COMPRESSIVE BRAIN NEURAL ACTIVITY DETECTION USING FUNCTIONAL MAGNETIC RESONANCE IMAGES

by

Chuan Li

Department of Electrical and Computer Engineering
The University of Alabama

Date: ____________________________

Committee:

___________________________
Dr. Qi Hao, Supervisor

___________________________
Dr. Fei Hu

___________________________
Dr. Jeff Jackson

___________________________
Dr. Shuhui Li

___________________________
Dr. Jingyuan Zhang

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering in the Graduate School of The University of Alabama

2010
Abstract

Functional Magnetic Resonance Imaging (fMRI) measures the hemodynamic response related to neural activity in the brain or spinal cord of human subjects. The goal of our research is to develop a framework of data processing and logic inference which can identify the neural activity patterns using fMRI measurements with high accuracy at low computational cost. Throughout the development of our fMRI brain neural activity detection/identification framework, the following issues have been focused and investigated:

1) **Selection and modeling of a priori.** The prior knowledge in the context of fMRI data analysis includes the sparsity of neural activity, the spatial, temporal, inter-subject correlations and structural invariance of fMRI measurements. Proper mathematical models of prior knowledge should be developed and utilized to reduce the data volume, restore image distortions, and remove false alarms, which will lead to higher detection accuracy and lower processing cost. We have explored and developed a series of *a priori* models ranging from wavelet/contourlet hidden Markov tree, to hidden Markov chain and to multi-layer/multi-scale sparse neural activity model. Based on those models, the acquired fMRI data can reveal more statistical patterns of brain neural activity with respect to specific subject behavior.

2) **Algorithmic trade-off between performance and cost.** The high resolution and large volume of fMRI data demand a balance between accuracy of data
analysis and computational cost of data processing. This requirement necessitates efficient representation of information of interest and fast convergent optimization/learning algorithms. In this respect, we have explored several high-efficiency information representation methods including linear predictive coding, contourlet decomposition, principal/independent component analysis, and random projection to reduce data dimensionality without losing information of interest. In addition, various approaches including expectation-maximization (EM), maximum a posteriori (MAP), $L_1$ norm regularization, and variational Bayesian (VB) approximation have been investigated to increase the convergence speed as well as robustness of optimization/learning algorithms.

3) **Combination of information processing techniques in the logic and data layers.** Most conventional neural activity detection techniques process information in data layer such as filtering, statistical analysis and optimization. We further generate logical events, probabilistic structure and sparsity regularization out of original fMRI measurements through data modeling and learning. The advantages of information processing in the logic layer include fast speed and high robustness against measurement noise, motion artifacts and false alarms. The extra cost is the training of logic models. By using various logic models and inference methods, conventional fMRI data processing and analysis techniques such as deformable image restoration and independent component analysis (ICA), temporal cluster analysis (TCA) can be enhanced in terms of neural activity detection performance and robustness.

The main accomplishments of this thesis include the following three components. 1) Image restoration using wavelet/contourlet hidden Markov tree (HMT)
models, which can restore statistical properties of fMRI data afflicted with motion artifacts. 2) Independent component analysis (ICA) with $L_1$ norm regularization, which can efficiently decompose fMRI data into four components: low frequency neural activity, high frequency neural activity, motion artifacts and other noise. 3) Graphical model inference for compressive neural activity detection, which can reject inconsistent signal events and hence reduce false alarms in neural activity detection.

Through developing these three components, this dissertation provides a complete robust neural activity identification/detection framework, from image restoration to signal decomposition and to signal inference, with high accuracy and low computational cost.
Acknowledgements

I want to express my gratitude to all those people who have helped me to accomplish this dissertation.

My deepest gratitude is to my advisor, Dr. Qi Hao. I have been amazingly fortunate to have an advisor who guided me with his passion and infectious enthusiasm. His unlimited zeal have been major driving forces through my graduate career at the University of Alabama. He taught me how to question thoughts and express ideas.

My co-advisor, Dr. Fei Hu, has been always there to listen and give advice. I am deeply grateful to him for the long discussions that helped me sort out the details of my work. I am also thankful to him for encouraging the use of correct grammar and consistent notation in my writings and for carefully reading and commenting on countless revisions of this manuscript.

I am grateful to Dr. Jeff Jackson. His patience and support helped me overcome many crisis situations and finish this dissertation.

I would like to sincerely thank Dr. Shuihui Li, and Dr. Alex Jingyuan Zhang, for providing constructive comments and useful help on this dissertation.

It is my privilege to have the chance to work with the cognitive sensor and network security lab. The working experience shared with Dr. Qi Wei, and Qingquan Sun has been pleasant and unforgettable.

This proposal is dedicated to Chenwei Li, my dear wife who supports me with
her endless optimism during the process of my PhD study.


Contents

Abstract ii

Acknowledgements v

List of Figures xii

List of Tables xvii

1 Introduction 1
  1.1 Background ................................................. 1
    1.1.1 Neuroimaging ........................................... 1
    1.1.2 Neuroimaging Techniques ............................... 2
    1.1.3 Significance of fMRI Study ............................ 5
  1.2 Challenges and Motivation ................................. 7
    1.2.1 Inter-Subject Variabilities ............................ 7
    1.2.2 Motion Artifacts ..................................... 9
    1.2.3 Low SNR and High Data Volume ....................... 9
  1.3 Goals and Achievements .................................. 11
    1.3.1 Main Goal ............................................. 11
    1.3.2 Main Achievements .................................... 12
      Motion Artifacts Correction through HMT Modeling .... 12
      Sparse Neural Activity Detection Using NHR Model .... 13

vii
Compressive Neural Activity Detection Using Logical Layer Inference ........................................... 13

1.4 Structure of The Dissertation .......................................................... 14

2 Experiment Setup and State-of-the-Art Approaches ........................................... 15

2.1 Experiment Setup .................................................................................. 15

2.1.1 Simulation Data .................................................................................. 15

2.1.2 Experimental Data ............................................................................. 16

2.2 State-of-the-Art Approaches ................................................................... 17

2.2.1 Image Preprocessing: Deformable Image Registration .................. 20

Similarity Measures .................................................................................. 20

Deformation Models ................................................................................. 23

2.2.2 Image Preprocessing: Contourlet Transform .................................. 24

Pyramid Frame ......................................................................................... 26

Directional Filter Bank ............................................................................... 27

Contourlet Transform Framework .................................................................. 28

2.2.3 Data Analysis and Modeling: ICA .................................................... 28

2.2.4 Statistical Approaches: Temporal Clustering Analysis .................. 32

2.2.5 State-of-the-Art Technique Comparison .......................................... 32

3 Hidden Markov Tree Based Image Restoration ........................................... 34

3.1 Introduction ......................................................................................... 34

3.2 Related Work ....................................................................................... 36

3.3 Wavelet and Contourlet HMT Model .................................................. 42

3.4 Preprocessing and Motion Detection ................................................... 45

3.4.1 Processing Framework ....................................................................... 45
3.4.2 Rigid Image Registration
3.4.3 WHMT Model Based Denoising
3.4.4 Motion Detection

3.5 Motion Correction
3.5.1 Non-Rigid Registration
3.5.2 HMT Restoration
3.5.3 Robust Training via Tying
3.5.4 Loop Between Motion Detection and Correction

3.6 Experiment Results and Discussion
3.6.1 Simulation Results
3.6.2 Real fMRI data Results

3.7 Discussion
3.8 Conclusion

4 Independent Component Analysis with $L_1$ Regularization
4.1 Introduction
4.2 Related Work
4.3 Modeling and Problem Statement
4.3.1 Multi-Scale Neural Activity Model
List of Figures

1.1 Inter-subject variabilities. ......................................... 8
1.2 Motion artifacts and proposed solutions. .......................... 10
1.3 Sample fMRI’s temporal signal compared with background noise. .  11
1.4 Contributions of this dissertation. ................................. 12
2.1 Sample simulation data: (a) Baseline image; (b) Noise affected image; (c) Motion affected image. ................................. 16
2.2 Sample fMRI’s data for testing .................................... 17
2.3 fMRI neural activity detection methods ................................ 18
2.4 Deterministic approaches ........................................... 19
2.5 Statistical approaches .............................................. 20
2.6 Compare with rigid and deformable registration ................... 25
2.7 Laplacian pyramid. One level of decomposition. The outputs are a coarse approximation a and a difference b between the original signal and the prediction. ................................. 26
2.8 Contourlet transform reconstruction scheme for the Laplacian pyramid. 27
2.9 Iteration of two-channel filter banks ................................ 28
2.10 Multi scale decomposition and DFB .................................. 29
2.11 Independent component analysis .................................... 30
2.12 TCA time window detection ...................................... 33
3.1 Gaussian mixture model for wavelet/contourlet coefficient histogram.  
(a) motion affected image (b) Region of Interest (ROI) of the motion affected image Image (c) Baseline image (d) ROI of the motion affected image Image. .......................................................... 38

3.2 Gaussian mixture model for wavelet/contourlet coefficient histogram.  
(e) Wavelet/contourlet coefficients histogram of the motion affected image (f) Gaussian mixture approximation of the histogram (g) Wavelet/contourlet coefficients histogram of the baseline image (h) Gaussian mixture approximation of the histogram ........................................... 39

3.3 (a) Interscale dependencies among wavelet coefficients and (b) Interscale dependencies among contourlet coefficients .......................... 41

3.4 Hidden Markov tree model: value denotes to wavelet/contourlet coefficients. ................................................................. 43

3.5 Processing framework of the proposed image restoration method. .... 46

3.6 Tying in the HMT model: tying across the wavelet/contourlet trees, and tying within a wavelet/contourlet tree. .............................. 54

3.7 Flowchart of our approach. ......................................................... 55

3.8 Image restoration and denoising for simulation data: (a) Baseline image; (b) Noise affected image; (c) Motion affected image (d) HMT denoising result. (e) Wavelet HMT restoration result. (f) Contourlet HMT restoration result. ................................................................. 57

3.9 Wavelet/contourlet coefficients histogram for (a) Noise affected image (b) Wavelet histogram for motion affected image (c) Contourlet histogram for motion affected image.(d) De-noised wavelet coefficients histogram (e) WHMT restoration wavelet histogram (f) CHMT restoration contourlet histogram ........................................... 58

3.10 Activation map for the simulation data (a) Baseline motion free activation map (b) Non-rigid registration of motion affected data’s activation map (c) WHMT restored activation map (d) CHMT restored activation map ................................................................. 61

3.11 Image restoration with water drinking activity: the right column are the image section contains the neural activities, and the right column are the wavelet/contourlet coefficients histograms. ............................... 63
3.12 Image restoration result comparison: (a) Rigid registration result (b) Non-rigid registration result (c) Wavelet HMT restoration result (d) Contourlet HMT restoration result (e-h) Average residue between restored image and baseline image; 

3.13 Image restoration result comparison: (a) Mutual Information of the four methods; (b) Signal to noise ratio of the four methods; 

3.14 (a) The average residue for 6 data sets between restored images and baseline image. (b) The difference of the gaussian mixture model for 6 data sets between restored images and baseline image 

3.15 fNmax value after image restored by (a) Mean square registration (b) 2d-deformable registration (c) 2d-deformable registration on wavelet domain (d) Wavelet HMT based restoration 

3.16 Activation map of (a) Rigid registration. (b) Non-rigid registration. (c) Wavelet base HMT restoration (d) Contourlet based HMT restoration.

4.1 The proposed model for the random projection and ICA with neural activity detection model. 

4.2 Proposed framework for the ICA with multi-scale, sparsity-exploiting constrained $L1$ regularization approach. 

4.3 Three-layer NHR model. 

4.4 Neural activity detection results at three layers: (a) Layer 1, (b) Layer 2, and (c) Layer 3. 

4.5 Simulation data: (a) Baseline image; (b) baseline image with noises; (c) baseline image with motion artifacts. 

4.6 Different Neural Impulse Response Model. 

4.7 ICA decomposing the compressed signal into neural activities, motion artifacts and noises. 

4.8 Sparse neural activities detection using ICA with constrained $L1$ regularization. 

4.9 Residue between ICA detected signal and the original signal.
4.10 Residue between ICA detected signal and the original signal.

4.11 Multi-layer neural activity detection through constrained $L_1$ regularization.

4.12 Multi-layer neural activity detection through constrained $L_1$ regularization.

4.13 Multi-layer neural activity detection through constrained $L_1$ regularization.


4.15 Water drinking neural activity patterns.

4.16 Activation map detection for the four approaches.

4.17 Multi-layer neural activity detection through constrained $L_1$ regularization.

4.18 Multi-layer neural activity detection through constrained $L_1$ regularization.

5.1 System diagram of the proposed hybrid approach.

5.2 TCA signals. (a) TCA signals generated by one subject drinking water. (b) zoomed-in signals to illustrate the generation of time windows.

5.3 The procedure of event prototype generation.

5.4 Compressive linear predicative coding of signal events.

5.5 Clustering of event prototypes using the VBGMM approach.

5.6 Graphical model of neural activities in fMRI data.

5.7 detection of neural activities. (a) fMRI temporal signals. (b) time windows generated by temporal cluster analysis. (c) labeled event sequence (Red, Blue, Green, and Pink: cluster 1, 2, 3, and 4). (d) detected neural activity status (Dark: activation; Gray: normality; White: deactivation).
5.8 neural activity detection performance comparison between TCA (upper) and the proposed Hybrid approach (lower) in the area of hypothalamus.
List of Tables

2.1 Compare the State-of-the Art Techniques ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ......
Chapter 1

Introduction

Detecting neural activities associated with certain brain function has been a popular research area in medical image processing. With the development of a brand new imaging method called functional magnetic resonance imaging (fMRI), it is possible to measure the changes in blood flow and blood oxygenation level which directly associates with the specific functions. In this chapter, the neuroimaging technologies, the properties of fMRI will be briefly reviewed. Then the challenges and motivations for accurately and efficiently detect the brain neural activities will be discussed. Goals and achievement will be highlighted accordingly. At last, the structure of this dissertation will be presented.

1.1 Background

1.1.1 Neuroimaging

Neuroimaging includes the use of various techniques to either directly or indirectly image the structure and the function of the brain. It is a relatively new discipline within medicine and neuroscience/psychology [1] [2].

Neuroimaging falls into two broad categories [3]:

**Structural Imaging**: which deals with the structure of the brain and the
diagnosis of gross (large scale) intracranial disease (such as tumor), and injury [4].

**Functional Imaging:** which is used to diagnose metabolic diseases and lesions on a finer scale (such as Alzheimer’s disease) and also for neurological and cognitive psychology research and building brain-computer interfaces. Functional imaging enables, for example, the processing of information by centers in the brain to be visualized directly. Such processing causes the involved area of the brain to increase metabolism and "light up" on the scan [5].

### 1.1.2 Neoroimaging Techniques

Brain imaging techniques allow doctors and researchers to view activity or problems within the human brain, without invasive neurosurgery. There are a number of accepted, safe imaging techniques in use today in research facilities and hospitals throughout the world.

**Computed Axial Tomography:** Computed Tomography (CT) or Computed Axial Tomography (CAT) scanning uses a series of x-rays of the head taken from many different directions [6]. Typically used for quickly viewing brain injuries, CT scanning uses a computer program that performs a numerical integral calculation (the inverse Radon transform) on the measured x-ray series to estimate how much of an x-ray beam is absorbed in a small volume of the brain [7]. Typically the information is presented as cross sections of the brain. In approximation, the denser a material is, the whiter a volume of it will appear on the scan (just as in the more familiar ”flat” X-rays). CT scans are primarily used for evaluating swelling from tissue damage in the brain and in assessment of ventricle size. Modern CT scanning can provide reasonably good images in a matter of minutes.

**Diffuse Optical Imaging:** Diffuse Optical Imaging (DOI) or Diffuse Optical
Tomography (DOT) is a medical imaging modality which uses near infrared light to generate images of the body. The technique measures the optical absorption of haemoglobin, and relies on the absorption spectrum of haemoglobin varying with its oxygenation status [8].

