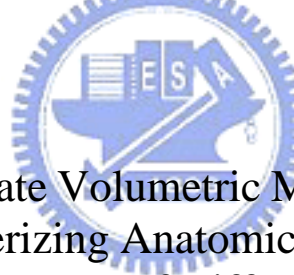


國立交通大學

資訊科學與工程研究所

碩士論文

多變量體積型態法用於磁共振影像
群組間結構差異性之定量分析



Multivariate Volumetric Morphometry
for Characterizing Anatomical Discrepancy
in MR Images of Different Groups

研究生：楊承嘉

指導教授：陳永昇 博士

中華民國九十五年七月



多變量體積型態法用於磁共振造影影像
群組間結構差異性之定量分析

Multivariate Volumetric Morphometry for Characterizing
Anatomical Discrepancy in MR Images of Different Groups

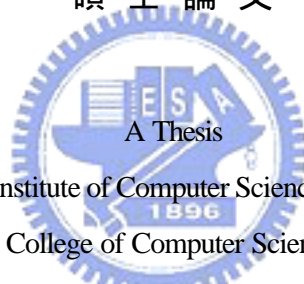
研究生：楊承嘉

Student : Cheng-Chia Yang

指導教授：陳永昇

Advisor : Yong-Sheng Chen

國立交通大學
資訊科學與工程研究所
碩士論文



Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年七月



摘要

以體素為基礎的型態計量學 (voxel-based morphometry, VBM)，近年來已被廣泛應用在許多腦部結構研究上，以統計的方式量化、比較兩組人腦結構每一體素之型態是否具有顯著的差異。體素型態學使用單變異量統計，對於較為集中且巨大的差異具有簡單有效的優點。但是由於此方法沒有考慮相鄰體素間的連帶關係，因此可能無法找出差異量較為分散且微小的腦部區域。

本研究中我們提出了一個新的腦部結構評估技術—多變量體積形態法 (multivariate volumetric morphometry, MVM)，以運用於磁振造影影像群組間結構差異性之定量分析。相對於以體素為基礎的型態計量學，多變量體積形態法採用線性鑑別度分析 (linear discriminant analysis, LDA) 作為多變量的估計方法來取代單變量的分析。此方法能同時考慮全部的體素，以找出對於群組間結構差異性最具鑑別力的投影軸。位於該投影軸上之每一元素即代表相對應體素具有的鑑別力比重 (discrimination weight)，且此鑑別力比重可被視為用來評估磁振造影影像群組間結構差異之程度等級 (significance level)。此種多變量的計量方法能突顯群組間較為細微的結構性差異，很適合用於分析腦部結構差異的研究。此外，我們也證明了不論使用原始的磁振造影影像或是平滑化過後的影像，此方法所能達到的最大區辨能力是一樣的，因此可以直接在原始影像而非在人工處理平滑化後的影像上推論分析結果。相反地，以體素為基礎的型態計量學卻必須使用平滑化的處理來增加分析的正確性，同時加強相鄰體素間的連帶關係。然而使用平滑化處理時，要決定其恰當的影響範圍是很困難的一個問題，因為較大範圍的平滑化處理雖然可降低影像雜訊，但卻必須付出細部資訊被模糊的代價。

透過模擬小腦周圍區域萎縮的實驗，我們驗證了多變量體積形態法的有效性與正確性。比起以體素為基礎的型態計量學，該方法確實更有能力可偵測到群組間細微的腦部結構差異之處。應用在脊髓小腦運動失調症 (spinocerebellar ataxia, SCA) 的結構差異分析上，多變量體積形態法也比以體素為基礎的型態計量學更明顯地找出和病理相關的腦部結構組織。



誌 謝

感謝很多很多的人，有大家的互相扶持與陪伴，我很快樂也很幸福。碩士兩年，最感謝就是兩位指導老師—陳永昇老師和陳麗芬老師。不僅給予我學業及研究上很大的幫助，同時也關心我的生活，鼓勵我繼續向前，真的很感謝。實驗室的所有伙伴們，謝謝你們讓我的生活多采多姿，特別是我們這一屆的大家。每次的實驗室出遊，都載著滿滿的歡笑與回憶，我們的 Pingu 版也同樣精彩。最後，請大家在發揚 BSP 的精神時，繼續維持良好的實驗室傳統—學姊唯大。

也感謝一直支持我、給我溫暖、為我加油打氣的家人，和已經很久很久的劉大博士元平，謝謝你們總是讓我又充滿力量。還有一起來交大讀書的同學和高中死黨們，你們讓我在這裡一樣有家的感覺。舊的情感堆積上了新的，歷久彌堅。

感謝台北榮總神經內科的宋秉文醫師和王伯山醫師，提供我研究上的協助。再次感謝所有我所認識的大家，謝謝你們。





Multivariate Volumetric Morphometry for Characterizing Anatomical Discrepancy in MR Images of Different Groups

A thesis presented

by

Cheng-Chia Yang

to

Institute of Computer Science and Engineering
College of Computer Science

in partial fulfillment of the requirements

for the degree of

Master

in the subject of

Computer Science

National Chiao Tung University

Hsinchu, Taiwan

2006

**Multivariate Volumetric Morphometry for Characterizing Anatomical Discrepancy
in MR Images of Different Groups**

Copyright © 2006

by

Cheng-Chia Yang



Abstract

Recently, voxel-based morphometry (VBM) has been widely applied to statistically infer the structural anomalies between the brains of two subject groups, in a voxel-by-voxel manner. This method is effective for mapping massive and centralized discrepancy. However, it may suffer from the poor sensitivity to subtle and widely-distributed discrepancy in brain structures.

In this work, we propose a novel multivariate morphometry (MVM) method that can be used to delineate the anatomical discrepancy between two groups of MR images. Rather than voxel-by-voxel manner in VBM, the proposed MVM simultaneously considers all of the voxels in MR volumes and map the group differences by using the linear discriminant analysis to determine the most discriminant projection vector. Each element in the projection vector represents the discrimination weight of the corresponding voxel involved in the combination of the most discriminant components. This weight can thus be regarded as the significance level of the corresponding voxel when differentiating two groups of MR volumes. This multivariate approach is appropriate to characterize group discrepancy, particularly when the brain atrophy distributes widely. Moreover, we prove that the discriminability remains the same no matter the projection vector is calculated from the original MR volumes or from the smoothed ones. Hence we can simply use the original data without the interference of the blurring artifact caused by the smoothing operation. On the contrary, VBM method applies the Gaussian smoothing filter to reduce image noise as well as to incorporate spatial support from neighboring voxels. It is difficult to determine an appropriate kernel size for the smoothing filter because larger kernel can reduce more noise, but with the penalty of more smeared image.

According to our experiments, we demonstrate the effectiveness of the proposed method by using the simulation data set containing artificial atrophy around the cerebellum area. Compared to the VBM method, the proposed MVM method can achieve a better sensitiv-

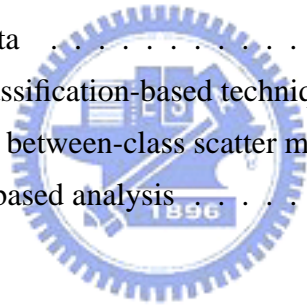
ity to subtle and widely-distributed variation of brain structure. When applied to a clinical study of SCA3 disease, the MVM method clearly reveals more significant atrophy in the disease-related areas within the brain volumes of the patient group, than the VBM method does.



Contents

List of Figures	v
List of Tables	vii
1 Introduction	1
1.1 Brain Structures	2
1.2 Magnetic Resonance Imaging	5
1.3 Morphometrics	8
1.4 Motivation	9
2 Voxel-Based Morphometry	11
2.1 Introduction to VBM	12
2.2 Optimized VBM Protocol	15
2.3 Drawbacks of VBM	22
3 Multivariate Volumetric Morphometry	25
3.1 Ideas of the Proposed Method	26
3.2 Framework of Multivariate Volumetric Morphometry	28
3.3 Multivariate Analysis using a Reformatory LDA-Based Method	32
3.3.1 Conventional Linear Discriminant Analysis and Its Potential Problem	32
3.3.2 Discriminative Common Vector Method	33
3.3.3 Efficient Implementation for Computing Discriminative Common Vectors	35
3.4 Multiresolution Analysis using Wavelet Transform	38

3.4.1	3-D Discrete Wavelet Transform	38
3.4.2	Wavelet Thresholding	42
4	Experiments	45
4.1	Capability Assessment for Discrepancy Revelation	46
4.1.1	Materials	46
4.1.2	Accuracy Evaluation	49
4.1.3	Comparisons between MVM and VBM	52
4.2	Structural Atrophy Analysis for Patients Suffering Spinocerebellar Ataxia Type 3	64
4.2.1	Materials	64
4.2.2	Results and Comparison between MVM and VBM	65
5	Discussion	83
5.1	Why we do not smooth data	84
5.2	Comparison with other classification-based techniques	89
5.3	Weighted within-class and between-class scatter matrices	92
5.4	Multivariate deformation-based analysis	93
6	Conclusions	95
	Bibliography	99



List of Figures

1.1	Main structures of the human brain	3
1.2	Brodmann's maps	4
1.3	A typical MR scanner	6
1.4	A 3-D magnetic resonance image of a human head	7
2.1	The normalization	13
2.2	The segmentation	14
2.3	Flowchart of basic VBM steps	16
2.4	Preprocessing of optimized VBM protocol	18
2.5	Concept of a MR image lying in a high-dimensional space	22
2.6	Schematic illustration of the significant bias of VBM	23
3.1	How a high-dimensional classification technique can be used to measure group differences	27
3.2	Flowchart of multivariate analysis and visualization in MVM	31
3.3	Analysis by 3-D discrete wavelet transform with filter banks	41
3.4	A MR image by discrete wavelet transform with different levels	42
4.1	The distribution of control points for TPS simulation	48
4.2	The source and simulated images by TPS	50
4.3	Example of a ROC curve	53
4.4	ROC curves of MVM and VBM results with the simulation data	54
4.5	Analysis result of MVM for 1mm simulated cerebellum atrophy	57
4.6	Analysis result of VBM for 1mm simulated cerebellum atrophy	58

4.7	Analysis result of MVM for 2mm simulated cerebellum atrophy	60
4.8	Analysis result of VBM for 2mm simulated cerebellum atrophy	61
4.9	Analysis result of MVM for 6mm simulated cerebellum atrophy	62
4.10	Analysis result of VBM for 6mm simulated cerebellum atrophy	63
4.11	Strategy for choosing a compatible threshold of MVM by VBM t values . .	67
4.12	Volumetric atrophy of gray matter in SCA3 patients by MVM analysis method	68
4.13	Volumetric atrophy of gray matter in SCA3 patients by VBM analysis method	71
4.14	3-D rendering of GM atrophy in SCA3 by the MVM analysis method . . .	73
4.15	3-D rendering of GM atrophy in SCA3 by the VBM analysis method	75
4.16	Volumetric atrophy of white matter in SCA3 patients by MVM analysis method	76
4.17	Volumetric atrophy of white matter in SCA3 patients by VBM analysis method	77
4.18	Volumetric enlargement of CSF in SCA3 patients by MVM analysis method	80
4.19	Volumetric enlargement of CSF in SCA3 patients by VBM analysis method	81
5.1	ROC curves of MVM results with the non-smoothed and smoothed simu- lation data	88
5.2	How SVM determines the separating hypersurface	90
5.3	Vectors used to obtain a single map for nonlinear classification	92

List of Tables

4.1	Definitions of TP, FP, TN, and FN	51
4.2	PAUC indices for ROC curves of MVM and VBM results with the simulation data	55
4.3	Clinical data of patients carrying SCA3	65
4.4	Atrophy of gray matter in SCA3 patients by MVM analysis method	69
4.5	Atrophy of gray matter in SCA3 patients by VBM analysis method	72
4.6	Detected GM atrophy of MVM and VBM in SCA3 patients	74
4.7	Atrophy of white matter in SCA3 patients by MVM analysis method	75
4.8	Atrophy of white matter in SCA3 patients by VBM analysis method	78
4.9	Detected WM atrophy of MVM and VBM in SCA3 patients	79
5.1	PAUC indices for ROC curves of MVM results with the non-smoothed and smoothed simulation data	89



Chapter 1

Introduction



In this chapter, we will briefly introduce the human brain structures, the magnetic resonance imaging (which is an imaging tool often used to detect pathologic tissues from normal tissues), and then the current morphometric methods based on medical images to analyze differences of brain structures. One of the most popular morphometric approach is the voxel-based morphometry, which have been applied in many researches of brain structures, but it has an inherent defect while detecting subtle and distributed changes. Our goal is to overcome this drawback and to propose a better morphometric method in this work. In the final of the chapter, we will guide the organization of this thesis.

1.1 Brain Structures

Brain is the most sophisticated and elegant organ of human beings. It plays an important role in the control of human mind and behavior. Several involuntary activities, such as heartbeat, respiration, and digestion, and conscious activities, such as thought, reasoning, and abstraction, are all operated by the brain. In the 3rd century B.C., Doctor Herophilus in Alexandria, the "Father of Anatomy", is considered as the first person to dissect human body for the purpose of scientific research. He obtained a lot of scientific discoveries, and one of his main contributions is to discover four rooms of the brain, that is, ventricles. Until now, people have done various researches on the brain and understand many the tissues and structures of the human brain.

A brain consists of three parts, which are the cerebrum, cerebellum, and brain stem. Brain stem is below the cerebrum and above the spinal cord, and is the major route to connect the cerebrum, cerebellum, and spinal cord. Its main function is to maintain individual life, including heartbeat, breath, digestion and the other important physiological faculties. Cerebellum is below the cerebrum and behind the brain stem, and composed of left and right two hemispheres. Cerebellum can balance the body and the posture, and also control the motion of muscle with the cerebral cortex. Cerebrum is considered as the most impor-

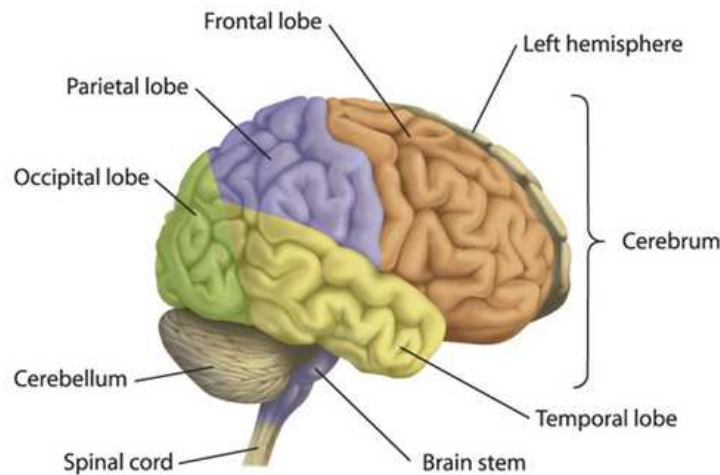


Figure 1.1: Main structures of the human brain. There are three parts, which are the cerebrum, cerebellum, and brain stem. According to sulci and gyri of cerebral hemispheres, brain cortex is divided into four lobes: the frontal lobe, the parietal lobe, the temporal lobe, and the occipital lobe. Photo courtesy of the website of Heart and Stroke Foundation (<http://ww2.heartandstroke.ca/>).

tant nerve center, and divides into left and right two cerebral hemispheres. Between two cerebral hemispheres is the corpus callosum to communicate left cerebral hemisphere and right cerebral hemisphere. Moreover, according to sulci and gyri of the exterior of cerebral hemispheres, brain cortex can be segmented into four lobes: frontal lobe, parietal lobe, temporal lobe, and occipital lobe. The frontal lobe is understood as the central administration of thought. The parietal lobe receives and handles kinds of feeling signals. The temporal lobe has relations with perception and recognition of auditory signals and memory. And the occipital lobe is the center of visual processing. Figure 1.1 shows main structures of the human brain.

According to the type of brain tissues, they can be generally separated into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Gray matter is composed of nerve cell bodies responsible for processing information, and white matter is composed of the axons responsible for transmitting information. Gray matter forms the exterior part of the brain, and is referred to as the cortex; white matter forms the interior part of the brain,

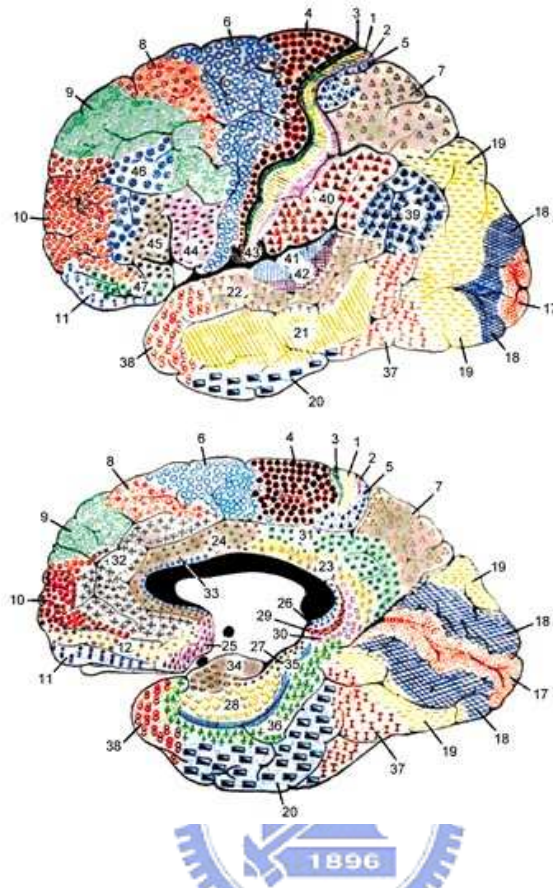


Figure 1.2: Brodmann's maps. The human brain is classified into 52 discrete cortical areas in a cytoarchitectonic way. The partitions are referred to as the Brodmann's areas.

and referred to as the medulla. Cerebrospinal fluid, which is the colorless and transparent fluid, fills ventricles and surrounds the brain and the spinal cord. Cerebrospinal fluid can absorb the shock to the brain or to the spinal cord, and also can drain out waste materials from the brain or from the spinal cord.

In 1909, Brodmann cytoarchitectonically classified brain into 52 discrete cortical areas using a light microscope, and sketched the anatomical maps of the human brain [1]. Each and every area is labeled with a number. These are known as the Brodmann's areas (BAs). Figure 1.2 is the famous Brodmann's maps. Many BAs were later shown that they are associated to specific functions, such as BA 17 and BA 18 in the occipital lobe (associated to vision). Brodmann's areas have become a common classification for scientists to refer to

a particular region of the brain cortex and the related nervous functions. In 1988, Talairach and Tournoux drew a 3-D stereotaxic atlas and defined the Talairach coordinates of the human brain, by anatomizing the brain of a European female aged 60 [2]. They used Brodmann's maps as the basis for the architectonic parcellation in their atlas. It is very useful for localization of brain tissues. Thus, when given a 3-D coordinate in Talairach space, we can indicate that which brain structure it is located at and which BA it belongs to, and then know broadly about its associated functions. The Talairach brain is usually taken as the standard stereotaxic space when investigating human brain structures.

Along with progress of science and technology, the first neuroimaging technique, the pneumoencephalography (PEG), was developed in the early 1900s. Invention of the brain imaging technology makes observing the human brain on living beings come true. By these medical images, scientists and doctors can investigate or make a diagnosis about those diseases resulting from some brain disorder as the patients are still alive, rather than dissect patients' bodies as they were died. Up to now, there are many functional brain imaging technologies, such as positron emission tomography (PET) and single photon emission computed tomography (SPECT); as well as structural imaging technologies, such as X-ray computer tomography (CT) and magnetic resonance imaging (MRI). In this thesis, we used magnetic resonance images as experimental materials to find the structural differences of different brains. In the next section, we will briefly introduce this technology, magnetic resonance imaging.

1.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is one of popular imaging tools for clinical diagnosis in recent years. It was developed by Paul Lauterber in 1972 [3]. The technique is based on the principles of nuclear magnetic resonance (NMR) to produce data images of internal physical and chemical characteristics of an object. The original name of this technique



Figure 1.3: A typical MR scanner. Photo courtesy of Lab. of Integrated Brain Research, Department of Research and Education, Taipei Veterans General Hospital.

was nuclear magnetic resonance imaging (NMRI), but people called it magnetic resonance imaging (MRI) because of the word nuclear with the negative connotations of radiation exposure in the late 1970's.

An MR scanner is shown in Figure 1.3. An scanner includes three main hardware devices: a main magnet, a magnetic field gradient system, and an RF system [4]. The main magnet generates a strong uniform magnetic field for polarization of nuclear spins in an object; the magnetic field gradient system produce controlled time-varying gradient fields in different directions to make each of the regions of spin experiences a unique magnetic field for signal localization; the RF (radio frequency) system generates a rotating magnetic field in a pulse sequence to excite spins and detects signals from the spins. All the components of the scanner are placed in a scan room to segregate outside interference. After analysis and reconstruction of signals by a computer, a magnetic resonance image representing the spatial distribution of the inside of living organisms is obtained like Figure 1.4.

MR imaging has many advantages. One is that it is a noninvasive way to detect signals inside the body, so people are unnecessary to bear with pain resulted from invaders of med-

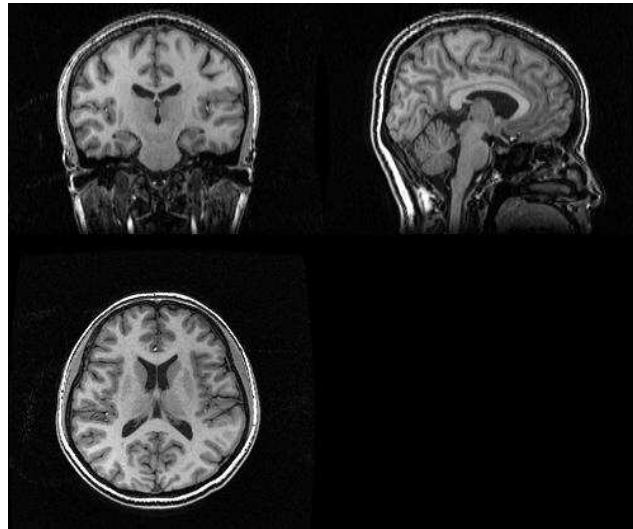


Figure 1.4: A 3-D magnetic resonance image of a human head. It is shown in the coronal, sagittal and axial views.

ical treatments. Moreover, this imaging uses magnetic fields and non-ionizing radiation. According to current knowledge, they do not have potential harmful effects to humans. In comparison with some other scanning methods like CT, it is very safe. Another advantage of MR imaging, probably the most important character, is the flexibility of data acquisition and the outstanding contrast resolution. Therefore, it can be used as spectroscopic imaging, diffusion-weight imaging, angiographic imaging, and functional imaging. That makes MR images able to provide much respectable information and endow the technique with superior scientific and diagnostic values [4].

Because of the clear contrast resolution, MR images are often used to observe pathologic tissues from normal tissues, and help doctors to diagnose medical conditions and disorders of the brain. However, such a manual diagnosis is very subjective and time-consuming, especially when the amount of images is large. Thanks to the advances in computer, computerized approaches can help to deal with the huge and complex data. Many morphometric analysis methods were proposed to quantitatively analyze MR images by computers.

1.3 Morphometrics

By using MR images, a number of *in vivo* anatomical studies of the human brain have been done. Most studies are based on the defined regions of interests (ROIs), and then analyze each tissue volumes [5–7] in ROI. However, this method has some limitations. It wastes a lot of time to define the ROIs, especially when there are large amount of subjects [8]. In addition, when analyzing a certain disease, users have to know the most concerned regions [9] before selecting ROIs. It makes ROI-based analysis inconvenient to be used in practice.

Therefore, another kind of automatic morphometric methods, involving the technique of spatial normalization, to characterize neuroanatomical differences has been developed. These methods broadly fall under two categories: (1) ones handle macroscopic differences in shape of brain, and (2) ones handle microscopic differences in brain tissue as the shape differences have been discounted. When connecting these with the technique of spatial normalization, the first kind of methods analyzes the parameters or the deformation fields used during the normalization; and the second kind of methods analyzes resulting normalized images after normalization.

The first family of morphometric method includes the pattern-theoretic approach [10], deformation-based method [11, 12], tensor-based method [13, 14], and factor analytic approach [15]. These methods quantify brain shape by using deformation fields obtained from nonlinear registration. This kind of methods can potentially obtain a precise estimation of the brain shape, but it is very sensitive to the accuracy of the underlying normalization approach. Consequently, there are some limitations in practice [9].

The second family of morphometric methods characterizes anatomy in brain tissue by estimating voxel intensities of normalized images. Because this type of methods makes use of images after normalization, the differences in the brain shape are eliminated. Thus, it is

suitable for analyzing local and subtle differences in brain tissue. A common-used method, voxel-based morphometry [16], is belong to this family. Besides, the RAVENS [9] is also a kind of these methods.

This thesis emphasizes the second family of morphometric methods. The targeted images are all normalized. Now, the voxel-based morphometry (VBM) is the most popular method applied to analysis of structural brain discrepancy between different groups of images. For each and every voxel from the normalized images, it makes a standard statistical test to examine if there exists a significant difference of brain structure on the location of this voxel. Although VBM is an intuitional and simple approach, it has a fatal defect so that its sensitivity to some kind of group differences is bad. Our goal is to propose a method to ameliorate this lack. In the following, we will briefly indicate the main drawbacks of the VBM analysis, and try to improve according to the fundamental cause of it. It goes into details in chapter 2.



1.4 Motivation

Although VBM is one of most popular morphometric method and has been applied successfully in many instances, there are still limitations that make VBM disable to detect particular anatomical differences in some situations. These limitations are caused by the inherent defect of this approach. It is because VBM is a voxel-by-voxel manner, i.e. a univariate method, to analysis differences by using standard statistical tests at each distinct voxel. That means when VBM tests group difference at a particular voxel, it only takes measurements of images at this voxel in account at a time, and discards potential information carried by other voxels adjacent to this voxel. The way of VBM to analyze the brain structures makes this method simple to use. However, from the spatial point of view, such the voxel-wise manner to find anatomical differences appears improper, because it treats each voxel as an independent object. Adjacent brain tissues should have relations to each

other. As a result, this method is criticized for its capability to estimate widely-distributed, continuous and subtle changes in brain structure [17].

In this work, we proposed a novel method that can consider interrelations between voxels, called the multivariate volumetric morphometry (MVM). In this method, a high-dimensional classification technology is employed. It seeks the most discriminative hyper-plane that separates populations by minimizing the scatter within each individual group and simultaneously maximizing the scatter between groups. The discriminative hyper-plane not only has the ability to classify different groups, but also is appropriate to be used in this application of characterizing the anatomical group discrepancy. Besides, before using this classification technique to find brain differences, a recombination of the spatial and frequency signals is performed for the multiresolution analysis. Our method is built on the classification and the data recombination techniques.

In this thesis, we not only demonstrate the effectiveness of the proposed method, but also develop an efficient computational implementation to save time for analysis. Moreover, a part of idea of this method has been proved in this work. Experimental results showed that the multivariate volumetric morphometry (MVM) indeed has a better sensitivity to subtle and distributed changes of brain structures. So, it is very useful to characterize early symptoms of a disease especially. The details of the reason why we need a multivariate approach and the proposed method are described in the chapter 2 and 3, respectively.

In the following chapters, we introduce the voxel-based morphometry and its drawbacks in chapter 2, and then our method in chapter 3. In chapter 4, some experiments are used to estimate the accuracy of the proposed method, and the comparison between the results of MVM and VBM is performed. Finally, we will bring up some issues about our method MVM in chapter 5, and conclude this work in chapter 6.

Chapter 2

Voxel-Based Morphometry



This chapter is about one of the most popular morphometry—the voxel-based morphometry (VBM). After the introduction, there is a brief interpretation of the basic concept and the optimized protocol of VBM. In the end of this chapter, we statement the inherent drawback of such a voxel-based morphometric analysis, and that is the motive for us to develop another method characterizing anatomical differences.

2.1 Introduction to VBM

The voxel-based morphometry (VBM) is a technique measuring concentrative or volumetric group differences of brain tissues through a voxel-wise analysis of MR images [16]. It is an unbiased and objective method, which explores whole brains rather than specific regions to find the significant structural discrepancy between different groups of subjects. That means people can use the method without the need for the background knowledge of where the discrepancy may exist.

Due to its simplicity, feasibility, and effectiveness, VBM has been widely applied to structural brain studies in the recent years. It is shown in earlier researches that many diseases are related to the abnormal brain structures. The defect, damage, or irregularity of the brain structure will cause irregular behavior of patients. Several studies using VBM characterizing brain differences in a certain disease had good outcomes consistent with observations of those previous researches, such as schizophrenia [18–20], Alzheimer’s disease (AD) [21–23], autism [24,25], spinocerebellar ataxia (SCA) [26], and attention deficit hyperactivity disorder (ADHD) [27].

