Analysis of fMRI time series with mutual information

Vanessa Gómez-Verdejo a, Manel Martínez-Ramón a,⇑, José Florensa-Vila b, Antonio Oliviero b

a Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Leganés, Madrid 28911, Spain
b Hospital Nacional de Parapléjicos de Toledo, SESCAM, Toledo 45004, Spain

ARTICLE INFO

Article history:
Received 15 February 2011
Received in revised form 7 November 2011
Accepted 8 November 2011
Available online 25 November 2011

Keywords:
Mutual information
Statistical parametric mapping
fMRI

ABSTRACT

Neuroimaging plays a fundamental role in the study of human cognitive neuroscience. Functional magnetic resonance imaging (fMRI), based on the Blood Oxygenation Level Dependent (BOLD) signal, is considered as a consolidated technique for a system level understanding of the human brain. The problem of identifying regionally specific effects in neuroimaging data is usually solved by applying Statistical Parametric Mapping (SPM). Here, a mutual information (MI) criterion is used to identify regionally specific effects produced by a task. In particular, two MI estimators are presented for its use in fMRI data. The first one uses a Parzen probability density estimation, and the second one is based on a K Nearest Neighbours (KNN) estimation. Additionally, a statistical measure has been introduced to automatically detect the voxels which are relevant to the fMRI task. Experiments demonstrate the advantages of MI estimators over SPM maps; firstly, providing more significant differences between relevant and irrelevant voxels; secondly, presenting more focalized activation; and, thirdly, detecting small areas related to the task. These findings, and the improved performance of KNN MI estimator in multisubject and multistimuli studies, make the proposed methods a good alternative to SPM.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Neuroimaging plays a fundamental role in the study of human cognitive neuroscience. Functional magnetic resonance imaging (fMRI), based on the Blood Oxygenation Level Dependent (BOLD) signal, is considered as a consolidated technique for a system level understanding of the human brain. Since its first description by Ogawa et al. (1990), the number of BOLD fMRI reports has sharply increased (Bandettini, 2007), underlining the importance of this technique. fMRI allows a characterization of the functional anatomy, the localization of functional areas and the localization of distributed networks necessary for the functional integration.

Today the problem of identifying regionally specific effects in neuroimaging data is usually solved by applying Statistical Parametric Mapping (SPM) (Friston et al., 2007), previously introduced for PET data analysis (Friston et al., 1990) which combines two stages: (1) a General Linear Model (GLM) and (2) a statistical test as a linear combination of an unknown brain response $\mathbf{X}$, which stands for the effects of interest (one or more reference vectors or some psychophysical score), and term $\mathbf{e}$ is added to account for the model error. When used in fMRI, this model is extended by temporal smoothing of the haemodynamic function $\mathbf{X}$ as

$$
\mathbf{Y} = \mathbf{S}\mathbf{X}\mathbf{f} + \mathbf{e} 
$$

(2)

$\mathbf{S}$ being a filtering matrix emulating the dispersion associated to the haemodynamic response function. Then, $\mathbf{f}$ is estimated from (2).

In the second stage, the error covariance matrix of estimator $\mathbf{f}$ is evaluated using ReML (Harville, 1977) and, then, statistical inference methods based on this matrix such as $t$ of $F$ Student tests are applied. A full description of these techniques can be found in Friston et al. (2007).

Techniques based on Independent Component Analysis (ICA) (Hyvärinen et al., 2001) have been recently applied to fMRI. These methods consider that in the observed brain responses $\mathbf{X}$, both stimulus signals and biological activities have been linearly mixed. Thus, the goal of these approaches is “unmixing” these signals. In this case, due to the specific model involving the overall brain signals (both external stimulus and biological phenomena) is unknown, the problem is usually blindly solved with an ICA algorithm (Calhoun et al., 2002), which makes ICA approach an exploratory technique, since unexpected signals can be uncovered. More recent works (Calhoun et al., 2005; Barriga et al., 2011) propose a semi blind ICA approach, where the information about the external stimuli is included in the unmixing process.

In this paper, a different point of view to solve the fMRI problem is addressed. The proposed approach is intended to find the voxels...
that are necessary to rebuild the stimulus signals of the design matrix or, at least, the most useful ones. For this purpose, a Mutual Information (MI) criterion (Cover and Thomas, 1991) is considered. MI is a concept borrowed from Information Theory and widely accepted in the Machine Learning field as an accurate measure to find the most relevant data features (see among others Battiti, 1994; Kwak and Choi, 2002a,b; Peng et al., 2005; Torkkola et al., 2003).