Event-Related Optical Signal: Event-Related Optical Signal (EROS) is a brain-scanning technique which uses infrared light through optical fibers to measure changes in optical properties of active areas of the cerebral cortex [9]. Whereas techniques such as Diffuse Optical Imaging (DOT) and Near Infrared Spectroscopy (NIRS) measure optical absorption of haemoglobin, and thus are based on blood flow, EROS takes advantage of the scattering properties of the neurons themselves, and thus provides a much more direct measure of cellular activity. EROS can pinpoint activity in the brain within millimeters (spatially) and within milliseconds (temporally) [10]. Its biggest downside is the inability to detect activity more than a few centimeters deep. EROS is a new, relatively inexpensive technique that is non-invasive to the test subject.

Magnetic Resonance Imaging: Sagittal MRI slice at the midline [11]. Magnetic Resonance Imaging (MRI) uses magnetic fields and radio waves to produce high quality two or three dimensional images of brain structures without use of ionizing radiation (X-rays) or radioactive tracers [12]. During an MRI, a large cylindrical magnet creates a magnetic field around the head of the patient through which radio waves are sent. When the magnetic field is imposed, each point in space has a unique radio frequency at which the signal is received and transmitted. Sensors read the frequencies and a computer uses the information to construct an image. The detection mechanisms are so precise that changes in structures over time can be detected [12]. Using MRI, scientists can create images of both surface and subsurface structures with a high degree of anatomical detail. MRI scans can
produce cross sectional images in any direction from top to bottom, side to side, or front to back [13].

**Electroencephalography:** Electroencephalography (EEG) is an imaging technique used to measure the electric fields in the brain via electrodes placed on the scalp of a human. EEG offers a very direct measurement of neural electrical activity with very high temporal resolution but relatively low spatial resolution [14].

**Functional Magnetic Resonance Imaging:** Functional magnetic resonance imaging (fMRI) relies on the paramagnetic properties of oxygenated and deoxygenated hemoglobin to see images of changing blood flow in the brain associated with neural activity [15]. This allows images to be generated that reflect which brain structures are activated (and how) during performance of different tasks. Most fMRI scanners allow subjects to be presented with different visual images, sounds and touch stimuli, and to make different actions such as pressing a button or moving a joystick [16]. Consequently, fMRI can be used to reveal brain structures and processes associated with perception, thought and action. The resolution of fMRI is about 2-3 millimeters at present, limited by the spatial spread of the hemodynamic response to neural activity [17]. It has largely superseded PET for the study of brain activation patterns. PET, however, retains the significant advantage of being able to identify specific brain receptors (or transporters) associated with particular neurotransmitters through its ability to image radio labeled receptor "ligands" (receptor ligands are any chemicals that stick to receptors) [18].

The problem with original MRI technology was that while it provides a detailed assessment of the physical appearance, water content, and many kinds of subtle derangements of structure of the brain (such as inflammation or bleeding), it fails to provide information about the metabolism of the brain (i.e. how actively it is functioning) at the time of imaging [19].
1.1.3 Significance of fMRI Study

fMRI is increasingly used for the medical diagnosis of disease. Because fMRI is exquisitely sensitive to blood flow, it is extremely sensitive to early changes in the brain resulting from ischemia (abnormally low blood flow), such as the changes which follow stroke [20]. Early diagnosis of certain types of stroke is increasingly important in neurology, since substances which dissolve blood clots may be used in the first few hours after certain types of stroke occur, but are dangerous to use afterwards. Brain changes seen on fMRI may help to make the decision to treat with these agents. With between 72% and 90% accuracy where chance would achieve 0.8%, fMRI techniques can decide which of a set of known images the subject is viewing [21].

MRI (Magnetic Resonance Imaging) for brain imaging, in general, is used for view the structure of the brain [22]. It is especially useful for detecting small anatomical changes as a result of disease processes or trauma that cannot be resolved in a CT scanner [23]. It has great research utility for correlating structural changes/differences with behavior. fMRI (functional MRI) is a recent development in MR technology that allows you to obtain a 'functional' image of the brain by measuring blood flow (blood oxygenation actually in certain cases). This is a hot area because the hope is that you would eventually be able to take advantage of the high spatial resolution of MRI and the far superior temporal resolution of fMRI - relative to PET(PET takes about 1 min / scan - fMRI 500 millisecond to 10 seconds). So the major difference is the MRI images structure and fMRI images function [24].

The major advantages of fMRI are listed below [25]:

1) **Noninvasive Sensing Modality:** It can non-invasively record brain signals
without risks of radiation inherent in other scanning methods, such as CT or PET scans.

2) **Whole Brain Measurement:** It can record signal from all regions of the brain, unlike EEG/MEG which are biased towards the cortical surface.

3) **Illustrative ”Activation” Images:** fMRI is widely used and standard data-analysis approaches have been developed which allow researchers to compare results across labs and produce compelling Images of Brain ”Activation”.

The Disadvantages of fMRI are listed below:

1) **Expensive to Collect Data:** single machine costs hundreds of thousands of dollars.

2) **Sensitive to Motion Artifacts:** The BOLD signal is only an indirect measure of neural activity, and is therefore susceptible to influence by non-neural changes in the body [26].

3) **Non-Localized to Region of Interest:** The images produced must be interpreted carefully, since correlation does not imply causation, and brain processes are complex and often non-localized.

4) **Poor Temporal Resolution:** fMRI has poor temporal resolution [27]. The BOLD response peaks approximately 5 seconds after neuronal firing begins in an area. This means that it is hard to distinguish BOLD responses to different events which occur within a short time window. Careful experimental design can reduce this problem. Also, some research groups are attempting to combine fMRI signals that have relatively high spatial resolution with signals recorded with other techniques, electroencephalography (EEG) which have higher temporal resolution but worse spatial resolution.
For these reasons, functional MR imaging provides insights into neural processing that are complementary to insights of other studies in neurophysiology. And methods for detect the fMRI neural activity has been widely studied and developed.

1.2 Challenges and Motivation

Various methods leads to multiple ways to categorize them. Some methods utilize the repeatable and stable patterns in neural activities to train neural activity model for detection. Others just use the statistical properties as a reference without involve using prior knowledge. Some methods incorporate the traditional image processing techniques into fMRI while others focus on improve techniques specifically based on the unique feature of fMRI.

Despite many achievements we have, there are still many challenges current approaches commonly face as all algorithms has to deal the trade-off between performance and cost and the selection or modeling of a priori.

Out of all challenges, there are three main issues matter the most in fMRI neural activity detection. They are: (1) inter-subject variabilities. (2) motion artifacts and (3) low signal-to-noise ratio and high data volume.

1.2.1 Inter-Subject Variabilities

When different people or one people perform same cognitive tasks at different time. The activations region may occur at slight different in spatial and temporal domain. This signal distortion cause inter-subject variability maybe greatly reduced using image normalization for example using statistical properties restoration or using the inter-subject correlation to resume the image based on statistical modeling. Yet it is still hard to tackle the issue by totally eliminate it. Fig.1.1 shows the inter-subject
Variability for four different subjects.

Current state-of-the-art approaches generally have two ways to eliminate the inter-subject variabilities:

1) **Normalization** changes the range of pixel intensity values by dynamic range expansion to achieve the statistical properties restoration.

2) **Statistical Modeling** utilize inter-subject correlation/variability to resume the image correspondence.
1.2.2 Motion Artifacts

Brain fMRI motion artifacts could be caused by muscle movement like swallowing and sometime the patient’s head moving. Resulting in Image Blurring, reduce the Signal-to-Noise Ratio and generate the false alarm. A common approach to solve this issue is through image registration. Yet Image registration, no matter for rigid registration which translating and rotating and scaling, or for more advanced non-rigid registration frequently deform one image until it matches together more closely, which involves the use of additional constraints. It is still hard to totally eliminate them as if the artifact cause certain region of the fMRI’s intensity spike, as the fMRI’s histogram will change. Both rigid and non-rigid image registration techniques just involves point matching and transformation without future manipulate the intensity of the images by modeling the motion artifacts. De-noising techniques like Gaussian smoothing may eliminate the spike yet smoothing suppresses noise while it also changes the intensity variation of the underlying image. Solution for eliminate the motion artifacts generally can be categorized into three types:

1) Using rigid image registration to eliminate linear distortion;
2) Using non-rigid image image registration to eliminate the non-linear (elastic) distortion
3) Using Image Restoration to deal with more complex distortion.

Fig.1.2 shows different types of motion artifacts.

1.2.3 Low SNR and High Data Volume

Neural activity in brain fMRI data could as low as only 2% higher rise from the background noise. There low SnR if polluted with artifacts would be hard for many
traditional signal processing techniques to be applied in. Also now a day, one single brain fMRI image often features resolution higher than 256*256. A set of simple cognitive task (e.x: water drinking) often include more than 100 pics. Locate the region of interest, detect the neural activity in real time will be a real challenge when dealing with these big amount of data. To solve this Issue, we must utilizing more prior knowledge to pre-processing the data in order to increase the SNR. Also, compressive signal techniques need to be applied to reduce the computational cost. Fig.1.3 shows the fMRI’s signal in temporal domain compared with back ground noise. Current solution for increasing the SNR and reduce the data volume focus on the following two disciplines:

1) Using Prior Knowledge to Establish model to Increase the Signal-to-Noise Ratio.
2) Using Compressive Signal Processing Techniques to reduce the data volume.

1.3 Goals and Achievements

1.3.1 Main Goal

Based on the challenges we faced and the common solutions to those challenges. The main goal for this dissertation is to design an efficient and effective approach which could solve the previous listed challenges while outperform the common solutions. By holding this goal, in this dissertation. Several achievements have been made.
1.3.2 Main Achievements

There are three main achievements have been made in this dissertation. Fig. 1.4 shows these contributions in the general fMRI neural activity detection framework: At preprocessing stage, a hidden markov tree model based method is developed for image restoration. At data analysis stage, a neural hemodynamic response model was established for independent component analysis. At decision making stage, a logical layer event prototype is developed for graphical model inference.

Motion Artifacts Correction through HMT Modeling

Motion artifacts impact the quality of fMRI data in three aspects: (1) pixel displacement within the region of interest (2) image blurring within motion affected regions, and (3) gray level intensity spikes. Spatial filters have been applied to de-blurring (sharpening) and smoothing to reduce the gray-level intensity distortion. Unfortunately, the goals of de-blurring and spike reduction usually contradict to each other. By explore the structural information among image data sets in hidden Markov tree model. A gaussian mixture model is established to approximate the
coefficient with Laplacian distribution such like wavelet/contourlet transformation. When image displacement/blurring/intensity spikes occurs, the interscale/intrascle dependencies of the coefficients are corrupted and result in the skewness of the spectrum distribution. By restore the statistical dependencies, motion artifacts could be effectively eliminated.

**Sparse Neural Activity Detection Using NHR Model**

Independent component analysis could detect the neural activities without model yet lack of detecting accuracy if more than one gaussian components within the source fMRI. By establish a neural hemodynamic response (NHR) model which accounts for temporal, spatial coherence of fMRI data. The sparse nature of neural activity could be explored, hence reduce the computational cost. A multi-layer, multi-scale NHM that can describe both high-frequency and low-frequency components of fMRI measurements which convert the detecting of neural activities into finding solutions in an ill-conditioned matrix. Through $L_1$ regularization, the ICA detecting accuracy could be effectively increased by incorporating the NHR model.

**Compressive Neural Activity Detection Using Logical Layer Inference**

By using data driven approach provided *prior* information, with signal compression techniques, the high data volume, low SNR fMRI images can be converted to discrete event prototypes and further classified into logical layer event sequences. Based on a graphical-model representation, the neural activities can be detected and classified at reduced computation cost with robustness against the inter-subject variability.
1.4 Structure of The Dissertation

The rest of the dissertation will be organized in the following way: Chapter 2 briefly overviews the experimental setup and review the state-of-the-art approaches related to this dissertation. Chapter 3 introduces the image restoration using hidden Markov tree model. Chapter 4 presents sparse neural activity detection using independent component analysis with $L_1$ regularization. Chapter 5 introduces the graphical model based inference with linear predictive coding. Finally, chapter 6 highlights the conclusion, summarizes the contributions of this dissertation and the future work.
Chapter 2

Experiment Setup and State-of-the-Art Approaches

In this chapter, the experimental and environmental setup will be presented and the state-of-the-art approaches related to brain neural activities detection will be introduced.

2.1 Experiment Setup

All experiments have been done at three phases: first the proposed approach is tested using simulation data to verify its correctness and effectiveness. Then real fMRI data be applied on the same task. Finally the results will be studied and compared with other popular state-of-the-art approaches to demonstrate the advantages of the proposed approach.

2.1.1 Simulation Data

Simulation data will be designed at relatively low resolution to reduce the computational cost. The base simulation data have a roughly human brain shape. The hypothalamus area of human brain is painted with separate gray-level with distinctive edges around as in real data, most neural activities related to performed tasks
occur in this area. Artificial neural activities will be added in using impulse response model to approximate the neural dynamics within fMRI. Motion artifacts including pixel level displacement, gray level intensity spikes, and blurring are added. White noise has also applied into simulation data. Details of the different simulation data related to specific proposed methods will be presented in chapter 3 and 4. Fig.2.1 shows one sample of the simulation data.

2.1.2 Experimental Data

The fMRI data were collected from two different tasks: water drinking and glucose intaking. Patients were laid down on the fMRI machine before performing the task. Water drinking and glucose intaking will start after several minutes. The patients drink water through a straw and glucose will be injected into patient’s body in order to minimize the involuntary mascele movement caused motion artifacts. In each task, 6 sets of data were collected. Each set contains 183 frames of pictures at the resolution of 256*256 each. Both water drinking and glucose were sampled at the rate of 15 sec/frame. Water drinking times for 6 patients are vary from each other.(54”, 45”, 1’40”, 1’, 57”, 1’36”, respectively). Glucose intaking times are
Figure 2.2: Sample fMRI’s data for testing.

1’28”, 1’, 1’, 50”, 1’2”, 1’10”, respectively. Fig.2.2 shows the sample of the fMRI test data.

2.2 State-of-the-Art Approaches

After extensive study for more than a decade, scientists have come up with a variety of different methods for neural activities detecting using fMRI. With all of the methods, a popular way to categorize is to put them into two schemes: model based methods and data driven methods [28] [29] [30]. Model-based methods demand repeatable and stable patterns in brain activities and they can provide insights into how a particular cognitive process is implemented in a specific brain area instead of merely identifying where such a process is located. Data driven method have more flexibility and low computational cost as no model were trained during the process yet leads to false detection. Fig.2.3 shows popular way to categorize fMRI neural activity detection methods.

In this dissertation, these methods was studied based on their functionality and they were classified into two categories: deterministic approach and statistical approach.
Figure 2.3: fMRI neural activity detection methods.
1) **Deterministic Approach**: Deterministic approaches utilize the spatial, temporal and inter-subject correlations to detect the neural activities. Furthermore, these methods can be divided into three sub-categories: (a) Impulse Response Modeling: Such like Maximum Correlation Method. (b) Logic Signal Estimation: Such Like E-M Optimization. (c) Spatial Smoothing: Gaussian Filtering. Fig.2.4 shows the flowchart for deterministic approaches.

2) **Statistical Approaches**: Statistical approaches utilizes the statistical method to locates/detect the neural activities. Further sub-categories could be divided as below: (a) Statistical Analysis: ICA,PCA,TCA (b) Statistical Modeling: GLM, graphical modeling, random markov field. (c) Statistical Inference: variation bayesian Inference, belief propagation, monte carlo markov chain. Fig.2.5 shows the flowchart for Statistical approaches.

In the following sections, state-of-the-art approaches related to the dissertation will be introduced and compared.
2.2.1 Image Preprocessing: Deformable Image Registration

Deformable image registration has been studied since the early 80s and for many years. Brain surgery and neuroscience has been the driving applications for developing an abundant number of techniques. Despite the significant progress that has been made, deformable registration in fMRI is still not clinically accepted and remains a challenging problem [25].

Similarity Measures

Registration methods that use voxel similarity measures determine the registration transformation by optimizing the similarity function directly from the voxel values rather than from points or surfaces derived from the image.

