Tree–Dependent and Topographic Independent Component Analysis for fMRI Analysis

Oliver Lange[1,2], Anke Meyer–Bäse[1], Axel Wismüller[1,2],

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

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

Monica Hurdal, DeWitt Sumners, and

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

Dorothee Auer

Max Planck Institute of Psychiatry, Munich 80804, Germany

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). Recently, a new paradigm in ICA emerged, that of finding "clusters" of dependent components. This striking philosophy found its implementation in two new ICA algorithms: tree–dependent and topographic ICA. For fMRI, this represents the unifying paradigm of combining two powerful exploratory data analysis methods, ICA and unsupervised clustering techniques. For the fMRI data, a comparative quantitative evaluation between the two methods, tree–dependent and topographic ICA was performed. The comparative results were evaluated by (1) task–related activation maps, (2) associated time–courses and (3) ROC study. It can be seen that topographic ICA outperforms all other ICA methods including tree–dependent ICA for 8 and 9 ICs. However, for 16 ICs topographic ICA is outperformed by both FastICA and tree–dependent ICA (KGV) using as an approximation of the mutual information the kernel generalized variance.

Keywords: Tree–dependent ICA, topographic ICA, 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.[1–5] As is known from both theoretical estimations and experimental results,[4,6,7] 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. There are two kinds of model-free methods. The first method, principal component analysis (PCA)[8,9] or independent component analysis (ICA),[10–13] transforms original data into high-dimensional vector space to separate functional response and various noise sources from each other.

Independent Component Analyses, Wavelets, Unsupervised Smart Sensors, and Neural Networks II, edited by Harold H. Szu, et al., Proc. of SPIE Vol. 5439 (SPIE, Bellingham, WA, 2004) 0277-786X/04/$15 · doi: 10.1117/12.541779

Among the data–driven techniques, ICA has been shown to provide a powerful method for the exploratory analysis of fMRI data.[11,13] ICA is an information theoretic approach which enables to recover 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 applied to fMRI both temporal[14] or spatial.[11] 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[15–18] or self-organizing map,[18–20] 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 for fMRI among the tree–dependent and topographic ICA with standard ICA techniques. 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 out 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[11] 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, is representing 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:

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

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[11] 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.

