ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 37 (2007) 1250–1259

Unsupervised analysis of fMRI data using kernel canonical correlation

David R. Hardoon,[a,*] Janaina Mourão-Miranda,[b] Michael Brammer,[b] and John Shawe-Taylor[a]

[a]*The Centre for Computational Statistics and Machine Learning, Department of Computer Science, University College London, UK*
[b]*Brain Image Analysis Unit, Centre for Neuroimaging Sciences (PO 89), Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK*

Received 22 February 2007; revised 9 June 2007; accepted 25 June 2007
Available online 3 July 2007

We introduce a new unsupervised fMRI analysis method based on kernel canonical correlation analysis which differs from the class of supervised learning methods (e.g., the support vector machine) that are increasingly being employed in fMRI data analysis. Whereas SVM associates properties of the imaging data with simple specific categorical labels (e.g., −1, 1 indicating experimental conditions 1 and 2), KCCA replaces these simple labels with a label vector for each stimulus containing details of the features of that stimulus. We have compared KCCA and SVM analyses of an fMRI data set involving responses to emotionally salient stimuli. This involved first training the algorithm (SVM, KCCA) on a subset of fMRI data and the corresponding labels/label vectors (of pleasant and unpleasant), then testing the algorithms on data withheld from the original training phase. The classification accuracies of SVM and KCCA proved to be very similar. However, the most important result arising form this study is the KCCA is able to extract some regions that SVM also identifies as the most important in task discrimination and these are located manly in the visual cortex. The results of the KCCA were achieved blind to the categorical task labels. Instead, the stimulus category is effectively derived from the vector of image features.

Keywords: Machine learning methods; Kernel canonical correlation analysis; Support vector machines; Classifiers; Functional magnetic resonance imaging data analysis

Introduction

Recently, machine learning methodologies have been increasingly used to analyse the relationship between stimulus categories and fMRI responses (Cox and Savoy, 2003; Carlson et al., 2003; Wang et al., 2003; Mitchell et al., 2004; LaConte et al., 2005; Mourao-Miranda et al., 2005, in press; Haynes and Rees, 2005; Davatzikos et al., 2005; Kriegeskorte et al., 2006). In this paper, we introduce a new unsupervised machine learning approach to fMRI analysis, in which the simple categorical description of stimulus type (e.g., type of task) is replaced by a more informative vector of stimulus features. We compare this new approach with a standard support vector machine (SVM) analysis of fMRI data using a categorical description of stimulus type.

The methodology underlying the present study originates from earlier research carried out in the domain of image annotation (Hardoon et al., 2006), where an image annotation methodology learns a direct mapping from image descriptors to keywords. Previous attempts at unsupervised fMRI analysis have been based on Kohonen self-organising maps, fuzzy clustering (Wismuller et al., 2004; Ngan and Hu, 1999) and non-parametric estimation methods of the hemodynamic response function, such as the general method described in Ciuciu et al. (2003), kernel-PCA (Thirion and Faugeras, 2003) and probabilistic ICA/PCA analysis (Beckmann and Smith, 2004). A more recent attempt has been undertaken by Faisan et al. (2005) with the application of hidden Markov event sequence models to fMRI. These Markov events are a special class of hidden Markov models (HMMs) dedicated to the modeling and analysis of event-based random processes. O'Toole et al. (2005) have reported an interesting study which showed that the discriminability of PCA basis representations of images of multiple object categories is significantly correlated with the discriminability of PCA basis representation of the fMRI volumes based on category labels.

The current study differs from previous approaches to fMRI analysis principally in that we do not apply categorical labels (e.g., −11 contrasts) to stimuli. We employ natural images rather than simple low level objects and transform each image to a vector representation summarising its main features. We then employ kernel canonical correlation analysis to associate the vector representations of image features with their corresponding fMRI image volumes. In general, canonical correlation analysis can be seen as the problem of finding basis vectors for two sets of variables such that the correlations of the projections of the variables onto corresponding basis vectors are maximised. KCCA differs from this in that it first projects the data into a higher dimensional feature space before performing CCA. CCA (Friman et al., 2001, 2003) and KCCA (Hardoon et al., 2004a) have been

* Corresponding author. Fax: +44 20 7387 1397.
E-mail address: D.Hardoon@cs.ucl.ac.uk (D.R. Hardoon).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.06.017

used in previous fMRI analysis, but using only conventional categorical stimulus labels. In contrast, in this work we are interested in learning the association between complex image representations and fMRI responses to characterise these associations. The fMRI data used in the following study originated from an experiment in which the responses to stimuli were designed to evoke different types of emotional responses, pleasant or unpleasant. The pleasant images consisted of women in swimsuits while the unpleasant images were a collection of images of skin diseases. Each stimulus image was represented using Scale Invariant Feature Transformation (SIFT) (Lowe, 1999) features.

We have shown that KCCA is able to extract some of the brain regions identified by supervised methods such as SVM in task discrimination (mainly in the visual cortex) and to achieve similar levels of accuracy. We discuss some of the challenges in interpreting the results given the complex input feature vectors used by KCCA in place of categorical labels.

The paper is structured as follows. Section 2 gives a review of the fMRI data acquisition as well as the experimental design and the pre-processing. These are followed by a brief description of the scale invariant feature transformation in Section 2.5. The SVM is briefly described in Section 2.6.1 while Section 2.6.2 elaborates on the KCCA methodology. Our analysis procedure is given in Section 2.7 and the results in Section 3. We conclude with a discussion in Section 4.

Materials and methods

Subjects

fMRI data were acquired from 16 right-handed healthy US college male students (aged 20–25). According to self-report, participants did not have any history of neurological or psychiatry illness. All subjects had normal vision. All subjects gave written informed consent to participate in the study after the study was explained to them. The study was performed in accordance with the local Ethics Committee of the University of North Carolina.

Data acquisition

The data for this study were collected at the Magnetic Resonance Imaging Research Center at the University of North Carolina on a 3-T Allegra Head-only MRI system (Siemens, Erlangen, Germany). The fMRI runs were acquired using a T2∗ sequence with 43 axial slices (slice thickness, 3 mm; gap between slices, 0 mm; TR = 3 s; TE = 30 ms; FA = 80°; FOV = 192 × 192 mm; matrix, 64 × 64; voxel dimensions, 3 × 3 × 3 mm). In each run 254 functional volumes were acquired.

