UNIVERSITY OF SOUTHAMPTON

KCCA Feature Selection for fMRI Analysis

by

David R. Hardoon, John S. Shawe-Taylor & Ola Friman[1]

Technical Report

Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science
Image, Speech and Intelligent Systems Group

August 25, 2004

[1]Harvard Medical School, Harvard University, Boston, USA

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE
IMAGE, SPEECH AND INTELLIGENT SYSTEMS GROUP

by David R. Hardoon, John S. Shawe-Taylor & Ola Friman[1]

We use Kernel Canonical Correlation Analysis (KCCA) to infer brain activity in functional MRI by learning a semantic representation of fMRI brain scans and their associated activity signal. The semantic space provides a common representation and enables a comparison between the fMRI and the activity signal. We compare the approach against Canonical Correlation Analysis (CCA) and the more commonly used Ordinary Correlation Analysis (OCA) by localising "activity" on a simulated null data set. We also compare performance of the methods on the localisation of brain regions which control finger movement and regions that are involved in mental calculation. Finally we present an approach to reconstruct an activity signal from an "unknown" testing-set fMRI scans. This is used to validate the learnt semantics as non-trivial.

[1]Harvard Medical School, Harvard University, Boston, USA

Contents

List of Figures

List of Tables

Acknowledgements

The first & second authors would like to acknowledge the financial support of EU Project LAVA, No. IST-2001-34405.

%

Chapter 1

Introduction

Understanding the functional processes of the brain is still a new and difficult task. Functional Magnetic Resonance Imaging (fMRI) is a relatively new tool for the purpose of mapping the sensor, motor and cognitive tasks to specific regions in the brain. The underlying mechanics of this technique is in the regulation of the blood flow, as an excess of oxygen is supplied to active neurones via an increase in oxygenated blood surrounding the tissue of the active brain region. The oxygenation difference can be measured using an MR scanner which uses the different magnetic properties of oxygenated and deoxygenated blood. This effect is referred to as the Blood Oxygenation Level Dependent signal (BOLD). A more detailed description of the BOLD signal and its usage in MR scanners can be found in Friman (2003).

We present a Kernel-CCA (KCCA) approach to measure the active regions of the brain using fMRI scans and their activity signal. Friman et al. (2001b) have shown that CCA can gives us the ability to introduce several time-courses as the BOLD response has been shown to vary both between people and brain regions. In previous work (Hardoon and Shawe-Taylor, 2003; Vinokourov et al., 2003) we have shown that applying kernel methods (Cristianini and Shawe-Taylor, 2000) can increase the performance and computational power of CCA. Finally we show that due to the properties of KCCA (Shawe-Taylor and Cristianini, 2004) we are able to use this approach to validate the learnt semantics as non-trivial by reconstructing the activity signal from an "unknown" testing-set fMRI scans.

The paper is divided as follows; In section 2.1 we give an overview of the time frequency used to emulate the brain-task activity where in section 3.2 a background description of the theory is given. In section 4 we present three experiments which have been conducted comparing the baseline CCA and OCA approach to our KCCA method. We present the statistical reconstruction of a signal in section 5 and in section 6 we draw our final conclusions.

Chapter 2

fMRI

Functional magnetic resonance imaging or fMRI, is a relatively new imaging technique with the goal of mapping different sensor, motor and cognitive functions to specific regions in the brain. fMRI allows one to carry out specific non-invasive studies within a given subject while providing an important insight to the neural basis of masses of brain processes. Neurons, which are the basic functional unit of the brain, consume a higher level of oxygen when active. To achieve this, blood with a higher level of oxygenation is supplied. fMRI makes an indirect use of this effect to detect areas of the brain which have an elevated consumption of oxygen. This effect can be used to pin-point areas of the brain functions. In order to determine the elevation of oxygen consumption during a task, images acquired during a resting state are required. We obtain several slices from the brain as demonstrated in figure 2.1, each slice has a series of images which correspond to the time-sequence (figure 2.2). In order to keep the alternation between activity, a reference time-course is needed, where the resting and active states are embedded. A commonly used reference time-course is the square-wave time-course as plotted in figure 2.3 while in the following next subsection we present a more elaborate time-course which can take into account various delays that occur in the brain while performing a task. Although we only make use of the more complicated time-course for the experiments presented in section 4.2, in all other experiments we use the square-wave time-course (figure 2.3).

2.1 Time Frequency

The BOLD response evoked by a stimulus varies between different brain areas and between different test subjects (Aguirre et al., 1998; Glover, 1999). A common approach to the BOLD response is the usage of a square-wave signal, although in order to maximise detection sensitivity, such variations must be accounted for. For example, there is a delay between stimulus presentation and the onset of the BOLD response which lie in a range

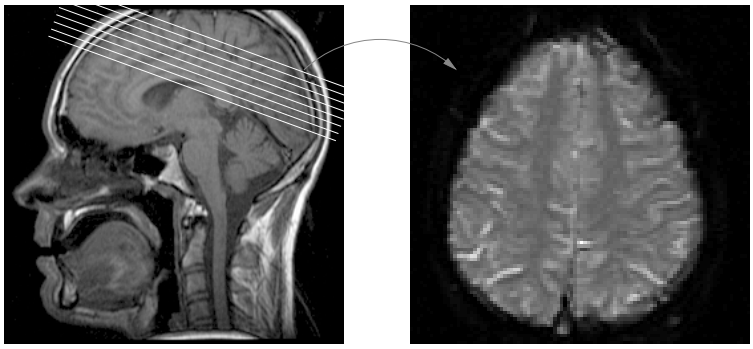

FIGURE 2.1: Extracted slices (from different depths) of the brain.

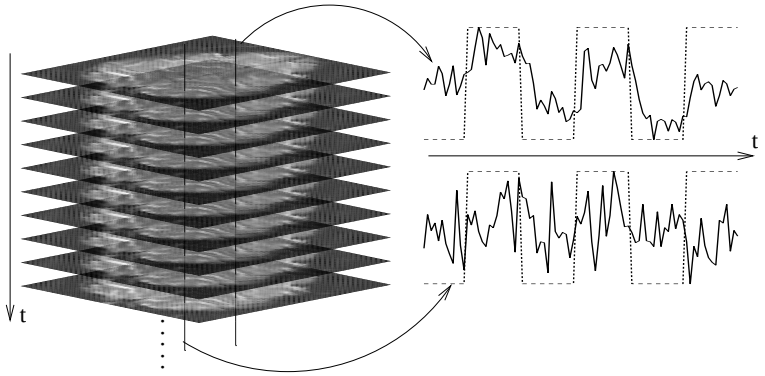

FIGURE 2.2: The time-sequence as an image sequence.

