Magnetic Resonance Imaging, Vol. 17, No. 6, pp. 817-826, 1999

Printed in the USA.
0730-725X/99 $–see front matter

ELSEVIER

PII S0730-725X(99)00014-4

● *Original Contribution*

A HIERARCHICAL CLUSTERING METHOD FOR ANALYZING FUNCTIONAL MR IMAGES

PETER FILZMOSER,* RICHARD BAUMGARTNER,† AND EWALD MOSER‡

*Department of Statistics and Probability Theory, Vienna University of Technology, A-1040 Vienna and †NMR Group, Institute for Medical Physics, and ‡Clinical MR-Unit, University of Vienna, A-1090 Vienna, Austria

We introduce a novel method for detecting anatomic and functional structures in fMRI. The main idea is to divide the data hierarchically into smaller groups using *k*-means clustering. The separation is halted if the clusters contain no further structure that is verified by several independent tests. The resulting cluster centers are then used for computing the final results in one step. The procedure is flexible, fast to compute, and the numbers of clusters in the data are obtained in a data-driven manner. Applying the algorithm to synthetic fMRI data yields perfect separation of "anatomic," i.e., time-invariant, and "functional," i.e., time-varying, information for a standard off-on paradigm and a typical functional contrast-to-noise ratio of two and higher. In addition, an EPI-fMRI data set of the human motor cortex was analyzed to demonstrate the performance of this novel approach in vivo. © 1999 Elsevier Science Inc.

Keywords: **fMRI; Clustering; k-means; Synthetic data; Human brain; Motor cortex.**

INTRODUCTION

Recently, functional magnetic resonance imaging (fMRI) has been established as a potent brain mapping method (for review see, e.g., ref. 1-3). Upon physiological stimulation of the various senses (e.g., acoustic, motor, olfactory, sensory, visual, etc.), a focal hemodynamic response may be observed via changes in cerebral blood flow, volume and/or oxygenation. The actual signal enhancement and the size of activated areas depends, in addition to the stimulation paradigm and the actual hemodynamic response function of the subjects brain, on parameters such as magnetic field strength (currently about 1-4 T), measurement sequence design, (mostly gradient-recalled echo, echo-planar or spiral imaging) and measurement parameters. In particular, functional contrast depends on the spatial resolution, the exitation flip angle, the echo time, and the repetition time, which defines the temporal resolution. Nevertheless, the signal enhancement in cortical brain areas may be only in the lower percent range, whereas artifactual signal changes from a) large draining veins, b) gross head motion, or c) technical artifacts may be much higher.[4-6] Furthermore, physiologic motion also contributes to the noise level, however, the noise level varies across the brain.[7-9]

Consequently, the extraction of information primarily related to focal neuronal activation is still a formidable task. In addition, due to the large number of non-parametric distributed data points per study, typically about 5-100 million, statistical treatment is non-trivial. Historically, simple statistical tests were applied to visualize so called "activated brain maps" and corresponding signal time courses were calculated separately.[10] However, these simple approaches, although broadly used, require a high degree of prior knowledge, and have a low specificity to differentiate contributions from large, draining veins and cortical activation,[11-14] as well as stimulus-correlated brain motion.[6] Recently, more elaborate, multivariate statistical methods have been introduced.[15-18] However, they still require a high degree of prior knowledge; not only about the paradigm but also on the actual (i.e., individual and maybe local) hemodynamic response function which may not be generally known.

RECEIVED 8/30/98; ACCEPTED 3/7/99.

Address correspondence to Dr. Ewald Moser, Arbeitsgruppe NMR, Institut f. Medizinische Physik, Universität Wien, Währingerstrasse 13, A-1090 Wien, Austria. E-mail: ewald.moser@univie.ac.at

Present address of R. Baumgartner c/o: Institute for Biodiagnostics, NRC, Winnipeg, Canada.

The latest development in advanced signal processing in fMRI concerns the application of so-called explorative, paradigm-free approaches such as factor analysis,[19,20] and fuzzy-cluster analysis.[13,14,21,22] These methods aim to extract the full amount of information available from the acquired data sets without model bias. This may include actual brain activation as well as physiologic or technical artifacts leading to time-varying signal changes in fMRI time series. However, every single approach has advantages and disadvantages, and has to be selected properly in connection with the actual data set and quality.

fMRI data is analyzed to detect the main structures which are caused by different signal intensities and temporal changes. This leads to variations in the data matrix where groups of data points describing different (main) signal intensities in the image are expected. The analysis of grouped data is not advisable with correlation-based methods since, in general, the assumption of normal distribution is not fulfilled. A better procedure may be to separate the various groups using the actual data directly, which may be achieved by cluster analysis, applying one of the numerous clustering techniques published (see, e.g., ref. 23). Since in fMRI the number of observations or pixel points n is very large, we can exclude methods which are based on the matrix of distances (dimension $n \times n$) instead of directly using the data matrix. Another objective is to set up a procedure that leads to the final result in real-time, so time consuming clustering algorithms should be avoided.

Here we present a novel approach for an explorative, model-free fMRI analysis strategy based on hierarchical clustering employing crisp k-means clustering. The method is being evaluated using synthetic, and in vivo fMRI data.

MATERIALS AND METHODS

Synthetic fMRI Data

Thirty-five images, representing transversal brain slices (full matrix size 128×128), were simulated by a time invariant anatomy (Fig. 1a) with signal variations at the location of activated areas as shown in Fig. 1b ("A" denotes "cortical" and "B" denotes "venous" activation.) The number of pixels in area "A" was NA = 35 and in area "B" was NB = 14. The anatomy consists of two structures; 1) a checkered pattern representing lower signal intensity (SI) (SI = 204, n = 2,268) and higher signal intensity components (SI = 241, n = 2,268) simulating gray and white matter, respectively, and 2) the (fluid filled) ventricles also represented by a checkered pattern with lower and higher intensity component (SI = 156, n = 145 and SI = 192, n = 145, respectively). The data were corrupted by Gaussian white noise resulting in a difference in functional contrast-to-noise ratio (CNR) in the range of 1.33-2.33 between anatomy (Fig. 1a) and regions "A" and "B" (Fig. 1b). In addition, a random baseline shift simulating thermal shifts in a temporal image series was introduced, not exceeding 1% of the baseline level. For more details about the synthetic data design see ref.[20]