Experimental design

The stimuli were presented in a block fashion. There were three different active conditions: viewing unpleasant (dermatological diseases), neutral (people) and pleasant images (female models in swimsuits) and a control condition (fixation). There were 42 imagers per category. Examples of pleasant and unpleasant are given in Tables 1 and 2, respectively (we do not show natural and fixation as we do not use these instances in our work). During the experiment, there were 6 blocks of each active condition (each consisting of 7 images volumes) alternating with control blocks (fixation) of 7 images volumes. It is important to note that throughout the paper we associate pleasant with positive and unpleasant with negative.

Pre-processing

The data was pre-processed using SPM2 (Wellcome Department of Cognitive Neurology, London, UK). We used the default SPM2 pre-processing settings. All the scans were realigned to remove residual motion effects, transformed into standard space (Talairach and Tournoux, 1988) and smoothed in space using an 8-mm Gaussian filter (FWHM). The time series of each voxel was detrended using a straight-line fit linear function. In addition, we applied a mask to select voxels defining intracerebral voxels over the whole group.

Table 1
Examples of pleasant image stimulus

Table 2
Examples of unpleasant image stimulus

Scale invariant feature transformation

The image representation is an important part of the analysis as we would like to extract as much detailed information as possible for the learning process. Various approaches have previously been suggested such as color moments and Gabor texture descriptors (Sebe et al., 2003) as well as scale invariant interest points (Mikolajczyk and Schmid, 2001) and the affine invariant interest point detector (Mikolajczyk and Schmid, 2002).

Scale Invariant Feature Transformation (SIFT) was introduced by Lowe (1999) and shown to be superior to other descriptors (Mikolajczyk and Schmid, 2003). This is due to the SIFT descriptors being designed to be invariant to small shifts in position of salient (i.e., prominent) regions. SIFT transforms the image data into scale invariant coordinates relative to local features. The underlying idea is to extract distinctive invariant features from an image such that they can be used to perform reliable matching between different views of an object or scene. An example of SIFT is given in Fig. 1 where SIFT features on a rotated object are linked to matching SIFT features on a new image. This representation was ideal for our earlier research, which was aimed at learning the association between keywords to an object. These could appear in different angles and scenes. Our current problem has a similar aim, of learning the association between image features and activity patterns. Therefore we believe SIFT to be an appropriate image representation.

Calculation of the SIFT vector begins with a scale space search in which local minima and maxima are identified in each image (so-called key locations). The properties of the image at each key location are then expressed in terms of gradient magnitude and orientation. A canonical orientation is then assigned to each key location to maximise rotation invariance. Robustness to reorientation is introduced by representing local image regions around key locations in a number of orientations. A reference key vector is then computed over all images and the data for each image are represented in terms of distance from this reference. Interestingly, some of the properties of the SIFT representation have been modeled on the properties of complex neurons in the visual cortex. Although not specifically exploited in the current paper, future studies may be able to utilise this property to probe aspects of brain function such as modularity.

Image processing

Let $\boldsymbol{f}_i^l$ be the SIFT features vector for image i, where l is the number of features. Each image i has a different number of SIFT features l, making it difficult to directly compare two images. To

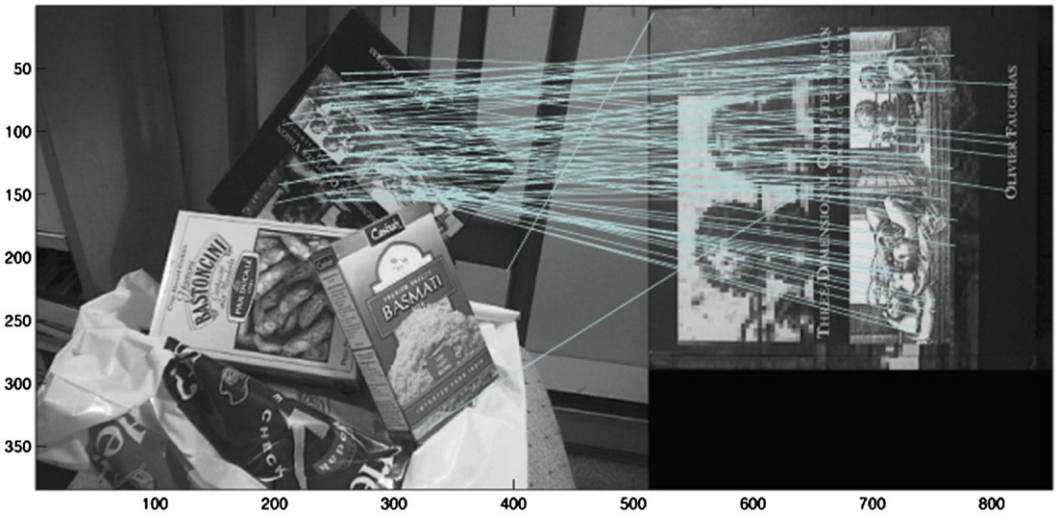

Fig. 1. Example of SIFT features similarity mapping.

overcome this problem, we apply K-means to cluster the SIFT features into a uniform frame. Using K-means clustering, we find K classes and their respective centers o_j, where $j=1,\ldots, K$. The feature vector $\boldsymbol{x}_i$ of an image stimuli i is K dimensional with jth component $\boldsymbol{x}_{i,j}$. The feature vectors are computed as the Gaussian measure of the minimal distance between the SIFT features $\boldsymbol{f}_i^l$ to the centre o_j. This can be represented as

$$\boldsymbol{x}_{i,j} = \exp^{-\left(\min_v \in \boldsymbol{f}_i^l d(v,o_j)^2\right)} \quad (2.1)$$

where $d(.,.)$ is the Euclidean distance.

For simplicity, the number of centres is set to be the smallest number of SIFT features computed, which was found to be 300. We use all images to compute the 300 centres and therefore after processing each image, we will have a 300-dimensional feature vector representing its relative distance from the cluster centres.