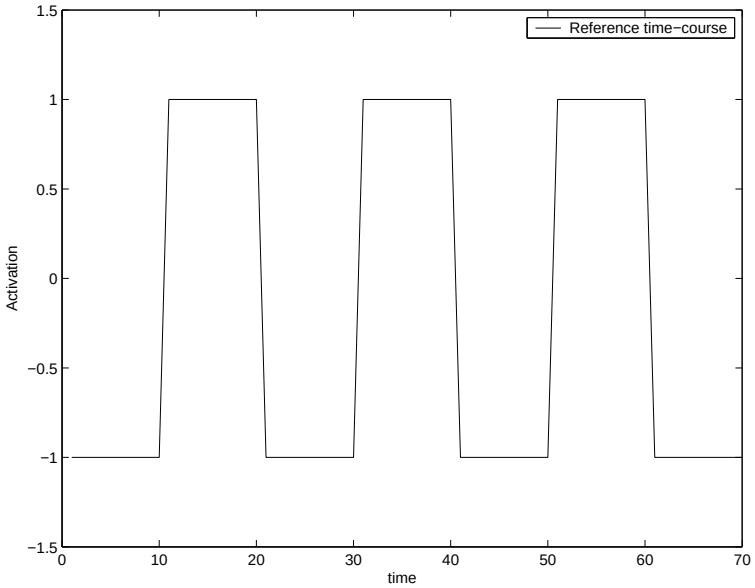

FIGURE 2.3: The commonly used square-wave reference time-course.

of several seconds (Saad et al., 2001). A computationally and theoretically tractable approach is to capture the BOLD response variations in a linear subspace spanned by a set of temporal basis functions. An example of such a subspace is the truncated Fourier subspace, which is spanned by a set of sine and cosine functions. These basis functions are especially useful when a blocked experimental design is utilized. The frequencies of the sine and cosine functions are then chosen to the fundamental frequency of the paradigm and a few harmonic frequencies. Within this subspace we find smooth models of the BOLD response in any phase position. Hence, we can find good matches to BOLD responses with any time delay.

Let

$$y(t) = \begin{pmatrix} sin(wt) \\ cost(wt) \\ \vdots \\ sin(kwt) \\ cos(kwt) \end{pmatrix} w = \frac{\pi}{T}, t = 1 \dots N$$

where T is the period of time-course reference, N is the number of observation (i.e. frames per slice) and $k = 1 \dots 4$.

Chapter 3

Theory

3.1 Ordinary Correlation Analysis

The most commonly used approach for fMRI analysis is known as ordinary correlation analysis. In this approach we assume N acquisitions of several brain slices. We define a model time-course $y(t)$, a possible choice is the square-wave which is used as a reference time-course. We obtain for each pixel a time-course of the different intensity values $x(t)$. Using $x(t)$ and $y(t)$ we compute the correlation using Pearson's correlation for zero mean random variables

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}}.$$

Therefore the sample correlation is calculated as

$$\rho = \frac{\sum_{t=1}^{N} x(t)y(t)}{\sqrt{\sum_{t=1}^{N} x(t)^2 \sum_{t=1}^{N} y(t)^2}}$$

this results in a correlation map of the fMRI scan. A drawback of this approach is that we are confined to model the BOLD response by a single time-course. The BOLD response has been shown to vary between brain regions and people, and may have delays. We would like to be able to model a BOLD response with any delay. We address this issue in the following section where we introduce multidimensional variables on both sides.

3.2 Canonical Correlation Analysis

Proposed by Hotelling (1936), Canonical correlation analysis can be seen as the problem of finding basis vectors for two sets of variables such that the correlation between the

projections of the variables onto these basis vectors are mutually maximised. Correlation analysis is dependent on the co-ordinate system in which the variables are described, so even if there is a very strong linear relationship between two sets of multidimensional variables, depending on the co-ordinate system used, this relationship might not be visible as a correlation. Canonical correlation analysis seeks a pair of linear transformations one for each of the sets of variables such that when the set of variables are transformed the corresponding co-ordinates are maximally correlated (Hardoon et al., 2003).

Definition 3.1. We use A' to denote the transpose of a vector or matrix A.

Let

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \rho = \frac{\mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y}} \tag{3.1}$$

the maximum canonical correlation is the maximum of ρ with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$.

We represent the two time-courses as a linear combination of pixel time-course

$$\mathbf{w}_{\mathbf{x}}'\mathbf{x}(\mathbf{t}) = w_{x_m} x_1(t) + \ldots + w_{x_m} x_m(t)$$

and basis functions, as described in the previous section 2.1

$$\mathbf{w}_{\mathbf{y}}'\mathbf{y}(\mathbf{t}) = w_{y_m} y_1(t) + \ldots + w_{y_m} y_m(t).$$

Observe that the covariance matrix of $(\mathbf{x}, \mathbf{y})$ is

$$C(\mathbf{x}, \mathbf{y}) = \hat{\mathbb{E}} \left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}' \right] = \begin{bmatrix} C_{\mathbf{xx}} & C_{\mathbf{xy}} \\ C_{\mathbf{yx}} & C_{\mathbf{yy}} \end{bmatrix} = C. \tag{3.2}$$

The total covariance matrix C is a block matrix where the within-sets covariance matrices are $C_{\mathbf{xx}}$ and $C_{\mathbf{yy}}$ and the between-sets covariance matrices are $C_{\mathbf{xy}} = C_{\mathbf{yx}}'$.

Figure 3.1 shows how CCA can be applied on the fMRI scans Friman et al. (2001a). We create a correlation mask by computing the correlation value for each 3×3 region and assign the correlation value to the middle pixel x_5. After computing the correlation mask we can threshold it to extract interesting pixels, though we could display the correlation mask without thresholding. The threshold can be computed using statistical approaches as described in Friman (2003).

Although there have been various methods of using feature selection techniques as an analysis tool for fMRI (McIntosh et al., 1996; Friston et al., 1995; Lange et al., 1999) and what seems to be a direct extension of CCA as described in Friston et al. (1995); Lange et al. (1999). We choose to extend CCA based on Kernel methodology in an attempt to

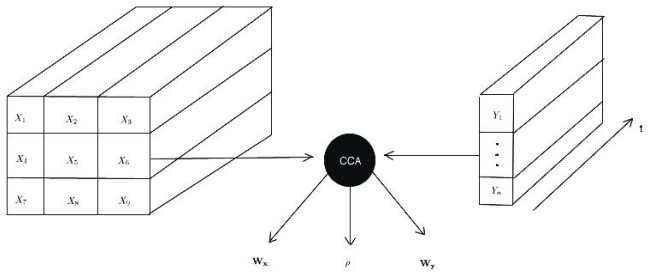

FIGURE 3.1: CCA between a set of fMRI time-courses and a region of pixels.

combine Machine Learning with fMRI analysis techniques. We describe the essence of Kernel Methods in the following section.

