How to control the false positive rate in fuzzy clustering: Application to fMRI activation detection using a hemodynamic-based feature space

Hesam Jahanian[1, 2], Gholam A. Hossein-Zadeh,[1] Hamid Soltanian Zadeh[1, 2, 3]

[1]Control and Intelligent Processing Center of Excellence, Elec. Eng. Dept., University of Tehran, Tehran, Iran
[2]Signal and Image Processing Group, School of Intelligent Systems, IPM, Tehran, Iran
[3]Image Analysis Lab. Radiology Dept., Henry Ford Health System, Detroit, MI, USA

ABSTRACT

FCM suffers from some drawbacks such as a priori definition of number of clusters, unidentified statistical significance of results, and instability of results when it is applied on raw fMRI time series. Using the randomization we developed a method to control the rate of false positive detection in FCM which gives a meaningful statistical significance to the results. Making use of it, we derived the optimum number of clusters. In this study we applied the FCM on a feature space that takes the variability of hemodynamic response function into account and compared it with the cross correlation feature space.

KEY WORDS

fMRI, fuzzy clustering, statistical test, Hemodynamic response, randomization, cluster validity.

1. INTRODUCTION

Deoxygenated hemoglobin acts as an endogenous paramagnetic agent. Therefore, a reduction in the concentration of deoxy-hemoglobin increases the T2* weighted magnetic resonance signal. Based on this, functional magnetic resonance imaging (fMRI) measures changes in blood oxygenation and blood volume brought about by neural activity of the brain while a subject is performing some cognitive or motor task.

The majority of fMRI practitioners currently use statistical techniques such as t-test or cross correlation to determine whether voxels of the brain show task related signal variation. In statistical methods, the resulting activation map is usually characterized with a significance level which determines the rate of false alarm occurrence (type I error). To compare such statistical methods, one should compare the results obtained with the same false positive rate.

The main drawback of these methods is their assumptions or models of the noise structure, the statistical behavior of fMRI data, or activation procedure which may not be true. These assumptions bias the results obtained by such methods to the specific results especially in the cases where experimental conditions become more complex or when applied to data of different subjects [7].

Beside these model-based statistical methods, some model-free methods such as PCA, ICA, cluster analysis, and self organizing maps have been used [8]. In neuroimaging model-free analysis has been mostly carried out using clustering methods. The aim of clustering techniques is identifying regions with similar patterns of activation. They partition the brain voxels in some predefined number of clusters and one cluster will be chosen as the active cluster. Different clustering methods, such as k-means, Kohonen clustering neural network, and hierarchial clustering have been used in this field, but the most popular method has been fuzzy C-means (FCM) [2]. FCM gives the membership map of brain voxels in different clusters. After FCM convergence, the cluster with the most similar centroid to stimulation pattern is selected as the active cluster and the membership degrees of image voxels to this cluster (u) is compared with a threshold u_a in order to detect activated voxels.

Defining the right number of clusters is one of the main issues in clustering brain voxels. For this purpose some cluster validity measures have been proposed but this intensive search for a standard index has not yet succeeded [13]. Another drawback of FCM and other clustering techniques is their inability to assign statistical significance to the results. For example choosing different number of clusters, or thresholding the membership degree with different u_a's, lead to considerably different activation maps. Each result corresponds to a specific but unknown level of confidence. In other words, choosing a high u_a or a high number of clusters decreases the probability of false detection. As a result, one cannot compare the results obtained by statistical methods and clustering methods.

In order to limit number of false positives in clustering of fMRI data, Jarmasz *et. al* assumed a linear model for time series of each cluster. Each time series of a cluster is

considered as the center of cluster multiplied by a correlation coefficient plus a residual sequence. Then significance of correlation coefficient is checked [14]. Baumgartner *et. al* did the same significance test through resampling the centers of clusters in time domain to avoid the model assumption [11]. Aufferman *et. al* proposed a method using bootstrap and Fisher's linear discriminant function, which relies on the multivariate normal assumption to assess the statistical significance associated with partitioning one cluster into two clusters or the inverse problem of combining two clusters into one cluster [12].

