Exploratory Data Analysis Methods Applied to fMRI

Oliver Lange, Anke Meyer–Bäse, Uwe Meyer–Bäse,

Department of Electrical and Computer Engineering,
Florida State University, Tallahassee, Florida 32310-6046 USA

Axel Wismüller, and

Department of Clinical Radiology, Ludwig–Maximilians University,
Munich 80336, Germany

Monica Hurdal

Department of Mathematics,
Florida State University, Tallahassee, Florida 32306-4510 USA

ABSTRACT

Exploratory data–driven methods such as unsupervised clustering and independent component analysis (ICA) are considered to be hypothesis–generating procedures, and are complementary to the hypothesis–led statistical inferential methods in functional magnetic resonance imaging (fMRI). In this paper, we present a comparison between unsupervised clustering and ICA in a systematic fMRI study. The comparative results were evaluated by a very detailed ROC analysis. For the fMRI data, a comparative quantitative evaluation between the three clustering techniques, SOM, "neural gas" network, and fuzzy clustering based on deterministic annealing, and the three ICA methods, FastICA, Infomax and topographic ICA was performed. The ICA methods proved to extract features relatively well for a small number of independent components but are limited to the linear mixture assumption. The unsupervised clustering outperforms ICA in terms of classification results but requires a longer processing time than the ICA methods.

Keywords: Minimal free energy vector quantization, "neural gas" network, self–organizing map, Infomax, FastICA, topographic ICA, PCA, fMRI

1. INTRODUCTION

Functional magnetic resonance imaging with high temporal and spatial resolution represents a powerful technique for visualizing rapid and fine activation patterns of the human brain.[1] As is known from both theoretical estimations and experimental results,[2] an activated signal variation appears very low on a clinical scanner. This motivates the application of analysis methods to determine the response waveforms and associated activated regions. Generally, these techniques can be divided into two groups: Model-based techniques require prior knowledge about activation patterns, whereas model-free techniques do not. However, model-based analysis methods impose some limitations on data analysis under complicated experimental conditions. Therefore, analysis methods that do not rely on any assumed model of functional response are considered more powerful and relevant. We distinguish two groups of model-free methods: transformation-based and clustering-based.

The first method, principal component analysis (PCA)[3, 4] or independent component analysis (ICA),[5–8] transforms original data into high-dimensional vector space to separate functional response and various noise sources from each other.

Among the data–driven techniques, ICA has been shown to provide a powerful method for the exploratory analysis of fMRI data.[6, 8] ICA is an information theoretic approach which enables recovery of underlying signals, or independent components (ICs) from linear data mixtures. Therefore, it is an excellent method to be applied for the spatial localization and temporal characterization of sources of BOLD activation. ICA can be

Intelligent Computing: Theory and Applications III, edited by Kevin L. Priddy,
Proceedings of SPIE Vol. 5803 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.601004

applied to fMRI both temporal[9] or spatial.[6] Spatial ICA has dominated so far in fMRI applications because the spatial dimension is much larger than the temporal dimension in fMRI. However, recent literature results have suggested that temporal and spatial ICA yield similar results for experiments where two predictable task–related components are present.

The second method, fuzzy clustering analysis[10–13] or self-organizing map,[13–15] attempts to classify time signals of the brain into several patterns according to temporal similarity among these signals.

In this paper, we perform a detailed comparative study among unsupervised clustering methods ("neural gas" network,[16] fuzzy clustering based on deterministic annealing,[13] and Kohonen's self–organizing map (SOM)) and spatial ICA techniques (FastICA,[17] topographic ICA,[18] Infomax,[19] PCA) for fMRI. In a systematic manner, we will compare and evaluate the results obtained based on each technique and present the benefits associated with each paradigm.

