beta = a^2/6$ in all experiments, except in the test with a modified β. The number of false positive voxels detected in 500 simulations were counted for each a, and divided by the number of truly nonactive voxels (500 × 31 782) to estimate voxel-wise probabilities for false positives (ϵ_0). The probabilities for false negatives (ϵ_1) were computed in a similar fashion. The error rates for thresholding were computed in the same way, but the used threshold values T were computed by taking the inverse of the normal cumulative distribution function at ϵ_0 of contextual clustering. Thus, the ϵ_1 were evaluated using the same ϵ_0 for both thresholding and contextual clustering. In certain simulations, the spatial filtering was used to mimic the spatial correlations of the fMRI data. The results are shown in the form of receiver operating characteristic (ROC) curves. The curves reflect the trade-off between sensitivity and specificity.

First, simulations were made using a Gaussian distribution model for the activations with four different means (s_o =0.5, 1.5, 2.5, 3.5) and unit variance [Fig. 4(a)–(d)]. When s_o was very small, both the thresholding and contextual clustering failed to detect the activation at a reasonable ϵ_0. Conversely, when s_o was very high, thresholding similarly performed well. The largest difference in performance was found with intermediate s_o. The discrepancies between the ϵ_0 in Fig. 4 and fprs in Table I were due to the higher error probability near the activation edges than in an empty image. This effect is noticeable when the fraction of the edge voxels from the volume is high.

In the subsequent simulation, a non-Gaussian distribution model for activations was used [Fig. 4(e)–(f)], whereby two Gaussian densities ($s_{o_1} = 1.5$, $s_{o_2} = 3.5$) formed a single mixed, non-Gaussian density ($s_o = 2.5$). The detection rate was between the cases of Fig. 4(b) and (d). It can be seen that the correlation in the noise slightly lowered the power of the contextual clustering.

Subsequently, the experiments were conducted using a Gaussian distribution with differing variances and uniform distributions with varied intervals (Fig. 5). Although the differences in power between contextual clustering and thresholding decreased as the variance increased, they still existed. As the variance increased, the ϵ_1 of contextual clustering and