3.3 Kernel Methods

Kernel methodology is an approach to deal with data, which cannot be separated in a linear fashion and with algorithms that are restricted to handling linearly separable data. Kernels are a method of implicitly mapping data into a higher dimensional feature space, a method known as the "kernel trick". The implicit approach to the mapping implies that the functional form of the mapping does not need to be known. With suitable choice of kernel, non-separable data can become separable in the feature space, although we may map the data into an alternative dimensional space to obtain a better representation.

Definition 3.2. $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product of the vectors $\mathbf{x}$,$\mathbf{y}$ and is equal to $\mathbf{x}'\mathbf{y}$.

A kernel is a function K, such that for all $x, z \in \mathcal{X}$

$$K(x, z) = \langle \phi(x), \phi(z) \rangle \tag{3.3}$$

where ϕ is a mapping from $\mathcal{X}$ to a feature space F

$$\phi : \mathcal{X} \to F.$$

The transformation to feature space F allows us to deal with the problem using linear algebra. It lets us define a similarity measure from the dot product in F. The projection mapping also gives us the freedom to modify the mapping ϕ so as to change the representation of the input data into one that is more suitable for a given problem and learning algorithm. Kernels offer a great deal of flexibility, as they can be generated from other kernels. In the kernel the data only appears through entries in the Gram

matrix, therefore this approach gives a further advantage as the number of tuneable parameters and updating time does not depend on the number of attributes being used.

3.4 Kernel canonical correlation analysis

CCA may not extract useful descriptors of the data because of its linearity. Kernel CCA offers an alternative solution by first projecting the data into a higher dimensional feature space (where $n < N$)

$$\phi : \mathbf{x} = (\mathbf{x}_1, \ldots \mathbf{x}_n) \mapsto \phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_N(\mathbf{x}))$$

before performing CCA in the new feature space, essentially moving from the primal to the dual representation approach. Hence we obtain from CCA (for details Hardoon et al. (2003)) the dual representation

$$\rho = \max_{\alpha,\beta} \frac{\alpha' K_x K_y \beta}{\sqrt{\alpha' K_x^2 \alpha \beta' K_y^2 \beta}}. \tag{3.4}$$

In Hardoon et al. (2003) we observe that with full rank kernel matrices maximal correlation can be obtained, suggesting that learning is trivial. To force non-trivial learning we introduce a control on the flexibility of the projection mappings using Partial Least Squares (PLS) to penalise the norms of the associated weights. We convexly combine the PLS term with the KCCA term in the denominator (detailed description of CCA and KCCA can be found in Hardoon et al. (2003)), we obtain from equation 3.4

$$\rho = \max_{\alpha,\beta} \frac{\alpha' K_x K_y \beta}{\sqrt{(\alpha' K_x^2 \alpha + \kappa \alpha' K_x \alpha)(\beta' K_y^2 \beta + \kappa \beta' K_y \beta)}}. \tag{3.5}$$

Each fMRI frame is a $l \times l$ image matrix. We transform the matrix representation to a l^2 vector representation. Let X_i be each fMRI slice, of size $l^2 \times M$ where M is the number of frames in the slice (i.e. our number of samples). Let $\hat{X}_i$ be our sample in the feature space, we are able to compute a kernel per slice or we can combine all the slices such that let $\hat{X} = [\hat{X}_1 \ldots \hat{X}_i]$ and kernelise the global information of the brain and similarly let $\hat{Y}$ be the selected time sequence that corresponds to the fMRI frames. Therefore we can encapsulate all the information in the fMRI by kernelising $\hat{X}$ and similar with the time-courses $\hat{Y}$

$$\begin{aligned} K_x &= \left\langle \hat{X}, \hat{X} \right\rangle \\ K_y &= \left\langle \hat{Y}, \hat{Y} \right\rangle \end{aligned}$$

due to the kernel trick we obtain K_x and K_y are of size $M \times M$. Due to that α, β and the correlation values ρ are in the dual representation we need to "step back" to the primal weighting (Hardoon et al., 2003) such that we can recreate the original slice image with weights. Pixels that are associated with high correlation will have a high weight value while those associated with low correlation will have a corresponding low weight value. Let

$$\mathbf{w}_{x_i} = X_i \alpha \tag{3.6}$$

KCCA has a direct advantage over CCA in the ability of computing the correlation, activation weights values, on all slices simultaneously due to the usage of the kernel methodology, which represent the scans as a Gram matrix. As we are now transforming the fMRI data to the dual representation using the kernel trick we are not limited only to the usage of a linear kernel as used in our experiments. We are able to create tailored kernels to better extract the different image elements, though as we are interested in preserving the weight vector, we are limited to the usage of inner product kernels.

Chapter 4

Experiments

We compare the KCCA approach to the baseline CCA and OCA as presented in Friman et al. (2001a,b); Friman (2003) using three separate experiments, which are presented in the following subsections. We start by demonstrating the performance of KCCA on simulated data in subsection 4.1 as it is extremely hard to tell which method is better when using real fMRI analysis. In subsection 4.2 we present two tasks where the first was a flexing of the index finger of the right hand and the second was of a mental calculation. In these two tasks the time frequency used was as described in section 2.1. In all the figures the right side of the image is the left hemisphere and vice versa. This is the radiology convention of showing brain images (looking at the images from beneath or from the patient's feet). We compute the regularisation parameter κ from equation (3.5) for the experiments a priori as described in Hardoon and Shawe-Taylor (2003); Hardoon et al. (2003).

In the following experiments we use the correlation values computed by CCA and OCA. Although in KCCA we prefer to compute the weights associated with the pixels, as we hypothesise that this can give us more information. This step can also be done with CCA although we do not do so as this approach is not as intuitive and straightforward as with KCCA. In OCA this can not be done as there are no weights.

4.1 Simulated Data

As it is impossible to tell which method is better when real data is used, we first experiment with controlled simulated data. We embed square-wave "activity" in a null-data set (no brain activity), figure 4.1 shows the square-wave that was embedded. The paradigm of the applied activity is 10 images rest, 10 images activity and so forth, resulting with 200 time points. As the frame size was 100×100 we create a $100^2 \times 200$ matrix and compute the kernel as described in section 3.4. As we know the activation period we

use for our time sequence a square-wave representation of activity 1 and rest -1 over the 200 time-course. As we are using a simulated data with no delays in the time-course (we use the square-wave) the true potential of KCCA and CCA will not be apparent.