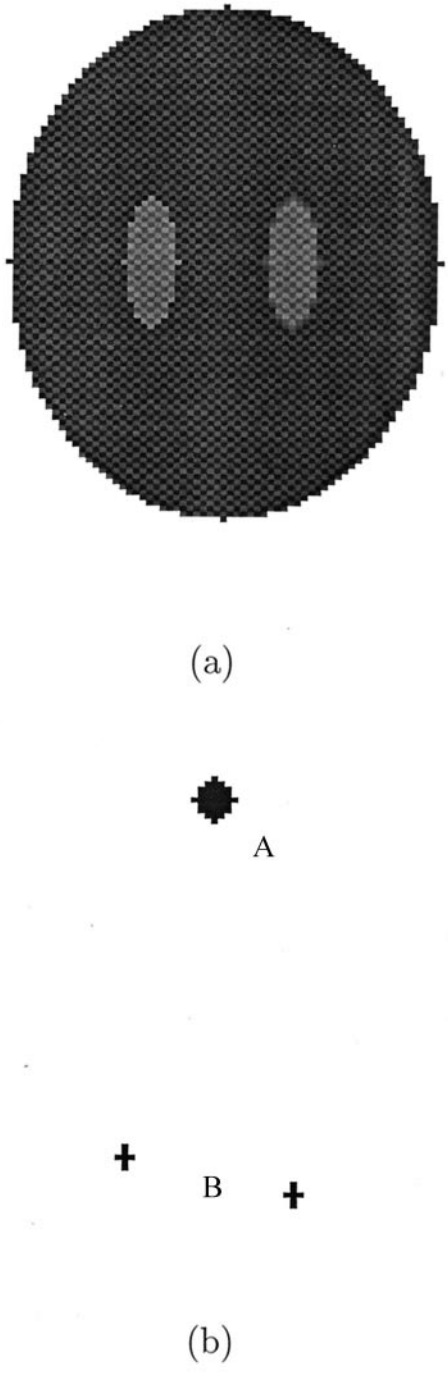


Fig. 1. In a, anatomic details of the synthetic data set; b, simulated activated areas.

In Vivo fMRI Data

Functional MRI data from the human motor cortex were measured on a whole body MR-system at 3T (BRUKER Medspec 30/80-DBX, Ettlingen, Germany), equipped with a high performance gradient coil (B-GA 55; 19 mT/m in $< 300\ \mu s$) using a quadrature bird-cage head coil. A single-shot blipped GRE-EPI sequence was applied to collect 70 T_2*-weighted sequential images. Sequence parameters were as follows: TE = 82 ms, TR = 1000 ms, RBW = 200 KHz, MA = 128 × 128, FOV = 256 × 256 mm, thk = 3 mm. The reference gray scale image (see Fig. 5) has been created by averaging all 70 EPI images (without any registration). This also confirms that no distorting motion artifacts are present in the whole image series.

Clustering Algorithm

A clustering technique that is very fast to compute is the *k*-means method.[24] The objective of *k*-means is to find a partition of the observations into *G* groups by minimizing a predefined criterion. Starting from *G* initial guesses for the cluster centers, each object is transferred to another group until an "error measure" (e.g., the sum over the squared distances from each observation point to its cluster center) cannot be further reduced. As it would be computationally infeasible to calculate the overall minimum of the objective function, a procedure for finding a local minimum may be used.[25] In general, the *k*-means method is not sensitive to the choice of the initial cluster centers. The simplest approach is to take the first *G* observations of a data matrix. It is, however, important to know a priori the number *G* of clusters, which, in general, is not the case with fMRI data. The problem of the choice of *G* can be avoided if the whole *k*-means clustering is done for an arbitrary number of clusters. Thereafter, one may select the "best" or most satisfying result by whatever criteria. This procedure takes a lot of time, especially for large data sets and produces a lot of output that needs to be further analyzed. Therefore, we suggest a kind of hierarchical *k*-means clustering technique that divides the data into two groups in each clustering step. The procedure is halted after there is no structure left in the data, i.e., it is a purely data driven approach.

Suppose we have acquired a time sequence of p images. Each gray scale image consists of 128 × 128 pixels. The $n = 128 \times 128 = 16384$ pixels can be seen as the observation points for the time sequence of images, resulting in a data matrix $\boldsymbol{X}$ of size $n \times p$. The goal now is to find the main structure contained in $\boldsymbol{X}$. Either different signal intensities in pixels located in different tissue types (i.e., amplitude or "anatomy" based) or changes in signal intensities with time (i.e., "function" based) leading to detectable variations in the data matrix.

The algorithm proposed is divided in two main parts: the *clustering* and the *merging* part. In the clustering part we step-by-step divide clusters which contain structures into two smaller clusters, whereas in the merging part we re-combine clusters. Finally, the clusters are re-computed using the previously obtained cluster centers.

Initially, all objects (data points) of the data matrix $\boldsymbol{X}$ are combined in one big cluster. In the *first step* this cluster is divided into two smaller clusters. This is performed by the *k*-means method where the first two rows of $\boldsymbol{X}$ can be taken as initial cluster centers. As a result a "clustering vector" is obtained $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_n)^T$ containing the numbers 1 and 2, which indicate the membership to cluster 1 or 2 for each observation. Next, $\boldsymbol{X}$ is subdivided into two matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The matrix $\boldsymbol{X}_1$ is of size $n_1 \times p$ and contains all observations $\mathbf{x}_{i.}$ (i-th row of $\boldsymbol{X}$) for which $\nu_i = 2$ ($i = 1,\ldots,n$). Complementary, the matrix $\boldsymbol{X}_2$ is of size $n_2 \times p$, where $n_1 + n_2 = n$ and contains all observations $\mathbf{x}_j$ (j-th row of $\boldsymbol{X}$) for which $\nu_j = 2$ ($j = 1,\ldots n$). Commonly, for fMRI data, these two clusters correspond to the *region of interest*, i.e., the brain, and to the *background*.