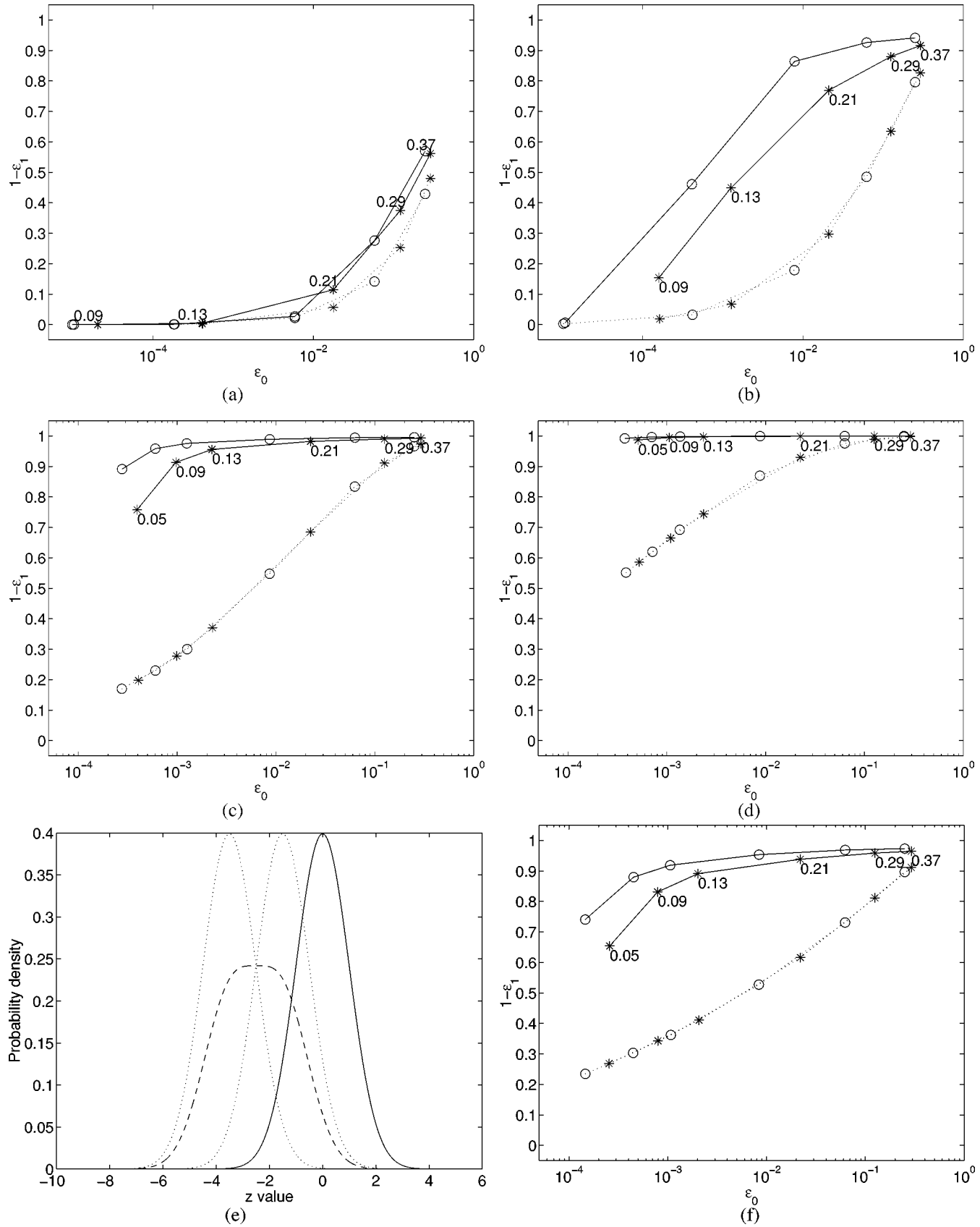

Fig. 4. ROC curves of contextual clustering (solid lines) and thresholding (dotted lines). The ROC curves show the probability of correctly classifying the activation voxels $(1-\epsilon_1)$ on the vertical scale and the probability of misclassifying the background class (ϵ_0) on the horizontal scale. $\circ$ marks refer to the spatially uncorrelated noise and $*$ to the correlated noise ($f_{w,3\text{-}\mathrm{D}} = 0.6$). The numbers for each point on the curves correspond to the values of α_n. (a) Case activation separation $s_\circ = 0.5$, (b) case $s_\circ = 1.5$, (c) case $s_\circ = 2.5$, (d) case $s_\circ = 3.5$, and (e)–(f) mixed density case. (e) density functions of background (solid line), two Gaussian activation densities $s_{\circ_1} = 1.5$, and $s_{\circ_2} = 3.5$ (dotted lines) and the mixed density (non-Gaussian) of activations (dashed line). (f) ROC curves in the mixed-density case.