Here we propose a method based on randomization to evaluate the statistical significance of activation and to control the false detection rate in the fuzzy clustering analysis of fMRI. Making no specific assumption about the noise structure, the randomization procedure can provide the distribution of "the membership degree to the active cluster (u)" under the null hypothesis (resting state condition). Using this probability density function, we can determine u_a in order to control false positive rate.

We also suggest a method for determining the number of clusters using the procedure we introduced for false positive control. However, the procedure of controlling the false positive rate is independent from the number of clusters and the number of clusters can be found via any other approach.

Clustering on the raw time series is potentially able to separate cognitive or hemodynamic effects without precisely modeling them. However, due to high noise level in fMRI experiments, the results of clustering on the raw time series is often unsatisfactory and does not necessarily group data according to the similarity of their pattern of response to stimulation. An associated concern is that increasing the size of clustering space leads to practical difficulties such as curse of dimensionality [2, 10, 16].

Goutte *et. al* considered a feature space based on correlation between time pattern of stimulus and time series. They showed that clustering this feature space yields noise reduction, improved performance and robustness [2,10]. Therefore they assumed a fixed reference as the time pattern of activation to construct the feature space. However, the actual functional response which may differ in various brain areas, different subjects, and under different conditions even in a simple visual or motor task, is far more complicated than the usually assumed boxcar waveform [7]. Here we have used a feature subspace which takes into account these variability and compared it systematically with cross correlation feature space.

2. DATA

A. Simulated Dataset

For a realistic simulation of fMRI data, computer generated 'activation' time series were added to the measured time series of a single slice of a resting state experimental fMRI data in 116 voxels and with different contrasts (1%, 1.5%, 2%, and 2.5%). The activation time series was obtained by convolving a stimulation pattern (a boxcar function with five periods of 60 seconds ON and 90 seconds OFF) with a Gamma function that models the hemodynamic response function (HRF). In order to model the variability of the HRF, the parameters of the Gamma function were varied randomly between different activated voxels. Fig. 1 shows the spatial location of the active voxels.

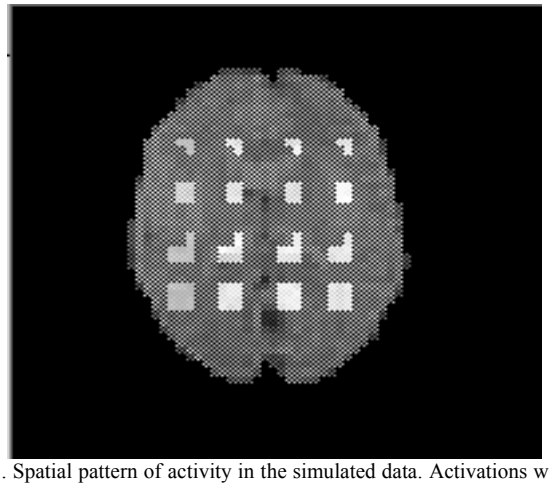

Fig. 1. Spatial pattern of activity in the simulated data. Activations were added to the dataset in the regions shown. The activation contrasts for the columns (from left to right) are 1%, 1.5%, 2% and 2.5% , respectively.

B. experimental Dataset

Functional images were acquired from 6 normal volunteers using a single-shot GRE spiral scan sequence (TR=2 sec, TE=30 ms, FOV=220×220×96 mm^3, matrix size=64×64×24) on a 3 Tesla GE MRI scanner (General Electric, Milwaukee, WI, USA). The subject performed a finger tapping task with both hands. The task consisted of 12 periods of 36 seconds, where each period contained 18 seconds of finger tapping, followed by 18 seconds of rest. The first four volumes of the functional images were discarded and the remaining volumes were motion corrected using the AFNI software package [6]. Linear drifts and mean components were then removed from each voxel time-series.

