Using Image Stimuli to Drive fMRI Analysis

David R. Hardoon[1], Janaina Mourão-Miranda[2], Michael Brammer[2], and John Shawe-Taylor[1]

[1] {D.Hardoon,jst}@cs.ucl.ac.uk
The Centre for Computational Statistics and Machine Learning
Department of Computer Science
University College London
Gower St., London WC1E 6BT

[2] {Janaina.Mourao-Miranda,Michael.Brammer}@iop.kcl.ac.uk
Brain Image Analysis Unit
Centre for Neuroimaging Sciences (PO 89)
Institute of Psychiatry, De Crespigny Park
London SE5 8AF

Abstract. We introduce a new unsupervised fMRI analysis method based on Kernel Canonical Correlation Analysis which differs from the class of supervised learning methods that are increasingly being employed in fMRI data analysis. Whereas SVM associates properties of the imaging data with simple specific categorical labels, 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, 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 from this study is that KCCA in able in part to extract many of the brain regions that SVM identifies as the most important in task discrimination blind to the categorical task labels.

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

1 Introduction

Recently, machine learning methodologies have been increasingly used to analyse the relationship between stimulus categories and fMRI responses [1–10]. 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 technology of the present study originates from earlier research carried out in the domain of image annotation [11], 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 [12] and nonparametric estimation methods of the hemodynamic response function, such as the general method described in [13]. [14] 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 conventional unsupervised approaches in that it makes use of the stimulus characteristics as an *implicit* representation of a complex state label. We use kernel Canonical Correlation Analysis (KCCA) to learn the correlation between an fMRI volume and its corresponding stimulus. 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 first projects the data into a higher dimensional feature space before performing CCA in the new feature space. CCA [15, 16] and KCCA [17] have been used in previous fMRI analysis using only conventional categorical stimulus descriptions without exploring the possibility of using complex characteristics of the stimuli as the source for feature selection from the fMRI data.

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) [18] features. 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 utilize this property to probe aspects of brain function such as modularity.

In the current study, we present a feasibility study of the possibility of generating new activity maps by using the actual stimuli that had generated the fMRI volume. We have shown that KCCA is able to extract brain regions identified by supervised methods such as SVM in task discrimination and to achieve similar levels of accuracy and discuss some of the challenges in interpreting the results given the complex input feature vectors used by KCCA in place of categorical labels. This work is an extension of the work presented in [19].

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 2.1. The SVM is briefly described in Section 2.2 while Section 2.2 elaborates on the KCCA methodology. Our results in Section 3. We conclude with a discussion in Section 4.

2 Materials and Methods

Due to the lack of space we refer the reader to [10] for a detailed account of the subject, data acquisition and pre-processing applied to the data as well as to the experimental design.

2.1 Scale Invariant Feature Transformation

Scale Invariant Feature Transformation (SIFT) was introduced by [18] and shown to be superior to other descriptors [20]. This is due to the SIFT descriptors being designed to be invariant to small shifts in position of salient (i.e. prominent) regions. 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 maximize rotation invariance. Robustness to reorientation is introduced by representing local image regions around key voxels 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 utilize this property to probe aspects of brain function such as modularity.

Image Processing Let $\mathbf{f}_i^l$ 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 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 $\mathbf{o}_j$ where $j = 1, \dots, K$. The feature vector $\mathbf{x}_i$ of an image stimuli i is K dimensional with j'th component $\mathbf{x}_{i,j}$. The feature vectors is computed as the Gaussian measure of the minimal distance between the SIFT features $\mathbf{f}_i^l$ to the centre $\mathbf{o}_j$. This can be represented as

$$\mathbf{x}_{i,j} = \exp^{-\left(\min_{\mathbf{v}\in\mathbf{f}_i^l} d(\mathbf{v},\mathbf{o}_j)^2\right)} \tag{1}$$

where $d(.,.)$ is the Euclidean distance. The number of centres is set to be the smallest number of SIFT features computed (found to be 300). Therefore after processing each image, we will have a 300 dimensional feature vector representing its relative distance from the cluster centres.