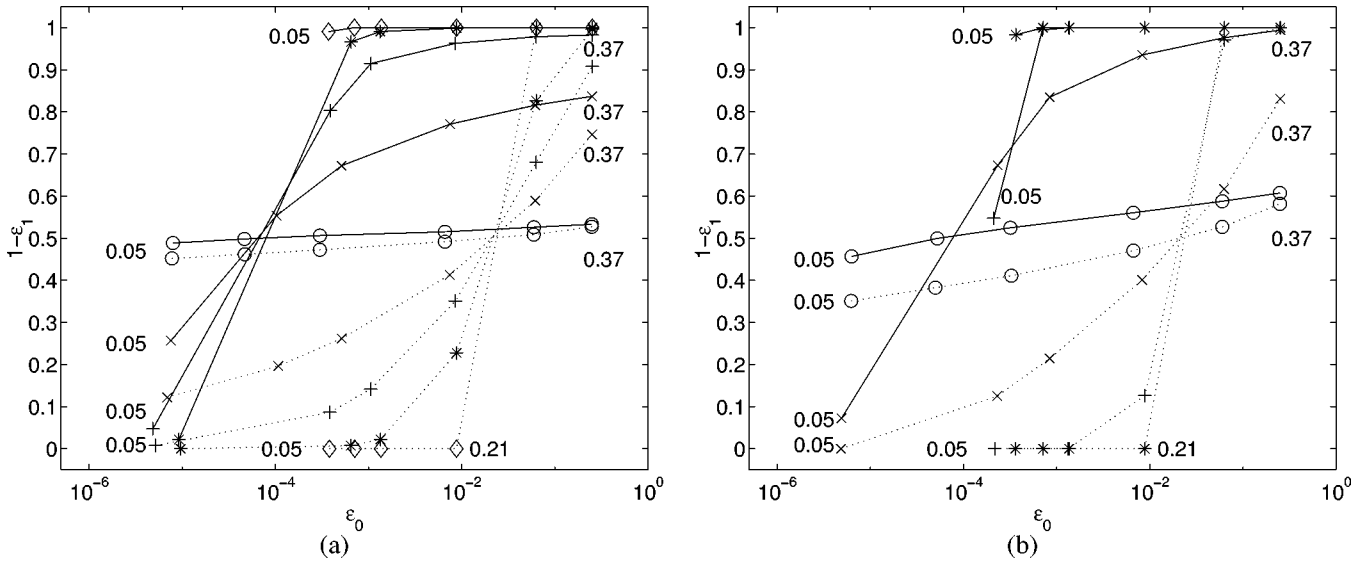

Fig. 5. ROC curves for the simulated activation following Gaussian distribution with varied variances and uniform distribution with varied intervals. $s_\circ = 2.0$ in all cases. Solid lines refer to contextual clustering and dotted lines to thresholding. (a) Gaussian distribution with varied standard deviation, $\circ$ marks refer to $\sigma = 20$, $\times$ to $\sigma = 2.0$, $+$ to $\sigma = 1.0$, $*$ to $\sigma = 0.5$, $\diamond$ to $\sigma = 0.05$. (b) Uniform distribution with varied intervals. $\circ$ marks: interval $[-10, 6]$; $\times$:l $[-4,0]$; $+$: interval $[-2.5; -1.5]$; and $*$: interval $[-2.6, -2.4]$.

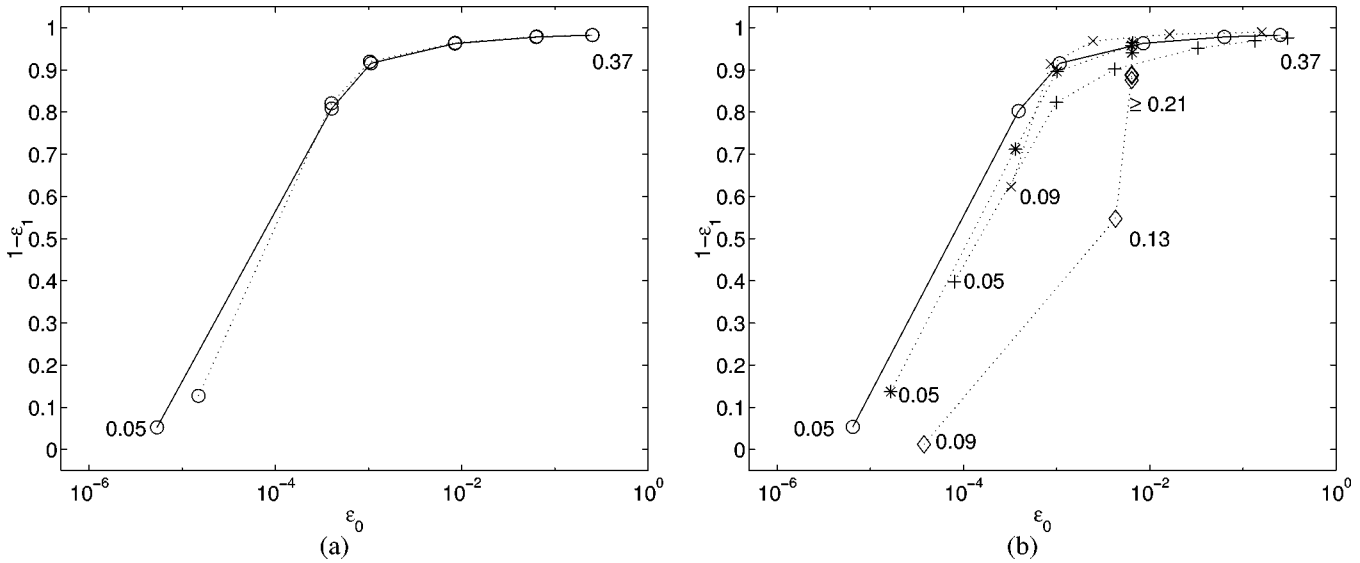

Fig. 6. ROC curves for modified algorithms. Solid lines with circle marks refer to unmodified algorithm ($\beta = a^2/6$ and uniformly weighted neighborhood) and dotted lines to modified algorithm. In all cases, the simulation activation follows Gaussian distribution with unit variance, $s_\circ = 2.0$. In (a), the dotted line refers to a distance weighted neighborhood; in (b), to varied β. $\times$ marks: $\beta = a^2/4$; $+$: $\beta = a^2/9$; $*$: $\beta = 0.4$; and $\diamond$: $\beta = 0.9$.

thresholding approached the value 0.5. The type of activation distribution (Gaussian vs. uniform) does not seem to have a significant effect on the results.