This criterion has the advantage, over other statistical measures, of being able to measure nonlinear relationships between variables. This fact may provide enhanced performance in fMRI time series analysis as suggested in (Boynont et al., 1996; Friston et al., 1998) for event related experiments. However, due to the fact that the underlying data probability distribution is unknown, probability distribution estimators (Duda et al., 2001; Scott, 1992) or methods that directly estimate MI (Kraskov et al., 2004; Gómez-Verdejo et al., 2009) are needed.

MI techniques have been largely applied to medical image pre-processing. In Viola and Wells (1995), Wells et al. (1996), Plum et al. (2003), Mellor and Brady (2005), Andronache et al. (2008), medical image registration techniques using MI techniques can be found, and its application to fMRI is proposed in Kim et al. (1999), Freire and Mangina (2001), Otte (2001). Several MI measures have been recently used to assess the connectivity between pairs of brain regions in fMRI (see, for instance, Gretton et al., 2006, Hinrichs et al., 2006), or to describe the degree of connectivity between clusters of brain regions (Salvador et al., 2007). Nevertheless, there are few attempts (Tsai et al., 1999; Li et al., 2004; Tedeschi et al., 2005; Alpert et al., 2007) to introduce a MI criterion for the identification of regionally specific effects in fMRI. In all these cases, discrete reference vectors have been used as explanatory variables, leading to an unidimensional MI formulation which is easily solved with histogram based estimators. Despite of the simplicity of these approaches, this discrete formulation together with an histogram based MI estimation is one of the main flaws of these approaches, since it prevents them from including the HRF in the signal modelling. Using a continuous reference vector (as in the present proposal) leads to a 2-dimensional space. If N volumes are needed for an accurate estimation in a one dimensional space, O(N^2) volumes may be needed to obtain a similar performance in a 2-dimensional space. Another approach based on Parzen estimator is used in (Tsai et al., 1999), which has not the limitation of the above proposals, but that keeps using a discrete function.

Two additional drawbacks can be pointed out: firstly, no statistical test (similar to the t or F statistics used in SPM) has been proposed to analyze the significance of MI results; and, secondly, their application to multisubject studies does not achieve the desired outcome because both histogram and Parzen estimators significantly degrade their performance as data dimensionality is increased.

In this paper, the above limitations are overcome by introducing two new MI approaches: the first one is based on the K-Nearest Neighbours (KNN) MI estimator proposed by Kraskov et al. (2004) and the second one is an extension of the Parzen probability density estimator used by Tsai et al. (1999). Both estimators use a formulation which allows the use of continuous explanatory variables in order to account for the haemodynamic response function, making them more appropriate for its use in fMRI in single subject and multisubject studies. The introduced KNN estimator is robust to Parzen one when MI. An additional novelty of this work, with respect to the previous attempts, consists of including a statistical test based on the fact that the MI measurement can be approximated by a χ^2 distribution function. This test provides a threshold to remove those voxels which are unrelated to the stimulus signal with a given probability, and standard tools as the Family Wise Error criterion can be straightforwardly applied to the procedure.

Experiments suggest that MI maps outperform SPM ones providing more focalized activation and more significant differences between relevant and irrelevant voxels in a consistent way across all subjects. A thresholding process based on a statistical criterion can be applied showing that the presence of voxels unrelated to the task in MI maps is much less frequent than in thresholded SPM maps, thus obtaining more confident results. Additionally, the KNN MI estimator allows not only to carry out multisubject studies but also to detect small activated areas that are not detected with the standard SPM multisubject study.

2. Methods

2.1. An introduction to mutual information

MI criteria measure the statistical relationship among two or more variables. That is, it provides a measure on the relevance of a variable to predict the other one. Let U and V be two dependent variables and let us denote their marginal density probability distributions as p_u(u) and p_v(v) respectively, and their joint probability distribution as p_{uv}(u,v); then, the MI between U and V is given by (Cover and Thomas, 1991):

\[
I(U;V) = \int \int p_{uv}(u,v) \log \frac{p_{uv}(u,v)}{p_u(u)p_v(v)} \, du \, dv
\]

(3)

When U and V are independent, p_{uv}(u,v) = p_u(u)p_v(v) and the MI value is zero, whereas if both variables are closely related the MI value is large.

