THE
UNIVERSITY OF
STRATHCLYDE
IN GLASGOW

University of Strathclyde
Department of Electronics and
Electrical Engineering
& Bioengineering Unit

MRI BRAIN TUMOUR CLASSIFICATION USING IMAGE PROCESSING AND DATA MINING

by

SHAN SHEN

A thesis presented in fulfilment of the requirements for the degree of

Doctor of Philosophy

September 2004

Copyright Statement

Acknowledgement

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

I am deeply indebted to my direct supervisor Dr William Sandham. I owe him lots of gratitude for his excellent supervision, invaluable support and encouragement which helped me in all the time of my research. I also would like to thank my second supervisor Dr Malcolm Granat who gave me great guidance and provided constructive comments during my thesis time as well as on the preliminary version of this thesis.

I want to thank Dr Mary Frances Dempsey and Dr James Patterson of the Southern General Hospital, Glasgow, UK for providing data and expert clinical advice. Without their help, my research cannot be carried out even from the beginning.

Special thanks go to Prof J. Barbenel, from Department of Electronics and Electrical Engineering for helping me obtain the opportunity to start my PhD study. Financial support from overseas research student (ORS) awards, international Scholarship of University of Strahtclyde, Faculty of Engineering and Bioengineering Unit is gratefully acknowledged.

I am grateful for Prof. J. Courtney from Bioengineering Unit who always concerned my work progress and supported me as a good friend.

Many thanks are due to Amy Saunders and Phil Dean who spent a lot of time improving the English in this thesis.

A deep sense of gratitude is for my parents who formed part of my vision and taught me the good things that really matter in life, and my husband, Jian, for his love and support. I also would like to thank all my friends for always being there for me.

Many more persons participated in various ways to ensure my research succeeded and I am thankful to them all.

To my beloved parents and husband

ABSTRACT

Detecting and diagnosing brain tumour types quickly and accurately is essential to any effective treatment. The general brain tumour diagnosis procedure, biopsy, not only causes a great deal of pain to the patient but also raises operational difficulty to the clinician. In this thesis, a non-invasive brain tumour diagnosis system based on MR images is proposed. The first part is image preprocessing applied to original MR images from the hospital. Non-uniformed intensity scales of MR images are standardized relying on their statistic characteristics without requiring prior or post templates. It is followed by a non-brain region removal process using morphologic operations and a contrast enhancement between white matter and grey matter by means of histogram equalization. The second part is image segmentation applied to preprocessed MR images. A new image segmentation algorithm named IFCM is developed based on the traditional FCM algorithm. Neighbourhood attractions considered in IFCM enable this new algorithm insensitive to noise, while a neural network model is designed to determine optimized degrees of attractions. This extension can also estimate inhomogenities. Brain tissue intensities are acquired from segmentation. The final part of the system is brain tumour classification. It extracts hidden diagnosis information from brain tissue intensities using a fuzzy logic based GP algorithm. This novel method imports a fuzzy membership to implement a multi-class classification directly without converting it into several binary classification problems as with most other methods. Two fitness functions are defined to describe the features of medical data precisely.

The superiority of image analysis methods in each part was demonstrated on synthetic images and real MR images. Classification rules of three types and two grades of brain tumours were discovered. The final diagnosis accuracy was very promising. The feasibility and capability of the non-invasive diagnosis system were testified comprehensively.

Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 1

INTRODUCTION

1.1 BACKGROUND

Brain tumours are composed of cells that exhibit unrestrained growth in the brain. It is a complex disease that affects more than 100,000 annually in the USA, and it is still on the increase. Worldwide this number is much higher. Although the incidence rate of brain tumours is lower than other cancers, it is still the most serious disease threatening human lives. The survival rate of brain tumours is significantly lower than that for most other types of cancers, as their cause is unknown. It is not clear how to reduce the risk of having brain tumours [The Brain Tumour Society, 2002].

Brain tumours are generally divided into two categories, primary and secondary. Primary tumours are tumours that begin in the brain. Secondary tumours start elsewhere in the body such as the lung, kidney, breast, or skin and spread, or metastasize, to the brain. Gliomas are the most common type of primary brain tumours which arise from the connective tissue of the brain. Types of gliomas include astrocytomas, glioblastomas, oligodendrogliomas and ependymomas. These tumours make up approximately 50% of all primary brain tumours. Meningiomas comprise approximately 25% of brain tumours, with pituitary tumours making up 10% and acoustic neuromas comprising 7.5%. Other tumour types including chondrosarcomas, germinomas, hemangiomas, teratomas, and chordomas, make up the remaining 7.5% [The Wallace-Kettering Neuroscience Institute, 2003].

Detecting and diagnosing brain tumours quickly and accurately is essential to the effective treatment. Clinical brain tumour diagnosis normally includes three steps: neurological exam, brain image analysis and biopsy. When a brain tumour is suspected from the patient's medical history and symptoms, a neurological exam is used to check, for example, the eye movement, eye reflex, pupil reaction, sensation, hearing, balance and coordination of the patients. These can imply the signs of diminished mental function. The next step involves a brain image scan of the patients. It may detect the presence of a brain tumour and its exact size and location. The most common imaging techniques for brain tumour diagnosis are:

Computed Tomography (CT) CT is an established technique which uses X-rays or gamma rays to reconstruct cross-sectional image of the human body. It uses the different attenuation coefficients of different tissues after absorbing the rays to produce the anatomical structure of the body [Cho, 1975]. A CT scan often gives the location of the tumour and can detect swelling, bleeding and other associated conditions.

Magnetic Resonance Imaging (MRI) MRI does not involve ionizing radiation but provides information on the number and position of hydrogen nuclei within the body by detecting the Radio Frequency (RF) signals which are emitted following excitation by magnetic fields. It can often distinguish more accurately between healthy and abnormal tissue than CT [Liang, 2000]. MRI can generate images from various angles which enable 3D brain images to be constructed. It can supply a better view of tumours located near bone, smaller tumours, brainstem tumours, and low-grade tumours. However, the running cost of MRI scanner is high and a long scanning time is required.

Positron Emission Tomography (PET) PET yields transverse tomographic images of the distribution of positron-emitting radionuclides systemically administered to the human body. It provides an image of brain activity rather than structure by measuring levels of injected glucose (sugar) or methionine (amino acid) that have been labeled with a radioactive tracer [Robb, 1985]. Recent studies have shown that it may be easier to detect recurrent brain tumours with PET rather than CT and MRI. Fig.1.1 shows CT, MRI and PET images of the brain.

The final step of brain tumour diagnosis is a biopsy. A biopsy is a surgical procedure used to take a sample tissue from the suspected tumour to determine an exact diagnosis such as malignance and tumour types [The Brain Tumour Society, 2002]. Generally, biopsies can be performed "open" or "closed". An open biopsy usually implies that a small window of bone is temporarily removed from the skull to allow the surgeon to remove a small portion of the suspected tumour under direct visualization. A closed biopsy is usually performed through a tiny hole, often no larger than one eighth of an inch. The tumour is not directly seen by the surgeon, rather a CT scan or MRI is used to determine from where the biopsy should be obtained. The suspected tumour sample can then be viewed under the microscope and a diagnosis is made.

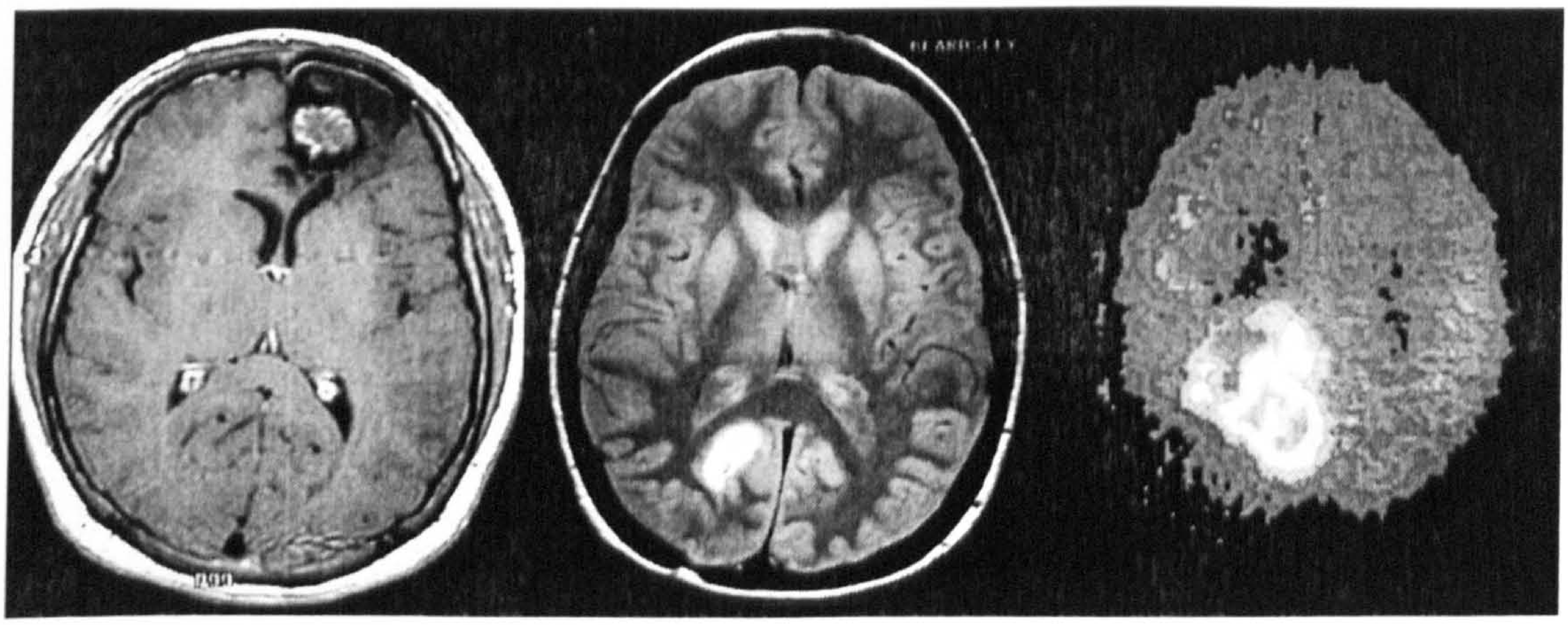


Figure 1.1 (Left) CT [Alvira 1997], (middle) MRI [Beardsley 2004] and (right) PET [Medica.de 2004] images.

A biopsy, however, can cause a great deal of pain to the patients and an open or closed biopsy is also a dangerous procedure, particularly the open biopsy. This raises the difficulty of diagnosis. Further more, not all brain tumours can be biopsied, for example brain stem gliomas, as it is too hazardous, as removing any healthy tissue from the brain stem can affect vital functions. So, a basic question arises "Is it possible to diagnose brain tumour types without an invasive procedure?" There are new developing techniques that may provide an answer to the question.

1.2 AIMS

The development of imaging techniques has greatly extended the range of human vision into realms that would otherwise be inaccessible, such as the anatomical structure inside the human body. In fact, much of what we know about ourselves and the world around us has been derived from images produced by various imaging devices. These imaging techniques may supply much more knowledge than we can imagine.

To answer the question we stated above 'how to diagnose brain tumour types without an invasive procedure?' again, the consideration needs to be given to imaging techniques.

The main brain imaging techniques include CT, MRI, PET, functional MRI (fMRI) and magneto-encephalography (MEG). CT and MRI provide images of the brain structure. PET mainly focuses on the chemistry of brain activities and fMRI records physiological changes like blood flow in the brain. MEG supplies images of nerve activity. All these techniques need to be considered, but which one is the most suitable to our problem?

Siromoney *et al.* (2000) first discriminated two types of brain tumours, meningioma and astrocytoma, using MR images analysis. Hence the possibility of using MRI for brain tumour diagnosis has been established. Besides the successful precedent, MRI has advantages over other imaging techniques. MRI scanner produces multi-dimensional images representing the spatial distribution of some measured physical quantity like other imaging techniques. However, unlike many of them, it can generate, two-dimensional sectional images in any orientation, three-dimensional volumetric images or even four-dimensional images representing spatial-spectral distributions [Liang and Lauterbur, 2000]. Secondly, MRI does not use ionizing radiation like CT and requires no injection of radioactive isotopes like PET. It is considered very safe for biological systems as there are no known harmful effects [Siromoney *et al.*, 2000]. Finally, the most important advantage of MRI over other techniques is its high spatial resolution and contrast resolution [Wells *et al.*, 1996]. MR images can supply excellent discrimination of soft tissues, such as white matter and gray matter in the brain. Taking into account these factors, MRI is the most appropriate and most promising technique for non-invasive diagnosing brain tumours.

Signal intensities are important information in an MRI scan. Different tissues in the brain MRI have different signal intensities and therefore have distinctive displays. Siromoney *et al.* (2000) utilized the measurements of MR signal intensities of brain tumour and white matter to diagnose tumour types. A pre-prepared data set of signal intensities for white matter, gray matter, cerebrospinal fluid (CSF) and brain tumours measured manually were applied in diagnosis. Two types of brain tumours were diagnosed merely just by using these intensities. Their results demonstrated the ability of MRI to diagnose brain tumours. The aim of this thesis was to set up an automatic system to diagnose the brain tumour types from MR images, instead of taking a biopsy. This system would be much more comprehensive than the work of Siromoney *et al.* It would use original MR images and make a decision on the tumour types from them.

According to above statement, it is reasonable to assume that the brain tissue intensities and brain tumour intensities, or their relations in MR images may indicate the brain tumour type. To obtain the intensities of MR images, an image segmentation procedure and a preprocessing procedure are required which enable the original images to be segmented and meet other related requirements. The acquired intensities can constitute the classification rules for each tumour type. The structure of the automatic diagnosing system may therefore contain three major parts: image preprocessing, image segmentation and brain tumour classification (Fig.1.2).

Segmenting brain tissues precisely is a key step which determines the direction of our further work. However, the original MR images usually have various signal intensities. This makes the intensities in different images impossible to compare. Therefore the preprocessing procedure before segmentation is crucial, as it must standardize the signal intensities. This is the first major difficulty that needs to be addressed in this work. In addition, the preprocessing procedure must remove the non-brain regions from MR images due to the redundancy of the non-brain regions in the brain analysis. Presenting an easily manipulated method for the accurate removal of the non-brain regions should also be part of this system. Contrast enhancement is another important consideration in preprocessing because it may reduce the difficulty of the following segmentation.

Tissue segmentation is an imperative step before measuring tissue intensities. Manual segmentation by an expert operator is too time-consuming. It involves not only a large amount of work, but also a mass of data and its accuracy may be degraded by different human operators. An automatic segmentation method is therefore an important part of this work. A number of algorithms based on approaches such as histogram analysis, region growing, edge detection and clustering have been proposed in the literature [Pal and Pal, 1993; Robb, 2000]. However, they

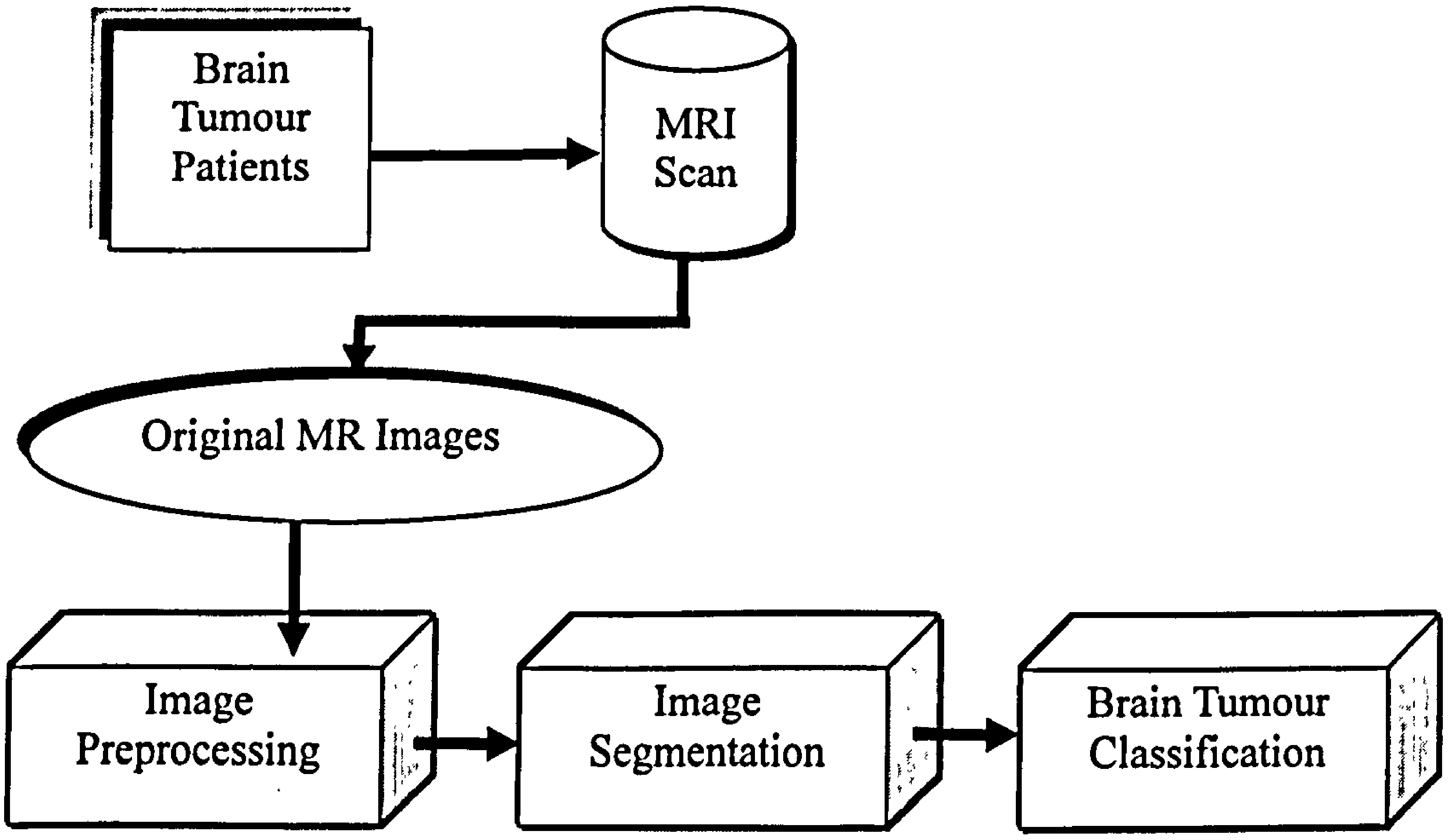


Figure 1.2 Structure of brain tumour diagnosing system.

all have some shortcomings. Therefore, the most important task in this thesis is to develop a robust algorithm for brain MR image segmentation, which should be robust for most types of MR images. The average intensities of different tissues should be easy to obtain in each image after segmentation.

The initial purpose of this thesis was to discover the rules for predicting the brain tumours using the tissue intensities acquired. A large number of data (tissue intensities) can be collected from previous processing. In order to discover useful rules from this data, data mining and knowledge discovery techniques would be employed. Tumour diagnosis may be categorized as a classification problem which aims to find classification rules for each brain tumour type and classify the tumour correctly. Classification methods used by data mining and knowledge discovery techniques include for example, decision tree induction, neural networks, Bayesian classification, genetic algorithms, genetic programming, and rough sets. Selecting a proper classification method can greatly improve the accuracy of tumour diagnosis and save computing time. This is therefore another important aspect of the thesis. Necessary improvements to match the features of the medical data must also be considered.

In summary, the primary goal of this thesis is to present a new method to diagnose brain tumours non-invasively. Three major parts will be needed for this. The first part is image preprocessing, from which the signal intensities of the brain tissues can be prepared for the following process of segmentation, by standardizing, non-brain removal and tissue contrast enhancement. The second one is segmentation. A powerful segmentation method needs to be developed to overcome the shortcomings of other segmentation methods. The last part is a data mining and knowledge discovery step. This is required to discover the classification rules for each brain tumour type using an appropriate classification method. It is expected that good accuracy of tumour diagnosis would be achieved.

1.3 THESIS ORGANIZATION

The first four chapters of this thesis cover the fundamental description of all fields involved in this thesis. The next three chapters review present work and propose novel procedures to diagnose brain tumours non-invasively. The last chapter presents conclusions.

Chapter 2 reviews the basic principle of MRI technique, the characteristics of MRI and the artifacts in MR images.

Chapter 3 introduces an overview of brain tumours. It includes tumour definition, types of brain tumours and the common diagnosis methods.

Chapter 4 presents a general introduction of data mining and knowledge discovery techniques. Fuzzy clustering and genetic programming, two essential methods in this approach, are fully described in this chapter and then provides the foundation of the later work.

Chapter 5 details the methodology applied in image preprocessing. It comprises three main sections: image standardization, the non-brain region removal and tissue contrast enhancement. The first section describes the standardization process for image intensities by means of statistical features of intensities. The second section presents the techniques for the removal of non-brain regions using morphological operations. The third section proposes the technique for tissue contrast enhancement, especially between white matter and gray matter, using the histogram equalization.

Chapter 6 introduces an improved fuzzy clustering algorithm developed for MR image segmentation. Neural networks are also used to decide the optimized parameters for this algorithm. The robustness of the proposed method to noise is evaluated by various types of images and also compared to other segmentation methods described in the literature.

Chapter 7 presents a fuzzy logic based genetic programming method to deal with the data mining task of brain tumour classification. It overcomes the drawback in present classification methods in multi-class classification problems. The characteristics of medical data are comprehensively considered during the classification. Its performance is compared and evaluated with other classification methods by applying it to Iris data and Wisconsin breast cancer data. The classification rules for different MRI modalities and different brain tumour types are described.

Chapter 8 summarizes the thesis with the main conclusions and recommendation for future work.

Chapter 2

MAGNETIC RESONANCE IMAGING

2.1 INTRODUCTION

2.1.1 History of MRI

Magnetic resonance imaging (MRI) has become the primary technique in the routine diagnosis of many diseases, replacing and sometimes surpassing computed tomography (CT). MRI has particular advantages because it is non-invasive, using non-ionizing radiation, and has a high soft-tissue resolution and discrimination in any imaging plane. It may also provide both morphological and functional information. The resultant MR image is based on multiple tissue parameters used, any of which can modify tissue contrast. In its development MRI has incorporated a multi-disciplinary team of radiologists, technicians, clinicians and scientists who have made, and are continuing to make combined efforts to further extend the clinical usefulness and effectiveness of this technique.

The first successful nuclear magnetic resonance (NMR) experiment was made in 1946 independently by two scientists in the United States [Ellard, 2003]. Felix Bloch, working at Stanford University, and Edward Purcell, from Harvard University, found that when certain nuclei were placed in a magnetic field they absorbed energy in the radiofrequency range of the electromagnetic spectrum, and re-emitted this energy when the nuclei returned to their original state. The strength of the magnetic field and the radiofrequency matched each other as earlier demonstrated by Sir Joseph Larmor (Irish physicist 1857-1942) and is known as the Larmor relationship (i.e., the angular frequency of precession of the nuclear spins being proportional to the strength of the magnetic field). This phenomenon was termed as NMR.

With this discovery, NMR spectroscopy was invented and soon became an important analytical method in the study of the composition of chemical compounds. For this discovery Bloch and Purcell were awarded the Nobel Prize for Physics in 1952. Interestingly, Dr Isidor Rabi, an American physicist who was awarded the Nobel Prize for Physics in 1944 for his invention of the atomic and molecular beam magnetic resonance method of observing atomic spectra, came across the NMR phenomenon in the late 1930's but considered it to be an artifact of his apparatus and disregarded its importance [Ellard, 2003]. During the 50's and 60's NMR spectroscopy became a widely used technique for the non-destructive analysis of small samples. Many of its applications were at the microscopic level using small (a few centimeters) bore high field magnets.

In the late 60's and early 70's Raymond Damadian, an American medical doctor at the State University of New York in Brooklyn, demonstrated that a NMR tissue parameter (termed T_1 relaxation time) of tumour samples, measured in vitro, was significantly higher than normal tissue. Although not confirmed by other workers, Damadian intended to use this and other NMR tissue parameters not for imaging but for tissue characterization (i.e., separating benign from malignant tissue). This has remained the "Holy Grail" of NMR which is yet to be achieved due mainly to the heterogeneity of tissue. His description of relaxation time changes in cancer tissue was one of the main impetuses for the introduction of NMR into medicine.

In 1973, a short paper was published in Nature entitled "Image formation by induced local interaction; examples employing magnetic resonance" by Paul Lauterbur, a professor of Chemistry at the State University of New York. In this seminal paper Lauterbur proposed using magnetic field gradients to distinguish between NMR signals originating from different locations. He termed this new imaging technique as zeugmatography (from the Greek zeugmo meaning yoke or a joining together) [McRobbie *et al.*, 2003]. This referred to the joining together of a weak gradient magnetic field with the stronger main magnetic field allowing the spatial localization of two test tubes of water. This imaging experiment moved from the single dimension of NMR spectroscopy to the second dimension of spatial orientation being the foundation of MRI. The technique was called NMR imaging rather than MRI. The 'N' was dropped because the term 'nuclear' implied a connection with nuclear energy.

Lauterbur's idea revolutionized NMR because it opened the field to imaging. As a result, selective excitation or sensitization of tomographic image slices was invented at the University of Nottingham, England in 1974 by Alan N. Garroway, Peter K. Grannell, and Peter Mansfield [EMRF Foundation, 2003]. In 1975, Richard Ernst's group in Zurich invented two-dimensional Fourier transform imaging (2D FT). The first practical 2D FT imaging method, dubbed 'spin warp', was developed by Bill Edelstein and Jim Hutchison at the University of Aberdeen, Scotland in 1980.

There are other milestones in the development of MRI for medical applications. By 1975, Peter Mansfield and Andrew A. Maudsley proposed a line technique which, in 1977, led to the first image of in vivo human anatomy, a cross section through a finger. In 1978, Mansfield presented his first image through the abdomen. In 1977, Hinshaw, Paul Bottomley, Neil Holland, Moore, and Brian Worthington and collaborators succeeded with an image of the wrist. More human thoracic and abdominal images followed, and by 1978, Hugh Clow and Ian R. Young, working at EMI, reported the first transverse NMR image through a human head. Two years later, William Moore and colleagues presented the first coronal and sagittal images through a human head.

All of the research described above contributed to a completely new imaging technique. Nowadays, MRI is one of the most popular techniques for clinical imaging where the potential contrast between normal tissues and abnormal tissues is many times greater than that offered by X-ray technology and ultrasound.

2.1.2 Basic Principles of MRI

Clinical MRI uses the magnetic properties of hydrogen and its interaction with both a large external magnetic field and a radio frequency (RF) to produce highly detailed images of the human body. Fig.2.1 shows a sketch of MRI device. MRI is a dynamic and flexible technology that allows one to tailor MR images to the anatomical part of interest and to abnormal changes in the body.

In order to understand MRI, it is necessary to understand the properties of atoms. The atom consists of a central nucleus and orbiting electrons. The nucleus contains nucleons which are subdivided into protons and neutrons; protons are positive charged, neutrons have no charge and electrons are negatively charged. The atomic number is the number of protons in the nucleus which determines the type of element the atoms make up. Mass number is the sum of the neutrons and protons in the nucleus. In a stable atom, the number of negatively charged electrons equals the number of positively charged protons. Atoms with a deficit or excess number of electrons are called ions [Westbrook, 2002].

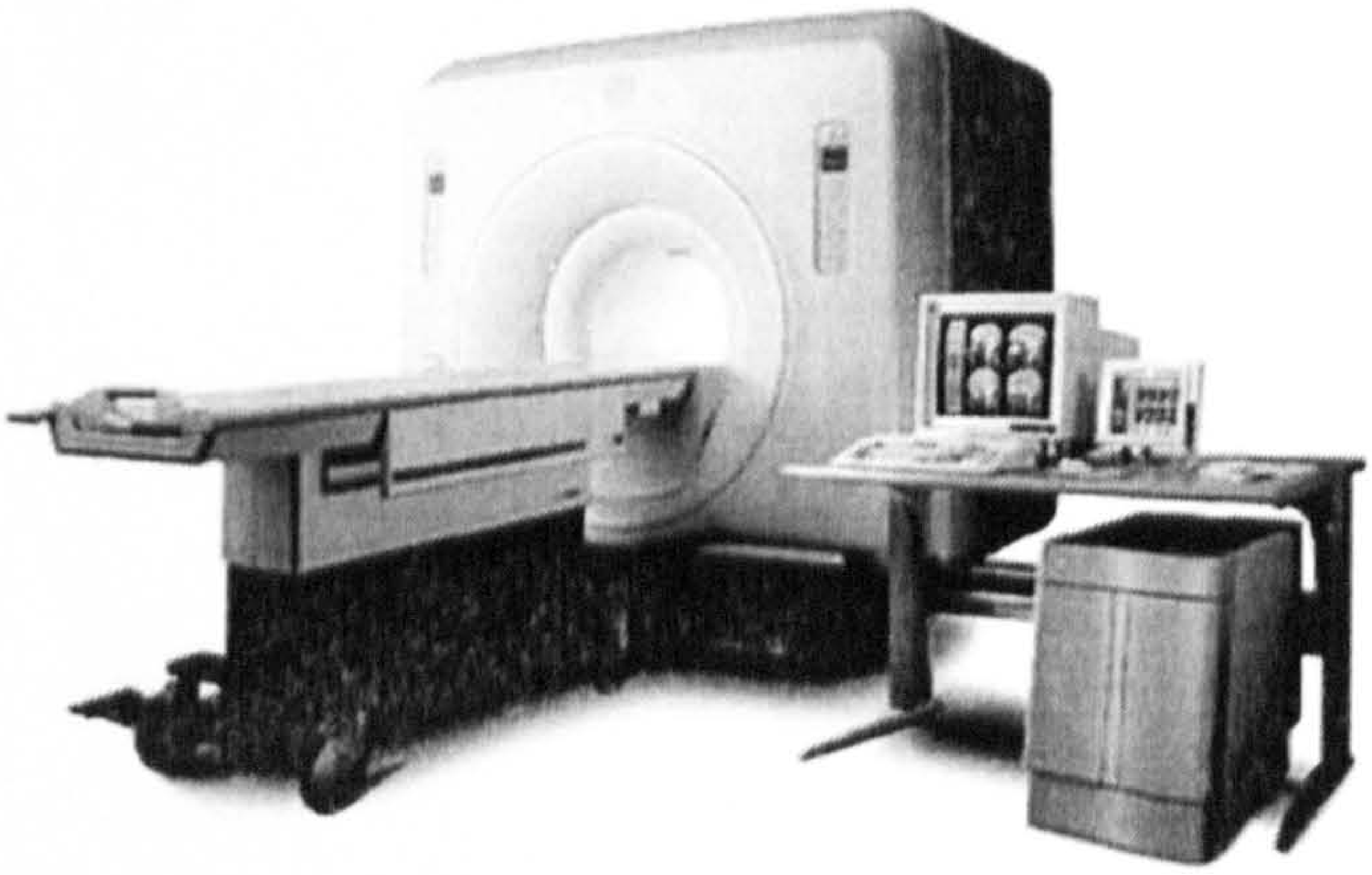
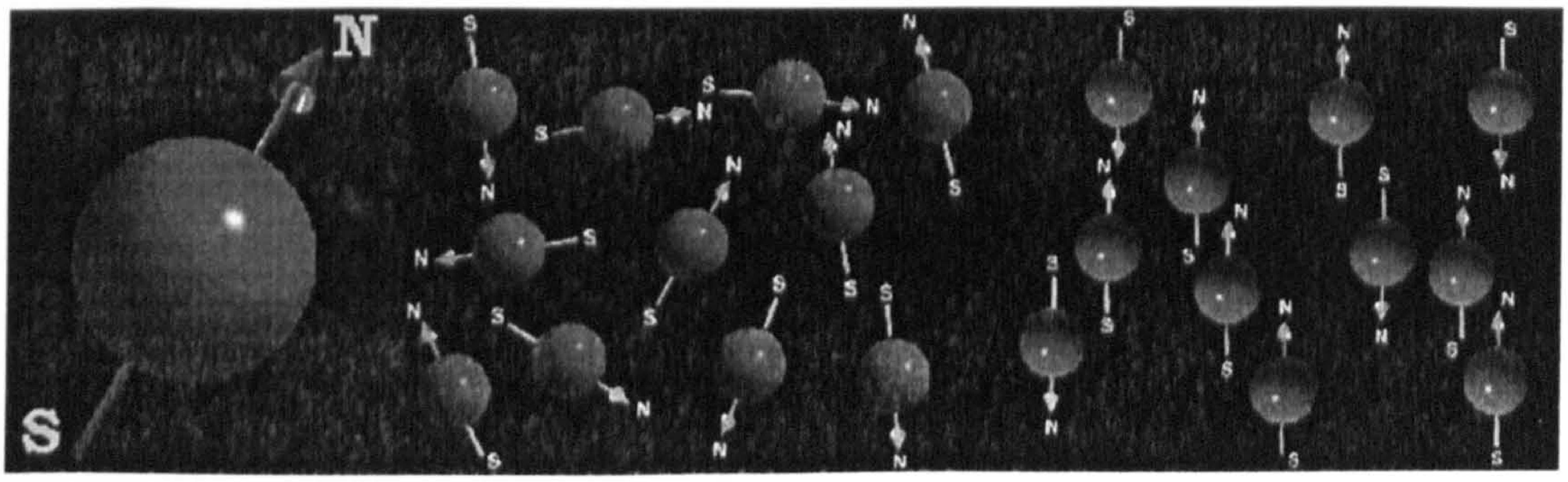


Figure 2.1 MRI machine [Wilson Memorial Hospital 2003].

Three types of motion are present within the atom. Electrons spin on their own axis and orbit the nucleus, the nucleus itself spin on its own axis. The principles of MRI rely on the spinning motion of specific nuclei in biological tissues, known as MR active nuclei. The hydrogen nucleus is the MR active nucleus used in clinical MRI because it is abundant in the human body (in fat and water, 70% of the body is made up of water). The simple hydrogen nucleus consists of one proton, and no neutrons. Due to the presence of only one proton, the hydrogen atom has a positive charge and an atomic number of 1. Any electrically charged particle which moves creates a magnetic field called a magnetic moment. Therefore the hydrogen nucleus induces a magnetic field around itself. The other reason for the use of hydrogen nucleus is that its solitary proton provides a relatively large magnetic moment. The magnetic moment of hydrogen is called the net magnetization vector (NMV) [Westbrook and Kaut, 2002].

Usually the magnetic moments are randomly orientated without an external magnetic field. When a human body is placed in a large magnetic field B_0, many of the free hydrogen nuclei align themselves with the direction of the magnetic field and constitute the NMV of the human body (Fig.2.2). Each hydrogen nucleus that makes up the NMV spins on its axis. The influence of the external magnetic field B_0 produces an additional spin, or wobble of the NMV around B_0. This secondary spin is called precession. The speed at which the NMV wobbles around B_0 is called the precessional frequency. The value of the precessional frequency is proportional to the



(a) (b) (c)

Figure 2.2 (a) Spinning proton (b) No magnetic field present (c) Magnetic field present [King 2003].

strength of the applied magnetic field. The stronger the magnetic field, the higher the precessional frequency. It is described as follows:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

This expression is stated as the Larmor equation, so the precessional frequency is often called the Larmor frequency, where γ is the gyromagnetic ratio, which is a constant and expressed as the precessional frequency of a specific MR active nucleus at 1.0 T. The gyromagnetic ratio of hydrogen is 42.57MHz/T [Mackiewich, 1995].

Resonance refers to the property of an atom to absorb energy only at the Larmor frequency. Energy at the Larmor frequency of hydrogen at all magnetic field strengths in clinical MRI corresponds to the radio frequency (RF). To induce resonance in hydrogen, an RF pulse of energy must be applied. The energy must also be delivered at 90° to the NMV and main magnetic field (B_0). In the first part of resonance, the hydrogen nuclei absorb energy from the RF pulse. As a result, NMV moves out of alignment, 90° away from B_0, and lies in the transverse plane. Then, the magnetic moment of the hydrogen nuclei within the transverse NMV moves into phase with each other.

A receiver coil is situated in the transverse plane. As the NMV rotates around the transverse plane as a result of resonance, it passes across the receiver coil inducing a voltage in it according to Faraday's law. This voltage is the MR signal. Once the RF pulse is removed, the energy of the NMV given by the RF pulse starts to decrease because the NMV tries to realign with B_0. The amplitude of the MR signal consequently decreases. This is called free induction decay (FID).

Gradients

The imaging system must be able to locate the detected signal spatially in three dimensions, so that it can position each signal at the correct point on the image. This task is performed by magnetic field gradients. Gradients are alterations to the main magnetic field and are generated by coils of wire located within the bore of the magnet through which current is passed. The passage of current through a gradient coil induces a gradient magnetic field around it, which either subtracts from or adds to the main static magnetic field strength B_0. The strength of B_0 is altered in a linear fashion by the gradient coils, so that the magnetic field strength and therefore the precessional frequency experienced by nuclei situated along the axis of the gradient can be predicted. This is called spatial encoding.

Slice selection

Nuclei experience an increased magnetic field strength whose precessional frequency increases and vice versa. Therefore the position of a nucleus along a gradient can be identified according to its precessional frequency. There are three gradient coils situated within the bore of the magnet and these are named according to the axis along which they act when they are switched on, i.e. The Z gradient alters the magnetic field strength along the Z axis of the magnet. The magnetic centre is the centre point of the axis of all three gradients and the magnetic field strength always remains the same as B_0. If the Z gradient is on, the magnetic field strength and therefore the precessional frequency of nuclei located along the Z axis vary linearly from one end of the magnet to the other. Therefore a specific point along the Z axis has a specific precessional frequency. In this way, a single transverse slice can be selectively excited without the excitation of other slices. If a gradient along the X axis were used instead of the Z gradient, the slices selected would be a sagittal slice. Similarly, a Y gradient would select a coronal slice. By the combination of X, Y and Z gradients, any plane may be chosen.

Frequency encoding

Once a slice has been selected, the signal coming from the slice must be located along both axes of the image. The signal is usually located along the long axis of the anatomy by a process known as frequency encoding. When the frequency encoding gradient is switched on, the magnetic field strength and therefore the precessional frequency of signal along the axis of the gradient vary linearly. The gradient produces a frequency difference or shift of signal along its axis. The signal can now be located along the axis of the gradient according to its frequency. The direction of frequency encoding can be selected by the operator so that it encodes the signal along the long axis of the anatomy. In coronal and sagittal images, the Z gradient performs frequency encoding. In transverse images, the X gradient performs frequency encoding, but the Y gradient for brain images.

Phase encoding

The signal is located along the short axis of the anatomy by a process known as phase encoding. When the phase encoding gradient is on, the precessional frequency of nuclei along the axis of the gradient is altered. As the speed of precession of the nuclei changes, so does the accumulated phase of the magnetic moments along their precessional path. Nuclei that have sped up, due to the presence of the gradient, move further around their precessional path. Nuclei that have slowed down, due to the presence of the gradient, move further back around their precessional path. There is now a phase difference or shift between nuclei positioned along the axis of the gradient. After the phase encoding gradient is switched off, the magnetic field strength experienced by the nuclei returns to the main field strength B_0. Therefore the precessional frequency of all the nuclei returns to the Larmor frequency. However, the phase difference between the nuclei remains. This difference in phase between the nuclei is used to determine their position along the phase encoding gradient.

The basic principles of MRI have been described [Westbrook and Kaut, 1994; Bushong, 1995] and the details of different types of RF pulse sequences are described in the next section.

2.2 MRI PARAMETERS

Many parameters are introduced in MRI. The most common ones are the strength of magnetic field, the Larmor frequency, the relaxation times (T_1, T_2), the repetition time (TR), the echo time (TE). These are all relevant to MR image generation.

The magnetic field and Larmor frequency has already been described in section 2.1. The strength of the magnetic field used in most clinical MRI units is 1.5 Tesla or above and the corresponding Larmor frequency of the protons is 64 MHz.

As stated above, after the removal of the RF pulse, the energy of NMV decreases. The process is called relaxation. Relaxation then results in the recovery of magnetization in the longitudinal plane called T_1 recovery and decay of magnetization in the transverse plane called T_2 decay. Different tissues undergo different rates of relaxation which create image contrast. T_1 recovery is caused by the nuclei emitting their energy to the surrounding lattice known as spin lattice relaxation. Energy released to the surrounding lattice causes the nuclei to recover their magnetization. The rate of recovery is an exponential process, with a recovery time constant called T_1. T_1 is the time that 63% of the longitudinal magnetization is recovered. T_2 decay is caused by nuclei exchanging energy with neighbouring nuclei known as spin-spin relaxation. It results in a decay of transverse magnetization. The rate of decay is also an exponential process with a decay time constant T_2. T_2 is the time that 63% of the transverse magnetization decayed.

The repetition time (TR) is the time from the application of one RF pulse to the application of the next RF pulse, measured in milliseconds (ms). It determines the amount of T_1 recovery that will occur. The echo time (TE) is the time from the application of the RF pulse to the peak of the signal induced in the coil and is measured in ms. TE controls the amount of T_2 decay that will occur. Fig.2.3 shows the sketch of TR and TE.

Three types of MR images are commonly used, T_1-weighted, T_2-weighted and proton density (PD) weighted. A T_1-weighted image is an image whose contrast is predominated by T_1 signals. To achieve this, TR must be short in order to exaggerate T_1 and TE must be short in order to diminish T_2. A T_2-weighted image is created by a long TE and a long TR whose contrast in predominately determined by T_2 signals, because TR is long to diminish T_1 and TE is long to exaggerate T_2. A PD weighted image is an image dependent primarily on the density of protons by minimizing the effects of T_1 and T_2. A long TR and short TE are chosen to diminish both T_2 and T_1

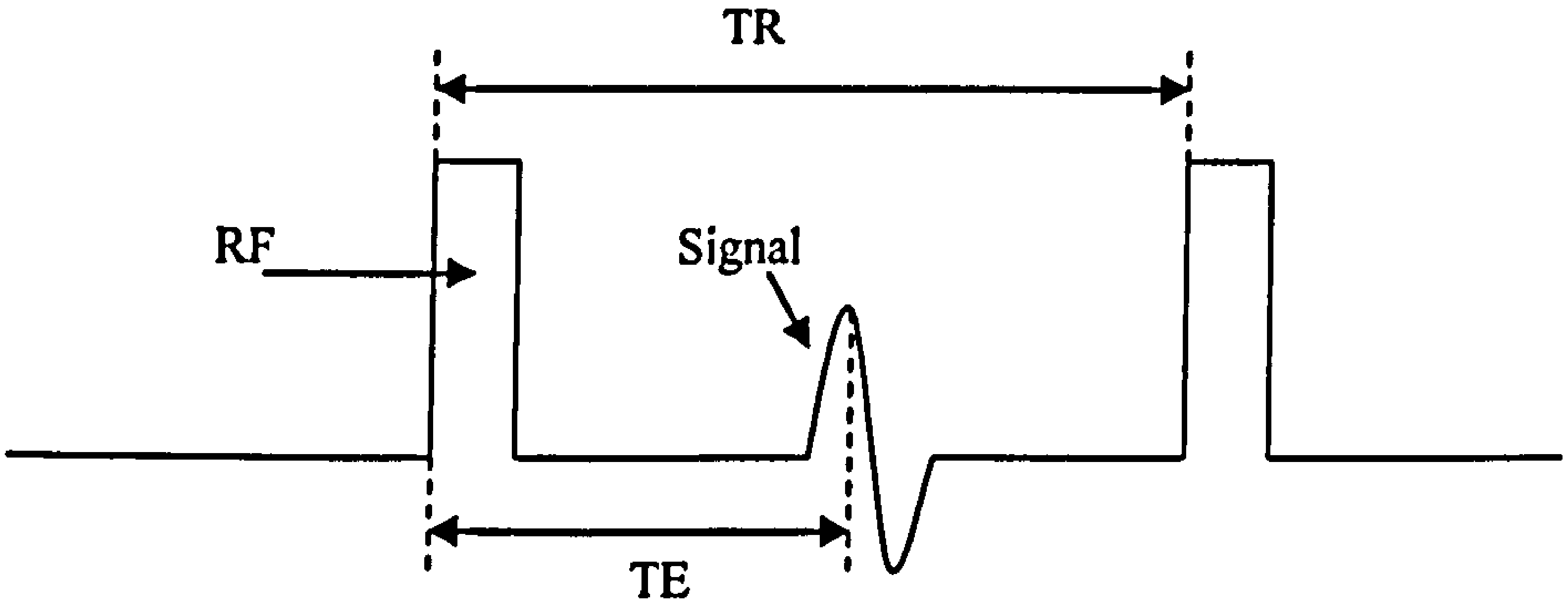


Figure 2.3 TE and TR.

[Fonar Corporation, 2003]. Typical values of TR and TE are [Westbrook and Kaut, 2002]:

Long TR: over 2000 ms
Short TR: 250-700 ms
Long TE: over 60 ms
Short TE: 10-25 ms

In T_1-weighted images, tissues with short T_1 recovery time such as fat are bright (high signal). Because they recover most of their longitudinal magnetization during TR, more magnetization is available in order to allow them to be flipped into the transverse plane by the next RF pulse. In contrast, they are dark (low signal) in T_2-weighted images. Tissues with long T_2 decay time are bright, such as water, because they retain most of their transverse coherence during the TE period. In PD-weighted images, tissues with a high proton density, such as cortical bone are bright because the high number of protons results in a large component of transverse magnetization. Table 2.1 shows the brightness of some typical tissues in different types of MR images.

Table 2.1 Brightness of typical tissues in different MR images [Hesselink, 2003].

Image types / Tissue	T_1-weighted	T_2-weighted	PD-weighted
Fat	**Bright**	**Dark**	**Bright**
Cyst	**Dark**	**Bright**	**Grey**
White matter	**Bright**	**Grey**	**Grey**
Gray matter	**Grey**	**Bright**	**Bright**
CSF	**Dark**	**Bright**	**Grey**

2.3 MR SIGNALS

Based on the introduction above, the implementation of MRI required the body to be placed in a main, uniform magnetic field which is then is excited with another oscillating magnetic field at the Larmor frequency. An MR signal of a slice is thus generated. The signal expression is as following assuming that the nuclear magnetization is fully relaxed before the following 90° RF pulse applied:

$$S(t) = k\int\int \rho(x,y)e^{-t/T2}e^{-i\omega_0 t}dxdy \tag{2.2}$$

Where ω_0 is the Larmor frequency, $\rho(x,y)$ is the spin spectral density function which is specific to different tissues. (x,y) are the spatial coordinates. k is a proportional coefficient. The detected magnitude of MR signal is obtained from the real part:

$$S_r(t) = k\int\int \rho(x,y)e^{-t/T2}\cos\omega_o t dxdy \tag{2.3a}$$

and the imaginary part:

$$S_i(t) = k\int\int \rho(x,y)e^{-t/T2}\sin\omega_o t dxdy \tag{2.3b}$$

A set of RF pulses applied to produce a specific form of MR signal is called a pulse sequence. Signals are determined by variable parameters based on pulse sequences used [Hornak, 1996]. Spin echo (SE) pulse sequences are one of the most basic pulse sequences used in MRI. It consists of a 90° excitation pulse followed by a 180° rephrasing pulse and an echo. An echo is in fact the signal generated in the receiver coil as shown in Fig.2.4.

The MR signal of repeated SE pulse sequences can be defined as a function of T_1, T_2, TR and TE [Stark and Bradley, 1999]. Its maximum amplitude is:

$$S = k\rho(1 - e^{-TR/T1})e^{-TE/T2} \tag{2.4}$$

Fast spin echo (FSE) sequence is a simple extension of SE sequence. It uses a 90° pulse followed by a series of 180° rephrasing pulses to produce multiple echoes in a given TR [Westbrook and Kaut, 1994].

Inversion recovery (IR) sequences are another type of pulse sequence. They begin with a 180° inverting pulse followed by a 90° pulse and a FID is generated after them. It is shown in Fig.2.5. The maximum MR signal of no repeated IR sequence can be expressed as follows:

$$S = k\rho(1 - 2e^{-TI/T1}) \tag{2.5}$$

Where, TI is the time between the 180° inverting pulse and the following 90° pulse, and is known as inversion time. If the IR sequence repeats at every TR seconds, the signal becomes:

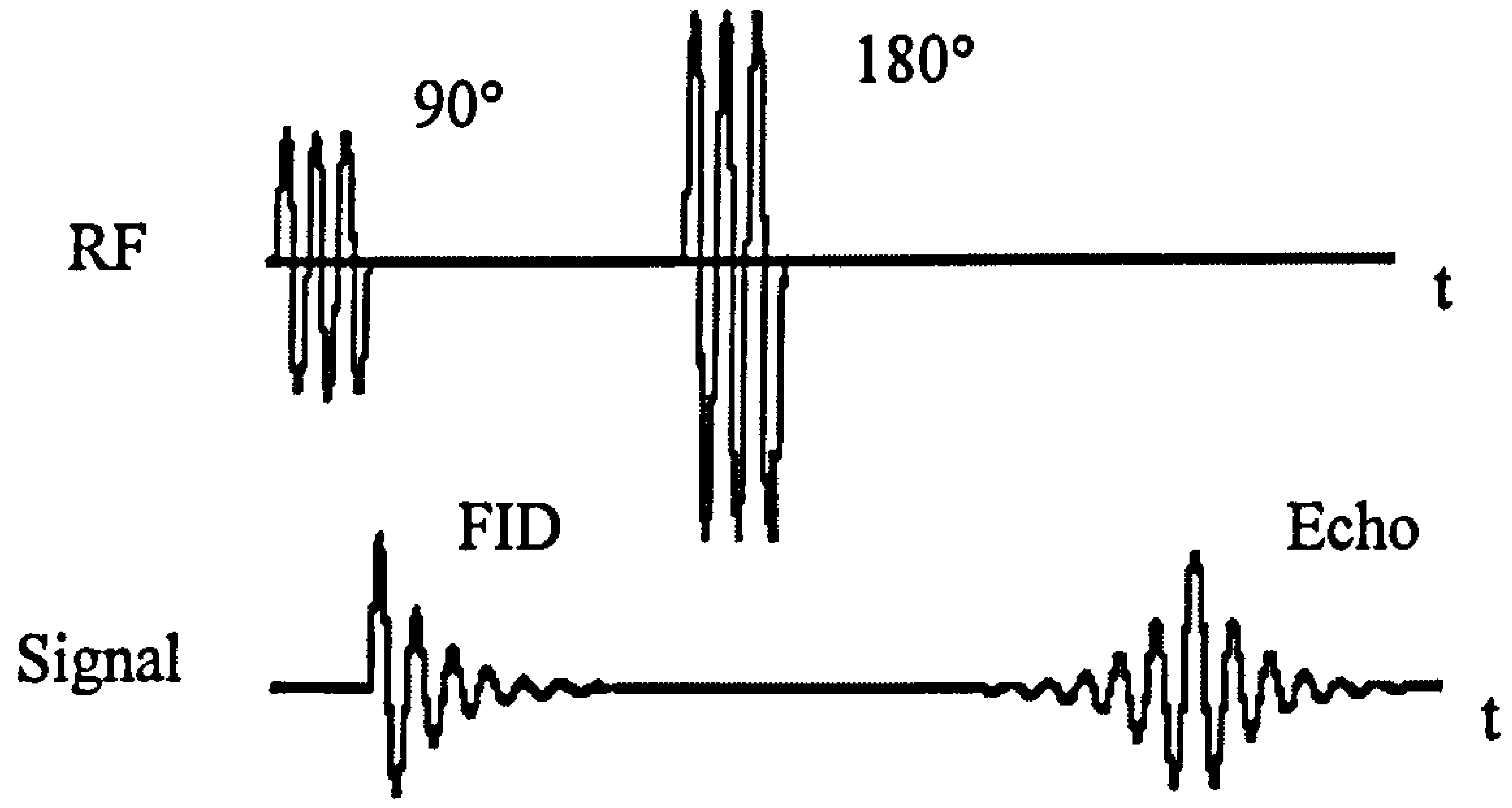


Figure 2.4 Spin echo sequence and signal.

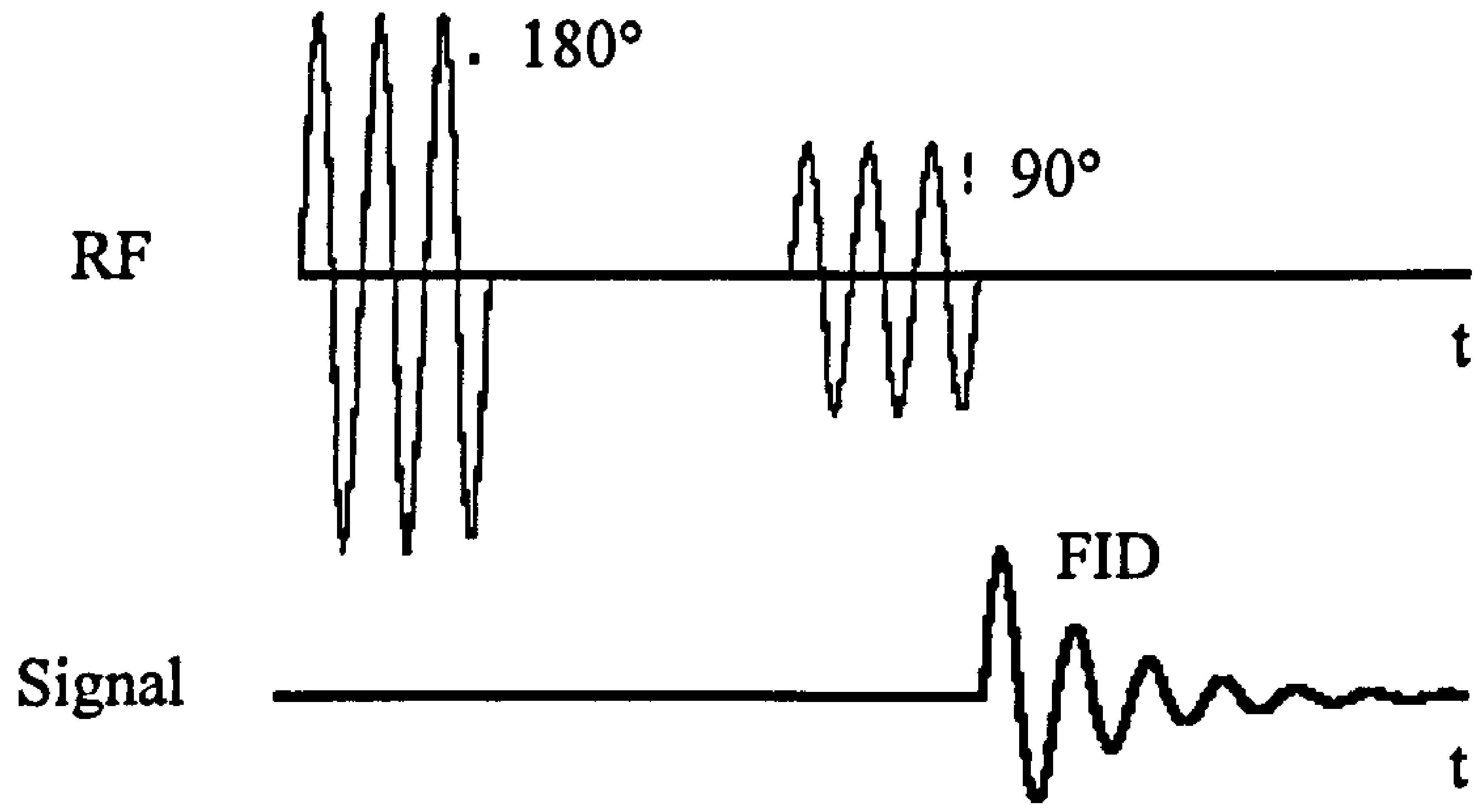


Figure 2.5 Inversion recovery sequence.

$$S = k\rho(1 - 2e^{-TI/T1} + e^{TR/T1}) \tag{2.6}$$

Fluid attenuated inversion recovery (FLAIR) is a variation of IR sequence. It is an MRI pulse sequence in which fluids such as CSF appears dark and most lesions including MS plaques, other white matter lesions, tumours, edema, and acute infarcts appear bright.

However, this MRI signal is a sum from all parts of the human body. Because the human body is not spatially homogeneous, it is necessary to differentiate the signals from different parts and reconstruct MR images. By far the most common method to reconstruct MR images is 2D FT (2-dimenional Fourier Transform), also called 'spin-warp'. The details of MR image reconstruction can be found in [Liang, 2003].

2.4 MRI ARTIFACTS

Artifacts are features in MR images produced by various complications of the imaging process. They result in an image that does not portray (in the simple visual sense) an accurate representation of a slice of tissue [Savoy and Jovicih, 2001]. Examples of sources of artifacts include chemical shift, aliasing, RF inhomogeneity, motion, flow, truncation, and partial volume. Almost all MR images have to some extent artifacts and some of these may significantly affect the quality of MR images and even cause incorrect diagnosis. This section presents the most common artifacts encountered in MRI [Ballinger, 1996].

Chemical shift artifact:

The chemical shift artifact is commonly noticed in the spine at the vertebral body end plates, in the abdomen, and in the orbits where fat and other tissues form borders. It is caused by the different chemical environment of fat and water. Although fat and water are both made up of hydrogen protons, fat consists of hydrogen linked to carbon, whereas hydrogen in water is lined to oxygen. As a result, hydrogen in fat resonates at a lower Larmor frequency than that in water. Therefore a frequency shift is inherently presented between fat and water, known as chemical shift. Its magnitude depends on the magnetic strength. A low magnetic strength for scanning can reduce this type of artifact. In Fig.2.6, the arrow shows the location of a chemical shift artifact.

Aliasing or wrap around artifact:

This is a common artifact produced when the field of view (FOV) is smaller than the anatomy being imaged. The FOV is the physical size of the imaged region [Siemens, 2001]. The tissue outside the selected FOV still produces a signal if it is close to the receiver coil. This signal is mis-mapped into pixels within the FOV. The primary solution to this artifact is oversampling, which is the digitization of a time domain signal at a frequency much greater than necessary to record the desired FOV [Aziz and Uetani, 2002]. Fig.2.7 shows the wrap-around of the back of the head on to the front of the head.

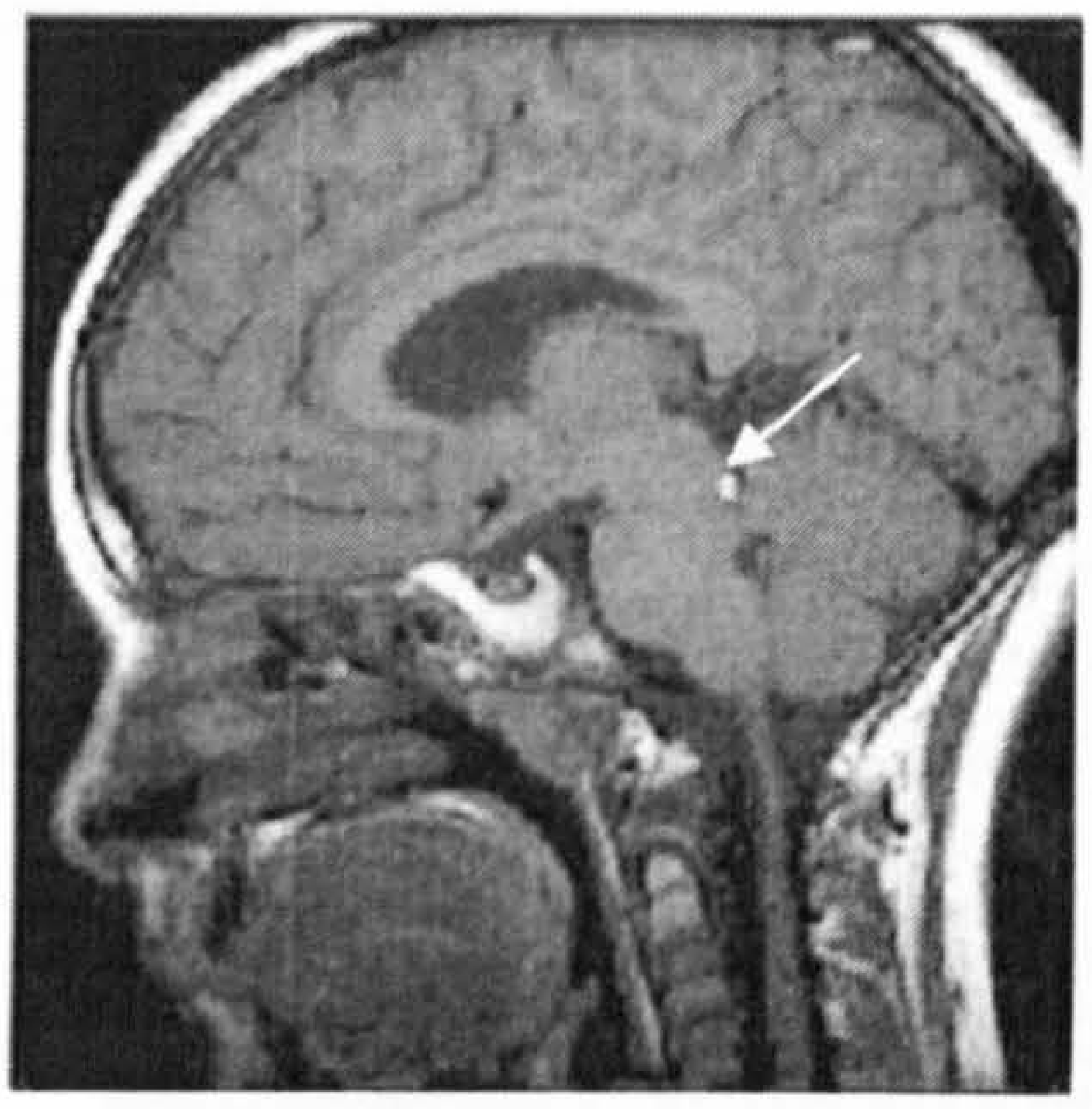


Figure 2.6 Chemical shift artifact [MRI-CHUQ 2003].

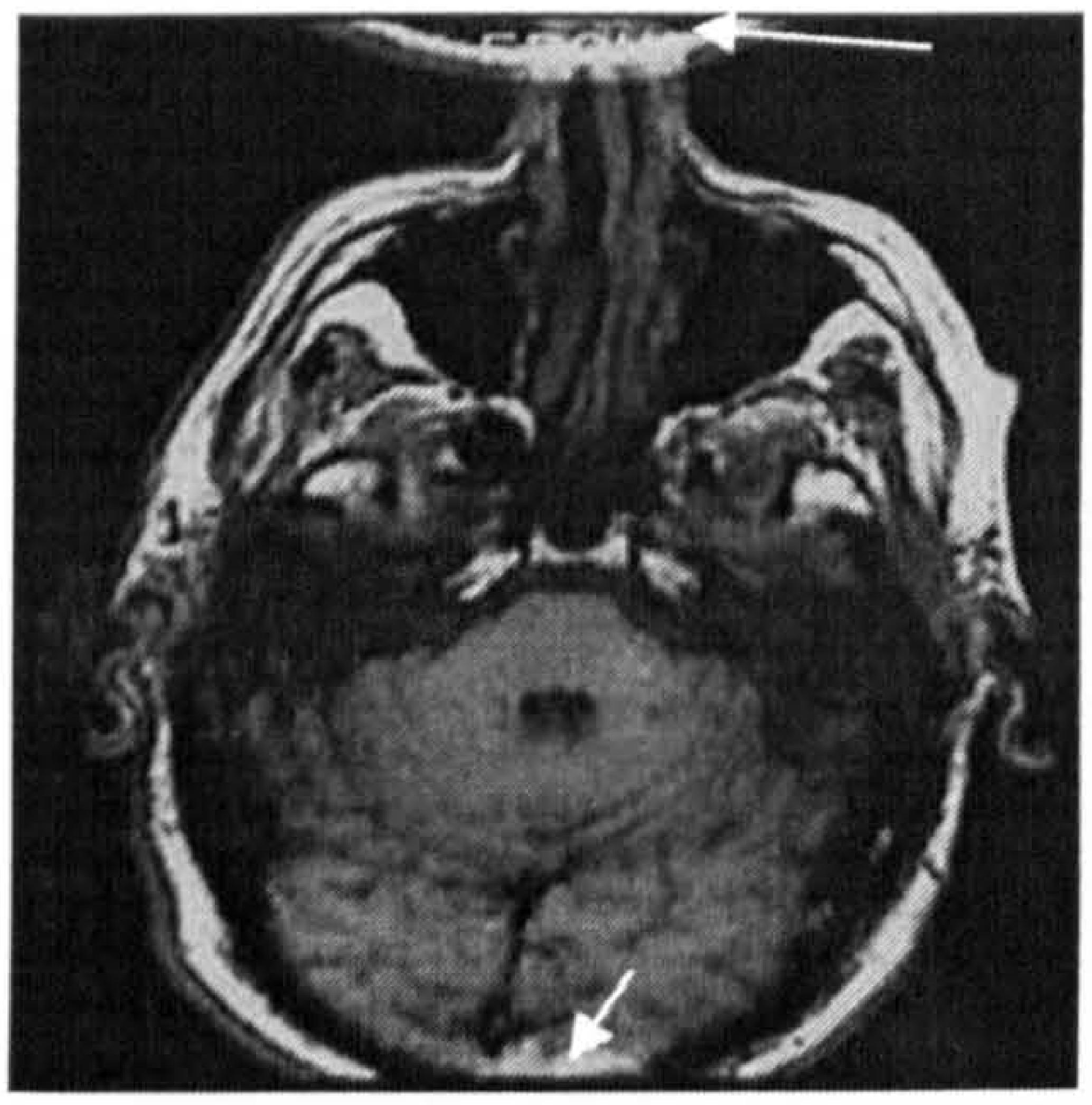


Figure 2.7 Aliasing artifact [Ballinger 1996].

RF inhomogeneity:

RF inhomogeneity is an intensity variation across an image. It is caused by variable sensitivity in an RF coil. Some RF coils, such as surface coils, have natural variations in sensitivity and will always display this artifact. The presence of this artifact in other coils represents the failure of an element in the RF coil or the presence of non-ferromagnetic material in the imaged object. Therefore, the selection of RF coils and magnets may reduce this artifact. Superconductive magnets are the most commonly used magnets which can maintain a homogenous magnetic field over time [Aziz and Uetani, 2002; Douglas 1998]. The transverse image of the brain shown in Fig.2.8 contains an RF inhomogeneity artifact in anterior region of the brain.