Methods

Support vector machines

Support vector machines are kernel-based methods that find functions of the data that facilitate classification. They are derived from statistical learning theory (Vapnik, 1995) and have emerged as powerful tools for statistical pattern recognition (Boser et al., 1992). In the linear formulation an SVM finds, during the training phase, the hyperplane that separates the examples in the input space according to their class labels. The SVM classifier is trained by providing examples of the form (x_a, y), where x_a represents an input and y as its class label. Once the decision function has been learned from the training data it can be used to predict the class of a new test example. In the present study, $\boldsymbol{x}_a$ represents an fMRI observation and y is the task performed ($y=1$ for task 1 (pleasant) and $y=-1$ for task 2 (unpleasant)). For a detailed description of SVMs, see Cristianini and Shawe-Taylor (2000). We used a linear kernel SVM that allows direct extraction of the weight vector as an image (i.e., the discriminating spatial pattern). A parameter C that controls the trade-off between training errors and smoothness was fixed at $C=1$ for all cases (default value).[1]

Kernel canonical correlation analysis

Proposed by Hotelling in 1936, Canonical Correlation Analysis (CCA) is a technique for finding pairs of basis vectors that maximise the correlation between the projections of paired variables onto their corresponding basis vectors. Correlation is dependent on the chosen coordinate system; therefore, even if there is a very strong linear relationship between two sets of multidimensional variables this relationship may not be visible as a correlation. CCA seeks a pair of linear transformations one for each of the paired variables such that when the variables are transformed the corresponding coordinates are maximally correlated.

Consider the linear combination $x=w_a'x$ and $y=w_b'y$. Let x and y be two random variables from a multidimensional distribution, with zero mean. The maximisation of the correlation between x and y corresponds to solving $\max_{wa,wb}\ \rho=w_a'C_{ab}w_b$ subject to $w_a'C_{aa}w_a=w_b'C_{bb}w_b=1$. C_{aa} and C_{bb} are the non-singular within-set covariance matrices and C_{ab} is the between-sets covariance matrix.

We suggest using the kernel variant of CCA (Fyfe and Lai, 2001) since due to the linearity of CCA useful descriptors may not be extracted from the data. This may occur as the correlation could exist in some non-linear relationship. The kernelising of CCA offers an alternate solution by first projecting the data into a higher dimensional feature space ϕ: $x=(x_1,\ldots, x_n)\rightarrow\phi(x)=(\phi_1(x),\ldots, \phi_N(x))$ $(N\geq n)$ before performing CCA in the new feature space.

Given the kernel functions κ_a and κ_b let $K_a=\mathbf{X}_a\mathbf{X}_a'$ and $K_b=\mathbf{X}_b\mathbf{X}_b'$ be the kernel matrices corresponding to the two representations of the data, where $\mathbf{X}_a$ is the matrix whose rows are the vectors $\phi_a(\boldsymbol{x}_i)$, $i=1,\ldots,\ell$, from the first representation (fMRI Volume) while $\mathbf{X}_b$ is the matrix with rows $\phi_b(\boldsymbol{x}_i)$ from the second representation (image stimulus). The weights w_a and w_b can be expressed as a linear combination of the training examples $w_a=\mathbf{X}_a\alpha$ and $w_b=\mathbf{X}_b\beta$. Substituting into the primal CCA equation gives the optimisation $\max_{\alpha,\beta,\rho}=\alpha' K_aK_b\beta$ subject to $\alpha' K_a^2\alpha=\beta' K_b^2\beta=1$. This is the dual form of the primal CCA optimisation problem given above, which can be cast as a generalised eigenvalue problem and for which the first k generalised eigenvectors can be found efficiently. Both CCA and KCCA can be formulated as an eigenproblem.

The theoretical analysis shown in Hardoon et al. (2004b) (and later in Shawe-Taylor and Cristianini (2004), Hardoon (2006)) suggests the need to regularise kernel CCA as it shows that the quality of the generalisation of the associated pattern function is controlled by the sum of the squares of the weight vector norms. We refer the reader to Hardoon et al. (2004b), Shawe-Taylor and Cristianini (2004) and Hardoon (2006) for a detailed analysis and the regularised form of KCCA.

Although there are advantages in using kernel CCA, which have been demonstrated in various experiments across the literature. We must clarify that in this particular work, as we are using a linear kernel[2] in both views, regularised CCA is exactly the same as regularised a linear KCCA (since the former and latter are linear). Although using KCCA with a linear kernel has advantages over CCA, the most important of which is in our case speed, together with the regularisation as well.[3]

Analysis

Using linear kernels as to allow the direct extraction of the weights, KCCA performs the analysis by projecting the fMRI volumes into the found semantic space defined by the eigenvector corresponding to the largest correlation value (these are outputted from the eigenproblem). We classify a new fMRI volume as follows; Let α_i be the eigenvector corresponding to the largest eigenvalue, and let $\phi(\hat{\boldsymbol{x}})$ be the new volume. We project the fMRI into the semantic space $w=\mathbf{X}_a'\alpha_i$ (these are the training weights, similar to that of the SVM) and using the weights we are able to classify the new example as

$$\hat{w} = \phi(\hat{\boldsymbol{x}})w$$

where $\hat{w}$ is a weighted value (score) for the new volume. The score can be thresholded to allocate a category to each test example. To avoid the complications of finding a threshold, we zero-mean the outputs (independently from the training or testing data) and

[1] The LibSVM toolbox for Matlab was used to perform the classifications http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[2] i.e., $K_a=\mathbf{X}_a\mathbf{X}_a'$.

[3] The KCCA toolbox used was from http://www.homepage.mac.com/davidrh/Code.html.

