

# Statistical Segmentation of fMRI Activations Using Contextual Clustering

Eero Salli<sup>1,2</sup>, Ari Visa<sup>3</sup>, Hannu J. Aronen<sup>1,2</sup>, Antti Korvenoja<sup>2,4</sup>, and Toivo Katila<sup>1,4</sup>

<sup>1</sup> Laboratory of Biomedical Engineering, Helsinki University of Technology,  
P.O.B. 2200, FIN-02015 HUT, Espoo, Finland

{Eero.Salli, Toivo.Katila}@hut.fi

<sup>2</sup> Department of Radiology, Helsinki University Central Hospital, P.O.B. 380,  
FIN-00029 HYKS, Helsinki, Finland

Hannu.Aronen@huch.fi, Antti.Korvenoja@helsinki.fi

<sup>3</sup> Department of Information Technology, Lappeenranta University of Technology,  
FIN-53851 Lappeenranta, Finland

Ari.Visa@lut.fi

<sup>4</sup> BioMag Laboratory, Helsinki University Central Hospital, P.O.B. 503, FIN-00029  
HYKS, Helsinki, Finland

**Abstract.** A central problem in the analysis of functional magnetic resonance imaging (fMRI) data is the reliable detection and segmentation of activated areas. Often this goal is achieved by computing a statistical parametric map (SPM) and thresholding it. Cluster-size thresholds are also used. A new contextual segmentation method based on clustering is presented in this paper. If the SPM value of a voxel, adjusted with neighborhood information, differs from the expected non-activation value more than a specified decision value, the contextual clustering algorithm classifies the voxel to the activation class, otherwise to the non-activation class. The voxel-wise thresholding, cluster-size thresholding and contextual clustering are compared using fixed overall specificity. Generally, the contextual clustering detects activations with higher probability than the voxel-wise thresholding. Unlike cluster-size thresholding, contextual clustering is able to detect extremely small area activations, too. Moreover, the results show that the contextual clustering has good segmentation accuracy, voxel-wise specificity and robustness against spatial autocorrelations in the noise term.

## 1 Introduction

The most common goal of the fMRI study is the recognition and segmentation of brain areas that respond to a given task or stimulus. Usually, it is required that the overall specificity of the recognition is high. On the other hand, the sensitivity of the process and good specificity in the neighborhood of activated areas are also important. The segmentation task is challenging due to low signal-to-noise ratio of activation related signal changes.

The analysis of an fMRI image series is most often based on the computation of the SPM and making statistical inferences from it. The SPM can be computed using general linear model [1]. Also non-parametric approaches like Kolmogorov-Smirnov [2] statistics are frequently used. Inferences from the SPMs are usually based on voxel-wise intensities above a threshold [3], on the spatial extent of contiguous voxels above a threshold (cluster-size thresholds) [4, 5] or on the combination of these two [6]. Usually the parameters of tests are chosen so that the probability of detecting a false activation in the whole volume (or in the search volume) is relatively small, e.g. 0.05.

In this paper, we present a new method for making statistical inferences and segmentation, called as contextual clustering, which uses both the original statistical value and current voxel classification in the neighborhood to determine the new voxel classification. Contextual clustering, voxel-wise and cluster-size thresholding methods are compared. Especially, sensitivity, segmentation accuracy and robustness against spatial correlations are studied.

## 2 Materials

The results of this study are based on simulated and measured fMRI data. First, noisy background data (*Data A*) was created by using Matlab's (The Mathworks, inc., Natick, MA, USA) pseudorandom generator. Voxel values were chosen from the Gaussian distribution with mean zero and variance one (Fig. 1(a)). For each simulation, 120 three-dimensional samples (images) were created. The matrix size of each sample was  $32 \times 32 \times 16$ . In total, 500 fMRI experiments were simulated, so that the total number of random values generated was  $500 \times 120 \times 32 \times 32 \times 16$ . The *Data A* was used to determine segmentation parameters and to estimate voxel-wise specificity of each segmentation method. Next, a simulation phantom of 12 spheres (Fig. 1(b)) was added into the samples 21...40, 61...80 and 101...120 of *Data A*. This operation formed *Data B* which was used to estimate the sensitivity and segmentation accuracy of the segmentation methods. The *Data* sets *C* and *D* were formed by spatially filtering *Data A* using a filter with  $3 \times 3 \times 3$  Gaussian kernel of full width at half magnitude (FWHM) 0.6 voxels and 1.2 voxels, respectively. Data sets *C* and *D* were used to evaluate robustness of the various methods against spatial autocorrelations.

