The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Ji, Yongnan (2001) Data-driven fMRI data analysis based on parcellation. PhD thesis, University of Nottingham.

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/12645/1/546702.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see: http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Data-driven fMRI data analysis based on parcellation.

Yongnan Ji MSc.

Thesis submitted to The University of Nottingham

for the degree of Doctor of Philosophy

School of Computer Sciences

University of Nottingham

May 2011

GEORGE GREEN LIBRARY OF
SCIENCE AND ENGINEERING

Abstract

Functional Magnetic Resonance Imaging (fMRI) is one of the most popular neuroimaging methods for investigating the activity of the human brain during cognitive tasks. As with many other neuroimaging tools, the group analysis of fMRI data often requires a transformation of the individual datasets to a common stereotaxic space, where the different brains have a similar global shape and size. However, the local inaccuracy of this procedure gives rise to a series of issues including a lack of true anatomical correspondence and a loss of subject specific activations.

Inter-subject parcellation of fMRI data has been proposed as a means to alleviate these problems. Within this frame, the inter-subject correspondence is achieved by isolating homologous functional parcels across individuals, rather than by matching voxels coordinates within a stereotaxic space. However, the large majority of parcellation methods still suffer from a number of shortcomings owing to their dependence on a general linear model. Indeed, for all its appeal, a GLM-based parcellation approach introduces its own biases in the form of a priori knowledge about such matters as the shape of the Hemodynamic Response Function (HRF) and task-related signal changes.

In this thesis, we propose a model-free data-driven parcellation approach to single- and multi-subject parcellation. By modelling brain activation without an relying on an a priori model, parcellation is optimized for each individual subject. In order to establish correspondences of parcels across different subjects, we cast this problem as a multipartite graph partitioning task. Parcels are considered as the vertices of a weighted complete multipartite graph. Cross subject parcel matching becomes equivalent to partitioning this graph into disjoint cliques with one and only one parcel from each subject in each clique. In order to solve this NP-hard problem, we present three methods: the OBSA algorithm, a method with quadratic programming and an intuitive approach. We also introduce two quantitative measures of the quality of parcellation results.

We apply our framework to two fMRI data sets and show that both our single- and multi-subject parcellation techniques rival or outperform model-based methods in terms of parcellation accuracy.

Acknowledgements

I would like to thank my supervisor Dr. Alain Pitiot. Without his help I would not have had the opportunity to focus on this subject. I would also like to thank Professor Uwe Aickelin for his support and guidance. Additionally, I want to express my thanks to Dr. Pierre-Yves Hervé for his suggestions and advice.

Thanks also go to the committee of Collaborative Medical Image Analysis On Grid (CMIAG) group. I have been fortunate to receive funding from Marie Curie Action to complete the research. I also deeply appreciate the help I have received from the School of Computer Science, Brain & Body Centre and School of Psychology in the University of Nottingham.

Lastly but certainly not least, I would like to thank my family for their support over the years. I would especially like to thank my wife Xiaojie Song. Without her encouragement and support I would not have been able to concentrate and complete this thesis.

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Background and motivation

Due to the fact that neurons do not store energy, the firing of neurons leads to changes in both local blood flow and local deoxyhemoglobin content in the blood. The dynamic regulation of the blood flow in the brain is called hemodynamic. The use of magnetic resonance (MR) scans to measure hemodynamic responses provides a non-invasive approach to study the functions of the human brain.

The hemodynamic response corresponding to neural activity in the brain alters the contrast of T2* weighted magnetic resonance images (MRI) [Ogawa et al., 1990a,b; Turner et al., 1991]. This is called Blood Oxygen Level Dependence (BOLD). The precise nature of the relationship between neural activity and the BOLD signal is still a subject of research. However, in general, they are well related. Functional Magnetic Resonance Imaging (fMRI) uses BOLD signals as an indirect approach to the measurement of neural activity in the brain.

The use of fMRI to measure BOLD signals has provided neuroscientists with a powerful tool to examine brain activity. It has been widely used in various fields of neuroscience [Achard et al., 2006; Cohen et al., 2008; Simon et al., 2004]. As the acquisition of fMRI signals is complicated and BOLD signals have a very low signal-noise-ratio, preprocessing is an important step for fMRI data analysis.

According to different research aims, many methods have been proposed for fMRI data analysis. The General Linear Model (GLM) introduced by Friston et al. [1994] is one of the most popular model-driven methods in fMRI data analysis. In this regression method, a model is first set up to describe the BOLD signals corresponding to the stimulation, after which the model is applied to the data. Statistical analysis is performed on each voxel with the null hypothesis that the model does not match the time course on that voxel. Therefore, for each voxel, GLM provides a statistical measure (e.g. a t-value or an F-value) to present the possibility that the brain structure corresponding to that voxel is activated during the stimulation.

In GLM-based analysis, the shapes of the BOLD responses are presumed to be the same for all subjects and voxels. The variability of the BOLD responses is ignored. However, as shown by Aguirre et al. [1998] and Handwerker et al. [2004], the BOLD responses of the human brain vary across subjects, trials, days, and even different regions in the brain. Therefore, several methods have been proposed to overcome the influence of BOLD variability in model-driven study. For instance, Friston et al. [1999]; Woolrich et al. [2004] have proposed different basis sets for the Hemodynamic Response Function (HRF). The authors first define an HRF basis set, which could describe the reasonable HRF shapes. The BOLD signals from different subjects

can be modelled with different HRFs from the predefined basis set. These methods relax the assumption of BOLD signals from a fixed model to a model set.

Some other analysis methods do not require the assumption about the shapes of BOLD. For instance, Backfrieder et al. [1996] use Principal Component Analysis (PCA) for fMRI data analysis. With visual and motor stimulation experiments, they show that their method yields accurate absolute quantification of in vivo brain activity. Besides PCA, McIntosh et al. [2004] have proposed Partial Least Squares (PLS) as an effective multivariate analytic tool to identify brain activity patterns. In this work, they use event-related fMRI data to demonstrate that their method could provide robust statistical assessment without making assumptions about the shape of the HRFs.

Data-driven analysis is another type of method widely used in the area of fMRI data processing, in which brain activation is detected using only information contained in the fMRI signal itself. Techniques, such as Independent Component Analysis (ICA) [Beckmann and Smith, 2004; Li et al., 2007; Wang and Peterson, 2008] and clustering [Gao and Yee, 2003; Goutte et al., 1999], have been successfully used to extract the main components of responses from the fMRI time series. As for most data-driven techniques, the components of activation are extracted individually from each subject; the cross-subject variability of the BOLD signals does not influence the analysis results. Therefore, data-driven approaches can also be considered to be a means of solving the problem of cross-subject HRF variability.

In recent years, pattern-based classification analyses appear with increasing frequency in the functional neuroimaging area [Haynes et al., 2007; Kamitani and Tong,

2005, 2006; Mitchell et al., 2004]. These methods use machine learning algorithms to decode different mental states, behaviour and other variables from fMRI data. Compared with other methods, a machine learning classifier is complex to implement but, it makes a fundamental advance in the state of the art by linking patterns of brain activity to experiment design variables [O'Toole et al., 2007].

No matter which analysis approach is used, the study of the relationship between function and structure in the human brain relies on the analysis of groups of subjects. Therefore, voxel-based spatial normalization is also required for multi-subject analysis in order to bring fMRI images of different subjects into the same coordinate system, such as Talairach space [Talairach and Tournoux, 1988] and MNI space [Evans et al., 1993]. After spatial normalization, it is generally assumed that for all subjects registered to the standard space, the same coordinates correspond to the same brain structure. Further analysis can be applied in the standard space.

This method relies heavily on the assumption that for all spatially normalized subjects, the same coordinates in the standard space correspond to brain structure with the same function. However, even though many registration methods have been introduced [Brown, 1992; Zitova, 2003], due to the limitation of algorithms and the complexity of human brain anatomy, the problem of mis-registration still exists. For instance, in Figure 1.1, images from two subjects were registered to MNI space with rigid and affine registration. Both of these subjects were scanned under the same stimulation paradigm. Slice timing correction and motion correction were applied to the data. Then, the fMRI data was processed with GLM. The images in the red, green and blue dashed line rectangles are the images of the transverse, coronal and

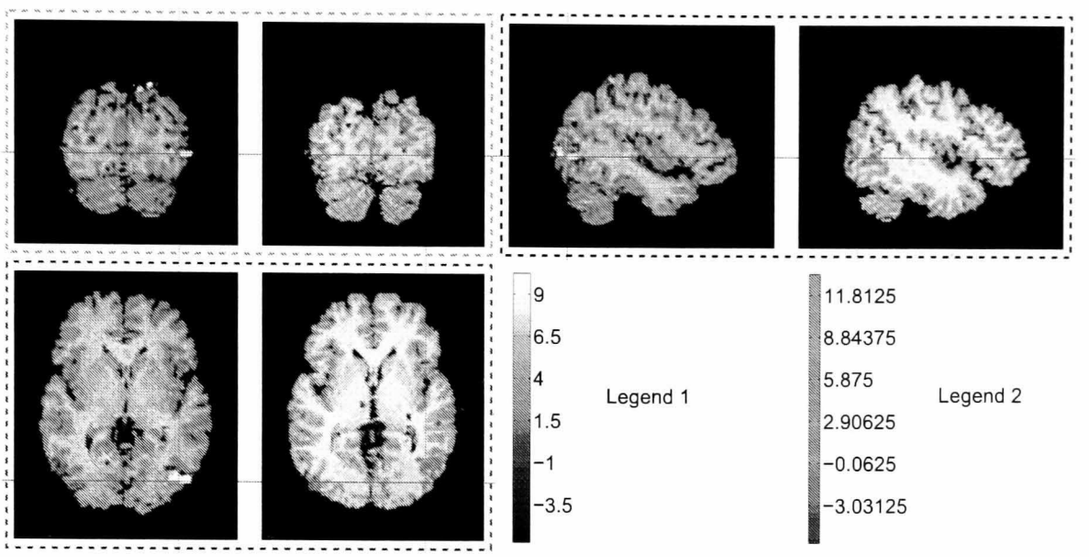


Figure 1.1: An example of mis-registration. The fMRI images of both subjects were scanned under the same stimulation paradigm. Activation with $p < 0.05$ is shown. The mis-registration problem is discussed in section 2.2.1. More details of the experiment are presented in Chapter 5

sagittal planes. In each dashed line rectangle there are two images. The left is from subject 1 and the right is from subject 2. The activation detected in subject 1 is presented on the red and yellow t-value map as shown in legend 1. The activation detected in subject 2 is presented on the blue and green t-value map as shown in legend 2. Comparing the activation maps from both subjects on the right occipital gyrus, the activation of the subject 1 is about 5mm posterior to the activation in subject 2. In these two subjects, the same coordinates in the standard space do not correspond to the same function. If the activation regions are small, this activation may be missed in a group analysis.

Parcellation could be used for group fMRI analysis to deal with the mis-registration

problem and overcome the limitation of spatial normalization. The parcellation of the human cerebral cortex into functionally distinct areas is an important area of neuroscience. Brodmann has parcellated the human brain into 52 different fields, based upon its cytoarchitecture. Using modern neuroimaging techniques, many methods have been proposed to parcellate the brain noninvasively[Peltier et al., 2009; Pohl, 2005; Shen et al., 2010].

Coulon et al. [2000] have proposed a method that uses hierarchical grey-level blobs to describe individual activation maps in terms of structures. A comparison graph is constructed based on these blobs for group analysis. This method can be considered as one of the earliest studies to use parcellation for the analysis of functional activation maps. Later, Flandin et al. [2002] presented parcellation as a way of dealing with the shortcomings of spatial normalization for model-driven analysis. They parcellate the brain of each subject into about 1000 functionally homogenous parcels with GLM parameters and group analysis is implemented on the parcels. However, this method is specifically designed for GLM analysis.

We consider that parcellation based analysis can be improved in the following two ways at least. (1) For individual subject parcellation, we need to overcome the variability of HRFs and provide the parcellation that is optimised for each individual subject. So that the parcellation accuracy could be increased. (2) We need to widen the scope of parcellation based analysis, so that data-driven analysis or machine learning classifiers can also be constructed on the parcels.

1.2 Aim and contribution

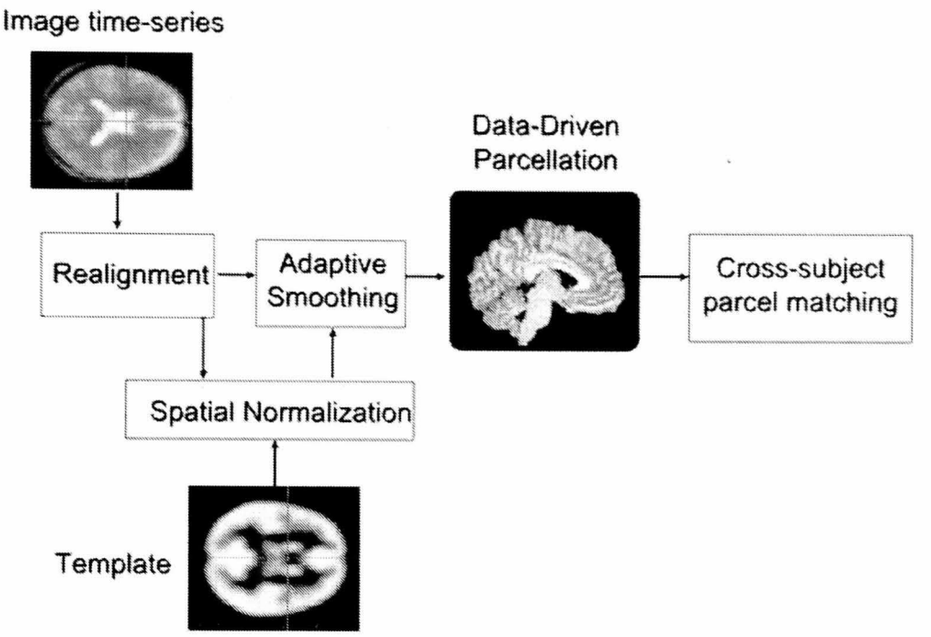


Figure 1.2: Data-driven parcellation framework.

The aim of this research is to develop a flexible fMRI data analysis framework based on parcellation. This framework should be able cope with the problem of mis-registration and HRF variability and can be used for data-driven analysis and machine learning based analysis.

Figure 1.2 shows this framework. In order to alleviate the issue of cross-subject parcel matching, images of all subjects are first aligned into standard space. After that a novel data-driven parcellation method based on adaptive smoothing for manifold embedding is applied to each subject [Ji et al., 2009]. Figure 1.3 shows the data flow of this method. Because no prior information on HRF is required in this approach, the cross-subject variability of HRF does not influence the parcellation results.

Next, the parcels from all subjects are matched for further analysis. Here, we try to

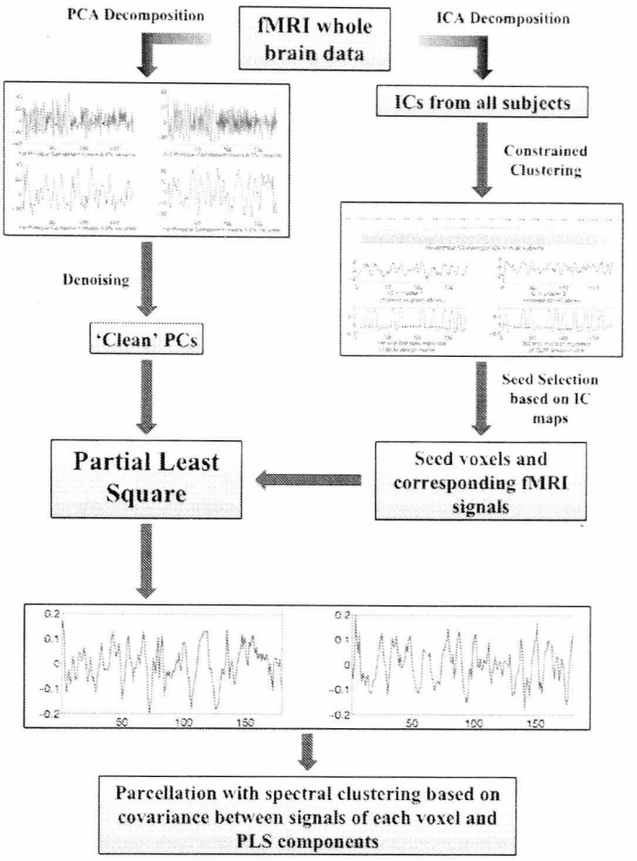


Figure 1.3: Single subject data-driven parcellation.

answer the question that, given only a suitable definition of the similarity between parcels from different subjects, is it possible to use the group information to find the best parcel correspondence? In order to answer this question, we formalize the problem of parcel matching as a multi-partite graph partitioning problem. Matching the parcels across all subjects is the same as partitioning a weighted graph into disjoint cliques by cutting some edges. The matching is optimized by minimizing the weights of the cut edges. We propose an order-based annealing method to solve this problem effectively and we discuss the similarity between the parcel matching problem and permutation invariant analysis [Jebara, 2003]. Therefore, in order to accelerate the optimization, we formalize the problem of parcel matching

into quadratic programming. We test the parcel matching algorithms with one toy dataset and one fMRI real dataset.

Hypothesis

The main hypothesis of this thesis is that our multi-subject data-driven parcellation approach improves over (1) standard voxel-wise fMRI analysis in terms of both robustness and sensitivity to normalization issues and (2) model-based parcellation techniques in terms of parcellation accuracy.

1.3 Overview of the thesis

Chapter 2 presents a review of the related work on fMRI data analysis and parcellation. We discuss previous work on model-driven and data-driven fMRI data analysis. Moreover, the problems of mis-registration and variability of HRF are addressed in this chapter.

Then, in Chapter 3, we first discuss spectral clustering and its application to parcellation. Next, the impact of noise on manifold embedding is discussed. Due to these factors, we suggest adaptive smoothing as a preprocessing step for parcellation with spectral clustering. Using one group of subjects as an example, the thesis shows the data structure of independent components from groups of subjects. Following that, combining independent component analysis and partial least square, we propose a novel, data-driven single-subject parcellation procedure. Finally, we proposed several methods to measure the parcellation quantitatively.

Chapter 4 describes how the cross-subject parcel matching problem could be considered as a graph partitioning problem. We compare three methods in order to partition multi-partite graphs effectively and efficiently.

In the next step, the data-driven framework proposed in this thesis is applied to two real fMRI data sets in Chapter 5.

Finally, in Chapter 6, we give the conclusion and discuss the directions of future work.

CHAPTER 2

Literature Review

In this chapter, we give a review of fMRI data analysis and human brain parcellation. First, we give an introduction to functional magnetic resonance imaging in section 2.1. In section 2.2, we review the state-of-art data analysis methods for fMRI. After that, in section 2.3, human brain parcellation is introduced in short. Finally, we give a summary in section 2.4.

2.1 Introduction to fMRI

Magnetic Resonance Imaging (MRI) is an imaging method that uses strong magnetic fields to create images of biological tissue [Huettel et al., 2004; Lauterbur, 1973]. During a MR scan, the subject is placed in a powerful static magnetic field to align the magnetization of some atoms in the body. To create an image, the scanner uses a series of changing magnetic gradients and oscillating electromagnetic fields, known as a pulse sequence, to systematically alter the alignment of this magnetization. This causes the nuclei to produce a magnetic signal detectable by the scanner. Accord-

ing to these signals the scanner can construct an image of the scanned area of the body. Using different pulse sequences, the scanner can provide images with different properties for a variety of research purposes [Bernstein et al., 2004].

Functional magnetic resonance imaging (fMRI) uses MR imaging to measure the metabolic changes in blood flow which is related to neural activities in the brain or spinal cord of humans or other animals. As neuron cells do not reserve energy, the energy consumed for neuronal activity is supplied by chemical reactions of glucose and oxygen. During this chemical action, oxygenated haemoglobin in the blood flow turns to deoxygenated haemoglobin. This transformation supplies the needed oxygen. Pauling and Coryell [1936] found that oxygenated haemoglobin and deoxygenated haemoglobin have different magnetic properties. Consequently, the magnetic resonance signal of blood flow is slightly different according to its level of oxygenation. Ogawa et al. [1990a,b] have demonstrated that the presence of deoxygenated blood decreases the measured MR signal on T2* images. The proportion of deoxygenated haemoglobin leads to the signal change on T2*-weighted images. Such a change is called blood-oxygenation-level dependent (BOLD) contrast.

Based on BOLD contrast, three groups published the first BOLD fMRI studies in 1992. Kwong et al. [1992] used 1.5 T MRI to study the activity in the human primary visual (V1) and motor (M1) cortex. During the scan, brain activation was evoked by visual simulation and hand squeezing. They found that in both areas the MR signal changes agree with the corresponding stimulation. About one month later, using 4T MRI, Ogawa et al. [1992] published a similar experiment to evaluate changes in gradient-echo signal resulting from longer visual stimuli. In addition, by using

different image-acquisition echo time, they further proved that the BOLD signal is produced by T2* effects. At almost the same time, Bandettini et al. [1992] studied a motor task in which subjects were instructed to touch each finger to thumb repetitively. They showed local signal increase of $4.3 \pm 0.3\%$ in the human primary motor cortex.

Since these studies, the BOLD-fMRI has been applied to researching into different brain functions in several ways in order to understand the workings of the human brain. The most popular topics are the task related fMRI studies. These studies attempt to find the patterns of brain activity associated with the mental processes of interest. In this type of experiment, during an fMRI scan, subjects are required to do certain tasks. These tasks are designed according to the research interests. Using fMRI data, one could construct statistical maps of task-dependent activation. For instance, Christensen et al. [2006]; Spalek and Thompson-Schill [2008] have studied the BOLD responses under visual and language tasks.

Besides task-related fMRI studies, some researchers are interested in using resting-state fMRI experiment to investigate the functional connectivity of the human brain [van den Heuvel and Pol, 2010]. Functional connectivity is defined as the temporal dependency between spatially remote neurophysiologic events [Friston et al., 1993; van den Heuvel and Pol, 2010]. During resting-state experiments, volunteers are instructed to relax and not to think of anything particular. Biswal et al. [1995, 1997] have demonstrates that, during rest, the left and right hemispheric regions of the primary motor network show a high correlation between their fMRI BOLD time series. Subsequently, many researchers have successfully shown the functional connectiv-

ity of other known functional networks, such as visual, auditory network and higher order cognitive networks [Achard et al., 2006; Bassett et al., 2006; Cordes et al., 2001, 2000; Fox and Raichle, 2007; Lowe et al., 2000]. These and subsequent studies have revealed new fundamental insights in the organization of the human brain.

In recent years, there has been growing interest in the use of machine learning for analyzing fMRI data. An increasing number of studies have shown that machine learning can be used to extract exciting new information from neuroimaging data [Norman et al., 2006; O'Toole et al., 2007; Pereira et al., 2009]. These studies cover a wide range of research topics, such as predicting conscious visual perceptions [Haynes and Rees, 2005a,b], decoding different mental states [Haynes et al., 2007; Mitchell et al., 2004; Mourao-Miranda et al., 2005], and classifying brain activity patterns for lie detection [Davatzikos et al., 2005]. Almost all of the techniques developed for pattern classification and data mining can be applied to fMRI data analysis. Therefore, researchers have been attracted to the use of machine-learning techniques to analyze fMRI data.

2.2 fMRI data analysis

Each volume of fMRI data can be considered as a three-dimensional matrix, whose elements are voxels $v(x, y, z)$. The volumes are sampled repeatedly over time. The whole fMRI data is a four-dimensional matrix with elements $v(x, y, z, t)$. For an fMRI experiment, a volume, which could have $64 \times 64 \times 32$ voxels, is sampled every 3 seconds (TR $= 3$) for about 100 time points. Ideally, $\forall t$, $v(x, y, z, t)$ corresponds

to the same location in the brain. However, almost all fMRI data suffers from distortion caused by subject head motion, physiological oscillations (e.g. heartbeats and respiration), inhomogeneities in the static field, and/or differences in the timing of image acquisition. Due to these distortions, preprocessing is necessary to reduce variability in the data that is unrelated to the experimental task. In this section, we will first introduce the preprocessing of fMRI data. After that, we will review the state-of-art fMRI data analysis methods.

2.2.1 Preprocessing of fMRI data

Slice-timing correction

Most fMRI data are acquired using two-dimensional pulse sequences to generate thin image planes (slices) [Huettel et al., 2004]. The number of slices required to cover the whole brain depends on the capabilities of the scanner. A typical scan, for instance the one used to generate the data sets in this thesis, needs 32 slices. These slices are acquired with equal spacing across the repetition time (TR), but in different orders.

Figure 2.1 illustrates an example volume with four slices. In order to avoid cross-slice excitation, most pulse sequences use interleaved slice acquisition, in which the odd slices are scanned first, followed by the even slices. For instance, in Figure 2.1, there are four slices in one volume, and each volume is scanned within TR = 3s. The four slices are acquired at $0.75s$ (red), $1.5s$ (green), $2.25s$ (blue) and $3s$ (yellow). However, in data analysis, it is commonly assumed that all these slices in this volume are

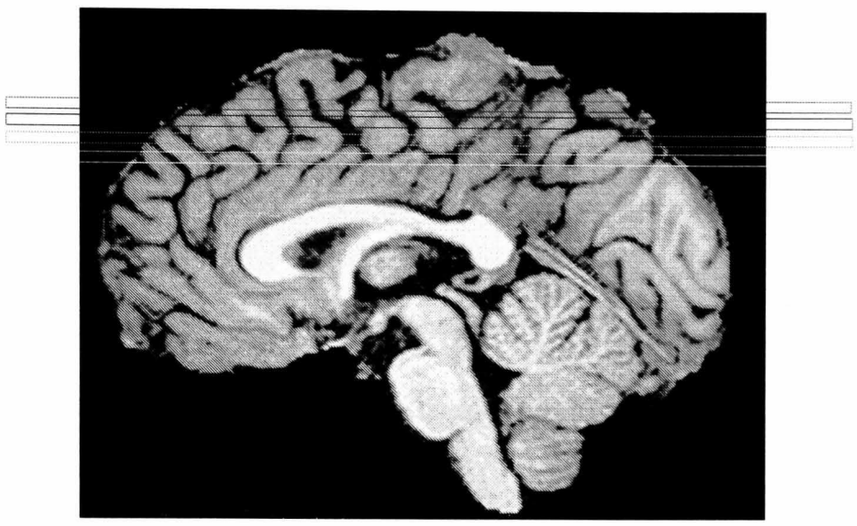


Figure 2.1: Slice-timing correction for fMRI data.

acquired at time 0*s*. Such difference in the timing of acquiring each slice is called the slice-timing problem. Henson et al. [1999]; Moortele et al. [1997] have described this slice-timing problem and have demonstrated its influence on the statistical analysis.

The most commonly used method to correct slice-timing errors is temporal interpolation. In this method, using the information from nearby time points, different interpolation techniques are used to estimate amplitude of the MR signal at the onset of the TR. Thus, for each volume, the intensity of any voxel in that volume is corrected to its intensity values at 0*s*. Although some researchers (e.g. Calhoun et al. [2000]) have proposed more advanced algorithms for slice-timing correction, no method could perfectly recover the missing information from samples. The accuracy of correction depends on the variability in the experimental data and the rate of sampling. Generally, when the variability is low or TR is short, accuracy is higher. For the fMRI data sets with typical temporal variability, slice timing correc-

tion is more effective for data acquired at relatively short TRs. For the data sets with longer TRs, slice timing correction could introduce errors. Therefore, this step could be skipped when the TR is long.

Motion Correction

In fMRI analyses, it is assumed that each voxel represents a fixed location of the brain. If the volunteer's head moves, each voxel's time course is derived from more than one brain location. Even small head motion may cause very large damage to raw signal over time. Despite the widespread use of head restraints during fMRI scans, it is hardly possible to keep the head perfectly still. The goal of motion correction is to adjust the time series of images so that $\forall t$, the voxels $v(x, y, z, t)$ in every image correspond to the same position in the brain.

Generally, the process of establishing spatial correspondences between two images is called coregistration. Let M and N be two image volumes. $\mathscr{F}$ denotes the spatial transformation that maps voxel coordinates in image M to the coordinates in image N. The coregistration between M and N can be described as an optimisation problem:

$$\widehat{\mathscr{F}} = \arg\max_{\mathscr{F}} \Big(sim(\mathscr{F}(M), N) + \lambda \cdot R(\mathscr{F}) \Big), \tag{2.2.1}$$

where $sim(\mathscr{F}(M), N)$ represents the similarity between the image N and the deformed image $\mathscr{F}(M)$. $R(\mathscr{F})$ is the regularisation on the deformation $\mathscr{F}$.

Many coregistration methods have been developed for different image modalities [Ashburner, 1999; Ashburner and Friston, 1999; Essen et al., 1998; Gee et al., 1997; Park et al., 2003]. In motion correction, the images of the time series are from the

same brain. Therefore, all the volumes in the time series are coregistered to a single reference volume with rigid-body transformation [Bannister, 2004; Frackowiak et al., 2004; Friston et al., 1996]. When using rigid-body transformations for coregistration of two images, it is assumed that the size and shape of the two objects are identical. By a combination of *translations* and *rotations*, one image can be superimposed exactly upon the other.

Here, translation is defined as the movement of the whole image volume along the axes. Let $\mathbf{m} = [x\ y\ z]'$ be a point in image volume M, where x, y, z are the coordinates in three-dimensional space. The transformation is:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \alpha_x \\ 0 & 1 & 0 & \alpha_y \\ 0 & 0 & 1 & \alpha_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

where $\mathbf{m}$ is translated α_x, α_y, α_z units along the axis x, y and z.

Rotation is defined as the turning of the entire image volume around the axes. The Rotation of θ_x radians around axis x is normally described by:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x & 0 \\ 0 & -\sin\theta_x & \cos\theta_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

Similarly, rotations around axis y and z can be implemented by the following matri-

ces:

$$\begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \cos\theta_z & \sin\theta_z & 0 & 0 \\ -\sin\theta_z & \cos\theta_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let $\Omega = \{\alpha_x, \alpha_y, \alpha_z, \theta_x, \theta_y, \theta_z\}$ be the set of parameters in translation and rotation. We denote the rigid-body transformation with parameter Ω on image volume M as $\mathscr{F}(M|\Omega)$. The realignment parameters are determined as:

$$\widehat{\Omega} = \arg\max_{\Omega} \Big(sim(\mathscr{F}(M|\Omega), N) \Big). \tag{2.2.2}$$

The sum of squared differences or mutual information can be used to measure similarity between the reference and corrected volume. As there is a large number of parameters in Ω, it is challenging to optimise equation 2.2.2. Thus, realignment algorithms use iterative approaches for head-motion correction. Gauss-Newton optimization is commonly used in rigid registration [Woods et al., 1998].

Spatial normalization

In fMRI analysis, it is sometimes desirable to analyze the functional data from a group of subjects. For instance, some experiments need to examine cross-subject consistency of results. Some researchers try to establish the difference in fMRI responses between healthy and diseased subjects. To analyze fMRI data across subjects, each subject must be transformed into a standard space so that it is the same size and shape as the others. This process is known as spatial normalisation, which is an important preprocessing step for most voxel-based fMRI studies

[Frackowiak et al., 2004; Huettel et al., 2004]. After registration into the standard space, it is generally assumed that the same Euclidean co-ordinates correspond to approximately the same brain region in all subjects. Although many brain atlases have been proposed [Collins et al., 1993; Dimitrova et al., 2006; Mazziotta et al., 1995], Talairach space [Talairach and Tournoux, 1988] and MNI space [Evans et al., 1993] are the most commonly adopted co-ordinate systems for spatial normalization.

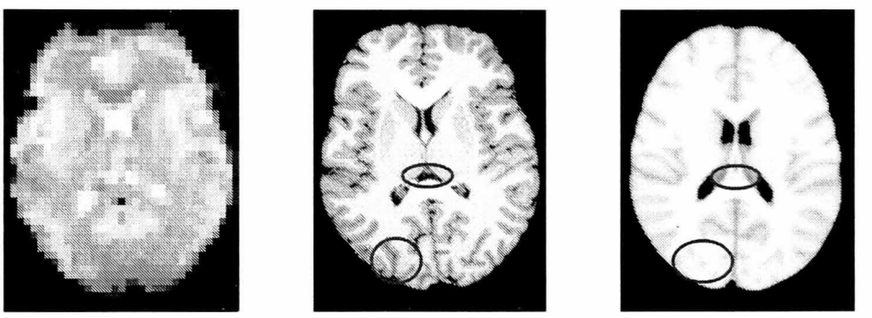


Figure 2.2: One slice of functional image, structural image and MNI atlas.

Figure 2.2 shows a slice of the functional image (left), the structural image (middle) and the MNI atlas(right). A typical functional image has a relatively low resolution. With this type of image, it is difficult to identify anatomical structures or boundaries and match them with the atlas. On the contrary, high-resolution structural images provide more details. Thus, it is common to acquire a structural image with an fMRI scan. The reference volume of the functional image is first mapped with the structural image using affine registration.

Affine transformations can be described as:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

let matrix **A** be

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Since most motions for medical imaging applications are reversible, invertibility is a natural requirement for image registration. An affine transformation is invertible if and only if the matrix **A** is invertible. The rigid body transformations introduced previously are a subset of affine transformations. As affine transformations are linear, they can only model the global geometric differences between images. However, as the functional and structural images are acquired with the same brain at almost the same time, affine registration is sufficient to align them with each other.

After registering the functional volumes onto the structural image, the structural image is normalised into a standard space. Then, the same transformations are applied to the functional volumes to bring them into the standard space.

Many different registration algorithms can be used to map the structural image with the standard space. However, as the standard space is generated as an average of hundreds of subjects, the image is generally lacking in detail. Some local information is lost. For instance, from the right image of Figure 2.2 we cannot determine the size of the ventricle or the boundary of the gyrus, which are marked with red circles.

Due to this limitation, global registration methods are commonly used for spatial normalization. And affine registration is one of the most popular and reliable global registration methods. There are a variety of affine registration toolboxes. With respect of the accuracy and computational requirement, [Zhilkin and Alexander, 2004] have compared the performance of several affine registration programs. The comparison includes FSL [FMRIB, 2007] and SPM [SPM8, 2009] which are the most commonly used toolboxes in fMRI data analysis.

Beside linear registration methods, there are also many local linear and non-linear registration methods that could provide more accurate mapping. Brown [1992]; Maintz and Viergever [1998]; Pluim et al. [2003]; Zitova [2003] give surveys of these techniques from different perspectives. Klein et al. [2009] gives a comprehensive evaluation of nonlinear deformation algorithms. More than 45,000 registrations between 80 manually labeled brains were performed with 14 nonlinear algorithms. And 8 different error measures are used to compare the performance to these algorithms. However, due to the large variability in brain features and the limitation of algorithms, after normalisation, the same coordinates may still correspond to different brain structure in different subjects. In fMRI data analysis, spatial smoothing is commonly used to deal with this problem by increasing the overlap between subject-specific activated regions. This approach may mask important cross-subject differences. Thirion et al. [2006] stated this problem and proposed to use parcellation to overcome this disadvantage of spatial normalization. However, this method considers only the model-based data analysis methods.

2.2.2 General Linear Model

The General Linear Model (GLM) is one of the most popular techniques for fMRI data analysis. Grinband et al. [2008] identified that in the first six months of 2007 alone, 170 papers published in leading journals used this approach.

In the GLM model, the observed data $\mathbf{y}_j$ from voxel $j, j = 1, 2, ..., J$ is modelled as a weighted combination of several explanatory variables $\mathbf{x}_n, n \in \{1, 2, ..., N\}$ plus an additive error term ϵ_j:

$$\mathbf{y}_j = \beta_{0j} + \beta_{1j}\mathbf{x}_1 + \beta_{2j}\mathbf{x}_2 + \cdots + \beta_{Nj}\mathbf{x}_N + \epsilon_j. \tag{2.2.3}$$

Here, the vectors $\mathbf{x}_n$ are the models to describe the hypothesised changes in BOLD activity, corresponding to the experiment process or other known sources of variability. β_{0j} is the parameter that reflects the total contribution of all constant factors. The parameters $\beta_{nj}, n \in \{1, 2, ..., N\}$ indicate how much each explanatory variable contributes to the data of voxel j. These parameters are calculated to minimise the error term. After that, in order to test the significance of a model factor ($\mathbf{x}_n$) for a given voxel (j), the corresponding parameter (β_{nj}) is divided by the residual error (ϵ_j). Statistical significance can be evaluated from this quantity.

Almost all major fMRI statistical analysis packages, such as SPM, FSL, AFNI Brain [Cox, 1996], include this model with the specific implementation dependent on the program. They share the same assumption that the observed fMRI data can be modelled as the sum of separate factors along with additive Gaussian noise. This assumption limits the performance and application of the GLM model.

A potential problem is that the GLM model requires an accurate estimate of the fMRI

signals corresponding to the performance of the task. However, for many reasons it is difficult to provide precise models. For instance, during the scan, the subjects may have been doing the task incorrectly. Even if the volunteers perform perfectly in the experiment, different subjects may still give different BOLD signals to the same stimuli. The same subject may also give different response signals at different time.

Another limitation is that for some experiments, it may be impossible to specify a model to describe the waveform of activated voxels. One example is the research on the *default mode* of brain function. Raichle et al. [2001] argue that there might be an organised mode of brain function. This mode is present as a baseline or default state, which is suspended during specific goal-directed behaviors. Following this work, the research on the default mode network becomes a very active topic. For this type of studies, the subjects are mostly in a resting state during the scan. We cannot provide an accurate model for the BOLD signals. Thus, it is not convenient to use the GLM model to describe the default mode network.

Another example is that, when the experimental process is too complicated, we do not know how BOLD signal will change corresponding to these tasks. Thus, it is impossible to use GLM for data analysis. For instance, Hasson et al. [2008] find the patterns of brain activation correlating with long-term memory formation during the viewing of extensive movie stimulus. Haynes et al. [2007] decode the mental states from fMRI data sets and found the brain regions that encode this information.

Beside the above studies, there are many other experiments, in which, researchers cannot determine the BOLD models. For these studies, it is not suitable to use GLM

to process the data. Under this situation, data-driven analyses and machine learning techniques provide complementary approaches.

2.2.3 Data-Driven analyses (DDA)

Data-Driven analyses explore the structure of the data under the assumption that with suitable approaches, the signals of interest (e.g. task-related activation or signals associated with default state network) have distinctive data structures. Based on that, many model-free methods have been successfully applied to fMRI data analysis. Clustering and Independent Component Analysis (ICA) are the most popular techniques in this area.

Clustering methods

Fuzzy C-means (FCM) is the most commonly used clustering method. It is also one of the first clustering methods to have been applied to fMRI data analysis. Baumgartner et al. [1997] and Moser et al. [1997] applied FCM to detect the activation in the human visual cortex. In this method, the time course of each voxel is considered as a vector in T-dimensional Euclidean space, where T is the number of time instances. The FCM analysis is performed directly in the time domain. The significant intensity changes are represented by different cluster centroids. Furthermore, they compared the performance of their method with three previous approaches from the perspectives of reproducibility and quantification. One problem of the conventional FCM is its sensitivity to noise and the clustering result is dependent upon the random initialization. In order to improve clustering results, Chuang et al.

[1999] proposed a method that combines Kohonen clustering network and FCM to increase the detection sensitivity and decrease the computation demand.

The above clustering studies directly use time courses as feature vectors. Thus, the feature vectors are in very high dimensional spaces (usually more than 100). Using such high dimensional feature vectors, clustering results would be less robust. Therefore, in later studies, different dimension reduction techniques are used before clustering. For instance, Liu et al. [2000] developed temporal clustering analysis, in which the three-dimensional brain was collapsed into a one-dimensional space. In this one-dimensional space, they could detect brain activity without a priori knowledge concerning when and where would be a response. However, this method can only detect the largest peak of the activation. Therefore, in their later work Gao and Yee [2003] improved their method and proposed iterative temporal clustering analysis for multiple response peaks.

Lange and Zeger [1997] applied a more commonly used dimension reduction method. They showed that the BOLD response to a periodic stimulus can be well characterised by Fourier coefficients. According to this discovery, Meyer and Chinrungrueng [2005] proposed a method for the clustering of fMRI time series in the spectral domain. In order to improve the detection of brain activity, this method explicitly takes into account the intrinsic spatiotemporal correlations of fMRI time courses. Later, Wang et al. [2005] proposed the use of Support Vector Clustering (SVC) for activation detection. This method could give high quality detection results without specifying the number of clusters. Afterwards, they extended SVC to ESVC (Ellipsoidal support vector clustering) in order to find the clusters that are more consistent with

the true data structure [Wang et al., 2007]. Although these Fourier transformation-based methods are limited to the experimental designs with periodic stimuli, they could be extended to analyse non-periodic fMRI data by replacing the spectral analysis with other feature extraction methods (e.g. wavelet analysis).

Besides activation detection, clustering methods are also widely applied to studying the default mode network (DMN) and brain connectivity detection. Cordes et al. [2002] applied a hierarchical clustering algorithm to find clusters whose voxel members have high cross correlation coefficients that represent a synchronous fMRI signal. One general problem of resting state fMRI analysis is that it is difficult to validate the DMN derived in a particular experiment. Bellec et al. [2010] proposed a framework called Bootstrap Analysis of Stable Clusters (BASC) to study the stability of resting-state networks in fMRI. In another interesting work, Mezer et al. [2009] used short time frequency analysis and clustering to study the spatial signal characteristics of resting state fMRI time series. In addition, they scanned non-functional T1-weighted time series and used them to examine the contribution of the non-functional fluctuation in BOLD signal. Using T1 image series as a baseline to study the fMRI image series is a new and interesting topic.