One of the simplest voxel similarity measures between a transformed image $I_2$ and a fixed image $I_1$ is the sum of squared grey value differences:

$$SSD = \sum_{i \in M} (I_1(i) - I_2(i))^2$$
Where $M$ is the region of overlap of the images $I_2$ and $I_1$. Sum of the Squared Difference (SSD) is very sensitive to voxels with large intensity differences (outliers) which makes SSD only applicable in single-modality registration contexts, or more precisely, in cases where the images to be registered only differ by noise when registered. The least-squares form of SSD makes the measure computationally very attractive since fast optimization schemes such as Gauss-Newton or Levenberg-Marquardt can be applied.

If a linear relationship between the grey values of the images can be assumed, correlation-based similarity measures such as the cross-correlation:

$$CC = \frac{\sum (I_1(i) - \bar{I}_1)(I_2(i) - \bar{I}_2)}{\sqrt{\sum (I_1(i) - \bar{I}_1)^2(I_2(i) - \bar{I}_2)^2}}$$

can be applied. As this is a quadratic form, the same highly efficient numerical methods can be applied as for the optimization of SSD-based measures. Usually $CC$ is not suited for multi-modality registration since a global linear transformation function of the grey values cannot be presumed [31]. However, in a number of small neighborhoods the assumption of a linear relationship is valid and the cross-correlation coefficient can be used as an indicator of image similarity. If we square and accumulate the local CC values (allowing for positive as well as negative correlated transitions) then also multi-modality images can be registered. The measure is denoted as local correlation:

$$LC = \sqrt{\frac{1}{n} \sum_{S_j \in M} CC^2(S_j)}$$

Where $CC^2$ is the square cross correlation coefficient for the $j$th subregion $S_j$, and $N$ is the number of subregions contained in $N$. $LC$ has been successfully used
for various medical rigid and deformable registration tasks.

Image registration can also be considered within an information theoretic framework [32]. The basic idea is to exploit a statistically significant relationship between the grey values of the input images. This relationship does not have to be explicitly known. The only fact used is that proper registration means proper alignment of significant grey value structures that via their statistical relationship lead to pronounced peaks in the joint grey value distribution detected as maxima of its mutual information or entropy. The mutual information:

$$MI = - \sum_{j,k} \left( \frac{P_{j,k}^{2D}}{V} \log \left( \frac{P_{j,k}^{2D}}{P_{1,j}P_{2,k}} \right) \right)$$

Where $V$ denotes the volume of overlap, $P_{1}^{j}$ and $P_{2}^{k}$ are the probabilities of grey values $j$ and $k$ in the two images respectively, and $P_{1,2}^{j,k}$ is the probability that grey values $j$ and $k$ occur in the fixed and at the corresponding position in the transforming image. $MI$ has become the accepted standard for image registration, in particular for multi-modality applications. Over the last years, a large amount of publications demonstrate that $MI$ can be used without need for pre-processing, user initialization and parameter tuning.

The normalization of $MI$ (NMI) with respect to the image overlap has proven as a useful extension of the measure. A drawback is that $MI$ is not a least-squares criterion and the calculation of derivative information is not straightforward [33]. Steepest decent or simplex optimization schemes are frequently applied which may result in prohibitive computational costs for elastic transformations with a larger number of parameters. A dedicated Levenberg-Marquardt method for MI optimization can be found in. This approach has recently been extended to higher order MI.
Deformation Models

**Parametric Transformations:** A common method for describing parametric elastic transformations is to view the transformation as a deformation field defined by a linear combination of a class of basis functions. Common choices of basis functions are thin-plate spline (TPS) models using radial basis functions, elastic body splines (EBS), or B-splines, and irregular grids parametric registration. The latter have the advantage of local support (basis functions with local support have also recently been applied for registration using TPS and EBS. Using basis functions with compact support, a change of a parameter only affects the transformation in a spatially limited neighborhood while other parts of the deformation remain unchanged. Hence, with respect to image reformation, only the relevant part of the image has to be resampled, which significantly improves the computational performance [34].

**Non-Parametric Transformations:** Non-parametric transformations rely on physical properties and functions to guide the registration process. Solving the transformation may be less efficient, but offer increased flexibility. Common Approach includes:

**Linear Elastic:** the deformation a body undergoes, when subjected to a stress, the force per unit area. There is a linear relationship described by \( F = -kx \), where \( x \) is the change in length of the object, \( F \) is the restoring force exerted by the body, and \( k \) is the spring or force constant. Hookes law can be rewritten, in terms of stress and strain, as \( \sigma = E\epsilon \) or as

\[
\Delta L = \frac{1}{E} \times F \times \frac{L}{A} = \frac{1}{E} \times L \times \sigma
\]

**Visco Elastic:** In some materials, the relationship between stress and strain is not linear. A viscoelastic material exhibits hysteresis in the stress-strain curve and
stress relaxation and creep occurs. As the linear elastic model is represented as a spring, a viscoelastic material is presented using springs and dashpots, connected in a series, a Maxwell material, or in parallel and series, a Kelvin material. In a viscoelastic model the stress and strain are a function of time.

**Hyper Elastic:** The hyper elastic model, the most general type of nonlinear elastic behavior, assumes a strain energy density potential, \( U \), which defines the stresses. The strain equation becomes:

\[
\sigma = \frac{\partial U}{\partial \varepsilon}
\]

, where \( \sigma \) and \( \varepsilon \) are the work conjugate strain and stress measures.

Deformable image registration is one of the key technologies in fMRI analysis [35]. It has been increasingly used in health care for diagnosis, radiation treatment planning, and tumor growth monitoring. Unlike traditional rigid image registration purely focuses point matching, on rotation, and scaling, where the estimation of the geometric deformation is reduced to the search for the 'best' parameters. Deformable Image Registration are non-parametric transformation and the registration is achieved by locating the minimum energy state in an iterative fashion. Below shows an sample result for rigid registration and deformable registration: The first row displays the rigid registration result while the second row shows the non-rigid result.

**2.2.2 Image Preprocessing: Contourlet Transform**

Wavelet transform provides an optimal representation way for one-dimensional piecewise smooth signals. However, natural images are not simply 1-D piecewise smooth scan-lines; discontinuity points like edges are often located along smooth
contours owing to smooth boundaries of the objects. Wavelet transform in 2-D are good at isolating the discontinuities at edge points yet can not identify the smoothness along the contours. Also, wavelets basis can capture only limited directional information. These disappointing behaviors indicate that more powerful transformation are needed in higher dimensions. Contourlet transform is an efficient representation way starts with a discrete-domain construction and with the following features:

1) **Multiresolution:** The representation allows images to be successively approximated, from coarse to fine resolutions.

2) **Localization:** The basis elements in the representation is localized in both the spatial and the frequency domains.

3) **Critical sampling:** the representation forms a basis with small redundancy.

4) **Directionality:** The representation contains basis elements oriented at a variety of directions, much more than the few directions that are offered by separable wavelets.
5) Anisotropy: To capture smooth contours in images, the representation contains basis elements using a variety of elongated shapes with different aspect ratios.

The Contourlet transform has a fast implementation based on a Laplacian Pyramid decomposition followed by directional filter banks applied on each bandpass subband.

Pyramid Frame

Laplacian pyramid (LP) can be used in a multiscale decomposition. The LP decomposition at each level generates a down sampled low pass version of the original and the difference between the original and the prediction, resulting in a bandpass image. Fig. 2.7 depicts this decomposition process, where $H$ and $G$ are called (low pass) analysis and synthesis filters, respectively, and $M$ is the sampling matrix. A drawback of the LP is the implicit over sampling. Similar to LP decomposition the contourlet transform is using the theory of frames and over sampled filter banks. The LP with orthogonal filters, for example, the analysis and synthesis filters are time reversal and synthesis filters is orthogonal to its translates with respect to the sampling lattice by $M$, provides a tight frame with frame bounds are equal
to 1. Hence, Using optimal linear reconstruction using the dual frame operator as shown in Fig.2.8. The new reconstruction differs from the usual method, where the signal is obtained by simply adding back the difference to the prediction from the coarse signal, and was shown to achieve significant improvement over the usual reconstruction in the presence of noise.

**Directional Filter Bank**

The DFB is a shift-invariant version of the critically sampled DFB in the contourlet transform. The building block is also a two-channel non-sub sampled filter bank. To obtain finer directional decomposition, we iterate the DFB for the next level and up sample all filters by a quincunx matrix given by

\[
S_k^{(l)} = \text{diag}(2^{l-1}, 2) \text{ for } 0 \leq k \leq 2^{l-1}, \text{ or}
\]

\[
S_k^{(l)} = \text{diag}(2, 2^{l-1}) \text{ for } 2^{l-1} \leq k \leq 0
\]

The iteration of two-channel filter banks in the analysis part of a LP is showed in Fig.2.9.
Contourlet Transform Framework

The contourlet transform combines pyramids and directional filter banking. LP provide multiscale decomposition and DFB provides directional information. Fig.2.10 illustrate the process. (a) First, LP split the input into a lowpass subband and a highpass subband. Then a nDFB decomposes the highpass subband into several directional subbands. The scheme is iterated repeatedly on the lowpass subband. (b) Resulting frequency division, where the number of directions is increased with frequency.

2.2.3 Data Analysis and Modeling: ICA

Independent component analysis is a recently developed popular method for functional connectivity detection using fMRI [36]. Since it needs no prior information about the spatial or temporal patterns of source signals, ICA is well suited for resting-state fMRI study. Therefore, there is increasing interest in applying ICA algorithm to resting-state fMRI study for functional connectivity detection. Like PCA/SVD, ICA seeks to find a linear combination of components. The difference
is that ICA would find components that are as statistically independent as possible, while PCA/SVD would find orthogonal components. For fMRI data $X$ ($T$ time points × $N$ voxels), the ICA model can be expressed as:

$$X = AC = \sum_{i=1}^{N} A_i C_i$$

Where $C_i$ is the $i$th underlying signal source (IC component); $A$ is the mixing matrix with a dimension of $T \times N$. Different sources are independent from each other:

$$P(C_1, C_2, ... C_N) = \prod_{i=1}^{N} P(C_i)$$

Here, $P(C_i)$ is the probability of the $i$th underlying signal source. Denoting $W$ as the pseudo reverse of $A$ ($W$ also called unmixing matrix), we can obtain the independent components (ICs) simply by:

$$C = WX$$

Figure 2.10: Multi scale decomposition and DFB.
According to whether to decompose the data into spatially independent components and spatially independent time course (sICA), or temporarily independent components and temporarily independent time course (tICA), ICA could be divided into spatial ICA (sICA) and temporal ICA (tICA). Then the question is which type one should choose for functional connectivity detection.

Since the introduction of ICA into fMRI study, both sICA and tICA have been widely used. However, the criterion for which one to use seems to be task dependent. Researchers reported that sICA and tICA can have diverging results, depending upon the characteristics of the underlying signals to be estimated. If the underlying signals are spatial correlated but not temporarily, one may want to choose tICA instead of sICA since sICA would probably not yield the correct activation pattern if the null spatial correlation is strongly violated, and vice versa for tICA.

Despite the increasing popularity of applying ICA algorithm to fMRI study, especially on resting-state fMRI data, there are some pitfalls that need mentioning.
Firstly, ICA is grounded on the assumption of components (signal sources) independence, whether spatially or temporally. Violation of this assumption would decrease the effectiveness of ICA considerably.

Secondly, how to choose the number of independent components and how to threshold the IC maps have become open questions. Ma et al. studied these questions and concluded that when the number of ICs is smaller than that of the source signals, ICA results become highly dependent on the number. Actually, threshold IC maps directly is difficult. In practice, it is common to convert an independent map with a non-Gaussian distribution into a z-map with a Gaussian distribution [37] and [36]. Ma et al.’s results show that the z-map conversion tends to overestimate the false-positive rate (FPR) [38]. This overestimation, however, is not very severe and may be acceptable in many cases.

Last but not least, ICA is a noise-free generative model [39]. The observed fMRI data sets are completely explained by the source signals contained in matrix C and the mixing matrix A, and thus precludes the assessment of statistical significance of the source estimates within the framework of null-hypotheses. To solve this problem, Beckmann et al. recently developed a new model called probabilistic ICA or pICA [40], which assumes that the observed p-dimensional time series are generated from a set of \( q \) \((q \neq p)\) statistically independent non-Gaussian sources (spatial maps) via a linear and instantaneous mixing process corrupted by additive Gaussian noise \( \eta(t) \):

\[
X_i = AS_i + \mu + \eta_i
\]

Where \( X_i \) refers to the p-dimensional column vector of individual measurements at voxel location \( i \); \( A \) is mixing matrix; \( S_i \) denotes the \( q \) dimensional column vector of non-Gaussian source signals contained in the data; \( \mu \) is constant part; and \( \eta_i \)
denotes Gaussian noise $\eta_i \sim N(0,2i)$.

### 2.2.4 Statistical Approaches: Temporal Clustering Analysis

Temporal clustering analysis (TCA) is an exploratory data-driven technique that has been proposed for the analysis of resting fMRI to localize epileptiform activity without need for simultaneous EEG [41]. Conventionally, fMRI of epileptic activity has been limited to those patients with subtle clinical events or frequent introital epileptiform EEG discharges, requiring simultaneous EEG recording, from which a linear model is derived to make valid statistical inferences from the fMRI data [42]. TCA detects the Saliency Window (Temporal Region of Interest) in temporal domain by counting the pixel in each frame reach its maximal value [?]. TCA is a very fast and efficient approach yet it subjects to false detection caused by motion artifacts from different experimental conditions as it is not fully utilize the spatial correlation inherited within a set of fMRI data. Below we show an example of using TCA to detect the time windows in an water drinking activity response in human brain.

### 2.2.5 State-of-the-Art Technique Comparison

The following table shows the major advantage and disadvantage for each of the previous mentioned state-of-the-art techniques. We could see that each of the previous stated method has their disadvantages and their common drawbacks bring us the motivations for further investigation. From next chapter, methods are presented to specially targeting on overcoming these common disadvantages the state-of-the-art approaches poses.
Figure 2.12: TCA time window detection.

Table 2.1: Compare the State-of-the Art Techniques

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICA</td>
<td>no prior knowledge needed</td>
<td>incapable of separate modeling noise and Motion Artifacts with high computational cost</td>
</tr>
<tr>
<td>Deformable Registration</td>
<td>can eliminate part of motion artifacts</td>
<td>incapable of resume the statistical characters of the Image</td>
</tr>
<tr>
<td>TCA</td>
<td>fast, no prior knowledge needed</td>
<td>not fully utilize the spatial/temporal coherence</td>
</tr>
<tr>
<td>Graphical Model Inference</td>
<td>accurate</td>
<td>need prior knowledge</td>
</tr>
<tr>
<td>Maximum Correlation Method</td>
<td>fast and easy to use</td>
<td>lack of using temporal correlation</td>
</tr>
<tr>
<td>General Linear Model</td>
<td>fast and easy to use</td>
<td>tend to yield false activation</td>
</tr>
<tr>
<td>PCA</td>
<td>reduce the data dimensionality</td>
<td>subject to motion artifacts</td>
</tr>
</tbody>
</table>
Chapter 3

Hidden Markov Tree Based Image Restoration

3.1 Introduction

Functional MR Imaging technology measures the hemodynamic response (change in blood flow) related to neural activities in the brain with low invasiveness and high spatial resolution [43]. Compared with CT/EEG, it can collect signals from all regions of the brain and produces illustrative brain “activation” map. However fMRI can only capture a clear image if the subject stays completely still [44]. Motion artifacts and noises often cause blurring and inter-subject variabilities which will lead to false alarms [45]. In brain fMRI, motion artifacts can fall into two major categories [46], (1) involuntary muscle movement such as breath, eye movement [47] etc, and (2) unpredictable changes in blood flow irrelevant to testing tasks. These actions cause image distortions, reduce information acquisition fidelity, and compromise medical evaluations [48].