Due to the fact that computing (3) entails a double integral over the domains of U and V, the continuous expressions of p_u(u), p_v(v) and p_{uv}(u,v) need to be known. This drawback can be solved by taking into account that (3) defines the expected value of the quotient logarithm between joint and marginal probabilities, i.e.,

\[
I(U;V) = \int \int p_{uv}(u,v) \log \frac{p_{uv}(u,v)}{p_u(u)p_v(v)} \, du \, dv = E \left\{ \log \frac{p_{uv}(u,v)}{p_u(u)p_v(v)} \right\}
\]

(4)

Thus, this expectation can be directly evaluated over a set of N identically distributed samples and a sample-based estimation of the actual MI value is obtained as

\[
\hat{I}(U;V) = \frac{1}{N} \sum_{n=1}^{N} \log \frac{p_{uv}(u_n,v_n)}{p_u(u_n)p_v(v_n)}
\]

(5)

To compute this measure the probability density functions of the involved signals must be known. However, in any practical application, such as fMRI problem, the probability density functions are unknown and only a finite data set is available. To deal with this limitation, a probability distribution estimate using Parzen and a direct MI estimation using KNN distances are considered in this paper.

In the following sections, these approaches will be explained in detail, as well as their application to the fMRI problem.

2.2. MI estimation based on Parzen probability density

Parzen probability density approach (Parzen, 1962) is a nonparametric way of estimating the unknown probability density function of a random variable from which only a set of sampled data is available. Let \{(u_1, v_1), \ldots, (u_N)\} be a set of identically distributed sampled data from the random variable U, then, the Parzen approximation of its density function, \(p_u(u)\), is given by

\[
p_u(u) = \frac{1}{N} \sum_{n=1}^{N} K \left( \frac{u - u_n}{h} \right)
\]

(6)

where \(K(\cdot)\) is the window or kernel function and \(h\) is its width parameter. Usually, a Gaussian kernel with covariance matrix \(\Sigma\) is considered as window function.
d being the dimension of the random variable $U$. Thus, $p_{\mathbf{u}}(\mathbf{u})$ is obtained with the following expression:

$$
\check{p}_{\mathbf{u}}(\mathbf{u}) = \frac{1}{N(2\pi)^{d/2}|\Sigma|^2} \sum_{\mathbf{u}_i=1}^{N} \exp \left( -\frac{(\mathbf{u} - \mathbf{u}_i)^T \Sigma^{-1} (\mathbf{u} - \mathbf{u}_i)}{2\sigma^2} \right)
$$

(8)

The Parzen estimator places a window function in each data sample location and, then, evaluates the probability density by summing up these kernels. When infinite length windows (as Gaussian windows) are used, they avoid obtaining probability estimates equal to zero in those locations where there are no data. This is particularly important to avoid numerical problems when computing (5) (note that $p_{\mathbf{u}}(\mathbf{u})$ and $p_{\mathbf{v}}(\mathbf{v})$ are in the denominator).

A direct method to estimate the MI consists of estimating the different probability distributions, $p_{\mathbf{u}}(\mathbf{u}), p_{\mathbf{v}}(\mathbf{v})$ and $p_{\mathbf{uv}}(\mathbf{v},\mathbf{u})$, according to (8) and then plugging them into (5).

For the particular case of the fMRI problem, the aim is to measure the MI value between the $l$th voxel of the smoothed brain response and the smoothed reference vector. Considering that for the $l$th voxel of the smoothed brain response of a single subject, the time series $(\mathbf{u}_n^l)$, $1 \leq n \leq N$ is available together with the smoothed reference time series $(\mathbf{v}_n)$, $1 \leq n \leq N$, then, the MI value between the $l$th voxel of the time series and the stimulus can be estimated as

$$
\hat{I}(\mathbf{U}^l, \mathbf{V}) = \frac{1}{N} \sum_{n=1}^{N} \log \frac{N}{\sigma_{\mathbf{u}_n} \sigma_{\mathbf{v}}}
$$

$$
\times \sum_{n=1}^{N} \exp \left( -\frac{(\mathbf{u}_n^l - \mathbf{v}_n)^T \Sigma_{\mathbf{uu}_n}^{-1} (\mathbf{u}_n^l - \mathbf{v}_n)}{2\sigma_{\mathbf{u}_n}^2} \right)
$$

$$
\times \sum_{n=1}^{N} \exp \left( -\frac{(\mathbf{v}_n - \mathbf{v}_n)^T \Sigma_{\mathbf{vv}}^{-1} (\mathbf{v}_n - \mathbf{v}_n)}{2\sigma_{\mathbf{v}}^2} \right)
$$