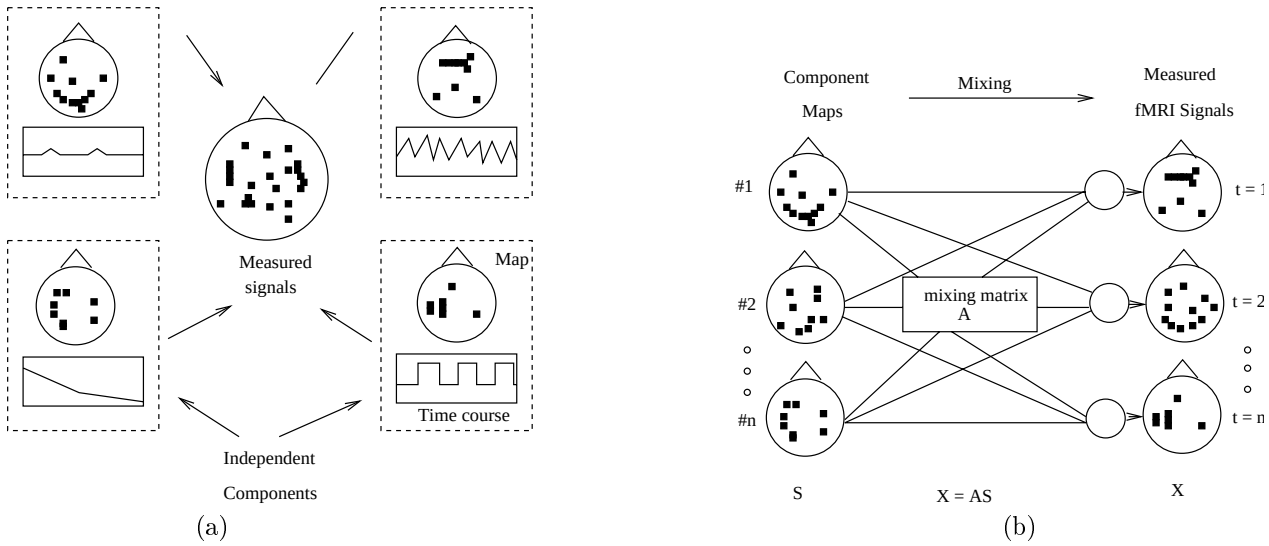

Figure 1. Visualization of ICA applied to fMRI data. (a) Scheme of fMRI data decomposed into independent components, and (b) fMRI data as a mixture of independent components where the mixing matrix **M** specifies the relative contribution of each component at each time point.[11]

2.2. Models of Spatial ICA in fMRI

In the following we will assume that $\mathbf{X}$ is a $T \times M$ matrix of observed voxel time courses (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.

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 employed ICA algorithms are the TDSEP, JADE, the FastICA approach based on minimization of mutual information but using the negentropy as a measure of non–gaussianity,[21] and topographic ICA which combines topographic mapping with ICA.[22]

2.3. Tree–Dependent Component Analysis

The tree–dependent component analysis (TCA) concept is based on the idea of weakening the assumption of independence in ICA. Here, we search for a transformation such that the independent components can be modeled by a tree–structured graphical model.[23] The most important feature is that the trees can have more than a single connected component. The connected components of the graphical model can be viewed as "clusters" of dependent components, and thus the decomposition of the source variables yields to dependent components within a cluster and independent outside a cluster. This approach is different from the topographic ICA described in the next section since it does not require to specify the topology (number and sizes of components) in advance.

2.3.1. Beyond Independent Components: Trees and Clusters

The paradigm of TCA is derived from the theory of tree–structured graphical models. In[24] was shown a strategy to approximate optimally an n–dimensional discrete probability distribution by a product of second–order distributions, or the distribution of the first–order tree dependence. A tree is an undirected graph with at most a single edge between two nodes. This tree concept can be easily interpreted with respect to ICA. A graph with no edges means that the random variables are mutually independent and this pertains to ICA. On the other hand, if no assumptions are made about independence, then the corresponding family of probability distributions represents the set of all distributions.

In[24] was developed a strategy of the best approximation of an nth–order distribution by a product of $n-1$ second–order component distributions:

$$P_i(\mathbf{x}) = \prod_{i=1}^{n} P(x_{m_i}|x_{m_{j(i)}}), \quad 0 \leq j(i) < i \tag{4}$$

where $P(\mathbf{x})$ is a joint probability distribution of n discrete variables with $\mathbf{x} = x_1, \cdots, x_n$ being a vector, $(m_1, \cdots, m_n)$ is an unknown permutation of integers 1,2, ..., n and $P(x_i|x_0)$ is by definition equal to $P(x_i)$. The above introduced probability distribution is named a probability distribution of first–order tree dependence.

A probability distribution can be approximated in several ways. Here, we look into approximations based on a product of $n-1$ second–order component distributions. To determine the goodness of an approximation, it is necessary to define a closeness as

$$I(P, P_a) = \sum_{\mathbf{x}} P(\mathbf{x}) \log \frac{P(\mathbf{x})}{P_a(\mathbf{x})} \tag{5}$$

where $P(\mathbf{x})$ and $P_a(\mathbf{x})$ are two probability distributions of the n random variables $\mathbf{x}$. The quantity $I(P, P_a)$ has the property $I(P, P_a) \geq 0$.