VBM includes a preprocessing and a voxel-based parametric statistical analysis. Basically, the preprocessing involves spatial normalization, segmentation, and smoothing [16]. The spatial normalization is responsible for registering brain images of different subjects into the same stereotactic space defined by a template image. In the space, we assume

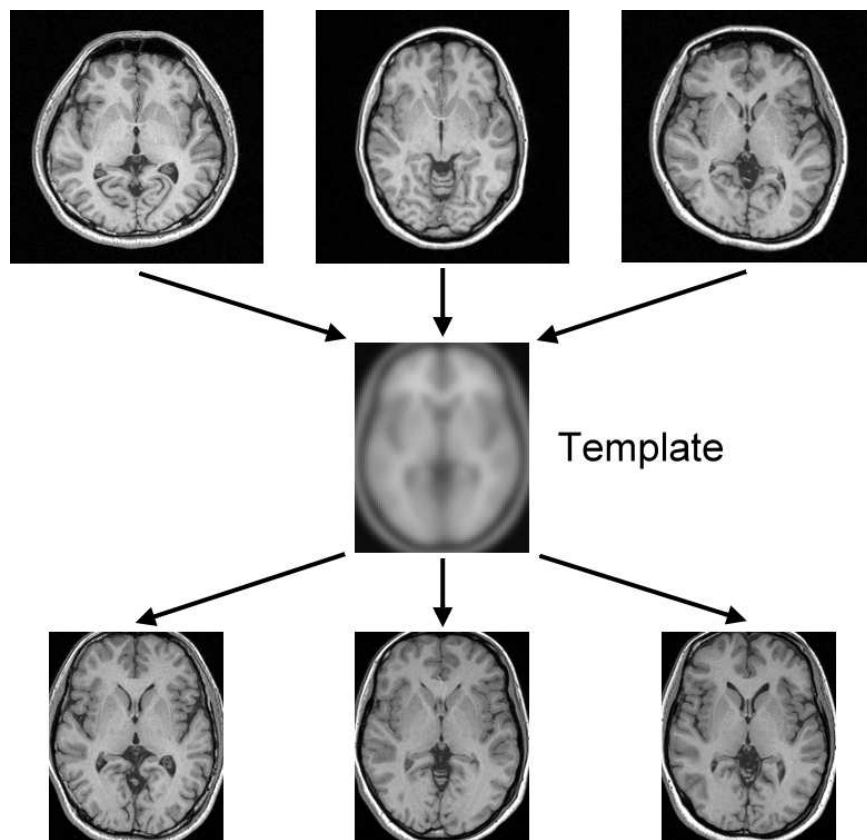


Figure 2.1: The normalization. Images in the upper row, the middle row and the lower row are native MR images, template image and normalized images respectively. Before normalization, the scales and shapes of heads of different subjects are dissimilar. Normalizing images with a standard brain template makes all images in the same stereotactic space where the voxel-wise comparison can be performed.

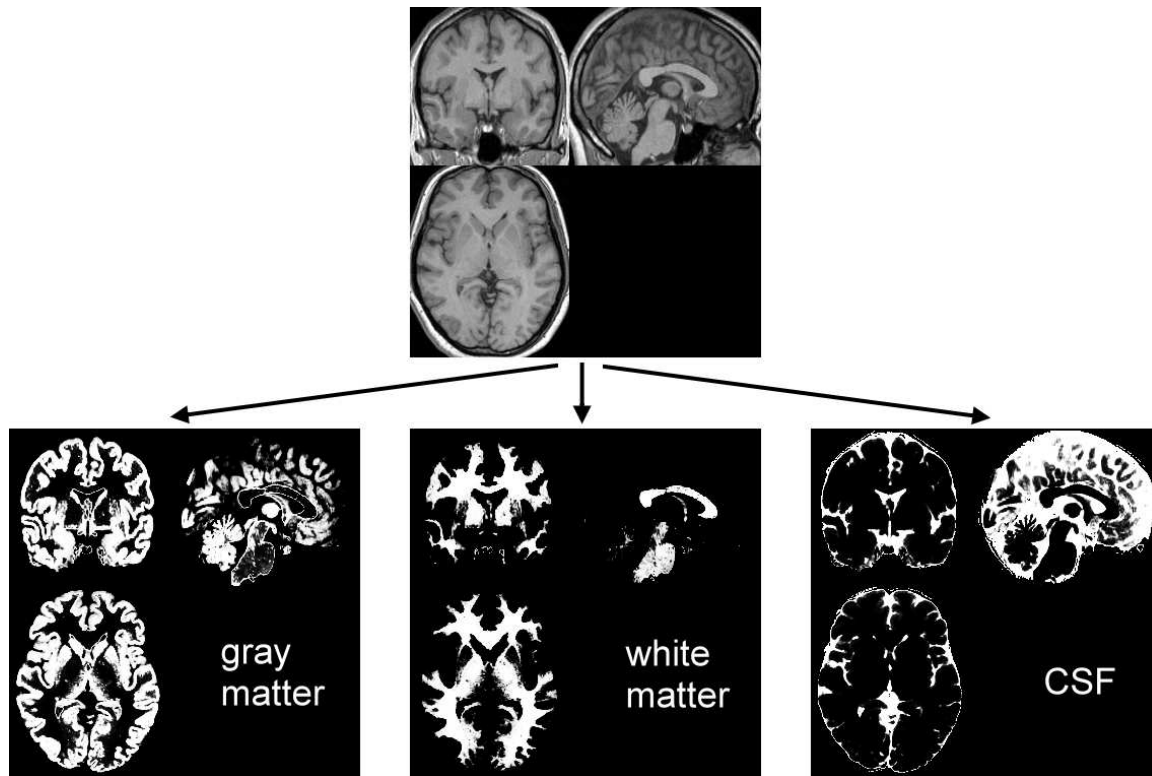


Figure 2.2: The segmentation. The figure shows segmentation of a normalized image into different tissue classes. The resulting segments includes a gray matter (GM) image, a white matter (WM) image, and a cerebrospinal fluid (CSF) image.

that measurements of one certain voxel in different normal brain images should represent the same brain tissue. In the segmentation, the images are segmented into different tissue classes as the gray matter (GM), white matter (WM) or cerebrospinal fluid (CSF). That makes statistical analysis can be performed on different brain tissues. The smoothing is necessary for the following statistical analysis. It conditions the data more normally-distributed and reduces the registration error resulted from the normalization, to increase the validity of inferences based on parametric tests. Figure 2.1 and Figure 2.2 illustrate the concept of normalization and segmentation respectively. After preprocessing, the voxel-based statistical analysis is performed by comparing the normalized and smoothed GM or/and WM images of different groups of subjects. That is, it uses a standard (univariate)

statistical test with a null hypothesis at each and every voxel, to evaluate whether the voxel values of different groups reach the significant level in statistic. If reaching the significant level, we can say there is a difference between the groups at this position of this voxel. The resulting statistical parameters are assembled into an image. Finally, voxels with the statistical parameters preceding the significant level form the regions representative of the detected group differences.

The following is the summary of basic VBM steps [16], and its corresponded flowchart is shown in Figure 2.3:

1. Spatially normalization of all images to the same stereotactic space
2. Segmentation of normalized images into GM, WM, and CSF
3. Smoothing
4. Voxel-based statistical analysis
5. Making inferences about group differences



Actually, there are various methodological implementations of voxel-based morphometric analysis. For example, the RAVENS [9] applied segmentation first, and then normalization, smoothing and statistical analysis. In next section we will introduce one of the most popular the implementations—the optimized VBM protocol [28]. It was used and compared to our method in this thesis.

2.2 Optimized VBM Protocol

There are several cases that some structural differences found by VBM do not really exist between groups of subjects when using the basic VBM steps. The misinterpreted

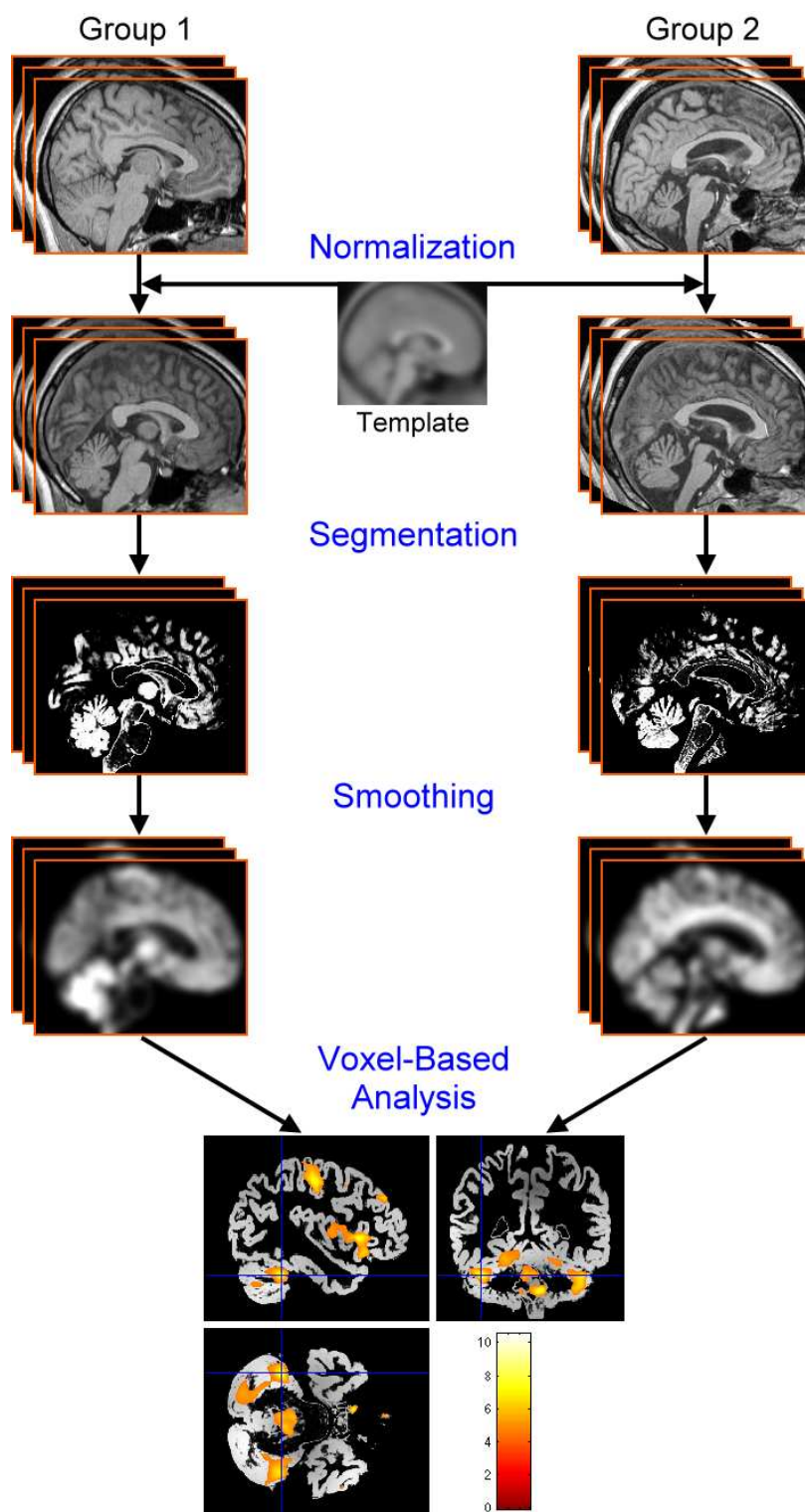


Figure 2.3: Flowchart of basic VBM steps.

differences may be resulted from bad results of normalization which induce subsequently inappropriate comparison between dissimilar brain structures [29]. Although missegmentation leads to the incorrect comparison seemingly, in fact, missegmented tissues are often caused from badly normalized images. It is because in the implement of segmentation of normalized images into gray matter, white matter and cerebrospinal fluid (CSF), Ashburner and Friston use a mixture model cluster analysis designating the distributions of voxel intensity of specific tissue types, and use a priori probability maps describing a priori knowledge of the distribution of different brain tissues in normalized normal subjects to accomplish tissue segmentation [16]. Notice that the a priori probability maps are in the normalized stereotactic space. The efficiency of segmentation is influenced by effect of normalization, because better normalization makes a priori knowledge of the brain tissue distribution can be used more validly in the segmentation. Thus, to minimize the probability of inappropriate comparison between dissimilar brain tissues and structures is to minimize potential error of normalization. For the reason, when wanting to measure group differences of GM/WM, normalization is performed on the segmented GM/WM images rather than on the whole brain images to increase correctness of GM/WM registration results. If the GM/WM images used for normalization are well segmented, then normalization will most likely be fine. It becomes interesting that, a good outcome of normalization could lead to a good outcome of segmentation, and vice versa. Accordingly, the optimized VBM protocol proposed by Good et al. [28] adopts an iterative version of segmentation and normalization to improve effects of preprocessing.

The following is the optimized VBM protocol for measuring group differences of gray matter [28], and its flow diagram is shown in Figure 2.4:

- 1. Creation of customized T1 template and a prior probability maps of GM, WM and CSF**

Customized template is used to reduce potential bias which results from the scanner and the subject population differing from the existing template. All brain images are

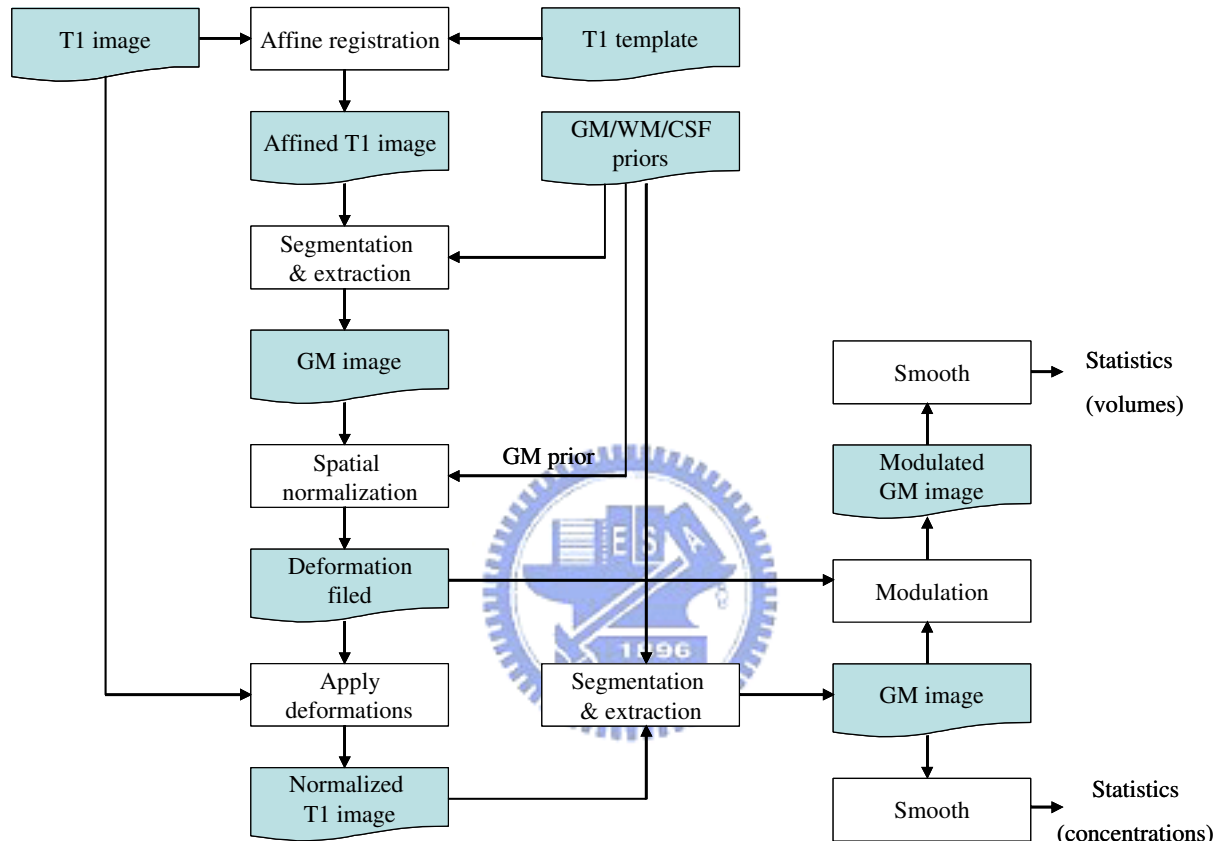


Figure 2.4: Flowchart of the preprocessing of optimized VBM protocol for estimating brain discrepancy of gray matter. It is executed in order of (1) creation of customized T1 template and a prior probability maps of GM, WM and CSF, (2) segmentation and extraction of affine-registered whole brain images, (3) obtaining optimized normalization parameters by normalizing GM images into GM template, (4) normalization of whole brain T1 images with optimized normalization parameters, (5) segmentation and extraction of normalized whole brain images, (6) modulation (if need), and (7) smoothing.

first normalized to the ICBM 152 template (Montreal Neurological Institute), and then segmented into different brain tissues as gray matter, white matter and CSF. Finally, the normalized T1, gray matter, white matter and CSF images are smoothed with an 8mm FWHM isotropic Gaussian kernel and then averaged to create the customized T1/gray/white/CSF mean images (template and priors).

2. Segmentation and extraction of affine-registered whole brain images

In the step, all original structural MR images are affine-registered to the customized T1 template and then segmented with the GM/WM/CSF a priori probability maps derived from above step in native space, followed by morphological operations to remove unconnected non-brain tissues of segmented images. Notice that there is a caveat that the segmentation is performed in native space, but the a priori probability maps are in the normalized stereotactic space. Therefore, there will be another segmentation of normalized images in the following to produce better segmented images.

3. Obtaining optimized normalization parameters by normalizing GM images into GM template

To obtaining the optimized normalization parameters, segmented gray matter images are normalized to the customized gray matter template, which is the GM a priori probability map derived from the first step, by applying 12-parameter affine transformation and nonlinear spatial warping using discrete cosine basis functions. As a result of normalization of gray matter images rather than whole brain images, the normalization parameters of gray matter are optimized by preventing any deformation contributions of non-GM tissues. When wanting to measure group differences in white matter instead, we obtain the optimized normalization parameters just by normalizing WM images into the WM template.

4. Normalization of whole brain T1 images with optimized normalization parameters

All original MR images then are normalized with the optimized normalization parameters. The resolution of normalized images should be relatively high for reducing partial volume effects, which means there is a mixture of different tissue types at one voxel and confounds the subsequent tissue segmentation. In common cases, the voxel size of the normalized image is isotropic in three dimensions.

5. Segmentation and extraction of normalized whole brain images

This step involved segmentation of optimally-normalized images. The images are divided into gray matter, white matter and CSF partitions in the normalized stereotactic space. Non-brain tissues are removed by using morphological operations and a brain mask. In addition, this step also incorporates a correction of image intensity nonuniformity [16] which is mainly caused by magnetic field inhomogeneity of the RF coils during image acquisition. The resulting images are extracted gray matter partition. When estimating WM group differences, extracted WM images are the ticket.

6. Correction for volume changes (optional)

In the segmented image, the value of each voxel is assigned the a posteriori probability that the voxel is classified into this particular tissue type, ranging between 0 and 1. Thus, the segmented GM/WM/CSF data will represent the concentration of GM/WM/CSF. To preserve the total amount of brain tissue, a correction for volume changes, which is usually referred to as “modulation”, is performed by multiplying a voxel value by its Jacobian determinant, which is the relative volume before and after normalization (in step 4). After this correction, these modulated images represent the volume of brain tissues. And, to analyze the modulated images is to test group structural differences in the absolute amount of brain tissues. In this thesis, we always used modulated data to estimate volumetric group differences rather than used non-modulated images to consider the differences in concentration.

7. Smoothing

The normalized, segmented, and modulated images are smoothed with a Gaussian kernel in this step. Smoothing is necessary for the following voxel-based parametric tests. It substitutes the value of each of voxels with a weighted average of surrounding voxels. By the central limit theorem, this action conditions the data more normally distributed such that the validity of inferences based on statistical tests can be increased. Smoothing also reduces the registration error of spatial normalization. However, it is worth noticing that, the choice of the smoothing kernel size should be corresponding to the size of the expected regional differences [30]. Many studies adopt an 8-mm or 12-mm FWHM smoothing kernel when using the VBM method.

8. Statistical analysis

After the preprocessing, the final step is a voxel-wise statistical analysis of normalized and smoothed gray matter images. Statistical analysis employs the general linear model, which is a flexible framework allowing many different tests to be applied, to distinguish significant differences in brain structures of different groups under study [31]. In this thesis, we applied two-sample t -test as the fitting model to describe data of two groups, and used the standard (univariate) parametric t tests to evaluate the residuals at each and every voxel. The resulting statistical parameters of t tests are assembled into an image called the t -test map. By setting a significance level and a minimum cluster size to the t -test map, voxels with the statistical parameters preceding the significant level and in the clusters whose size is larger than the minimum cluster size form the regions representative of the detected significant group differences.

In this work, we not only applied the optimized voxel-based morphometry to compare the capability of revealing structural brain discrepancy between different groups with one of our method, but also used the preprocessing of the optimized VBM protocol to deal with MR images before analyzing them by the proposed method. That is, the implementation of this image preprocessing was also applied in our method. Here we used the SPM2

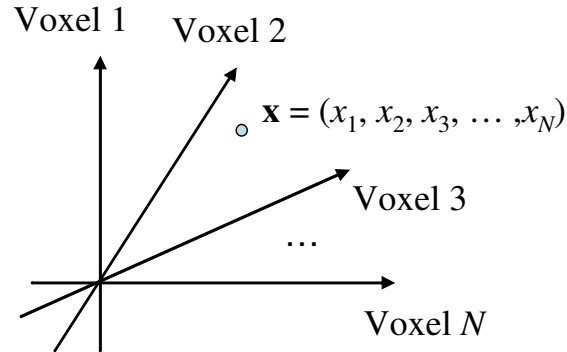
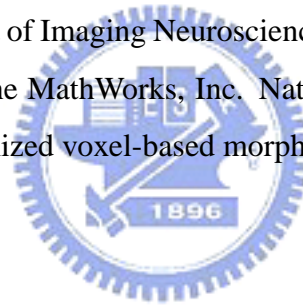


Figure 2.5: Concept of a MR image lying in a high-dimensional space. \mathbf{x} represents a image with N voxels. If we rearranged voxels of the 3-D image to produce an unique 1-D vector by a particular fixed order, then \mathbf{x} can be regarded as one point in a N -dimensional space, where each dimension i stands for voxel i , for $i = 1, \dots, N$.

software (the Wellcome Department of Imaging Neuroscience, University College London, UK) implemented in Matlab 6.5 (the MathWorks, Inc. Natick, MA, USA) to accomplish all procedures involved in the optimized voxel-based morphometry.



2.3 Drawbacks of VBM

Since an individual MR image, a kind of morphological profiles, is commonly described as a collection of voxel-wise morphological measurements, it can be placed in a high-dimensional space where each dimension presenting a voxel. That is, a MR image is thought as one sample point in a high-dimensional space whose dimensionality is equal to the number of voxels of the MR image. Figure 2.5 graphically illustrates this concept. When all images have the same sizes and have been normalized into the same stereotactic space, where voxels at the same position in all the images should contain the same brain tissue, the patient morphological profiles and the normal morphological profiles will form two distributions in the high-dimensional space, like Figure 2.6 shows.

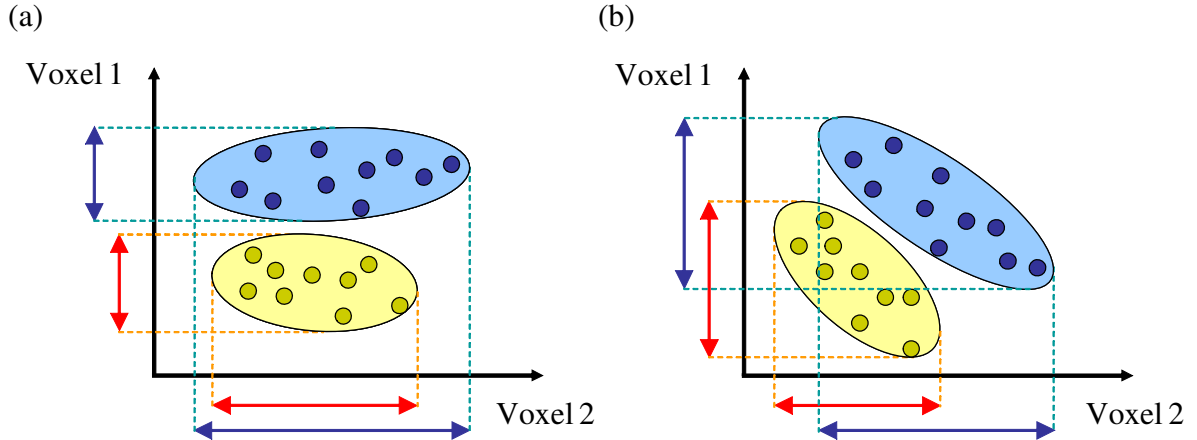


Figure 2.6: Schematic illustration of the significant bias of VBM. Each ellipse represents a population of one group. In the case of (a), we can see that the group discrepancy almost centralizes at the position of voxel 1, so it is very probable that VBM detects a difference at voxel 1 but has no finding at voxel 2. In contrast, because the group discrepancy is spread at voxel 1 and voxel 2 in the case of (b), the difference at voxel 1 and voxel 2 is obscure. Therefore, the voxel-based morphometry may fail to find any differences, even though the discrepancy between the two groups has the same overall magnitude with case (a). These simple two cases show the instability that, the ability of VBM to detect group discrepancy is influenced by the distribution, or pattern, of the discrepancy.

However, Davatzikos [17] pointed out that there is a significant bias of VBM to render inferences about group differences. Let consider two cases where the morphological differences between two groups have the same overall magnitudes in Figure 2.6 (a) and (b). For the purpose of display, there are only two dimensions in the figure, but the dimensionality is much higher in practice. Because voxel-based morphometry detect group differences voxel-by-voxel, only the values along one dimension are taken into account at a time. In the Figure 2.6 (a), there is a significant group difference at the voxel 1, since the two distributions along voxel 1 are easily separated. It is probable that VBM can detect a difference at voxel 1. Along the voxel 2, the situation is opposite, thus VBM may fail to find any difference at voxel 2. Now we focus on the case in the Figure 2.6 (b). It is clear that there is a group discrepancy spreading at voxel 1 and voxel 2. But in the voxel respect, a large overlap of distributions of two groups exists along both voxel 1 and voxel

2. This situation may cause VBM to be unable to detect any differences at either voxel. This simple example in Figure 2.6 (b) reveals that when applying voxel-based morphometry to estimate group discrepancy, some subtle and complex patterns of brain differences, which are widely distributed over many voxels, may not be significant at each single voxel for VBM to detect.

From the cases in Figure 2.6, we know there is a bias in VBM that it detects relatively localized differences much easier than relatively distributed differences involved with several brain structures [17]. This bias is an unavoidable and fatal problem to VBM, and it makes the analysis result of VBM forced to be relied upon the disease characteristics. The problem results from that VBM analyzes the group discrepancy in a voxel-by-voxel manner rather than considers the entirety of voxels simultaneously. In such the voxel-wise analysis method, related information carried by the neighboring voxels are not considered, so it may cause the disability to measure the subtle and widely-distributed discrepancy located in a large region composed of many voxels. Therefore, we proposed another unbiased and automatic method, using a multivariate analysis approach, called the multivariate volumetric morphometry (MVM) to break this limitation of univariate analysis.

Chapter 3

Multivariate Volumetric Morphometry



Owing to the congenital problem of this voxel-wise comparison approach, in this chapter we will introduce the proposed method called the multivariate volumetric morphometry (MVM) which can assess anatomical brain differences. The MVM includes a preprocessing and an analysis step as VBM does. The image preprocessing of MVM is the same with one of VBM (shown in Figure 2.4), in which the modulation is required to characterize volumetric group differences of brain tissues, but the data smoothing is omitted. In the statistical analysis step, MVM adopts a reformatory LDA-based method as the basis of multivariate analysis, and conjugates the wavelet transform, which is used to rearrange the spatial and frequency information for the multiresolution analysis, to measuring group differences. Because each and every voxel represents one variate in analysis, thus MVM is a multivariate approach. This method overcomes the drawback of voxel-based analysis, and is appropriate for estimating the structural brain discrepancy between different groups.

3.1 Ideas of the Proposed Method



The multivariate volumetric morphometry (MVM) is the proposed method which characterizes volumetric anatomical discrepancy between different groups through a multivariate analysis of MR images of particular brain tissues. It is an unbiased, objective and whole-brain measurement. The multivariate volumetric morphometry contains several processes like VBM does, and the chief breakthrough of this thesis is the multivariate analysis. Thus, in the following, we only focus on the statistical analysis step of MVM.