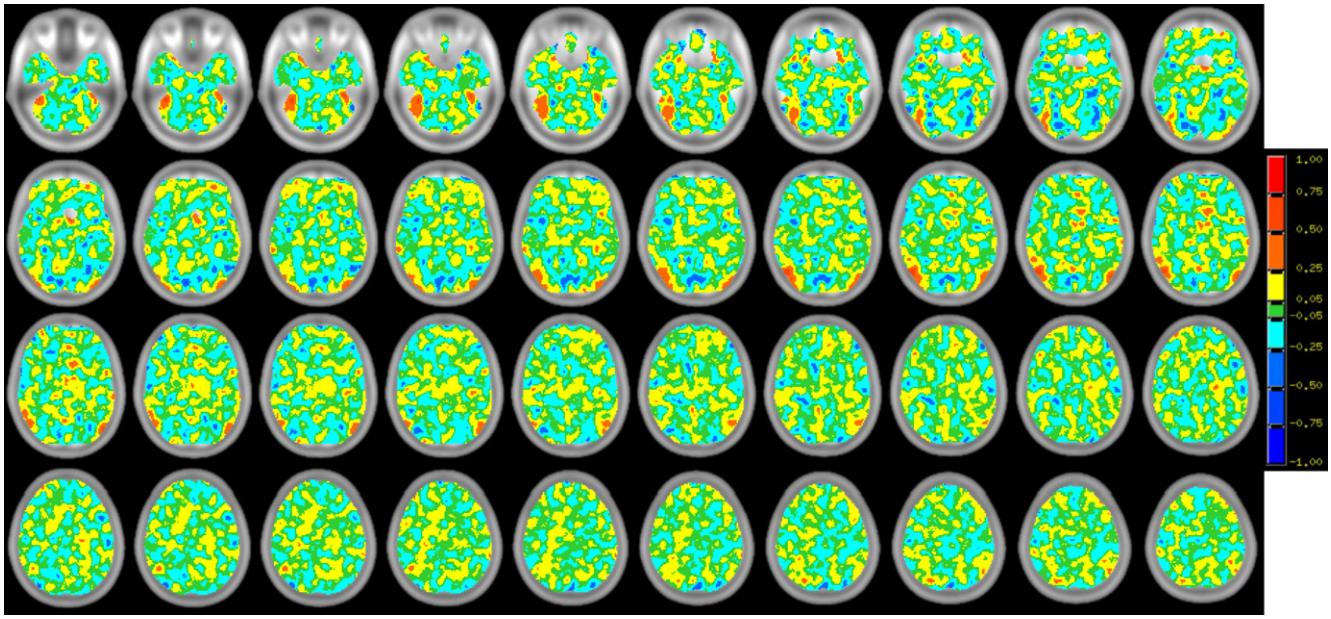

Fig. 2. The unthresholded weight values for the SVM approach showing the contrast between viewing pleasant vs. unpleasant. We use the blue scale for negative (unpleasant) values and the red scale for the positive values (pleasant). The discrimination analysis on the training data was performed with labels (+1/−1).

threshold the scores at zero, where $\hat{w}<0$ will be associated with unpleasant (a label of −1) and $\hat{w}\geq 0$ will be associated with pleasant (a label of 1).

We hypothesis that KCCA is able to derive additional activities that may exist a priori, but possibly previously unknown, in the experiment. By projecting the fMRI volumes into the semantic space using the remaining eigenvectors corresponding to lower correlation values. We have attempted to corroborate this hypothesis on the existing data but found that the additional semantic features that cut across pleasant and unpleasant images did not share visible attributes. We have therefore confined our discussion here to the first eigenvector.

KCCA and SVM

In order to compare the weight maps of the KCCA and SVM we evaluated the overlap between the peaks of both maps. The peaks were defined as areas above the threshold. For each approach the threshold was computed by training the KCCA/SVM with random labels[4] to generate the distribution of weight values under the null hypothesis of no relationship between the class labels and the global structure of the fMRI volumes. The threshold then was defined as the value correspondent to the 1st percentile of the null distribution, i.e., using this threshold one assumes a probability of 0.01% of getting a weight values greater or equal to the threshold by chance.

Results

Experiments were run on a leave-one-out basis where in each repeat a block of positive and negative fMRI volumes was withheld for testing. Data from the 16 subjects was combined. Giving a sum total of 96 blocks in each category and each block consisting of 7 fMRI volumes. This amounted, per run, in 1330 training and 14 testing fMRI volumes, each set evenly split into positive and negative volumes (these positive/negative splits were not known to KCCA but simply ensured equal number of images with both types of emotional salience). The analyses were repeated 96 times. Centralised linear kernels were used throughout. The KCCA regularisation parameter was found using 2-fold cross validation on the training data, where the optimisation criterion was classification performance.

Initially, we describe the fMRI activity analysis (the maps were computed on the training data). After training the SVM, we are able to extract and display the SVM weights as a representation of the brain regions important in the pleasant/unpleasant discrimination. A thorough analysis is presented in Mourao-Miranda et al. (in press). We are able to view the results in Figs. 2 and 3 where in both figures the weights are not thresholded and show the contrast between viewing pleasant vs. unpleasant. We use the blue scale for negative (unpleasant) values and the red scale for the positive values (pleasant). The weight value of each voxel indicates the importance of the voxel in differentiating between the two brain states. In Fig. 2, the unthresholded SVM weight maps are given. Similarly with KCCA, once learning the semantic representation we are able to project the fMRI data into the learnt semantic feature space producing the primal weights. These weights, like those generated from the SVM approach, could be considered as a representation of the fMRI activity. Fig. 3 displays the KCCA weights.

Figs. 4 and 5 display the overlap of voxels above the threshold in the SVM weight and in the KCCA weight. Voxels with positive weights in both KCCA and SVM are shown in Fig. 4 and voxels with negative weights in both are shown in Fig. 5. Values above the threshold in both approaches are colored in red, voxels above the threshold only in the KCCA weight are colored in blue and voxels above the threshold only in the SVM are colored in green. The brain areas corresponding to the clusters above the threshold are listed in Table 3 (positive values) and Table 4 (negative values).

[4] In KCCA, we randomise the image stimulus for the training data and for SVM we randomise the categorical labels.