In clustering *step 2*, we check for each of the clusters separately, derived from the previous step (i.e., cluster 1 and 2), whether they include further structure or just noise. There are different possibilities for this examination:

Visual inspection. Visualize the data set by, e.g., plotting the scores of the first and second principal component (see ref.[26]) of $\boldsymbol{X}$. The principal component scores correspond to the data values transformed to the coordinate system of the principal components. The (two) clusters may be displayed by assigning the elements of the clustering vector $\boldsymbol{\nu}$ to the data objects (see Fig. 2). If the data set is well structured, we can easily decide if cluster 1 and/or 2 should be further divided.

Eigenvalues. Calculate the eigenvalues of the covariance matrix of data set $\mathbf{X}_l$ ($l = 1,2$) and sort them by their magnitude. If the eigenvalues are roughly of the same magnitude, the content of information is approximately the same, which means that the data contain essentially noise (see, e.g., ref. 27).

Within cluster sum of squares (WSS). Compute the sum over the squared distances from each observation to its cluster center. This value can be adjusted by the number of objects in cluster l ($l = 1,2$). A low value of the (adjusted) WSS indicates that the data are "compact" and should, therefore, not be further divided.

Statistical analysis. After each single step in hierarchical clustering, statistical tests may be performed to compute significance levels for the different clusters obtained. However, note that simple tests to calculate

e.g., *p* values, assume normally distributed data which, in general, is not the case. Therefore, non-parametric tests such as the Kolmogorov-Smirnov test are more advisable (see, e.g., ref. 28).

Based on the guidelines given above, we can decide whether cluster l ($l = 1,2$) should be further decomposed. In the positive case we divided the cluster(s) by the *k*-means method and step 2 is finished. The final result fully describes the data by *m* clusters where *m* is 2, 3 or 4 and statistical tests confirm the significance level(s).

Experience has shown that, in general, it is not possible to give exact thresholds to the criteria for eigenvalues and WSS, which are valid for all different kinds of data. This is the reason why, in addition, a visual inspection is suggested to get a better impression of data quality and structure.

All computations were performed on a workstation HP 9,000, (Hewlett Packard, California, USA), model C180 (180 MHz) with a RISC PA 8,000 processor and 288MB RAM. We used the statistical software package S-Plus (see, e.g., ref. 29).

RESULTS AND DISCUSSION

We analyzed five synthetic fMRI data sets of size 16,384 × 35 and varying "functional" contrast-to-noise ratio (CNR) (see also ref. 30) and performed *k*-means clustering with two clusters and obtained the matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. In this example (Fig. 1), we show the data set with delta CNR = 1.66 between anatomy and regions "A" and "B." The five largest eigenvalues of the covariance matrix of $\boldsymbol{X}_1$ are

$$16020.1 \quad 82.1 \quad 54.6 \quad 52.3 \quad 52.1$$

whereas for $\boldsymbol{X}_2$ we obtained

$$17.9 \quad 17.8 \quad 17.5 \quad 17.5 \quad 17.4$$

From the magnitude of these eigenvalues it is clear that $\boldsymbol{X}_2$ includes just noise, i.e., it is the matrix describing the background, whereas $\boldsymbol{X}_1$ includes the region of interest, i.e., the brain.

The same differentiation is possible via the within cluster sum of squares (WSS):

$$WSS_1 = 8.48e + 7,\ n_1 = 4809,\ \frac{WSS_1}{n_i} = 17627.0$$

$$WSS_2 = 6.57e + 6,\ n_2 = 11575,\ \frac{WSS_2}{n_2} = 567.8$$

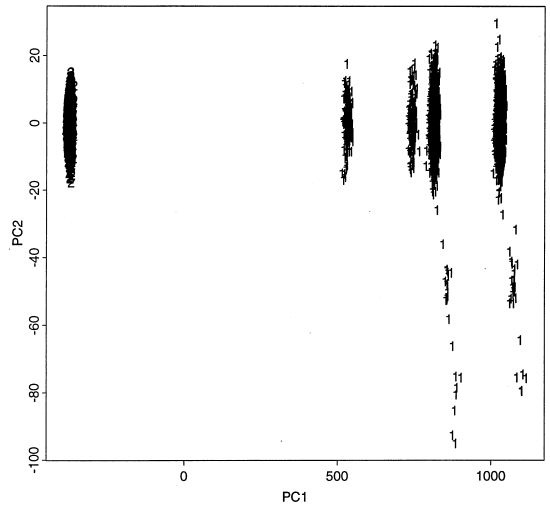
A Kolmogorov-Smirnov test of the two cluster centers gives a *p* value of 0. This means that the null hypothesis for equality of the empirical distributions of the cluster centers has to be highly significant rejected. This again confirms the previous results.

As mentioned before, this synthetic data set is composed of groups with normally distributed data points. So, for some situations, a *t*-test could also be applied since the statistical assumptions might be fulfilled. However, non-parametric tests can be used for each kind of fMRI data, and the fulfillment of normal distribution such as in this case is, of course, no contradiction for applying the Kolmogorov-Smirnov test.

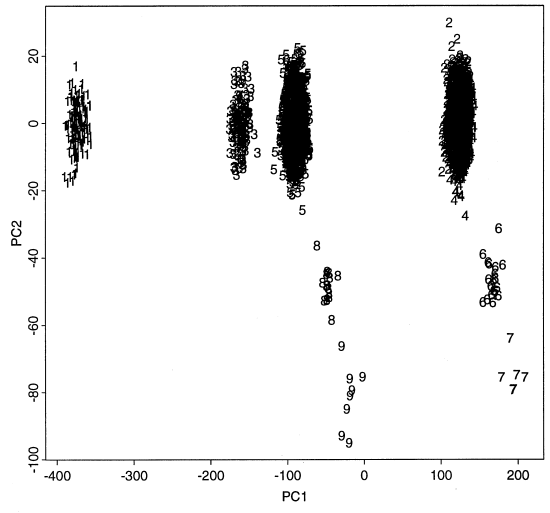
To visualize the clustering, Fig. 2a shows the scores of the first and second principal component of $\boldsymbol{X}$, with the digits representing the objects corresponding to the clustering vector. This plot confirms the above conclusions: cluster 2, denoted by the digits 2, includes no further structure (i.e., represents the background), and cluster 1, denoted by the digits 1, obviously contains information which may be further investigated. Note that principal component 1 (PC1) differentiates between signal intensities constant in time, i.e., "anatomic" structures, and principal component 2 (PC2) allows differentiation for time varying signal intensities, e.g., function.