Translated to random variables, the above definition is named mutual information and is always nonnegative:

$$I(x_i, x_j) = \sum_{x_i, x_j} P(x_i, x_j) \log \left(\frac{P(x_i, x_j)}{P(x_i)P(x_j)} \right) \tag{6}$$

In the following, we will state the solution to the approximation of the probability distribution. We are searching for a distribution of tree dependence $P_\tau(x_1, \cdots, x_n)$ such that $I(P, P_\tau) \leq I(P, P_t)$ for all $t \in T_n$ where T_n represents the set of all possible first–order dependence trees. Thus, the solution τ is defined as the optimal first–order dependence tree.

In parlance of graph theory, every branch of the dependence tree is assigned a branch weight $I(x_i, x_{j(i)})$. Thus being given a dependence tree t, the sum of all branch weights becomes a useful quantity.

In[24] was shown that a maximum–weight dependence tree is a dependence tree t such that for all t' in T_n

$$\sum_{i=1}^{n} I(x_i, x_j(i)) \geq \sum_{i=1}^{n} I(x_i, x_{j'(i)}) \tag{7}$$

In other words, a probability distribution of tree dependence $P_t(\mathbf{x})$ is an optimum approximation to $P(\mathbf{x})$ if and only if its dependence tree t has maximum weight. Or, minimizing the closeness measure $I(P, P_t)$ is equivalent to maximizing the total branch weight.

2.3.2. Tree–Dependent Component Model

The idea of approximating discrete probability distributions with dependence trees described in the last section and adapted from,[24] can be easily translated to ICA.[23]

In classic ICA, we want to minimize the mutual information of the estimated components $\mathbf{s} = \mathbf{W}\mathbf{x}$. Thus, the result derived in,[24] can be easily extended and becomes the tree–dependent ICA.

The objective function for TCA is given by $J(\mathbf{x}, \mathbf{W}, t)$ and includes the demixing matrix $\mathbf{W}$. Thus, the mutual information for TCA becomes

$$J(\mathbf{x}, \mathbf{W}, t) = I^t(\mathbf{s}) = I(s_1, \cdots, s_m) - \sum_{(u,v)\in t} I(s_u, s_v) \tag{8}$$

$\mathbf{s}$ factorizes in a tree t.

In TCA as in ICA, the density $p(\mathbf{x})$ is not known and the estimation criteria have to be substituted by empirical contrast functions. As described in,[23] we will employ three types of contrast functions: (i) approximation of the entropies being part of equation (8) via kernel density estimation (KDE), (ii) approximation of the mutual information based on kernel generalized variance (KGV), and (iii) approximation based on cumulants using Gram–Charlier expansions (CUM).

2.4. Topographical Independent Component Analysis

The topographic independent component analysis[22] represents a unifying model which combines topographic mapping with ICA.

Achieved by a slight modification of the ICA model, it can at the same time be used to define a topographic order between the components, and thus has the usual computational advantages associated with topographic maps.

The paradigm of topographic ICA has its roots in[25] where a combination of invariant feature subspaces[26] and independent subspaces[27] is proposed. In the following, we will describe these two parts, which substantially reflect the concept of topographic ICA.[28]

2.4.1. Invariant Feature Subspaces

The principle of invariant-feature subspaces was developed by Kohonen[26] with the intention of representing features with some invariances. This principle states that an invariant feature is given by a linear subspace in a feature space. The value of the invariant feature is given by the squared norm of the projection of the given data point on that subspace.

A feature subspace can be described by a set of orthogonal basis vectors $\mathbf{w_j}, j = 1, \ldots, n$, where n is the dimension of the subspace. Then the value $G(\mathbf{x})$ of the feature G with the input vector $\mathbf{x}$ is given by

$$G(\mathbf{x}) = \sum_{j=1}^{n} < \mathbf{w_j}, \mathbf{x} >^2 \tag{9}$$

In other words, this describes the distance between the input vector $\mathbf{x}$ and a general linear combination of the basis vectors $\mathbf{w_j}$ of the feature subspace.[26]

2.4.2. Independent Subspaces

Traditional ICA works under the assumption that the observed signals $x_i(t), (i = 1, \ldots, n)$ are generated by a linear weighting of a set of n statistically independent random sources $s_j(t)$ with time-independent coefficients a_{ij}. In matrix form, this can be expressed as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{10}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_n(t)]^T, \mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]$, and $\mathbf{A} = [a_{ij}]$.