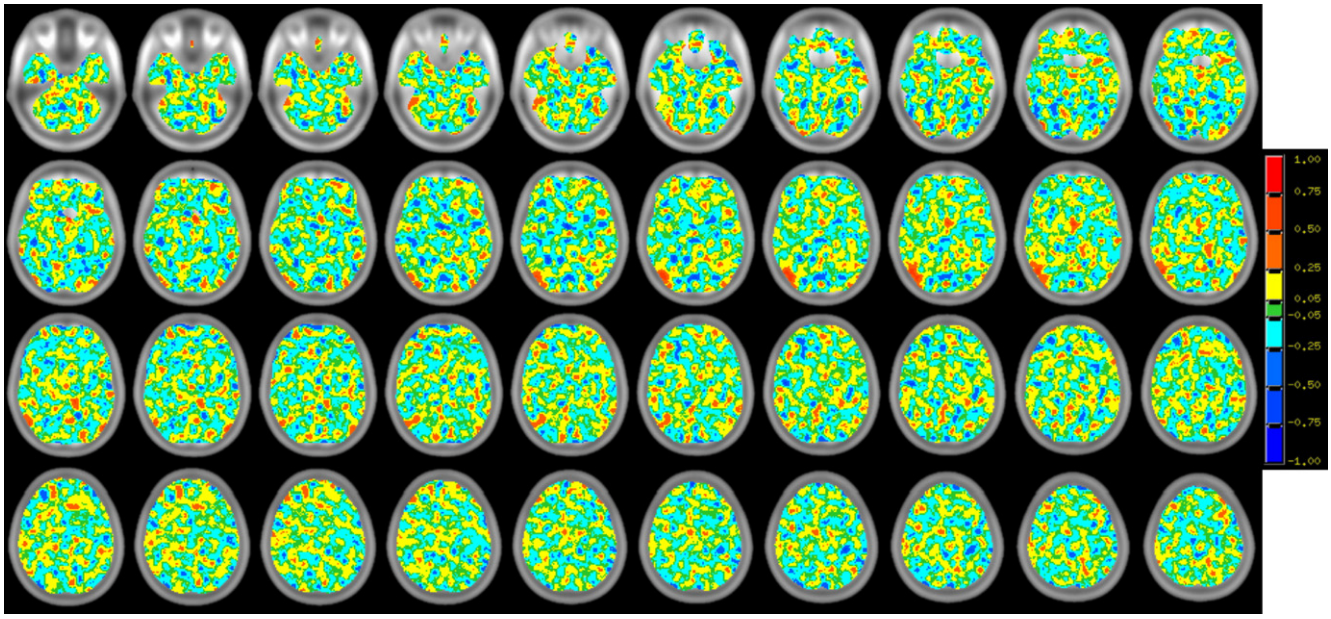

Fig. 3. The unthresholded weight values for the KCCA approach showing the contrast between viewing pleasant vs. unpleasant. We use the blue scale for negative (unpleasant) values and the red scale for the positive values (pleasant). The discrimination analysis on the training data was performed without labels. The class discrimination is automatically extracted from the analysis.

As the KCCA weights are not driven by simple categorical image descriptors (pleasant/unpleasant) but by complex image feature vectors it is of great interest that in the visual cortex, many regions, found by SVM are also highlighted by the KCCA. We interpret this similarity as indicating that many important components of the SIFT feature vector are associated with pleasant/unpleasant discrimination. Other features outside of the visual cortex are much less reproducible between SVM and KCCA indicating that many brain regions detect image differences not rooted in the major emotional salience of the images.

In order to validate the activity patterns found in Fig. 3, we show that the learnt semantic space can be used to correctly discriminate withheld (testing) fMRI volumes.

Table 5 shows the average and median performance of SVM and KCCA on the testing of pleasant and unpleasant fMRI blocks. Our proposed unsupervised approach had achieved an

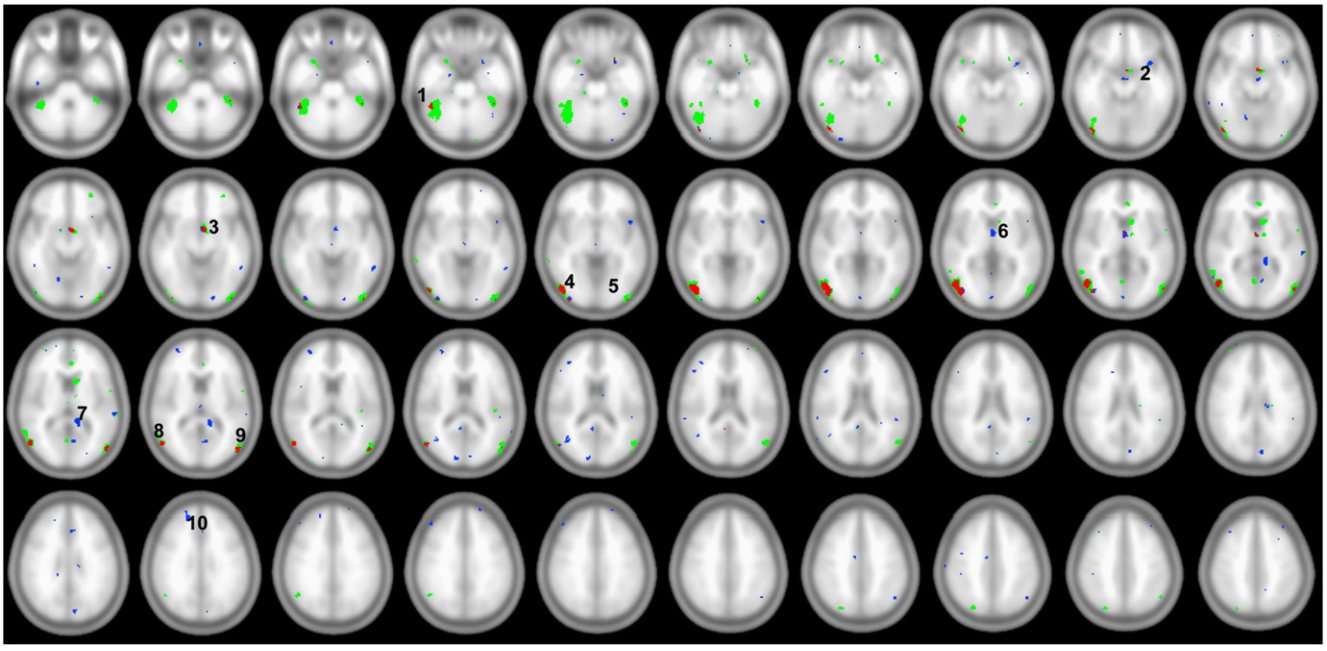

Fig. 4. Overlap of voxels with positive values above the threshold in the SVM weight and in the KCCA weight. Values above the threshold in both approaches are colored in red, voxels above the threshold only in the KCCA weight are colored in blue and voxels above the threshold only in the SVM are colored in green.