Finally, tests were made using a modified contextual clustering algorithm. In Fig. 6(a), the results were obtained using a distance weighted neighborhood. The second-order and third-order neighbors were downweighted by a factor of $\sqrt{2}$ and $\sqrt{3}$, respectively. Next, tests using a modified β were performed [Fig. 6(b)]. The differences between $\beta = a^2/4, a^2/6, a^2/9$ were not particularly large. When β had a fixed value ($\beta = 0.4, 0.9$) and α_n was sufficiently large ($\alpha_n \geq 0.21$), the activation was detected in most cases but the empty space inside the activation object was filled. The reason for this was that the parameter β was too large compared to α_n, i.e., to the parameter a. It is recommended, therefore, that the parameter β be related to the parameter a as in (13).

As an illustrative example, results using a Gaussian noise distribution for background and activation ($s_\circ = 1.5$) with unit variance are shown in Fig. 7. The simulated z map is shown in Fig. 7(a). When the testing was performed by thresholding the z map, only a few voxels of the activation were recognized [Fig. 7(b)]. Fig. 7(c) shows the initialization image ($\alpha_n = 0.21$) for the contextual clustering, Fig. 7(d) shows the clustering after the first cycle, and Fig. 7(e) after the second cycle. After the tenth cycle [Fig. 7(f)], the clustering converged. This image has approximately the same fpr (see Table I) as the image in Fig. 7(b) but the sensitivity is better.

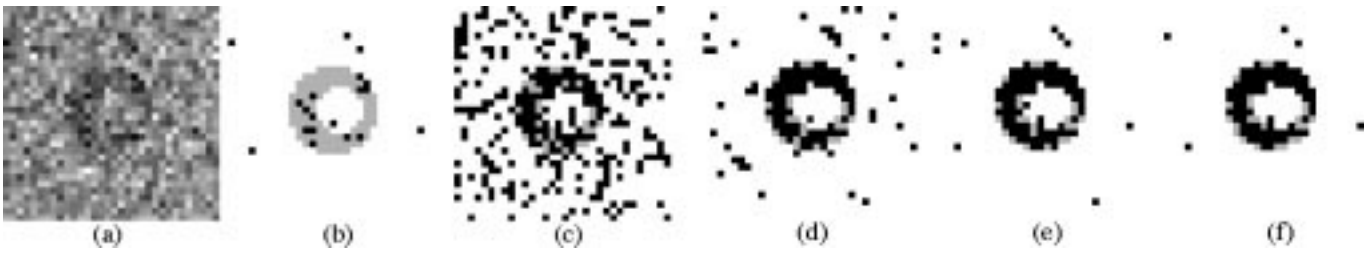

Fig. 7. Detection of the simulated activation from a noisy background. (a) The z map of slice 15 of the activation (see Fig. 3). In this z map, the activation voxels follow a Gaussian distribution with unit variance, $s_o = 1.5$. (b) Thresholded z map with false positive rate $fpr \approx 0.006,\ T \approx -2.52$. (c) Initialization image for contextual clustering $a \approx -0.806$. (d) Contextual clustering after the first cycle. (e) Contextual clustering after the second cycle. (f) Contextual clustering after the tenth cycle (convergence), $fpr \approx 0.006$. In (b)–(f), the activation voxels erroneously classified as background are shown as light gray.