In the multivariate analysis stage, it employs a high-dimensional classification technique, which considers all voxels of MR images at one time, to identify the most discriminative hyper-plane that well separates the populations of groups in the high-dimensional space. This hyper-plane goes along with a unique normal vector, the most discriminant projection vector w , which is the direction shifted from one population to another population. The appearance of a shift might be resulted from some factors of interest cause the group

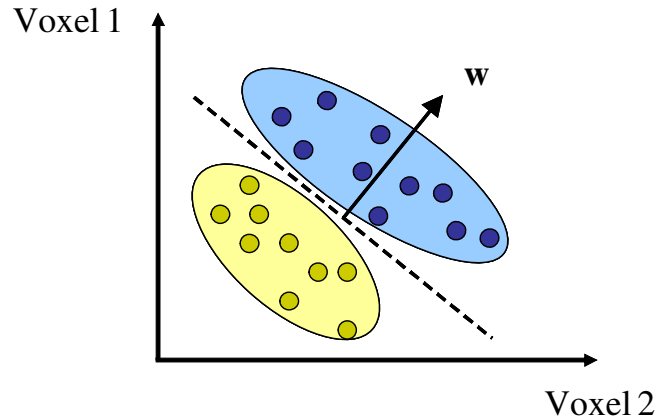


Figure 3.1: Sketchily showing how a high-dimensional classification technique can be used to measure group differences. Assume the yellow points are the patients' morphological profiles, and the corresponding yellow elliptic area is the patients' distribution; either are the blue ones for normal subjects. A classification technique determines the most discriminative hyper-plane, which is presented by the dotted line, and the corresponding most discriminant projection vector w . In such the projection vector w , each element denotes the discrimination weight of group discrepancy. Therefore, the vector w can be considered as a spatial map containing the regions representative of group differences.

discrepancy under study. The most discriminant projection vector is also an image which has the same size of all sample images. The way of using a classification technique to find such the most discriminant projection vector does not need to coincide along any voxels (dimensions), that voxel-based analyses are unable to achieve. Further, each of the parameters from the most discriminant projection vector w denotes the weighting, or the discrimination of characterizing group discrepancy, so the most discriminant projection vector can regarded as the analytic image containing the resulting analysis parameters. In this way, we can quantify differences throughout the whole brain between different groups by such a high-dimensional classification technique. Figure 3.1 illustrates the idea schematically.

3.2 Framework of Multivariate Volumetric Morphometry

Before the multivariate analysis, we used the first six steps of the optimized VBM protocol mentioned in the chapter 2 to obtain individual normalized and modulated gray/white matter images. Notice that the smoothing was disused. The reason why we do not smooth the images in the multivariate volumetric morphometry will be explained in chapter 5, discussion. Then, in the central multivariate method, we used a reformatory LDA-based method, the discriminative common vector method [32], to find the most discriminant projection vector which minimizes the scatter within each individual group and simultaneously maximizes the scatter between groups without the small sample size problem. The resulting projection vector forms a spatial map, whose image size is the same with all gray/white matter images used for the analysis, containing the regions which are most representative of group differences. The details of the method and its efficient implementation we proposed for implementation will be interpreted in the next section.

Besides the discriminative common vector method, we also used the wavelet transform rearranging the spatial and frequency information of MR images to improve the effect of MVM upon catching significant group differences, in several varied scales. The reason we applied the discriminative common vector method in the wavelet space is that, although the method considers all voxels of images simultaneously when estimating group differences, there are the same forces of relationships between all pairs of two voxels in the method; no matter the two voxels are adjacent to or far away from each other. Thus, to increase spatial correlations between neighboring voxels, the 3-D wavelet transformation is used to restructure voxel data into space-scale features in a hierarchical representation way. After that, we then apply the discriminative common vector method on these wavelet features. Moreover, the wavelet transform also makes the analysis become a multivariate multiresolution analysis.

After getting the most discriminant projection vector of two groups' features, the weight

of each feature from the projection vector represents the degree of importance for characterizing the group differences. The number of features, equal to the number of voxels in a MR image, is usually a huge amount. Only the features with larger weights in the most discriminant projection vector are used when performing the inverse 3-D wavelet transformation, to obtain the final discriminating map in the original voxel-based space. Discarding the features with small weights helps to remove trifling differences and to improve accuracy of the multivariate analysis.

Finally, for the purpose of determining and displaying which regions representative of the significant group differences, a smoothing and thresholding of discrimination weights of the parameters in the discriminating map are needed. As mention before, each parameter of the discriminating map denotes the discrimination of characterizing the group discrepancy, so in an intuitively thinking, the changes of the weights of neighboring parameters should be slight. However, in practice, it does not often look smooth as we think. It may result from the noise or the variation within groups, or the error produced during the preprocessing like a wrong image registration or tissue segmentation. It happens especially when we abandon the smoothing step in MVM preprocessing. Therefore, to constrain the smoothness of discrimination weights in the discriminating map regionally, we use a smoothing for the discriminating map. In addition, a thresholding is done before displaying the discriminating map to show the detected regions most representative of group discrepancy. Only voxels with a parameter value preceding the threshold in the discriminating map are considered to reach the significant level and to shall be showed. The minimum cluster size of the parameters also can be set to reject the too small regions. Although the smoothing and thresholding are not the parts of the multivariate analysis step, they are need for visualizing the discrepancy pattern between the groups. Of course, both of the procedures can be regulated by users. In the end of the MVM analysis, we can also obtain a whole-brain confidence in explaining whether the detected group discrepancy is correct, by caculating the p -value associated with the T-statistic on the two groups of projected

images onto the discriminating map.

Summarily, the multivariate volumetric morphometry (MVM) contains the following steps:

1. Spatially normalization of all images to the same stereotactic space
2. Segmentation of normalized images into GM, WM, and CSF
3. Correction for volume changes of segmented images
4. Multivariate analysis
 - (a) Forward 3-D wavelet transformation to the multiresolution space
 - (b) Discriminative common vector method to obtain the most discriminant projection vector
 - (c) Discarding unimportant wavelet features with small discrimination weights in the most discriminant projection vector
 - (d) Inverse 3-D wavelet transformation to obtain the discriminating map in the stereotactic normalization space
5. Visualization of the discrepancy pattern
 - (a) Smoothing
 - (b) Thresholding

Figure 3.2 is the flowchart of the multivariate analysis part of MVM. In implementation, we used the first six steps of the optimized VBM protocol to accomplish the step 1 and 2 of MVM. In the following sections we will introduce the techniques used in the multivariate analysis step, including the discriminative common vector method, the efficient implementation for the discriminative common vector, and the 3-D wavelet transform.

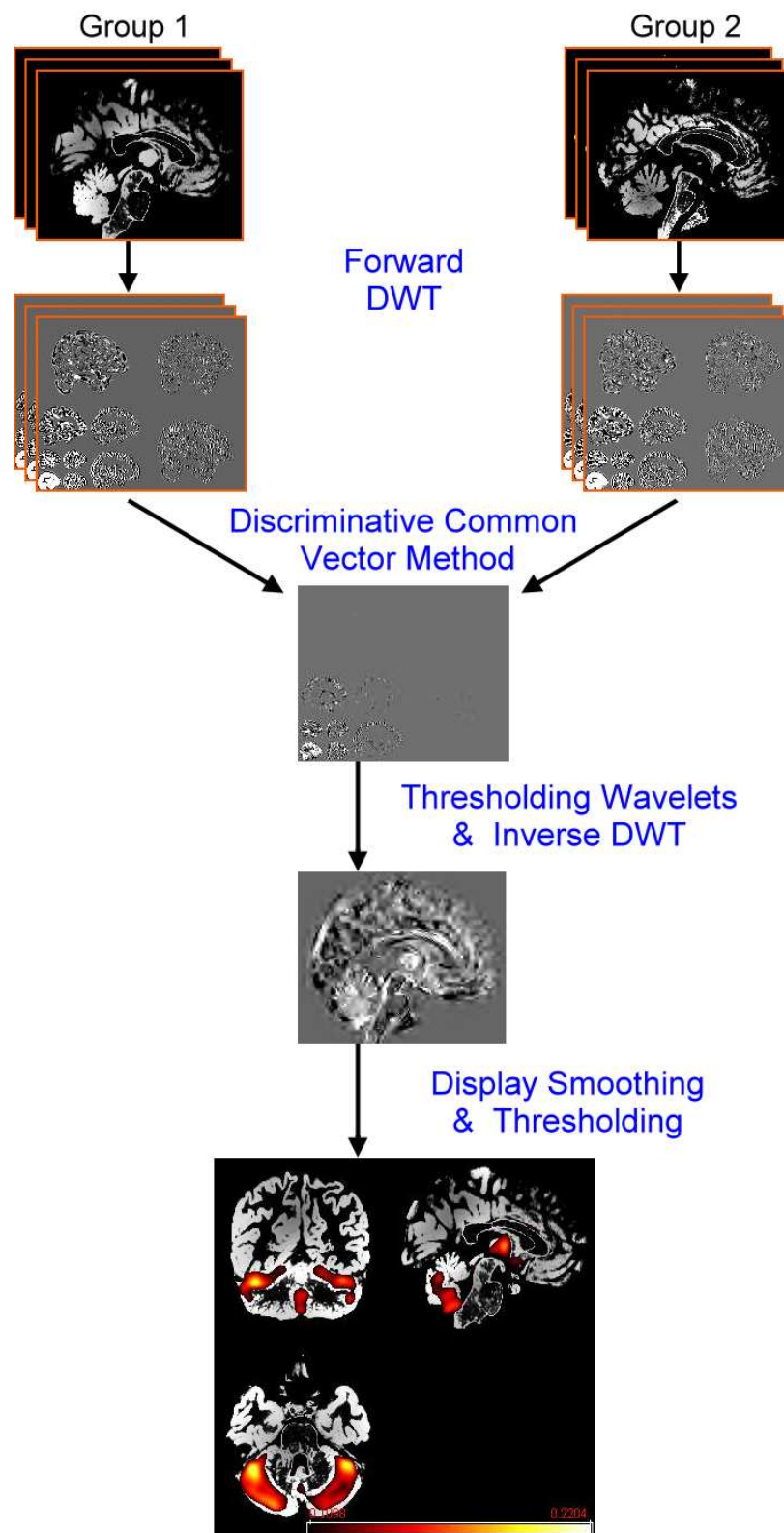


Figure 3.2: Flowchart of multivariate analysis and visualization in MVM.

3.3 Multivariate Analysis using a Reformatory LDA-Based Method

3.3.1 Conventional Linear Discriminant Analysis and Its Potential Problem

The linear discriminant analysis (LDA) is one of the most popular linear projection techniques. It was invented by Ronald A. Fisher in 1936 [33], and has been successfully applied in many classification problems such as image recognition, multimedia information retrieval and so on. Its goal is to find the most discriminant projection vector \mathbf{w} , in which direction groups can be separated with the maximum between-class scatter and the minimum within-class scatter.

Let K be the number of classes (groups), where the k th class contains M_k samples, and let \mathbf{x}_m^k be a N -dimensional column vector which denotes the m th sample of the k th class. There is a total of $M = \sum_{k=1}^K M_k$ samples. The within-class scatter matrix \mathbf{S}_W and the between-class scatter matrix \mathbf{S}_b are defined as

$$\mathbf{S}_W = \sum_{k=1}^K \sum_{m=1}^{M_k} (\mathbf{x}_m^k - \mu^k)(\mathbf{x}_m^k - \mu^k)^T, \quad (3.1)$$

and

$$\mathbf{S}_b = \sum_{k=1}^K M_k (\mu^k - \mu)(\mu^k - \mu)^T, \quad (3.2)$$

where $\mu^k = 1/M_k \sum_{m=1}^{M_k} \mathbf{x}_m^k$ as the mean of samples in the k th class, and $\mu = 1/M \sum_{k=1}^K \sum_{m=1}^{M_k} \mathbf{x}_m^k$ as the mean of all samples. The objective of LDA is to find a projection matrix \mathbf{P}_{lda} that maximizes the Fisher's linear discriminant criterion, that is

$$\mathbf{P}_{lda} = \arg \max_{\mathbf{P}} F(\mathbf{P}) = \arg \max_{\mathbf{P}} \frac{|\mathbf{P}^T \mathbf{S}_b \mathbf{P}|}{|\mathbf{P}^T \mathbf{S}_W \mathbf{P}|}. \quad (3.3)$$

According to linear algebra, the ratio is maximized when the column vectors of \mathbf{P}_{lda} are the eigenvectors of $\mathbf{S}_W^{-1} \mathbf{S}_b$. In implementation, each individual morphological profile is

first reshaped into a sample vector by arranging the 3-D volume in some consistent order before applying the linear discriminant analysis. Moreover, there are only two groups in our case, i.e. $K = 2$, so we can immediately obtain the most discriminant projection vector \mathbf{w} , which is the only one eigenvector composing \mathbf{P}_{lda} , by the formula $\mathbf{w} = \mathbf{S}_W^{-1}(\mu^1 - \mu^2)$.

However, LDA encounters difficulties when the number of samples is much smaller than the dimensionality of the sample space. This situation causes the within-class scatter matrix singular and not invertible, so the LDA cannot be applied directly. It is known as the small sample size (SSS) problem [34]. Therefore, we employ the discriminative common vector method [32], which was proposed by Cevikalp and Wilkes for face recognition, to solve this problem.

3.3.2 Discriminative Common Vector Method

The discriminative common vector method for the small sample size problem is based on a variation of the LDA by maximizing the modified Fisher's linear discriminant criterion [35]. The general idea of the common vector is to find a vector which can represent a class by extracting common properties of the class, or saying that, by eliminating differences between the samples in the class. After getting each common vector of every class, we can use the principal components analysis (PCA) [36] to find the principal components which actually equate the most discriminant projection vectors of LDA.

Let us use all previous definitions and let the total scatter matrix be defined as

$$\mathbf{S}_t = \sum_{k=1}^K \sum_{m=1}^{M_k} (\mathbf{x}_m^k - \mu)(\mathbf{x}_m^k - \mu)^T = \mathbf{S}_w + \mathbf{S}_b. \quad (3.4)$$

The modified Fisher's linear discriminant criterion

$$\hat{F}(\mathbf{P}) = \frac{|\mathbf{P}^T \mathbf{S}_b \mathbf{P}|}{|\mathbf{P}^T \mathbf{S}_t \mathbf{P}|} = \frac{|\mathbf{P}^T \mathbf{S}_b \mathbf{P}|}{|\mathbf{P}^T \mathbf{S}_w \mathbf{P} + \mathbf{P}^T \mathbf{S}_b \mathbf{P}|} \quad (3.5)$$

has been proved that it is exactly equivalent to the original Fisher's criterion by Liu et al. [35], saying that

$$\arg \max_{\mathbf{P}} \hat{F}(\mathbf{P}) = \arg \max_{\mathbf{P}} F(\mathbf{P}). \quad (3.6)$$

The modified criterion will attain a maximum in the special case, proved in [37], where $\mathbf{p}^T \mathbf{S}_w \mathbf{p} = 0$ and $\mathbf{p}^T \mathbf{S}_b \mathbf{p} \neq 0$, for all projection vectors $\mathbf{p} \in \mathbb{R}^N \setminus \{0\}$. Under these conditions for \mathbf{p} , a better criterion [38] will be

$$\arg \max_{|\mathbf{P}^T \mathbf{S}_w \mathbf{P}|=0} |\mathbf{P}^T \mathbf{S}_b \mathbf{P}| = \arg \max_{|\mathbf{P}^T \mathbf{S}_w \mathbf{P}|=0} |\mathbf{P}^T \mathbf{S}_t \mathbf{P}|.$$

That is to say, if we transform all samples onto the null space of \mathbf{S}_w to restrict the projected within-class scatter matrix to be a zero matrix, and then calculate the principal components that maximize $|\mathbf{P}^T \mathbf{S}_t \mathbf{P}|$ by performing PCA, we will obtain the most discriminant projection vectors without the small sample size problem. It is called the null space method proposed by Chen et al. [37].

The transformation matrix from the original sample space to the null space of \mathbf{S}_w is $\bar{\mathbf{Q}}\bar{\mathbf{Q}}^T$, where the column vectors of $\bar{\mathbf{Q}}$ are the vectors spanning the null space of \mathbf{S}_w . Cevikalp and Wilkes [32] have proved that, projecting every samples \mathbf{x}_m^k (which denotes the m th sample of the k th class) in the k th class onto the null space of \mathbf{S}_w will produce exactly one vector $\mathbf{x}_{\text{com}}^k = \bar{\mathbf{Q}}\bar{\mathbf{Q}}^T \mathbf{x}_m^k$, which is referred to the common vector; moreover, because of $\bar{\mathbf{Q}}\bar{\mathbf{Q}}^T \mathbf{x}_m^k = \mathbf{x}_m^k - \mathbf{Q}\mathbf{Q}^T \mathbf{x}_m^k$, the common vector $\mathbf{x}_{\text{com}}^k$ of the k th class can be calculated without $\bar{\mathbf{Q}}$ by using

$$\mathbf{x}_{\text{com}}^k = \mathbf{x}_m^k - \mathbf{Q}\mathbf{Q}^T \mathbf{x}_m^k, \quad (3.7)$$

where \mathbf{Q} is the matrix whose column vectors are the orthonormal vectors spanning the range space of \mathbf{S}_w . Since the number of columns in \mathbf{Q} is about M and the number of columns in $\bar{\mathbf{Q}}$ is about $N - M$, the size of \mathbf{Q} is much smaller than the size of $\bar{\mathbf{Q}}$. It states that the method can greatly reduce the computational burden than the null space method.

After obtaining the common vector for each and every group, the principal components of those common vectors will be the most discriminant projection vectors. It is because

there is exactly one class over the common vectors now. These principal components of the common vectors are called the discriminative common vectors. Again, in our practice, samples are divided into only two groups, so there is only one discriminative common vector in the event. We obtain the most discriminant projection vector \mathbf{w} by directly subtracting of the two common vectors, namely, $\mathbf{w} = \mathbf{x}_{\text{com}}^1 - \mathbf{x}_{\text{com}}^2$.

So the steps of the discriminative common vector method are as follows:

1. Compute the eigenvectors $\alpha_1, \alpha_2, \dots, \alpha_r$ corresponded to the nonzero eigenvalues of \mathbf{S}_W , where r is the rank of \mathbf{S}_W , and set $\mathbf{Q} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_r]$.
2. Obtain the common vectors for each class by choosing any sample from each class and projecting it onto the null space of \mathbf{S}_W , those are

$$\mathbf{x}_{\text{com}}^1 = \mathbf{x}_m^1 - \mathbf{Q}\mathbf{Q}^T \mathbf{x}_m^1, \quad m \in \{1, \dots, M_1\}, \quad (3.8)$$

and

$$\mathbf{x}_{\text{com}}^2 = \mathbf{x}_m^2 - \mathbf{Q}\mathbf{Q}^T \mathbf{x}_m^2, \quad m \in \{1, \dots, M_2\}. \quad (3.9)$$

3. Compute the only one discriminative common vector, i.e. the most discriminant projection vector \mathbf{w} by

$$\mathbf{w} = \mathbf{x}_{\text{com}}^1 - \mathbf{x}_{\text{com}}^2. \quad (3.10)$$

3.3.3 Efficient Implementation for Computing Discriminative Common Vectors

Although the discriminative common vector method solves the small sample size problem, there are still come difficulties in implementation. It is because the dimensionality of the sample space is a very huge amount. For example, a 3-D image with the size $157 \times 189 \times 156$ has more than 4.6×10^6 voxels. Therefore, we proposed an efficient implementation for computing the discriminative common vector.

Since the within-class scatter matrix is defined as $\mathbf{S}_W = \sum_{k=1}^K \sum_{m=1}^{M_k} (\mathbf{x}_m^k - \mu^k)(\mathbf{x}_m^k - \mu^k)^T$, it can be rewritten as

$$\mathbf{S}_W = \mathbf{A}\mathbf{A}^T, \quad (3.11)$$

where the matrix $\mathbf{A} = [\mathbf{x}_1^1 - \mu^1 \ \cdots \ \mathbf{x}_{M_1}^1 - \mu^1 \ \mathbf{x}_1^2 - \mu^2 \ \cdots \ \mathbf{x}_{M_2}^2 - \mu^2]$. Rather than directly calculating the large N -by- N matrix \mathbf{S}_W , we used a computationally feasible method [39,40] to compute the eigenvectors of $\mathbf{A}\mathbf{A}^T$ by multiplying the matrix \mathbf{A} by the matrix $\tilde{\mathbf{Q}}$ whose columns are the eigenvectors of $\mathbf{A}^T\mathbf{A}$. So the matrix representation of the subsequent operations is written as

$$\begin{aligned} \mathbf{Q} &= \mathbf{A}\tilde{\mathbf{Q}} \\ \mathbf{x}_{\text{com}}^1 &= \mathbf{x}_1^1 - \mathbf{Q}(\mathbf{Q}^T \mathbf{x}_1^1) \\ \mathbf{x}_{\text{com}}^2 &= \mathbf{x}_1^2 - \mathbf{Q}(\mathbf{Q}^T \mathbf{x}_1^2) \\ \mathbf{w} &= \mathbf{x}_{\text{com}}^1 - \mathbf{x}_{\text{com}}^2 \end{aligned} \quad (3.12)$$

where we choose the first sample of each class to obtain the common vector.

However, there is still a heavy computational cost if we translate these equations into programming codes without simplifying them. It is known that, a matrix multiplication \mathbf{BC} needs $n_1 \times n_2 \times n_3$ multiplications when the matrix \mathbf{B} is n_1 -by- n_2 and the matrix \mathbf{C} is n_2 -by- n_3 . So, how many multiplications it will take if we do not change the computation way? For this purpose, we developed an efficient implementation to achieve the above objective (3.12). The following is the pseudo-code:

```

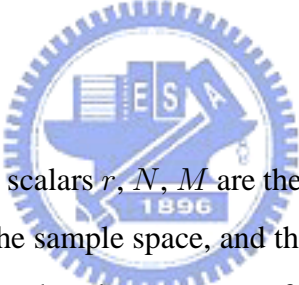
1 for  $j := 1$  to  $r$  do
2     for  $i := 1$  to  $N$  do
3          $\mathbf{q}_j(i) := 0$ 
4         for  $l := 1$  to  $M$  do
5              $\mathbf{q}_j(i) := \mathbf{q}_j(i) + \mathbf{A}(i, l) * \tilde{\mathbf{Q}}(l, j)$ 
6         end

```

```

7      end
8       $\mathbf{q}_j := \text{NormalizeVector}(\mathbf{q}_j)$ 
9       $\text{dot}_1 := \text{Dot2Vectors}(\mathbf{q}_j, \mathbf{x}_1^1)$ 
10      $\text{dot}_2 := \text{Dot2Vectors}(\mathbf{q}_j, \mathbf{x}_1^2)$ 
11     for  $i := 1$  to  $N$  do
12          $\mathbf{QQtX}_1(i) := \mathbf{QQtX}_1(i) + \mathbf{q}_j(i) * \text{dot}_1$ 
13          $\mathbf{QQtX}_2(i) := \mathbf{QQtX}_2(i) + \mathbf{q}_j(i) * \text{dot}_2$ 
14     end
15 end
16 for  $i := 1$  to  $N$  do
17      $\mathbf{w}(i) := (\mathbf{x}_1^1(i) - \mathbf{QQtX}_1(i)) + (\mathbf{x}_1^2(i) - \mathbf{QQtX}_2(i))$ 
18 end

```



In this code fragment, the scalars r , N , M are the number of column vectors composed of $\tilde{\mathbf{Q}}$, the dimensionality of the sample space, and the number of all samples, respectively. The vector \mathbf{q}_j represents the j th column vector of \mathbf{Q} . The N -by- M matrix \mathbf{A} and the M -by- r matrix $\tilde{\mathbf{Q}}$ represent as the definitions before. And, the vector \mathbf{QQtX}_k , used to calculate the common vector $\mathbf{x}_{\text{COM}}^k$, represents the projected sample of \mathbf{x}_1^k of the range space of \mathbf{S}_W , for $k = 1, 2$. Furthermore, the function *NormalizeVector*() makes the input vector turning out a vector with the norm equal to 1 in the same direction. The function *Dot2Vectors*() returns the scalar product of the input vectors. Finally, the resulting vector \mathbf{w} is the most discriminant projection vector.

3.4 Multiresolution Analysis using Wavelet Transform

In image processing, there are many methods and theories to be developed. Owing to convenience of analysis, it usually transforms domain of signals. In engineering application, the most popular method is Fourier transform. Fourier transform can transform the signals from spatial domain to frequency domain, but the information of spatial domain is lost after applying Fourier Transform on an image. In many applications, however, it needs to analyze both frequency and spatial information at the same time. To avoid the lack of spatial information, Haar, Goupillaud, Grossman, and Morlet proposed and improved wavelet transform [41].

Wavelet transform is one of multi-resolution analysis. Wavelet transform not only can transform an image from spatial domain to frequency domain, but also has information of both two domains. Similar to the Fourier transform, wavelet transform consists of continuous wavelet transform (CWT), and discrete wavelet transform (DWT). However, continuous wavelet transform is limited by the redundancy and impracticality in image processing. Therefore, the discrete wavelet transform is applied in this thesis.

3.4.1 3-D Discrete Wavelet Transform

The most important parameter in wavelet transform is called wavelet function, which is also called mother wavelet. Wavelet function $\psi(x)$, where x is the parameter in the spatial domain, has to satisfy two properties as follows:

1. The integration of wavelet function has to be zero,

$$\int_{-\infty}^{\infty} \psi(x) dx = 0. \quad (3.13)$$

2. Wavelet function has finite energy,

$$\int_{-\infty}^{\infty} |\psi(x)dx|^2 < 0. \quad (3.14)$$

First property represents that wavelet function is oscillating, so wave function is always like an oscillatory wave. Second property represents finite energy, so wavelet function decays to zero in both positive and negative directions. Compared with harmonic waveform, wavelet function is relative smaller. This is the underlying reason that it is called "wave-let".

In this thesis, wavelet transform is used for the multiresolution analysis. Wavelet transform based multiresolution analysis is to analyze the signals or images under different scales and resolutions. Utilizing the multiresolution analysis, an image with complex frequencies can be decomposed into many images with simple frequencies. The decomposed images can be analyzed independently or in community. To discuss the method of multiresolution analysis, besides wavelet functions, the scaling functions, usually inferred to the father wavelet, have to be introduced. Define the scaling function $\phi(x)$, which have to satisfy three basic properties:

1. The integration of scaling function has to be 1,

$$\int_{-\infty}^{\infty} \phi(x)dx = 1. \quad (3.15)$$

2. The energy of scaling function is equal to 1,

$$\int_{-\infty}^{\infty} |\phi(x)dx|^2 = 1. \quad (3.16)$$

3. The scaling function $\phi(x)$ and its transformation by shifting n , $\phi(x - n)$, compose of an orthogonal set,

$$\langle \phi(x), \phi(x - n) \rangle = \delta(n), \quad (3.17)$$

where $\delta(n)$ is Kronecker delta symbol. $\delta(n) = 1$ as $n = 0$, and $\delta(n) = 0$ as $n \neq 0$. Wavelets can be defined by the wavelet function and the scaling function. Moreover, the

wavelet function and the scaling function are usually set to satisfy the orthogonal relation. When the scaling function and the scaling function are orthogonal, besides computing easily, image energy can be divided in the spatial domain and the frequency domain. This is really a useful property in practice.

The forward 1-D discrete wavelet transform and the inverse 1-D discrete wavelet transform are defined respectively that

$$W_\phi(a, b) = \frac{1}{\sqrt{a}} \sum_x f(x) \phi\left(\frac{x-b}{a}\right), \quad (3.18)$$

$$W_\psi(\bar{a}, b) = \frac{1}{\sqrt{\bar{a}}} \sum_x f(x) \psi\left(\frac{x-b}{\bar{a}}\right), \quad (3.19)$$

and

$$f(x) = \frac{1}{\sqrt{a}} \sum_b W_\phi(a, b) \phi\left(\frac{x-b}{a}\right) + \frac{1}{\sqrt{\bar{a}}} \sum_{\bar{a}=a}^{\infty} \sum_b W_\psi(\bar{a}, b) \psi\left(\frac{x-b}{\bar{a}}\right), \quad (3.20)$$

where $f(x)$ is a spatial signal to decompose, b is the spatial shift or translation factor, a and \bar{a} are called the spatial scaling or dilatation factor which determine the size of the spatial axis translation, $W_\phi(a, b)$ is the scaling coefficient, and W_ψ is the wavelet coefficient.

In practice, all discrete wavelet transforms make use of filter banks that contain band-pass filters separating the input signal into several components. Each component carries a single frequency sub-band of the original signal. In the 3-D case, the 1-D analysis filter bank is applied in turn to each of the three dimensions. Usually, the overwhelming majority of the signal information appears in the component with the lowest frequency band. So it is often to separate the existing component with the lowest frequency band in a recursive way. The level of a discrete wavelet transformation represents how many times the signal separation is preformed. Figure 3.3 interprets this idea schematically.