Figure 4.2 presents the found correlation values on the simulated data, we are able

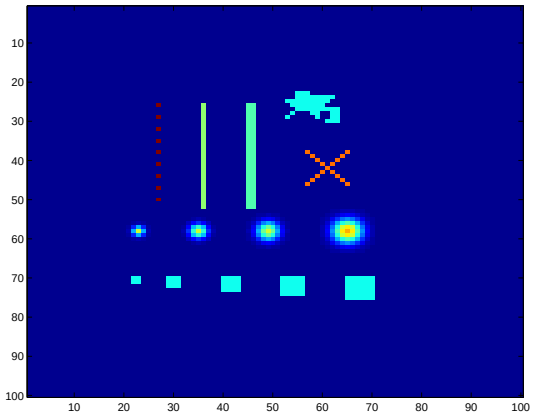

FIGURE 4.1: The square-wave "activity" embedded in the null brain scan.

to observe that the embedded activity was found. In figure 4.3 we view the correlation

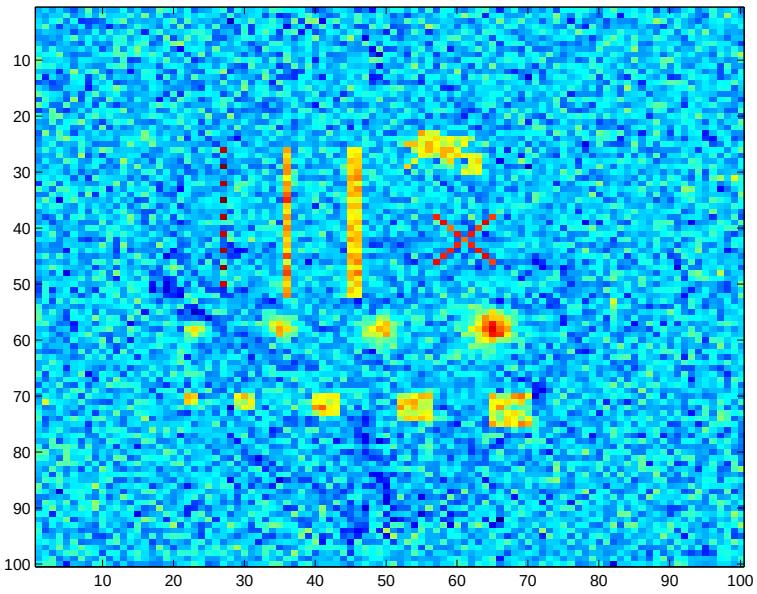

FIGURE 4.2: OCA correlation on the test brain scan

result from CCA. Active regions according to the embedded square-wave in figure 4.3(a) have been found, though we observe that the active pixels have a tendency to overflow to nearby pixels and that some objects get distorted from their original shape. This problem may be solved by thresholding. In figure 4.4 we view the KCCA found pixel weights. We find that our inital assumption that using the weights rather than the correlation is somewhat reassured when comparing the KCCA pixel weights to the OCA

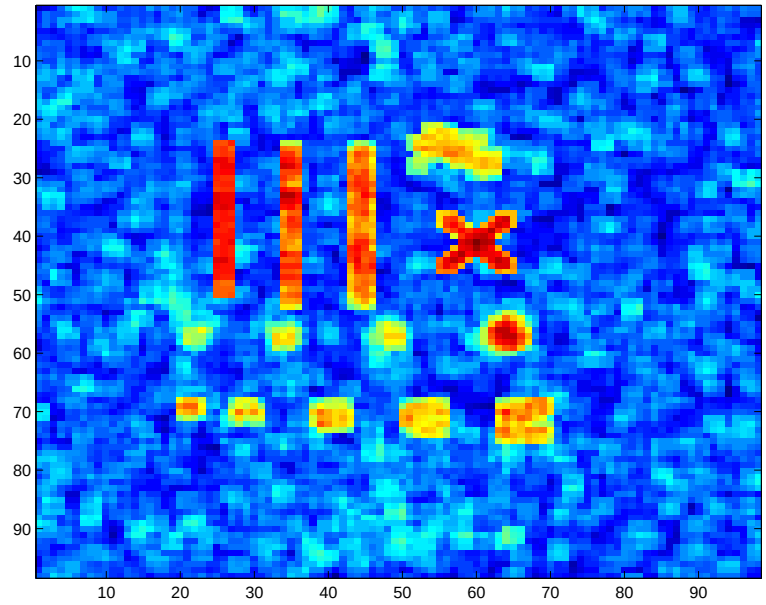

FIGURE 4.3: CCA correlation on the test brain scan

pixel correlation. The presented figures 4.2, 4.3 and 4.4 are to give us a visual result of

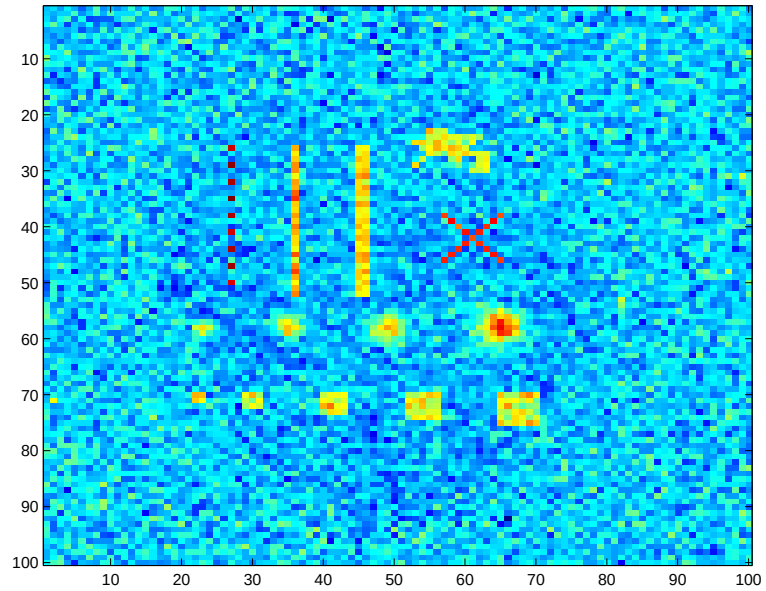

FIGURE 4.4: KCCA weights on the test brain scan

the different methods. As due to the fact that we use correlation in the baseline methods and weights in KCCA may imply that we are unable to compare them directly visibly. Therefore we attempt to statistically verify the advantage of KCCA over baseline methods by computing the true-positive voxels, the pixels which are found and really are active, and the false-positive voxels, the pixels which are found to be active but in fact are not. This comparison has to be done via thresholding of the image. We compute the

threshold values by taking all the positive pixel values from both images by decreasing order.

In the following figures we plot the true-positive and false-positive pixels found.