(9)

where $\Sigma_{\mathbf{uu}_n}$ is a vector with components $(\mathbf{u}_n^l, \mathbf{v}_n)$, $\sigma_{\mathbf{u}_n}^2$ and $\sigma_{\mathbf{v}}^2$ are the variances of the smoothed response of the $l$th voxel of the time series and the stimulus, respectively, and $\Sigma_{\mathbf{vv}}$ is their covariance matrix. Finally, forcing $\Sigma_{\mathbf{uu}_n}$ to be a diagonal matrix1 (their covariance terms are set to zero), expression (9) is simplified as follows:

$$
\hat{I}(\mathbf{U}^l, \mathbf{V}) = \sum_{n=1}^{N} \log \frac{N}{\sigma_{\mathbf{u}_n} \sigma_{\mathbf{v}}}
$$

$$
\times \sum_{n=1}^{N} \exp \left( -\frac{(\mathbf{u}_n^l - \mathbf{v}_n)^2}{2\sigma_{\mathbf{u}_n}^2} \right)
$$

$$
\times \sum_{n=1}^{N} \exp \left( -\frac{(\mathbf{v}_n - \mathbf{v}_n)^2}{2\sigma_{\mathbf{v}}^2} \right)
$$

(10)

2.3. MI estimation based on K-Nearest Neighbour distances

Another way to proceed is to directly estimate MI values, without first estimating the probability distributions. This can be easily carried out with the Karskow’s estimator (Karskow et al., 2004) based on Nearest Neighbours distances. The MI value between $\mathbf{U}^l$ and $\mathbf{V}$ can be obtained by means of

$$
H(\mathbf{U}^l, \mathbf{V}) = H(\mathbf{U}^l) + H(\mathbf{V}) - H(\mathbf{U}^l, \mathbf{V})
$$

(11)

$H(\cdot)$ being an entropy estimator. This entropy can be computed using the well-known Kozachenko-Leonenko estimator for differential Shannon entropy (Kozachenko and Leonenko, 1987):

$$
\tilde{H}(\mathbf{U}^l) = -\psi(K) + \psi(N) + \log c_d - \frac{1}{N} \sum_{n=1}^{N} \log \xi(n, K)
$$

(12)

where $\psi(\cdot)$ is the digamma function, $K$ is the number of Nearest Neighbours (a parameter of the algorithm), $N$ is the number of samples in the data set, $d$ is the dimensionality of $\mathbf{U}^l$, $c_d$ is the volume of a unitary ball in a $d$-dimensional space, and $\xi(n, K)$ is twice the distance from $\mathbf{u}^l_n$ to its $K$th neighbour. Substituting (12) into (11) and after some algebraic manipulations, the desired MI estimator is given by (see Kraskov et al. (2004) for further details):

$$
\tilde{H}(\mathbf{U}^l, \mathbf{V}) = \psi(N) + \psi(K) - \frac{1}{N} \sum_{n=1}^{N} \left[ \psi(\xi(n, n)) + \psi(\tau_s(n)) \right]
$$

(13)

where $\tau_s(n)$ and $\tau_s(n)$ are the number of neighbors within $\epsilon(n, K)$/2 in the subspaces spanned by $\mathbf{u}^l_n$ and $\mathbf{v}_n$, respectively, $\epsilon(n, K)$ being twice the distance from the point $\mathbf{u}^l_n, \mathbf{v}_n$ to its $K$th neighbour.

The Parzen MI estimator is computed as a quotient of estimated densities, which tends to be affected by errors in the estimation of prior probabilities (see Eq. (5)). However, the above KNN-based method has the advantage of simply replacing entropies with their estimates (see Eq. (11)), which facilitates the error cancellation in the individual entropy estimation. This feature provides practical advantages to the KNN MI estimator with regard to Parzen one, some more accurate results can be obtained, making easier the application of a statistical test over its results and allowing us applying different multisubject strategies.