More specifically, motion artifacts impact the quality of fMRI data in three aspects [49]: (1) pixel displacement within the region of interest (2) image blurring within motion affected regions, and (3) gray level intensity spikes. Most fMRI pre-processing techniques for motion correction utilize spatial and temporal coherence
among image data sets. For example, image registration transforms the different sets of data into one coordinate system with pixel re-alignment [50]. However, it cannot eliminate the blurring and intensity spikes. Spatial filters have been applied to de-blurring (sharpening) and smoothing to reduce the gray-level intensity distortion. Unfortunately, the goals of de-blurring and spike reduction usually contradict to each other [46]. In fact, the intrinsic problem of the above image restoration techniques is insufficient utilization of the structural information among image data sets. Actually, wavelet (contourlet) coefficients of most structured images follow Laplacian distributions [51]. Their joint statistics can be sufficiently represented by a hidden Markov tree (HMT) model [52]. For those images affected with motion artifacts, such pixel displacement will cause the interscale/intrascle dependencies of wavelet (contourlet) coefficients are corrupted. Image blurring and intensity spikes will cause the skewness of the spectrum distribution. Hence, it is rather reasonable to use the HMT model of the baseline image to restore the statistical dependencies of the distorted image in the wavelet (contourlet) domain while performing the conventional image registration procedures.

In this paper, we present an image statistical restoration method based on HMT models. Combined with non-rigid image registration algorithms, the proposed method can restore the following two image properties simultaneously: (1) intensity distribution in the spatial domain and (2) statistical distribution in the wavelet (contourlet) domain. Under this scheme, the wavelet (contourlet) coefficients of the target image are filtered using the HMT model of the baseline image to minimize statistical divergence between two images. An iterative algorithm between image registration and HMT filtering is developed to achieve a tradeoff between least mean square error (in spatial domain) and minimum statistic divergence (in spectrum domain). In order to reduce the computational cost, the images are
processed in blocks. Only those blocks with high statistic divergence are put into the iterative processing. We used several sets of fMRI data collected with different experimental configurations to test the effectiveness of the proposed method. Although this method is developed for neural activity detection, it indeed can be used for a variety of medical imaging applications.

The rest of the paper is organized as follow: Section II reviews the related work. Section III introduces wavelet and contourlet hidden Markov tree models. Section IV describes the preprocessing with HMT models and motion detection based on KL divergence. Section V describes the image statistical restoration with HMT models. Section VI presents the experimental results. Section VII concludes the paper and outlines the future works.

### 3.2 Related Work

Medical image restoration requires both rigid and nonrigid image registration techniques [53] [54] [55]. Rigid registration assumes that the image distortion is caused by rotation and translation of objects being scanned. The rotation angles and displacement values of the distorted image can be estimated through least mean square error optimization [56]. Nonrigid registration, on the other hand, builds the correspondence between the distorted images and the baseline image at the pixel level. It usually contains four steps: locally structured point selection, correspondence construction, displacement estimation, and gray level interpolation [57] [58]. Nonrigid registration approaches have been applied to various medical imaging applications where the objects under examination are deformable or partially invisible [59] [60]. However, both of the above two image registration methods have some limits: (1) They cannot reduce image blurring caused by motion artifacts [61]. (2) They can not
eliminate the gray-level intensity spikes related to motion artifacts [62]. (3) They can not correct the skewness in statistics of images related to motion artifacts [63].

Generally speaking, spatial filtering techniques could be used for eliminating the gray-level intensity spikes and image blurring [64]. Gaussian smoothing filter is commonly used to remove the gray-level intensity spikes caused by false alarms, but it also reduces the signal-to-noise ratio and causes image blurring [65]. Adaptive image sharpening methods such as Wiener deconvolution could reduce image blurring but can not be used to remove intensity spikes [66]. Inference based methods that utilize structured prior knowledge have also been proposed to remove both the intensity spikes and image blurring at high computational cost [67] [68].

Actually, both blurring and intensity spikes change the statistics of images. This change can be better observed in the spectrum domain, as wavelet(contourlet) coefficients contain both localized spatial and frequency information. Fig.3.1 and Fig.3.2 illustrate the wavelet coefficients histogram of an image before and after distortion mainly caused by motion artifacts. The difference in two histograms can be highlighted after fitting the histograms with the Gaussian mixture model. This problem is especially important for fMRI studies. In many data-driven approaches such as independent component analysis (ICA) and temporal clustering analysis (TCA), neural activities are detected through a statistical analysis of data sets [69] [70]. Therefore, in order to reduce detection false alarms, a procedure is required for normalization of statistics among image data sets in addition to regular image registration.

Wavelet hidden Markov tree (WHMT) model has been extensively studied to represent the joint statistics of wavelet coefficients. It can describe the inter-scale/intrascale dependencies of wavelet coefficients in terms of clustering and persistency properties [71]. Various WMHT models have been successfully used for
Figure 3.1: Gaussian mixture model for wavelet/contourlet coefficient histogram. (a) motion affected image (b) Region of Interest (ROI) of the motion affected image Image (c) Baseline image (d) ROI of the motion affected image Image.
Figure 3.2: Gaussian mixture model for wavelet/contourlet coefficient histogram. (e) Wavelet/contourlet coefficients histogram of the motion affected image (f) Gaussian mixture approximation of the histogram (g) Wavelet/contourlet coefficients histogram of the baseline image (h) Gaussian mixture approximation of the histogram.
de-noising and data compression [72] [73]. In the context of fMRI, we believe that the neural activities and image noise/distortions should not change the structural information among images given a certain fMRI experiment setup and measurement protocol. Therefore, the goal of this work is to investigate the possibility of using the WHMT model that can preserve the structural information of fMRI image for high-fidelity neural activity detection. It is expected that the histogram skewness upon wavelet coefficients caused by motion artifacts could be restored though a recursive filtering procedure based on a WHMT model.

Moreover, when motion artifacts cause distortions along curves, a large amount of wavelet coefficients have to be used to represent those curvature features. Recently, contourlet transform has been proposed and studied to represent curvature features with high data efficiency using pyramid filtering and directional filter banks [52]. Therefore, in this work we also study the possibility of using contour hidden Markov tree (CHMT) model to restore structural information of fMRI image data sets. Furthermore, in order to reduce the computational workload, the proposed HMT based image statistical restoration is performed in blocks. For each distorted image, only those blocks with high statistics divergence are processed with the proposed method.

In summary, the proposed image statistical restoration method for neural activity detection with fMRI images include four stages: (1) Preprocessing: remove rigid motion artifacts with rigid image registration and denoising with pre-trained WHMT model. (2) Motion Detection: divide each fMRI image into 8x8 blocks and only select those block with high statistics divergence for restoration. (3) Non-rigid Registration: perform pixel level registration. (4) Statistical Restoration: filter wavelet or contourlet coefficients with HMT models. The whole procedure involves a loop between stages (3) and (4).
Figure 3.3: (a) Interscale dependencies among wavelet coefficients and (b) Interscale dependencies among contourlet coefficients
3.3 Wavelet and Contourlet HMT Model

The discrete wavelet transform (DWT) is a linear transform that represents an image with both the spatial and frequency characterizations. Wavelet Transform coefficients have two unique properties which can be utilized for multi-resolution models to capture the joint statistical structure [71):

1) Clustering: If a particular wavelet coefficient is large/small, adjacent coefficients are very likely also large/small.

2) Persistence: Large/small values of wavelet coefficients tend to propagate across scales.

The wavelet coefficients of an image can be obtained by

\[ W_w = T_w I \]  

(3.1)

where \( W_w \in R^{n \times 1} \) is the wavelet coefficient, \( I \in R^{n \times 1} \) is the source image, and \( T_w \) is the \( n \times n \) linear wavelet transform.

However, the wavelet transform is inefficient to represent 2-D singularities such as edges and curves [52]. Therefore, the contourlet transform has been proposed to efficiently represent image data for its two properties:

1) Directionality: contourlet basis functions can be developed in many directions, as opposed to only three directions of wavelet basis functions. [52]

2) anisotropy: the contourlet basis functions can have various aspect ratios, whereas the wavelet basis functions’ aspect ratio equals only to one.

The contourlet coefficients of an image can be obtained by
Figure 3.4: Hidden Markov tree model: value denotes to wavelet/contourlet coefficients.

$$W_c = T_c I$$  \hspace{1cm} (3.2)

where $W_c \in \mathbb{R}^{m \times 1}$ is the contourlet coefficient, and $T_c$ is the $m \times n$ linear contourlet transform.

In fact, both wavelet and contourlet coefficients follow Laplacian distributions [51]. There are also interscale/intrascale dependency among those coefficients, as illustrated in Fig 2. (a) and (b). Such a statistical dependency can be represented by a HMT model [71].

A HMT model is composed of an observed random tree $W = w_1, ..., w_N$ (wavelet/contourlet coefficients) and a hidden random tree $S = s_1, ..., s_N$, which has the same indexing structure as the observed tree. $S$ has $k$ discrete states, which are referred as $1, ..., K$. 
The joint distribution of \( W \) and \( S \) satisfies the HMT properties if and only if

\[
P(W, S) = P(s_1) \prod_{t \neq 1} P(s_t | s_{\rho(t)}) \prod_t P(w_t | s_t)
\]  

where \( \rho(t) \) represents the parent of node \( t \).

The HMT model’s parameters include:

1) \( \pi_j = P(s_1 = j) \) is the initial hidden state prior.

2) \( A_{ij} = P(s_t = j | s_{\rho(t)} = i) \) is the transition probability matrix.

3) \( \mathcal{N}(w_t | \mu, \sigma^2) = P(w_t | s_t = i) \) is the emission probability density.

4) \( \theta = (\pi, A, \mu, \sigma) \) is the HMT model.

Hence the posterior probability is

\[
P(W | S, \theta) = \frac{1}{P(W, S | \theta)} \prod_{t \neq 1} P(s_t | s_{\rho(t)}) \prod_t P(w_t | s_t)
\]  

where \( P(W | \theta) = \sum_S P(W, S | \theta) \).

The estimation of model parameters can be performed through the Expectation-maximization (EM) iterations [75]. In each E step, the distribution of \( n^{th} \) hidden state is estimated by its posterior distribution, that is

\[
Q_{s_n}^{new} (s_n) = P(S | W, \theta^{old}), \forall_n.
\]

In each M step, the model parameters are updated by

\[
\theta^{new} = \arg \max_\theta \sum_{n=1}^N \sum_{s_n} Q_{s_n}^{new} (s_n) \log P(w_n, s_n | \theta^{old}).
\]
The details of HMT Maximum Likelihood (ML) learning algorithm can be found in [71].

3.4 Preprocessing and Motion Detection

3.4.1 Processing Framework

The whole framework of the proposed image restoration method is shown in Fig.3.5. It contains three components: (1) preprocessing, (2) motion detection, and (3) motion correction.

The preprocessing component performs a rigid image registration and a WHMT model based denoising. The former minimizes the mean square error of gray-level intensity between distorted images and the baseline image caused by subject rotation and translation. The later increases the signal-to-noise ratio of image data by utilizing the persistence and clustering properties of wavelet coefficients [71]. During the motion detection phase, the whole image is divided into $8 \times 8$ blocks. The statistical divergence of those blocks from the baseline image is estimated and then used to localize those regions inflicted with deformable motion artifacts, which actually cannot be removed through the rigid image registration. The motion correction component contains two procedures: non-rigid image registration and HMT model based image statistical restoration. Besides, a recursive loop between motion detection and motion correction is developed to guarantee the total removal of motion artifacts. The neural activity detection algorithms are then performed using those restored images.
3.4.2 Rigid Image Registration

The goal of image registration is to geometrically align two or more images in order to superimpose pixels representing the same underlying structure. Since the seminal works of Viola and Wells [76] and Maes et al. [77] the maximization of the mutual information measure between a pair of images has gained an increasing popularity as a criterion for image registration. The estimation of both marginal and joint probability density functions of the involved images is a key element in mutual information based image alignment. And the calculate of the a least mean square error between images is the fastest way to reach the maximization of the mutual information.

As subject’s head movement, during the experimental and/or clinical procedure is an inevitable part of the functional magnetic resonance imaging (fMRI) brain mapping methods despite the availability of a large variety of head fixation devices employed in these studies. Thus, image registration is an essential processing step in fMRI. An additional challenge is the explicit geometrical deformations associated with fMRI [78]. It is known that orientational changes are problematic in fMRI in the presence of susceptibility differences especially between bone-tissue and air-tissue interfaces. In this paper, A least mean square error based 2-D rigid registration has been applied to get the maximization of mutual information of fMRI data. The details of the algorithm could be found [79].
\[ [R^*, D^*] = \min_{R, D} \|RI_d + D - I_b\|_2 \] (3.7)

Mutual information (MI)-based image registration has been proved to be very effective in multi-modal medical image applications. For computing the mutual information between two images, the joint histogram needs to be estimated. As we know, the joint histogram estimation through linear interpolation and partial volume (PV) interpolation methods may result in the emergency of the local extreme in mutual information registration function. As the local extreme is likely to hamper the optimization process and influence the registration accuracy. We utilize a histogram estimation method (HPV) proposed by [80] and we use an approximate function of Hanning windowed \( \text{sinc} \) as kernel function of partial volume interpolation. And the maximization of the joint entropy matrix is carried out using Powel’s Direction set method [81].

3.4.3 WHMT Model Based Denoising

The Discrete Wavelet Transform (DWT) has proved remarkably successful for estimating signals corrupted by additive white Gaussian noise. The superior results of HMT model denoising have demonstrated that significant performance gains can be achieved by exploiting dependencies between wavelet coefficients [71]. Since the orthogonal DWT of zero-mean WGN is again zero-mean WGN of the same power, the signal estimation problem can be posed in the wavelet domain as: Estimate the wavelet coefficients \( \hat{w}_i \) of a signal given the noisy measurements \( w_i = \hat{w}_i + n_i \), with \( n_i \) has a WGN of variance \( \sigma^2 \). As in [71], an “empirical” Bayesian approach has been adopted and modeled the signal wavelet coefficients \( Y \), using a two-component Gaussian mixture \( (M = 2) \) with \( \mu_i, 1 = \mu_i, 2 = 0 \). If we knew the hidden state \( S_i \)
of $\hat{w}_i$, then the minimum-mean-squared-error (MMSE) estimate would be the conditional mean estimate of a Gaussian signal in Gaussian noise, that is

$$E(\hat{w}_i|w_i, S_i = m) = \frac{\sigma_{i,m}^2}{\sigma_{i,m}^2 + \sigma_n^2} w_i$$  \hfill (3.8)

### 3.4.4 Motion Detection

As motion artifacts could cause pixel displacement, blurring and intensity spikes vary from different places of the fMRI. In order to better detect the motion artifacts as well as reduce the motion correction computational cost, we divided each frame of our fMRI data into a set of blocks with $8 \times 8$ pixels. Motion artifacts could be observed in two different categories: pixel displacement could be observed as non-rigid distortion in spatial domain and motion caused blurring and intensity spikes could be observed as the statistical distortion in spectrum domain. We calculate the KL divergence for the statistical distortion of distorted image and baseline image and the motion affected region will be applied for motion correction.

$$KL(f, g) = \sum_i f_i \log \frac{f_i}{g_i}$$  \hfill (3.9)

where $\{f_i\}$ and $\{g_i\}$ are the discrete distribution densities of wavelet coefficients in distorted image blocks and corresponding baseline image blocks, respectively.

The divergence satisfies three properties, hereafter referred to as the divergence properties.

1) Self similarity: $D(f||f) = 0$.
2) Self identification: $D(f||g) = 0$ only if $f = g$.
3) Positivity: $D(f||g) \geq 0$ for all $f, g$. 


For two Gaussians \( f \) and \( g \) the KL divergence has a closed formed expression.

\[
D(f||g) = \frac{1}{2} \left\{ \log \frac{\Sigma_f}{\Sigma_g} + \text{Trace}[\Sigma_g^{-1}\Sigma_f] - d \\
+ (\mu_f - \mu_g)^T \Sigma_g^{-1} (\mu_f - \mu_g) \right\}
\]  

(3.10)

where \( f \) and \( g \) are \( d \) dimension Gaussian Mixture Models.

\[
f(x) = \Sigma_a \pi_a N(x, \mu_a, \Sigma_a) \text{ and } g(x) = \Sigma_b \pi_b N(x, \mu_b, \Sigma_b)
\]

(3.11)

where \( \pi_a \) is the prior probability of each state, and \( N(x, \mu_a, \Sigma_a) \) is a gaussian in \( x \) with mean \( \mu_a \) and variance \( \Sigma_a \).