3. Methods

Our proposed method consists of three steps. First, a set of features are extracted for each fMRI time series. This step will be explained in Section A. In the second step, FCM will be applied on proposed feature space for different number of clusters in order to select the optimum number of clusters using the method described in Section C. Finally, FCM will be applied with the optimum number of clusters. After FCM convergence, the cluster with the most similar centroid to the stimulation pattern is selected as the active cluster. Then, a statistical

membership threshold (u_a) corresponding to desired false alarm rate will be computed using the method proposed in Section B. Then, the membership degree of each voxel to the active cluster (u) is compared with threshold u_a and voxels which have greater "membership degree to the active cluster" than u_a will be considered as active voxels.

A. Feature Extraction

Clustering raw fMRI time series may lead to stability problems and the risk of clustering on the noise rather than on the activation because of poor fMRI signal to noise ratio. Therefore, feature spaces based on cross correlation of a fixed reference time pattern and fMRI time series has been used as a proper feature space for cluster analysis of fMRI [10]. However the hemodynamic response function (HRF) of brain has been shown to vary significantly between different brain areas or subjects [5].

The Gamma hemodynamic response function, commonly used in statistical analysis of fMRI, includes two unknown shape parameters that are usually selected a priori by the analyst. Hossein-Zadeh *et. al* [1] proposed a new method that approximates the Gamma HRF over a wide range of parameters by a linear combination of three elementary signals. These elementary functions were derived from singular value decomposition of a large number of signals generated by systematically varying the parameters of gamma function. The elementary signals together accounted for 99% of the total variation in the data. Figure 2 shows these signals. Convolving these elementary signals with the stimulation pattern provides three basis functions ($z_1(t)$, $z_2(t)$, $z_3(t)$) for signal subspace. Therefore each fMRI time series may be considered as eq (1) where $e(t)$ is the error term considered as noise.

$$y(t) = \alpha_1 z_1(t) + \alpha_2 z_2(t) + \alpha_3 z_3(t) + e(t) \qquad (1)$$

The unknown coefficients α_1, α_2, and α_3 may be obtained for each voxel through least squares (LS) estimation. These coefficients along with a conventional cross correlation coefficient cc, (the cross correlation between $y(t)$ and the stimulation pattern) is proposed as a feature space for FCM clustering. We call this feature space HRF-based feature space. Considering the ability of the elementary functions to model the hemodynamic response variability the coefficients α_1, α_2, and α_3 are supposed to provide appropriate features for clustering.

B. False Alarm Rate Control

After FCM convergence the cluster with the most similar centroid to stimulation pattern is selected as the active cluster and the membership degrees of each voxels to this cluster (u) is compared with a threshold u_a in order to detect activated voxels. This threshold strongly affects the results significance. But it has been chosen a priori and heuristically by investigators till now. By comparing u at each voxel with u_a one tests the null hypothesis H_0:'no activation', and rejects it if $u>u_a$. For controlling the type I error of this test at level α, the threshold u_a must be found such that $prob(u>u_a \mid H_0) = \alpha$. This requires the probability density function (pdf) $f_u(u|H_0)$, which is difficult to derive theoretically. We propose a method based on randomization for finding this pdf. In this research, we use the resampling procedure introduced by Bullmore *et. al* [9], which permutates the wavelet coefficients of fMRI time series in order to make surrogate data under the null hypothesis. The wavelet coefficients (obtained using Daubechies basis with 4 vanishing moments) of the fMRI time series are permutated at different levels of resolution (in 4 levels), and then an inverse wavelet transform is applied on them to generate various realizations of data under null hypothesis

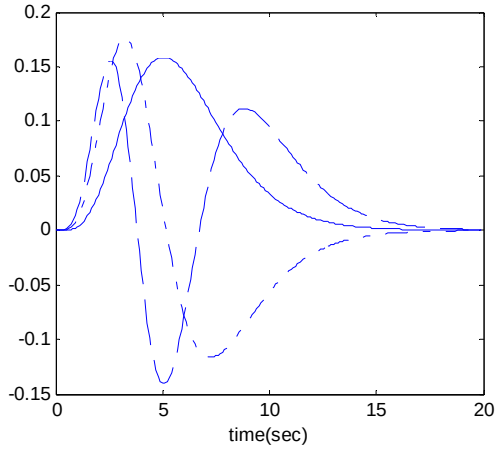

Fig. 2. Convolving the above elementary signals with the stimulation pattern produces three basis functions ($z_1(t)$, $z_2(t)$, $z_3(t)$).