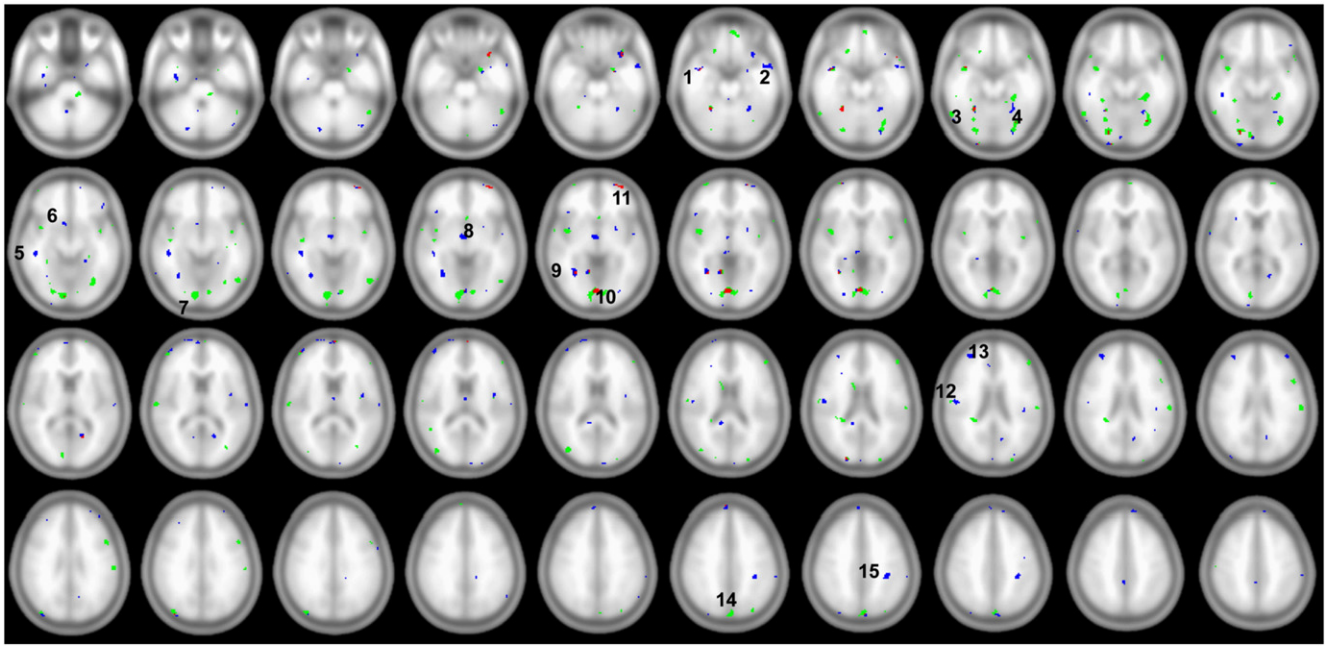

Fig. 5. Overlap of voxels with negative values above the threshold in the SVM weight and in the KCCA weight. Values above the threshold in both approaches are colored in red, voxels above the threshold only in the KCCA weight are colored in blue and voxels above the threshold only in the SVM are colored in green.

average accuracy of 87%, slightly less than the 91% of the SVM. Although, both methods had the same median accuracy of 92, the results demonstrate that the activity analysis is meaningful. To further confirm the validity of the methodology we repeat the experiments with the image stimuli randomised, hence breaking the relationship between fMRI volume and stimuli. Table 5 shows that the randomisation reduced the KCCA average performance to 49% and the SVM to 52% (both had a median result of 50%). This is equivalent to the performance of a random classifier.

In Table 6, we further validated the method with a leave-one-subject-out approach to verify the generalisation ability of the proposed method. These are repeated for all subjects and averaged across. The results demonstrate that it is possible to generalise to a subject-based approach.

Table 3
Clusters suprathreshold weights (positive values)

Cluster	Area	Talairach coordinates	Method
1	Right fusiform gyrus	45, −46, −26	SVM and KCCA
2	Inferior frontal gyrus	−34, 10, −16	KCCA
3	Nucleus accumbens	−6, 2, −10	SVM and KCCA
4	Right inferior occipital gyrus	45, −80, −4	SVM and KCCA
5	Left inferior occipital gyrus	−43, −88, −4	SVM
6	Fornix	−2, −4, 2	KCCA
7	Hippocampus gyrus	−11, −42, 6	SVM and KCCA
8	Right medial temporal gyrus	53, −70, 10	SVM and KCCA
9	Left medial occipital gyrus	−53, −78, 10	SVM and KCCA
10	Right superior frontal gyrus	18, 46, 30	KCCA

In Table 7, we present the number of overlapping and non-overlapping voxels between the methods. There are more non-overlapping than overlapping voxels. From Figs. 4 and 5, we can see that the most of the overlapping voxels are in visual area. The fact that we are detecting only 12% and 6.8% overlap is not surprising given that the process of SIFT vector construction is in no way dependent on emotional salience per se. The latter is a complex human interpretation of image features that is not incorporated in any way in the SIFT is computed. The SIFT

Table 4
Cluster suprathreshold weights (negative values)

Cluster	Area	Talairach coordinates	Method
1	Right medial temporal gyrus	40, 2, −22	SVM and KCCA
2	Left medial temporal gyrus	−53, 8, −22	KCCA
3	Right medial temporal gyrus	26, −52, −18	SVM
4	Left fusiform gyrus	−28, −50, −18	SVM
5	Right fusiform gyrus	47, −34, −12	KCCA
6	Anterior cingulate	6, 10, −12	KCCA
7	Right lingual gyrus	8, −88, −10	SVM
8	Hypothalamus	2, −8, −6	KCCA
9	Right fusiform gyrus	30, −56, −4	SVM and KCCA
10	Lingual gyrus	2, −80, −4	SVM and KCCA
11	Right medial frontal gyrus	30, 48, 22	SVM and KCCA
12	Right postcentral gyrus	53, −14, 22	KCCA
13	Right medial frontal gyrus	32, 44, 24	KCCA
14	Right parieto-occipital sulcus	0, −86,40	SVM
15	Right lateral parietal lobule	−34, −36, 42	KCCA