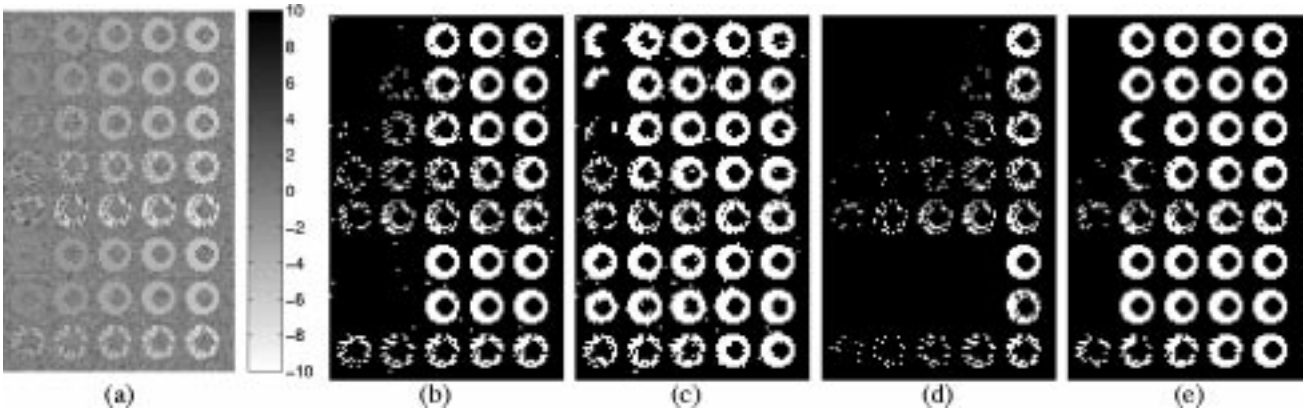

Fig. 8. Detection of simulated activations following Gaussian or uniform distribution. (a) In the z map, the distribution of the activation is Gaussian in rows 1 to 5 and uniform in rows 6 to 8. The standard deviation is 0.25, 0.5, 1.0, 2.0, 4.0, ~ 0.0722, ~ 0.5774 and ~ 4.6188 in rows 1, 2, 3, 4, 5, 6, 7, and 8, respectively. The widths of uniform distributions are 0.25, 2 and 16 in rows 6, 7, and 8, respectively. The mean of the activation, $-s_o$, varies column-wise as -1, -2, -3, -4, and -5, starting from the left. (b) Thresholded z map ($fpr \approx 0.006,\ T \approx -2.52$). (c) Contextually clustered z map ($fpr \approx 0.006,\ a \approx -0.81$). (d) Thresholded z map ($fpr \approx 1.5 \times 10^{-6},\ T \approx -4.67$). (e) Contextually clustered z map ($fpr \approx 1.5 \times 10^{-6},\ a \approx -1.48$).

In Fig. 8, an image with several activation objects following several distributions is clustered using two different significance levels. It can be seen that the contextual clustering is more sensitive than the thresholding regardless of the type and variance of a distribution. It should be noted that the behavior of the contextual clustering algorithm is local. Therefore, the existence of activation in one location does not affect the results at distant locations. Moreover, by dividing the Fig. 8(a) into 40 regions, each having one activation object, and clustering each region separately, it is likely that the same result as in Fig. 8 would be obtained.

E. Measured fMRI Data Studies

Fig. 9(a) illustrates the computed z map of the measured fMRI study. The activation of the visual cortex can be seen on the low part of the slices. In Fig. 9(b), the threshold is set to the value $T = -4.24$, which corresponds to $fpr \approx 0.000\,01$. In Fig. 9(c), the contextual clustering is applied to the z map with decision parameter value $a \approx -1.341$, which similarly corresponds to $fpr \approx 0.000\,01$. With the high number of repetitions, the thresholding technique finds the core of the activation while the contextual clustering appears to find the weaker outer areas, too. However, it is difficult to determine which one of the results is more accurate. In Fig. 9(d)–(f), only every sixth sample (in time) was used. In these cases, significantly larger activation areas were detected using contextual clustering [Fig. 9(f)] than using the thresholding technique [Fig. 9(e)]. In other words, when the activations are weak, or the z map is computed using a small number of samples, the contextual clustering has more acute sensitivity than the thresholding technique. This result is consistent with the results obtained using the simulated activation.

The experiments were repeated using sequential updating of the class values. The results were very similar with synchronous updating. When all time samples were used to compute a z map, 24 active voxels out of the original 1431 changed to nonactive voxels, and two nonactive voxels out of 64 105 changed to active. When every sixth sample was used, 29 active voxels out of 551 changed to nonactive, and four nonactive voxels out of 64 985 changed to active. However, the false positive rates (Tables I and II) should be re-estimated if sequential updating (or other modification) is used.

IV. Discussion and Conclusion

The utilization of contextual information was studied for the analysis of fMRI data. The approach consisted of two phases: computation of a statistical parametric map and activation detection by contextual clustering. The iterative clustering algorithm presented in this work is based on Besag's ICM algorithm [29]. Our contribution has been to apply and evaluate the ICM in the context of hypothesis testing and statistical parametric maps.