Motion artifact:

Any movement of the patient can cause this artifact. The patient motion during the imaging sequence generally results in a blurring of the entire image with ghost images in the phase encoding direction. Movement of a part of the patient results in a blurring of the corresponding part across the image. The voluntary motion of patients can usually be prevented, but the involuntary motion such as heart beating, breathing, bowel motion etc cannot be eliminated. The solution for the first two cases is to gate the imaging sequence to the cardiac or respiratory cycle of the patient. Bowel motion can be reduced by giving the patient an anti-spasmodic agent prior to the scan when imaging the abdomen or pelvis [Westbrook and Kaut, 2002b; Aziz and Uetani, 2002]. Fig.2.9 shows a head axial image with motion artifact. A blood vessel in the posterior

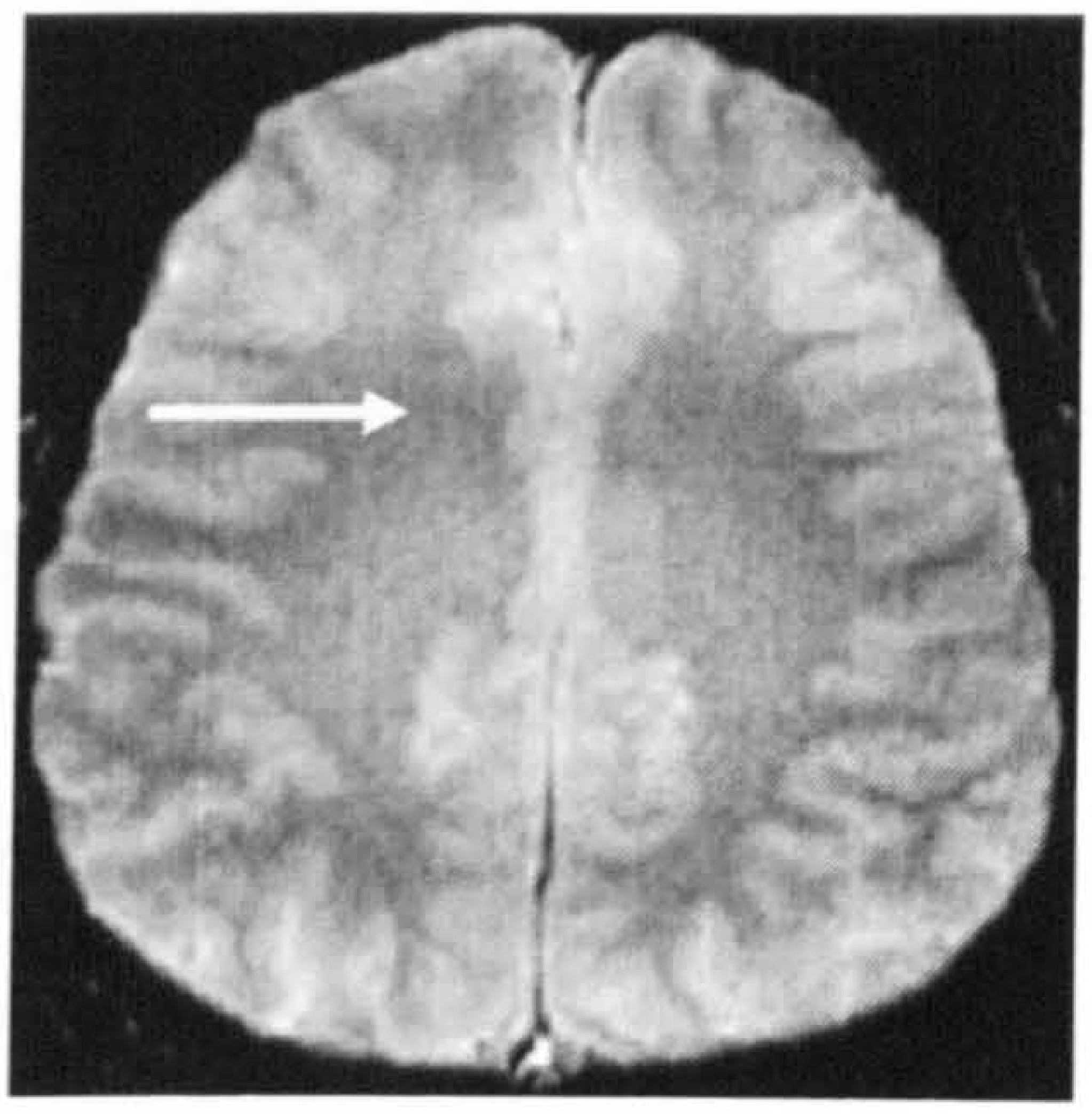


Figure 2.8 RF inhomogenity [Hornak 1996].

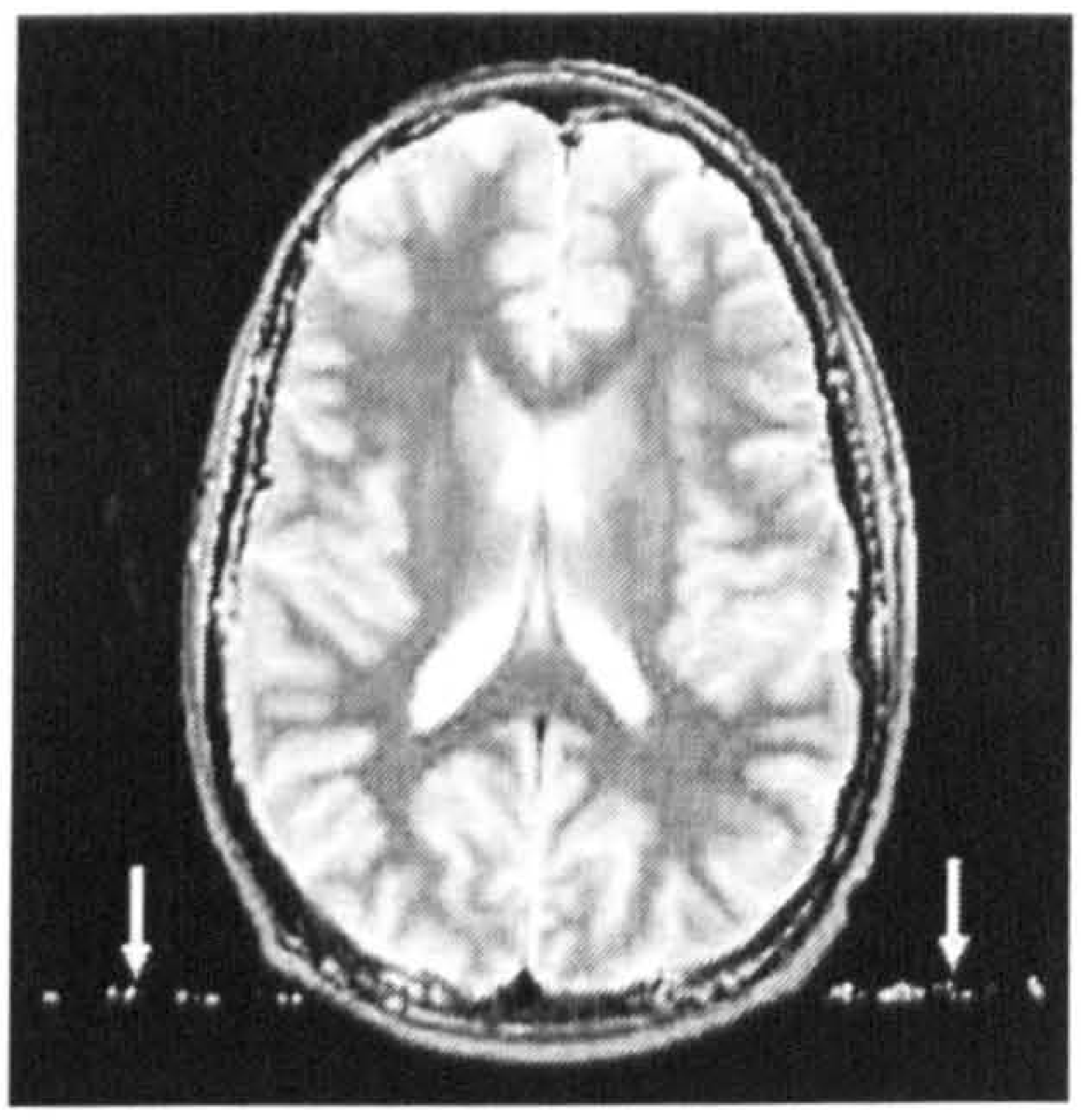


Figure 2.9 Motion artifact [Hornak 1996].

side of the head moved in a pulsating motion during the acquisition. This motion caused a ghosting across the image.

Flow artifact:

This is caused by flow of blood or fluids in the body. A liquid flowing through a slice can be subjected to an RF pulse and disappear by the time the signal is recorded. It may result in different signal intensities of blood vessels. For example, the intensity of a vessel perpendicular to the image plane changes periodically due to pulsatile blood flow [Douglas, 1998; Siemens, 2001]. Fig.2.10 shows a T_2-weighted axial image of the spine. Note the appearance of two spinal cords. The artifacts produced extra spinal cord is due to pulsatile flow of the CSF.

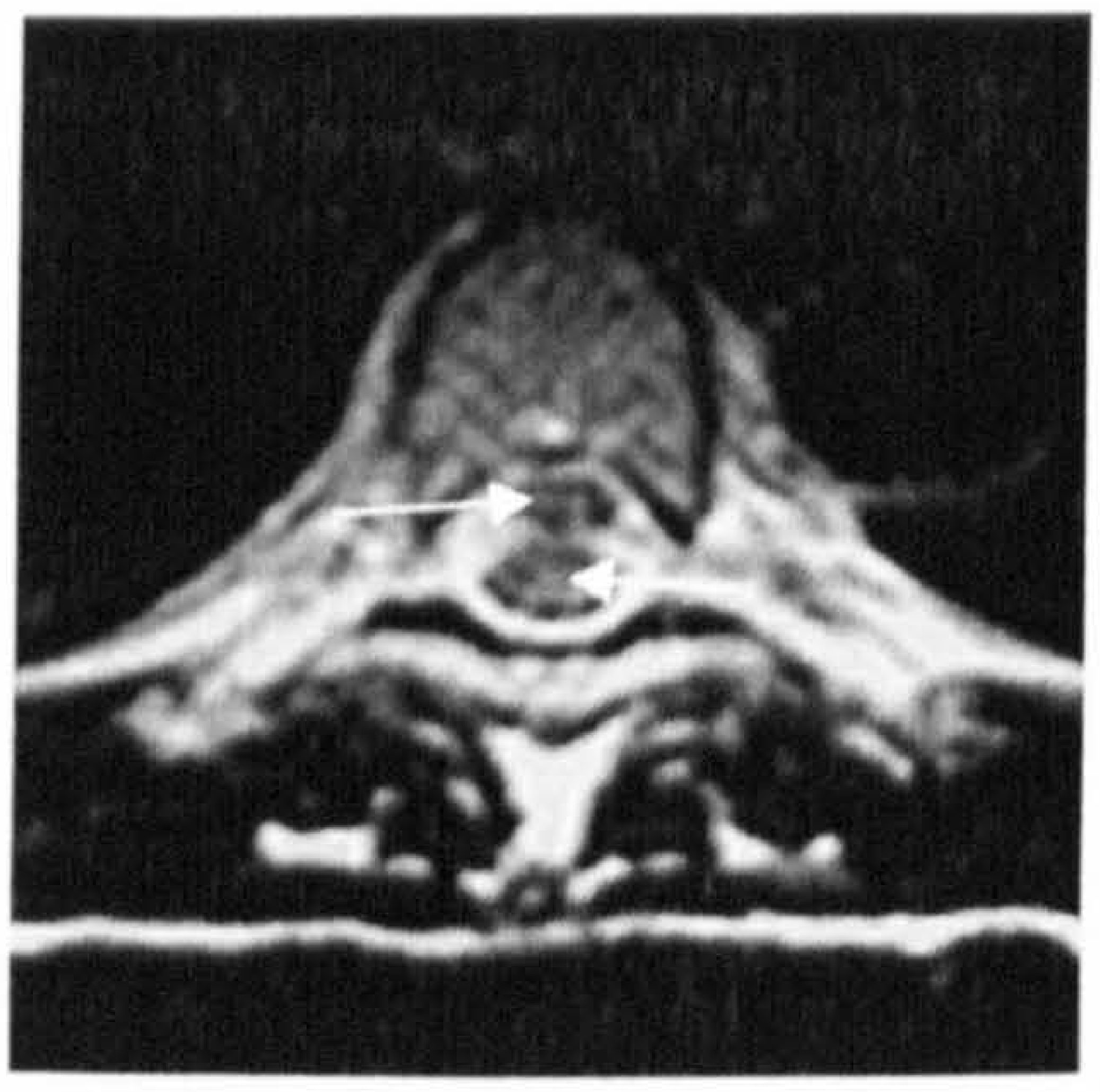


Figure 2.10 Flow artifact [Patola and Coulter 1997].

Figure 2.11 Truncation artifact [Patola and Coulter 1997].

Truncation artifact:

Truncation artifact results from under sampling of data so that interfaces of high and low signal are incorrectly represented on the image. Usually bright or dark lines are seen parallel and adjacent to borders of abrupt intensity change. A common site for this artifact is in the T_2-weighted image, where there is bright CSF next to dark spinal cord. It is also seen in other locations such as the brain and calvarium interface. To reduce this artifact, under sampling should be avoided [Westbrook and Kaut, 2002b; Ballinger 1996]. The fine lines visible in Fig.2.11 are due to under sampling of the high spatial frequencies. Sharp edged borders between areas of high contrast are represented by high spatial frequency data.

Partial volume:

A partial volume artifact is any artifact which is caused by the size of the image voxel (the smallest discrete spatial component of the image). For example, a small voxel may contain a signal from one tissue type, and a larger voxel may contain a combination of two types, whose signal intensity is equal to the weighted average of the quantity of two tissues in the voxel. It may be manifest as a loss of resolution caused by multiple features presenting in the image voxel [Douglas, 1998]. The solution to this artifact is to use a smaller voxel; however this may result in poorer signal-to-noise ratios in the image. Fig.2.12 shows a comparison of two axial slices through the same location of the head. One is taken with a 3 mm slice thickness and the other with a 10 mm slice thickness. The loss of resolution in the 10 mm image is obvious and the detail of some structures disappears.

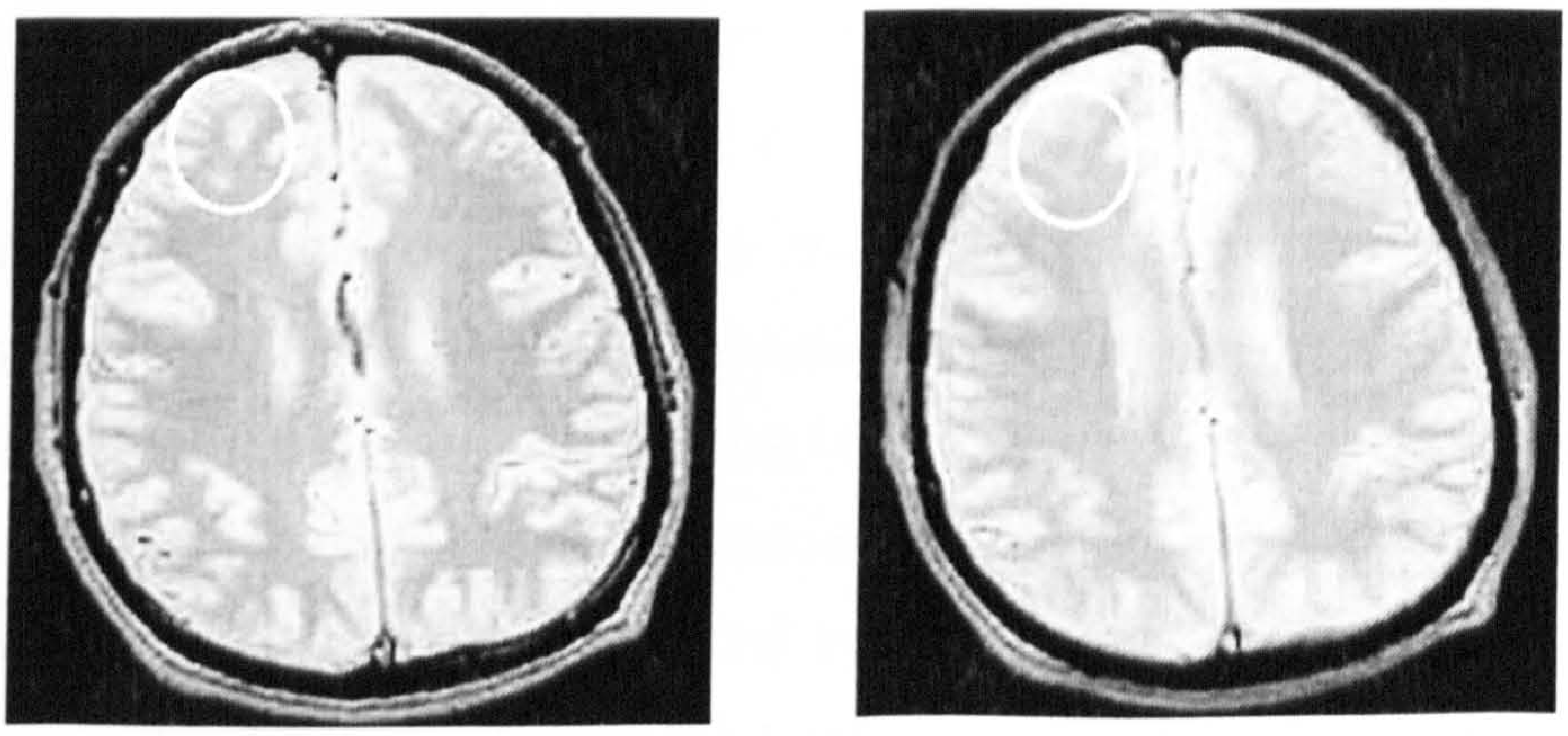


Figure 2.12 Resolution comparison (left): 3mm slice thickness (right): 10mm slice thickness [Hornak 1996].

Chapter 3

BRAIN TUMOUR DIAGNOSIS

3.1 INTRODUCTION

The brain is the most complex organ in the human body. It is a major part of the central nervous system (CNS), which controls our personality – memory, intelligence, speech, emotions; senses – vision, hearing, taste, smell and touch; basic functions – breathing, heart-beat, blood pressure, movement and balance. The brain is a soft, spongy mass of tissue and it is protected by the bones of the skull and three thin membranes called meninges. Watery fluid, called cerebrospinal fluid (CSF), cushions the brain. This fluid flows through spaces between the meninges and through spaces within the brain called ventricles. The simple anatomy of brain is shown in Fig.3.1.

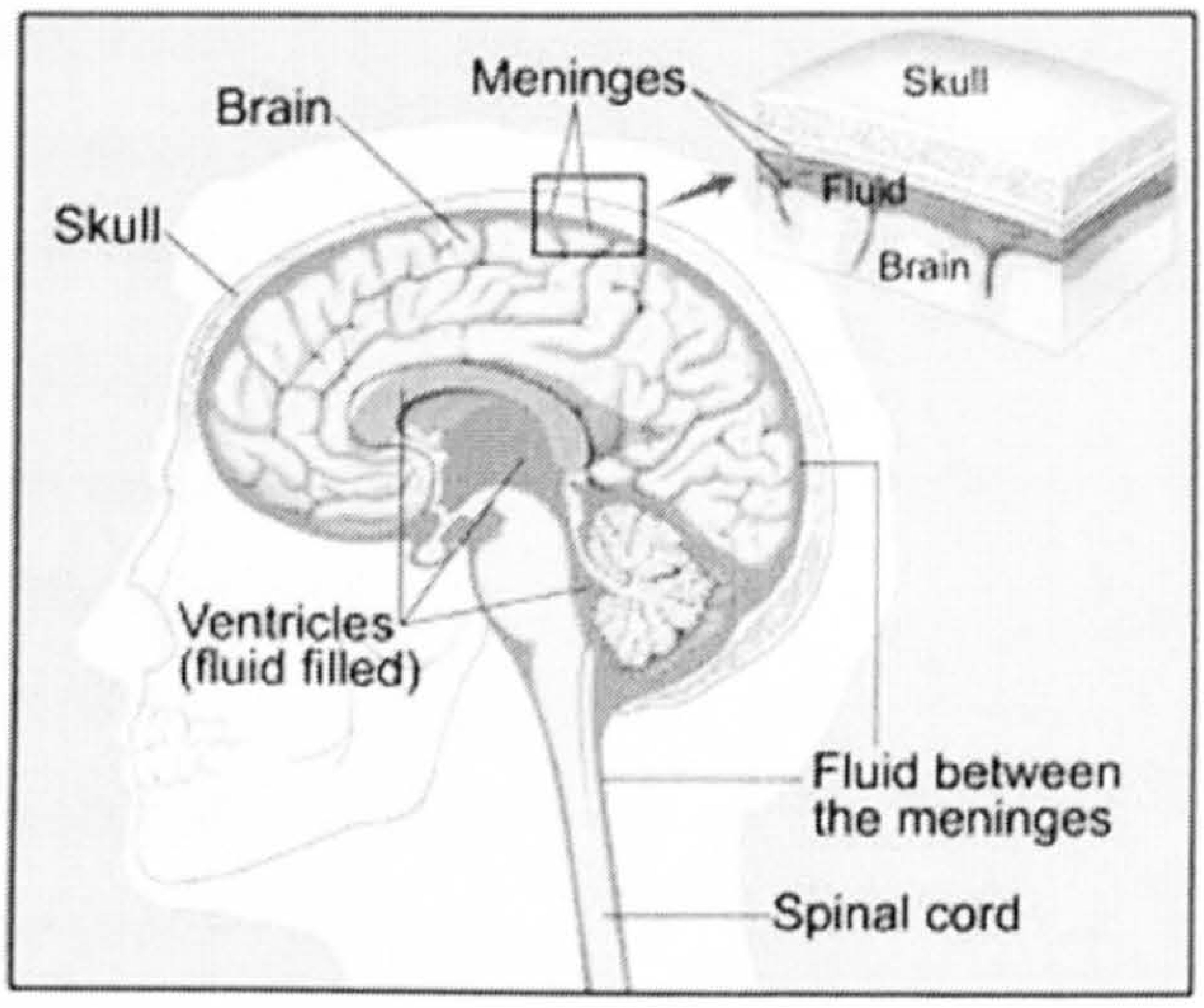


Figure 3.1 The brain [National Cancer Institute 2003]

A network of nerves carries messages back and forth between the brain and the rest of the body. Some nerves go directly from the brain to the eyes, ears, and other parts of the head. Other nerves run through the spinal cord to connect the brain with the other parts of the body. The brain is mainly made up of two types of cells, nerve cells (also called neurons) and glial cell. There are two types of nervous tissue. One type of tissue is made up of connected cell bodies known as gray matter. The other type of tissue containing mainly long, myelinated (sheathed) axons is known as white matter. Fig.3.2 shows the structure of white matter and gray matter. Within the brain and the spinal cord, glial cells are the building-block cells of the connective or supportive tissues. They surround neurons to hold them and help them carry out their

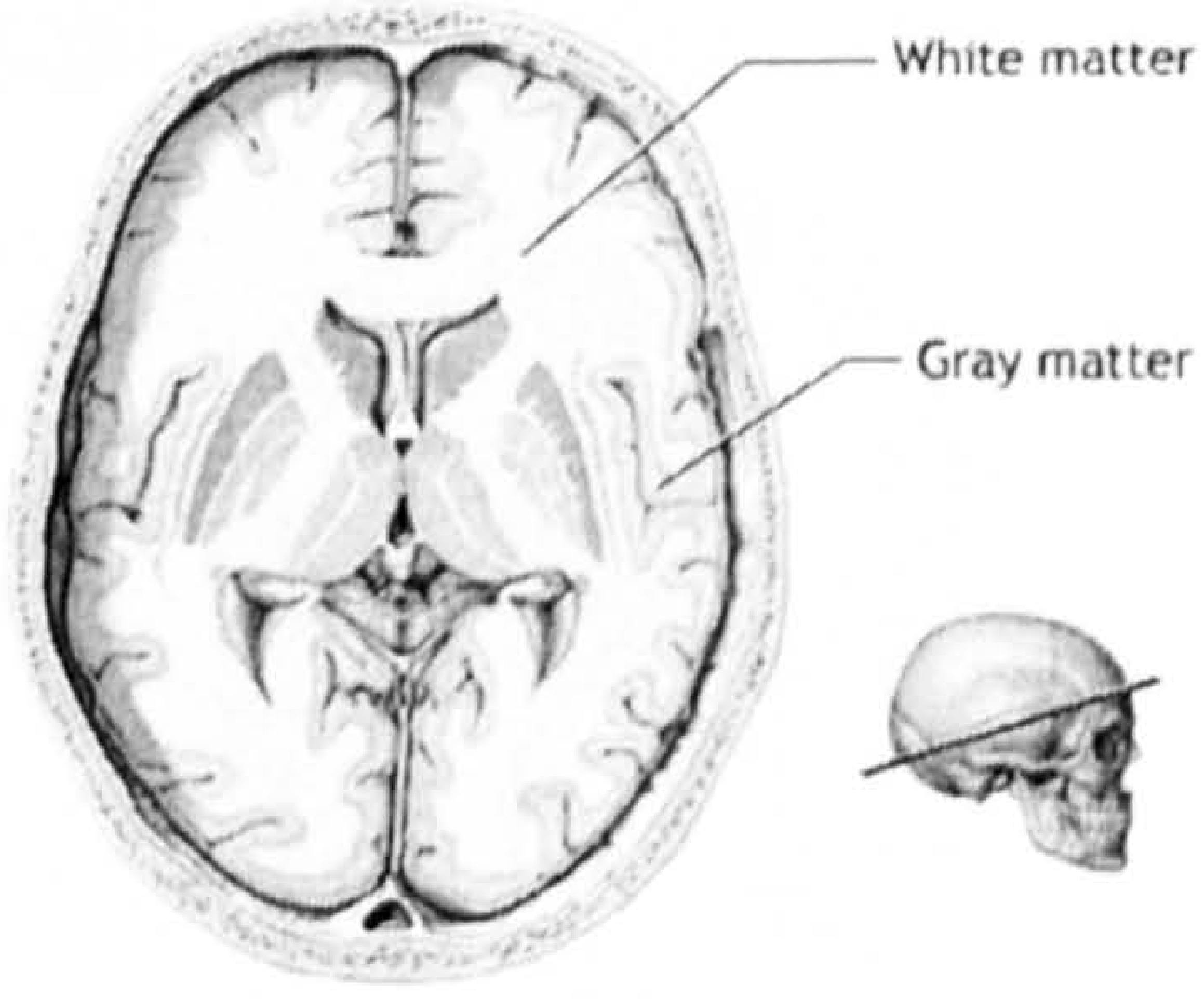


Figure 3.2 White matter and gray matter [Galit 2001].

functions.

A brain tumour is a mass of unnecessary, abnormal cells growing in the brain as shown in Fig.3.3. It is more dangerous and more difficult to treat than other tumours because of the brain's vital functions. Brain tumours are the second most common type of tumour in children. They account for almost 21% of tumours, and are a leading cause of death in children younger than 15 years of age. Most brain tumours develop by age 10. The incidence is 2.2-2.5 cases per 100,000 with the peak being between birth and 10 years of age. Brain tumours are slightly more common in boys [Pennstate Children's Hospital, 2003].

The particular severity of brain tumours are expressed in terms of the following aspects. Firstly, brain tumours occur in the brain which is enclosed in a bony canal. It allows little room for growth of the tumour without compressing and damaging the normal brain. Secondly, many brain tumours extensively invade normally functioning brain tissues, making complete surgical removal impossible. Thirdly, in their early stages, brain tumours are protected behind a blood-brain barrier; even when this barrier is disrupted in the bulk of the tumour, infiltrating tumour cells at the growing edge remain protected. Fourthly, disruption of the blood-brain barrier leads to oedema, which the brain tolerates poorly because of the limited intracranial space and the lack of lymphatics to rid itself of the products of oedema and other debris. Fifthly, the brain is itself rich in expressed genes and therefore is a fertile field for the growth of both primary tumours and metastases. Finally, the brain and brain tumours appear to be less susceptible to attack by the immune system than are tumours in other organs [Louis *et al.*, 2000].

Brain tumours produce a variety of symptoms ranging from headache to stroke.

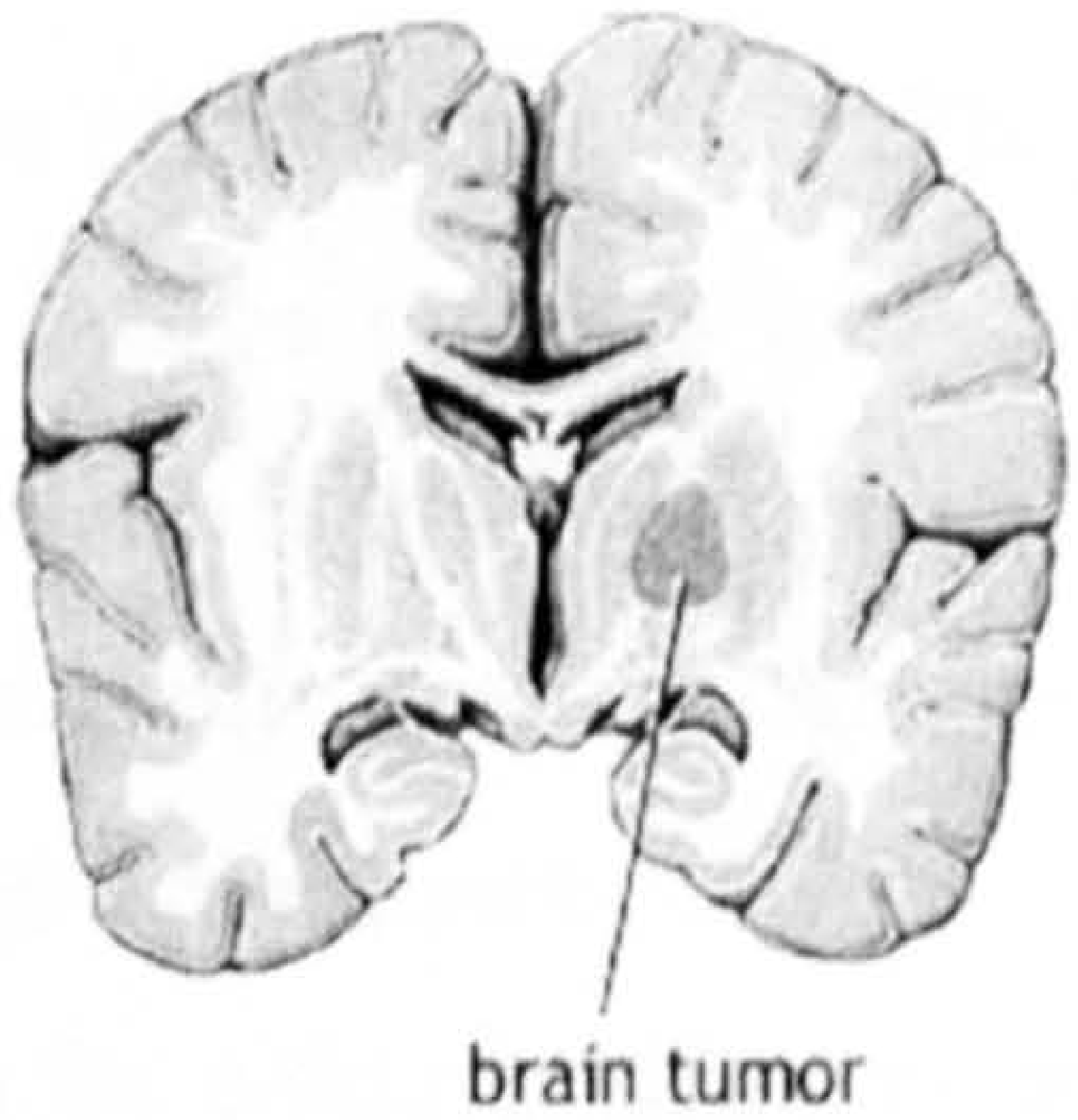


Figure 3.3 Brain tumour [ThirdAge 2003].

A persistent headache can often be the first symptom of a brain tumour. Although very rarely a headache is due to a brain tumour, severe, persistent or more frequent headaches should be investigated further. It occurs because of the raised intracranial pressure. Nausea and vomiting can accompany the headaches, and so can mental confusion. Other symptoms of a brain tumour include disturbed vision, impaired speech and hearing, drowsiness, reduced movement or loss of balance and personality changes and these symptoms depend on the part of the brain that is affected. Another sign that sometimes occurs is seizures, which can be as mild as a loss of consciousness for a few seconds or may involve severe shaking of the limbs. This is due to a build up of abnormal electrical activity in the brain when the nerve cells are irritated by the brain tumour [Slevin and Ryan, 1988].

3.2 BRAIN TUMOUR TYPES

Brain tumours can be classified as either benign (non-cancerous) or malignant (cancerous). Benign tumours do not invade tissues around them or spread to other parts of the body. The border or edge of a benign brain tumour can be clearly seen. Malignant tumours contain cancerous cells. They are likely to grow rapidly and crowd or invade the surrounding healthy brain tissue.

Brain tumours can also be divided into primary and secondary. Primary tumours are tumours that begin in the brain. Secondary tumours are tumours that start out elsewhere in the body and spread, or metastasize, to the brain. For example, secondary brain tumours could have begun as breast cancer or lung cancer [Musella Foundation, 2003].

All benign tumours are primary tumours. They represent half of all primary brain tumours. The cells of benign tumours look normal and grow slowly. Benign tumours are not particularly harmful in most parts of the human body, however they are dangerous in the brain since any abnormal growth in the brain can place pressure on sensitive tissues and impair brain functions.

Malignant tumours can be either primary or secondary. Secondary malignant brain tumours are about three times more common than primary malignant brain tumours. They occur when cancer cells spread to the brain from a primary cancer in another part of the body. Mostly, the primary cancer causing the secondary brain tumours originates in the lung, breast, kidney or skin. Primary malignant brain tumours that originate in the brain rarely spread to other parts of the body [American Accreditation HealthCare Commission, 2002].

Primary brain tumours are named according to the type of cells or the part of the brain in which they begin.

About half of all primary brain tumours are known collectively as gliomas, which come from glial cells. There are several types of gliomas such as astrocytomas from astrocytes, oligodendrogliomas from oligodendrocyte glial cells, ependymomas from ependymal cells etc. Gliomas can be categorized into different grades depending on the degree of their malignancy. They can be either low grade or high grade (Other systems: Kernohan grades these tumours on a scale of I to IV and the WHO grades on a scale of I to III). Low grade (I and II) is less malignant and high grade (III and IV) are more dangerous. Low grade gliomas are almost normal-shaped and grow slowly over a long period as solid masses. High grade gliomas grow rapidly and can invade surrounding brain tissues. Therefore, low grade gliomas are easier to treat and high grade gliomas require more intense therapy [The Brain Tumour foundation, 2003]. Other tumour types can also be graded as gliomas. In a tumour that contains a mixture of different cell grades, the tumour is graded using the highest grade cells even when they are very few.

Meningioma is another common type of brain tumour. It is usually a benign tumour originating from the meninges, or membranes, which cover the brain and the spinal cord. Typically, a meningioma is not recognized until it has become relatively large in size. In fact, meningiomas comprise 25% of all primary brain tumours but less than 2% of all childhood brain tumours. A meningioma may be classified by its constituent cell types (i.e. syncytial, fibroblastic, angioblastic, transitional) [American Accreditation HealthCare Commission, 2002; The Hyman-Newman Institute for Neurology and Neurosurgery, 2002].

Pituitary tumours comprise about 10% of primary brain tumours. They mostly develop in the front, or anterior portion of the pituitary gland (the adenohypophysis). The term, "pituitary adenoma", describes a tumour of the pituitary gland, the majority of which are benign and curable. Only about 10% of all identified pituitary tumours are found in children. There are several types of functioning pituitary adenomas, including those which produce prolactin (PRL), growth hormone (GH), corticotropin (ACTH), and mixed varieties.

Other types of primary brain tumours include medulloblastomas, neuronomas, craniopharingiomas, pineal tumours, germ cell tumours, and schwannomas. Fig.3.4 shows the distribution of brain tumour types.

3.3 TRADITIONAL DIAGNOSIS METHODS

As described above there are a great number of brain tumour types. Different brain tumours may have different treatments. For instance, a low grade glioma is usually treated by surgery and radiotherapy, and a high grade glioma requires surgery, radiotherapy, chemotherapy and possibly investigational treatments. Hence, brain tumour diagnosis is very important when choosing effective therapies.

A neurological exam is usually the first test given after checking the medical history of the patients when they complain of symptoms that suggest a brain tumour. Part of the exam includes checking the power and strength of the arms and legs, knee

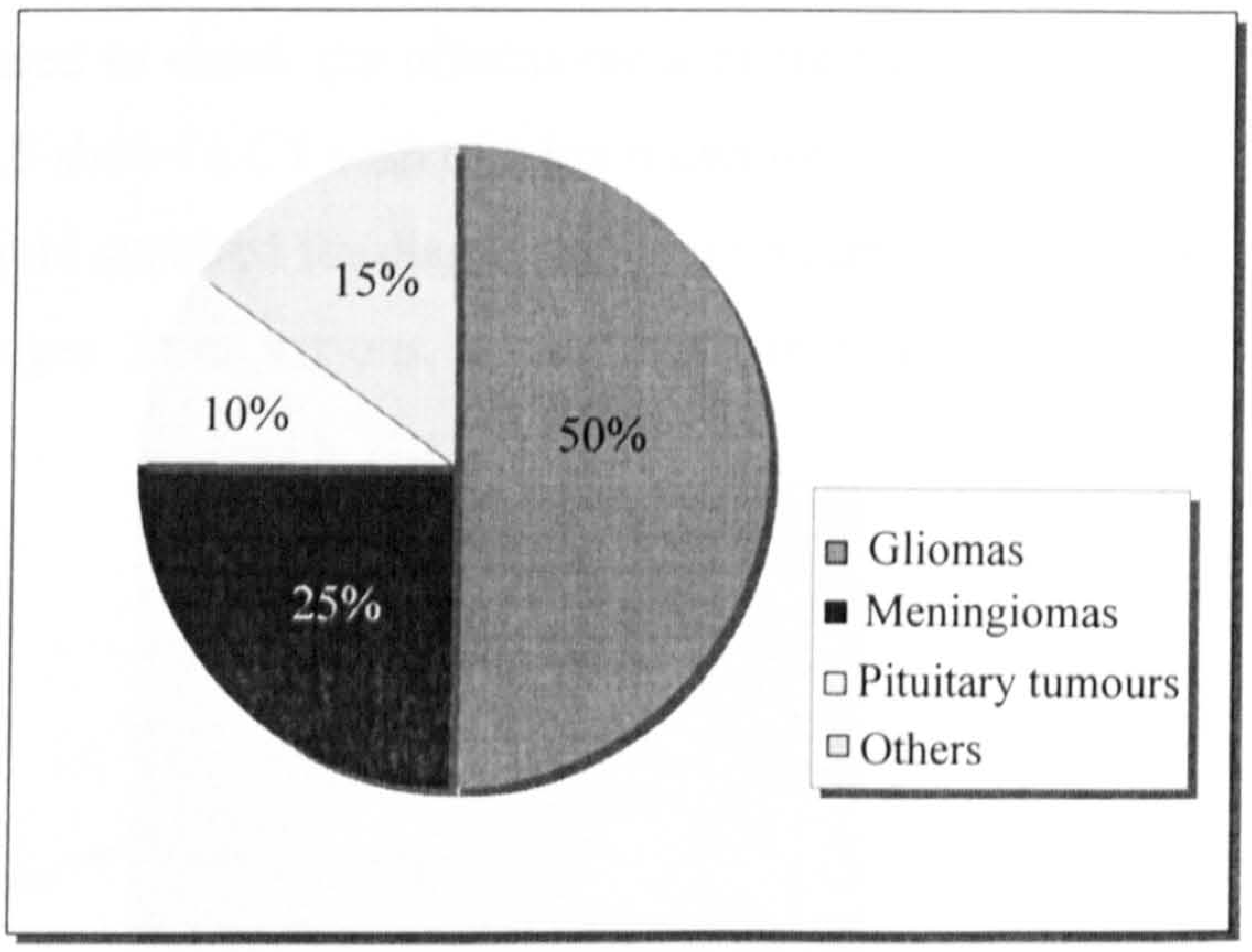


Figure 3.4 Distribution of brain tumour types.

jerks and other reflexes, feeling pin-pricks, distinguishing between heat and cold, and also performing some mental exercises. These can indicate whether there are signs of diminished mental function. In addition to these, an eye exam using an ophthalmoscope can be included. The ophthalmoscope shines light into the lining of the eye (the retina) and the optic nerve, which connects the eye to the brain. A brain tumour causing raised intracranial pressure may swell part of the optical nerve [Slevin and Ryan, 1988].

A further test involves a brain image scan of the patient. It may detect the presence of a brain tumour and its exact size and location. X-rays of the skull were once standard diagnostic tools but are now performed only when more advanced procedures are not available. Unusually, some tumours may be shown on an X-ray image due to them containing calcium (bones). Advanced imaging techniques have dramatically improved the diagnosis of brain tumours in recent years. The most common imaging techniques for brain tumour diagnosis are CT, MRI and PET.

CT uses a sophisticated X-ray machine and a computer to create a detailed image of the body's tissues and structures. It is not as accurate as an MR image and does not detect about half of low-grade gliomas. However, it is still useful in certain situations. Often, doctors will inject the patient with an iodine dye, called contrast material, to make it easier to see abnormal tissues. A CT scan helps locate the tumour and can sometimes help detect swelling, bleeding, and associated conditions. In addition, CT is used to check the effectiveness of treatments and watch for tumour recurrence. Fig.3.5 shows a CT scan of a brain tumour.

MRI is the gold standard for diagnosing a brain tumour. It does not use radiation and provides images from various angles that can enable doctors to construct a

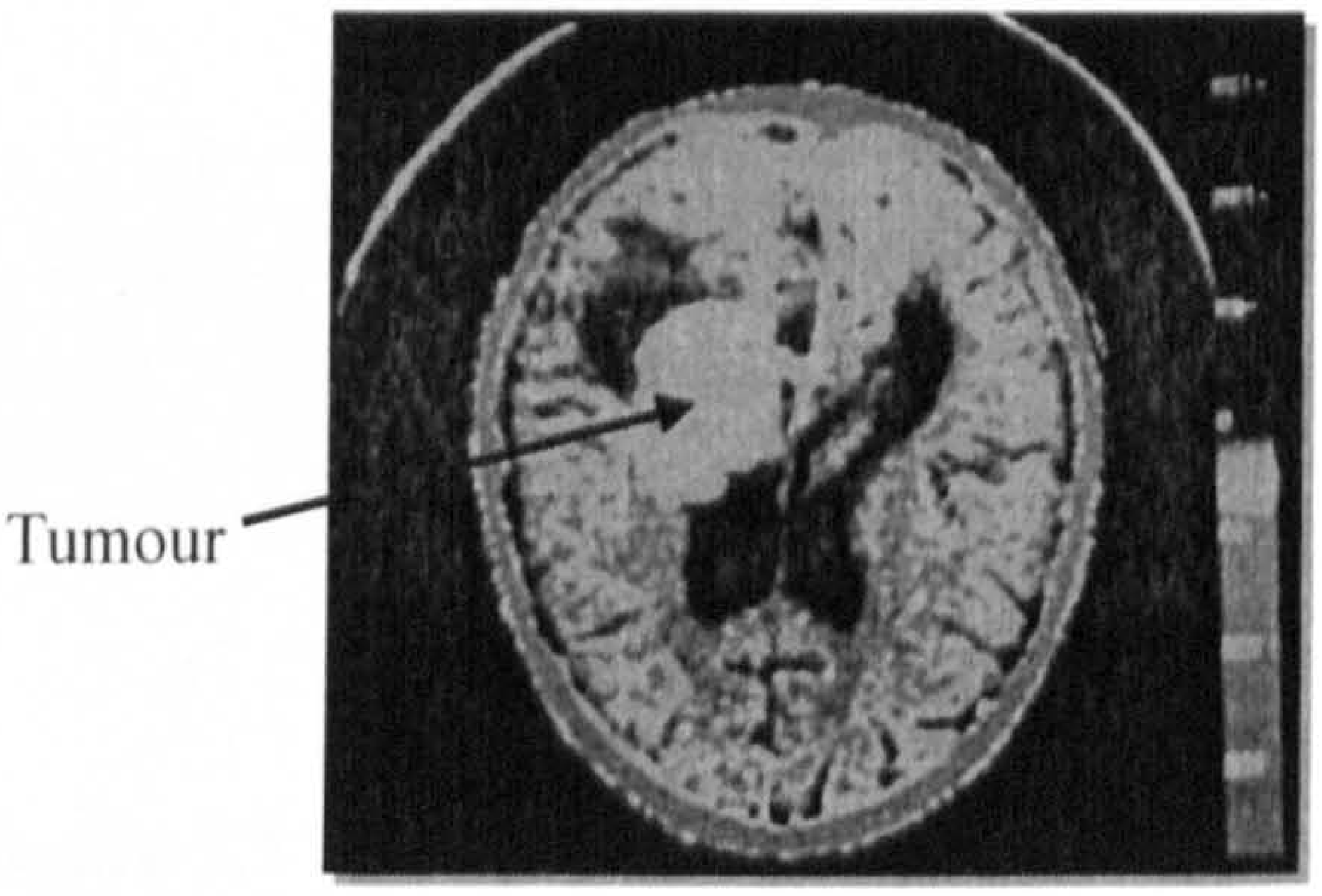


Figure 3.5 CT scan of brain tumour [Rich and Lasley 1998].

three-dimensional image of the tumour. It gives a clear view of tumours near bones, smaller tumours, brainstem tumours, and low-grade tumours. MRI is also useful during surgery to show tumour bulk, for accurately mapping the brain and for detecting response to therapy. Fig.3.6 is an MR image of a brain tumour. A variant called magnetic resonance spectroscopy (MRS) is capable of providing information on the activity of the brain using MRI. MRS is proving to be accurate for distinguishing dead (necrotic) tissue caused by previous radiation treatments from recurring tumour cells in the brain, which is a difficult diagnostic issue.

PET provides an image of the brain's activity rather than its structure by tracking substances that have been labeled with a radioactive tracer. PET is not routinely used for diagnosis, but it may supplement MRI to help determine tumour grade after a diagnosis. Fig.3.7 shows a brain tumour image of PET using fluorodeoxyglucose (FDG) as the tracking substance. As with MRS, it is also able to

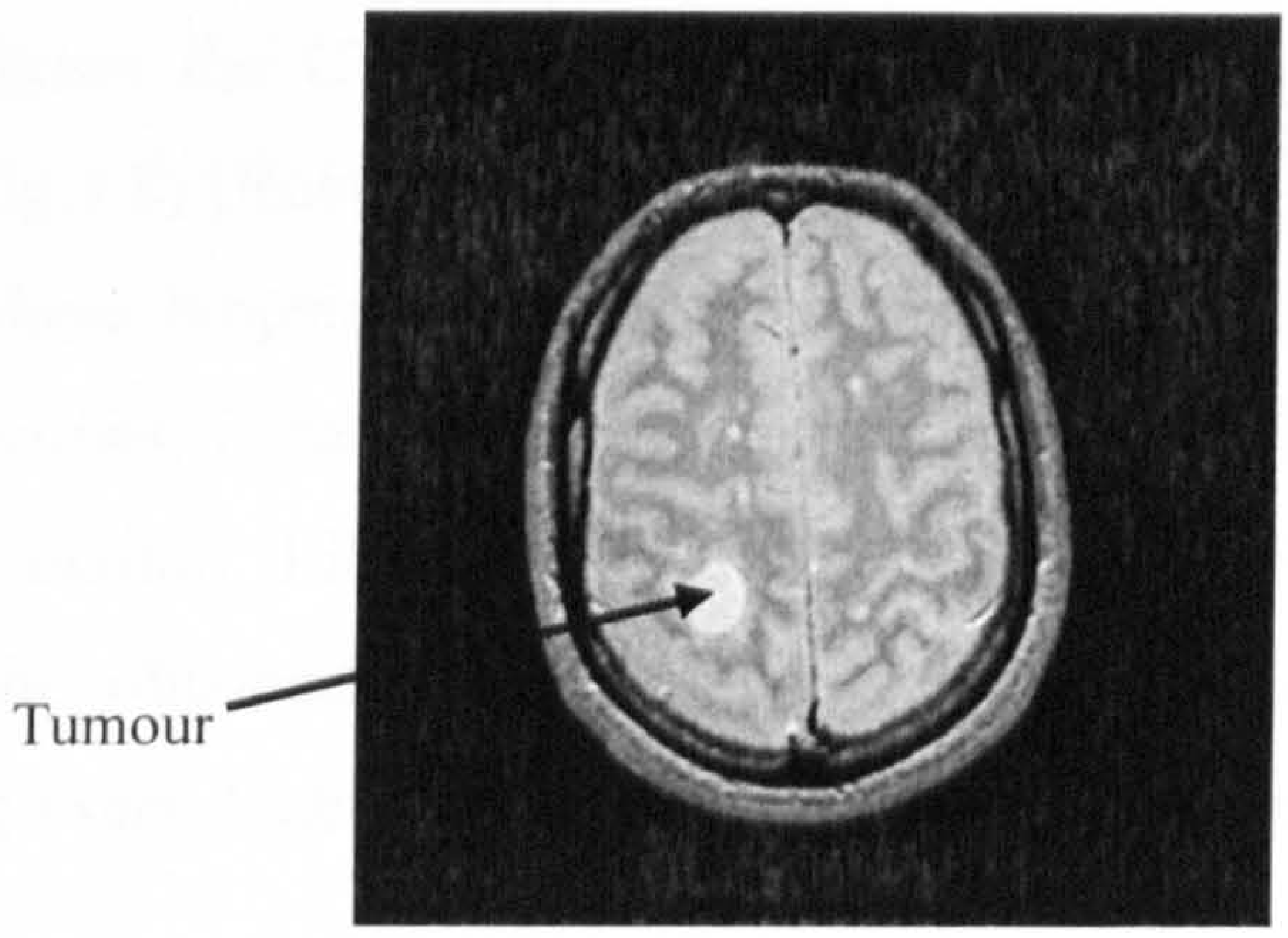


Figure 3.6 T_2-weighted MR image of brain tumour [Srdanovic 1998].

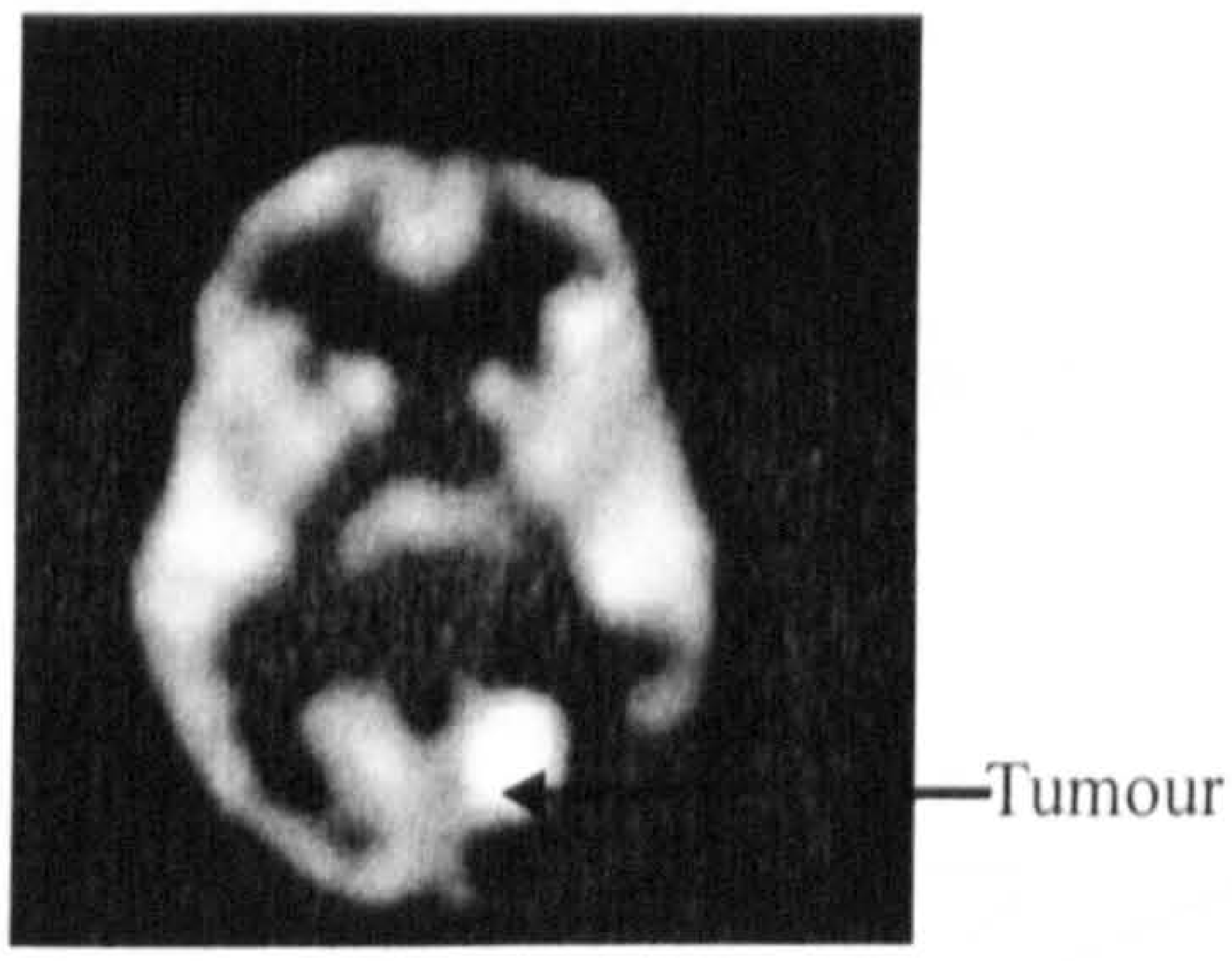


Figure 3.7 PET image of brain tumour [Sabbatini 1997].

distinguish between recurrent tumour cells from dead cells or scar tissue, although MRS is more widely available.

The final step for brain tumour diagnosis is a biopsy. A biopsy is a surgical procedure to take a sample tissue from the suspected tumour [The Brain Tumour Society, 2002]. An imaging scan such as CT, MRI and PET is taken prior to the biopsy to determine the position where the biopsy will be performed. Generally, biopsies can be performed "open" or "closed" through a craniotomy. An open biopsy usually implies that a small window of bone is temporarily removed from the skull to allow the surgeon to remove a small portion of the suspected tumour under direct visualization. A closed biopsy, also called stereotactic biopsy is a new kind of biopsy. It relies on computer guidance to locate the brain tumour and avoid serious complications. A tiny hole, no larger than an eighth of an inch, is drilled into the skull and a needle is inserted into the brain tissue guided by computer-assisted imaging techniques (CT or MRI). A special head frame is applied to direct the probe into the brain and allows the CT and/or MRI to be used along with a highly specialized computer (Fig.3.8) [Rokahr, 1996]. Since the early nineties, it has also been possible to perform these biopsies without the frame. Because the frame is attached to the skull with screws, it can cause the patient some additional suffering [AccessMed Health Information Library, 2002; York Neurosurgical Associate, 2003]. By examining the obtained tissue sample under a microscope, the pathologist can determine an exact diagnosis such as malignance and tumour types and discriminate

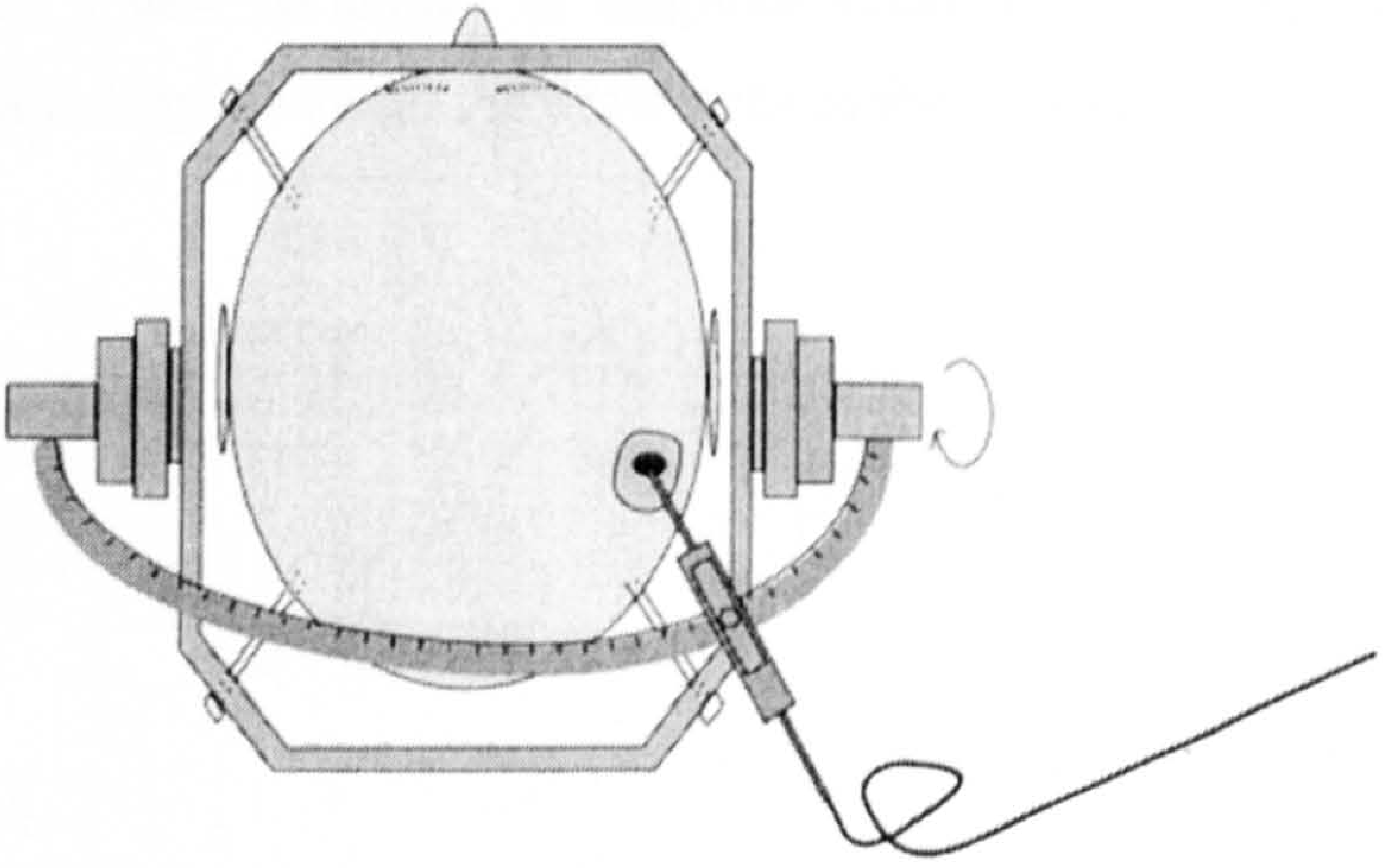


Figure 3.8 Stereotactic brain biopsy [Rokahr 1996].

a tumour from a brain abscess.

3.4 DIAGNOSIS PROBLEMS

According to description above, it is clear that the biopsy is an essential step to confirm the diagnosis of brain tumours. However, its drawbacks are also very explicit. First of all, although the advance of the biopsy means that it is less invasive and better tolerated by the patient, it is still an invasive procedure and can cause a great deal of pain. Secondly brain injury may occur due to the removal of brain tissue. Because of the vital function of the brain, removing any healthy tissue may affect normal functioning. Thirdly, an anesthesia process is necessary before the biopsy to relieve the pain. Usually, a general anesthesia is applied to the patient, which further increases the risks from the biopsy. Fourthly, the resulting scar, left on the brain has the potential to trigger seizures. Fifthly, the patient has to be monitored after the biopsy for several hours in case of unexpected complications and is usually required to spend a few days in the hospital. Sixthly, because the biopsy is a very precise cranial procedure, it must be performed by a specified neurosurgeon. An anesthetist and other staff are required as well. It is obvious a very expensive surgical procedure. In addition, not all brain tumours can have a biopsy, such as brain stem gliomas, because it may be too hazardous. Diagnosis is therefore complicated, risky and can result in complications.

Consequently, a new technique to diagnose brain tumours non-invasively is required, which can compensate for the drawbacks in the present diagnosis method.

Chapter 4

DATA MINING AND KNOWLEDGE DISCOVERY

4.1 INTRODUCTION

In this information age, advances in our capability to both generate and collect data can lead to a flood of data. The amount of data is growing exponentially in a wide variety of fields. It has been estimated that the amount of data in the world doubles every 20 months [Cios, Pedrycz and Swiniarski, 1998]. Databases today can range in size up to the terabytes - more than 1,000,000,000,000 bytes of data [Two Crows Corporation, 1999]. A great deal of useful knowledge is hidden within this mass of data, however such volumes of data clearly overwhelm the traditional manual methods of data analysis. A new generation of techniques and tools is therefore required to analyze this amount of data intelligently and automatically. Data mining and knowledge discovery techniques are rapidly emerging to satisfy this need.

The term "knowledge discovery" was coined in 1989. It refers to the overall process of discovering useful knowledge from data. It is defined as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [Cios *et al.*, 2000]. Here, it implies that knowledge discovery comprises several steps which involve such as data preparation, pattern searching, and knowledge evaluation. Extracting a pattern means fitting a model to data, finding structure from data or making decisions using data. Data mining is a major step in the process of knowledge discovery, consisting of particular data mining algorithms under some acceptable computational efficiency limitations, producing a particular enumeration of patterns [Fayyad, Shapiro and Smyth, 1996a].

Basically, knowledge discovery is the process of using data mining methods (algorithms) iteratively to extract what is deemed knowledge according to the specifications of measures and thresholds, using the database along with any required preprocessing, sub-sampling and transformation. The components of knowledge discovery are variable in different description. Here, the process is divided into five steps as follows:

1. Data cleaning and data preprocessing

Generally, data cleaning deals with removing noise or outliers if appropriate. Data preprocessing covers many fields such as handling missing or wrong data, developing an understanding of the application domain and the relevant prior knowledge, collecting the necessary information to model, accounting for time-sequence information.

2. Data selection

Each database may consist of many data sets and each data set may have a considerable number of variables or attributes. The objective of this step is to select a data set or to focus on a subset of variables or data samples from the data provided according to task requirements. It should greatly decrease the quantity of work so that task-relevant data can be obtained.

3. Data mining

Data mining is the essential step in the knowledge discovery process. It searches for patterns of interests according to the particular task, such as classification, regression and summarization. The most commonly used techniques of data mining are evolutionary computing, machine Learning (ML), neural networks, rough sets, fuzzy logic, and Bayesian method. Many of these have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data. Each technique has its particular features tailored to different tasks.

4. Pattern evaluation

Since patterns are discovered after data mining, their significance in solving the proposed tasks must then be evaluated. This includes understanding and interpreting the patterns, from which the novel, particular and interesting ones can be found. Sometimes if the patterns do not make sense, advice is sought from experts for evaluation.

5. Knowledge discovery

The most novel and interesting patterns are chosen from the evaluation. Useful knowledge or effective solutions are therefore provided in the final step of knowledge discovery.

The process of knowledge discovery can involve significant iteration due to no satisfactory knowledge perceived. Fig.4.1 shows the diagram of this whole process [Cios, Pedrycz and Swiniarski, 1998; Cios *et al.*, 2000; Fayyad *et al.*, 1996b; Shaw *et al.*, 2001; Ramirez, 2000; Feelders, Daniels and Holsheimer, 2000].

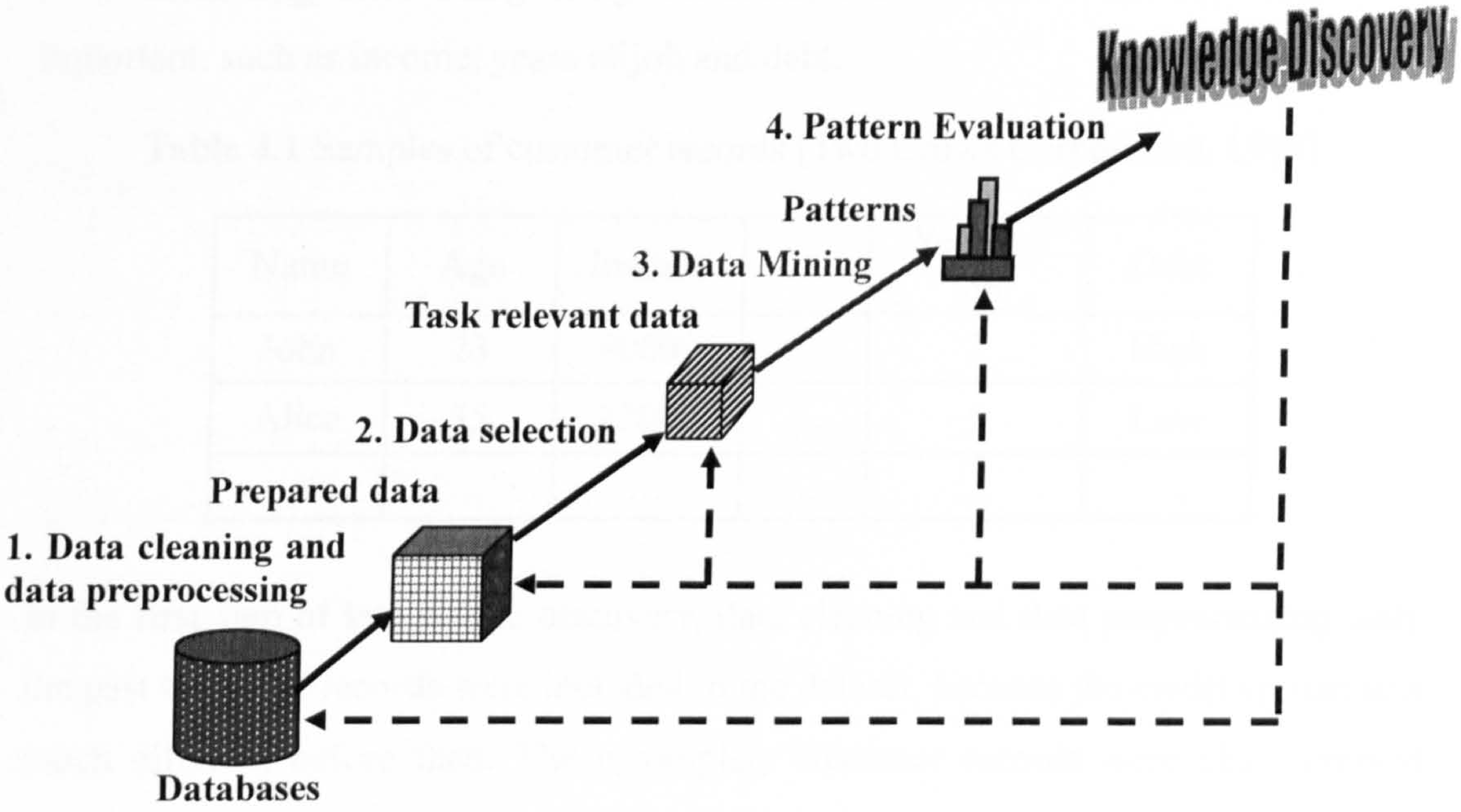


Figure 4.1 Knowledge discovery process.

4.2 SIMPLE ILLUSTRATIONS FOR DISCOVERED PATTERNS

For the sake of complete understanding, some simple examples in data mining and knowledge discovery are described in this section to demonstrate the power of these techniques.

4.2.1 Finance

Credit risk is a crucial criterion for credit card or finance companies in dealing with their customers. Prompt and exact evaluation of the credit risk leads to successful company operation. However, there is not always a pattern in deciding how to determine a credit risk to an individual. Although hundreds and thousands of customer records with many attributes are available, it is not easy to find relationships between them. Table 4.1 lists samples of customer records. Notwithstanding there being many attributes, only a few of them are considered important, such as income, years of job and debt.

Table 4.1 Samples of customer records [Two Crows Corporation, 1999].

Name	Age	Income	...	Years of job	Debt
John	23	40000	...	3	High
Alice	35	32000	...	5	Low
...	...	...	...	...	...

In the first step of knowledge discovery, data cleaning and data preprocessing, only the past ten years records were included in the dataset, because the credit system was much different before then. The incomplete customer records were also removed from the data. Then, the attributes in each customer records such as Name, Telephone number were excluded from the dataset in data selection. A particular data mining method, such as ID3 [Quinlan, 1986] or C4.5 [Quinlan, 1993], was applied to obtain the patterns. The patterns acquired were evaluated according to their accuracy of representing the dataset. This is the fourth step, pattern evaluation. The patterns with the highest accuracy were selected, known as knowledge discovery. Two best IF-THEN patterns (rules) were obtained as follows:

Pattern I

IF Income>£40,000 AND Age<30 AND Debt=High
THEN Credit risk= Bad Risk

Pattern II

IF Income<£40,000 AND Years of Job>5 AND 28<Age<42
THEN Credit risk= Good Risk

From the patterns, it is seen that age is a determinant of risk which may have been ignored by the analysts [Two Crows Corporation, 1999]. This situation may happen in many different circumstances in that some unconsidered attributes are disclosed as being very important. This is the significance of data mining and knowledge discovery techniques.

4.2.2 Medicine

Example 1

A historical set of 9,714 medical records describes pregnant women over time. The task was to discover rules that predict which future patients will be at high risk of requiring an emergency cesarean-section delivery [Mitchell, 1999]. Table 4.2 gives an example of the patient records.

Table 4.2 N times records of Patient103.