Table 5
KCCA and SVM results on the leave one out block experiment

Method	Average	Median
KCCA	0.87	0.92
SVM	0.91	0.92
Random KCCA	0.49	0.50
Random SVM	0.52	0.50

Average and median performance over 96 repeats. The value represents accuracy, hence higher is better.

process reveals that a small, but significant (in terms of overlap with visual SVM weight vector concentrations) set of image features in the SIFT are correlating with emotional salience. As SIFT is based on describing orientation and gradient information around prominent image features and there is no clear way that we can envisage a simple transformation between emotional content of an image and the SIFT type of representation. In view of this fact, there must be many elements of the SIFT vector that probably reflect image features that have no correlation whatsoever with emotional salience. It is also clear however that SIFT must contain some image features that, although not directly computed from emotional salience of images, must correlate emotional salience and thus that there are features of the pleasant/unpleasant images which correlate with gradient/orientation information around major image features. We thus regard the regions "detected" by KCCA as being derived from many features of the SFIT vector not associated or correlated with emotional salience and a smaller number that do have such associations/correlations.

Discussion

In this paper we present a novel unsupervised methodology for fMRI activity analysis in which a simple categorical description of a stimulus type is replaced by a more informative vector of stimulus (SIFT) features.

Previous studies investigated the pattern of response of the visual cortex to difference categories of stimuli (Haxby et al., 2001; O'Toole et al., 2005). They aimed to investigate modular versus distributed neural hypotheses. The modular hypothesis proposes that ventral temporal cortex contains a limited number of areas that are specialised categories of stimuli (e.g., Kanwisher et al. (1997)). In contrast, the distributed hypothesis proposes that the representations of different categories of visual stimuli (e.g., faces, objects, etc.) are widely distributed and overlapping (Haxby et al., 1999; Ishai et al., 2000). To address this issue (Haxby et al., 2001) measured the pattern of response with fMRI in six subjects while they viewed pictures of faces, cats, five categories of man-made objects and scrambled images. The data of each subject were split in two sets (namely even and odd runs) and the correlations between the patterns of fMRI responses within and between categories were computed. They found higher within-category correlations and using this procedure were able to predict the type of stimuli that the subjects were viewing at a level significantly better than chance. O'Toole et al. (2005) re-analysed the (Haxby et al., 2001) data set and found a high correlation between brain map and stimulus discriminability. The discriminability of brain maps and stimuli were computed independently. In each case, the discriminability was evaluated by projecting the data (e.g., brain maps or stimuli) onto PCA bases and training a linear discriminant (LD) using a subset of the PCs (i.e., the most accurate PCs). Our work differs from these previous works in a number of points. First, we used a non-supervised method (i.e., no explicit information about the stimulus category was provided to KCCA) aiming to find the correlations between the fMRI volumes and its corresponding stimulus features described by the SIFT feature vector. KCCA "finds" areas in the brain that are correlated with the features in the SIFT vector regardless of the stimulus category. Because many features of the stimuli were associated with the pleasant/unpleasant categories, we were able to use the KCCA results to classify the fMRI images between these categories. In the current study, it is difficult to address the issue of modular versus distributed neural coding as the complexity of the stimuli (and consequently of the SIFT vector) is very high.

The most interesting aspect of KCCA is its ability to extract visual regions very similar to those found to be important in categorical image classification using supervised SVM (red clusters in Figs. 4 and 5). According to the SVM results the most discriminating regions are in the visual cortex. However, there are also dissimilarities (blue and green clusters in Figs. 4 and 5). KCCA is able to extract visual processing features characteristic of the differences between the unpleasant/pleasant images blind to the existence of this categorical difference and to assign the same directional weights to these features (i.e., negative for the striate visual cortex, positive for the extrastriate visual cortex). We take this to indicate that the emotional saliences of the pleasant/unpleasant images are associated with basic visual features identified by KCCA. KCCA and SVM also show concordance in identifying other regions in the frontal (cluster 11 in Table 4 and Fig. 5) and temporal cortices (cluster 8 in Table 3 as well as the nucleus accumbens (cluster 6 in Table 3 and Fig. 4) that might reflect rewarding aspects of the pleasant images. The identification of these regions by KCCA as well as SVM suggests that the visual aspects of these emotional saliences contribute significantly to the feature set used by KCCA. There are however many differences between the regions given large weights by KCCA (blue clusters in Figs. 4 and 5) and those with high weights on the SVM pleasant/unpleasant discrimination (green clusters in Figs. 4 and 5). It can be speculated that they relate to processing of differences not

Table 6
KCCA and SVM results on the leave-one-subject-out experiment

Method	Average	Median
KCCA	0.79	0.79
SVM	0.84	0.86
Random KCCA	0.48	0.47
Random SVM	0.48	0.48

Average and median performance over 16 repeats. The value represents accuracy, hence higher is better.

Table 7
Overlap analysis

	Positive values	Negative values
Voxels above the threshold in SVM and KCCA (overlapping voxels)	402	177
Voxels above the threshold only in SVM (non-overlapping voxels)	1735	1379
Voxels above the threshold only in KCCA (non-overlapping voxels)	758	1033
Total of voxels used in the analyses	211,034	

determined by the basic stimulus categories (pleasant/unpleasant) but to other properties of the images. The answer to these questions will require association of particular brain regions with particular elements of the SIFT vector.