In multidimensional ICA,[27] the sources s_i are not assumed to be all mutually independent. Instead, it is assumed that they can be grouped in n-tuples, such that within these tuples they are dependent on each other, but are independent outside. This newly introduced assumption was observed in several image processing applications. Each n-tuple of sources s_i corresponds to n basis vectors given by the rows of matrix $\mathbf{A}$. A subspace spanned by a set of n such basis vectors is defined as an independent subspace. In[27] two simplifying assumptions are made: (1) Although s_i are not at all independent, they are chosen to be uncorrelated and of unit variance, and (2) the data are preprocessed by whitening (sphering) them. This means the $\mathbf{w_j}$ are orthonormal.

Let J be the number of independent feature subspaces and $S_j, j = 1, \ldots, J$ the set of indices that belong to the subspace of index j. Assume that we have T given observations $\mathbf{x}(t), t = 1, \ldots, T$. Then the likelihood L of the data based on the model is given by

$$L(\mathbf{w_i}, i = 1, \ldots, n) = \prod_{t=1}^{T}[|\det\mathbf{W}| \prod_{j=1}^{J} p_j(< \mathbf{w_i}, \mathbf{x}(t) >, i \in S_j)] \tag{11}$$

with $p_j(\cdot)$ being the probability density inside the jth n-tuple of s_i. The expression $|\det\mathbf{W}|$ is due to the linear transformation of the pdf. As always with ICA, $p_j(\cdot)$ need not be known in advance.

2.4.3. Fusion of Invariant Feature and Independent Subspaces

In[25] it is shown that a fusion between the concepts of invariant and independent subspaces can be achieved by considering probability distributions for the n-tuples of s_i being spherically symmetric, that is, depending on the norm. In other words, the pdf $p_j(\cdot)$ has to be expressed as a function of the sum of the squares of the $s_i, i \in S_j$ only. Additionally, it is assumed that the pdfs are equal for all subspaces.

The log-likelihood of this new data model is given by

$$\log L(\mathbf{w_i}, i = 1, \ldots, n) = \sum_{t=1}^{T}\sum_{j=1}^{J} \log p(\sum_{i \in S_j} < \mathbf{w_i}, \mathbf{x}(t) >^2) + T \log|\det\mathbf{W}| \tag{12}$$

$p(\sum_{i \in S_j} s_i^2) = p_j(s_i, i \in S_j)$ gives the pdf inside the jth n-tuple of s_i. Based on the prewhitening, we have $\log|\det\mathbf{W}| = 0$.

For computational simplification, set

$$G(\sum_{i \in S_j} s_i^2) = \log p(\sum_{i \in S_j} < \mathbf{w_i}, \mathbf{x}(t) >^2) \tag{13}$$

Since it is known that the projection of visual data on any subspace has a super-Gaussian distribution, the pdf has to be chosen to be sparse. Thus, we will choose $G(u) = \alpha\sqrt{u} + \beta$ yielding a multidimensional version of an exponential distribution. α and β are constants and enforce that s_i is of unit variance.

2.4.4. The Topographic ICA Architecture

Based on the concepts introduced in the preliminary subsections, this section describes the topographic ICA.

To introduce a topographic representation in the ICA model, it is necessary to relax the assumption of independence among neighboring components s_i. This makes it necessary to adopt an idea from self-organized neural networks, that of a lattice. It was shown in[22] that a representation which models topographic correlation of energies is an adequate approach for introducing dependencies between neighboring components.

In other words, the variances corresponding to neighboring components are positively correlated while the other variances are in a broad sense independent. The architecture of this new approach is shown in Figure 2.