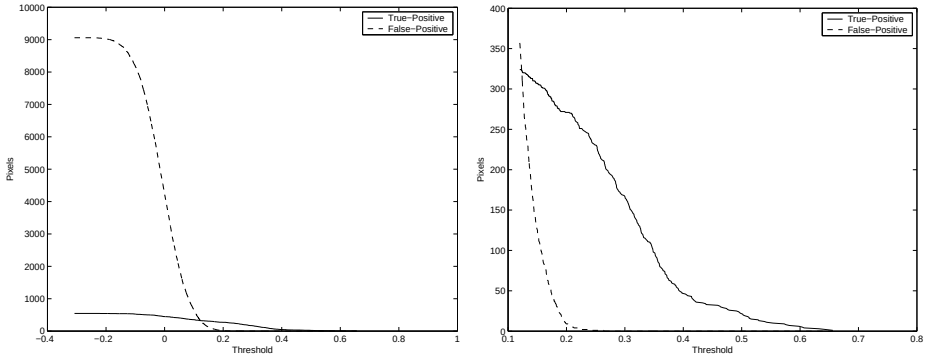

FIGURE 4.5: OCA Found True/False-Positives.

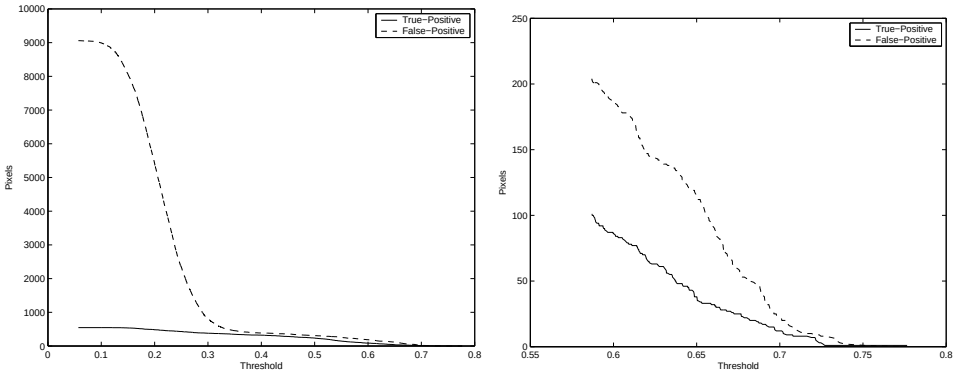

FIGURE 4.6: CCA Found True/False-Positives.

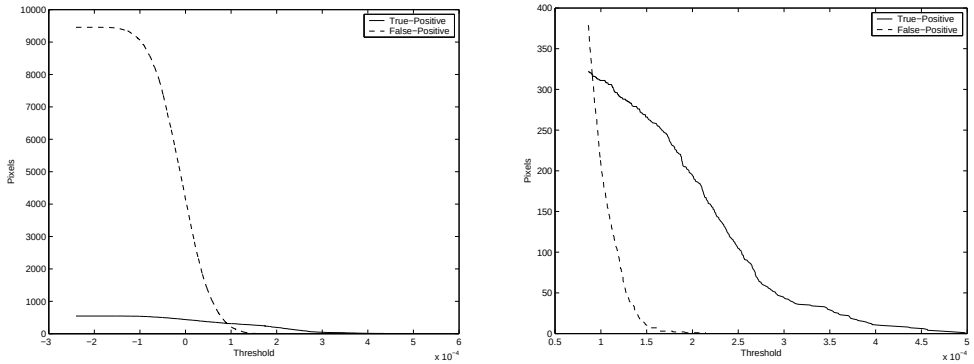

FIGURE 4.7: KCCA Found True/False-Positives.

In figure 4.5 we plot the results as found by OCA, in figure 4.6 by CCA and last in figure 4.7 the found true-positive and false positive pixels using KCCA are plotted. We

are able to observe that OCA and KCCA are very similar to each other where using a certain threshold, the number of false-positive pixels drops below that of the true-positive. This does not happen with CCA where the ratio of false-positive pixels are greater then true-positive.

Finally we plot a recall vs. precision of the positive pixels extracted by the different methods. Here we look at the true-positive pixel ratio and hence expect CCA to arise more successfully then other methods as the 'overflowing' will assure that all true-positive pixels are located. Although as observed in figure 4.6 this does not assure us that the false-positive will also be low, which as visible is not the case. We compute the recall and precision as follows

$$\begin{aligned} recall &= \frac{true - positive}{all - positive} \\ precision &= \frac{true - positive}{(true - positive) + (false - positive)} \end{aligned}$$

As we observe in figure 4.8 OCA and KCCA share the same pattern of behaviour

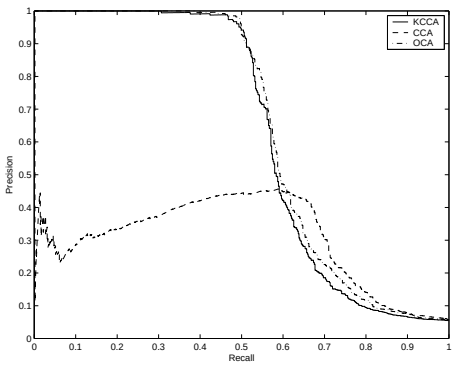

FIGURE 4.8: Recall vs. Precision.

where OCA is able to find a little more true-positive pixels then KCCA. CCA presents an interesting behaviour pattern which could be interpreted as a sharp drop in true-positive pixels to false-positive pixel followed with a steady line of true-positive found pixels while the false-positive pixels drop. This reaches a crossing point with OCA and KCCA, which also passes the other two methods. This informs us that the overall number of true-positive pixels found from this point are greater then those found in OCA and CCA, although as stated already this does not indicate the number of false-positive pixels found as well.

4.2 Finger Flexing & Mental Calculation

In the following two real experiments we present the visual results as the statistical verification, as done with the simulated data, is unavailable (we have no ground truth). It is important to emphasis that a direct comparison may not be adequate as KCCA uses weights while CCA and OCA use the actual correlation values. Although we feel that it may still be reasonable to present this as the weights represent the information given in the correlation (throughout the figures presented blue represents the negative side of the spectrum while red represents activity). As real fMRI is used in the experiments we use for the KCCA and CCA methods the time-course as a set of basis functions as described in subsection 2.1 and the regular square-wave time-course (figure 2.3) for the OCA method.

In the first experiment we attempt to find the active regions for a finger flexing task. As described in Friman et al. (2001a), a volunteer flexed his finger inside a MR scanner while image slices of the brain were acquired. The time-course is 10 frames rest, 10 frames flexing and so forth. Resulting in 200 acquisitions per depth slice, there were an overall of 12 slices.