The data used to be analyzed by the discriminative common vector method, which are illustrated in Figure 3.4, are these wavelet coefficients from the 3-D discrete wavelet transform. After the discriminative common vector method analysis and a wavelet thresholding

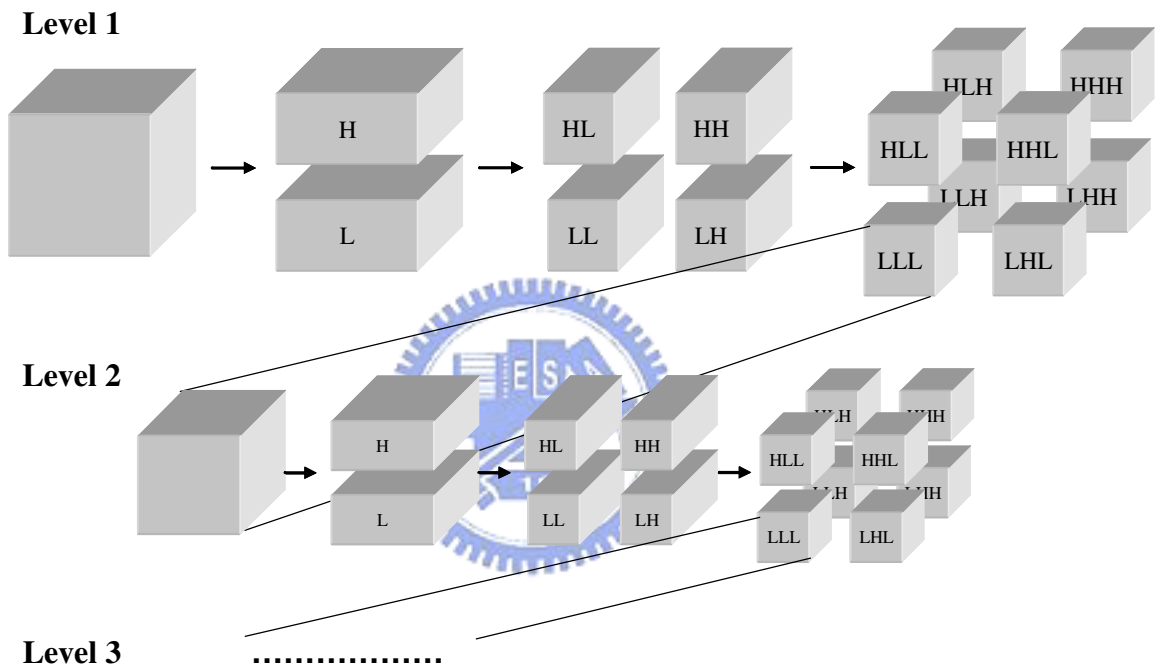


Figure 3.3: Analysis by 3-D discrete wavelet transform with filter banks. The 1-D analysis filter bank is applied in turn to each of dimensions. And the resulting component with the lowest frequency band continues to decompose in the next level of DWT.

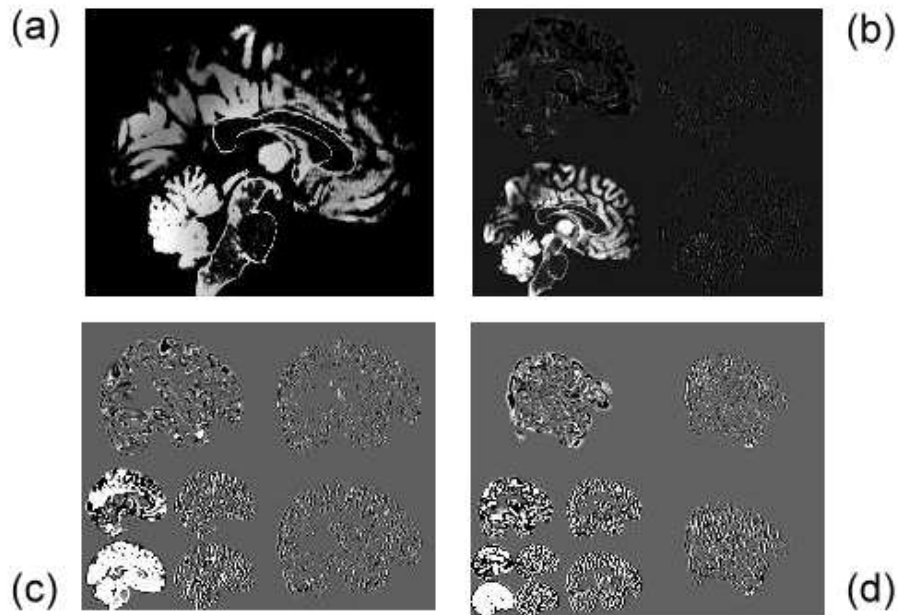


Figure 3.4: A MR image by discrete wavelet transform with different levels. (a) Original MR image. Decomposed images, or the wavelet coefficients, from (b) 1-level DCT, (c) 2-level DCT, (d) 3-level DCT.

which is introduced in the next section, we perform the inverse 3-D discrete wavelet transform to obtain the discriminating map of MVM.

3.4.2 Wavelet Thresholding

As the mention before, a thresholding of wavelets coefficients is performed before the inverse wavelet transform, in order to preserve the large discrimination weights for a good discriminating map. There are three simple thresholding approaches in common use: the hard thresholding, soft thresholding, and percentage thresholding [42]. We used the percentage thresholding approach to preserve certain quota of the total discrimination to characterize the discrepancy between groups. By setting the percentage $\alpha\%$, we keep the wavelet coefficients with larger absolute discrimination weights from the most discrim-

inant projection vector, so that the sum of these absolute weights comprises α percents of the total of the absolute weights from the LDA projection vector. And, we discard other wavelet coefficients which have the smaller absolute weights. After doing the α percentage thresholding, the inverse wavelet transform is now performed to obtain the discriminating map in the original voxel-based space.





Chapter 4

Experiments



This chapter shows the experimental results which utilize MVM and VBM respectively to analyze the anatomical differences between two different groups. It includes the experimental parameters, the results and comparison between the two analysis methods, and some other related methods about the experiments. Simulation data were used to assess the capability for MVM and VBM to reveal the group discrepancy. These two methods were also applied on real SCA3 data. In both simulation and SCA experiments, the proposed method expresses a better sensitivity to subtle and widely-distributed variation of brain structure.

4.1 Capability Assessment for Discrepancy Revelation

To validate whether the proposed multivariate method can detect more distributed atrophy than voxel-based analysis, simulation MR images were generated. We used thin-plate splines (TPS) [43], which is an interpolation tool and suited for image morphing, to simulate normal and patient groups of images. The most parts of advantages of using the simulation data are that we can control the difference patterns between two groups, and we have the ground truth to verify and compare the results of the two analysis approaches. The methods of data generation and accuracy evaluation, and the experiment results and comparisons will be presented in order.

4.1.1 Materials

We used a normalized gray matter MR image of a normal subject as the source image to simulate one normal group and eight patient groups by the thin-plate spline (TPS) method. The simulated atrophy of patients is around the cerebellum area. Eight patient groups have different degrees of simulated atrophy to compare with the normal group. The general idea of TPS is first choosing several control points on the source image, and then moving

these control points to produce a new image by interpolation with thin-plate splines. The simulation procedure is described below:

1. Select one optimized-normalized and modulated GM image from normal subjects to be the simulation source image. All simulation data are produced from this image. The reason why the simulation was performed on a normalized and segmented image rather than on a native T1-weighted image is that, we wanted to avoid other factors, such as normalization and segmentation, which might affect accuracy of simulation data for analysis and which should be not considered under this study. The sizes of the source image and simulation data are all $78 \times 94 \times 78$ with a voxel size $2 \times 2 \times 2\text{mm}^3$.
2. Choose control points for the thin-plate spline (TPS) transformation. 68 control points were manually selected on the surface of cerebellar gray matter, and 140 control points were automatically selected on the surface of cerebral cortex by a program. Specifically, 30 of the 68 control points, which are on the exterior surface of cerebellar gray matter, were used to simulate atrophy around the cerebellum; Other 38 of the 68 control points, which are on the interior surface of cerebellar gray matter or on the interface between the cerebellum and cerebrum, were used to avoid global motion effects of the atrophy simulation as using TPS interpolation. We used these two different groups of (30 and 38) control points on the cerebellum to vary the thickness of cerebellar gray matter, rather than the size of cerebellum (in that way it only needs the 30 control points on the exterior surface of cerebellum). Figure 4.1 sketches the distribution of these control points.
3. Move the 30 control points on the exterior surface of cerebellar gray matter towards the individual cerebellar hemisphere centers to create thickness atrophy of the cerebellum. The distance of movement is one of control parameters in this simulation. By setting different distances, we can produce various degrees of simulated cerebellum atrophy. This step should be skipped when generating simulated normal images.

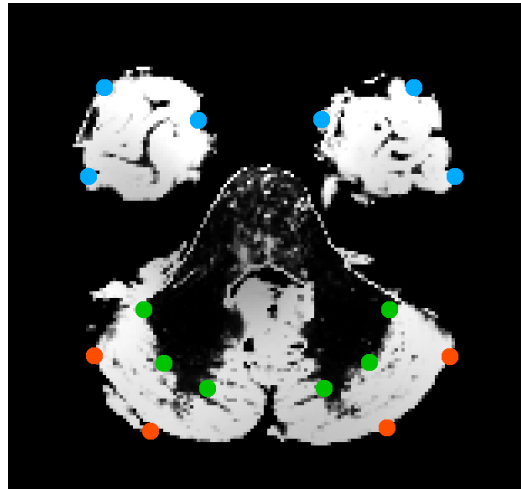


Figure 4.1: The distribution of control points for TPS simulation. The orange points represent these control points on the exterior surface of cerebellar gray matter, which were used to generate cerebellar atrophy; The green points represent ones on the interior surface of cerebellar gray matter, which were used to avoid global motion effects of TPS interpolation; The blue points represent ones on the surface of cerebral cortex. Finally, to simulate the variance within a group, all points were used to render the cerebral and cerebellar surfaces with a white noise.

4. Shift all control points in arbitrary directions with the small distances by using a white noise to randomly render the cerebral and cerebellar surfaces. The shifting distance is also a control parameter in this simulation to produce the inter-subject variation.
5. Perform the TPS transformation on the source image to generate simulated images with or without the atrophy around the cerebellum area. Images with the simulated atrophy form the patient group, and ones without the simulated atrophy form the normal group.

To generate the normal group and eight patient groups with different degrees of cerebellum atrophy, we used the TPS simulation procedure with a white noise, which has standard deviation 2mm along each dimension, on all control points for all the normal and patient groups, and with the eight atrophy distances from 1mm to 8mm on these control points on

the exterior surface of cerebellum for the patient groups. In every group, there were 20 simulated images. Thus we had eight pairs of the simulated normal and patient groups to perform eight analyses with distinct degrees of structural differences by using both MVM and VBM. Figure 4.2 shows the source image and simulated images with 1mm, 3mm, 5mm and 7mm cerebellum atrophy.

4.1.2 Accuracy Evaluation

Because the simulation parameters and data are under our control, we can use the ground truth of the discrepancy pattern we already knew to test and verify the results with some evaluation methods. The ground truth is an image containing the region with the structural discrepancy between the simulated normal and patient groups. It was obtained by subtracting an image, which was generated by the same TPS procedure without noise appended on the control points but only with the atrophy distance producing the cerebellar atrophy, from the source image. If the simulated images were well generated with good parameters that make the distribution within the group of the images near the normal distribution, the ground truth is good for representation of the actual difference between two simulated groups.

After analyzing the simulation data and determining the significant level (threshold) of a result image with the analytic parameters, i.e. a discriminating map of MVM or a t -test map of VBM, a voxel-wise comparison between the result image and the ground truth is preformed to find the regions of true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN). The four terms TP, FP, TN, and FN mean that:

- True-positive (TP): the analysis result predicts the voxel is in the atrophy area, and in reality it is in the atrophy area of the ground truth;
- False-positive (FP): the analysis result predicts the voxel is in the atrophy area, but

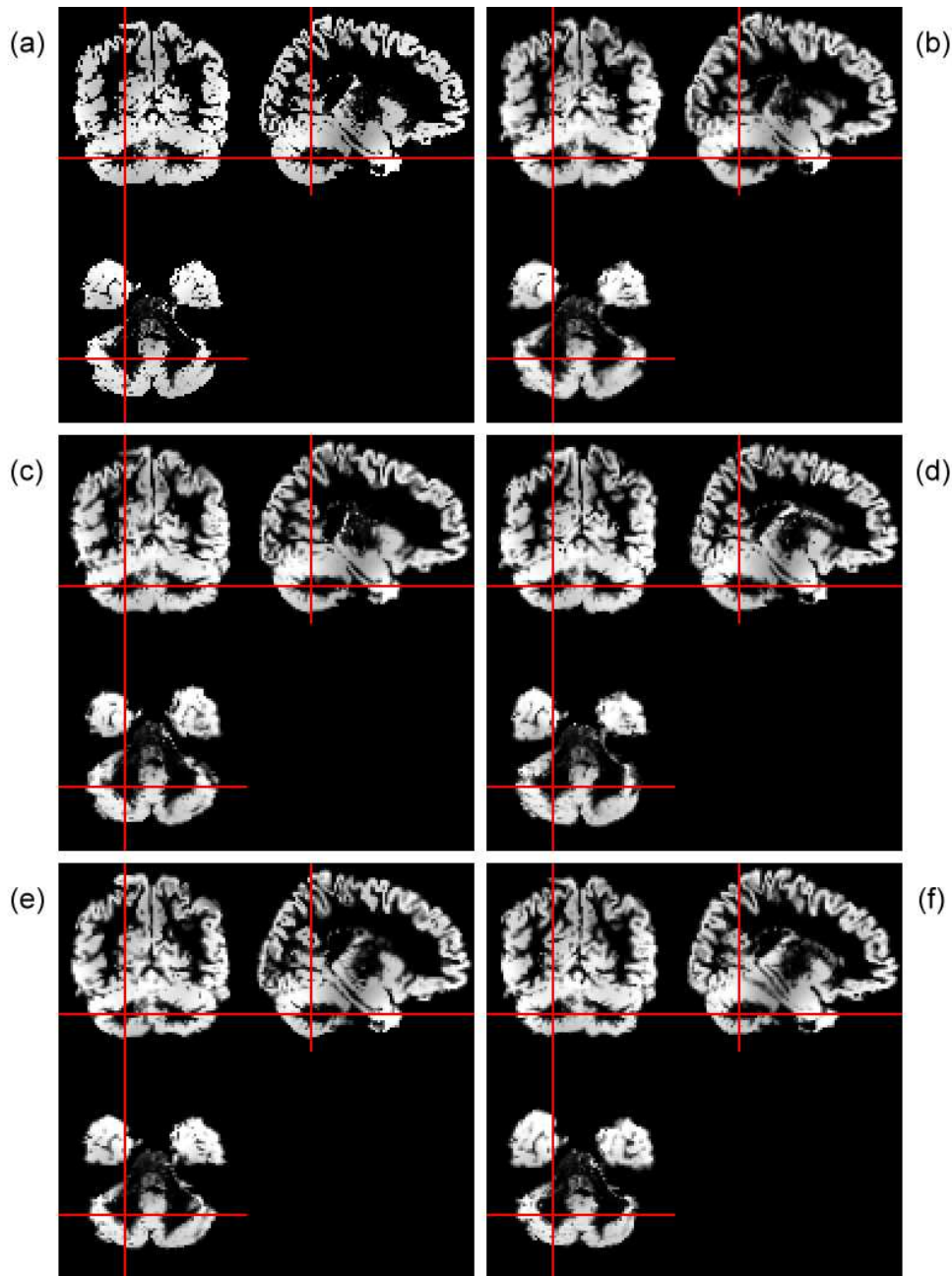


Figure 4.2: The source and simulated images by TPS shown in the coronal, sagittal and axial views. (a) is the source image used to generated simulation data; (b) is a simulated normal image; and (c), (d), (e), (f) are the simulated patient images with atrophy around the outer cerebellar cortex with a atrophy distance of 1mm, 3mm, 5mm, and 7mm respectively. All simulated images were rendered with a white noise whose standard deviation is 2mm along each dimension. We can see the thickness of the cerebellum in the simulated images is thinner and thinner from (b) to (f). The sizes of these images are all $78 \times 94 \times 78$ with a voxel size $2 \times 2 \times 2\text{mm}^3$, and the red cross is placed at the voxel (52, 25, 17)

Table 4.1: Definitions of TP, FP, TN, and FN.

Actual \ Predicted	Yes	No
	Yes	TP
No	FP	TN

in reality it is **not** in the simulated atrophy area of the ground truth;

- True-negative (TN): the analysis result predicts the voxel is **not** in the atrophy area, and in reality it is **not** in the simulated atrophy area of the ground truth;
- False-negative (FN): the analysis result predicts the voxel is **not** in the atrophy area, but in reality it is in the simulated atrophy area of the ground truth.

Table 4.1 shows the brief definitions of them. The analysis result could be true in two respects TP and TN; and it could be false in two respects FP and FN. In statistics, we call FN the type I error, and call FP the type II error. Moreover, the rates of TP, FP, TN and FN are defined as:

$$TPrate = \frac{TP}{TP + FN}, \quad (4.1)$$

$$FPrate = \frac{FP}{TN + FP}, \quad (4.2)$$

$$TNrate = \frac{TN}{TN + FP}, \quad (4.3)$$

$$FNrate = \frac{FN}{TP + FN}. \quad (4.4)$$

After we got these regions, we can assess an analysis result by displaying the result image with labeled TP, FP, FN regions. More TP regions and fewer FP, FN regions indicate the result is more accurate. The comparison between two morphometric analysis methods MVM and VBM is done by showing their labeled result images and TP/FN rates, in the case of that their FP rates are equal to each other.

Another more precise assessment is to use the receiver operating characteristic (ROC), which employs *sensitivity* and *specificity* to present achievements of a prediction method in a graphic way. In this study, the sensitivity means the ability of a method to identify which voxels has actually atrophy, so it is simply the TP rate; and the specificity means the ability to identify which voxels do not have actually atrophy, so it is the TN rate. The ROC curve of a method is obtained by plotting sensitivity against 1-specificity via varying the value of the parameter of this method. In our case, the varied parameter of MVM is the display threshold of the discrimination weights in the discriminating map, and the varied parameter of VBM is the display threshold of the t values in the t -test map. In general, a method with a ROC curve closer to the top left corner represents that it has a better performance and is more accurate when a good parameter value is chosen. When quantifying ROC curves, one of the common summary indices is the area under the curve (AUC), which is defined geometrically, as its name suggests. The AUC index can be interpreted as the overall probability of correct identification. In this work, we used the partial area under the curve (PAUC) [44] as the basis of the comparison between different ROC curves, because the curves were not long enough to separate the area under the curve and the area over the curve in practice. Only a specific region of the ROC curves was employed. Although the computed areas of AUC and PAUC are distinct, their concepts are similar that: an AUC/PAUC index with a larger area indicates a better achievement of the prediction method. Figure 4.3 gives an example of a ROC curve.

4.1.3 Comparisons between MVM and VBM

Simulation data were analyzed by both MVM and VBM (excluding the preprocessing part). Before the VBM analyses, all images were smoothed with a 4mm FWHM isotropic Gaussian kernel, and then we applied the two-sample t -test to obtain the t -test maps. In contract, MVM analyzed images without smoothing them. We used the wavelets proposed by Abdelnour and Selesnick [45] to transform the simulated images to the 3-level wavelet

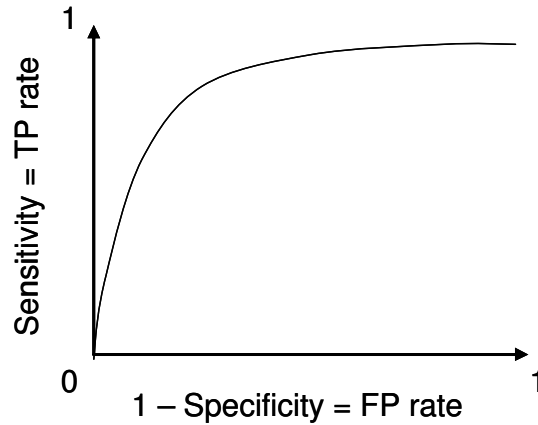


Figure 4.3: Example of a ROC curve. It is obtained by varying the value of the parameter of a method and then plotting sensitivity (TP rate) against 1-specificity (FP rate) under this parameter value. In general, methods with ROC curves closer to the top left corner indicate better performances. When wanting to compare performances of different methods, we often use the summary statistics for ROC curves, such as AUC and PAUC indices.

space, then preformed the discriminative common vector method, and set $\alpha = 20$ to preserve coefficients with larger absolute weights of the most discriminant projection vector and to discard other coefficients before the inverse wavelet transform. Finally, a 4mm FWHM isotropic Gaussian kernel was applied to the discriminating map for the MVM analysis result. Averagely, the MVM cost 51 seconds to analyze a set of data, and the VBM cost averagely 65 seconds. Both were on the same PC equipped Windows XP with a processor 1.83GHz and 1GB RAM.

The ROC curves of analyses results of MVM and VBM for detecting simulated group differences with the atrophied thickness of cerebellum from 1mm to 8mm are shown in Figure 4.4. The corresponding PAUC indices, in a specific region where the TP rate ranges form 0.8 to 1 and the FP rate ranges form 0 to 0.2, are listed in Table 4.2. By comparing ROC curves of two approaches in Figure 4.4, we found that the MVM curves are closer to the top left corner than VBM, and this appearance is obvious especially in the cases of 1mm and 2mm atrophy analyses. That is, the MVM produced a more accurate result than

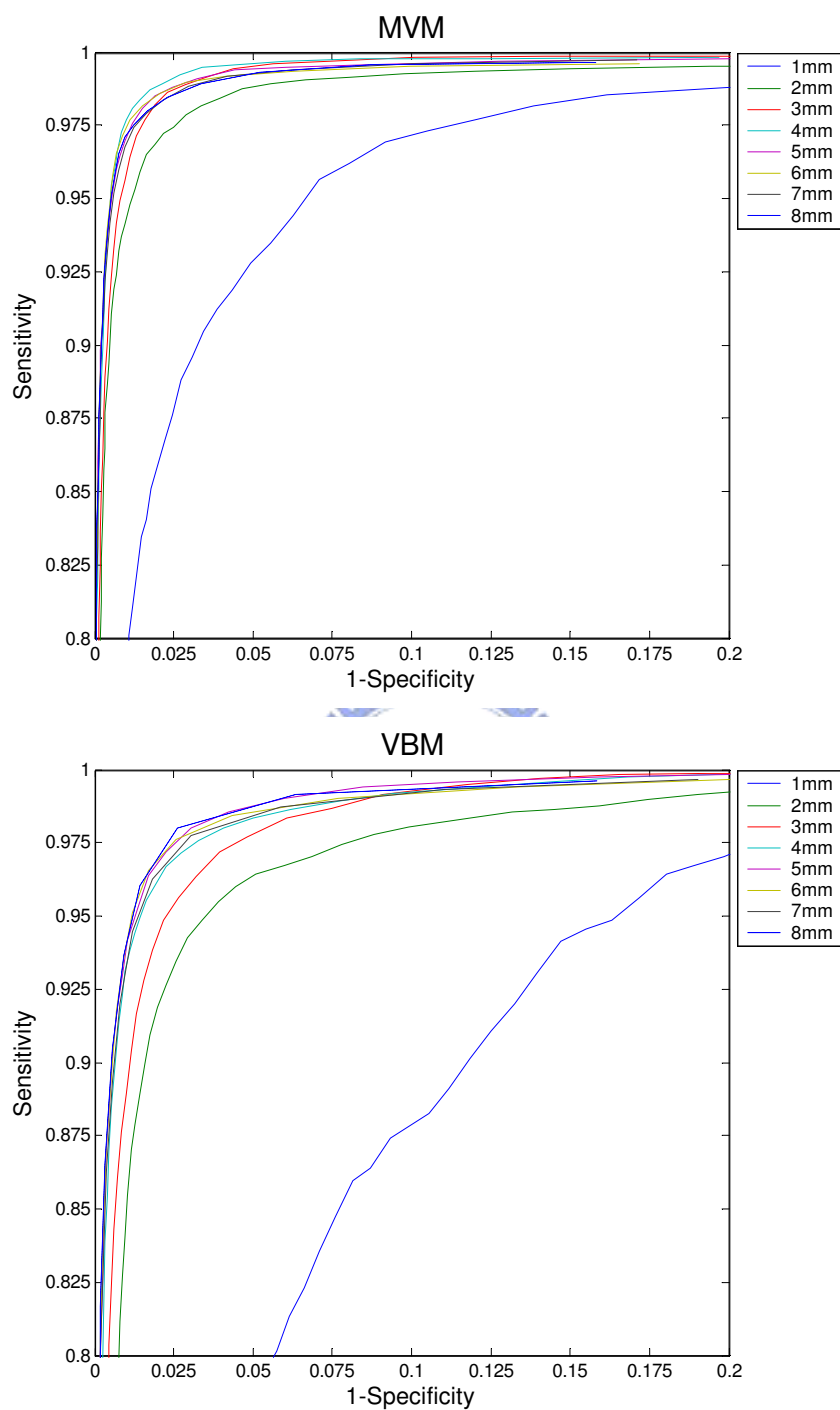


Figure 4.4: ROC curves of MVM and VBM results with the simulation data. Notice that both horizontal axes are limited from 0 to 0.2 and both vertical axes are limited from 0.8 to 1 for more specific display. It is manifest that MVM curves are closer to the top left corner than VBM in the same analysis data, especially when the simulated atrophy distance is small.

Table 4.2: PAUC indices for ROC curves of MVM and VBM results with the simulation data. The partial area was calculated in a specific region where the TP rate ranges from 0.8 to 1 and the FP rate ranges from 0 to 0.2. As this specific region denotes 1, we have $0 \leq \text{PAUC} \leq 1$. It is manifest that MVM PAUC indices are greater than VBM ones in the same analysis data, especially when the simulated atrophy distance is small.

Atrophy size	MVM	VBM
1mm	0.73466	0.38676
2mm	0.92286	0.81895
3mm	0.95521	0.88867
4mm	0.96613	0.91171
5mm	0.95836	0.92719
6mm	0.95398	0.91616
7mm	0.95465	0.91302
8mm	0.95357	0.92445

VBM when the atrophy was subtle and refined, as well as what we supposed. Besides, there is a general trend of the two results that, as the atrophy size increases, the ROC curve moves toward the top left corner. This trend is reasonable since large group differences are more easily detected by nature. But, the situation ends when the simulated atrophy size is bigger than 4mm. The curves almost go forward to the corner on longer but tend to fall into a pattern there. It might be evidence that the capability of two analysis methods for finding the group discrepancy will achieve an extremity when the group discrepancy is large enough. Similar findings were observed in their PAUC indices in Table 4.2. We could see that MVM have greater PAUC than VBM in the same analysis case, and the difference between MVM and VBM PAUC indices is large especially when the simulated atrophy is small. Both Figure 4.4 and Table 4.2 illustrate our method has a better sensitivity and accuracy than the VBM analysis.

To demonstrate the analysis results of MVM and VBM for detecting the cerebellar atrophy with the same FP rate, the result images are presented by displaying several slices, overlaying the same slices of the source image, with labeled TP, FP and FN regions. First we chose a reasonable t value of the VBM t -test map, and then selected the display threshold of the MVM discriminating map by the strategy that, this MVM threshold should make the FP rate of MVM analysis result equal to the FP rate of VBM. By comparing the result image with the ground truth of discrepancy pattern, each and every voxel of the result image was decided as a particular type of TP, TN, FP or FN. Finally these voxels were stained by different colors: green, yellow and red, that indicate the TP, FN and FP regions respectively.

In the case of 1mm cerebellar atrophy, the FP rate is approximate 0.0013 when choosing the threshold of VBM t value as 3.32 (i.e. p value = 0.001 uncorrected as degree of freedom = 38). The TP rates of MVM and VBM are approximate 0.4155 and 0.1632 respectively. Analysis results of MVM and VBM are shown in Figure 4.5 and Figure 4.6. From these two figures, we can see that there are more TP regions in the MVM result than in the VBM, as well as that their TP rates reveal. Moreover, a rough shape of the exterior cerebellum is shown in the MVM TP result in lower slices, but in the VBM result TP regions are too loose to form a shape. Besides, since the simulated atrophy was widely-distributed and was a very small quantity all over on the exterior surface of cerebellum, the FN regions of the two approaches are both large. It is instinctive that group difference is difficult to be detected when it is small. However, it is still visible that FN regions of MVM are fewer than ones of VBM. On the side, both results had some FP regions. Some were in the cerebrum, and some were around the cerebellum. Ones in the cerebrum were thought to be the false outcomes of the analyses, where should not have atrophy. But the appearance of FP regions around the cerebellum could be thought resulted from the smoothing, which is either the data smoothing in VBM or the display smoothing in MVM. This kind of FP false is not very taken to heart.