2. EXPLORATORY DATA ANALYSIS METHODS

Functional organization of the brain is based on two complementary principles, localization and connectionism. Localization means that each visual function is performed mainly by a small set of the cortex. Connectionism, on the other hand, expresses that the brain regions involved in a certain visual cortex function are widely distributed, and thus the brain activity necessary to perform a given task may be the functional integration of activity in distinct brain systems. It is important to stress that in neurobiology the term "connectionism" is used in a different sense that that used in the neural network terminology.

The following sections are dedicated to presenting the algorithms and evaluate the discriminatory power of the two main groups of exploratory data analysis methods.

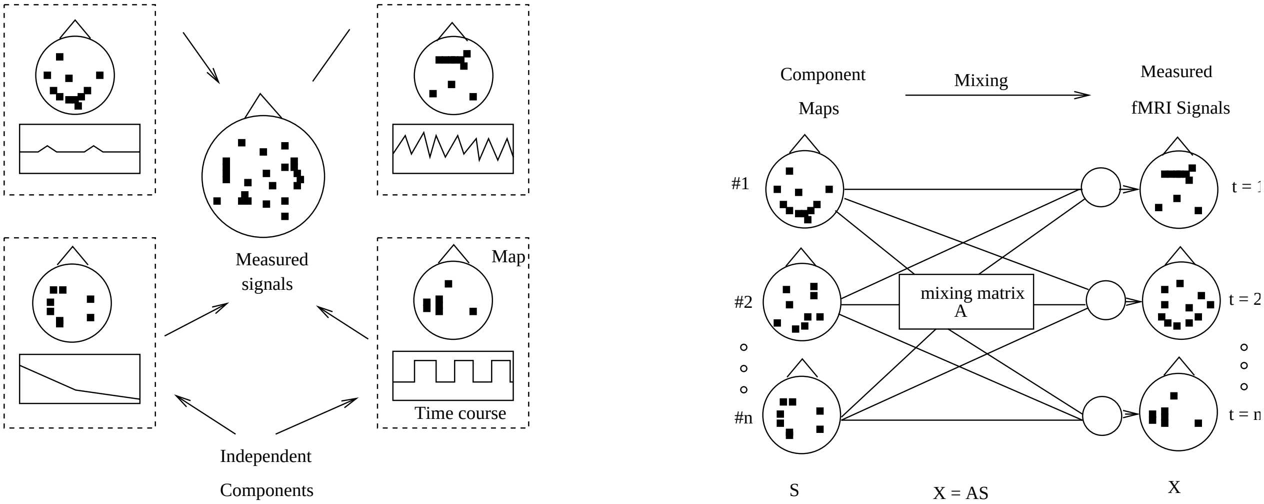
2.1. The ICA Algorithms

According to the principle of functional organization of the brain, it was suggested for the first time in[6] that the multifocal brain areas activated by performance of a visual task should be unrelated to the brain areas whose signals are affected by artifacts of physiological nature, head movements, or scanner noise related to fMRI experiments. Every single above mentioned process can be described by one or more spatially–independent components, each associated with a single time course of a voxel and a component map. It is assumed that the component maps, each described by a spatial distribution of fixed values, represent overlapping, multifocal brain area of statistically dependent fMRI signals. This aspect is visualized in Figure 1. In addition, it is considered that the distributions of the component maps are spatially independent, and in this sense uniquely specified. Mathematically, this means that if $p_k(C_k)$ specifies the probability distribution of the voxel values C_k in the kth component map, then the joint probability distribution of all n components yields:

$$p(C_1, \cdots, C_n) = \prod_{k=1}^{n} p_k(C_k) \tag{1}$$

where each of the component maps C_k is a vector $(C_{ki}, i = 1, 2, \cdots, M)$, where M gives the number of voxels. Independency is a stronger condition than uncorrelatedness. It was shown in[6] that these maps are independent if the active voxels in the maps are sparse and mostly nonoverlapping. Additionally it is assumed that the observed fMRI signals are the superposition of the individual component processes at each voxel. Based on these assumptions, ICA can be applied to fMRI time–series to spatially localize and temporally characterize the sources of BOLD activation.