FCM clustering is then applied on each set of randomized data while we hold the center of active cluster found before randomization unchanged, and then the membership degrees of all voxels in the active cluster will be computed. These values construct an empirical histogram which estimates the required pdf $f_u(u|H_0)$. Using this histogram one finds a proper threshold corresponding to the desired α. Thresholding the active cluster membership degree map of brain voxels with this threshold generates statisticaly meaningful results.

C. Number of Clusters

Logically choosing the optimum number of clusters in FCM leads to the most accurate detection of fMRI activation. The area under the Receiver Operating Characteristics (ROC) curve is commonly considered as a good criterion for characterizing the detection accuracy. We are facing two issues in using ROC curves in fMRI data analysis with fuzzy clustering: first we can not control the false alarm rate in activation detection via fuzzy clustering; second, there is no way to measure true positive detections when applying the method on experimental fMRI data. The first issue has been addressed with the method described in pervious section. To overcome the second issue, we used the fact that truly activated voxels tend to be spatially clustered, while falsely activated voxels will tend to be scattered so that one does not expect random spatial activations. These scattered voxels mainly appear as single voxels which are

treated in many investigations as false detections and they are removed from the results [15]. We used the number of detected single voxels (voxels with no activated neighbors) as a criterion for estimating the false positive detection in experimental data. In fact based on spatial connectivity of active voxels, we are looking for the number of clusters that produces the most compact activation regions with less single voxels.

For a particular number of clusters, we do the following setps; first we apply the method proposed in the previous section for various amounts of α in order to find their corresponding thresholds; using these thresholds then we find the corresponding active regions by thresholding the active cluster membership map obtained from fuzzy c-means clustering (FCM); Next an estimate of true positive detections is made by excluding the single voxels and counting the remaining voxels. We use these estimates in order to derive an estimate of β for different values of α. This produces a ROC curve for the specified cluster number. The area under this ROC curve in the interval [0 0.1] (the common interval for alpha used in fMRI) is used as the cluster validity measure. By performing these steps one can measure the cluster validity for different number of clusters and then select the optimum number which has the maximum measure.

4. Results

An estimate of the false alarm rate of an fMRI detection method can be made by applying the method to the resting state data. In order to provide the resting state data, time series of activated voxels were discarded from each of 6 fMRI experimental data. After computing the cross-correlation map for each data, the active voxels were detected for false alarm rate of 0.1, and their time series were discarded from the data. This ensures us that the remaining voxels are in the resting state. The proposed method, explained in Section 3-B, was applied on each resting state data, and activated voxels were detected by assuming different false alarm rates. An estimate of the actual (occurred) false alarm rate is then made in each case by dividing the number of detected voxels to the number of voxels in the analyzed resting state data.

Fig 3 graphs the expected false alarm rate versus the observed (measured) false alarm rate for one of the 6 subjects. Table 1 shows the numerical values of theses parameters for all 6 subjects. This figure demonstrates the ability of our proposed method to control the false positive rate. In fact, using the pdf of *u* under the null hypothesis for choosing the threshold is the main foundation of false positive control. One of the estimated pdf's has been shown in Fig 4.

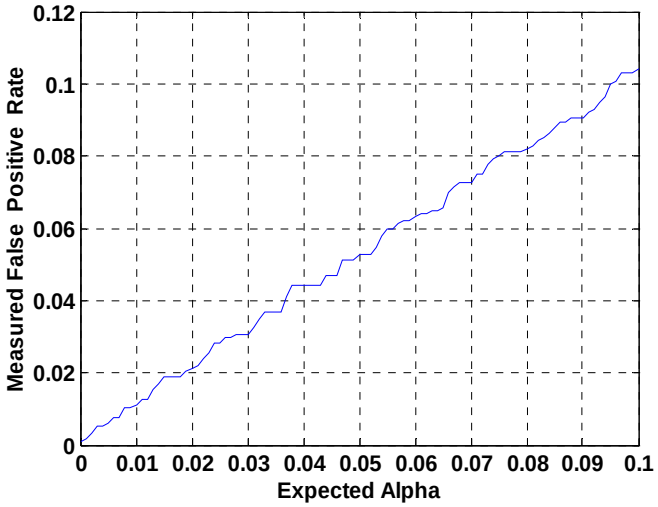

Fig. 3.The measured false positive rate versus its expected value in one of the 6 subjects.