2.2 Methods

Support Vector Machines Support vector machines[21] are kernel-based methods that find functions of the data that facilitate classification. They are derived from statistical learning theory [22] and have emerged as powerful tools for statistical pattern recognition [23]. In the linear formulation a 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 $(\mathbf{x}, y)$ where $\mathbf{x}$ represents a input and y it's 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. We used a linear kernel SVM that allows direct extraction of the weight vector as an image. A parameter C, that controls the trade-off between training errors and smoothness was fixed at $C = 1$ for all cases (default value).[3]

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 $x = \mathbf{w}_a'\mathbf{x}$ and $y = \mathbf{w}_b'\mathbf{y}$. Let $\mathbf{x}$ and $\mathbf{y}$ be two random variables from a multi-dimensional distribution, with zero mean. The maximisation of the correlation between x and y corresponds to solving $\max_{\mathbf{w}_a,\mathbf{w}_b} \rho = \mathbf{w}_a' C_{\mathbf{ab}} \mathbf{w}_b$ subject to $\mathbf{w}_a' C_{\mathbf{aa}} \mathbf{w}_a = \mathbf{w}_b' C_{\mathbf{bb}} \mathbf{w}_b = 1$. $C_{\mathbf{aa}}$ and $C_{\mathbf{bb}}$ are the non-singular within-set covariance matrices and $C_{\mathbf{ab}}$ is the between-sets covariance matrix.

We suggest using the kernel variant of CCA [24] 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 $\boldsymbol{\phi} : \mathbf{x} = (x_1, \ldots, x_n) \rightarrow \boldsymbol{\phi}(\mathbf{x}) = (\boldsymbol{\phi}_1(\mathbf{x}), \ldots, \boldsymbol{\phi}_N(\mathbf{x}))\ \ (N \geq n)$ before performing CCA in the new feature space. Given the kernel functions κ_a and κ_b let $K_{\boldsymbol{a}} = \mathbf{X}_a \mathbf{X}_a'$ and $K_{\boldsymbol{b}} = \mathbf{X}_b \mathbf{X}_b'$ be the kernel matrices corresponding to the two representations of the data, where $\mathbf{X}_a$ is the matrix whose rows are the vectors $\boldsymbol{\phi}_a(\mathbf{x}_i)$, $i = 1, \ldots, \ell$ from the first representation while $\mathbf{X}_b$ is the matrix with rows $\boldsymbol{\phi}_b(\mathbf{x}_i)$ from the second representation. The weights $\mathbf{w}_a$ and $\mathbf{w}_b$ can be expressed as a linear combination of the training examples $\mathbf{w}_a = \mathbf{X}_a \boldsymbol{\alpha}$ and $\mathbf{w}_b = \mathbf{X}_b \boldsymbol{\beta}$. Substituting into the primal CCA equation gives the optimisation $\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \rho = \boldsymbol{\alpha}' \mathbf{K_a} \mathbf{K_b} \boldsymbol{\beta}$ subject to $\boldsymbol{\alpha}' \mathbf{K_a^2} \boldsymbol{\alpha} = \boldsymbol{\beta}' \mathbf{K_b^2} \boldsymbol{\beta} = 1$. 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 [25, 26] suggests the need to regularise kernel CCA as it shows that the quality of the generalisation of the associated

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

pattern function is controlled by the sum of the squares of the weight vector norms. We refer the reader to [25, 26] 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 in both views, regularised CCA is the same as regularised 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.[4]

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 $\phi(\hat{x})$ be the new volume. We project the fMRI into the semantic space $\mathbf{w} = \mathbf{X}_a' \alpha_i$ (these are the training weights, similar to that of the SVM) and using the weights we are able to classify the new example as $\hat{w} = \phi(\hat{x})\mathbf{w}$ 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 and 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.

3 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. This amounted, per run, in 1330 training and 14 testing fMRI volumes, each set evenly split into positive and negative volumes (these pos/neg 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. Similarly, we run a further experiment of leave-subject-out basis where 15 subjects were combined for training and one left for testing. This gave a sum total of 1260 training and 84 testing fMRI volumes. The analyses was repeated 16 times. The KCCA regularisation parameter was found using 2-fold cross validation on the training data.

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

Initially we describe the fMRI activity analysis. 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 [10]. We are able to view the results in Figures 1 and 2 where in both figures the weights are not thresholded and show the contrast between viewing *Pleasant* vs. *Unpleasant*. The weight value of each voxel indicates the importance of the voxel in differentiating between the two brain states. In Figure 1 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. Figure 2 displays the KCCA weights.

In Figure 3 the unthresholded weights values for the KCCA approach with the hemodynamic function applied to the image stimuli (i.e. applied to the SIFT features prior to analysis) are displayed. The hemodynamic response function is the impulse response function which is used to model the delay and dispersion of hemodynamic responses to neuronal activation[27]. The application of the hemodynamic function to the images SIFT features allows for the reweighting of the image features according to the computed delay and dispersion model. We compute the hemodynamic function with the SPM2 toolbox with default parameter settings.

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 many regions, especially in the visual cortex, 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 in the frontal cortex are much less reproducible between SVM and KCCA indicting that many brain regions detect image differences not rooted in the major emotional salience of the images.