Different methods for performing ICA decompositions have been proposed which employ different objective functions together with different criteria of optimization of these functions, and it is assumed that they can produce different results.



(a) (b)

Figure 1. Visualization of ICA applied to fMRI data. (a) Scheme of fMRI data decomposed into independent components, and (b) fMRI data as a mixture of independent components where the mixing matrix $\mathbf{M}$ specifies the relative contribution of each component at each time point.[6]

2.2. Models of Spatial ICA in fMRI

In the following we will assume that $\mathbf{X}$ is a $T \times M$ observed fMRI signal data matrix, $\mathbf{C}$ is the $N \times M$ random matrix of component map values, and A is a $T \times N$ mixing matrix containing in its columns the associated time–courses of the N components. Furthermore, T corresponds to the number of scans, and M is the number of voxels included in the analysis. Matrix $\mathbf{X}$ is the so–called matrix of observed voxel time courses (VTC).

The spatial ICA (sICA) problem is given by the following linear combination model for the data:

$$\mathbf{X} = \mathbf{AC} \tag{2}$$

where no assumptions are made about the mixing matrix $\mathbf{A}$ and the rows $\mathbf{C_i}$ being mutually statistically independent.

Then the ICA decomposition of $\mathbf{X}$ can be defined as an invertible transformation:

$$\mathbf{C} = \mathbf{WX} \tag{3}$$

where $\mathbf{W}$ is an unmixing matrix providing a linear decomposition of data. $\mathbf{A}$ is the pseudoinverse of $\mathbf{W}$.

The algorithmic description of the Infomax and FastICA can be found in.[20]

2.3. The Topographic ICA Approach

Topographic ICA represents a generative model which combines topographic mapping with ICA. As in all topographic mappings, the distance in the representation space given by the topographic grid is related to the distance of the represented components. This distance is defined for topographic ICA by the mutual information implied by higher–order correlations.[18] Thus, a natural distance measure is given in the context of ICA. Traditional topographic mapping methods define distance either based on the Euclidean distance or correlation. The ICA distance measure enables the definition of a topography even if the Euclidean distances are all equal as it is the case with an orthogonal vector space.

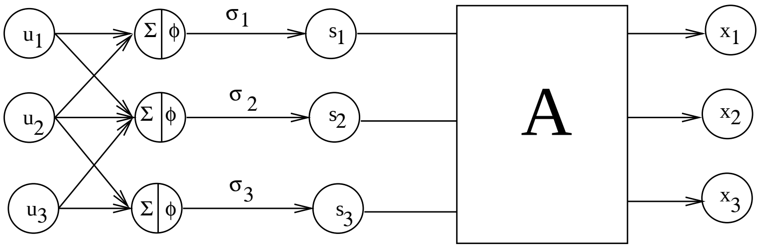


Figure 2. Topographic ICA model.[18] The variance generated variables u_i are randomly generated, and mixed linearly inside their topographic neighborhoods. This forms the input to nonlinearity ϕ, thus giving the local variance σ_i. Components s_i are generated with variances σ_i. The observed variables are x_i are obtained as with standard ICA from the linear mixture of the components s_i.

In the generative model described in[18] and shown in Figure 2, s_i represent the unknown sources and are independent given their variances σ_i^2. Dependence among the s_i is enforced by the variance dependence. Obeying the principle of topography, the variances of only neighboring components are positively correlated, while the others are independent. By using a neighborhood function $h(i,j)$, the variance σ_i is given by:

$$\sigma_i = \phi\left(\sum_{k=1}^{n} h(i,k)u_k\right) \tag{4}$$

u_i are the higher order independent components used to generate the variances, while ϕ describes some nonlinearity. The neighborhood function $h(i,j)$ can be either a two–dimensional grid or have a ring–like structure. The components s_i are given by the following relationship:

$$s_i = z_i\sigma_i \tag{5}$$

where z_i is a random variable having the same distribution as s_i while σ_i^2 is fixed to unity. u_i and z_i are mutually independent.