Table 1.Numerical values for Expected alpha value versus observed false alarm rate for 6 subjects.

alpha	subject 1	subject 2	subject 3	subject 4	subject 5	Subject 6
0.01	0.0102	0.0108	0.0102	0.0119	0.0111	0.0111
0.02	0.0196	0.0197	0.0196	0.0222	0.0209	0.0213
0.03	0.0307	0.0307	0.0299	0.0324	0.0298	0.0307
0.04	0.041	0.0418	0.0392	0.0469	0.0444	0.0444
0.05	0.0503	0.0512	0.0496	0.0503	0.0529	0.0518
0.06	0.0597	0.0614	0.0597	0.064	0.0631	0.0631
0.07	0.07	0.069	0.07	0.0694	0.0725	0.0725
0.08	0.0802	0.0811	0.0785	0.0833	0.0811	0.0819
0.09	0.0896	0.0904	0.0887	0.093	0.0904	0.0904
0.1	0.0998	0.1024	0.099	0.1058	0.1038	0.1041

We have also examined our method for defining number of clusters on experimental dataset, and compared it to the results of SCF cluster validity measure proposed by fadili *et. al* [3]. In 4 out of 6 subjects two methods derived the same number of clusters, whereas in 2 subjects their proposed "number of clusters" were different by 1. However, by repeating the procedure, our method shows less sensitivity to the initial values of FCM. Fig 5 shows the ROC curves, corresponding to one of experimental data, obtained by different number of clusters. This graph suggests N=6 as the optimum number of clusters.

Although our proposed method for false positive control can be used in applying the FCM on any kind of feature space, we have shown that the HRF-based feature space provides improved detection sensitivity over

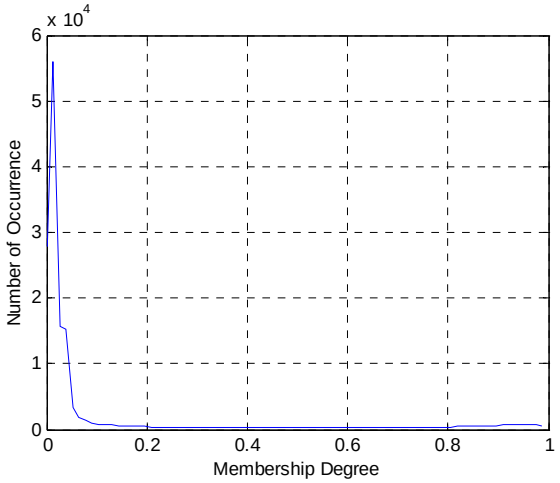

Fig. 4. Empirical histogram of "membership degrees to the active cluster" under the null hypothesis, obtained by randomization in one of experimental data sets. This histogram has been used as an estimate for $f_u(u|H_0)$ in that subject.

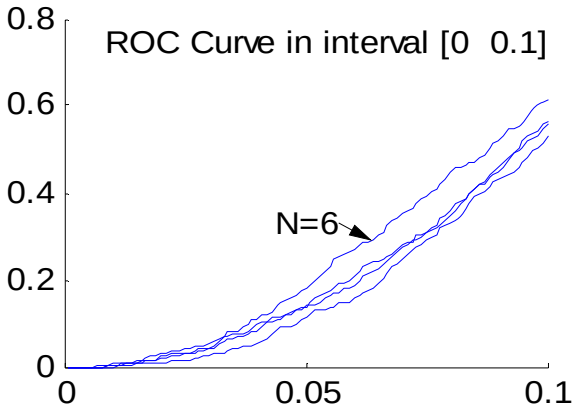

Fig. 5. The ROC curve for N=2,4,6,7.

the cross-correlation feature. FCM clustering activation detection with controlled rate of false alarm was applied on both simulated and experimental fMRI data using both feature spaces. m=2 suggested in [4] was used as fuzziness index of FCM. In simulated data, where an ROC curve can be derived, the HRF-based feature space demonstrates an improved sensitivity (Fig. 5).