ICA analysis

Independent Component Analysis (ICA) is another popular and successful technique in data-driven fMRI analysis. Calhoun et al. [2003] gave a brief overview of the basic motivation and of several early works using ICA on fMRI data. A principal advantage of this approach is that it can be applied to experimental paradigms

in which models of brain activity are not available.

The basic assumption of the ICA method is that the observed signals are linear mixtures of hidden sources. In ICA these hidden sources are called independent components (ICs). ICs are non-Gaussian and statistically independent of each other. Using ICA algorithms, the independent components can be estimated from the observed data. Commonly used ICA algorithms include Infomax, FastICA and JADE.

Cardoso and Souloumiac [1993] have presented Joint Approximate Diagonalization of Eigenmatrices (JADE). This algorithm performs joint approximate diagonalization on fourth order cumulant matrices to archive spatial independence among sources. One problem of this algorithm is that it assumes the distributions of the unknown sources to be close to Gaussian. If the sources are non-Gaussian, the performance of JADE decreases rapidly. In addition, JADE requires a very complex and large amount of matrix computation. Thus, this algorithm has very large memory requirements. Consequently, JADE can be prohibitive when dealing with high dimensional data like fMRI time series.

Infomax is another way to estimate ICs. Bell and Sejnowski [1995] developed this approach which is based on entropy maximization in a feedforward neural network. This method is especially suited to separate sources that have higher kurtosis than the Gaussian distribution (super-Gaussian). They later extended their algorithm in Lee et al. [1999], so that it would be able to separate both sub-Gaussian and super-Gaussian sources. Due to the fact that Infomax is based on neural networks and gradient-based optimization technique, Infomax suffers from several typical problems. Firstly, the algorithm may converge to a local minimum of the contrast func-

tion and consequently obtain a sub-optimal estimation. Secondly, the convergence speed is much lower than other ICA techniques and the convergence is critically dependent on the correct choice of the learning rate parameters.

In order to overcome these problems, Hyvarinen and Oja [1997] have developed an efficient algorithm called FastICA. This approach uses fixed-point iteration scheme to maximise the non-Gaussianity of the estimated sources. Compared to gradient-based methods, the fixed-point iteration technique converges much faster. Contrary to JADE, FastICA is computationally simple and requires little memory space. However, the problem of the sub-optimal results still exists.

All of the above algorithms have been applied to fMRI data analysis. With a simulated data set and an event related audio-visual task data, Ghasemi and Mahloojifar [2010] compared these three algorithms from the perspectives of robustness and reliability. They conclude that Infomax emerged as a more reliable choice for extracting task-related activation maps and time-courses from fMRI data sets. JADE and FastICA gave a similar performance. However, in terms of convergence Infomax was the slowest. Although, this comparison is based on very limited experiments (e.g. only one set of data, only one type of fMRI experiment), it gives a general clue as to how to choose the algorithm.

In fMRI data analysis, ICA can be applied in two approaches: spatial ICA (SICA) and temporal ICA (TICA) [Calhoun et al., 2001b]. In SICA, it is assumed that each fMRI image volume is a mixture of spatially independent components and each independent component is an image volume. On the other hand, TICA considers that the temporal signal of each voxel is a mixture of temporally independent time

courses. No direct and thorough comparison between these two approaches has appeared in the literature. However, most applications on fMRI data analysis use the SICA.

In ICA analysis there are several general problems. Firstly, the number of ICs is a free parameter. ICA does not naturally estimate the number of hidden sources. Usually, it is either empirically determined or estimated with other methods. Secondly, the ICA decomposition result is not unique. For the same set of data, different runs of the ICA algorithm could give different sets of ICs. In order to solve these problems, Beckmann and Smith [2004] proposed an integrated approach for fMRI data analysis named Probabilistic ICA (PICA). Differing from the classical ICA framework, PICA allows for non-square mixing in the presence of Gaussian noise. Using Bayesian analysis, this method first estimates the amount of Gaussian noise and the true dimensionality of the data. After that, it carries out probabilistic modeling and achieves a unique decomposition of the data. Thus, PICA provides an effective solution to the above two problems and reduces problems of interpretation. Nowadays, this model is one of the most commonly used ICA techniques in fMRI data analysis.

Another issue of ICA analysis is that it does not provide a method for inference regarding groups of subjects. Unlike GLM, where individuals in the group share the same models, in ICA different individuals in the group have different ICs and are sorted differently. Due to this issue, several multi-subject ICA analysis methods have been proposed. These methods can be generally grouped into three categories. The first type of group analyses performs single-subject ICA and then combines the output into groups afterwards. For instance, Esposito et al. [2005] proposed a frame-

work to study the natural self-organizing clustering of many independent components from multiple individual data sets in the subject space. Another example is Wang and Peterson [2008], who developed 'Partner-Matching' to identify the ICs that are reproducible within or across individuals. Generally speaking, this class of methods allows for unique spatial and temporal features for each subject. But, since the data is noisy, the components are not necessarily unmixed in the same way for each subject.

Another type of approaches concatenates the data from all subjects together either spatially or temporally with the independent components decomposed from concatenated data. Svensen et al. [2002] presented a method that produces a set of time courses common to the whole group. Corresponding to each time course, this method gives a separate image for each of the subjects. By contrast, [Calhoun et al., 2001a] proposed another way of concatenation. This method first reduces the dimension of data from each subject via PCA. Then, the data from all the subjects is concatenated into one matrix. After that, a second PCA reduction further reduces the data before the final ICA decomposition. Schmithorst and Holland [2004] compared these two methods with the conclusion that subject-wise concatenation produced the best overall performance. To summarise, for this type of group ICA, since all subjects share one set of ICs, the comparison of subject difference within a component is straightforward. However, due to the concatenation, these methods require large computation and PC memory.

Finally, the tensorial approach introduced in Beckmann and Smith [2005] factors data of all subjects as a combination of two outer products of loadings in the tempo-

ral and spatial domains. This method is a natural extension of PICA. But, differing from the widely used PICA, the performance of this method is still under exploration.

For more thorough reviews on group ICA for fMRI data, readers could refer to Leibovici et al. [2001] and Calhoun et al. [2009]. Generally, these group analysis methods increase the power of ICA-based fMRI analysis.

Although data-driven methods have been widely accepted for fMRI data analysis, their primary disadvantage is that the interpretation of the derived components is left completely to the experimenter. On the contrary, pattern-based classifiers can overcome fatal flaws in the inferential and exploratory multivariate approaches. We introduce this type of analysis in the following section.

2.2.4 Machine learning classifier for fMRI data analysis

In the last few years, pattern-based classification analyses are appearing with increasing frequency in the functional neuroimaging literature. These methods cover a wide range of applications from activation detection to mental state recognition. For instance, Liang et al. [2006] presented an application of support vector machine (SVM) methodology for fMRI activation detection. Later, Song et al. [2007] formulated the problem of activation detection as an outlier detection problem of the one-class support vector machine. Another example is that Wang [2009] proposed a hybrid exploratory and hypothesis-driven fMRI data analysis method through combining conventional GLM with the support vector machine.

Besides brain activation detection, mental state recognition is another important ap-

plication of machine learning classifiers to fMRI analysis. Mitchell et al. [2004] used machine learning methods to classify the cognitive state of human subjects based on fMRI data sets. They have successfully distinguished cognitive states such as whether the subject is looking at a picture or a sentence. Haynes et al. [2007] used SVM to predict hidden intentions in the human brain. According to the prediction accuracy, they found the brain regions that encode these intentions. Similar applications also include those of Kamitani and Tong [2005, 2006]; Lee et al. [2009] and others. As mental state recognition is a complex process, it is very difficult to study it by classical GLM or data-driven approaches. The successful applications of machine learning algorithms increase the potential of using fMRI as a powerful tool to research brain functions.

Figure 2.3 presents the general process of applying a supervised machine learning classifier to a practical problem. The first step is to define the problem. According to this problem, researchers need to design experiments and collect data. The second step is data preprocessing. In most cases, the original data contains noise and irrelevant components. In this step, noise and other redundant information should be identified and removed. Apart from that, high data dimensionality is also a problem for pattern analysis. Reducing data dimension is also an important task in this step. The third step is to choose a learning algorithm, which can automatically generate classifiers according to the training data. Finally, researchers need to adjust the parameters in each step so that the resulting classifier could give the best prediction rate. Kotsiantis [2007] provides a detailed summary of this process and a brief review of the most commonly used algorithms such as multilayered perceptrons

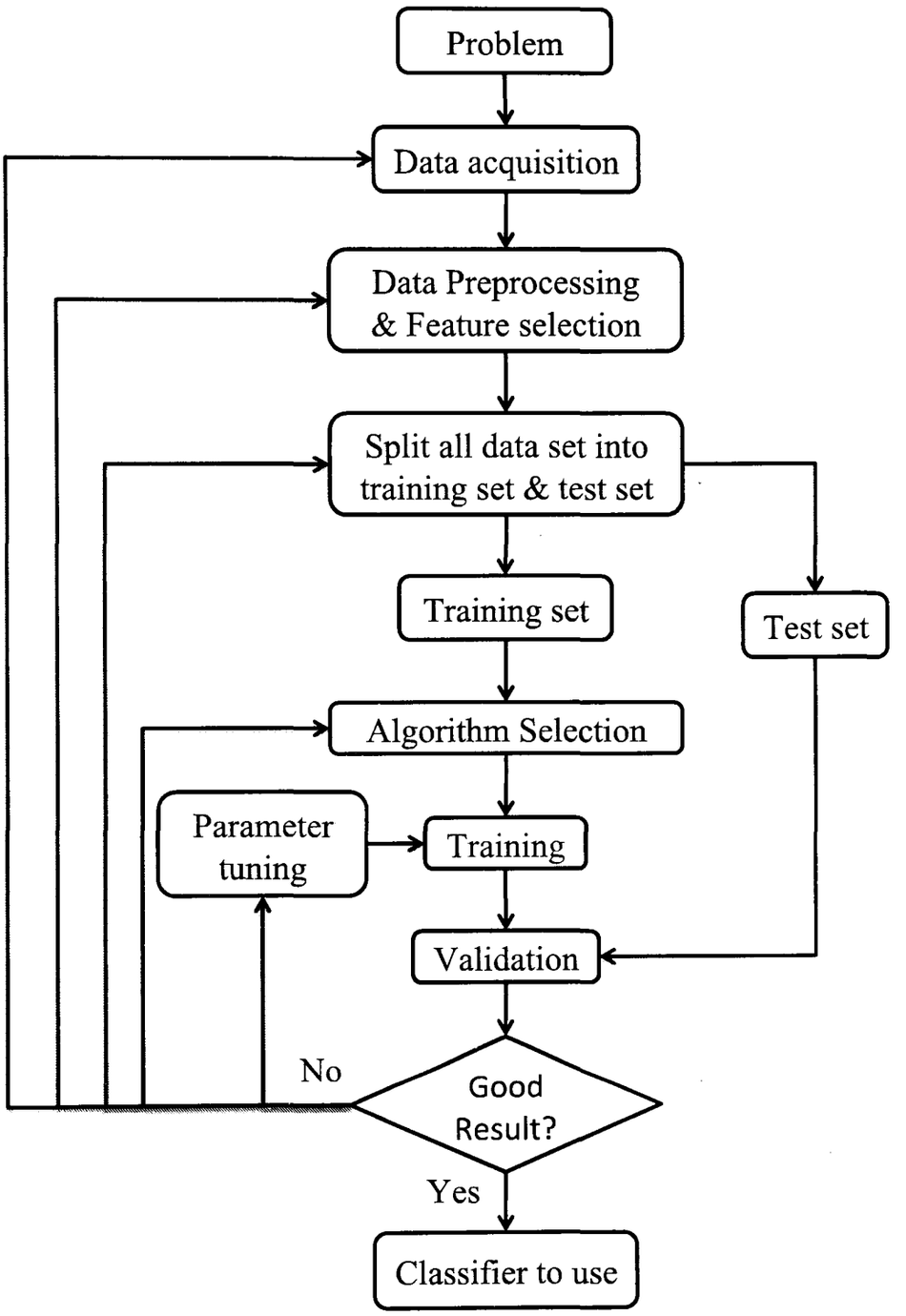


Figure 2.3: Process of constructing a machine learning classifier.

[Rumelhart et al., 1986] and SVM [Vapnik, 1995].

O'Toole et al. [2007] stated that there are three main reasons why pattern-based classification analysis is attracting attention. Firstly, it overcomes flaws of voxel-based

inferential (e.g. GLM) and exploratory multivariate approaches (e.g. ICA). The voxel-based methods take the data as a union of independent voxels. The correlation across voxels is usually ignored. On the other hand, the exploratory multivariate analyses lack effective ways of providing quantifiable links to experimental design variables. Secondly, pattern-based classification methods help with understanding of neural representation. By appropriately framing the experimental question, pattern based classifiers can offer insight into the neural codes that underlie different mental states. The third advance is that these approaches make fMRI data analysis a more interdisciplinary subject and attract research expertise from a wider range of behavioural and brain science.

Although pattern-based classification has the above advantages, one limitation of this method is the complexity of implementation and result interpretation. The application of a classifier is not as straightforward as the statistical and exploratory method. Different experimental designs and data samples require different classification approaches and parameter tuning methods to avoid overfitting and to keep the results reliable. Otherwise, the high dimensionality and limited number of samples could easily bias the analyses. Successful application of a classifier to fMRI data relies on tight cooperation between neuroscientists and experts in machine learning techniques. The neuroscientist needs to propose appropriately framed questions and the machine-learning specialist must ensure the accuracy and reliability of the data analysis. Under such circumstances, these approaches could open a door towards advancing functional neuroimaging studies and replacing the state-of-art analyses.

Machine learning for fMRI data analysis is a complex and quickly developing area. We suggest to readers Pereira et al. [2009] for a more detailed discussion of classifier methods in fMRI.

2.3 Human Brain Parcellation

Research on human brain mapping can be dated back to the times of the phrenologists. They believed that the amount of brain tissue devoted to a cognitive function determined its influence on behaviour. Accordingly, they divided the brain into several regions corresponding to different cognitive functions. Although the phrenologists' mapping is now considered to be a pseudoscience on scientific grounds, it introduced the idea of localization of function and established a functional atlas of the entire human brain that could be used to label each brain area with a specific function. Even today, the making of such an atlas remains one of the main aims in human brain mapping.

In 1934, Kleist published an atlas of the human brain by correlating the location of brain lesions with the a behavioural examination [Bartsch et al., 2000]. The next development came with Brodmann's observation of the neurons in the cerebral cortex, using the Nissl stain in 1909. Based on the cortical cytoarchitectonic organisation of these neurons, he defined 52 Brodmann areas [Guillery, 2000]. His maps of the cortical areas in humans form a fundamental step in the process of human brain parcellation.

The recent development of neuroimaging techniques provides new effective tools

for human brain parcellation in vivo. From our point of view, with the use of neuroimaging techniques, there are two major types of methods on parcellating the whole brain: *top-down approaches* and *bottom-up approaches*. We will discuss these two types of methods in the following sections.

2.3.1 Top-down approaches

In a top-down approach, researchers start the parcellation from the whole brain and use one or more imaging modalities to gradually partition the whole brain into smaller sub-regions, providing clear evidence. For example, using very high-resolution structural MRI and fMRI, Bridge et al. [2005] investigated the anatomical and functional borders between the primary and secondary human visual areas (V1 and V2). In order to find the anatomical boundary, they used three separate scanning sessions, in each of which, anatomical images were collected with different slice orientations. The hypo-intense band in the middle of the cortical grey matter was used as the anatomical signature of V1. In contrast, the functional borders were mapped with fMRI. They used visual stimulation to generate retinotopic maps, which were used to measure the location of V1/V2. They showed an excellent correspondence between the anatomical and functional borders.

Behrens et al. [2003] used DTI images to parcellate the thalamus, according to its connectivity to different cortex regions. They manually outlined the whole thalamus and a number of cortical zones. Using their probabilistic tractography algorithm with diffusion imaging data, they identified specific connections between the human thalamus and the cortex. According to this connection, the thalamus was

parcellated into several sub-regions. Later, Draganski et al. [2008] use probabilistic tractography on magnetic resonance diffusion weighted imaging data to segment basal ganglia and thalamus in 30 healthy subjects. They also found strong correlation between tractography-based basal ganglia parcellation and anatomical data from previously reported invasive tracing studies in nonhuman primates.

Also using DTI, Klein et al. [2007] parcellated BA 44/45 and SMA/pre-SMA. To parcellate SMA and pre-SMA, they first used a mask corresponding to these areas. Next, probabilistic tractography was run from every voxel within this mask to access connectivity with every voxel in the whole brain volume. According to this connectivity, they computed the cross-correlation matrix between the connectivity patterns of all voxels in the mask area. Next, in order to divide all voxels in the mask into two clusters, spectral reordering and k-means clustering were applied to the cross-correlation matrix. A similar process was also performed on BA 44/45. They found that the results of the two clustering methods agreed with each other. In addition, they used cytoarchitectonic probability data from SPM Anatomy to further examine their findings, .

Beckmann et al. [2009] used a similar method to parcellate the human cingulate cortex. They manually drew a mask, named the cingulate seed masks (CSM). Then, using the same method as Klein et al. [2007], a cross-correlation matrix was calculated based on the basis of probabilistic tractography. According to this matrix, CSM was first parcellated into five sub-regions. These sub-regions were then used as seed masks for the second iteration of the parcellation procedure, which divided CSM into 9 parcels. It was then seen that the third iteration did not lead to parcellations

that were reliably similar across subjects. Finally, CSM was parcellated into 9 sub-regions with two iterations. In order to assess the relationship between anatomical connectivity and function, they performed a meta-analysis of 171 functional studies reporting cingulate activation.

Unlike previous studies, Cohen et al. [2008] developed a method that used functional connectivity MRI (fcMRI) to define functional areas in individual human brains. This method consisted of a surface-based analysis. Image volume was transformed into cortical surface with Caret software and a grid of seed points was sampled on the surface. These seeds could be considered as pixels on a 2D image (cortical surface). For each seed point, a volumetric correlation map was generated by correlating the time course of the seed point and all other voxels over the entire volume of the brain. Next, they defined the similarity, eta^2, between two seed points, which is calculated between the two volumetric correlation maps generated from these seed points. For each seed, there was matrix of eta^2 representing the similarity between that seed and all other seeds. This matrix could be considered as a 2D image. Finally, Canny edge detection algorithm was applied to each seed's eta^2 "image" for edge detection. Combining this method with functional MRI, Nelson et al. [2010] divided the left lateral parietal cortex into sub-areas based on the presence (or absence) of memory-retrieval-related activity.

This type of parcellation method focuses on solid evidence that the proposed sub-regions exist and the parcellation is highly reproducible within and across subjects. However, the limitation of this approach is that it can only be applied to some specific brain regions. Due to the high complexity and individual variability of the hu-

man brain, whole brain parcellation is very difficult using this type of approaches. However, the continued endeavours of researchers enable this type of parcellation to cover more human brain areas.

2.3.2 Bottom-up approaches

In contrast to the first approach, this type of parcellation methods first defines a measurement that assesses the similarity between voxels. According to this measurement the whole brain is parcellated into a certain number of homogeneous regions. The main aim of this kind of parcellations is to facilitate further analysis.

Coulon et al. [2000] produced one of the earliest studies to use parcellation for the analysis of functional activation maps. They aimed to process a group analysis while preserving individual information. In this research, they computed grey-level blobs from three-dimensional activation maps. These blobs, which can also be termed parcels, are calculated in the following way: from each local maximum, a growing region is constructed around this maximum, until it meets another region or a point that belongs to the 'background'. With different scale-space blobs, they constructed a comparison graph that included all subjects. This inter-subject comparison graph was used in a labelling process for activation detection.

With a similar research aim, Thirion et al. [2006] proposed a multi-subject whole brain parcellation. This method parcellates the whole brain into a certain number of parcels according to the parameters of General Linear Models (GLM). In this approach, voxels from all subjects are first pooled together. A C-means clustering algorithm is used to derive parcel prototypes on GLM parameters. The clustering

process is under the spatial constraint that voxels can only be assigned to prototypes that are closer than a predefined distance. Next, for each subject, seed voxels are found that correspond to the parcel prototypes defined in the previous step. These seed voxels should be functionally and spatially (in standard space) close to the parcel prototypes. Moreover, the warp from these seed voxels to the prototypes should be regular. In the third step, other voxels in each subject are assigned to these seed voxels with a spectral clustering algorithm. The voxels assigned to the same seed voxel form a parcel. All the parcels whose seed voxels correspond to the same parcel prototype are matched with each other. Statistical analysis is constructed on the matched parcels. This work also demonstrates a method to improves the sensitivity of group analyses and functional activity representation.

Unlike the above studies which have focused on parcellation for activation detection. Hutchinson et al. [2009] have proposed a method that integrates parcellation with the classifier model. In this research, they use a Hidden Process Model (HPM) to model the fMRI data. This model assumes that the observed data is generated by a sequence of underlying mental processes the timing of which may be unknown. For each voxel, a set of parameters is used to describe the mental process. Machine learning algorithms are applied to estimate model parameters. In order to improve the accuracy of this estimation, they proposed a method that reduces the effective number of parameters. This algorithm uses a nested cross-validation hierarchical approach to undertake two tasks at the same time. The first task is to partition the brain into clusters of voxels that will share parameters and the second task is to estimate these parameters simultaneously.

Although this application is successful, we do not prefer this type of parcellation. One problem is that the integration of parcellation and parameter estimation increases the complexity of the model and the risk of overfitting. Aonther disadvantage is the difficulty of combining other image modalities into this parcellation process.

To sum up, all of parcellation methods in this class try to parcellate the whole brain at once. For each subject, they divide the brain into hundreds of regions. Therefore, compared to $top - down$ approaches, it is difficult to validate rigorously every parcel and every boundary from the parcellation results. However, a noticeable advantage of these methods is that they can give meaningful and reasonable parcellation of the whole brain in a very short time. As discussed before, the high spatial dimensionality of neuroimaging data is a common problem for many analysis methods, especially for multivariate methods like machine learning classifiers. Parcellation could provide an effective tool to solve this problem. Therefore, it is meaningful to develop effective and efficient individual parcellation methods and as well as cross-subject parcel analysis methods.

2.4 Summary

In this chapter, we reviewed the fMRI imaging theory, data analysis processes and human brain parcellation.

After introducing the imaging theory, we presented several fMRI experiments and data analysis methods. Although we introduced the experiment design and data

analysis separately, they are very closely related. The experiment design determines corresponding data processing methods.

Next, we gave a survey on current parcellation methods and categorised these methods into two classes: top-down approaches and bottom-up approaches. We described the corresponding advantages and disadvantages of these two classes.

According to our survey, we found that there were only a limited number of parcellation methods in the bottom-up approach. On the other hand, multivariate analysis methods (e.g. machine learning classifiers) need an efficient whole brain parcellation process to reduce data dimension and improve reliability of analysis. The multi-subject parcellation framework developed in Thirion et al. [2006] is specially designed for GLM based analysis. There is a need for a parcellation framework that can be used for data-driven and multivariate analyses.

In order to fill this gap in knowledge, we propose a flexible fMRI data analysis framework based on parcellation in this thesis. We will first develop a data-driven parcellation method for individual subject. Then, we will use a graph partitioning method to find the correspondence of parcels in different subjects. The multivariate analysis or other models can be constructed on matched parcels. In the next chapter, we will introduce our data-driven individual parcellation method.

CHAPTER 3

Parcellation of Individual Subjects

3.1 Introduction

As mentioned in Section 2.3, Thirion et al. [2006] present parcellation as a solution to the problem of mis-registration and individual variance of brain anatomy. In the previously proposed parcellation framework, voxels are first clustered into functionally homogeneous regions, or *parcels*, for each subject. After that, the parcellations are homogenized across subjects, so that statistics can be computed at parcel level rather than at voxel level.

In this chapter, we will present a data-driven, model-free, parcellation technique, based on Principle Component Analysis (PCA), Independent Component Analysis (ICA) and Partial Least Squares (PLS) (see Figure 3.1). Instead of using GLM, the feature space is generated with ICA and PLS for each subject. Thus, parcellation results are not biased by the assumption of a particular HRF shape.

As in most parcellation methods [Kim et al., 2010a; Neumann et al., 2006; Peltier et al.,

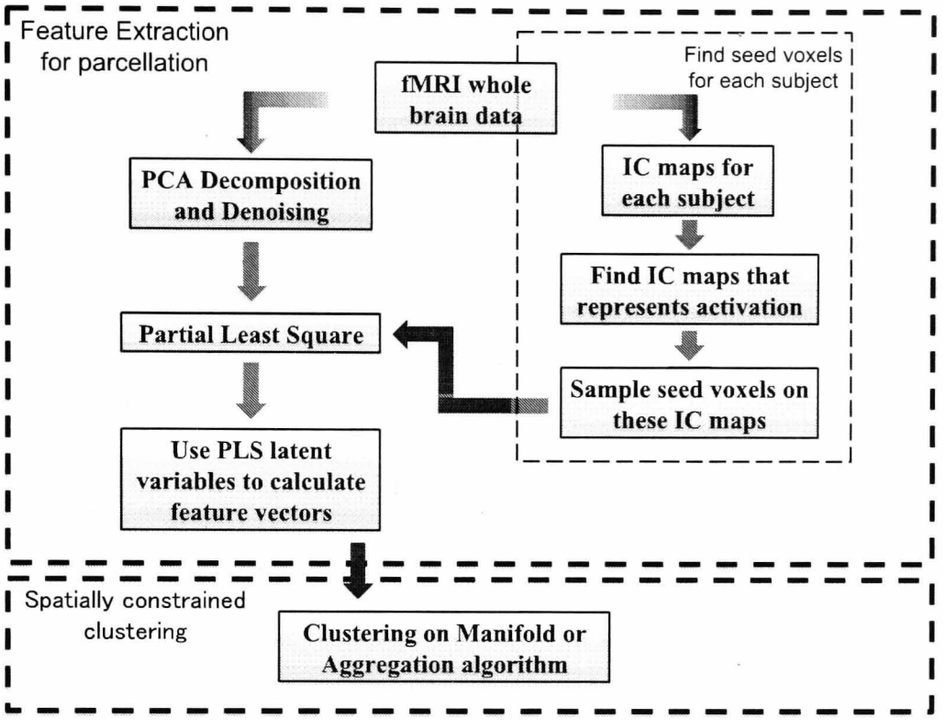


Figure 3.1: A data-driven approach to parcellation. This method has two steps: feature extraction and spatially constraint clustering. In the first step, seed voxels are selected on IC maps of each subject. After denoising with Principal Component Analysis (PCA), Partial Least Square (PLS) latent variables are calculated with signals from seed voxels and Principal Components (PCs) from the whole brain. The covariance between signals of each voxel and PLS latent variables are used as feature vectors. In the second step, spatially constrained clustering is applied on these feature vectors for parcellation

2009; Shen et al., 2010; Thirion et al., 2006], our parcellation process can be divided into two steps: (1) feature extraction and (2) spatially constrained clustering . In the feature extraction step, pre-processing steps, such as slice timing and realignment, are first applied to the data. We implement linear interpolation and affine registration with FSL. Then, according to the histogram of the functional images, some

invalid voxels are removed. Finally, in our scheme, ICA and PLS are applied to calculate a vector for each voxel in order to describe the functional behaviour. In the second step, we discuss several methods for spatially constrained clustering. We first introduce spectral clustering. After that, we propose an aggregation method.

In the following sections, we first propose a novel feature extraction method for parcellation. In Section 3.3, we introduce spectral clustering for parcellation and suggest a statistical adaptive smoothing as a preprocessing step. In addition, we propose a fast parcellation method based on aggregation. In Section 3.4, two criteria for validating parcellation results are introduced and some experiment results are shown. We give the conclusion of this chapter in Section 3.5.

3.2 Feature extraction for parcellation

In this section, we introduce a data-driven method to extract brain activation features from fMRI data. ICA is a popular and effective tool to detect activation in fMRI experiment. However, there are two problems that make ICA maps not ideal for parcellation. First, the Independent Components (ICs) from ICA decomposition are not ordered. It is difficult to tell which ICs are activation related and which ones are not, especially when considering the individual variability of BOLD responses. Another problem is that ICs are optimized for maximizing the independence between ICs, rather than describing the BOLD signals. Using IC maps to describe the functional behaviour of each voxel may lose some activation-related information in the fMRI data.

McIntosh et al. [1996, 2004] have shown that Partial Least Square (PLS) is sensitive to the detection of task-related activity changes. Thus, we propose a new method which uses PLS scores as the functional measurement of each voxel. In this method, PLS is used to predict signals in the activated parts of the brain with the PCs of signals in the whole brain. The latent variables describe signals in the activated parts of the brain. If dealing carefully with the PCs and activated seed voxels, PLS could be a better way of characterizing the BOLD signal.

Here, we explain the approach proposed in Figure 3.1. This method selects several seed voxels in activated regions. The seed voxels should contain activation signals. In addition, they should also be located in different regions to account for BOLD signal variability. These seed voxels represent the fMRI signals of interest. However, there is also noise in addition to artefacts in the time courses of these voxels. Thus the Partial Least Square latent variables are calculated using the PC of the whole brain to predict signals from seed voxels. The covariances between the fMRI signal in each voxel and the latent variables are used as the feature space for further parcellation. Therefore, the problem of extracting feature space for parcellation becomes a question of how to select appropriate the seed voxels.

Although many methods could be used in this step, we choose ICA due to the fact that it is one of the most effective data-driven approaches for fMRI data analysis. There are many methods to apply ICA on fMRI data analysis. In this thesis, Probabilistic Independent Component Analysis (PICA) [Beckmann and Smith, 2004] is applied to individual subjects. For single subject data, the ICs of interest have to be selected either manually or according to their correlation with the experiment

design. For multi-subject data we propose the use of the cross-subject reproducibility of ICs for IC selection. As the pre-processing steps of slice timing and realignment have been thoroughly discussed in previous work [Andersson et al., 2000; Frackowiak et al., 2004; Grootoonk et al., 2000; Henson et al., 1999], they are not covered in this section. The discussion in this section focuses on the step of using ICA and PLS for feature extraction and using different methods for parcellation.

3.2.1 Histogram of functional images

Slice-timing and realignment are first implemented to the fMRI data as preprocessing steps. After that, the 4-D fMRI signal can be presented as $f(x,y,z,t)$, where x,y,z are the coordinates of the voxels and t is the time index. The intensity of the image time series is first normalized to $0-100$, according to the equation:

$$f_n(x,y,z,t) = \frac{f(x,y,z,t) - \min_{x,y,z,t} f(x,y,z,t)}{\max_{x,y,z,t} f(x,y,z,t) - \min_{x,y,z,t} f(x,y,z,t)} * 100. \tag{3.2.1}$$

Let $f_{mean}(x,y,z)$ be the mean of the normalized image series:

$$f_{mean}(x,y,z) = \frac{1}{T}\sum_{t=1}^{T} f(x,y,z,t), \quad \text{where} \quad 0 \leq t \leq 1. \tag{3.2.2}$$

A typical histogram of the mean image f_{mean} is shown in Figure 3.2. This data set was acquired on a Philips Intera 1.5T scanner with a TR of 3s. During the scan, the subject was undertaking a sequential finger-tapping task auditorily paced with a metronome. More detail of this data set is introduced in Section 5.2.

The histogram is divided into four colour-coded parts. As shown in Figure 3.3, we overlaid each part of the brain images with a different colour. The blue part of the histogram corresponds to the background of the images. The green part of

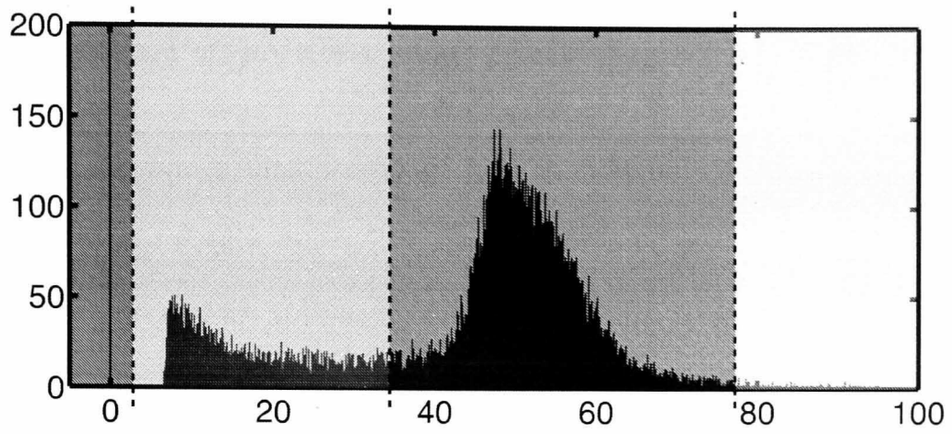


Figure 3.2: Histogram of normalized fMRI images. The whole histogram is divided into four parts by three thresholds (dash lines). The four parts of the histogram are illustrated with the mean fMRI images in Figure 3.3

the histogram corresponds to the voxels on the boundary between the background and grey matter. In these boundary areas, the process of realignment biases the signals. The bias reduces the mean of the signal intensity and lowers the reliability of the analysis made on these voxels. The yellow part of the histogram mainly corresponds to the ventricles in the brain. The signals from the ventricles are random and have great similarity across subjects. These voxels could also bring bias to further analysis.

Following the above discussion, only the voxels corresponding to the red part of the histogram are considered as valid voxels. These voxels are brought to the next processing steps. As different fMRI images have different histograms, attention should be paid to the selection of the threshold for the histogram. As shown in Figure 3.3, there are three thresholds to determine. Here we choose the threshold manually. For convenience of description, in the rest of the thesis, this step of pre-processing

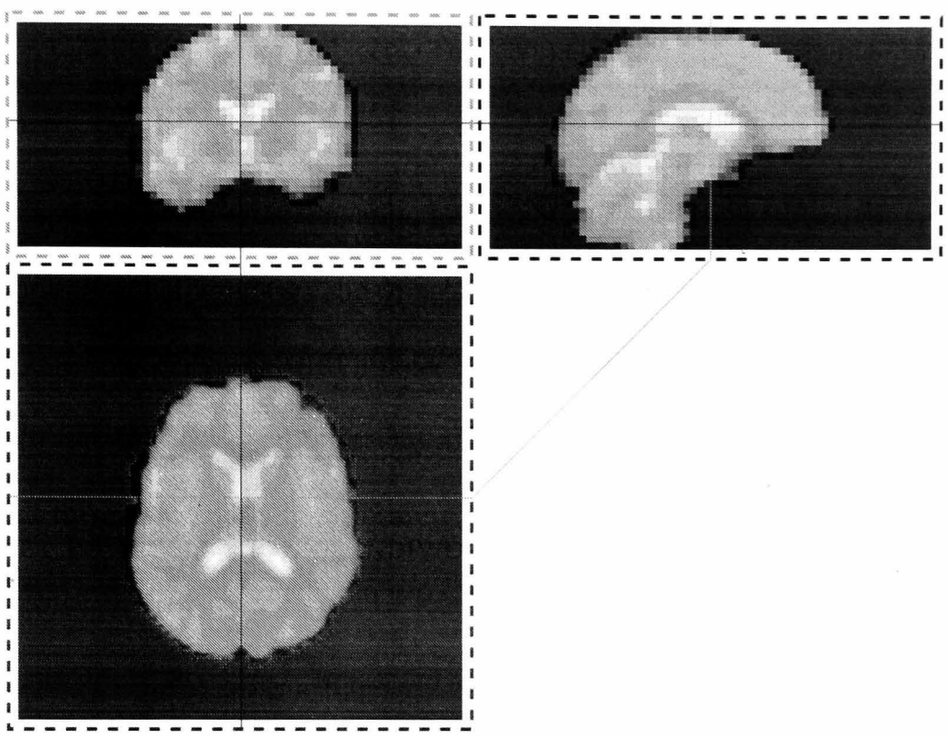


Figure 3.3: Different parts of brain corresponding to different parts of histogram. The images in the red, green and blue dashed rectangles are the images on the transverse, coronal and sagittal plane. The red, green and blue lines on the images are the axes of transverse, coronal and sagittal plane.

is called *histogram-filtering*.

3.2.2 Independent Components Analysis for fMRI Group Analysis

As introduced in the Section 3.2, the aim of ICA analysis in the proposed scheme is to select seed voxels in all subjects. Therefore, we are looking for the ICs that are reproducible across subjects. Here, we focus on the reproducibility of IC time courses.

For each subject, we first use PICA to decompose fMRI signals into several ICs. We

assume that, from the subjects who are under the same stimulation paradigm, ICs that represent neuronal signals are more similar to each other than to noise and artefact ICs. Thus, the corresponding IC time courses representing independent fMRI signals of neuronal origin can be clustered and identified.

Here, we propose the use of a constrained clustering approach. Similar in spirit to the approach of Partner Matching [Wang and Peterson, 2008], this method can be considered to be a means of finding the ICs that best capture the Blood Oxygen Level Dependent Hemodynamic (BOLD) response to the stimuli. We aim to group those IC time courses that are associated with the responses to the same task features in different subjects into one cluster. The other ICs, which do not contain relevant (i.e. task-related) information, should be grouped into other clusters.

Let N_a and N_b be the number of ICs for subjects A and B respectively, with $\mathbf{IC}_i^{\mathrm{A}}$ and $\mathbf{IC}_j^{\mathrm{B}}$ the ith IC of subject A and jth IC of subject B. $t = 1, \ldots, T$ is the time index. Their correlation coefficients are given by:

$$\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) = \frac{\sum_{t=1}^{T} \left(\mathbf{IC}_i^{\mathrm{A}}(t) - \overline{\mathbf{IC}_i^{\mathrm{A}}}\right)\left(\mathbf{IC}_j^{\mathrm{B}}(t) - \overline{\mathbf{IC}_i^{\mathrm{B}}}\right)}{\sqrt{\sum_{t=1}^{T} \left(\mathbf{IC}_i^{\mathrm{A}}(t) - \overline{\mathbf{IC}_i^{\mathrm{A}}}\right)^2} \sqrt{\sum_{t=1}^{T} \left(\mathbf{IC}_j^{\mathrm{B}}(t) - \overline{\mathbf{IC}_i^{\mathrm{B}}}\right)^2}}, \tag{3.2.3}$$

The normalized correlation coefficients ρ_{norm} is:

$$\rho_{\mathrm{norm}}(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) = \frac{\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) - \mathrm{mean}\left(\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}})|_{j=1,2,\ldots,N_b}\right)}{\mathrm{std}\left(\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}})|_{j=1,2,\ldots,N_b}\right)} \tag{3.2.4}$$

Since the aim of the clustering is to put similar ICs from different subjects into one cluster, all the ICs of the same cluster should come from different subjects. Therefore we need to set the similarity between ICs of the same subject to 0. The similarity

between two ICs is then defined as

$$S(\mathbf{IC}_i^A, \mathbf{IC}_j^B) = \begin{cases} 0 & \text{if } A = B \\ \min\left(\rho_{\text{norm}}(\mathbf{IC}_i^A, \mathbf{IC}_j^B), \rho_{\text{norm}}(\mathbf{IC}_j^B, \mathbf{IC}_i^A)\right) & \text{other wise .} \end{cases} \tag{3.2.5}$$

The similarity matrix defined in 3.2.5 is based on the Correlation Coefficient. It could be easily modified to accommodate other measurements for similarity. For instance, if we change equation 3.2.3 to

$$\rho(\mathbf{IC}_i^A, \mathbf{IC}_j^B) = I(\mathbf{IC}_i^A, \mathbf{IC}_j^B) = \sum_{u \in \mathbf{IC}_i^A} \sum_{v \in \mathbf{IC}_j^B} p(u,v) \log \frac{p(u,v)}{p(u)p(v)}, \tag{3.2.6}$$

and brought equation 3.2.6 into equation 3.2.4 and equation 3.2.5, we could get similarity indices based on mutual information.

Following this process, we use clustering techniques to separate ICs that correspond to brain activation. Three clustering and analysis methods are used here to find these ICs: hierarchical clustering, Principal Components Analysis (PCA) and manifold embedding. Hierarchical clustering is calculated directly on the similarity defined in equation 3.2.5. PCA is applied to the similarity matrix to examine the structure of these ICs. Considering ICs as vertices of a graph, based on similarity, the weights between two vertices can be calculated as:

$$d(\mathbf{IC}_i^A, \mathbf{IC}_j^B) = \begin{cases} 1/S(\mathbf{IC}_i^A, \mathbf{IC}_j^B) & \text{if } S(\mathbf{IC}_i^A, \mathbf{IC}_j^B) > 0, \\ \infty & \text{if } S(\mathbf{IC}_i^A, \mathbf{IC}_j^B) <= 0. \end{cases} \tag{3.2.7}$$

Next, the embedding method introduced in Section 3.3, is applied to the graph for spectral clustering. The results are further discussed in Section 5.3.

3.2.3 Seed Selection

Due to the fact that BOLD responses of the human brain vary across different regions of the brain, in order to calculate the PLS latent variables that best capture the BOLD response, a number of seeds representing different active regions should be selected. For instance, when using GLM to measure functional responses, we could select seeds according to the map of the t-values. The voxels selected as seeds should have high t-values, so that the time course of these voxels can represent the functional responses. In addition, these seeds should be spatially separated from each other and be located in different activated regions to contain the variance of BOLD across regions.

In this thesis, in order to parcellate the whole brain in a data-driven approach, we use the IC maps in Beckmann and Smith [2004] to select the seeds. First, the IC maps corresponding to the brain activations are selected, using the method introduced in the last section. Then, within each IC map, the first seed is chosen as the voxel with the largest value, which presents the strongest response. The second seed is then chosen amongst the voxels at least R voxels away from the first seed voxel. Of these voxels, the second seed should correspond to the largest IC map value. The iterative process is repeated until all the seeds have been selected. Therefore the seeds are located in different active regions.