Figure 4.9 presents the correlation values found using OCA. We are able to observe that although the expected active regions for the finger flexing task are highlighted, there are many regions highlighted as well, making the distinguishing between the active and non-active region difficult. In figure 4.10 the correlation image using CCA is presented where the image is clearer then that of OCA. The active region is highlighted in contrast to those that are not, making the separation clear. In Figure 4.11 the

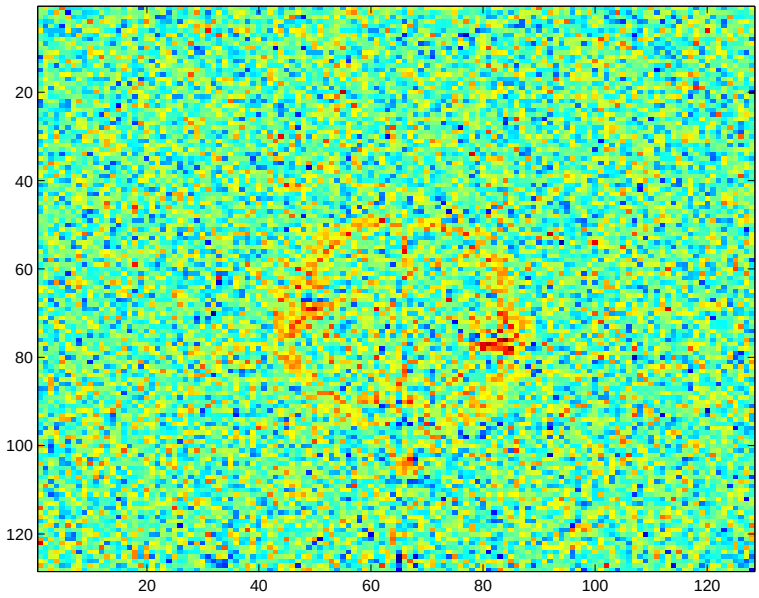

FIGURE 4.9: for slice 12.

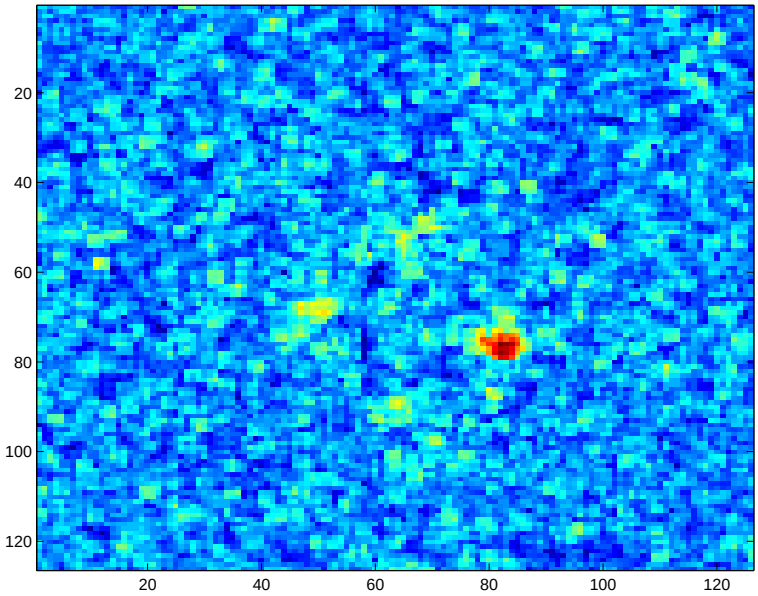

FIGURE 4.10: Finger flexing activity detection using CCA for slice 12.

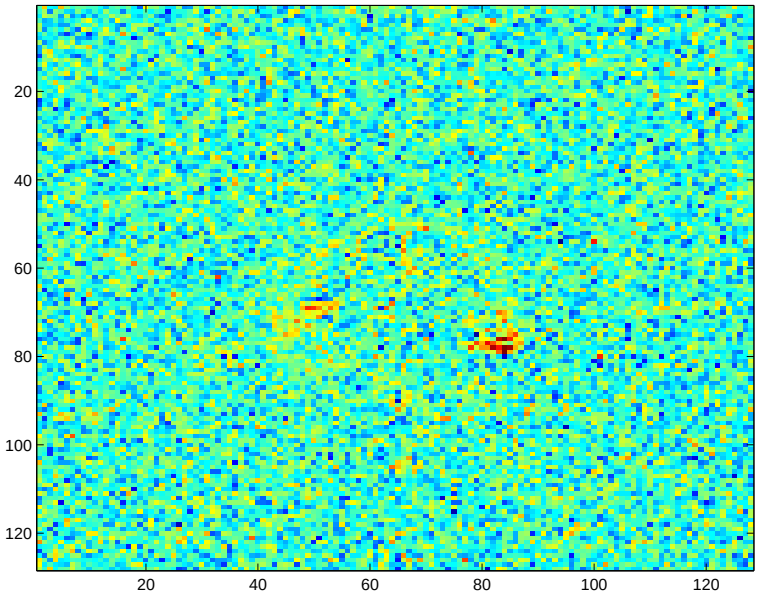

FIGURE 4.11: Finger flexing activity detection using KCCA for slice 12.

weighted image is displayed computed by KCCA. We view that the expected active region is also found with more information than that displayed in figure 4.10 though clearer than that displayed in figure 4.9.

In the second experiment, the task given to the volunteer was to compute the sum of two numbers that were projected onto a wall in the scanner room. The time-course is 15 frames rest, 15 frames performing the task and so forth, resulting in 180 acquisitions per slice. There are 12 slices. We attempt to infer the active brain regions during a

mental calculation task.

Figure 4.12 presents the results as obtained by OCA, as we observe no useful information can be extracted from the image. In Figure 4.13 we can view the CCA correlation mask of slice 12 for the mental calculation task, we find that in this image a clear separation between active and non-active regions are found. In these types of experiments neurological interpretation is hard to obtain. In Figure 4.14 we observe the active regions

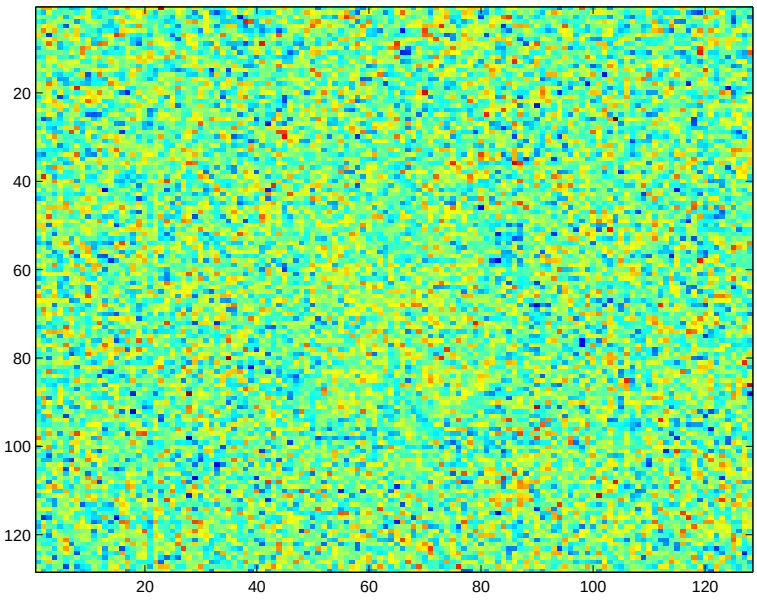

FIGURE 4.12: Mental task activity detection using OCA for slice 12.