Finger-tapping paradigm regularly produces activation in the sensorimotor cortex (SMC), supplementary motor area (SMA), and cerebellum. Activity in the sensorimotor cortex produces transient neural activity in subcortical regions [5]. Moritz et. al reported activation detection in subcortical regions by changing the temporal duration of the reference function [5].In the experimental fMRI data, using HRF-based feature space revealed activation in sub-cortical regions where the cross-correlation feature failed to detect them. Table 2 shows the activated regions of both feature spaces, and Fig. 6 shows an example of such a case. These results are consistent with the study performed by Moritz *et. al* [5].

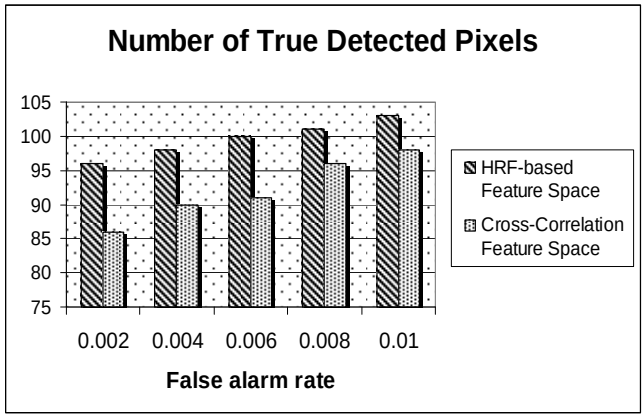

Fig. 5. Comparison of the number of correctly detected active voxels (true positives) for HRF-based feature space compare to cross-correlation coefficient feature space at different false alarm rates.

Table 2. Number of subjects who showed activation in specific regions for different feature spaces.

Detected Activation Region	HRF-based Feature Space	Cross Correlation Feature Space
SMA	6	5
SMC	6	6
Cerebellum	6	6
Putamen	2	0
Thalamus	3	0
Temporal Gyrus	2	0

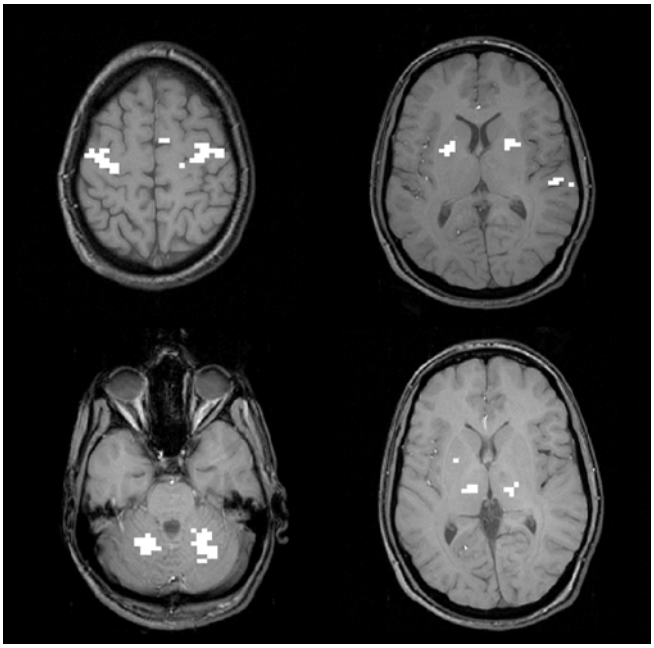

Fig. 6. Activation regions detected by the proposed method, overlaid on the corresponding anatomical slices. Activation is detected in SMC, SMA, thalamus, cellebrum, putamen, and temporal gyrus at α =0.005.