The number of seed voxels and the size of radius R depend on the experiment design of the fMRI scan. The rule of thumb is that seed voxels should be located in different activated regions, such that the corresponding fMRI signals could represent the variance of BOLD in different activated regions. For simple task data, for

instance, the finger tapping data introduced in Section 5.2, 30 seeds are selected from one IC map with $R = 6$. When the experiment design is more complex, we need to select the seeds from more than one IC map. For instance, in the multi-subject case described in Section 5.3, the subjects were under four types of stimulation: angry hand gesture, neutral hand gesture, angry face expression and neutral face expression. According to the result of the clustering analysis introduced in section 3.2.2, two IC maps are used for each subject. Thus, we pick $R = 6$ voxels and select $N_{seed} = 15$ seeds for each map.

3.2.4 PCA for fMRI denoising

Due to the fact that the BOLD signal is very weak and the fMRI experiment process is complex, the Signal Noise Ratio (SNR) of the fMRI time series is very low. Various noise sources contribute to this low SNR, such as physical noise, machine drift from the scanner [Weisskoff, 1996], respiratory and heart beat noise [Biswal et al., 1996], and so on. Correspondingly, many methods have been introduced for denoising fMRI data [Flandin and Penny, 2007; Mohamed et al., 2007; Monir and Siyal, 2009; Song et al., 2006], either for general denoising or for removing specific noise structure. Here, we use Principal Component Analysis (PCA) as a general denoising tool [Kerrouche et al., 2006; Zuendorf et al., 2003].

PCA is one of the most popular data analysis tools for dimensionality reduction [Jolliffe, 2002]. It transforms the data into a new coordinate system. In the new coordinate system, the greatest variance by any projection of the data comes to lie on the first coordinate (the first principal component). The second coordinate (the

second principal component) is the direction that is orthogonal to the first principal component and covers the second greatest variance. The third principal component is orthogonal to previous components and covers the third greatest variance. The rest of the principal components are calculated accordingly.

Let $\mathbf{X}_{V\times T} = x_{ij}$ denote the data matrix after time-slicing, realignment and histogram-filtering, where each row corresponds to the fMRI signal of a given voxel. After that, we propose to denoise the signal with PCA in our parcellation scheme. We first centre the signal at each voxel by subtracting its mean as in equation 3.2.8.

$$x_{ij}^{\text{centered}} = x_{ij} - \frac{\sum_{j=1}^{T} x_{ij}}{T} \quad i = 1, 2, ..., V. \tag{3.2.8}$$

After that, $\mathbf{X}_{V\times T}$ is decomposed into $\mathbf{P}_{\text{PCA}}$ and $\mathbf{T}_{\text{PCA}}$:

$$\mathbf{X}_{V\times T} = \mathbf{P}_{\text{PCA}} \cdot \mathbf{T}'_{\text{PCA}} \tag{3.2.9}$$

$\mathbf{T}'_{\text{PCA}}$ is the transpose of the PCA score matrix of $\mathbf{X}$ (the matrix whose columns are the Principal Components (PCs) of the fMRI data), and $\mathbf{P}_{\text{PCA}}$ is the PCA loading matrix. As the size of $\mathbf{X}$ is very large (about $20000 \times T$ for 3T $64 \times 64 \times 32$ fMRI data), the decomposing method introduced in Zuendorf et al. [2003] is used here. So that, $\mathbf{X}_{V\times T}$ can be decomposed as:

$$\mathbf{P}_{\text{PCA}} = \mathbf{X}_{V\times T} \cdot \mathbf{e} \cdot \mathbf{L}^{-1/2}, \tag{3.2.10}$$

where, $\mathbf{e}$ is the matrix of eigenvector of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{L}$ is a diagonal matrix and its nonzero elements are the eigenvalues of $\mathbf{X}^T\mathbf{X}$. As $\mathbf{P}'_{\text{PCA}} * \mathbf{P}_{\text{PCA}} = \mathbf{I}$, the principal components can be calculated as

$$\mathbf{T}_{\text{PCA}} = \mathbf{P}'_{\text{PCA}} * \mathbf{X}_{V\times T}. \tag{3.2.11}$$

Ranking the PCs according to the variance they cover, the first few PCs usually have very high variances. These PCs correspond to high frequency noise and are removed manually. The last PCs are slow-variance artefacts. PCs that cover the last 10% of the variance are removed.

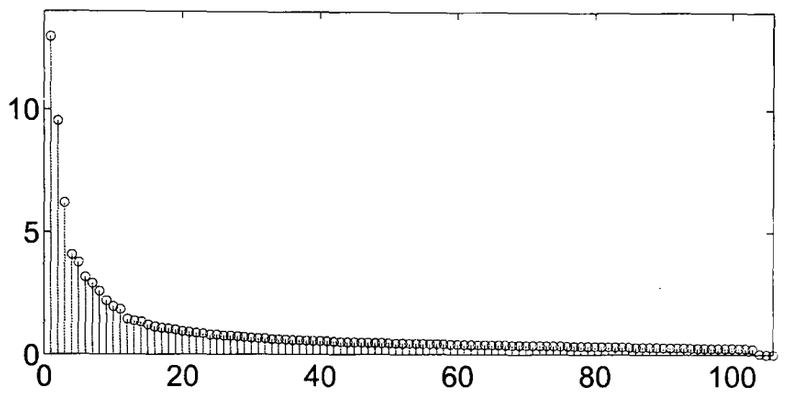


Figure 3.4: Variance explained by each principal component.

Here we still use the single-subject finger tapping data as an example to demonstrate the step of denoising with PCA. Figure 3.4 shows the variance covered by each PC, ranked from high to low. The horizontal axis presents the order of PCs, and the vertical axis presents the percentage of variance each PC covers. In Figure 3.5, the six PCs that cover the largest variance are shown. As claimed above, the first and second PCs present the high frequency noise. This is different from most other applications of PCA, in which the first few components provide the most useful information. It can be understood by the fact that, in fMRI, there is a large amount of noise in the data. These noises form the principal variance of the data.

In Figure 3.6, the six PCs that cover the smallest variance are shown. These components are slow-variant artefacts and meaningless noise. Thus, in this example, only the $3rd$ to the $40th$ PCs are used in the next step of analysis.

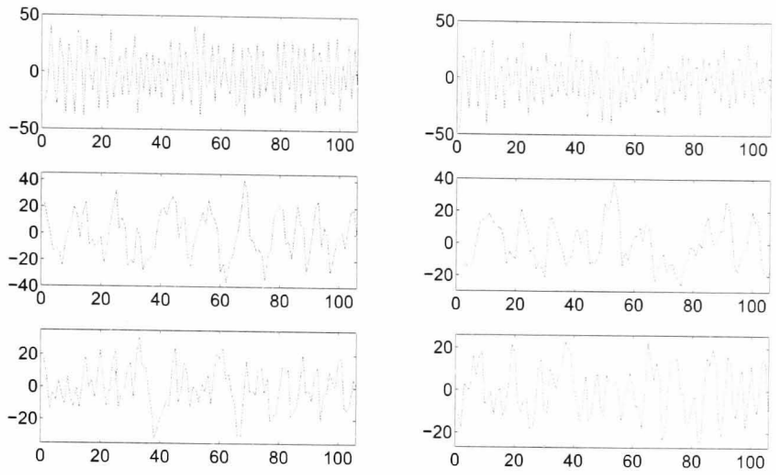


Figure 3.5: First six principal components

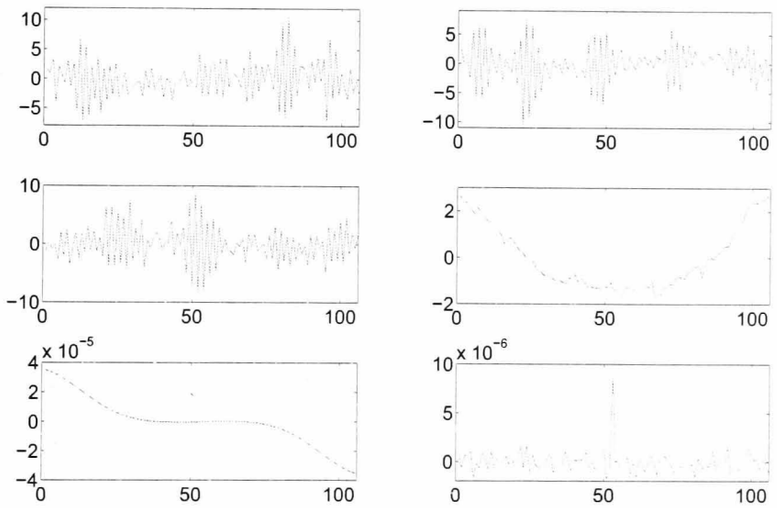


Figure 3.6: Last six principal components.

With different data, PCA results may be slightly different from each other. For instance, in some trials of the data introduced in Section 5.3, only the first PC presents the high frequency noise. However the general rule is still the same. For simplicity, in the rest of the thesis, $\mathbf{T}_{PCA}$ is used to denote the matrix that contains only PCs of interest.

3.2.5 Partial Least Square (PLS) for feature extraction

Let $\mathbf{D}_{T\times N_{seed}}$ represent the fMRI signals of the seed voxels selected as in Section 3.2.3. Each column of $\mathbf{D}_{T\times N_{seed}}$ corresponds to the fMRI signal in a given seed. Then, we use the PCs in matrix $\mathbf{T}_{\text{PCA}}$ for the prediction of $\mathbf{D}$ with Partial Least Square (PLS). These components, the *latent variables*, should contain information from both $\mathbf{T}_{\text{PCA}}$ and $\mathbf{D}$. Here PLS is used to calculate the time series components that represent the individual specific functional activity signals. We decompose $\mathbf{T}_{\text{PCA}}$ as:

$$\mathbf{T}_{\text{PCA}} = \mathbf{T}_{\text{PLS}}\mathbf{P}'_{\text{PLS}} \quad \text{where} \quad \mathbf{T}'_{\text{PLS}}\mathbf{T}_{\text{PLS}} = \mathbf{I} \tag{3.2.12}$$

And $\mathbf{D}$ is predicted as:

$$\widehat{\mathbf{D}} = \mathbf{T}_{\text{PLS}}\mathbf{B}\mathbf{C}', \tag{3.2.13}$$

where the columns of $\mathbf{T}_{\text{PLS}}$, $\mathbf{t}_i, i = 1,2,...,$ are the latent vectors of size $T \times 1$. $\mathbf{B}$ is a diagonal matrix with the "regression weights" as diagonal elements and $\mathbf{C}$ is the "weight matrix" of the dependent variables [Abdi, 2003].

Given $\mathbf{T}_{\text{PCA}}$ and $\mathbf{D}$, the latent vectors could be chosen in several different ways. The canonical way is to find the latent vectors that maximize the covariance between the columns of $\mathbf{T}_{\text{PLS}}$ and $\mathbf{D}$ [Wold et al., 2001]. Specifically, the first latent vector is calculated as:

$$\mathbf{t}_1 = \mathbf{T}_{\text{PCA}}\mathbf{w}_1, \tag{3.2.14}$$

$$\mathbf{u}_1 = \mathbf{D}\mathbf{c}_1, \tag{3.2.15}$$

with the constraint that

$$\mathbf{t}_1'\mathbf{t}_1 = 1, \tag{3.2.16}$$

$$\mathbf{w}_1'\mathbf{w}_1 = 1, \tag{3.2.17}$$

and $\mathbf{t}_1'\mathbf{u}_1$ be maximal.

After that, the first component is subtracted from $\mathbf{T}_{PCA}$ and $\mathbf{D}$, and the rest of the latent variables are calculated iteratively as the above until $\mathbf{T}_{PCA}$ becomes a null matrix. The first PLS latent variables are signals of interest. Let $\mathbf{X}_0$ be derived from $\mathbf{X}$ after the signal variance has been removed: $\mathbf{x}_0 = \mathbf{x}/||\mathbf{x}||$, where $\mathbf{x}$ and $\mathbf{x}_0$ are the row vectors of $\mathbf{X}$ and $\mathbf{X}_0$. We use the covariances between fMRI signals and latent variables, as shown in equation 3.2.18 , as the feature space for parcellation.

$$\mathbf{r}_i = \mathbf{X}_0\mathbf{t}_i \tag{3.2.18}$$

3.3 Spatially constrained clustering for parcellation

Given the feature space that represents neural activity, the next step is to divide the whole brain into several spatially connected and functionally homogeneous regions, according to the feature space. Thirion et al. [2006] have proposed a solution to this problem with spectral clustering. In this section, to solve the parcellation problem, two different ways are discussed. We first implement parcellation using a manifold embedding trick similar to that of Thirion et al. [2006]. After that, the *Aggregation* and *Boundary Competition* methods are described. In this section, we use a toy example to illustrate the performance of different methods. The discussion and comparison of these methods are given in Section 3.3.3.

3.3.1 Clustering on the manifold for parcellation

After feature extraction, each voxel in the spatio-functional space can be represented as a vector

$$\mathbf{v}(i) = [x(i) \quad y(i) \quad z(i) \quad f_1(i) \quad f_2(i) \quad ... \quad f_n(i)]', \qquad i = 1, 2, ..., V, \tag{3.3.1}$$

where x, y, z are the coordinates of the voxel; $f_1, f_2, ..., f_n$ are the measurement of functional behaviour; and V is number of valid voxels. For instance, when the parcellation is based on the GLM parameter [Thirion et al., 2006] , if there are n regressors to model the functional behaviour, the vector can be written as:

$$\mathbf{v}(i) = [x(i) \quad y(i) \quad z(i) \quad \beta_1(i) \quad \beta_2(i) \quad ... \quad \beta_n(i)]', \qquad i = 1, 2, ..., V. \tag{3.3.2}$$

We denote the set of these vectors as $\mathbf{V}$. These vectors can be considered as points sampled from a 3-dimensional manifold embedded in $\mathbb{R}^{n+3}$. According to Whitney's embedding theorem [Hirsch, 1994], n must conform to the constraint $n + 3 \leqslant 2 \times 3 + 1$. In other words, there can be, at most, $n = 4$ functional features. Otherwise the topology of the manifold is broken and unpredictable results may appear [LaValle, 2006].

Now, we explain how to implement the parcellation based on clustering on the manifold. For simplicity, a toy example generated with the equation 3.3.3 is used to illustrate the discussion.

$$f(x, y) = 5 \cdot (e^{-\frac{(x-15)^2+(y-35)^2}{90}} + e^{-\frac{(x-35)^2+(y-15)^2}{90}}) \tag{3.3.3}$$

Figure 3.7 shows this double Gaussian toy example, which is a 2-dimensional manifold embedded in $\mathbb{R}^3$. In this case, the voxels are $\mathbf{v}(i) = [x(i) \; y(i) \; f(i)]$, where

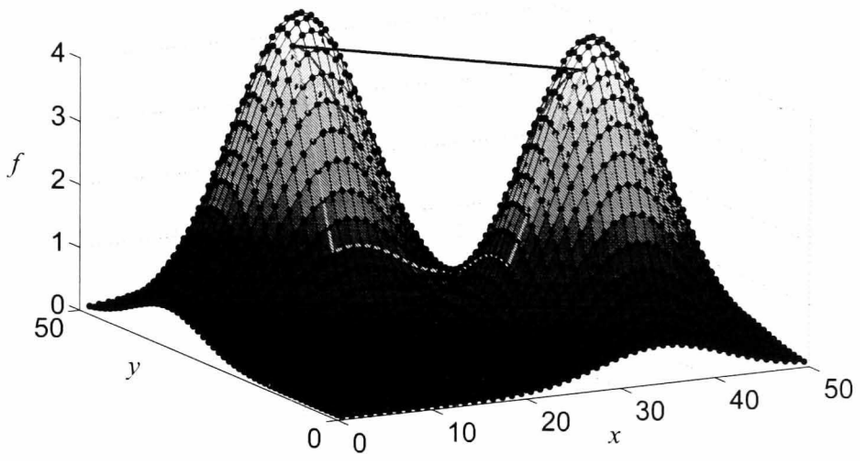


Figure 3.7: The manifold is generated by sampling points on double Gaussian function introduced in equation 3.3.3. The black dots are the sampled points. The blue line represents the Euclidean distance between two points in $\mathbb{R}^3$. The red line is the distance between these points on the 2-dimensional manifold embedded in $\mathbb{R}^3$. The green line illustrates that, considering the sampled points as vertices of graph, distances on the manifold can be estimated with the shortest path between the two points on the graph.

$f(i) = f(x(i), y(i))$. Given two voxels on the manifold, the Euclidean distance between them is illustrated by the blue line. The geodesic distance, which is the shortest possible line between these two voxels on the manifold, is shown by the red line.

Classic clustering methods use the Euclidean distances for clustering. For the parcellation problem, when using Euclidean distances for clustering, the voxels attributed to the same cluster may be spatially separate. This problem can be solved by using the geodesic distances for clustering [Thirion et al., 2006]. Due to the fact that

voxels are considered as sampled points on the manifold, we cannot directly calculate the geodesic distance on the manifold. Therefore, many methods have been developed to map the structure of the manifold into Euclidean space with lower dimensionality [Lafon and Lee, 2006; Roweis and Saul, 2000; Tenenbaum et al., 2000; Zhang and Zha, 2002], so that the geodesic distances can be calculated in the low-dimensional Euclidean space.

The general ideas of these methods are the same. Given a set of data sampled on a manifold, for instance $\mathbf{v}(i), i = 1, 2, ...,$, the geodesic distance between any two points$\mathbf{v}(i)$ and $\mathbf{v}(j)$ is denoted as $d_g(\mathbf{v}(i), \mathbf{v}(j))$. The aim of these methods is to find a new coordinate $\mathbf{x}_n(i)$ for each $\mathbf{v}(i)$, so that for any two points on the manifold, the geodesic distance between them $d_g(\mathbf{v}(i), \mathbf{v}(j))$ is equal to the Euclidean distance between their new coordinates $||\mathbf{x}_n(i) - \mathbf{x}_n(j)||$. The new coordinates are the embedding coordinates of the data points on the manifold. The Euclidean space described by the new coordinates is the embedded space of the manifold.

Here we introduce two methods. They are based on different measurement of distances defined between sampled points on the manifold. These two methods are discussed and compared using the toy example above. As fMRI functional maps are usually very noisy, we further discuss the influence of noise on the parcellation with the manifold.

Isomap

Isomap is used for computing low-dimensional embedding coordinates of high-dimensional data points on the manifold by viewing the data sets as a weighted

graph. Each data point is a vertex on the weighted graph. The geodesic distance between any two points on the manifold is estimated by the shortest path between these two vertices on the graph. The green line in Figure 3.7 shows an example of this estimation. MultiDimensional Scaling (MDS) is then applied to calculate the embedding coordinates. The Isomap algorithm can be described in three steps.

1. *Defining neighbourhood graph.* Tenenbaum et al. [2000] introduced two ways of constructing the neighbourhood graph. In one approach, two data points i and j are connected if the distance between them $d(i,j)$ is closer than ϵ. The resulting Isomap is an ϵ-Isomap. In the other approach, two data points i and j are connected if i is one of the K nearest neighbours of j. The resulting Isomap is a K-Isomap. On the constructed graph, the weight of the edge between i and j is $d(i,j)$.

   In our case, as the data points $\mathbf{v}(i)$ are evenly sampled in 3-dimensional Euclidean space, the connectivity of the graph is defined on the 3-dimensional space of $[x(i)\ y(i)\ z(i)] \in \mathbb{R}^3$. The spatial distance between two voxels $\mathbf{v}(i)$, $\mathbf{v}(j)$ is defined as:

$$d_s\Big(\mathbf{v}(i),\mathbf{v}(j)\Big) = \sqrt{\Big(x(i)-x(j)\Big)^2 + \Big(y(i)-y(j)\Big)^2 + \Big(z(i)-z(j)\Big)^2} \quad (3.3.4)$$

   The commonly used 6-connectivity, 18-connectivity or 26-connectivity in 3-dimensional space can be used. (However, due to the influence of the noise discussed in Section 3.3.1, a large neighbourhood is not suggested.)

   The weights on the graph are defined as $d_f(\mathbf{v}(i),\mathbf{v}(j))$, the functional difference between $\mathbf{v}(i)$ and $\mathbf{v}(i)$. For instance, when using PLS as the measurement

of functional behaviours and spatial 6-connectivity to construct the graph, the weights can be defined as:

$$d_f\Big(\mathbf{v}(i),\mathbf{v}(j)\Big) = \begin{cases} \sqrt{\sum_{k=1}^{n} \frac{\big(r(i)_k - r(j)_k\big)^2}{b_k}} & \text{if } d_s{=}1, \\ \infty & \text{otherwise} \end{cases}, \tag{3.3.5}$$

where, $r_k(i)$ is the covariance between the time course on voxel $[x(i)\ y(i)\ z(i)]$ and the k^{th} PLS latent variable as introduced in equation 3.2.18, and b_k is the k^{th} element on the diagonal of matrix $\mathbf{B}$ in equation 3.2.13.

2. *Calculating geodesic distances on the graph.* In Isomap, the distance on the manifold is estimated as:

$$\widetilde{d}\Big(\mathbf{v}(i),\mathbf{v}(j)\Big) = \min_{p} \sum_{i} d\Big(f(i), f(i+1)\Big)^2, \tag{3.3.6}$$

where, p is a sequence of points of length $l \geqslant 2$ with $p_1 = \mathbf{v}(i)$, $p_l = \mathbf{v}(j)$, $p_i \in \mathbf{V}, \forall i \in \{2,...,l-1\}$ and $d_f(p_i, p_{i+1}) < \infty$. Thus, these points, p, form the shortest path between $\mathbf{v}(i)$ and $\mathbf{v}(j)$ on the constructed graph in step 1. Also $\widetilde{d}\Big(\mathbf{v}(i),\mathbf{v}(j)\Big)$ is the corresponding distance on the graph which can be calculated with the Dijkstra algorithm [Dijkstra, 1959]. The green line in Figure 3.7 illustrates the shortest path between two vertices as the estimation of the geodesic distances on the manifold. $\mathbf{D}_G$ is the $V \times V$ matrix, which is defined as $\mathbf{D}_G(i,j) = \widetilde{d}\Big(\mathbf{v}(i),\mathbf{v}(j)\Big)$. Thus, $\mathbf{D}_G = \mathbf{D}'_G$. In addition, if $\exists\ k$, $\mathbf{D}_G(k,l) = \infty, \forall l \in \{1,2,...V\}$, the k^{th} row and column of $\mathbf{D}_G$ should be removed, because $v(k)$ is not connected to any other voxel on the graph.

3. *Constructing a lower dimension embedding.* In this step, in order to embed the data into lower dimensional Euclidean space, MultiDimensional Scaling (MDS)

[Borg and Groenen, 2005] is applied to $\mathbf{D}_G(k,l)$, so that the estimated distances on the manifold are preserved. Let $\mathbf{y}_i$ be the new coordinates of $\mathbf{v}(i)$ in the lower d-dimensional Euclidean space and $\mathbf{Y}$ be the set that includes all $\mathbf{y}_i$. In MDS, the new coordinates are chosen to minimize

$$E = \sum_{i,j} \Big(||\mathbf{y}_i - \mathbf{y}_j||^2 - \mathbf{D}_G(i,j) \Big). \tag{3.3.7}$$

Let

$$\delta(\mathbf{D}_G) = -\mathbf{H}\mathbf{D}_G\mathbf{H}' \tag{3.3.8}$$

where, $\mathbf{H}$ is $V \times V$ matrix with each element

$$H(i,j) = \begin{cases} 1 - 1/v & \text{if } i = j, \\ -1/v & \text{otherwise.} \end{cases} \tag{3.3.9}$$

Singular Value Decomposition (SVD) is then applied to $\delta(\mathbf{D}_G)$. As $\delta(\mathbf{D}_G)$ is symmetric, $\delta(\mathbf{D}_G)$ can be decomposed as follow:

$$\delta(\mathbf{D}_G) = \mathbf{U}\mathbf{\Sigma}\mathbf{U}', \tag{3.3.10}$$

where the elements Σ_{ii} in the diagonal of $\mathbf{\Sigma}$ are the Singular Values and the columns of $\mathbf{U}$, written as $\mathbf{u}_i$, are the singular vectors. Thus, the new coordinates in the embedded d-dimensional Euclidean space are:

$$\mathbf{y} = [\Sigma_{11}\mathbf{u}_1 \quad \Sigma_{22}\mathbf{u}_2 \quad \ldots \quad \Sigma_{dd}\mathbf{u}_d]'. \tag{3.3.11}$$

Following this, clustering can be applied to the new d-dimensional space. Figure 3.8 shows the Gaussian data (the left figure) and the data embedded in 3-dimensional Euclidean space with Isomap.

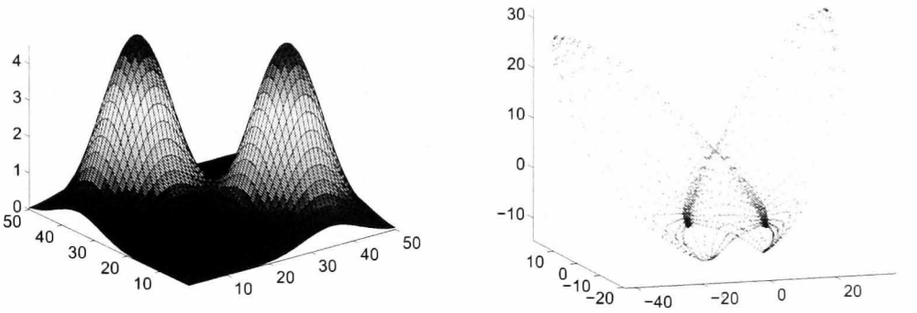


Figure 3.8: Double Gaussian data and the data embedded in 3-dimensional Euclidean space with Isomap.

Diffusion Map

In Diffusion Map, the main goal is to find the low dimensional coordinates of data points on a 3-dimensional manifold embedded in an n-dimensional space. Unlike Isomap, which uses the shortest path to estimate the distance on the manifold, Diffusion Map builds random walks on the data set, based on the connectivity and similarity between the data points. It calculates the *diffusion distance* as an estimate of the distance on the manifold.

Given a data set $\mathbf{V} = \{\mathbf{v}(1),\ \mathbf{v}(2),\ ...,\ \mathbf{v}(V)\}$, a kernel $k : \mathbf{V} \times \mathbf{V} \rightarrow \mathbb{R}$ could be defined so that it satisfies:

- k is symmetric: $k\Big(\mathbf{v}(i), \mathbf{v}(j)\Big) = k\Big(\mathbf{v}(j), \mathbf{v}(i)\Big)$,
- k is positivity preserving $k\Big(\mathbf{v}(i), \mathbf{v}(j)\Big) \geqslant 0$.
- k represents the spatial connectivity of voxels by setting:

$$k\Big(\mathbf{v}(i), \mathbf{v}(j)\Big) = 0, \forall d_s\Big(\mathbf{v}(i), \mathbf{v}(j)\Big) > \epsilon \tag{3.3.12}$$

where d_s is the same as defined in equation 3.3.4 and ϵ is the threshold for constructing spatial connectivity. It is similar as the ϵ in ϵ-Isomap. This kernel k represents the similarity between any two voxels in the dataset $\mathbf{V}$. As in Isomap, the voxels can be considered as the nodes of a weighted symmetric graph. The weight and connectivity of the graph is defined by k. For instance, isotropic diffusion could be defined as:

$$k\Big(\mathbf{v}(i),\mathbf{v}(j)\Big) = \begin{cases} e^{-d_f\left(\mathbf{v}(i),\mathbf{v}(j)\right)^2/\delta^2} & \text{if } d_s\Big(\mathbf{v}(i),\mathbf{v}(j)\Big)=1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3.13}$$

After defining the graph by (**V**,k), we can construct a reversible Markov Chain on **V**. Setting the degree of a vertex in graph **V** as

$$d\Big(\mathbf{v}(i)\Big) = \sum_{j=1}^{V} k\Big(\mathbf{v}(i),\mathbf{v}(j)\Big), \tag{3.3.14}$$

a new kernel is defined as:

$$p\Big(\mathbf{v}(i),\mathbf{v}(j)\Big) = \frac{k\Big(\mathbf{v}(i),\mathbf{v}(j)\Big)}{d\Big(\mathbf{v}(i)\Big)}. \tag{3.3.15}$$

The new matrix is not symmetric. However, due to the constraint,

$$\sum_{j=1}^{V} p\Big(\mathbf{v}(i),\mathbf{v}(j)\Big) = 1, \forall i, \tag{3.3.16}$$

p can be considered as the transition kernel of a Markov chain on **V**. Thus, we can define a row-normalized diffusion matrix, **P** , with each element as:

$$\mathbf{P}(i,j) = p\Big(\mathbf{v}(i),\dot{\mathbf{v}}(j)\Big). \tag{3.3.17}$$

Each element $\mathbf{P}(i,j)$ represents that given a random walk defined on **P**, if the current state is i, the next state is j with probability $\mathbf{P}(i,j)$. Given a time series $t = 1, 2, \ldots$, we could get a series of kernel $\mathbf{P}^t$. Denoting the element in matrix $\mathbf{P}^t$ as $\mathbf{P}_t(i,j)$, for

a random walk defined on $\mathbf{P}$, given the current state i, after t times of transition, the possibility of that the state turns to j is $\mathbf{P}_t(i,j)$. Denoting π a column vector, in which each element $\pi(i)$ is the stationary distribution of the state $\mathbf{v}(i)$ in the Markov Chain, then

$$\lim_{t\to\infty} \mathbf{P}_t(i,j) = \pi(j). \tag{3.3.18}$$

We define a new symmetric matrix $\mathbf{A}$, with each element $a(i,j)$ in $\mathbf{A}$ as

$$a(i,j) = \frac{\sqrt{\pi(i)}}{\sqrt{\pi(j)}}\mathbf{P}(i,j). \tag{3.3.19}$$

The elements of set $\{\lambda_l\}_{l\geqslant 1}$ are the eigenvalues of $\mathbf{A}$. They satisfy the condition $|\lambda_1| \geqslant |\lambda_2|, \ldots$. The corresponding eigenvectors are $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots$. After that, we have

$$\mathbf{P}_t(i,j) = \sum_{l\geqslant 0} \lambda_l \Big(\frac{\boldsymbol{\phi}_l(i)}{\sqrt{\pi(i)}}\Big)\Big(\boldsymbol{\phi}_l(i)\sqrt{\pi(j)}\Big). \tag{3.3.20}$$

Therefore, the *diffusion distances* at time t, $\{\mathbf{D}_t\}_{t\in\mathbb{N}}$ are:

$$\mathbf{D}_t(i,j) = \sum_{k=1}^{V} \frac{1}{\pi(k)}\Big(\mathbf{P}_t(i,k) - \mathbf{P}_t(j,k)\Big) \tag{3.3.21}$$

$$= \sqrt{\sum_{l\geqslant 1} \lambda_l^{2t}\Big(\boldsymbol{\psi}_l(i) - \boldsymbol{\psi}_l(j)\Big)^2}, \tag{3.3.22}$$

where, $\boldsymbol{\psi}_l(i) = \boldsymbol{\phi}_l(i)/\pi(i)$. The detailed proof of equation 3.3.22 can be found in Coifman and Lafon [2006]. Finally, we can embed diffusion distances into a lower d-dimensional Euclidean space by setting the new coordinated $\mathbf{y}_t(i)$ as:

$$\mathbf{y}_t(i) = [\lambda_1^t\boldsymbol{\psi}_1(i) \quad \lambda_2^t\boldsymbol{\psi}_2(i) \quad \ldots \quad \lambda_d^t\boldsymbol{\psi}_d(i)]'. \tag{3.3.23}$$

The new coordinates embedded in the d-dimensional Euclidean space are termed *diffusion coordinates*.

In brief, given matrix $\mathbf{P}$ and vector π, Diffusion Map can be calculated in the following steps:

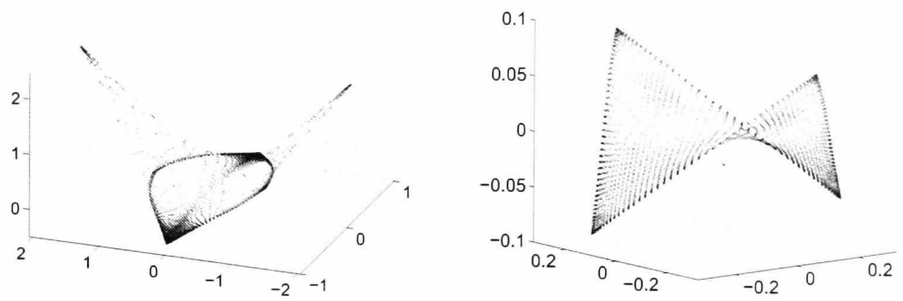


Figure 3.9: Double Gaussian data embedded in a 3-dimensional Euclidean space using Diffusion Map with $\delta^2 = 0.05$ (left) and $\delta^2 = 1$ (right). For both of them diffusion time $t = 2048$.

1. *Calculate matrix* $\mathbf{A}$

$$\mathbf{A} = \widehat{\mathbf{\Pi}}\mathbf{P}\widetilde{\mathbf{\Pi}}, \tag{3.3.24}$$

where, $\widehat{\mathbf{\Pi}}$ and $\widetilde{\mathbf{\Pi}}$ are diagonal matrices with $\widehat{\mathbf{\Pi}}(i,i) = \sqrt{\pi(i)}$ and $\widetilde{\mathbf{\Pi}}(i,i) = 1/\sqrt{\pi(i)}$.

2. *Apply SVD on* $\mathbf{A}$

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}'. \tag{3.3.25}$$

3. *Calculate diffusion coordinates*

$$\mathbf{y}_t(i) = [\frac{\Sigma_{22}^t \mathbf{u}_2(i)}{\sqrt{\pi(i)}} \quad \frac{\Sigma_{33}^t \mathbf{u}_3(i)}{\sqrt{\pi(i)}} \quad \ldots \quad \frac{\Sigma_{dd}^t \mathbf{u}_d(i)}{\sqrt{\pi(i)}}]', \tag{3.3.26}$$

where, Σ_{ii} are the elements on the diagonal of $\mathbf{\Sigma}$. As $\Sigma_{11}^t \mathbf{u}_1(i)/\sqrt{\pi(i)}$ is constant $\forall i$, it is omitted here.

If, using a Gaussian kernel, there are two parameters in Diffusion Map: diffusion time t and the kernel width δ. The diffusion time t is easier to select. It can be

explained in two ways. First, there is a wide range for t. When t is in this range, the results are similar. For instance, for the double Gaussian data, the results are almost the same $\forall t \in (1000, 4000)$. Secondly, it can be seen from equation 3.3.26 that, after the SVD decomposing $\mathbf{A}$, computationally, it is very easy to calculate diffusion coordinates with different diffusion times.

On the other hand, the kernel is an important issue in Diffusion Map. There is much discussion on the selection of different isotropic and anisotropic diffusion kernels for different problems[Coifman and Lafon, 2006; Yen et al., 2009; Yu et al., 2008]. The Gaussian kernel is the most commonly used isotropic kernel. Here, we use the double Gaussian data in Figure 3.8 to demonstrate how to choose the kernel width δ of the Gaussian kernel.

We first show the results of Diffusion Map with two different kernel widths $\delta^2 = 0.05$ and $\delta^2 = 1$ in Figure 3.13. The toy data is embedded into a 3-dimensional Euclidean space with the kernel presented in equation 3.3.13. When using $\delta^2 = 0.05$, the distances on the manifold are well presented in the embedded Euclidean space. After being embedded in the Euclidean space, the two peaks in the Gaussian data are far from each other and from the rest of the data points. However, when $\delta^2 = 1$ the embedding loses the structure of the manifold. The two peaks cannot be found the embedded Euclidean space. This can be explained by the fact that when δ is very large, transition probability in the random walk $\mathbf{P}(i, j)$ is the same for all j. Therefore, the diffusion matrix can no longer represent the local geometry of the manifold.

On the other hand, when δ is too small, transition probability $\mathbf{P}(i, j)$ is very sensitive

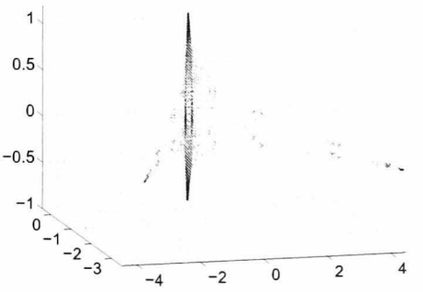


Figure 3.10: First three diffusion coordinates when $\delta^2 = 0.03$

to $d_f(\mathbf{v}(i), \mathbf{v}(j))$. Therefore, two peaks are embedded to points that are very far from the rest of the data. The corresponding parcellation results would be more functional than spatial. The embedded data is shown in Figue 3.10.

The influence of noise on parcellation with the manifold

For the double Gaussian toy example, the above methods effectively embedded the manifold into Euclidean space. However the toy example used above and by Thirion et al. [2006] is sampled under very low noise. With the finger tapping data, Figure 3.11 shows a slice of GLM parameter map and the corresponding manifold, based on which parcellation was made [Thirion et al., 2006]. It is obvious that the map is not smooth.

The topological stability of the Isomap algorithm under noise has been discussed before. Tenenbaum et al. [2000] proposed and tested the Isomap algorithm on a Swiss Roll data without any noise. Balasubramanian and Schwartz [2002] questioned the stability of this algorithm under noise. Later, Tenenbaum et al. [2002] explained that, if we carefully choose the connection of the graph (for instance, optimizing the pa-

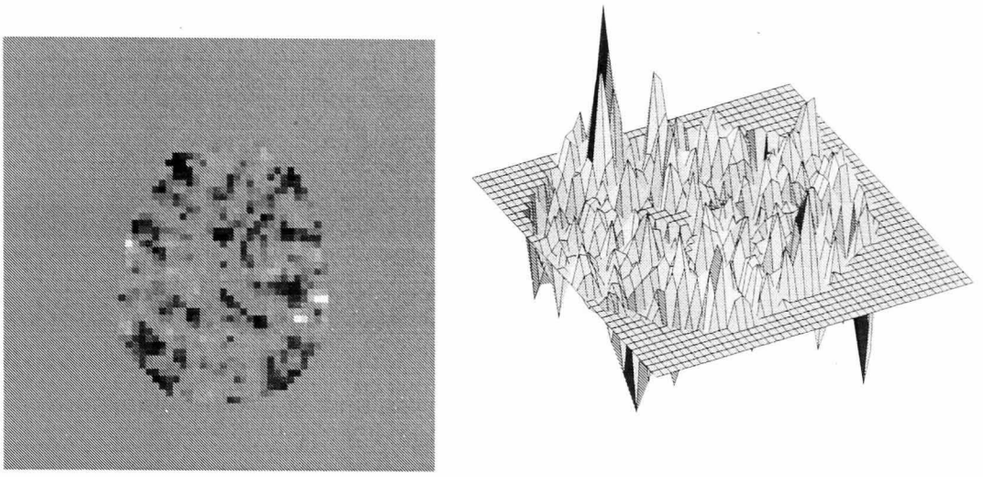
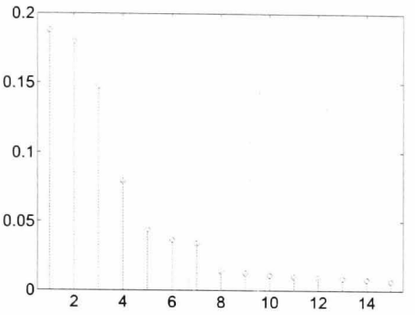


Figure 3.11: Illustration of noise level on the manifold of a GLM parameter map.

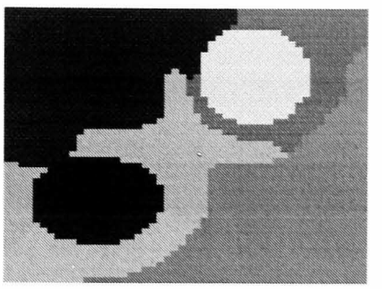
rameter ϵ in the ϵ-Isomap), the algorithm is still reliable. However, in our case, this is not an effective way of dealing with the noise.

Here, with the double Gaussian toy example, we argue that with parcellation methods based on a manifold, noise influences the results in at least two ways: the dimension of the embedded Euclidean space and the spatial connectivity of the parcels.

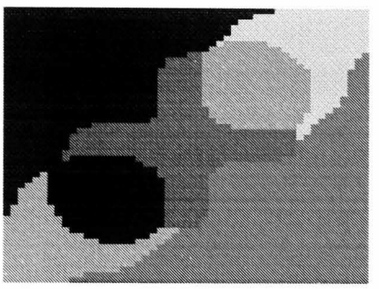
Tenenbaum et al. [2000] and Thirion et al. [2006] have suggested that the dimension of the embedded Euclidean space should be the dimension of the manifold. Here, we use the double Gaussian toy example to show a different case. The Isomap spectrum in Figure 3.12.1 indicates that the dimension of the embedded Euclidean space should be at least 3. This is also supported by Figure 3.12.2 – 3.12.4. When increasing the dimension of the embedded space from 2 to 3, there is a significant improvement in the parcellation results. Continuing to increase the number of dimensions to 8, there is very little difference in the results. This is because of the structure of the manifold. For the Swiss Roll data, it is enough to use 2-dimensional Euclidean



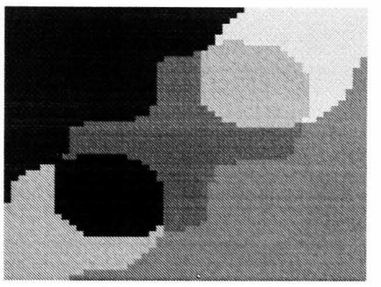
3.12.1: The spectrum of Isomap



3.12.2: Parcellation result with first two embedded dimensions



3.12.3: Parcellation result with first three embedded dimensions



3.12.4: Parcellation result with first eight embedded dimensions

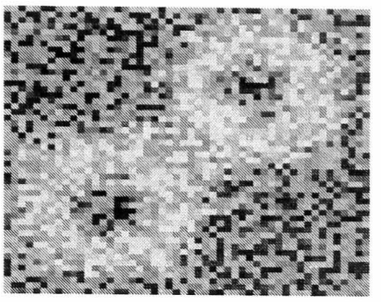
Figure 3.12: Parcellation results from toy data in different embedded spaces

space to represent the structure of the manifold. However, it is not the case for the double Gaussian data.