This idea leads to the following representation of the source signals:

$$s_i = \sigma_i z_i \tag{14}$$

where z_i is a random variable having the same distribution as s_i, and the variance σ_i is fixed to unity.

The variance σ_i is further modeled by a nonlinearity:

$$\sigma_i = \phi\left(\sum_{k=1}^{n} h(i,k)u_k\right) \tag{15}$$

where 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 either be a two-dimensional grid or have a ring-like structure. Further u_i and z_i are all mutually independent.

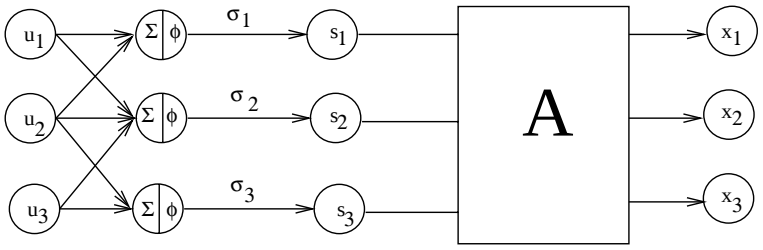

Figure 2. Topographic ICA model.[22] The variance generated variables u_i are randomly generated and mixed linearly inside their topographic neighborhoods. This forms the input to nonlinearity ϕ, thus giving the local variance σ_i. Components s_i are generated with variances σ_i. The observed variables x_i are obtained as with standard ICA from the linear mixture of the components s_i.

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

$$\log L(\mathbf{w_i}, i = 1, \ldots, n) = \sum_{t=1}^{T}\sum_{j=1}^{n} G(\sum_{i=1}^{n}(\mathbf{w_i}^T\mathbf{x}(t)^2) + T\log|\det\mathbf{W}| \tag{16}$$

The update rule for the weight vector $\mathbf{w}_i$ is derived from a gradient algorithm based on the log-likelihood assuming $\log|\det\mathbf{W}| = 0$:

$$\Delta\mathbf{w}_i \propto E\{\mathbf{x}(\mathbf{w}_i^T\mathbf{x})r_i\} \tag{17}$$

where

$$r_i = \sum_{k=1}^{n} h(i,k) g(\sum_{j=1}^{n} h(k,j)(\mathbf{w}_j^T \mathbf{x})^2) \tag{18}$$

The function g is the derivative of $G = -\alpha_1 \sqrt{u} + \beta_1$. After every iteration, the vectors $\mathbf{w}_i$ in eq. (17) 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.

The classic ICA results from the topographic ICA by setting $h(i,j) = \delta_{ij}$.

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.[18] The first scans were discarded for remaining saturation effects. Motion artifacts were compensated by automatic image alignment (AIR,[29]).

The clustering results were evaluated by (1) task–related activation maps, (2) associated time–courses and (3) ROC curves.

3.1. ESTIMATION OF THE ICA MODEL

To decide to what extent spatial ICA of fMRI time–series depends on the employed algorithm, we have first to look at the optimal number of principal components selected by PCA and used in the ICA decomposition. ICA is a generalization of PCA. In case no ICA is performed, then the number of independent components equals zero, and this means there is no PCA decomposition performed.

In the following we will give the set parameters. For PCA, no parameters had to be set. For FastICA we choose: (1) $\epsilon = 10^{-6}$, (2) 10^5 as the maximal number of iterations, and (3) the nonlinearity $g(u) = \tanh u$. And last, for topograph ic ICA we set: (1) stop criterium is fullfilled 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.

It is significant to find a fixed number of ICs that can theoretically predict new observations in same conditions, assuming the basic ICA model actually holds. To do so, we compared the six proposed algorithms for 8, 9, and 16 components in terms of ROC analysis using correlation map with a chosen threshold of 0.4. The obtained results are plotted in Figure 3. It can be seen that topographic ICA outperforms all other ICA methods for 8 and 9 ICs. However, for 16 ICs topographic ICA is outperformed by both FastICA and tree–dependent ICA (KGV) using as an approximation of the mutual information the kernel generalized variance.