2.4. Extension to multisubject studies

The proposed MI estimator formulations can be straightforwardly extended to be used in multisubject studies. Let us consider $\mathbf{U}^l = [\mathbf{u}_1^l, \ldots, \mathbf{u}_J^l]^T$ to be a vector built up by the smoothed brain responses in the $l$th voxel of $J$ subjects under study. Then, the Parzen estimator in this multisubject study can be computed just extending (10) by including a multidimensional estimator of $p_{\mathbf{uv}}(\mathbf{u}^l)$ under the assumption of independence among subjects, i.e.,

$$
\tilde{I}(\mathbf{U}^l, \mathbf{V}) = \sum_{j=1}^{N} \log \frac{N}{\sigma_{\mathbf{u}_j} \sigma_{\mathbf{v}}}
$$

$$
\times \sum_{n=1}^{N} \exp \left( -\frac{(\mathbf{u}_j^l - \mathbf{v}_j)^2}{2\sigma_{\mathbf{u}_j}^2} \right)
$$

$$
\times \sum_{n=1}^{N} \exp \left( -\frac{(\mathbf{v}_j - \mathbf{v}_j)^2}{2\sigma_{\mathbf{v}}^2} \right)
$$

(14)

This extension to multisubject studies applies the MI estimation has to be carried out in a $J$th dimensional space, whereas the number of data remains being $N$. This translates into a degradation in the probability density estimation accuracy, thus dramatically reducing the MI estimator performance, which may preclude its usage in practice. However, as it will be checked in the experimental section, the nature of the K-NN estimator provides itself of a higher robustness and the dimensional increase does not degrade its performance.

The above approach has an inherent limitation in the fact that all subjects into study must share the same design matrix, since all time series are introduced in a parallel form. An alternative approach is possible where design matrices can be independently constructed by every single subject, consisting of introducing time series of all subject in series, or in a single extended vector. That is, let the time series for subject $j$ be $\mathbf{u}_j = [\mathbf{u}_1^j, \ldots, \mathbf{u}_d^j]^T$. Then, an extended column vector containing all subjects can be constructed as $\mathbf{u} = [\mathbf{u}_1^j, \ldots, \mathbf{u}_d^j]$. Then, BOLD responses can just be formatted as

1 Although the covariance terms of matrix $\Sigma_{\mathbf{uu}_n}$ may be relevant, preliminary results have corroborated that similar results can obtained forcing them to be zero.

2 The digamma function, $\psi$, is the logarithmic derivative of the gamma function.
the same way and the MI estimation process can be performed as in the single subject analysis.

2.5. A statistical analysis over MI measurements

The MI measurement can be seen as a likelihood ratio test (Kim et al., 2000; Suzuki et al., 2008). For this purpose, let us define the following hypothesis test:

\[ H_0 : \mathbf{U}^0 \text{ and } \mathbf{V} \text{ are independent} \quad P(\mathbf{U}^0, \mathbf{V}) = P(\mathbf{U}^0)P(\mathbf{V}) \]

\[ H_1 : \mathbf{U}^0 \text{ and } \mathbf{V} \text{ are dependent} \quad P(\mathbf{U}^0, \mathbf{V}) \neq P(\mathbf{U}^0)P(\mathbf{V}) \]

Then, the MI measure which allows us decide whether \( f^0 \) and \( V \) are independent or not from a set of \( N \) independently distributed samples, \( \{u^0_n, v_n\}_{n=1}^N \), is given by:

\[
\log \frac{P_{U,V}(u^0_n, v_n|H_1)}{P_{U,V}(u^0_n, v_n|H_0)} = \frac{1}{n} \sum_{n=1}^{N} \log \left( \frac{P_{U,V}(u^0_n, v_n)}{P_{U}(u^0_n)P_{V}(v_n)} \right)
\]

which is equal to \( \frac{1}{N} \log \left( \frac{P_U(u^0_n|V)}{P_U(u^0_n)} \right) \). Furthermore, considering that this log-likelihood ratio, multiplied by a factor \( 2 \log 2 \), approximate a \( \chi^2 \) distribution, we can use this relationship to establish an appropriate threshold. The number of freedom degrees of the \( \chi^2 \) distribution is estimated by computing the averaged value of the above set of samples. Then each sample (associated to a single voxel) can be compared to the specified significance level of the \( \chi^2 \) distribution.

2.6. Synthetic data

Experiments with synthetic data have been run in order to show the advantage of MI methods with respect to linear methods in presence of stimuli that is not linearly related to the BOLD responses used to construct the design matrix.

The used HRF signal \( f[n] \) consisted of the canonical function used in real fMRI experiments, representing exactly the block design used in the motor activation detection experiment described below.