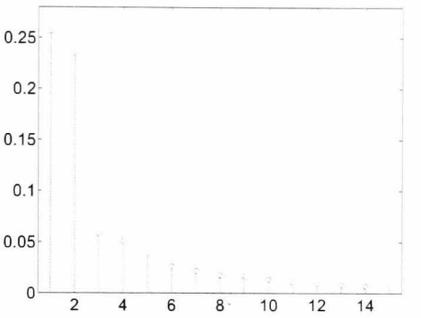
In order to examine the performance of Isomap under noise, we apply Isomap to a noisy double Gaussian data generated by:

$$f(x,y) = 5 \cdot (e^{-\frac{(x-15)^2+(y-35)^2}{90}} + e^{-\frac{(x-35)^2+(y-15)^2}{90}}) + 3\epsilon. \tag{3.3.27}$$

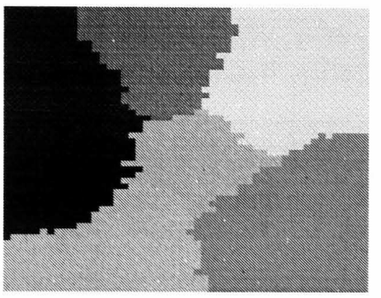
where, $\epsilon \sim \mathcal{N}(0,1)$. The data is shown as an image in Figure 3.13.1. Both the Isomap spectrum (Figure 3.13.2) and the parcellation results (Figure 3.13.3 – 3.13.4) indicate that 2-dimensional Euclidean space is enough to embed the manifold. However the



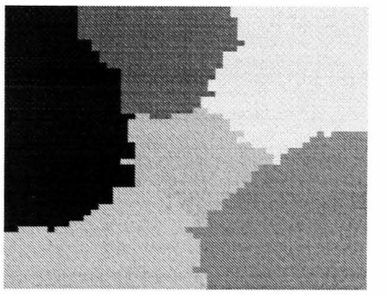
3.13.1: The double Gaussian data with noise



3.13.2: The spectrum of Isomap



3.13.3: Parcellation result with first two embedded dimensions



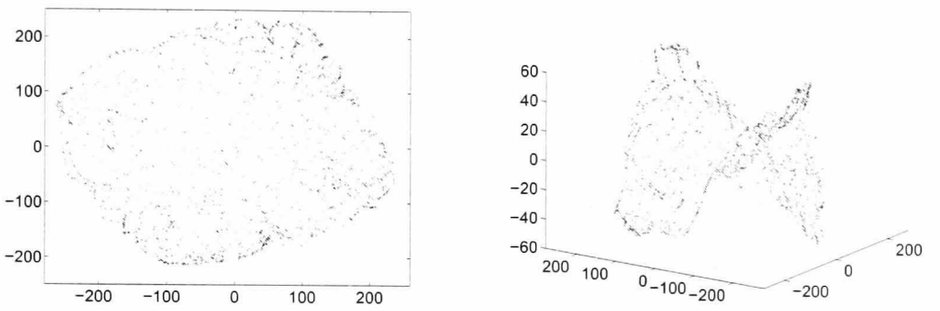
3.13.4: Parcellation result with first eight embedded dimensions

Figure 3.13: Parcellation results from toy data with noise

parcellation results are much worse than the ones in Figure 3.12. The two peaks of the double Gaussian model cannot be presented. The parcellation is almost purely spatial.

In order to explain these results, in Figure 3.14, we show the data embedded in 2-dimensional and 3-dimensional Euclidean space. Comparing Figure 3.14 and Figure 3.8, we can see that, without noise, at least 3-dimensional Euclidean space is needed to embed the data. When high-level noise is added on the data, 2-dimensional Euclidean space is enough. The new coordinates cannot however represent the struc-

ture of the manifold. The reason is that, when noise is added to the data rather than on the manifold, the data is scattered around it. The estimation of the geodesic distance is no longer accurate. Therefore, the Isomap cannot present the structure of the manifold.

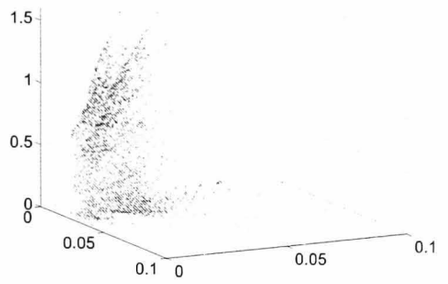


3.14.1: Embedding on 2-dimensional space 3.14.2: Embedding on 3-dimensional space

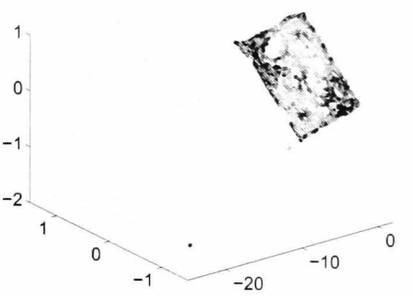
Figure 3.14: Embeddings of the toy data with noise

It is claimed in Coifman and Lafon [2006] that Diffusion Map could a provide more robust estimation of the geodesic distance. Therefore, we apply Diffusion Map to the noisy double Gaussian data. Different kernel widths are used to find the best embedding. However, the results are disappointing.

Figure 3.15 shows the results of Diffusion Map with two different kernel widths. The kernel width $\delta^2 = 0.05$ is a suitable kernel width for the data without noise. However when noise is added to the same data, embedding results become very unreliable. Some points (in this case, 2% of all sampled points) become outliers in the embedded Euclidean space. They are very far from other embedded data. After removing the outliers, the data embedded in the 3-dimensional space is shown in Figure 3.15.1. Apparently, the structure of the manifold can no longer be found.



3.15.1: Embedding on 3-dimensional space with $\delta^2 = 0.05$



3.15.2: Embedding on 3-dimensional space with $\delta^2 = 0.4$

Figure 3.15: Embeddings of the toy data with noise

From the above discussion, we can conclude that the parcellation methods based on the manifold are effective only when the manifold is smooth. When there is a high-level of noise in the data, the low dimensional embedding of the data cannot represent the structure of the manifold. The parcellation results are similar to the results of clustering based on the spatial location of the voxels. In order to deal with this problem, we propose two methods: *adaptive smoothing* and *Aggregation and Boundary Competition*. We first introduce adaptive smoothing as a preprocessing step for manifold based parcellation. After that, in Section 3.3.2, we introduce a seed based parcellation method, the Aggregation and Boundary Competition method, which could give fast and reasonable parcellation results.

Adaptive Smoothing for parcellation based on the manifold

As discussed above, parcellation based on the manifold is effective only on a smooth manifold. However, there is high-level noise in fMRI data set. The noise in the

fMRI data could decrease the functional homogeneity of the final parcellation results. Here, we propose an adaptive smoothing method to deal with this problem. We first introduce this method for the general manifold embedding problem. After that, using the double Gaussian toy example, we discuss adaptive smoothing for parcellation.

Given that $\mathbf{x}_i = [x_1(i)\ x_2(i)\ ...x_n(i)]'$, where $i = 1, 2, ..., M$, are M points evenly sampled from a d-dimensional manifold embedded in n-dimensional space with noise. A Gaussian filter is implemented on this data, so that the smoothed data point $\widehat{\mathbf{x}}_i$ is

$$\widehat{\mathbf{x}}_i = \sum_{j=1, j\neq i}^{M} \left(\frac{1}{G_i} e^{-\frac{||\mathbf{x}_j - \mathbf{x}_j||^2}{2\sigma^2}} \mathbf{x}_j\right), \tag{3.3.28}$$

where,

$$G_i = \sum_{j=1}^{M} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}}. \tag{3.3.29}$$

As the Gaussian filter is isotropic, there is only one parameter σ in this smoothing. Therefore, the next problem is how to choose σ. In order to solve this problem, we propose a method similar to the k-fold cross validation in statistics. The data is first randomly divided into k equal (or roughly equal, when there is a remainder in M/k) parts, $C_1, C_2 ... C_k$. Next, the error E_k is defined as:

$$E_k = \sum_{\mathbf{x}_i \in C_k} \left(\mathbf{x}_i - \sum_{\mathbf{x}_j \notin C_k} \left(\frac{1}{G_i} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}} \mathbf{x}_j\right)\right), \tag{3.3.30}$$

where,

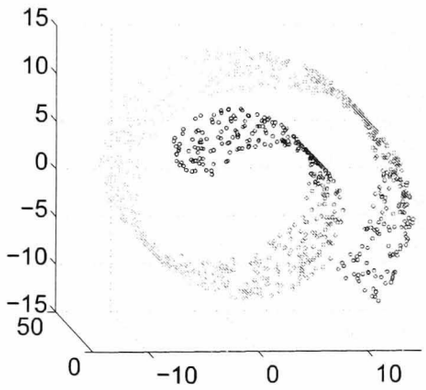
$$G_i = \sum_{\mathbf{x}_j \notin C_k} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}}. \tag{3.3.31}$$

After that, the cross validation error E_{CV} is defined as:

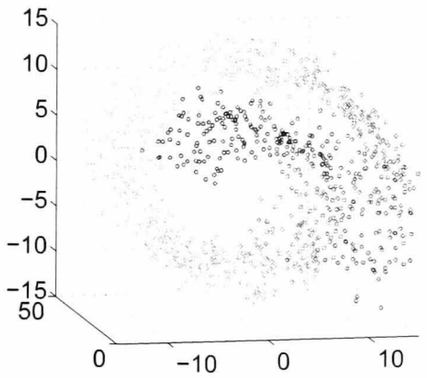
$$E_{CV} = \sum_k E_k \tag{3.3.32}$$

Along with the increasing of σ, E_{CV} first decrease and then increase. We choose σ as corresponding to the minimal E_{CV} as the optimal parameter.

We first use 'Swiss Roll' data set as an example to show the performance of the above smoothing method.



3.16.1: Swiss Roll data.



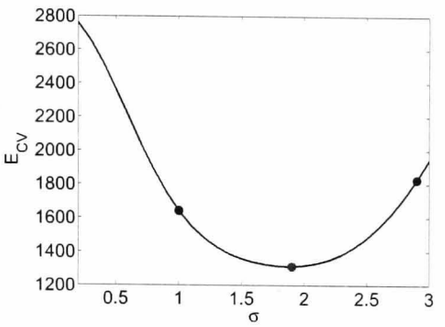
3.16.2: Swiss Roll data with noise.

Figure 3.16: Embeddings of the toy data with noise

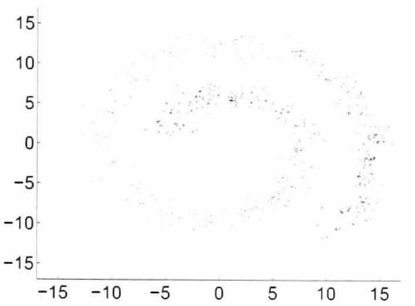
Figure 3.16.1 shows the original Swiss Roll data. Given that, $\mathbf{x}_i = [x_i(1)\ x_i(2)\ x_i(3)]$ is a point in the data set, Gaussian noise is added to the data by

$$xn_i(i) = x_i(i) + \epsilon, \text{ where } \epsilon \sim \mathcal{N}(0,1) \tag{3.3.33}$$

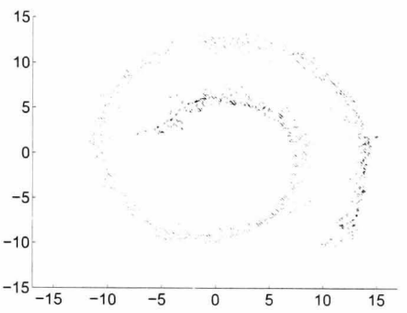
Figure 3.16.2 shows the data with Gaussian noise. With different kernel width σ, E_{CV} is calculated. Here, we use each sample as one part of the cross-validation. The corresponding E_{CV} are shown in Figure 3.17.1. In order to show the smoothing results with different kernel width, we use three different values for σ (corresponding to the three dots in Figure 3.17.1) to smooth the data. The smoothing results are shown in Figure 3.17.2 – 3.17.4. From these figures, it can be seen that $\sigma = 1.9$ gives the best result.



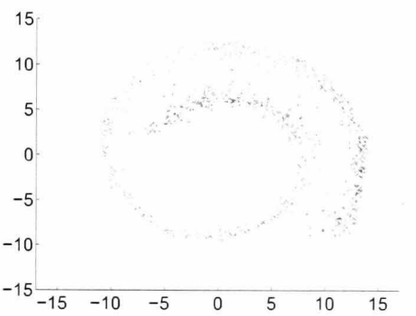
3.17.1: The cross validation Error.



3.17.2: Smoothing result with $\sigma = 1$.



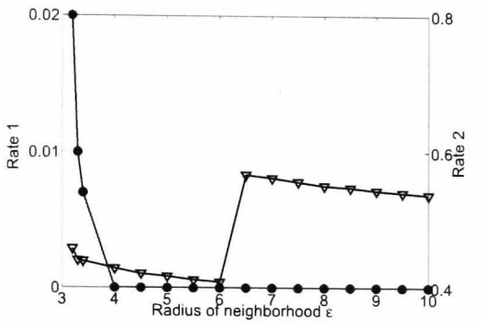
3.17.3: Smoothing result with $\sigma = 1.9$.



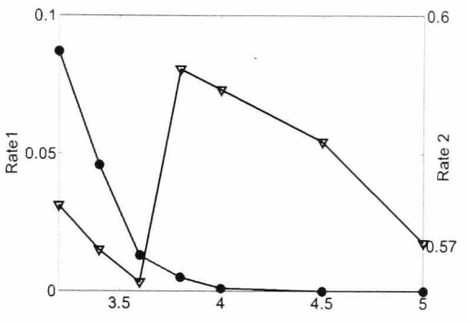
3.17.4: Smoothing result with $\sigma = 2.9$.

Figure 3.17: Smoothing of Swiss Roll data.

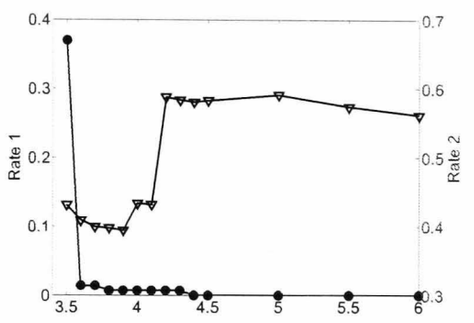
In order to further prove the validity of this smoothing method, we use the graphs in Figure 3.18 to illustrate the improvement brought by the adaptive smoothing. Figure 3.18.1 shows a commonly used graph to select the size of neighbourhood for Isomap [Balasubramanian and Schwartz, 2002; Tenenbaum et al., 2000]. It can also be used to evaluate the results of embedding. In this graph, the horizontal axis represents the radius ϵ in ϵ-Isomap. The ϵ determines the neighbourhood size in Isomap. The vertical axis presents two different rates: 'Rate 1', which is marked with a blue line and dots, represents a fraction of the points that are not included in the largest connected component of the neighbourhood graph; 'Rate 2', which is marked with a black line and triangles, represents the fraction of the variance in



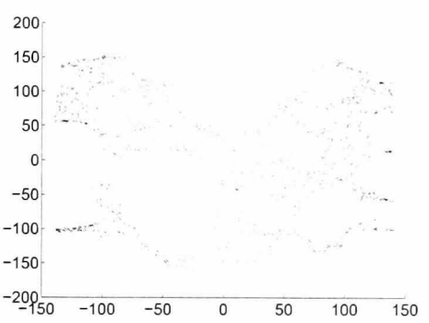
3.18.1: Cost function for noiseless Swiss Roll data.



3.18.2: Cost function for Swiss Roll data with noise.



3.18.3: Cost function for smoothed noisy Swiss Roll data.



3.18.4: The embedding of smoothed data

Figure 3.18: Results of tests on Swiss Roll data

geodesic distance estimates not accounted for in the Euclidean embedding. The desired results should have both low 'Rate 1' and 'Rate 2'. However when the neighbourhood size is small, there must be some points that are connected to the neighbourhood graph, and thus not included in the Euclidean embedding. When the neighbourhood size is too large, estimated geodesic distances are not sufficiently represented by a 2D Euclidean embedding. This problem leads to high 'Rate 2'. The optimal neighbourhood size is determined on the basis of a trade-off between these two rates.

Figure 3.18.1 shows the results from the noiseless Swiss Roll data. The optimal ϵ locates between 4 and 6. When ϵ is in this range, all the data points are connected to the neighbourhood graph. 'Rate 2' is also very low. Figure 3.18.1 shows the result when noise is added to the data. The noise level used here is higher than the one in Balasubramanian and Schwartz [2002]. Therefore, this graph shows that there is no optimal range for ϵ. Both 'Rate 1' and 'Rate 2' are very high.

After adaptive smoothing, the results of the smoothed data are shown in Figure 3.18.3. Compared to the results in Figure 3.18.2, the smoothing decreases both 'Rate 1' and 'Rate 2'. In addition, it broadens the stable range of neighbourhood size, in which Isomap could yield reasonable results. The 2D embedding of the smoothed Swiss Roll data is shown in Figure 3.18.4.

After that, we use the double Gaussian toy examples to test the performance of the adaptive filter on the parcellation problem. First, using a double Gaussian we generate 10 sets of data, with different levels of noise with the equation below:

$$f(x,y) = 5 \cdot (e^{-\frac{(x-15)^2+(y-35)^2}{90}} + e^{-\frac{(x-35)^2+(y-15)^2}{90}}) + p\epsilon, \tag{3.3.34}$$

where, $\epsilon \sim \mathcal{N}(0,1)$ and $p = 0.5,\ 1,\ 1.5,\ 2,\ ...,\ 5$. The data, with different levels of noise, is shown in Figure 3.19. The adaptive smoothing method proposed above is implemented to smooth the data. The results are shown in Figure 3.20.

After adaptive smoothing, the smoothed data is embedded into 3D Euclidean space with Isomap. Figure 3.21 shows the embedding and parcellation of the data generated with $p = 5$.

As demonstrated above, the manifold embedding methods are not reliable when

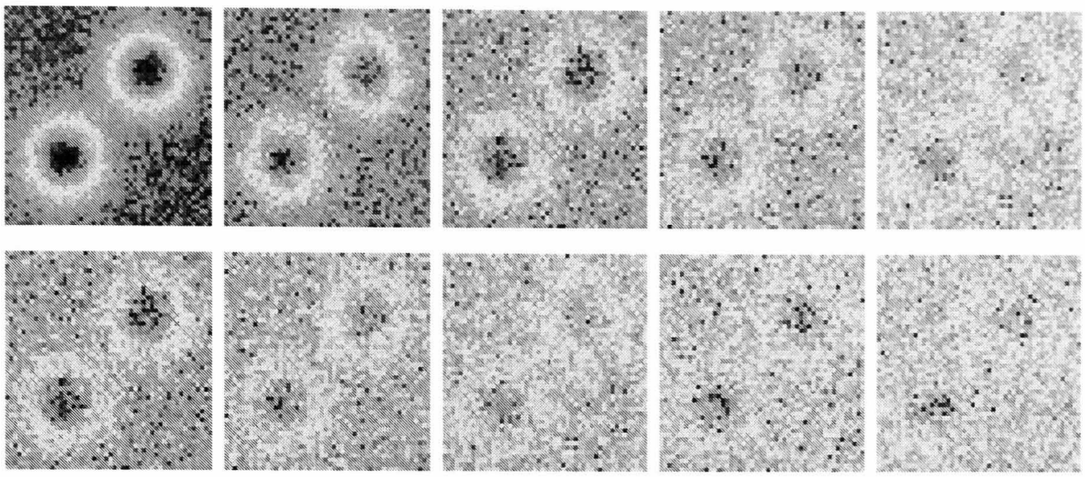


Figure 3.19: Toy data with different levels of noise.

the data is sampled under high-level noise. We use two examples to show that adaptive smoothing could effectively improve the results of manifold embedding techniques. The smoothing proposed here is different from the one used in classic fMRI group analysis. In these analyses, a large kernel width is used to increase the overlapping of the activated regions in different subjects. In contrast, the reason we use smoothing here is to make the manifold embedding techniques applicable. Therefore, as shown in Chapter 5, the automatic selected kernel widths are much smaller than the ones commonly used in classic fMRI data processing.

3.3.2 Aggregation and Boundary Competition

In this section, we propose another spatially constrained clustering method for parcellation. Unlike the methods introduced above, this method is based on seed voxels and parameter-controlled aggregation. The algorithm can be divided into two parts: *Aggregation* and *Boundary Competition*. In *Aggregation*, voxels are aggregated to neighbouring seed voxels. Following that, in *Boundary Competition*, the voxels that

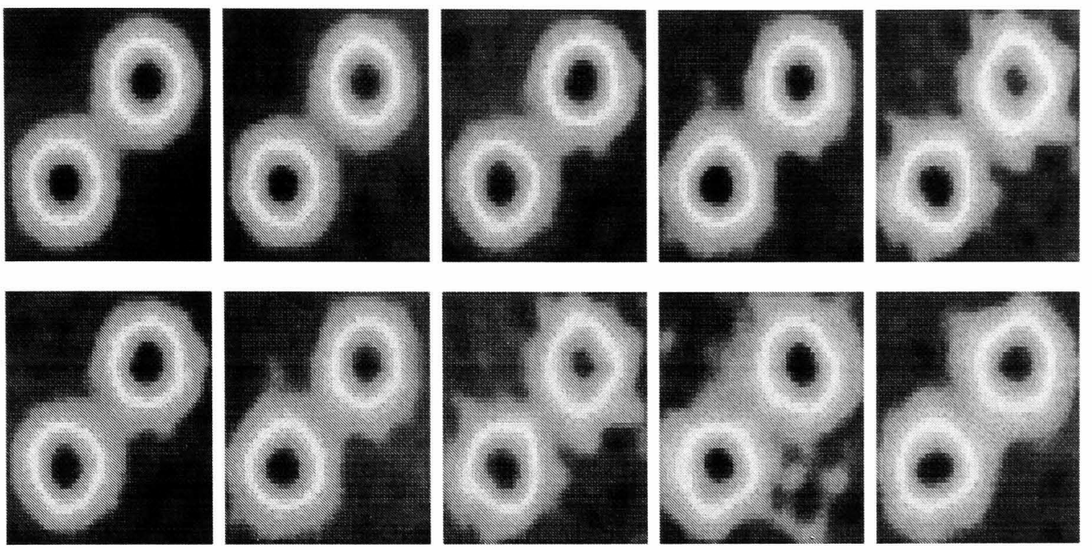


Figure 3.20: Toy data with different levels of noise.

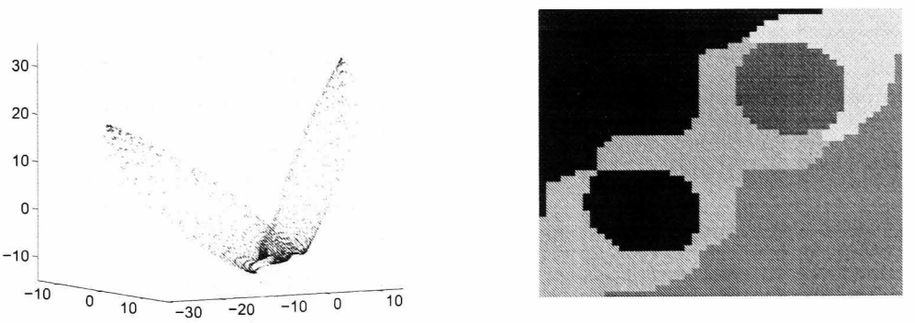


Figure 3.21: Embedding and parcellation on smoothed data.

locate at the boundary of different parcels are reassigned to improve the functional homogeneity and the spatial connectivity of the parcellation results.

We still use the notation in equation 3.3.1 to represent the fMRI data. The vector set $\{\mathbf{v}(i), i = 1, 2, ..., V\}$ is the set of all the voxels to be parcellated. The number i is the index of the voxels and V is the number of the voxels in the vector set. In addition, we further denote the number of the parcels as P. And $\mathbf{P}_i$ is a set of voxel indices, such that for any index $k \in \mathbf{P}_i$, the voxel $\mathbf{v}(k)$ is labelled as the ith

parcel. Therefore, we have $\mathbf{P}_i \cap \mathbf{P}_j = \emptyset, \forall i \neq j$. At the end of parcellation, we have $\cup_{j=1}^{P} \mathbf{P}_j = \{1, 2, ..., V\}$.

At the beginning of *Aggregation*, the number of the parcels P is first determined and P seed voxels are selected. The seed voxels are the prototypes of parcels. Hence, they should be spatially and functionally far from each other. The seed selection method introduced in section 3.2.3 can be used here. However the setting of radius R is different.

Here, the *Aggregation* is presented in two steps: *Setting Seeds* and *Aggregation*.

1. Setting Seeds:

   Let the indices of p seed voxels for aggregation be $S_1, S_2 ... S_p$. The first seed voxel $\mathbf{v}(S_1)$ is selected as,

   $$S_1 = \arg\max_i \left|\left|[f_1(i) \quad f_2(i) \quad ... \quad f_n(i)]\right|\right| \tag{3.3.35}$$

   The second seed voxel $\mathbf{v}(S_2)$ is selected as,

   $$S_2 = \arg\max_i \left|\left|[f_1(i) \quad f_2(i) \quad ... \quad f_n(i)]\right|\right| \tag{3.3.36}$$

   $$\text{subject to } \left|\left|[x(i) \;\; y(i) \;\; z(i)] - [x(S_1) \;\; y(S_1) \;\; z(S_1)]\right|\right| > R.$$

   And the kth seed voxel $\mathbf{v}(S_k)$ is selected as,

   $$S_k = \arg\max_i \left|\left|[f_1(i) \quad f_2(i) \quad ... \quad f_n(i)]\right|\right| \tag{3.3.37}$$

   $$\text{subject to } \left|\left|[x(i) \;\; y(i) \;\; z(i)] - [x(S_j) \;\; y(S_j) \;\; z(S_j)]\right|\right| > R, j = 1, 2, ..., i-1$$

   It continues until p seeds are selected. R should be large enough, so that for any voxel there should be a seed voxel with the distance smaller than R. On

the other hand, if R is too large, there may not be enough voxels for P seeds. Therefore, we suggest R be selected as $R^3 < V/P$.

2. Aggregation:

The step of Aggregation is similar to seed region-grow algorithms in image segmentation [Shapiro and Stockman, 2001]. Seeds $S_1, S_2 ... S_P$ are the initial state of the P parcels. Therefore, at the beginning of the parcellation, $\mathbf{P}_i = \{S_i\}$. After that, in each step of aggregation, the algorithm allocates additional voxels to $\mathbf{P}_i$. Let H be the set of all unlabelled voxels that are the neighbours of the labelled voxels:

$$H = \{i \notin \bigcup_{j=1}^{P} \mathbf{P}_j | N(i) \cap \bigcup_{j=1}^{P} \mathbf{P}_j \neq \emptyset\}, \tag{3.3.38}$$

where, $N(i)$ is the set of voxel indices. The corresponding voxels of $N(i)$ are in the neighbourhood of voxel $\mathbf{v}(i)$. The neighbourhood of a voxel could be defined as 6-connected or 28-connected in 3D space. For unlabelled voxel $\mathbf{v}(i), i \in H$, the distance between voxels $\mathbf{v}(i)$ and the neighbouring parcel $\mathbf{P}_j$ is defined as,

$$d(\mathbf{v}(i), \mathbf{P}_j) = \frac{\left|\left| [f_1(i) \;\; f_2(i) \;\; \cdots \;\; f_n(i)] - \sum_{j \in \mathbf{P}_j} \left([f_1(j) \;\; f_2(j) \;\; \cdots \;\; f_n(j)] \right) / V_j \right|\right|}{(|N(i) \bigcap \mathbf{P}_j| / |N(i)|)^{\delta}}, \tag{3.3.39}$$

where, $\mathbf{v}(i)$ is the voxel in the set H; V_j is the number of voxels in region $\mathbf{P}_j$; $|N(i) \bigcap \mathbf{P}_j|$ is the number of voxels labelled as parcel $\mathbf{P}_j$ in the neighbourhood of $\mathbf{v}(i)$ and δ is the parameter to control the aggregation. When δ is 0, the aggregation is based on functional similarity. On the other hand, when $\delta \rightarrow \infty$

the aggregation is based on spatial connectivity. At each step of the aggregation, a certain number (s_v) of voxels are labelled to the nearest parcel according to the distance defined in equation 3.3.39. These voxels should be the voxels that are closest to the parcels. Consequently, for each voxel $\mathbf{v}(i), i \in H$, we find the parcel that is nearest to this voxel by

$$\hat{j} = \arg\min_j d(\mathbf{v}(i), \mathbf{P}_j). \tag{3.3.40}$$

After that, for all $i \in H$, the distances $d(\mathbf{v}(i), \mathbf{P}_{\hat{j}})$ are ranked. The first s_v voxels associated with the smallest $d(\mathbf{v}(i), \mathbf{P}_{\hat{j}})$ are labelled as $\mathbf{P}_{\hat{j}}$.

The aggregation is repeated until all the voxels are labelled.

During the process of aggregation, there are two parameters that control the results: δ in equation 3.3.39 and the number of voxels to aggregate at each step, s_v. The parameter δ is used to control the balance between the spatial connectivity and functional homogeneity of the parcels. When $\delta = 0$, the $d(\mathbf{v}(i), \mathbf{P}_j)$ is the functional distance. The voxel may be assigned to a spatially unconnected the parcel. Thus, the resulting parcels may have very high functional homogeneity but the voxels in the same parcel may be spatially separated from each other. On the other hand, when δ is too big, the voxel may be assigned to a parcel that is spatially well connected but functionally very far. The resulting parcels may have low functional homogeneity. Intuitively, we suggest using δ between 0 and 1. The other parameter s_v controls the speed of aggregation.

After Aggregation, Boundary Competition is implemented. In Boundary Competition, the voxels whose neighbouring voxels are labelled differently to them are

considered as voxels on the boundary. B is denoted as the set of the indices of all the voxels on the boundary. For each voxel i in B, i is first marked as unlabelled. Then, the distance between $\mathbf{v}(i)$ and its neighbour parcels are calculated, as in equation 3.3.39. Following this, $\mathbf{v}(i)$ is labelled as its neighbour region with the minimal $d(\mathbf{v}(i), \mathbf{P}_j)$. This is repeated until stop criterion is met. The stop criterion could be that label of all voxels are no longer changed.

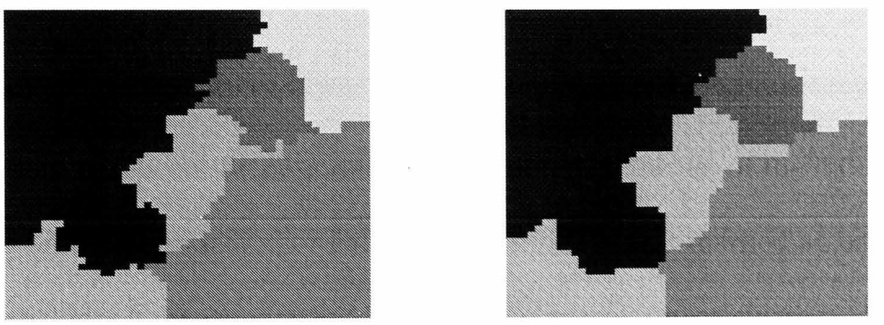


Figure 3.22: Results of Aggregation (left) and Boundary Competition

The left graph of Figure 3.22 shows the result of aggregation with $\delta = 0.2$ and $s_v = 28$. The boundary of the parcel is noisy. During boundary competition, $\delta = 1$ is used. The boundary of the parcels is more regular.

3.3.3 Discussion

In the above sections, we introduced three spatially constrained clustering methods for parcellation. Two of them were based on the manifold approach. The last one was based on aggregation and boundary competition.

In methods based on the manifold approach, the feature space is first embedded into Euclidean space, which could best represent the geodesic distances of the data

in the feature space. After that, the data is parcellated with k-mean algorithm in the embedded Euclidean space. Alternatively, in the method based on aggregation, parcels are calculated according to designed algorithms.

Both of these methods share one common point in that seed voxels have to be used. In manifold based methods, seed voxels are used in the last step, k-mean algorithm. In aggregation, seed voxels are selected first. The major difference is that, in manifold based approaches, voxels in the whole brain are allocated to the seed voxels according to the geodesic distances represented by the embeddings. Unlike this, in aggregation, voxels are parcellated according to the parameter controlled automatic algorithm.

Considering the effectiveness and reliability of these algorithms, manifold based methods are effective only when the noise level is low and the manifold is smooth. Under high level noise, even though the first eigenvalues cover most of the variance in the spectrum of Figure 3.13.2, the embeddings cannot represents the geodesic distances sufficiently. Under these circumstances, the adaptive smoothing is proposed. The result of experiments on two toy examples shows that this method could improve the embedding results and increase the reliability against noise.

On the other hand, the aggregation algorithm is controlled by two parameters. Choosing appropriate parameters could make the algorithms more robust to noise. However, this algorithm could be sensitive to different settings of parameters and seeds. Using different seed voxels and parameters on the same data, the result could be very different from each other. Thus the choice of seeds and parameters is important for aggregation. At the end of aggregation, the central voxel of each parcel

can be used as new seeds and the aggregate is iterated in order to increase the convergence of the algorithm. In our examples, due to the fact that the seed voxels are selected carefully, such iteration does not give much improvement.

Next, we compared these methods from a computational point of view. Due to the fact that calculating the embeddings of the geodesic distance needs large computation, the efficiency of the manifold embedding algorithms is not very high. Especially when compared them to the aggregation algorithms, the difference is obvious. However, embedding methods give better global parcellation results, on the condition that the manifold is smooth.

Therefore, if computational efficiency is an important requirement in the analysis, the aggregation and boundary competition method for parcellation is to be recommended. If time permits, manifold based methods could give better parcellation results. When using manifold based parcellation for 3D image parcellation, the use of at least 4 last embeddings for parcellation is suggested. If the manifold is not smooth, adaptive smoothing can be used as a preprocessing step. Otherwise, the embeddings cannot represent the geometry of the manifold.

3.4 Validation of Intra-subject Parcellation

3.4.1 Intra-parcel functional variance

The aim of parcellation is to divide the brain into functionally homogeneous regions. Therefore, the functional homogeneity of parcels is naturally a criterion by which to examine parcellation methods. Thirion et al. [2006] present intra-parcel

functional homogeneity with intra-parcel variance of the GLM parameter. However, due to the fact that there is noise as well as other artefacts in the fMRI data, the GLM parameters cannot sufficiently represent the functional behaviour of the corresponding voxels. Therefore, in this section, we propose some other methods to measure intra-parcel functional homogeneity.

Intra-parcel variance of GLM t-values

Let Nr be the number of regressors in the design matrix of GLM and $\mathbf{f}^i \in R^{Nr\times 1}$ be the vector of t-values for voxel i. Np is the number of parcels in the whole brain. For any parcel $\mathbf{P}_j, j = 1, 2, ..., Np$, the functional variance of $\mathbf{P}_j$, $v(\mathbf{P}_j)$, is:

$$v(\mathbf{P}_j) = \sqrt{\sum_{k=1}^{Nr}(\text{std}(f_k^i))^2}, \quad \text{where, } \mathbf{f}^i = [f_1^i\ f_2^i\ \dots\ f_{Nr}^i]' \text{ and } i \in \mathbf{P}_j \tag{3.4.1}$$

The mean and distribution of $v(\mathbf{P}_i)$ across all parcels is used to compare the accuracy of the parcellations.

Intra-parcel variance of PLS t-values

Given a design matrix $\mathbf{Y} \in R^{T\times Nr}$, where $\mathbf{y}_k \in R^{T\times 1}$ is the kth column of $\mathbf{Y}$, instead of using the matrix $\mathbf{D}$ as in section 3.2.5, the regressor $\mathbf{y}_k$ is used to calculate latent variables as in section 3.2.5. If r_k is the covariance between the fMRI time series and the first latent variable, then

$$t_k = \frac{r_k\sqrt{T-2}}{1-r_k^2} \tag{3.4.2}$$

has a t-distribution with $T-2$ degrees of freedom. The null hypothesis of this test is that the signal of that voxel is not covariant with the PLS components. Thus, we

can generate statistical maps to represent the significance of the covariance between the signals in each voxel and the first latent variable.

3.4.2 Nearest Silhouette Coefficient

Silhouette is a method proposed in Rousseeuw [1987] to evaluate the clustering validity. In this method, each sample is given a so-called *silhouette* to represent whether the sample lies well within the cluster it is assigned to.

Given that all the N samples, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$, are clustered into P clusters, $\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_P$, and $d(\mathbf{x}_i, \mathbf{P}_j)$ is the distance between the sample $\mathbf{x}_i$ and the cluster $\mathbf{P}_j$, the silhouette coefficient for the sample $\mathbf{x}_i$ is defined as:

$$s(i) = \begin{cases} 1 - a(i)/b(i) & \text{if } a(i) < b(i) \\ 0 & \text{if } a(i) = b(i) \\ b(i)/a(i) - 1 & \text{if } a(i) > b(i) \end{cases} \tag{3.4.3}$$

where,

$$a(i) = d(\mathbf{x}_i, \mathbf{P}_k), \mathbf{x}_i \in \mathbf{P}_k \tag{3.4.4}$$

$$b(i) = \min_{j, \mathbf{x}_i \notin \mathbf{P}_j} d(\mathbf{x}_i, \mathbf{P}_j). \tag{3.4.5}$$

The distance $d(\mathbf{x}_i, \mathbf{P}_j)$ is defined as:

$$d(\mathbf{x}_i, \mathbf{P}_j) = \frac{1}{|\mathbf{P}_j|} \sum_{\mathbf{x}_j \in \mathbf{P}_j} ||\mathbf{x}_i - \mathbf{x}_j|| \tag{3.4.6}$$

The $a(i)$ is the distance between the sample $\mathbf{x}_i$ and the cluster it is assigned to, Whilst $b(i)$ is the distance between the sample $\mathbf{x}_i$ and the cluster that is the second best choice for $\mathbf{x}_i$ to be clustered to. The silhouette coefficient of that sample, $s(i)$, is calculated by comparing these two distances.

Silhouette has been successfully applied to validate the result of many different clustering problems [Kim et al., 2010b; Maulik and Mukhopadhyay, 2010; Mitra et al., 2010]. However Silhouette cannot be effective in validating the results of parcellation. In parcellation, for any voxel $\mathbf{v}(i)$, there may be a parcel that is spatially far but functionally very close to the voxel $\mathbf{v}(i)$. Although due to the spatial constraint, this parcel cannot be the second best choice for $\mathbf{v}(i)$, the Silhouette calculates $b(i)$ with that parcel. Such a situation gives a very low $s(i)$ no matter whether $\mathbf{v}(i)$ is well clustered or not. Therefore, with the same spirit of Silhouette Coefficient, we propose an adapted version of Silhouette Coefficient called *Nearest Silhouette Coefficient*.

Here, we use the same denotation in section 3.3.2. For any voxel $\mathbf{v}(i), i \in \mathbf{P}_l$, the spatial distance between voxel $\mathbf{v}(i)$ and parcel $\mathbf{P}_m, m \neq l$ is defined as:

$$d(\mathbf{v}(i), \mathbf{P}_m) = \begin{cases} \min\limits_{j \in \mathbf{P}_m} \left|\left| [x(i)\ y(i)\ z(i)] - [x(j)\ y(j)\ z(j)] \right|\right| & \text{if } \mathbf{P}_m, \mathbf{P}_l \text{ are neighbours,} \\ \infty & \text{otherwise} \end{cases} \tag{3.4.7}$$

According to the distance defined above, the parcel spatially closest to $\mathbf{v}(i)$ is considered as the the second best choice for $\mathbf{v}(i)$. Therefore, $b(i)$ is calculated as:

$$b(i) = \frac{1}{|\mathbf{P}_k|} \sum_{j \in \mathbf{P}_k} \left|\left| [f_1(i)\ \ f_2(i)\ \ \ldots\ \ f_n(i)] - [f_1(j)\ \ f_2(j)\ \ \ldots\ \ f_n(j)] \right|\right| \tag{3.4.8}$$

where,

$$k = \arg\min_{m, i \notin \mathbf{P}_m} d(\mathbf{v}(i), \mathbf{P}_m). \tag{3.4.9}$$

In addition, $a(i), i \in \mathbf{P}_l$ is defined as :

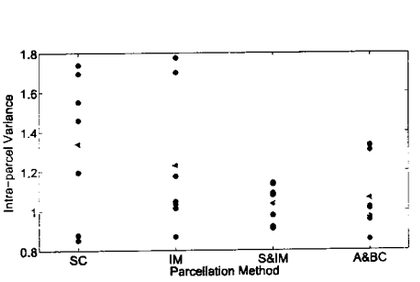
$$a(i) = \frac{1}{|\mathbf{P}_l|} \sum_{j \in \mathbf{P}_l} \left|\left| [f_1(i)\ \ f_2(i)\ \ \ldots\ \ f_n(i)] - [f_1(j)\ \ f_2(j)\ \ \ldots\ \ f_n(j)] \right|\right|. \tag{3.4.10}$$

Finally, $s(i)$ is calculated as in equation 3.4.3. The mean value of all $s(i)$ can be used to evaluate the parcellation results.