The most important properties of the topographic ICA are: (1) all the components are uncorrelated, (2) components far from each other are independent, (3) neighboring components tend to be active (non–zero) at the same time, and thus have positively correlated energies s_i^2 and s_j^2. The classic ICA results from the topographic ICA by setting $h(i,j) = \delta_{ij}$.

The learning rule is based on the maximization of the likelihood. First, it is assumed that the data is preprocessed by whitening and that the estimates of the components are uncorrelated.

The update rule for the weight vector $\mathbf{w}_i$ is derived from a gradient algorithm:

$$\Delta\mathbf{w}_i \propto E\{\mathbf{x}(\mathbf{w}_i^T\mathbf{x})r_i)\} \tag{6}$$

where

$$r_i = \sum_{k=1}^{n} h(i,k)g(\sum_{j=1}^{n} k(k,j)(\mathbf{w}_j^T\mathbf{x})^2) \tag{7}$$

The function g is the derivative of $G = -\alpha_1\sqrt{u} + \beta_1$. After every iteration, the vectors $\mathbf{w}_i$ in equation (6) are normalized to unit variance and orthogonalized. This equation represents a modulated learning rule, where the learning term is modulated by the term r_i.

It's useful to point out some differences between topographic ICA and other topographic mappings: (1) topographic ICA finds a decomposition into independent components, while topographic mappings find cluster centers or codevectors, and (2) the similarity of two vectors in topographic ICA is based on higher–order correlations and not defined by an Euclidean distance or dot–product. However, if the data is prewhitened, the dot–product in the data space is equivalent to correlation in the original space.[18]

Topographic ICA represents a new paradigm for fMRI signal analysis since the strict independence condition imposed by standard ICA techniques is relaxed among neighboring components, such that neighboring components (voxels) are positively correlated.

2.4. The Clustering Algorithms

The previous sections showed that ICA techniques can be applied to fMRI by considering brain function as consisting of sets of non–systematically overlapping networks. In other words, ICA works by assuming that during a given fMRI experiment there are a number of brain regions (networks) that are spatially independent from one another (sources) and are mixed together via a network specific hemodynamic time course.

In this section, we will review cluster analysis as an alternative technique which is based on grouping image voxels together based on the similarity of their intensity profile in time (i.e., their time courses).

Let T denote the number of subsequent scans in a fMRI study, and let M be the number of voxels. The dynamics of each voxel $\mu \in \{1, \cdots, M\}$, i.e. the sequence of signal values $\{\mathbf{x}^{\mu}(1), \cdots, \mathbf{x}^{\mu}(T)\}$ can be interpreted as a vector $\mathbf{x}^{\mu}(i) \in \mathbf{R}^{T}$ in the T–dimensional feature space of possible signal time–series at each voxel.

Cluster analysis groups image voxels together based on the similarity of their intensity profile in time. In the clustering process, a time course with T points is represented by one point in an T–dimensional Euclidean space which is subsequently partitioned into clusters based on the proximity of the input data.

Here, we employ several vector quantization (VQ) approaches as a method for unsupervised image time–series analysis. VQ clustering identifies several groups of voxels with similar VTC, while these groups or clusters are represented by prototypical time–series called codebook vectors (CV) located at the center of the corresponding clusters. The CVs represent prototypical VTCs sharing similar temporal characteristics. Thus, each VTC can be assigned in the crisp clustering scheme to a specific CV according to a minimal distance criterion, while in the fuzzy scheme according to a membership to several CVs. Accordingly, the outcomes of VQ approaches for fMRI data analysis can be plotted as "crips" or "fuzzy " cluster assignment maps.