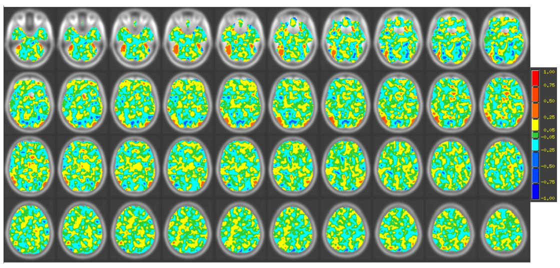

Fig. 1. The unthresholded weight values for the SVM approach showing the contrast between viewing *Pleasant* vs. *Unpleasant*. We use the blue scale for negative (*Unpleasant*) values and the red scale for the positive values (*Pleasant*). The discrimination analysis on the training data was performed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

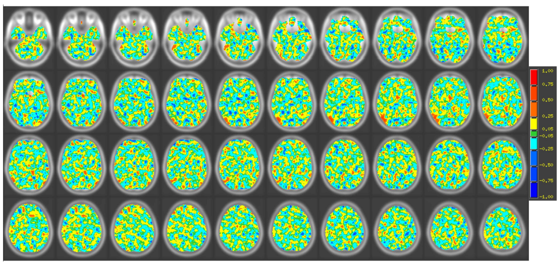

Fig. 2. The unthresholded weight values for the KCCA approach showing the contrast between viewing *Pleasant* vs. *Unpleasant*. We use the blue scale for negative (*Unpleasant*) values and the red scale for the positive values (*Pleasant*). The discrimination analysis on the training data was performed without labels. The class discrimination is automatically extracted from the analysis.

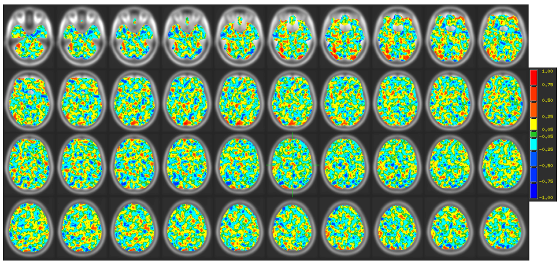

Fig. 3. The unthresholded weight values for the KCCA approach with the hemodynamic function applied to the image stimuli showing the contrast between viewing *Pleasant* vs. *Unpleasant*. We use the blue scale for negative (*Unpleasant*) values and the red scale for the positive values (*Pleasant*).

In order to validate the activity patterns found in Figure 2 we show that the learnt semantic space can be used to correctly discriminate withheld (testing) fMRI volumes. We also give the $2-$norm error to provide an indication as to the quality of the patterns found between the fMRI volumes and image stimuli from the testing set by $\|K_{\boldsymbol{a}}\boldsymbol{\alpha} - K_{\boldsymbol{b}}\boldsymbol{\beta}\|_2$ (normalised over the number of volumes and analyses repeats). The latter is especially important when the hemodynamic function has been applied to the image stimuli as straight forward discrimination is no longer possible to compare with.

Table 1 shows the average and median performance of SVM and KCCA on the testing of pleasant and unpleasant fMRI blocks for the leave-two-block-out experiment. Our proposed unsupervised approach had achieved an average accuracy of 87.28%, slightly less than the 91.52% of the SVM. Although, both

Table 1. KCCA & SVM results on the leave-two-block-out experiment. Average and median performance over 96 repeats. The value represents accuracy, hence higher is better. For norm-2 error lower is better.

Method	Average	Median	Average $\|\cdot\|_2$ error	Median $\|\cdot\|_2$ error
KCCA	87.28	92.86	0.0048	0.0048
SVM	91.52	92.86	-	-
Random KCCA	49.78	50.00	0.0103	0.0093
Random SVM	52.68	50.00	-	-
KCCA with HF	-	-	0.0032	0.0031
Random KCCA with HF	-	-	1.1049	0.9492

Table 2. KCCA & SVM results on the leave-one-subject-out experiment. Average and median performance over 16 repeats. The value represents accuracy, hence higher is better. For norm-2 error lower is better.

Method	Average	Median	Average $\|\cdot\|_2$ error	Median $\|\cdot\|_2$ error
KCCA	79.24	79.76	0.0025	0.0024
SVM	84.60	86.90	-	-
Random KCCA	48.51	47.62	0.0052	0.0044
Random SVM	48.88	48.21	-	-
KCCA with HF	-	-	0.0016	0.0015
Random KCCA with HF	-	-	0.5869	0.0210

methods had the same median accuracy of 92.86%. The results of the leave-subject-out experiment are given in Table 2, where our KCCA has achieved an average accuracy of 79.24% roughly 5% less than the supervised SVM method. In both tables the Hemodynamic Function is abbreviated as HF. We are able to observe in both tables that the quality of the patterns are better than random.