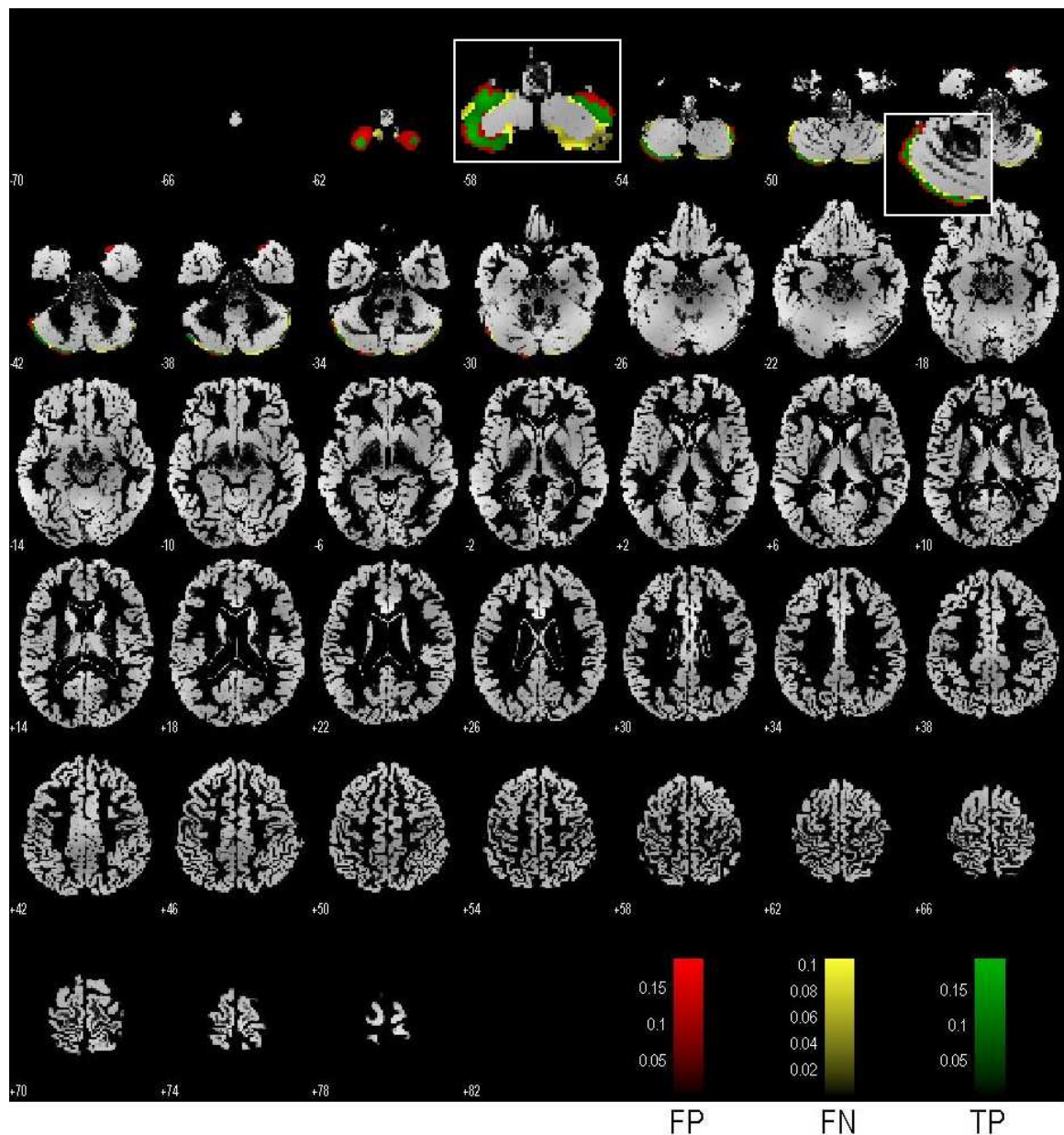


Figure 4.5: Analysis result of MVM for 1mm simulated cerebellum atrophy. The FP rate and TP rate are approximate 0.0013 and 0.4155 respectively. White-edge images are zoomed in for exquisite demonstration. Compared with the VBM result, the MVM had a better accuracy to detect the simulated atrophy around the cerebellum area, in the case of the 1mm atrophy simulation.

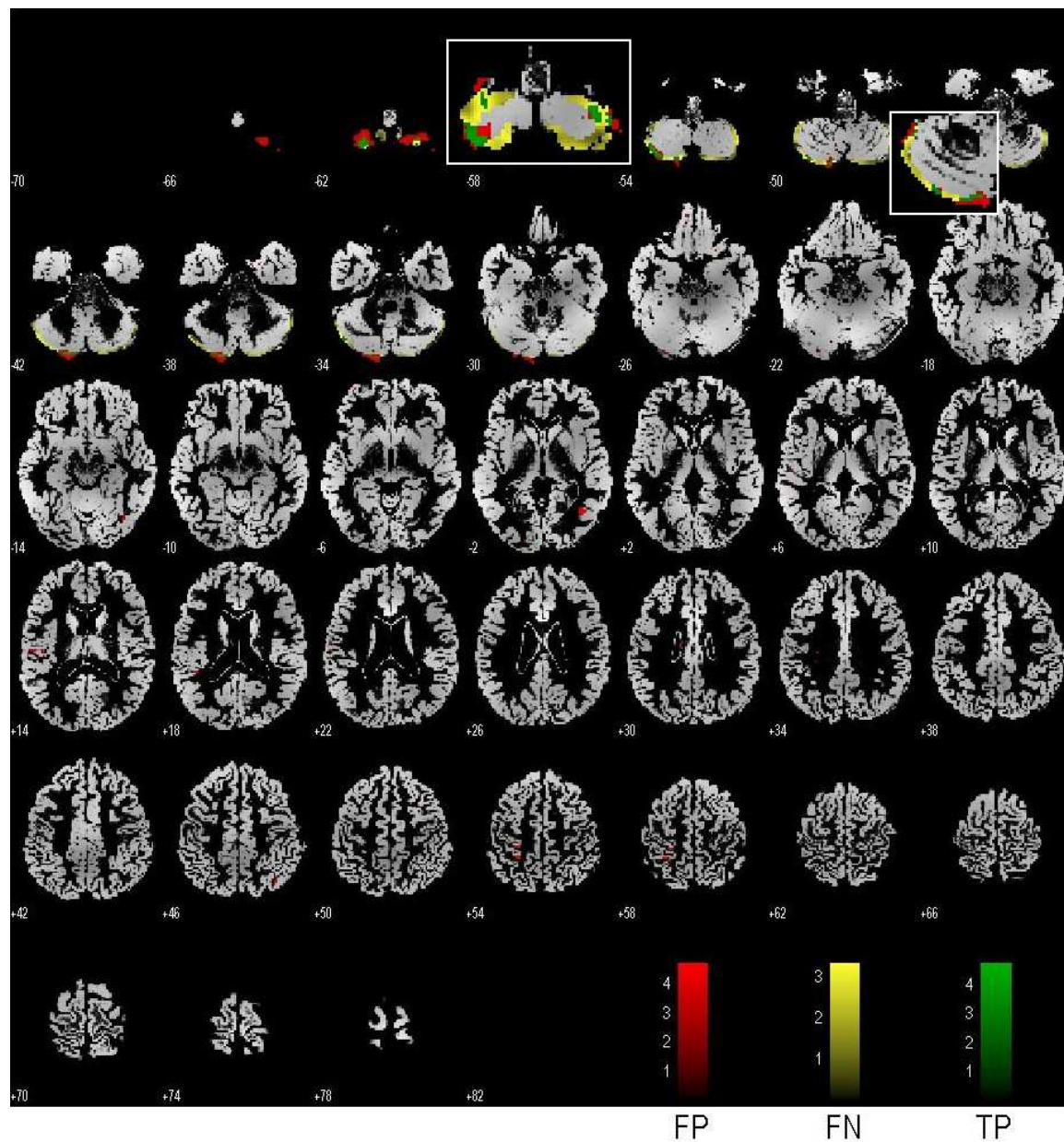


Figure 4.6: Analysis result of VBM for 1mm simulated cerebellum atrophy. The FP rate and TP rate are approximate 0.0013 and 0.1632 respectively. White-edge images are zoomed in for exquisite demonstration. Compared with the MVM result, the VBM had a worse accuracy to detect the simulated atrophy around the cerebellum area, in the case of the 1mm atrophy simulation.

Figure 4.7 and Figure 4.8 are similar overlays of analysis results of MVM and VBM for 2mm cerebellar atrophy. Both results have the same FP rate approximately 0.0021, when choosing the threshold of VBM t value as 4.12 (i.e. p value = 0.0001 uncorrected as degree of freedom = 38). The TP rates of MVM and VBM are approximate 0.8184 and 0.5061 respectively. Since the simulated atrophy size is only 2mm, an inference is similar to one of 1mm atrophy simulation: there are more TP regions and fewer FN regions in MVM than in VBM. It illustrates that, when atrophy is subtle and distributed, MVM can produce a more accurate result than VBM, once again. Furthermore, by comparing 2mm atrophy results with 1mm atrophy results, we found that amounts of TP regions increased and amounts of FN regions decreased in individual analysis approach. That means the analysis results with 2mm atrophy had a higher accuracy than 1mm ones. The same discovery was found in their ROC curves that Figure 4.4 displays.

In the analysis of simulation data with 6mm atrophy around the cerebellum, both MVM and VBM had good outcomes. Figure 4.9 and Figure 4.10 show the MVM and VBM results with the same FP rates approximate 0.0011, when choosing a **strict** threshold of VBM t value as 6.01 (i.e. p value = 0.1 FWE corrected as degree of freedom = 38) for the purpose of display. Under this strict threshold, TP rates of MVM and VBM are approximate 0.8498 and 0.7598 respectively. Both approaches found the majority of simulated atrophy regions correctly as TP parts shown in Figure 4.9 and Figure 4.10. Only a few atrophy regions were not detected, which occurred in the thin cortex or fringe of the cerebellum. And in these subtle regions, MVM still had a better ability to recognize the differences than VBM does.

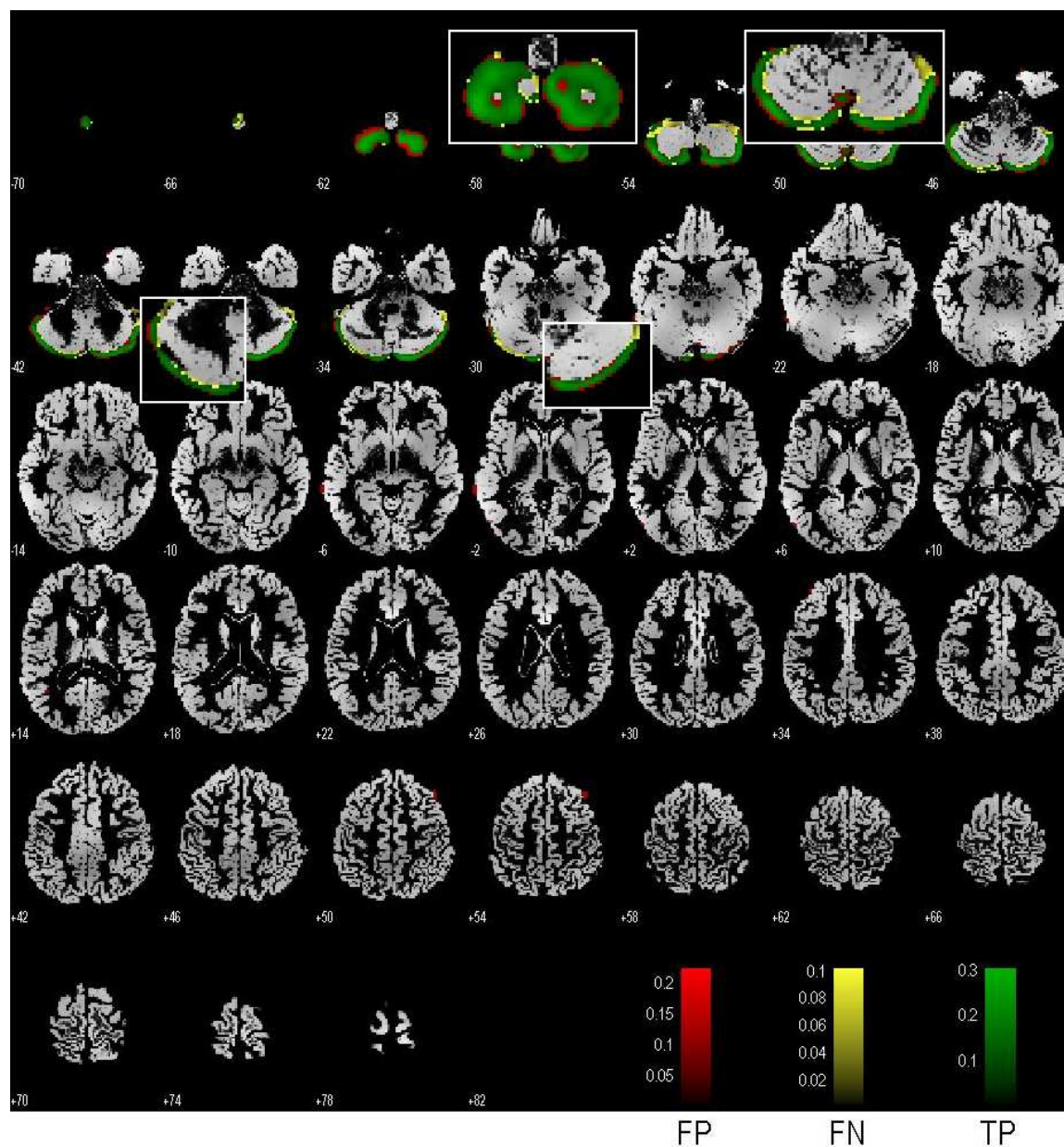


Figure 4.7: Analysis result of MVM for 2mm simulated cerebellum atrophy. The FP rate and TP rate are approximate 0.0021 and 0.8184 respectively. White-edge images are zoomed in for exquisite demonstration. Compared with the VBM result, the MVM had a better accuracy to detect the simulated atrophy around the cerebellum area, in the case of the 2mm atrophy simulation. Besides, it was also a more correct result than the result of 1mm atrophy. It is because the atrophy size became larger so that the atrophy area became more easily detected.

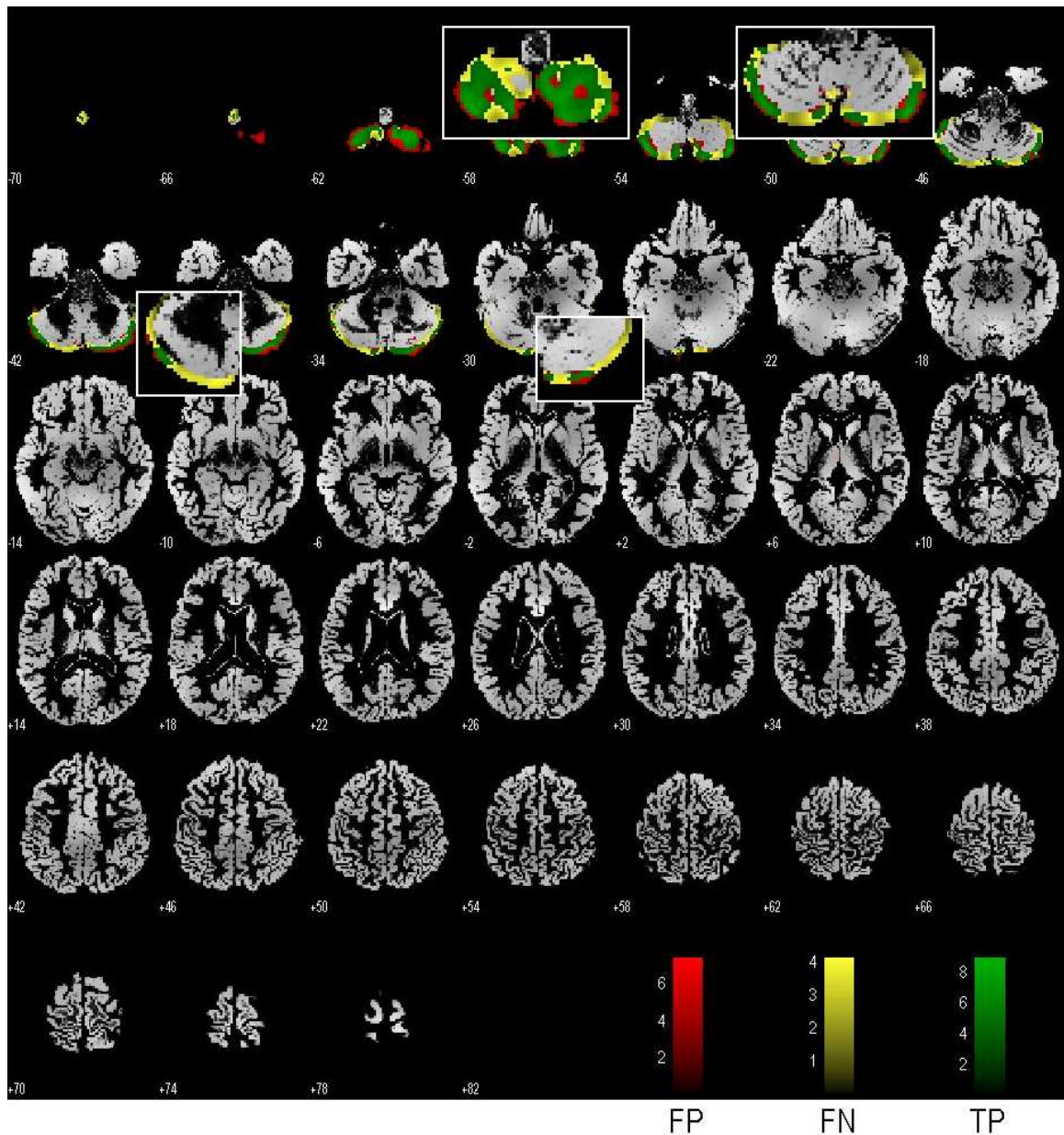


Figure 4.8: Analysis result of VBM for 2mm simulated cerebellum atrophy. The FP rate and TP rate are approximate 0.0021 and 0.5061 respectively. White-edge images are zoomed in for exquisite demonstration. Compared with the MVM result, the VBM had a worse accuracy to detect the simulated atrophy around the cerebellum area, in the case of the 2mm atrophy simulation. Nevertheless, it was also a more correct result than the result of 1mm atrophy. It is because the atrophy size became larger so that the atrophy area became more easily detected.

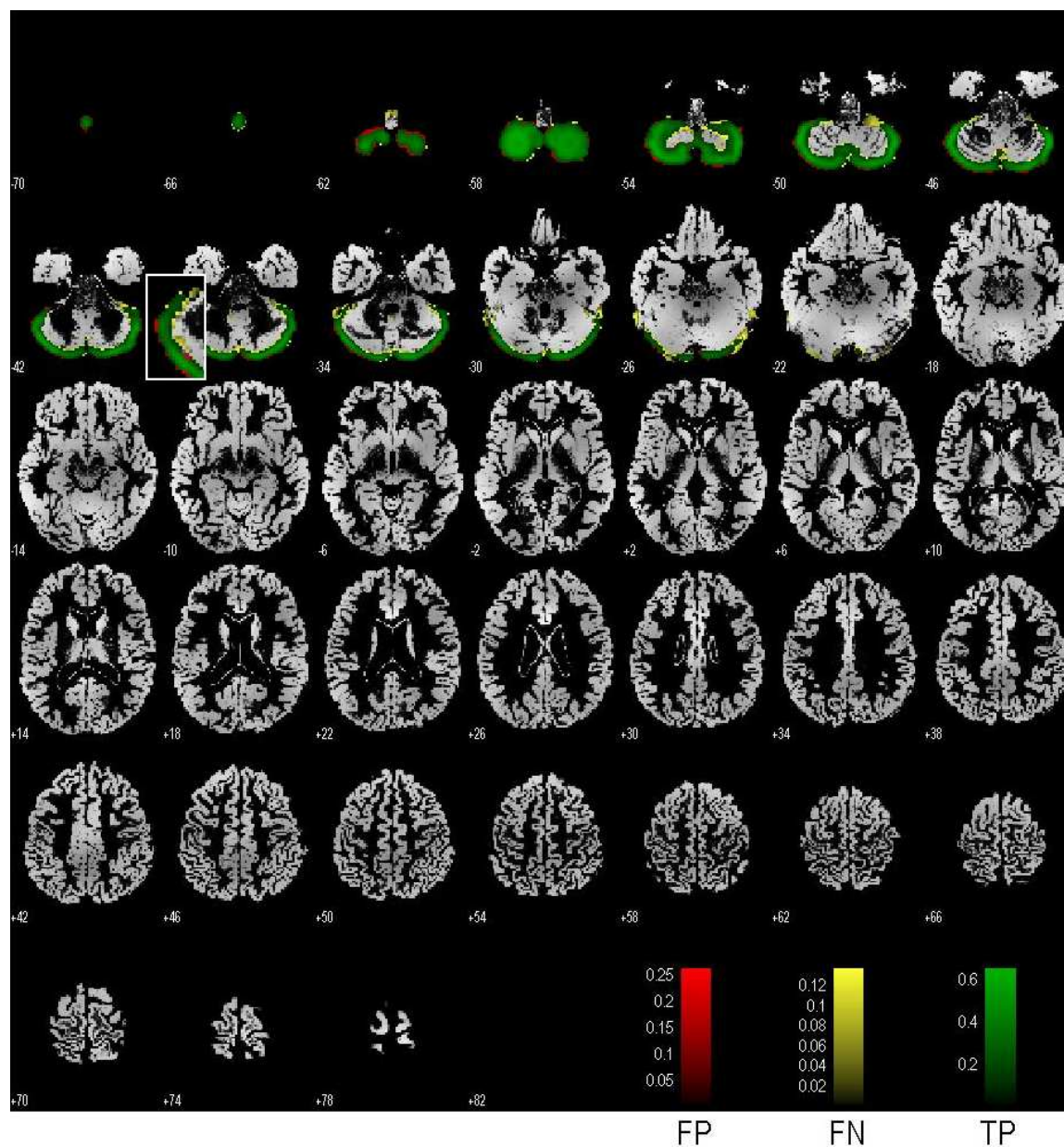


Figure 4.9: Analysis result of MVM for 6mm simulated cerebellum atrophy. The FP rate and TP rate are approximate 0.0011 and 0.8498 respectively. The white-edge image is zoomed in for exquisite demonstration. The majority of simulated atrophy regions were correctly detected because the atrophy size was large enough. Therefore there are only a few FN regions occurred in the thin cortex or fringe of the cerebellum. And, in these subtle regions, such as the white-edge image, MVM still had a better ability to recognize the differences than VBM does.

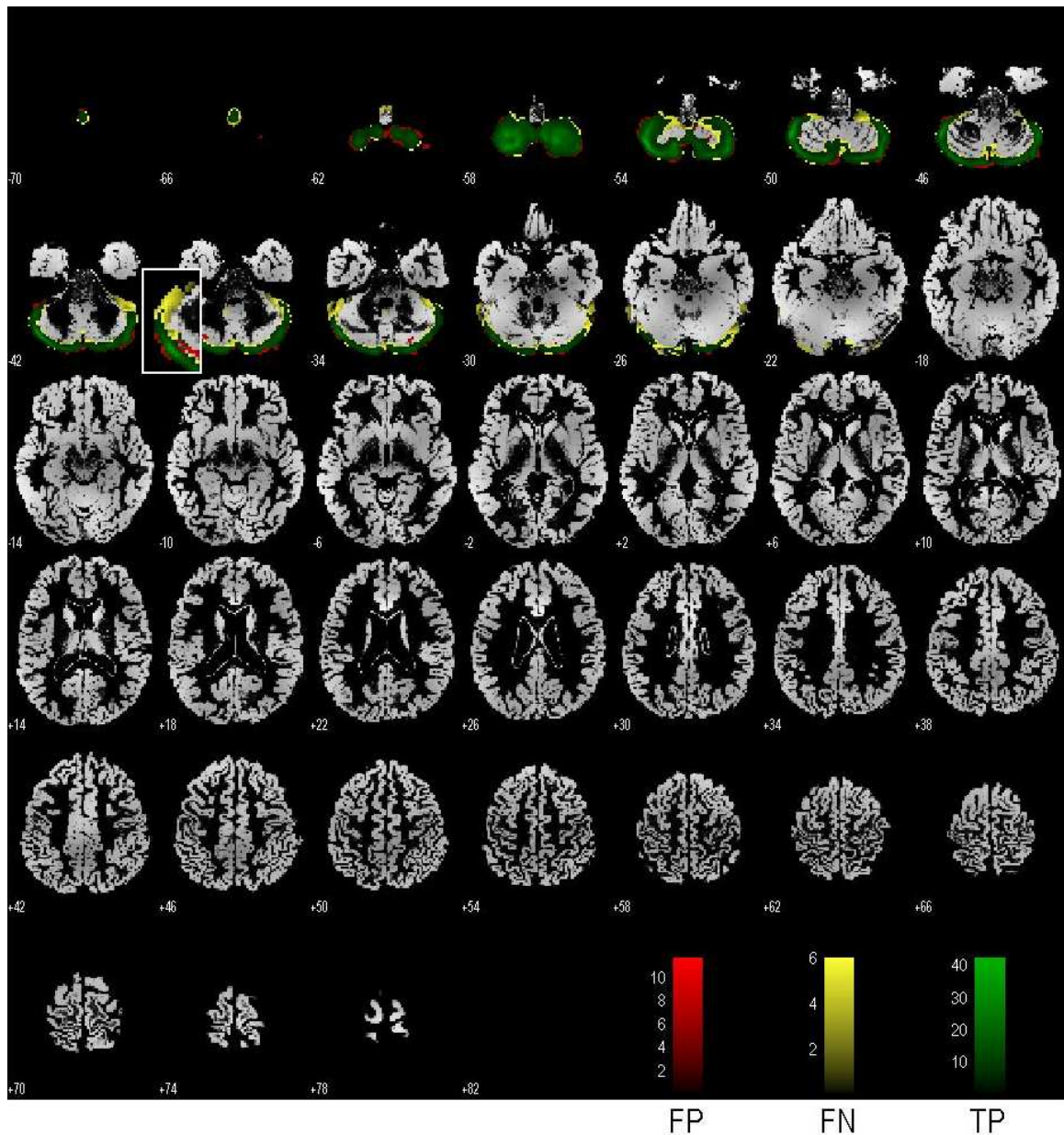


Figure 4.10: Analysis result of VBM for 6mm simulated cerebellum atrophy. The FP rate and TP rate are approximate 0.0011 and 0.7598 respectively. The white-edge image is zoomed in for exquisite demonstration. The majority of simulated atrophy regions were correctly detected because the atrophy size was large enough. Therefore there are only a few FN regions occurred in the thin cortex or fringe of the cerebellum. However, in these subtle regions, such as the white-edge image, VBM still had a worse ability to recognize the differences than MVM does.

4.2 Structural Atrophy Analysis for Patients Suffering Spinocerebellar Ataxia Type 3

Spinocerebellar ataxia (SCA) is a kind of inherited neurological disorders, whose clinical characteristics are the progressive neurodegeneration of the cerebellum, spinal cord, and brain stem. The symptoms of SCA include walking unsteadily, incongruity of limbs, dysarthria, hyperreflexia, and postural tremor, etc. The clinical behaviors of SCA patients are various, that the patients usually have different combination of different symptoms and different ages at falling ill. There is still no effective way to cure SCA completely now. It can only use medicine or physiatrics to alleviate the symptoms of SCA. Up to now, SCA is already classified into above 26 types. Most researches indicate that SCA type 3 is the most popular type of SCA in the world. In Taiwan, SCA3 is also the most epidemic type of spinocerebellar ataxia [46]. This experiment is accordingly to analyze the structural atrophy for the patients suffering SCA3, by applying two morphometric analysis methods, MVM and VBM. The analysis results and comparison of the two methods will be presented in the following.

4.2.1 Materials

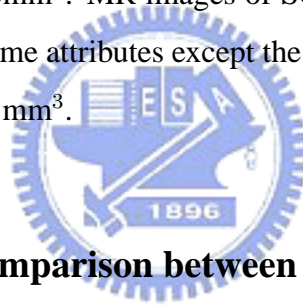
The study group was composed of six patients carrying spinocerebellar ataxia type 3 mutations and of eighteen normal subjects from the Taipei Veterans General Hospital. The clinical data of the patients are summarized in Table 4.3, where the international cooperative ataxia rating scale (ICARS) [47] is a pharmacological assessment of the cerebella syndrome, which involves quantification of postural and stance disorders, limb ataxia, dysarthria and oculomotor disorders.

Magnetic resonance images of all normal subjects were acquired from the same 1.5T Siemens scanner at the Taipei Veterans General Hospital, which used a T1-weighted 3-D

Table 4.3: Clinical data of patients carrying SCA3 under study.

Patient	Gender	Age	ICARS score
A	Male	27	19
B	Female	31	19
C	Male	37	22
D	Female	54	19
E	Male	57	16
F	Female	60	37

IR sequence with TR = 9.7ms, TE = 4ms, FA = 12°, matrix size = 256 × 256, slices = 128, voxel size = 0.9 × 0.9 × 1.5mm³. MR images of SCA3 patients were also acquired from the same scanner with the same attributes except the matrix size = 512 × 512, slices = 160, voxel size = 0.47 × 0.47 × 1mm³.