The measured data of a volunteer was acquired with a 1.5-T Siemens Magnetom Vision system (Siemens, Erlangen, Germany) using a gradient-echo echo-planar (EPI) sequence (TE 76 ms, TR 3.5 s, flip angle  $90^\circ$ ), and a standard head coil. The right median nerve of a right-handed male volunteer was stimulated electrically with a 0.2 ms constant voltage pulse. The stimulus intensity was adjusted slightly above the motor threshold. The stimulation rate was 4 Hz. The subject kept eyes closed during the experiment. To minimize head movements, bite bar and supporting vacuum pillow were used.

A set of 240 images (matrix  $128 \times 128 \times 16$ , pixel size  $2 \times 2$  mm, slice thickness 5 mm) was acquired with the baseline and stimulation condition alternating in

blocks of ten images. The first three samples of each block were not taken into statistical analysis to account for the delay in the hemodynamic response.

### 3 Methods

#### 3.1 Computation of a statistical parametric map

For the purposes of this segmentation study the simple  $t$  statistics is an adequate method to compute an SPM. In voxel  $i$ , the test statistics can be formulated [7]:

$$t_i = \frac{\bar{X}_{iB} - \bar{X}_{iA}}{\sqrt{\frac{(n_A-1)\sigma_{iA}^2 + (n_B-1)\sigma_{iB}^2}{n_A+n_B-2} \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}, \quad (1)$$

where  $\bar{X}_{iA}$  and  $\bar{X}_{iB}$  are the means of the observed intensity values,  $\sigma_{iA}^2$  and  $\sigma_{iB}^2$  estimated variances and  $n_A$ ,  $n_B$  the sizes of the task state (subindex  $A$ ) and control state sets (subindex  $B$ ). The  $t$  statistics can be formulated easily also using the general linear model [1]. Variable  $t_i$  has the Student's  $t$  distribution with  $n_1 + n_2 - 2$  degrees of freedom under null hypothesis. The  $t$ -map is transformed to Gaussian distributed  $z$ -map using  $z_i = q_{norm}(p_t(t_i, r))$ , where  $q_{norm}$  is the normal inverse distribution function at  $x$  and  $p_t(x, r)$  is the cumulative distribution function for  $t$  distribution with  $r$  degrees of freedom at  $x$ .

Often, prior to the computation of SPM, the image data are spatially filtered in order to increase the signal-to-noise ratio and sensitivity [3]. However, artifacts such as blurred edges result after low pass filtering. As our study concentrates on the segmentation of an already computed SPM, the filtering of image data is not studied here.

#### 3.2 Segmentation methods

**Voxel-wise thresholding (VWTH).** The simplest segmentation method of an SPM is thresholding applied to each voxel separately. The voxel at location  $i$  is considered as active if and only if  $z_i < T$ , where  $T$  is a threshold. In our notation the activations have negative mean and  $T$  is chosen negative.

**Cluster-size thresholding (CSTH).** A cluster  $c$  is considered as active if and only if for all voxels within the cluster  $z_i < T$  and the size of cluster  $c$  is at least  $T_{size}$  voxels [5]. In our study, the method is used in 3D with 26-connectivity.

**Contextual clustering (CC).** The algorithm of contextual clustering is: 1) Label the voxels with  $z_i < T$  as *active* and other voxels as *non-active*. Voxels outside the image volume are considered as *non-active*. 2) Compute for all voxels  $i$  the number of *active* neighbor voxels,  $u_i$ . 3) Relabel the voxels for which

$$z_i + \frac{\beta}{T}(u_i - N/2) < T \quad (2)$$

as *active* and other voxels as *non-active*. The number of neighbor voxels equals to  $N$ . Using the 26-connectivity  $N = 26$ . The parameter  $\beta$  determines the weighting of the contextual information and is usually positive. If we set  $\beta = 0$  we have the conventional context-free thresholding rule. 4) If the current labeling is same as the labeling in the previous cycle or in the cycle before that, then stop the iterations, otherwise return to the step 2).