We use kernel canonical correlation analysis using an implicit representation of a complex state label to make use of the stimulus characteristics. The results are very promising. In terms of ability to correctly classify stimuli into different categories KCCA performs almost equivalently to SVM. We show that some of the regions detected by SVM with the largest weight vectors are also present in the KCCA maps. In the main, these are in the visual cortex. This is not unprecedented as there is a considerable number of papers showing modulation of responses in the fusiform gyrus and other visual regions by the strength of emotion represented in a stimulus (for a recent review, see Vuilleumier and Driver, 2007). Potentially, there is the possibility that there are visual areas that process fairly basic image features but are also modulated by emotion (Vuilleumier and Driver, 2007). It might thus be the case that there is not a clear boundary between areas that process basic visual data and areas modulated by emotion. Future studies should be done to clarify this. The current study has indicated that this might be an issue by classifying on two different dimensions (emotion and image properties) and showing overlaps. As two images cannot have different emotional content and be identical, then the possibility always exists that SIFT will include any differences that are present.

A further interesting possible application of KCCA relates to the detection of "inhomogeneities" in stimuli of a particular type (e.g., happy/sad/disgusting emotional stimuli). If KCCA analysis revealed brain regions strongly associated with substructure within a single stimulus category this could be valuable in testing whether a certain type of image was being consistently processed by the brain and designing stimuli for particular experiments.

There are many open-ended questions that have not been explored in our current research, which has primarily been focused on fMRI analysis and discrimination capacity. KCCA is a bi-directional technique and therefore are also able to compute a weight map for the stimuli from the learned semantic space. This capacity has the potential of greatly improving our understanding as to the link between fMRI analysis and stimuli by potentially telling us which image features were important. Finally, KCCA also has the potential of performing unsupervised multiactivity analysis, we have taken the eigenvector corresponding to the largest correlation value although remaining eigenvectors may correspond to further sub-tasks (if such exist in the stimulus).

Acknowledgments

This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. David R. Hardoon is supported by the EPSRC project Le Strum, EP-D063612-1. This publication only reflects the authors views. Janaina Mourão-Miranda and Michael Brammer thank Unilever plc (UK) for financial support for part of this project.

References

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE TMI 23, 137–152.

Boser, B.E., Guyon, I., Vapnik, V., 1992. A training algorithm for optimal margin classifiers. D. Proc. Fifth Ann. Workshop on Computational Learning Theory, ACM, pp. 144–152.

Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. 15, 704–717.

Ciuciu, P., Poline, J., Marrelec, G., Idier, J., Pallier, C., Benali, H., 2003. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. IEEE TMI 22, 1235–1251.

Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19, 261–270.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. NeuroImage 28, 663–668.

Faisan, S., Thoraval, L., Armspach, J.-P., Foucher, J.R., Metz-Lutz, M.-N., Heitz, F., 2005. Hidden Markov event sequence models: toward unsupervised functional MRI brain mapping. Acad. Radiol. (Acad. Radiol.) ISSN1076-6332 12, 25–36.

Friman, O., Carlsson, J., Lundberg, P., Borga, M., Knutsson, H., 2001. Detection of neural activity in functional MRI using canonical correlation analysis. Magn. Reson. Med. 45 (2), 323–330.

Friman, O., Borga, M., Lundberg, P., Knutsson, H., 2003. Adaptive analysis of fMRI data. NeuroImage 19, 837–845.

Fyfe, C., Lai, P.L., 2001. Kernel and nonlinear canonical correlation analysis. Int. J. Neural Syst. 10, 365–377.

Hardoon, D.R., 2006. Semantic Models for Machine Learning. PhD thesis, University of Southampton.

Hardoon, D.R., Shawe-Taylor, J., Friman, O., 2004a. KCCA for fMRI analysis. Proceedings of Medical Image Understanding and Analysis, London, UK.

Hardoon, D.R., Szedmak, S., Shawe-Taylor, J., 2004b. Canonical correlation analysis: an overview with application to learning methods. Neural Comput. 16, 2639–2664.

Hardoon, D.R., Saunders, C., Szedmak, S., Shawe-Taylor, J., 2006. A correlation approach for automatic image annotation. Springer LNAI, vol. 4093, pp. 681–692.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., Martin, A., 1999. The effect of face inversion on activity in human neural systems for face and object perception. Neuron 22, 189–199.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pitrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691.

Ishai, A., Ungerleider, L.G., Haxby, J.V., 2000. Distributed neural systems for the generation of visual images. Neuron 28, 979–990.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialised for face perception. Neuroscience 17 (11), 4302–4311.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. PANAS 103, 3863–3868.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. NeuroImage 26, 317–329.

Lowe, D., 1999. Object recognition from local scale-invariant features. Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra Greece, pp. 1150–1157.

Mikolajczyk, K., Schmid, C., 2001. Indexing based on scale invariant interest points. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hawaii USA, pp. 525–531.

Mikolajczyk, K., Schmid, C., 2002. An affine invariant interest point

detector. Proceedings of the 2002 European Conference on Computer Vision, Copenhagen Denmark, pp. 128–142.
Mikolajczyk, K., Schmid, C., 2003. Indexing based on scale invariant interest points. International Conference on Computer Vision and Pattern Recognition, pp. 257–263.
Mitchell, T., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., Just, M., Newman, S., 2004. Learning to decode cognitive states from brain images. Mach. Learn. 1–2, 145–175.
Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, S., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. NeuroImage 28, 980–995.
Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., Brammer, M., in press. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. NeuroImage 33 (4), 1055-1065.
Ngan, S.-C., Hu, X., 1999. Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity. Magn. Reson. Med. 41 (5), 939–946.
O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. 17 (4), 580–590.
Sebe, N., Tian, Q., Loupias, E., Lew, M., Huang, T., 2003. Evaluation of salient point techniques. Image Vis. Comput. 21, 1087–1095.
Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.
Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York.
Thirion, B., Faugeras, O., 2003. Dynamical components analysis of fMRI data through kernel PCA. NeuroImage 20 (1), 34–49.
Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.
Vuilleumier, P., Driver, J., 2007. Philos. Trans. R. Soc., B 362, 837–855.
Wang, X., Hutchinson, R., Mitchell, T.M., 2003. Training fMRI classifiers to detect cognitive states across multiple human subjects. Proceedings of the 2003 Conference on Neural Information Processing Systems.
Wismuller, A., Meyer-Base, A., Lange, O., Auer, D., Reiser, M.F., Sumners, D., 2004. Model-free functional MRI analysis based on unsupervised clustering. J. Biomed. Inform. 37, 10–18.