In clustering *step t* ($t > 2$) we have to decide, on the basis of the above guidelines, for each of *m* clusters whether they should be decomposed into (two) smaller clusters. If none of the *m* clusters is further divided, we have finished the clustering part by obtaining *m* clusters with the corresponding cluster centers. During this hierarchical clustering procedure it may happen that objects belonging to one group become divided into different clusters. Note that a potential disadvantage of *k*-means clustering is that this algorithm tries to find clusters with about the same size. As an extreme case we may consider groups with very different sizes that are not well separated. This situation often appears with fMR images when only small parts of the brain are activated. Here, the larger group may be divided unequally into two clusters such that one of the new clusters will contain also points from the small, activated part. Although this smaller group can be separated from the other points in subsequent steps, it is clear that the clusters containing the objects of the original large group should be recombined.

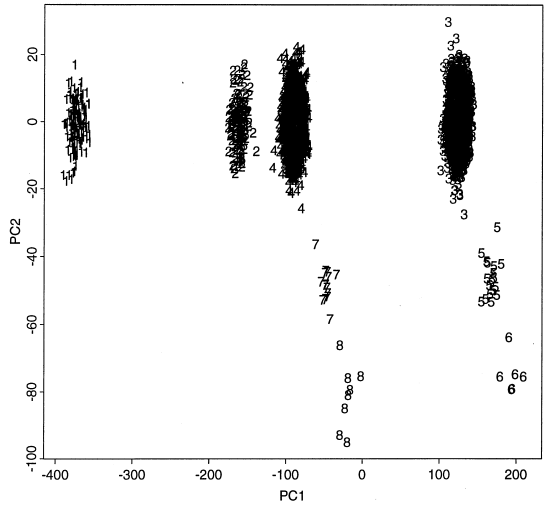
The above requirement is achieved during the second part, or merging part, of the proposed procedure. We compute all (Euclidean) distances between the *m* cluster centers obtained from the clustering part and sort them in increasing order. Small numbers indicate that the clusters are strongly related and they should, therefore, be merged. Merging clusters means that one of the two cluster centers is deleted. The merging part should be



(a)



(b)



(c)

Fig. 2. To visualize the potential of the hierarchical k-means clustering approach, principal component analysis was performed. The scores of the first and second principal component are displayed for different levels of the clustering hierarchy: a) after the first step cluster 2, denoted by digits 2, shows no further structure, whereas cluster 1, denoted by the digits 1, suggests further clustering; b) After finishing the splitting part, several anatomic (clusters denoted by digits 2-5) and functional (clusters 6-9, respectively) clusters could be separated. c) By merging cluster 2 and 4, selecting the cluster centers, and repeating the k-means clustering on the full data set, perfect separation of all simulated structures, whether anatomic or functional, could be obtained.

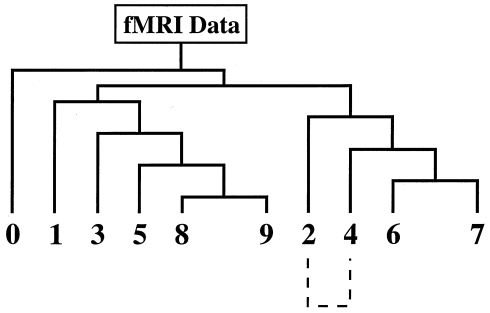
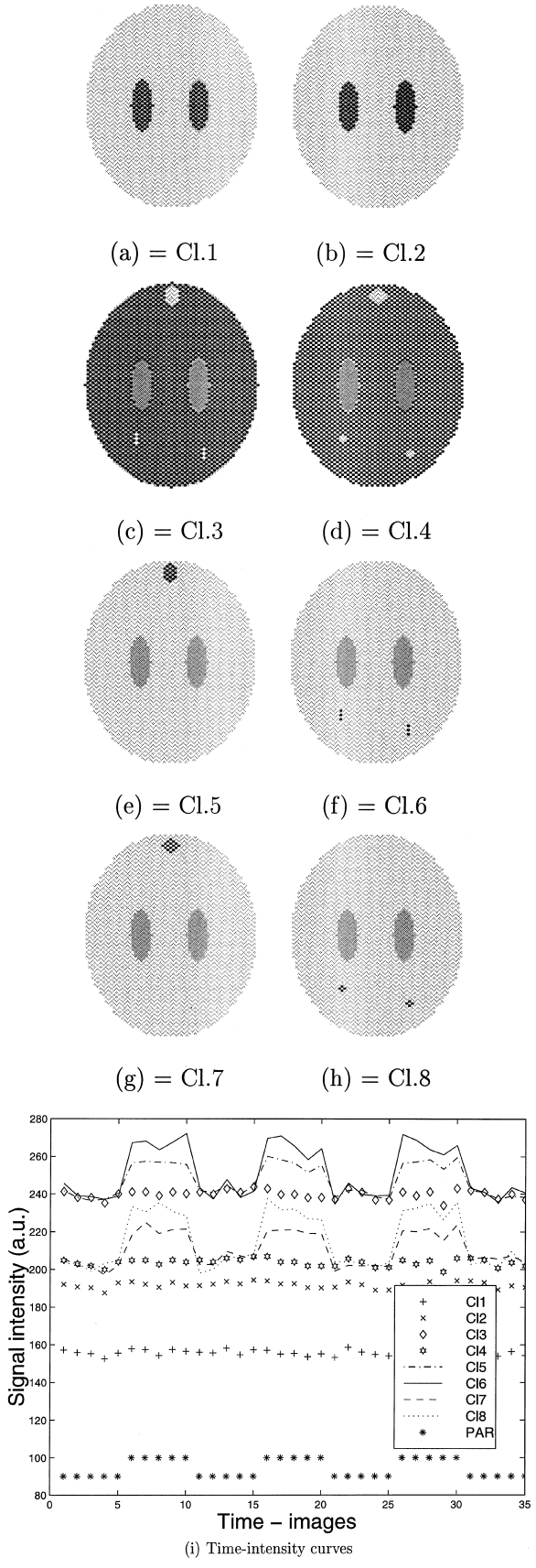