Patient 103	1st	2nd	...	Nth
Age	23	23	...	23
FirstPregnancy	No	No		No
Anemia	No	No		No
Diabetes	No	No		No
PreviousPrematureBirth	No	No		No
Ultrasound	?	Abnormal		?
Elective c-section	?	?		No
Emergency c-section	?	?		Yes

The records with missing attributes were removed in data cleaning and data preprocessing step. Some presumably unimportant attributes were excluded in data selection. The data mining method used in this example was CN2 [Clark and Boswell, 1989]. Pattern evaluation evaluated the ability that a pattern could predict whether an emergency c-section was required. The pattern would help to predict the probability that pregnant women will need an emergency c-section. This could reduce the risk of giving birth.

Pattern learned:
IF No previous vaginal delivery AND Abnormal second trimester Ultrasound AND Malpresentation at admission
THEN Probability of Emergency c-section is 0.6

7% of all pregnant women in the data set received emergency c-sections, however the pattern identified a subclass at 60% risk for requiring c-sections.

Example 2

This example is for early diagnosis of rheumatic disease. To simplify the description, only the pattern discovered is given which was:

IF Sex = male AND Age>46 AND Number of painful joints>3 AND Skin manifestations = psoriasis
THEN Diagnosis = Crystal induced synovitis

It assigns the diagnosis of crystal induced synovitis to male patients older than 46 years that have more than three painful joints and psoriasis as a skin manifestation [Lavrac, 1999].

All examples given above are concerned with one type of pattern: rule, which is the simplest and the most understandable. Data mining methods will be introduced later.

4.3 DATA MINING ALGORITHMS

As stated above, data mining is extremely important in the knowledge discovery process. Many kinds of data mining algorithms exist in the literature which can be sorted into different categories by their function or by their method [Cios, Pedrycz and Swiniarski, 1998; Fayyad, Shapiro and Smyth, 1996].

4.3.1 By Function

Classification is learning a function that classifies a data item into one of several predefined classes [Weiss and Kulikowski, 1991]. The term originated from pattern recognition, for which a large number of classifiers were developed. Examples of classification using data mining algorithms range from diagnosing diseases [Bojarczuk, Lopes and Freitas, 2000] to classifying trends in financial markets [Apte and Hong, 1996].

Regression is mapping a data item to a real-valued prediction variable. The underlying idea is to construct a linear function explaining the data. The applications in regression, which include estimating the probability that a patient will survive, are estimated by the results of a set of diagnostic tests, or predicting consumer demand for a new product as a function of advertising expenditure.

Summarization involves methods for finding a compact description for a subset of data. It is an approach that characterizes data using a small number of attributes. It is often applied to interactive exploratory data analysis and automated report generation.

Dependency consists of finding a model that describes significant dependencies between variables. In medicine, discovering related symptoms or must-combined physical tests for a disease is an application of this domain. Investigating connected retail goods, such as nails and hammers, is another application of dependency analysis.

Deviation detection focuses on discovering the most significant changes in the data from previously measured or normative values. Its prominent application is predicting the amelioration and deterioration of diseases, and the quality of products.

4.3.2 By Method

Evolutionary computing can be seen as an optimization method driven by a biological principle of the survival of the fittest. It exploits an entire population of potential solutions and evolves them according to some genetically driven principles.

There are three main algorithms used as the basis for evolutionary computing, genetic algorithm (GA), genetic evolution strategy (ES), and evolutionary programming (EP). GAs use fixed-length character strings to represent the genetic information of a population of individuals which undergo genetic operations (for example, crossover, reproduction, mutation) in order to find interesting patterns. ES and EP share many similar features of GA. However, ES employs real-coded parameters and relies on mutation as the genetic operators with a population size one. EP dispenses with both genomic representations and with crossover as the genetic operator. Genetic programming (GP) is an extension of GA, which is highly valued by scientists nowadays. It continues the trend of dealing with the problem of representation in GA by increasing the complexity of the individuals in the population. Further details of GP are described below [Miettinen, 1999; Cios 1998; Hiker and Beasley, 2000].

Machine Learning (ML) models the environment or generates a new data structure that is different from the old one. The ability of a computer program improves its own performance and aims at revealing the relationships within the dataset (observations or experimentation), exploiting either deterministic or non-deterministic methods [Cios, 1998]. The result of ML is provided in the form of decision trees or production rules. ML methods can be classified into two groups: 1. inductive learning of symbolic rules, such as induction of rules (CN2, C4.5), decision trees (ID3, Assistant-R) and logic programming (FOIL, Progol and Claudien); 2. statistical or pattern-recognition methods, such as *k*-nearest neighbours or instance-based learning (IBL) [Cios, 1998; Kukar *et al.*, 1999; Bergadano, Giordana and Saitta,. 1991].

Neural networks is an information processing method which models the biological nervous system to process numeric data and building nonlinear relationships between input and output. It was inspired by the mimicing of the structure and function of the biological nervous system, such as the brain. The model used in neural networks is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. It is able to solve large and complex problems in which there may be hundreds of predictor variables that have many interactions. It is very effective in dealing with image data and solving many pattern recognition problems [Cios, 1998; Pudi, 2003; Stergiou and Siganos, 1996].

Rough sets algorithms provide rigorous mathematic techniques for discovering regularities in data and are particularly useful for dealing with imprecise and inconsistent information. The principle of rough sets is lowering the precision in data representation and making it possible to uncover patterns which may otherwise be obscured by too many details. It is a discrete technique, requiring discrete types of attributes and providing granular computations of approximations, classification. The attribute-oriented rough sets technique reduces the computational complexity of learning processes and eliminates the unimportant or irrelevant attributes. This method has been shown to be very effective in revealing relationships within imprecise data, discovering dependencies among objects and attributes, evaluating the classificatory importance of attributes, removing data redundancies and generating decision rules [Cios, 1998; Jagielska, Matthews and Whitfort, 1999].

Fuzzy logic is a mathematical methodology (and a philosophical ideology) that is similar in construct to Boolean algebra and similar in appearance to probability, but more general than both in fundamental ideas [Pulo, 1999]. Fuzzy logic extends Boolean logic to handle the expression of vague concepts and, as a result, solve the problems with imprecise and incomplete data. To express imprecision quantitatively, it introduces a membership function which indicates the degree of truth. This membership function ranges from zero to one inclusively, where zero implies totally untrue and one indicates perfectly true. It concentrates on representation of data at a nonnumeric level [Krantz, 1999].

Bayesian method is the technique learning the Bayesian networks from the data which represents the probabilistic relationships among a set of attributes. It provides statistical methods for handling the required probability descriptors of the problem and offers an efficient and principled approach for avoiding the over fitting of data. It can handle incomplete data and learn about casual relationships [Cios, 1998; Heckerman, 1997].

4.4 GENETIC PROGRAMMING

4.4.1 Introduction

Genetic Programming (GP) is a relatively recent technology and is an important branch in data mining techniques. It was proposed by Koza in 1987 and has grown exponentially since then [Koza, 1992; Koza, 1994; Spector *et al.*, 1999].

GP is based on the Darwinian principle of reproduction and survival of the fittest and analogy of naturally occurring genetic operations such as reproduction, crossover and mutation. It starts with an initial population of randomly generated computer programs (individuals) composed of functional sets and terminal sets appropriate to the problem domain. The individuals are usually expressed in tree shapes. The functional sets may be, for example, standard arithmetic operations, standard programming operations, standard mathematical functions, logical functions, or domain-specific functions. Depending on the particular problem, the computer program may be Boolean-valued, integer-valued, real-valued, complex-valued, vector-valued, symbolic-valued, or multiple-valued [Kenneth and Kinnear, 1994]. Each individual (computer program) in the population is then evaluated with respect to its capability performing in the particular problems. This evaluation is called the fitness function. The nature of the fitness function varies with the problem. Thereafter, the individuals undergo the genetic operations depending on their fitness. After applying different genetic operators on the individuals according to given probabilities, a new generation of the population is created. The fitness evaluation, genetic operation and creation of the new generation are executed iteratively until the maximum number of generations is reached. The result of GP is the fittest individuals (with highest or lowest fitness) produced along all generations [Koza, 1992; Bojarczuk, Lopes and Freitas, 1999; Langdon and Poli, 1997].

GP has been applied successfully to a large number of difficult problems like automatic design, pattern recognition, robotic control, marketing, image processing, and medical diagnosis. Koza *et al.* (1997) showed the ability of GP to automatically design electrical circuits. Handley (1993) used GP to predict the shape of proteins using the composition of the proteins. Andre (1994) applied GP in optical character recognition problems. Lee *et al.* evolved mobile robot controllers using GP and a simulator and demonstrated the controller running on the physical robot [Lee, Hallam and Lund, 1997]. GP was also employed by Andrews and Prager (1994) to create strategies which for trading in simulated commodity and future markets. Poli (1996) presented an image analysis method using GP. It was based on the idea that image enhancement, feature detection and image segmentation can be re-framed as filtering problems. Bojarczuk *et al.* diagnosed twelve types of chest pains using genetic programming which had high accuracy levels [Bojarczuk, Lopes and Freitas, 2000]. Other applications of GP include using GP to find programs to do location independent pattern recognition [Breunig, 1995] and using GP to discover the cellular automata rule [Andre, Bennett III and Koza, 1996].

4.4.2 Genetic Operators

Several types of genetic operators have emerged during the evolution. Prior to introducing genetic operators in details, the selection methods according to the fitness of individuals are described. They determine which individuals can be selected to do the particular genetic operation.

The most popular selection method is fitness-proportionate selection, first introduced by Holland (1975). If $f(i)$ is the fitness of an individual i in the population, the probability that the individual i will be selected to perform a particular genetic operation is as following (assuming the higher the fitness, the better):

$$Probability = \frac{f(i)}{\sum_{i=1}^{M} f(i)} \tag{4.1}$$

Where, M is the number of individuals in the population.

Another selection method is rank selection. In rank selection, all individuals are sorted by ascending fitness:

$$f(i) \leq f(j), \text{if } i < j$$

Each individual is assigned a rank $r(i)$, relying on its order in the population, where $r(1)=1$, $r(i)=i$ and $r(M)=M$. So the probability for selection is determined by the ranks:

$$Probability = \frac{r(i)}{\sum_{i=1}^{M} r(i)} \tag{4.2}$$

This method can reduce the potential domination of individuals with high fitness and also exaggerate the difference between individuals with close fitness. However, this may lead to slower convergence [Obitko, 1998; Koza, 1992].

Tournament selection can be viewed as an extension of rank selection. It randomly picks N ($N \geq 2$) individuals from the population. A rank selection is then carried out on the subgroup with N individuals to select an individual [Smith, 2002; Bennett, 2000]. The probability of selection is as follows:

$$Probability = \frac{N}{M} \frac{r(i)}{\sum_{i=1}^{N} r(i)} \tag{4.3}$$

It may sometimes be described differently as selecting the individuals with the best fitness from the subgroup instead of using rank selection [Koza, 1992; Burgess, 1999].

4.4.2.1 Reproduction

Reproduction is the basic genetic operator of Darwinian natural selection and survival of the fittest. It is an asexual operator. A single individual is selected from the population according to the selection method and copied into the new generation. The individuals with the best fitness in each generation are usually selected to do reproduction. This guarantees that the best fitness from a subsequent generation will never be worse than the best fitness from the prior generation.

4.4.2.2 Crossover

Crossover is a sexual operator which starts with two individuals and produces two offspring (new individuals in next generation). The first individual (the first parent) is selected from the population by the pre-decided selection method. The second parent is generally chosen by means of the same selection method. The operation

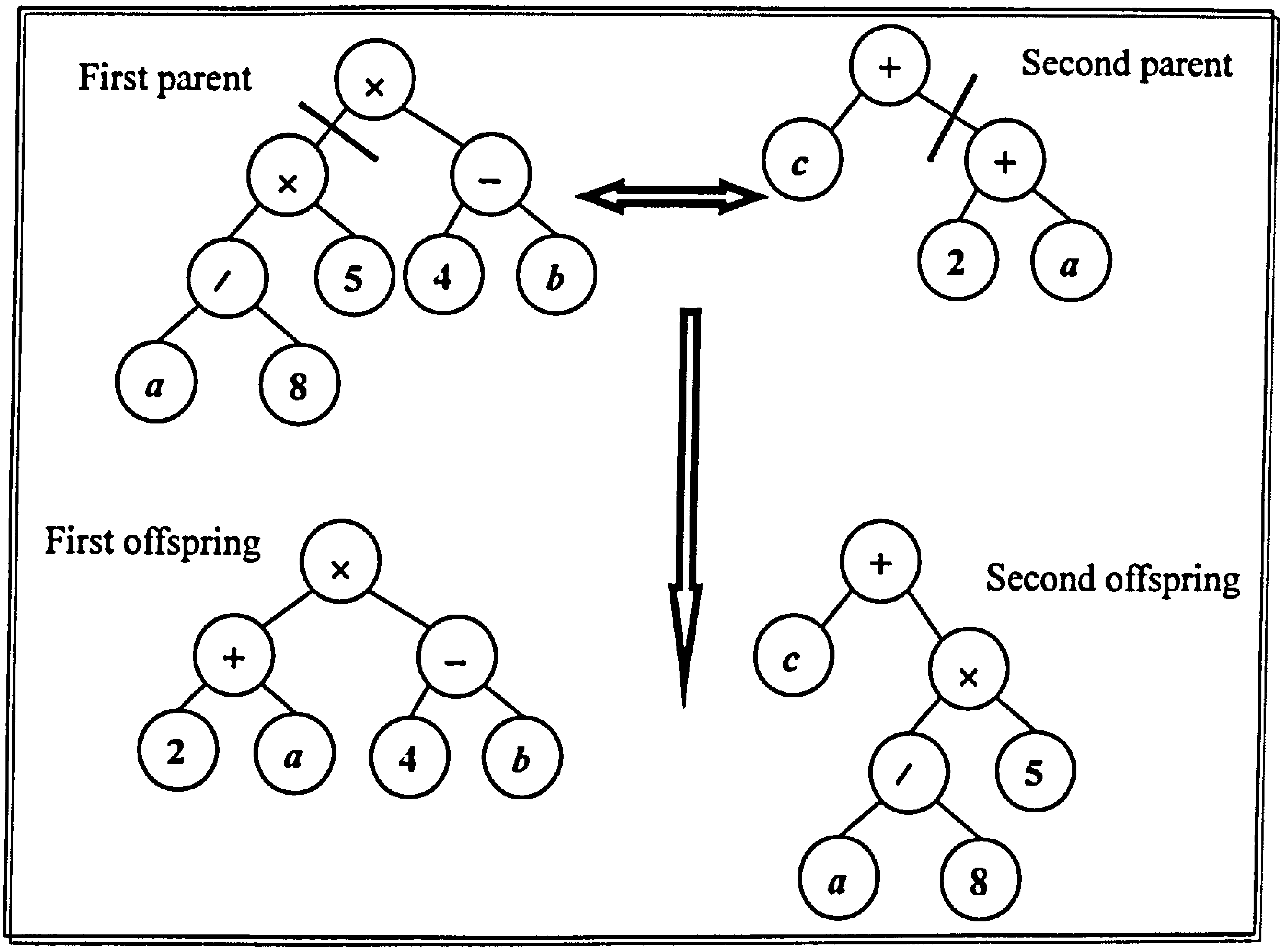


Figure 4.2 Crossover

then chooses a random point in each parent as the crossover point and swaps the branches under the crossover points. An example of crossover is shown in Fig.4.2. The first parent is an expression: $((a/8)\times 5)\times(4-b)$; the second parent is: $c+(2+a)$. Both of them are described tree-shaped as in Fig.4.2 (top). The crossover point of the first parent is chosen to be the first '×' and the crossover point of the second parent is the second '+'. The branches below the crossover points (inclusive the crossover points) are separated from the parents and swapped. Two offspring (Fig.4.2 bottom) are finally produced and stored for the next generation.

4.4.2.3 Mutation

A mutation operator introduces random changes into the individuals in the population. It is an asexual operator with only one parent involved and only one offspring is produced for the next generation. The parent is selected with the particular selection method. The mutation point of this parent is chosen randomly. Thereafter, the branch below the mutation point (including the mutation point) is

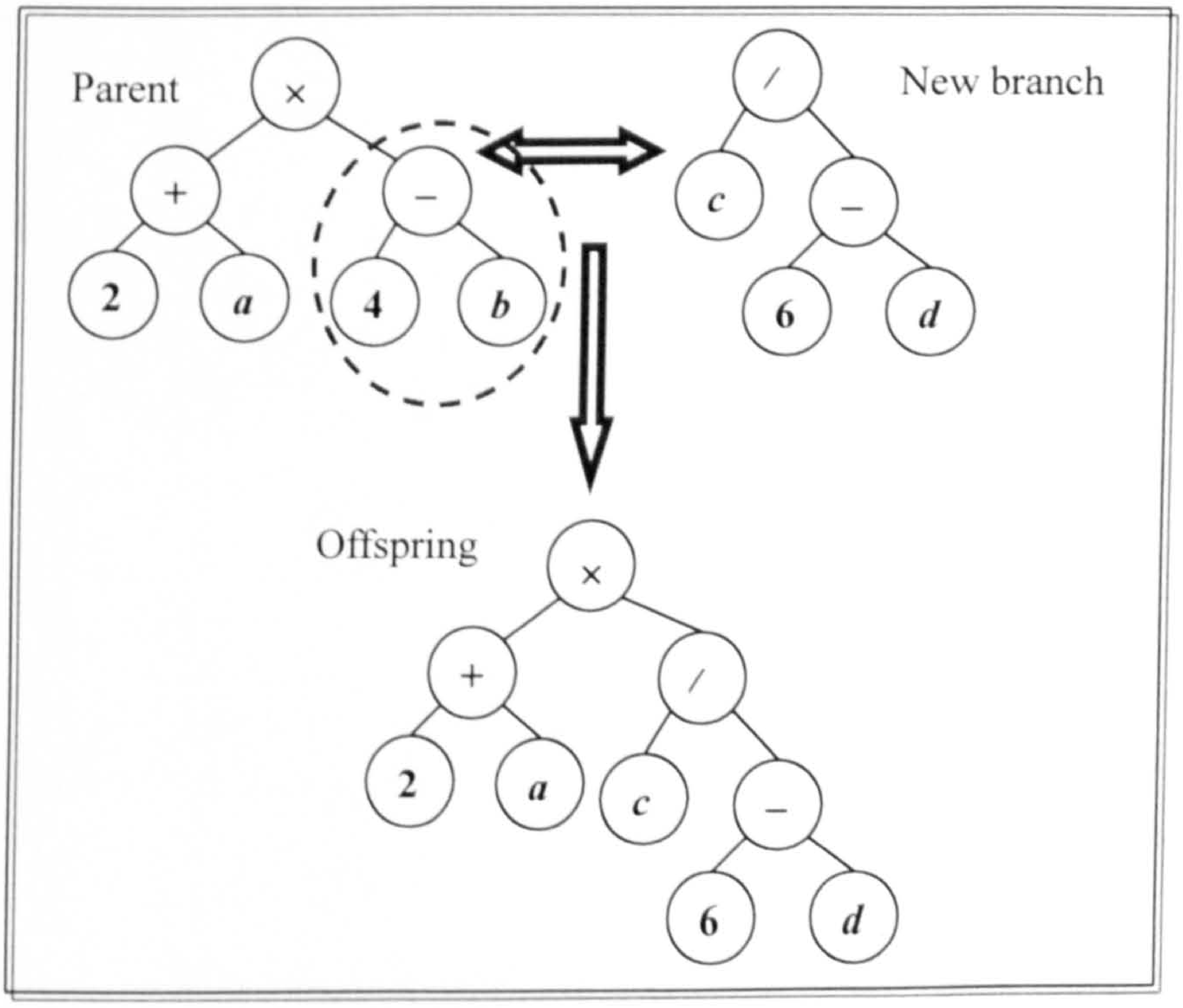


Figure 4.3 Mutation

discarded from the parent. A new branch is randomly generated and inserted at the position of the discarded branch. Fig.4.3 shows an example of mutation. The selected *parent is:* $(2+a)\times(4-b)$ *and the mutation point is* '-'. The old branch, $(4-b)$ is *replaced by a new generated branch,* $c/(6-d)$ *in this example. Its offspring is then* produced for the population of the next generation.

4.4.3 Individual Structure and Initial Population

The structure of individuals is a major point in GP which determines the effect of the GP algorithm. Each individual is made up of functional sets and terminal sets according to the grammar of the target problem. Different problems have specific grammar, which represent the way of explaining and solving the problems.

Functional sets are a set of elementary operators, appropriate to the problem, which are available as inner nodes in the tree-shaped individuals of GP. A function can have no input but must have an output or outputs. Typical functional sets may include [Koza, 1992; Beyer *et al.*, 2002]:

- Arithmetic operations: {+, -, ×, /}
- Mathematic operations: {sin, cos, exp, log}
- Boolean operations: {And, Or, Not}
- Logical operations: {If-Then-Else, Do-Until}
- Comparative operations: {<, >, =, ≤, ≥, ≠}

Terminal sets are a set of variables or constants which are like leaves in the tree-shaped individuals of GP. The format of the terminal sets is depended on the various tasks. Terminals must have an input but no output.

A tree-shaped individual is shown in Fig.4.4, where '–' and '/' are functions; '*c*',

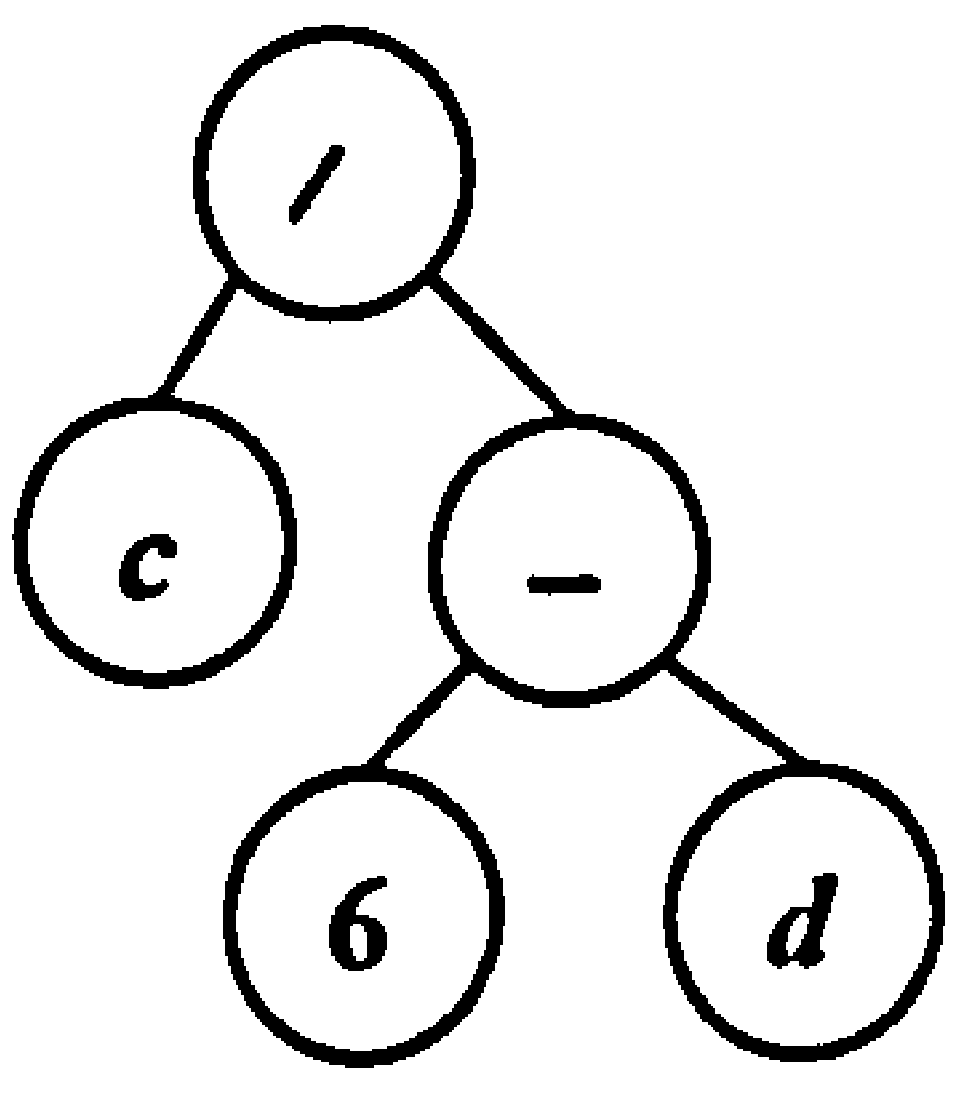


Figure 4.4 Structure of individuals.

'6' and '*d*' are terminals.

Creating an initial population is the first operation in GP. In this stage, functional sets and terminal sets are already decided but are required to be connected. Three popular techniques of creating the initial population are presented here [Koza, 1992; Walker, 2001]. One technique is called full method; this method involves creating tree-shaped individuals with a certain depth. It does not specify the number of nodes in an individual but requires its final depth to be equal to a certain value *d*. The depth means the number of layers in the tree-shaped individual. The individual in Fig.4.4 has a depth of three but it is not a full tree. If a node in each individual has a depth less than *d*, the element of the node is randomly selected from the functional sets. If the node has a depth equal to *d*, the element of the node then is randomly selected from the terminal sets. In other words, only the nodes in the deepest layer of the individual are selected from the terminal sets and other nodes are chosen from the functional sets.

The grow method is another technique of creating an initial population. It generates tree-shaped individuals with variable shapes and depths up to a specified maximum. Every node of the individual can be chosen from either the functional sets or the terminal sets. The selection of nodes continues until a terminal is chosen or the maximum depth is reached. This method provides a range of structures throughout the population and may even produce individuals containing only one (terminal) node. Fig 4.3 (top left) gives an individual created by the full method with a depth of three; the individual in Fig 4.4 is created by the grow method with a depth of three too, however the maximum depth can be greater than three. Noticeably, the grow method can still create a full tree, but not vice versa.

The ramped half-and-half method is a mixture of both 'full' method and 'grow' method. It therefore creates individuals having a wide variety of shapes and depths which have more possibility of including appropriate solutions for the problems. It involves creating an equal number of individuals using a specified depth ranging from 2 to the maximum depth. For each value of depth, 50% of individuals are created by full method and the other 50% are produced by grow method.

4.4.4 Process of Genetic Programming

Fig.4.5 is a flowchart of GP. The process of GP breeds computer programs to solve problems by executing the following five steps:

1. Generate an initial population by means of a creating method (full, grow, ramped half-and-half).
2. Compute and evaluate the fitness of each individual in the population.
3. Create a new population for the next generation using genetic operators (reproduction, crossover, and mutation) which are applied to the individuals chosen by a specified selection method (fitness proportion, rank selection, and tournament selection) and increase the number of generation.
4. Repeat steps 2 and 3 until the termination criteria are satisfied (The termination

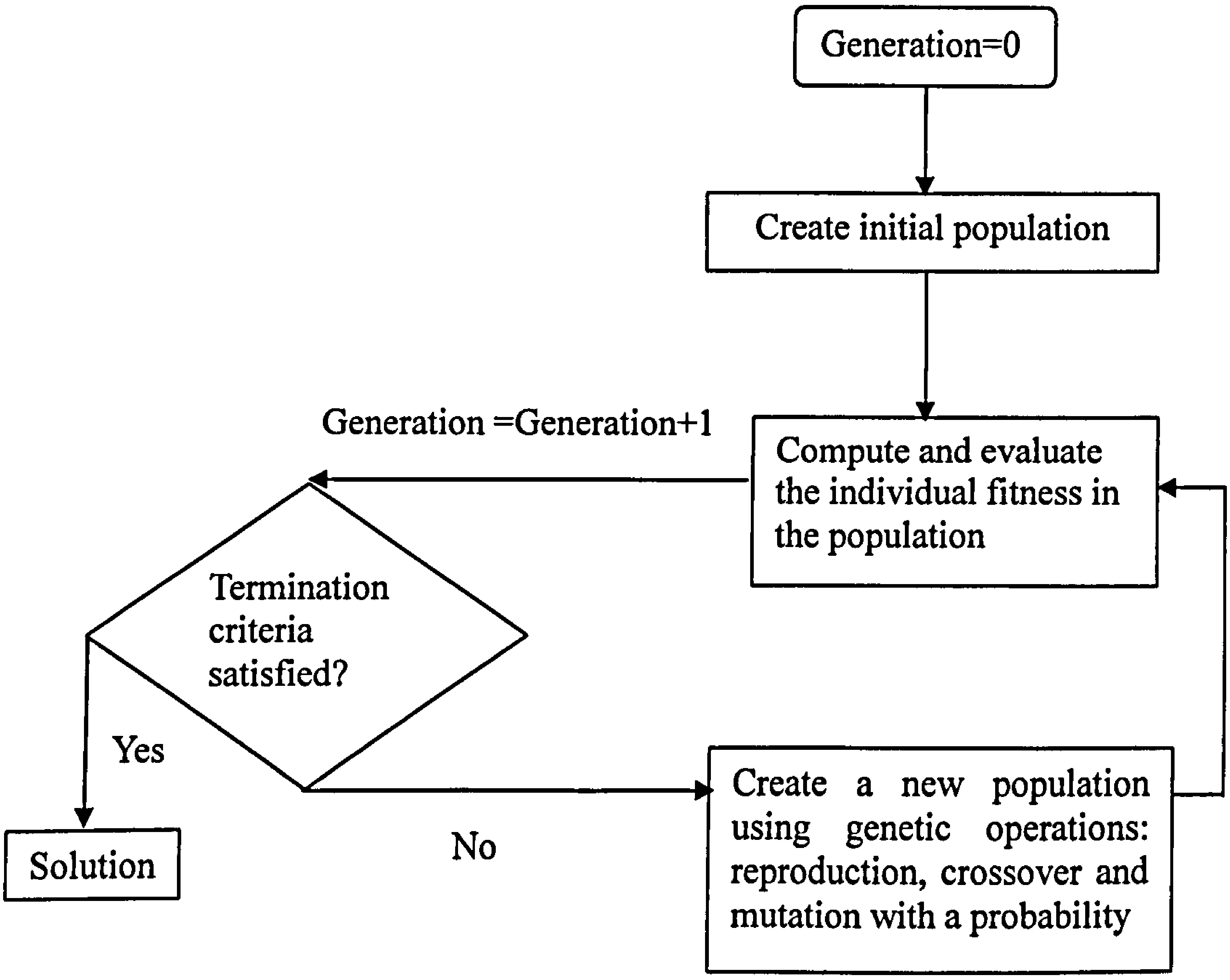


Figure 4.5 Flowchart of GP.

criteria can be the approaching of the maximum generation or the obtaining of the best fitness).

5. Acquire the solution of the target problem which is the individuals with the best fitness appearing in any generation.

4.5 FUZZY LOGIC

4.5.1 Introduction

Fuzzy Logic was initiated by Zadeh in 1965 [Zadeh, 1965] at the University of California in Berkeley as a way of processing data. However, it did not receive much attention until 1974, when Mamdani and Assilian used this technique to regulate a steam engine. The next major commercial development occurred in 1985, when researchers at Bell laboratories developed the first fuzzy logic chip. This chip led to a wide range of products such as cameras, camcorders and rice cookers. In 1993, OMRON built the first fuzzy computer [Dutta, 1993]. Nowadays, fuzzy logic has already become one of the fastest growing techniques of applied artificial intelligence technology [Krantz, 1999].

Classic logical is based on Boolean logic, which assumes that every fact is either entirely true or false (never both). Unfortunately, the inherent restriction of this technique is that it is incapable of representing the imprecise and incomplete concepts. For example, suppose that Boolean logic is used to identify whether a person is 'tall' or 'short'. If a threshold is given, that over 185cm is regarded as 'tall' otherwise is 'short'. Therefore a person is regarded as 'short' if his height is 184cm or 184.9cm, but 0.1cm or even 1 cm is too short to be noticeable.

Fuzzy logic extends Boolean logic to handle this type of problems and provides the means of identifying an intermediate value. To express imprecision quantitatively, it introduces a membership function that describes the degree of truth using real values between zero and one (inclusive). A membership value of zero indicates totally untrue and one indicates completely true. Any value between the two extremes indicates a degree of partial membership to the truth. Reconsider the example discussed above. If fuzzy logic is used to represent the height of a person, 185cm would have a membership value of one and 150cm would have a membership value of zero. Hence, 184cm would have a membership value very close to one, such as 0.97 and 155cm would have a membership value close to zero, such as 0.1.

If Boolean logic is called a binary logic, then fuzzy logic may be called a multi-valued logic. Using this logic, notions like ‘rather’, ‘very’, ‘a little’ can be formulated mathematically and processed by computers. It enables the computers to deal with problems in a more human-like way.

Fuzzy logic may appear similar to probability and statistics, however the term 70% has a different meaning in fuzzy logic and statistics. For example, Tom is 70% tall. In the probability statement, it means that "There is a 70% chance that Tom is tall" which supposes that there is a 70% chance that we know Tom is either tall or he is not. But in the fuzzy logic statement, it means that "Tom has a 70% degree of membership to be a tall people" which indicates that Tom is rather tall.

4.5.2 Membership Function

The membership function in fuzzy logic is a mathematic function which is variable according to the particular problems. For example, the membership function of tallness above can be described as:

$$Membership = \begin{cases} 0 & height < 150cm \\ \dfrac{height - 150}{35} & 150cm \le height \le 185cm \\ 1 & height > 185cm \end{cases}$$

Fig 4.6 shows a graphic representation of this membership function which is a very general type.

Another type of membership function is shown in Fig 4.7. It shows an example of the membership function of warmness. Suppose 18-25°C is warm, the temperature under this range is considered as cold and over this range is as hot. The membership function has a maximum one in the centre and decreases on both sides. The membership function is expressed as follows:

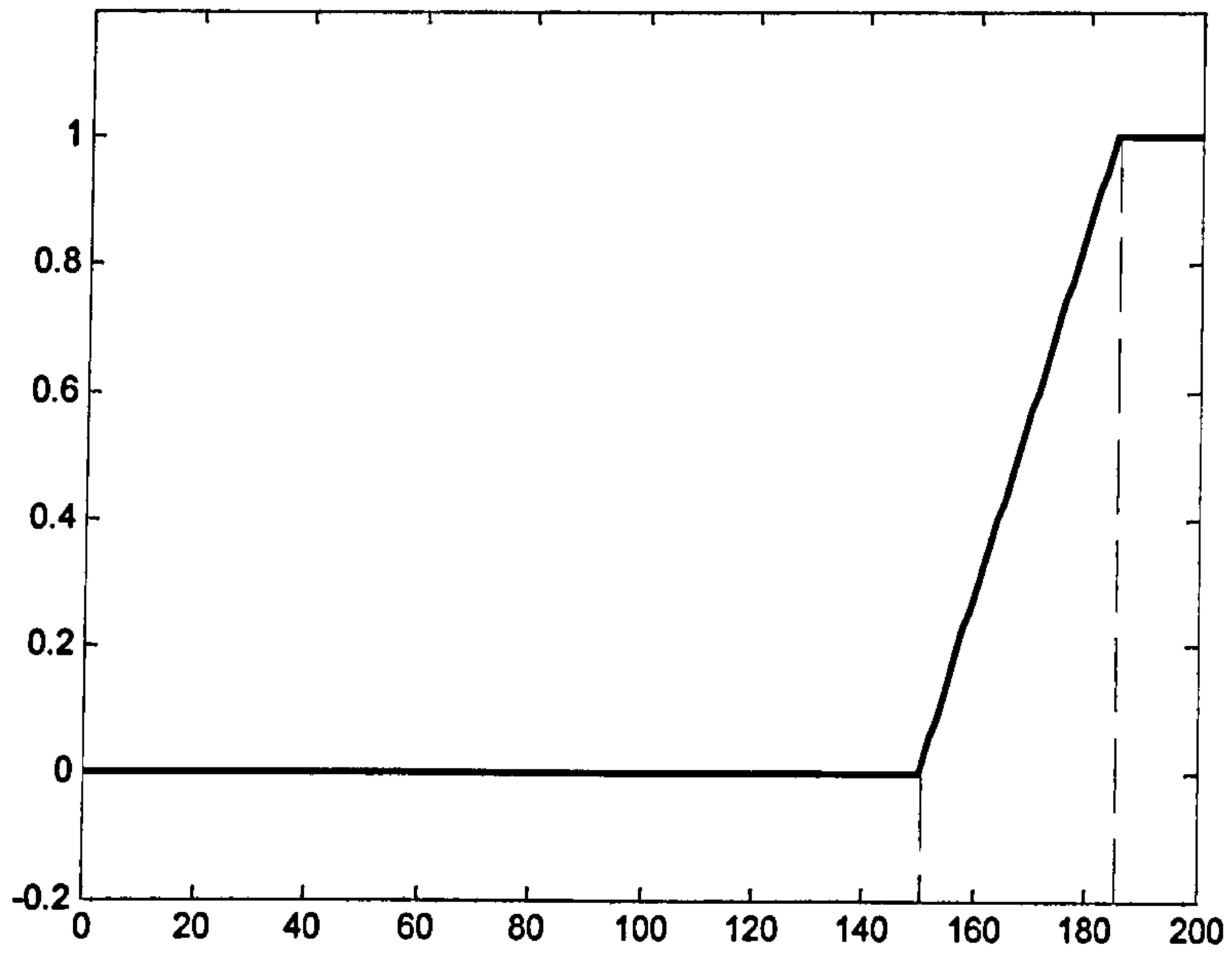


Figure 4.6 Membership function of height.

$$Membership = \begin{cases} 0 & T < 0^{\circ}C \\ \dfrac{T}{18} & 0^{\circ}C \le T \le 18^{\circ}C \\ 1 & 18^{\circ}C < T \le 25^{\circ}C \\ 1-\dfrac{T-25}{25} & 25^{\circ}C < T \le 50^{\circ}C \\ 0 & T > 50^{\circ}C \end{cases}$$

where T represents temperature.

In practice, the membership functions are much more complicated than in the examples above. They are not all linear but non-linear which depends on the type of problems. Moreover, the membership functions may be decided by a variety of factors rather than one. Many of them are even unable to be described using graphics.

4.5.3 Applications

As stated above, fuzzy logic is a well-suited technique to handle vague data and model imprecise reasoning procedures. Many commercial applications of fuzzy logic

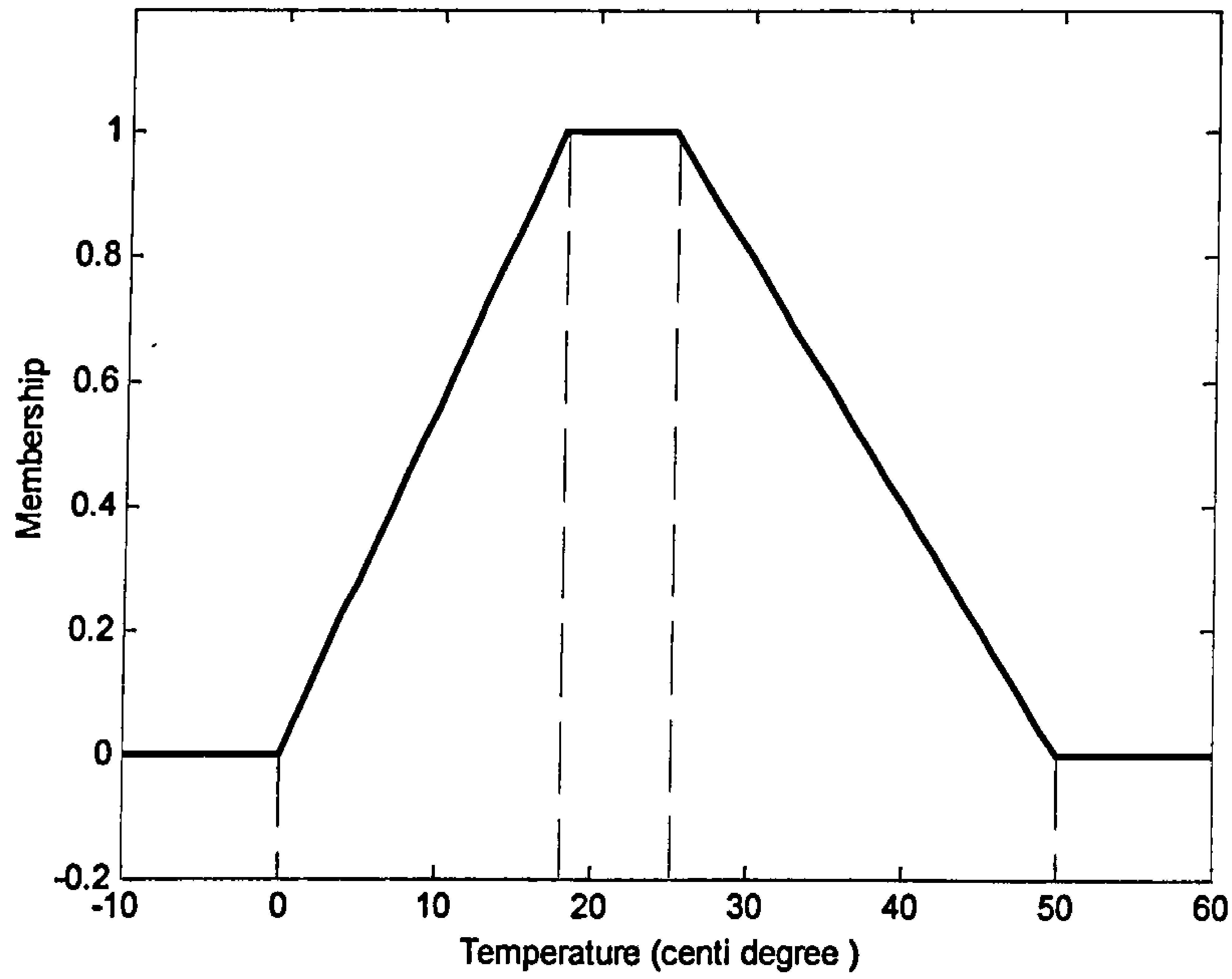


Figure 4.7 Membership function of warmness.

relate to control system, which refer to the management of a mechanical or environmental process [Krantz, 1999]. Practical applications of fuzzy logic are [Bauer, Nouak and Winkler, 1996]:

- Automatic control of dam gates for hydroelectric-power plants (Tokio Electric Pow.)
- Simplified control of robots (Hirota, Fuji Electric, Toshiba, Omron)
- Camera aiming for the telecast of sporting events (Omron)
- Substitution of an expert for the assessment of stock exchange activities (Yamaichi, Hitachi)
- Preventing unwanted temperature fluctuations in air-conditioning systems (Mitsubishi, Sharp)
- Efficient and stable control of car-engines (Nissan)
- Cruise-control for automobiles (Nissan, Subaru)
- Improved efficiency and optimized function of industrial control applications (Aptronix, Omron, Meiden, Sha, Micom, Mitsubishi, Nisshin-Denki, Oku-Electronics)

Besides the practical applications listed above, fuzzy logic is also widely employed in many other technical fields such as neural networks, genetic programming, and clustering. It is also a major technique in this thesis and will be used in later chapters.

Chapter 5

MR IMAGE PREPROCESSING

5.1 IMAGE SOURCE AND IMAGE TYPES

Before starting the description of image processing techniques developed for this thesis, the details of the MR images obtained are introduced in this section. All images used in this thesis were acquired from a 1.5T GE NVI scanner in the Southern General Hospital, Glasgow, UK. They were from 46 brain tumour patients. Each of them has one of four tumour types, gliomas (high/low), meningiomas, pituitary tumour, and medullablastma. There were 5 MR image modalities, which were described as FLAIR-FSE, T_1-SE, T_2-FSE, PD-FSE and T_1-contrast. T_1-SE represents a T_1-weighted MR image produced by the conventional spin echo (SE) pulse sequence. T_2-FSE and PD-FSE images are T_2-weighted and PD-weighted MR images generated by a fast spin echo (FSE) pulse sequence. T_1-contrast is a T_1-weighted acquisition with contrast enhancement i.e. via an injection of gadolinium-diethylene triaminopentoacetic acid (Gd-DTPA). A FLAIR-FSE image is therefore an MR image acquired by a special FLAIR with FSE MRI sequence [Liang, 2003; Fonar, 2003]. Not every patient had all five types of images. All techniques introduced in this thesis were applied to every image except blurred ones. As images were generated independently, the basic parameters such as TR, TE and IR were various. Therefore no such details were given in this thesis.

5.2 INTRODUCTION

Image preprocessing was the first part in the automatic brain tumour diagnosis system developed for this thesis and comprises three major steps.

As stated above, the main goal of this thesis was to diagnose brain tumours using image techniques alone without using an invasive procedure. The pivotal point was to obtain brain tissue intensities through segmentation. However, segmentation on raw MR images is not usually available due to their non-standard intensity values. Even displaying each MR image using the same window setup is not possible. The lack of a meaning for intensities also poses problems in image segmentation and quantification. These problems lead to difficulties in continuing this work; thereby the original MR images acquired directly from the hospital had to be preprocessed before segmentation. The first and the foremost step in image preprocessing part is the standardization of intensity scales for MR images.

The second step is non-brain region removal. Generally, the brain MR images contain the non-brain regions which are useless in this analysis since only the brain is of interest. However, the existence of the non-brain regions may greatly affect the segmentation results because they may have the similar range of intensities to tissues of interest. So, the removal of the non-brain regions was also indispensable to the image preprocessing part. The removal follows standardization because each image may require a different threshold if the intensity scales were not standardized.

To measure the tissue intensities, the brain MR images must be segmented into different tissues, such as white matter, gray matter, CSF and brain tumours. High contrast images enable much easier segmentation processes to be implemented, especially between white matter and gray matter which often have similar intensity values; otherwise it may cause failure of the segmentation. A contrast enhancement step was therefore as essential as the standardization and the non-brain removal before the segmentation.

In this chapter, three sections describe these steps respectively.

5.3 IMAGE STANDARDIZATION

As we know, one of the great advantages of MRI is that a variety of MRI modalities are available to set up the different contrasts to different tissues. However, this advantage also causes a major problem in that the signal intensities of MRI do not have a fixed value, not even in the same body region of the same patient obtained on the same scanner [Nyúl and Udupa, 1999]. It indicates that the same tissue type cannot be represented by the same scale of signal intensities in the different MR images. Due to this, the intensities in different images are not comparable with each other. It also implies that MR images cannot be displayed using similar formats and images are unable to be segmented automatically. The comparison of tissue intensities can quantify the analysis, which is crucial in many image analyzing systems. Intensity normalization is thus necessary. The automation of the analyzing systems also requires normalized intensity scales, otherwise the system parameters may be modified for each image. So, the standardization of intensity scales for MR images must firstly be applied.

Attempts have been proposed to calibrate MR signals using phantoms during acquisition. However, post-processing on MR images is more attractive using a number of ever stored MR images. A standardizing method was presented by Nyúl and Udupa (1999). It offers a two step process consisting of a training step for each type of MR image and body region, and a transformation step on each given image.

It scales the original MR images into the same maximum and minimum intensity values. However, this method is not suitable to images which include pathological abnormalities especially in intensity-based analysis. For example, as we are working on diagnosing brain tumour types based on tissue intensities, if each image has the same maximum and minimum intensity, then the tumours in different images are unable to be represented because they have similar intensities. To overcome this problem, an easy post-standardization method is presented in this thesis which can retain the basic intensity features of the images and standardize the intensity scales.

Fig.5.1 shows the histograms of two general MR images. One image had a range of intensities from 0 to 800 and the other one was from 0 to 1800. Without standardization, the first image may be displayed clearly by means of either unsigned 8-bit or 16-bit format. However, using the former the image may be too bright and using the latter the image may be too dark. The second image probably has the same type of problems, but it is obvious that the tissues in both images have rather distinct intensity values.

From viewing of the histograms, the first peak usually represents the number of background pixels, which is of no interest. Additionally, due to the high number of background pixels in MR images and their zero or very low intensities, they must be excluded in standardization otherwise their existence may dominate the standardization. The background pixels compose of the first peak of the histogram which generally ends at around 5% of the maximum intensity value in the image. However, this percentage may be variable for different images. It must be adjusted interactively to improve the accuracy. Assume that the threshold percentage of the maximum ranges from 1% to 10%. The histogram of the foreground pixels is initially obtained without the intensities lower than the 1% of the maximum. If the number of pixels at the first several low intensity values is not greater than 100 (this number is especially used for 256x256 images and 400 for 512x512 images) in the new histogram, the background peak is considered to be eliminated completely. Otherwise, by increasing the percentage of the maximum intensity that the first peak ends, 1% per loop, until the pixel number at the first several low intensities is lower than 100. A simple implementation of discarding the background pixels is given by Fig.5.2.

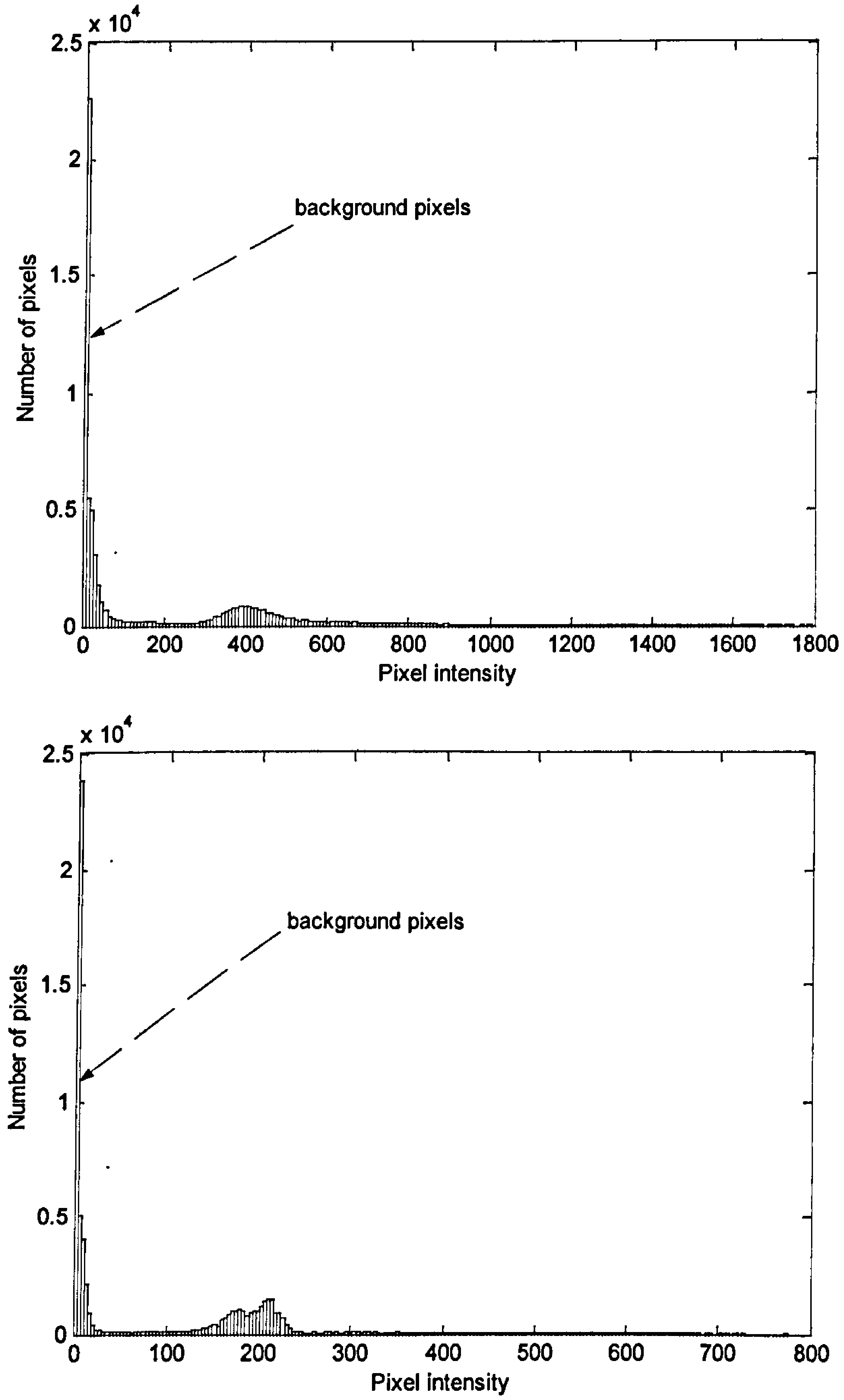


Figure 5.1 Histograms of two MR images.

```
$i=0$;
Threshold percentage= 0; New histogram= Original histogram;
While
[The total number of pixels of the lowest ten intensity values in the new histogram >100]
and [Threshold percentage <10%],
    $i=i+1$;
    Threshold percentage=1% + ($i$-1)×1%;
    Threshold= Threshold percentage × maximum intensity
    New histogram= The histogram starting at the value of Threshold;
End;
```

Figure 5.2 Implementation of removing background pixels.

With the background pixels excluded, the histograms of foreground pixels in Fig.5.1 are shown in Fig.5.3. In order to find a curve to fit the distribution trend, the number of pixels is normalized to between 0 and 1. Let σ be the standard deviation and μ be the mean of the foreground pixel intensities in an MR image. Assume the fitting line of the intensity distribution satisfy a normal distribution ('*' line) with the standard deviation σ_0 and the mean μ_0. Their relationship can approximately be expressed as follows:

$$\sigma_0 = \frac{\sigma}{n}, \mu_0 = \mu + \delta \tag{5.1}$$

where n is the amplitude and typically chosen to be 4. δ is a linear displacement. On that account, this specific type of distribution is named "sub-normal distribution" in this thesis. A data set which is satisfied as a normal distribution is that where about 95% of the data are within a distance of two standard deviations on either side of the mean. The standard deviation therefore can be interpreted as the degree of the data spreading from the mean [Siegel and Morgan, 1996]. As stated above, the foreground pixel intensities in MR images satisfy a sub-normal distribution. Its histogram has a

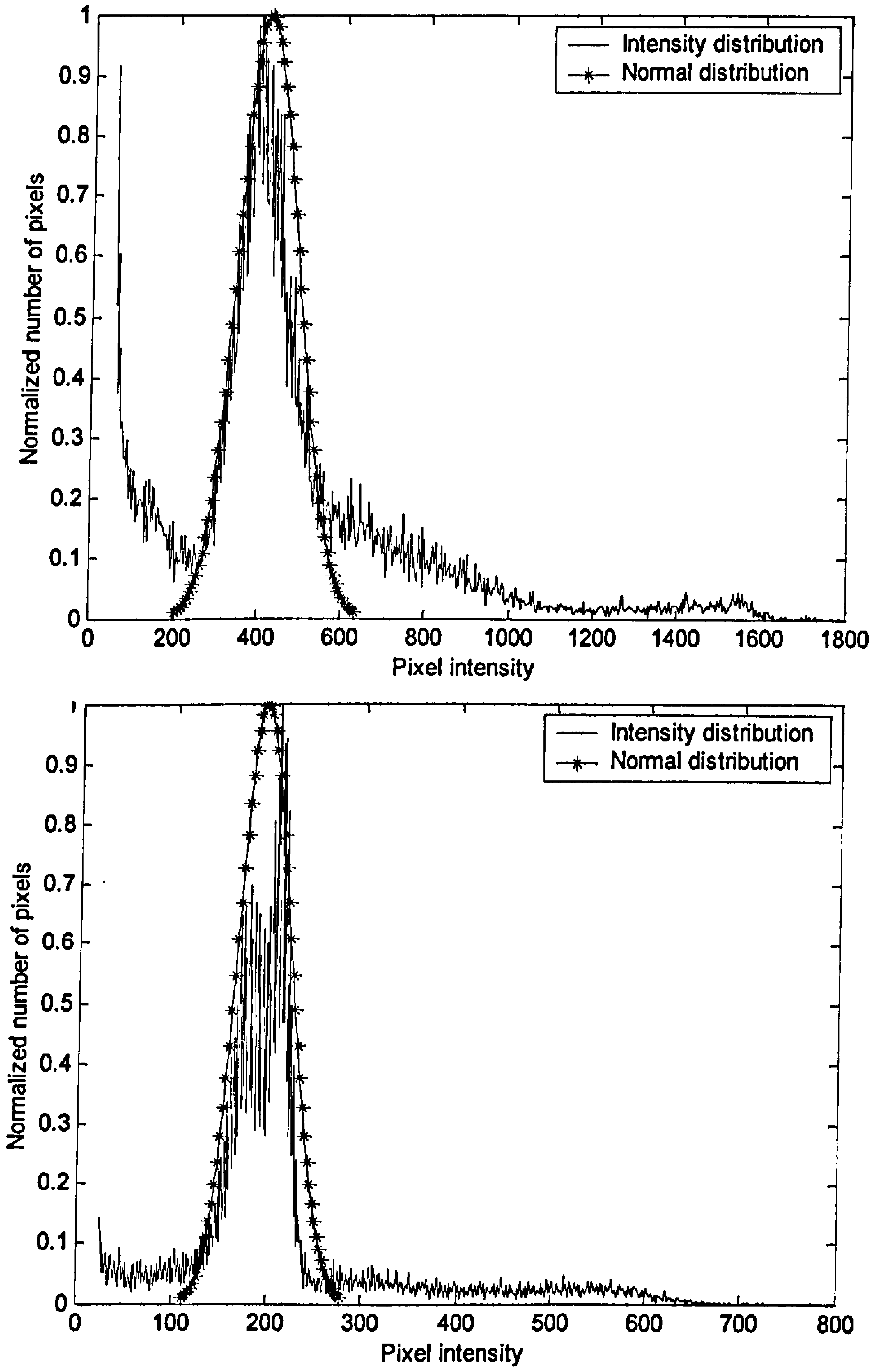


Figure 5.3 Histograms of foreground pixels in Fig.5.1 (top) (bottom).

similar shape to a normal distribution with a standard deviation four times higher than that of the sub-normal distribution. It implies that the sub-normal distribution of the foreground pixel intensities have about 95% of the intensities falling between half the standard deviations on either side of the mean. The distribution can, however, sometimes be displace to some extent from the mean. Therefore, the standard deviation can also be described as the degree of spread of intensities from the mean.

The formula of the standard deviation is as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} \tag{5.2}$$

Where x_i (i=1...N) is the intensity of any foreground pixel; N is the number of foreground pixels in each image. σ represents the standard deviation and μ represents the mean of the data set. If the values of σ and μ are known, the shape of the intensity histogram can then be determined which then indicates that the intensity scale is approximately fixed. According to this idea, the standardization of intensities can be implemented by transforming the foreground pixel intensities which will result in each MR image having the same σ and μ.

Initially, the mean μ and the standard deviation σ are transformed as zero and one. Dividing σ on both sides of Eq. (5.2) gives:

$$1 = \sqrt{\frac{\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma})^2}{N}} \tag{5.3}$$

It can also be expressed as follows:

$$1 = \sqrt{\frac{\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma} - 0)^2}{N}} \tag{5.4}$$

This equation can be regarded as a new intensity value which has a zero mean and a standard deviation of one. The new intensity has a form of:

$$x_{new} = \frac{x_i - \mu}{\sigma} \qquad (i=1...N) \tag{5.5}$$

However, the new intensities mostly cannot supply good contrasts in MR images. More appropriate values of the mean and the standard deviation must be chosen for the intensity scale standardization of all MR images.

Suppose $X_s \supset x_{si}$ (i=1...N) are the final intensities after the standardization and they have a mean as μ_s and a standard deviation as σ_s. They can be also expressed similarly as in Eq. (5.4):

$$1 = \sqrt{\frac{\sum_{i=1}^{N}(\frac{x_{si} - \mu_s}{\sigma_s} - 0)^2}{N}} \tag{5.6}$$

Equaling Eq. (5.4) and Eq. (5.6),

$$\frac{x_i-\mu}{\sigma}=\frac{x_{si}-\mu_s}{\sigma_s}$$

$$x_{si}=\frac{\sigma_s}{\sigma}(x_i-\mu)+\mu_s \tag{5.7}$$

So, Eq. (5.7) gives the relationship between the original intensities and the standardized intensities. The chosen values of μ_s and σ_s should be those that give the best contrasts in most MR images.

The MR images acquired for this thesis were all unsigned 16 bit images; the range of the intensities was between 0 and 65535. However, most images only had a maximum intensity lower than 5000. To enhance the image contrast while standardizing, μ_s and σ_s were chosen to be 32500 and 10000 which enabled the images to occupy the full range of intensities. After the transformation for standardization, the intensities of foreground pixels in each MR image satisfy a sub-normal distribution with the mean of 32500 and the standard deviation of 10000. The process of standardization is summarized as follows:

(1) Determine the value of σ_s and μ_s (σ_s =10000 and μ_s =32500 for unsigned 16 bit images)

(2) Eliminate the background pixels from an original MR image as in Fig.5.2.

(3) Compute the mean and the standard deviation of the foreground pixel intensities for the original MR image.

(4) Standardize the foreground pixel intensities according to Eq. (5.7).

(5) Add the background pixels back to the standardized pixels and obtain a new MR image.

Fig.5.4 shows the standardized histograms of Fig.5.3. It is clear that these two histograms still retain their original shapes and features as in Fig.5.3; but the intensities in two images have similar scales.

Two example images are given to clarify these results. A T_1-weighted MR image is shown in Fig.5.5, displayed as a binary image (0 to 1) and an unsigned 8-bit image (0 to 255). However, this image has an intensity scale from 0 to 619. Both displays cannot show its details. If the image is displayed as unsigned 16-bit (0 to 65535), it becomes completely dark and does not show any contours. Fig 5.6 shows

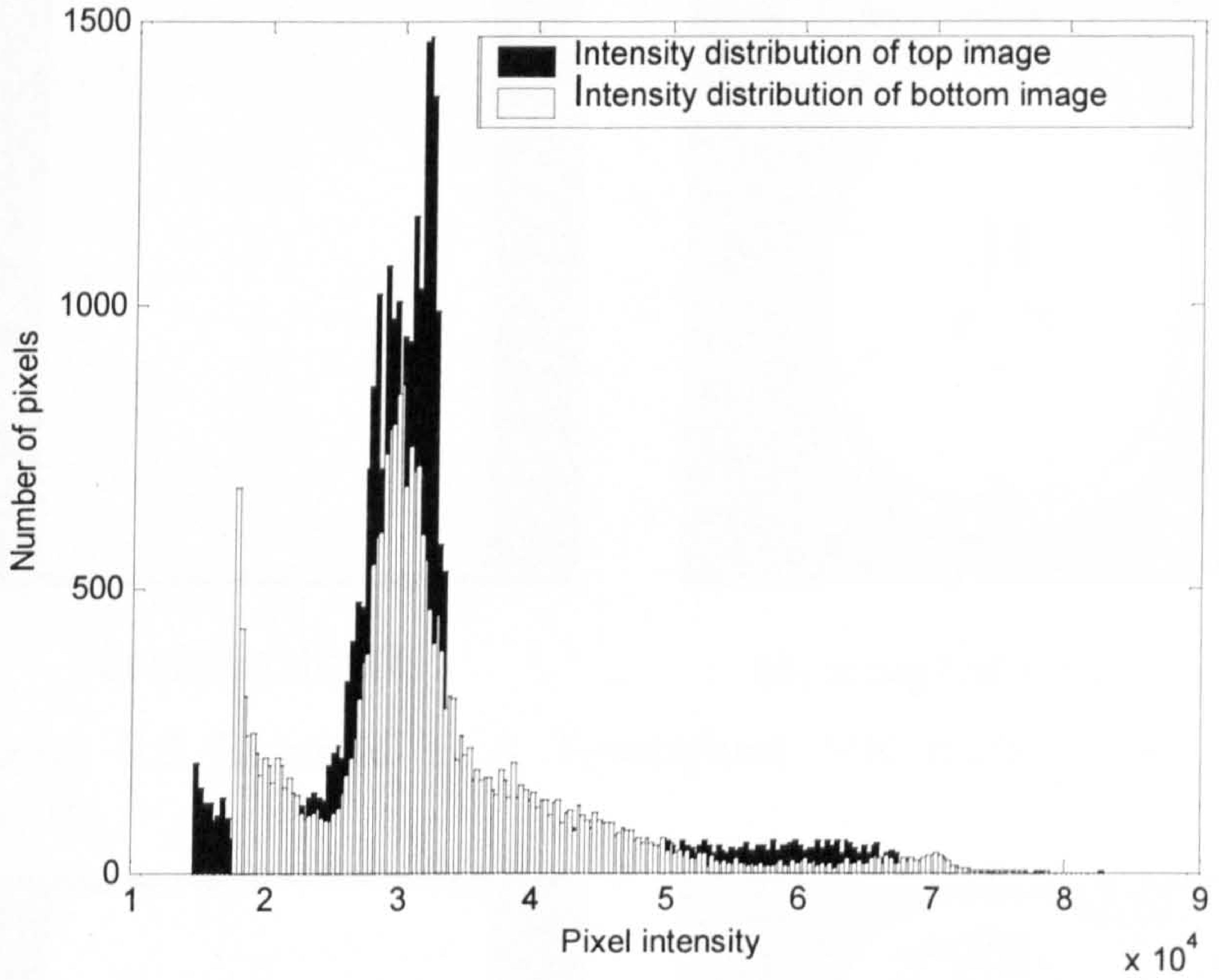
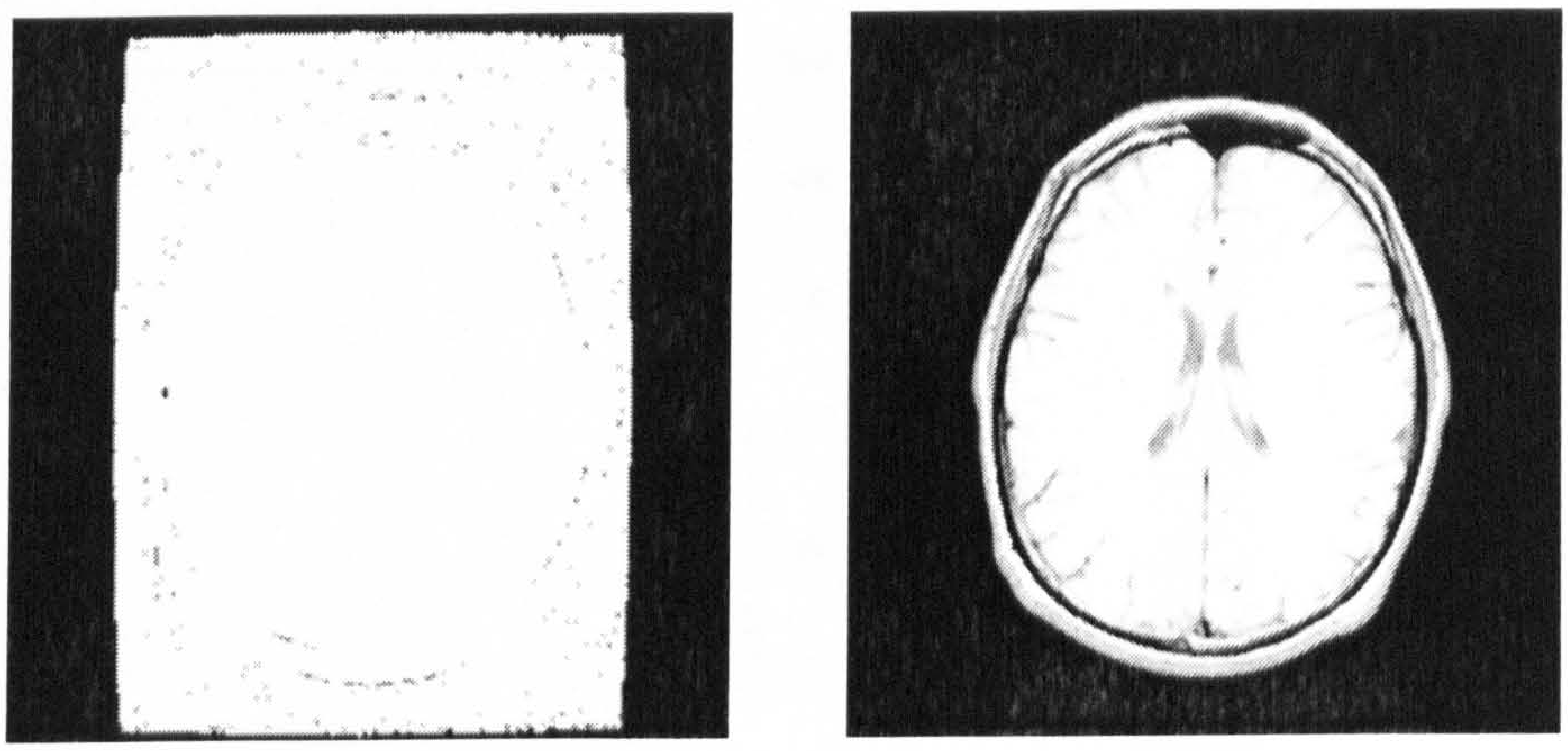


Figure 5.4 Standardized histograms of Fig.5.3.

another T_1-weighted MR image also displayed as a binary and an unsigned 8-bit image. It has intensities ranging from 0 to 904. The latter image has a relatively good contrast and it is too dark if displayed as an unsigned 16-bit image. It demonstrates that MR images cannot be displayed uniformly as described above.