It should be noted that from the mathematical point of view the algorithm is identical with iterated conditional modes (ICM) algorithm of two classes [8]. However, the 2-class ICM algorithm with on-line parameter updating is not applicable because each activation has its own density function. Therefore, actually the 4-class ICM should be used in our case of three different activation and one background distributions. In practice, the number of activation distributions is unknown and therefore the ICM algorithm cannot be used conventionally. Instead of trying to classify every voxel into most probable class, as is done with ICM, the algorithm is used in a completely new way: as a hypothesis testing technique. In other words, if the statistical parametric value of a voxel adjusted with contextual information differs from the null distribution more than a preset threshold, the null hypothesis is rejected. Using the parameter  $\beta$  the trade off between sensitivity and segmentation accuracy can be adjusted. One way to set the parameter  $\beta$  is to write  $\beta = T^2/s$ . Then  $s$  specifies the excess of activated voxels ( $u_i - N/2$ ) in the neighborhood required to classify a non-active voxel to the *active* class with the probability of 0.5.

### 3.3 Estimation of parameter values for common overall false activation probability and voxel-wise specificity

In order to compare the sensitivity of the methods the probability of observing false activation voxel(s) in a whole volume was fixed to value of 0.05. The parameter estimation was done by applying the algorithms to the *Data A* with different decision parameter values. A decision parameter value that gave a false activation in 25 images out of 500 (i.e. overall p-value 0.05) was chosen. Also the probability of false activation at voxel level was computed from the measured number of false activation voxels.

### 3.4 Sensitivity, segmentation accuracy and robustness

The 500 SPMs computed from the *Data B* were segmented with each segmentation method. The segmented images of each method were averaged separately. The resulting *mean image* shows the probability of detection at voxel-level.

In addition, in order to estimate the segmentation accuracy, the number of voxels falsely classified as active in the neighborhood of activation *R3C4* (see Fig. 1(b)) was counted.

The robustness against spatial autocorrelations in the noise term was evaluated by segmenting *Data C* and *Data D*. If the overall probability of detecting false activation is significantly larger than the expected value of 0.05, the robustness of the method against spatial autocorrelations is low.

## 4 Results

The estimated decision parameter values and measured probabilities of false activations at voxel level for various segmentation methods are given in Table I. It is seen that although the overall specificity is fixed, there are differences in specificity at voxel level. Best voxel level specificity was achieved with voxel-wise thresholding and contextual clustering. An interesting observation was that all false activations detected with contextual clustering were only one voxel in size. We also studied the overall specificity with larger volume  $64 \times 64 \times 16$  using the same decision parameters. In this case the achieved overall false activation rate was  $0.23 \pm 0.05$ .