The clustering results for the two method, the tree–dependent and topographic ICA are shown in Figure 4. Figure 4 (a) and (c) illustrates the so–called assignment maps where all the pixels belonging to a specific cluster are highlighted. The assignment between a pixel and a specific cluster is given by the minimum distance between the pixel and a IC from the established codebook. On the other hand, each IC shown in Figure 4 (b) and (d) can be viewed as the cluster–specific weighted average of all pixel time courses.

3.2. CHARACTERIZATION OF TASK–RELATED EFFECTS

For all subjects, and runs, unique task–related activation maps and associated time–courses were obtained by the tree–dependent and topographic ICA techniques. The correlation of the component time course most closely associated with the visual task for the these two techniques is shown in Table 1 for IC=8,9, and 16. From the Table, we see for the tree–dependnet ICA a continuous increase for the correlation coefficient while for the topographic ICA this correlation coefficient decreases for IC=16.

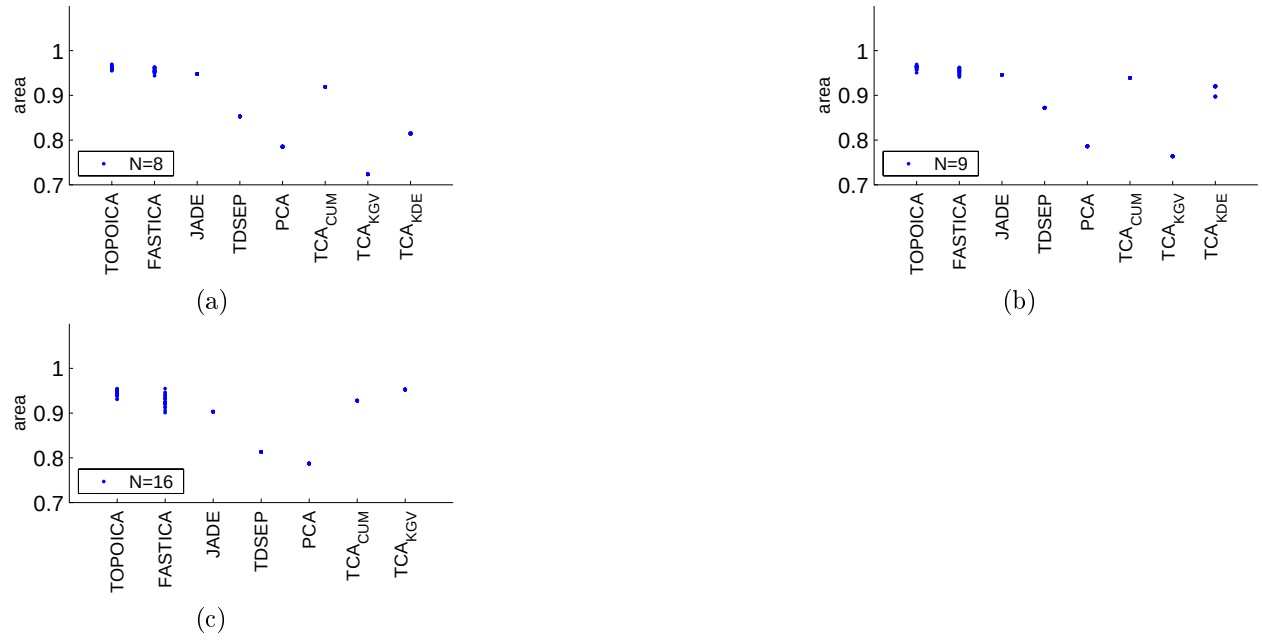

Figure 3. Results of the comparison between tree–dependent ICA, topographic ICA, Jade, FastICA, TDSEP, and PCA on fMRI data. Spatial accuracy of ICA maps is assessed by ROC analysis using correlation map with a chosen threshold of 0.4. The number of chosen independent components for all techniques is in (a): IC=8, (b): IC=9, and (c): IC=16.