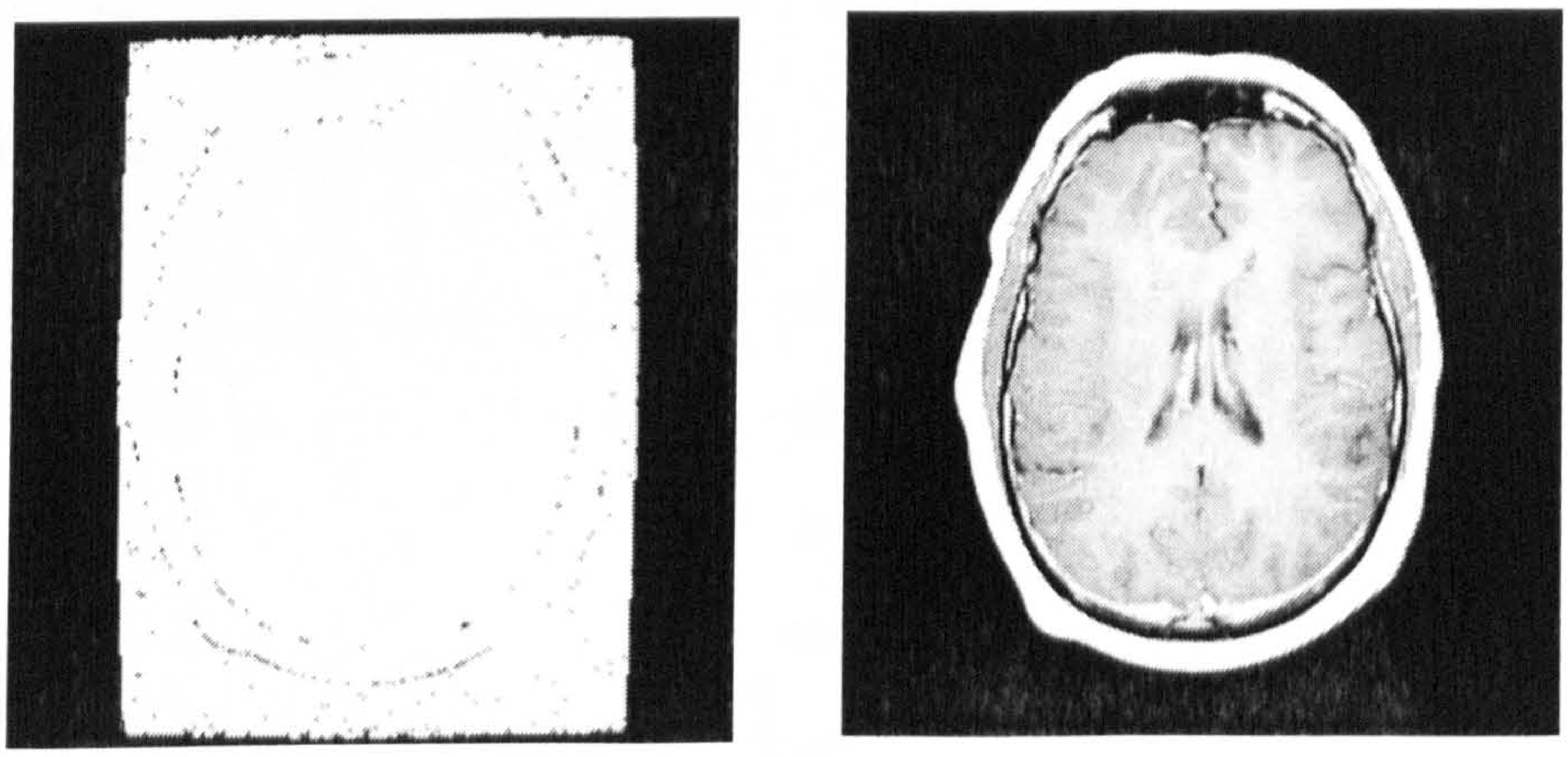
Using the standardization process, both images above can be standardized and displayed as unsigned 16-bit images. Their standardized images are shown in Fig.5.7. Both of them can be displayed similarly with good contrasts. Fig.5.8 compares the histogram of the original image in Fig.5.5 with the histogram after the standardization. It also shows the histograms of the foreground pixels before and after the standardization. It demonstrates the capability of this standardization method which can standardize the intensity scales without affecting the original features of MR images.

All MR images in our study were displayed as unsigned 16-bit images after standardization with good definition. Their intensities therefore have similar scales which will be useful for later comparison.



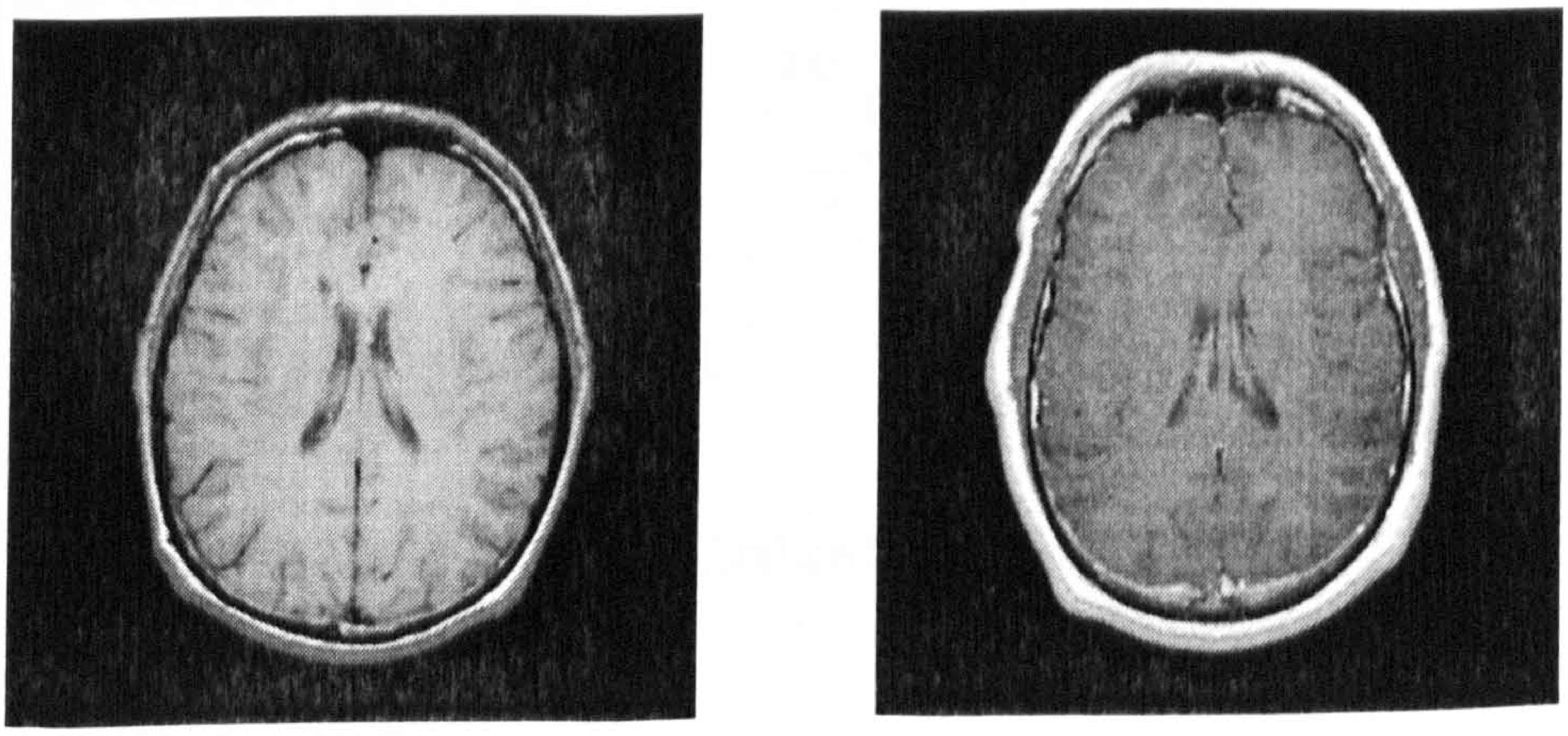
(a) binary display (b) unsigned 8-bit display

Figure 5.5 Example 1: A T_1-weighted MR image using different display.



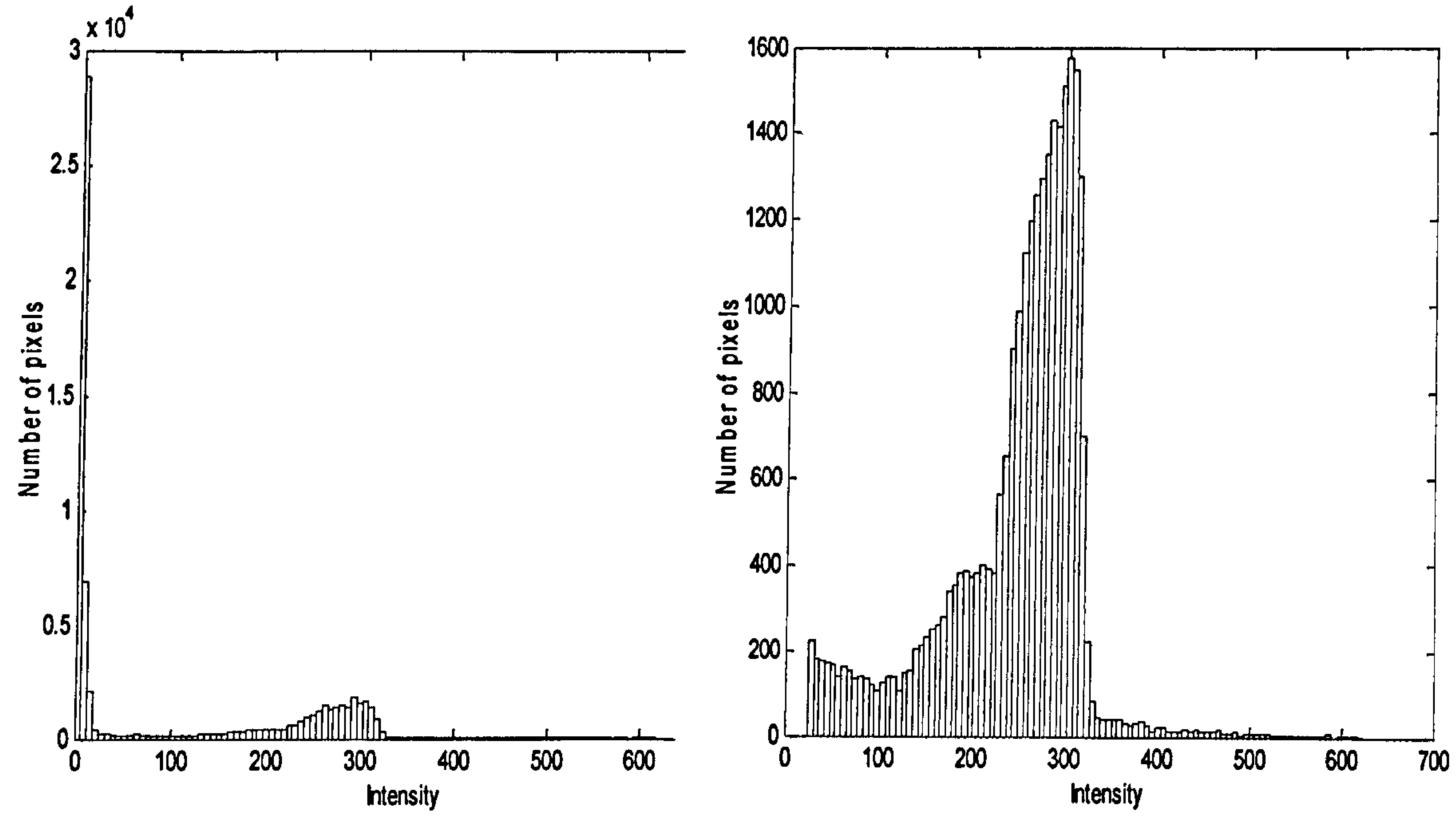
(a) binary display (b) unsigned 8-bit display

Figure 5.6 Example 2: A T_1-weighted image using different display.



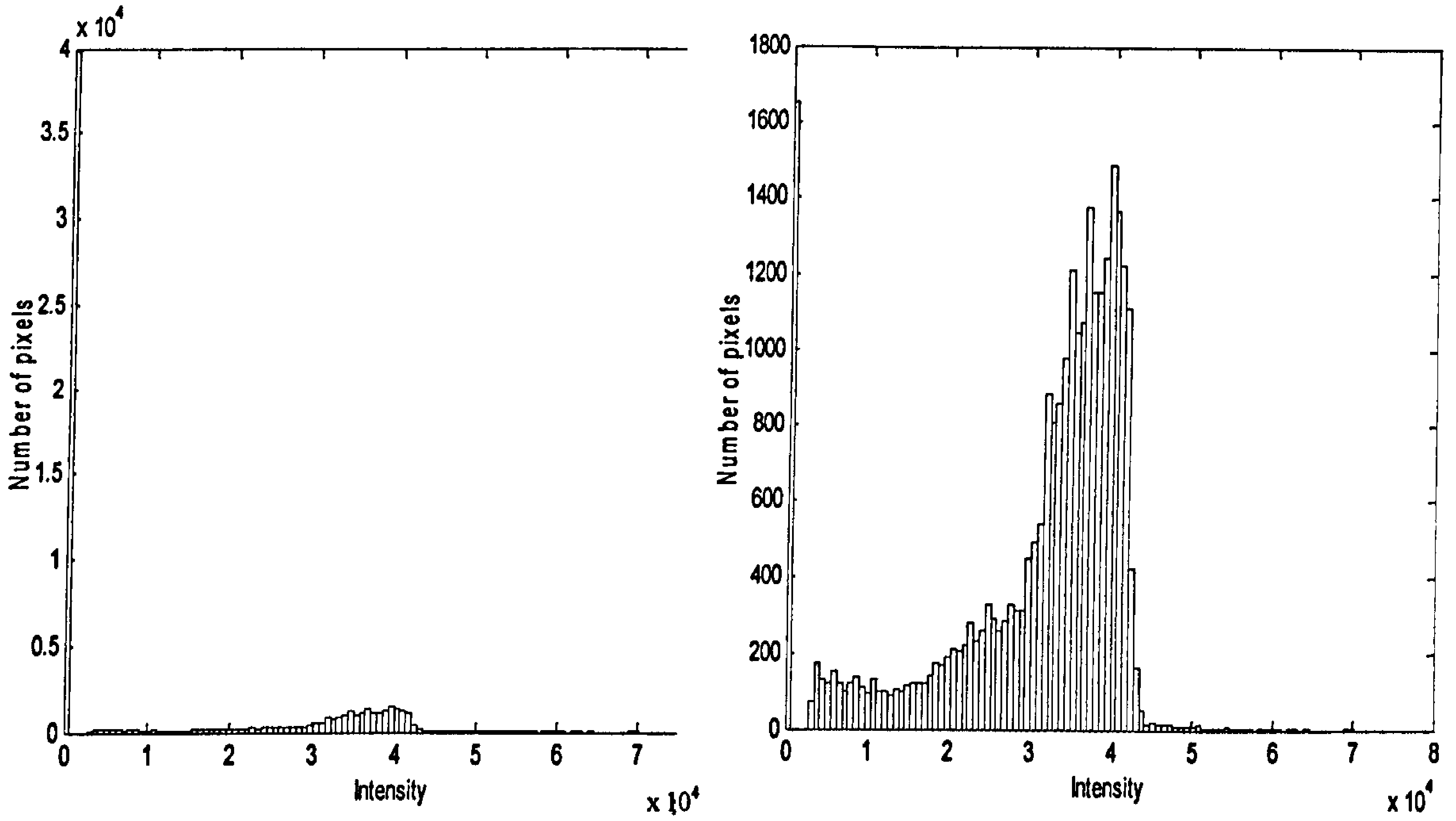
(a) standardized image of Fig.5.5 (b) standardized image of Fig.5.6

Figure 5.7 Standardized results.



(a) Histogram of the original image

(b) Histogram of the foreground pixels in the original image



(c) Histogram of the standardized image

(d) Histogram of the foreground pixels in the standardized image

Figure 5.8 Histograms of original and standardized MR image.

5.4 REMOVAL OF NON-BRAIN REGIONS

Non-brain regions include structures such as the skull, fat, meninges and scalps, which are not of interest in this thesis. However, their intensities may lie in a similar range to other brain tissues such as brain tumours and consequently will affect the segmentation results. Therefore the removal of the non-brain regions is an essential step in image preprocessing. The most popular technique proposed for this task is morphological processing.

5.4.1 Morphological Processing

Morphological processing is a non-linear image processing technique which is built on the foundation of set theory. The basic idea is to extract relevant structures of the image by probing the image with another set of known shape called structuring element (SE) [Soille, 1999]. The shape of SE is usually symmetric, such as squares, rectangles and circles. Commonly, this technique is applied to binary images whose pixel values are only 0 (black) or 1 (white).

Logic operators, 'AND' and 'OR' are the basic blocks in morphological processing. They consist of major morphological operators such as erosion and dilation.

5.4.1.1 Erosion

Erosion is a morphological operator which is used to reduce the size of an image by removing 'on' pixels from the boundaries of objects and also to increase the size of holes by removing pixels around the perimeter of the hole. Consequently, it can help to break down the connections between objects [Gaborski, 2001]. The symbol of erosion is AΘB where A is the image and B is the SE. Let A be an image with pixels a, and B be the SE with elements b. The erosion of A by B is defined as:

$$A \Theta B = \{a \in A \mid a + b \in A, b \in B\} \qquad (5.8)$$

The function of erosion includes removing noise and other small objects, breaking the weak connections between objects, and increasing the size of holes within an object.

Since it usually operates on binary images, erosion can be simply described as follows:

$$A\Theta B = \{\forall a : a \cap b \neq b \Rightarrow a = 0, othwise, a = 1\} \tag{5.9}$$

where A represents a binary image and B is still the SE.

Fig.5.9 shows an illustration of erosion. The object in the image is finally eroded by the SE into a smaller one.

5.4.1.2 Dilation

Dilation is another major morphological operator. It is used to enlarge the size of images by adding 'on' pixels to the boundaries of objects and decrease the size of holes by adding the pixels around the perimeter of the holes. Dilation can be regarded as a dual operation to erosion which can enhance the connections between objects. The symbol used to denote dilation is A⊕B, where A is the image and B is the SE. Let A be an image with pixels a, B be the SE with elements b, and c represent the new pixels in A after dilation. The dilation of A by B is defined as:

$$A \oplus B = \{c \in A \mid c = a + b, a \in A, b \in B\} \tag{5.10}$$

The function of dilation is enlarging the objects, filling the holes and thickening the connections between objects.

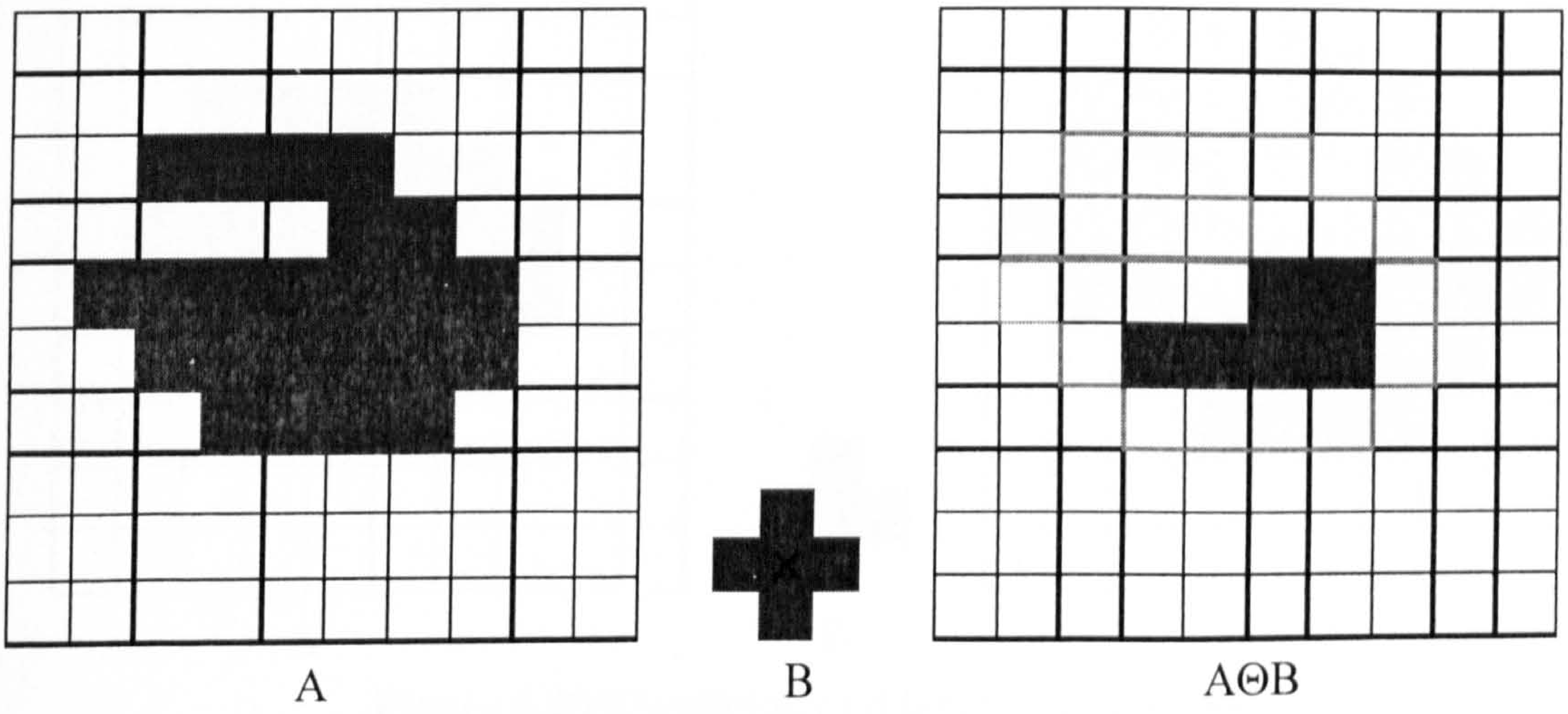


Figure 5.9 Illustration of erosion [Larsen 2000].

The simple expression for binary dilation can be described as following:

$$A \oplus B = \{\forall a : a \cap b \neq \Phi \Rightarrow a = 1, otherwise, a = 0\} \tag{5.11}$$

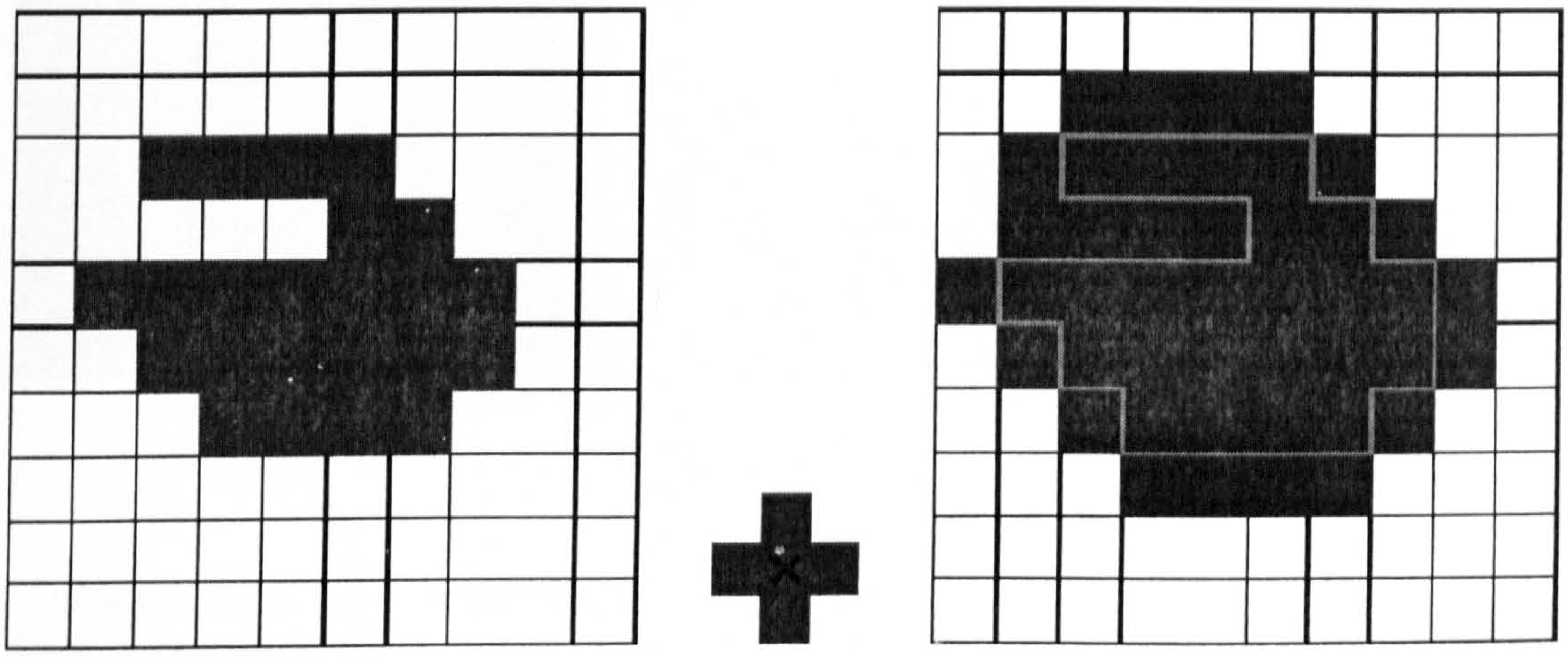
where Φ is an empty set. Fig.5.10 shows an illustration of dilation. The object in the image is enlarged after dilation.

5.4.2 Removal of Non-brain Regions Using Morphological Processing

Generally, there always exists some space between the non-brain regions and the brain which has lower intensities than other parts of MR images. It leads to a dark gap between the non-brain regions and the brain and the non-brain regions can be removed using the morphological processing.

Since the morphological processing is commonly applied to binary images, a threshold must be selected initially to obtain the binary mask of the images. The optimal value of the threshold can enhance the gap between the non-brain regions and the brain in the binary mask, thereby simplifying removal.

As stated above, the distribution of foreground pixel intensities in MR images is a sub-normal distribution. The center of the histogram approximately represents the mean of the intensities which approaches the intensity value with the maximal number of pixels. As the gap between the non-brain and the brain always has lower



A B A⊕B

Figure 5.10 Illustration of dilation [Larsen 2000].

intensities than the mean, the threshold for the binary mask is chosen practically to be the intensity value corresponding to the most number of pixels in the histogram of foreground pixels. The ideal threshold can let a binary mask have a pixel value of 0 for the gap and not have connected pixels within it. Fig.5.11 shows a standardized T_1-weighted MR image and its foreground pixel histogram pointed to the threshold.

The process of the non-brain region removal can be summarized in five steps: thresholding, erosion, labeling, dilation and masking.

(1) The threshold, the intensity value corresponding to the most number of pixels in the foreground pixel histogram, is applied to create a binary MR image.

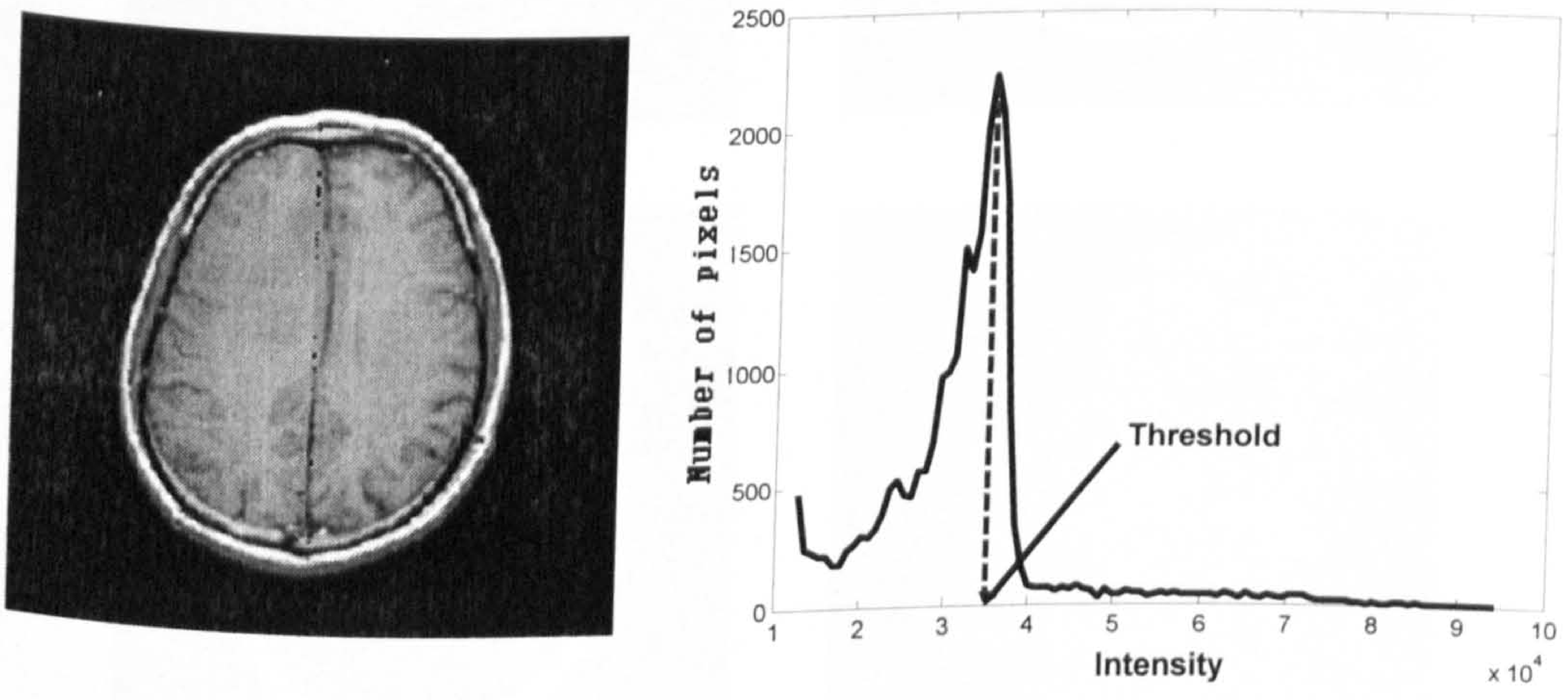
(2) Erosion is then employed to break down the connections between the brain and the non-brain regions.

(3) Labeling finds the non-brain regions and isolates them from the binary image.

(4) Dilation recovers the non-brain regions as they are thinned down by erosion.

(5) Masking removes the dilated non-brain regions from the original MR image.

In this thesis, the SE of the erosion is composed of two 'on' pixels horizontally as [1 1]. The SE of the dilation is a little larger than that of erosion. It is a square

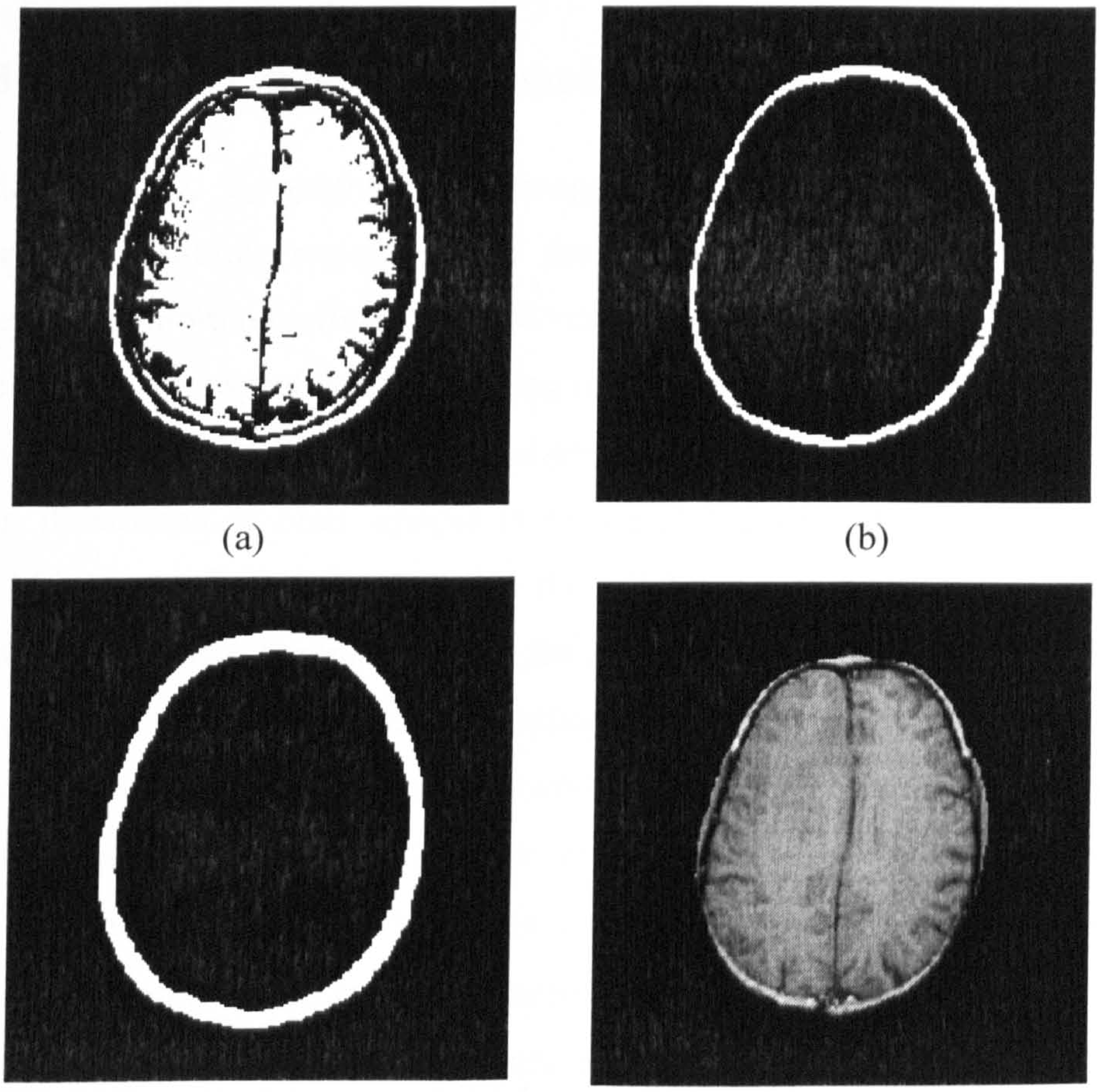


(a) (b)

Figure 5.11 (a) a standardized T_1-weighted MR image (b) histogram of foreground pixels.

composed of four pixels $\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$. So the dilated non-brain region is slightly larger than the corresponding one in the binary mask. It assures the entire non-brain region can be removed completely.

Fig.5.12 shows the process of the non-brain region removal for the MR image in Fig.5.11. A binary mask (Fig.5.11 (a)) was obtained using the threshold 32128. Then, the non-brain region was labeled (The peripheral section was the non-brain) followed by an erosion operation (Fig.5.11 (b)); thereafter it was expanded through the dilation (Fig.5.11 (c)). The dilated non-brain region is clearly larger than the real non-brain region. The brain region only image was acquired after removing the non-brain region from the original image (Fig.5.11 (d)).



(a) (b)

(c) (d)

Figure 5.12 Process of non-brain region removal.
(a) binary mask (b) non-brain region after erosion (c) non-brain region after dilation (d) brain only

The morphologic processing can deal with most MR images but it may fail in certain types of images. For example, T_2-weighted MR images display CSF with high intensities, thus the number of pixels with high intensities accounts for a higher proportion of an image. However, the standardized mean was 10000 for any type of image and this makes the non-brain regions in T_2-weighted images exhibit a rather thin layer. As a result, the non-brain regions are easily broken apart in the erosion and may not be labeled completely. For this, a higher threshold may be chosen which was the mean plus a standard deviation of the foreground pixel intensities. In addition, a simple supplementary method was developed for this thesis to handle the non-brain region removal for the unsuccessful examples. Some other types of problems encountered in the non-brain region removal are also described.

5.4.3 Removal Using A Supplementary Method

As stated above, T_2-weighted MR images have high intensities for CSF which thinned the non-brain regions due to the equal standardized mean for every MR image. The non-brain regions are hence very difficult to label if they are broken apart in the erosion. In respect that the labeling is the main reason causing the failure of the removal, its exclusion from the removal process may solve the problem.

An illustration of brain images is shown in Fig.5.13. It represents a 256x256 brain MR image. It was assumed that the non-brain regions have higher intensities than the background; the intensities of the gap are lower than both the non-brain and the brain. A threshold was chosen according to the determination of the threshold for binary masks in morphological processing. The image was divided into four quadrants as labeled in Fig.5.13. The pixel intensities were compared with the threshold. The comparison started at the first column of the first quadrant from top to bottom, and then from left to right; whereas in the second quadrant it was from top to bottom, right to left. In the third quadrant the comparison started from bottom to top, left to right and in the fourth quadrant it was from bottom to top, right to left. In each column, the comparison stopped when the border of the quadrant was reached or when the first pixel with a lower intensity than the threshold, right after a pixel with a higher intensity than the threshold, was met. The reason was that the pixel intensities

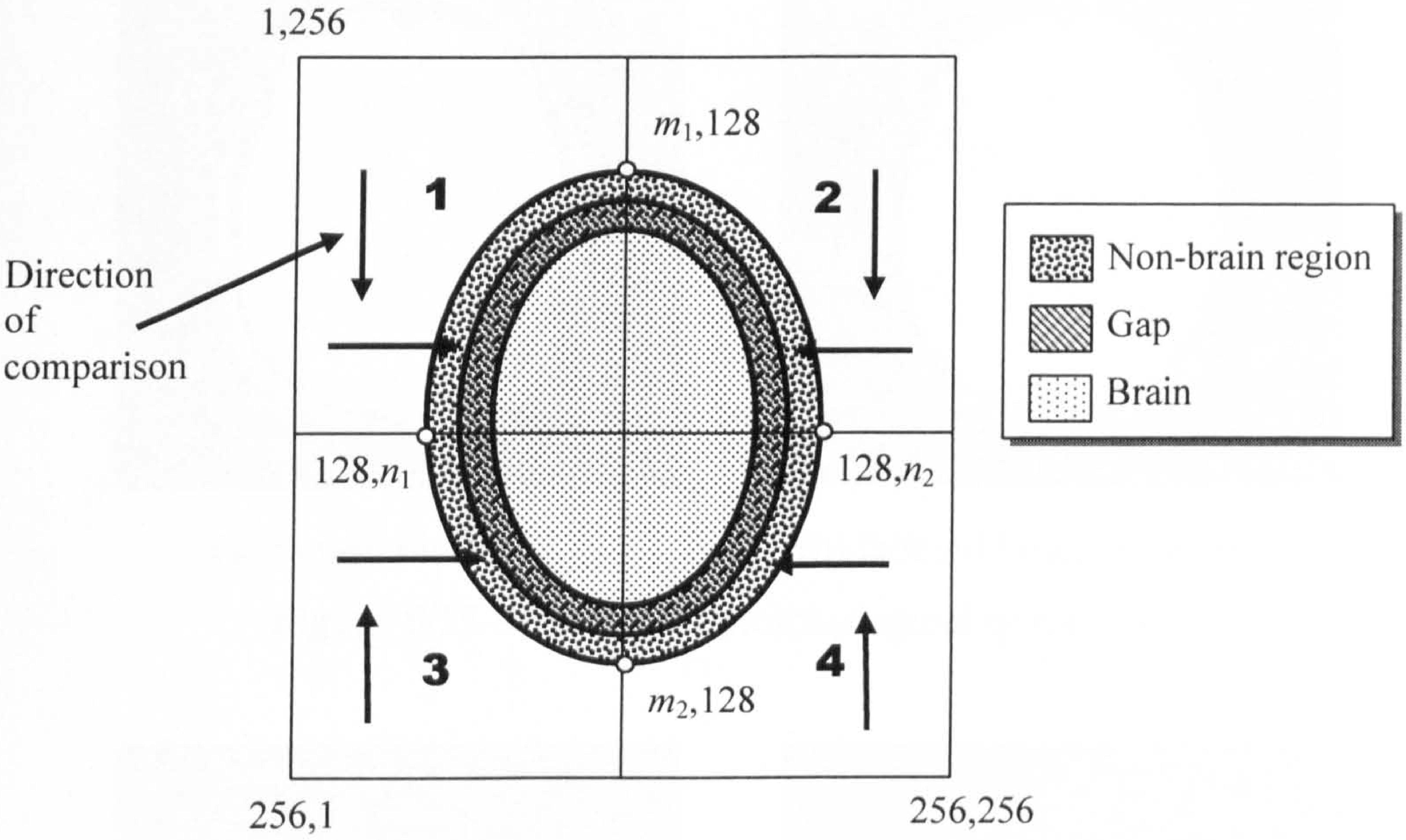
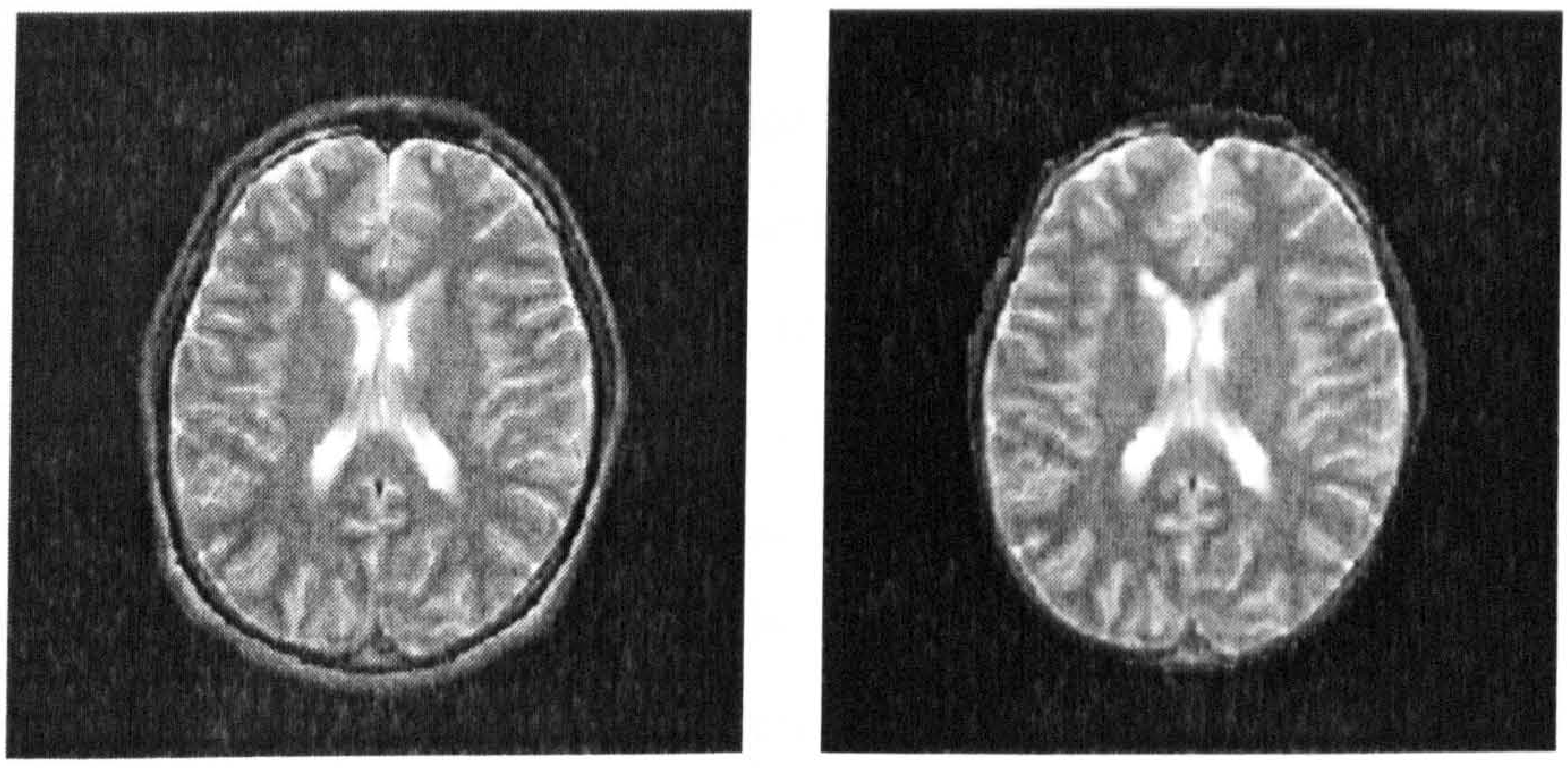


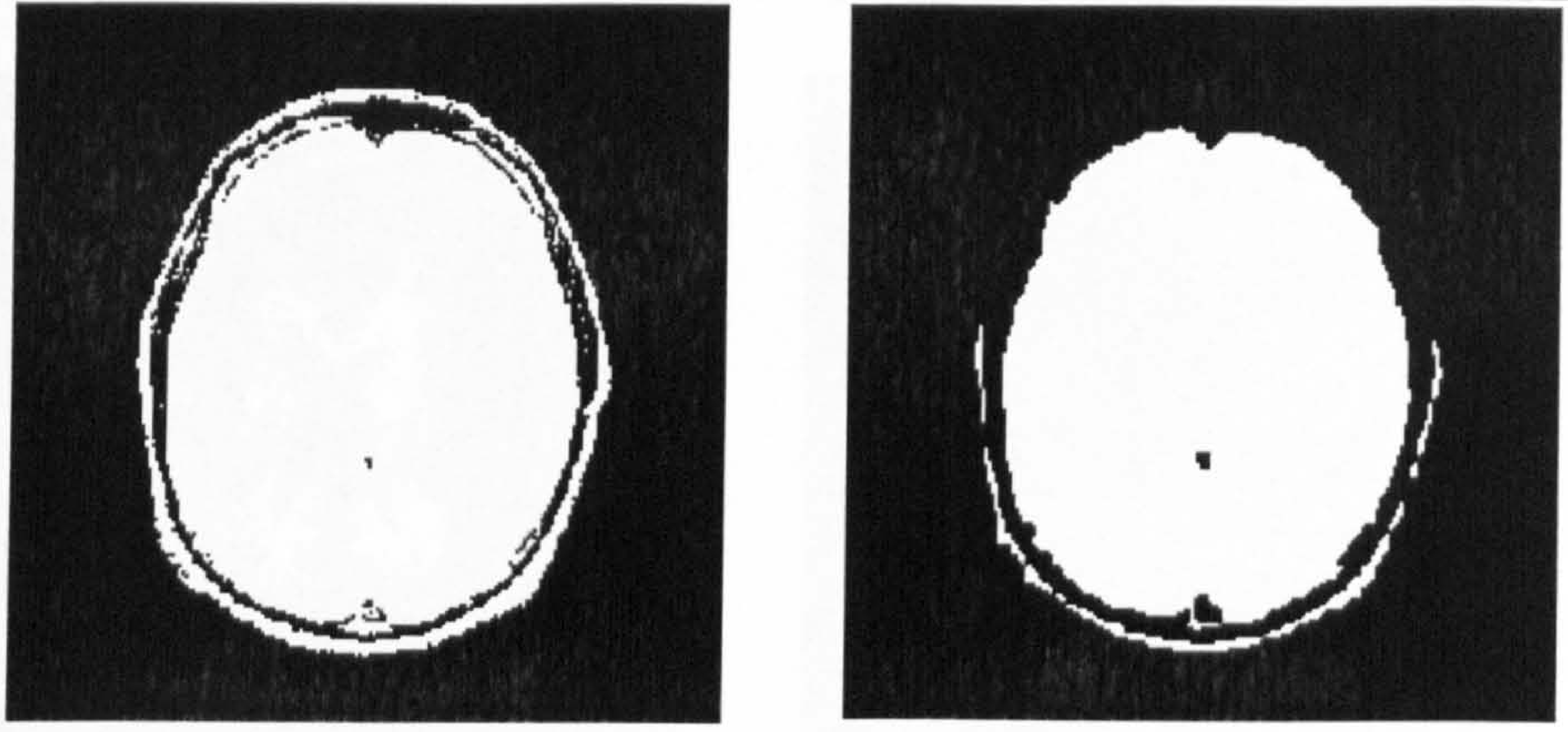
Figure 5.13 Illustration of brain image.

initially were all lower than the threshold when compared with the background pixels. Then, the pixel intensities were higher than the threshold when compared with the non-brain pixels. When the pixel intensity became lower again than the threshold, it indicated that the previous pixel was the last pixel of the non-brain regions. Finally, all pixels already compared were set to zero and other pixels remained at their intensities in the image.



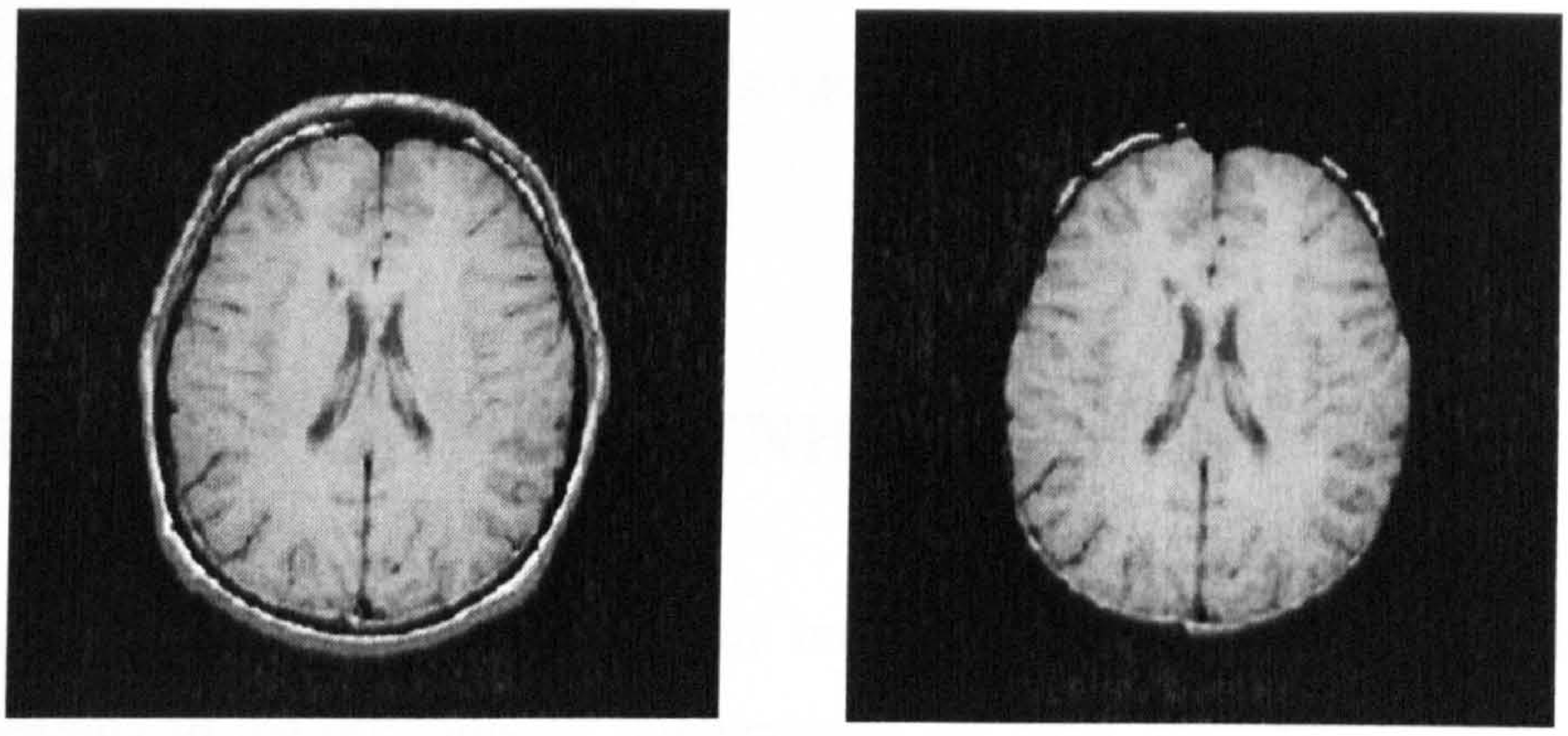
(a) original T_2-weighted MR image (b) brain region only

Figure 5.14 Non-brain region removal by the supplementary method.



(a) binary mask (b) labeled image after erosion

Figure 5.15 Removal by morphological processing.

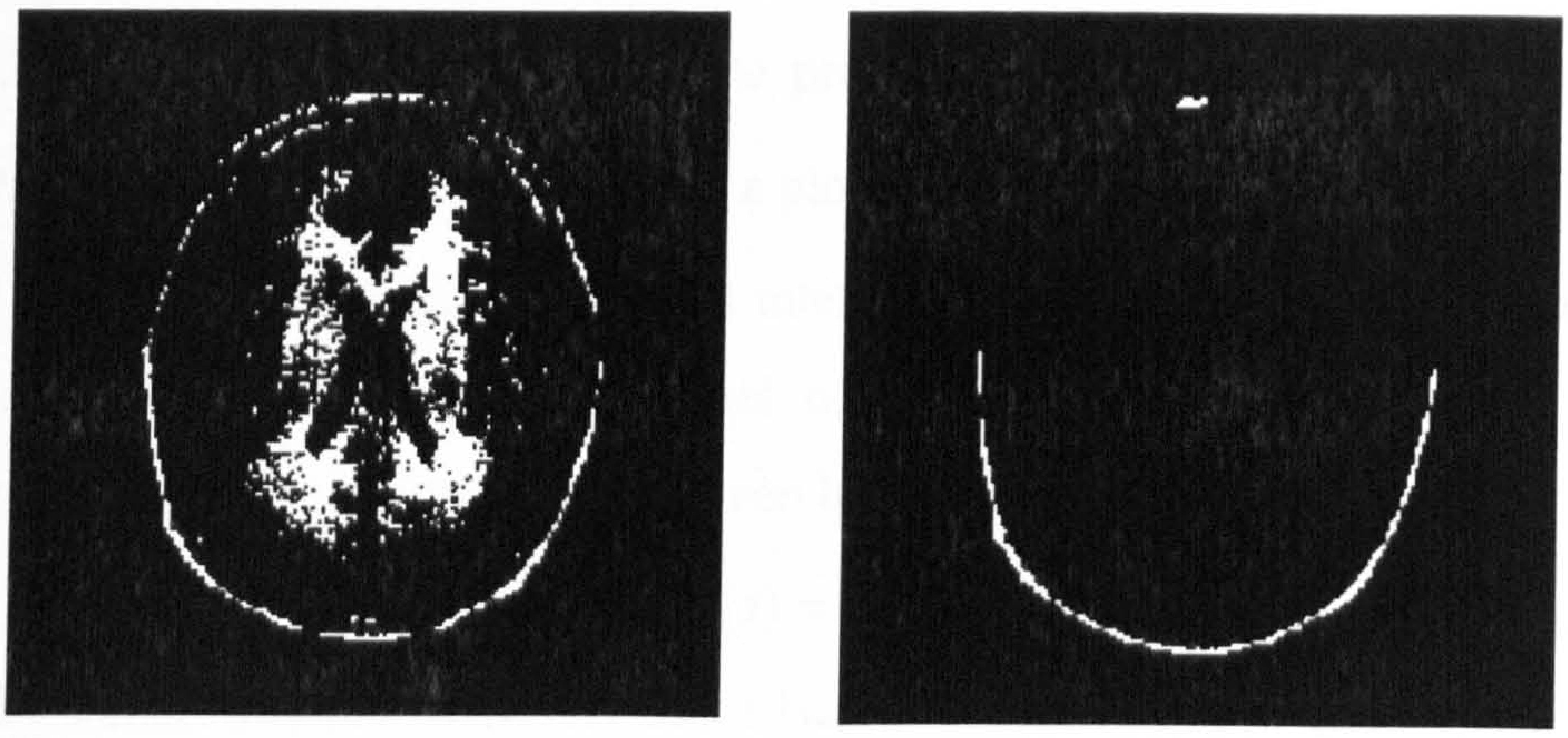


(a) original T_1-weighted MR image (b) brain region only

Figure 5.16 Non-brain region removal by the supplementary method.

An example is given in Fig.5.14 to demonstrate the supplementary method. It shows a standardized T_2-weighted MR image and its brain only image obtained through the supplementary removal method. After the removal, all background pixels are set to zero. Fig.5.15 shows the morphological processing applied to the same T_2-weighted MR image whose non-brain regions could not be labeled completely.

Beside T_2-weighted MR images, the supplementary method can be applied to other images with thin or non-uniformed non-brain regions. Fig 5.16 shows a T_1-weighted MR image and the result of its removal using the supplementary method. Since the anterior part of the non-brain region has lower intensities than its posterior part, it cannot be removed using the morphological processing. Fig.5.17 shows the failed removal using the morphological processing. This supplementary



(a) binary mask (b) labeled image after erosion

Figure 5.17 Removal by morphological processing.

method could be applied to the most images whose non-brain regions cannot be labeled as a whole.

5.5 TISSUE CONTRAST ENHANCEMENT

The intensities of white matter and gray matter in MR images are usually quite similar leading to the difficulty of segmentation for most brain MR images. In this section, histogram equalization is applied to enhance the tissue contrasts particularly between white matter and gray matter.

5.5.1 Histogram Equalization

A histogram is a special type of graph that shows the number of pixels with specific intensity values in an image. It provides a description of the intensity distribution of the image [Gonzalez and Wintz, 1977]. If r represents the specific intensity value of pixels in the image, then the histogram of an image can be expressed in the following mathematical term:

$$p_r(r_k) = \frac{N_k}{N} \quad (k = 0,1,...,L-1) \tag{5.12}$$

Where, L is the number of intensity values in the image, N_k is the number of pixels with the k-th intensity value in the image, and N is the total number of pixels in the image. $p_r(r_k)$, therefore describes the probability of that the k-th intensity value appears in the image. The histogram is a plot of $p_r(r_k)$ against r_k.

It is assumed that the transformed intensity value is s and its density function is $p_s(s)$ in the desired image. The target of histogram equalization is to obtain a uniformed density function which is given by

$$p_s(s) = 1/L \tag{5.13}$$

The transformation between r and s has the form as following:

$$s = T(r) \tag{5.14}$$

To obtain the uniformed density as in Eq. (5.13), the discrete form of the transformation can be described as:

$$\begin{aligned} s_k &= T(r_k) = \sum_{j=0}^{k} \frac{N_j}{N} \\ &= \sum_{j=0}^{k} p_r(r_j) \quad (k = 0,1,\ldots L-1) \end{aligned} \tag{5.15}$$

Ideally, the equalized histogram is expected to be perfectly flat. Although this is impossible in practice, considerable improvement over the original image can still be achieved.

The histograms appeared in this thesis are not plotted by $p_r(r_k)$. They are composed of N_k, the number of pixels with specific intensity values, not the probabilities. Figure 5.18 shows an ideal transformation of histogram equalization.

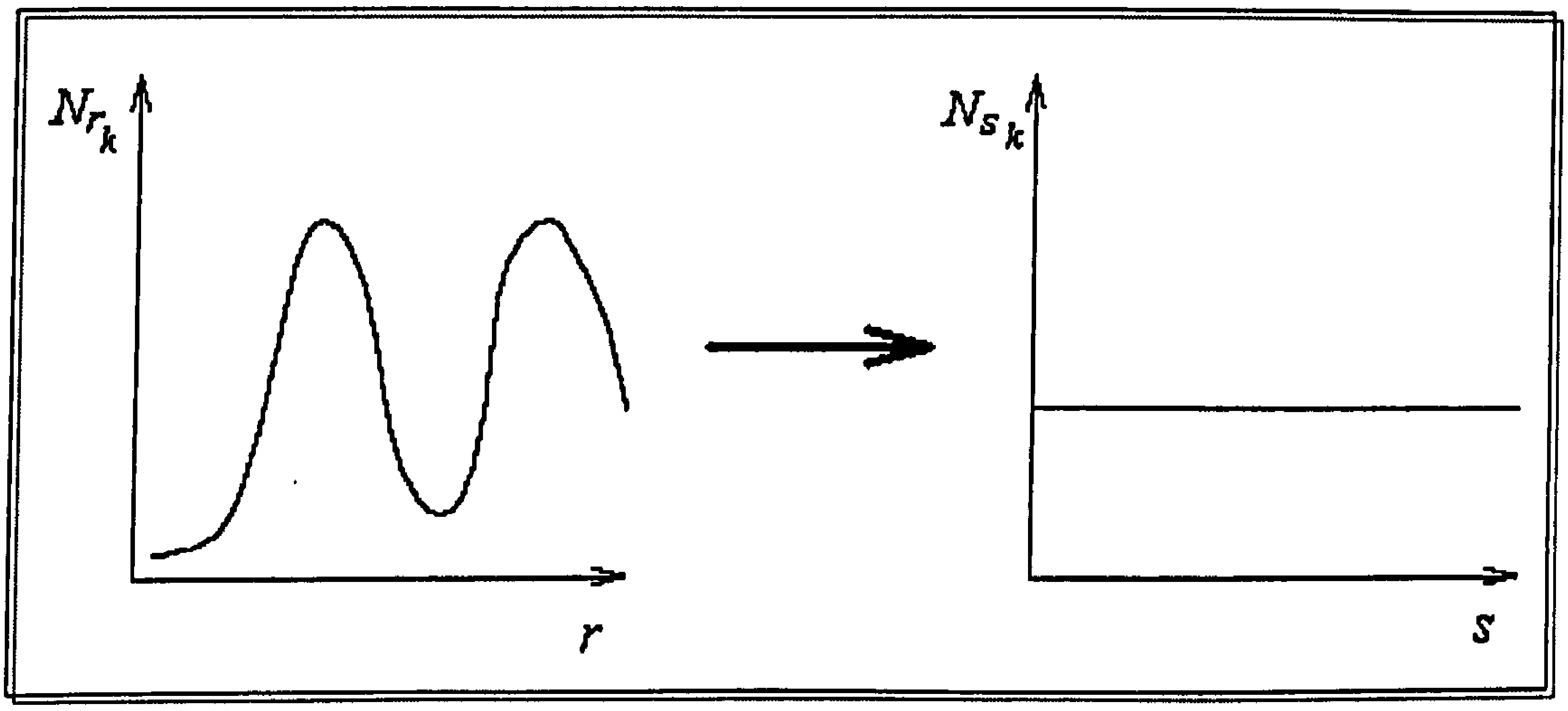


Figure 5.18 Histogram equalization (left) original histogram (right) equalized histogram.

5.5.2 Tissue Contrast Enhancement in MR Images

In most brain MR images, the intensities of white matter and gray matter are very similar. This leads to the difficulty of segmentation. Therefore, a tissue contrast enhancement process was developed in this thesis.

In most types of MR images, the background pixels have the lowest intensities. The CSF has almost the lowest intensities in T_1-weighted and PD-weighted MR images, but the highest intensities in T_2-weighted MR images. Brain tumours usually have the highest intensities in T_1-weighted and T_2-weighted images. Whatever the MR image types, the intensities of white matter and gray matter are in the middle of the intensity range. In this thesis, to maintain the basic features of the images, only the intensity range including white matter and gray matter was enhanced using histogram equalization which did not affect other tissue intensities. The intensity range is specified as follows:

$$Threshold - \sigma_{s'} \le x_j \le Threshold + \sigma_{s'} \qquad (5.16)$$

where x_j is the pixel intensity in the image. Threshold refers to the same value for obtaining the binary brain mask during the removal of non-brain regions. $\sigma_{s'}$ is the

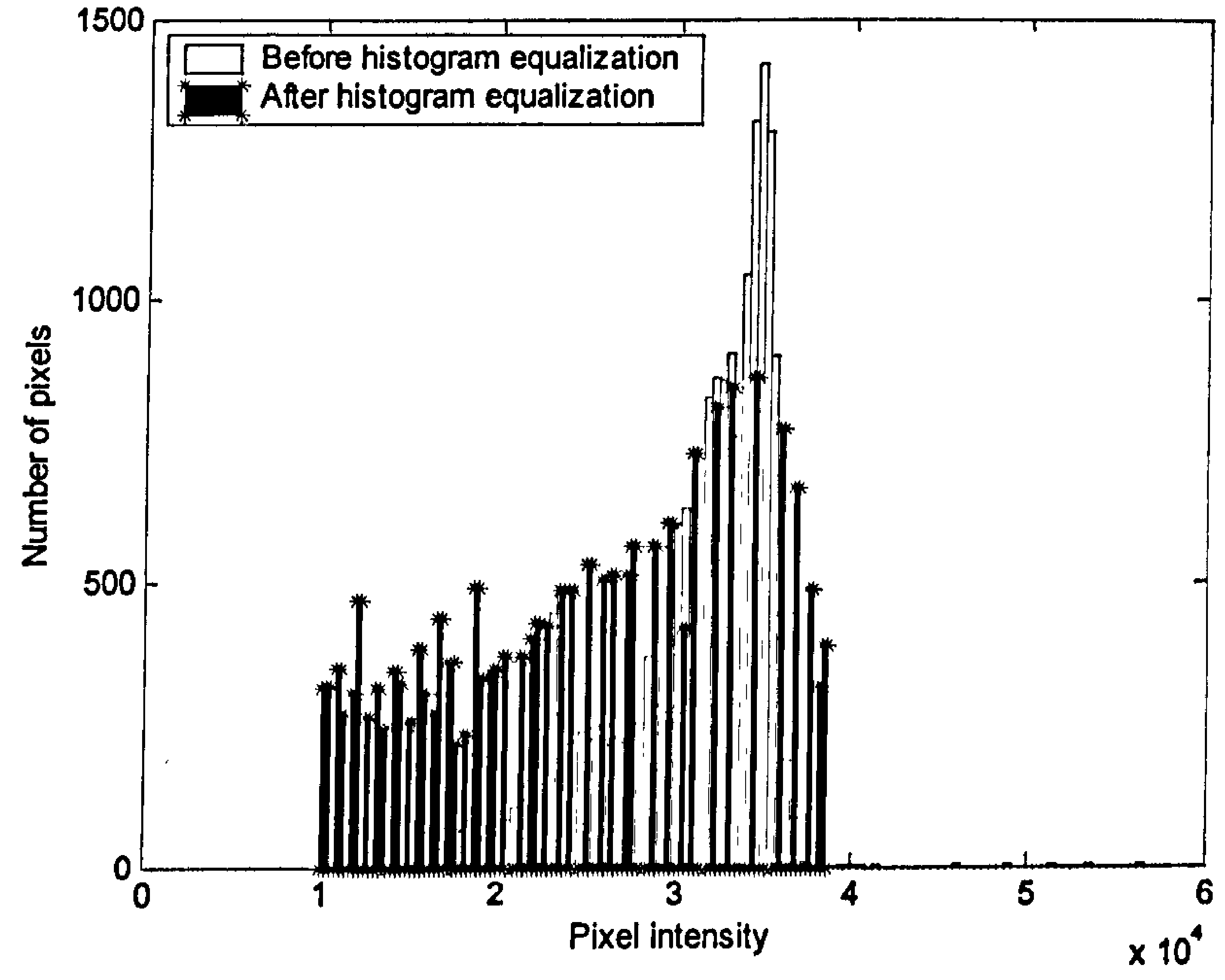
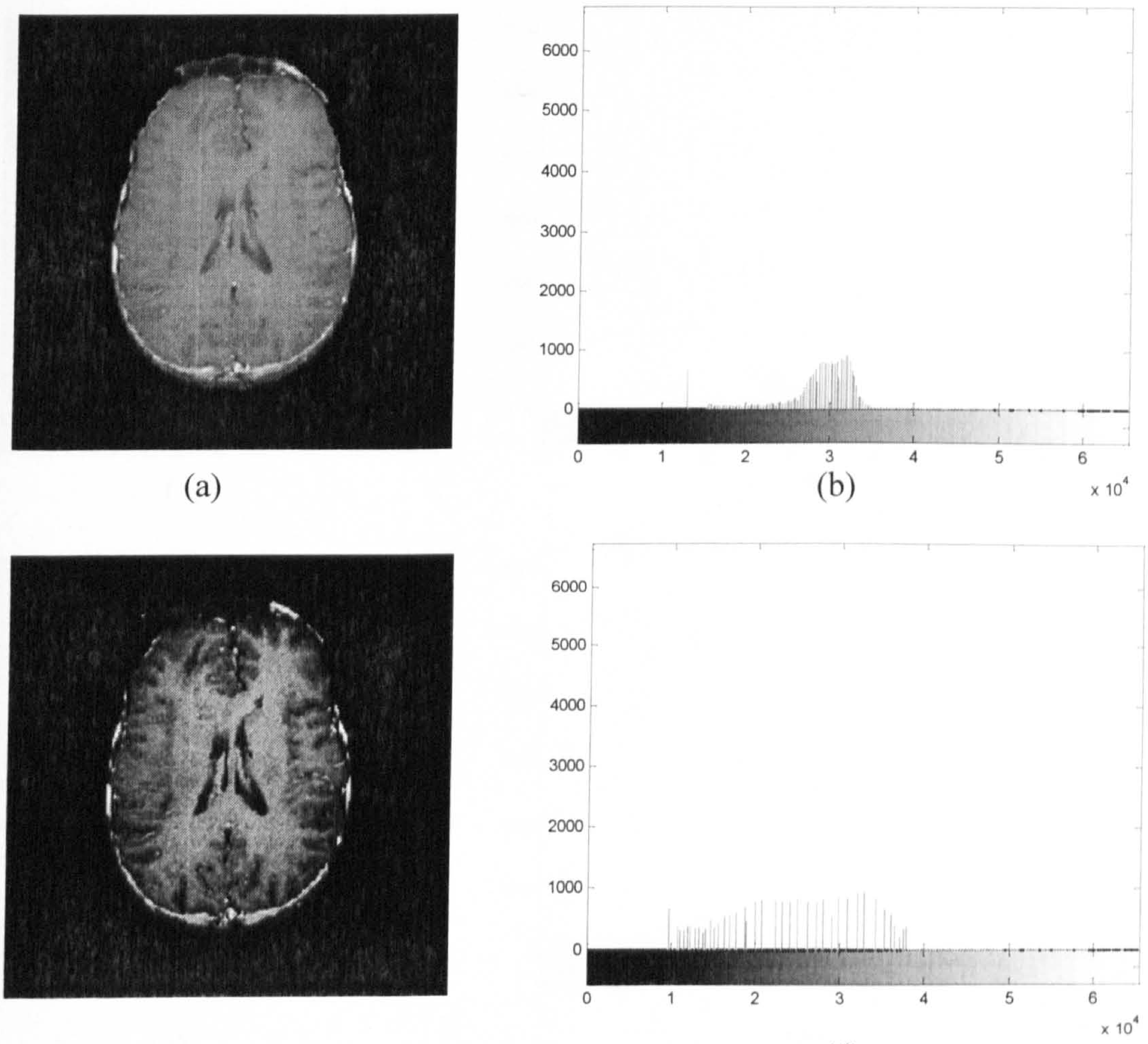


Figure 5.19 Before and after histogram equalization.

standard deviation of the standardized image.