The simulated brain response consisted of series of 160 volumes within a brain shape. We simulated two kinds of response in two balls with a diameter of 8 voxels. The first ball holds a time series linearly related to the HRF signal. The second ball provides the signal:

\[ u[n] = 0.4v[n] + 0.6 \times v[n] \sin(f_0n + \phi_0) \]

(15)

where \( f_0 = \frac{\pi}{20} \) and \( \phi_0 = 1.6\pi \). The frequency has been chosen so the sinusoidal function is synchronous with \( f[n] \) and the phase makes its maxima be aligned (see Fig. 1). Additionally, 10 random responses plus additive noise have been added. Random responses consisted of Gaussian noise filtered with 6th order Butterworth filters with cutoff frequencies uniformly distributed between 0 and 0.5. Each voxel time series was corrupted with independent realizations of white plus AR(1) Gaussian noises. All experiments had identical data, except for the noise amplitudes. Experiments were run for different SNR levels.

2.7. Subjects and paradigm 1: finger tapping fMRI experiment

Ten healthy subjects (5 female, 5 male, mean age 33.8 (±7.7) years, range 24–46, all right handed) were studied on a 3.0 Tesla Siemens Trio. Stimuli (start-stop) were presented via MR-compatible headphones (VisuaStim XGA; Resonance Technology, Northridge, CA). The paradigm consisted of a block with 20 s of activity and 20 s of resting. Subjects were asked to perform alternative finger tapping with maximum extension of the finger onto a button-response pad (VisuaStim XGA; Resonance Technology, Northridge, CA).

2.8. Subjects and Paradigm 2: sensorimotor and cognitive fMRI experiment

Fourteen healthy subjects were studied on a 4.0 Tesla Bruker MedSpec Scanner. Informed consent based on institutionally reviewed guidelines was obtained prior to participation in the study. Stimuli were presented via MR-compatible LCD goggles and headphones (Resonance Technology Inc., Northridge, CA). The paradigm consists of three interleaved tasks: visual (8 Hz checkerboard stimulation), motor (2 Hz right index finger tapping) and cognitive (mental calculation). These tasks are arranged in a randomized block design (8 s per block), with a cross-hair serving as baseline for a total of 132 s per scan. The total duration for each condition was thus approximately 27 s. Visual stimulation consisted of 8 Hz reversing black and white checker boards. Finger tapping in the motor task was paced with an auditory tone (1 kHz). Subjects were asked to tap with maximum extension of the finger onto a button-response pad (Cedrus corp., San Pedro, CA). The cognitive task consisted of mental calculations. Subjects were asked to sum three aurally presented numbers and divide the sum by three, responding with a button press when the sum was divisible by three without remainder (50% of trials). Subjects were instructed to attend to each task with a constant effort across scans and field-strengths.

fMRI data were acquired using single-shot echo-planar imaging with TR: 2000 ms, TE: 30 ms, flip angle: 90°, matrix size: 64 × 64 pixels, 32 slices. Slices were 3 mm thick, with 25% gap. Voxel dimensions were 3 × 3 × 3.75 mm. Hundred and sixty volumes were collected for a total measurement time of 320 s per scanning session.

2.9. Data analysis

Experiments were run in a computer grid running Matlab 7.4.0 (The Matworks, Inc.). T-maps and MI maps were computed for all available data: 160 volumes in the synthetic data and in the finger tapping experiment, and 66 in the sensorimotor experiment. Statistical parametric mapping using SPM8 (Kiebel and Friston, 2004a,b) was performed to generate t-maps representing brain activation changes. Preprocessing steps included motion correc-
3.2. Analysis of motor stimulus experiment

Raw maps from the Parzen method, parameter h was set to 1/log(N), N being the number of considered volumes; \( \sigma_n \) and \( \sigma_r \) were approximated by their sample estimations. For the KNN estimator, the number of neighbors \( K \) was set to 6. These parameters were adjusted as suggested in Kwak and Choi (2002a). Kraskov et al. (2004).

Results maps from SPM and MI algorithms were thresholded at a confidence level of 95% applying the Family Wise Errors approach (\( \alpha_{FWE} = 0.05 \)).

3.3. Analysis of sensorimotor and cognitive stimuli experiments

Fig. 4 compares the performances of the parallel Parzen and KNN and serial KNN MI methods against SPM. In this experiment, a Parzen estimator with all data sorted in parallel is no able to detect any kind of activation in multi-subjects studies as a consequence of the high dimensionality of the multisubject study (for this reason, parallel Parzen maps have not been included in this analysis). Our data may suggest that KNN MI methods are useful and outperform the SPM in the multisubject studies, as they can reduce the specific activation unrelated to the task.