Table 1. Comparison of the correlations of the component time course most closely associated with the visual task for tree–dependent and topographic ICA for IC=8,9, and 16.

	Tree–dependent ICA	Topographic ICA
IC=8	0.78	0.85
IC=9	0.91	0.87
IC=16	0.92	0.86

4. CONCLUSION

In the present paper, we have experimentally compared four standard ICA algorithms already adopted in the fMRI literature with two new algorithms, the tree– dependent and topographic ICA. 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 can be seen that topographic ICA outperforms all other ICA methods for 8 and 9 ICs. However, for 16 ICs topographic ICA is outperformed by both FastICA and tree–dependent ICA (KGV) using as an approximation of the mutual information the kernel generalized variance. The applicability of the new algorithms is demonstrated on experimental data.

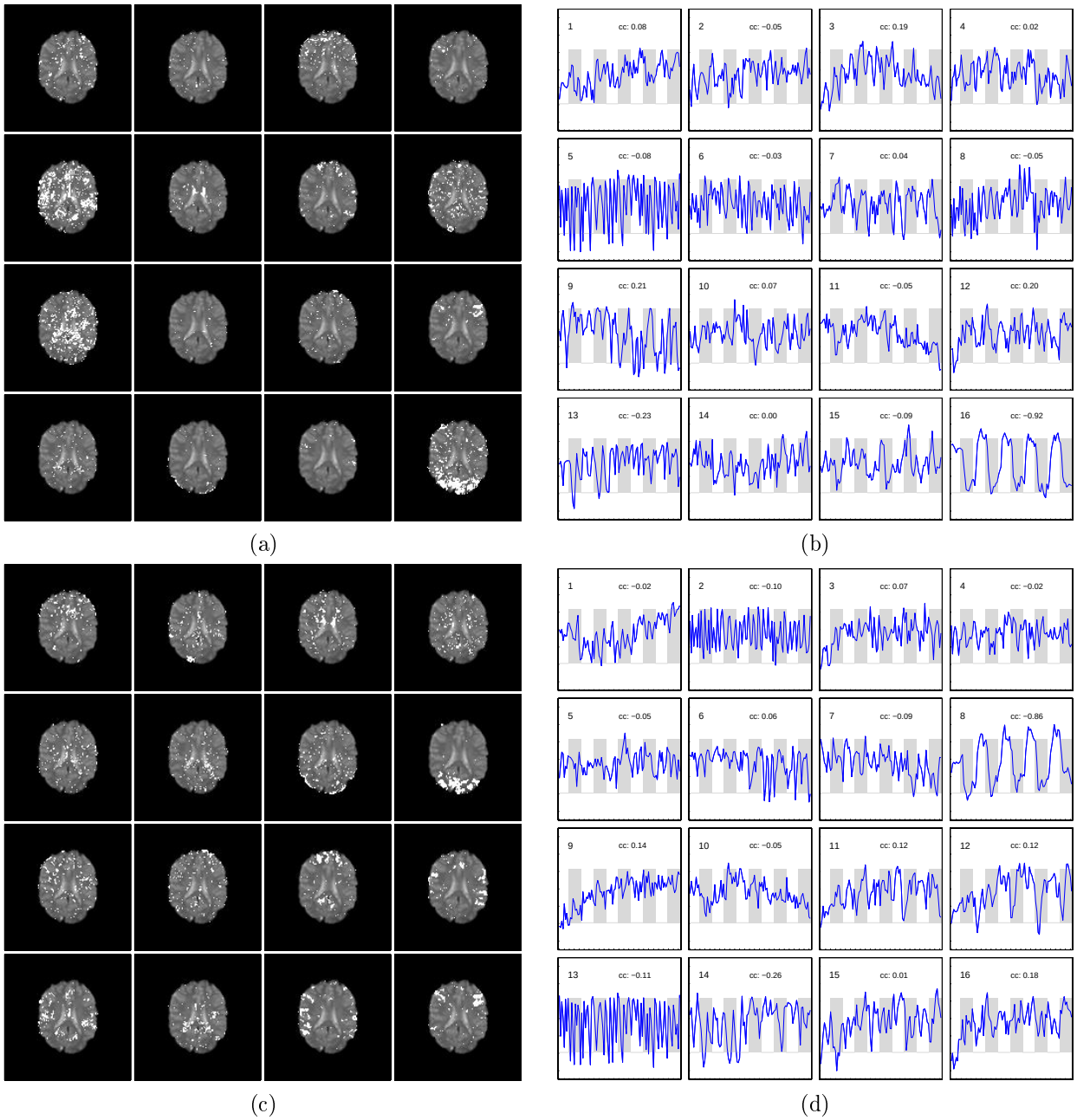

Figure 4. Results for a visual stimulation fMRI experiment and obtained for 16 ICs for tree–dependent ICA: (a) cluster assignment maps and (b) associated ICs. Results for topographic ICA: (c) cluster assignment maps and (d) associated ICs.

REFERENCES

1. P. Bandettini, E. Wong, R. Hinks, R. Tikofski, and J. Hyde, "Time course epi of human brain function during task activation," *Magn. Reson. Med.* **25**, pp. 390–397, 8 1992.
2. J. Frahm, K. Merboldt, and W. Hanicke, "Functional mri of human brain activation at high spatial resolution," *Magn. Reson. Med.* **29**, pp. 139–144, 8 1992.
3. K. Kwong, "Functinal magnetic resonance imaging with echo planar imaging," *Magn. Reson. Q.* **11**, pp. 1–20, 8 1992.
4. K. Kwong, J. Belliveau, and D. Chesler, "Dynamic magnetic resonance imaging of human brain activity during primary sensor stimulation," *Proceedings of the National Academy of Science* **89**, pp. 5675–5679, 8 1992.