4.2.2 Results and Comparison between MVM and VBM

We used the first six steps of the optimized VBM including modulation to preprocess all MR images by the software SPM2, and then statistically analyzed structural brain differences between the normal and patient groups by both MVM and VBM methods. In the preprocessing stage, the nonlinear spatial normalization part was performed by 7 × 9 × 7 discrete cosine basis functions, and the output normalized images had a high resolution with the voxel size of 1 × 1 × 1mm³. Again, before the VBM analysis, all images were smoothed with an 8mm FWHM isotropic Gaussian kernel. Then we applied a two-sample *t*-test, and set the significance level at $p < 0.00005$ uncorrected (i.e. *t* value > 4.74 as degree of freedom = 22) for the final *t*-test map. Parallel to VBM, MVM analyzed those images without smoothing them. We used the same wavelets in the simulation experiment to transform images to the 3-level wavelet space, and set $\alpha = 20$ to discard coefficients

with smaller weights of discrimination. After obtaining the discriminating map, an 8mm FWHM isotropic Gaussian kernel was used for the analysis result display. The display threshold of MVM was the mean of discrimination weights of some regions corresponding to those regions in the same locations of VBM t -test map whose t values were above 4.74 and below 5.04. It made the MVM result was not far off the result of VBM. This strategy for choosing a compatible threshold of MVM is illustrated with Figure 4.11. To restrict the minimum cluster size when showing analysis results, both extension thresholds of the two methods were set for 40 voxels. The p -value representing $1 -$ the whole-brain confidence of MVM discriminating map was about 5.83×10^{-10} . It took 335 and 403 seconds in the MVM and VBM analysis step (excluding the preprocessing) respectively, on the same PC equipped Windows XP with a processor 1.83GHz and 1GB RAM.

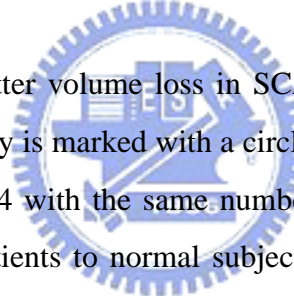


Figure 4.12 illustrates gray matter volume loss in SCA3 patients by MVM analysis method. Detected significant atrophy is marked with a circle and a number in Figure 4.12, furthermore, it is listed in Table 4.4 with the same number in the no. column. Comparing brain structure of SCA3 patients to normal subjects, volume loss of gray matter was detected in the regions of anterior and posterior lobes of cerebellar hemispheres, vermis, basal ganglia (including lentiform nucleus, caudate nucleus, thalamus), frontal areas (including inferior, middle and superior frontal gyri), right occipital lobe, and left middle temporal area. Most of these observed regions are symmetrical in both sides. The MVM analysis result accentuated the atrophy of cerebellar area, and it is a fine outcome because SCA is characterized as the neurodegeneration of the cerebellum. Basal ganglia, which is associated with motor and learning functions, is also noticeable and reasonable finding, as a result of that it is a main abnormal region for Parkinson's symptom involved with SCA3. Frontal areas are known about the cognitive functions that may be influenced by the cerebellar dysfunction and result in frontal-executive deficits [48], so those observed frontal areas were thought as a reasonable result. The atrophy of middle temporal area was also reported in [26]. Since the frontal and temporal cortex were both the target organ of

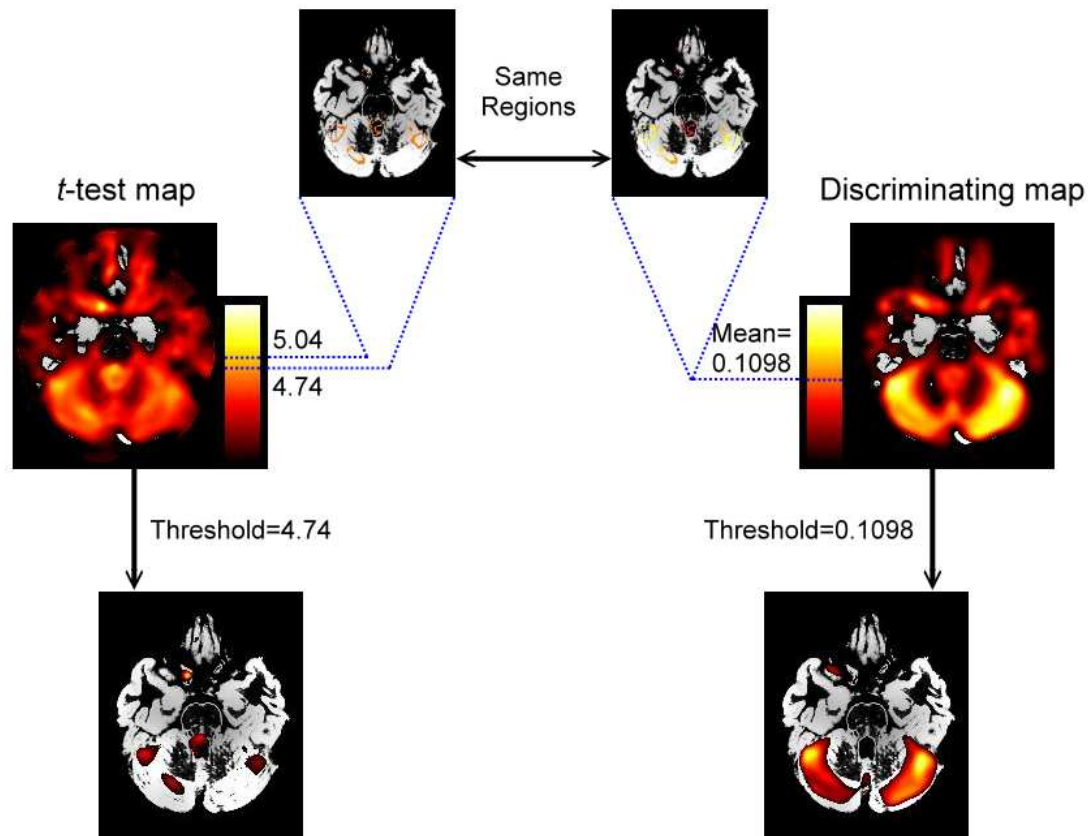


Figure 4.11: Strategy for choosing a compatible threshold of MVM by VBM t values. First we decided the threshold of t value in VBM, then found regions whose t values are above 4.74 and below 5.04, and finally chosen mean of discrimination weights of MVM discriminating map in the same regions as the MVM threshold.

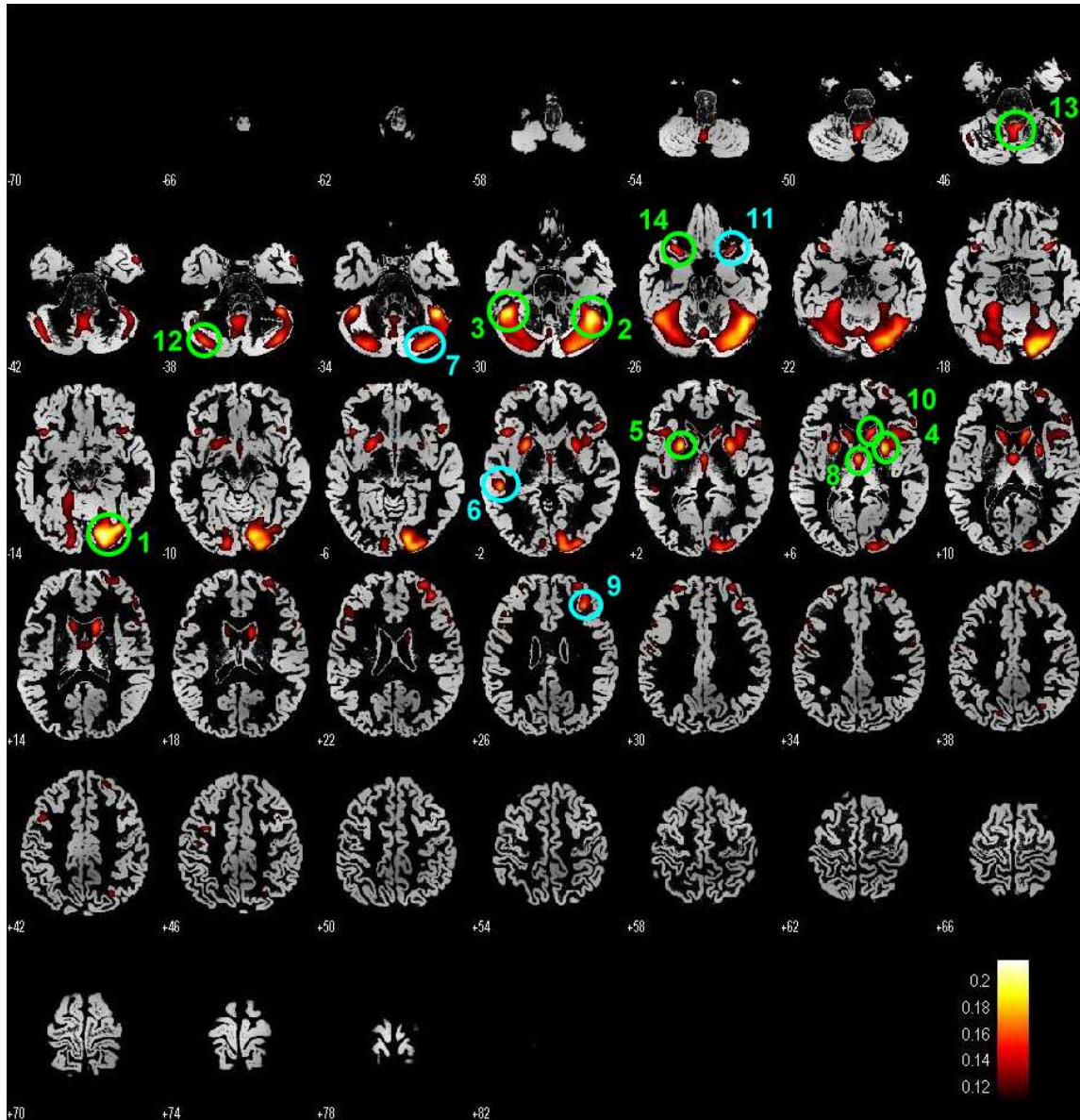


Figure 4.12: Volumetric atrophy of gray matter in SCA3 patients by MVM analysis method. Regions detected as significant group differences are marked with circles. One marked with a green circle is also found in the VBM analysis result; one marked with a blue circle is only revealed in the MVM result. Numbers of these circles represent the importance of characterizing volume loss in SCA3 patients. Specifically, smaller number represents higher importance and larger number represents lower importance. More information of each region is listed in Table 4.4.

Table 4.4: Atrophy of gray matter in SCA3 patients by MVM analysis method. These atrophy regions are arranged in order by their peak discrimination weights of each cluster, and the numbers denote the sequence of peak discrimination weights. Some clusters are very large so that there is more than one peak listed in a cluster. The serial numbers are informative for revealing the significance of gray matter atrophy in SCA3.

Location	Side	Talairach coordinate			Peak discrimination	Cluster size	Fig. 4.12 no.
		<i>x</i>	<i>y</i>	<i>z</i>			
Occipital Lobe, Lingual Gyrus	R	19	-83	-8	0.2157	74086	1
Cerebellum, Anterior Lobe	R	42	-53	-23	0.2089		2
Cerebellum, Anterior Lobe	L	-44	-48	-23	0.2021		3
Cerebellum, Posterior Lobe	R	26	-81	-23	0.1760		7
Cerebellum, Posterior Lobe	L	-36	-75	-28	0.1638		12
Cerebellum, Vermis	R	4	-48	-37	0.1618		13
Sub-lobar, Lentiform Nucleus	R	26	5	4	0.1946	11083	4
Frontal Lobe, Inferior Frontal Gyrus (BA47)	R	29	15	-21	0.1652		11
Frontal Lobe, Inferior Frontal Gyrus	R	50	25	-4	0.1534		
Sub-lobar, Lentiform Nucleus	L	-24	7	1	0.1889	9333	5
Temporal Lobe, Superior Temporal Gyrus	L	-29	15	-23	0.1585		14
Frontal Lobe, Inferior Frontal Gyrus	L	-42	19	-3	0.1554		
Temporal Lobe, Middle Temporal Gyrus	L	-50	-33	0	0.1772	982	6
Thalamus		0	-6	6	0.1755	3137	8
Frontal Lobe, Middle Frontal Gyrus	R	35	34	23	0.1704	6205	9
Sub-lobar, Caudate Nucleus	R	13	20	6	0.1658	2777	10
Temporal Lobe, Middle Temporal Gyrus	R	51	5	-34	0.1508	632	
Frontal Lobe, Middle Frontal Gyrus	L	-33	-1	43	0.1387	270	
Frontal Lobe, Superior Frontal Gyrus (BA10)	L	-28	62	-6	0.1387	195	
Frontal Lobe, Superior Frontal Gyrus	L	-27	50	26	0.1380	870	
Frontal Lobe, Middle Frontal Gyrus (BA9)	L	-41	10	37	0.1378	1157	

the cerebellar efferent tracts, the atrophy of the frontal and temporal regions in our study may be secondary to the degeneration of the cerebellum. Besides, a significant atrophy was found in the occipital lobe, owing to some existing clinical vision problem in our patients accordingly.

Figure 4.13 illustrates detected gray matter volumetric atrophy in SCA3 patients by VBM analysis method. Observed significant discrepancy is marked with a circle and a number in Figure 4.13, furthermore, it is listed in Table 4.5 with the same number in the no. column. Comparing brain structure of SCA3 patients to normal subjects, volume loss was detected in the regions of the basal ganglia (including lentiform nucleus and caudate nucleus), vermis, anterior lobe of cerebellar hemisphere, frontal areas (including rectal, inferior, superior, and precentral frontal gyri), and right occipital lobe. Some reasonable atrophy regions that have been described in last paragraph are also detected by the VBM approach, such as the cerebellum, basal ganglia, and areas in the frontal and occipital lobes. However, there is a strange finding in the lateral ventricle. It may result from an incorrect orientation around the thalamus.

By comparing the two analysis results of MVM and VBM to each other, we found MVM provided a better outcome in catching the structural atrophy patterns of SCA3 than VBM. The most apparent distinction is about the cerebellum, which is mildly affected in neuropathological findings but distributes widely. MVM not only detected the posterior cerebellar area that could not be found in VBM, but also detected the atrophy area in the whole cerebellum with a larger degree of differences than other atrophy regions in the brain. By the serial number of cerebellar areas in Figure 4.12 and Figure 4.13, it is illustrated that MVM more highlighted the atrophy of the cerebellum than VBM does. More specifically, when characterizing the brain structural differences between the SCA patients and normal subjects, the cerebellum is an important area for MVM. But VBM indicates other regions first, such as the basal ganglia and frontal lobe, and then mentions the cerebellum. Such the analysis results may result from the situation that, there is not an obvious difference of

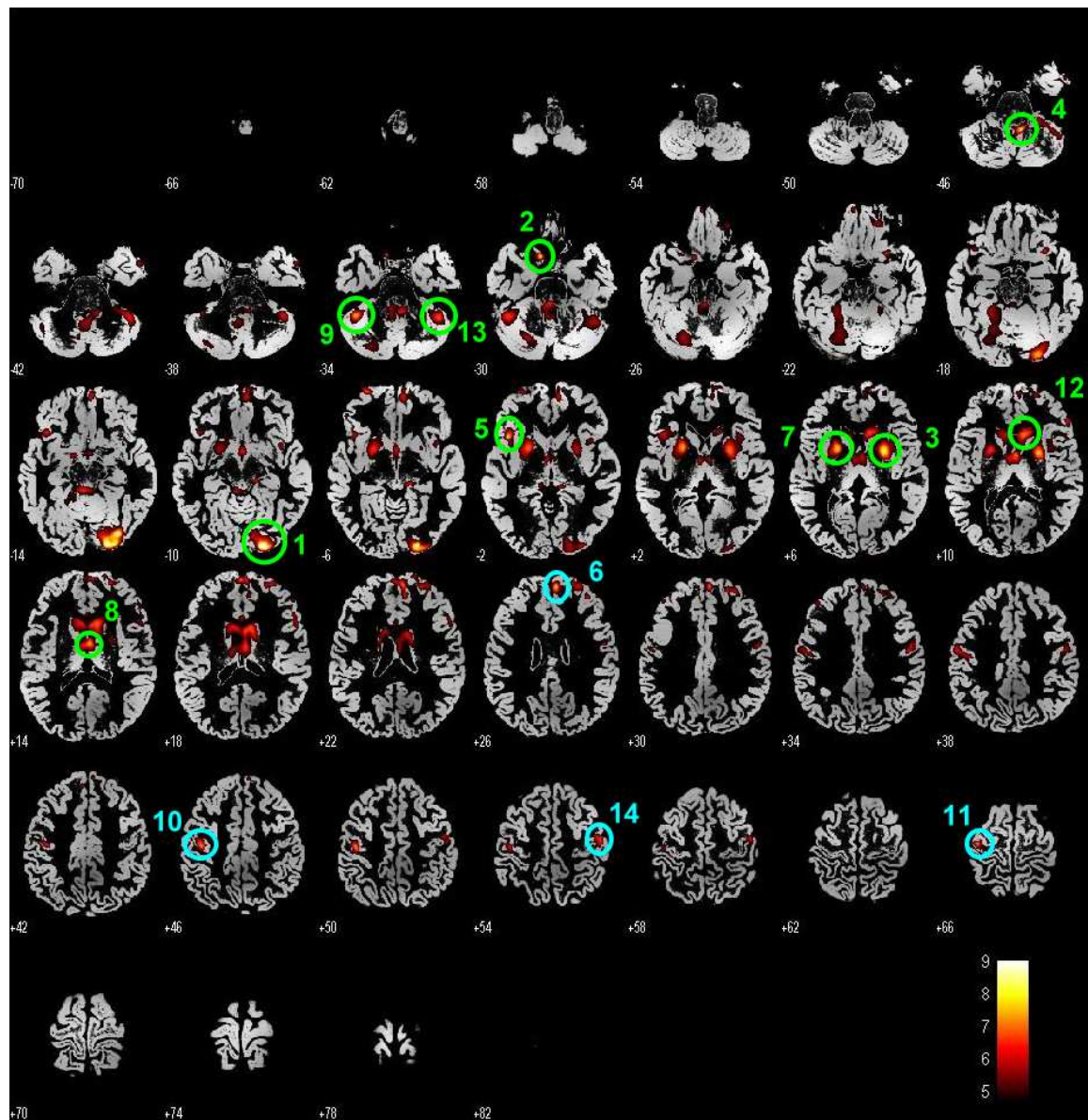


Figure 4.13: Volumetric atrophy of gray matter in SCA3 patients by VBM analysis method. Regions detected as significant group differences are marked with circles. One marked with a green circle is also found in the MVM analysis result; one marked with a blue circle is only revealed in the VBM result. Numbers of these circles represent the importance of characterizing volume loss in SCA3 patients. Specifically, smaller number represents higher importance and larger number represents lower importance. More information of each region is listed in Table 4.5.

Table 4.5: Atrophy of gray matter in SCA3 patients by VBM analysis method. These atrophy regions are arranged in order by their peak t values of each cluster, and the numbers denote the sequence of peak t values. Some clusters are very large so that there is more than one peak listed in a cluster. The serial numbers are informative for revealing the significance of gray matter atrophy in SCA3.

Location	Side	Talairach coordinate			Peak t value	Cluster size	Fig. 4.13 no.
		x	y	z			
Occipital Lobe, Lingual Gyrus	R	20	-90	-4	10.553	7196	1
Frontal Lobe, Rectal Gyrus	L	-10	11	-27	8.356	422	2
Sub-lobar, Lentiform Nucleus	R	25	3	6	8.226	5665	3
Cerebellum, Vermis	R	9	-47	-35	8.066	13243	4
Brainstem, Midbrain	L	-5	-35	-8	6.582		
Frontal Lobe, Inferior Frontal Gyrus	L	-42	17	-3	7.876	2282	5
Frontal Lobe, Superior Frontal Gyrus (BA9)	R	5	55	22	7.696	8032	6
Inter-Hemispheric		10	59	15	6.805		
Frontal Lobe, Medial Frontal Gyrus	R	8	57	-6	6.386		
Sub-lobar, Lentiform Nucleus	L	-21	7	4	7.634	5689	7
Sub-lobar, Lateral Ventricle	R	2	-1	12	7.560	14168	8
Sub-lobar, Caudate Nucleus	R	10	20	7	7.217		12
Cerebellum, Anterior Lobe	L	-41	-47	-25	7.557	1783	9
Frontal Lobe, Precentral Gyrus (BA4)	L	-40	-17	45	7.439	2579	10
Frontal Lobe, Precentral Gyrus	L	-36	-14	62	7.255	314	11
Cerebellum, Anterior Lobe	R	44	-49	-27	6.935	3981	13
Frontal Lobe, Precentral Gyrus	R	50	-11	50	6.847	976	14
Frontal Lobe, Superior Frontal Gyrus	L	-29	62	-7	6.429	469	
Frontal Lobe, Middle Frontal Gyrus	L	-42	39	28	6.317	159	
Brainstem, Midbrain	R	15	-27	-6	6.297	593	

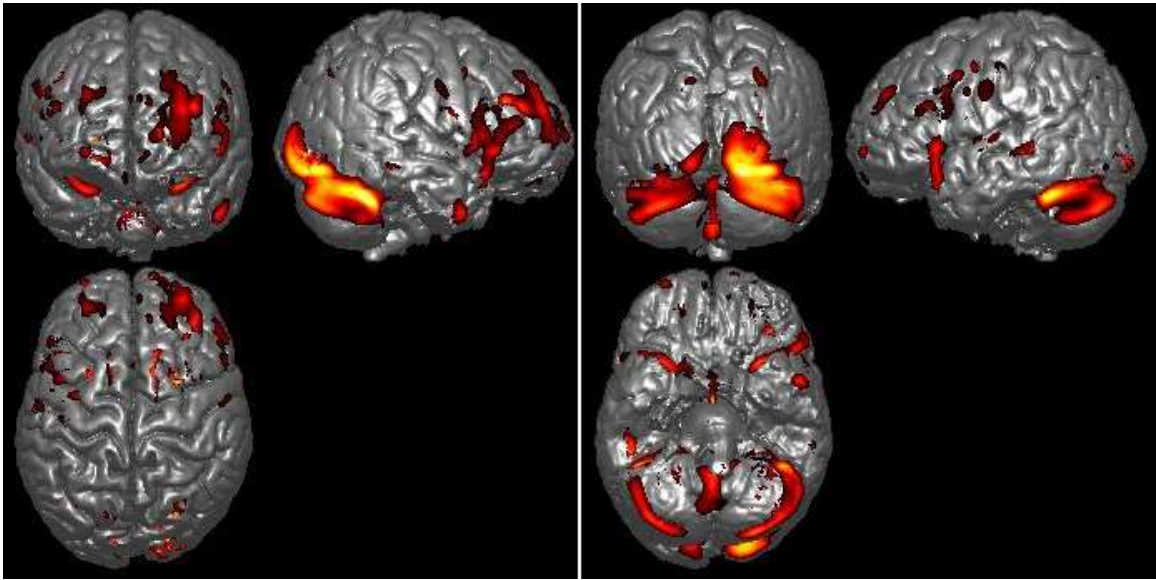


Figure 4.14: 3-D rendering of GM atrophy in SCA3 by the MVM analysis method. The detected atrophy regions showed in this figure include the cerebellum, vermis, right occipital lobe, left middle temporal lobe, and inferior, middle and superior frontal gyri.

brain tissue at each voxel in the cerebellum, and it had been verified by manual comparison between the MR images of SCA3 patients and the normal subjects. However, the MVM analysis can collect the information from the adjacent voxels to estimate the differences, so the widely-distributed atrophy in the cerebellum can be detected strongly. To other piecemeal and local regions with differences, each of the performances of MVM and VBM has its merits. Table 4.6 put the detected GM atrophy regions of MVM and VBM in SCA3 patients together in an order of anatomical location of brain. Figure 4.14 and Figure 4.15 are the MVM and VBM analysis results presented by the 3-D rendering, respectively.

In the white matter analysis, we used the same parameters except $p < 0.001$ uncorrected (i.e. t value > 3.505 as degree of freedom = 22) for the VBM t -test map and the corresponding MVM display threshold 0.0915 for the discriminating map with a whole-brain significance p -value about 2.35×10^{-7} . Figure 4.16 and Table 4.7 illustrate the observed white matter volumetric atrophy in SCA3 patients by MVM analysis method. Volume

Table 4.6: Detected GM atrophy of MVM and VBM in SCA3 patients. Anatomical location, side, Talairach's coordinates (x, y, z), and peak of the discrimination weights or t values are listed for each of detected atrophy regions.

Location	Side	MVM				VBM			
		x	y	z	Discrimination	x	y	z	t value
Cerebellum									
Anterior Lobe	R	42	-53	-23	0.2089	44	-49	-27	6.935
	L	-44	-48	-23	0.2021	-41	-47	-25	7.557
Posterior Lobe	R	26	-81	-23	0.1760				
	L	-36	-75	-28	0.1638				
Vermis		4	-48	-37	0.1618	9	-47	-35	8.066
Basal Ganglia									
Lentiform Nucleus	R	26	5	4	0.1946	25	3	6	8.226
	L	-24	7	1	0.1889	-21	7	4	7.634
Caudate Nucleus	R	13	20	6	0.1658	10	20	7	7.217
Thalamus		0	-6	6	0.1755				
Lateral Ventricle	R					2	-1	12	7.560
Midbrain	R					15	-27	-6	6.297
	L					-5	-35	-8	6.582
Occipital Lobe	R	19	-83	-8	0.2157	20	-90	-4	10.553
Frontal Lobe									
Superior Frontal Gyrus	R					5	55	22	7.696
	L	-28	62	-6	0.1387	-29	62	-7	6.429
Middle Frontal Gyrus	R	35	34	23	0.1704	8	57	-6	6.386
	L	-41	10	37	0.1378	-42	39	28	6.317
Inferior Frontal Gyrus	R	29	15	-21	0.1652				
	L	-42	19	-3	0.1554	-42	17	-3	7.876
Precentral Gyrus	R					50	-11	50	6.847
	L					-40	-17	45	7.43
Rectal Gyrus	L					-10	11	-27	8.356
Middle Temporal Gyrus	L	-50	-33	0	0.1772				

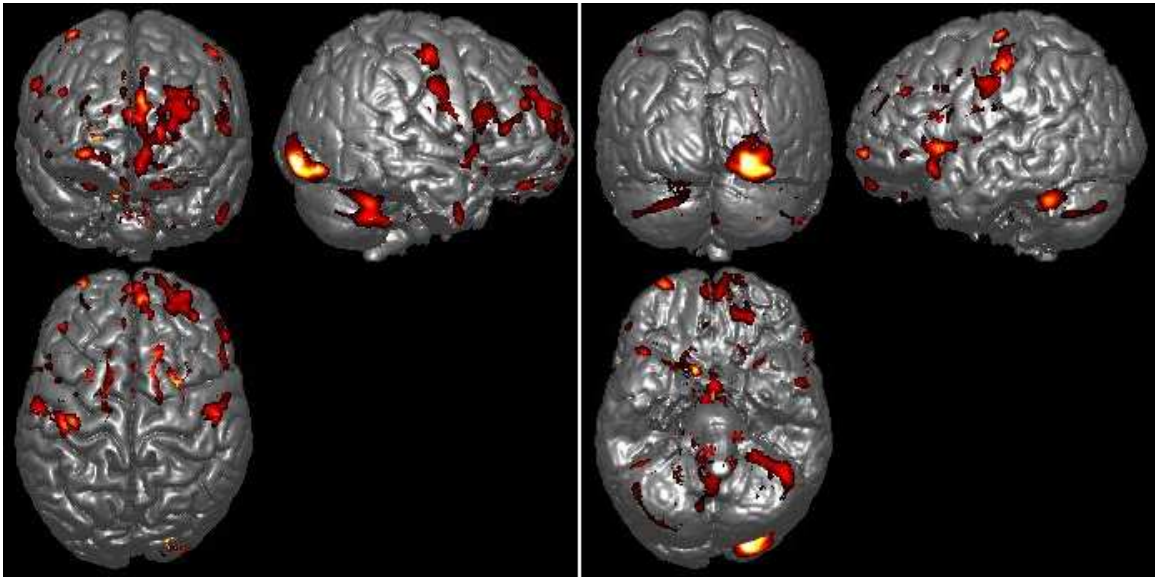


Figure 4.15: 3-D rendering of GM atrophy in SCA3 by the VBM analysis method. The detected atrophy regions showed in this figure include the cerebellum, right occipital lobe, and rectal, inferior, superior, and precentral frontal gyri.