Fig.5.19 shows a histogram of foreground pixel intensities from a standardized T_1-weighted MR image after the removal of the non-brain regions, compared with the new histogram after applying histogram equalization to enhance the contrast between white matter and gray matter.

Two images are given in Fig.5.20 and Fig.5.21 to demonstrate the enhanced results. Fig.5.20(a) shows a processed T_1-weighted MR image and its histogram. This image has already been standardized and the non-brain regions have been removed. The intensity range selected, [12547 41709], was equalizated using Eq.



(a) (b)

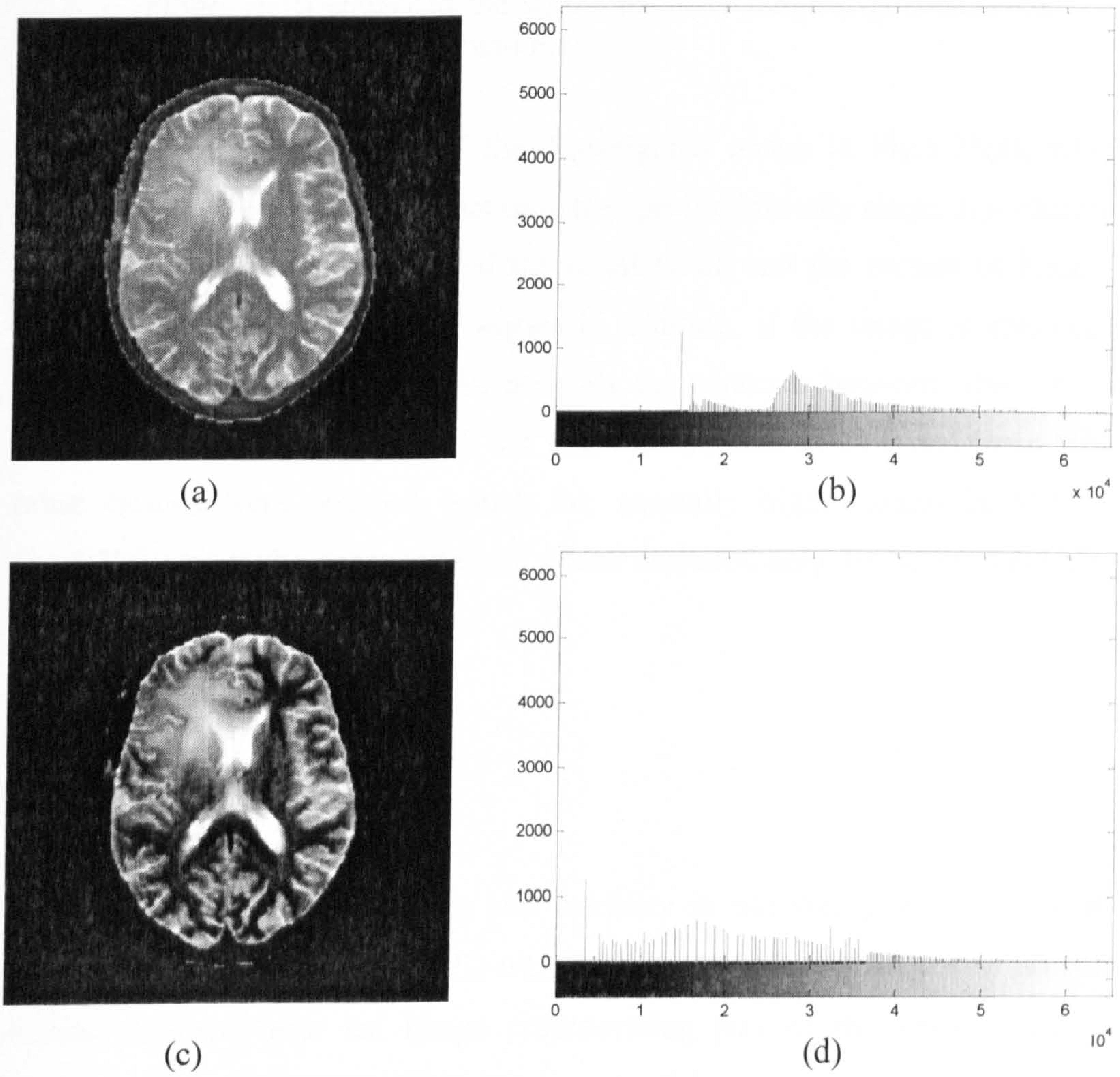
(c) (d)

Figure 5.20 Contrast enhancement of T_1-weighted MR image.
(a) original T_1-weighted image (b) original histogram (c) enhanced image (d) equalized histogram

(5.17) and other intensities were not changed. The enhanced image is shown in Fig.5.20(c), in which the contrasts between white matter and gray matter are greatly improved. Fig.5.20(d) gives the equalized histogram.

Fig.5.21 shows a T_2-weighted MR image after the standardization and the removal of non-brain regions. It was equalized in the range of [3630, 36533]. Both examples demonstrated that the histogram equalization enhaced the contrast between white matter and gray matter in MR images.

Assume the image is enhanced using the histogram equalization in the full range of intensites, the result will be dominated by the background pixels. Fig.5.22(a)



(a) (b)

(c) (d)

Figure 5.21 Contrast enhancement of T_2-weighted MR image.
(a) original T_2-weighted image (b) original histogram (c) enhanced image (d) equalized histogram

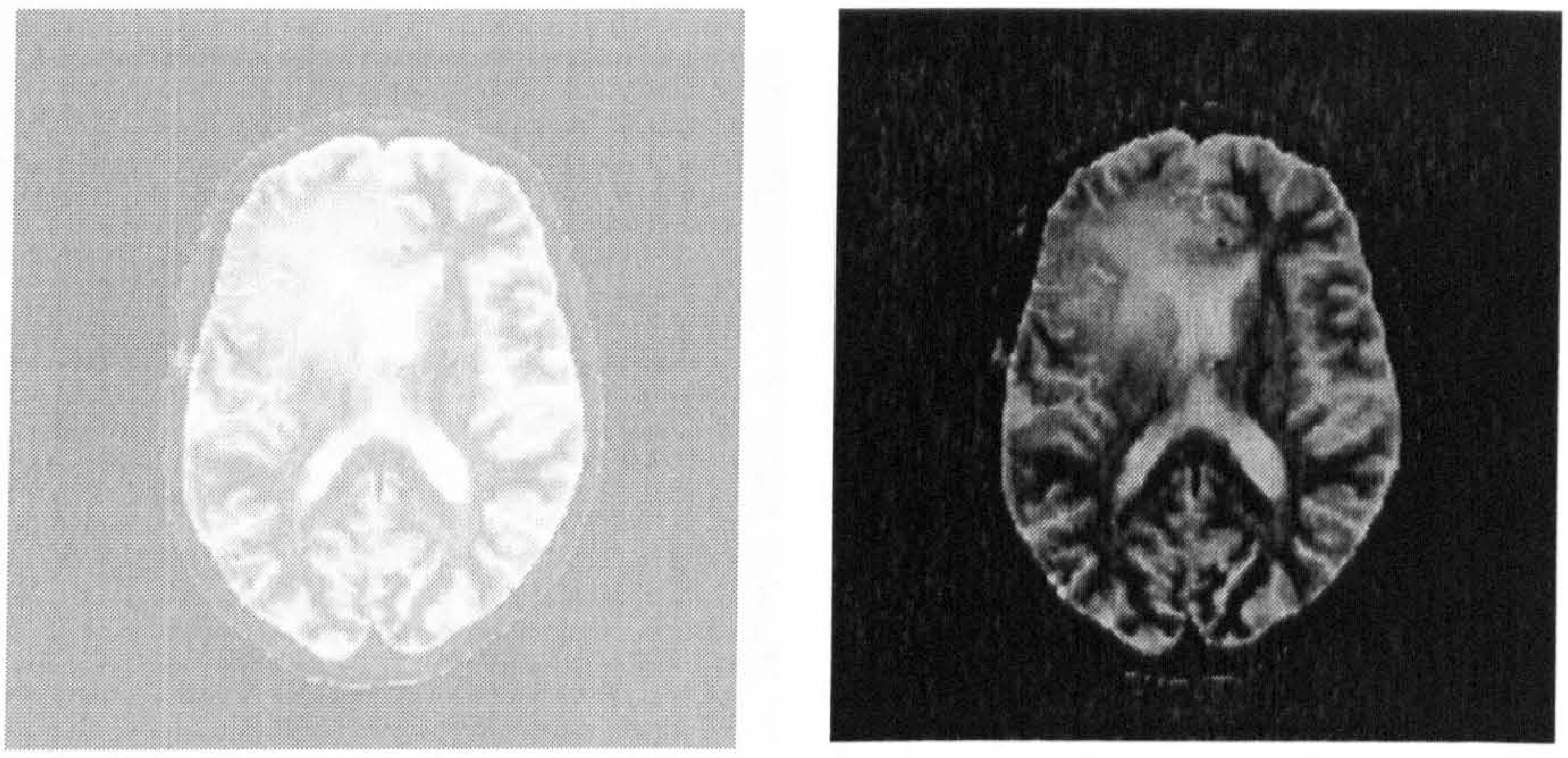
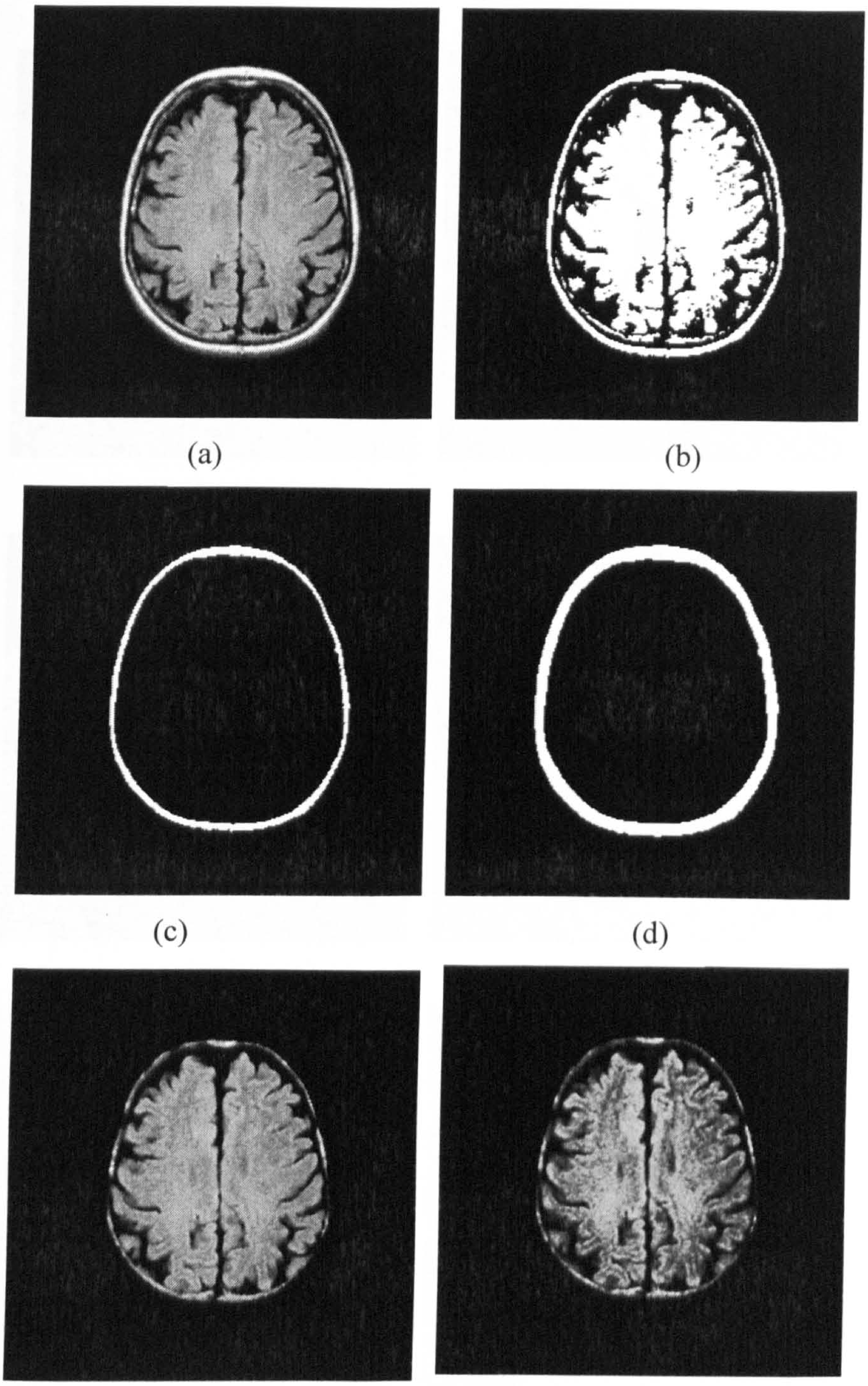


Figure 5.22 Contrast enhancement without specifying an intensity range. (left) enhanced the whole intensity range (right) enhanced only excluding background pixels

shows the enhanced results of the T_2-weighted image in Fig.5.21(a), which has equalized the whole histogram not only the specific intensity range. It is clear that the background intensities dominated the equalization and the contast of brain tissues had not improved but became worse. In addition, if the image is enhanced only excluding the background pixels, not only the contrasts between white mattar and gray matter cannot be enhanced, but also the contrasts of CSF and brain tumour to other tissues were reduced, which are normally high enough in MR images. Fig.5.22(b) gives the enhanced result which excluded only the background pixels for enhancement.

5.6 RESULTS

A typical MR image from each MR modality in our study, such as FLAIR-FSE, T_1-weighted, T_2-weighted and PD-weighted is shown in this section to illustrate the results obtained from the image preprocessing part of the whole brain tumour diagnosis system (Fig.5.23-5.26). Since original images from the hospital are unable to display uniformly, only the standardized images are shown.



(a) (b)

(c) (d)

(e) (f)

Figure 5.23 Illustration of image preprocessing on FLAIR-FSE MR image. (a) standardized image (b) binary mask (c) non-brain region after erosion (d) non-brain region after dilation (e) brain region only (f) brain region after contrast enhancement

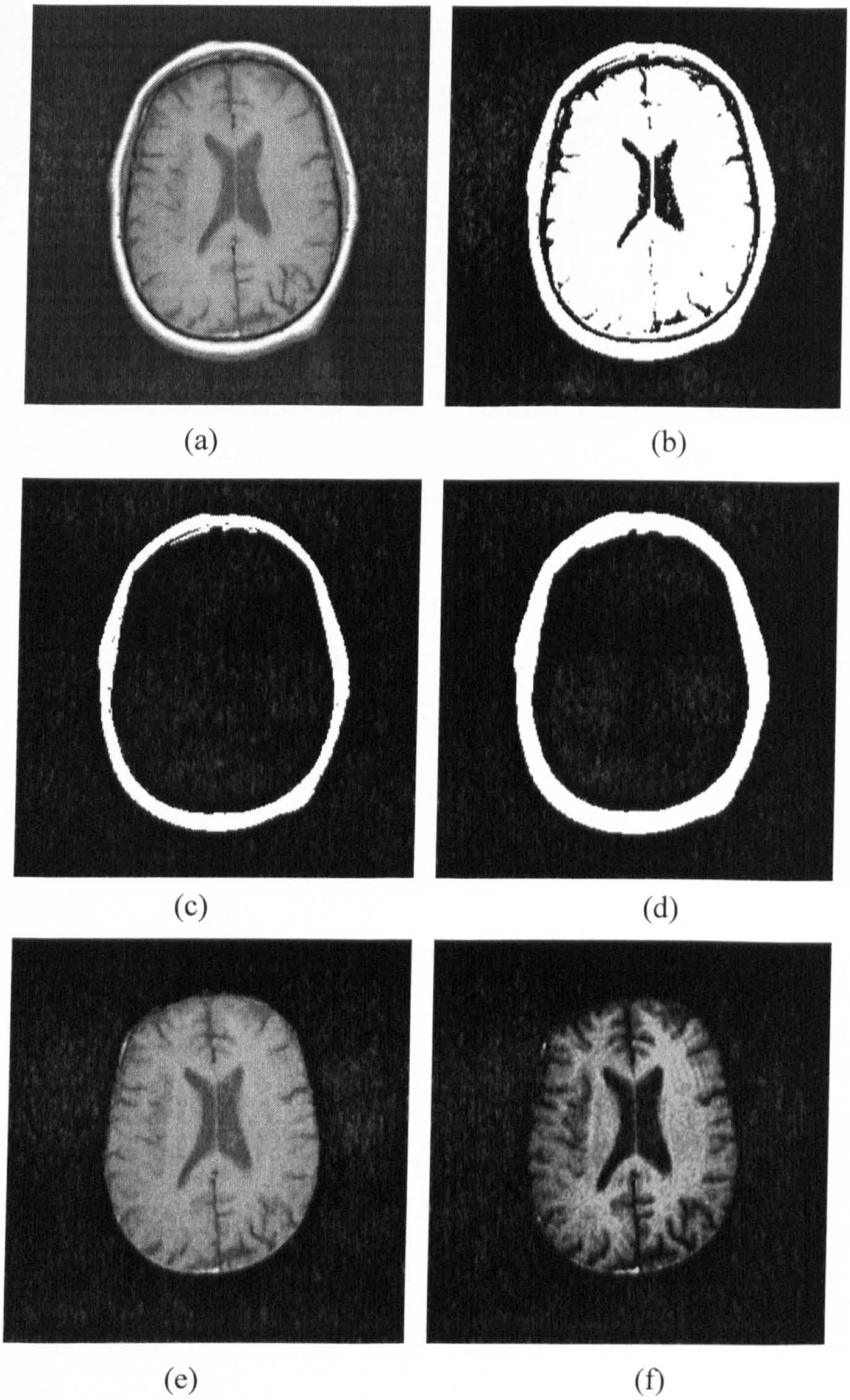
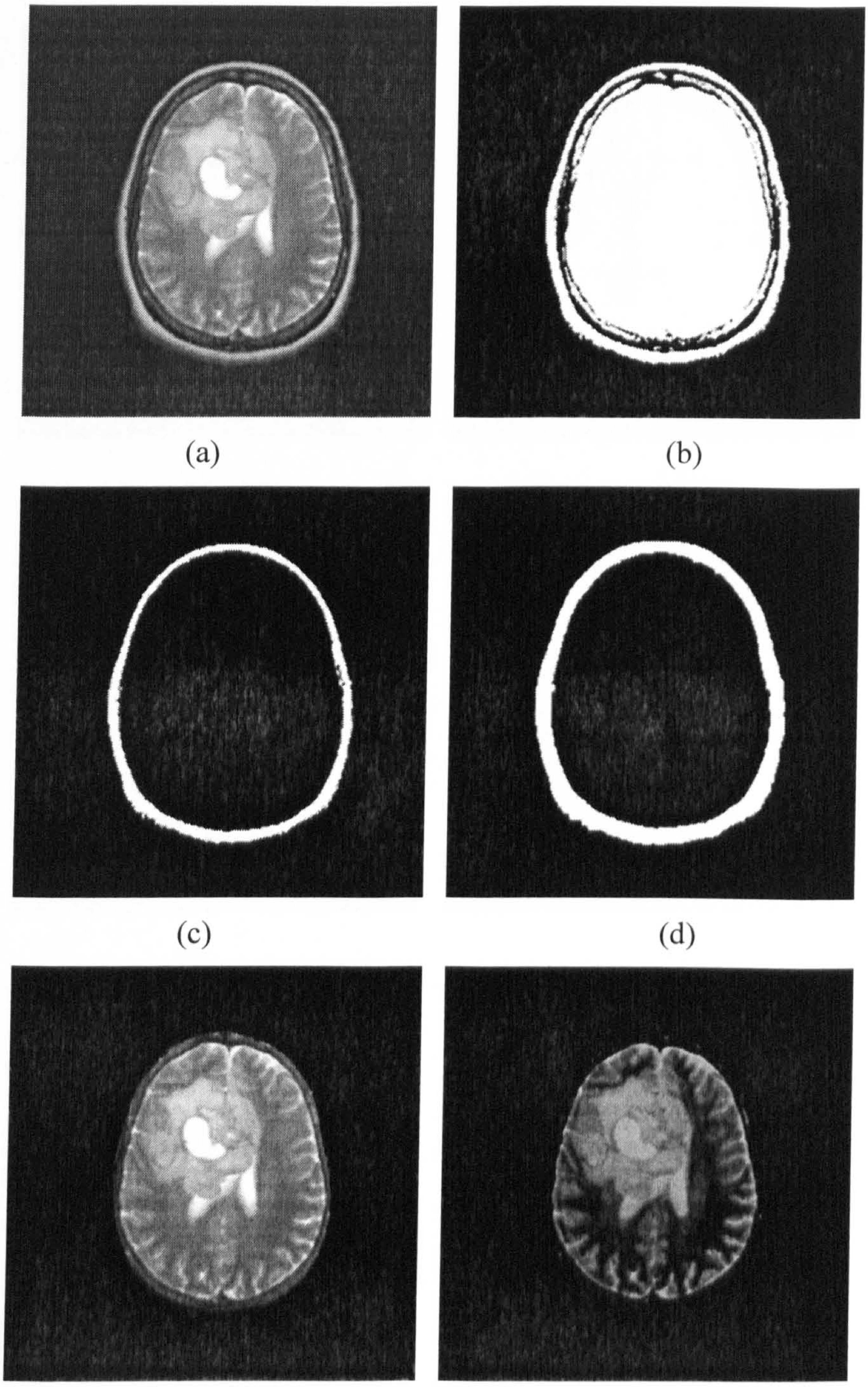


Figure 5.24 Illustration of image preprocessing on T_1-weighted image. (a) standardized image (b) binary mask (c) non-brain region after erosion (d) non-brain region after dilation (e) brain region only (f) brain region after contrast enhancement



(a) (b)

(c) (d)

(e) (f)

Figure 5.25 Illustration of image preprocessing on T_2-weighted image. (a) standardized image (b) binary mask (c) non-brain region after erosion (d) non-brain region after dilation (e) brain region only (f) brain region after contrast enhancement

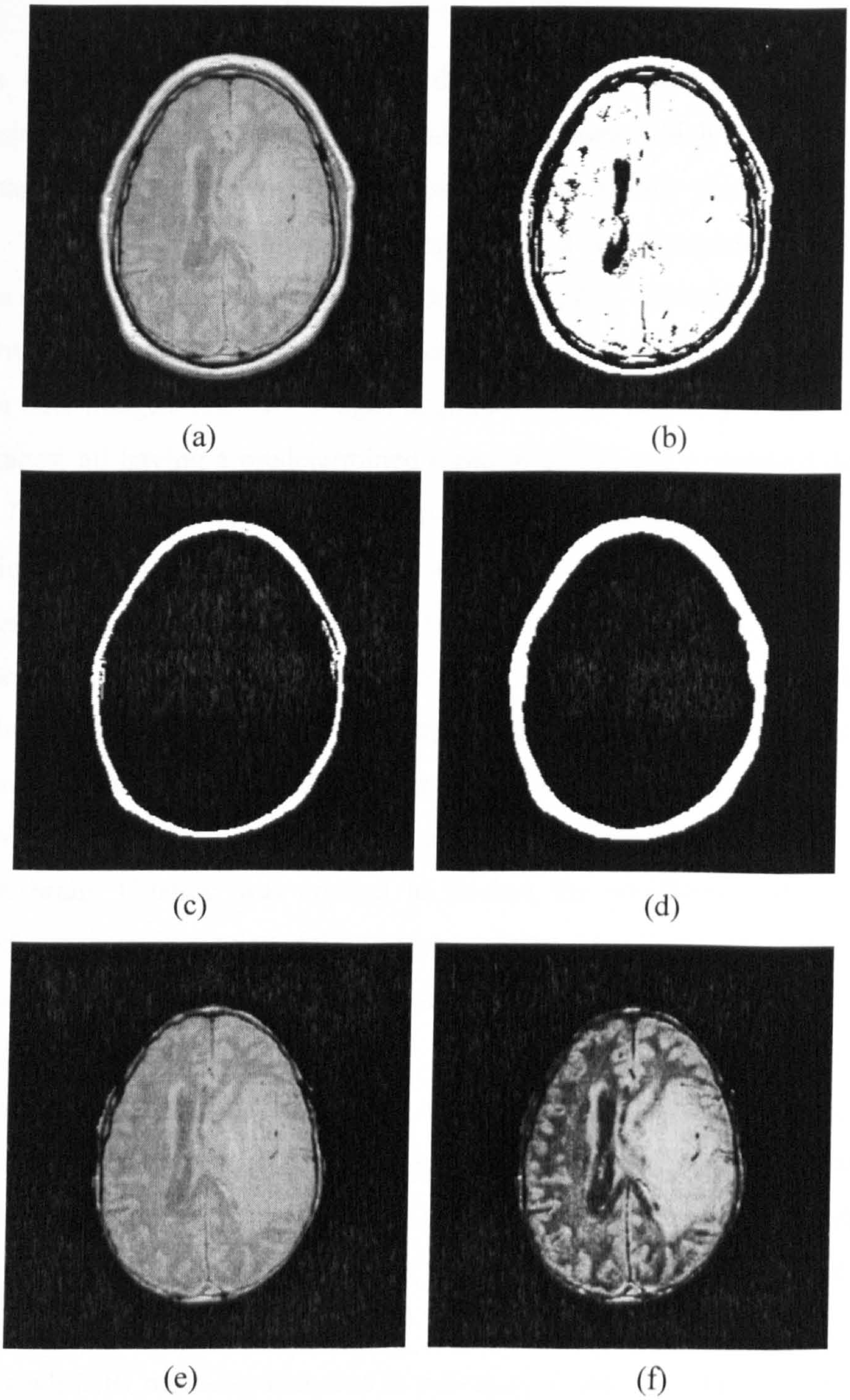


Figure 5.26 Illustration of image preprocessing on PD-weighted image. (a) standardized image (b) binary mask (c) non-brain region after erosion (d) non-brain region after dilation (e) brain region only (f) brain region after contrast enhancement

5.7 CONCLUSIONS

In this chapter, image preprocessing, the first major part in the brain tumour diagnosis system was presented. It included three steps which enabled the original MR images from the hospital to be prepared for the further processing used in the system. The first step was image standardization; it was completed using statistical analysis depending on the statistical features of the pixel intensities in MR images. A sub-normal distribution was specifically defined to describe the intensity distribution of brain MR images. All MR images acquired for this thesis, were transformed to new images, all having a predetermined mean of 32500 and a standard deviation of 10000. These specific values of mean and standard deviation cannot only scale the intensities of MR images into similar distribution shapes but also unify the images displayed as unsigned 16-bit format with full resolution.

The second step was non-brain removal by means of morphological processing. A threshold corresponding to the intensity value having the most number of pixels was chosen to generate the binary image from the standardized image. The binary image was then eroded until the non-brain regions could be labeled and separated from the brain. Dilation was applied to thicken the non-brain regions that were thinned down through erosion, followed by masking the whole non-brain regions out of the standardized MR image. The occasional examples which were failed analysed by a supplementary method proposed.

Finally, the contrast between white matter and gray matter, in which there are always have difficulties in segmentation due to their similar intensities, was enhanced using histogram equalization. Only the intensities lying in a specific range referring to white matter and gray matter were treated without changing other pixels in the image.

For each MRI modality included in our study a typical image was presented to demonstrate the effectiveness of the processes stated above.

Chapter 6

MR IMAGE SEGMENTATION

6.1 INTRODUCTION

Image segmentation is a procedure that spatially partitions an image into a certain number of homogeneous regions or isolates specific objects in an image [Robb, 2000; Kurz and Benteftifa, 1997]. These homogeneous regions are determined by some criteria, such as grey level, color and texture. The segmentation results are various depending on the features of the image.

Image segmentation is a crucial step in image processing and has been applied in many fields. In aerial photography recognition, its objective is to exactly identify the objects in the scene. In medicine, it is used to distinguish different tissues so that a specific measurement can be done automatically [Li and Gray, 2000]. In industry, it is used to recognize the appropriate components for assembly systems on which the quality of the final output depends largely on the accuracy of segmentation. Inaccurate results can be useless, even damaging.

A variety of segmentation methods have been proposed in the literature. The principal methods include thresholding, region growing, edge-based segmentation and clustering. Each of them has its own features. The description of segmentation in this chapter specifically refers to grey level images.

6.1.1 Thresholding

Thresholding is one of the old, simple and popular techniques for image segmentation. Initially, a threshold is chosen based on image histogram or co-occurrence matrix [Pal and Pal, 1993] to separate the image into two sub-regions. Two thresholds are then chosen for these two sub-regions. It is an iterative process, in which a specific threshold is always determined for each sub-region. The appropriateness of the threshold is a decisive factor to the success of segmentation. If the region of interest has obviously different grey levels from other regions in the image, it is easy to find the threshold just by adjusting its maximum and minimum and displaying the thresholded image iteratively. However, in most cases, the distribution of grey level is complicated, leading to difficulty in threshold determination. Cheriet *et al.* (1998) proposed a thresholding technique which divides the pixels into the region of interest and background iteratively. It continuously separates the largest peak in the grey level histogram until there are no clear peaks left. Nevertheless, in some situations, thresholding is still a difficult method because the histogram may have only one peak, which is called unimodal. This is due to the large number of background pixels and often happens in medical images.

Many applications of thresholding on MR image segmentation have been developed by researchers all over the world. Suzuki *et al.* (1991) used iterative thresholding to distinguish brain tissues in axial MR slices. The thresholds for the brain are iteratively adjusted based on the geometry of the resulting mask. Lemieux *et al.* (1999) presented a three-dimensional technique which used both automated thresholding and morphological operations. It calculated the thresholds of different tissues based on the histogram of the whole image and combined the morphological operations to disconnect the segmented regions. Since the distribution of tissue intensities in brain images is complicated, this leads to difficulty in threshold determination and so restricts the application of thresholding.

6.1.2 Region Growing

Region growing is an extended method of thresholding. The connectivity condition (four or eight connected pixels) is combined with thresholding to perform the segmentation [Robb, 2000]. The number of regions and the location of a single pixel or several pixels in each region should be known initially. These pixels are called seeds and they are decided by the users manually or by computer-based methods using some criteria. Starting from the seeds for each region, the pixels are classified into this region if they satisfy the thresholding and connectivity condition. The number of seeds increases until no connected pixels satisfy the thresholding condition. The success of region growing method is decided by seed selection. If suitable seeds are chosen, segmentation results will be much better; and if seeds are unsuitable, segmentation results will be poor. Generally, seeds can be the brightest pixels or can be found from the peaks of histogram. Region growing performs segmentation using seeds as well as thresholding. It is therefore more accurate than the thresholding method. However, a shortcoming of region growing is that it requires rather long computation times. Also segmentation methods in region growing category are all sensitive to the initial seeds [Gonzalez and Wintz, 1977]. Inappropriate seeds can lead to the complete failure.

Using the region growing method on medical images requires precise anatomical information to choose the location of seeds, for each region and the region homogeneity. Pohle and Toennies (2001) presented a region growing algorithm based on a model of regions and their homogeneity, assuming that the deviation of the grey levels within regions is smaller than that between regions and learns its homogeneity criterion. The reliability of the method largely depends on the model's assumption on region homogeneity which is not robust to the noisy environment. Law and Heng (2000) developed a segmentation method iteratively incrementing the deviation of the regions to search for the optimal thresholds until the number of pixels in the regions suddenly increased. However, none of these methods can overcome the disadvantages of region growing itself.

6.1.3 Edge-based Segmentation

Edge-based segmentation represents a large group of methods based on information about edges in the image. It relies on edge detecting operators, such as Roberts, Prewitt, Sobel and Laplacian, to extract the edge by initially identifying the grey level discontinuity in an image. However, the image resulting from edge detection cannot be used as a segmentation result. Supplementary steps must follow to combine edges into edge chains. If the edge chains are close, they may correspond to the borders in the image. The segmentation can finally be achieved by partitioning the region enclosed by these edge chains. A large number of edge-based segmentation methods are available nowadays which differ in strategies which leading to final border construction, and also differ in the amount of prior information that can be incorporated into these methods. The most common problems of all edge-based segmentation methods, caused by image noise or unsuitable information in an image, are an edge presence in locations where there is no border, and no edge presence where a real border exists [Zhang, 1997].

Edge-based segmentation is a common approach for medical image segmentation. Many researchers are working towards more robust methods. Gibbs *et al.* (1996) developed a method combining morphological edge detection and region growing to complete MR image segmentation. Another method using fuzzy object extraction and edge detection applied on MR images was proposed by [Lin, Tian and He, 2002]. Although these methods are quite promising, they greatly depend on methods in other categories and still cannot avoid the disadvantages that exist in most edge-based segmentation methods.

6.1.4 Clustering

Clustering is an unsupervised technique which automatically partitions the data into meaningful subgroups. It performs the same function as classification without the training data. A cluster in the image means a region composed of pixels with similar features. Clustering methods are based on the minimization of the variance between the pixel feature and the so-called cluster means. Feature similarity is thought of as similar grey level values in grey level images. The clustering algorithm partitions an image into a number of clusters to complete the segmentation. The pixels in the same cluster have similar grey level values and those in different clusters have dissimilar grey level values. Clustering methods can be generally categorized into three types, deterministic, fuzzy and statistical methods. The representative algorithms of these three classes are *k*-means clustering, fuzzy *c*-means (FCM) clustering, and expectation-maximization (EM) algorithm [Cannon *et al.*, 1986]. The *k*-means clustering algorithm clusters data by iteratively computing a centre of grey level for each cluster and segments the image by partitioning each pixel into the cluster with the closet centre. FCM is the fuzzy equivalent of the *k*-means clustering algorithm, which developed on fuzzy set theory. It takes into account the overlapping of the clusters and allows partial belongingness of the pixels to all the clusters [Kettaf *et al.*, 1996]. Hence, the FCM algorithm is superior to the *k*-means algorithm for handling images with uncertainties, such as medical images. EM algorithm assumes that data follows a Gaussian mixture model. It considers the clusters as missing information and computes the maximum likelihood estimates of the mean, covariance, and mixing coefficients of the mixture model [Coleman and Andrews, 1979; Power, 2000].

Clustering is the most popular category for medical image segmentation, especially FCM and EM. Wells *et al.* (1996) reported a new segmentation method called adaptive segmentation. It applies the EM algorithm on intensity inhomogeneities of different tissues to segment brain images. Leemput *et al.* (1999) also described an automated model-based segmentation method for brain MR images based on an iterative EM procedure that interleaves tissue classification with the estimation of tissue-class-specific intensity models. The common disadvantage of the EM algorithm is that the intensity distribution of brain images is modeled as a normal distribution which is not usually true especially in noisy images. This considerably limits the application of the EM algorithm. The FCM algorithm is also employed by many researchers on brain MR images. Li *et al.* (1993) presented a knowledge-based classification and tissue labeling approach to initially segment MR brain using the FCM algorithm. This was followed by an expert system to locate a landmark tissue by matching it with a prior model. Hall *et al.* (1992) compared the FCM algorithm with neural networks in segmenting brain MR images and demonstrated FCM performed better on normal brains, but worse on abnormal brains with oedema or tumour. Pham and Prince (1999) extended the traditional FCM algorithm to deal with MR images corrupted by intensity inhomogeneities. However, the greatest shortcoming of FCM is its high sensitivity to noise which is also a flaw of many other intensity based segmentation methods.

6.2 FUZZY C-MEANS CLUSTERING ALGORITHM

The traditional FCM algorithm introduced by Bezdek is an improvement of earlier clustering methods [Bezdek, 1981]. It is based on minimizing an objective function, with respect to fuzzy membership *U*, and a set of cluster centroids *V*:

$$J_m(U,V)=\sum_{j=1}^{N}\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}d^2(\mathbf{x}_j,\mathbf{v}_i) \tag{6.1}$$

In the above equation, $X=\{\mathbf{x}_1,\mathbf{x}_2,\ldots\mathbf{x}_j\ldots\mathbf{x}_N\}$ is a $p\times N$ data matrix, where p represents the dimension of each $\mathbf{x}_i$ "feature" vector (grey level, colors etc in images), and N represents the number of feature vectors (pixel numbers in images). C is the number of clusters. $\mathbf{u}_{ij}\subseteq U\ (p\times N\times C)$, is the membership function of vector $\mathbf{x}_j$ to the i-th cluster, which satisfies: $\mathbf{u}_{ij}\in[0,1]$ and $\sum_{i=1}^{C}\mathbf{u}_{ij}=1$, ($j$=1, 2,...$N$). The higher the membership, the higher the probability that the pixel belongs to the cluster. The membership function is expressed as follows:

$$\mathbf{u}_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d(\mathbf{x}_j,\mathbf{v}_i)}{d(\mathbf{x}_j,\mathbf{v}_k)}\right)^{\frac{2}{m-1}}} \tag{6.2}$$

$V=\{\mathbf{v}_1,\mathbf{v}_2,\ldots\mathbf{v}_i\ldots\mathbf{v}_C\}$, which is a $p \times C$ matrix, and denotes the cluster feature centre:

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}(\mathbf{u}_{ij})^m \mathbf{x}_j}{\sum_{j=1}^{N}(\mathbf{u}_{ij})^m} \qquad (i = 1,2\ldots C) \tag{6.3}$$

$m \in (1,\infty)$ is a weighting exponent on each fuzzy membership, which controls the degree of fuzziness. If m=1, clustering becomes hard. The fuzziness of clustering increases with the value of m. $d^2(\mathbf{x}_j,\mathbf{v}_i)$ is a measure of similarity between $\mathbf{x}_j$ and $\mathbf{v}_i$:

$$d^2(\mathbf{x}_j,\mathbf{v}_i) = \left\|\mathbf{x}_j - \mathbf{v}_i\right\|^2 \tag{6.4}$$

$\|\cdot\|$ can be defined as either a straightforward Euclidean distance or its generalization such as the Mahalanobis distance [Taguchi, Chowdury and Wu, 2002].

The feature vector X in the FCM algorithm represents grey levels of pixels because of our interest in grey level images. So, the feature vector only has one dimension where p is equal to one. The membership U and the cluster centre V are both reduced to $N \times C$.

The FCM algorithm iteratively optimizes $J_m(U,V)$ with the continuous update of $U^{(l)}$ and $V^{(l)}$, where l is the number of iterations. The basic steps of this algorithm are described below:

(1) Determine the number of clusters C, $2 \le C \le N$ and m.

(2) Randomly initialize the fuzzy membership $\mathbf{u}_{ij}^{(0)}$ of $\mathbf{x}_j$ belonging to the i-th cluster.

(3) At the l-th iteration (l=0,1,2,...), calculate the cluster feature centre $\mathbf{v}_i^{(l)}$ for $i = 1,2\ldots C$.

(4) Calculate $d^2(\mathbf{x}_j,\mathbf{v}_i^{(l)})$.

(5) Update $\mathbf{u}_{ij}^{(l)}$ with $d^2(\mathbf{x}_j,\mathbf{v}_i^{(l)})$.

(6) Compare $\mathbf{u}_{ij}^{(l)}$ and $\mathbf{u}_{ij}^{(l-1)}$, if $\left\|\mathbf{u}_{ij}^{(l)} - \mathbf{u}_{ij}^{(l-1)}\right\| < \varepsilon$, then stop, otherwise, $l=l+1$, go to step (3) and repeat.

Finally, each vector has a membership to each cluster and it is allocated to the cluster for which it has the highest membership.

In image segmentation, the cluster number C refers to the number of sub-regions into which the image is segmented. The pixel most likely belongs to the sub-region in which it has the highest membership. Fig.6.1 shows an illustration of the FCM clustering algorithm. The red labeled $\mathbf{u}_{cj}$ refers to the fact that the pixel $\mathbf{x}_j$ has the highest membership to the c-th cluster; because $\mathbf{x}_j$ has the closest grey level to the c-th cluster centre. To clarify the illustration, the colours in Fig.6.1 represent the different grey levels.

The drawback of the FCM algorithm for image segmentation is very obvious. Firstly, it is clear from Eq. (6.1) that the objective function of FCM does not take into consideration any spatial dependence among X, but deals with images the same as

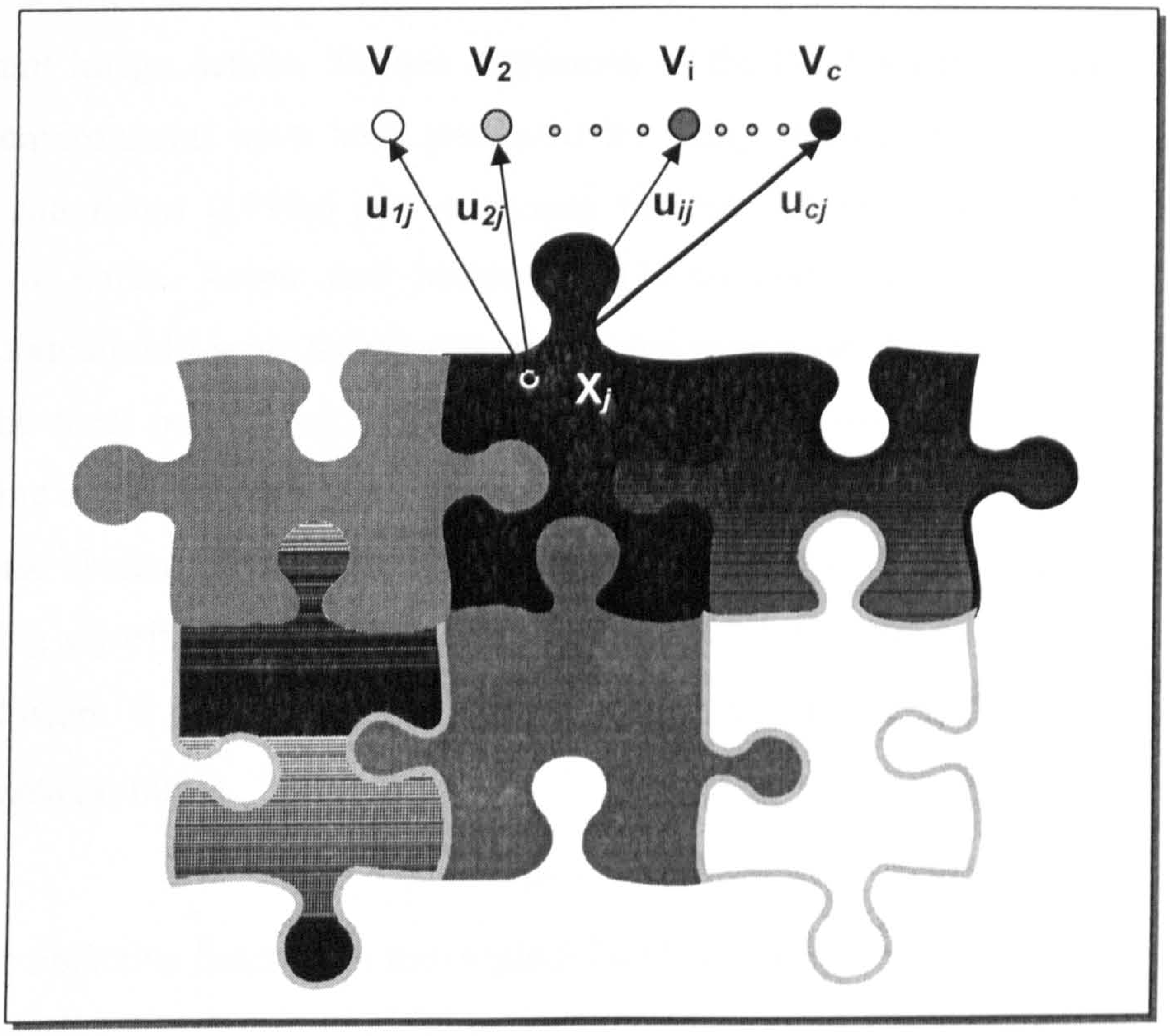


Figure 6.1 Illustration of FCM clustering algorithm.

separate points. Secondly, the membership function U (Eq. (6.2)) is mainly decided by $d^2(\mathbf{x}_j, \mathbf{v}_i)$, which measures the similarity between the pixel intensity and the cluster centre. Higher membership depends on closer intensity values to the cluster centre, thus indicating the sensitivity of the membership function to noise. Suppose an MR image contains noise or is affected by artifacts, the pixel intensities may be changed by the noise or artifacts. It can result in the incorrect membership and improper segmentation. The FCM algorithm must be improved in order to solve this problem.

6.3 RELATED EXTENSIONS TO FUZZY C-MEANS CLUSTERING ALGORITHMS

The most direct way to compensate for the drawback of FCM is to smooth the image before segmentation. However, standard smoothing filters can lead to a loss of important image details. Various extensions of the FCM algorithm considering the noisy environment have been presented by many researchers all over the world. Tolias and Panas (1998a) post-processed the membership function to smooth the effect of noise. Acton and Mukherjee (2000) and Tolias and Panas (1998b) incorporated multi-scale information to enforce spatial constraints.

The most popular approach for increasing the robustness of FCM to noise in clustering is to modify the objective function directly. Dåve (1991) proposed the idea of a noise cluster to deal with noisy clustering data in the approach known as nosing clustering algorithm (NC). The noise was clustered separately and excluded from the good clusters. It was represented by a prototype that has a constant distance, δ, from all the data points.

$$d^2(\mathbf{x}_j, \mathbf{v}_i) = \delta^2 \tag{6.5}$$

The objective function in the original FCM was modified in NC as follows:

$$J_m(U,V) = \sum_{j=1}^{N}\sum_{i=1}^{C} {}_{ij}^{m} d^2(\mathbf{x}_j, \mathbf{v}_i) + \sum_{j=1}^{N} \delta^2 \left(1 - \sum_{i=1}^{C} \mathbf{u}_{ij}\right)^m \tag{6.6}$$

where

$$\delta^2 = \lambda \left[\frac{\sum_{i=1}^{C-1}\sum_{j=1}^{N} d^2(\mathbf{x}_j,\mathbf{v}_i)}{N(C-1)} \right] \tag{6.7}$$

λ is the value of the multiplier used to obtain δ from the average of the distances. The C-th cluster refers to the noise.

The membership function is updated from Eq. (6.2) to:

$$\mathbf{u}_{ij} = \frac{1}{\sum_{k=1}^{C}\left[\frac{d^2(\mathbf{x}_j,\mathbf{v}_i)}{d^2(\mathbf{x}_j,\mathbf{v}_k)}\right]^{1/(m-1)} + \left[\frac{d^2(\mathbf{x}_j,\mathbf{v}_i)}{\delta^2}\right]^{1/(m-1)}} \tag{6.8}$$

However, the NC algorithm is not suitable for image segmentation since the noisy pixels should not be separated from other pixels, but are assigned to the most appropriate clusters while reducing the effect of noise. Each pixel can only belong to one cluster. The result of segmentation results using NC cannot reflect the complete image because noisy pixels are eliminated.

Another similar method delivered by Krishnapuram and Keller (1993) is called possibilistic *c*-means algorithm (PCM) which interprets the clustering as a possibilistic partition. It constructs a new objective function based on (6.1) as following:

$$J_m(U,V) = \sum_{j=1}^{N}\sum_{i=1}^{C} \mathbf{u}_{ij}^m d^2(\mathbf{x}_j,\mathbf{v}_i) + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} (1-\mathbf{u}_{ij})^m \tag{6.9}$$

where η_i are suitable positive numbers. The first term in Eq. (6.9) is equal to the objective function of FCM which requires that the distance from the feature vectors to the cluster centers be as low as possible. The second term forces the membership to be as high as possible without having the maximum limit of one. The value of η_i is defined as:

$$\eta_i = K \frac{\sum_{j=1}^{N} \mathbf{u}_{ij}^m d^2(\mathbf{x}_j,\mathbf{v}_j)}{\sum_{j=1}^{N} \mathbf{u}_{ij}^m} \tag{6.10}$$

where K is typically chosen to be 1. Differentiating Eq. (6.9) leads to the expression of the membership function in this approach as:

$$\mathbf{u}_{ij} = \frac{1}{1+\left(\dfrac{d^2(\mathbf{x}_j,\mathbf{v}_i)}{\eta_i}\right)^{1/(m-1)}} \tag{6.11}$$

Although the PCM algorithm can improve the robustness of clustering in a noisy environment, it is incapable of dealing more widely with image segmentation. Since the limit of $\sum_{i=1}^{C}\mathbf{u}_{ij} = 1$ is less strict in PCM, it is very likely that the pixels will be stuck in one cluster. This means the PCM algorithm needs specific requirements for raw data if clustering is to be successful. This therefore weakens its application.

The Robust Fuzzy C-means Algorithm (RFCM) was presented by Pham (2001). In this method, a new objective function was proposed for incorporating spatial context into FCM. It "includes a penalty term that is reminiscent of MRF priors but is consistent with the desired behaviour of the membership function..." [Pham, 2001]. The modified objective function is as follows:

$$J_{RFCM} = \sum_{j=1}^{N}\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}\left\|x_j - v_i\right\|^2 + \frac{\beta}{2}\sum_{j=1}^{N}\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}\sum_{l\in N_j}\sum_{k\in M_i}\mathbf{u}_{lk}^{m} \tag{6.12}$$

where N_j is the set of neighbours of pixel j, and $M_i=\{1,\ldots,C\}\backslash\{i\}$ representing a set of clusters except the i-th one. The parameter β controls the trade-off between the standard FCM objective function and the smooth membership functions. Its value depends on the brightness of the image as well as the variation of the intensity values relative to the centroid value of each class. The new penalty term is minimized when the membership value for a particular cluster is high and the membership values for the other clusters at neighbouring pixels are low, and vice versa.

The membership function in this approach is therefore expressed as follows:

$$\mathbf{u}_{ij} = \frac{\left(\left\|\mathbf{x}_j,\mathbf{v}_i\right\|^2 + \beta\sum_{l\in N_j}\sum_{k\in M_i}\mathbf{u}_{lk}^{m}\right)^{-1/(m-1)}}{\sum_{i=1}^{C}\left(\left\|\mathbf{x}_j,\mathbf{v}_i\right\|^2 + \beta\sum_{l\in N_j}\sum_{k\in M_i}\mathbf{u}_{lk}^{m}\right)^{-1/(m-1)}} \tag{6.13}$$

The RFCM considers the neighbouring pixels to smooth the effect of noise which is more appropriate than NC and PCM for image segmentation. However, the modification of the objective function results in a complex variation of the membership function. It inevitably complicates the computation and weakens the continuity to the FCM algorithm.

6.4 IMPROVED FUZZY C-MEANS CLUSTERING (IFCM) ALGORITHM

A new approach is proposed in this thesis in order to overcome the drawbacks of the methods (FCM, NC, PCM and RFCM) described above. It considers the attraction between neighbouring pixels to reduce the effect of noise and artifacts in image segmentation, and is named here improved fuzzy c-means clustering (IFCM).

Usually, one pixel is too small to represent a part in an image. Assume a pixel has a completely different intensity from its surrounding pixels, it is reasonable to consider that this pixel is affected by noise. Its real intensity should be identical to its neighbour pixels. Hence, an attraction is supposed to exist between the pixels nearby. During clustering, each pixel attempts to attract its neighbouring pixels towards its own cluster. This attraction is entitled neighbourhood attraction in this thesis. Fig.6.2 is an illustration of the neighbourhood attraction.

The neighbourhood attraction depends on two factors, the pixel feature and the structure of the neighbourhood. The pixel feature is the pixel intensity in grey level images. If a pixel has a very similar intensity to one of its neighbours, the attraction between them is stronger than the attraction between the pixel and another neighbour with rather different intensities. The structure of the neighbourhood is the other important factor determining the neighbourhood attraction. Obviously, the closer neighbour pixel has a stronger attraction than the farther neighbour. The components

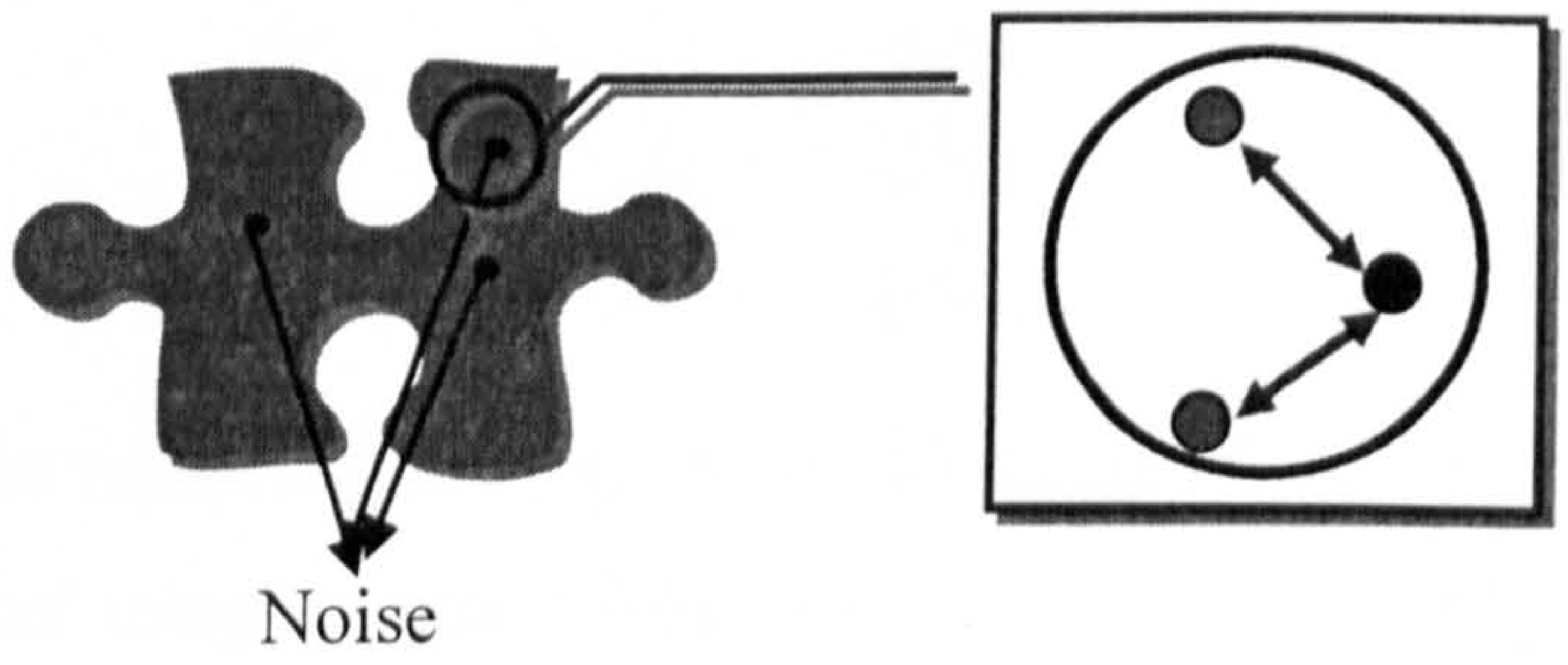


Figure 6.2 Illustration of neighbourhood attraction.

of the neighbourhood can also influence the attraction. In other words, who the neighbours are is also decisive to the neighbourhood attraction. To summarize, the first part of the neighbourhood attraction, depending on the pixel intensities, is called feature attraction while the other part of the neighbourhood attraction, depending on the position of the neighbours, is called distance attraction.

From the above description, the membership value of FCM decides the segmentation results, while the membership value is determined by the similarity measurement $d^2(\mathbf{x}_j,\mathbf{v}_i)$ [Eq. (6.4)]. It is clear that this measurement is the key to segmentation. In FCM, $d^2(\mathbf{x}_j,\mathbf{v}_i)$ measures the difference between the intensity of the pixel itself and the cluster centre without considering the effects of noise. In this approach, IFCM considers the neighbourhood attraction in $d^2(\mathbf{x}_j,\mathbf{v}_i)$ directly. The extended expression is as follows:

$$d^2(\mathbf{x}_j,\mathbf{v}_i)=\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2\left(1-\lambda H_{ij}-\xi F_{ij}\right) \tag{6.14}$$

where H_{ij} represents the feature attraction and F_{ij} represents the distance attraction.

$$H_{ij}=\frac{\sum_{k=1}^{S}\mathbf{u}_{ik}g_{jk}}{\sum_{k=1}^{S}g_{jk}} \tag{6.15}$$

In Eq. (6.15), g_{jk} is the intensity difference between the study pixel j and its neighbour pixel k.

$$g_{jk}=\left|\mathbf{x}_j-\mathbf{x}_k\right|$$

$\mathbf{u}_{ik}$ is the membership of the neighbour pixel k to the i-th cluster. S is the number of neighbour pixels.

$$F_{ij}=\frac{\sum_{k=1}^{S}\mathbf{u}_{ik}^2 q_{jk}^2}{\sum_{k=1}^{S}q_{jk}^2} \tag{6.16}$$

where q_{jk} is the relative location between the pixel j and its neighbour pixel k. In a two-dimensional image, j and k only refer to the name of the pixel, not the actual location. Fig.6.3 shows that for neighbourhood attraction the lower the intensity difference is, the stronger the feature attraction and vice versa; the closer the pixels are, the stronger the distance attraction and vice versa. The width of the arrows

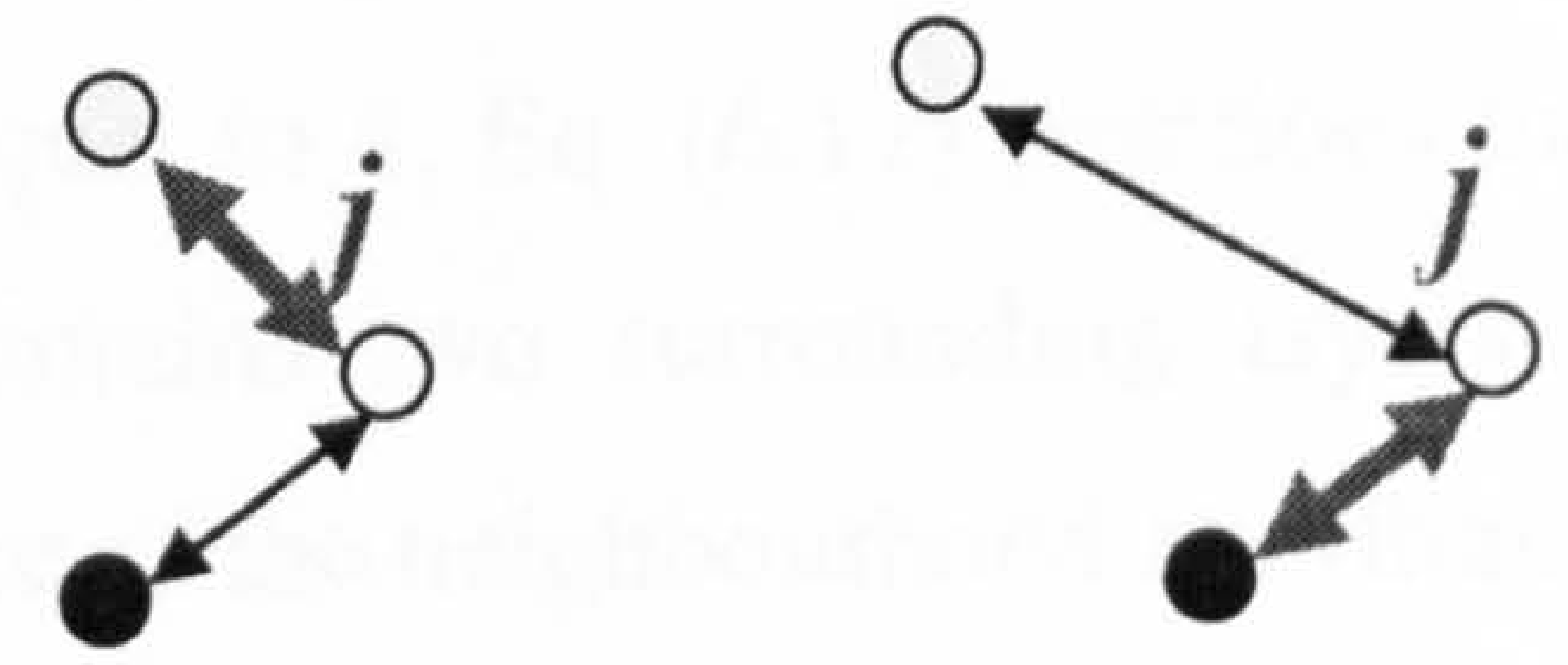


Figure 6.3 Two factors of neighbourhood attraction.

(left) feature attraction (right) distance attraction

denotes the strength of attraction.

The structure of neighbourhood is in the form:

$$K_j = \{k \in N \mid 0 < (a_j - a_k)^2 + (b_j - b_k)^2 \le Q\} \tag{6.17}$$

Where $(a_j, b_j), (a_k, b_k)$ denote the coordinates of pixel *j*, *k* and *N* is the total number of pixels in the image. *Q* is a constant, equal to $2^{(L-1)}$. *L* is the level of the neighbourhood. For example, if *Q* is equal to 1 (*L*=1), then Eq. (6.17) describes the first level neighbourhood. It includes four pixels on the top, bottom, left and right adjacently to the study pixel. If *Q* is equal to 2 (*L*=2), the eight pixels surrounding the

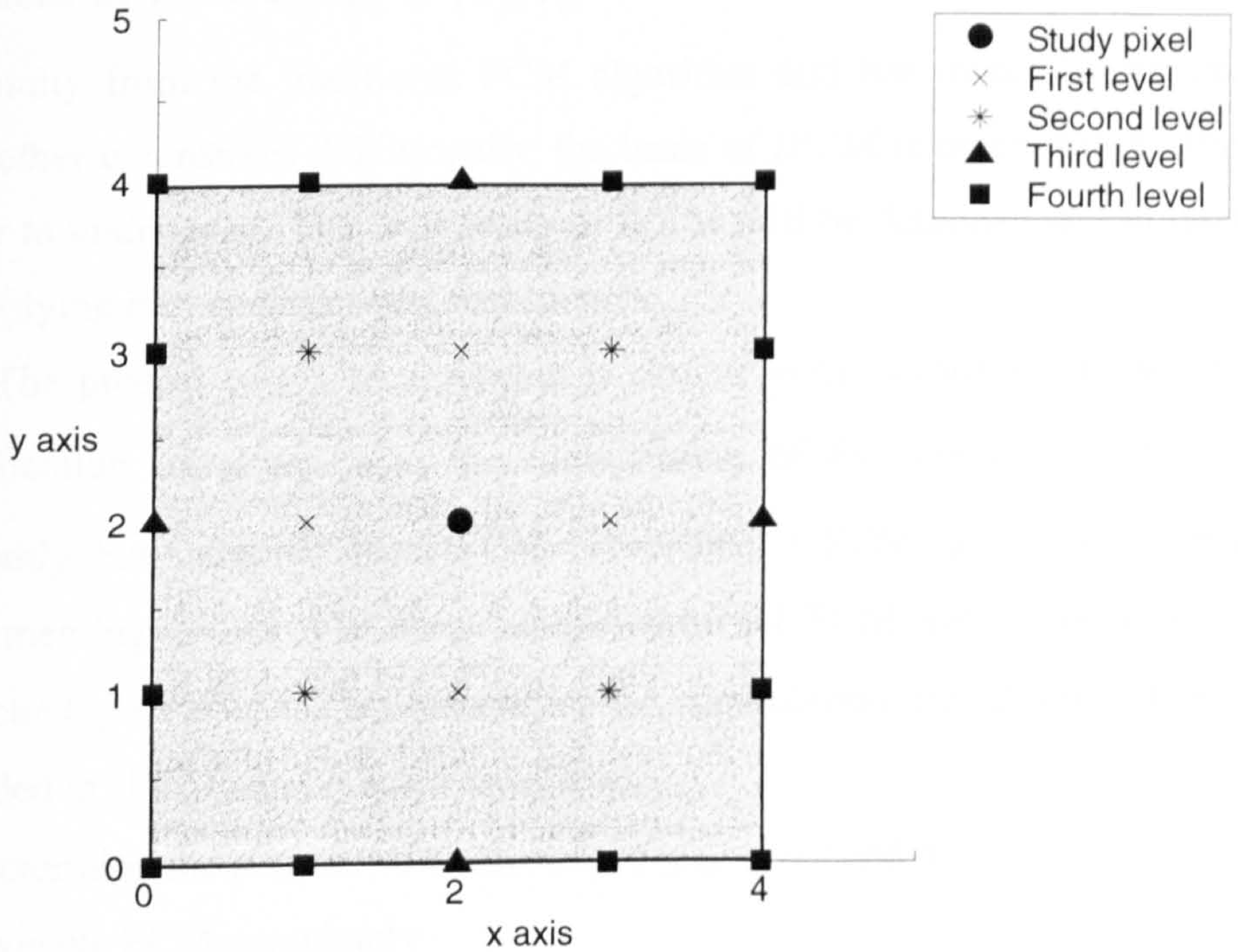


Figure 6.4 Structure of neighbourhood.

study pixel compose the neighbourhood, which is called the second level neighbourhood. If Q is equal to 8, Eq. (6.17) describes the fourth level (L=4) of the neighbourhood which contains two surrounding layers of 24 pixels, and so on. Fig.6.4 shows the structure of the neighbourhood at different levels.