VQ approaches determine the cluster centers $\mathbf{w}_i$ by an iterative adaptive update based on the following equation:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \epsilon(t) a_i(\mathbf{x}(t), C(t), \kappa)(\mathbf{x}(t) - \mathbf{w}_i(t)) \tag{8}$$

where $\epsilon(t)$ represents the learning parameter, a_i a codebook $C(t)$ dependent cooperativity function, κ a cooperativity parameter, and $\mathbf{x}$ a randomly chosen feature vector. For fMRI, the feature vector represents the VTC.

The algorithmic description of the "neural gas" network, fuzzy clustering based on deterministic annealing and Kohonen's self–organizing map can be found in[21] and.[20]

3. RESULTS AND DISCUSSION

FMRI data were recorded from six subjects (3 female, 3 male, age 20-37) performing a visual task. In five subjects, five slices with 100 images (TR/TE=3000/60msec) were acquired with five periods of rest and five photic simulation periods with rest. Simulation and rest periods comprised 10 repetitions each, i.e. 30s. Resolution was $3 \times 3 \times 4$ mm. The slices were oriented parallel to the calcarine fissure. Photic stimulation was performed using an 8 Hz alternating checkerboard stimulus with a central fixation point and a dark background with a central fixation point during the control periods.[13] The first scans were discarded for remaining saturation effects. Motion artifacts were compensated by automatic image alignment (AIR,[22]).

The clustering results were evaluated by ROC analysis.

In the following we will give the set of parameters chosen for the comparative evaluation for both exploratory data analysis techniques. For PCA, no parameters had to be set. For Infomax we choose: (1) the learning rate $\eta = 10^{-6}$, and (2) 10^5 as the maximal number of iterations. For FastICA we choose: (1) 10^5 as the maximal number of iterations, and (2) the nonlinearity $g(u) = \tanh u$. And last, for topographic ICA we set: (1) stop criterion is fulfilled if the synaptic weights difference between two consecutive iterations is less than $10^{-5} \times$ number of IC, (2) the function $g(u) = u$, and (3) 10^4 is the maximal number of iterations.

For SOM we employed the SOMPAK (http://www.cis.hut.fi/research/sompak) and we choose: a rectangular grid, the neighborhood function is a bubble (step function), and for the initial ordering training phase we set $\epsilon_i = 0.05$, $\sigma = 3.0$, and the maximum iteration number equals the number of data points. For the fine–tuning training phase we set $\epsilon_i = 0.01$, $\sigma = 1.5$, and the maximum iteration number equals $1.5 \times$ the number of data points. For "neural gas" network we choose: (1) the learning parameters $\epsilon_i = 0.5$ and $\epsilon_f = 0.005$, and (2) the lattice parameters λ_i equals half the number of classes and $\lambda_f = 0.01$ and (3) the maximal number of iterations equals 1000. And last, for fuzzy clustering based on deterministic annealing we set: (1) neurons' initialization with principal components, (2) learning parameter $\rho_{final} = 0.01$ and updating based on a linear annealing scheme, and (3) the maximal number of iterations equals 100.

It is important to determine the differences and implications on the analysis of fMRI time–series. The two main groups which are subject to our comparative study are (1) the transformation–based methods such as PCA and ICA, and (2) the clustering methods such as SOM, "neural gas" network, and the fuzzy clustering based on deterministic annealing.

The main difference between these two model–free fMRI analysis techniques lies in their categorization properties. The transformation–based techniques allow us to determine the contribution of the consistently task–related (CTR) component and the artifacts components (head movement) to a single voxel. While based on ICA it is possible to separate the artifacts from the CTR, clustering techniques try to identify clusters of similar voxel time courses within the data space. In other words, they try to determine clusters describing similar voxel time courses based on a minimal distance criterion, such that the resulting codebook vectors are then prototypes of voxel time courses of similar temporal characteristics.