5. CONCLUSION

A method for controlling false positive rate in FCM was proposed and its efficiency was evaluated by activation detection with FCM on 6 rest fMRI data. Fixing the false positive rate in activation detection using FCM, makes it possible to compare the FCM with other fMRI activation detection methods. One can also evaluate the performance of different FCM-based methods, such as using different feature spaces. An exact comparison between the above methods can not be made without considering the statistical significance of the results. The proposed method controls the rate of false positive occurrence without any assumption about the noise or activation pattern at the expense of computational complexity of randomization. Using this method, we compared two feature spaces: cross correlation feature space; and HRF-based feature space. Our comparison using simulated and experimental data showed improved sensitivity of HRF-based feature space over the cross correlation feature space. In the analysis of 6 finger-tapping fMRI data, activation was detected in sub-cortical regions using HRF-based feature space, where the cross-correlation feature space failed to detect them.

REFERENCES

Proceedings Papers:

[1] G. Hossein-Zadeh, and B. A. Ardekani, A signal subspace for modeling the Hemodynamic Response Function in fMRI, *Proceedings of the 10th ISMRM Conference,* Honolulu, Hawaii, USA, 2002, 1430.

Journal Papers:

[2] C Goutte, L. Hansen, M. Liptrot, & E. Rostrup, Feature-Space Clustering for fMRI Meta-Analysis, *Human Brain Mappin ,13*, 2001, 165-183.

[3] M. J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer, A Multistep Unsupervised Fuzzy Clustering Analysis of fMRI Time Series, *Human Brain Mappin ,10*, 2000, 160-178.

[4] M. J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer, On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series, *Medical Image Analysis* 5 (2),2001,55-67.

[5] C.H. Moritz, M.E. Meyerand, D. Cordes, V.M. Haughton, Functional MR Imaging Activation after Finger Tapping Has a Shorter Duration in the Basal Ganglia Than in the Sensorimotor Cortex, *AJNR 21,*2000,1228-1234.

[6] R.W. Cox and J.S. Hyde, Software tools for analysis and visualization of fMRI data, *NMR Biomed.*, 10, 1997, 171-178.

[7] K. Chuang, M. Chiu, C. Lin, and J. Chen, Model-Free Functional MRI Analysis Using Kohonen Clustering Neural Network and Fuzzy C-Means, *IEEE Trans. Medical Imaging, 18*(12), 1999, 1117-1128.

[8] R. Baumgartner, L. Ryner, W. Richter, R. Summers, M. Jarmsaz, R. Somorjai, Comparision of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis, *Magnetic Resonance Imaging, 18,* 2000, 89-94.

[9] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter, and M. Brammer, Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains, *Human Brain Mappin ,12*, 2001, 61-78.

[10] C. Goutte, P. Toeft, E. Rostrupt, F. Nielsen, L. Hansen, On Clusteing fMRI time series, *Neuroimage, 9,* 1999, 298-310.

[11] R. Baumgartner, R. Somorjai, R. Summers, W. Richter,L. Ryner, and M. Jarmasz, Resampling as a Cluster Validation Technique in fMRI, *journal of Magnetic Resonance Imaging, 11,* 2000, 228-231.

[12] William F. Auffermann, S. Ngan, and X. Hu, Cluster Significance Testing Using the Bootstrap, *NeuroImage, 17,* 2002, 583–591.

[13] U. Mo¨ller, M. Ligges, P. Georgiewa, C. Grunling, W. A. Kaiser, H. Witte, and B. Blanz, How to Avoid Spurious Cluster Validation? A Methodological Investigation on Simulated and fMRI Data, Neuroimage, *17*, 2002, 431-446.

[14] M. Jarmasz, R. Somorjai, and R. Baumgartner, A new statistical inference test for fMRI time-series, *Neuroimage, 13,* 2001, s163.

[15] O. Friman, J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson, Detection of neural activity in functional MRI using canonical correlation analysis, *Magn. Reson. Med. (45),* 2001, 323-330.

Books:

[16] S. Theodoridis, K. Koutroumbas, *Pattern Recognition* (San Diego: Academic Press, 1998).