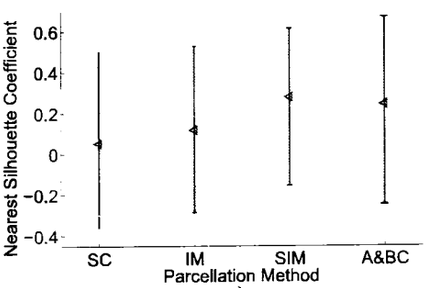
3.4.3 Results from the toy data

Using the evaluation methods proposed above and the double Gaussian toy example shown in Figure 3.13.1, we examine the parcellation methods introduced in this chapter. Here, three methods are compared: Isomap parcellation, Isomap parcellation with Adaptive Smoothing and lastly Aggregation and Boundary Competition. The parcellation result of these methods are illustrated in Figure 3.13.3, Figure 3.21, and Figure 3.22. *K*-means Clustering on the spatial coordinates of the voxels are used as a baseline. The intra-parcel variances and nearest silhouette coefficients for each parcellation result are calculated and shown in Figure 3.23.

Figure 3.23.1 shows the intra-parcel variances. The horizontal axis denotes different parcellation methods: clustering on the spatial coordinates (SC), Isomap parcellation (IM), Isomap parcellation with Adaptive Smoothing (SIM) and Aggregation and Boundary Competition (A&BC). The vertical axis represents the intra-parcel variance of each parcel. Each black dot shows the intra-parcel variance of a parcel. The mean intra-parcel variance of each parcellation method is presented with the corresponding red triangle.



3.23.1: Intra-parcel variances



3.23.2: Nearest Silhouette Coefficients

Figure 3.23: Comparison of parcellation results.

The nearest silhouette coefficients of each parcellation method are shown in Figure 3.23.2 with error bars. The red triangles and the bars illustrate the mean, the first and the third quartile of nearest silhouette coefficients of each parcellation.

From Figure 3.23, we can see that, according to both evaluation methods, the Isomap parcellation with adaptive smoothing gives the best results. The Aggregation and Boundary Competition is also effective.

3.5 Summary

The main contribution of this chapter is that we used the signals from seed voxels and the principal components of all signals to calculate PLS components as the activation signal for individual subjects. Clustering methods were applied to automatically select a number of independent components that are reproducible across all the subjects. Then, seed voxels were obtained from the associated ICA maps. After that, we computed the PLS latent variables between the fMRI signal of the seed voxels and the principal components of the signal across all voxels. The PLS maps were used as feature space.

Next, we introduced two ways of parcellation: manifold based methods and aggregation based method. Using a toy example, we demonstrated that, due to noise in the feature space, the parcellation results did not have desired parcel characteristic introduced in Thirion et al. [2006]. Therefore, we proposed an adaptive smoothing method to solve this problem. The smoothing method was applied to two sets of data: the Swiss Roll data and the double Gaussian data. The results showed that

our smoothing method improves the results of manifold embedding.

In order to evaluate parcellation results, we proposed two criteria for validation. The first one is the intra-parcel functional variance, which is similar to that proposed in Thirion et al. [2006]. However, in order to reduce the influence of the noise and other artefacts, we used GLM t-values and PLS t-values to represent the functional behaviour of each voxel. In addition, we proposed Nearest Silhouette Coefficient as another validation method.

Using the toy example and these validation methods, we examined these parcellation techniques: Aggregation and Boundary Competition Parcellation, Isomap parcellation with and without Adaptive Smoothing. The results of both validation approaches give the same conclusion: Isomap parcellation is an effective method for parcellation. However, the performance is limited by noise. Adaptive Smoothing can solve this problem for Isomap parcellation and improve the final results. The Aggregation and Boundary Competition Parcellation gives fast parcellation results and the accuracy is similar to the Isomap parcellation with Adaptive Smoothing. These methods are further compared with two fMRI data sets. The results will be shown and discussed in Chapter 5.

In this chapter, we focused on the parcellation methods for individual subject. In the next chapter, we will discuss how to find the correspondence of parcels across subjects.

CHAPTER 4

Cross Subject Comparison of Parcels

4.1 Introduction

In the previous chapter, we introduced parcellation methods for individual subjects. Due to the fact that many fMRI studies rely on the analysis of groups of subjects, it is essential to establish parcel correspondence across subjects. As mentioned in Chapter 2, there are only few cross-subject parcellation methods. Most of these methods are indirect. For instance, Thirion et al. [2006] constructed the correspondence during individual parcellation. Jbabdi et al. [2009] built hierarchical models on the parcels from all subjects.

In this chapter, we consider the parcel-matching as an independent step. We propose a more general and direct way of solving this problem. Given several individually parcellated subjects, we look for the best correspondence for the parcels across

all subjects.

We formularize the cross-subject parcel-matching problem as a multipartite graph partitioning problem. The main advantage of this approach is that one only needs to define a similarity (or distance) measure between two parcels from two different subjects for the algorithm. The algorithm can optimize the overall parcel matching. The optimization is based on the information of all subjects to increase accuracy. As the parcel matching step is independent of individual parcellation, this method can be applied to match parcels generated with any technique and any imaging modalities (e.g. structural image, DTI).

In this chapter, we introduce the Order Based Simulated Annealing (OBSA) method as a way of solving the multipartite graph partitioning problem. Using one toy example and the multi-subject data, we show that this method gives fast and reliable performance. In addition, we discuss the similarity between our methods and the 'Bags of Pixels' introduced in Jebara [2003], which recast a special case of our problem. Furthermore, we illustrate that the optimization subject used in the 'Bags of Pixels' approach is ill posed as a general method for the task of image recognition.

This chapter is organized as follows. The model-based multi-subject parcellation method is introduced in section 4.2. In section 4.3, the cross-subject parcel matching problem is formalized into a multipartite graph partitioning problem. Then, we describe how to solve this graph partitioning problem. After that, we discuss the idea of representing images as 'Bags of Voxels' in this section. Finally, the experimental results and discussion are given in section 4.4 and section 4.5.

4.2 Multi-subject parcellation

Thirion et al. [2006], proposed a three step approach for multi-subject parcellation: (1) find parcel prototypes, (2) identifiy subject-based instances of these prototypes and (3) parcellate each individual subject with the prototypes. In the following paragraphs, we explain these three steps.

Let us first assume that there are S subjects, each of which has $V_s, s = 1, 2, ..., S$ voxels and each subject is to be parcellated into P parcels. All these subjects have been registered to standard space.

In the first step, all the voxels are pooled together. The set $\mathbb{V}$ has all the $V = \sum_s V_s$ voxels. Each voxel $v(i) \in \mathbb{V}$ is associated with two vectors: (1) the location vector $\mathbf{c}(i)$ which represents the coordinates of the voxels in the standard space, (2) the feature vector $\mathbf{f}(i)$ which is the functional measurement of $v(i)$. These two vectors are represented as:

$$\mathbf{c}(i) = [x(i) \quad y(i) \quad z(i)]', \tag{4.2.1}$$

$$\mathbf{f}(i) = [f_1(i) \quad f_2(i) \quad ,..., \quad f_n(i)]', \quad i = 1, 2, ..., V, \tag{4.2.2}$$

Given any two voxels $v(i)$ and $v(j)$ in the set $\mathbb{V}$, Thirion et al. [2006] defined a spatial distance and a functional distance between them. The spatial distance is

$$d_{spatial}(v(i), v(j)) = \sqrt{[(x(i) - x(j)]^2 + [y(i) - y(j)]^2 + [z(i) - z(j)]^2}. \tag{4.2.3}$$

The functional distance is defined as

$$d_f(v(i), v(j)) = \sqrt{\sum_{k=1}^{n} (f_k(i) - f_k(j))^2}. \tag{4.2.4}$$

After pooling all voxels together, they are clustered to P clusters according to their functional measurement. In addition, the clustering is under the spatial constraint that for any voxel in a cluster, the spatial distance between the voxel and the centre of the corresponding cluster must be less than d_c. The centres of these clusters are used as parcel prototypes. Let $p_i, i = 1, 2, ..., P$ be the prototypes of the P parcels. As in the case of a voxel, each parcel prototype has a location vector $\mathbf{c}^P$ and a feature vector $\mathbf{f}^P$ associated with it, which are

$$\mathbf{c}^P(i) = [x^P(i) \quad y^P(i) \quad z^P(i)]', \tag{4.2.5}$$

$$\mathbf{f}^P(i) = [f_1^P(i) \quad f_2^P(i) \quad ,..., \quad f_n^P(i)]', \quad i = 1, 2, ..., V, \tag{4.2.6}$$

The voxel $\mathbf{v}_s(i)$ is allocated to $p_{\hat{j}}$ according to the following equation:

$$\hat{j} = \arg\min_j d_f(\mathbf{v}_s(i), p_j) \quad \text{subject to} \quad d_s(v_s(i), p_j) < d_c \tag{4.2.7}$$

Constraint C-means is used for this step. More discussion on constraint clustering methods can be found in Basu et al. [2008].

In the second step, given a parcel prototype p_i and a subject s, we need to find the corresponding subject-based instance p_i^s. p_i^s is a voxel in subject s that satisfies two conditions: firstly, it must be spatially and functionally closest to the prototype p_i; secondly, for any parcel prototype p_i, in each subject s, there is one and only one voxel p_i^s that corresponds to p_i. Thirion et al. [2006] model the correspondence between parcel prototypes and subject-based instances by image warping. They propose an iterative algorithm to calculate the these instances.

In the last step, $p_i^s, i = 1, 2, ...P$, are used as seeds for the individual parcellation of the subject. For instance, when using Isomap, $p_i^s, i = 1, 2, ...P$ are used as the

seeds of the k-means clustering algorithm implemented in the embedded Euclidean space. The parcels corresponding to the same parcel prototype are matched with each other.

In this method, the correspondence of the parcels across subjects is constructed during parcellation based on the functional measurement and anatomy space. It fits the analysis framework proposed by Flandin et al. [2002]. However, in this method, the parcel matching step is closely combined with the individual parcellation step. Thus, it is difficult to apply this parcel matching approach to other parcellation methods or parcellation with other image modalities. To deal with this problem, in the next section, we propose a direct way to match parcels across all subjects.

4.3 Cross-subject matching as a multipartite graph partitioning problem

4.3.1 Multipartite graph partitioning for cross-subject matching

In this model, we first assume that the distance between any two parcels from different subjects can be well defined. This distance can effectively measure the dissimilarity between parcels from different subjects. The second assumption is that the parcel matching between any two subjects is one to one. It means that, for any parcel in a subject, it can be matched to one and only one parcel from another subject.

Then, we consider each parcel as a vertex on a graph. Each vertex is connected to all

parcels that are not from the same subject. Each edge is weighted with the distance between the parcels that it connects. Therefore, the parcels and the edges connecting them form a weighted complete multipartite graph K_P^S.

Next, we consider a way to partition the graph K_P^S into a disjoint union of S cliques by removing some edges. After partitioning, each clique is a complete graph. It has one and only one vertex from each subject. In addition, the sum of the weights in all cliques is minimized. According to our assumptions, the parcels in each clique can be considered as matched parcels.

Let p_i^s be the ith parcel of subject s, where $s \in 1, 2, ..., S$ and $i \in 1, 2, ...P$. According to our model, they are the vertices of a multipartite graph. The weight connecting vertices p_i^m and p_j^n is denoted as $w(p_i^m, p_j^n)$ which represents the distance between these two parcels . We want to minimize:

$$W_s = \frac{1}{2} \sum_{m=1}^{S} \sum_{n=1}^{S} \sum_{i=1}^{P} \sum_{j=1}^{P} c_{ij}^{mn} w(p_i^m, p_j^n) \tag{4.3.1}$$

$$\text{Subject to} \quad \sum_{i=1}^{P} c_{ij}^{mn} = 1, \quad \forall m, n \in 1, ...S$$

$$\sum_{j=1}^{P} c_{ij}^{mn} = 1, \quad \forall m, n \in 1, ...S$$

$$c_{ij}^{mn} = 0 \text{ or } 1, \quad \forall m, n \in 1, ...S$$

This problem is essentially a multidimensional assignment problem [Garey and Johnson, 1990]. When there are only two subjects in the data set ($S = 2$), the problem turns into a classic assignment problem:

$$\begin{aligned} \text{Minimize} \quad & W_s = \sum_{i=1}^{P}\sum_{j=1}^{P} c_{ij} w(p_i^1, p_j^2) \\ \text{Subject to} \quad & \sum_{i=1}^{P} c_{ij} = 1 \\ & \sum_{j=1}^{P} c_{ij} = 1, \\ & c_{ij} = 0 \text{ or } 1. \end{aligned} \tag{4.3.2}$$

This problem can be solved with the Munkres algorithm [Munkres, 1957; Riesen et al., 2007], which is illustrated in Figure 4.1. We denote $\mathbf{W}$ as the weight matrix with $w(p_i^1, p_j^2)$ as the element in row i and column j. Given $\mathbf{W}$, Munkres algorithm provides the matrix $\mathbf{C}$. The element c_{ij} in $\mathbf{C}$ is the solution of the assignment problem presented in equation 4.3.2. We give more details of this algorithm in Section 4.3.2.

Unfortunately, the Munkres algorithm can only find the optimal match of the parcels from two subjects. When there are more than two subjects, the corresponding multi-partite graph partitioning problem is generally difficult to solve. For instance, when $S = 3$, as introduced in Garey and Johnson [1990] the problem is generally NP-hard. Burkard et al. [1996]; Crama and Spieksma [1992]; Spieksma and Woeginger [1996] proposed approximation methods for some special cases. When $S \geq 4$, the problem is less studied. Haley [1963] and Pierskalla [1968] mentioned this problem but did not find a good solution. Crama and Spieksma [1992]; Jebara [2003] and Bandelt et al. [1994] proposed algorithms that could give approximations for some special situations.

Similar to the multi-subject image registration problem, one intuitive solution is that we take one subject as reference and match all the other subjects to the reference sub-

ject. The parcels assigned to the same parcel of the reference subject are considered as one clique. When using subject s_0 as reference, rather than minimize equation 4.3.1, this method minimizes

$$\sum_{s=1}^{S}\sum_{i=1}^{P}\sum_{j=1}^{P} c_{ij}^{ss_0} w(p_i^s, p_j^{s_0}). \tag{4.3.3}$$

In each clique, this approach only minimizes the weights of the edges connecting the parcels from the reference subject. Thus, the solution is obviously suboptimal. Especially when the reference subject has a poor imaging or parcellation quality, the matching becomes less reliable. On the other hand, the parcel-matching step is a fundamental step for further analysis. The quality of the matching results directly influences the accuracy of the analysis in the next step. Therefore, it is important to find the optimal match of the parcels. Unfortunately, to the best of our knowledge, there is no efficient algorithm to solve this problem. The discussion of this problem mostly focuses on how to approximate the optimal solution.

Another possible solution for this problem is to make further assumptions on the graph to simplify the problem. For instance, in Jebara [2003]; Kuroki and Matsui [2009], it is assumed that each vertex in the graph can be described as a point in an Euclidean space. The weight of an edge is the Euclidean distance between the corresponding vertices. Under this assumption, the multipartite partitioning problem can be formularized into a quadratic programming problem.

However, this assumption limits the application of this model. In many situations, it is not convenient to describe the weights as the Euclidean distances between the vertices, especially in analysis with a different imaging modality. For instance, in

fMRI analysis, for any two parcels from two subjects in standard space, if the centres of these two parcels are more than 5cm away from each other, it is not sensible to match them together. Therefore, we could set the weight connecting them as infinity. But the consequence of this setting is that it is difficult to consider the vertices as points embedded in a Euclidean space. When combining other imaging modalities, for instance, DTI, it would be common to encounter such a problem.

In order to provide a way of solving this problem, in Section 4.3.2, we propose a novel algorithm that could approximate the optimal matching results, without making further assumptions on the graph .

4.3.2 Order Based Simulate Annealing (OBSA)

As mentioned in section 4.3.1, Munkres algorithm gives optimal results to the assignment problem but it cannot deal with high dimensional problems. In this section, we propose Order Based Simulate Annealing (OBSA) to solve the multipartite graph partitioning problem. The OBSA provides an approximation of the optimal solution based on the Munkres algorithm.

Munkres Algorithm

The Munkres algorithm (also known as Kuhn-Munkres Algorithm or the Hungarian method) was first published in Kuhn [1955]. After that, Munkres [1957] found that this method is polynomial. In Figure 4.1, we show the flow chart of this algorithm. Given a weight matrix $\mathbf{W}$ 4.3.2, the algorithm returns a modified matrix $\mathbf{W}'$. The c_{ij}

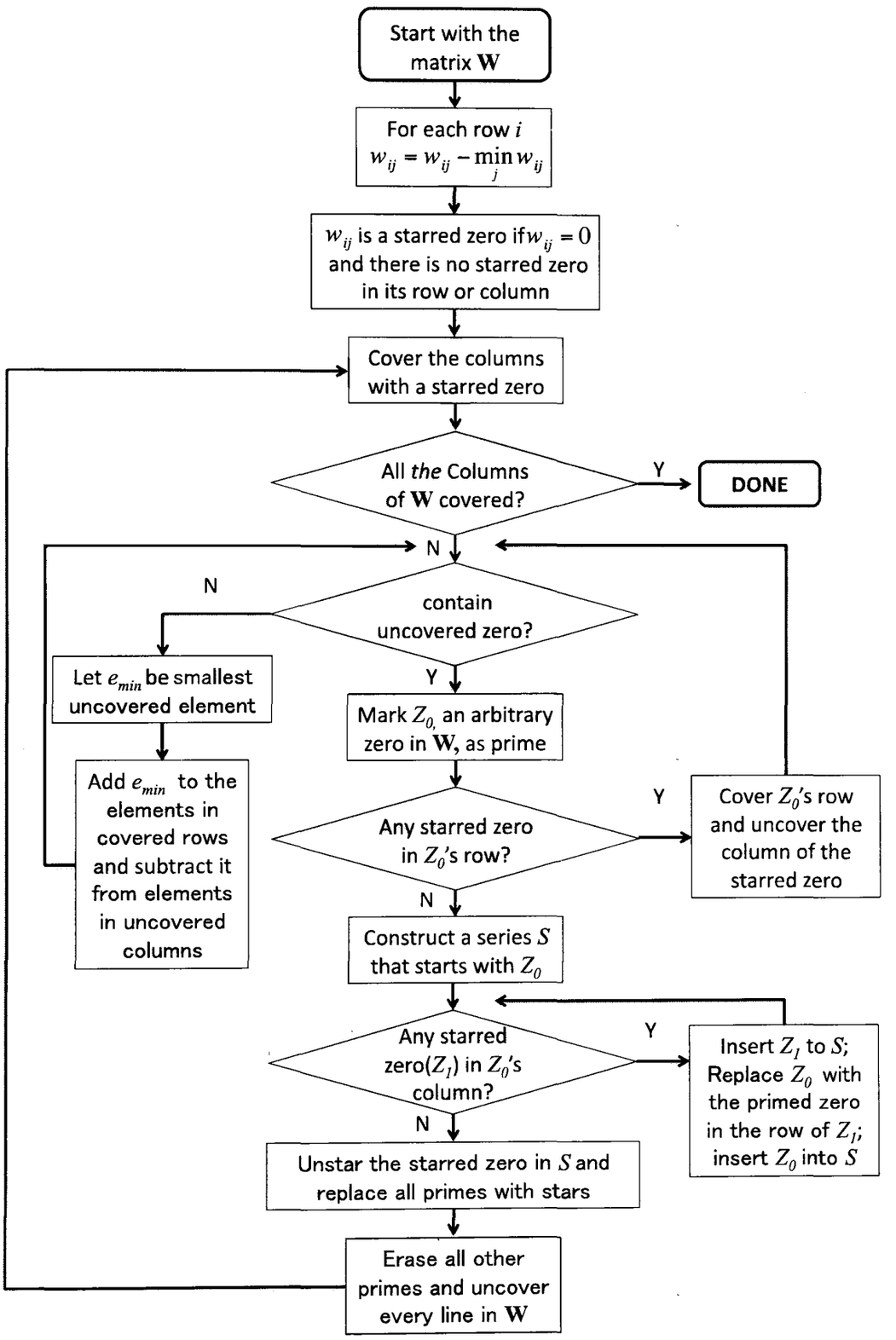


Figure 4.1: Munkres Algorithm.

can be calculated as:

$$c_{ij} = \begin{cases} 1 & \text{if } w'_{ij} \text{ is a starred zero} \\ 0 & \text{otherwise} \end{cases} \tag{4.3.4}$$

The corresponding sum of weights is:

$$W_s = \sum_{ij} c_{ij} w(p_i^1, p_j^2). \tag{4.3.5}$$

W_s can be used to evaluate the quality of this assignment.

Order Based Simulate Annealing (OBSA)

Using the same notation as in section 4.3.1, this method can be described by the following algorithm:

1. *Initialization*

   Set $W_{\min} = \infty$.

2. *Randomly permute the order of all the subjects*

   Let a sequence $\mathbf{S}$ be a random permutation of the integers from 1 to S, which is the number of all subjects. $\mathbf{S}(i)$ is the i^{th} element of the sequence. It represents the index of a subject.

3. *Match the first two subjects in the sequence*

   Use Munkres algorithm to match subject $\mathbf{S}(1)$ and $\mathbf{S}(2)$ and set $i = 3$. $\mathbf{C}^2$ is the matrix for the result assignment. G^2 is this matched bipartite graph. In this graph, only the edges correspond to $c_{ij}^2 \geq 0, c_{ij}^2 \in \mathbf{C}^2$ are kept.

4. *Match the $i^{th}(i \geq 3)$ subjects in the sequence*

   If $i > S$, go to next step, otherwise, set $j = 1$, $W_t = \infty$

   (a) If $j \geq i$, set $i = i + 1$ and go to the beginning of step 4, otherwise, use Munkres algorithm to match subject $\mathbf{S}(i)$ to $\mathbf{S}(j)$. The $\mathbf{C}^i$ is the result assignment matrix.

   (b) According to matrix $\mathbf{C}^i$, match the new subject $\mathbf{S}(i)$ to the graph G^{i-1} to formalize a i-partite graph. This new graph is denoted as G_t^i. In G_t^i, the subject $\mathbf{S}(i)$ only connects to $\mathbf{S}(j)$ by the edges corresponding to the non-zero elements in $\mathbf{C}^i$. G_t^i partitions all vertices into P disjoint sets. Each set has one vertex from each subject. Let K_t^i be the completed weighted multipartite graph with subjects $S(1), ...S(i)$. So it has the same vertices as G_t^i. We partition K_t^i into P cliques so that each clique has the same vertices as each set in G_t^i. Then, calculate the sum of weights $\widetilde{W}$ of all the cliques in K_t^i. If $\widetilde{W} < W_t$, set $W_t = \widetilde{W}$ and $G^i = G_t^i$. Then set $j = j + 1$ and go to step (a).

5. *Update the matched graph*

   Until now we could get an S-partite graph G^S and the sum of the weights of this graph W_t. W_t is minimized according to the order of subjects determined in step 2. If $W_t < W_{\text{min}}$, $W_{\text{min}} = W_t$ and $G_{\text{opt}} = G^s$. Return to step 2.

The algorithm should stop when W_{min} stops decreasing. Practically, this algorithm converges quickly. We show this in Section 4.4.

4.3.3 Bags of Pixels and Bags of Parcels

Jebara [2003] proposed a method that represents an image as a Bag of Pixels. In this section, we demonstrate that this representation of images can be considered as a special case of our problem. Furthermore, we show that this representation of the images is ill posed. However, the quadratic programming formularized in Jebara [2003] could be used to solve the multipartite graph partitioning problem. Therefore, we introduce the method in this section.

When representing images as Bags of Pixels, each pixel in a image is described as *n-tuple*. For instance, given a grey scale image with N pixels, each pixel in the image is presented as a 3-tuple $\mathbf{x} = [x\ y\ i]$, in which, x and y represent the location of the pixel and i is the associated intensity value. Then, the image turns into a vector of $3N \times 1$:

$$[\mathbf{x}_1\ \mathbf{x}_2, ...,\ \mathbf{x}_N]' = [x_1\ y_1\ i_1\ x_2\ y_2\ i_2, ...,\ x_N\ y_N\ i_N]'. \tag{4.3.6}$$

Let $\mathbf{X}$ be a $N \cdot D \times 1$ vector representing a image with N pixels. Each pixel is described with a D-tuple: $\mathbf{x}_i = [x\ y\ f(1)\ f(2), ...,\ f(D-2)], i \in \{1, 2, ..., N\}$. The matrix $\mathbf{A}$ is a $(N \cdot D) \times (N \cdot D)$ permutation matrix, such that:

$$\mathbf{A} = \begin{pmatrix} a_{11}\mathbf{I}_D & a_{12}\mathbf{I}_D & \dots & a_{1N}\mathbf{I}_D \\ a_{21}\mathbf{I}_D & a_{22}\mathbf{I}_D & \dots & a_{2N}\mathbf{I}_D \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}\mathbf{I}_D & a_{N2}\mathbf{I}_D & \dots & a_{NN}\mathbf{I}_D \end{pmatrix}$$

where, $\mathbf{I}_D$ is a $D \times D$ identity matrix. $\mathbf{A}$ is a doubly stochastic matrix with $a_{ij} \in \{0, 1\}$, $\sum_i a_{ij} = 1$ and $\sum_j a_{ij} = 1$.

Due to the fact that each tuple contains all the information of a pixel, the tuples can be concatenated in an arbitrary order. Therefore, for any permutation matrix $\mathbf{A}$, $\mathbf{AX}$ is a different configuration of the same image. All the configurations of $\mathbf{AX}$ form a manifold.

For a set of K images $\mathbf{X}^k, k \in 1, 2, ..., K$ that are from the same class, a Gaussian model is constructed to describe these images:

$$\mathbf{A}^k\mathbf{X}^k \sim N(\mathbf{A}^k\mathbf{X}^k; \mu, \Sigma) \tag{4.3.7}$$

For simplicity, we assume that the covariance matrix Σ is an identity matrix. Then, the log-likelihood can be expressed as

$$\begin{aligned} l(\mathbb{A}, \hat{\mu}) &= -\frac{KD}{2}\log(2\pi) - -\frac{1}{2}\sum_k ||\mathbf{A}^k\mathbf{X}^k - \hat{\mu}||^2 \\ \text{where} \quad & \hat{\mu} = \frac{1}{K}\sum_k \mathbf{A}^k\mathbf{X}^k, \\ \text{and} \quad & \mathbb{A} = \{\mathbf{A}^1, \mathbf{A}^2, ..., \mathbf{A}^K\}. \end{aligned} \tag{4.3.8}$$

The log-likelihood $l(\mathbb{A}, \hat{\mu})$ can be further maximized over $\mathbb{A}$ by minimizing

$$C(\mathbb{A}) = \sum_k ||\mathbf{A}^k\mathbf{X}^k - \hat{\mu}||^2. \tag{4.3.9}$$

It should be noticed that equation 4.3.9 is equivalent to the multipartite graph partitioning problem. Let $\mathbf{x}_i^k = [x_i^k\ y_i^k\ f_i^k(1)\ f_i^k(2), ...,\ f_i^k(D-2)]$ be a vertex of a graph. It connects to all the tuples in other $K-1$ images. The weights of edges connecting any two tuples is defined as:

$$\begin{aligned} w(\mathbf{x}_i^m, \mathbf{x}_j^n) &= ||\mathbf{x}_i^m - \mathbf{x}_j^n||^2 \\ &= (x_i^m - x_j^n)^2 + (y_i^m - y_j^n)^2 + (f_i^m(1) - f_j^n(1))^2 + \\ &\quad + ... + (f_i^m(D-2) - f_j^n(D-2))^2 \end{aligned} \tag{4.3.10}$$

All the tuples form a K-partite graph. The multipartite graph partitioning problem can be represented by

$$\text{Mininize} \quad W_s(\mathbb{A}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}||\mathbf{A}^i\mathbf{X}^i - \mathbf{A}^j\mathbf{X}^j||^2 \tag{4.3.11}$$

The minimization object $W_s(\mathbb{A})$ and $C(\mathbb{A})$ are equivalent. Therefore, the idea of Bag of Pixels can be understood as a two step process. First, the pixels from different images are first matched to each other according to their positions and intensity values. This step is similar to our multipartite graph partition model. Further analysis, such as PCA [Jebara, 2003] and kernel PCA [Kondor and Jebara, 2003], is then applied to the matched pixels. From a contrary viewpoint, the cross subject parcel-matching can be considered as an idea for representing each subject as a Bag of Parcels. However the similarity between parcels can be more flexibly and reliably defined.

In image recognition, it is not common to take the co-registration of all the images as a preprocessing step. Without a careful preprocessing, it is difficult to be convinced that the correspondence of two pixels in two images can be determined only by the position and the intensity values of these two pixels.

In addition, when the scale of the image intensity changes, even for the same data set, the result would be different. For instance, given a set of the grey scale images with intensity $0 \sim 1$, if we linearly enhance the intensity scale to $0 \sim 100$, the matching result would depend more on the intensity. Vice versa, if we linearly change the intensity scale to $0 \sim 0.1$, the matching result would depend more on the location of the pixel.

Following the above discussion, when images are carefully preprocessed, represent-

ing them as bags of pixels may be effective but as a general method for image recognition, this formalization is ill posed.

4.3.4 Quadratic programming for multipartite graph partitioning

Jebara [2003] proposed a fast method to solve equation 4.3.8 using quadratic programming. According to the discussion in the last section, when the vertices of a multipartite graph can be presented as points embedded in a Euclidean space, minimizing W_s is equivalent to minimizing $C(\mathbb{A})$. Therefore, the algorithm proposed by Jebara [2003] could solve a sub-class of the multipartite graph partitioning problem. In this method, the constraints on the permutation matrix $\mathbf{A}^k$ are relaxed. The optimization problem formularized as:

$$\begin{aligned} &\text{Minimize} \quad \hat{C}(\mathbb{A}) = \sum_k ||\mathbf{A}^k\mathbf{X}^k - \hat{\mu}||^2 + \lambda \sum_{ijk} (a_{ij}^k - \frac{1}{N})^2 \\ &\text{subject to} \quad a_{ij}^k \geq 0, \sum_i a_{ij}^k = 1, \sum_j a_{ij}^k = 1, \end{aligned} \tag{4.3.12}$$

In order to avoid degeneracy, in equation 4.3.12, a penalty term is added to the minimization object. This term penalizes the a_{ij}^k close to $1/N$ and favours the situation when a_{ij}^k is close to 0 or 1. The parameter λ balances the convexity of the problem and the optimality of the result.

The quadratic programming is solved with the SMO algorithm introduced in Platt [1999]. This algorithm iteratively updates the elements of matrix $\mathbf{A}^k$, where k is randomly selected. At each iteration step, only four elements are updated. If denoting these 4 elements as a_{mn}^k, a_{mq}^k, a_{pn}^k and a_{pq}^k, due to the linear constraints in equation

4.3.12, we have:

$$\begin{aligned} a^k_{mn} + a^k_{mq} &= a \qquad & a^k_{mn} + a^k_{pn} &= b \\ a^k_{mq} + a^k_{pq} &= c \qquad & a^k_{pn} + a^k_{pq} &= d = b + c - a \end{aligned} \tag{4.3.13}$$

where $0 \leq a, b, c \leq 1$.

If further denoting H_1,H_2,H_3,H_4 as:

$$\begin{aligned} H_1 &= \sum_{j \neq n,q}^{N} a^k_{mj} \mathbf{x}^k_j \qquad & H_2 &= \sum_{k=1}^{K} \sum_{j \neq n,q}^{N} a^k_{mj} \mathbf{x}^k_j \\ H_3 &= \sum_{j \neq n,q}^{N} a^k_{pj} \mathbf{x}^k_j \qquad & H_4 &= \sum_{k=1}^{K} \sum_{j \neq n,q}^{N} a^k_{pj} \mathbf{x}^k_j, \end{aligned} \tag{4.3.14}$$

the updating rule for the new a^k_{mn} is:

$$a^k_{mn}{}^{\text{new}} = \frac{\text{num}}{\text{den}}, \tag{4.3.15}$$

where,

$$\begin{aligned} \text{num} &= [H_2 - KH_1 + KH_3 - H_4 + (c - 2a)(K-1)\mathbf{x}^k_q + \\ &\quad + b(K-1)\mathbf{x}^k_n] \cdot (\mathbf{x}^k_n - \mathbf{x}^k_q) - \lambda K(2a + b - c) \qquad (4.3.16) \\ \text{den} &= 2(K-1)(\mathbf{x}^k_n - \mathbf{x}^k_q) \cdot (\mathbf{x}^k_n - \mathbf{x}^k_q) - 4\lambda K. \qquad (4.3.17) \end{aligned}$$

This means that the updating rule for new a^k_{mq}, a^k_{pn} and a^k_{pq} can be calculated according to the constraints in equation 4.3.13.

The entries in the permutation matrix are updated iteratively, until the cost function $\hat{C}(\mathbb{A})$ drops below the threshold steadily. More details of this algorithm can be found in Jebara [2003] and Guo and Gao [2006].

4.4 Experiment results

In this section, we examine the multipartite graph partitioning algorithms with two sets of data. One is the toy data generated with the double Gaussian model as in equation 4.4.1. The other one is the multi-subject face and gesture data introduced in section 5.3. Three parcel-matching methods are tested and compared in this section. The first one takes each subject as reference and matches others to the reference subject with Munkres algorithm. As this method matches other subjects to the reference subject, it cannot fully use the information in all subjects to find the best match. The result of this algorithm relies heavily on the selection of the reference subject. Next, we apply our OBSA algorithm and the quadratic programming formalized in Jebara [2003] to the data sets. The accuracy of the matching is evaluated by the sum of weights in all the cliques as the W_s in equation 4.3.1.

4.4.1 Results from toy data

In this section, we first use the double Gaussian model to generate 20 images as a toy data set. The images are generate according this equation:

$$f(x,y) = 5 \cdot (e^{-\frac{(x-x_1)^2+(y-y_1)^2}{90}} + e^{-\frac{(x-x_2)^2+(y-y_2)^2}{90}}) + p\epsilon, \tag{4.4.1}$$

where, $\epsilon \sim \mathcal{N}(0,1)$. In order to simulate the inter-subject variability, for each image the location of the Gaussian peaks and the noise level are randomized. Each image

is 50×50. The parameters x_1, y_1, x_2, y_2, p are calculated as:

$$x_1 = 25 - 10\sqrt{2}\cos(\pi/4 - \theta) + \gamma_1 \quad y_1 = 25 + 10\sqrt{2}\sin(\pi/4 - \theta) + \gamma_1$$

$$x_2 = 25 + 10\sqrt{2}\cos(\pi/4 - \theta) + \gamma_2 \quad y_2 = 25 - 10\sqrt{2}\sin(\pi/4 - \theta) + \gamma_2$$

$$\gamma_1 \sim U(0,8) \quad \gamma_2 \sim U(0,8) \quad p \sim U(0.1, 0.3) \quad \theta \sim U(-\pi/4, \pi/4) \quad (4.4.2)$$

The generated images are shown in Figure 4.2.

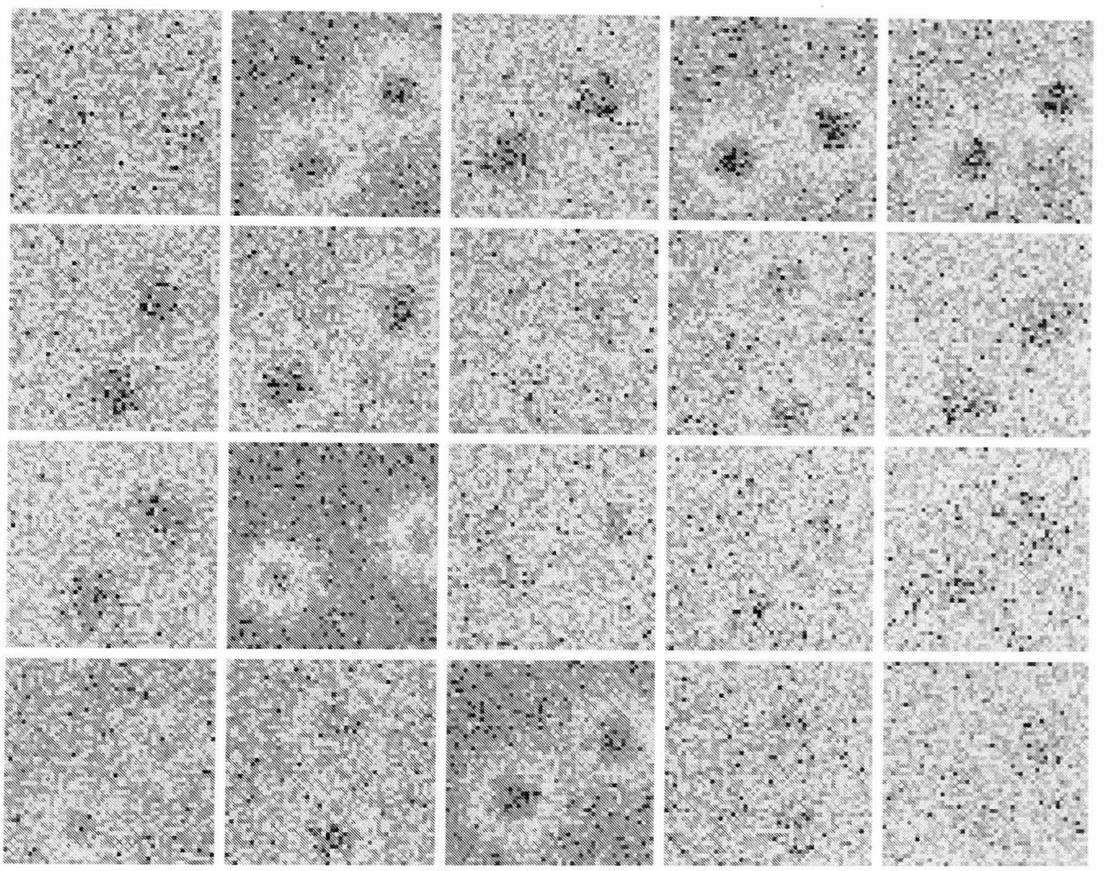


Figure 4.2: Toy data.

Adaptive smoothing and Isomap are used for the parcellation. Each image is parcellated into 7 parcels. The indices of the parcels are randomly labelled. Figure 4.3 shows the parcellation results. We use these parcels as a toy example to test the parcel-matching algorithms.

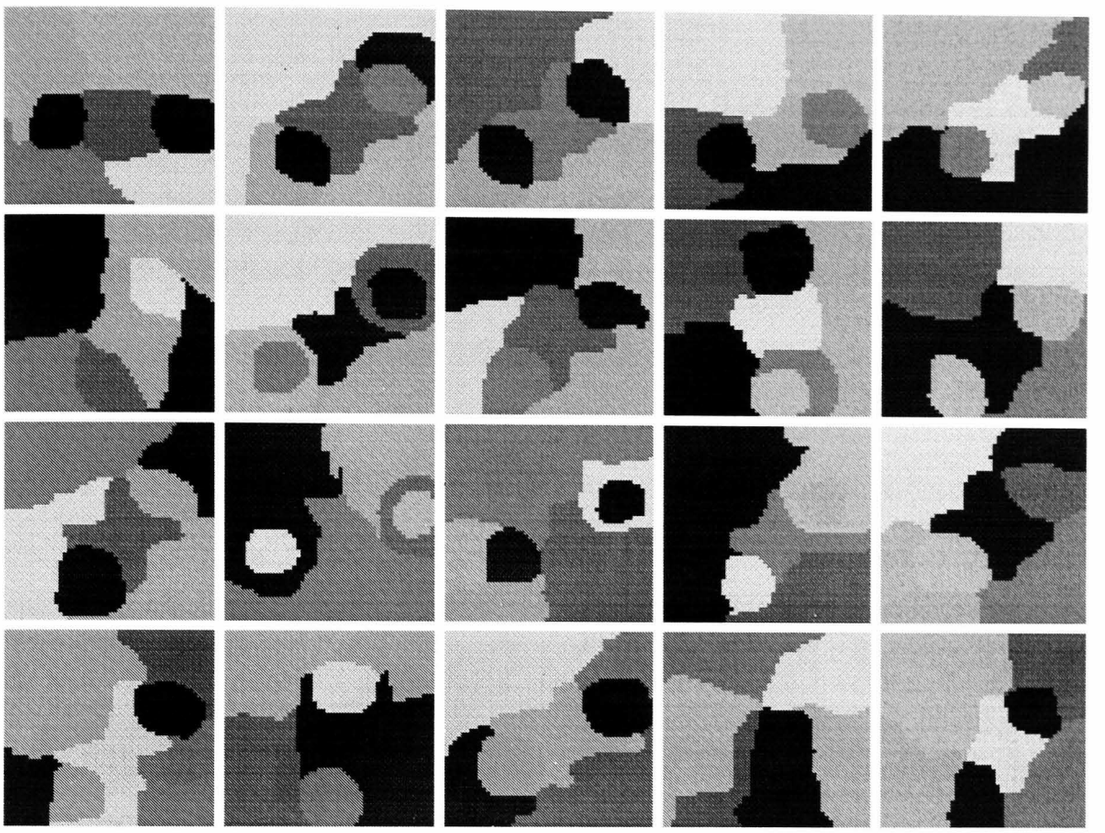


Figure 4.3: The parcellation of the toy data.