3.1. ROC Analysis

It is important to perform a quantitative analysis of the relative performance of the introduced exploratory data analysis techniques. To do so, we compared the proposed algorithms for 8 and 36 components in terms of ROC analysis using correlation map with a chosen threshold of 0.4. We report the ROC performances for the five subjects in Figure 3. The Figure illustrates the average area under the curve and its deviations for 20 different ROC runs using the same parameters but different algorithms' initializations. From this Figure, we can see that the clustering methods outperform for 36 components the ICA methods including PCA for all five subjects. For 8 components, we see that for most subjects SOM is outperformed by topographical ICA, while the other two clustering techniques achieve the best results.

4. CONCLUSION

In the present paper, we have experimentally compared two exploratory data analysis methods for fMRI: the ICA techniques versus unsupervised clustering. The ICA techniques were two standard ICA algorithms, the Infomax and the FastICA, and a new algorithm, the topographic ICA. The unsupervised clustering techniques were two proven clustering algorithms, the SOM and the fuzzy clustering based on deterministic annealing, and a less known algorithm, the "neural gas" network.

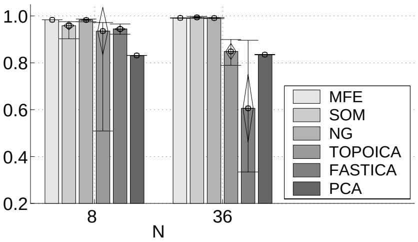
The goal of the paper was to determine the robustness and reliability of extracting task–related activation maps and time–courses from fMRI data sets. The success of ICA methods is based on the condition that the spatial distribution of brain areas activated by task performance must be spatially independent of the distributions of areas affected by artifacts. It was also shown that unsupervised clustering techniques represent a successful strategy for the analysis of time–courses from fMRI data sets. The increasing cluster resolution proved to reveal extremely well the structure of the data set. From the ROC analysis we observe, that for 36 components the clustering methods outperform the transformation–based methods for all five subjects. For 8 components, we see that for most subjects SOM is outperformed by topographical ICA, while the "neural gas" network and fuzzy clustering based on deterministic annealing achieve the best results.

Subject 1

Subject 2

Subject 3

Subject 4



Subject 5

Figure 3. Results of the comparison between the two different exploratory data analysis methods on fMRI data. Spatial accuracy of the different maps is assessed by ROC analysis using correlation map with a chosen threshold of 0.4. It is illustrated the average area under the curve and its deviations for 20 different ROC runs using the same parameters but different algorithms' initializations. The number of chosen independent components or codebook vectors for all techniques is equal to 8 and 36 and results are plotted for all five subjects.

Both the "neural gas" network and fuzzy clustering based on deterministic annealing outperform ICA in terms of classification results and sensitivity range of the CTR–component but require a longer processing time than the ICA methods. Another important aspect is that topographical ICA represents a unifying paradigm between transformation–based and clustering techniques and bridges thus the discriminatory capabilities of FastICA and "neural gas" network. The relaxation of the independence condition for neighboring components leads to an increase in sensitivity range compared to standard ICA, and achieves in most cases a higher correlation coefficient compared to the other ICA techniques.

The applicability of the new algorithm is demonstrated on experimental data.