Fig. 3. All steps of clustering and merging are reported in this clustering tree. At the beginning (top of the diagram), all points are combined in one single cluster. After the first clustering step, this cluster is divided into "background" (cluster 0) and brain (all other points), and so on. The numbers 0 to 9 indicate the results after having finished the clustering part (see also Fig. 2b). In the merging part, cluster 2 and 4 are re-combined (dashed line).

stopped when there is a sudden increase of the distances for different cluster pairs.

We now continue the clustering procedure with the above-described data set. The hierarchical clustering steps are performed until none of the clusters is to be (significantly) divided any more (see *guidelines*). Figure 2b shows the scores of the first and second principal component of our synthetic fMRI data set, by just showing the data points from the region of interest (i.e., represented by the digits 1 in Fig. 2a). The digits representing the points in Fig. 2b correspond to the actual clustering vector. We note that cluster 2 and 4 are strongly related and they should, therefore, be merged into one cluster. The reason for dividing this cluster can be duplicated by inspecting the hierarchical clustering tree shown in Fig. 3. The digits 1 to 9 in Fig. 3 correspond to the clusters 1 to 9 of Fig. 2b, the digit 0 represents the cluster "background." *k*-means clustering has the tendency to extract clusters of about the same size, which caused a separation of the cluster with the



(i) Time-intensity curves

points labeled by 2, 4, 6 and 7 into one cluster with the points 2 and another cluster including the remaining points (see Fig. 2b). The relationship between the clusters may be judged by inspecting the (sorted) distances of the cluster centers:

Cluster pairs **2–4**	6–7	8–9	4–6	5–8	2–6	
	3–5	4–7	···			
Distance 12.00	41.93	46.99	60.12	65.40	68.35	
	71.75	99.70	···			

The smaller the distance between two cluster centers, the stronger the similarity between these clusters. The distance between the centers of cluster 2 and 4 is rather small in comparison to all other distances. Moreover, there is a "jump" after the first value of the sorted distances. The conclusion, therefore, is that we have to merge cluster 2 with cluster 4. In the clustering tree (Fig. 3) this is marked by the dashed line. Note again that PC1 differentiates between signal intensities constant in time, and PC2 allows differentiation for temporally and spatially varying signal intensities (Fig. 2b).

After completing the clustering and merging part we obtain $k \leq m$ clusters with the corresponding cluster centers. These centers are taken now as initial guesses for *k*-means clustering. This means that the previous procedure (clustering and merging part) just provides *k* initial cluster centers for the final clustering where *k* is the final number of clusters. The reason why we do not take the merged clusters as final clusters is that in the hierarchical clustering part some objects, e.g., due to low SNR or (functional) CNR, may be assigned into the wrong cluster. A final *k*-means clustering is able to correct for such mistakes. Figure 2c shows the results of this procedure. We note that besides merging of clusters 2 and 4 in Fig. 2b nothing else has changed. The final *k* clusters reflect the full structure of the data set. The objects of the clusters correspond to the simulated pixel points of the synthetic fMR image, whereas the cluster centers (*p*-dimensional) give information about the magnitude of this activation. In Fig. 4a-h the detected groups (clusters)

Fig. 4. Final result of the clustering procedure applied to synthetic fMRI data and overlaid on the initial ("anatomic") image (a-h). Corresponding clusters (extracted activated pixels) are given in black, overlaid on the basic anatomic structures (light and dark gray). In (i) the corresponding time-intensity curves are given. Four clusters represent time invariant structures (cl. 1-4) and clusters 5-8 represent "stimulus correlated" signal changes.

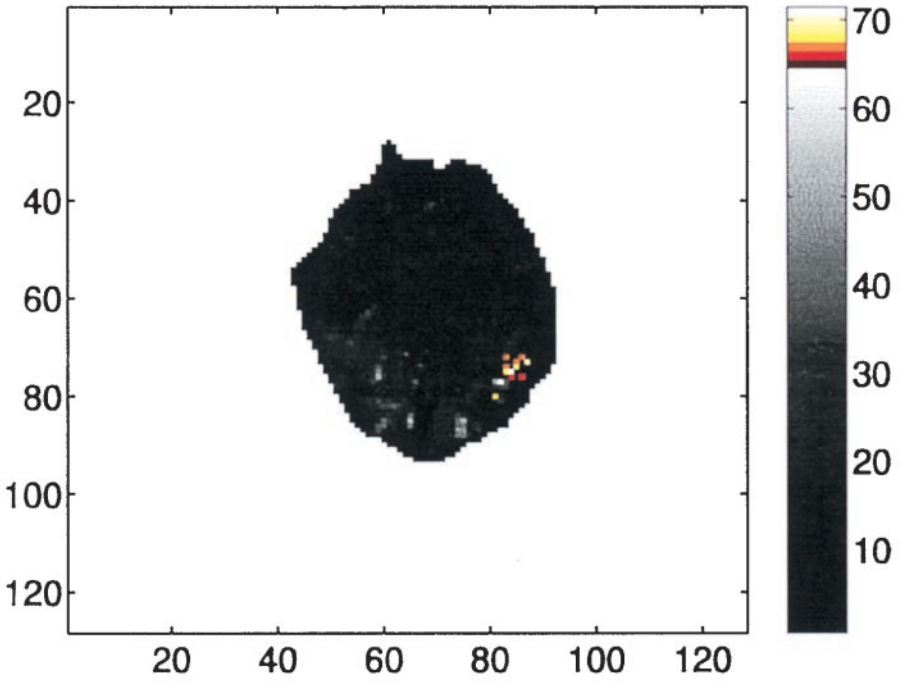


Fig. 5. Pixels of six clusters or sub-clusters reflecting the stimulation paradigm are overlaid on the anatomy. The hot-color map shows the signal enhancement (ΔS/S) of the respective cluster (red pixels belong to the cluster with the lowest signal enhancement (i.e., Cluster 4 in Fig. 6b) and the lighter the color the higher the relative signal enhancement of the corresponding cluster. See Fig. 6 and Table 1).