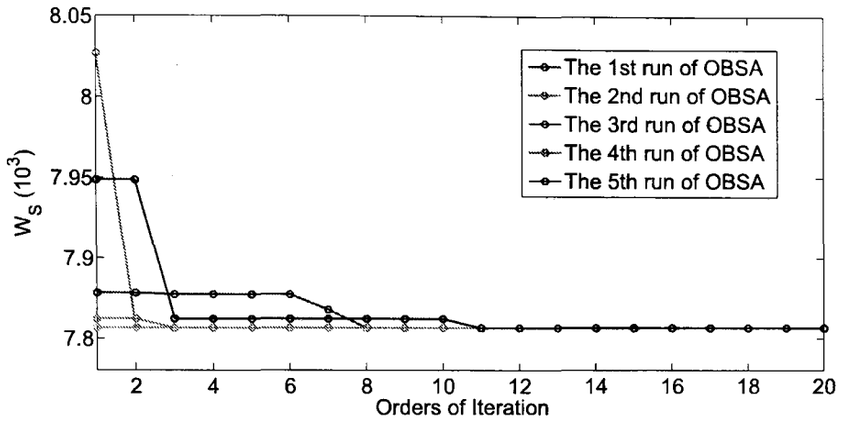
As we need to test the algorithm based on quadratic programming, the j^{th} parcel of subject i is represented as a point in 3-dimensional Euclidean space as:

$$p_j^i = [x_j^i \quad y_j^i \quad \sqrt{\gamma} I_j^i], \tag{4.4.3}$$

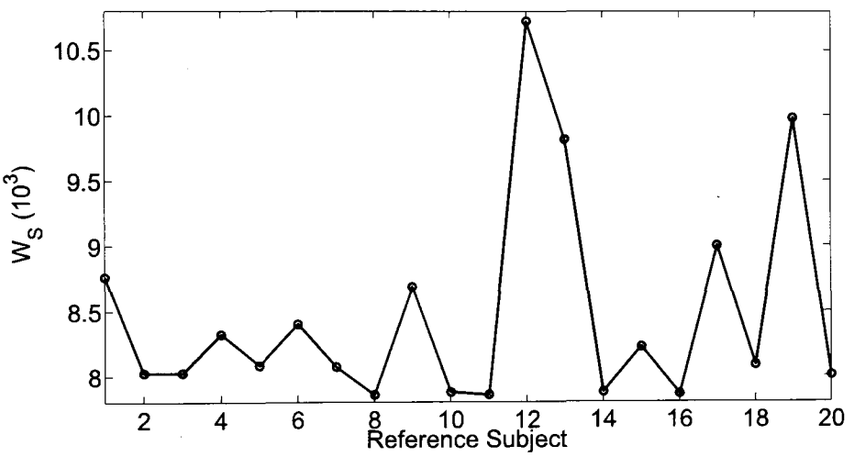
and the weight of the edge connecting two parcels p_a^m and p_b^n is :

$$w(p_a^m, p_b^n) = (x_a^m - x_b^n)^2 + (y_a^m - y_b^n)^2 + \gamma \cdot (I_a^m - I_b^n)^2 \tag{4.4.4}$$

In the above equations, the x and y are the locations of the geometric centre of the parcel and I is the mean intensity of the pixels in the parcel. The parameter γ is used to adjust the optimization object in parcel matching. When γ is large, the algorithms match the parcels according to the image intensity. On the contrary,



4.4.1: Matching with OBSA.



4.4.2: Matching with each subject as reference.

Figure 4.4: Comparison of parcel matching methods with toy data.

when γ is small, the algorithms match the parcels whose location are closer to each other. The parameter γ is determined by the size of the image and the scale of the intensity. Here, the γ is selected as:

$$\gamma = \frac{1}{P^2S^2} \sum_{m,n,a,b} \frac{(x_a^m - x_b^n)^2 + (y_a^m - y_b^n)^2}{(I_a^m - I_b^n)^2}, \tag{4.4.5}$$

where, $P = 7$ is the number of the parcels in each subject and $S = 20$ is the number of subjects.

In Figure 4.4, W_s is used as a criterion to measure the matching results of different matching methods. Figure 4.4.2 shows the results of an intuitive matching method, which takes each subject as a reference and matches others to the reference subjects. The horizontal axis represents the subject taken as reference. According to the black line, when taking subject 12, 13, 19 as the reference, the corresponding W_S is higher. This is due to the fact the parcellation results of these subjects are not as good as others. This can be seen from Figure 4.3. These three subjects are shown in row 3 column 2, row 3 column 3 and row 4 column 4. Due the the high noise in the images, the parcellation results of these subjects are different from others. In subject 12 and 13, the parcels connecting the two Gaussian peaks are lost. In subject 19, one Gaussian peak is lost.

Figure 4.4.1 shows the results of OBSA algorithm. In order to test the stability of the OBSA algorithm, we run OBSA five times. At each time, the indices of the parcels are randomly permuted, so that the initialization point of the algorithm is randomized. We show the results of these five runs with different colour lines. The horizontal axis represents the iteration in the OBSA algorithm. In the first four runs, we iterated 20 times and in the last run, the algorithm was iterated 200 times. In this run, after 20 times of iteration, W_S did not decrease. These test results show that the OBSA algorithm converges quickly. In the worst case of our experiment, the algorithm converges at the 11th iteration.

In the quadratic programming, formalized in equation 4.3.12, parameter λ is used to control convergence and performance. In order to achieve the best result of this algorithm, we used several λ values to evaluate its performance. However, the

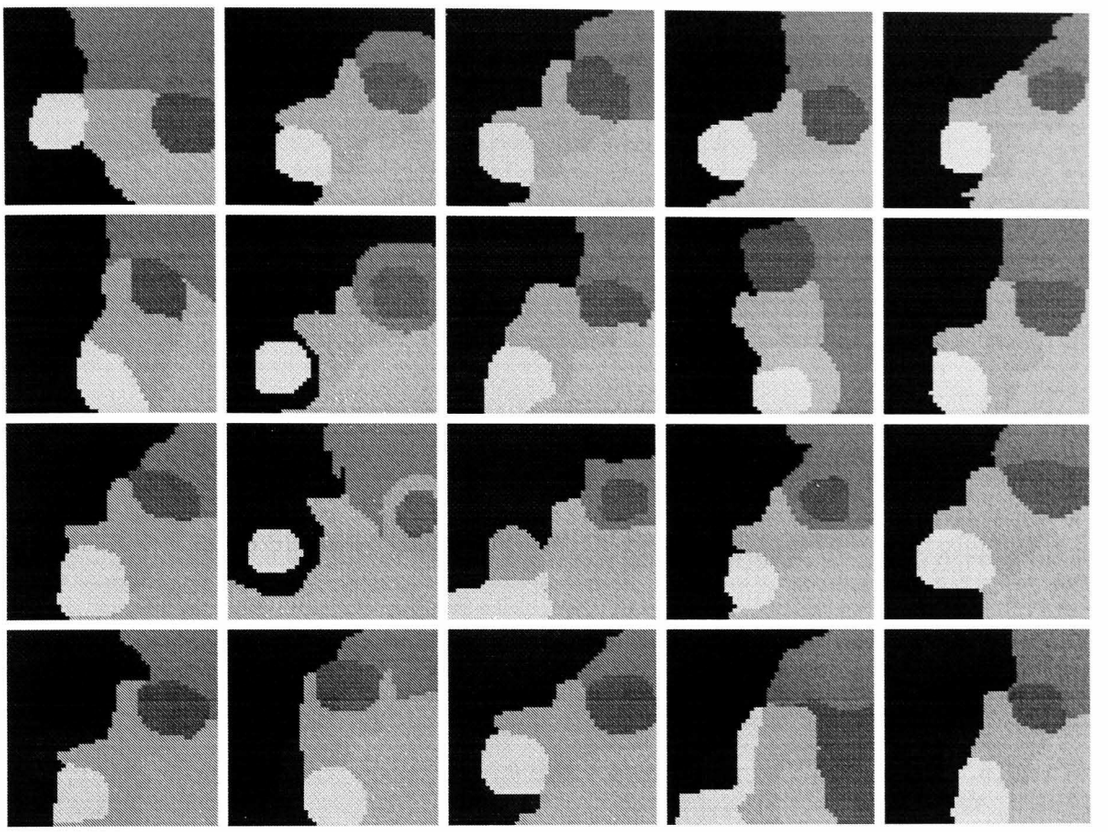


Figure 4.5: Matched parcels. The parcels matched to each other are marked with the same colour in this figure. The parcels that represents the Gaussian peaks (orange and green parcels) are matched well.

result is disappointing. We first use small λ for fast convergence, but the result W_S is very large. After that, we gradually increase λ. As λ increases, the algorithm takes a longer time to converge but the matching result keeps the same. The best result we could get from this algorithm is $W_S = 13.3 \times 10^3$, which is worse than the previous two methods.

To conclude, for this toy data, the matching results from the intuitive method and our approach are the same and the OBSA algorithm shows a fast convergence. The quadratic programming formularized in Jebara [2003] does not give a comparable

result.

According to the best matching result, we show the matched parcels in Figure 4.5. The same colour represents the parcels in the same clique.

4.4.2 Results from multi-subject fMRI data

In this section, we use the multi-subject face and gesture data to test the parcel matching algorithms. As in the experiment on the toy data, the method is based on soft permutation. As quadratic programming could not give a stable and satisfactory result, we apply only the other two methods this data set.

When dealing with real fMRI data,the situation is more complicated. The distances between the parcels should be defined according the specific application. However, the aim of this section is to compare the parcel matching methods. Therefore, we focus on a numerical comparison of the matching results. More discussion on the application of this data set is discussed in the next chapter.

There are 25 subjects in the data set with each subject parcellated into 600 parcels. The parcel i from subject s is represented by:

$$p_i^s = [x_i^s \quad y_i^s \quad z_i^s \quad \mathbf{t}_i^s], \tag{4.4.6}$$

where, vector $[x_i^s \quad y_i^s \quad z_i^s]$ represents the centre of this parcel in MNI space, and $\mathbf{t}_i^s$ is a vector of GLM t-values that describes the functional measurement of this parcel. For any two parcels p_i^m and p_j^n, $w(p_i^m, p_j^n)$, the weights of the edge connecting these

two parcels is defined as:

$$w(p_i^m, p_j^n) = \begin{cases} d_s(p_i^m, p_j^n) + \gamma||\mathbf{t}_i^m - \mathbf{t}_j^n|| & \text{if } d_s(p_i^m, p_j^n) < 50mm \\ \infty & \text{otherwise} \end{cases} \tag{4.4.7}$$

where,

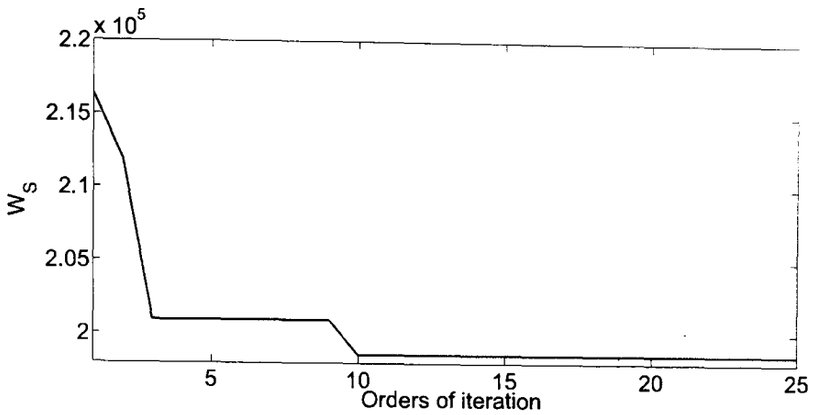
$$d_s(p_i^m, p_j^n) = ||[x_i^m \quad y_i^m \quad z_i^m] - [x_j^n \quad y_j^n \quad z_j^n]||. \tag{4.4.8}$$

Here, we add a spatial constraint to the definition of the edge weights. We assume that for any two parcels from two subjects, if their distance in the MNI space is more than 50mm, they do not correspond to the same brain structure. It is a very common assumption for fMRI data analysis. However, on account of this assumption, the vertices in the multipartite graph cannot be presented as points embedded in Euclidean space. Thus, it excludes several fast algorithms, including the quadratic programming method discussed in Section 4.3.4.

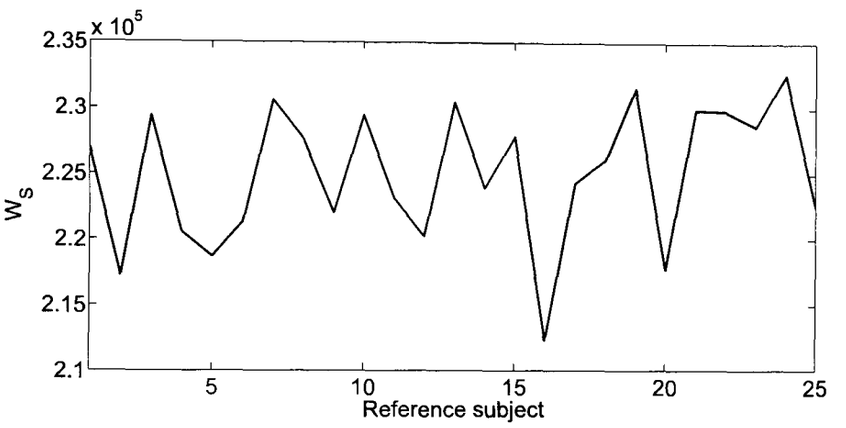
However, the other two methods are still effective. The results are shown in Figure 4.6. Similarly to Figure 4.4, the black line in Figure 4.6.2 shows the results when taking each subject as a reference and matching the rest to it. The horizontal axis represents the indices of the reference subject. The blue line Figure 4.6.1 shows the result W_S from the OBSA algorithm and the horizontal axis denotes the order of the iteration. The algorithm stabilizes after 10 iterations.

4.5 Discussion

In this chapter, the multi-subject parcel-matching problem is considered as a multi-partite graph partitioning problem. We demonstrate the similarity between our idea



4.6.1: Matching with OBSA.



4.6.2: Matching with each subject as reference.

Figure 4.6: Comparison of parcel matching methods with multi-subject fMRI data.

and the method of representing images as bags of pixels. Based on that discussion, we conclude that representing images as bags of pixels is ill-posed for the general image recognition problem.

In order to solve the multipartite graph partitioning problem, the method of Order Based Simulated Annealing (OBSA) is proposed. Based on two data sets, this method is compared with the approach based on convex programming introduced in Jebara [2003] and a heuristic method using Munkres algorithm. Unfortunately,

the soft permutation algorithm formularized in Jebara [2003] could not give satisfactory results. This further supports the theoretical argument presented in Guo and Gao [2006]. Compared with the heuristic method, the OBSA improves the accuracy of parcel-matching problem.

The direct way of solving the multipartite graph partitioning problem is to search the space of c_{ij}^{mn} to find the minimal W_S. The OBSA changed the searching space to the order of the sequence of all subjects $\mathbf{S}$. In this space, for most practical problems, more than one order corresponds to the minimal W_S. Thus it is easier to find a reasonable solution. The experiment results also support this hypothesis. The algorithm converges quickly during the iteration.

In the application of multi-subject matching to neuroscience it is important that similarity between the parcels can be defined freely, according to the requirements of the research. Thus, sometimes it is not convenient to model the parcels as points embedded in a Euclidean space. For instance, in fMRI analysis, the weights defined in equation 4.4.7 are reasonable and necessary but such a definition excludes many fast algorithms, such as the ones proposed by Bandelt et al. [1994] and Kuroki and Matsui [2009]. Under these circumstances, the OBSA can be used as a suitable tool to solve these problems encountered in neuroscience.

CHAPTER 5

Application to fMRI data sets

5.1 Introduction

In this chapter, we apply our methods to two sets of fMRI data: single-subject motor cortex stimulation data and multi-subject face and gesture data. We use the experiment results to test our hypotheses and examine the performance of our methods.

In the following section, we first present the experiment results with the single-subject motor cortex data set. The discussion focuses on the performance of our data-driven parcellation for individual subjects. Using a statistical test, we examine the hypothesis that our individual data-driven parcellation approach improves over model-based parcellation techniques in terms of parcellation accuracy.

In Section 5.3 we use the multi-subject face and gesture data to test this hypothesis. When applying our individual parcellation method on on multi-subject data, we need to first find seed voxels that represent the desired functional responses. With this data set, we start by examining whether we can find reproducible ICs with

clustering methods. Then we compare the performance of our parcellation methods with the model-based individual parcellation method.

In the next section, we examine our second hypothesis: our multi-subject data-driven parcellation approach improves over standard voxel-wise fMRI analysis in terms of both robustness and sensitivity to normalization issues. In addition, we also show that the use of group information for parcel matching increases the sensitivity of the analysis.

Finally, in Section 5.5, we discuss the result of these analyses and give a summary of this chapter.

5.2 Experiment on single-subject motor cortex stimulation data

5.2.1 Data

In Chapter 3, we used this dataset to illustrate several data processing steps. Here we give more details about this set of data and the corresponding experiment.

This set of single-subject data was acquired on a Philips Intera 1.5T scanner. Each volume was scanned within 3 seconds (TR = 3s). There were altogether 108 volumes. The task consisted of a sequential finger-tapping task paced with auditory signals from a metronome. These auditory signals were given every 0.6 seconds. The digit order of the tapping was 1 - 3 - 2 - 4 repeated 6 times (14.4 seconds) in each period with a 14.4 second rest afterwards. Therefore, the period of one on-and-off block

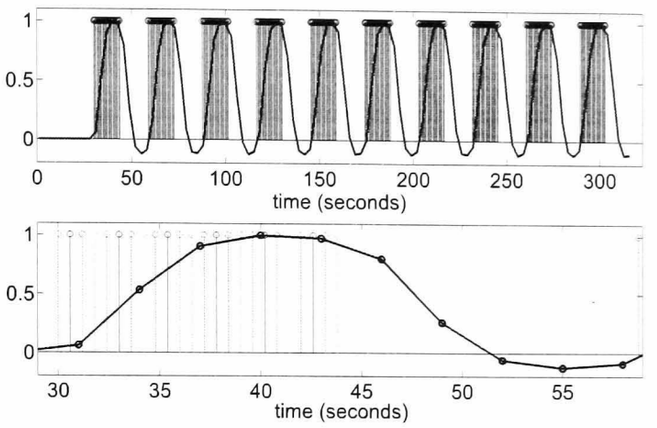


Figure 5.1: The fMRI scan of the single-subject motor cortex stimulation data.

was 28.8 seconds. This period was repeated 10 times during the experiment.

Figure 5.1 shows the tasks during the scan and the convolutional model of the corresponding BOLD response. The lower graph shows one period of the experiment, in which different colours illustrate tapping different fingers. The blue line shows the hemodynamic model. At the beginning of the experiment, there was a rest period of 30 seconds for the scanner to stabilize.

5.2.2 Parcellation

Preprocessing

The data was first preprocessed with slice-timing correction and motion correction. We used the linear interpolation and affine registration in these steps which were implemented with the FSL toolbox [FMRIB, 2007]. To remove invalid voxels, we then applied histogram filtering on the preprocessed data. Smoothing is not used in the preprocessing step.

PCA denoising

As introduced in section 3.2.4, we used PCA to increase the sensitivity of the analysis. Here, based on this single-subject data set, we used an experiment to illustrate the performance of this denoising method.

We first calculated the GLM t-values from the centred data matrix $\mathbf{X}_{V\times T} = x_{ij}$. Then, $\mathbf{X}_{V\times T}$ was decomposed as in equation 3.2.10 and equation 3.2.11. For this set of data, the first two PCs, which represents high-frequence noise, are first removed. After that, the last PCs, which altogether cover that last 20% of the variance, are removed. Let $\mathbf{T}_{\text{PCA}}^{clean}$ and $\mathbf{P}_{\text{PCA}}^{clean}$ be the PCA score matrix and loading matrix after removing the components corresponding to the noise. The 'clean' data matrix is:

$$\mathbf{X}_{V\times T}^{\text{clean}} = \mathbf{P}_{\text{PCA}}^{\text{clean}} \cdot \mathbf{T}_{\text{PCA}}^{\text{clean}'}. \tag{5.2.1}$$

Then, we used matrix $\mathbf{X}_{V\times T}^{\text{clean}}$ to calculate another set of t-values. Figure 5.2 shows the comparison of both sets of t-values. From this figure, we can see that after PCA denoising, the sensitivity of the GLM model increases.

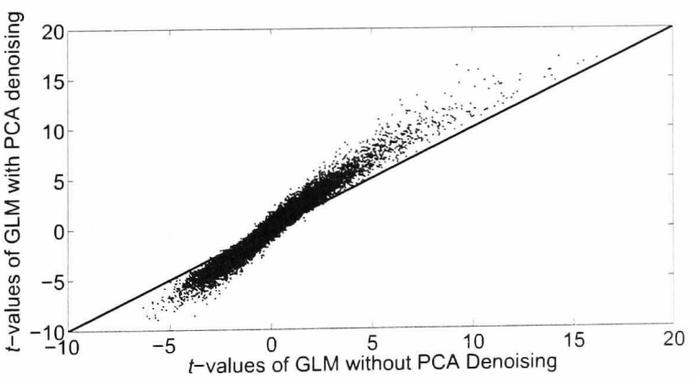


Figure 5.2: Comparison of GLM t-values with and without PCA denoising.

In order to further illustrate the effectiveness of the denoising, we show eight slices

of t-statistics maps in Figure 5.3. The first row shows four slices of GLM t-map without using PCA denoising. The second row shows these slices when PCA denoising is applied. We can see that the t-maps in the second row show more activation. Especially on the images in the fourth column, the activation near the right superior parietal lobule becomes more noticeable after denoising.

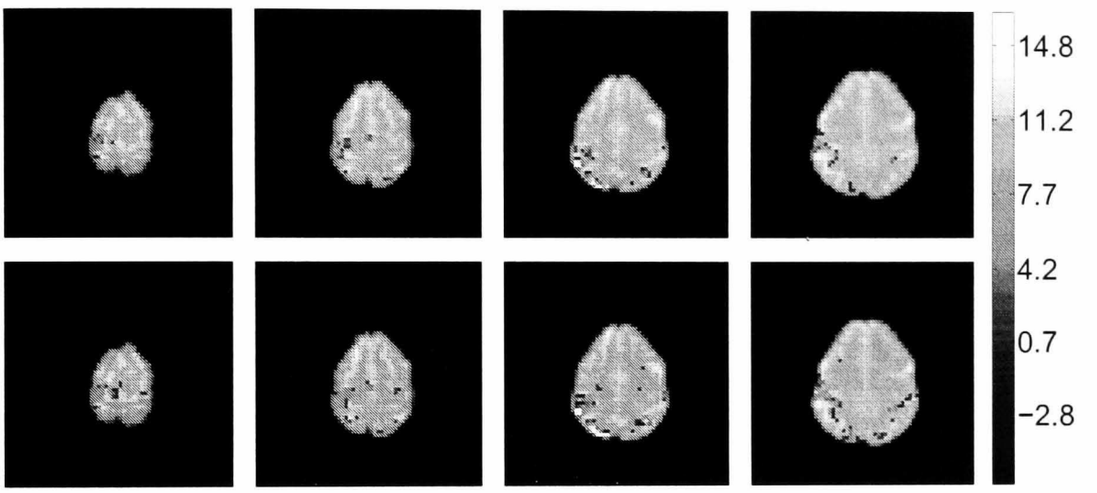


Figure 5.3: Statistical maps ($t > 5$) with and without PCA denoising. The slices in the first row show the result of t-test without using PCA denoising. The ones in the second row show the activation detection when PCA denoising is used. We use a large threshold ($t > 5$) to demonstrate the increase of detection sensitivity.

Parcellation

In order to examine our hypothesis and parcellation methods, we use different feature spaces and clustering methods to parcellate the whole brain into 600 parcels.

First, we directly apply k-means clustering with the spatial coordinates of all voxels. Each cluster is a parcel. In this parcellation, the voxels are grouped into parcel

according to Euclidean distance in the MNI space. The functional measurement of each voxel is not taken into consideration so the parcellation is purely spatial. We consider the results of this parcellation as a baseline. For convenience, we refer to this parcellation approach as Direct Clustering in the following parts of this thesis.

The second parcellation method is the one proposed by Thirion et al. [2006]. In this method, GLM parameters are used as feature space and Isomap is applied for parcellation. We argued in Chapter 3 that the high level noise may influence the parcellation. Thus we also use adaptive smoothing to improve this parcellation method.

Next, we implement our data-driven individual parcellation on this data set. In this approach, we use PLS correlation coefficients as feature space. In order to calculate these coefficients, we first decompose the fMRI data into ICs with the FSL toolbox. We then find the IC map whose corresponding time course has the maximal correlation coefficient with the BOLD model and 15 seed voxels are sampled from this IC map. Finally, PLS correlation coefficients are calculated from these seed voxels and clean PCs. After calculating these feature vectors, as in previous parcellation method, we use adaptive smoothing and Isomap for parcellation.

Finally, we implement our aggregation algorithm on the PLS map for parcellation. These parcellation results are compared in the next section.

5.2.3 Result analysis

In Section 3.4, we introduced two quantitative measurements to evaluate the parcellation results. In this section, we use the single-subject data to compare individual parcellation methods based on these quantitative measurements.

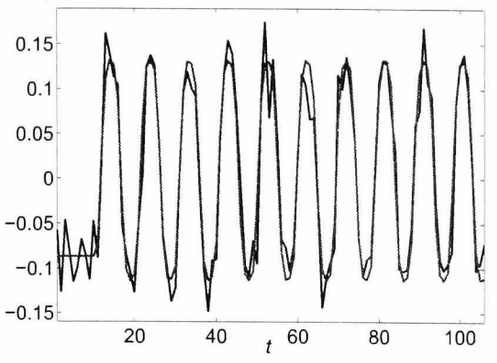
For both of these quantitative measurements, we first need to find a way to measure the functional behaviour of each voxel. We propose the use of GLM t-values and PLS t-values. In the following paragraphs, we compare these two approaches before discussing the comparison of parcellation results.

GLM t-values vs PLS t-values

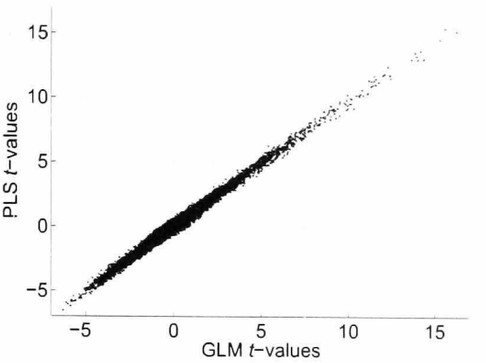
In Section 3.4.1, we introduce GLM t-value to evaluate intra-parcel homogeneity. In addition, we propose PLS t-values as a novel approach to the measurement of parcellation performance. As reviewed in Chapter 2, the GLM t-value is a well-studied criterion to detect functional activity. On the other hand, PLS t-values are comparatively new [McIntosh and Lobaugh, 2004; Rayens and Andersen, 2006]. Using the single-subject data, Figure 5.4 illustrates the performance of PLS t-values against GLM t- values.

The blue line in Figure 5.4.1 shows the PLS latent variable for PLS t-test and the red line is the GLM design. Figure 5.4.2 shows the scatter plot of GLM t-values and PLS t-values. In this plot, each dot corresponds to a voxel in the single subject data. The horizontal coordinate corresponds to the GLM t-values and the vertical coordinate represents the PLS t-values. According to this graph, PLS t-values give a performance similar to the GLM t-values. It can be considered to be an alternative approach to evaluating the intra-parcel homogeneity.

In the following paragraphs, we compare the parcellation results with intra-parcel homogeneity and Nearest Silhouette Coefficient. Based on the comparison, we try to answer the following three questions:



5.4.1: The PLS latent variable. The blue line shows the PLS latent variable and the red line shows the convolutional HRF model.



5.4.2: GLM t-values against PLS t-values. Each dot represent a voxels, whose corresponding GLM t-values and PLS t-values are the coordinates in this graph.

Figure 5.4: Comparison between GLM t-values and PLS t-values.

1. Does our feature extraction method improve the parcellation process?

2. Could smoothing on the manifold improve the final results?

3. Comparing the manifold based algorithm, how does aggregation algorithm perform?

Intra-parcel homogeneity results

We use different feature extraction methods and parcellation methods to generate a group of parcellation results, in which each subject is parcellated into 600 parcels. For each parcellation result, we calculate the intra-parcel functional variance $v(\mathbf{P}_i)$ for each parcel $\mathbf{P}_i$, where $i = 1, 2, ..., 600$. For each parcellation result, we consider these 600 values as a sampling distribution that represents the overall intra-parcel homogeneity of this parcellation result. In the following discussion, we compare the

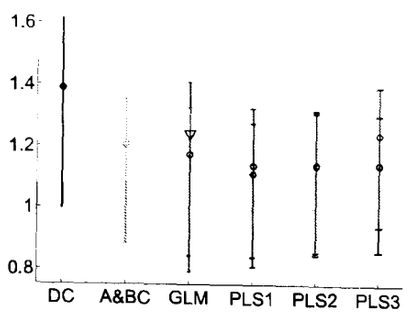
distributions corresponding to different parcellation results in two ways. We use error bars for visual comparison and t-test to compare the mean of these distributions.

In addition, we use four criteria to measure the corresponding intra-parcel homogeneity. These criteria are: GLM parameters, as in Thirion et al. [2006]; PLS correlation coefficients r, defined in equation 3.4.2, GLM t-values and PLS t-values, defined in section 3.4.1.

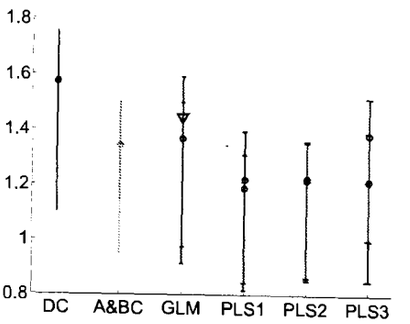
Figure 5.5 shows the visual comparison of parcellation results with the criteria stated above. Figure 5.5.1 shows the comparison based on GLM t-values. Each error bar represents one parcellation result. The dot in the middle of the error bar represents the mean value of $v(\mathbf{P}_i), i = 1, 2, ..., 600$. The top and bottom of the bar correspond to the first and third quartile of these intra-parcel variance values. The parcellation that gives high homogeneous parcels corresponds to the error bar with lower values.

The horizontal axis and the colours of the error bars represent the parcellation methods. The black bar corresponding to 'DC' shows the result of Direct Clustering. The green error bar above 'A&BC' shows the result of the Aggregation and Boundary Competition method with PLS latent variables. The rest of the error bars show the parcellation results based on Isomap. We use red colour to mark the results of Isomap with adaptive smoothing, whilst blue error bars represent the Isomap based parcellation results without the use of adaptive smoothing. 'GLM' means GLM parameters are used for parcellation based on Isomap. 'PLS1', 'PLS2' and 'PLS3' represent Isomap-based parcellation with the first 1, 2 and 3 PLS latent variables.

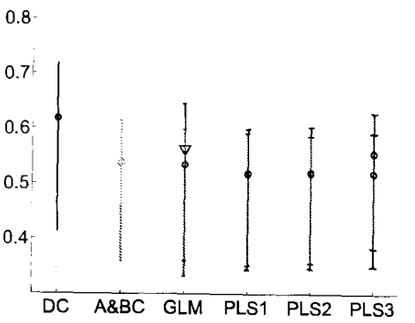
The error bars in Figure 5.5.2 – 5.5.4 use the same notation. In Figure 5.5.3, $v(\mathbf{P}_i)$ is calculated with GLM parameters. However, we consider that the comparison



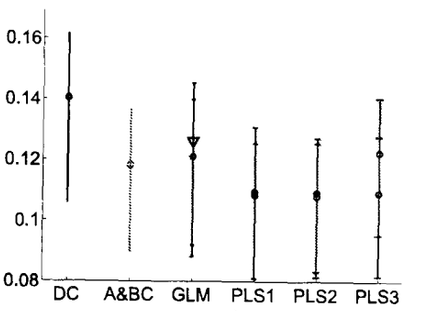
5.5.1: Comparison with GLM t-values.



5.5.2: Comparison with PLS t-values.



5.5.3: Comparison with GLM parameter β.



5.5.4: Comparison with PLS correlation coefficient r.

Figure 5.5: Parcellation results comparison with four different criteria. In each graph, the black bar represents the result of direct clustering in Euclidean space, which can be considered as a base line. Green error bar shows the result of the Aggregation and Boundary Competition method with PLS latent variables. The blue bars show results of Isomap with adaptive smoothing. And the red bars show results of Isomap with adaptive smoothing.

results in Figure 5.5.1 and Figure 5.5.2 are more reliable. This is due to the fact that these criteria bring noise into consideration. According to these graphs, all of the parcellation methods give better results than baseline.

Here, we try to answer the first question proposed earlier in this section: Does our

feature extraction method improve parcellation accuracy?

We first assume that when using the same spatially constrained clustering method to parcellate the whole brain into the same number of parcels, a better feature extraction method leads to parcellation results with higher functional intra-parcel homogeneity.

We first compare the blue error bars. These are the results when using Isomap directly on different feature spaces without adaptive smoothing. In Figure 5.5, the parcellation based on PLS gives a lower intra-parcel variance than the one based on GLM. We can see that when using an extra latent variable, the performance of PLS decreases. The reason is that, as the stimulation of the experiment is simple, one PLS latent variable is enough to describe the BOLD signal. Redundant latent variables bias the parcellation process.

In addition, we use the t-test to compare the parcellation results based on GLM and PLS. In this test, we consider results corresponding to two blue error bars in each figure: the one corresponds to 'GLM' and the other corresponds to 'PLS1'. The null hypothesis is that $v(\mathbf{P}_i), i = 1, 2, ...600$ of these two sets of parcels are from the same distribution. If the hypothesis is rejected, it means that these two parcellation results are statistically different. As can be seen from Figure 5.5, 'PLS1' has lower mean values. Therefore, a rejection of this hypothesis means that our feature extraction method leads to higher intra-parcel functional homogeneity. The results of the t-test are listed in Table 5.1.

Each column in Table 5.1 represents a way of measuring the intra-parcel variances. The corresponding p-values are listed in this table. They are all smaller than 0.05,

	Measurement of intra-parcel variances			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value	0.002	8.85×10^{-7}	0.029	6×10^{-10}

Table 5.1: Comparison of GLM and PLS based parcellation with t-test (Using Isomap parcellation without adaptive smoothing).

which means that the null hypothesis should be rejected. Our feature extraction method provides higher intra-parcel functional homogeneity.

Then, we check the red error bars. They represent the results when using adaptive smoothing and Isomap for parcellation. We can reach the same conclusion from these red error bars as from the blue ones. However, the bias caused by the redundant latent variables is more obvious. It can be explained as that, when using adaptive smoothing to improve the parcellation algorithm, the results are more sensitive to the bias brought by the feature vectors.

We also apply the same t-test on these parcellation results. The results of the t-test are listed in Table 5.2.

	Measurement of intra-parcel variances			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value	0.35	0.001	0.4	3.3×10^{-6}

Table 5.2: Comparison of GLM and PLS based parcellation with t-test (Using Isomap parcellation with adaptive smoothing).

According to this test, when using GLM t-values and GLM parameters to measure the intra-parcel functional variances, these two feature spaces give similar results.

When using PLS t-values and PLS correlation coefficients to measure the intra-parcel functional variances, our feature extraction method increases the intra-parcel homogeneity.

Combining the above analyses, we can conclude that generally our feature extraction method improves the parcellation result from the perspective of intra-parcel homogeneity. Although, in some situations, the improvement is not significant, we are able to answer the first question: when choosing the right number of latent variables for parcellation (for this set of data, 1 latent variable), our feature extraction method gives a better performance than GLM parameters.

Next, we try to answer the second question: Could smoothing on the manifold improve the final results?

In order to answer the second question, we compare the blue bars and the red bars in Figure 5.5. It should be noticed that all measurements for intra-parcel variances are calculated on the images without smoothing. Therefore, it should give some bias to the results without smoothing. However, for parcellation based on GLM parameters, adaptive smoothing gives a noticeable improvement. Table 5.3 shows the p-values of a t-test. The null hypothesis is that adaptive smoothing does not decrease the intra-parcel variance.

For the results based on one PLS latent variable, except for Figure 5.5.3, smoothing increases the functional homogeneity. However, the improvement is not obvious. Similarly we also use the t-test to examine these two distributions. Table 5.4 shows the p-values of the t-test. The null hypothesis is that adaptive smoothing does not decrease the intra-parcel variance.

	Measurement of intra-parcel variances			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value	0.042	0.048	0.16	0.08

Table 5.3: Comparison of GLM based parcellation with and without adaptive smoothing.

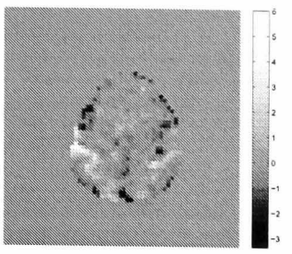
	Measurement of intra-parcel variances			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value	0.62	0.72	0.93	0.89

Table 5.4: Comparison of PLS based parcellation with and without adaptive smoothing.

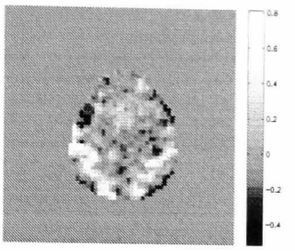
According to these t-values, adaptive smoothing does not provide effective improvement when using one PLS latent variable for parcellation. We use Figure 5.6 to explain this difference. Figures 5.6.1 and 5.6.2 show the images of GLM parameters and PLS correlation coefficients. Figures 5.6.3 and 5.6.4 are the corresponding histograms. From these figures, we can see that the manifold of PLS correlation coefficients is smoother than that of GLM parameters. Therefore, adaptive smoothing is more effective on parcellation with GLM parameters.

When using more than one PLS latent variable, due to the bias introduced by the redundant latent variables, the optimization object deviates from the desired one. The more optimal the algorithm could reach, the worse the final result would be. Therefore, adaptive smoothing cannot offer any advantage.

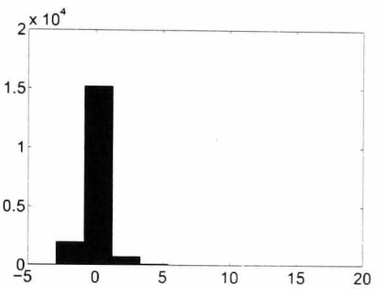
According to the above analysis, we can answer the second question: adaptive



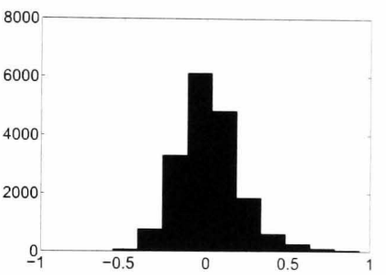
5.6.1: Image of GLM parameters.



5.6.2: Image of PLS correlation coefficients.



5.6.3: Histogram of GLM parameters β.



5.6.4: Histogram of PLS correlation coefficients.

Figure 5.6: Comparison between GLM parameters and PLS correlation coefficients.

smoothing improves the manifold based parcellation algorithm when the noise level is high. In addition, we want to mention that the automatically selected Gaussian kernel widths for smoothing are between 1mm and 2mm. These kernel widths are much smaller than the ones commonly used after spatial normalization.

Finally, we answer the third question: comparing the manifold based algorithm, how does our aggregation algorithm perform?

The green bar shows the results of the Aggregation and Boundary Competition algorithm. We use one PLS latent variable to calculate feature vectors. The parameters

of this algorithm are selected with the intuitive method introduced in section 3.3.2. In all the graphs in Figure 5.5, the results from this algorithm reach the performance of Isomap based method. Table 5.5 shows the results of the t-test. The null hypothesis is that the result of the Aggregation and Boundary Competition algorithm based on one PLS latent variable gives the same intra-parcel variance as the Isomap parcellation based on GLM parameters. According to these tests, this method gives similar or better results in comparison with the method based on GLM and Isomap.

	Measurement of intra-parcel variances			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value	0.23	0.031	0.22	0.016

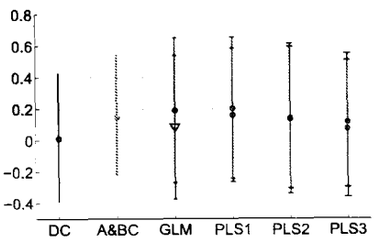
Table 5.5: Comparison between Aggregation parcellation with PLS and Isomap parcellation with GLM.

From another point of view, the computation required for this algorithm is much less than the Isomap algorithm. Therefore, compared with the manifold base methods, this algorithm gives an efficient and reasonable result.

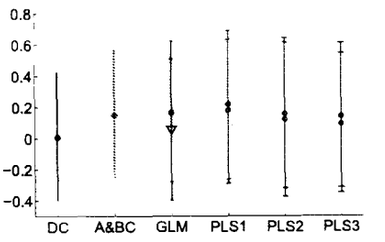
Nearest Silhouette Coefficient Results

In section 3.4.2, we propose Nearest Silhouette Coefficient (NSC) as a novel method to validate the results of parcellation. This method gives each voxel a coefficient to measure how well this voxel is assigned. For each parcellation result, we have N_v coefficients, where N_v is the number of voxels. A high coefficient means that the corresponding voxel is functionally closer to the assigned parcel than the second-nearest parcel. Each parcellation result is presented as a distribution of N_v NSCs.