REFERENCES

1. S. Ogawa, D. Tank, and R. Menon, “Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging,” *Proceedings of the National Academy of Science* **89**, pp. 5951–5955, 8 1992.
2. S. Ogawa, T. Lee, and B. Barrere, “The sensitivity of magnetic resonance image signals of a rat brain to changes in the cerebral venous blood oxygenation activation,” *Magn. Reson. Med.* **29**, pp. 205–210, 8 1993.
3. J. Sychra, P. Bandettini, N. Bhattacharya, and Q. Lin, “Synthetic images by subspace transforms i. principal components images and related filters,” *Med. Phys.* **21**, pp. 193–201, 8 1994.
4. W. Backfrieder, R. Baumgartner, M. Samal, E. Moser, and H. Bergmann, “Quantification of intensity variations in functional mr images using rotated principal components,” *Phys. Med. Biol.* **41**, pp. 1425–1438, 8 1996.
5. M. McKeown, T. Jung, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, “Spatially independent activity patterns in functional magnetic resonance imaging data during the stroop color-naming task,” *Proc. Natl. Acad. Sci.* **95**, pp. 803–810, 8 1998.
6. M. McKeown, T. Jung, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, “Analysis of fmri data by blind separation into independent spatial components,” *Human Brain Mapping* **6**, pp. 160–188, 8 1998.
7. F. Esposito, E. Formisano, E. Seifritz, R. Goebel, R. Morrone, G. Tedeschi, and F. D. Salle, “Spatial independent component analysis of functional mri time–series: to what extent do results depend on the algorithm used?,” *Human Brain Mapping* **16**, pp. 146–157, 8 2002.
8. K. Arfanakis, D. Cordes, V. Haughton, C. Moritz, M. Quigley, and M. Meyerand, “Combining independent component analysis and correlation analysis to probe interregional connectivity in fmri task activation datasets,” *Magnetic Resonance Imaging* **18**, pp. 921–930, 8 2000.
9. B. Biswal and J. Ulmer, “Blind source separation of multiple signal sources of fmri data sets using independent component analysis,” *Journal of Computer Assisted Tomography* **23**, pp. 265–271, 8 1999.
10. G. Scarth, M. McIntrye, B. Wowk, and R. Samorjai, “Detection novelty in functional imaging using fuzzy clustering,” *Proc. SMR 3rd Annu. Meeting* **95**, pp. 238–242, 8 1995.
11. K. Chuang, M. Chiu, C. Lin, and J. Chen, “Model-free functional mri analysis using kohonen clustering neural network and fuzzy c-means,” *IEEE Transaction on Medical Imaging* **18**, pp. 1117–1128, 8 1999.
12. R. Baumgartner, L. Ryder, W. Richter, R. Summers, M. Jarmasz, and R. Somorjai, “Comparison of two exploratory data analysis methods for fmri: fuzzy clustering versus principal component analysis,” *Magnetic Resonance Imaging* **18**, pp. 89–94, 8 2000.
13. A. Wism”uller, O. Lange, D. Dersch, G. Leinsinger, K. Hahn, B. P”utz, and D. Auer, “Cluster analysis of biomedical image time–series,” *International Journal on Computer Vision* **46**, pp. 102–128, 2 2002.
14. H. Fisher and J. Hennig, “Clustering of functional mr data,” *Proc. ISMRM 4rd Annu. Meeting* **96**, pp. 1179–1183, 8 1996.
15. S. Ngan and X. Hu, “Analysis of fmri imaging data using self-organizing mapping with spatial connectivity,” *Magn. Reson. Med.* **41**, pp. 939–946, 8 1999.
16. T. Martinetz, S. Berkovich, and K. Schulten, “Neural gas network for vector quantization and its application to time-series prediction,” *IEEE Transactions on Neural Networks* **4**, pp. 558–569, 4 1993.
17. A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks* **10**, pp. 626–634, 5 1999.
18. A. Hyvarinen and P. Hoyer, “Topographic independent component analysis,” *Neural Computation* **13**, pp. 1527–1558, 7 2001.
19. A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation* **7**, pp. 1129–1159, 12 1995.
20. A. Meyer-B”ase, *Pattern Recognition for Medical Imaging*, Elsevier Science/Academic Press, 2003.
21. A. Wism”uller, A. Meyer-B”ase, O. Lange, and D. Auer, “Comparison of two exploratory data analysis methods for fmri: Unsupervised clustering vs. independent component analysis,” *Invited Paper, Proceedings of SPIE, Vol. 5103* , pp. 104–115, 4 2004.
22. R. Woods, S. Cherry, and J. Mazziotta, “Rapid automated algorithm for aligning and reslicing pet images,” *Journal of Computer Assisted Tomography* **16**, pp. 620–633, 8 1992.