3.3. Analysis of sensorimotor and cognitive stimuli experiments

Fig. 5 shows the visual, motor, and cognitive responses estimated using the 3 methods under study in a representative subject. In this figure four cuts in the axial plane corresponding to axial Talairach coordinates of 6 and 4 mm respectively. Raw maps and their responses after thresholding them (at a 95% confidence level) are displayed.

Raw maps of the visual stimulus present activity in the occipital lobe as expected. KNN approach showed the highest difference between activated and not activated voxels. After thresholding, only KNN presented voxels in the visual area whose value was over the 95% confidence level. Regarding to the motor activation, the three raw maps showed activity in the motor cortex. After thresholding, all voxels were removed from the Parzen MI estimator, where t-map and KNN showed significant voxels. KNN estimator shows a larger number of activated voxels in the motor cortex. Parzen and KNN MI estimators show similar performance identifying focal brain activation. When raw maps are compared, t-maps and Parzen estimator maps give high significance levels to areas probably unrelated to the task, while KNN estimator maps provide a remarkable difference between significant and probably irrelevant voxels. Furthermore, the activated area shows more focalization in KNN maps. After the thresholding process, both MI maps show that the presence of activity in voxels probably unrelated to the task is less frequent, thus obtaining more confident results than those of SPM.

4. Discussion and conclusion

Neuroimaging studies are playing an important role in characterizing brain states and cognitive processes. To establish fMRI role in this setting, characterization of functional patterns and their correlates with behavioral data are critical.

The most usual method to identify regionally specific effects in neuroimaging is SPM, which combines the GLM (an invariant linear estimator) with statistical inference. The use of MI has also been reported in the literature. The underlying idea of the MI approach is to estimate the MI between the brain response measured in each voxel and the stimuli reference vector. This has the advantage of being able to capture the possibly nonlinear relationships between the brain response and the reference vector, and the Gaussian assumption over the noise is drop down. Until now, these approaches have been limited to binary reference vectors. Furthermore, the lack of a

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>False alarm and detection rates of all algorithms and experiments.</td>
</tr>
<tr>
<td>t-map (%)</td>
</tr>
<tr>
<td>( \sigma_n=0.02 )</td>
</tr>
<tr>
<td>Rf</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
The statistical test associated to the MI approach has avoided its use as an alternative to SPM.

The aim of this work is to overcome these limitations by: (1) introducing a formulation that extends MI methods to real valued reference vectors that can account for the haemodinamic response dispersion; (2) proposing the use of a robust KNN MI estimator that can be used in multisubject studies with a reasonable sensitivity; (3) presenting a practical statistical measure based on the same principles of the applied in SPM test which allows the user to detect the significant voxels of the obtained MI maps. This method does not need to apply a ReML procedure.

Our findings evidence that MI estimators outperforms SPM maps from different points of view. Experiments with synthetic data reveal that MI approaches have better detection capabilities in presence of activation related to the task but with a low cross-correlation due to a nonlinear relationship between the BOLD response and the used HRF. MI maps provide more significant differences between voxels related and unrelated to the task, both in

Fig. 2. Brain response for linear and nonlinear stimulus estimated using the three methods under study in a synthetic experiment with different noise levels. Raw maps (left panel) and their responses after thresholding them at a 95% confidence level with a FWE approach (right panel) are displayed.

Fig. 3. Brain response estimated using the three methods under study in a representative subject. Raw maps (left panel) and their responses after thresholding them (at a 95% confidence level) are displayed. Talairach coordinates of axial slices are 66 mm, 55 mm, -20 mm and -23 mm.
raw and thresholded maps. Some studies present more focalized activation, in comparison with SPM, to detect cortical activation contralateral to finger movement, and to detect visual and cognitive activity. Finally, MI approaches are able to detect smaller areas related to the task. These advantages are probably due to the nonlinear nature of MI estimators.

Although Parzen and KNN estimators present similar performances when applied on a single subject study, Parzen performance shows lower sensitivity in activation detection compared to KNN. KNN estimator becomes a good alternative to SPM when a multisubject study is planned due to its improved performance. The last contribution of the presented work relies on the proposed significance
test, which provides an automatic procedure to remove all irrelevant
voxels, similar to t-map used in SPM and to which... Medical Image
Analysis 1 (1), 35–51.

458 V. Gómez-Verdejo et al. / Medical Image Analysis 16 (2012) 451–458

Acknowledgments

This work has been supported by FISCAM and Research Grants
TEC2011-22480 and TIN2011-24533 (Spanish MICINN). Authors
would like to acknowledge Dr. Stefan Possee (School of Medicine,
The University of New Mexico) for his useful comments and sug-
gestions and for his permission to use his sensorimotor and cogni-
tive fMRI data in this work.