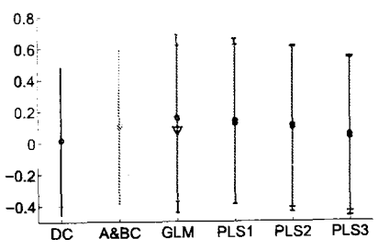
Figure 5.7 shows the evaluation results based on NSCs. The annotation is the same as that used in Figure 5.5. We also use four functional measurements to evaluate the results. They are listed in the four sub-figures. The black error bars represent the results from purely spatial clustering. As the parcellation does not include any functional information, the mean values of the corresponding NSCs are close to 0. Other parcellation methods give higher mean NSC values.



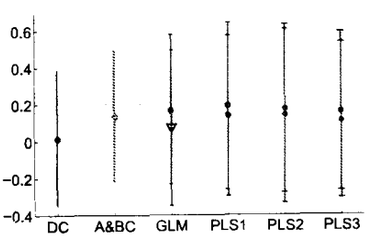
5.7.1: Comparison with GLM t-values.



5.7.2: Comparison with PLS t-values.



5.7.3: Comparison with GLM parameter β.



5.7.4: Comparison with PLS correlation coefficient r.

Figure 5.7: Parcellation results comparison based on Nearest Silhouette Coefficient with four different criteria. The graph legend is the same as Figure 5.5.

According to Figure 5.7, the evaluation based on Nearest Silhouette Coefficient generally indicates the same conclusion as the evaluation based on intra-parcel homogeneity. Here, we also use the t-test to compare the NSCs from different parcellation results. Based on these comparisons, we answer these three questions proposed

earlier this section.

	Measurement of NSCs			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value1	0	0	0	0
p-value2	0.0212	0.0372	0.3394	0.0002

Table 5.6: Comparison between PLS and GLM based parcellation with NSCs.

We compare the results of the Isomap parcellation with GLM and one PLS latent variable. First, the parcellation is directly applied to the feature vectors without adaptive smoothing. These results are presented in Figure 5.7 with blue error bars corresponding to 'GLM' and 'PLS1'. The null hypothesis is that the results from 'GLM' and 'PLS1' are sampled from the same distribution. Four functional measurements provide four p-values. They are listed in the first row of Table 5.6. In the next step, we compare the results when adaptive smoothing is used. These results are shown with the red error bars corresponding to 'GLM' and 'PLS1' in Figure 5.7. We use the same null hypothesis. The p-values are listed in the second row of Table 5.6. These comparisons show that our feature extraction method increases NSCs of the parcellation.

Next, we compare the results with and without using adaptive smoothing to answer the second question. We first examine the performance of adaptive smoothing on GLM based Isomap parcellation. Using four functional measurements, we test the hypothesis that the use of adaptive smoothing does not increase the result from NSCs. The resulting p-values are shown in the first row of Table 5.7. These

p-values support the conclusion that adaptive smoothing does improve the parcellation results. Then, we investigate whether adaptive smoothing can still increase NSCs when using one PLS latent variable for parcellation. We compare the blue and red error bars corresponding to 'PLS1' and hypothesize that they are from the same distribution. The results are shown in the second row of 5.7. For PLS based parcellation, adaptive smoothing does not provide the same improvement.

	Measurement of NSCs			
	GLM t-values	PLS t-values	GLM parameter β	PLS coefficient r
p-value1	0	0	0	0
p-value2	0.0527	0.0011	0.0416	0.0202

Table 5.7: Comparison of Isomap parcellation with and without adaptive smoothing.

Finally, we compare the result of our Aggregation and Boundary Competition (green error bars) with the method based on GLM and Isomap (blue error bars corresponding to 'GLM'). The aggregation method is applied to the feature vector calculated with one PLS latent variable. We still use the t-test and four functional measurements to compare these distributions. All of them are smaller than 10^{-4}. Thus, we can conclude that our method provides effective and efficient parcellation.

5.3 Experiment on multi-subject face and gesture data

5.3.1 Data

Grosbras and Paus [2006] have used fMRI to identify the brain regions engaged during the observation of hand action and facial expression, performed either in a neutral or an angry way. Here we use the same data. The data was scanned in the Brain and Body centre, University of Nottingham. Fourteen adults participated this experiment. Some of them were scanned twice at different times. Altogether, there were 25 scans treated as 25 different subjects. The data was also acquired on a Philips Intera 1.5T scanner. The TR of the scan was 3 seconds. We used four types of stimuli: neutral hand gesture, angry hand gesture, neutral face expression and angry face expression. The stimuli consisted of short (2 - 5 seconds) black-and-white video clips depicting either a hand action or a face in movement.

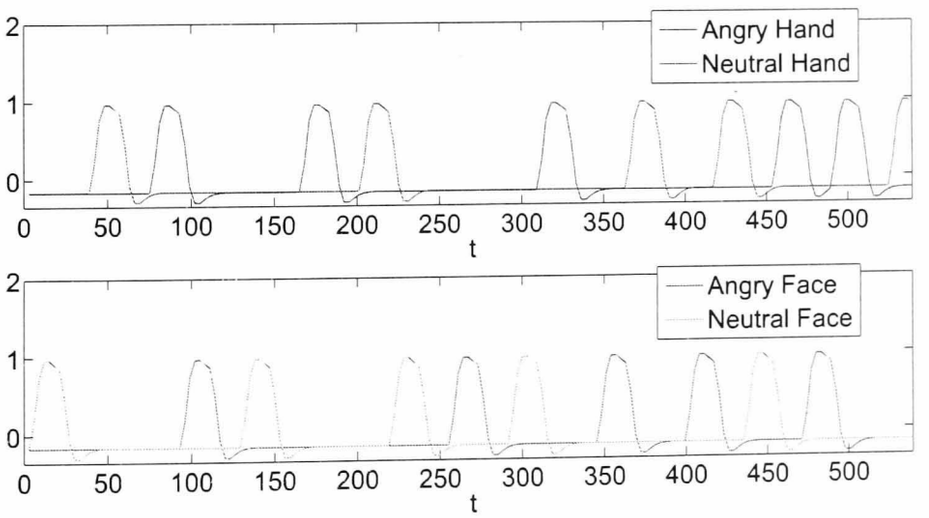


Figure 5.8: Convolutional HRF model for the multi-subject face and gesture data.

Figure 5.8 shows the convolution model of this experiment. Four solid lines rep-

resent the models of the BOLD signals corresponding to the four types of stimuli. More details of the experiment and scanning can be found in Grosbras and Paus [2006].

5.3.2 Parcellation

The individual parcellation process for single-subject data and multi-subject data is generally the same. The key difference is the way of choosing the seed voxels. In multi-subject parcellation, we try to find the IC time courses that are reproducible across all subjects. In this section, we focus on the discussion of whether the proposed IC clustering method could find the reproducible ICs.

The clustering of ICs

We first use the FSL toolbox for slice timing correction, motion correction and PICA decomposition. After that, we pool these ICs together and apply PCA to the pooled ICs. Figure 5.9 shows the eigenvalues and the projection of the ICs in the direction of the first three PCs. Unfortunately, there is no obvious structure in the data. It means that clustering the ICs directly according to Euclidean distance cannot give satisfactory results.

Then, we use the method introduced in section 3.2.2 to define the similarity between the ICs. Based on this definition of similarity, we apply hierarchical clustering on these ICs. Figure 5.10 shows the dendrogram of the clustering result with the vertical axis representing the distance between the clusters. From this figure, we find that two clusters are very far from the rest of the ICs. Cluster 1 is marked in green

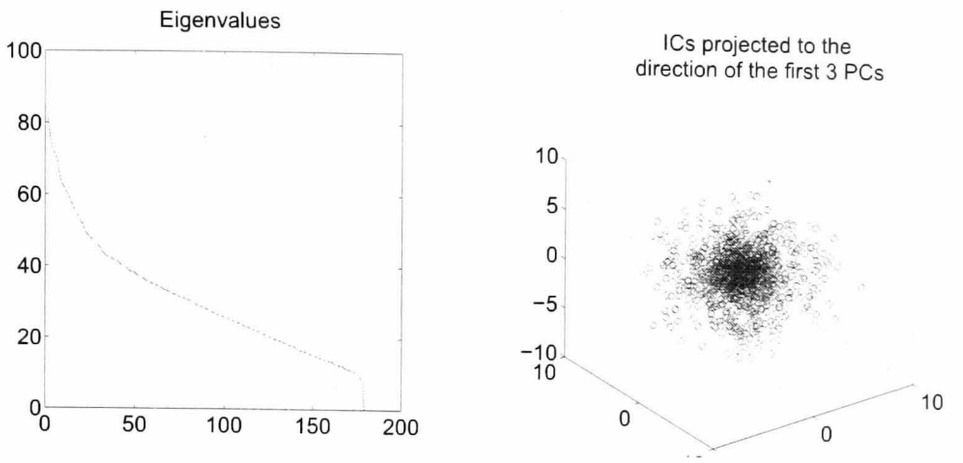


Figure 5.9: PCA analysis of pooled ICs.

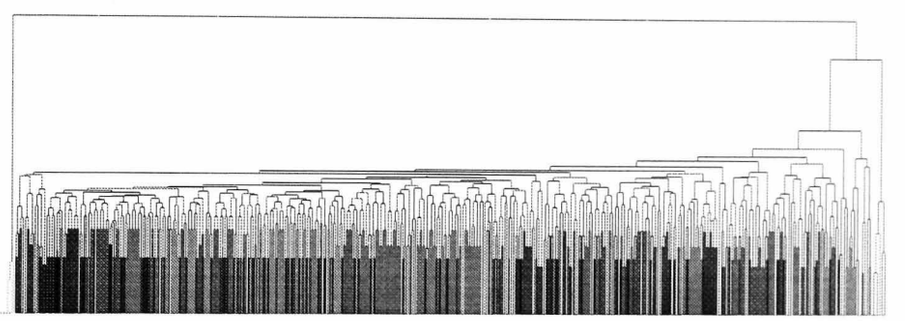
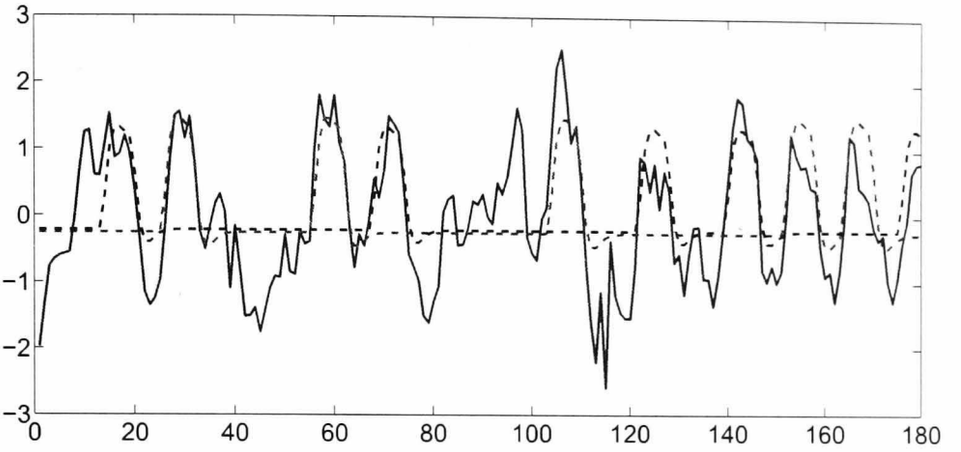


Figure 5.10: The dendrogram of hierarchical clustering the ICs from all subjects. There are two clusters (red and green) of ICs that are far from other ICs (blue).

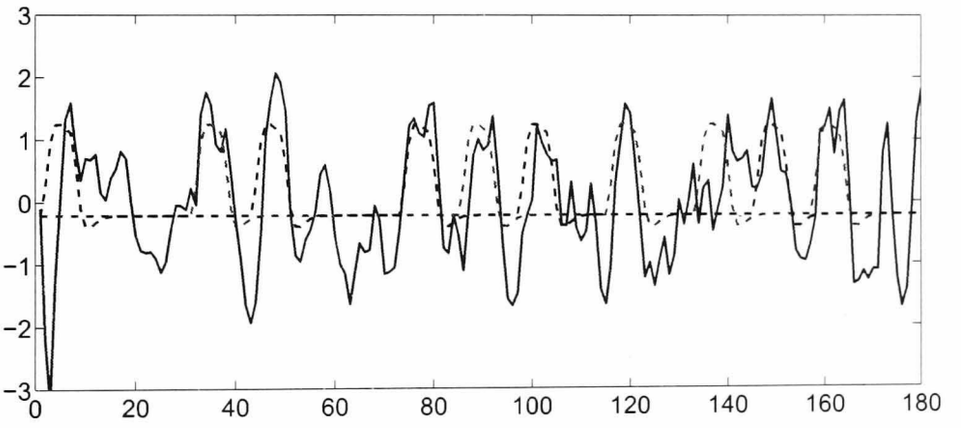
and Cluster 2 is marked in red. In Cluster 1, there are 22 ICs each of which is from a different subject. In Cluster 2, there are 20 ICs. As in Cluster 1, the ICs in Cluster 2 are all from different subjects.

We further examine these two clusters in Figure 5.11. Figure 5.11.1 shows an IC from Cluster 1 (the blue solid line) which matches the BOLD models of the viewing the angry and neutral hand gesture (the red and black dashed lines). In Figure 5.11.2, the blue solid line shows an IC from Cluster 2. This IC matches the HRF models of

viewing the angry and neutral facial expression (the red and black dashed lines).



5.11.1: An IC in Cluster 1, the dashed red and black lines represent the convolutional HRF models of neutral and angry hand gesture stimulation.



5.11.2: An IC in Cluster 2, the dashed red and black lines represent the convolutional HRF models of neutral and angry face expression stimulation.

Figure 5.11: ICs in different clusters.

In addition, we apply manifold embedding to the ICs according to the distance defined in equation 3.2.7. Figure 5.12 shows the embedding results. The left figure shows the first 20 eigenvalues. In the right figure, each point represents an IC embedded in two dimensional Euclidean space. In this figure, we can see two distinc-

tive clusters. Compared with the above hierarchical clustering results, we find that these two clusters are the same as the green and red clusters in Figure 5.10.

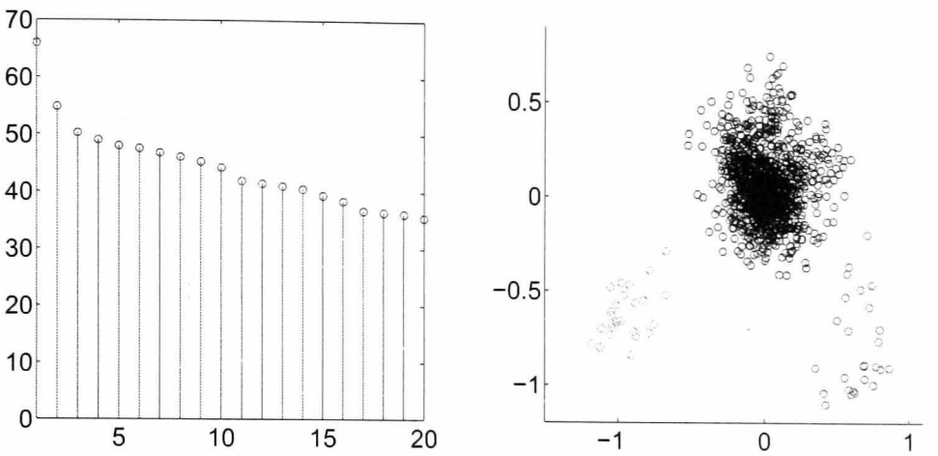


Figure 5.12: The clustering of ICs based on manifold.

According to the clustering results, we have two sets of ICs each of which corresponds to a BOLD response. However, there are a few subjects who have no ICs that are organized into these two clusters. In such a case, we attribute the IC that is closest to each IC cluster to the corresponding cluster. This means that in each cluster, there is one IC from each subject. Seed voxels are sampled according to the IC maps of these ICs. From each IC map, we sample 10 seed voxels. These seed voxels are used to calculate the PLS latent variables. Finally, the parcellation algorithms are implemented on the corresponding PLS correlation coefficients.

5.3.3 Result analysis

In section 5.2.3, we used the single-subject data to discuss whether the use of PLS as feature extraction method could increase the intra-parcel homogeneity. Here, we use the multi-subject to compare PLS and GLM based parcellation. The comparison is

based on two evaluation methods: intra-parcel homogeneity and Nearest Silhouette Coefficient (NSC). We use this comparison to further examine the question: does our feature extraction method improve the parcellation process?

Intra-parcel homogeneity results

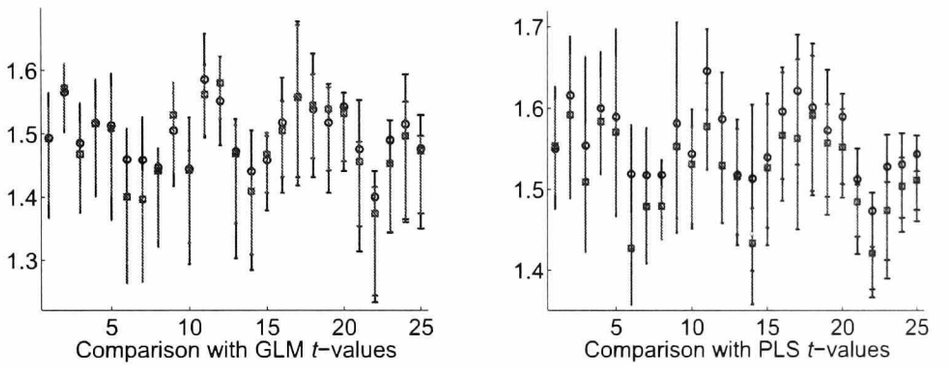


Figure 5.13: Comparison of functional intra-parcel homogeneity. The blue error bars represent results of parcellation with GLM parameters. The red error bars represent results of parcellation with PLS covariance coefficients.

First, using the GLM parameters as feature vectors, we parcellate each subject into 600 parcels with Isomap parcellation. Before parcellation, we apply adaptive smoothing to the feature space. The automatically selected kernel widths are between 1.2mm and 2.2mm. Then, we use PLS to calculate feature vectors and apply the same process for parcellation. After that, the intra-parcel homogeneity is evaluated with the variance of GLM t-values and PLS t-values within each parcel. We compare the results in Figure 5.13.

Similar to section 5.2.3, Figure 5.13 shows the intra-parcel variance of GLM t-values

Functional Measurements	p-values						
GLM t-values	0.961	0.712	0.345	0.962	0.811	0.019	0.007
	0.788	0.171	0.926	0.131	0.060	0.874	0.126
	0.652	0.499	0.947	0.736	0.239	0.559	0.357
	0.269	0.109	0.369	0.865			
PLS t-values	0.843	0.183	0.012	0.293	0.269	0.000	0.024
	0.008	0.106	0.432	0.000	0.001	0.895	0.000
	0.434	0.076	0.001	0.550	0.309	0.015	0.081
	0.002	0.002	0.091	0.029			

Table 5.8: Comparison of GLM and PLS parcellation with multi-subject data.

(left) and PLS t-values (right). Each error bar shows one parcellation result. The blue and red bars show the results of parcellation based on GLM and PLS. The horizontal axis represents the indices of subjects. The lower the error bar the more homogeneous is the corresponding parcellation result.

In addition, we use the t-test to compare these parcellation results. Using GLM t-values and PLS t-values as functional measurements, we implement the t-test on the two parcellation results of each subject. The hypothesis is that both of the parcellation results are sampled from the same distribution. The corresponding p-values are listed in Table 5.8. The first four rows show the p-values when using GLM t-values to measure the intra-parcel functional variance. The last four rows correspond to the p- values when PLS t-values are used to measure the intra-parcel functional variance. The p-values marked in blue colour show that, for the corresponding sub-

ject, GLM parcellation gives lower intra-parcel functional variance. Others show that PLS parcellation gives lower intra-parcel functional variance.

Combining Figure 5.13 and Table 5.8, we can give the following summary of this comparison. When using GLM t-values to examine the intra-parcel homogeneity, PLS gives similar results to those of GLM. When using PLS t-values to examine the results, PLS performs better.

Nearest Silhouette Coefficient (NSC) results

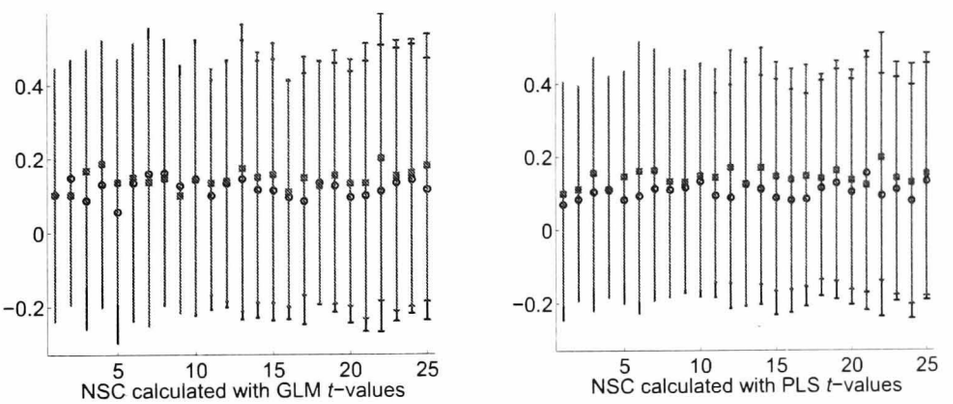


Figure 5.14: Comparison of parcellation results with NSC. The blue error bars represent results of parcellation with GLM parameters. The red error bars represent results of parcellation with PLS covariance coefficients.

Figure 5.14 shows the comparison between the GLM and PLS parcellation based on NSC values. As in the last section, the distribution of the NSC is shown with an error bar. Higher NSC values indicate better parcellation results. The left graph in Figure 5.14 shows NSCs calculated with GLM t-values. The one on the right shows NSCs based on PLS t values.

We also compare these results with the t-test. The null hypothesis is that GLM and

Functional Measurements	p-values						
GLM t-values	0.665	0.000	0.000	0.000	0.000	0.003	0.000
	0.000	0.000	0.462	0.000	0.188	0.000	0.000
	0.000	0.000	0.000	0.029	0.000	0.000	0.000
	0.000	0.000	0.000	0.000			
PLS t-values	0.000	0.000	0.000	0.000	0.574	0.000	0.000
	0.000	0.000	0.000	0.000	0.000	0.981	0.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	0.000			

Table 5.9: Comparison of GLM and PLS parcellation with multi-subject data.

PLS parcellation give the same results. We list all p-values in Table 5.9. For each subject, if GLM parcellation gives higher mean value of NSCs, we mark the corresponding p-values in Blue colour.

To start with, we examine the first four rows of Table 5.9. The p-values in these rows represent the comparison results when using GLM t-values to calculate NSCs. For 18 of the 25 subjects, PLS parcellation leads to higher NSCs. And for 17 of these18 subjects, PLS parcellation provides a statistically significant improvement.

The last four rows show the p-values when using PLS t-values to calculate NSCs. According to these p-values, for 24 of the 25 subjects, PLS parcellation gives higher NSCs, and for 22 of these 24 subjects, the improvement is statistically significant.

Therefore, we can conclude that, when using NSC to measure the parcellation results, our PLS parcellation provides similar or better results for these 25 subjects.

5.4 Cross-Subject parcel matching

The data set was acquired from 25 subjects viewing angry gestures or expressions. Scanning was performed on a Philips Intera 1.5T, with TR=3s. During the scan, four types of visual stimuli are given to the sub jects, which are angry hand gestures, neural hand gestures, angry facial expression and neural facial expression. Using PLS latent variables, we parcellate each subject into 600 parcels with adaptive smoothing and Isomap. Using the OBSA algorithm, we match the parcels from all of the 25 subjects. The distance is defined with the method introduced in the last chapter. In Section 4.4.2, we showed the numerical result of this matching. Here, we illustrate the application of t-statistics on the matched parcels.

We define a *clique* as a set of parcels. Each clique has 25 parcels from different subjects. The parcels in the clique are matched with each other with the OBSA algorithm. For each clique, we calculate two values: group parcel t-value and intra-clique weight. The group parcel t-value represents the functional activity of the clique corresponding to each stimulus. The intra-clique weight is the sum of weights connecting all the parcels in the clique. This weight represents how well the parcels in this clique are matched with each other.

Figure 5.15 shows the parcels that are activated with respect to the stimulus of the 'angry hand gesture'. We first calculate t_i, the GLM t-values for each voxel i. The voxels in each clique is considered as a group and the t-values for each clique C are calculated as:

$$t_{\mathbf{C}} = \frac{1}{C}\sum_{i\in\mathbf{C}} t_i \times \frac{\sqrt{C-1}}{\sqrt{1+\frac{1}{C}\sum_{i\in\mathbf{C}}(t_i - \sum_{i\in\mathbf{C}} t_i)}}, \tag{5.4.1}$$

where C is the number of voxels in clique $\mathbb{C}$.

The first image (left top) in Figure 5.15 shows the group statistical map in standard space. Smoothing is applied after spatial normalization to increase the sensitivity. The blue colour bar indicates the corresponding t-values. Other images show the activation maps of matched parcels (with t-values greater than 4) in 25 subjects. The yellow-red colour bar shows the t-values of each clique. Due to the fact that the parcels in the same clique shares the same t-values, they are marked with the same colour on this map. Comparing the parcel-based analysis with the standard voxel-based analysis, we find that the activation near the inferior frontal gyrus is enhanced.

Figure 5.16 shows the weight of each clique. The first image is the same as the one in Figure 5.15. Other images illustrate the weight of each clique. For each parcel in each image, the lower the weight, the better the corresponding clique is matched.

Next, we compare two parcel-matching methods. We first take one subject as reference, and match all other subjects to the reference subject. When using this approach, taking subject 16 as reference gives the minimal cost, as shown in Figure 4.5. Thus, we match all other subjects with subject 16. Here, we consider all the parcels that matched to the same parcel in the reference subject as a clique and we calculate the t-values for each clique. The left image in Figure 5.17 shows an activation map of one subject.

In contrast, we use our OBSA method to calculate the clique and activation maps ($t > 4$). The corresponding activation map is shown in the right image of Figure 5.17. The right image shows one more parcel, which is not detected in the left one.

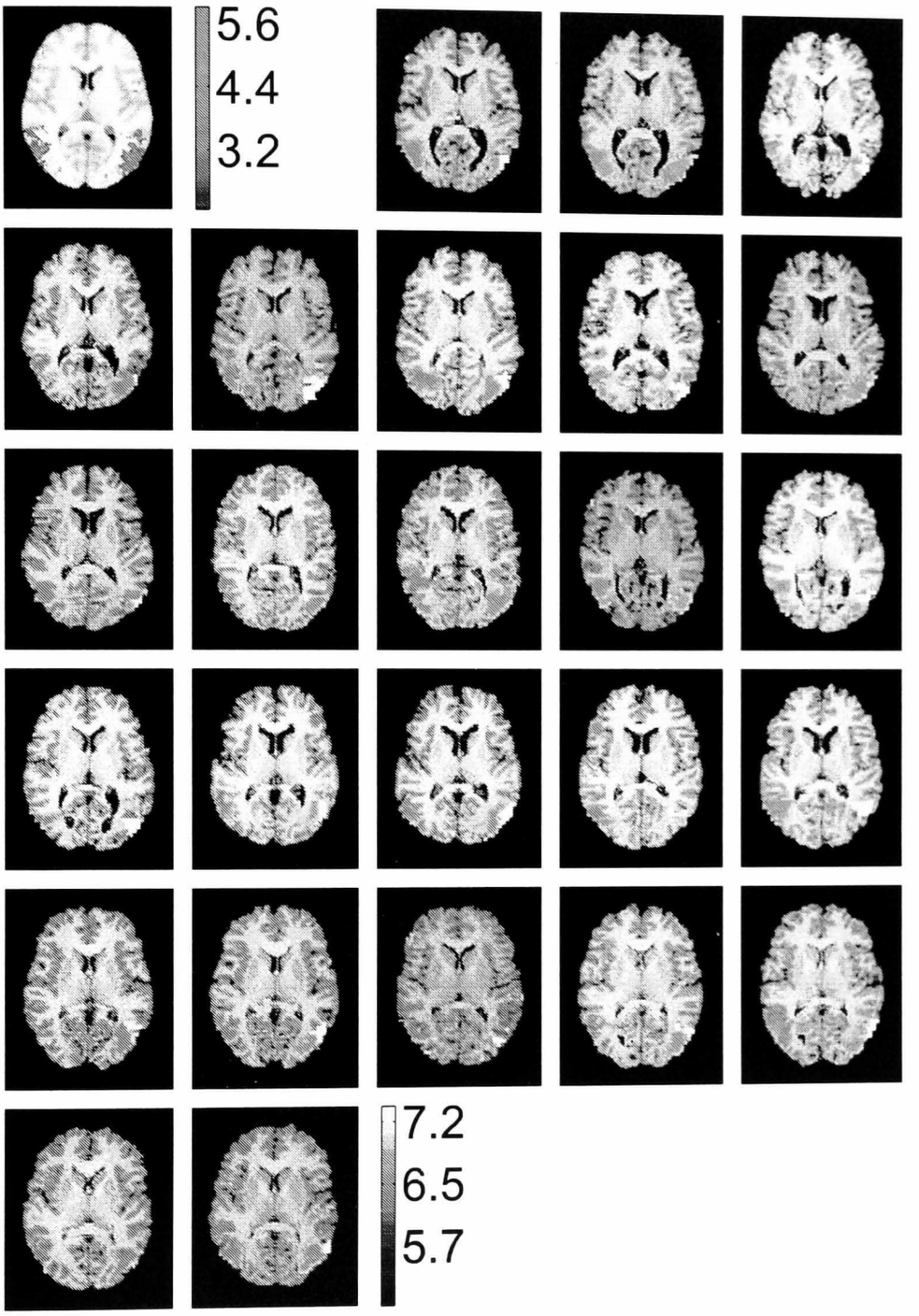


Figure 5.15: Two types of group analysis for the 'angry hand gesture' stimulation. The first image (left top) shows the a group t-statistical map in standard space. The blue-green bar represents t-values. Others show the activation maps of matched parcels in 25 subjects. The yellow-red bar shows the t-values of the matched parcels.

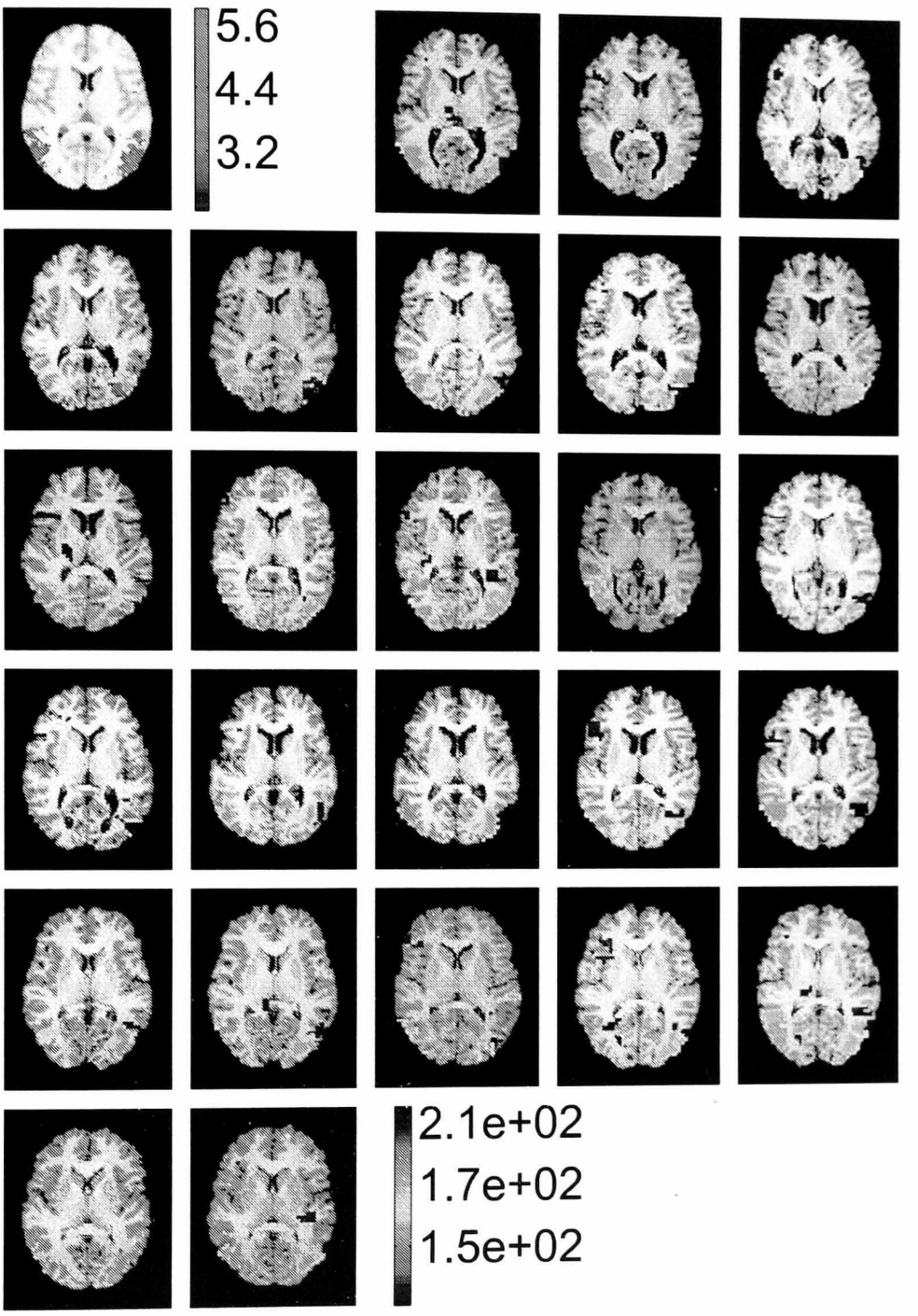


Figure 5.16: The sum of weights of the matched parcels. The first image and the blue-green bar are the same as in Figure 5.15. Other images show the sum of weights of cliques. The yellow-red bar shows the corresponding values. The smaller the sum of weights, the better the corresponding parcels are matched.

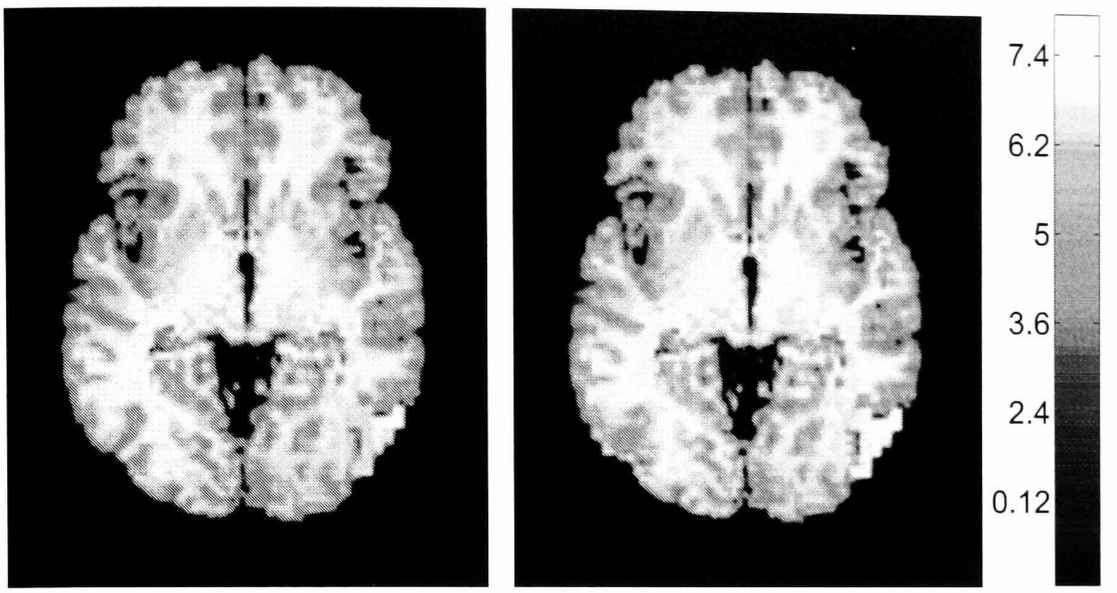


Figure 5.17: Comparison of parcel-matching. We use two methods for the parcel-matching. The first one is the intuitive way in which all subjects are matched to subject 16. The second method is the multi-partite graph partitioning method. After parcel-matching, t-statistical maps are constructed on the matched parcels. The left figure shows the result of intuitive matching and the right figure shows the result of matching method based on multi-partite graph partitioning.

It means that our OBSA method increases the sensitivity of the statistical analysis based on parcels.

5.5 Summary

In Chapter 3, we discussed parcellation methods for individual subjects. In this chapter, we used two fMRI data sets to examine these parcellation methods. The first question is whether our data-driven feature extraction method improves parcellation accuracy. To answer this question, we applied the same algorithm to our

feature vectors and GLM parameters. In addition, we used two different validation approaches to compare parcellation results. We found that parcellation based on our feature extraction method gives better parcellation accuracy in most cases.

In addition, we used adaptive smoothing to improve parcellation based on Isomap. Experimental results show that adaptive smoothing increases the functional homogeneity and NSC values, especially for the feature vectors that contain high-level noise.

Finally, we examined the performance of our Aggregation and Boundary Competition algorithm for parcellation. Our experimental results show that this algorithm could give reasonable parcellation accuracy with a reduced computation requirement.

Therefore, from these experimental results and analyses, we conclude that our data-driven parcellation approach improves over model-based parcellation from two perspectives: (1) it does not need to presume HRF model; (2) it can reach and even give better parcellation accuracy than the model-based parcellation method.

Next, we considered the methods that match the parcels from individual subjects. Here, we examined the hypothesis that the use of information from all subjects for parcel matching improves the accuracy of parcel-matching and further analysis. Our multipartite graph partitioning model and OBSA algorithm is a method that uses the information from all subjects. We compared the results from our method with results from the intuitive method that uses one subject as reference. The numerical comparison shows that our method provides a more effective and efficient result. In addition, we applied statistical analysis on the matched parcels. The results

indicate that our method increases the sensitivity of activation detection. Furthermore, when we compared our methods with the standard voxel-based analysis, we found that our method enhances the activation detection. Therefore, our approach improves over the standard method in terms of robustness and sensitivity.

CHAPTER 6

Conclusion and Future Work

This thesis proposes a framework for data-driven fMRI analysis based on parcellation. We argue that our multi-subject data-driven parcellation approach improves over (1) standard voxel-wise fMRI analysis in terms of both robustness and sensitivity to normalization issues and (2) model-based parcellation techniques in terms of parcellation accuracy. Two fMRI data sets were used to support our hypothesis. In the following sections, we first give a summary of this thesis. Then, we discuss future directions for fMRI data analysis and human brain parcellation.

6.1 Summary

As mentioned in Chapter 2, Functional MRI uses magnetic resonance to measure changes in blood oxygenation level, which is related to the brain activity. We presented three types of fMRI analysis methods: General Linear Models (GLM), data-driven analyses and machine learning classifiers. GLM is arguably the most popular method for fMRI data analysis. However, it relies heavily on a priori BOLD models.

When these models are unavailable, GLM is not an appropriate method for brain activation detection. For instance, when researching the default mode of brain function or decoding mental states, we do not know the corresponding BOLD signals. Data-driven analyses provide effective alternative approaches. In this thesis, we review two classes of data-driven analysis: clustering analyses and ICA analysis. These methods are widely used for detecting brain activity and perform analysis based on the data structure itself. The principal advantage of these methods is their applicability to experimental paradigms in the absence of an a priori model of brain activity. One flaw, however, is that they do not provide an interpretation of the results.

In recent years, the use of machine learning classifiers has grown in popularity for fMRI analysis. This method could overcome the flaws in voxel-based inferential and exploratory multivariate approaches and help with the understanding of neural representation. Compared with other methods, a machine learning classifier is often more complex to implement and requires full cooperation between experts from different areas. With appropriate application, this method could provide an opportunity for more advanced neuroscience studies.

Parcellation has been proposed as a way of dealing with the shortcomings of spatial normalization for fMRI analysis. In this thesis, we provide a taxonomy of brain parcellation methods, divided into two classes: *top-down* and *bottom-up*. In the first approach, parcellation starts with the whole brain and gradually divides it into smaller regions. When dividing a region into sub-regions, the focus is on the anatomical or functional evidence that could prove the existence of these sub-regions. Whole brain

parcellation is a long-term research subject. In contrast, in the second approach, researchers first define a measure of similarity between voxels. According to this similarity, the whole brain is then parcellated into a certain number of regions. The accuracy of this parcellation depends on many issues, such as imaging quality, the chosen similarity, the actual parcellation algorithm etc. From our review of parcellation methods, we find that there is very little work using the second approach. Yet, there is a need for a method that could effectively and efficiently parcellate the whole brain into an arbitrary number of parcels. Therefore, we have developed our data-driven parcellation methods. Our multi-subject data-driven parcellation method has two major parts: data-driven individual subject parcellation and cross-subject parcel matching.

Data-driven individual subject parcellation

Our parcellation method can be considered as a two-step process. The first step is feature extraction. The main idea behind this step is the use of seed voxels and the PCA components of whole brain data to calculate the PLS latent variables. The correlation coefficients between the time course of each voxel and these PLS latent variables are used as feature space. With this method, the choice of seed voxels is very important: the seeds should be located in different activated regions. In this thesis, we propose to sample the seeds from IC maps, which have been widely accepted as an effective way to study brain activations for individual subjects. For multi-subject analysis, we use IC maps whose IC time courses are reproducible in all subjects.