Therefore, q_{jk} in Eq. (6.16) can be described as follows:

$$q_{jk} = (a_j - a_k)^2 + (b_j - b_k)^2 \tag{6.18}$$

Two parameters λ and ξ in Eq. (6.14), with values between 0 to 1, are used to adjust the degree of the two neighbour attractions. Higher λ leads to a stronger feature attraction and higher ξ leads to a stronger distance attraction. The determination of these two parameters is described later.

The original FCM algorithm has been improved in IFCM in that the segmentation of images is not only decided by the pixel itself but also decided by its neighbouring pixels. Even the noise or artifact changes the intensities of some pixels; they can be segmented more appropriately with the help of their neighbour pixels. Compared to RFCM and other methods which modified the objective function also resulting in the complexity of the membership function, IFCM preserves most equations in FCM except $d^2(\mathbf{x}_j, \mathbf{v}_i)$. There is no doubt that IFCM has a better continuity from the traditional FCM algorithm and has more simple computation than other extensions. Additionally, the basis of IFCM is more straight forward and easier to understand. The superiority of IFCM will be demonstrated in later sections by applying it to synthetic and real images.

The process of IFCM algorithm is similar to the traditional FCM. Besides the modification of $d^2(\mathbf{x}_j, \mathbf{v}_i)$, the initialization of the membership is not created randomly but inherited from FCM. Therefore, a FCM is executed firstly before implementing IFCM. The final memberships of FCM are regarded as the initial membership of IFCM. This can reduce the computation time dramatically. The steps included in IFCM can be listed as follows:

(1) Determine the number of clusters C ($2 \le C \le N$) and m

(2) Execute FCM completely

(3) Utilize the final membership of FCM as the initial membership $\mathbf{u}_{ij}^{(0)}$ of IFCM

(4) At the l-th iteration (l=0,1,2,...), calculate the cluster centre $\mathbf{v}_i^{(l)}$ ($i = 1,2...C$)

using the membership $\mathbf{u}_{ij}^{(l)}$

(5) Calculate the improved similarity measurement $d^2(\mathbf{x}_j,\mathbf{v}_i^{(l)})$

(6) Update $\mathbf{u}_{ij}^{(l)}$ with $d^2(\mathbf{x}_j,\mathbf{v}_i^{(l)})$

(7) Compare $\mathbf{u}_{ij}^{(l)}$ and $\mathbf{u}_{ij}^{(l-1)}$, if $\left\|\mathbf{u}_{ij}^{(l)}-\mathbf{u}_{ij}^{(l-1)}\right\|<\varepsilon$, then stop, otherwise, $l=l+1$, go to step 4 and repeat

6.5 PARAMETER ESTIMATION

From Eq. (6.14), the similarity measurement is modified by considering the neighbourhood attraction. Two parameters λ and ξ are selected to adjust the degree of feature attraction and distance attraction. A higher parameter leads to a stronger attraction. However, different images require different parameter values. Inappropriate λ and ξ will not improve the segmentation results but can make them worse.

The minimum and maximum limits of λ and ξ are 0 and 1. Suppose the minimum value of them is 0.01, the number of possible combinations of λ and ξ is 10000. To determine the best value of the parameters, 10000 tests are required. Obviously, it is not practical that each image is segmented 10000 times in order to find the best results. A simple artificial neural network (ANN) model is therefore created for this thesis in order to search for the optimal values of the parameters in IFCM. Fig.6.5 shows the architecture of the model, where $\mathbf{w}_i$ (i= 0, 1, 2) denotes the interconnection weights of the network.

$$\mathbf{w}=\begin{bmatrix}\mathbf{w}_0\\\mathbf{w}_1\\\mathbf{w}_2\end{bmatrix}=\begin{bmatrix}1\\\lambda\\\xi\end{bmatrix}$$

The cost function E of the neural work is defined as:

$$E=\frac{1}{N}\sum_{j=1}^{N}\frac{MJ_m}{DoRate} \tag{6.19}$$

where $j=1\ldots N$ (pixel number).

$$MJ_m = \frac{Min(J_m)}{Max(J_m)} \tag{6.20}$$

J_m is the objective function in Eq. (6.1), which requires to be minimized. It gradually converges during IFCM iteration. Hence, MJ_m describes the degree of the convergence. A lower MJ_m represents a better convergence of J_m.

$$DoRate = \frac{1}{N}\sum_{j=1}^{N} MaxU_j \tag{6.21}$$

DoRate is defined as the abbreviation of 'Dominating Rate'. $MaxU_j$ stands for the highest membership of each pixel to all clusters, which is also the membership of each pixel to the cluster that into which is finally segmented. For example, one pixel has memberships to five clusters as [0.3 0.1 0.05 0.5 0.05]; another pixel has [0.05

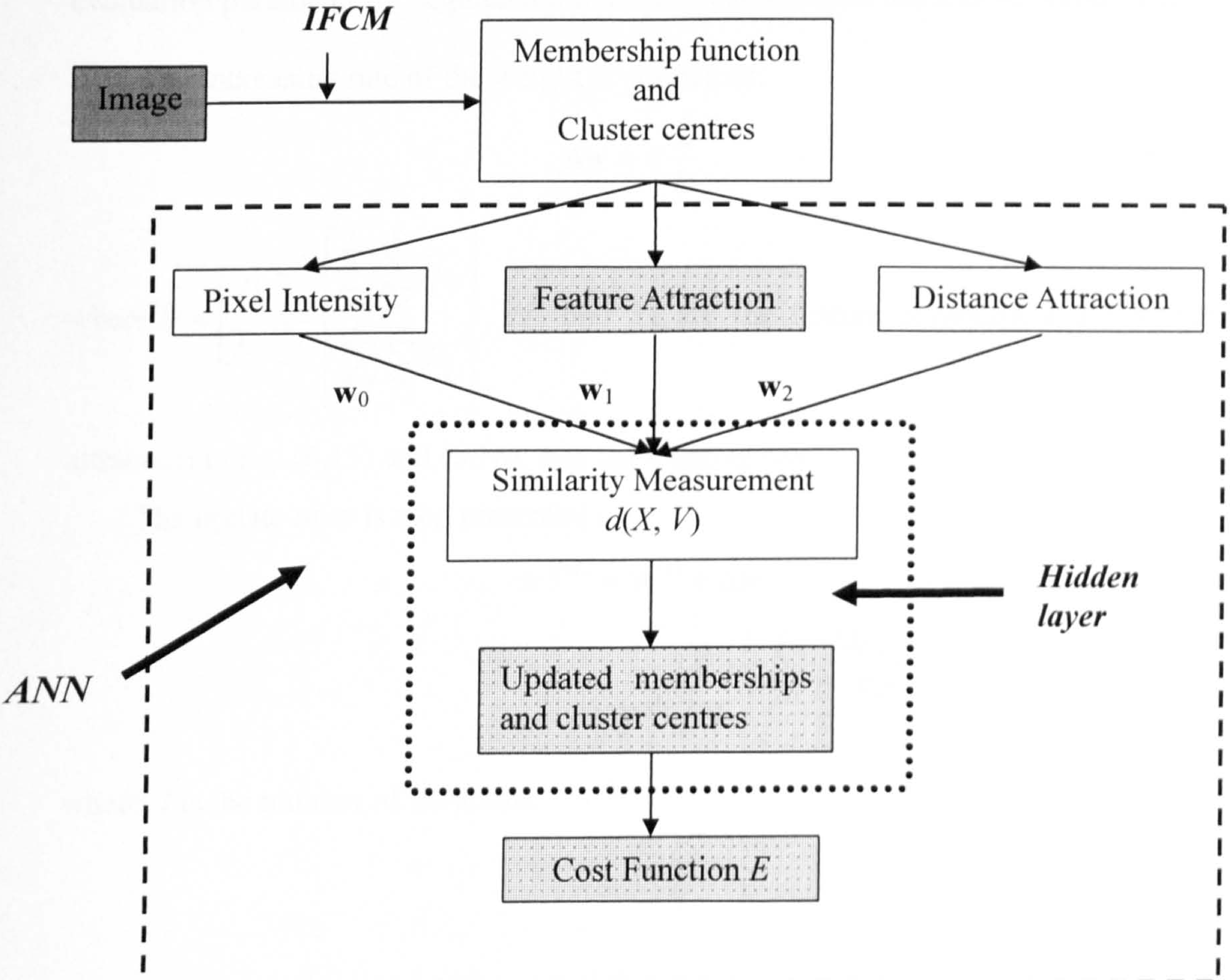


Figure 6.5 Architecture of ANN model.

0.01 0.02 0.91 0.01]. The former has a *MaxU* equal to 0.5 and *MaxU* of the latter is 0.91. Although both pixels can be segmented to the fourth cluster, it is clear the latter one is more dominant in its cluster. N is the number of pixels in the image; thus *DoRate* represents the average dominating rate in segmentation. Higher *DoRate* indicates a more definite segmentation result.

In the cost function E, the numerator MJ_m is proposed to be minimized and the denominator *DoRate* is meant to be maximized; hence E must be converged to a global minimum in the ANN model. $\frac{1}{N}\sum_{j=1}^{N}$ in Eq. (6.19) denotes this is an average minimum achieved for the whole image. E can be also regarded as a performance evaluation parameter for segmentation. Better segmentation has a lower value of E.

The increasing rate of the weight is defined as:

$$\Delta \mathrm{w} = \eta \frac{E}{D} \tag{6.22}$$

where $D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{C}\sum_{j=1}^{N} H_{ij} \\ \sum_{i=1}^{C}\sum_{j=1}^{N} F_{ij} \end{bmatrix}$, H_{ij} *and* F_{ij} are the feature attraction and distance attraction in Eq. (6.15) and (6.16). η is the learning rate.

The update of **w** is then presented as:

$$\begin{aligned} \mathrm{w}^{(l+1)} &= \mathrm{w}^{(l)} + \Delta \mathrm{w} \\ &= \mathrm{w}^{(l)} + \eta \frac{\frac{1}{N}\sum_{j=1}^{N}\frac{MJ_m}{DoRate}}{D} \end{aligned} \tag{6.23}$$

where, l is the number of iterations.

$$\begin{bmatrix} 1 \\ \mathbf{w}_1^{(l+1)} \\ \mathbf{w}_2^{(l+1)} \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{w}_1^{(l)} \\ \mathbf{w}_2^{(l)} \end{bmatrix} + \eta \begin{bmatrix} 0 \\ \dfrac{\frac{1}{N}\sum_{j=1}^{N}\frac{MJ_m}{DoRate}}{D_1} \\ \dfrac{\frac{1}{N}\sum_{j=1}^{N}\frac{MJ_m}{DoRate}}{D_2} \end{bmatrix} \tag{6.24}$$

The process of parameter estimation consists of the following steps:

(1) Initialize $\mathbf{w}^{(0)}$, where $\lambda=0$ and $\xi=0$, so that the segmentation method becomes identical to the traditional FCM

(2) Obtain the initial membership function and the cluster centers

(3) Compute the similarity measurement using the present $\mathbf{w}^{(l)}$

(4) Update the membership function and the cluster centers and compute the cost function E

(5) Update $\mathbf{w}^{(l)}$ to $\mathbf{w}^{(l+1)}$ using Eq. (6.23)

If $\mathbf{w}_r^{(l+1)} \geq 1$,

Then $\mathbf{w}_r^{(l+1)} = \mathbf{w}_r^{(l+1)} - 1$ where $r=1,2$

(6) Stop if both $\mathbf{w}_{1,2}^{l+1}$ reached the value 1 once, otherwise go to step 3

The termination criteria are defined as when both parameters have been searched completely from 0 to 1. This helps to find a global minimum rather than a local minimum of the cost function. The process usually iterates 20-30 times and reduces the validation time dramatically.

6.6 RESULT EVALUATION

Several segmentation methods based on FCM were described in section 6.3, such as NC, PCM and RFCM. They all are supposed to be robust in a noisy environment. In this section, the synthetic and real images were applied to such methods to compare their segmentation performance with our IFCM algorithm. For the sake of precise comparison, some definitions are addressed here:

N_{tp}: the number of pixels that belong to the cluster and are segmented into this cluster correctly.

N_{tn}: the number of pixels that do not belong to the cluster and are not segmented into this cluster.

N_{fp}: the number of pixels that do not belong to the cluster and are segmented into this cluster incorrectly.

N_{fn}: the number of pixels that belong to the cluster and are not segmented into the cluster.

N_p: the number of all pixels that belong to the cluster

N_n: the number of all pixels that do not belong to the cluster.

N: the number of all pixels in the image.

Five evaluation parameters are hence defined as follows:

First parameter of correct segmentation (C_1): $C_1 = \frac{N_{tp}}{N_p}$, representing the percentage of positive true segmentation.

Second parameter of correct segmentation (C_2): $C_2 = \frac{N_{tn}}{N_n}$, representing the percentage of negative true segmentation.

Under segmentation (UnS): $UnS = \frac{N_{fp}}{N_n}$, representing the percentage of negative false segmentation.

Over segmentation (OvS): $OvS = \frac{N_{fn}}{N_p}$, representing the percentage of positive false segmentation.

Incorrect segmentation (InC): $InC = \frac{N_{fp} + N_{fn}}{N}$, representing the percentage of false segmentation.

The first two parameters (C_1, C_2) describe correct evaluation, the higher the better. The latter three parameters (UnS, OvS, InC) describe incorrect evaluation, the lower the better.

6.6.1 Square Image

The first example is a square image. An image of dimension 256 by 256 was created. Every 64 pixels consisted of a square with equal grey levels. The structure of first four squares is shown in Fig.6.6.

There were 1024 squares in total in this image. The grey level of each square was randomly chosen from four grey levels: 0, 64, 128 and 196. The original image is shown in Fig.6.7 (a). Random noise uniformly distributed in the interval (0,100)

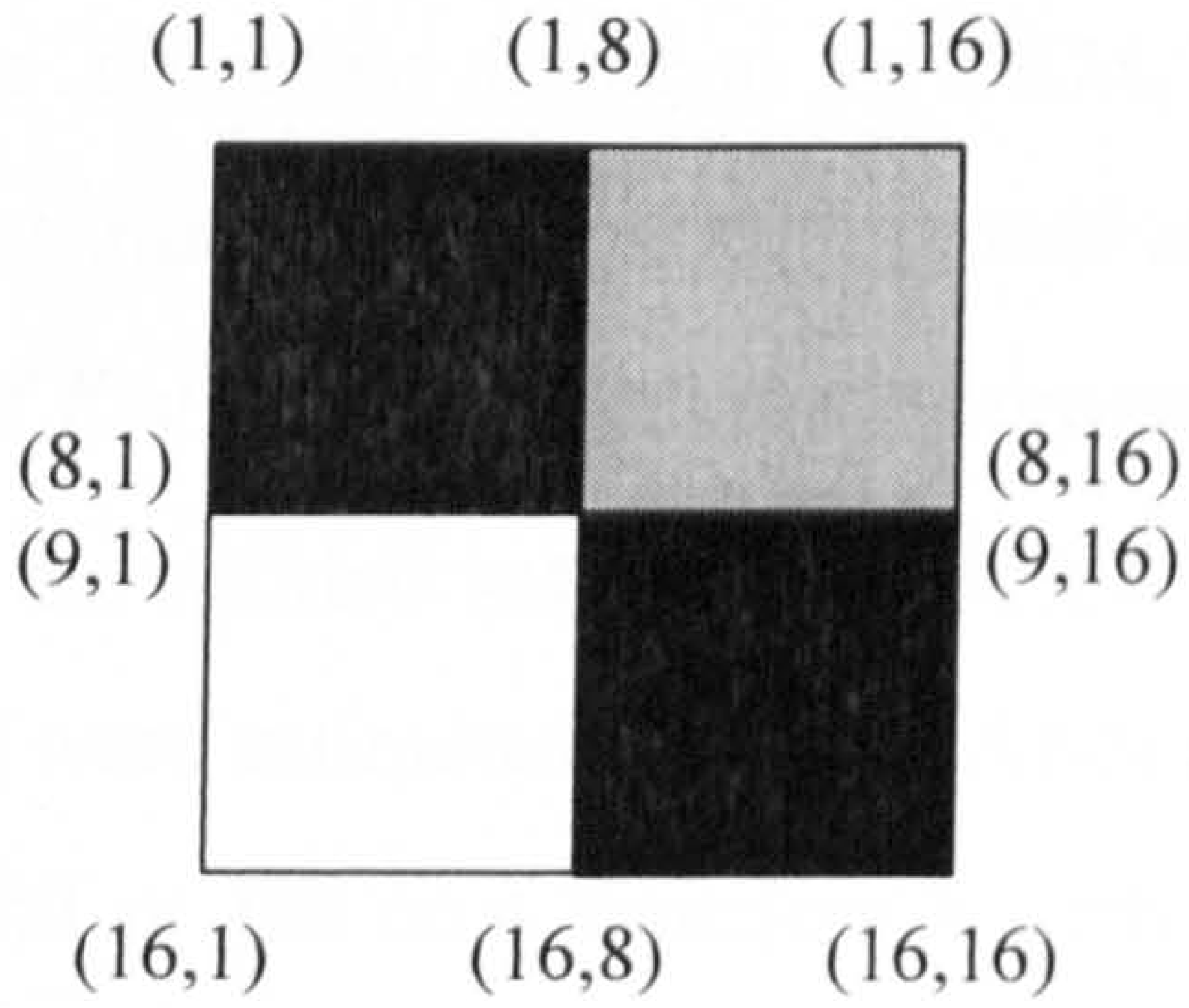
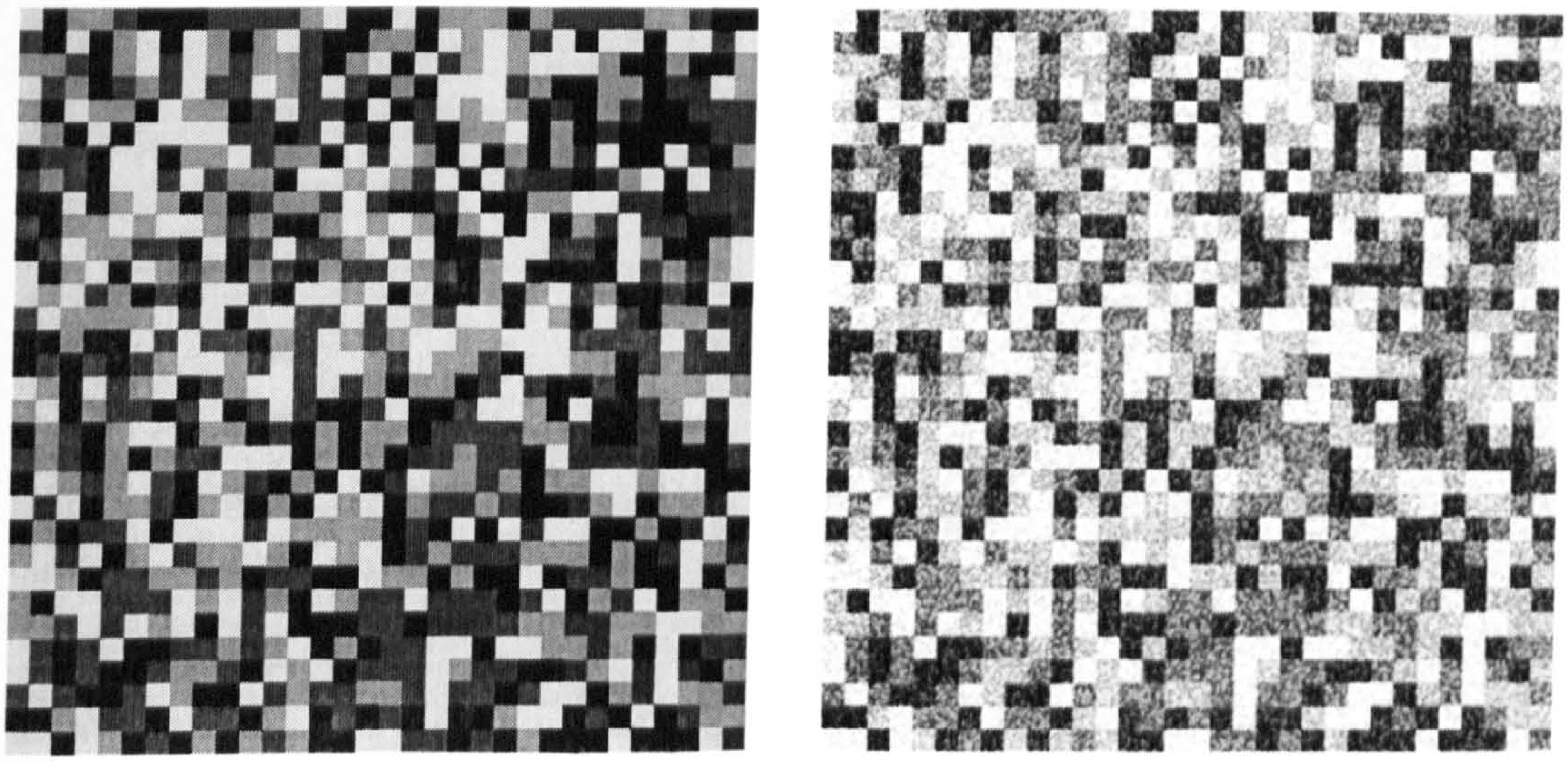


Figure 6.6 First four squares in square image.



(a) Original image

(b) Noisy image

Figure 6.7 1024-square image.

was added onto the original image and the noisy image is shown in Fig.6.7 (b).

The noisy image was segmented by FCM, RFCM and IFCM algorithms respectively. The implementation of PCM verified its fatal drawback in that pixels became trapped in one cluster, and this means that PCM cannot segment multi-class completely. Even a modified method in which the similarity measurement was assumed to be greater than a certain value did not solve the problem.

Optimized parameters in RCM were obtained by means of a cross validation method presented by Pham (2001). It allowed a fair comparison between RCM and IFCM as both methods were optimized. The value of the parameter β [Eq. (6.12)] can influence the segmentation results dramatically. A cost function *Ecv*, similar to the expression of the objective function in FCM, was defined to evaluate the adaptability of β to the image [Pham, 2001]. It was computed for the square image. Fig.6.8 shows a plot of *Ecv* against β and the minimum of *Ecv* was achieved at β=25, indicating that 25 was the optimal value of β in RFCM applied to the square image.

In IFCM, λ and ξ were estimated using the ANN model presented in section 6.5. Fig.6.9 shows the trend of the cost function E with respect to λ and ξ. The total number of iterations was 23. The global minimum of E was achieved when λ=0.5125 and ξ=0.6359 at the eleventh iteration. These values for IFCM were employed in segmentation.

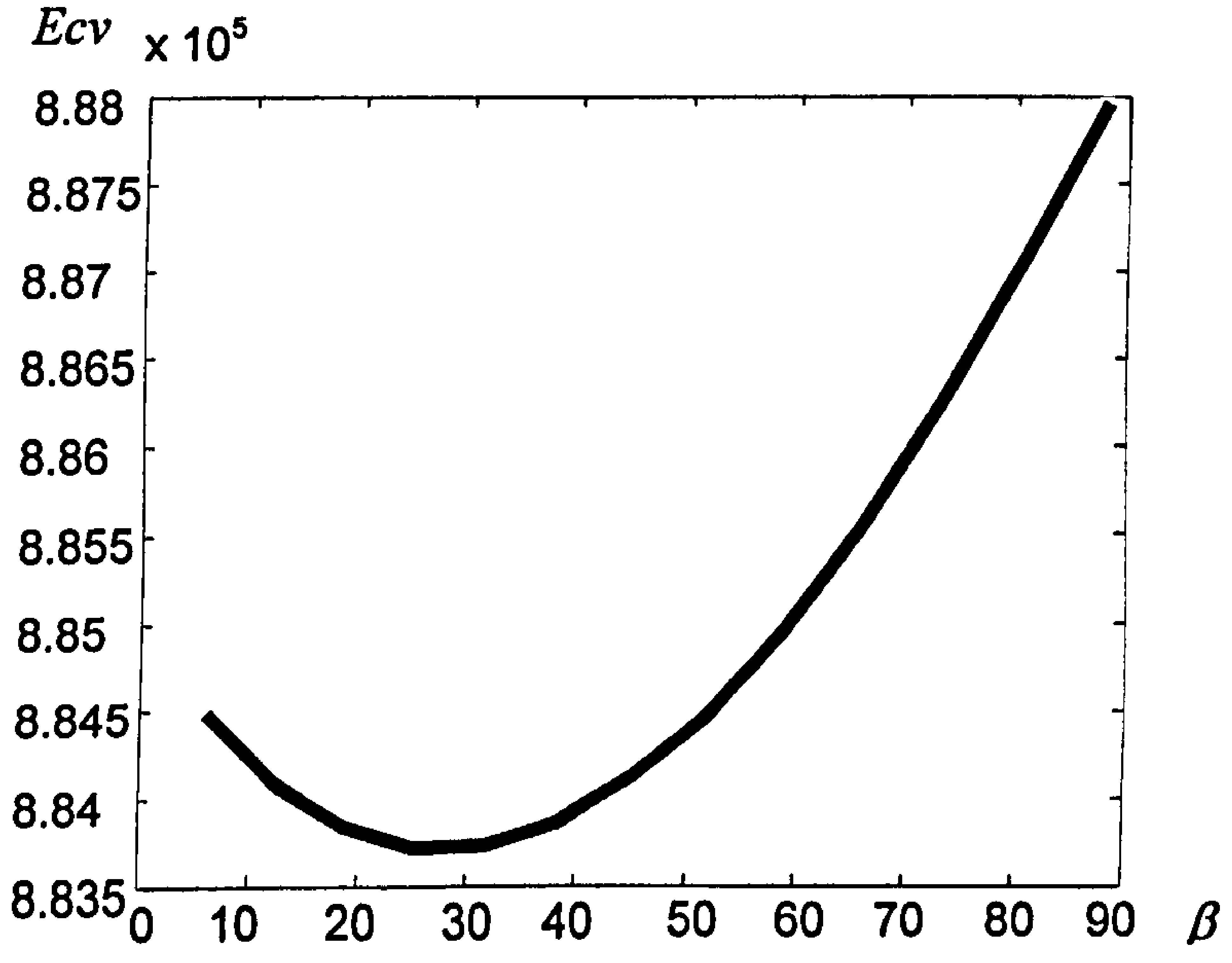


Figure 6.8 Plot of *Ecv* against β in RFCM.

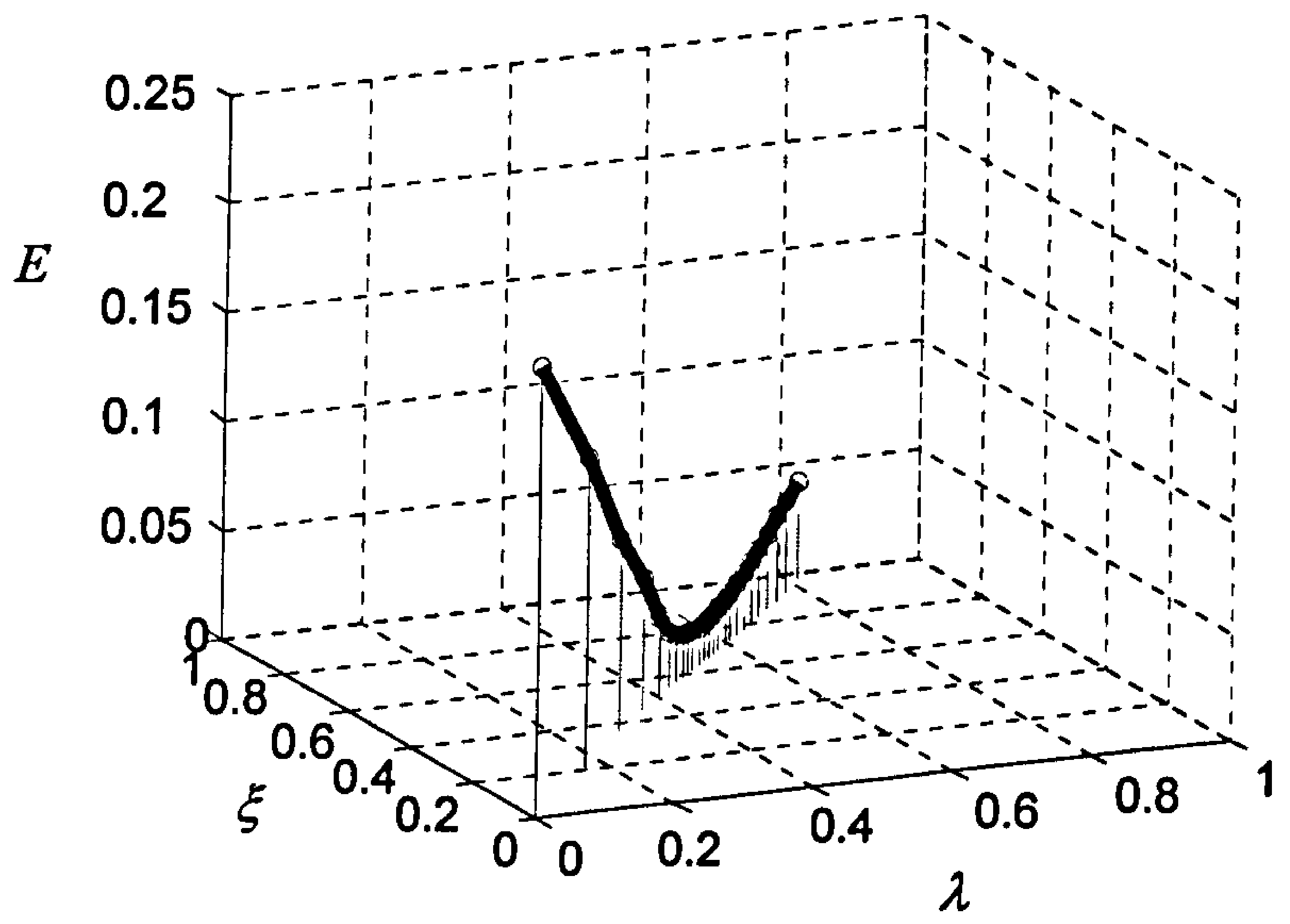
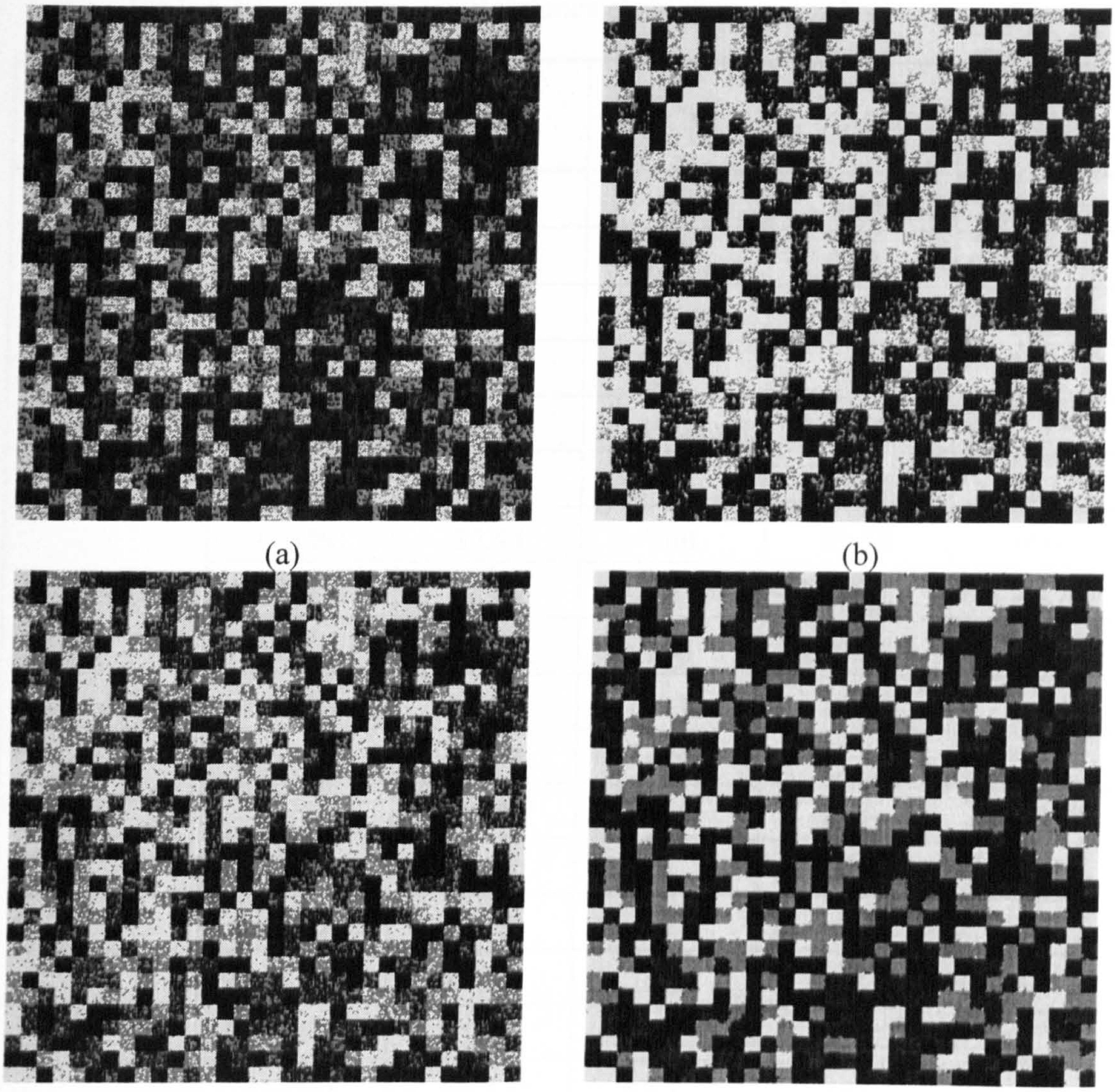


Figure 6.9 Cost function with respect to λ and ξ in IFCM.

The segmentation results of FCM, PCM, RFCM and the IFCM algorithms were obtained by assigning each pixel to the cluster in which it had the highest membership and this is shown in Fig.6.10 (The NC algorithm was not compared because of the flaw addressed above, may lead to the failure of segmentation).. A value of 2, unless otherwise stated, for the weighting exponent *m* was employed in all fuzzy clustering methods in this thesis, since this value was proven to yield good results in practice. The number of clusters (*C*) was 4 for all methods in this example. It was clear the former three results were seriously influenced by noise and only IFCM almost overcome this problem.

Table 6.1 lists the evaluation parameters of these four algorithms used to quantify the segmentation performance. The first column of the table represents each cluster within four clusters. RFCM had very similar evaluation parameters, both correct and incorrect ones, to the FCM algorithm, demonstrating the comparable performance of these two algorithms. PCM was the best for two of the clusters but the worst for the other two. There is no doubt suggesting that IFCM achieved much more improved performance for each cluster and also the best average performance. The best evaluation parameters were acquired, showing the highest accuracy of segmentation. Fig.6.11 is a chart displaying the parameters listed in Table 6.1.



(a) (b)

(c) (d)

Figure 6.10 Segmentation results of 1024-square image.

(a) FCM (b) PCM (c) RFCM (d) IFCM

Table 6.1 Segmentation performance evaluation on 1024-square image.

Method Class	Evaluation Parameters	FCM	PCM	RFCM	IFCM
1	C_1 (%)	57.27	100	91.91	98.40
	C_2 (%)	100	73.76	91.24	99.37
	UnS (%)	0	26.24	8.76	0.63
	OvS (%)	42.73	0	8.09	1.60
	InC (%)	10.64	19.71	8.60	0.87
2	C_1 (%)	59.78	0	72.94	95.77
	C_2 (%)	85.94	100	83.58	98.63
	UnS (%)	14.06	0	16.42	1.37
	OvS (%)	40.22	100	27.06	4.23
	InC (%)	20.42	24.32	19.01	2.06
3	C_1 (%)	72.27	100	60.21	95.76
	C_2 (%)	83.80	100	86.29	98.44
	UnS (%)	16.20	0	13.71	1.56
	OvS (%)	27.73	0	39.79	4.24
	InC (%)	19.22	26.17	20.54	2.26
4	C_1 (%)	91.13	0	58.87	97.57
	C_2 (%)	90.37	59.17	100	99.37
	UnS (%)	9.63	40.83	0	0.63
	OvS (%)	8.87	100	41.13	2.43
	InC (%)	9.44	30.78	10.12	1.07
Average	C_1 (%)	70.11	50.00	70.98	96.87
	C_2 (%)	90.03	83.23	90.28	98.95
	UnS (%)	9.97	16.77	9.72	1.05
	OvS (%)	29.89	50.00	29.02	3.13
	InC (%)	14.93	25.24	14.57	1.57

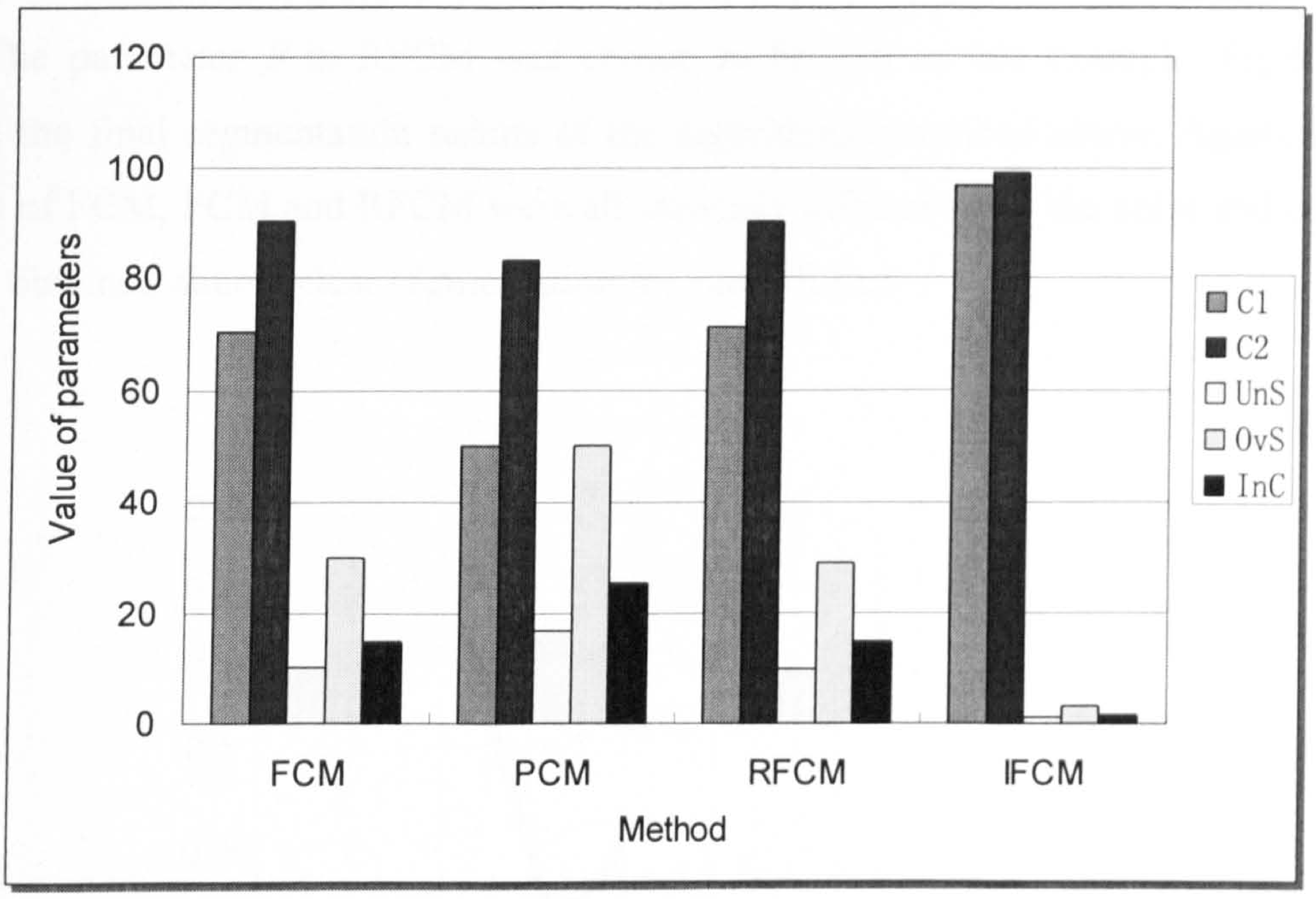


Figure 6.11 Chart of evaluation parameters of 1024-square image.

Another example was given to test the algorithm. It was a square image comprising 16 squares in total and including four grey levels: 0, 100, 200 and 300, shown in Fig.6.12. Noise, uniformly distributed in the interval (0, 120), was added on the image.

Fig.6.13 shows the trends of the cost function and the domination rate against λ and ξ during the neural network selection in IFCM for this image. The cost function achieved a global minimum at λ=0.4675 and ξ=0.5592 which also maximized the

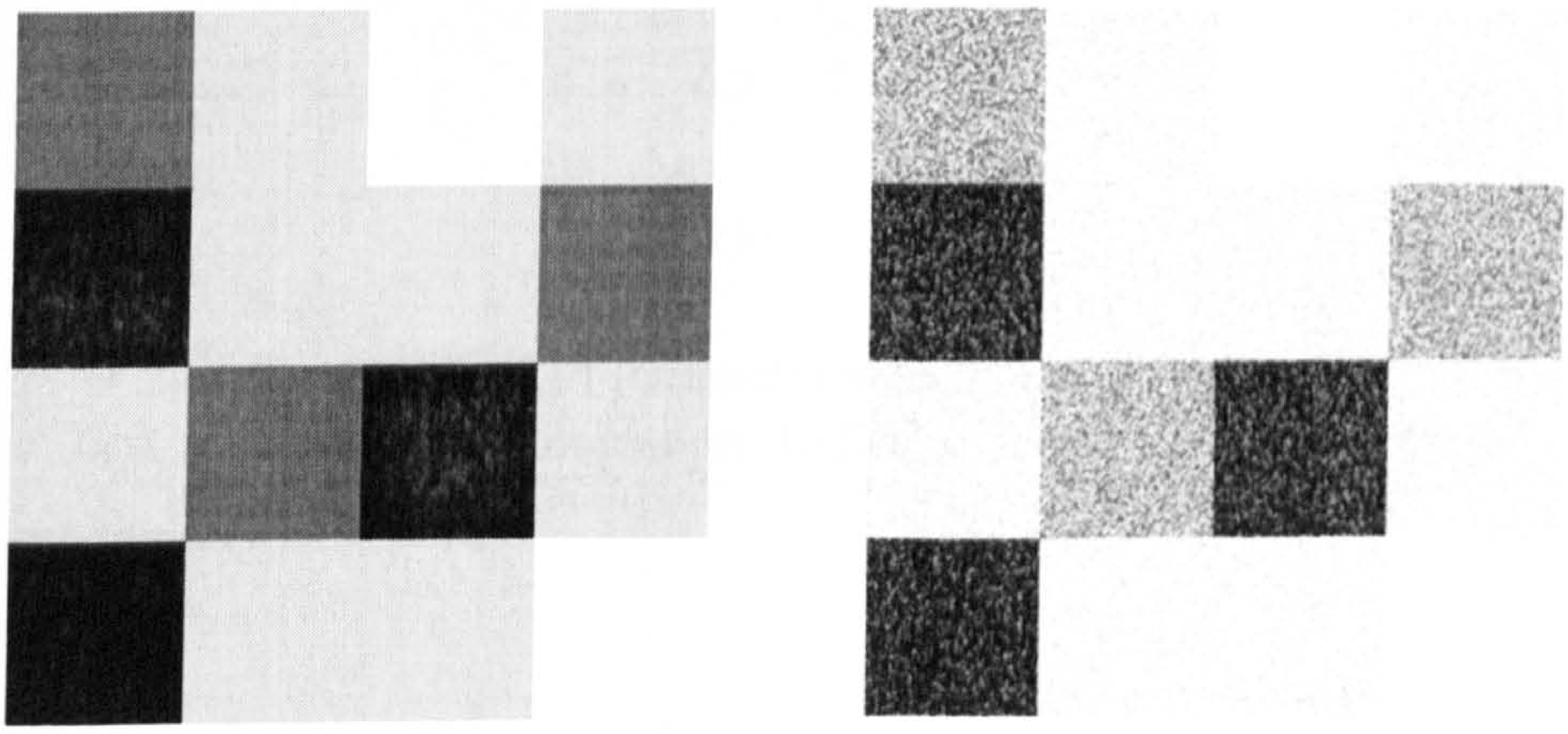
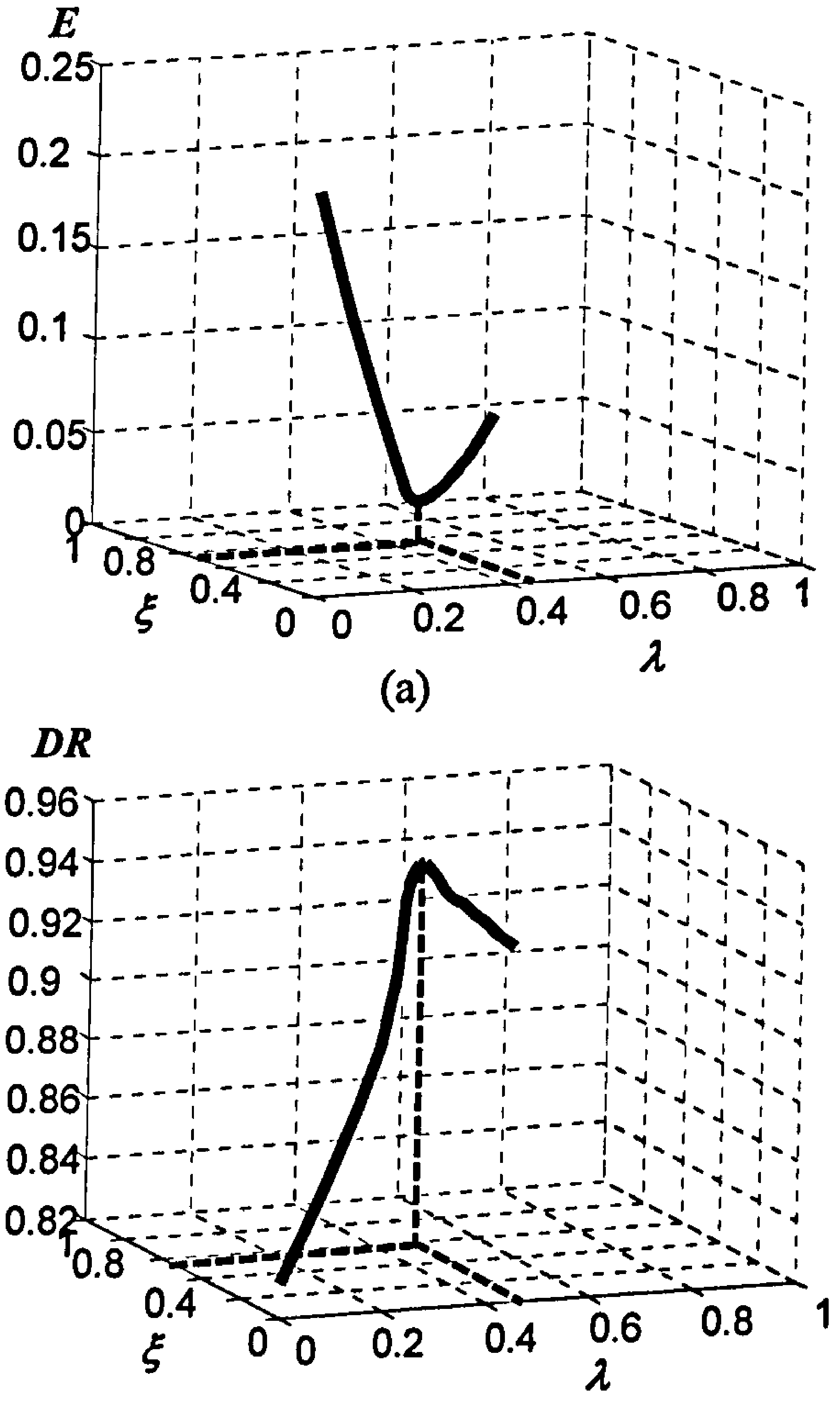


Figure 6.12 16-square image (left) original image (right) noisy image.

domination rate.

The parameter β in RFCM was chosen to be 328 in this example. Fig.6.14 shows the final segmentation results of the algorithms described above. Again, the results of FCM, PCM and RFCM were all seriously influenced by the noise and only IFCM obtained almost clear segmentation for each cluster.



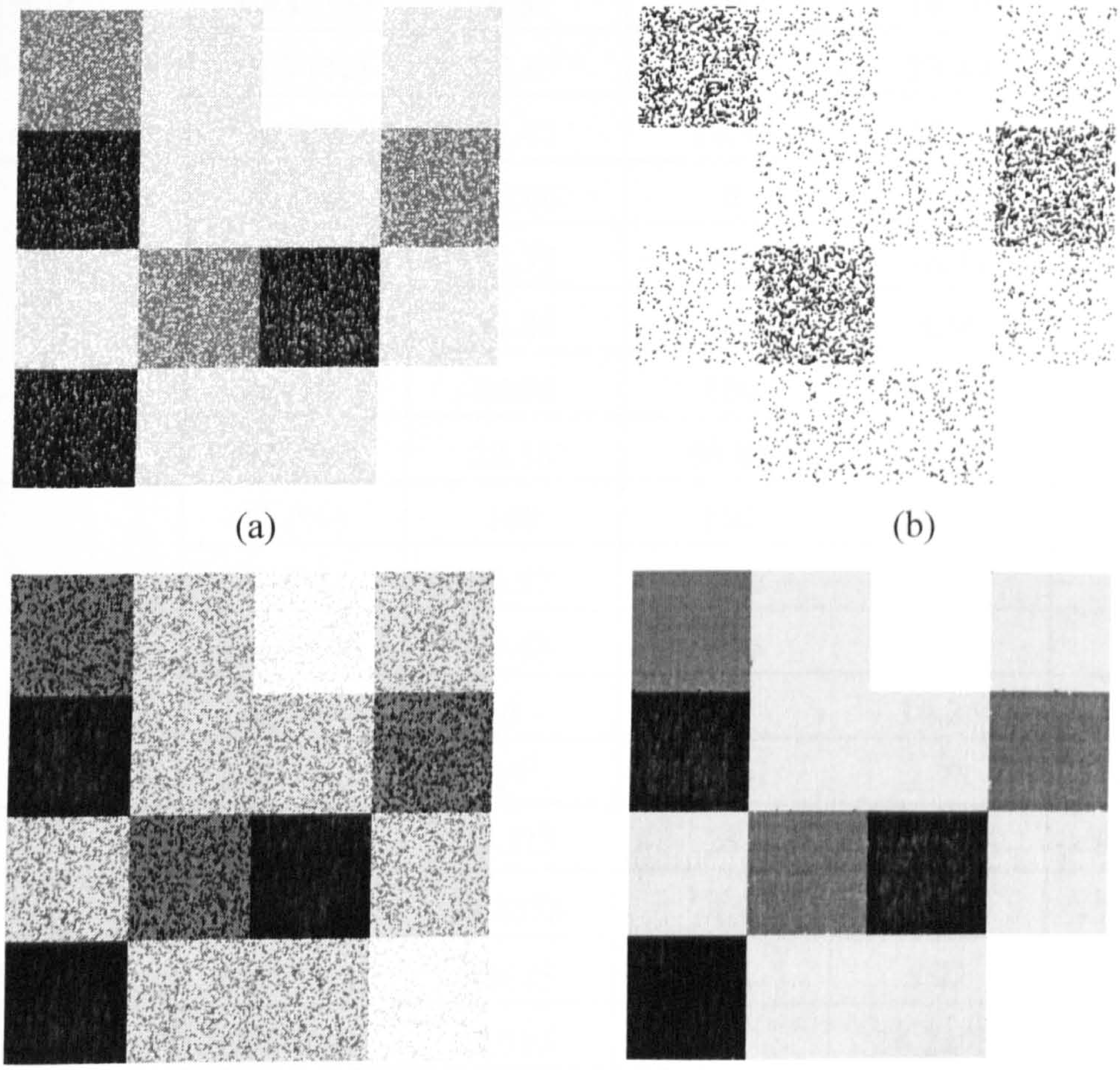
(a)

(b)

Figure 6.13 Parameter selection of IFCM.
(a) Cost function minimization (b) Domination rate maximization

Table 6.2 enumerates the evaluation parameters for each method. The accuracy of IFCM was much higher and the error rate was much lower than for the other methods. It again indicated the superiority of IFCM in a noisy environment compared to other methods.

A third example was tested which was a three grey level square image made up of 128 squares. It was segmented into 3 clusters by means of all algorithms described above. The noise added was Gaussian distributed in the interval (0, 100). β was 80 in RFCM, and λ=0.5466 and ξ=0.6844 in IFCM. Table 6.3 lists the average evaluation parameters of results for all algorithms. All these examples verified that any square image with any grey level and any noise can be segmented more accurately by the IFCM algorithm.



(a) (b)

(c) (d)

Figure 6.14 Segmentation results of 16-square image.
(a) FCM (b) PCM (c) RFCM (d) IFCM

Table 6.2 Segmentation performance evaluation on 16-square image.

Methods Class	Evaluation Parameters	FCM	PCM	RFCM	IFCM
1	C_1 (%)	71.53	0	100	99.98
	C_2 (%)	100	88.78	94.58	100
	UnS (%)	0	11.22	5.42	0
	OvS (%)	28.47	100	0	0.02
	InC (%)	5.34	27.87	4.40	0.006
2	C_1 (%)	67.33	0	76.51	99.91
	C_2 (%)	93.43	100	83.30	99.97
	UnS (%)	6.57	0	16.70	0.03
	OvS (%)	32.67	100	23.49	0.09
	InC (%)	11.47	18.75	17.97	0.038
3	C_1 (%)	66.00	0	72.87	99.95
	C_2 (%)	87.75	100	96.44	99.95
	UnS (%)	12.25	0	3.56	0.05
	OvS (%)	34.00	100	27.13	0.05
	InC (%)	23.13	50.00	15.35	0.050
4	C_1 (%)	100	100	85.75	99.95
	C_2 (%)	80.57	10.42	100	99.99
	UnS (%)	19.43	89.58	0	0.01
	OvS (%)	0	0	14.25	0.05
	InC (%)	17	78.38	1.78	0.011
Average	C_1 (%)	76.215	25	83.7825	99.9475
	C_2 (%)	90.4375	74.8	93.58	99.9775
	UnS (%)	9.5625	25.2	6.42	0.0225
	OvS (%)	23.785	75	16.2175	0.0525
	InC (%)	14.235	43.75	9.875	0.02625

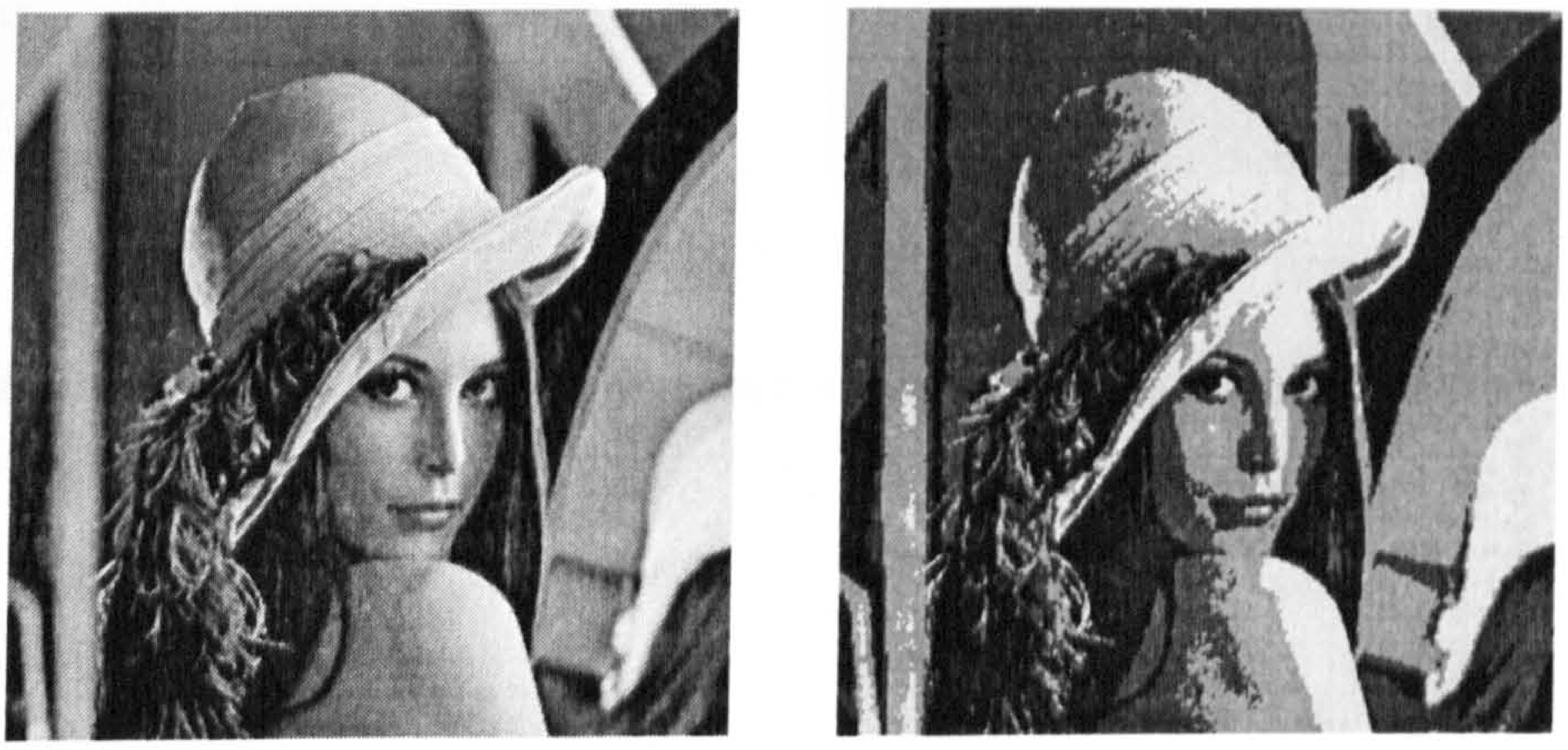
Table 6.3 Average performance evaluation on 128-square image.

Methods Parameters	FCM	PCM	RFCM	IFCM
C_1 (%)	**57.68**	**56.02**	**57.75**	**80.23**
C_2 (%)	**78.12**	**77.23**	**78.15**	**89.54**
UnS **(%)**	**21.88**	**22.77**	**21.85**	**10.46**
OvS **(%)**	**42.32**	**43.98**	**42.25**	**19.77**
InC **(%)**	**28.88**	**31.12**	**28.84**	**13.78**

6.6.2 Facial Image

Lena image is one of the most popular and famous images for image processing available to researchers [David and Munson, 1996]. It was employed in our work to test the superiority of IFCM on real images.

The image was a 256 by 256 grey level image shown in Fig.6.15 (a). The grey level was between 0 and 238. Unlike the synthetic square images, the ground truth of segmentation in Lena image is unknown. Therefore, FCM was applied to segment the original Lena image and the results were regarded as the ground truth of



(a) (b)

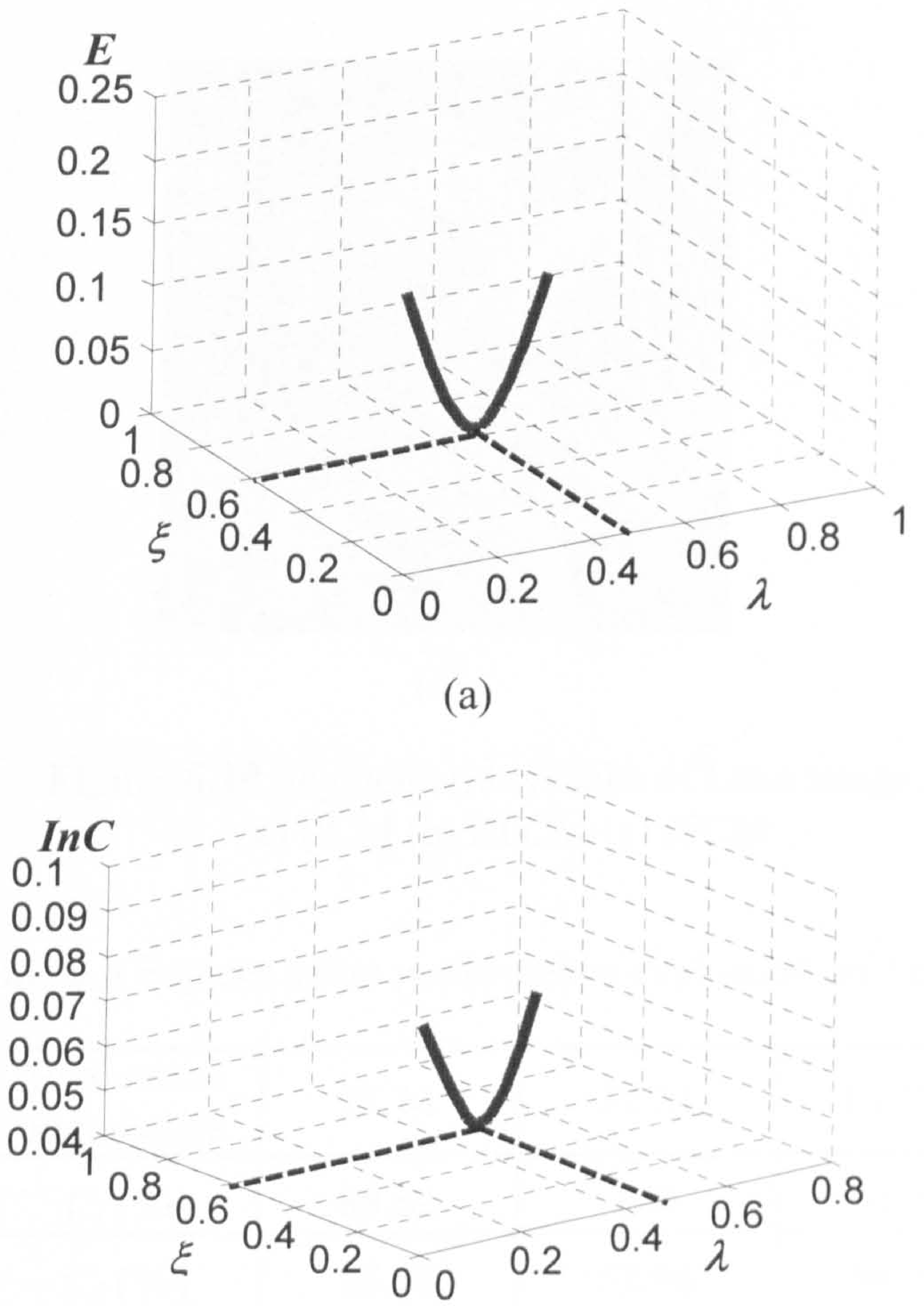
Figure 6.15 Lena image
(a) original Lena image (b) segmented image

segmentation. Fig 6.15 shows both the original Lena image and the segmented image. Thereafter, random noise which obeys the uniformed distribution between 0 and 100 was added to the original Lena image. The noisy image is given in Fig.6.16 (a) and the segmented noisy image by FCM is in Fig.6.16 (b).

The parameters λ and ξ in IFCM were estimated using the ANN model for the Lena image and were chosen to be 0.4976 and 0.5989, respectively. Fig.6.17 (a) gives the plot of the cost function *E* with respect to λ and ξ. Fig.6.17 (b) shows that the average *InC* of the segmentation results was minimized at the same value. The segmentation results of PCM, RFCM and IFCM are shown in Fig.6.18 and the evaluation parameters are listed in Table 6.4. This suggests that both visually and quantitatively, the segmentation result of IFCM had better quality and higher accuracy over all other methods. PCM was stuck in two clusters as it was when segmenting synthetic images and RFCM, with β=260 for this image, showed an improved performance over FCM.

(a) (b)

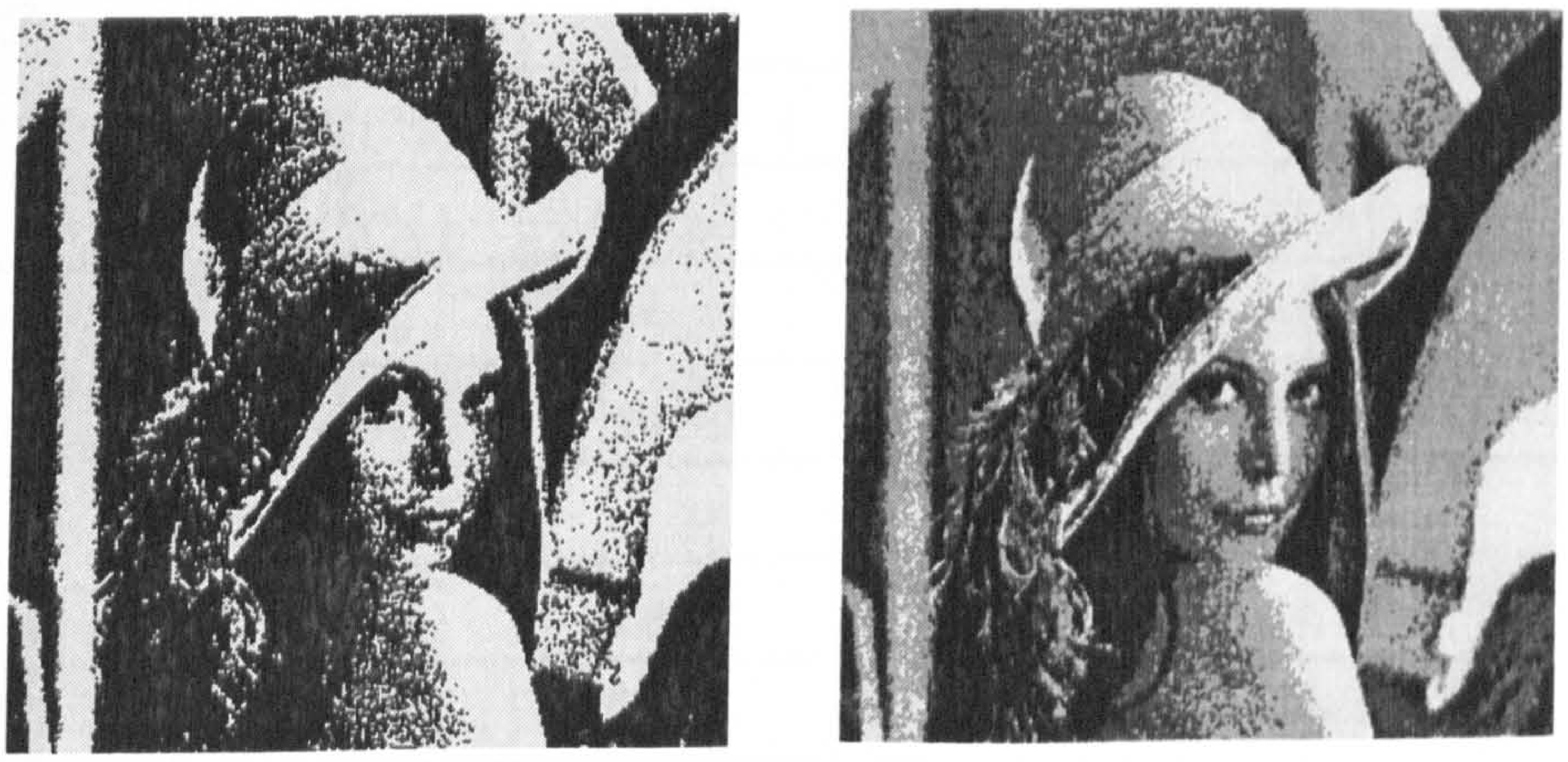
Figure 6.16 (a) Noisy Lena image and (b) segmented noisy image by FCM.