are overlaid onto the original synthetic fMR image. Each cluster is displayed in black in the image. Furthermore, Fig. 4i displays the corresponding signal time courses for the 8 different clusters. Four clusters represent time invariant structures (cl. 1-4) and the other four clusters (cl. 5-8) represent "stimulus correlated" signal changes (paradigm: PAR ***) on top of two different baselines (cl. 3 and 4). The comparison with the image from the originally simulated data set (Fig. 1) as well as the corresponding time courses (and signal intensities; see text) proofs that the procedure has found the exact solution. The same holds for all other data sets, i.e., for CNR $\geq$ 2.

In vivo example: In order to demonstrate the potential of hierarchical k-means clustering, an in vivo EPI fMRI data set was analyzed. During self-paced right hand finger tapping in an off/on manner (i.e., 10 images rest/10 images finger tapping etc.), a total of 70 single shot gradient-echo EPI images were acquired. Figure 5 displays the final result of the clustering (and merging) procedure described above with activated pixels overlaid on gray scale EPI image. In total six clusters of subclusters reflecting the stimulation paradigm were detected. Details, i.e., cluster time courses and corresponding PCA plots, are given in Fig. 6. Cluster 10 (subcluster 1), Figs. 6c and d, Cluster 11 (subcluster 3) Figs. 6e and f, and Cluster 16 (subcluster 1), Fig. 6g and h, represent 3 pixels with relative signal changes (ΔS/S) greater than 10%. In agreement with anatomic location (Fig. 5) and literature[4,11,13,14,31,32] this pixels very likely represent large through-plane vessels in the central sulcus. Eight more diffuse distributed pixels in the motor and premotor area (yellow and red pixels) are detected via clusters 9 (subcluster 3), 10 (subcluster 2) and 16 (subcluster 3), respectively. The results are summarized in Table 1 and Fig. 5. Note also that the temporal pattern of signal changes is quite similar for all clusters, indicating that statistical methods such as *t* test or correlation analysis[10-12,32] would not be able to allow the detailed differentiation shown with the proposed clustering technique.

SUMMARY

Cluster analysis belongs to the family of unsupervised pattern recognition techniques. It offers the possibility to recognize structures (similar observations) in a group of objects. If it is assumed that the objects or points are structured in such a way that they can be divided into some groups (clusters), these groups may be obtained by means of cluster analysis. The points should be classified on the basis of the degree of their similarity, i.e., points