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 1 and 2 KCCA and SVM both show performance equivalent to the performance of a random classifier. It is also interesting to observe that when applying the hemodynamic function the random KCCA is substantially different, and worse than, the non random KCCA. Implying that the spurious correlations are found.

4 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. We use kernel canonical correlation analysis using an implicit representation of a complex state label to make use of the stimulus characteristics. 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. 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.

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.

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. We would like to thank Karl Friston for the constructive suggestions.

References

1. Cox, D.D., Savoy, R.L.: Functional magnetic resonance imaging (fmri) 'brain reading': detecting and classifying distributed patterns of fmri activity in human visual cortex. Neuroimage **19** (2003) 261–270
2. Carlson, T.A., Schrater, P., He, S.: Patterns of activity in the categorical representations of objects. Journal of Cognitive Neuroscience **15** (2003) 704–717
3. Wang, X., Hutchinson, R., Mitchell, T.M.: Training fmri classifiers to detect cognitive states across multiple human subjects. In: In Proceedings of the 2003 Conference on Neural Information Processing Systems. (2003)
4. Mitchell, T., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., Just, M., Newman, S.: Learning to decode cognitive states from brain images. Machine Learning **1-2** (2004) 145–175
5. LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X.: Support vector machines for temporal classification of block design fmri data. NeuroImage **26** (2005) 317–329
6. Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, S.: Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data. NeuroImage **28** (2005) 980–995
7. Haynes, J.D., Rees, G.: Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience **8** (2005) 686–691
8. Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., Langleben, D.D.: Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. NeuroImage **28** (2005) 663–668

9. Kriegeskorte, N., Goebel, R., Bandettini, P.: Information-based functional brain mapping. PANAS **103** (2006) 3863–3868
10. Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., Brammer, M.: The impact of temporal compression and space selection on svm analysis of single-subject and multi-subject fmri data. NeuroImage (accepted) (2006)
11. Hardoon, D.R., Saunders, C., Szedmak, S., Shawe-Taylor, J.: A correlation approach for automatic image annotation. In: In Springer LNAI 4093. (2006) 681–692
12. Wismuller, A., Meyer-Base, A., Lange, O., Auer, D., Reiser, M.F., Sumners, D.: Model-free functional mri analysis based on unsupervised clustering. Journal of Biomedical Informatics **37** (2004) 10 –18
13. Ciuciu, P., Poline, J., Marrelec, G., Idier, J., Pallier, C., Benali, H.: Unsupervised robust non-parametric estimation of the hemodynamic response function for any fmri experiment. IEEE TMI **22** (2003) 1235–1251
14. O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V.: Partially distributed representations of objects and faces in ventral temporal cortex. Journal of Cognitive Neuroscience **17(4)** (2005) 580–590
15. Friman, O., Borga, M., Lundberg, P., Knutsson, H.: Adaptive analysis of fMRI data. NeuroImage **19** (2003) 837–845
16. Friman, O., Carlsson, J., Lundberg, P., Borga, M., Knutsson, H.: Detection of neural activity in functional MRI using canonical correlation analysis. Magnetic Resonance in Medicine **45**(2) (2001) 323–330
17. Hardoon, D.R., Shawe-Taylor, J., Friman, O.: KCCA for fMRI Analysis. In: Proceedings of Medical Image Understanding and Analysis, London, UK (2004)
18. Lowe, D.: Object recognition from local scale-invariant features. In: Proceedings of the 7th IEEE International Conference on Computer vision, Kerkyra Greece (1999) 1150–1157
19. Hardoon, D.R., Mourao-Miranda, J., Brammer, M., Shawe-Taylor, J.: Unsupervised analysis of fmri data using kernel canonical correlation. NeuroImage **In Press** (2007) –
20. Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: International Conference on Computer Vision and Pattern Recognition. (2003) 257–263
21. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press (2000)
22. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag, New York (1995)
23. Boser, B.E., Guyon, I., Vapnik, V.: A training algorithm for optimal margin classifiers. In: In D. Proc. Fifth Ann. Workshop on Computational Learning Theory, ACM. (1992) 144–52
24. Fyfe, C., Lai, P.L.: Kernel and nonlinear canonical correlation analysis. International Journal of Neural Systems **10** (2001) 365–377
25. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: an overview with application to learning methods. Neural Computation **16** (2004) 2639–2664
26. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press (2004)
27. Stephan, K.E., Harrison, L.M., Penny, W.D., Friston, K.J.: Biophysical models of fmri responses. Current Opinion in Neurobiology **14** (2004) 629–635