(a)

(b)

Figure 6.17 IFCM parameter estimation for Lena image.
(a) plot of the cost function E (b) plot of the average InC



(a)

(b)

(c)

Figure 6.18 Segmentation results of Lena image.
(a) PCM (b) RFCM (c) IFCM

Table 6.4 Segmentation performance evaluation on Lena image.

Method Class	Evaluation Parameters	FCM	PCM	RFCM	IFCM
1	C_1 (%)	**88.82**	**100**	**91.44**	**94.27**
	C_2 (%)	**95.06**	**52.96**	**96.15**	**97.50**
	UnS **(%)**	**4.94**	**47.04**	**3.85**	**2.50**
	OvS **(%)**	**11.18**	**0**	**8.56**	**5.73**
	InC **(%)**	**6.40**	**36.06**	**4.95**	**3.25**
2	C_1 (%)	**74.51**	**0**	**80.97**	**89.97**
	C_2 (%)	**90.57**	**100**	**92.83**	**94.95**
	UnS **(%)**	**9.43**	**0**	**7.17**	**5.05**
	OvS **(%)**	**25.49**	**100**	**19.03**	**10.03**
	InC **(%)**	**14.63**	**32.38**	**11.01**	**6.66**
3	C_1 (%)	**79.72**	**0.01**	**85.25**	**91.47**
	C_2 (%)	**91.72**	**99.99**	**93.83**	**96.61**
	UnS **(%)**	**8.28**	**0.01**	**6.17**	**3.39**
	OvS **(%)**	**20.28**	**99.99**	**14.75**	**8.53**
	InC **(%)**	**12.00**	**30.97**	**8.83**	**4.99**
4	C_1 (%)	**90.59**	**100**	**92.12**	**92.40**

	C_2 (%)	**97.10**	**68.53**	**98.02**	**99.35**
	***UnS* (%)**	**2.90**	**31.47**	**1.98**	**0.65**
	***OvS* (%)**	**9.41**	**0**	**7.88**	**7.60**
	***InC* (%)**	**3.77**	**27.28**	**2.77**	**1.58**
Average	**C_1 (%)**	**83.41**	**50.00**	**87.45**	**92.03**
	C_2 (%)	**93.61**	**80.37**	**95.21**	**97.10**
	***UnS* (%)**	**6.39**	**19.63**	**4.79**	**2.90**
	***OvS* (%)**	**16.59**	**50.00**	**12.56**	**7.973**
	***InC* (%)**	**9.2**	**31.67**	**6.89**	**4.12**

The chart in Fig.6.19 further illustrates the capability of IFCM in segmentation.

As a result of the above evaluation it was decided to add two types of noise were added onto the original Lena image. The first one was the uniformly distributed random noise in interval (0, 100) and the second type was Gaussian distributed random noise also in interval (0, 100). Their average segmentation performance is listed in Table 6.5.

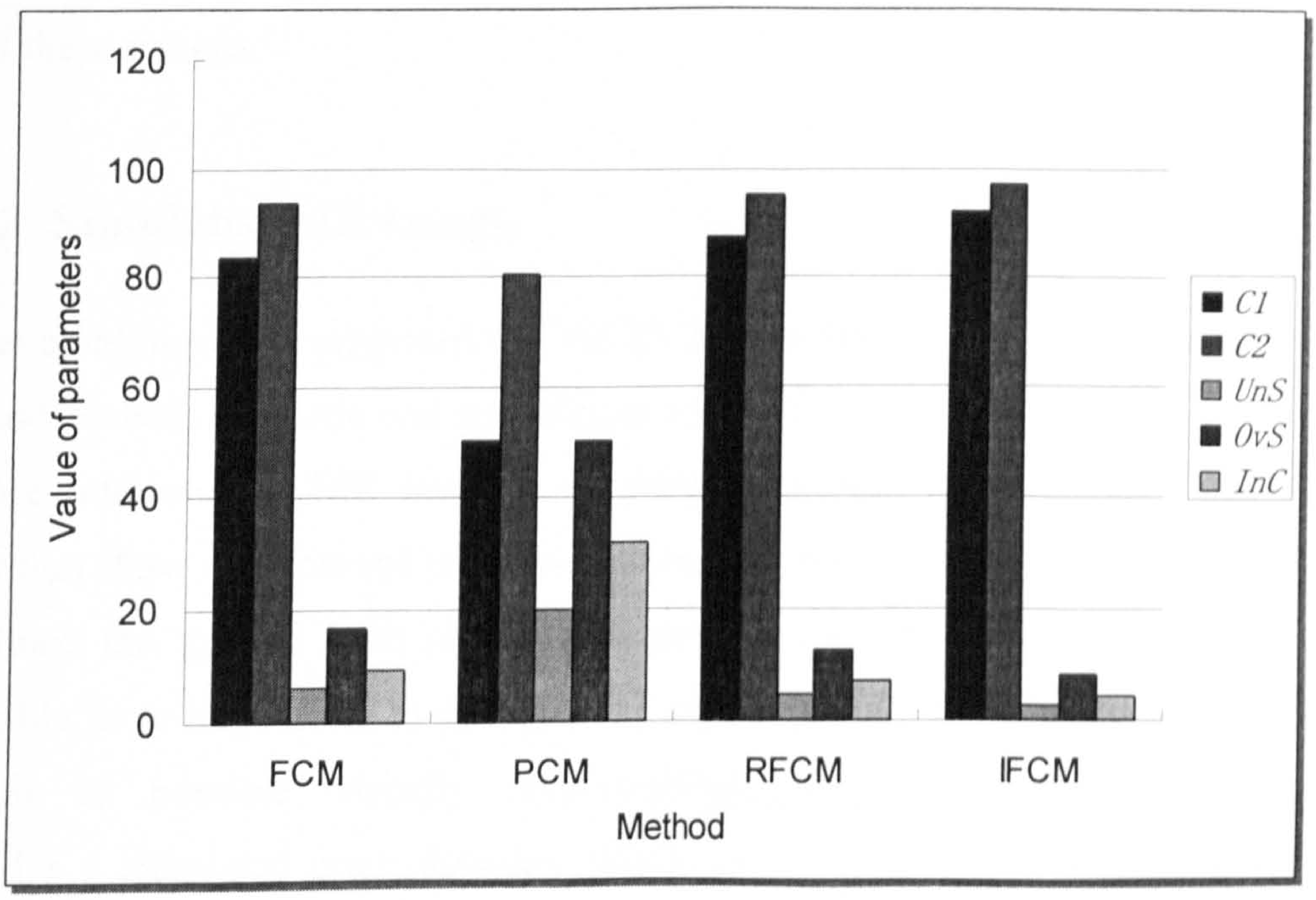


Figure 6.19 Chart of evaluation parameters of noisy Lena image.

Table 6.5 Evaluation parameters of Lena image with two types of noise.

Noise Type	Evaluation Parameters	FCM	PCM	RFCM	IFCM
Uniformly distributed noise	C_1 (%)	63.62	50	66.6925	92.1625
	C_2 (%)	86.41	80.5025	87.4225	97.07
	UnS (%)	13.59	19.4975	12.5775	2.93
	OvS (%)	36.38	50	18.0175	6.8375
	InC (%)	19.435	31.67	87.4225	2.10
Gaussian distributed noise	C_1 (%)	52.51	46.70	53.58	66.01
	C_2 (%)	82.855	79.78	83.19	87.75
	UnS (%)	17.15	20.22	16.81	12.25
	OvS (%)	47.49	53.31	46.42	34.00
	InC (%)	24.64	32.92	24.17	17.37

The parameters indicate Gaussian distributed noise affected segmentation more seriously than uniformly distributed one. However, IFCM conferred the best results in all these images.

6.6.3 Simulated MR Image

The evaluation above suggested that the IFCM algorithm was the more accomplished method on both synthetic and real images with different level of noise. However, as we are interested in MR images, we need to answer the question, 'Can IFCM algorithm show an improved performance on MR images?'.

Since the ground truth of segmentation for real MR images is not usually available, it is impossible to evaluate the segmentation performance quantitatively, but it is possible visually. Brainweb(http://www.bic.mni.mcgill.ca/brainweb) provides a simulated brain database (SBD) including a set of realistic MRI data volumes produced by an MRI simulator. These data enable us to evaluate the performance of various image analysis methods in a setting where the truth is known (The precise regions of white matter, gray matter, CSF) [Brainweb, 2003; Kwan, 1999; Collins, 1998; Cocosco, 1997].

Firstly, a simulated T_1-weighted MR image was downloaded from Brainweb. It is a 3D image with dimensions 181x217x181 (XxYxZ). Fig.6.20 shows several (Z axis) slices of the simulated image.

The discrete anatomical model of the simulated image consists of background, CSF, gray matter, white matter, fat, muscle, skin, skull, glial matter and connective. In this thesis, only CSF, white matter and gray matter are of interest and the anatomical model is shown in Fig.6.21.

7% noise was applied to each slice of the simulated image, and thereafter through the non-brain region process (The percentage of noise is the standard deviation of noise versus that of the signal of the brightest tissue). The brain region only and noisy 100-th slice is shown in Fig.6.22. The noisy slices were segmented

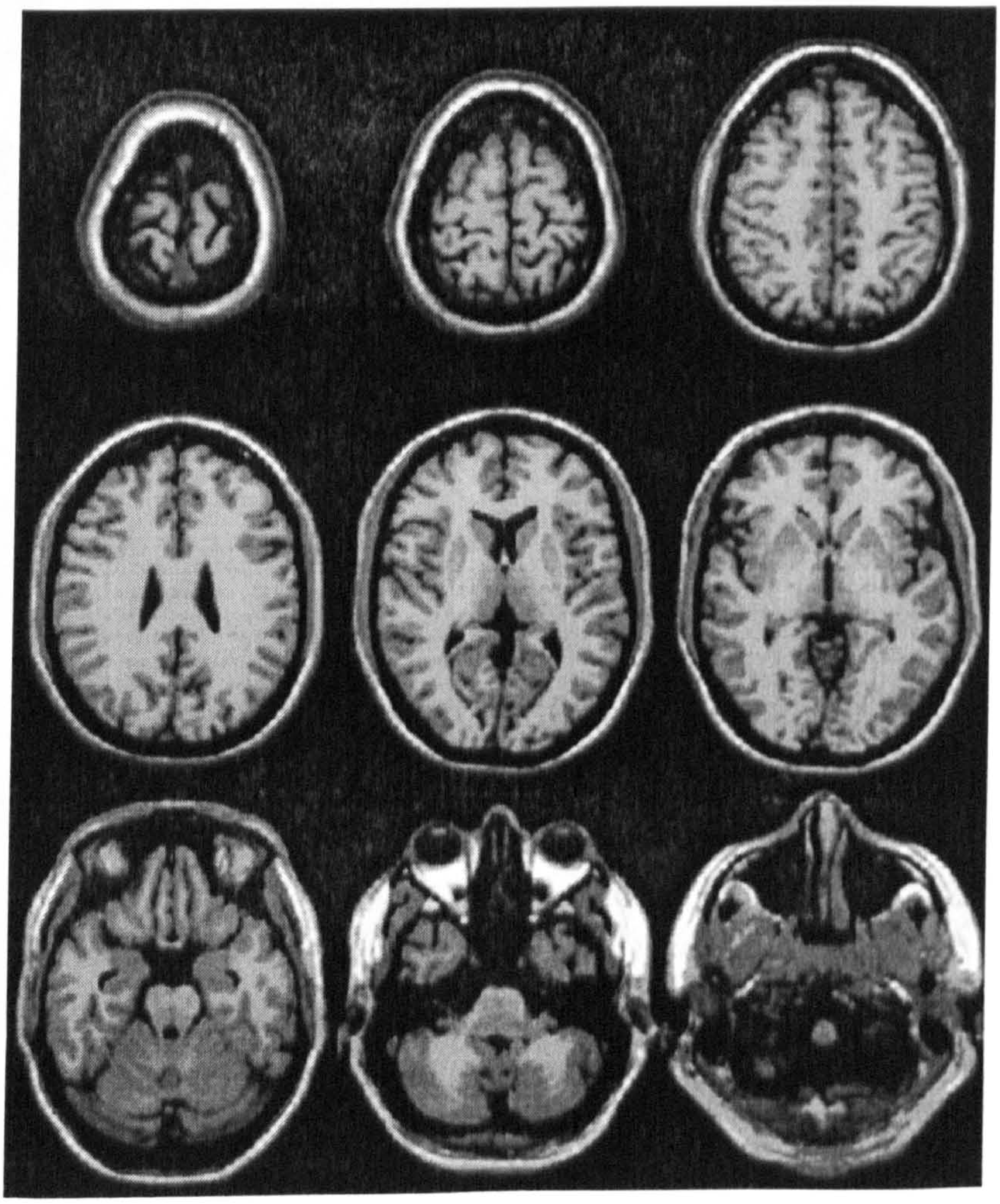


Figure 6.20 Simulated T_1-weighted MR image (from left to right, top to bottom) including slice 150, 140, 120, 100, 80, 70, 50 30 10.

into four clusters, background, CSF, white matter and gray matter using FCM, IFCM, PCM and RFCM respectively. Background was neglected from the viewing results. Fig.6.23 (a) plots the tendency of the cost function in IFCM based on the various values of λ and ξ. The cost function was optimized at λ=0.7257 and ξ=0.4055 where the learning rate was η=0.2. Fig.6.23 (b) shows a plot of the cross-validation error computed for various value of β applying RFCM on the noisy slices and this demonstrated that β=29 minimized the error function.

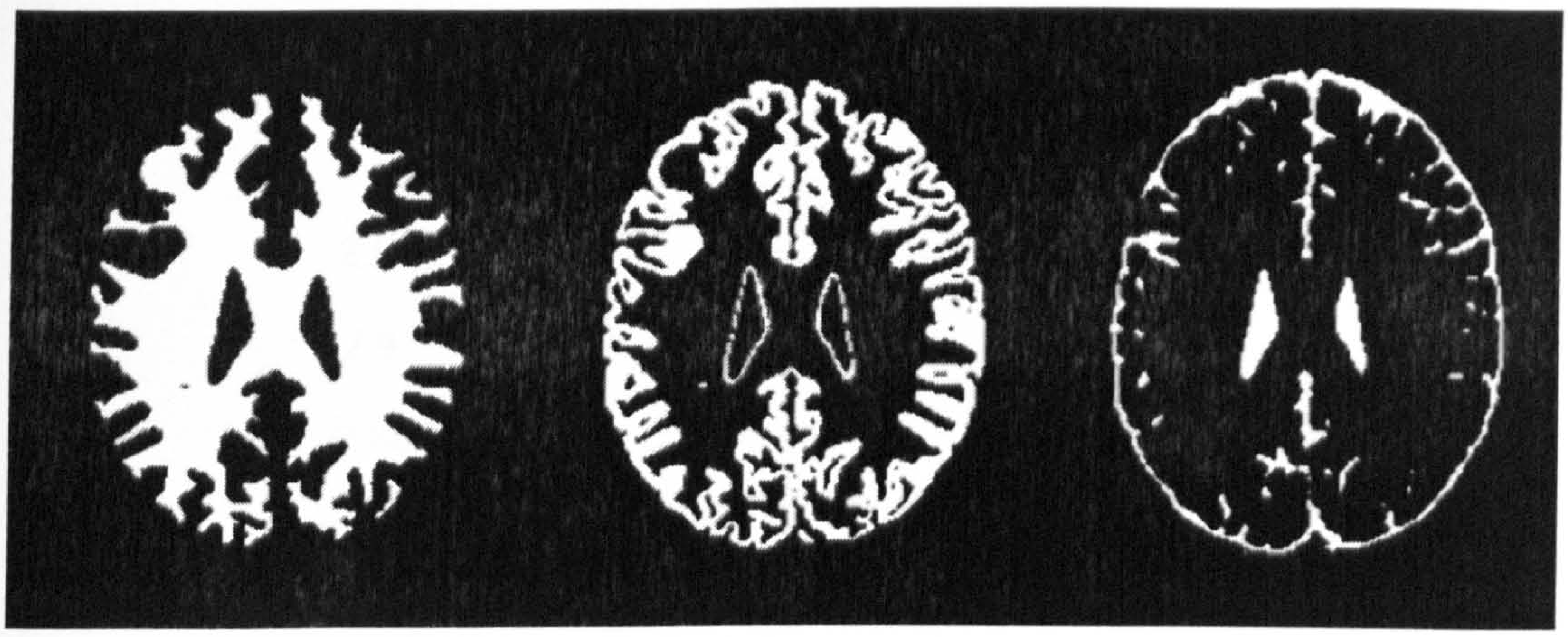


Figure 6.21 Discrete anatomical model.
(left) white matter (middle) gray matter (right) CSF

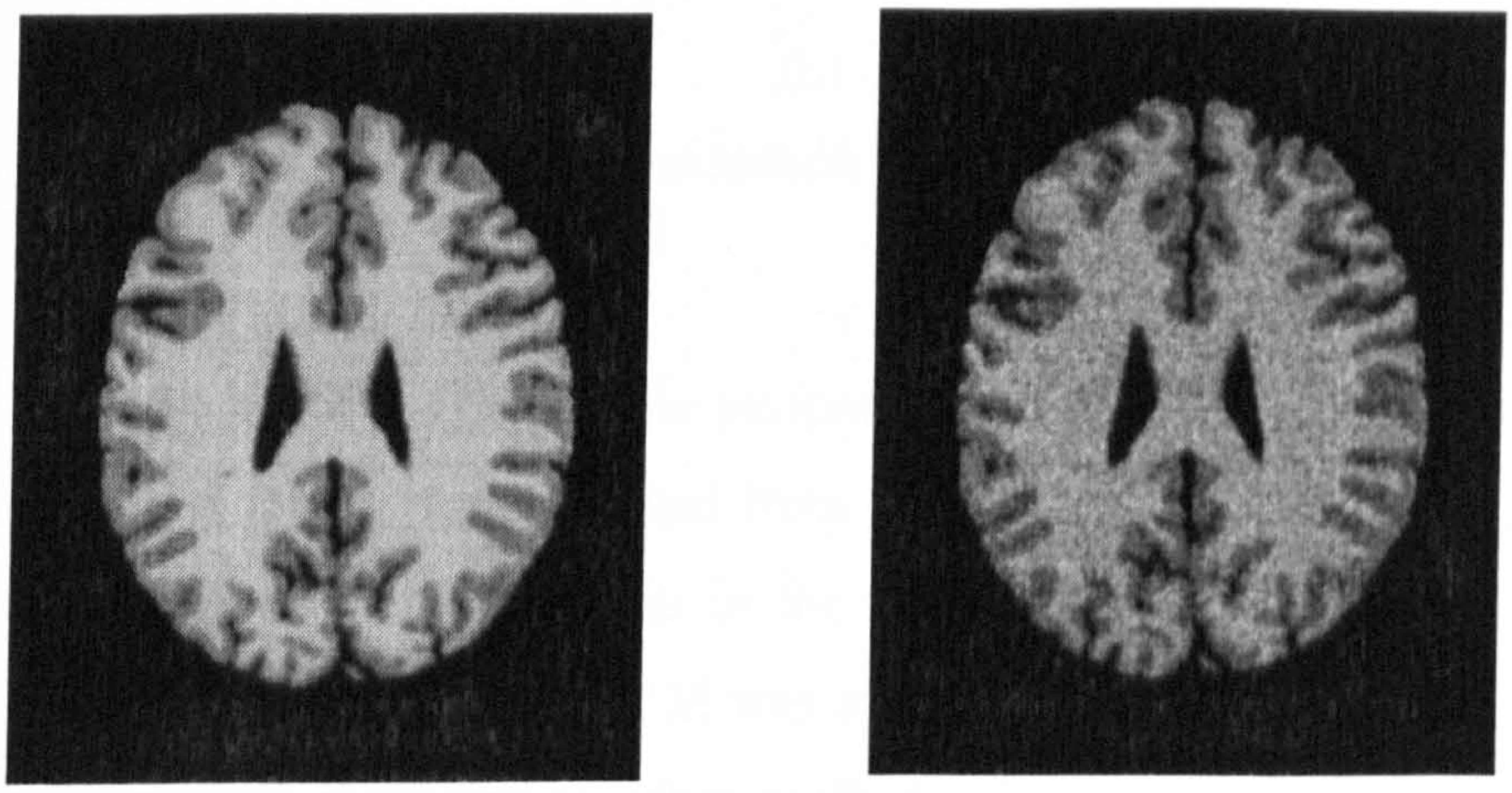
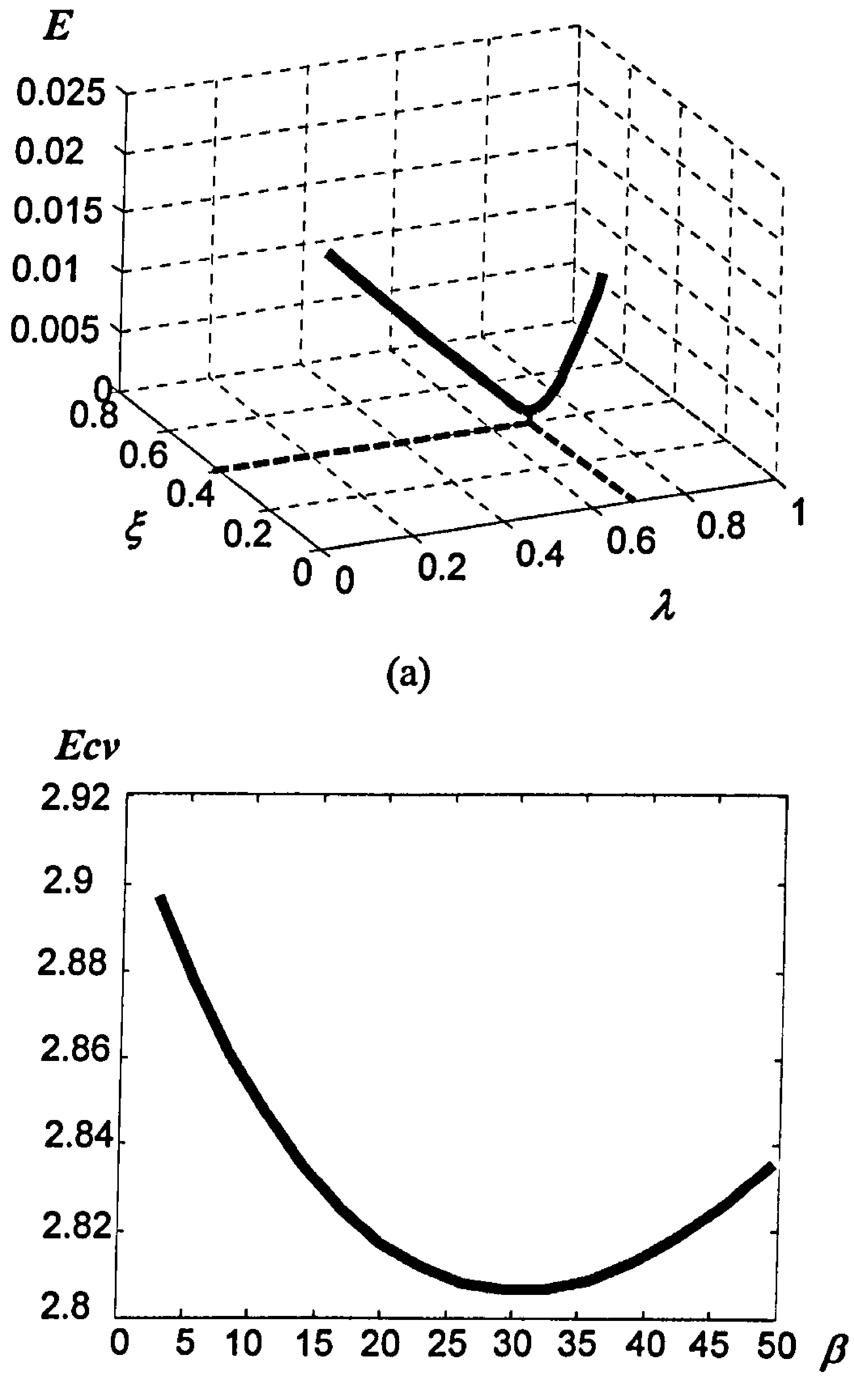


Figure 6.22 100-th slice of simulated T_1-weighted MR image.
(left) original slice (right) noisy slice with 7% noise

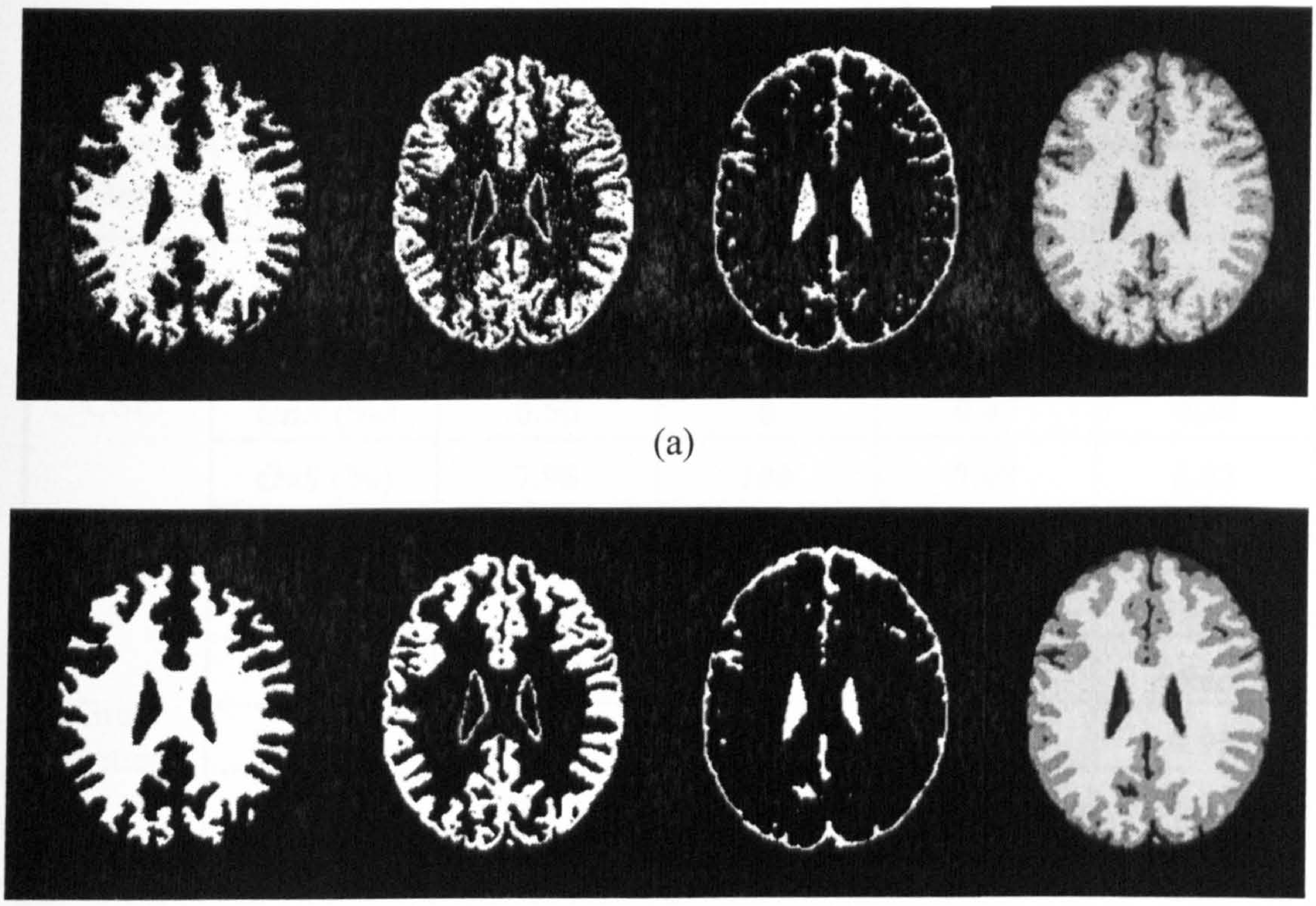


(a)

(b)

Figure 6.23 Parameter estimation for simulated T_1-weighted image. (a) IFCM (b) RFCM

Using the parameters selected, the performance evaluation parameters of FCM, RFCM, PCM and IFCM were obtained from the 100-th slice and they are listed in Table 6.6. Similar to the conclusions in the above sections, RFCM increased the accuracy of FCM marginally, but PCM was stuck in one cluster and IFCM showed the highest accuracy compared to other methods. The segmentation results of the noisy 100-th slice when applying RFCM and IFCM are shown in Fig.6.24. It demonstrated that although IFCM increased the accuracy by only several percent, it eliminated noise completely while RFCM could not overcome the influence of noise.



(a)

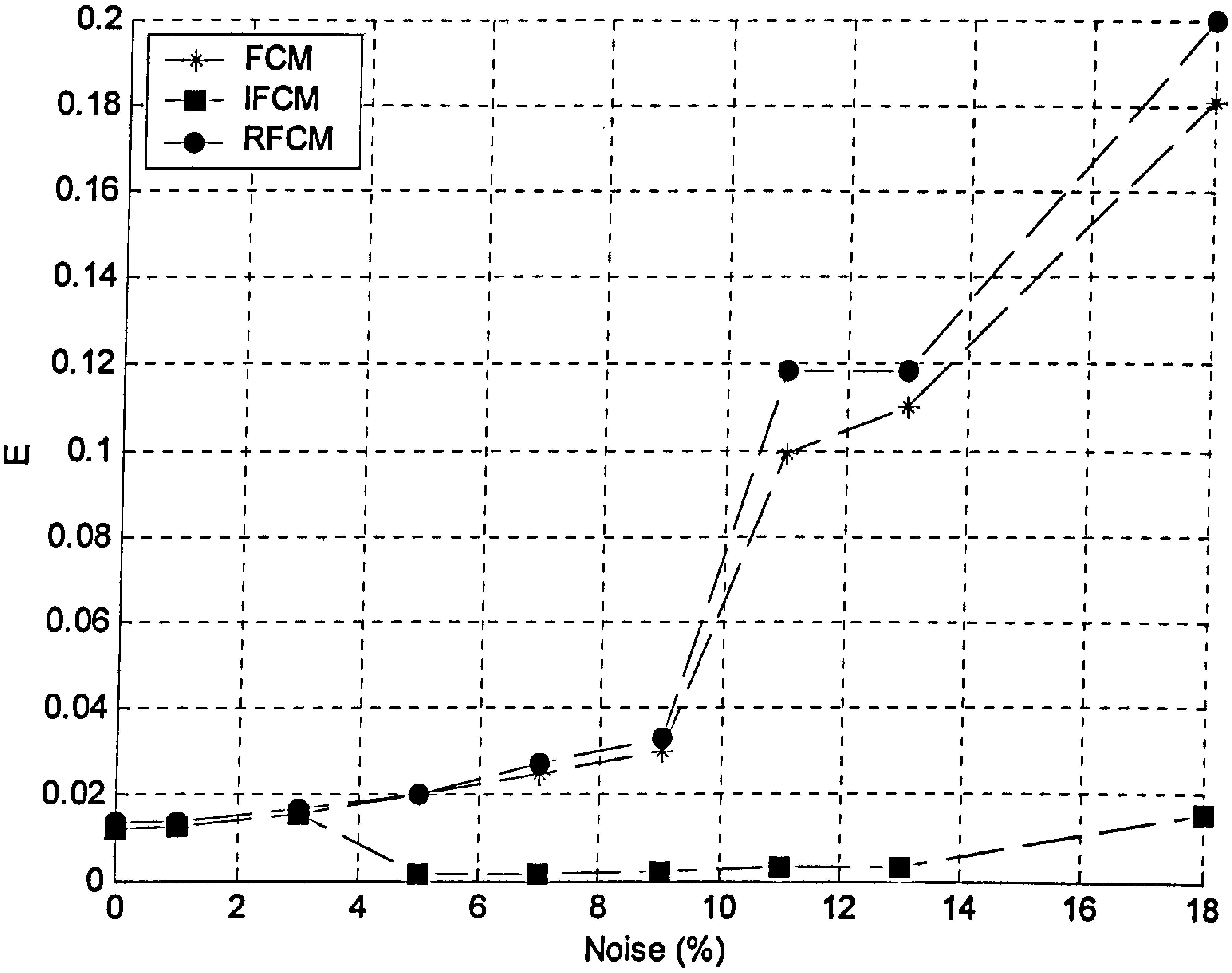
(b)

Figure 6.24 Segmentation results of simulated T_1-weighted MR image.
(a) RFCM (b) IFCM

Table 6.6 Segmentation performance evaluation on simulated T_1-weighted image.

Class	Evaluation Parameters	FCM	PCM	RFCM	IFCM
CSF	C_1 (%)	92.02	0	92.02	93.18
	C_2 (%)	99.50	100	99.53	99.80
	UnS (%)	0.50	0	0.47	0.20
	OvS (%)	7.98	100	7.98	6.82
	InC (%)	0.76	34.0	0.73	0.57
Gray matter	C_1 (%)	88.92	0	89.08	92.69
	C_2 (%)	98.65	100	98.89	99.05
	UnS (%)	1.35	0	1.11	0.95
	OvS (%)	11.08	100	10.92	7.31
	InC (%)	2.33	10.16	2.11	1.59
White Matter	C_1 (%)	92.77	100	94.28	97.35
	C_2 (%)	99.25	84.14	99.24	99.36
	UnS (%)	0.75	15.86	0.76	0.54
	OvS (%)	7.23	0	5.72	2.65
	InC (%)	1.68	13.57	1.47	0.93
Average	C_1 (%)	91.24	33.33	91.79	94.41
	C_2 (%)	99.13	94.71	99.22	99.40
	UnS (%)	0.87	5.29	0.78	0.56
	OvS (%)	8.76	66.67	8.21	5.59
	InC (%)	1.59	19.24	1.44	1.03

Different levels (0-18%) of noise were sequentially applied to the simulated T_1-weigthed MR image. The noisy images were segmented by means of all methods described. Fig.6.25 show their cost function *E* and *InC* obtained from FCM, RFCM and IFCM. An increase in the level of noise led to an increase of *E* and *InC* for all methods. Below the 3% noise level, the three methods had similar performance describing by *E* and *InC*. Above the 3% noise level, IFCM showed the most robustness to noise. It reduced *InC* greatly within the noise level 7% to 15%, and



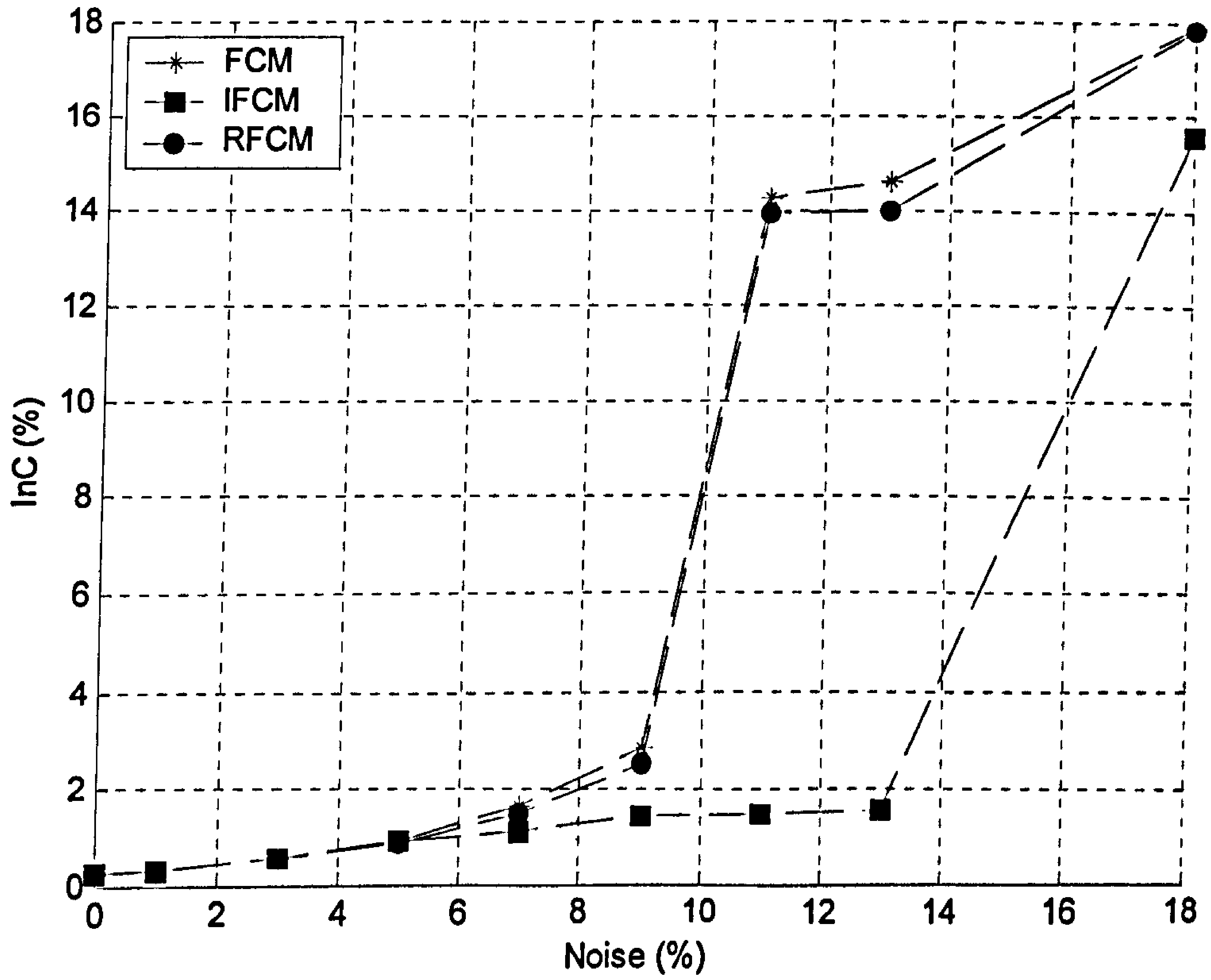
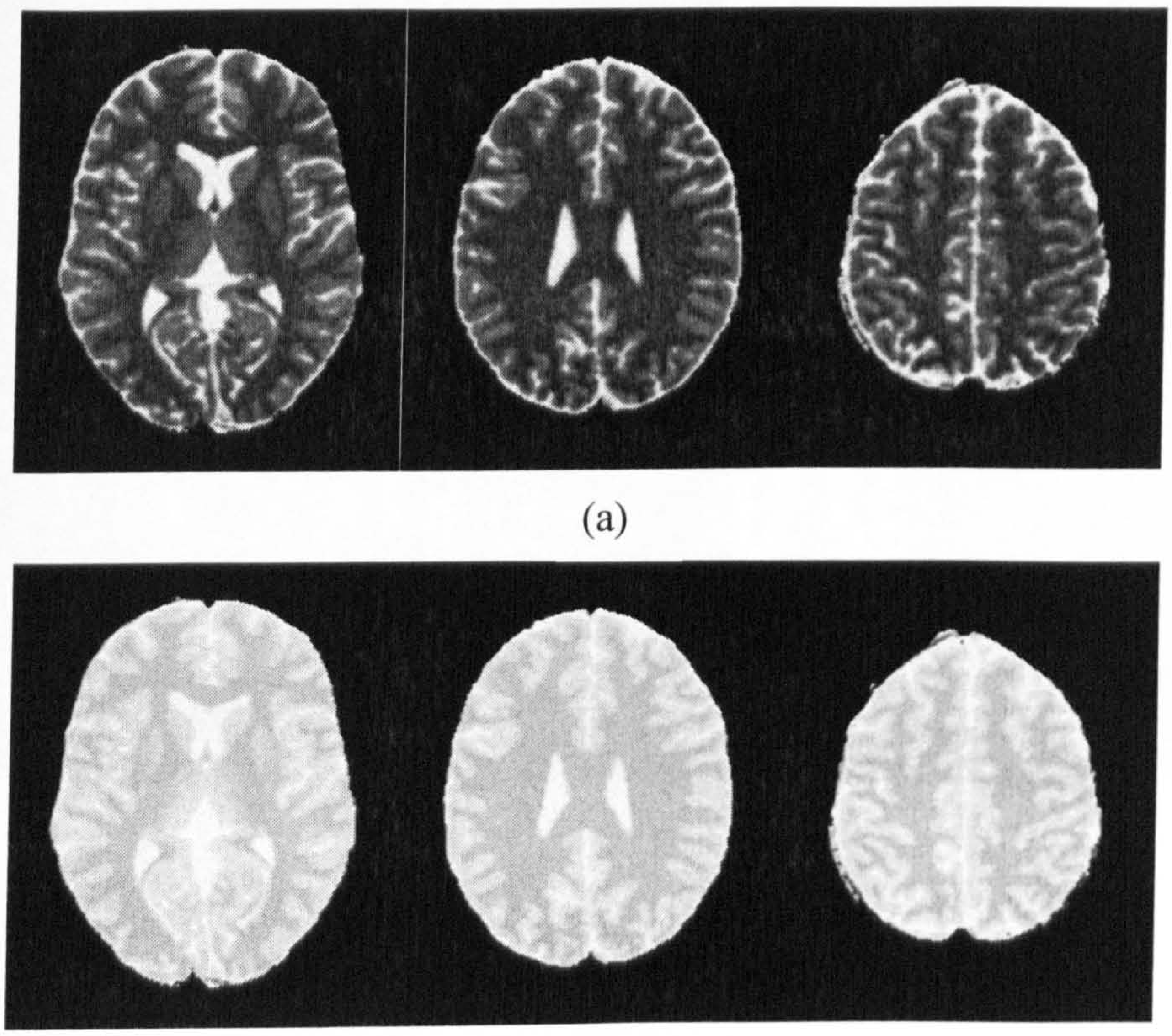


Figure 6.25 Cost function *E* (top) and *InC* (bottom) with respect to different noise levels in T_1-weighted MR image

even with other noise levels, it obtained a much lower *E* which reflected a more dominant segmentation. RFCM and FCM results were close; although RFCM had a lower *InC*, it was less positive in segmentation.

Two major modalities of MRI, T_2-weighted and PD-weighted shown in Fig.6.26 were also used to validate the advantages of IFCM on segmentation. 3% noise was added on the original images and they were segmented by both FCM and IFCM. Fig.6.27 and Fig.6.28 show the segmentation results. The influence of the noise on both images was overcome by IFCM, but FCM could not even segment the PD-weighted image correctly.

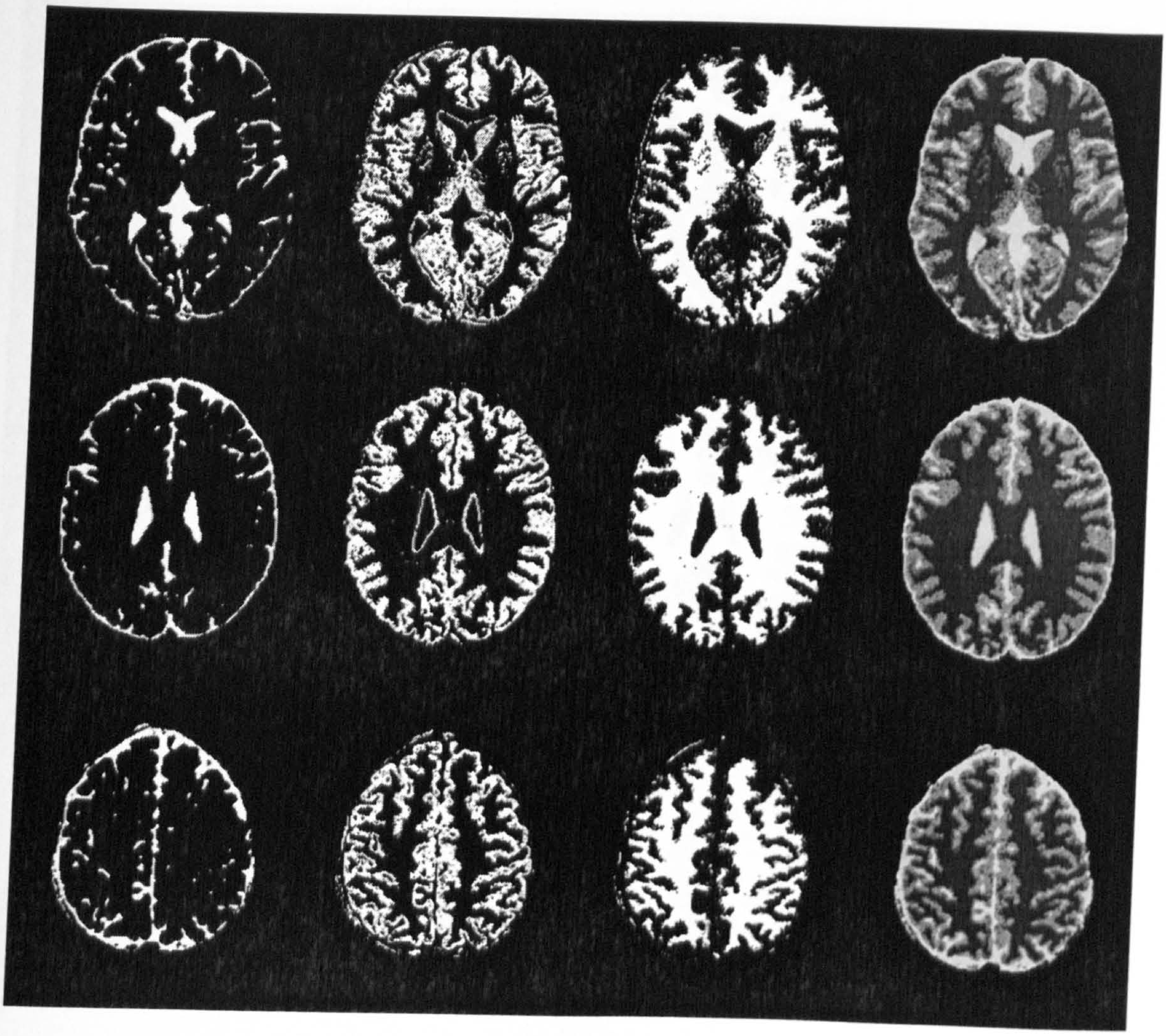
Fig.6.29 shows the chart of average evaluation parameters for different MRI modalities all with 3% noise. This further demonstrated the advantage of IFCM.



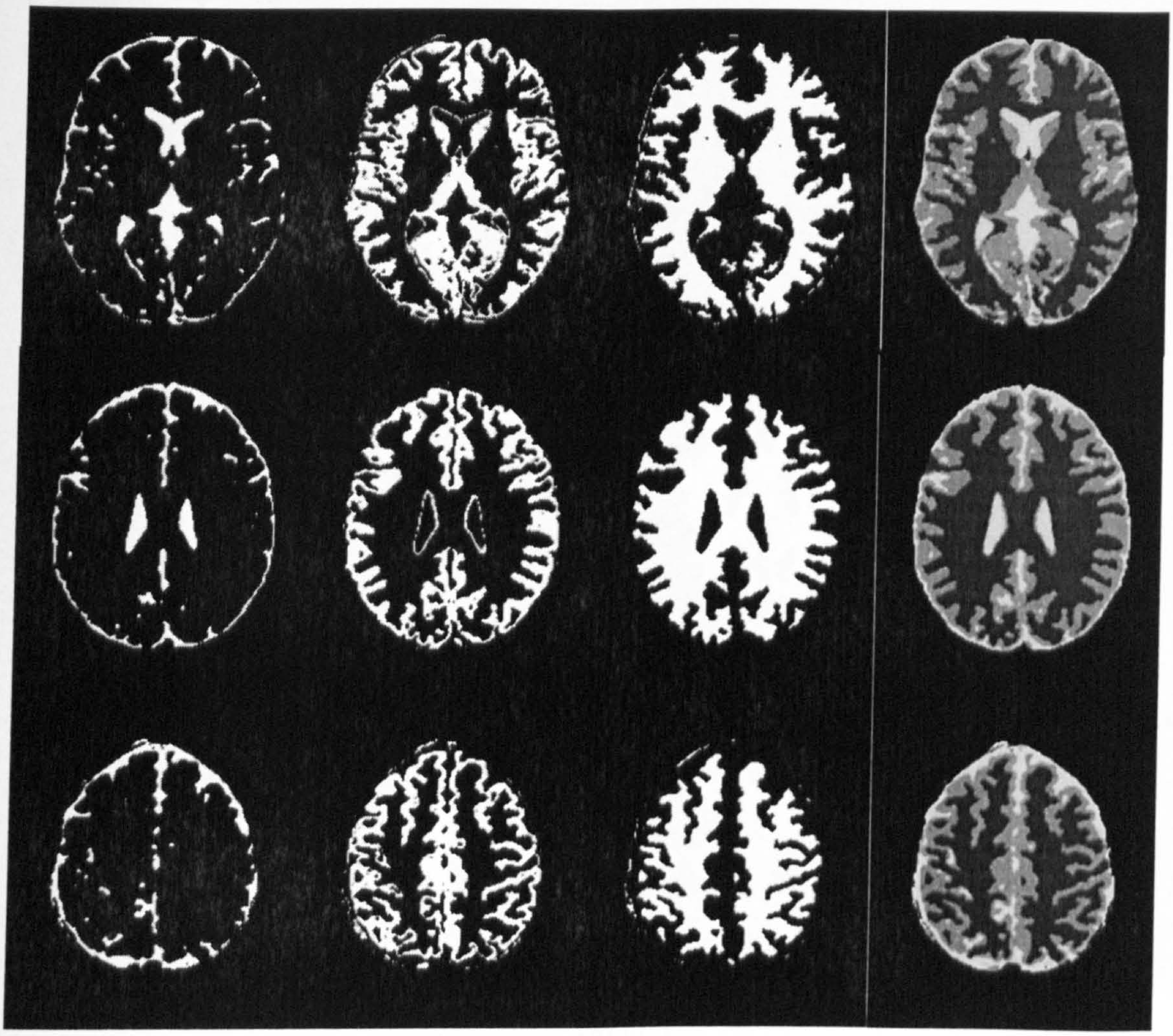
(a)

(b)

Figure 6.26 Simulated MR images with 3% noise: the 80-th, 100-th and 120-th slice (from left to right) (a) T_2-weighted (b) PD-weighted.

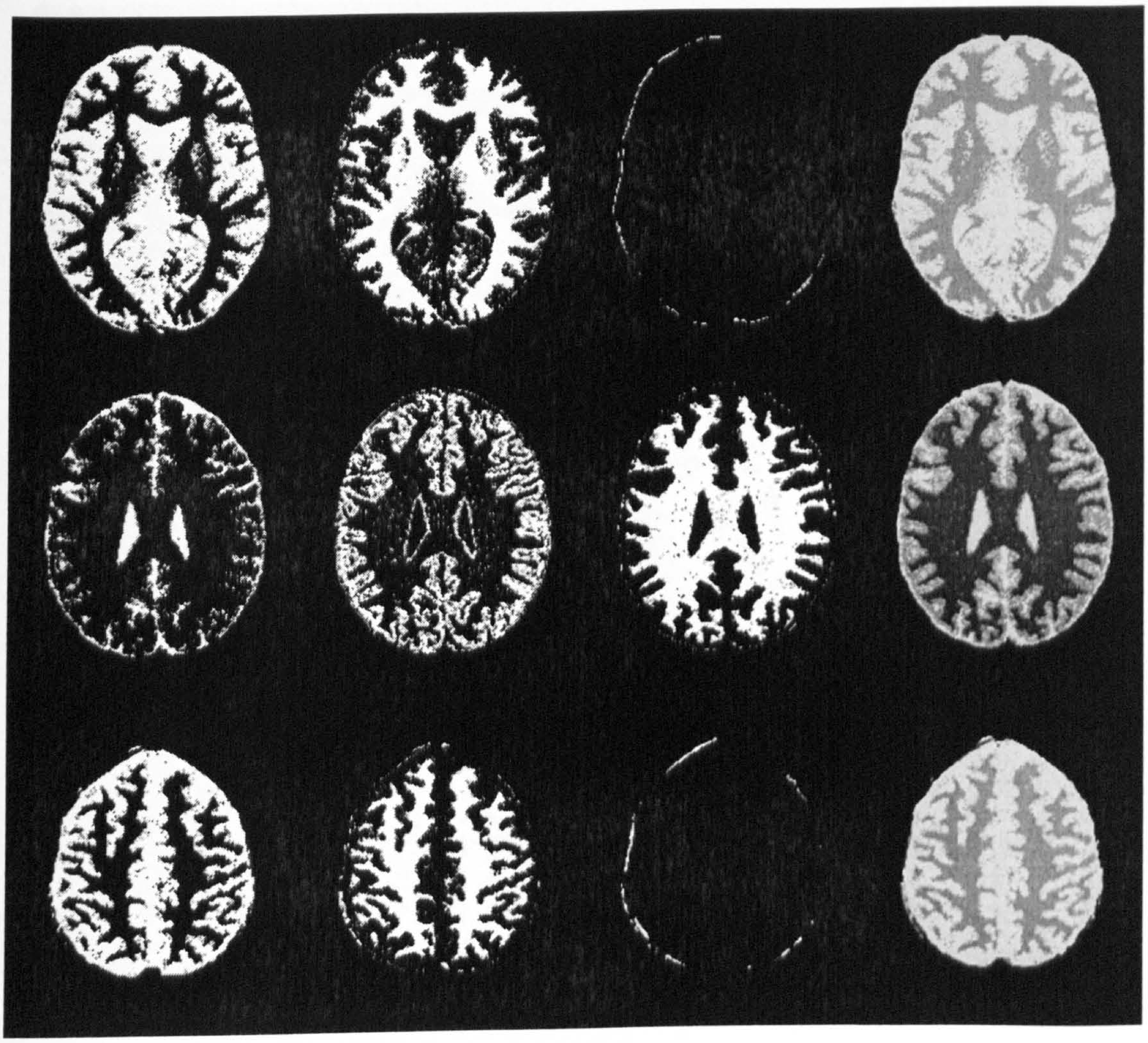


(a) FCM

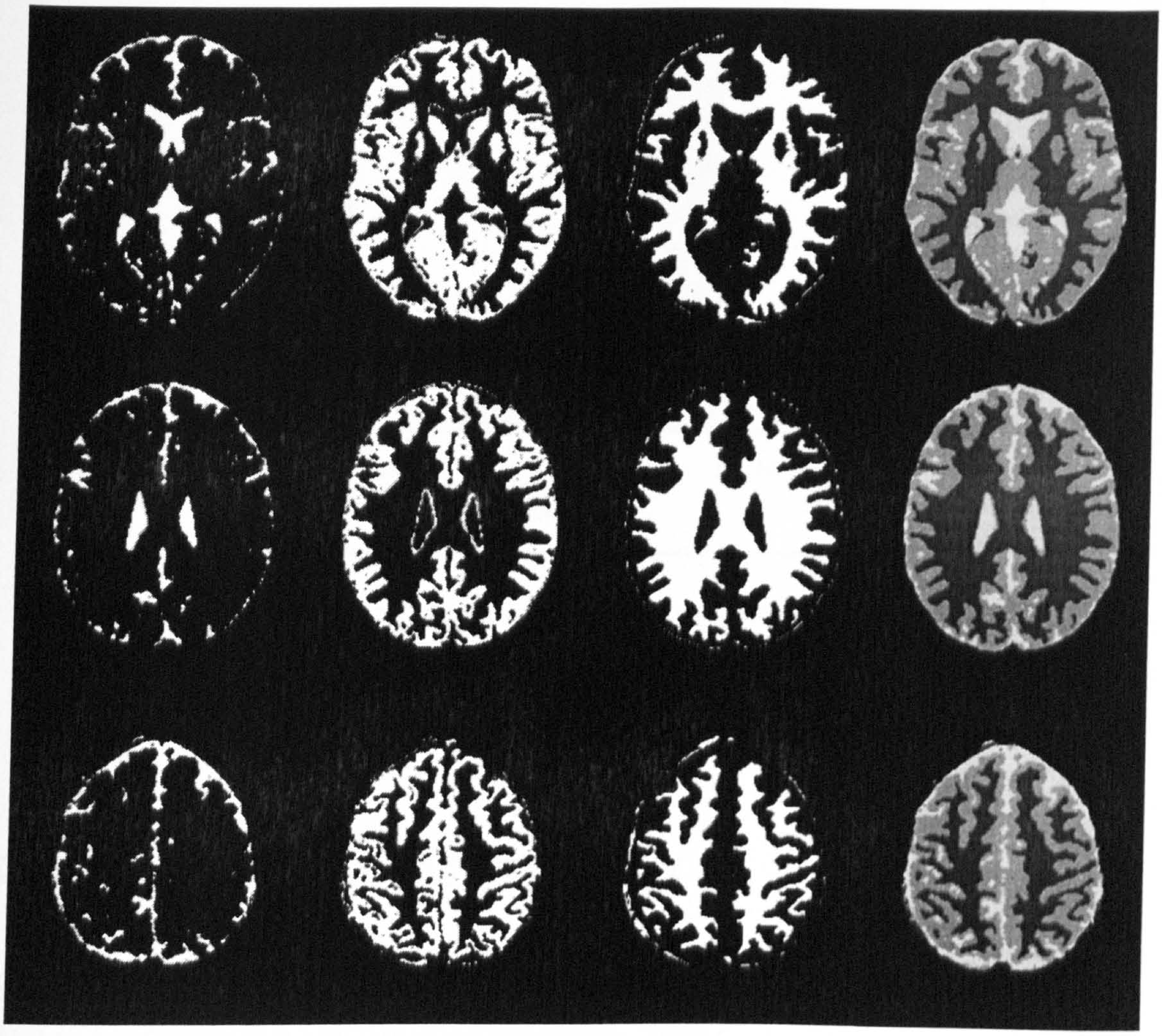


(b) IFCM

Figure 6.27 Segmentation results of T$_2$-weighted image.
The first row: the 80-th slice, the second row: the 100-th slice and the third row: the 120-th slice



(a) FCM



(b) IFCM

Figure 6.28 Segmentation results of PD-weighted image.
The first row: the 80th slice, the second row: the 100th slice and the third row: the 120th slice

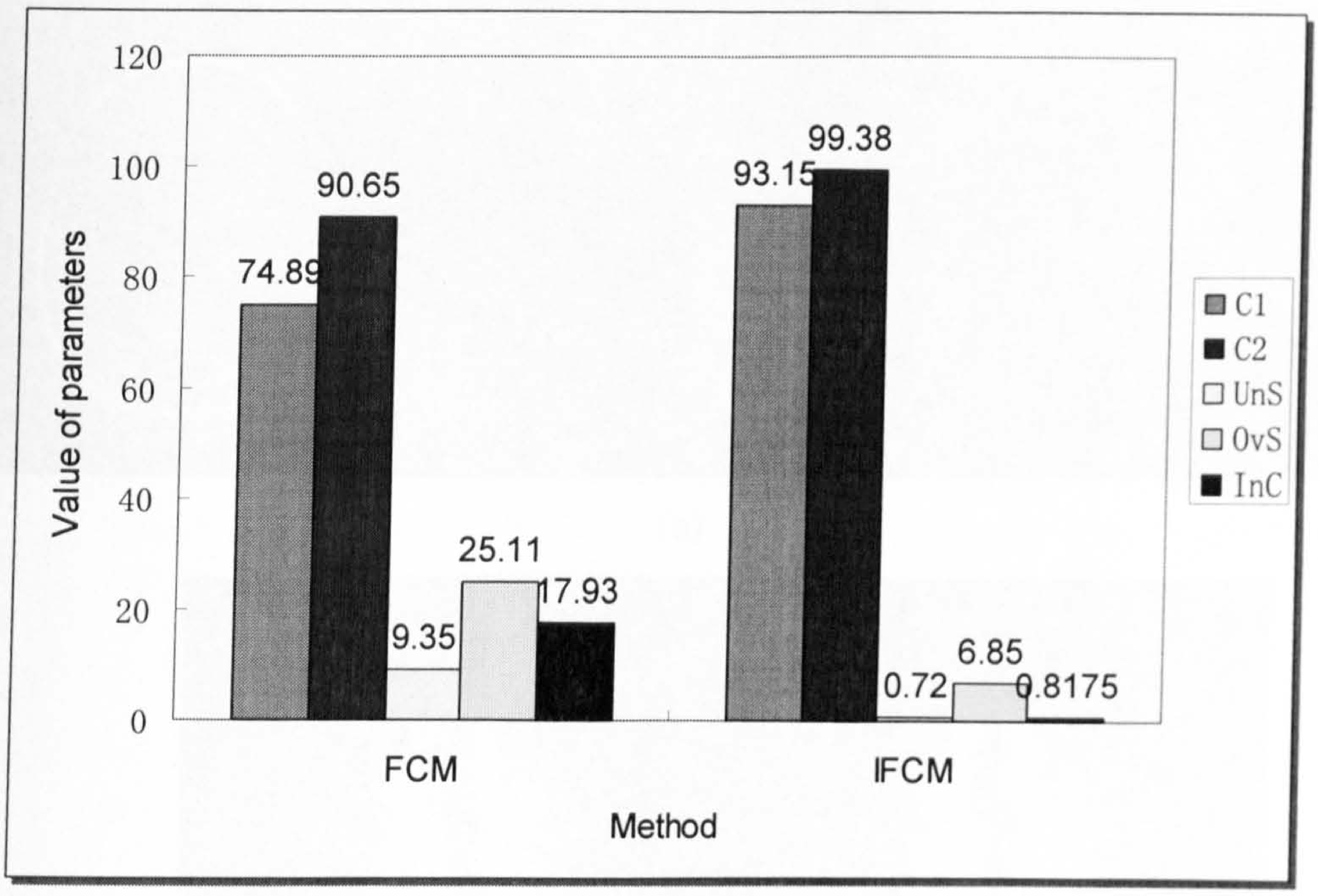
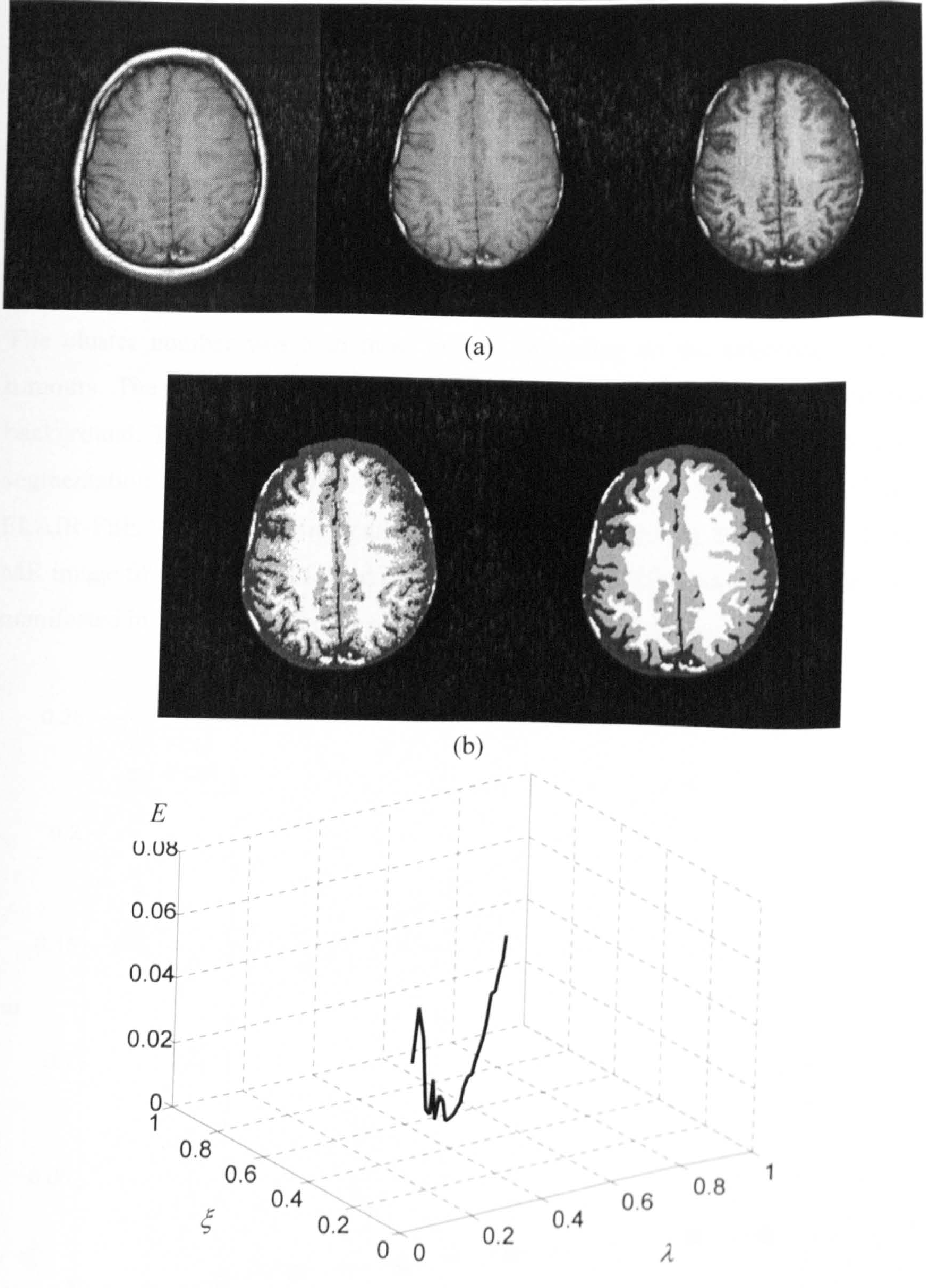


Figure 6.29 Chart of average evaluation parameters of MR images with 3% noise.

6.7 MR IMAGE SEGMENTATION

Through a series of tests on various synthetic, facial and simulated MR images, it maybe confirmed beyond all doubt that the IFCM algorithm is superior at segmentation. In this section, real MR images were tested using FCM and IFCM. The performance of IFCM was of most interest to us since our initial target was to develop an algorithm adaptive to real MR images for this thesis.

An example of a T_1-weighted MR image with 11% noise is shown in Fig.6.30. Along with the comparison of segmentation results from FCM and IFCM, the process of non-brain region removal and tissue contrast enhancement are also given (Fig.6.30 (a)). The parameters λ and ξ in IFCM were chosen as 0.5303 and 0.5707 by the ANN model. Clearly, IFCM acquired more smooth results than FCM.



(a)

(b)

(c)

Figure 6.30 T_1-weighted MR image with 11% noise. (a) (from left to right) original standardized image, brain region and enhanced image with 11 % noise (b) (from left to right) segmentation results of FCM and IFCM (c) parameter estimation

Several levels of noise were applied to the original image. As no ground truth was available, the cost function E was employed as a performance evaluation parameter; its capability has been explained above. Fig.6.31 is the plot of the cost function E with respect to the noise level. The value of E increased with the noise level. The much lower value of E implied a better performance by IFCM than FCM.

After the testing described above, we can be fairly certain that the IFCM algorithm can be applied to the real MR images that we have collected for our study. The cluster number was 5 in most images depending on the existence of brain tumours. The five clusters were white matter, gray matter, CSF, brain tumour and background. Typical images from each type of MRI modality in our study and their segmentation results using IFCM are shown in the following figures. Fig.6.32 is a FLAIR-FSE MR image from a patient having high level glioma. Fig.6.33 is a T_1-SE MR image from another patient having high level glioma. Brain tumour is not always manifested in T_1-SE image but has close intensities to white matter.