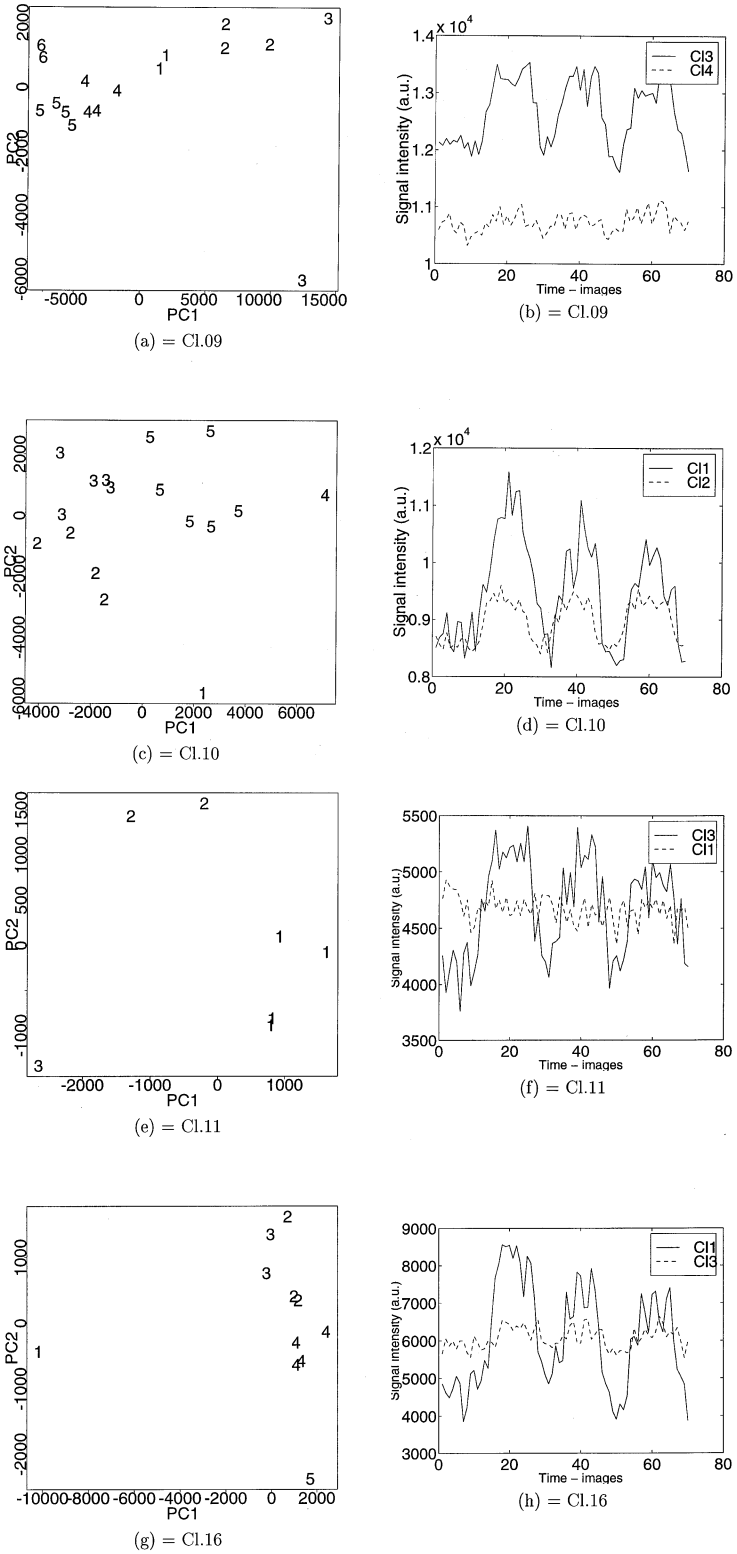


Fig. 6. Time courses and corresponding PCA plots of the clusters reflecting stimulation are presented. Cluster 10 (subcluster 1), Fig. 6c and d, Cluster 11 (subcluster 3) Fig. 6e and f, and Cluster 16 (subcluster 1), Fig. 6 g and h, represent three pixels with relative signal changes (ΔS/S) greater than 10%. These pixels may be attributed to large through plane vessels in the central sulcus. Eight more diffuse distributed pixels in the motor and pre-motor area (yellow and red pixels) belong to clusters 9 (subcluster 3), 10 (subcluster 2) and 16 (subcluster 3), respectively.

Table 1. Detailed results from hierarchical k-means cluster analysis of the EPI fMRI data set

Cluster (subcluster)	ΔS/S (%)	# of activated pixels
Cl 16 (1)	65	1
Cl 11 (3)	23	1
Cl 10 (1)	19	1
Cl 09 (3)	8.7	2
Cl 10 (2)	8.3	4
Cl 16 (3)	6	2

Clusters were selected on the basis of the (temporal) stimulation pardigm (see text and Fig. 6 for details).

of one cluster should be similar, points of different classes should be dissimilar. The Euclidean distance may be used as a distance or dissimilarity measure between two points. Proceeding from a distance matrix that indicates the distances between all points, hierarchical clustering methods (i.e., sequences of partitions) can be applied. The partitions may be constructed divisively (i.e., the clusters are subdivided in each step) or agglomeratively (i.e., clusters are merged step by step). Divisive methods are rarely applied because of the computational expense. Well-known agglomerative methods are single linkage, complete linkage, average linkage, and centroid methods.

For large data sets, the hierarchical clustering techniques mentioned above can not be further used since the distance matrix increases quadratically with the number of observation points. Fortunately, there are partitioning methods that are not based on the matrix of distances but directly on the data values. Since these algorithms are, in general, sensitive with respect to local minima, it is essential to estimate good starting values for the procedure. In *k*-means clustering, which is an iterative partitioning method, the number of starting values (i.e., initial guesses for cluster centers) equals the number of final clusters which is, in general, unknown.

The proposed procedure combines the ideas for hierarchical and *k*-means clustering. In the clustering part the data are split divisively. At the beginning all points are combined in one cluster, and in each step of the hierarchy a cluster is decomposed into two smaller clusters. The basis for the decomposition is not the distance matrix, as usual in divisive clustering, but the division is performed by *k*-means clustering (with two clusters in each step). Another difference to usual divisive clustering is that the procedure for dividing a cluster is stopped at a certain point. There are different parameters available for the decision whether a clustered group should be further divided. In addition, appropriate statistical tests are applied to prove significance levels, and the visualization of the data also is very helpful for this task. Once this divisive clustering procedure stops, clusters that have been divided unjustified can be re-combined. In this so-called merging part (statistical) parameters are supplied for the merging decision. As a result we obtain the final number of clusters, and, since all clusters have been determined by *k*-means clustering, the centers of these clusters. These cluster centers are assumed to be the optimal starting values for the final part of the proposed procedure, the (fast) *k*-means clustering.

We have shown that this procedure works well for analyzing both, synthetic and in vivo fMR images with sufficient functional CNR, i.e., ≥ 2. The algorithm requires no prior knowledge on the stimulation paradigm, the hemodynamic response function of the brain or the data quality. The computation time for a clustering step is about 10 ms to 1 s (on the previously mentioned computer), depending on the number of points to be clustered, and the final result is obtained in about 1 s. The parameters for the decision of dividing or merging clusters can be adjusted to the individual data quality and information content, if known. The data visualization enables a careful inspection of each step of the procedure. Therefore, it is also possible to improve the procedure by incorporating prior knowledge, which may be important with pilot studies or limited data quality (e.g., patient data). The method presented not only has the potential to separate time varying from time invariant structures but also, within the same analysis, allows the segmentation of various time invariant, i.e., anatomic, structures. This may help to better correlate anatomy and function in the human brain.

Since the computation time of the procedure is very short, it is especially suited for fMRI studies with a large number of sequential images, e.g., for EPI or spiral fMRI. The proposed method, in principal, may be improved by taking a clustering method that is more sensitive with respect to small clusters. This, however, could increase the computation time. For the analysis of large data sets of known data quality, the somewhat cumbersome procedure described above may be automated and made faster by using the same initial guesses for the clustering procedure estimated in a single pilot analysis. This reduces the computation time per data set to the time needed for the final clustering (i.e., about 1 s).

For in vivo fMRI data, usually providing less anatomic contrast, fuzzy clustering may have advantages over crisp clustering methods. This modified approach is currently tested and will be presented elsewhere.

Acknowledgements—R. B. was financially supported by the Austrian Science Foundation (Grant No. FWF P11438Med). E. M. acknowledges financial support from the L. Boltzmanninstitut f. radiol.–phys. Tumordiagnostik. In vivo fMRI data were acquired by C. Windischberger (Vienna).