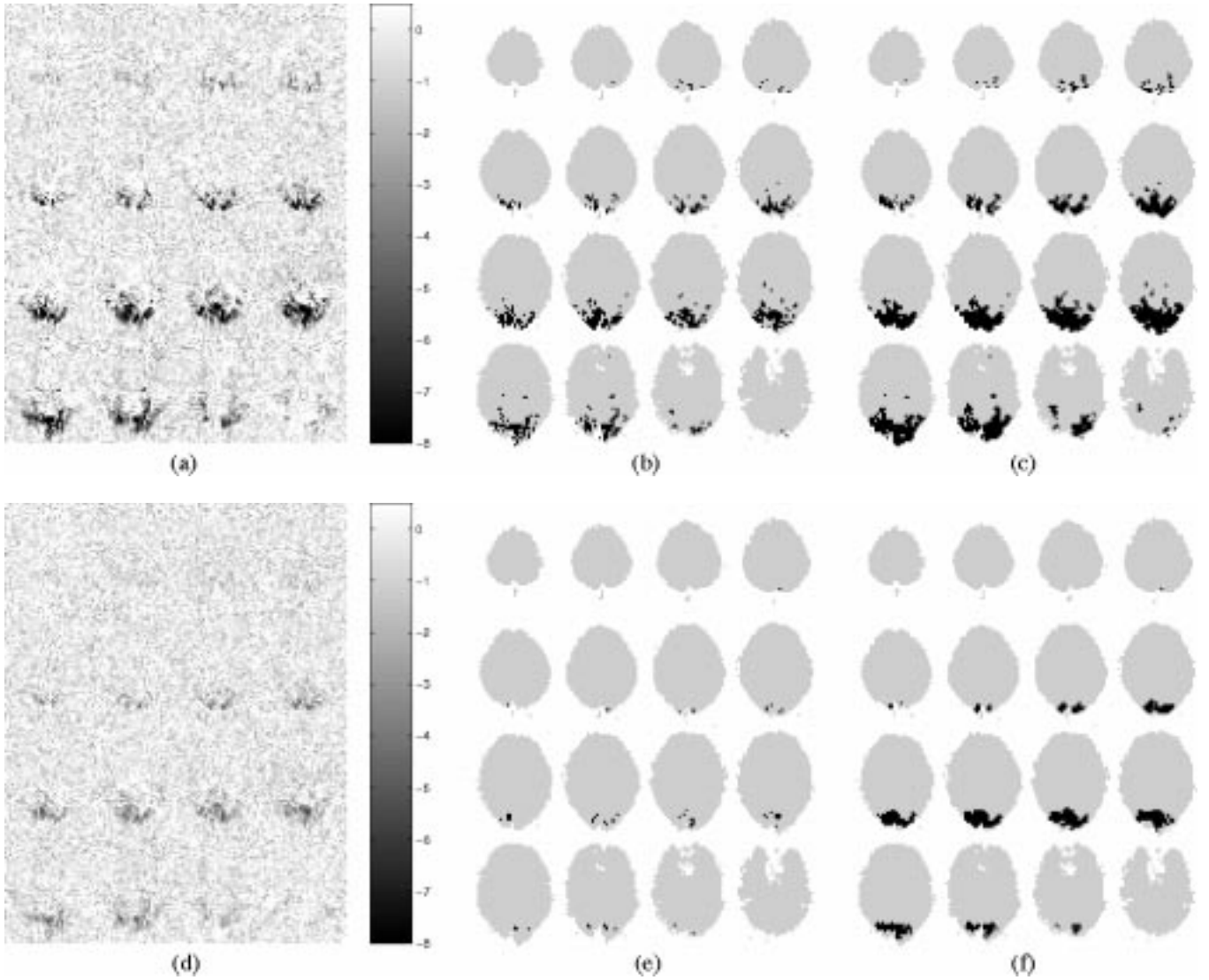

Fig. 9. Analysis of visual fMRI study, $fpr \approx 0.00001$ in all cases. (a) z map of the entire data, $N = 42 + 45$, (b) z map of the entire data thresholded, $T \approx -4.24$, (c) z map of the whole data contextually clustered, $a \approx -1.34$ (d) z map of the partial data, $N = 7 + 8$, (e) z map of the partial data thresholded, $T \approx -4.24$, and (f) z map of the partial data contextually clustered, $a \approx -1.34$. Although the brain slices are visualized here in planes, the data are analyzed in three dimensions.

The results presented here demonstrate that the context-free thresholding is more sensitive to random noise than is contextual analysis. Better sensitivity was achieved with contextual clustering. However, a strong spatial dependency in the noise term may increase the error rate. The results show that the spatial autocorrelations present in a typical fMRI study contribute little to the results. The effect is small particularly for family-wise tests. The spatial correlations may be more significant for smaller voxel size or for other imaging modalities, for example, positron emission tomography (PET).

Unlike median filters or spatial extent based methods, the contextual method presented here allows the possibility to preserve the smallest activations if the focal significance is sufficiently high. Furthermore, spatial smoothing or other procedures that result in a loss of information are not required.