References

Alpert, G.F., Sun, F.T., Handwerker, D., D’Esposito, M., Knight, R.T., 2007. Spatio-
temporal information analysis of event-related bold responses. Neuroimage 34
(4), 1545–1561.

registration of multi-modal images using both mutual information and cross-

Psychophysiology (63), 138–145.

Independent component analysis using prior information for signal detection in
a functional imaging system of the retina. Medical Image Analysis 15 (1), 35–44.

Battiti, R., 1994. Using the mutual information for selecting features in supervised

functional magnetic resonance imaging in human v1. Journal of Neuroscience
16, 4207–4221.

method for utilizing hypothesis-derived time courses in a spatial ICA analysis.
Neuroimage 25 (2), 527–538.

activation dynamics in multiple neural systems during simulated driving.


and Sons, New York.

Freire, L., Mangina, J.F., 2001. Motion correction algorithms may create spurious

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W., 2007. Statistical
Parametric Mapping. The Analysis of Functional Brain Images. Elsevier,
London, UK.

Friston, K., Holmes, A., Poline, J.-P., Grady, P.J., Williams, S., Frackowiak, R., Turner, R.,

Friston, K.J., Frith, C.D., Liddle, P.F., Dolan, R.J., Lammertsma, A.A., Frackowiak, R.S.J.,
1996. The relationship between global and local changes in pet scans. Journal of
Cerebral Blood Flow & Metabolism 10, 458–466.

in fMRI. Magnetic Resonance in Medicine 39, 41–52.

selection for functional data classification. Neurocomputing 72 (16–18), 3580–
3589.

Gretton, A., Belitski, A., Murayama, Y., Scholkopf, B., Logothetis, N., 2006. The effect of
artifacts on dependence measurement in fMRI. Magnetic Resonance Imaging 24,
401–409.

Harville, D.A., 1977. Maximum likelihood approaches to variance component
estimation and to related problems. Journal of the American Statistical
Association 72 (358), 320–338.

Hinrichs, H., Heinze, H., Schoenfeld, M.A., 2006. Causal visual interactions as
revealed by an information theoretic measure and fMRI. Neuroimage 31, 1051–
1065.

New York.


fMRI via registration of individual slices into an anatomical volume. Magnetic
Resonance in Medicine 41 (5), 964–972.

spatial priors into an information theoretic approach for fmi data analysis. In:
Delp, S.L., DiGioia, A.M., Jaramaz, B. (Eds.), MICCAI 2000. Pittsburgh,

random vector. Problems Information Transmission 23 (2), 9–16.

Physical Review E 69.

Kwak, N., Choi, C.H., 2002a. Input feature selection by mutual information based on
Parzen window. IEEE Transactions on Pattern Analysis and Machine Intelligence
24 (12), 1667–1671.

IEEE Transactions on Neural Networks 13 (1), 143–159.

Med. vol. 11, p. 1105.

Mellor, M., Brady, M., 2005. Phase mutual information as a similarity measure for

contrast dependent on blood oxygenation. In: Proceedings of the National


Parzen, E., 1962. On estimation of a probability density function and the model. The

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information:
criteria of max-dependency, max-relevance, and min-redundancy. IEEE
Transactions on Pattern Analysis and Machine Intelligence 27 (8), 1226–1238.

registration of medical images: a survey. IEEE Transactions on Medical Imaging
22 (8), 986–1904.

Salvador, R., Martin, S., Pomarol-Cloet, E., Sarrás, S., Suckling, J., Bullmore, E.,
2007. Frequency based mutual information measures between clusters of brain

Scott, D., 1992. Multivariable Density Estimation: Theory, Practice and


Tedeschi, W., Müller, H.P., de Araujo, D.B., Santos, A.C., Neves, U.P.C., Erné, S.N.,
Baffa, O., 2005. Generalized mutual information tests applied to fMRI analysis.

Torkkola, K., Guyon, I., Elisseeff, A., 2003. Feature extraction by non-parametric
mutual information maximization. Journal of Machine Learning Research
3, 1415–1438.

functional MRI data using Mutual Information. In: Medical Image Computing

Proceedings of the 5th International Conference on Computer Vision. IEEE,
Boston, MA.

volume registration by maximization of mutual information. Medical Image
Analysis 1 (1), 35–51.