### 3.5 Motion Correction

In order to deal with the two different motion artifacts, a non-rigid image registration has been applied on spatial domain to eliminate the non-rigid distortion. And an HMT model based image restoration has been applied on spectrum domain.

#### 3.5.1 Non-Rigid Registration

Deformable image registration allows localized transformations, and is able to account for internal organ deformations. Therefore, it has been increasingly used as it allows more precise targeting and normal tissue preservation. A deformable image registration is called inverse consistent, if the correspondence between two images is invariant to the order of choice of source and target. For example, let \( B \) and \( T \) be the baseline and target images, and \( F \) and \( R \) be the forward and reverse transformations, respectively, i.e. \( B \times F = T \) and \( T \times R = B \), then an inverse consistent
registration satisfies $F \times R = I_d$ and $R \times F = I_d$, where $I_d$ is the identity map. By applying an inverse consistent registration, measurements or segmentations from one image can be precisely transferred to the other. In the image guided radiation therapy, the inverse consistent deformable registration technique provides the pixel-to-pixel mapping between the reference phase and the test phase. This technique is referred to as “automatic re-contouring”.

The main idea of this method is modeling the forward and backward transformations as a one-parameter differ-morphism group. Then, a geodesic path connecting two images is obtained by minimizing an energy functional symmetric to the forward and backward transformations.

$$E(h) = M(S(h), T) + \lambda R(h) + \rho \int_\Omega |h - g^{-1}|^2 dx$$  \hspace{1cm} (3.12)

$$E(g) = M(S, T(g)) + \lambda R(g) + \rho \int_\Omega |g - h^{-1}|^2 dx$$

Where $M()$ is a dissimilarity measure between two images. $g$ is the backward mapping which deforms $T$ such that $T(g)$ is close to $S$ under measure $M$, and $g$ is expected to be the inverse of the forward mapping $h$. $R()$ is a regularity measure on deformation fields $h$ and $g$, $\lambda > 0$ and $\rho > 0$ are parameters balance the goodness of alignment, the smoothness of the deformation, and the consistence of invertibility.

The current framework of variational method finds the forward and backward transformations that deform a source image $S$ to match a target image $T$ and vice versa. In this work, we use a method which deforms $S$ and $T$ simultaneously, and the registration matches the deformed source and deformed target images. As the disparity between deformed $S$ and deformed $T$ is smaller than that between deformed $S$ and fixed $T$ or deformed $T$ and fixed $S$, the deformation by the bidirectional
simultaneous deformations is in general smaller than the deformation by unidirectional deformation that deforms $S$ full way to $T$ or $T$ full way to $S$. Therefore, as shown in our experimental results deforming $S$ and $T$ simultaneously leads to a faster and better alignment than deforming $S$ to the fixed $T$ or vice versa.

Details of the algorithm could be found in [82].

3.5.2 HMT Restoration

In the HMT model, the form for the marginal distribution of a wavelet/contourlet coefficient comes directly from the efficiency of the wavelet/contourlet transform in representing real-world images: a few wavelet coefficients are large, but most are small. This property can be captured by two-state zero-mean Gaussian mixture (GM). And the statistical restoration could be achieved by restored the wavelet/contourlet coefficients through Expectation/Maximization.

For the E-step we calculate the direct transposition to the Hidden Markov Tree (HMT) context of the forward-backward algorithm. The upward-downward recursions require marginal state distributions to be known. Hence the initial hidden state prior and the recursion $\pi_j = P(s_1 = j)$, and the transition probability $A_{ij} = P(s_t = j | s_{\rho(t)} = i)$ have to be known for the computation. The upward recursion for leaves of the tree:

$$\beta_t(j) = \frac{C_{jt}P(s_t = j)}{\sum_j C_{jt}P(s_t = j)}$$  \hspace{1cm} (3.13)

where $C_jt = \mathcal{N}(w_t|\mu, \sigma^2) = P(w_t|s_t = i)$ is the emission probability.
The upward recursion for non-leaves:

\[ \beta_t(j) = \frac{\sum \prod_{v \in C(t)} \beta_{t,v} C_{jt} P(S_t = j)}{\prod_{v \in C(t)} \beta_{t,v}} \]  

(3.14)

where

\[ \beta_{p(t),t(j)} = \sum_{i} \frac{\beta_t(i) P_{ji}}{P(s_t = i)} \]

The downward has the following expression:

\[ a_t(j) = \frac{1}{P(s_t = j)} \sum_{i} \frac{A_{ij} \beta_{p(t)}(i) a_{p(t)}(i)}{\beta_{p(t),t}(i)} \]  

(3.15)

At the M step, the maximization of the expectation of log likelihood of the complete data \( \log Q(W|\theta) \) is computed in order to re-estimate the model parameters to be used in the next iteration.

\[ \theta^{new} = \arg \max_{\theta} \log Q^{dd}(W|\theta^{old}) \]  

(3.16)

where \( \theta^{new} \) represents the model parameters at next iteration and \( \log Q(W|\theta) \) is the expectation of the likelihood of the complete data with respect to the current estimate of the distribution of latent variables. The \( \alpha \) and \( \beta \) probabilities determined at the E-step are used to find the expression of \( \log Q(W|\theta) \) as a function of parameter \( \theta \) of the hidden Markov tree

\[ \log Q(W|\theta) = \sum_{i} \sum_{j} P(S_t = j|W, \theta) \log A_{ij} \]  

(3.17)

The whole procedure can be listed below:

1) Set counter \( k = 0 \), initialize the parameters of HMT model \( \theta \), using the method
of within-scale scanning and across-scale counting

2) Compute $p_k(m)$, $S_i$ using the Upward-Downward algorithm;

3) Solve problem and obtain the solution using $w, k$ preconditioned conjugate gradient method;

4) If $\frac{1}{N} \sum_{i=1}^{N} |w_i^{k+1} - w_i^k| < \varepsilon$, go to step 5.

5) Reconstruct the recovered image by inverse wavelet transform.

The training algorithm in the HMT model is similar with that of [71]: First Calculate the probability mass function. Then update the $W$ with:

$$W_i = w_i \left( \frac{\sigma_{i,1}^2}{\sigma_{j,1}^2} P_1 + \frac{\sigma_{i,2}^2}{\sigma_{j,2}^2} P_2 \right)$$ (3.18)

where $P_1$, and $P_2$ are the probabilities of the wavelet/contourlet coefficients belonging to the specific gaussian distributions.

### 3.5.3 Robust Training via Tying

If we do not have enough training data to prevent “over-fitting”, we cannot expect to robustly estimate the marginal densities of the wavelet/contourlet coefficients and estimate the joint density for the entire wavelet/contourlet transform. As we just have limited training data are available, in order to make our modeling more robust, we modeled random variables that have similar properties by using a common density or a common set of density parameters. For example, if we expect two random variables to have roughly the same variability, we can describe them use a common variance. In this way, we obtain more reliable parameter estimates by
increasing the amount of training data associated with each parameter. This prac-
tice is known as tying in the HMM literature [71], and we use it to more robustly 
estimate the means, variances, and transition probabilities of our wavelet-domain 
HMMs.

In Fig.3.6, we distinguish between two different types of tying in the HMT model: 
tyling between wavelet trees and tying within wavelet trees. By tying across trees 
which assumes that the coefficients of these trees have the same density we can train 
as if we had multiple signal observations. We can also tie within trees by tying all 
coefficients within the same scale of a tree.

### 3.5.4 Loop Between Motion Detection and Correction

A loop is developed between detection and correction.

The main advantage for restoration using HMT is that the multi-layer signal 
model is capable of capturing the signal activities in coarse to fine scales. Hence we 
could able to first identify roughly where the motion and noise locates and then fine 
tune the motion affected region through motion correction through iterative updat-
Figure 3.7: Flowchart of our approach.
ing the wavelet/contourlet coefficients in the specific branches in the HMT model. This leads to reduce the computational complexity while keeps the restoration accuracy. After the mixture probabilities updating, the motion free image will be reconstructed through wavelet domain using the modified HMT parameters. And based on the restored images, various neural activity detection algorithms such as temporal clustering analysis, independent component analysis [83] can be applied for neural activity detection.

3.6 Experiment Results and Discussion

We present our experimental results in the following manner: First, we verify our approach through a set of simulation data generated by ourself in which it contains noises, neural activities and motion artifacts. After verify the effectiveness of our approach, we test it using the real fMRI data collected from water drinking and glucose intaking. Then we compare our method with other popular approaches to verify the accuracy as well as the the ability to eliminate the motion artifact of our approach. Finally we discuss the strengthens and weakness of our approach, and our challenges and future work.

3.6.1 Simulation Results

Simulation Data Description

First, we verify our proposed framework using simulation data. In order to reduce the computational cost, we generate some small sets of pictures: Each set of data contains 100 frames. Each frame has the resolution of 64*64. One sample frame of motion free simulation data can be seen in Fig.3.8 (a)
Figure 3.8: Image restoration and denoising for simulation data: (a) Baseline image; (b) Noise affected image; (c) Motion affected image (d) HMT denoising result. (e) Wavelet HMT restoration result. (f) Contourlet HMT restoration result.
Figure 3.9: Wavelet/contourlet coefficients histogram for (a) Noise affected image (b) Wavelet histogram for motion affected image (c) Contourlet histogram for motion affected image. (d) De-noised wavelet coefficients histogram (e) WHMT restoration wavelet histogram (f) CHMT restoration contourlet histogram
The base simulation data were draws roughly by a human brain shape. The hypothalamus area of human brain is painted with separate dark gray color. We did so because in our real data, most of the neural activities in real data related to our performed tasks occur in this area. We add gaussian white noise in each frame of the original simulation data and we add artificial distortions which cause both pixel level displacement, gray scale change and blurring. They can be seen at Fig.3.8 (b) and (c).

Neural activities within hypothalamus part were added in temporal domain to simulate the real data. The neural activities was generated using the impulse response model as the difference between two gamma functions:

\[ h(t; \tau_1, \tau_2, \delta_1, \delta_2, c) = \left( \frac{t}{\tau_1} \right)^{\delta_1} e^{-(\delta_1/\tau_1)(t-\tau_1)} - c \left( \frac{t}{\tau_2} \right)^{\delta_2} e^{-(\delta_2/\tau_2)(t-\tau_2)} \] (3.19)

Where \( \delta_1, \tau_1, c \) will determine the IR model shape and \( \delta_2, \tau_2 \) are set to have linear relationship to \( \delta_1, \tau_1 \).

Simulation Results

Fig.3.8 (a) in the first row on the shows one frame of the original simulation data which is supposed to be motion free. (b) shows gaussian noise blurred image and (c) shows motion distorted image. the second row from left (d) shows the denoised data using HMT denoising, (e)/(f) shows motion corrected through wavelet/contourlet HMT model based restoration.

Fig.3.9 shows the two variances of the gaussian mixtures in HMT model corresponding to Fig.3.8. As we could see, both noises and motion artifacts will change the statistical property of the image and we could observe this distortion by the
change of the variances of the gaussian mixtures of the images. After we use HMT model based restoration restored the image, we could see the restoration successfully corrected the statistical distortion by correcting the variance of the gaussian mixture model. This could be seen at Fig.3.9 (d), (e) and (f) compare with (a).

Fig.3.10(a) displays the activation map of the baseline data, which we thought work as a reference for our approach. As motion artifacts caused blurring and pixel spikes, traditional non-rigid registration can not eliminate the motion artifact through just simple pixel-level realignment and this leads to a false detection on the brain activation map. The false detection block is highlighted in white square in Fig.3.10 and (b). Our HMT model based method both works on wavelet and contourlet domain successfully restored the image and get the correct activation map shows on Fig.3.10 (c) and (d).

### 3.6.2 Real fMRI data Results

#### Real Data Description

Our real fMRI data are collected by performing two tasks: water drinking and glucose intaking. Each task contains six sets of fMRI data with each set contains 183 images. Each image in the dataset has the resolution of 255*256. Water drinking activities were taken vary from 54” to 1’40”. Glucose intake activities were taken vary from 50” to 1’28”. This time difference will cause degree of the reaction of the neural activity differ from each other, yet because they are performing the same task. Their neural activity response will be similar. For these test, we first carefully select the baseline image which we believe it has no motion artifacts. Rigid registration and HMT model based de-noising were applied to the dataset in order to minimize the from noise and rigid distortion caused by motion artifacts.
Figure 3.10: Activation map for the simulation data (a) Baseline motion free activation map (b) Non-rigid registration of motion affected data’s activation map (c) WHMT restored activation map (d) CHMT restored activation map
Motion Correction Results

Fig.3.11 shows the image restoration results on real data in the hypothalamus area. On the upper row from left to right are: baseline image; non-rigid registration result; wavelet HMT based restoration result; and contourlet HMT restoration result. The lower row shows the wavelet/contourlet coefficients histogram of the baseline image, non-rigid registration image, WHMT restored image and CHMT restored image. As we are not only focus on the spatial domain, our result yield better result on resuming the statistical properties by correct the variance of the gaussian mixtures of the tested images.

Results Compare with Other Methods

Fig.3.12 displays a sample motion distored image restored by both rigid and non-rigid registration, and our HMT model based approach. The first row shows the restoration result for (a) standard mean-square registration. (b) 2-D deformable registration. (c)HMT based Restoration on wavelet domain and (d) HMT-based Restoration on contourlet domain. The second row shows the residual between the source image and the restored motion image. As we could see, the residue of the both rigid/non-rigid registration could still see the roughly the brain shape. This is due to the motion caused blurring which works as a low pass filter. The high frequency edge and contour are still tend to me not well registrated through both rigid/non-rigid algorithm. We could see our result almost shows white noise which means we correct the motion cause blurring though statistical restoration in the wavelet/contourlet domain. Fig.3.13 (a) and (b) show the mutual information and signal to noise ratio between the baseline image and the restored image. Our results shows our image have higher mutual information with the baseline image and the
Figure 3.11: Image restoration with water drinking activity: the right column are the image section contains the neural activities, and the right column are the wavelet/contourlet coefficients histograms.
images restored by our method have higher signal to noise ratio than rigid/non-rigid registration.

Fig.3.14 (a) shows the average residue sum between restored images and baseline image for the six data sets. We could see our method shows smaller residue sum and yield the best performance. Fig.3.14 (b) shows the gaussian mixture difference of the wavelet/contourlet coefficients histogram difference between restored and baseline image. As both rigid and non-rigid registration did not consider the statistical characters of the image, their registered image still shows a big difference in the contourlet/wavelet HMT domain while our approach minimize the difference.

**Neural Activity Detection**

We use temporal clustering analysis to verify that our restored data are not only resumes the statistical character of the image, but also preserve the necessary neural activity signal with the data which we are interested. In Fig.3.15 shows the average $f_{N_{\text{max}}}$ value. $f_{\text{Nmax}}$ value means the number pixel reach it’s maximum value in the temporal domain at particular time. By counting the $f_{\text{Nmax}}$ value, we could be able to detect the neural activity time window and catch exactly where the neural activity occurs. However, because motion artifact and noise often distorted the image and cause pixel level intensity spikes. They will corrupt the $f_{\text{Nmax}}$ value and cause wrong time window detection and get the $f_{\text{Nmax}}$ value wrong. From Fig.3.15 we can see that we greatly reduce the $f_{\text{Nmax}}$ value yet still keep the spikes which contain the neural activity related to water drinking..