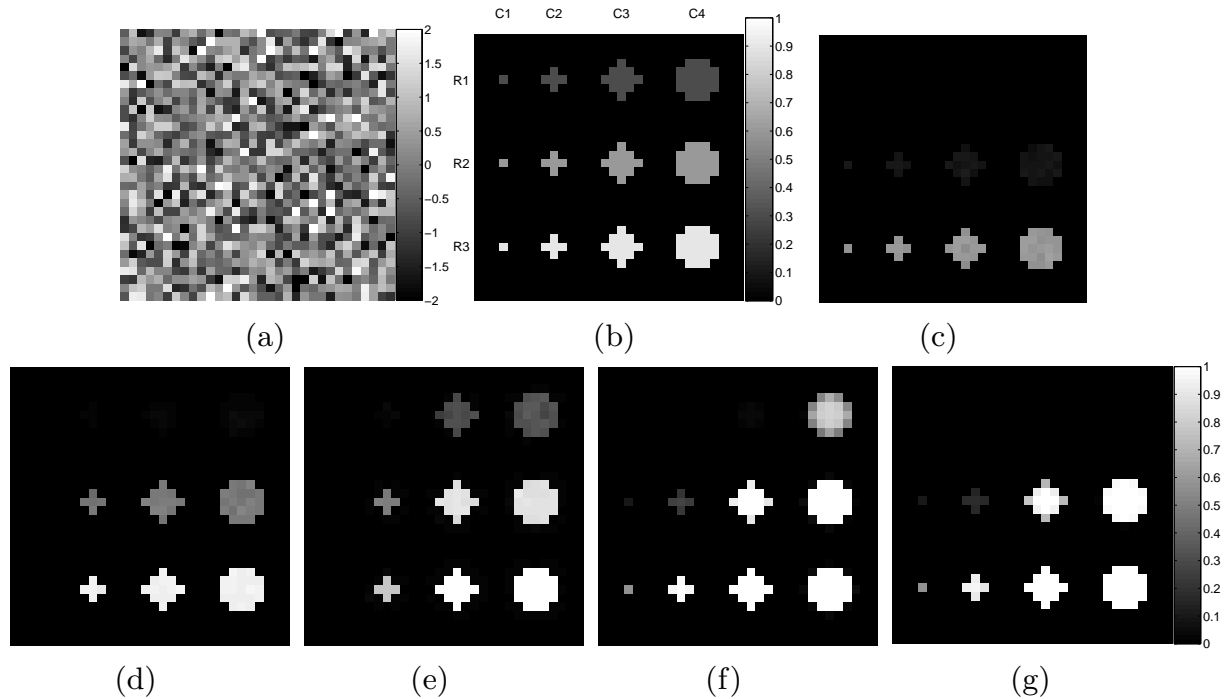
**Table 1.** Estimated decision parameter values  $T$  for overall false activation rate 0.05 and measured probability of false activation at voxel level  $P(f_v)$ , probability of false activation voxels  $P(f_{R3C4})$  in the neighborhood of activation  $R3C4$  at voxel level and overall probability of false activation in the presence of spatial autocorrelations  $P(f_{c1})$  (FWHM = 0.6 voxels) and  $P(f_{c2})$  (FWHM = 1.2 voxels).

Method	Abbrev.	$T$	$P(f_v) \cdot 10^{-6}$	$P(f_{R3C4})$	$P(f_{c1})$	$P(f_{c2})$
Voxel-wise thresholding	VWTH	-4.490	3.05	0.000	0.05	0.04
Cluster-size thr., $T_{size} = 2$	CSTH:2	-3.269	6.35	0.001	0.08	0.19
Cluster-size thr., $T_{size} = 8$	CSTH:8	-2.066	27.5	0.019	0.18	0.51
Context. clust., $\beta = T^2/2$	CC:2	-0.597	3.05	0.006	0.05	0.05
Context. clust., $\beta = T^2/6$	CC:6	-1.415	3.05	0.001	0.05	0.05

The sensitivity of each segmentation method is shown in Fig. 1(c)-(g). The images are computed by averaging the segmentation results. It can be seen that the voxel-wise thresholding and contextual clustering can detect activations as small as only one voxel in size. On larger activations the cluster-size thresholding with  $T_{size} = 8$  and contextual clustering are more sensitive than the voxel-wise thresholding. Cluster-size thresholding with  $T_{size} = 2$  seems to be only slightly more powerful than voxel-wise thresholding but misses the smallest activations.

Activation  $R3C4$  was used to evaluate segmentation accuracy, i.e. the specificity in the neighborhood of an activation. Also robustness against spatial autocorrelations was studied. The results are reported in Table I. Cluster-size thresholding with  $T_{size} = 8$  had worst segmentation accuracy and robustness against autocorrelations.

As the last experiment, we analysed the measured fMRI data. Results on two subsequent slices are shown in Fig. 2. All methods found activations of primary sensorimotor cortex (SMI) and supplementary motor area (SMA). Large  $\beta$  or cluster-size threshold  $T_{size}$  decreases the ability to distinguish closely separated activations as multiple foci. This is most clearly seen in the area of SMI, where the clusters of activated voxels in the central sulcus and postcentral sulcus merge with  $T_{size} = 8$  or  $\beta = T^2/2$ .