Table 4.7: Atrophy of white matter in SCA3 patients by MVM analysis method. These atrophy regions are arranged in order by their peak discrimination weights of each cluster, and the numbers denote the sequence of peak discrimination weights. The serial numbers are informative for revealing the significance of white matter atrophy in SCA3.

Location	Side	Talairach coordinate			Peak discrimination	Cluster size	Fig. 4.12 no.
		<i>x</i>	<i>y</i>	<i>z</i>			
Brainstem, Pons		3	-17	-35	0.2596	25410	1
Cerebellum, Anterior Lobe	L	-19	-52	-26	0.1602		3
Cerebellum, Anterior Lobe	R	26	-48	-29	0.1433		6
Corpus Callosum		1	11	16	0.1942	6860	2
Frontal Lobe, Inferior Frontal Gyrus	R	36	36	11	0.1590	2333	4
Sub-lobar, Lentiform Nucleus	L	-14	-2	-1	0.1552	2144	5
Sub-lobar, Lentiform Nucleus	R	16	-1	0	0.1422	1449	7

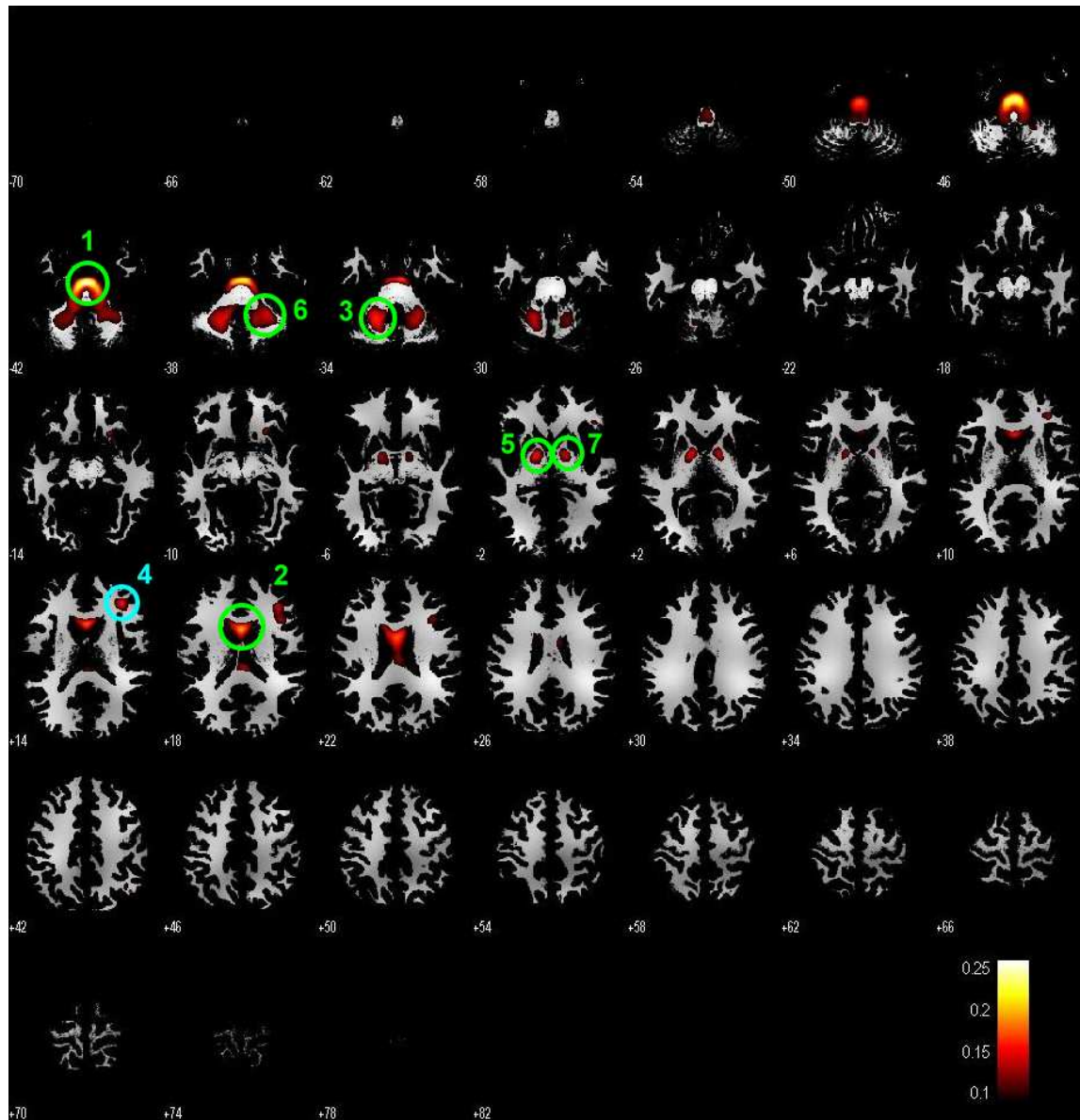


Figure 4.16: Volumetric atrophy of white matter in SCA3 patients by MVM analysis method. Regions detected as significant group differences are marked with circles. One marked with a green circle is also found in the VBM analysis result; one marked with a blue circle is only revealed in the MVM result. Numbers of these circles represent the importance of characterizing volume loss in SCA3 patients. Specifically, smaller number represents higher importance and larger number represents lower importance. More information of each region is listed in Table 4.7.

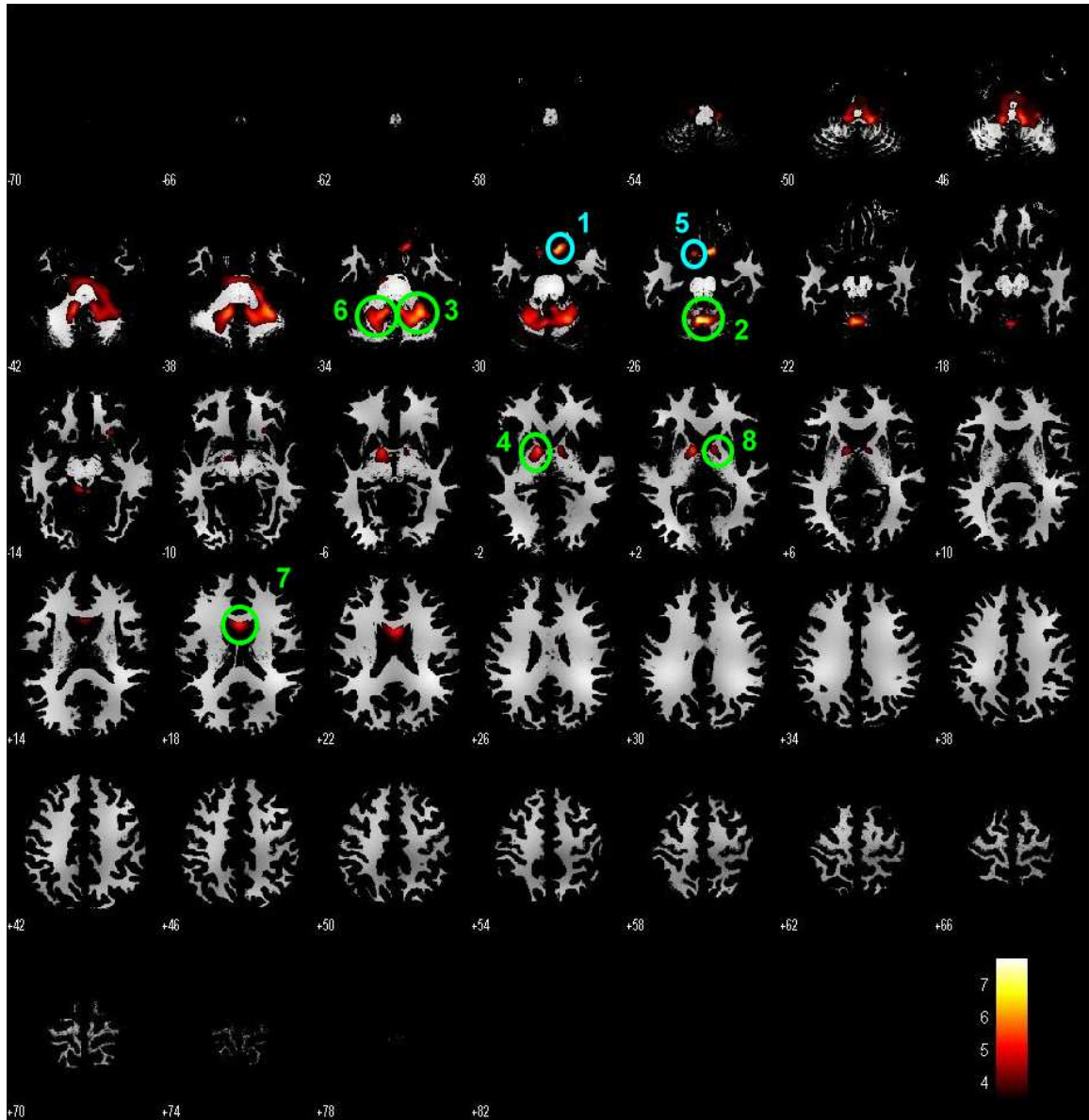


Figure 4.17: Volumetric atrophy of white matter in SCA3 patients by VBM analysis method. Regions detected as significant group differences are marked with circles. One marked with a green circle is also found in the MVM analysis result; one marked with a blue circle is only revealed in the VBM result. Numbers of these circles represent the importance of characterizing volume loss in SCA3 patients. Specifically, smaller number represents higher importance and larger number represents lower importance. More information of each region is listed in Table 4.8.

Table 4.8: Atrophy of white matter in SCA3 patients by VBM analysis method. These atrophy regions are arranged in order by their peak t values of each cluster, and the numbers denote the sequence of peak t values. The serial numbers are informative for revealing the significance of white matter atrophy in SCA3.

Location	Side	Talairach coordinate			Peak t value	Cluster size	Fig. 4.13 no.
		x	y	z			
Frontal Lobe, Rectal Gyrus	R	9	13	-25	7.806	1456	1
Cerebellum, Anterior Lobe		-3	-54	-19	7.685	30298	2
Cerebellum, Anterior Lobe	R	29	-47	-28	6.416		3
Cerebellum, Anterior Lobe	L	-20	-55	-24	5.576		6
Sub-lobar, Lentiform Nucleus	L	-10	2	-1	5.885	2651	4
Frontal Lobe, Rectal Gyrus	L	-9	10	-23	5.770	343	5
Corpus Callosum		2	12	18	5.487	2761	7
Sub-lobar, Lentiform Nucleus	R	11	3	1	4.569	1211	8

losses were in the pons, corpus callosum, anterior cerebellum, basal ganglia (lentiform nucleus) and inferior frontal gyrus. In contrast, the detected white matter volumetric atrophy by VBM analysis method are shown in Figure 4.17 and Table 4.8, and they are in the regions of the anterior cerebellum, basal ganglia (lentiform nucleus), corpus callosum, and frontal areas (including rectal and inferior frontal gyri). Two analysis results revealed similar WM atrophy regions. But it could be still caught clear sight of that MVM has a better detection of the pons atrophy, which takes part in the main neurodegeneration in spinocerebellar ataxia (SCA). Table 4.9 put the detected WM atrophy regions of MVM and VBM in SCA3 patients together in some order of anatomical locations of the brain.

Figure 4.18 and Figure 4.19 are the volumetric increase of CSF in SCA3 patients in the MVM and VBM analysis methods, respectively. Since CSF flows through the whole acquisition of MR image, the analysis of CSF has less validity than GM or WM. In our use, the analysis result of CSF is for testing and verifying the analysis results of GM and WM. By

Table 4.9: Detected WM atrophy of MVM and VBM in SCA3 patients. Anatomical location, side, Talairach's coordinates (x, y, z) , and peak of the discrimination weights or t values are listed for each of detected atrophy regions.

Location	Side	MVM			Discrimination	VBM			t value
		x	y	z		x	y	z	
Brainstem, Pons		3	-17	-35	0.2596				
Cerebellum									
Anterior Lobe	R	26	-48	-29	0.1433	29	-47	-28	6.416
	L	-19	-52	-26	0.1602	-20	-55	-24	5.576
Basal Ganglia									
Lentiform Nucleus	R	16	-1	0	0.1422	11	3	1	4.569
	L	-14	-2	-1	0.1552	-20	-55	-24	5.576
Corpus Callosum		1	11	16	0.1942	2	12	18	5.487
Frontal Lobe									
Inferior Frontal Gyrus	R	36	36	11	0.1590				
Rectal Gyrus	R					9	13	-25	7.806
	L					-9	10	-23	5.770

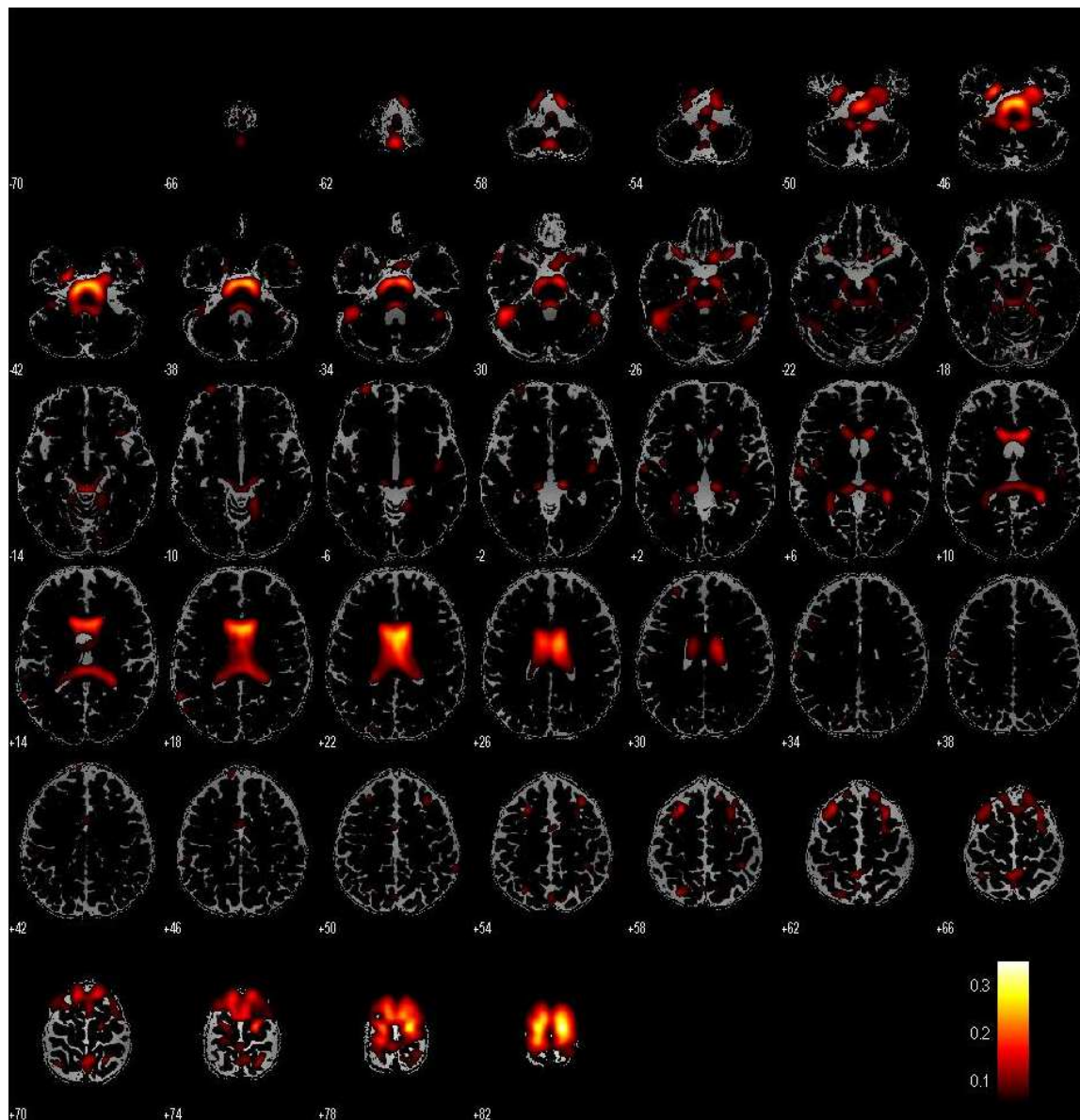


Figure 4.18: Volumetric enlargement of CSF in SCA3 patients by MVM analysis method. It is for testing and verifying the GM and WM analysis results. The increased space are detected around the pons, cerebellum, and ventricles, which are consistent with the GM and WM analysis results in MVM.

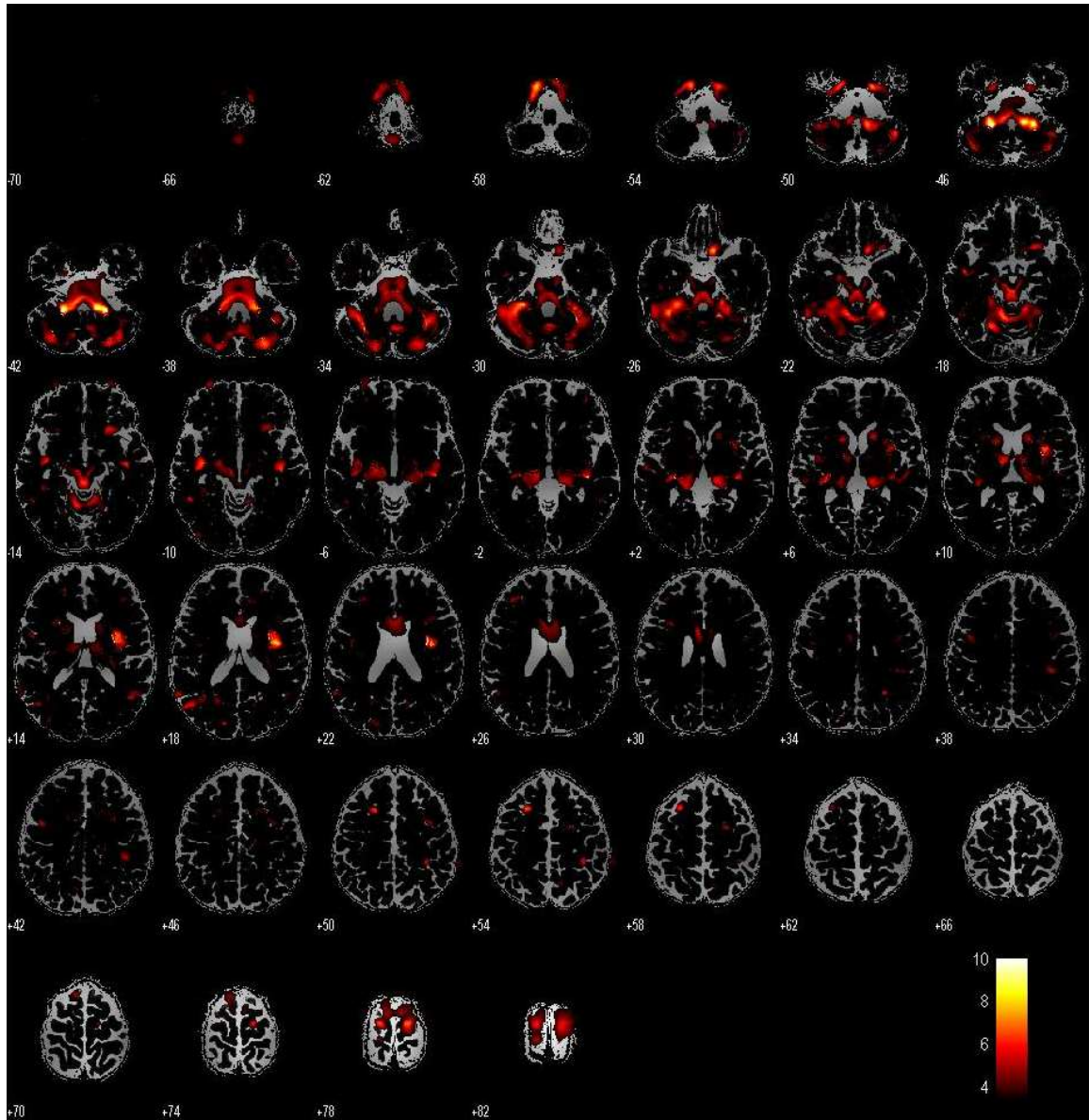


Figure 4.19: Volumetric enlargement of CSF in SCA3 patients by VBM analysis method. It is for testing and verifying the GM and WM analysis results. We can see there are some inconsistent revealed regions in VBM results.

putting GM, WM, and CSF analysis results together, we can see there are more consistent revealed regions in MVM than in VBM. MVM found the increased space around the pons, cerebellum, and ventricles. Most of these regions were also detected in atrophy of GM or WM. In contrast, The VBM CSF result seems some inconsistent. Even several regions should not appear in CSF. Therefore, we thought the MVM produced a more compelling analysis outcome than the VBM.



Chapter 5

Discussion



Some points are proposed and discussed in this chapter, concluding (1) the reason that data need not be smoothed, (2) the reason that why we employ LDA-based method as the classifier in the high dimensional space, and (3) the possible improvements in the future. Each point is discussed respectively in the following sections.

5.1 Why we do not smooth data

In the MVM, the method proposed in this thesis, data need not smooth before analysis to estimate the group discrepancy. Although there are some advantages of smoothing data, the bringing drawbacks also cannot be neglected. Specifically, these advantages are important and useful to VBM, but seem redundant and useless to MVM. That is, the smoothing step appears failed in the MVM analysis. Therefore the procedure of smoothing is omitted in the proposed method.

The contributions of smoothing are limited in MVM. As the mention in chapter 2, the advantages of smoothing are that: it can (1) render the data more normally distributed, (2) ensure that each voxel contains the average amount of gray matter or white matter from other adjacent voxels around this voxel, and (3) eliminate the error from spatial normalization. The profit of the first point is enhancement of the accuracy of the following voxel-by-voxel parametric statistical tests, and the second advantage helps compensate the voxel-wise analysis approach to catch some information being disregarded around this voxel. As to the third advantage, because the registration error of the underlying normalization method intimately influences the residual variability, which is estimated by final univariate tests in VBM, the profit is as well as the one of the first point. That is enhancing the correction of voxel-by-voxel parametric statistical tests. However, the MVM is a multivariate analysis approach itself, so information of all voxels is considered and analyzed at the same time rather than separately taken into account for each voxel. For this reason, benefit from the data smoothing is close to null for MVM.

The drawbacks from the data smoothing consist of three parts mainly. First, improper information spreads. Smoothing makes the information carried by a voxel disturbs other neighbor voxels on this voxel. Consequently, the information on those GM near the border between GM and non-GM tissues disperses to regions not belonging to GM, and the information on those WM near the border between WM and non-WM tissues disperse to regions not belonging to WM. This kind of smoothed result is unreasonable and inappropriate. Second, subtle differences are reduced. If the difference between two groups is subtle originally, this difference will disappear after smoothing the data. Thus, small variation of brain structures will be difficult to estimate. Third, problem to choosing the smoothing kernel size occurs. It is always a tough question about how large the size of smoothing kernel should be used. Analyzing data smoothed with different smoothing kernel sizes derives different conclusions about the presence and location of the result [30]. Although it is recommended that the size of smoothing kernel should be comparable to the size of the expected regional differences under study [49], it is unknown in advance how large the size of differences is in practical applications. To get a good analysis result subjectively, the smoothing kernel size is often need to be changed many times. Accordingly, due to this problem, it is required to analyze the data repeatedly, and the researchers also need the background knowledge of the underlying group discrepancy. That makes choice of smoothing kernel size adapting to various regional differences a troublesome matter.

Besides, we also have proved that, when based on Fisher's criterion to find the most discriminant projection vector of groups, using the smoothed data and the non-smoothed data as the samples to be analyzed can reach the same maximum of Fisher's criterion. Let \mathbf{A} be a N -by- N matrix that linearly combines elements of a N -dimensional column vector for performing the data smoothing, and let $\tilde{\mathbf{x}}_m^k$ be the smoothed sample of \mathbf{x}_m^k which is the m th sample of the k th class before smoothing. So $\tilde{\mathbf{x}}_m^k$ can be written as

$$\tilde{\mathbf{x}}_m^k = \mathbf{A}\mathbf{x}_m^k. \quad (5.1)$$

The mean of smoothed samples in the k th class and the mean of all smoothed samples are

easily derived from the following equations respectively, that are:

$$\tilde{\mu}^k = \frac{1}{M_k} \sum_{m=1}^{M_k} \tilde{\mathbf{x}}_m^k = \frac{1}{M_k} \sum_{m=1}^{M_k} \mathbf{A} \mathbf{x}_m^k = \mathbf{A} \left(\frac{1}{M_k} \sum_{m=1}^{M_k} \mathbf{x}_m^k \right) = \mathbf{A} \mu^k, \quad (5.2)$$

and

$$\tilde{\mu} = \frac{1}{M} \sum_{k=1}^K \sum_{m=1}^{M_k} \tilde{\mathbf{x}}_m^k = \frac{1}{M} \sum_{k=1}^K \sum_{m=1}^{M_k} \mathbf{A} \mathbf{x}_m^k = \mathbf{A} \left(\frac{1}{M} \sum_{k=1}^K \sum_{m=1}^{M_k} \mathbf{x}_m^k \right) = \mathbf{A} \mu. \quad (5.3)$$

Now, we are going to prove that it attains to the same maximum of Fisher's criterion no matter using the smoothed samples $\tilde{\mathbf{x}}_m^k$ or the non-smoothed samples \mathbf{x}_m^k . The maximum of Fisher's criterion of smoothed samples is

$$\max_{\mathbf{P}} \frac{|\mathbf{P}^T \widetilde{\mathbf{S}}_b \mathbf{P}|}{|\mathbf{P}^T \widetilde{\mathbf{S}}_w \mathbf{P}|} = \max_{\mathbf{P}} \frac{|\mathbf{P}^T \left(\sum_{k=1}^K M_k (\tilde{\mu}^k - \tilde{\mu})(\tilde{\mu}^k - \tilde{\mu})^T \right) \mathbf{P}|}{|\mathbf{P}^T \left(\sum_{k=1}^K \sum_{m=1}^{M_k} (\tilde{\mathbf{x}}_m^k - \tilde{\mu}^k)(\tilde{\mathbf{x}}_m^k - \tilde{\mu}^k)^T \right) \mathbf{P}|}, \quad (5.4)$$

where $\widetilde{\mathbf{S}}_b$ is the between-class scatter matrix of the smoothed samples, and $\widetilde{\mathbf{S}}_w$ is the within-class scatter matrix of the smoothed samples. By bringing Eq. (5.2) and Eq. (5.3) into Eq. (5.4), it becomes

$$\max_{\mathbf{P}} \frac{|\mathbf{P}^T \left(\sum_{k=1}^K M_k (\mathbf{A} \mu^k - \mathbf{A} \mu)(\mathbf{A} \mu^k - \mathbf{A} \mu)^T \right) \mathbf{P}|}{|\mathbf{P}^T \left(\sum_{k=1}^K \sum_{m=1}^{M_k} (\mathbf{A} \mathbf{x}_m^k - \mathbf{A} \mu^k)(\mathbf{A} \mathbf{x}_m^k - \mathbf{A} \mu^k)^T \right) \mathbf{P}|}. \quad (5.5)$$

This equation can continue to be derived and lead to the objective we want to prove, just by rearranging some orders of variables and setting a new projection matrix $\bar{\mathbf{P}}$ by $\bar{\mathbf{P}} = \mathbf{A}^T \mathbf{P}$.