An alternative to the hypothesis testing approach would be a maximum probability classification. This should include the estimation of the number of activations classes, their means, variances, the types of distributions and potential parameters (β). Several methods to estimate the parameters from data have been suggested. Besag suggests the estimation of unknown parameters β, μ_m, σ_m during the iterations of ICM [29]. At each cycle, the parameters are updated so that $l(\mathbf{r}|\hat{\mathbf{k}}; \mu_m, \sigma_m)$ and the pseudolikelihood $\prod_{i \in I} p_i(k_i|\mathbf{k}_{\partial i}; \beta)$ are maximized. Examples of how the number of classes might be estimated are provided in, for example, [39]–[41].

A contextual approach based on MRFs and Gibbs sampler was used by Holmes [32] to segment statistical parametric maps computed from PET data. One conclusion of [32] was that the approach is less powerful than the thresholding test. However, the tests were not "tuned" to the same significance level and spatial correlations in the noise of the simulated PET data were large.

In a related paper, spatio-temporal fMRI analysis by Descombes *et al.* utilizes contextual information and MRFs [33], [34]. The main conceptual difference between the method of Descombes *et al.* and the method presented here is that the

former performs data restoration and analysis for the original spatio-temporal fMRI data while the latter clusters a standard statistical parametric map. An advantage of the spatio-temporal analysis is that it uses more data and works on the original spatio-temporal data. In addition, simulated annealing (SA) optimization algorithm was used to find the global maximum of the objective function in [33], [34]. The ICM algorithm presented here converges to a local maximum (or with synchronous updating, may start oscillating between states). As a result of the use of the computationally demanding SA algorithm and original four-dimensional data, the spatio-temporal method is slower than the method presented here. A strong point of ICM is the exclusive dependence on local image characteristics. For fMRI data, the ICM algorithm converged in ten cycles. The small number of cycles assures that a deviating data distribution (either due to an activation or artifact) at one location does not increase the false positive probability or reduce the power at distant locations.

The algorithm using Matlab 5.3 was implemented. On a Pentium III/500 MHz PC, the contextual clustering, after the computation of a z map, took on average about 4 s CPU time for a $64 \times 64 \times 16$ volume. Compared to the 5 s of CPU time needed to compute the z map from 110 volumes (without preprocessing steps), this is a tolerable addition. Computational efficiency allows the data-analysis to be performed on-line to monitor the progress of imaging. The algorithm is highly parallel when the updating is carried out simultaneously on all voxels. This property allows the algorithm to run yet more efficiently on a computer with a parallel architecture.

The results indicate that the power of the developed contextual algorithm is superior to that of conventional voxel-by-voxel thresholding of a statistical parametric map. Although fMRI data were used to test the algorithm, the construction of the algorithm is general and it can be used to detect objects with unknown distribution from a known background distribution in other similar problems as well.

ACKNOWLEDGMENT

The authors would like to thank Prof. T. Katila and Prof. C. Tesche at Helsinki University of Technology for their contributions to this work. They would also like to thank C. Aschan and J. Lampinen of the University of Helsinki for helpful discussions on statistical analysis. Finally, they would like to thank the anonymous reviewers for valuable comments that helped to improve this paper.

REFERENCES

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

[2] K. K. Kwong, "Functional magnetic resonance imaging with echo planar imaging," *Magn. Reson. Q.*, vol. 11, pp. 1–20, 1995.

[3] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain," *Magn. Reson. Med.*, vol. 30, pp. 161–173, 1993.

[4] A. Kleinschmidt, M. Requardt, K.-D. Merboldt, and J. Frahm, "On the use of temporal correlation coefficients for magnetic resonance mapping of functional brain activation: individualized thresholds and spatial response delineation," *Int. J. Imaging Systems and Technology*, vol. 6, pp. 238–244, 1995.

[5] K. Kuppusamy, W. Lin, and E. M. Haacke, "Statistical assesment of crosscorrelation and variance methods and the importance of electrocardiogram gating in functional magnetic resonance imaging," *Magn. Reson. Imag.*, vol. 15, pp. 169–181, 1997.

[6] B. A. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno, "Activation detection in functional MRI using subspace modeling and maximum likelihood estimation," *IEEE Trans. Med. Imag.*, vol. 18, pp. 101–114, Feb. 1996.

[7] M. J. Brammer, "Multidimensional wavelet analysis of functional magnetic resonance images," *Human Brain Mapping*, vol. 6, pp. 378–382, 1998.

[8] J. L. Marchini and B. D. Ripley, "A new statistical approach to detecting significant activation in functional MRI," *NeuroImage*, vol. 12, pp. 366–380, 2000.

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

[10] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-B. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.

[11] 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. 214–220, 1994.

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

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

[14] K. J. Friston, A. P. Holmes, J.-B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak, and R. Turner, "Analysis of fMRI time series revisited," *NeuroImage*, vol. 2, pp. 45–53, 1995.