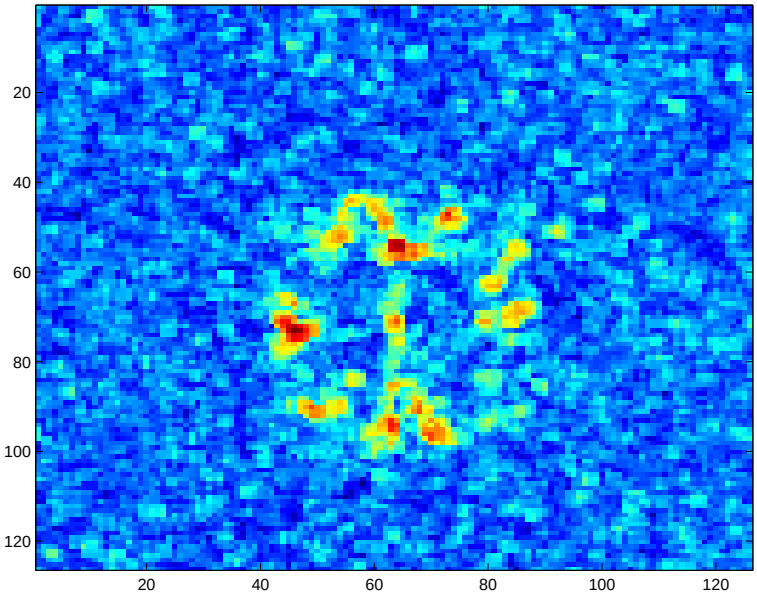

FIGURE 4.13: Mental task activity detection using CCA for slice 12.

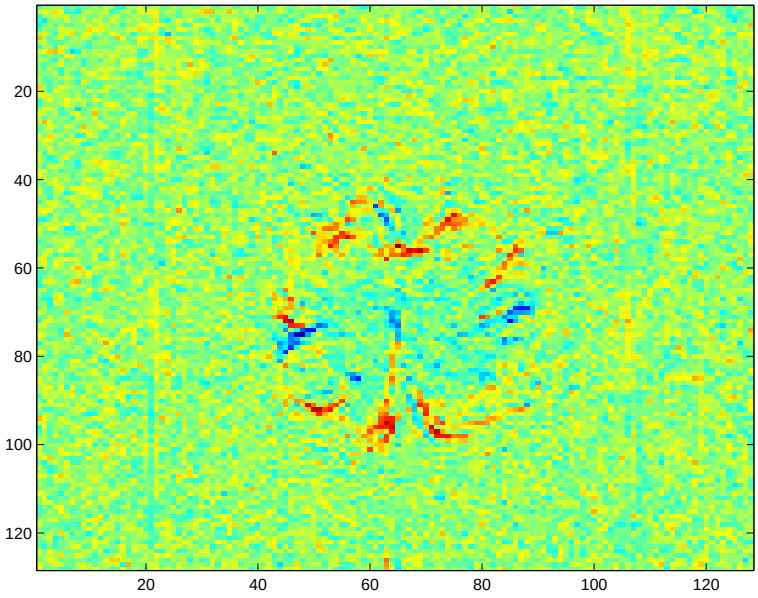

FIGURE 4.14: Mental task activity detection using KCCAfor slice 12.

for slice 12 as produced from KCCA. We find that the active regions located are similar to those located in figure 4.13 although $3-4$ regions are deep blue suggesting negative correlation. The negative correlation, or de-activation, found in the KCCA figure 4.14 is due to a higher activity during the rest periods than during the "active" periods. We believe that this is not located in CCA due to the small world view, analysing a section at a time, while in KCCA we apply to procedure on the whole brain.

Chapter 5

Signal Reconstruction

In the following section we present an approach of statistically reconstructing a signal from the fMRI scans. This reconstruction approach will allow us to determine the validity of our prior analysis for if we have learnt the appropriate function we will be able to reconstruct it. Let X be the fMRI training examples and X_t the fMRI testing examples, Y be the training activity time sequence and Y_t the testing activity signal we want to reconstruct. Let

$$g_{w_x,w_y} = \|Xw_x - Yw_y\|^2 \tag{5.1}$$

where $g_{w_x,w_y} \approx 0$ as we want the feature Xw_x from one view of the data to be identical to the feature Yw_y obtained from the second view of the data, this will be true on the training data if there is a high correlation between the two views. Therefore we can rewrite equation (5.1) as

$$\|Xw_x - Yw_y\|^2 \approx 0 \tag{5.2}$$

$$Xw_x \approx Yw_y \tag{5.3}$$

We divide the fMRI data into a training and testing set. On the training data we compute the KCCA coefficients α, β. Let $K_{x_t} = \langle X_t, X \rangle$ be the fMRI testing kernel and $K_{y_t} = \langle Y_t, Y \rangle$ be the time sequence testing kernel. Shawe-Taylor and Cristianini (2004) have shown that this equivalence can be held true also for the testing data using efficient regularisation. Hence we justify the usage of equation (5.3) and equation (3.6) to define

$$\begin{aligned} X_t w_x &\approx Y_t w_y \\ K_{x_t}\alpha &\approx Y_t' Y \beta. \end{aligned}$$

As we are interested in finding the testing-set unknown activity time sequence we can rearrange the equation to

$$Y_t \approx (K_{x_t}\alpha \cdot (Y\beta)^{-1})'. \tag{5.4}$$

As we are no longer interested in the weight vector but in the reconstruction of the signal, we are not confined to the usage of inner product kernels. In the following experiment we compare the success rate between the linear kernel as used, to the Gaussian kernel (defined in equation (5.5))

$$K(x_i, x_j) = \exp \frac{-\|x_i - x_j\|^2}{\sigma^2} \tag{5.5}$$

using σ as the minimum distance between the different labelled images.
We test our approach using the square-wave time sequence of 1 representing activity and -1 representing rest for both the simulated and real data experiments. The real data is comprised of mental calculation, the adding of two numbers, and right hand index finger flexing. For the simulated data and mental calculation we use the first 160 scans for training and the remaining 20 for testing, while with the finger flexing we use the first 180 for training and the remaining 20 for testing. We randomise the examples prior to the training and testing separation. Once we obtain the reconstructed Y_t we threshold it by $T = 0$, i.e. Let $\hat{Y}$ be the thresholded reconstructed signal

$$\hat{Y} = \begin{cases} 1 & if \quad Y_t \geq 0 \\ 0 & otherwise \end{cases}$$

Table 5.1 shows the average overall results of successfully reconstructing the activity time sequence for the testing fMRI data over 10 repeats using both a linear & Gaussian kernels. We are able to see that the linear kernel performs better then the Gaussian. It is important to state at this stage the difference between CCA and kernel CCA (KCCA) with a linear kernel. The former uses a larger number of features than in CCA, which are computed implicitly in the kernel.