That is,

$$\begin{aligned} & \max_{\mathbf{P}} \frac{|\mathbf{P}^T \left(\sum_{k=1}^K M_k (\mathbf{A} \mu^k - \mathbf{A} \mu)(\mathbf{A} \mu^k - \mathbf{A} \mu)^T \right) \mathbf{P}|}{|\mathbf{P}^T \left(\sum_{k=1}^K \sum_{m=1}^{M_k} (\mathbf{A} \mathbf{x}_m^k - \mathbf{A} \mu^k)(\mathbf{A} \mathbf{x}_m^k - \mathbf{A} \mu^k)^T \right) \mathbf{P}|} \\ &= \max_{\mathbf{P}} \frac{|\sum_{k=1}^K M_k \mathbf{P}^T (\mathbf{A} (\mu^k - \mu)) (\mathbf{A} (\mu^k - \mu))^T \mathbf{P}|}{|\sum_{k=1}^K \sum_{m=1}^{M_k} \mathbf{P}^T (\mathbf{A} (\mathbf{x}_m^k - \mu^k)) (\mathbf{A} (\mathbf{x}_m^k - \mu^k))^T \mathbf{P}|} \end{aligned}$$

$$\begin{aligned}
&= \max_{\mathbf{P}} \frac{\left| \sum_{k=1}^K M_k ((\mu^k - \mu)^T \mathbf{A}^T \mathbf{P})^T ((\mu^k - \mu)^T \mathbf{A}^T \mathbf{P}) \right|}{\left| \sum_{k=1}^K \sum_{m=1}^{M_k} ((\mathbf{x}_m^k - \mu^k)^T \mathbf{A}^T \mathbf{P})^T ((\mathbf{x}_m^k - \mu^k)^T \mathbf{A}^T \mathbf{P}) \right|} \\
&= \max_{\bar{\mathbf{P}}} \frac{\left| \sum_{k=1}^K M_k ((\mu^k - \mu)^T \bar{\mathbf{P}})^T ((\mu^k - \mu)^T \bar{\mathbf{P}}) \right|}{\left| \sum_{k=1}^K \sum_{m=1}^{M_k} ((\mathbf{x}_m^k - \mu^k)^T \bar{\mathbf{P}})^T ((\mathbf{x}_m^k - \mu^k)^T \bar{\mathbf{P}}) \right|} \\
&= \max_{\bar{\mathbf{P}}} \frac{\left| \sum_{k=1}^K M_k \bar{\mathbf{P}}^T (\mu^k - \mu) (\mu^k - \mu)^T \bar{\mathbf{P}} \right|}{\left| \sum_{k=1}^K \sum_{m=1}^{M_k} \bar{\mathbf{P}}^T (\mathbf{x}_m^k - \mu^k) (\mathbf{x}_m^k - \mu^k)^T \bar{\mathbf{P}} \right|} \\
&= \max_{\bar{\mathbf{P}}} \frac{\left| \bar{\mathbf{P}}^T \left(\sum_{k=1}^K M_k (\mu^k - \mu) (\mu^k - \mu)^T \right) \bar{\mathbf{P}} \right|}{\left| \bar{\mathbf{P}}^T \left(\sum_{k=1}^K \sum_{m=1}^{M_k} (\mathbf{x}_m^k - \mu^k) (\mathbf{x}_m^k - \mu^k)^T \right) \bar{\mathbf{P}} \right|} \\
&= \max_{\bar{\mathbf{P}}} \frac{|\bar{\mathbf{P}}^T \mathbf{S}_b \bar{\mathbf{P}}|}{|\bar{\mathbf{P}}^T \mathbf{S}_w \bar{\mathbf{P}}|} \tag{5.6}
\end{aligned}$$

It is to say, utilizing the smoothed data would not enlarge the maximum of Fisher's criterion. More precisely, the total discrimination of the discriminating map is the same regardless of smoothing or not. So the step of data smoothing in preprocessing becomes a redundant procedure in our method.

In addition to the theoretical proof, we also have the empirical validation that the MVM analysis result from non-smoothed data is indeed better than the result from smoothed data. Figure 5.1 and Table 5.1 show the ROC curves and the corresponding PAUC indices of MVM to analyze the same simulation data in section 4.1.1 without/with data smoothing.

It is obviously to see that, the ROC curves of the non-smoothed data are more near the left upper corner and their PAUC indices are greater than ones of the smoothed data, when applied to the MVM method. This represents that using non-smoothed data can obtain more correct results in MVM method. Smoothing the data reduces the accuracy instead.

For all the reasons mentioned above in this section, including the inferences, proof, and

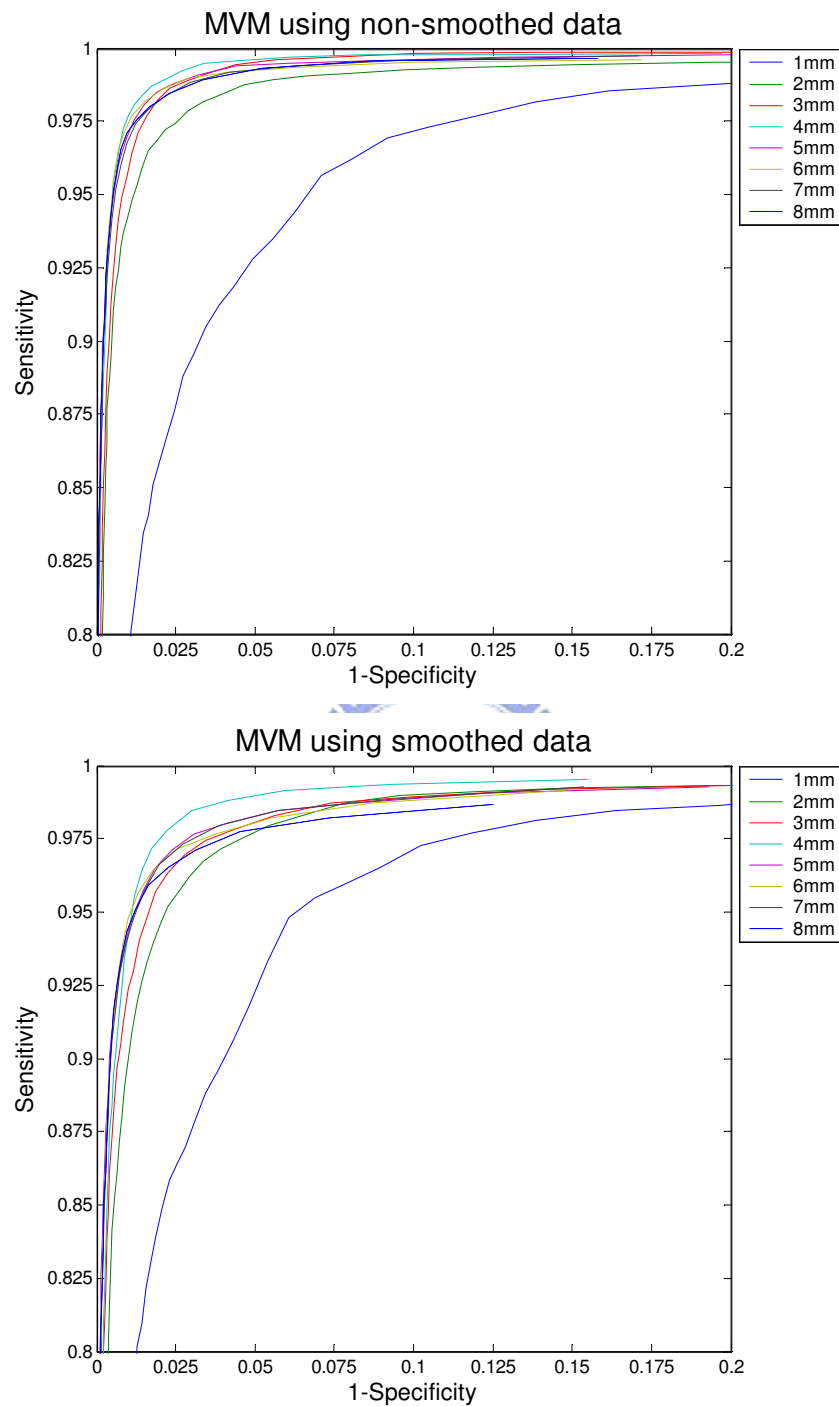


Figure 5.1: ROC curves of MVM results with the non-smoothed and smoothed simulation data. Notice that both horizontal axes are limited from 0 to 0.2 and both vertical axes are limited from 0.8 to 1 for more specific display. It is manifest that the curves of analysis with the non-smoothed data are closer to the top left corner than those curves of analysis with the smoothed data in the same analysis data.

Table 5.1: PAUC indices for ROC curves of MVM results with the non-smoothed and smoothed simulation data. The partial area was calculated in a specific region where the TP rate ranges from 0.8 to 1 and the FP rate ranges from 0 to 0.2. As this specific region denotes 1, we have $0 \leq \text{PAUC} \leq 1$. It is manifest that PAUC indices of the non-smoothed data are greater than ones of the smoothed data for the MVM method.

Simulated atrophy size	MVM	
	Non-smoothed data	Smoothed data
1mm	0.73466	0.71880
2mm	0.92286	0.87892
3mm	0.95521	0.89374
4mm	0.96613	0.92628
5mm	0.95836	0.90717
6mm	0.95398	0.90272
7mm	0.95465	0.90693
8mm	0.95357	0.88517

experiment, the proposed MVM analysis method discards the procedure of data smoothing in preprocessing. But it uses a smoothing of the discriminating map for visualization purpose, which has been explained in section 3.2.

5.2 Comparison with other classification-based techniques

Relative to the original paper which proposed we should replace the univariate approach by a multivariate approach when characterizing group differences, Lao et al. [50] did not adopt LDA as the analytic method, but adopt the support vector machine (SVM) as the classifier in the high-dimensional space. SVM [51] is a very powerful technique of classification. In recent years, SVM has been applied in many applications, such as face

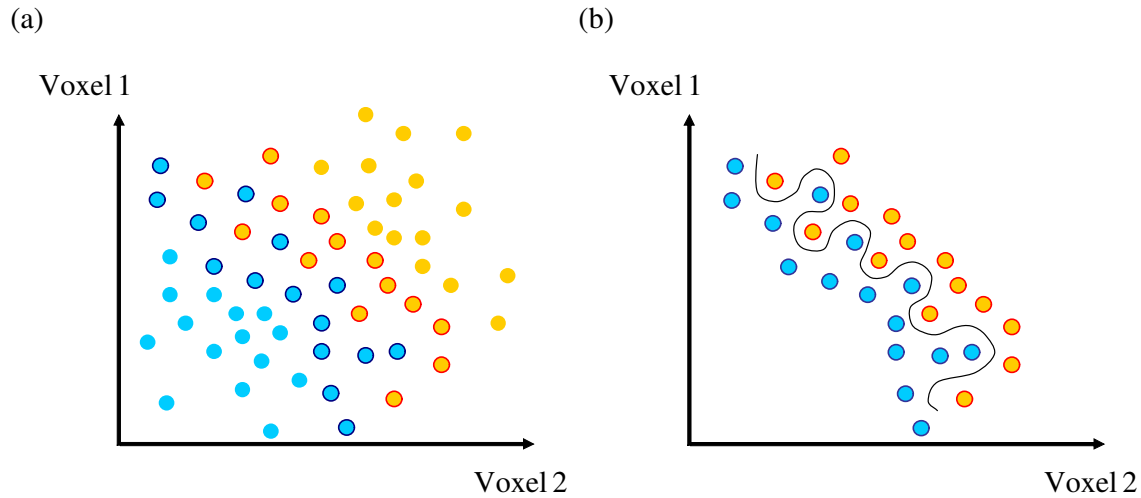


Figure 5.2: Sketch chart of how SVM determines the separating hypersurface. To display conveniently, only display in 2-D. (a) The blue points in the lower left area denote the samples and distribution of one group. The yellow points in the upper right area denote the samples and distribution of another group. Those blue points and yellow points which are emphasized by the circular frame represent the samples close to the interface between two groups. (b) SVM utilizes these support vectors to determine the hypersurface separating samples from different groups. Other samples which are not the support vectors are null for SVM.

recognition [52], and brain-computer interface system [53]. The way of SVM to determine a hypersurface dividing two groups only depends on those samples called support vectors that are close to the interface between groups. Other samples, far away from the interface between groups, are useless for the SVM classifier to determine the separating hypersurface. Figure 5.2 illustrates this thought. It makes SVM easily catch subtle changes of groups. Therefore, SVM works successfully in many classification problems.

On the other hand, SVM ignores samples far away which are subjects most representative of the group in most cases, so it may not be effective in catching significant group differences. This is why we did not use SVM to characterize anatomical differences between groups, even though the SVM has great ability of classification. Some other classification techniques, such as SVM-based nonparametric discriminant analysis [54] and

boosted LDA [55], have the same problem of depreciating samples far away from the interface between groups. This is inappropriate on our purpose. On the contrary, LDA adopts properties in statistics, such as group covariance, to determine the most discriminant projection vector. Therefore, it would not lay particular stress on the samples near the interface between groups. As a result, LDA is more suitable for analyzing group differences than the SVM in this application.

Moreover, LDA is a linear classifier, and consequently it is convenient for representation of the analysis result. The most discriminant projection vector is also an image, where each voxel has a weight of discrimination to distinguish group discrepancy, containing regions representing the spatial distribution and relative magnitudes of differences between different groups. Displaying the most discriminant projection vector directly, as well as we have done in chapter 4, is an intuitive and meaningful one of visualization ways.

But for other nonlinear classification techniques, such as SVM, there are difficulties in representation of detected significant differences. It is because nonlinear classifier determines a hypersurface rather than a hyperplane separating groups, and it is difficult to summarize the analysis result into an image from this hypersurface. In [50], the way of interpretation of the analysis result is that, for every support vector, they found the vector that is perpendicular to the separating hyperplane and passes through this support vector from one group to another group, and then averaged all vectors to obtain a single map, like Figure 5.3 shows. That is not a reasonable way to translate the group differences, because averaging all vectors does not make sense. There is still a representation problem when applying nonlinear classifiers.

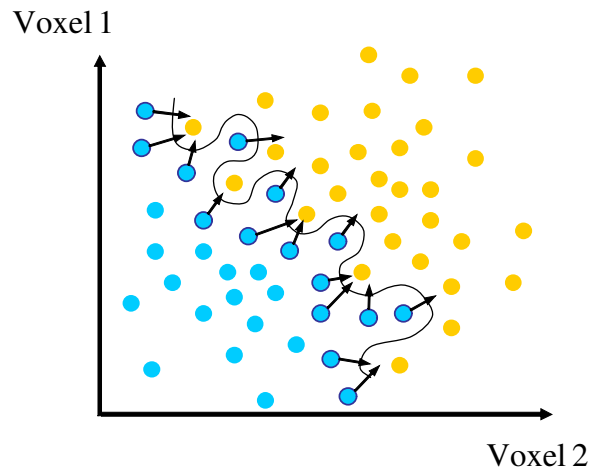


Figure 5.3: Vectors used to obtain a single map for nonlinear classification. It was proposed in [50]. First, for every sample near the separating hypersurface, the vector perpendicular to the hyperplane and through this sample is decided. In this figure only vectors derived from blue samples are drawn for the clear display. Then, a single map for nonlinear classification is obtained by averaging all those vectors.

5.3 Weighted within-class and between-class scatter matrices

In the proposed method, we use wavelet transformation to enhance the connection between nearby voxels, which can compensate the disadvantage that LDA treats relations between all voxel pairs equally. Wavelets improve this drawback, but the performance is limited. Thus, another improved approach, weighted within-class and between-class scatter matrices for LDA, is proposed. The idea of this method is quite easy. It directly uses a number of weights to change the relations between voxel pairs to achieve the enhancement of spatial information from MR images. It is intuitive to use a small weight for the combination of two voxels far away from each other, and a large weight for the combination of two voxels adjacent to each other. According to the distances in the space or correlations between brain structures, we can define, in advance, a symmetric weight matrix \mathbf{W} which has the same dimension (N-by-N) as within-class and between-class scatter matrices. Each

element $w_{i,j}$ in the i th row and j th column is the weight of spatial relation between the i th voxel and j th voxel. Then, element-to-element multiply the weight matrix \mathbf{W} by the within-class scatter matrix and the between-class scatter matrix respectively. The weighted within-class scatter matrix and between-class scatter matrix are defined as

$$\widehat{\mathbf{S}}_{\mathbf{w}} = \mathbf{S}_{\mathbf{w}} \cdot * \mathbf{W}, \quad (5.7)$$

and

$$\widehat{\mathbf{S}}_{\mathbf{b}} = \mathbf{S}_{\mathbf{b}} \cdot * \mathbf{W}, \quad (5.8)$$

where the operator $\cdot *$ represents the element-by-element multiplication of two matrices with the same size. And the Fisher's criterion yields that:

$$\arg \max_{\mathbf{P}} \frac{|\mathbf{P}^T \widehat{\mathbf{S}}_{\mathbf{b}} \mathbf{P}|}{|\mathbf{P}^T \widehat{\mathbf{S}}_{\mathbf{w}} \mathbf{P}|} \quad (5.9)$$

Although this idea is completed, until now we do not implement this method owing to the computation problem of a huge matrix. More specifically, the technique of solving eigen-problem of the within-class scatter matrix proposed in section 3.3.3 is failed here, so the newly most discriminant projection vector based on the weighted within-class and between-class scatter matrices cannot be solved. However, it is a worthily attempted method to improve the performance of MVM in the future.

5.4 Multivariate deformation-based analysis

The proposed multivariate analysis can be not only applied in the second family of morphometric methods for microscopic differences in brain tissue, but also applied in the first family of morphometric methods for macroscopic differences in brain shape. That is, we can use a multivariate analysis on the deformation fields rather than on the normalized

images. It is the multivariate deformation-based analysis. According to our experimental results, some wide-region differences in brain shape, which are hardly found by the conventional point-wise analysis method, are expected to be revealed by this kind of multivariate deformation-based analysis. It may produce a good outcome as measuring group differences of the brain shape.



Chapter 6

Conclusions



We propose a novel morphometric method, the multivariate volumetric morphometry (MVM), characterizing anatomical discrepancy by a multivariate way. MVM includes the preprocessing, the multivariate analysis, and the smoothing and thresholding for display. The multivariate analysis of MVM is the kernel breakthrough part in this work. In the multivariate analysis, it employs the discriminative common vector method to find the representative regions of group discrepancy, and arranges the wavelet transform in pairs to enhance the spatial relationship of two distinct voxels. In addition, we also proposed an efficient implementation to calculate the discriminative common vector. According to our experiments, the effectiveness of the proposed method is clearly demonstrated with a good ability to catch the anatomical discrepancy between different groups. Compared with one of most popular morphometric technique presently, voxel-based morphometry (VBM), MVM expresses a better sensitivity to subtle and widely-distributed variation of brain structure than VBM. Moreover, the time cost in MVM is less than VBM, even though MVM uses a multivariate analysis in a considerably high-dimensional space.

Other two important properties of MVM are that, it is unnecessary for MVM to smoothing the data or to reduce the features before the multivariate analysis. For the first point about data smoothing, we not only address the reasons that data smoothing has no benefits but is harmful to MVM, but also prove that the maximums of Fisher's criterion are equivalent no matter the data is smoothed or non-smoothed. Besides, the experiment also supported the point at issue that MVM does not need to smooth data before analysis. To the second point about feature reduction, because the discriminative common vector method can handle a good deal of data at one time, it is indeed no need to reducing the features for computation problem. That makes MVM superior to other multivariate methods as well as dealing with a huge amount of features, since no information is lost during the MVM analysis.

Somehow, there is a fly in the ointment in the MVM method. Unlike the t -test map of VBM, the voxel values of MVM discriminating map denotes the discrimination weight

distinguishing samples from groups, and they lack for absolute judgment standard to determine the significant level with a given confidence. Therefore, we chose the display threshold of MVM by a strategy mentioned in chapter 4 by consulting the VBM analysis result. Although discrimination weights of the MVM discriminating map do not provide confidences in whether there is a practical group difference at this location of each voxel, the discriminating map still provides a whole-brain confidence in explaining the detected group discrepancy. Our work demonstrated that the MVM analysis method has a good sensitivity to subtle and widely-distributed structural differences, and is a useful technique in neuroimage studies.





Bibliography

- [1] K. Brodmann. *Vergleichende lokalisationslehre der grohirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues*. Johann Ambrosius Barth Verlag, Leipzig, 1909.
- [2] J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain*. Thieme Medical Publishers, New York, 1988.
- [3] P. C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.
- [4] Z. P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging: A signal processing perspective*. IEEE Press series in biomedical engineering. IEEE Press, New York, 2000.
- [5] R. W. Buchanan, K. Vladar, P. E. Barta, and G. D. Pearlson. Structural evaluation of the prefrontal cortex in schizophrenia. *American Journal of Psychiatry*, 155(8):1049–1055, 1998.
- [6] R. W. Buchanan, A. Francis, C. Arango, K. Miller, D. M. Lefkowitz, R. P. McMahon, P. E. Barta, and G. D. Pearlson. Morphometric assessment of the heteromodal association cortex in schizophrenia. *American Journal of Psychiatry*, 161(2):322–331, 2004.
- [7] M. R. Herbert, G. J. Harris, K. T. Adrien, D. A. Ziegler, N. Makris, D. N. Kennedy,

- N. T. Lange, C. F. Chabris, A. Bakardjiev, J. Hodgson, M. Takeoka, H. Tager-Flusberg, and V. S. Caviness. Abnormal asymmetry in language association cortex in autism. *Annals of Neurology*, 52(5):588–596, 2002.
- [8] M. Kubicki, M. E. Shenton, D. F. Salisbury, Y. Hirayasu, and R. W. McCarley. Voxel-based morphometric analysis of gray matter in first episode schizophrenia. *Schizophrenia Research*, 53(3):96–96, 2002. Sp. Iss. SI Suppl. S.
- [9] C. Davatzikos, A. Genc, D. R. Xu, and S. M. Resnick. Voxel-based morphometry using the ravens maps: Methods and validation using simulated longitudinal atrophy. *Neuroimage*, 14(6):1361–1369, 2001.
- [10] M. Miller, A. Banerjee, G. Christensen, S. Joshi, N. Khaneja, U. Grenander, and L. Matejic. Statistical methods in computational anatomy. *Stat. Methods Med. Res.*, 6(3):267–299, 1997.
- [11] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsruce, C. Price, and K. Friston. Identifying global anatomical differences: Deformation-based morphometry. *Human Brain Mapping*, 6(5-6):348–357, 1998.
- [12] C. Gaser, H. P. Volz, S. Kiebel, S. Riehemann, and H. Sauer. Detecting structural changes in whole brain based on nonlinear deformations - application to schizophrenia research. *Neuroimage*, 10(2):107–113, 1999.
- [13] P. M. Thompson, D. MacDonald, M. S. Mega, C. J. Holmes, A. C. Evans, and A. W. Toga. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *Journal of Computer Assisted Tomography*, 21(4):567–581, 1997.
- [14] P. A. Freeborough and N. C. Fox. Modeling brain deformations in alzheimer disease by fluid registration of serial 3d mr images. *Journal of Computer Assisted Tomography*, 22(5):838–843, 1998.

- [15] A. M. C. Machado, J. C. Gee, and M. F. M. Campos. A factor analytic approach to structural characterization. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 219–225, Hilton Head, SC, 2000.
- [16] J. Ashburner and K. J. Friston. Voxel-based morphometry - the methods. *Neuroimage*, 11(6):805–821, 2000. Part 1.
- [17] C. Davatzikos. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *Neuroimage*, 23(1):17–20, 2004.
- [18] M. Kubicki, M. E. Shenton, D. F. Salisbury, Y. Hirayasu, and R. W. McCarley. Voxel-based morphometric analysis of gray matter in first episode schizophrenia. *Schizophrenia Research*, 53(3):96–96, 2002. Sp. Iss. SI Suppl. S.
- [19] H. Ananth, I. Popescu, H. D. Critchley, C. D. Good, R. S. J. Frackowiak, and R. J. Dolan. Cortical and subcortical gray matter abnormalities in schizophrenia determined through structural magnetic resonance imaging with optimized volumetric voxel-based morphometry. *American Journal of Psychiatry*, 159(9):1497–1505, 2002.
- [20] H. E. H. Pol, H. G. Schnack, R. C. W. Mandl, N. E. M. van Haren, H. Koning, D. L. Collins, A. C. Evans, and R. S. Kahn. Focal gray matter density changes in schizophrenia. *Archives of General Psychiatry*, 58(12):1118–1125, 2001.
- [21] J. C. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette, and F. Eustache. In vivo mapping of gray matter loss with voxel-based morphometry in mild alzheimer’s disease. *Neuroimage*, 14(2):298–309, 2001.
- [22] G. B. Frisoni, C. Testa, A. Zorzan, F. Sabbatoli, A. Beltramello, H. Soininen, and M. P. Laakso. Detection of grey matter loss in mild alzheimer’s disease with voxel based morphometry. *Journal of Neurology Neurosurgery and Psychiatry*, 73(6):657–664, 2002.

- [23] G. B. Karas, E. J. Burton, Sarb Rombouts, R. A. van Schijndel, J. T. O'Brien, P. Scheltens, I. G. McKeith, D. Williams, C. Ballard, and F. Barkhof. A comprehensive study of gray matter loss in patients with alzheimer's disease using optimized voxel-based morphometry. *Neuroimage*, 18(4):895–907, 2003.
- [24] G. M. McAlonan, V. Cheung, C. Cheung, J. Suckling, G. Y. Lam, K. S. Tai, L. Yip, D. G. M. Murphy, and S. E. Chua. Mapping the brain in autism. a voxel-based mri study of volumetric differences and intercorrelations in autism. *Brain*, 128:268–276, 2005. Part 2.
- [25] N. Boddaert, N. Chabane, H. Gervais, C. D. Good, M. Bourgeois, M. H. Plumet, C. Barthelemy, M. C. Mouren, E. Artiges, Y. Samson, F. Brunelle, R. S. J. Frackowiak, and M. Zilbovicius. Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry mri study. *Neuroimage*, 23(1):364–369, 2004.
- [26] C. Brenneis, S. M. Bosch, M. Schocke, G. K. Wenning, and W. Poewe. Atrophy pattern in sca2 determined by voxel-based morphometry. *Neuroreport*, 14(14):1799–1802, 2003.
- [27] S. Carmona, O. Vilarroya, A. Bielsa, V. Tremols, J. C. Soliva, M. Rovira, J. Tomas, C. Raheb, J. D. Gisbert, S. Batlle, and A. Bulbena. Global and regional gray matter reductions in adhd: A voxel-based morphometric study. *Neuroscience Letters*, 389(2):88–93, 2005.
- [28] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage*, 14(1):21–36, 2001. Part 1.
- [29] F. L. Bookstein. "voxel-based morphometry" should not be used with imperfectly registered images. *Neuroimage*, 14(6):1454–1462, 2001.

- [30] D. K. Jones, M. R. Symms, M. Cercignani, and R. J. Howard. The effect of filter size on vbm analyses of dt-mri data. *Neuroimage*, 26(2):546–554, 2005.
- [31] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1995.
- [32] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.
- [33] A. Fisher. *The mathematical theory of probabilities*. Macmillan, New York, 1923.
- [34] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, New York, 2nd edition, 1990.
- [35] K. Liu, Y. Q. Cheng, J. Y. Yang, and K. Liu. An efficient algorithm for foley-sammon optimal set of discriminant vectors by algebraic method. *International Journal of Pattern Recognition and Artificial Intelligence*, 6(5):817–829, 1992.
- [36] G. H. Dunteman. *Principal Components Analysis*. SAGE Publications, Thousand Oaks, 1989.
- [37] L. F. Chen, H. Y. M. Liao, M. T. Ko, J. C. Lin, and G. J. Yu. A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726, 2000.
- [38] Y. Bing, J. Lianfu, and C. Ping. A new lda-based method for face recognition. In *International Conference on Pattern Recognition*, Quebec City, Canada, 2002.
- [39] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):70–86, 1991.

- [40] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*, pages 586–591, Maui, HI, USA, 1991.
- [41] M. Farye. Wavelet transforms and their applications to turbulences. *Annual Review of Fluid Mechanics*, 24:395–457, 1992.
- [42] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelet-like bases for the fast solution of 2nd-kind integral-equations. *Siam Journal on Scientific Computing*, 14(1):159–184, 1993.
- [43] F. L. Bookstein. Principal warp: Thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.
- [44] M. L. Thompson and W. Zucchini. On the statistical analysis of roc curves. *Statistics In Medicine*, 8:1277–1290, 1989.
- [45] A. F. Abdelnour and I. W. Selesnick. Nearly symmetric orthogonal wavelet bases. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, 2001.
- [46] B. W. Soong, Y. C. Lu, K. B. Choo, and H. Y. Lee. Frequency analysis of autosomal dominant cerebellar ataxias in taiwanese patients and clinical and molecular characterization of spinocerebellar ataxia type 6. *Arch Neurol*, 58:1105–1109, 2001.
- [47] P. Trouillas, T. Takayanagi, M. Hallett, R. D. Currier, S. H. Subramony, K. Wessel, A. Bryer, H. C. Diener, S. Massaquoi, C. M. Gomez, P. Coutinho, M. BenHamida, G. Campanella, A. Filla, L. Schut, D. Timann, J. Honnorat, N. Nighoghossian, and B. Manyam. International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. *Journal of the Neurological Sciences*, 145(2):205–211, 1997.

- [48] J. D. Schmahmann and J. C. Sherman. The cerebellar cognitive affective syndrome. *Brain*, 121:561–579, 1998. Part 4.
- [49] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner. Voxel-based morphometry of the human brain: Methods and applications. *Current Medical Imaging Reviews*, 1(2):105–113, 2005.
- [50] Z. Q. Lao, D. G. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1):46–57, 2004.
- [51] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [52] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 130–136, Puerto Rico, 1997.
- [53] K. R. Muller, C. W. Anderson, and G. E. Birch. Linear and nonlinear methods for brain-computer interfaces. *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):165–169, 2003.
- [54] R. Fransens, J. De Prins, and L. Van Gool. Svm-based nonparametric discriminant analysis, an application to face detection. In *IEEE International Conference on Computer Vision*, volume 2, pages 1289– 1296, Belgium, 2003.
- [55] M. Tanigawa and T Kurita. Multi-class object recognition using boosted linear discriminant analysis combined with masking covariance matrix method. In *IEEE International Conference on Computer Vision Systems*, pages 33–38, New York, 2006.