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

[16] M. J. Lowe and J. A. Sorenson, "Spatially filtering functional magnetic resonance imaging data," *Magn. Reson. Med.*, vol. 37, pp. 723–729, 1997.

[17] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[18] J. J. Sychra, P. A. Bandettini, N. Bhattacharya, and Q. Lin, "Synthetic images by subspace transforms. I. Principal components images and related filters," *Med. Phys.*, vol. 21, pp. 193–201, 1994.

[19] W. Backfrieder, R. Baumgartner, M. Sámal, E. Moser, and H. Bergmann, "Quantification of intensity variations in functional MR images using rotated principal components," *Phys. Med. Biol.*, vol. 41, pp. 1425–1438, 1996.

[20] R. Baumgartner, G. Scarth, C. Teichtmeister, R. Somorjai, and E. Moser, "Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: Reproducibility," *J. Magn. Reson. Imag.*, vol. 7, pp. 1094–1101, 1997.

[21] E. Moser, M. Diemling, and R. Baumgartner, "Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: Quantification," *J. Magn. Reson. Imag.*, vol. 7, pp. 1102–1108, 1997.

[22] R. Baumgartner, C. Windischberger, and E. Moser, "Quantification in functional magnetic resonance imaging: Fuzzy clustering vs. correlation analysis," *Magn. Reson. Imag.*, vol. 16, pp. 115–125, 1998.

[23] C. Goutte, P. Toft, E. Rostrup, F. Å. Nielsen, and L. K. Hansen, "On clustering fMRI time series," *NeuroImage*, vol. 9, pp. 298–310, 1999.

[24] A. Baune, F. T. Sommer, M. Erb, D. Wildgruber, B. Kardatzki, G. Palm, and W. Grodd, "Dynamical cluster analysis of cortical fMRI activation," *NeuroImage*, vol. 9, pp. 477–489, 1999.

[25] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*, second ed. Pacific Grove: PWS Publishing, 1999.

[26] A. Visa and J. Iivarinen, "Evolution and evaluation of a trainable cloud classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 1307–1315, Sept. 1997.

[27] R. T. Constable, G. McCarthy, T. Allison, A. W. Anderson, and J. C. Gore, "Functional brain imaging at 1.5 T using conventional gradient echo MR imaging techniques," *Magn. Reson. Imag.*, vol. 11, pp. 451–459, 1993.

[28] J. Kittler and D. Pairman, "Contextual pattern recognition applied to cloud detection and identification," *IEEE Trans. Geosci. Remote Sensing*, vol. GRS-23, pp. 855–863, 1985.

[29] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.
[30] ——, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, vol. 36, pp. 192–236, 1974.
[31] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.
[32] A. P. Holmes, "Statistical Issues in Functional Brain Mapping," Ph.D., Univ. Glasgow, Glasgow, Scotland, 1994.
[33] X. Descombes, F. Kruggel, and D. Y. von Cramon, "Spatio-temporal fMRI analysis using Markov random fields," *IEEE Trans. Med. Imag.*, vol. 17, pp. 1028–1039, Dec. 1998.
[34] ——, "fMRI signal restoration using a spatio-temporal Markov random field preserving transitions," *NeuroImage*, vol. 8, pp. 340–349, 1998.
[35] J. S. Milton and J. C. Arnold, *Introduction to probability and statistics. Principles and applications for engineering and the computing sciences*, 3rd ed. New York: McGraw-Hill, 1995.
[36] K. Held, E. Rota Kops, B. J. Krause, W. M. Wells III, R. Kikinis, and H.-W. Müller-Gärtner, "Markov random field segmentation of brain MR images," *IEEE Trans. Med. Imag.*, vol. 16, pp. 878–886, Dec. 1997.
[37] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral MR images," *IEEE Trans. Med. Imag.*, vol. 16, pp. 176–186, Apr. 1997.
[38] E. Zarahn, G. K. Aguirre, and M. D'Esposito, "Empirical analyzes of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions," *NeuroImage*, vol. 5, pp. 179–197, 1997.
[39] V. E. Johnson, "A model for segmentation and analysis of noisy images," *J. Amer. Statist. Assoc.*, vol. 89, pp. 230–241, 1994.
[40] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.*, vol. 13, pp. 441–449, Sept. 1994.
[41] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd Jr, and R. E. Coleman, "Bayesian reconstruction and use of anatomical a priori information for emission tomography," *IEEE Trans. Med. Imag.*, vol. 15, pp. 673–686, Oct. 1996.