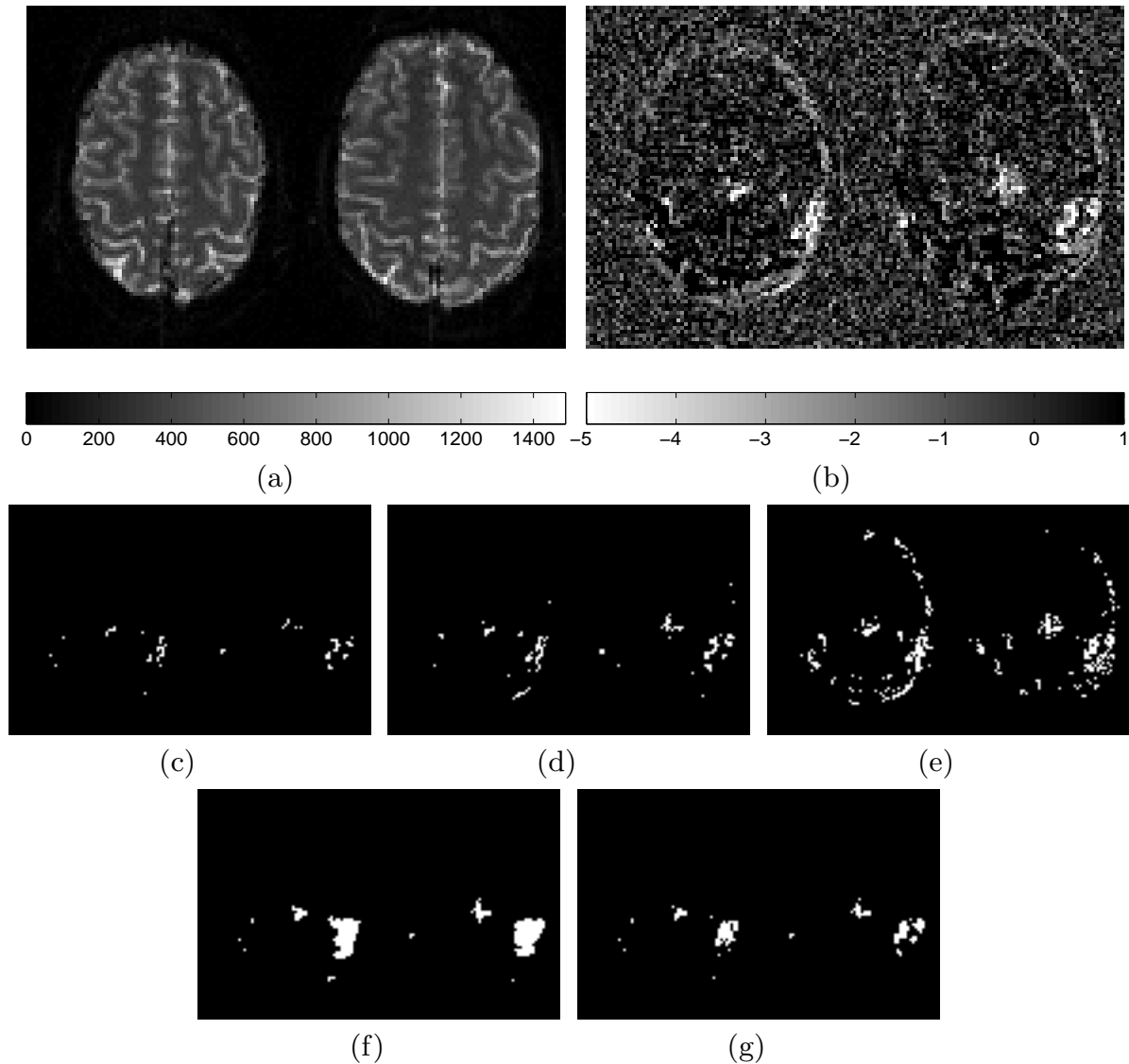
**Fig. 1.** (a) Gaussian noise image, (b) slice 8 of the phantom. (c)-(g) Probability of activation detection at voxel level. (c) VWTH, (d) Csth:2, (e) Csth:8, (f) CC:2, (g) CC:6. The scale for (c)-(g) is shown on the right side of (g).

## 5 Discussion

The methods for analysing fMRI data are often compared by power analysis only. Much less attention is paid to a segmentation accuracy and robustness. As we consider these properties equally important, they were included into this study.