Fig.3.16 shows the activation map detection results. As we can see as traditional registration methods (both rigid and non-rigid) did not fix the pixel level intensity spikes, their restored image are subject to false detection. (highlighted in white dot block in the Fig.3.16 (a) and (b)), and our methods shows better detection accuracy.
Figure 3.12: Image restoration result comparison: (a) Rigid registration result (b) Non-rigid registration result (c) Wavelet HMT restoration result (d) Contourlet HMT restoration result (e-h) Average residue between restored image and baseline image;
Figure 3.13: Image restoration result comparison: (a) Mutual Information of the four methods; (b) Signal to noise ratio of the four methods;
Figure 3.14: (a) The average residue for 6 data sets between restored images and baseline image. (b) The difference of the gaussian mixture model for 6 data sets between restored images and baseline image
Figure 3.15: fNmax value after image restored by (a) Mean square registration (b) 2d-deformable registration (c) 2d-deformable registration on wavelet domain (d) Wavelet HMT based restoration
3.7 Discussion

As we stated in the introduction part, modeling and eliminate the motion artifacts is still a huge challenge for fMRI data. As motion artifacts will cause pixel level displacement, intensity spikes and blurring. It is hard for us to tackle the three challenges in the same time as the methods to solve blurring and intensity spikes are contradict to each other. We argue that by utilize the structural information among image data sets, actually both distortion could be able to observed in the spectrum domain such as wavelet/contourlet domain as both obeys Laplacian distribution. Hence their joint statistics can be sufficiently represented by a hidden Markov tree (HMT) model. Image blurring and intensity spikes will cause the distortion of the spectrum distribution. By using HMT model combined with traditional non-rigid image registration, we could be able to correct the blurring and intensity spikes which cannot be able eliminated at the same time through traditional approaches.

Currently, our method is still not computational efficient as even we divide the test image into blocks for minimize the computational cost, the non-registration between blocks still takes long time to converge. We are hope we hope develop a registration method on the HMT tree to have a faster converge algorithm.

Also how to differentiate the neural activities from motion artifacts in the HMT model in wavelet/contourlet domain is still a problem. Certain prior knowledge (like the neural activity impulse response model, What kind of task it performs) could be used for preventing false eliminate the neural activity as motion artifacts.
Figure 3.16: Activation map of (a) Rigid registration. (b) Non-rigid registration. (c) Wavelet base HMT restoration (d) Contourlet based HMT restoration.
3.8 Conclusion

In this paper, a hidden markov tree (HMT) based contourlet transform model is established for modeling functional Magnetic Resonance Images (fMRI). Through wavelet/contourlet transform, the fMRI data could be represented at the least amount of coefficients with dependencies across the scales. As wavelet/contourlet coefficients could be statistically modeled by two-state zero-mean gaussian mixtures models in HMT model, motion artifacts cause the skew of the model could be observed in hidden markov tree structure. By iteratively resuming the statistical properties of the gaussian mixture, motion artifacts could be eliminated and fMRI can be restored with high accuracy. Experimental results show that our method yields superior motion artifacts removement and better signal-to-noise ratio than traditional methods such as elastic registration and pure HMT tree based image restoration. A list of the equation used in the different stages is listed below.
Chapter 4

Independent Component Analysis with $L_1$ Regularization

4.1 Introduction

Functional Magnetic Resonance Imaging (fMRI) measures the hemodynamic response related to neural activity in the brain or spinal cord of human beings [84]. In neuroscience studies, various data mining approaches have been developed to detect neural activities (i.e., activation, deactivation, and normality) through the fMRI measurements [85]. Data-driven approaches, such as temporal clustering analysis (TCA) [86] and independent component analysis (ICA) [87], generate statistical parameters of the noisy time-courses without using any prior knowledge [88]. Those approaches are robust against the variable nature of the hemodynamic response but tend to produce false alarms [89]. Model-based approaches, on the other hand, can capture repeatable and stable patterns in neural activities based on the knowledge of the hemodynamic response [90]. Those methods can provide insights into how a particular cognitive process is implemented in a specific brain area instead of merely identifying where such a process is located [91].

In general, the development of neural activity detection methods has to address the following challenges: (1) the variability of the hemodynamic response among
subjects and experimental contexts, (2) the large volume of fMRI data which may incur high computational complexity and cost, and (3) the artifacts caused by subject motions and noise under various experimental conditions.

To tackle the above challenges, it requires innovation and development of techniques for (1) an efficient representation of neural activity dynamics, (2) a proper exploitation of the sparse nature of the neural activity sources, and (3) an effective way to remove motion artifacts.

Previous research works has been working on these challenges by (1) studies neural hemodynamic response and established a NHR model for efficient neural activity representation [92]. (2) Using gaussian maximum likelihood spatial smoothing filter for sparsity exploration [93]. (3) Using independent component analysis and deformable image registration for motion artifacts removal [32].

However, also many achievements have already been done. There are still flaws for the previous works (1) The parameters of NHR model are unknown and need the procedure of regularization. (2) No approaches currently available for generating the activation and deactivation map of the neural activities at the same time. (3) It is difficult to separate the neural activities in different scales and provide more information associates with the brain function as no method uses the knowledge on neural dynamics.

In this paper, we present a sparsity-exploiting approach for brain neural activity detection. The sparse neural activity sources is modeled as a series of delta functions (‘1’ represents activation, ‘-1’ represents deactivation and ‘0’ represents normality). A multi-layer, multi-scale neural signal model is developed to represent the neural activities of different types (satisfaction, thirsty, muscle controlled functionally impulse response) based on NHR models under different parameterizations. Since both the neural activity sources and NHR model parameters are unknown,
we develop an expectation-maximization procedure to find their solutions with the maximum likelihood. In each E-step, the sparse neural activity sources are estimated with a set of NHR models. To explore the sparsity of neural activities, $L_1$ regularization is utilized under a Laplace distribution assumption upon the sparse signals. In each M-step, the parameters of NHR models are optimized for the given sparse sources in light of maximum likelihood. The proposed approach is developed under the independent component analysis framework, which decomposes the fMRI measurements into three components: neural activities (modeled signals), motion artifacts (un-modeled signals) and measurement noise (random signals).

The rest of the chapter is organized as follows. Section II reviews the related work. Section III describes the preliminaries and problem statement. Section IV presents the developed approach. Section V shows the experimental results. Section V concludes the paper.

### 4.2 Related Work

In fMRI data analysis, various independent component analysis (ICA) approaches have been applied to identify components common to the whole group as well as components manifested in single subjects only. ICA methods would decompose a set of measurements into different signals originating from the stimulation, slowly varying sources, motion artifacts and noise. Those methods don’t require a prior information on signals and experiment contexts. Compared to single subject ICA, components with similar spatial as well as temporal characteristics across subjects can be detected by group ICA.

ICA methods are known for their incapability of processing high-dimensional signals and more than one Gaussian components [94]. One solution is to use princi-
ple component analysis (PCA) [95] or other data dimensionality reduction methods (i.e. random projection) to achieve whitened, uncorrelated low-dimensional data first. Then the task of ICA becomes performing a proper rotation of signals in light of maximum non-Gaussianity (mutual information). Another solution is to associate more a priori information with each signal component. Actually, fMRI measurements include several components: high-frequency varying components that associated with stimulation [96], slowly varying components that associated with default brain neural activity (or activity unrelated to stimulation), artifacts caused by subject motion and measurement noise.

Neural hemodynamic models (NHM) can account for temporal, spatial coherence of fMRI data [30]. Hemodynamic models can fall into two categories: statistical and physical. The former includes the general linear model (GLM) and predictive linear coding (PLC) [97]. The later uses nonlinear (gamma) functions to describe the delay and undershoot in the hemodynamic response [93]. The physical models can provide more insight into understanding the fMRI signals but their parameters need to be determined. In this work, we develop a multi-layer, multi-scale NHR model that can describe both high-frequency and low-frequency components of fMRI measurements.

Meanwhile, the the sparsity of neural activity sources in the large volume of fMRI data needs further exploitation to reduce the computational cost and achieve computational stability. The neural activity sources can be modeled as sequences of ternary values. As a result, the neural activity detection becomes an ill-posed inverse problem. The efficiency and convergency of ICA approaches are determined by the underlying statistical assumption upon noise and signals of interest. As a matter of fact, when the noise is assumed as Gaussian and the ternary sources are Laplacian, the detection problem is equivalent to the $L_2$ regularization with $L_1$
constrain [98]. By minimizing the $L_1$ penalty of the results in variable selection, the coefficients of zero are effectively omitted from the model. An $L_2$ penalty pushes the neural activity toward zero with a force proportional to the value of the coefficient, and an $L_1$ penalty exerts the same force on all nonzero coefficients. Hence for variables that are most valuable will be located. This method could identify neural activities with high accuracy at low computational cost.

However, the $L_2$ and $L_1$ method cannot deal with motion artifacts that cannot be modeled as Gaussian/Laplace processes or NHR models. Therefore, in this work we develop a $L_1$ regularized ICA approach. The signal component associated with neural activity is filtered with the proposed multi-layer, multi-scale NHR model through a recursive $L_1$ regularization. The motion artifacts and measurement noise are extracted from the fMRI measurement in light of maximum non-Gaussianity of these components. This approach can be used to study both individual and group data to find specific and common neural activity patterns with respect to different stimulations.

4.3 Modeling and Problem Statement

4.3.1 Multi-Scale Neural Activity Model

The multi-layer neural activity model is developed to describe multi-scale hemodynamic responses in voxels. The evoked hemodynamic response is modeled by convolving a ternary reference time course $I(t)$, indicating stimuli presentations or onsets, with an impulse function $h(t)$. The impulse response can be modeled as a
difference between two gamma functions [93]:

\[ h(t; \tau_1, \tau_2, \delta_1, \delta_2, c) = \left( \frac{t}{\tau_1} \right)^{\delta_1} e^{-(\delta_1/\tau_1)(t-\tau_1)} - c \left( \frac{t}{\tau_2} \right)^{\delta_2} e^{-(\delta_2/\tau_2)(t-\tau_2)} \] (4.1)

Based on the multi-scale impulse response models, the neural activity signal model as shown in Fig. ?? . This linear model can be described by

\[ S = Lx = \begin{bmatrix} L_1 & L_2 & L_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \] (4.2)

where \( S \) is the fMRI time course data; \( x_1, x_2 \) and \( x_3 \) are the sparse signal sources at different layers; \( A_1, A_2, \) and \( A_3 \) represent the linear hemodynamic models of 3 layers, respectively. At each layer, the hemodynamic model contains an integrator and a impulse response model in a certain scale,

\[ L_i = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} h^i_0 & 0 & 0 & \cdots & 0 \\ h^i_1 & h^i_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h^i_n & \cdots & \cdots & h^i_0 \end{bmatrix}, \] (4.3)

where \([h^i_0, h^i_1, ..., h^i_n]\) is the hemodynamic impulse response in the \( i^{th} \) scale. In this model, the ternary neural activity signals are further sparsified through differentiation.

### 4.3.2 Independent Component Analysis

Independent Component Analysis (ICA) is a framework for separating out a mixture \( x \in R^m \) of independent sources \( s \in R^n \) with little or no prior information other than
the non-Gaussianity of the source distributions. If the number of sources is equal
to the number of mixtures the problem is one of identifying the mixing matrix $A$
and the source estimates are simply $s = A^{-1}x$. The common approach is to build a
probability density model for the observed data with the associated independence
constraints. Model estimation can then be done using simple learning algorithms
based on Maximum Likelihood (ML) principle.

We model the noisy data over complete ICA is as follows:

$$x_t = As_t + e_t$$  

(4.4)

where $A$ is an $m \times n$ matrix, whose columns are basis vectors with $n > m$, $x_t$ is
the observation vector, $s_t$ is the independent source vector.

The redundancy in the over complete representation is removed by defining a
density for the basis coefficients, $P(s)$, which specifies the probability of the alternative
representations. The most probable representations, $\hat{s}_t$, is found by maximizing
the posterior distribution:

$$\hat{s}_t = \max_s P(x|A, s_t) = \max_s P(s_t)P(x|A, s_t)$$  

(4.5)

$P(s)$ influences how the data are fit in the presence of noise and determines the
uniqueness of the representation. In this model, the data is a linear function of $s_t$.
If the basis function is complete ($A$ is invertible) then, assuming broad priors and
low noise, the most probable internal state can be computed simply by inverting $A$. 

78
4.3.3 Data Dimensionality Reduction

Principal Component Analysis (PCA) is a commonly used technique for data dimensionality reduction. In PCA, the eigenvalue decomposition of the data covariance matrix is computed as \( E(XX^T) = E \Lambda E^T \) where the columns of matrix \( E \) are the eigenvectors of the data covariance matrix \( E(XX^T) \) and \( \Lambda \) is a diagonal matrix containing the respective eigenvalues.

If dimensionality reduction of the data set is desired, the data can be projected onto a subspace spanned by the most important eigenvectors:

\[ X^{PCA} = E_k^T X \]  

(4.6)

where the \( d \times k \) matrix \( E_k \) contains the \( k \) eigenvectors corresponding to the \( k \) largest eigenvalues. PCA is an optimal way to project data in the mean-square sense: the squared error introduced in the projection is minimized over all projections onto a \( k \)-dimensional space. Unfortunately, the eigenvalue decomposition of the data covariance matrix is very expensive to compute. The computational complexity of estimating the PCA is \( O(d^2 N) + O(d^3) \).

A closely related method is Singular Value Decomposition (SVD): \( X = U S V^T \) where orthogonal matrices \( U \) and \( V \) contain the left and right singular vectors of \( X \), respectively, and the diagonal of \( S \) contains the singular values of \( X \). Using SVD, the dimensionality of the data can be reduced by projecting the data onto the space spanned by the left singular vectors corresponding to the \( k \) largest singular values:

\[ X^{SVD} = U_k^T X \]  

(4.7)

where \( U_k \) is of size \( d \times k \) and contains these \( k \) singular vectors. Like PCA,
SVD is also expensive to compute. However for sparse data matrices, there exists numerical routines such as the power or the Lanczos method that are more efficient than PCA. For a sparse data matrix $X_{d \times N}$ with about $c$ nonzero entries per column, the computational complexity of SVD is of order $O(dCN)$.

Random projection is a powerful technique for dimensionality reduction. The original $d$ dimensional data is projected to a $k$ dimensional ($k << d$) subspace through the origin, using a random $k \times d$ matrix $R$ whose columns have unit lengths. Using matrix notation where $X_{d \times N}$ is the original set of $N$ $d$ dimensional observations

$$X_{k \times N}^{RP} = R_{k \times d} X_{d \times N}$$

(4.8)

is the projection of the data onto a lower $k$ dimensional subspace.

The key idea of random mapping arises from the Johnson-Lindenstrauss lemma: if points in a vector space are projected onto a randomly selected subspace of suitably high dimension, then the distances between the points are approximately preserved. Random projection is computationally very simple: forming the random matrix $R$ and projecting the $d \times N$ data matrix $X$ into $k$ dimensions is of order $O(dkN)$, and if the data matrix $X$ is sparse with about $c$ nonzero entries per column, the complexity is of order $O(ckN)$.

By using the random projection, we greatly reduced our source data yet still preserve the neural activities within the brain fMRI for ICA.

### 4.3.4 Problem Statement

Assuming that we have observed data samples $x = x(1), x(2), \ldots, x(T)$; which are generated according to model $x = As + \epsilon$, the approximate conditional moments for
the gaussian posterior distribution, based on the Laplace estimation are given as:

\[ E\{s|x(t)\} = \hat{s} = \arg \max_s \log \{p|x(t), A\} \] (4.9)

From Fig.4.1 we know the source data could be able to converted into three different components which represents: neural activities, motion artifacts and noises. As our previous statement. Neural activities could be further divided into a ternary sparse sources (with 1 represents the activation and -1 represents the de-activation.) convolute with the hemodynamic models. In facts, as motion artifacts and noises are meaningless to the neural activities as long as they are independent with the neural activities signal. The brain neural activity detection problem could be described as follow for the temporal resolution of \( N \), to solve the problem, we need to detect the neural activity in the following way:

\[ D_r = D_s \times P = H_I \times [S; D; N] \times I_c \] (4.10)

Where \( D_r \) is the compressed data with the dimension of \( N \times 3 \). \( D_s \) is the observed source data with the dimension of \( N \times K \). \( P \) is the random projection matrix with the dimension of \( K \times 3 \). \( H_I \) is the \( N \times N \) matrix of the multi-scale hemodynamic responses convolute with the integrator. \( S, D, N \) represent sparse neural activity signals, motion artifacts vector and noise vector, respectively. Each of them are \( N \times 1 \) vector. \( I_c \) is the independent component analysis matrix with the dimension of \( 3 \times 3 \).