5. 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.
6. J. Boxerman, P. Bandettini, K. Kwong, and J. Baker, "The intravsacular contribution to fmri signal change: Monte carlo modeling and diffusion-weighted studies in vivo," *Magn. Reson. Med.* **34**, pp. 4–10, 8 1995.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. A. Wismüller, O. Lange, D. Dersch, G. Leinsinger, K. Hahn, B. Pütz, and D. Auer, "Cluster analysis of biomedical image time–series," *International Journal on Computer Vision* **46**, pp. 102–128, 2 2002.
19. H. Fisher and J. Hennig, "Clustering of functional mr data," *Proc. ISMRM 4rd Annu. Meeting* **96**, pp. 1179–1183, 8 1996.
20. 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.
21. A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks* **13**, pp. 411–430, 5 2000.

22. A. Hyvarinen and P. Hoyer, "Topographic independent component analysis," *Neural Computation* **13**, pp. 1527–1558, 7 2001.
23. F. R. Bach and M. I. Jordan, "Beyond independent components: Trees and clusters," *Journal of Machine Learning Research* **4**, pp. 1205–1233, 3 2003.
24. C. K. chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transaction on Information Theory* **14**, pp. 462–467, 3 1968.
25. A. Hyvarinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation* **12**, pp. 1705–1720, 7 2000.
26. T. Kohonen, "Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map," *Biological Cybernetics* **75**, pp. 281–291, 1 1996.
27. J. F. Cardoso, "Multidimensional independent component analysis," *Proc. IEEE ICASSP, Seattle* **4**, pp. 1941–1944, 5 1998.
28. A. Meyer-Bäse, *Pattern Recognition for Medical Imaging*, Elsevier Science/Academic Press, 2003.
29. 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.