Our parameter selection was based on a number of simulations. Although it is sometimes possible to approximately compute parameter values for chosen overall specificity [4], the simulation approach has some advantages. First, we are not forced to make any assumptions about the distribution of noise (e.g. smoothness or Gaussianity). The false activation rates can be estimated as precisely as needed, by simply increasing the number of simulations. The major drawback of the simulation approach, as compared to the theoretical one, is the large number of calculations required. However, with the relatively low matrix sizes of fMRI acquisitions, the parameter estimation can be done even with desktop computers in few hours.

One could criticize that in our simulations the time series of activation voxels follow simple box-car function corrupted with independent Gaussian noise. Some evidence exists that the assumptions about time series data being Gaussian distributed and time series samples being independent holds well in fMRI data [9], but the question is still debated. The hemodynamic response function is, of course, much more complicated than a simple box-car function. For the same reason, the SPMs are sometimes computed by using other methods than simple



**Fig. 2.** Segmentation of fMRI data in two subsequent slices. (a) Two EPI slices at  $t = 0$  s, (b) The corresponding SPM ( $z$ -map). Responses to stimulation of the right median nerve in the contra- and ipsilateral SMI and SMA can be seen as brighter areas in the SPM. In the SMI, activation is seen in two adjacent sulci (precentral and postcentral). Segmentation of the SPM by (c) VWTH, (d) Csth:2 (e) Csth:8, (f) CC:2 and (g) CC:6.

unpaired  $t$  test [10, 11]. We believe, however, that the limitations mentioned do not have any significant effects on the results of this segmentation study.

The spatial autocorrelations in the noise term may have effects on the performance of segmentation methods. If the structure of autocorrelations is known it can be taken into account in the estimation of decision parameters. In practice, the structure can not be estimated precisely. Therefore, a good robustness against spatial autocorrelations is a desirable feature of the segmentation method and this property was analysed in the present study.

## 6 Conclusions

Based on our results, the following conclusions can be made: 1) Generally, contextual clustering is more sensitive than voxel-wise thresholding. 2) Contextual clustering can detect, unlike cluster-size thresholding, activations that are only one voxel in size. 3) On detecting large activations, the precedence of contextual clustering and cluster-size thresholding depends on the parameter values chosen of a method and the size and strength of an activation. 4) Contextual clustering has extremely good voxel-wise specificity, segmentation accuracy, and robustness against spatial autocorrelations. 5) The detection rate of large but weak activations with contextual clustering is increased by increasing  $\beta$ -parameter, but at the same time, the segmentation accuracy decreases.

In summary, a new contextual method to statistically segment fMRI activations, called as contextual clustering, has been developed. This clustering method appears to be very competitive with the other methods studied.

## References

1. K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, R. S. J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
2. K. K. Kwong. Functional magnetic resonance imaging with echo planar imaging. *em Magn. Reson. Q.*, 11:1–20, 1995.
3. J. Xiong, J.-H. Gao, J. L. Lancaster, P. T. Fox. Assessment and optimization of functional MRI analyses. *Human Brain Mapping*, 4:153–167, 1996.
4. K. J. Friston, K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, A. C. Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:213–220, 1994.
5. S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, D. C. Noll. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magn. Reson. Med.*, 33:636–647, 1995
6. J.-B. Poline, K. J. Worsley, A. C. Evans, K. J. Friston. Combining Spatial Extent and Peak Intensity to Test for Activations in Functional Imaging. *Neuroimage*, 5:83–96, 1997.
7. J. S. Milton, J. C. Arnold. Introduction to probability and statistics. Principles and applications for engineering and the computing sciences. 3rd edn. McGraw-Hill, Inc. New York, 1995.
8. J. Besag. On the statistical analysis of dirty pictures. *J. R. Statist. Soc. B.*, 48:259–279, 1986.
9. D. Ekatodramis, G. Székely, G. Gerig. Detecting and inferring brain activation from functional MRI by hypothesis-testing based on the Likelihood ratio. In: M. W. Wells, A. Colchester, S. Delp (eds.). Medical Image Computing and Computer-Assisted Intervention - MICCAI'98. *Lecture Notes in Computer Science*, Vol. 1496. Springer-Verlag. Berlin 578–589, 1998.
10. K. J. Friston, P. Jezzard, R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1:153–171, 1994.
11. B. A. Ardekani, J. Kershaw, K. Kashikura, I. Kanno. Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. *IEEE Trans. Medical Imaging*, 18:101–114, 1999.