In order to find the sparse neural activities \( S \), as we only know the observation data \( D_s \), the problem for detecting neural activity becomes to find the sparse neural activity vector which independently with noise and motion artifacts that fits the
Figure 4.1: The proposed model for the random projection and ICA with neural activity detection model.

\[
E\{S|x(t)\} = \arg \max_H \log(P|x(t), H)
\]  
(4.11)

Where \( H \) is the hemodynamic model. The procedure for finding the sparse neural activity is described below:

### 4.4 Sparse Neural Activity Detection

#### 4.4.1 Convert the ICA to \( L_1 \) Regularization

As we previously stated, we use ICA to find the three independent components of the fMRI: neural activities, motion artifacts and noise. Then based on the hemodynamic response model, we convert the neural activities into sparse neural activity vectors. In fact, the two states can be emerged into one unique view: Convert the problem of finding independent component results into finding solutions of constrained \( L_1 \) regularization with the Laplacian prior:

Our initial model \( x_t = As_t + \epsilon \). where \( x \) is our observed data. \( s_t \) is the independent components. the \( \epsilon \) represents noise/motion artifacts. As we’ve established the hemodynamic response model. the problem hence becomes: \( x_t = A \times h[S, D, N] \)
Suppose we have \( K \) observations: \( x(1), x(2), \ldots, x(k) \). \( A = (a_1, a_2, \ldots, a_N)^T \) are the transformation matrix which convert the source into independent components. To calculate the likelihood:

\[
\mathcal{L}(A) = \sum_{t=1}^{K} \sum_{i=1}^{N} \log P_i(a_i^T x(t)) + K \text{lg}(\det |A|)
\]

(4.12)

as \( A \) has orthogonality. \( K \text{lg}(\det |A|) \) is constant. Hence to maximize \( \mathcal{L} \) equals to maximize

\[
J(A) = E \sum_{i=1}^{N} \log P_i(a_i^T x(t))
\]

(4.13)

To maximize eq.(13). It is same to say we minimize:

\[
\sum_{i=1}^{K} ||A \times x(t)||_1
\]

(4.14)

When \( P_i \) is a Laplacian distribution.

\[
P(w^T x | \lambda) \propto \prod_i \exp(-\lambda |a_i^T x|) = \exp(-\lambda ||a_i^T x||_1)
\]

(4.15)

where \( \lambda \) is the Laplacian distribution parameter.

### 4.4.2 Framework

As the previous argument, our proposed framework could be able to see at Fig.5.1. Raw fMRI data will be first compressed via random projection using through gaussian random matrix to reduce the data amount while keep the key information unchanged. An independent component analysis is then perform based on the compress data to reduce the noise and motion artifacts caused fMRI distortion/blurring and intensity spike in the spatial domain. By using our previously constructed
Figure 4.2: Proposed framework for the ICA with multi-scale, sparsity-exploiting constrained L1 regularization approach.

multi-layer neural activity detection model, we convert the brain neural activity into detection problem into a problem of finding the sparse solution using constrained L1 regularization. Finally a spatial smoothing filter is applied to further increase the signal-to-noise ratio of the fMRI data.

4.4.3 Constrained L1 Regularization

Let \((y_1, x_1), ..., (y_n, x_n)\) be \(n\) output/input pairs where \(y_i \in \mathcal{Y}\) and \(x_i = (x_{i1}, ..., x_{ip})' \in \mathcal{X}\) where \(\mathcal{X}\) is a subspace of \(\mathbb{R}^p\), the \(p\)-dimensional Euclidean space. Let \(\beta = (\beta_1, ..., \beta_p)'\) be the corresponding regression coefficients. Given loss function \(\mathcal{L}: \mathcal{Y} \times \mathcal{X} \times \mathbb{R}^p \rightarrow \mathbb{R}\), the objective of generalized LASSO is to find \(\beta\) which minimizes the empirical risk

\[
R(\beta) = \sum_{i=1}^{n} L(y_i, x_i, \beta) \tag{4.16}
\]

subject to \(||\beta||_1 \leq \lambda\) where \(\lambda > 0\) is the constrained parameter and \(||\beta||_1 = \sum_{k=1}^{p} |\beta_k|\) denotes the \(L_1\) norm on \(\mathbb{R}^p\).

Let \(w = \beta/\lambda\) and \(S = w : ||w||_1 \leq 1\). Then the regularization problem of gener-
alized LASSO is equivalent to minimizing $C(w) = R(\lambda w)$ subject to $w \rightarrow S$, and so we are to find $\hat{w}$ in $S$ such that

$$\hat{w} = \arg \min_{w \in S} C(w)$$

(4.17)

Hence, the desired direction is a vector in $R^p$ such that $\hat{k}$th element is $-\text{sign}(\nabla w)_k$ and the other elements are zeros, where

$$\hat{k} = \arg \min_{k=1,...,p} \{\nabla (w)_k, -\nabla (w)_k\}$$

(4.18)

The whole algorithm hence could be write in the following steps:

1) Initialize: $w = 0$ and $m = 0$.

2) Do until converge
   (a) $m = m + 1$.
   (b) Compute the gradient $\nabla (w)$
   (c) Find the $(\hat{k}, \hat{\gamma})$ that minimizes $\gamma \nabla (w)_k$ for $k = 1, ..., p$ and $\gamma = \pm 1$.
   (d) Let $v$ be the $p$ dimensional vector such that the $\hat{k}$-th element is $\hat{y}$ and the other column elements are zeros.
   (e) Find $\hat{\alpha} = \arg \min_{\alpha \in [0,1]} C(\alpha, v)$.
   (f) Update $w$: where $w_k = (1 - \hat{\alpha})w_k + \hat{y}\hat{\alpha}$, $k = \hat{k}$ or $(1 - \alpha)w_k, k \neq \hat{k}$

3) 3. Return $w$. 
Sparse Neural Activity Detection

Using the multi-layer neural activity signal model, given in Eq. (4.2), we can calculate the sparse signal sources. In this way, detecting the neural activities from fMRI data becomes solving the following convex regularization problem:

\[
\min_x ||x||_1 \quad \text{subject to } S = Ax. \tag{4.19}
\]

As both neural activity signals and impulse responses are unknown, an expectation-maximization (EM) procedure is performed to construct multi-scale models from the fMRI time course data. In each E-step, the ternary neural activity signals are estimated using the given impulse response model. In each M-step, the impulse response model parameters are optimized to maximize the correlation between the predicted and measured fMRI data. Based on experimental results, a three-layer impulse response model from coarse to fine scale can predict the fMRI time course data precisely. Fig.4.3 shows impulse response models in three scales and the corresponding ternary neural activity signals obtained from the given fMRI time course signals.

4.4.4 Iteration

We’ve bringing an iteration for reduce the ICA error. Before each loop, we update the transformation matrix \( A \) with

\[
A^+ = E\{x g(A^T x)\} - E\{g'(A^T x)\} A \tag{4.20}
\]

\[
A = A^+ / ||A^+|| \tag{4.21}
\]
where $g$ is the derivative of the non-quadratic function. Hence our proposed procedure contains three steps:

1. Iteration Step. 2. E-step: Estimate the Sparse Source. 3. M-Step: Maximize the correlation between source and hemodynamic model.

4.4.5 Spatial Smoothing

To reduce the false alarms caused by random human motions and noise, we use a set spatial filters to smooth the neural activity signals. In each 3 pixels $\times$ 3 pixels region, the ternary neural activity signals can be partitioned into 5 sets of symmetric components, that is,

$$I(t; w) = w_1I_1(t) + ... + w_5I_5(t) = w^TI(t). \quad (4.22)$$
The derived 5-element vector $w$ can account for different spatial coherence conditions. The spatial matched filters corresponding to these conditions can be used to estimate the neural activity status of each voxel: active if on active regions, normal if beside active regions or just noise. The spatial smoothing can reduce false alarms.

### 4.4.6 Neural Activity Signal Reconstruction

The main advantage of the multi-layer signal model lies in its capability of capturing neural activity in multiple scales. The neural activity signals in large scales represent major events, which may be associated more with the brain functionality under test. On the other hand, the neural activity signals in small scales represent some neural activities of little importance or interest. Therefore, at the stage of signal reconstruction, only those major signal events are useful to certain brain functionality under study. Fig. 4.4 shows the neural activation map at three layers with different gray scale. Layer 1 result describes the major neural activities. Layer 2 and 3 results describe the small and trivial neural activities.
4.4.7 Summary

We have described in this paper a new framework for detecting neural activities in fMRI data through ICA with constrained $L_1$ regularization. By exploring the hierarchial structure and sparse nature of the neural activities, we develop a multi-layer neural activity signal model. We compressed our data using random projection to preserved the key characteristic while reduce the data processing amount. In order to reduce the motion artifacts and noise caused false detection, we use ICA to isolate the neural activity vector. Then an constrained $L1$ regularization was performed to detect of the neural activities by finding the sparse solution of a linear inverse problem. This model has a higher detection accuracy than the data-driven approach and takes less computational cost than the two model-driven methods.

4.5 Experimental Results and Discussion

We present our experimental results in the following manner: First, we verify our approach through a set of simulation data generated by ourself in which it contains noises, neural activities and motion artifacts. After verify the effectiveness of our approach, we test it using the real fMRI data collected from water drinking and glucose intaking. Then we compare our method with other popular approaches to verify the accuracy as well as the efficiency for the neural activity detection of our approach. Finally we discuss the strengthens and weakness of our approach, and our challenges and future work.
4.5.1 Simulation Results

Simulation Data Description: First, we verify our proposed framework using simulation data. In order to reduce the computational cost, we generate some small sets of pictures: Each set of data contains 100 frames. Each frame has the resolution of 64*64. One sample frame of motion free simulation data can be seen in Fig.4.5 (a).

The base simulation data were draws roughly by a human brain shape. The hypothalamus area of human brain is painted with separate dark gray color. We did so because in our real data, most of the neural activities in real data related to our performed tasks occur in this area. We add gaussian white noise in each frame of the original simulation data and we add artificial distortions which cause both pixel level displacement, gray scale change and blurring. They can be seen at Fig.4.5 (b) and (c).

Neural activities within hypothalamus part were added in temporal domain to simulate the real data. The neural activities was generated using the impulse response model as the difference between two gamma functions stated in the previous chapter where $\delta_1, \tau_1, c$ will determine the IR model shape and $\delta_2, \text{tau}_2$ are set to have linear relationship to $\delta_1, \tau_1$. Fig.4.6 we show the different type of parameters
generates different impulse responses.

We’ve divided each frame of the simulation data into a set of 3 pixels × 3 pixels small blocks which overlap each other with 1 pixels for independent component analysis. After reduce the data dimensionality through random projection from 3 × 3 to 3 × 1, we apply ICA to separate the signal into three major components: neural activities, motion artifacts and noises. Fig.4.7 shows the decomposition result in one block.

Based on the previous statement in chapter III, the detecting of the neural activity problem could be able to converted into a constrained $L_1$ regularization problem with the help of the neural hemodynamic response model. By using the ICA generated neural activity source as the constraint for our $L_1$ regularization. The sparse neural activity detection could further verify and improve the ICA detecting results.

Fig.4.8 indicates the constrained $L_1$ regularization result for the simulation data in the same interested region. As the multi-layer neural hemodynamic response was established, our method not only captured the major neural activities associated with certain task (Level 1 in Fig.4.8) but also captured the minor activities (Level
Figure 4.7: ICA decomposing the compressed signal into neural activities, motion artifacts and noises.

2 in Fig.4.8). With these additional results, we could provide more possible information for further investigation.

Fig.4.9 shows the difference between original neural activities we generated for the simulation (RAW fMRI signal minus the motion artifacts and noises.) with the pure ICA technique and the constrained $L_1$ regularization based ICA. We could see by applying the $L_1$ regularization method. We could be able to captured more accurately where the neural activities exactly located.

Fig.4.10 shows the iteration of the ICA transformational matrix. With different set of converging force we chose. We could see the the residue decrease from the iteration. By minimizing the residue we further improve the accuracy of our detecting results.
Figure 4.8: Sparse neural activities detection using ICA with constrained $L_1$ regularization.

Figure 4.9: Residue between ICA detected signal and the original signal.
4.5.2 Real Data Results

Data Description

Our real fMRI data are collected by performing two tasks: water drinking and glucose intaking. Each task contains six sets of fMRI data with each set contains 183 images. Each image in the dataset has the resolution of 255*256. Water drinking activities were take vary from 54” to 1’40”. Glucose intake activities were taken var from 50” to 1’28”. This time difference will cause degree of the reaction of the neural activity differ from each other, yet because they are performing the same task. Their neural activity response will be similar.

Sparse Neural Activity Detection

Fig.4.16 shows the activation map and Fig.4.17 show the deactivation map of the multi-layer neural activities associated with water drinking task for the tested fMRI data. The upper row shows the pure $L_1$ regularization results and bottom row shows
the ICA based $L_1$ regularization results. As we use ICA separated neural activities as a priori and the L1 regularization results as feedback for iteration. We successfully eliminate some false alarms within the hypothalamus area.

Fig.4.11, Fig.4.12, and Fig.4.13 show the activation pixel numbers within each frame associated with glucose intaking neural activities over the temporal domain. Fig.4.11 shows the pure $L_1$ regularization method. Fig.4.12 shows the ICA with $L_1$ regularization, as ICA separates neural activities with noise and motion artifacts. Activation detection over the temporal domain could be detected with higher accuracy. (lower number showed on the activation pixel numbers indicates pure $L_1$ regularization have false alarms due to motion artifacts. Fig.4.13 shows the ICA with $L_1$ regularization with weight vector iteratively updated, this will leads to more accurately separate the interested signals with motion artifacts. Our results demonstrate that filtering out more false alarms over the temporal domain is quite important as most of the activation pixels with occur in more than more frame of fMRI image.

Fig.4.14 and Fig.4.15 shows the neural activity patterns for the 5 group of data. One of the innovation of our approach is it can be able to detect the common neural activity patterns related to specific function in brain. This is the thing ICA and other statistical approach can not do.

**Compare with Other Popular Approaches**

We use three other popular approaches as a reference to compare with our proposed method: Temporal Clustering Analysis (TCA), Maximum Correlation Method (MCM) and Graphical Model Inference (GMI).

Fig.4.16 shows neural activity detection results by the four approaches. As shown in Fig. 4.16 (a), TCA yields a false detection region (the white active region)
Figure 4.11: Multi-layer neural activity detection through constrained $L_1$ regularization.

Figure 4.12: Multi-layer neural activity detection through constrained $L_1$ regularization.
Figure 4.13: Multi-layer neural activity detection through constrained $L_1$ regularization.

Figure 4.14: Glucose intaking neural activity patterns.
although it consumes the least computational resource. MCM and GMI only detect the first layer of neural activities and they can not be extended to the multi-layer model. The proposed ICA with $L_1$ regularization based approach can detect neural activities of different layers with low computational cost. Using such a multi-scale scheme, the neural activities with all scales can be detected. In order to capture only the major signal events, the user can select the number of layers at the stage of signal reconstruction according to the task.

Fig.4.18 shows the computational time consumption for four methods. It can be seen that the proposed ICA with $L_1$ regularization approach saves about 40% of the time by using the MCM method. Both of them requires the procedure of training hemodynamic impulse response models. Although TCA has the shortest overall activation map detecting time. As no model was involved, the detection results is relatively poor and often leads to false alarm.
Figure 4.16: Activation map detection for the four approaches.
Figure 4.17: Multi-layer neural activity detection through constrained $L_1$ regularization.