REFERENCES

1. Ogawa, S.; Menon, R.S.; Tank, D.W.; Kim, S.G.; Merkle, H.; Ellermann, J.M.; Ugurbil, K. Functional mapping of blood oxygenation level-dependent contrast magnetic resonance imaging. Biophys. J. 64:803-812; 1993.
2. Kwong, K.K. Functional magnetic resonance imaging with echo-planar imaging. Magn. Res. Q. 11:1-20; 1995.
3. Turner, R. Functional mapping of the human brain with magnetic resonance imaging. Sem. Neurosci. 7:179-194; 1995.
4. Gomiscek, G.; Beisteiner, R.; Hittmair, K.; Müller, E.; Moser, E. A possible role of in-flow effects in functional MR-imaging. MAG*MA. 1:109-113; 1993.
5. Haacke, E.M.; Hopkins, A.; Lai, S.; Buckley, P.; Friedman, L.; Meltzer, H.; Hedera, P.; Friedland, R.; Klein, S.; Thompson, L. 2D and 3D high resolution gradient echo functional imaging of the brain: Venous contributions to signal in motor cortex studies. NMR Biomed. 7:54-62; 1994.
6. Hajnal, J.V.; Myers, R.; Oatridge A.; Schwieso, J.E.; Young, I.R.; Bydder, G.M. Artifacts due to stimulus correlated motion in functional imaging of the brain. Magn. Res. Med. 31:289-291; 1994.
7. Biswal, B.; DeYoe, E.; Hyde, J.S. Reduction of physiological fluctuations in fMRI using digital filters. Magn. Res. Med. 35:107-113; 1995.
8. Weisskoff, R.M.; O'Craven, K.M.; Savoy, R.; Brown, E.; Solo, V. Temporal correlation in fMRI of the visual cortex: Impact on imaging rate. NeuroImage (suppl 3):107; 1996.
9. Moser E.; Windischberger, C. Separation of physiological motion artifacts in single-shot EPI to improve reproducibility for functional MRI studies. Proceedings of the ISMRM, Sydney, Aus.; 1998: 133. vol. 1
10. Bandettini P.A.; Jesmanowicz A.; Wong E.C.; Hyde J.S. Processing strategies for time course data sets in functional MRI of the human brain. Magn. Res. Med. 30:161-173; 1993.
11. Frahm, J.; Merboldt, K.D.; Hänicke, W.; Kleinschmidt, A.; Boecker, H. Brain or vein. Oxygenation of flow? On signal physiology in functional MRI of human brain activation. NMR Biomed. 7:45-52; 1994.
12. Moser, E.; Teichtmeister, C.; Diemling, M. Reproducibility and post-processing of gradient-echo functional MRI to improve localization of brain activity in the human visual cortex. Magn. Res. Imaging 14:567-579; 1996.
13. Baumgartner, R.; Scarth, G.; Teichtmeister, C.; Somorjai, R.; Moser, E. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: Reproducibility. J. Magn. Res. Imaging 7:1094-1101; 1997.
14. Moser, E.; Diemling, M.; Baumgartner, R. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: Quantification. J. Magn. Res. Imaging 7:1102-1108; 1997.
15. Ding, X.; Masaryk, T.; Ruggieri, P.; Tkach, J. Detection of activation patterns in dynamic functional MRI with a clustering technique. Proceedings of the ISMRM, vol. 3, New York: 1996: 1798.
16. Friston, K.J.; Jezzard, P.; Turner, R. Analysis of fMRI time series. Human Brain Mapping 2:69-78; 1995.
17. Pedersen, F.; Bengtsson, E.; Hindmarsh, T.; Nordell, B.; Forssberg, H. Using principal component analysis to visualize the spatial distribution of functional areas of the brain as studied with MRI during motor and sensory activation. In: E.A. Hoffman, R.S. Acharya (Eds). Physiology and Function from Multidimensional Images (SPIE Proceedings 2168). Bellingham, WA: SPIE, 1994: pp. 422-437.
18. Sychra, J.J.; Bandettini, P.A.; Bhattacharya, N.; Lin, Q. Synthetic images by subspace transforms I. Principal component images and related filters. Med. Phys. 21:193-201; 1994.
19. Backfrieder, W.; Baumgartner, R.; Teichtmeister, C.; Bergmann, H.; Moser, E. Feature extraction from fMRI using orthogonal data decomposition. In: J. Richter (Ed). Medizinische Physik 95. Würzburg: DGMP, 1995: pp. 336-337.
20. Backfrieder, W.; Baumgartner, R.; Samal, M.; Moser, E.; Bergmann, H. Quantification of intensity variations in functional MR images using rotated principal components. Phys. Med. Biol. 41:1425-1438; 1996.
21. Scarth, G.; McIntyre, M.; Wowk, B.; Somorjai, R. Detection of novelty in functional images using fuzzy clustering. Proceedings of the SMR and ESMRMB, vol. 1, Nice (F); 1995: 238.
22. Baumgartner, R.; Windischberger, C.; Moser, E. Quantification in functional MRI: Fuzzy clustering versus correlation analysis. Magn. Res. Imag. 16:115-125; 1998.
23. Kaufman, L.; Rousseeuw, P.J. Finding Groups in Data. New York: Wiley & Sons; 1990.
24. Hartigan, J.A. Clustering Algorithms. New York: Wiley; 1975.
25. Hartigan, J.A.; Wong, M.A. A *k*-means clustering algorithm. Applied Statistics 28:100-108; 1979.
26. Jackson, J.E. A User's Guide to Principal Components. New York: Wiley & Sons; 1991.
27. Mardia, K.V.; Kent, J.T.; Bibby, J.M. Multivariate Analysis. London: Academic Press; 1979.
28. Afifi, A.A.; Azen, S.P. Statistical analysis: A Computer Oriented Approach. New York: Academic Press; 1979.
29. Venables, W.N.; Ripley, B.D. Modern Applied Statistics with S-Plus. New York: Springer; 1994.
30. Baumgartner, R.; Backfrieder, W.; Moser, E. Quantification of statistical type I and II errors in correlation analysis of simulated functional magnetic resonance imaging data. MAG*MA. 4:251-256; 1996.
31. Moser, E.; Baumgartner, R.; Barth, M.; Windischberger, C. Explorative signal processing in functional MR imaging. Int. J. Imag. Sys. Technol. 10:166–176, 1999.
32. Barth, M.; Reichenbach, J.R.; Venkatesan, R.; Moser, E.; Haacke, E.M. High-resolution, multiple gradient-echo functional MRI at 1.5 T. Magn. Res. Imaging 17:321-329; 1999.