The second step of this method is spatial constrained clustering. In previous parcellation studies, manifold based methods have been applied in this step. Thus, in this thesis, we investigate two manifold based methods: Isomap and Diffusion Map. Manifold based methods are sensitive to noise so that data sampled under high-level noise produces parcellation results with low intra-parcel homogeneity. In order to solve this problem, we propose an adaptive smoothing method as a preprocessing step. The experiment on 'Swiss Roll' data shows that this smoothing method is an effective way of dealing with the noise. Another disadvantage of manifold based methods is that they require a large amount of computation. Consequently, we develop a novel aggregation algorithm for parcellation which uses parameters to control the parcellation results. Correspondingly, we propose an intuitive method to estimate the optimal parameters. Our experiment results show that this method is more robust against noise. It can provide results comparable with manifold-based methods. Furthermore, the computation requirement is much less.

Another important issue for individual subject parcellation is the validation of parcellation results. In $bottom - up$ parcellation methods, each brain is parcellated into hundreds of parcels. It is difficult to examine every parcel and every boundary. Therefore, we need a qualitative measurement to make a global evaluation of parcellation results. Intra-parcel variance of the GLM parameters is the state-of-the-art method for this problem. However, due to the problem of noise and artefacts, GLM parameters do not always accurately represent the functional activation of each voxel. Therefore, we propose the use of GLM t-values and PLS t-values to cal-

culate intra-parcel functional homogeneity. We consider these validation methods to be more robust against noise. Besides intra-parcel homogeneity, we also propose another validation method: Nearest Silhouette Coefficient (NSC). This method gives each voxel a coefficient and the distribution of all voxels represents the quality of the parcellation. When the parcellation is implemented randomly, the distribution of NSCs should have a mean close to 0. The application of NSCs to a toy data set shows that this method is reliable.

Our main contribution in this part of this thesis are: (1) a data-driven method to calculate feature vectors, (2) an improvement of manifold based parcellation, (3) a novel aggregation based parcellation method, (4) an improvement on the measurement of intra-parcel homogeneity, (5) the use of NSC as a quantitative measurement for parcellation results.

Cross-subject parcel matching

After parcellation of each individual subject, the next step is to find the correspondence of the parcels across subjects. We propose a multipartite graph model which considers the parcels from all subjects as a complete weighted multipartite graph. In this graph, each parcel is a vertex that is connected only to the parcels from other subjects. Each edge is weighted by the dissimilarity between the parcels it connects. The problem of parcel matching becomes that of partitioning this graph into cliques, so that each clique has one parcel from each subject and the sum of weights in the clique is minimized. The main advantage of this technique is that it uses information from all subjects to find the best match of all parcels. According to the weights,

one could find outlier subjects with large dissimilarity to other subjects. However, according to previous studies (e.g. Crama and Spieksma [1992]; Garey and Johnson [1990]; Jebara [2003]), this multipartite partitioning problem is generally NP hard.

In order to solve this problem, we first present an intuitive method, which takes one subject as reference and matches other subjects with the reference subject. The parcels matched to the same parcel from the reference subject are considered as the vertices of a clique. However, the solution of this method is sub-optimal. It cannot use the information in all subjects to find the best match. To overcome this problem, we develop an Oder Based Simulated Annealing (OBSA) method to estimate the optimal result.

In addition, we discover that our multipartite partitioning model is similar to the concept of 'Bags of Pixels' [Jebara, 2003], which is designed for image recognition. We argue that the idea of 'Bags of Pixels' is a special case of our model. In Jebara's work, quadratic programming is used to solve a similar problem. Therefore, we also try to use quadratic programming to solve our multipartite graph partitioning problem. A toy data set and a real multi-subject fMRI data set are used to evaluate these parcel matching methods numerically. In this experiment, the method based on quadratic programming cannot give a satisfactory result. Our OBSA algorithm gives better numerical results than the intuitive method.

Although we give a negative report of the idea of 'Bags of Pixels', our aim is certainly not to deny the contribution of this idea. The goal of our discussion is to make a connection between the areas of fMRI data analysis and those of image recognition so that techniques for image recognition can be brought into the area of fMRI data

analysis in future studies.

Our main contributions in Chapter 4 include: (1) a multipartite graph partition model for cross-subject parcel matching problem,(2) an OBSA algorithm, (3) fMRI data as bags of parcels.

To examine the main hypothesis of this work, we apply our parcellation method to two fMRI data sets (Chapter 5).

We first show that our data-driven parcellation method is an improvement over model-based parcellation techniques, in terms of parcellation accuracy. In order to compare different feature extraction methods, we apply the same spatially constrained clustering methods to our PLS correlation coefficients and GLM parameters. The parcellation results are compared with different validation methods. From this comparison, we find that our feature extraction method gives a better or a similar performance. We also apply different clustering methods to the same feature space to compare the performance of these methods. The results of this comparison show that adaptive smoothing improves the manifold based clustering. In addition, our aggregation-based method could provide parcellation accuracy comparable to manifold based parcellation. To sum up, our parcellation method improves over model-based techniques in two ways. (1) It is data-driven, therefore we do not need to assume an HRF model for the analysis. This method can also be applied to other model-free analyses such as resting-state studies or mental state recognition researches. (2) It reaches and even exceeds the model-based parcellation accuracy.

The second hypothesis of this thesis is that our parcellation method is an improvement over standard voxel-wise fMRI analysis in terms of both robustness and sensi-

tivity to normalization issues. The main advantage of our parcel-matching method is that by defining the distance between two parcels, the algorithm gives a global optimal result using the information from all parcels. Using the multi-subject fMRI data set, we compare our approach with the intuitive method and demonstrate that our approach improves the accuracy of the next step in fMRI analysis (t-statistic analysis in our case). Finally, comparing the statistical analysis based on our parcellation method and standard voxel based method, we find that our method increases the sensitivity of the group statistical analysis. Therefore, we conclude that our new approach to parcellation and parcel matching increases the robustness of fMRI analysis based on parcellation.

6.2 Future Work

6.2.1 Further improvement to single-subject parcellation

Parcellation

Due to the limitations of functional imaging (e.g. low resolution, a high level of noise), it could be desirable to combine it with other imaging modalities for whole brain parcellation. For instance, in parcellation with fMRI, it is difficult to define accurately the boundaries between parcels. We could use a high resolution structural image to clarify these borders. DTI could also provide structural information. We would probably use a Bayesian model to combine them as a priori knowledge to functional parcellation.

The integration of *bottom-up* methods with *top-down* approaches would also most probably improve parcellation accuracy. For instance, we could use anatomical images to parcellate the brain into Brodmann areas and divide each region into subregions with *bottom-up* methods.

Validation

There are very few methods for validating the results of whole brain parcellation. In this thesis, we develop a novel quantitative approach. The use of other image modalities could also help here. For instance, DTI could be used to measure the connectivity of the voxels in each parcel with the variance of the connectivity as a way of estimating the structural intra-parcel homogeneity.

6.2.2 Cross-subject parcel matching

Distance between two parcels

In this thesis, we propose a method that could give optimal parcel matching on the condition that the similarity between parcels in different subjects is well defined. Thus, how the similarity could be defined is naturally a very important and interesting question. In our method, we use the functional distance and coordinates in a standard space. We still need linear spatial normalization as a preprocessing step. It would be interesting and challenging to combine other techniques or image modalities in this process. Since the similarity between parcels is directly constructed with functional and structural images, we could avoid the problem of mis-registration

and other flaws of spatial normalisation.

Parcel matching algorithms

Our parcel-matching algorithm has two main limitations. First, the matching takes place under the assumption that there is one and only one parcel from each subject in a clique. In most cases, this assumption does not present a problem for further analysis. However, occasionally, some parcels may be missing from a subject. In this situation, our algorithm may give a biased result. We could improve our parcel matching method by using, for instance, the k-cardinality assignment model Pentico [2007]. Rather than match all the parcels, this model can find only the first few cliques that have the maximal similarity. Thus, it could possibly eliminate the influence of missing parcels.

Another limitation of our parcel matching method is that this algorithm ignores the relationship of the parcels within each subject. For each parcel, the correlation between this parcel and other parcel in this subject is also important information. In our algorithm, this information was not used to match parcels from different subjects. Finding a way of incorporating this information would be an interesting topic for future study.

6.2.3 Application

Resting-state fMRI data analysis

In this thesis, our method is applied to task-related experiments. However, it could also be applied to resting-state fMRI to study the default mode network, which could provide valuable information on the physiological processes in the human brain. Parcellation based methods could provide more reliable analysis in these studies.

For some novel blind source separation algorithms (e.g. Wang et al. [2010]), the dimension of fMRI data is too high. Our data-driven parcellation method could reduce the dimension of the data and open a door to these types of analysis methods providing an alternative to the classic exploratory multivariate analyses (e.g. ICA).

Machine learning Classifier techniques

As introduced in Chapter 2, machine-learning algorithms are becoming more and more popular in fMRI data analysis. Here, as well, the dimension of the data is a problem. Parcellation could help to alleviate this problem and improve the effectiveness and efficiency of analyses. Furthermore, based on parcellation, many methods that have been successfully applied to the problem of object recognition (e.g. [Zhao et al., 2007]) could also be used in fMRI data analysis.

6.3 Closing Comment

We believe that the development of parcellation reflects the progress of understanding human brain functions. New projects (e.g. the *Human Connectome* Project) are proposed to elucidate the neural pathways that underlie brain function. The success of these projects could bring more understanding of the human brain, so that we could divide the brain into microstructures corresponding to more specific functions. This new knowledge could add extra dimensions (e.g. a connectivity dimension and a function dimension) to the state-of-art human brain atlas and merge the gap between these two types of parcellation.

References

Abdi, H., 2003. Partial least squares regression. Social Sciences, M. Lewis Beck, A. Bryman, and T. Futing, Eds. Thousand Oaks, CA: Sage 2003, 792–795.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E., January 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J. Neurosci. 26 (1), 63–72.

Aguirre, G. K., Zarahn, E., D'Esposito, M., November 1998. The variability of human, bold hemodynamic responses. Neuroimage 8 (4), 360–369.

Andersson, J., Ashburner, J., Friston, K., May 2000. A comprehensive framework for time series realignment. NeuroImage 11 (5, Supplement 1), S488.

Ashburner, J., 1999. High-Dimensional image registration using symmetric priors. NeuroImage 9 (6), 619–628.

Ashburner, J., Friston, K. J., 1999. Nonlinear spatial normalization using basis functions. Human Brain Mapping 7, 254—266.

Backfrieder, W., Baumgartner, R., Samal, M., Moser, E., Bergmann, H., Aug. 1996.

Quantification of intensity variations in functional MR images using rotated principal components. Physics in Medicine and Biology 41 (8), 1425–1438.

Balasubramanian, M., Schwartz, E. L., January 2002. The isomap algorithm and topological stability. Science 295 (5552), 7.

Bandelt, H.-J., Crama, Y., Spieksma, F. C. R., 1994. Approximation algorithms for multi-dimensional assignment problems with decomposable costs. Discrete Applied Mathematics 49 (1-3), 25 – 50.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., Hyde, J. S., Jun. 1992. Time course EPI of human brain function during task activation. Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 25 (2), 390–397, PMID: 1614324.
URL http://www.ncbi.nlm.nih.gov/pubmed/1614324

Bannister, P. R., 2004. Motion correction for functional magnetic resonance images. Ph.D. thesis, University of Oxford, UK.

Bartsch, A. J., Neumarker, K., Franzek, E., Beckmann, H., May 2000. Karl kleist, 1879-1960. Am J Psychiatry 157 (5), 703.

Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., Bullmore, E., December 2006. From the cover: Adaptive reconfiguration of fractal small-world human brain functional networks. PNAS 103 (51), 19518–19523.

Basu, S., Davidson, I., Wagstaff, K., 2008. Constrained Clustering: Advances in Algorithms, Theory, and Applications. Chapman & Hall/CRC.

Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., Moser, E., Dec. 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. part i: reproducibility. Journal of Magnetic Resonance Imaging: JMRI 7 (6), 1094–1101, PMID: 9400854.

Beckmann, C., Smith, S., Feb. 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. Medical Imaging, IEEE Transactions on 23 (2), 137–152.

Beckmann, C., Smith, S., Mar. 2005. Tensorial extensions of independent component analysis for multisubject FMRI analysis. NeuroImage 25 (1), 294–311.

Beckmann, M., Johansen-Berg, H., Rushworth, M. F. S., Jan. 2009. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29 (4), 1175–1190, PMID: 19176826.

Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., Barker, G. J., Sillery, E. L., Sheehan, K., Ciccarelli, O., Thompson, A. J., Brady, J. M., Matthews, P. M., Jul. 2003. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature neuroscience 6, 750–7, PMID: 12808459.

Bell, A. J., Sejnowski, T. J., Nov. 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7 (6), 1129–1159, PMID: 7584893.

Bellec, P., Rosa-Neto, P., Lyttelton, O. C., Benali, H., Evans, A. C., Jul. 2010. Multi-

level bootstrap analysis of stable clusters in resting-state fMRI. NeuroImage 51 (3), 1126–1139.

Bernstein, M. A., King, K. F., Zhou, X. J., September 2004. Handbook of MRI Pulse Sequences, 1st Edition. Academic Press.

Biswal, B., Deyoe, E. A., Hyde, J. S., 1996. Reduction of physiological fluctuations in fMRI using digital filters. Magnetic Resonance in Medicine 35 (1), 107–113.

Biswal, B., Yetkin, F. Z., Haughton, V. M., Hyde, J. S., Oct. 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 34 (4), 537–541, PMID: 8524021.

Biswal, B. B., Kylen, J. V., Hyde, J. S., Aug. 1997. Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps. NMR in Biomedicine 10 (4-5), 165–170, PMID: 9430343.

Borg, I., Groenen, P., 2005. Modern Multidimensional Scaling: Theory and Applications. Springer.

Bridge, H., Clare, S., Jenkinson, M., Jezzard, P., Parker, A. J., Matthews, P. M., 2005. Independent anatomical and functional measures of the V1/V2 boundary in human visual cortex. Journal of Vision 5 (2), 93–102, PMID: 15831070.

Brown, L. G., 1992. A survey of image registration techniques. ACM Comput. Surv. 24 (4), 325–376.

URL http://portal.acm.org/citation.cfm?id=146374

Burkard, R. E., Rudolf, R., Woeginger, G. J., 1996. Three-dimensional axial assignment problems with decomposable cost coefficients. Discrete Applied Mathematics 65 (1-3), 123 – 139, first International Colloquium on Graphs and Optimization.

Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001a. A method for making group inferences from functional MRI data using independent component analysis. Human Brain Mapping 14 (3), 140–151.
URL http://dx.doi.org/10.1002/hbm.1048

Calhoun, V. D., Adali, T., Hansen, L. K., Larsen, J., Pekar, J. J., 2003. ICA of functional MRI data: An overview. In Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation, 281—288.

Calhoun, V. D., Adali, T., Pearlson, G. D., Pekar, J. J., May 2001b. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human Brain Mapping 13 (1), 43–53, PMID: 11284046.

Calhoun, V. D., Golay, X., Pearlson, G., 2000. Improved fmri slice timing correction: Interpolation errors and wrap around effects. In: Proceedings, ISMRM, 9th Annual Meeting. pp. 81–0.

Calhoun, V. D., Liu, J., Adalı, T., 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage 45 (1, Supplement 1), S163 – S172, Mathematics in Brain Imaging.

Cardoso, J., Souloumiac, A., 1993. Blind beamforming for non-Gaussian signals. Radar and Signal Processing, IEE Proceedings F 140 (6), 362–370.

Christensen, M. S., Ramsoy, T. Z., Lund, T. E., Madsen, K. H., Rowe, J. B., Jul. 2006. An fMRI study of the neural correlates of graded visual perception. NeuroImage 31 (4), 1711–1725.

Chuang, K., Chiu, M., Lin, C., Chen, J., 1999. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy c-means. Medical Imaging, IEEE Transactions on 18 (12), 1117–1128.
URL 10.1109/42.819322

Cohen, A. L., Fair, D. A., Dosenbach, N. U. F., Miezin, F. M., Dierker, D., Essen, D. C. V., Schlaggar, B. L., Petersen, S. E., May 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage 41 (1), 45–57, PMID: 18367410.

Coifman, R. R., Lafon, S., Jul. 2006. Diffusion maps. Applied and Computational Harmonic Analysis 21 (1), 5–30.

Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., Peters, T. M., Evans, A. C., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Proc IEEE Nucl Sci Symp Med Imaging 3 (1-3), 1813–1817.

Cordes, D., Haughton, V., Carew, J. D., Arfanakis, K., Maravilla, K., May 2002. Hierarchical clustering to measure connectivity in fMRI resting-state data. Magnetic Resonance Imaging 20 (4), 305–317.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., Quigley, M. A., Meyerand, M. E., Aug. 2001. Frequencies contributing to func-

tional connectivity in the cerebral cortex in "Resting-state" data. AJNR Am J Neuroradiol 22 (7), 1326–1333.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., Quigley, M. A., Meyerand, M. E., Oct. 2000. Mapping functionally related regions of brain with functional connectivity MR imaging. AJNR. American Journal of Neuroradiology 21 (9), 1636–1644, PMID: 11039342.

Coulon, O., Mangin, J. F., Poline, J. B., Zilbovicius, M., Roumenov, D., Samson, Y., Frouin, V., Bloch, I., Jun. 2000. Structural group analysis of functional activation maps. NeuroImage 11 (6), 767–782.

Cox, R. W., Jun. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, an International Journal 29 (3), 162–173, PMID: 8812068.
URL http://www.ncbi.nlm.nih.gov/pubmed/8812068

Crama, Y., Spieksma, F. C. R., August 1992. Approximation algorithms for three-dimensional assignment problems with triangle inequalities. European Journal of Operational Research 60 (3), 273–279.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., Gur, R. C., Langleben, D. D., Nov. 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. NeuroImage 28 (3), 663–668, PMID: 16169252.
URL http://www.ncbi.nlm.nih.gov/pubmed/16169252

Dijkstra, E., 1959. A note on two problems in connection with graphs. Numerische Mathematik 1, 269–271.

Dimitrova, A., Zeljko, D., Schwarze, F., Maschke, M., Gerwig, M., Frings, M., Beck, A., Aurich, V., Forsting, M., Timmann, D., Mar. 2006. Probabilistic 3D MRI atlas of the human cerebellar dentate/interposed nuclei. NeuroImage 30 (1), 12–25.

Draganski, B., Kherif, F., Kloppel, S., Cook, P. A., Alexander, D. C., Parker, G. J. M., Deichmann, R., Ashburner, J., Frackowiak, R. S. J., Jul. 2008. Evidence for segregated and integrative connectivity patterns in the human basal ganglia. The Journal of Neuroscience 28 (28), 7143 –7152.
URL http://www.jneurosci.org/content/28/28/7143.abstract

Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., Tedeschi, G., Goebel, R., Seifritz, E., Salle, F. D., Mar. 2005. Independent component analysis of fMRI group studies by self-organizing clustering. NeuroImage 25 (1), 193–205.

Essen, D. C. V., Drury, H. A., Joshi, S., Miller, M. I., Feb. 1998. Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces. Proceedings of the National Academy of Sciences of the United States of America 95 (3), 788 –795.

Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., Peters, T., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. In: Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record. pp. 1813–1817 vol.3.

Flandin, G., Kherif, F., Pennec, X., Malandain, G., Ayache, N., Poline, J.-B., 2002. Improved detection sensitivity in functional mri data using a brain parcelling technique. In: Dohi, T., Kikinis, R. (Eds.), MICCAI (1). Vol. 2488 of Lecture Notes in Computer Science. Springer, pp. 467–474.

Flandin, G., Penny, W. D., 2007. Bayesian fmri data analysis with sparse spatial basis function priors. NeuroImage 34 (3), 1108 – 1125.

FMRIB, 2007. FMRIB Software Library.
URL http://www.fmrib.ox.ac.uk/fsl/

Fox, M. D., Raichle, M. E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8 (9), 700–711.
URL http://dx.doi.org/10.1038/nrn2201

Frackowiak, R. S., Ashburner, J. T., Penny, W. D., Zeki, S., 2004. Human Brain Function, Second Edition, 2nd Edition. Academic Press.

Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Human Brain Mapping 1, 153–171.

Friston, K., Zarahn, E., Josephs, O., Henson, R., Dale, A., 1999. Stochastic designs in event-related fMRI. NeuroImage 10, 607–619.

Friston, K. J., Frith, C. D., Liddle, P. F., Frackowiak, R. S., Jan. 1993. Functional connectivity: the principal-component analysis of large (PET) data sets. Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism 13 (1), 5–14, PMID: 8417010.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., Turner, R., Mar. 1996. Movement-related effects in fMRI time-series. Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 35 (3), 346–355, PMID: 8699946.

Gao, J., Yee, S., Jan. 2003. Iterative temporal clustering analysis for the detection of multiple response peaks in fMRI. Magnetic Resonance Imaging 21 (1), 51–53.

Garey, M. R., Johnson, D. S., 1990. Computers and Intractability; A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York, NY, USA.

Gee, J. C., Alsop, D. C., Aguirre, G. K., Neurology, 1997. Effect of spatial normalization on analysis of functional data.

Ghasemi, M., Mahloojifar, A., 2010. FMRI data analysis by blind source separation algorithms: A comparison study for nongaussian properties. In: Electrical Engineering (ICEE), 2010 18th Iranian Conference on. pp. 13–17.
URL 10.1109/IRANIANCEE.2010.5507112

Goutte, C., Toft, P., Rostrup, E., Nielsen, F., Hansen, L. K., Mar. 1999. On clustering fMRI time series. NeuroImage 9 (3), 298–310.

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., Hirsch, J., 2008. Detection of time-varying signals in event-related fmri designs. NeuroImage 43 (3), 509 – 520.

Grootoonk, S., Hutton, C., Ashburner, J., Howseman, A. M., Josephs, O., Rees, G., Friston, K. J., Turner, R., 2000. Characterization and correction of interpolation effects in the realignment of fMRI time series. NeuroImage 11 (1), 49–57.

Grosbras, M., Paus, T., Aug. 2006. Brain networks involved in viewing angry hands or faces. Cereb. Cortex 16 (8), 1087–1096.

Guillery, R. W., Apr. 2000. Brodmann's Localisation in the cerebral cortex. Journal of Anatomy 196 (Pt 3), 493–496.

Guo, Y., Gao, J., 2006. Manifolds of bag of pixels: A better representation for image recognition? In: Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on. Vol. 5. pp. 3618–3622.

Haley, K. B., Jun. 1963. The Multi-Index problem. Operations Research 11 (3), 368–379.

Handwerker, D. A., Ollinger, J. M., D'Esposito, M., Apr. 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21 (4), 1639–1651.

Hasson, U., Furman, O., Clark, D., Dudai, Y., Davachi, L., Feb. 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57 (3), 452–462, PMID: 18255037.

Haynes, J., Rees, G., May 2005a. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8 (5), 686–691.

Haynes, J., Rees, G., Jul. 2005b. Predicting the stream of consciousness from activity in human visual cortex. Current Biology: CB 15 (14), 1301–1307, PMID: 16051174.

Haynes, J., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R. E., Feb. 2007.

Reading hidden intentions in the human brain. Current Biology: CB 17 (4), 323–328, PMID: 17291759.

Henson, R., Buchel, C., Josephs, O., Friston, K., 1999. The slice-timing problem in event-related fMRI. In: NeuroImage. Vol. 9. p. 125.

Hirsch, M. W., 1994. Differential Topology. Springer-Verlag, Berlin.

Huettel, S. A., Song, A. W., Mccarthy, G., Apr. 2004. Functional Magnetic Resonance Imaging. Sinauer Associates, published: Hardcover.

Hutchinson, R. A., Niculescu, R. S., Keller, T. A., Rustandi, I., Mitchell, T. M., May 2009. Modeling fMRI data generated by overlapping cognitive processes with unknown onsets using hidden process models. NeuroImage 46 (1), 87–104, PMID: 19457397.

Hyvarinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. Neural Comput. 9 (7), 1483–1492.

Jbabdi, S., Woolrich, M. W., Behrens, T. E. J., Jan. 2009. Multiple-subjects connectivity-based parcellation using hierarchical dirichlet process mixture models. NeuroImage 44 (2), 373–384, PMID: 18845262.

Jebara, T., Oct. 2003. Images as bags of pixels. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. pp. 265–272 vol.1.

Ji, Y., Hervé, P.-Y., Aickelin, U., Pitiot, A., 2009. Parcellation of fmri datasets with ica and pls-a data driven approach. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (Eds.), Medical Image Computing and Computer-Assisted

Intervention – MICCAI 2009. Vol. 5761 of Lecture Notes in Computer Science. Springer, pp. 984–991.

Jolliffe, I. T., October 2002. Principal Component Analysis, 2nd Edition. Springer.

Kamitani, Y., Tong, F., May 2005. Decoding the visual and subjective contents of the human brain. Nature Neuroscience 8 (5), 679–685, PMID: 15852014.

Kamitani, Y., Tong, F., Jun. 2006. Decoding seen and attended motion directions from activity in the human visual cortex. Current Biology: CB 16 (11), 1096–1102, PMID: 16753563.

Kerrouche, N., Herholz, K., Mielke, R., Holthoff, V., Baron, J., Sep. 2006. 18FDG PET in vascular dementia: differentiation from alzheimer's disease using voxel-based multivariate analysis. Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism 26 (9), 1213–21, PMID: 16525414.

Kim, J., Lee, J., Jo, H. J., Kim, S. H., Lee, J. H., Kim, S. T., Seo, S. W., Cox, R. W., Na, D. L., Kim, S. I., Saad, Z. S., Feb. 2010a. Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: functional connectivity-based parcellation method. NeuroImage 49 (3), 2375–2386.

Kim, S. B., Rattakorn, P., Peng, Y. B., Aug. 2010b. An effective clustering procedure of neuronal response profiles in graded thermal stimulation. Expert Systems with Applications 37 (8), 5818–5826.

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M.,

Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., Parsey, R. V., Jul. 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46 (3), 786–802.

Klein, J. C., Behrens, T. E. J., Robson, M. D., Mackay, C. E., Higham, D. J., Johansen-Berg, H., Jan. 2007. Connectivity-based parcellation of human cortex using diffusion MRI: establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. NeuroImage 34 (1), 204–211, PMID: 17023184.
URL http://www.ncbi.nlm.nih.gov/pubmed/17023184

Kondor, R., Jebara, T., 2003. A kernel between sets of vectors. In: In International Conference on Machine Learning (ICML).

Kotsiantis, S. B., 2007. Supervised machine learning: A review of classification techniques. Informatica 31, 249–268.

Kuhn, H. W., 1955. The Hungarian method for the assignment problem. Naval Research Logistic Quarterly 2, 83–97.

Kuroki, Y., Matsui, T., 2009. An approximation algorithm for multidimensional assignment problems minimizing the sum of squared errors. Discrete Applied Mathematics 157 (9), 2124 – 2135, optimal Discrete Structures and Algorithms - ODSA 2006.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sen-

sory stimulation. Proceedings of the National Academy of Sciences of the United States of America 89 (12), 5675–5679.

Lafon, S., Lee, A. B., 2006. Diffusion maps and Coarse-Graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (9), 1393–1403.

Lange, N., Zeger, S. L., 1997. Non-Linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. Journal of the Royal Statistical Society. Series C (Applied Statistics) 46 (1), 1–29.
URL http://www.jstor.org/stable/2986204

Lauterbur, P. C., Mar. 1973. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. Nature 242 (5394), 190–191.

LaValle, S. M., 2006. Planning Algorithms. Cambridge University Press, Cambridge, U.K., also available at http://planning.cs.uiuc.edu/.

Lee, J.-H., Marzelli, M., Jolesz, F. A., Yoo, S.-S., 2009. Automated classification of fmri data employing trial-based imagery tasks. Medical Image Analysis 13 (3), 392 – 404.

Lee, T. W., Girolami, M., Sejnowski, T. J., Feb. 1999. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation 11 (2), 417–441, PMID: 9950738.

Leibovici, D., , Leibovici, D. G., Beckmann, C., 2001. An introduction to multiway methods for Multi-Subject fMRI experiment.

Li, Y.-O., Adali, T., Calhoun, V. D., 2007. A feature-selective independent component analysis method for functional mri. Journal of Biomedical Imaging 2007 (2), 1–6.

Liang, L., Cherkassky, V., Rottenberg, D. A., 2006. Spatial svm for feature selection and fmri activation detection. In: Neural Networks, 2006. IJCNN '06. International Joint Conference on. pp. 1463–1469.
URL http://dx.doi.org/10.1109/IJCNN.2006.246867

Liu, Y., Gao, J., Liu, H., Fox, P. T., Jun. 2000. The temporal response of the brain after eating revealed by functional MRI. Nature 405 (6790), 1058–1062.
URL http://dx.doi.org/10.1038/35016590

Lowe, M. J., Dzemidzic, M., Lurito, J. T., Mathews, V. P., Phillips, M. D., Nov. 2000. Correlations in Low-Frequency BOLD fluctuations reflect Cortico-Cortical connections. NeuroImage 12 (5), 582–587.

Maintz, J., Viergever, M. A., Mar. 1998. A survey of medical image registration. Medical Image Analysis 2 (1), 1–36.

Maulik, U., Mukhopadhyay, A., Aug. 2010. Simulated annealing based automatic fuzzy clustering combined with ANN classification for analyzing microarray data. Computers & Operations Research 37 (8), 1369–1380.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., Lancaster, J., Jun. 1995. A probabilistic atlas of the human brain: theory and rationale for its development. the international consortium for brain mapping (ICBM). NeuroImage 2 (2), 89–101, PMID: 9343592.
URL http://www.ncbi.nlm.nih.gov/pubmed/9343592

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., Grady, C. L., Jun. 1996. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 3 (3), 143–157.

McIntosh, A. R., Chau, W. K., Protzner, A. B., Oct. 2004. Spatiotemporal analysis of event-related fMRI data using partial least squares. NeuroImage 23 (2), 764–775.

McIntosh, A. R. R., Lobaugh, N. J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. NeuroImage 23 Suppl 1 (Supplement 1), S250–S263.

Meyer, F. G., Chinrungrueng, J., Feb. 2005. Spatiotemporal clustering of fMRI time series in the spectral domain. Medical Image Analysis 9 (1), 51–68.

Mezer, A., Yovel, Y., Pasternak, O., Gorfine, T., Assaf, Y., May 2009. Cluster analysis of resting-state fMRI time series. NeuroImage 45 (4), 1117–1125.

Mitchell, T., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., Just, M., Newman, S., Oct. 2004. Learning to decode cognitive states from brain images. Machine Learning 57 (1), 145–175.

Mitra, S., Pedrycz, W., Barman, B., Apr. 2010. Shadowed c-means: Integrating fuzzy and rough clustering. Pattern Recognition 43 (4), 1282–1291.

Mohamed, M., Abou-Chadi, F., Ouda, B., 2007. Denoising functional MRI: a comparative study of denoising techniques (2D). In: World Congress on Medical Physics and Biomedical Engineering 2006. pp. 913–920.

Monir, S. M. G., Siyal, M. Y., 2009. Denoising functional magnetic resonance imaging

time-series using anisotropic spatial averaging. Biomedical Signal Processing and Control 4 (1), 16–25.

Moortele, P. F. V., Cerf, B., Lobel, E., Paradis, A. L., Faurion, A., Bihan, D. L., Aug. 1997. Latencies in fMRI time-series: effect of slice acquisition order and perception. NMR in Biomedicine 10 (4-5), 230–236, PMID: 9430353.

Moser, E., Diemling, M., Baumgartner, R., Dec. 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. part II: quantification. Journal of Magnetic Resonance Imaging: JMRI 7 (6), 1102–1108, PMID: 9400855.

Mourao-Miranda, J., Bokde, A. L., Born, C., Hampel, H., Stetter, M., Dec. 2005. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. NeuroImage 28 (4), 980–995.

Munkres, J., 1957. Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5 (1), 32–38.

Nelson, S. M., Cohen, A. L., Power, J. D., Wig, G. S., Miezin, F. M., Wheeler, M. E., Velanova, K., Donaldson, D. I., Phillips, J. S., Schlaggar, B. L., Petersen, S. E., Jul. 2010. A parcellation scheme for human left lateral parietal cortex. Neuron 67 (1), 156–170, PMID: 20624599.

Neumann, J., von Cramon, D. Y., Forstmann, B. U., Zysset, S., Lohmann, G., Aug. 2006. The parcellation of cortical areas using replicator dynamics in fMRI. NeuroImage 32 (1), 208–219.

Norman, K. A., Polyn, S. M., Detre, G. J., Haxby, J. V., Sep. 2006. Beyond mind-

reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences 10 (9), 424–430.

Ogawa, S., Lee, T., Nayak, A. S., Glynn, P., 1990a. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magnetic Resonance in Medicine 14 (1), 68–78.
URL http://dx.doi.org/10.1002/mrm.1910140108

Ogawa, S., Lee, T. M., Kay, A. R., Tank, D. W., Dec. 1990b. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences of the United States of America 87 (24), 9868–9872.
URL http://www.pnas.org/content/87/24/9868.abstract

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., Ugurbil, K., July 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci U S A 89 (13), 5951–5955.

O'Toole, A. J., Jiang, F., Abdi, H., Penard, N., Dunlop, J. P., Parent, M. A., Nov. 2007. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. Journal of Cognitive Neuroscience 19 (11), 1735–1752, PMID: 17958478.

Park, H., Kubicki, M., Shenton, M. E., Guimond, A., McCarley, R. W., Maier, S. E., Kikinis, R., Jolesz, F. A., Westin, C., Dec. 2003. Spatial normalization of diffusion tensor MRI using multiple channels. NeuroImage 20 (4), 1995–2009.

Pauling, L., Coryell, C. D., 1936. The magnetic properties and structure of

hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. Proceedings of the National Academy of Sciences of the United States of America 22 (4), 210 –216.

Peltier, S., Hsu, M., Welsh, R., Bhavsar, R., Harris, R., Clauw, D., Symonds, L., Yang, L., Williams, D., Jul. 2009. Data-driven parcellation of the insular cortex using resting-state fMRI. NeuroImage 47 (Supplement 1), S169.

Pentico, D. W., 2007. Assignment problems: A golden anniversary survey. European Journal of Operational Research 176 (2), 774–793.

Pereira, F., Mitchell, T., Botvinick, M., Mar. 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45 (1, Supplement 1), S199–S209.

Pierskalla, W. P., Apr. 1968. The multidimensional assignment problem. Operations Research 16 (2), 422–431.

Platt, J. C., 1999. Using analytic qp and sparseness to speed training of support vector machines. In: Proceedings of the 1998 conference on Advances in neural information processing systems II. MIT Press, Cambridge, MA, USA, pp. 557–563.

Pluim, J. P. W., Maintz, J. B. A., Viergever, M. A., Aug. 2003. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging 22 (8), 986–1004, PMID: 12906253.
URL http://www.ncbi.nlm.nih.gov/pubmed/12906253

Pohl, K. M., 2005. Prior information for brain parcellation. Ph.D. thesis, Cambridge, MA, USA.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., Shul-

man, G. L., Jan. 2001. A default mode of brain function. Proceedings of the National Academy of Sciences of the United States of America 98 (2), 676–682, PMID: 11209064 PMCID: 14647.

Rayens, W. S., Andersen, A. H., Sep. 2006. Multivariate analysis of fMRI data by oriented partial least squares. Magnetic Resonance Imaging 24 (7), 953–958.

Riesen, K., Neuhaus, M., Bunke, H., 2007. Bipartite graph matching for computing the edit distance of graphs. In: Graph-Based Representations in Pattern Recognition. pp. 1–12.
URL http://dx.doi.org/10.1007/978-3-540-72903-7_1

Rousseeuw, P., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20 (1), 53–65.
URL http://portal.acm.org/citation.cfm?id=38772

Roweis, S. T., Saul, L. K., Dec. 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290 (5500), 2323–2326.
URL http://www.sciencemag.org/cgi/content/abstract/290/5500/2323

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning internal representations by error propagation, 318–362.

Schmithorst, V. J., Holland, S. K., Mar. 2004. A comparison of three methods for generating group statistical inferences from independent component analysis of fMRI data. Journal of magnetic resonance imaging : JMRI 19 (3), 365–368, PMID: 14994306 PMCID: 2265794.

Shapiro, L. G., Stockman, G. C., Feb. 2001. Computer Vision. Prentice Hall.

Shen, X., Papademetris, X., Constable, R., Apr. 2010. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. NeuroImage 50 (3), 1027–1035.

Simon, O., Kherif, F., Flandin, G., Poline, J., RiviÃÍre, D., Mangin, J., Bihan, D. L., Dehaene, S., Nov. 2004. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. NeuroImage 23 (3), 1192–1202.

Song, X., Iordanescu, G., Wyrwicz, A. M., 2007. One-class machine learning for brain activation detection. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 0, 1–6.

Song, X., Murphy, M., Wyrwicz, A., 2006. Spatiotemporal denoising and clustering of fMRI data. In: Image Processing, 2006 IEEE International Conference on. pp. 2857–2860.

Spalek, K., Thompson-Schill, S. L., Dec. 2008. Task-dependent semantic interference in language production: An fMRI study. Brain and Language 107 (3), 220–228.

Spieksma, F. C. R., Woeginger, G. J., 1996. Geometric three-dimensional assignment problems. European Journal of Operational Research 91 (3), 611 – 618.

SPM8, 2009. Statistical parametric mapping.
URL `http://www.fil.ion.ucl.ac.uk/spm/`

Svensen, M., Kruggel, F., Benali, H., Jul. 2002. ICA of fMRI group study data. NeuroImage 16 (3, Part 1), 551–563.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System : An Approach to Cerebral Imaging. Thieme Medical Publishers, published: Hardcover.

Tenenbaum, J. B., de Silva, V., Langford, J. C., Dec. 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290 (5500), 2319–2323.
URL http://www.sciencemag.org/cgi/content/abstract/290/5500/2319

Tenenbaum, J. B., de Silva, V., Langford, J. C., January 2002. The isomap algorithm and topological stability. Science 295 (5552), 7.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., Aug. 2006. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fmri datasets. Human brain mapping 27, 678–93, pMID: 16281292.

Turner, R., Le Bihan, D., Moonen, C. T., Despres, D., Frank, J., 1991. Echo-planar time course MRI of cat brain oxygenation changes. Magnetic Resonance in Medicine 22, 159–166.

van den Heuvel, M. P., Pol, H. E. H., Aug. 2010. Exploring the brain network: A review on resting-state fMRI functional connectivity. European Neuropsychopharmacology 20 (8), 519–534.

Vapnik, V. N., 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA.
URL http://portal.acm.org/citation.cfm?id=211359

Wang, D., Shi, L., Yeung, D. S., Heng, P.-A., Wong, T.-T., Tsang, E. C. C., 2005.

Support vector clustering for brain activation detection. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention 8, 572–9, pMID: 16685892.

Wang, D., Shi, L., Yeung, D. S., Tsang, E. C., Heng, P. A., Oct. 2007. Ellipsoidal support vector clustering for functional MRI analysis. Pattern Recognition 40 (10), 2685–2695.

Wang, F., Chi, C., Chan, T., Wang, Y., 2010. Nonnegative Least-Correlated component analysis for separation of dependent sources by volume maximization. IEEE Trans. Pattern Anal. Mach. Intell. 32 (5), 875–888.
URL http://portal.acm.org/citation.cfm?id=1749408.1749570

Wang, Z., 2009. A hybrid svm-glm approach for fmri data analysis. NeuroImage 46 (3), 608 – 615.

Wang, Z., Peterson, B. S., Aug. 2008. Partner-matching for the automated identification of reproducible ICA components from fMRI datasets: algorithm and validation. Human Brain Mapping 29 (8), 875–93, PMID: 18058813.

Weisskoff, R. M., 1996. Simple measurement of scanner stability for functional NMR imaging of activation in the brain. Magnetic Resonance in Medicine 36 (4), 643–645.

Wold, S., Sjostrom, M., Eriksson, L., 2001. Pls-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2), 109 – 130.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., Mazziotta, J. C., Feb. 1998. Automated image registration: I. general methods and intrasubject, intramodality validation. Journal of Computer Assisted Tomography 22 (1), 139–152, PMID: 9448779.

Woolrich, M. W., Behrens, T. E. J., Smith, S. M., Apr. 2004. Constrained linear basis sets for HRF modelling using variational bayes. NeuroImage 21 (4), 1748–1761.

Yen, L., Fouss, F., Decaestecker, C., Francq, P., Saerens, M., 2009. Graph nodes clustering with the sigmoid commute-time kernel: A comparative study. Data Knowl. Eng. 68 (3), 338–361.

Yu, J., Wang, Y., Shen, Y., 2008. Noise reduction and edge detection via kernel anisotropic diffusion. Pattern Recogn. Lett. 29 (10), 1496–1503.

Zhang, Z., Zha, H., 2002. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. SIAM Journal of Scientific Computing 26, 313—338.

Zhao, Z., Vashist, A., Elgammal, A., Muchnik, I., Kulikowski, C., 2007. Combinatorial and statistical methods for part selection for object recognition. Int. J. Comput. Math. 84 (9), 1285–1297.

Zhilkin, P., Alexander, M. E., Jan. 2004. Affine registration: a comparison of several programs. Magnetic resonance imaging 22 (1), 55–66.

Zitova, B., Oct. 2003. Image registration methods: a survey. Image and Vision Computing 21 (11), 977–1000.
URL http://dx.doi.org/10.1016/S0262-8856(03)00137-9

Zuendorf, G., Kerrouche, N., Herholz, K., Baron, J., 2003. Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional imaging data sets as applied to FDG-PET and normal aging. Human Brain Mapping 18 (1), 13–21.