Figure 4.18: Multi-layer neural activity detection through constrained $L_1$ regularization.
4.6 Conclusion

We have described in this paper a new framework for detecting neural activities in fMRI data using ICA with constrained $L_1$ regularization. By exploring the hierarchical structure and sparse nature of the neural activities, we develop a multi-layer neural activity signal model. With such a model, detection of the neural activities can be converted into finding the sparse solution of a linear inverse problem. We compressed the data using random projection and using ICA to separate the neural activity signals with noise and motion artifacts for $L_1$ regularization. The results of our approach were compared with other popular methods include TCA, MCM and GMI in terms of detection accuracy and computational cost. The experimental results showed that the proposed approach could detect the neural activities in multiple scales within a relatively short period of time. It has a higher detection accuracy than the data-driven approach and takes less computational cost than the two model-driven methods.
Chapter 5

Graphical Model Based Inference

5.1 Introduction

Data mining techniques used in fMRI studies can be characterized into two major schemes: model-based and data-driven. Model-based methods, such as general linear models (GLM) and graphical models, demand repeatable and stable patterns in brain activities. Those methods can provide insights into how a particular cognitive process is implemented in a specific brain area instead of merely identifying where such a process is located [28] [29] [30]. However, they cannot be used when the signal responses are not known as \textit{a priori}. Data-driven methods like temporal clustering analysis (TCA), independent component analysis (ICA) and principle component analysis (PCA), have more flexible frameworks for data analysis in the absence of \textit{a priori} model of brain activities, but not fully exploiting the spatial, temporal correlations between adjacent data.

Despite having many achievements, the existing fMRI data mining approaches have to address the following technical challenges:

1) artifacts caused by random subject movements under various experiment conditions;

2) insufficient utilization of temporal, spatial, and inter-subject correlations among
neural activities, typically associated with data-driven approaches; and

3) increased computational complexity and cost, typically associated with model-based approaches.

To tackle the above challenges, it requires innovation and development of techniques for:

1) effective image registration and segmentation to reduce artifacts and improve the SNR,

2) efficient representation of neural activity dynamics in the fMRI signal space, and

3) effective statistical inference mechanisms.

We present a hybrid approach to neural activity detection from fMRI data, as shown in Fig. 5.1. Under this framework, the data-driven approach predicts the
time windows in which events of interests will likely be detected; the model-based approach compresses the high volume data, and localize the information of interest efficiently. Through a proper combination of two approaches, the computational cost is reduced, and the efficiency and robustness of the neural activity detection process are improved. The whole procedure includes three stages: (1) fMRI images are enhanced through registration and segmentation. Neural activities are predicted by statistic data analysis, and signal events are captured through matched filtering; (2) signal events are represented by the linear predicative codes (LPCs) and classified into event prototypes through an variational bayesian approach; (3) graphical models are built from labeled event sequences, and are used to classify and detect neural activities from newly acquired data.

5.2 Methodology

The whole framework consists of three components: 1) image enhancement, event prediction and capture; 2) event feature extraction and modeling; and 3) graphical model based Bayesian inference.

5.2.1 Image Enhancement, Event Prediction and Detection

Image Registration and Segmentation

A 2D deformable image registration [82] and piecewise constant model for level set version of mumford-shah segmentation [34] are used to reduce the motion artifacts in fMRI data and isolate the gray matter, where neural activities are present.
Event Prediction

As a data-driven technique, TCA assumes that the number of pixels with extreme values will be greater during the activation periods than during the rest period. We use TCA to predict the time windows in which extraordinary neural activities arise. When the TCA signals are above a certain threshold, it can be regarded as a neural activation (deactivation) period. Fig. 5.2 (a) shows the TCA signals of one subject drinking water and Fig. 5.2 (b) shows the zoomed-in signals to illustrate the procedure of time window generation. Using the time windows generated by the data-driven approach, event signals will be detected and modeled.
Event Detection

Fig. 5.3 shows the procedure of event prototype generation. First the temporal sequence of the segmented signal by time windows is normalized by subtracting the mean of the signal. Then the sequenced signal is convoluted with a matched filter (band-pass filter) to reduce the SNR. A threshold manipulation is applied after the convolution to remove the low-information background activities. Then discrete events are isolated from the data sequence. Finally, signal event data sets are categorized into event prototypes for further logic processing.
5.2.2 Event Feature Extraction and Modeling

Feature extraction

When subjects perform similar activities such as water drinking and glucose intaking, their neural activities should exhibit similar patterns regardless of their anatomical differences. However, due to the inter-subject variabilities and artifacts caused by experiment conditions, the signal patterns may not be exactly the same. For example, some subjects drink water fast, their response signals could have a high-spike in a short period of time. When subjects drink water slowly, the response data might be very flat with little spike. In order to deal with such a disparity and only get the pattern information, we utilize linear predictive coding (LPC) which transmits spectral envelope information with tolerance of transmission errors. Fig. 5.4 shows activities performed by different subjects, resulted in transient event signals which can be represented in LPC coefficients. It can be seen that LPC can reserve the signals’ structural similarity and disparity.

Events Classification

After the event prototypes were generated. The event features could be classified using variational Bayesian Gaussian mixture model (VBGMM). Fig. 5.5 shows the event prototype clustering results. It can be seen that the event features are successfully clustered into four different clusters.

Graphical Model Based Inference

Probabilistic graphical models are graphs in which nodes represent random variables, and the arcs represent conditional independence assumptions. Hence they provide a compact representation of joint probability distributions. With the clas-
Figure 5.4: Compressive linear predicative coding of signal events.

Figure 5.5: Clustering of event prototypes using the VBGMM approach.
sified compressive transient events, our factor graph is showed in Fig. 5.6. Its hidden states include (1) activation, (2) deactivation, and (3) normality. Each hidden state has various emitted observations of event prototype.

5.3 Experimental Results

We tested our methodology with 6 subjects of fMRI data generated by water drinking activity. We used the data of five subjects to train the graphical model using Bayesian Network Toolbox, and used fMRI signals of the remaining subject as testing data. Fig. 5.8 shows the neural activity detection results in the area of hypothalamus using the proposed hybrid approach. Fig. 5.8 (a) illustrates a set of average fMRI signals in hypothalamus that is associated with the drinking activities, (b) shows the generation of time windows using TCA, (c) shows the event prototype sequences, and (d) gives the neural activity detection results. We also estimated the likelihood of each subject’s data fitting to event sequential models,
Figure 5.7: detection of neural activities. (a) fMRI temporal signals. (b) time windows generated by temporal cluster analysis. (c) labeled event sequence (Red, Blue, Green, and Pink: cluster 1, 2, 3, and 4). (d) detected neural activity status (Dark: activation; Gray: normality; White: deactivation).

Figure 5.8: neural activity detection performance comparison between TCA (upper) and the proposed Hybrid approach (lower) in the area of hypothalamus.
yielding little variance. Fig.10 shows the comparison between TCA and the proposed method detecting neural activity in the area of hypothalamus for 6 subjects. It can be seen that TCA yields a false alarm for subject 4 and the proposed approach shows robustness against the inter-subject variabilities. Another advantage of the hybrid approach is the reduced computational cost. As the fMRI data is of very high volume, this approach is suitable to analyze the brain activities in the long-time experiments such as detecting neural activities in virtual reality.

5.4 Conclusion

We have described in this section a new framework for detecting neural activities in fMRI data. By integrating data-driven and model-based approaches and using signal compression techniques, the high data volume, low SNR fMRI images can be converted to discrete event sequences. Based on a graphical-model representation, the neural activities can be detected and classified at reduced computation cost with robustness against the inter-subject variability. Future work will focus on finding better event prototype representations and improvement of the event clustering robustness in temporal domain. Using LPC to further compress the signal in hidden markov tree model. Spatial correlations among event sequences will also be further studied to extend the proposed frameworks capability of dealing with applications associated with complex neural activities (e.g. brain study in virtual reality). More effective structural learning methods will be developed for better training of graphical models.
Chapter 6

Summary

This dissertation introduced the background and state-of-the-art of functional magnetic resonance imaging (fMRI) technology and highlighted the significance of the research in brain neural activity identification/detection using functional magnetic resonance images with its applications in neuroscience and physiology. After reviewing the current state-of-the-art fMRI data processing and analysis techniques and analyzing their advantages and disadvantages, it can be found that the major challenges in fMRI study include: (1) Inter-subject variabilities. When different human subjects perform same cognitive tasks or one subject at different times, there will be variations in neural activation patterns in the spatial and temporal domains. (2) Motion artifacts. Brain fMRI motion artifacts could be caused by involuntary muscle movements like swallowing and sometime the patient’s head movements, resulting in image blurring, reducing the signal-to-noise ratio (SNR) and generating false alarms. (3) Low SNR and high data volume. Neural activity in brain fMRI data could as low as only 2% higher rise from the background noise. A typical set of 2-D fMRI measurement contains at least 1M bytes data, while the 3-D measurement generates hundreds of mega bytes data.

A complete set of successful neural activity identification/detection methods should be able to eliminate the inter-subject variabilities over signal patterns, re-
move the motion artifacts, increase SNR and reduce the data volume to be processed. It will require efficient representation of signal patterns, effective image structural model, accurate neural impulse response model and corresponding filtering, estimation, and learning/optimization algorithms. Most neural activity detection approaches, either deterministic or statistical, contain three stages: (1) Preprocessing: using techniques like image registration and normalization to align the fMRI data into one coordinate system within minimal difference and reduce the effect of the motion artifacts and noises. (2) Signal modeling: using statistical method to capture signal events from the repeatable patterns or establish neural activity impulse response models in multiple scales. (3) Decision making: using signal estimation or logic inference to determine the status of neural activities associated with a certain brain function.

This dissertation proposed two methods for motion artifacts removal during the preprocessing stage: (1) wavelet/contourlet hidden Markov tree based image restoration and (2) L1 regularized independent component analysis. The former can restore the statistical properties of the fMRI images afflicted with motion artifacts; the latter can decompose time-course data into four components: low frequency neural activity signal, high frequency neural activity signal, motion artifacts and measurement noise based on their independent statistics. In order to improve the SNR of data as well as reduce the data volume, this dissertation investigated two methods during the signal modeling stage: (1) linear predictive coding based signal events and (2) multi-layer, multi-scale neural hemodynamic response model. The former can compress and classify time-varying data into some logic event labels; the latter can decompose the neural activity signals into multiple bands while describing the sparsity of the neural activity sources. For the decision making stage, this dissertation developed two schemes (1) graphical model based inference and
expectation-maximization-like $L_1$ norm optimization. The former can take into account the logic consistence between discrete signal events and reject inconsistent events; the latter can find the common sparse neural activity source signals of a group of subjects through the up-to-date $L_1$ optimization techniques.

### 6.1 Contributions of This Dissertation

Chapter 2 described the experiment setup and fMRI data collection protocol. There are two groups of subjects under study, intaking water and glucose respectively. It also investigated the up-to-date techniques having been used to identify/detect neural activities and discussed their strengths and weaknesses in the context of fMRI data analysis.

Chapter 3 presented a hidden markov tree (HMT) model based fMRI image restoration technique. Conventional preprocessing techniques like image registration transform the different sets of fMRI data into one coordinate system through key points matching without considering fMRI data statistical characteristics in the spectral domain. Hence they cannot remove some specific image distortions such as spikes and blurring caused by motion artifacts. The hidden Markov tree (HMT) model can preserve the coefficients’ properties of clustering and persistency in the spectral domain. Combined with a non-rigid image registration algorithm, the proposed method restored the fMRI data’s statistical distributions in both the spatial and spectral domains, respectively. Under this scheme, the wavelet (contourlet) coefficients of the distorted image were filtered using the HMT model of the baseline image to minimize the statistical divergence between two images. An iterative loop between non-rigid image registration and HMT filtering was developed to achieve a tradeoff between the least mean square error in the spatial domain and the minimum
statistical divergence in the spectral domain. In order to reduce the computational cost, the images were processed in blocks. We compared our results with other popular image restoration approaches and showed that our method could eliminate the motion artifacts in the fMRI data more effectively and lead to more accurate neural activity detection. This method can also be used for image restoration in other medical imaging applications.

Chapter 4 presented a neural hemodynamic response model and use ICA with constrained $L_1$ regularization for data analysis. As neural activity detection with functional magnetic resonance images involves exploring large amount of data for finding sparse signal sources, Independent component analysis (ICA) methods have been developed to separate neural activities of interest from noise and motion artifacts without using any neural dynamics knowledge. In this chapter, a novel method was presented for detection of brain neural activities through $L_1$ regularized ICA. A multi-layer, multi-scale signal model was developed that can map sparse neural activity sources to fMRI measurements based on the neural hemodynamic response (NHR) model to enhance conventional ICA approaches. The parameters of the multi-scale NHR models and sparse signal sources are estimated by an expectation-maximization (EM) like procedure through $L_1$ regularized optimization. As a result, false alarms generated by conventional ICA approaches were reduced; the common neural activities of a group of subjects with similar experimental conditions were generated. Both simulation and experimental results demonstrated the computational efficiency and detection accuracy of the proposed approach.

Chapter 5 presented a framework for neural activity detection based on both statistical data analysis (data-driven) and graphical model inference (model-based). The data-driven approaches roughly predicted when an extraordinary amount of neural activities arise. By proper exploration of spatial, temporal, inter-subject
correlations, the model-based approaches provided more insights and details, and physiological meaning from high data volume, low signal-to-noise ratio (SNR) fMRI measurements. Through temporal cluster analysis (TCA), matched filtering, linear predictive coding (LPC), and variational Bayesian Gaussian mixture modeling (VBGMM), the temporal fMRI signals were converted into event prototypes associated with three neural statuses: activation, deactivation, and normality. As a result, the high volume fMRI data generated from multiple subjects were statistically modeled as coupled finite-state sequences. Based on the graphical-model representation, the neural activities captured through fMRI were classified and detected at reduced computational cost. The whole framework consists of three components: 1) image enhancement, event prediction and capture; 2) event feature extraction and modeling; and 3) graphical model based Bayesian inference. The experiment results demonstrated the advantages of the proposed hybrid, compressive signal processing approach in terms of computational cost and robustness against inter-subject variability as well as various artifacts.

6.2 Future Work

Neural activity classification/detection is still a new research area with various applications in the study of neuroscience and brain physiology. The presented work is just a tiny step over a wide terrain where several directions are worth further investigation.

3-D fMRI Measurement Based Neural Activity Detection. MRI is an intrinsic 3-D imaging technology, which can provide 3-D reconstruction data of the brain image. From the signal processing perspective, 3-D data can provide as many advantages and challenges. For example, motion artifacts caused by movements of a
3-D object can be easily removed using 3D rigid and deformable image registration techniques but will impose many difficulties for 2D image restoration. 3-D data also allows building more accurate spatial correlations among neural activities which will be helpful for understanding the collaborative functionality of brain regions. However, 3-D MRI measurements also generate a huge amount of data. For example, a typical set of 3-D MRI measurements could contain hundreds of Mb data. It requires more efficient data compression and compressive data processing techniques. Recent advances in compressive sensing can reduce the data volume of 3-D MRI and the corresponding image processing, signal modeling and data inference techniques should also be developed.

**Real-time Neural Activity Detection.** With the development of both hardware and software, modern MRI technology can generate real-time measurements on human subjects. The real-time MRI technology can provide more opportunities for study the anatomy and functionality of human brains. Besides, it can be used to develop a new human-machine interface. For example, in the virtual-reality study, the real-time MRI neural activity detection allows the development of the correlations between certain computer simulation scenarios and neural activity patterns. Based on those correlations, many human emotions and feelings associated with certain neural activity patterns can be trigger by designed computer simulations. On the other hand, the computer can interpret the intentions a subject through identification of his/her neural activity patterns and perform certain actions accordingly. However, real-time neural activity identification/detection demand the development of multi-layer, multi-scale neural hemodynamic model with more biological and physical parameters to capture accurate sparse neural activity signal source. Besides, efficient and fast data learning techniques for sparse signal sources to associate signal patterns with certain brain functionality should also be developed.
accordingly.

**Integrated Data Restoration and Neural Activity Detection.** Under the current fMRI data analysis frameworks, there are three separate processing stages: preprocessing, signal modeling, and decision making. In our study on statistical restoration of fMRI images, we have developed a method that integrates signal modeling and preprocessing. A hidden Markov tree model is used to represent the statistical properties of images in the spectral domain. The image restoration is performed based on the probabilistic model of image data. However, for neural activity detection, the gray-level images have to be reconstructed from the restored spectral coefficients. In fact, if the neural activity detection can be performed in the spectral domain, the computational cost on image reconstruction can be reduced. Therefore, the development of techniques that can detect neural activity in the spectral domain will be an interesting research direction. As a result, an integrated data restoration and neural activity detection can be developed based on hidden Markov tree model that can preserve the structural information of either 2-D or 3-D fMRI measurements.
Bibliography


125