TABLE 5.1: Success rate in reconstructing the test activity time sequence

Data-Set	Linear		Gaussian	
	Average	Standard Deviation	Average	Standard Deviation
Simulated data	96%	5.16%	93.5%	8.18%
Finger flexing	70.5%	10.12%	63.5%	8.83%
Mental calculation	51%	9.36%	44.5%	7.24%

We attempt to learn the relationship between the mental process prior to the motor process (finger flexing) by setting the square-wave sequence such that the three images before the actual finger flexing were considered as active and all the remaining images

were considered as inactive (see figure 5.1). We separate the data and train as before using only a linear kernel and attempt to reconstruct this new square-wave sequence over an average of 10 random repeats. We find that we can successfully reconstruct the signal with a success average of $68.5\% \pm 9.14\%$. This implies that the mental process prior to the actual motor process is sufficient to capture the functionality in the brain.

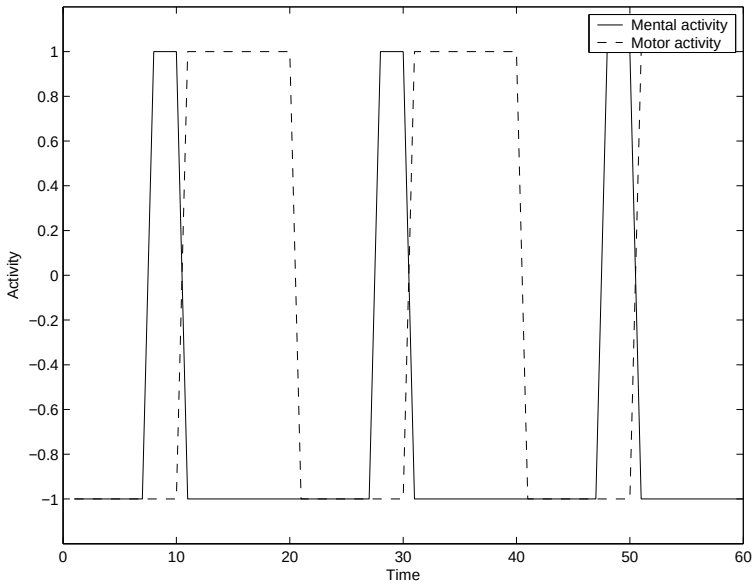

FIGURE 5.1: Activity plot of mental process prior to motor process.

Chapter 6

Conclusions

We have presented an extension to CCA analysis of fMRI based on Kernel Methods (KCCA). We show that KCCA has a computational advantage over CCA and OCA. The presented preliminary results show that KCCA can compute the activated regions with a finer precision even when using a simple linear kernel. We find that by using the found KCCA weights we return to the OCA original pixel by pixel analysis although unlike OCA where KCCA enables us to use new processing power within the kernel. Also as shown in the results OCA was not successful with real fMRI where noise is ramped in the fRMI while KCCA was able to distinguish active and non-active regions clearly.

We also present a methodology to reconstruct an activity signal from a fMRI scan that is of a similar activity process. This methodology uses the baseline concept that the two views used in KCCA are as close as possible. The proposed approach can be used as a validation technique for the original analysis. Further work on fMRI labelling using other machine learning methods is currently being completed in a different study.

For future work we would like to try more elaborate time basis functions and to experiment on different data types (emotional, mental and other motor fMRI data) and tailored kernels for better extracting the active regions in the brain. A further interesting avenue would be to observe the performance of applying our KCCA approach to other techniques of brain analysis and also to more complex tasks. We also speculate that KCCA would be able to handle a multiple task fMRI scenario (i.e. a scan with a few tasks at once) where baseline methods, such as CCA, require scans of individual tasks and can not handle a multiple task situation.

Bibliography

G.K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human, BOLD hemodynamic responses. *NeuroImage*, 8(4):360–369, 1998.

Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

O. Friman, M. Borga, P. Lundberg, and H. Knutsson. A correlation framework for functional MRI data analysis. In *Proceedings of the 12th Scandinavian Conference on Image Analysis*, Bergen, Norway, June 2001a. SCIA.

O. Friman, J. Carlsson, P. Lundberg, M. Borga, and H. Knutsson. Detection of neural activity in functional MRI using canonical correlation analysis. *Magnetic Resonance in Medicine*, 45(2):323–330, February 2001b.

Ola Friman. *Adaptive Analysis of Functional MRI Data*. PhD thesis, Linkoping Studies in Science and Technology, 2003.

K.J. Friston, C.D. Frith, R.S.J. Frackowiak, and R. Turner. Characterizing dynamic brain responses with fmri: A multivariate approach. *NeuroImage*, 2:166–172, 1995.

Colin Fyfe and Pei Ling Lai. Ica using kernel canonical correlation analysis, 2000.

G.H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4):416–429, 1999.

David R. Hardoon and John Shawe-Taylor. KCCA for different level precision in content-based image retrieval. In *Proceedings of Third International Workshop on Content-Based Multimedia Indexing*, IRISA, Rennes, France, 2003.

David R. Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis; an overview with application to learning methods. Technical Report CSD-TR-03-02, Royal Holloway University of London, 2003.

H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:312–377, 1936.

J. R. Ketterling. Canonical analysis of several sets of variables. *Biometrika*, 58:433–451, 1971.

N. Lange, S. Strother, J. Anderson, F. Nielsen, A. Hohnes, T. Kolenda, R. Savoy, and L. Hansen. Plurality and resemblance in fmri data analysis. *NeuroImage*, 10:280–303, 1999.

A. McIntosh, F. Bookstein, J. Haxby, and C. Grady. Spatial pattern analysis of functional brain images using partial least square, 1996.

A. Parry and P. M. Matthews. Function magnetic resonance imaging (fmri): A "window" into the brain, 2002.

Z. Saad, K. Ropella, R. Cox, and E. DeYoe. Analysis and use of FMRI response delays. *Human Brain Mapping*, 13(2):74–93, 2001.

John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

Alexei Vinokourov, David R. Hardoon, and John Shawe-Taylor. Learning the semantics of multimedia content with application to web image retrieval and classification. In *Proceedings of Fourth International Symposium on Independent Component Analysis and Blind Source Separation*, Nara, Japan, 2003.