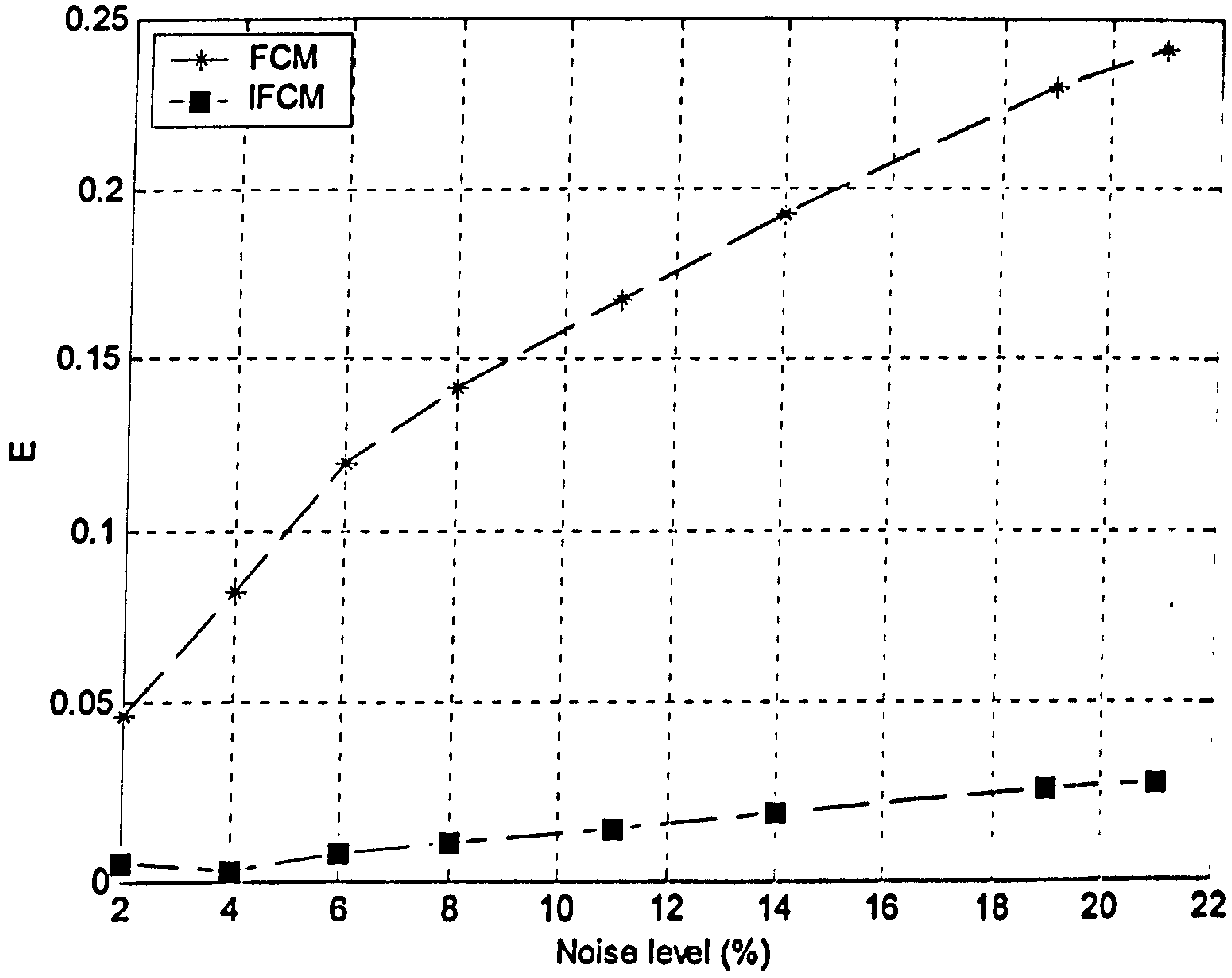
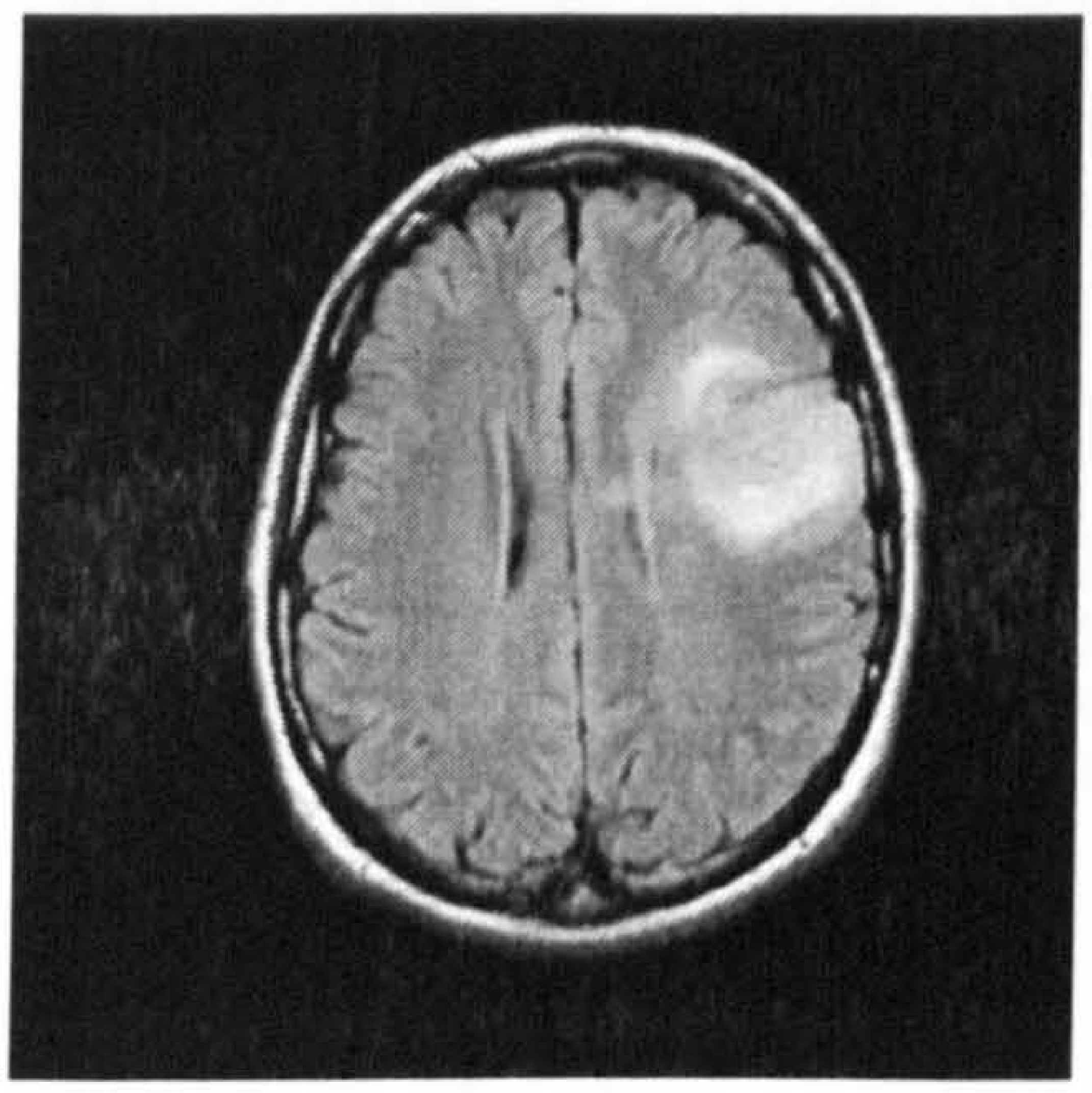
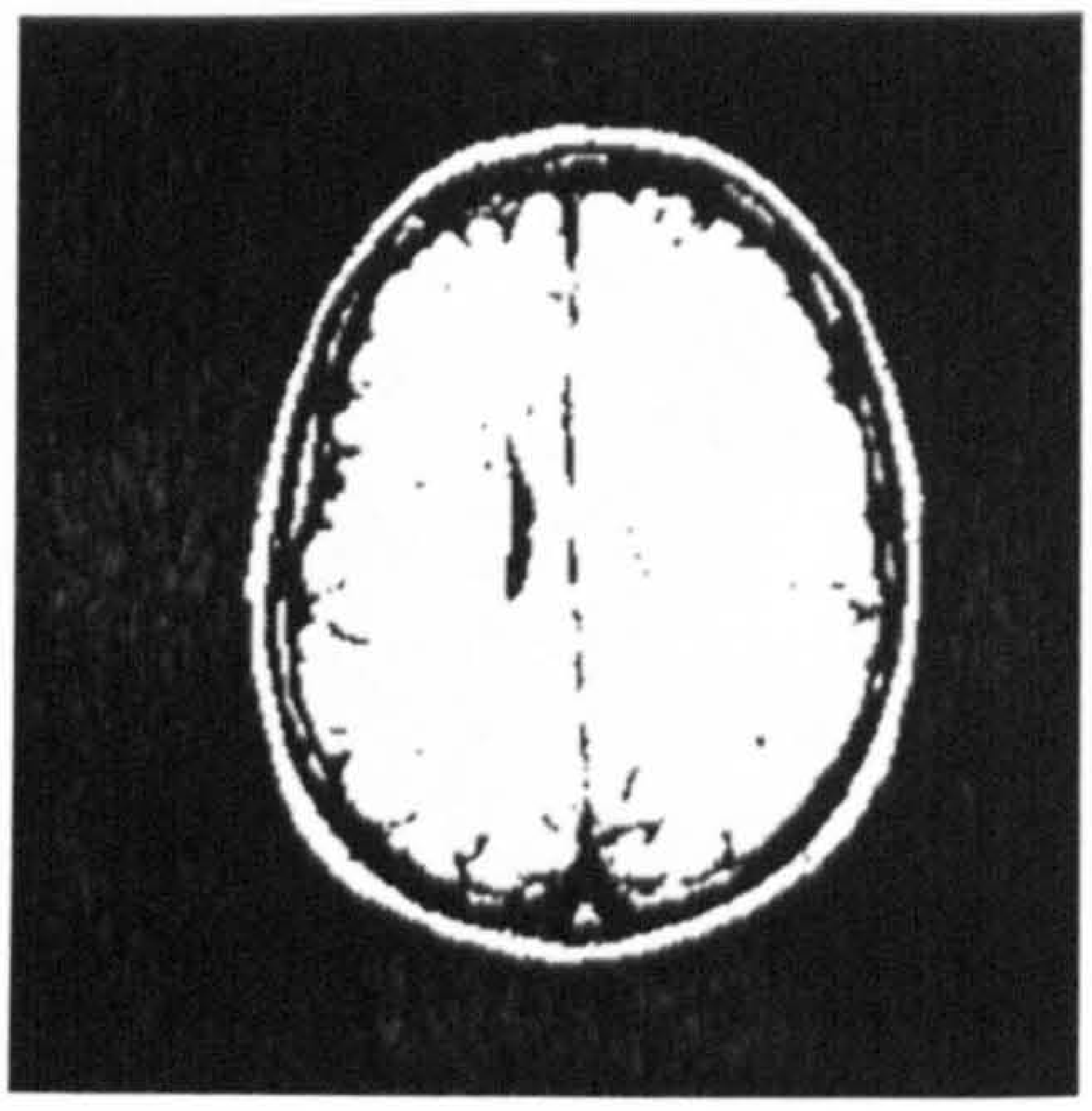


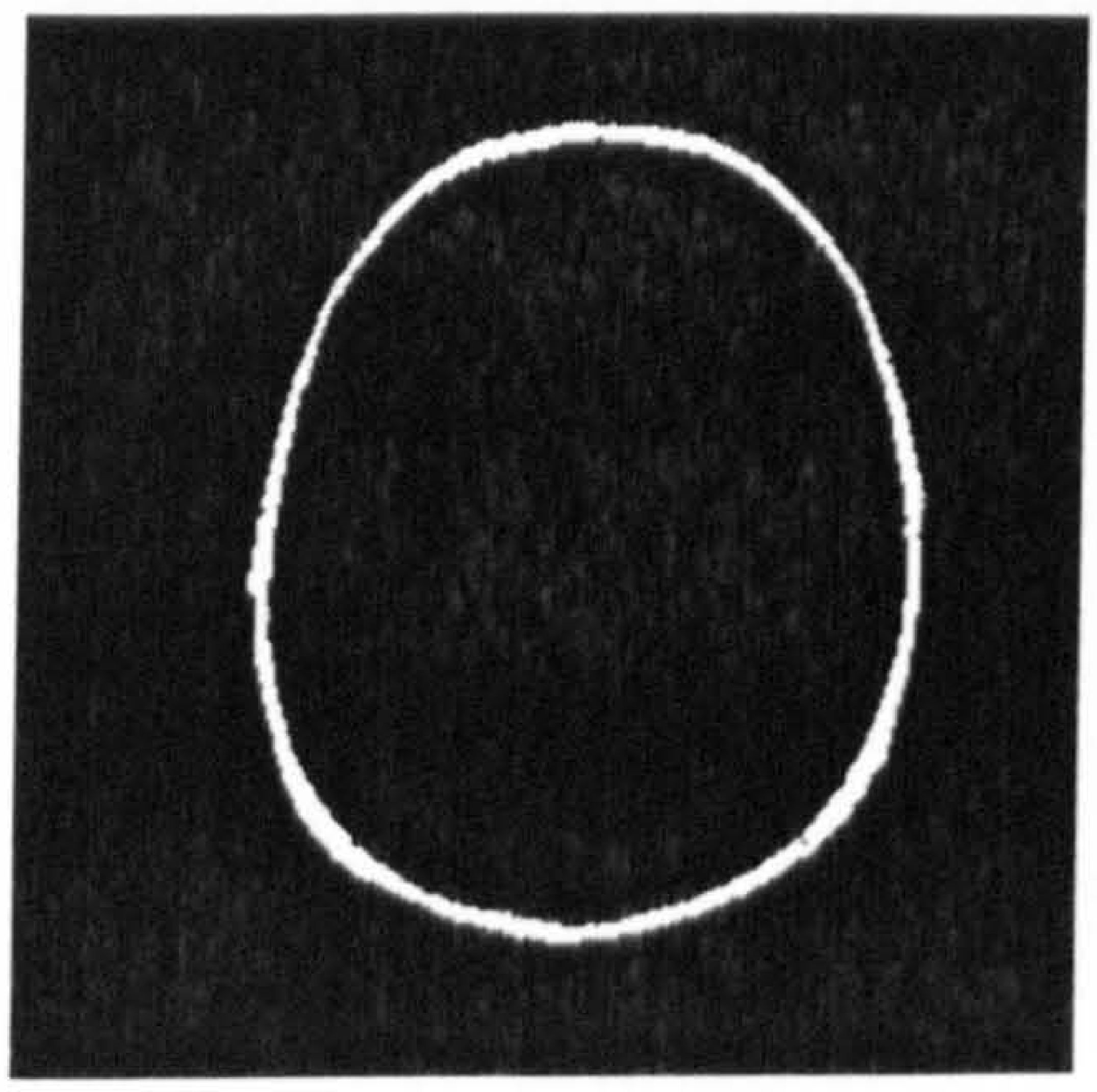
Figure 6.31 Plot of E with respect to noise level.



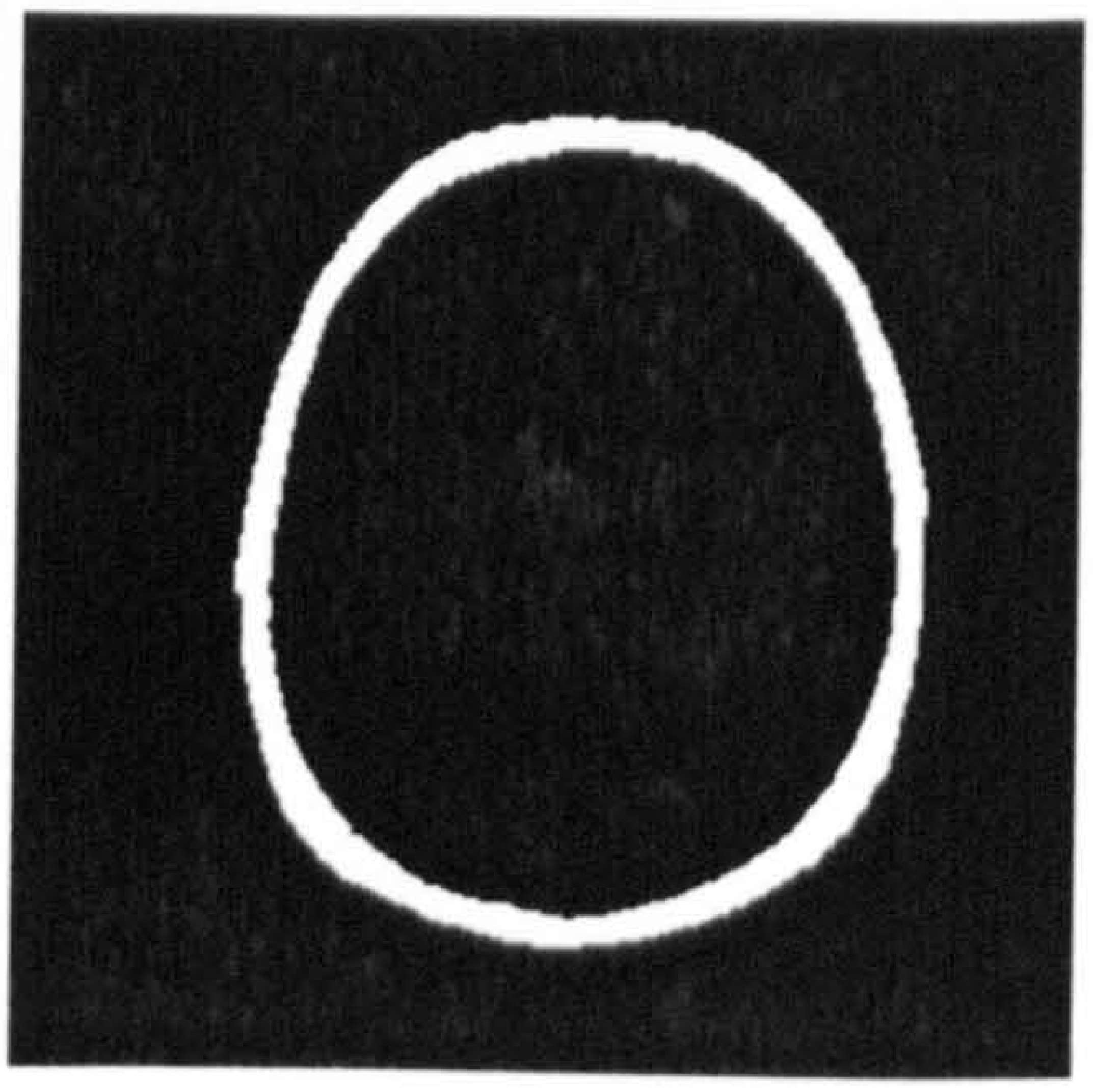
(a) standardized image



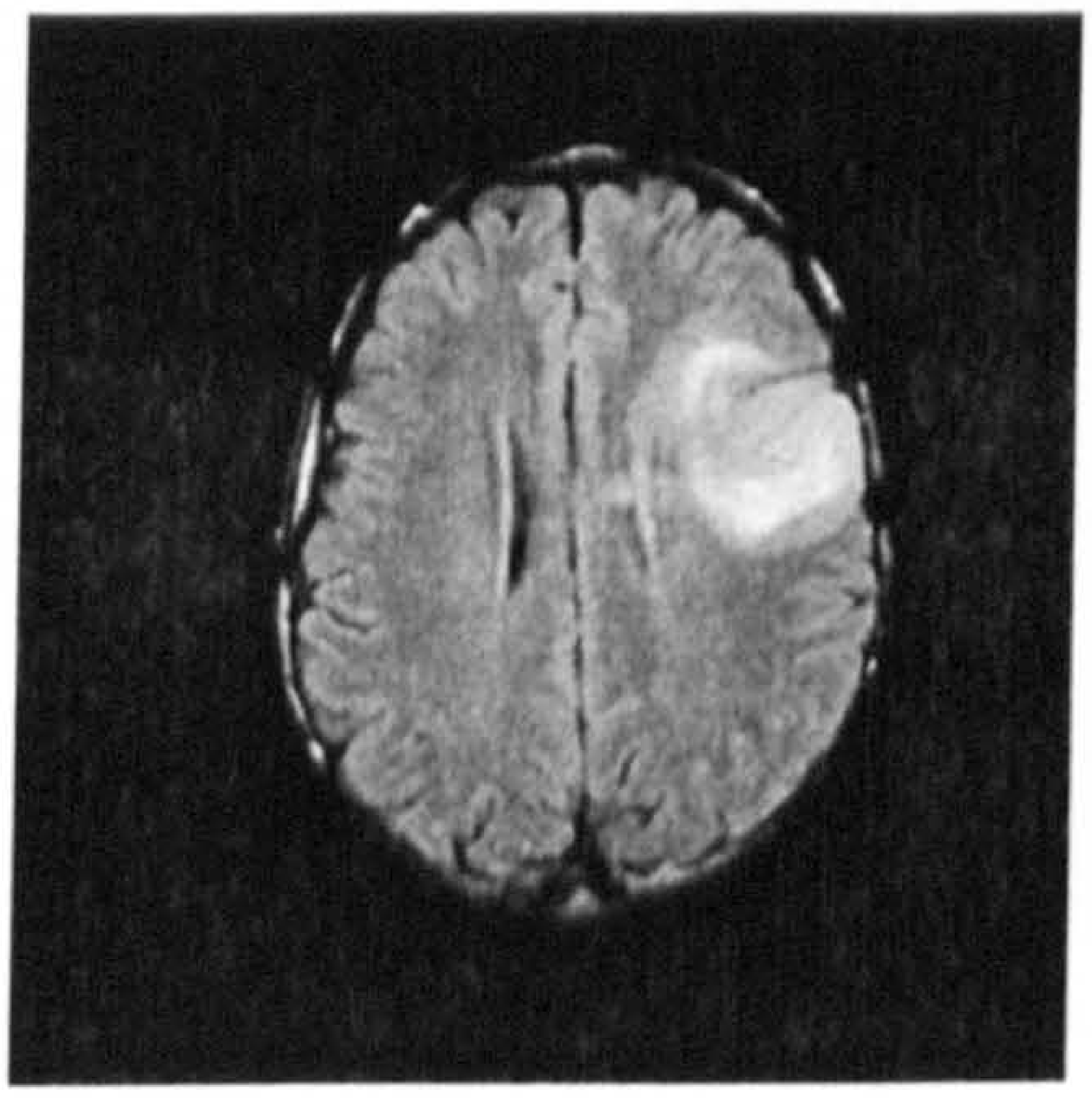
(b) binary mask



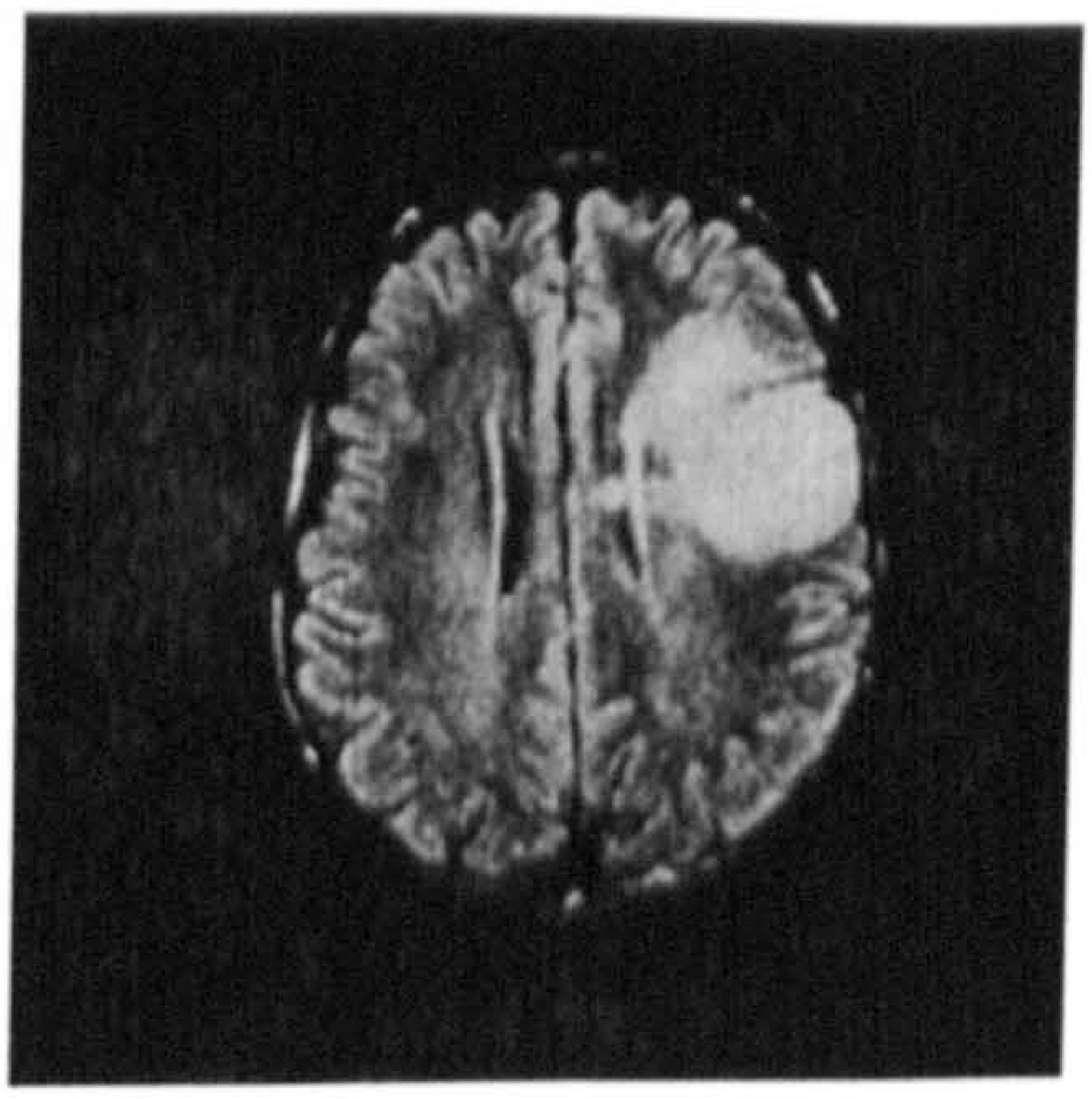
(c) non-brain region after erosion



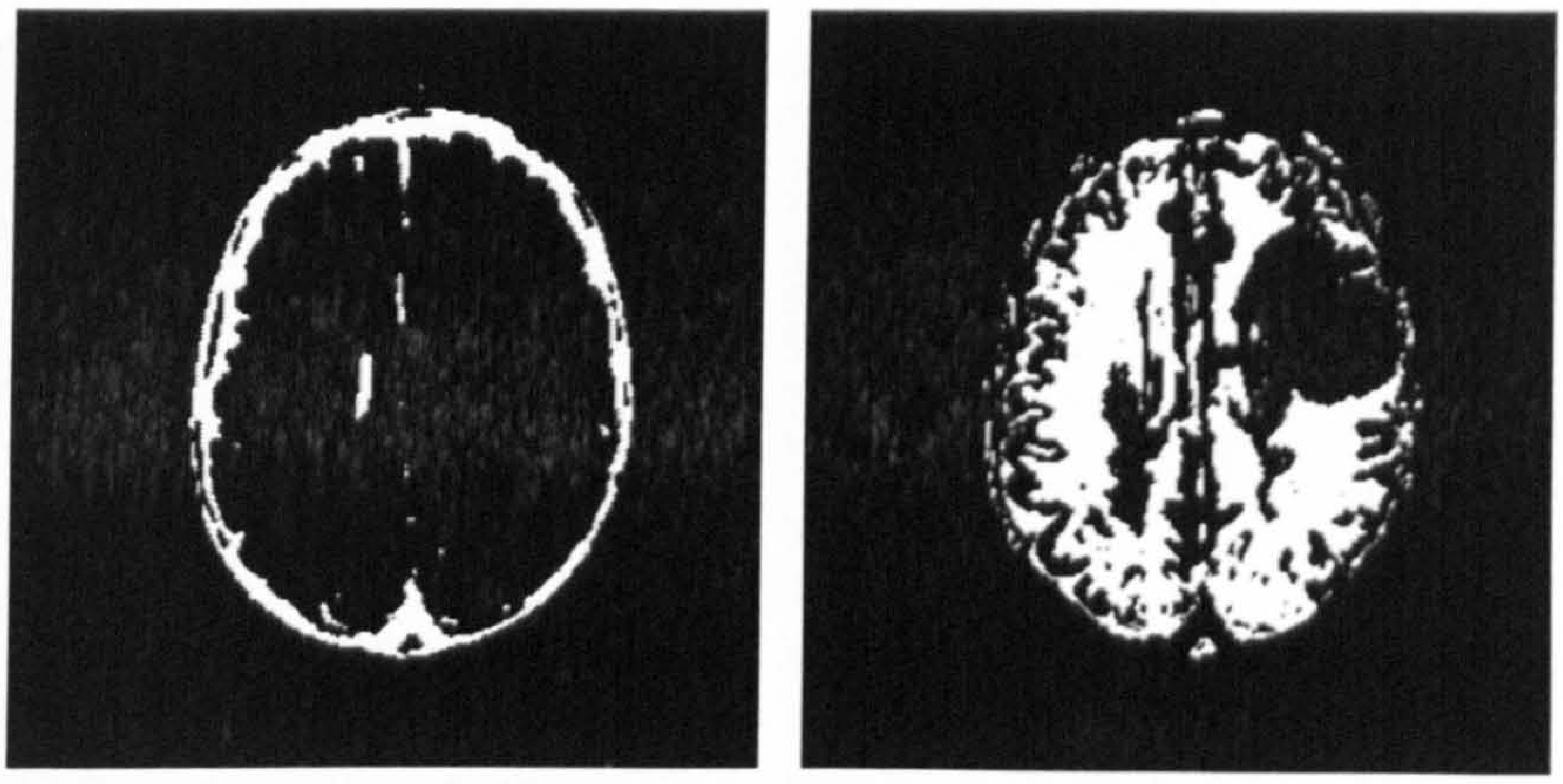
(d) non-brain region after dilation



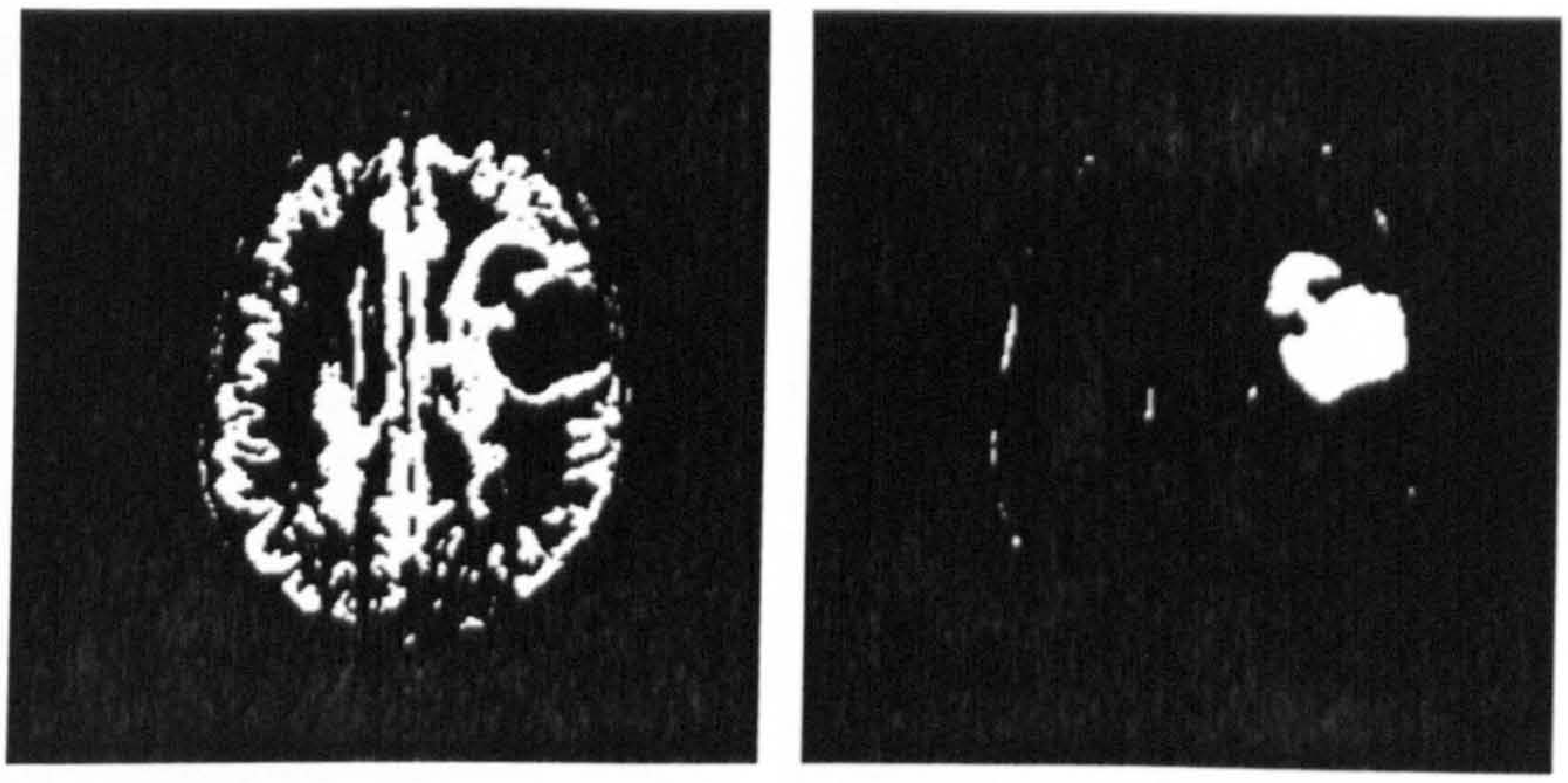
(e) brain region only



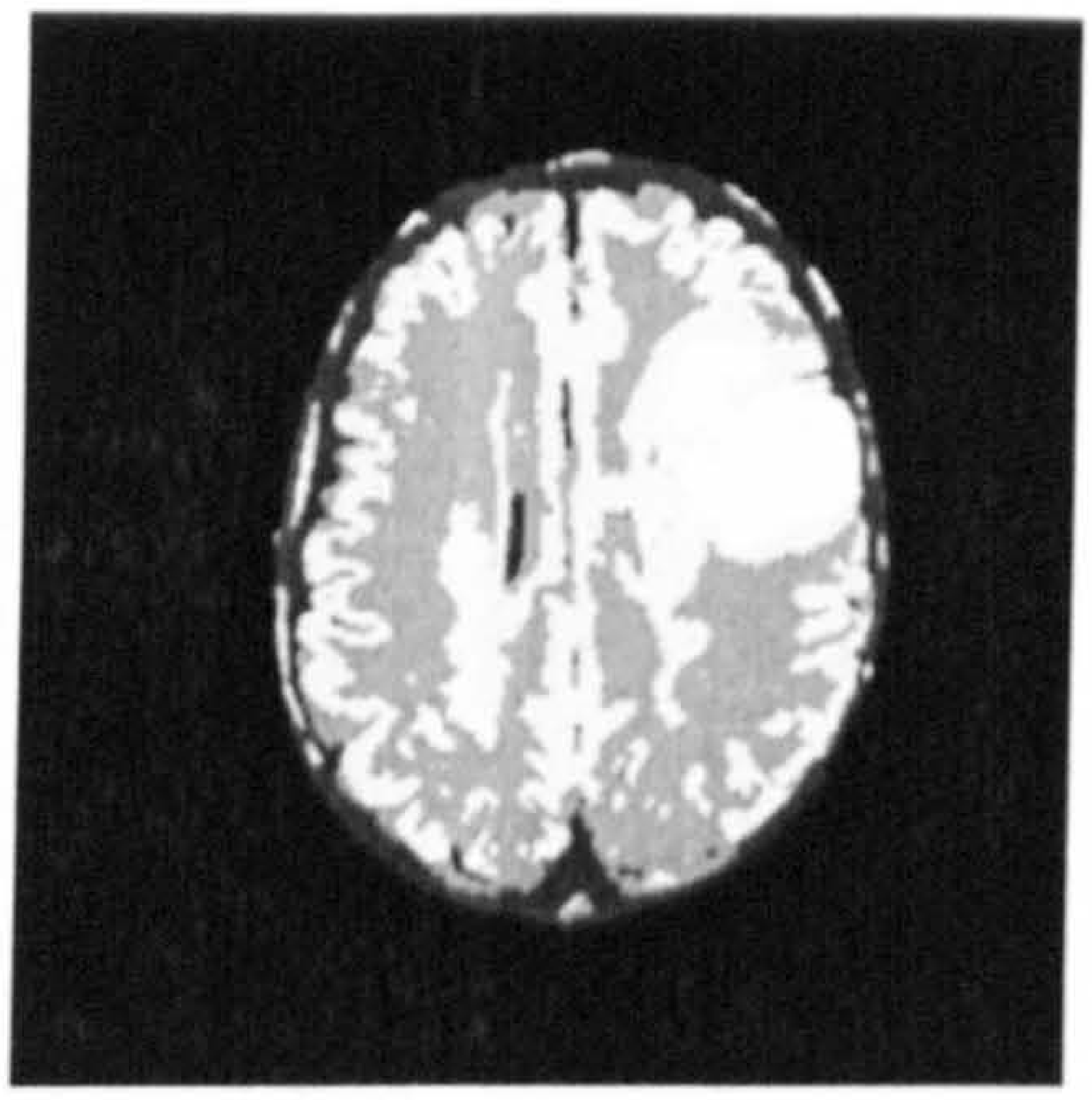
(f) enhanced image



(g) CSF (h) white matter

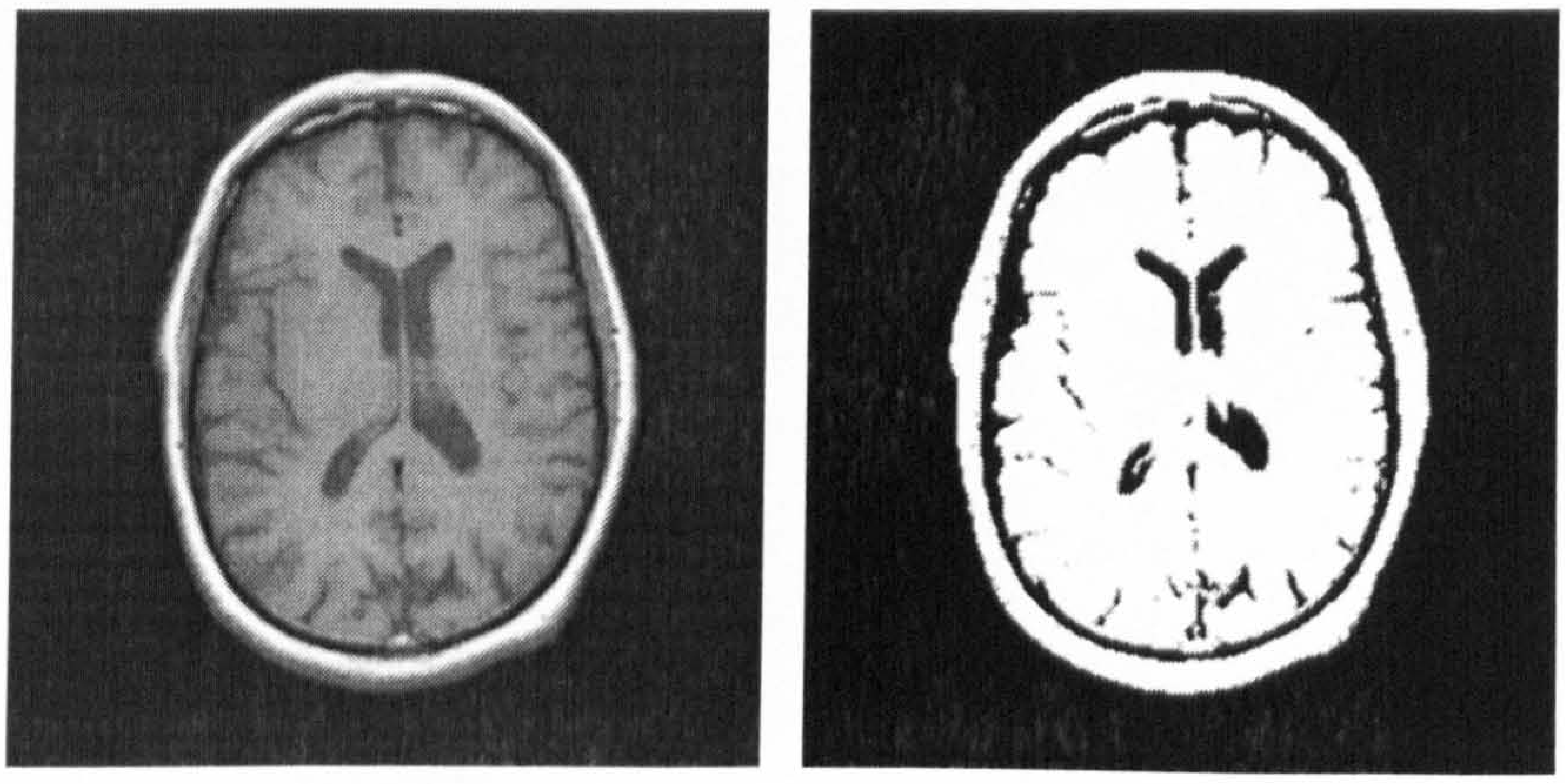


(i) gray matter (j) brain tumour



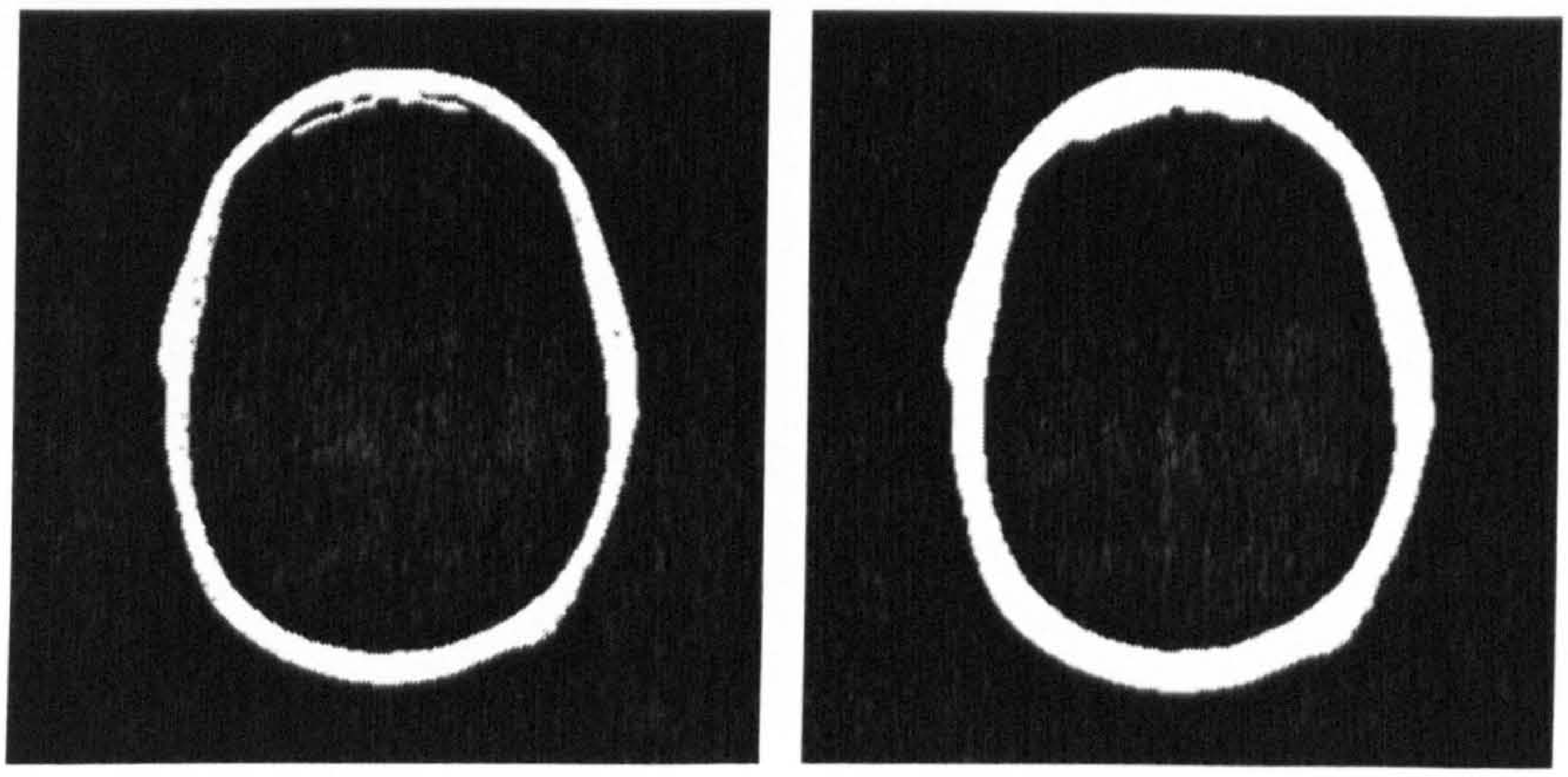
(k) segmentated image

Figure 6.32 Preprocessing and segmentation of FLAIR-FSE MR image.



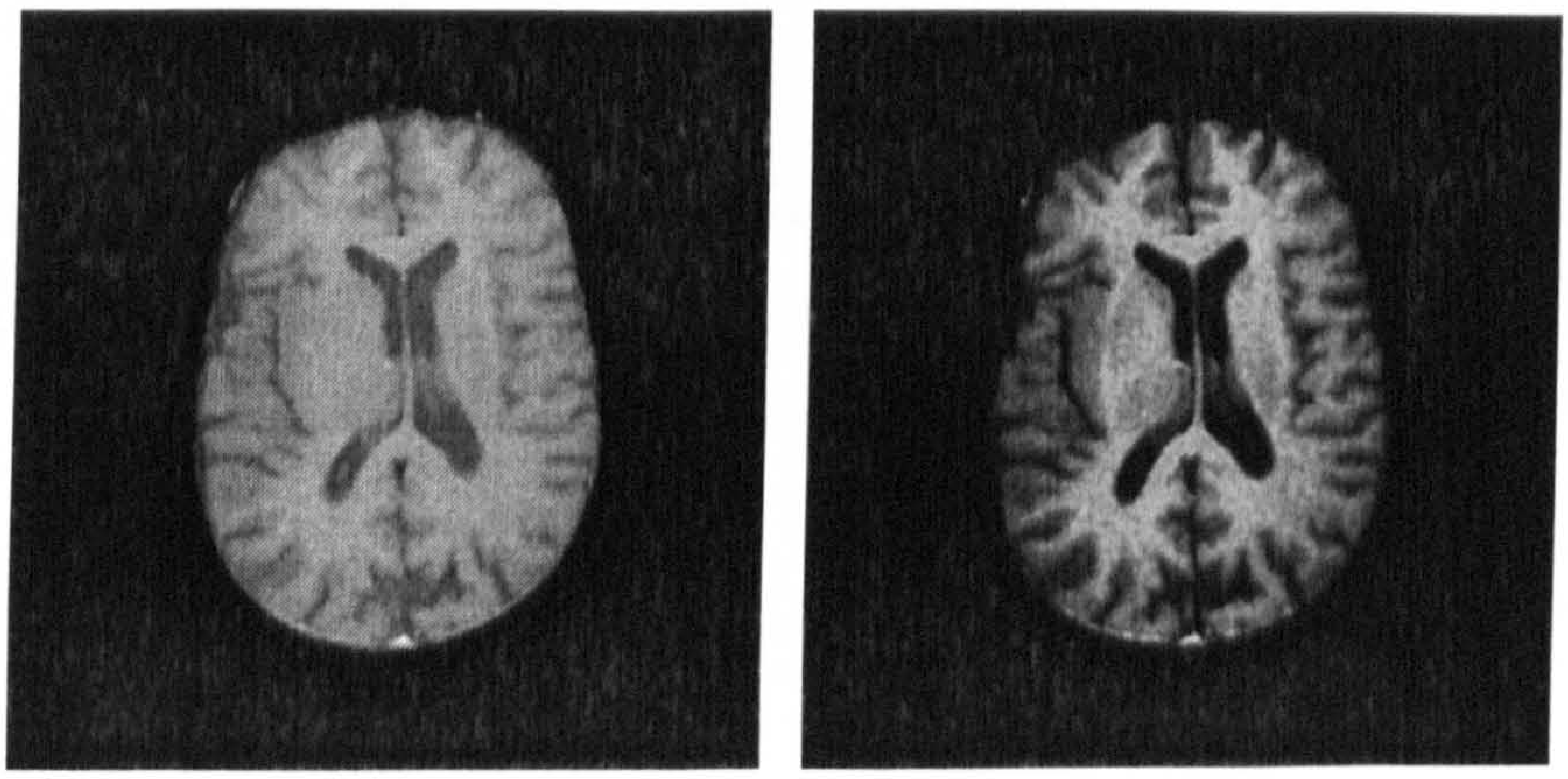
(a) standardized image

(b) binary mask



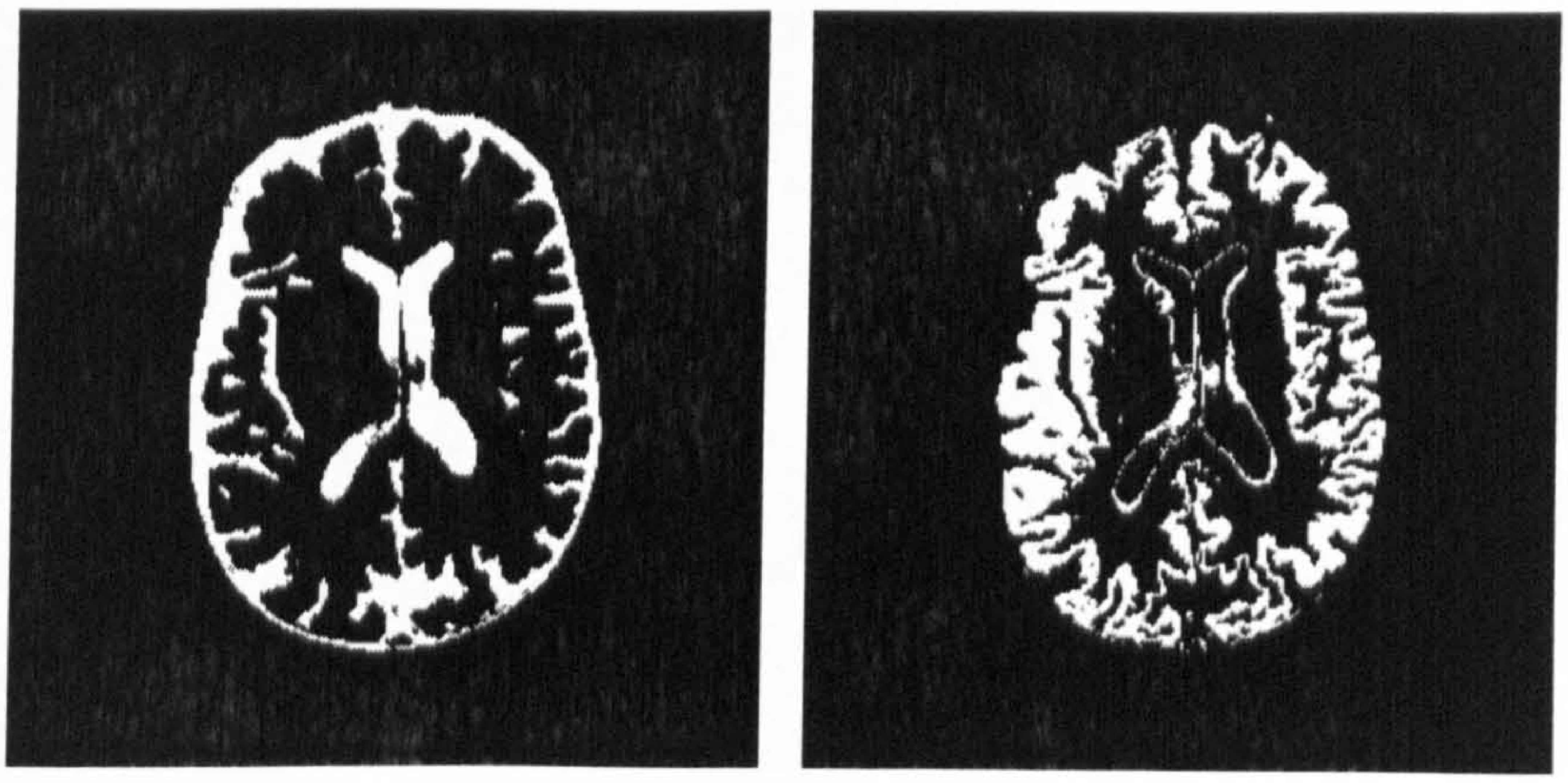
(c) non-brain region after erosion

(d) non-brain region after dilation

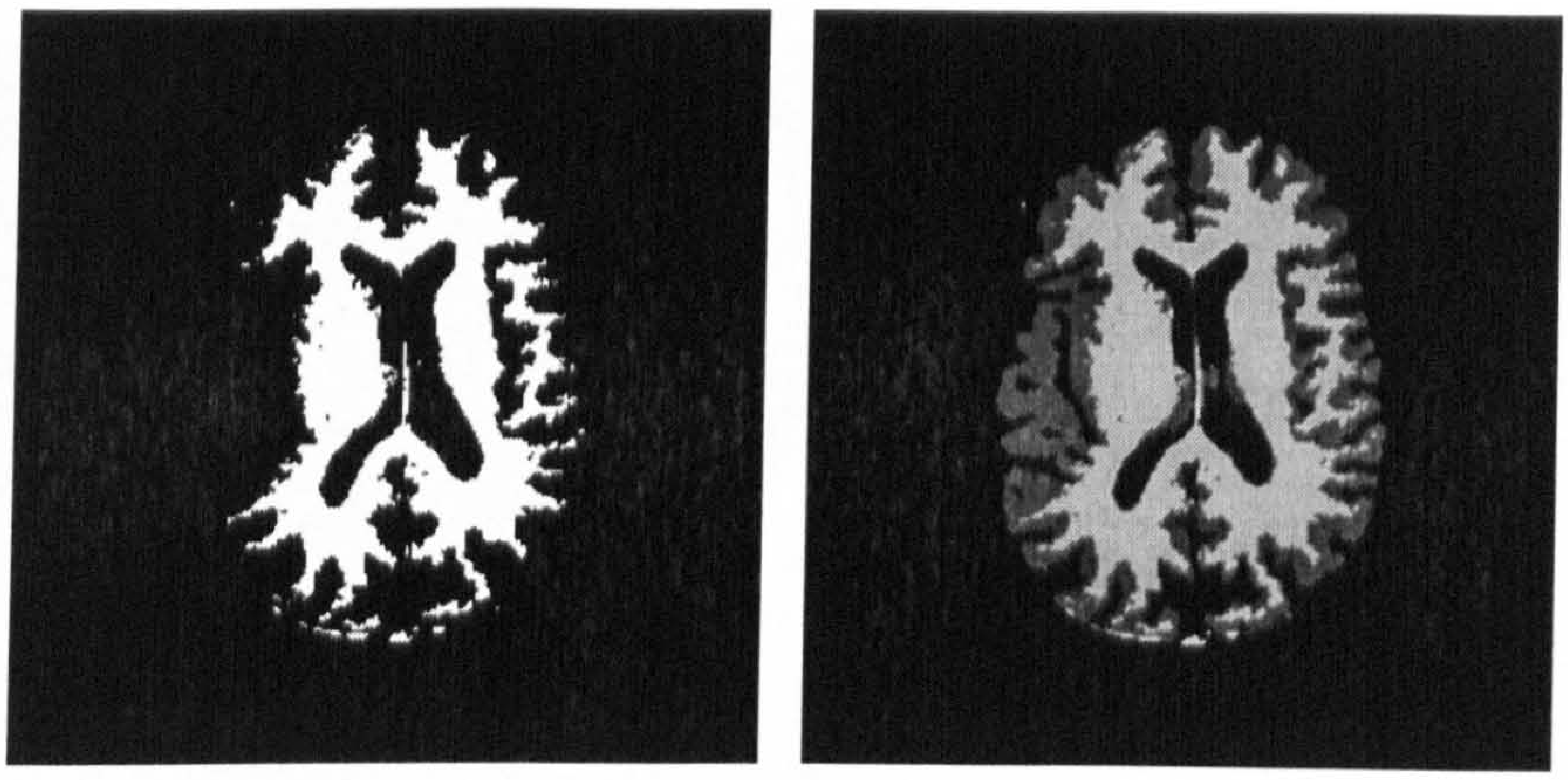


(e) brain region only

(f) enhanced image



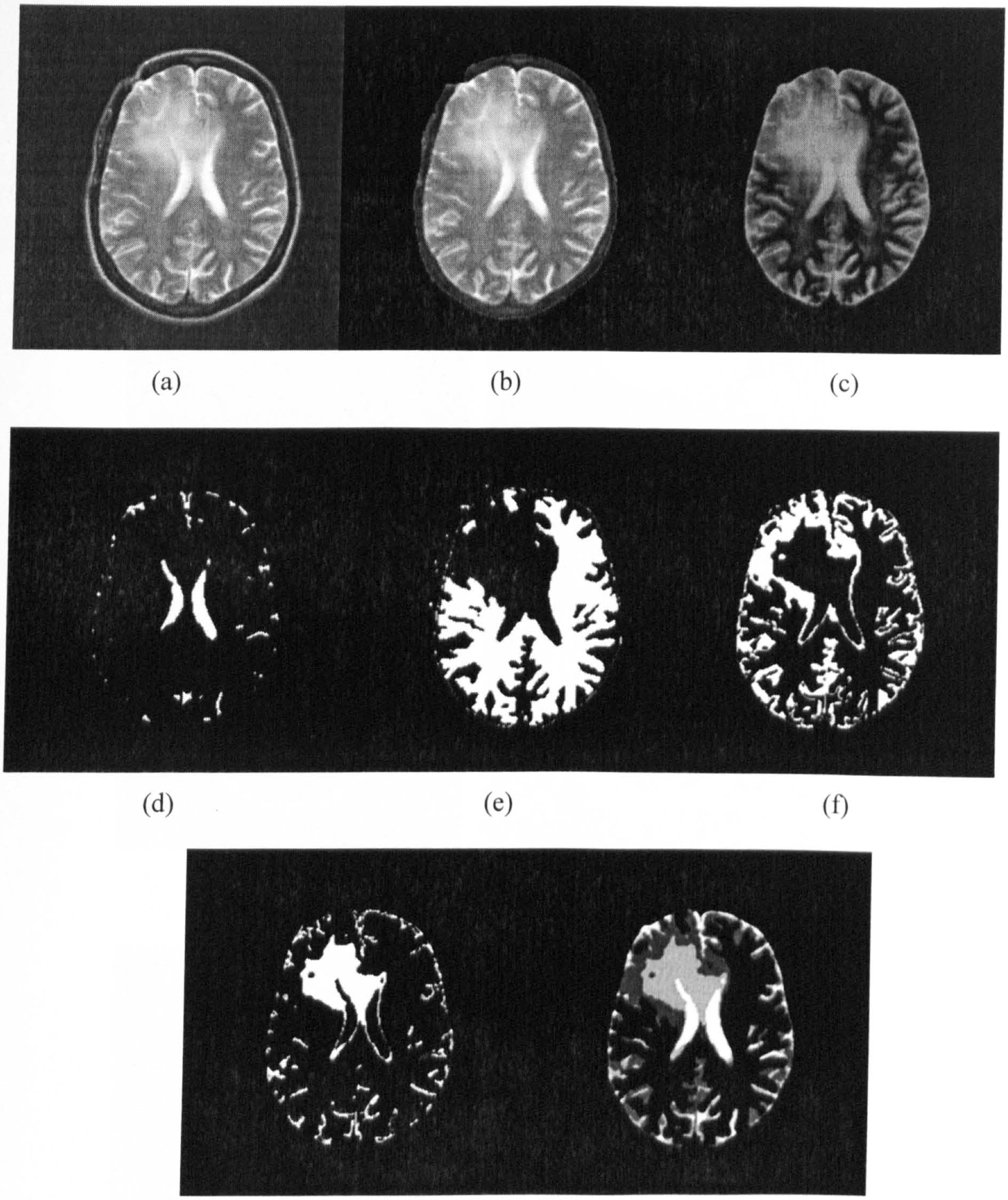
(g) CSF (h) gray matter



(i) white matter (j) segmented image

Figure 6.33 Preprocessing and segmentation of T_1-SE MR image.

The T_2-FSE image shown in Fig.6.34 was from a meningioma patient. As stated above, the non-brain regions in T_2-weighted images are quite difficult to remove due to the high intensities of CSF. The supplementary method presented in Chapter 5 was therefore applied to complete the non-brain removal.



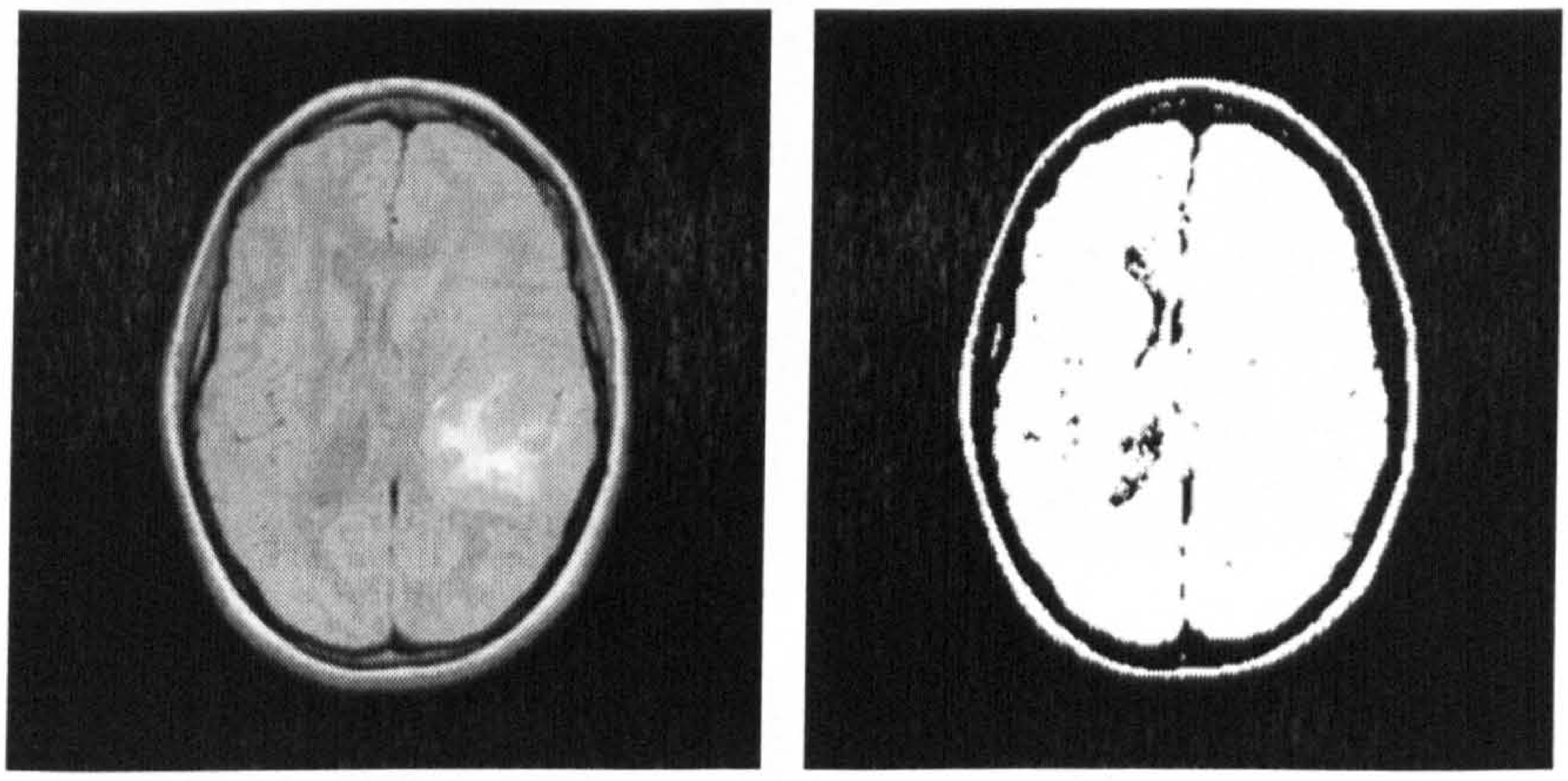
(a) (b) (c)

(d) (e) (f)

(g) (h)

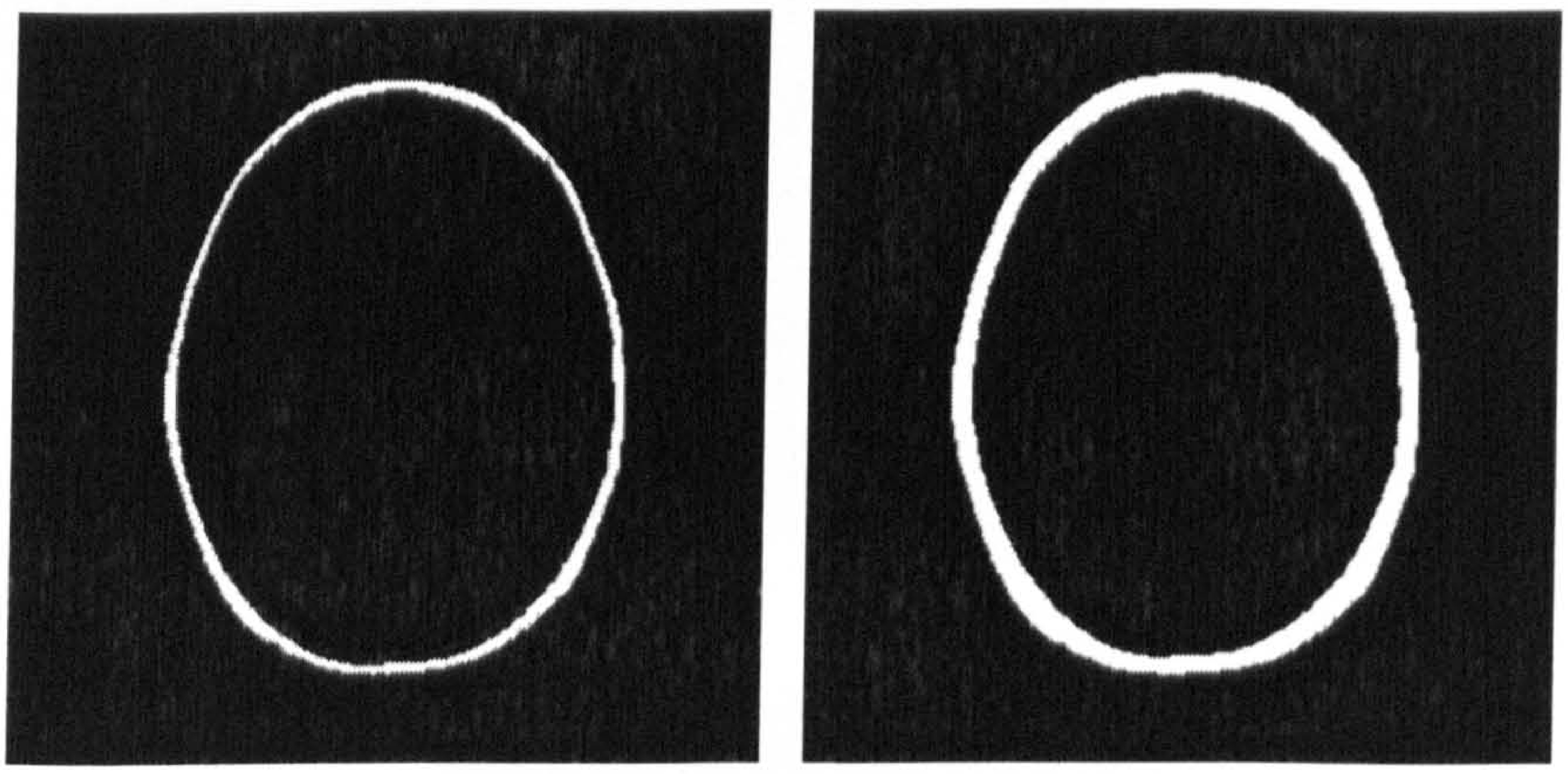
Figure 6.34 Preprocessing and segmentation of T_2-FSE MR image. (a) standardized image (b) brain region only (c) enhanced image (d) CSF (e) white matter (f) gray matter (g) brain tumour (i) segmented image

Fig.6.35 gives a PD-FSE MR image from a high level glioma patient and its preprocessing and segmentation results.



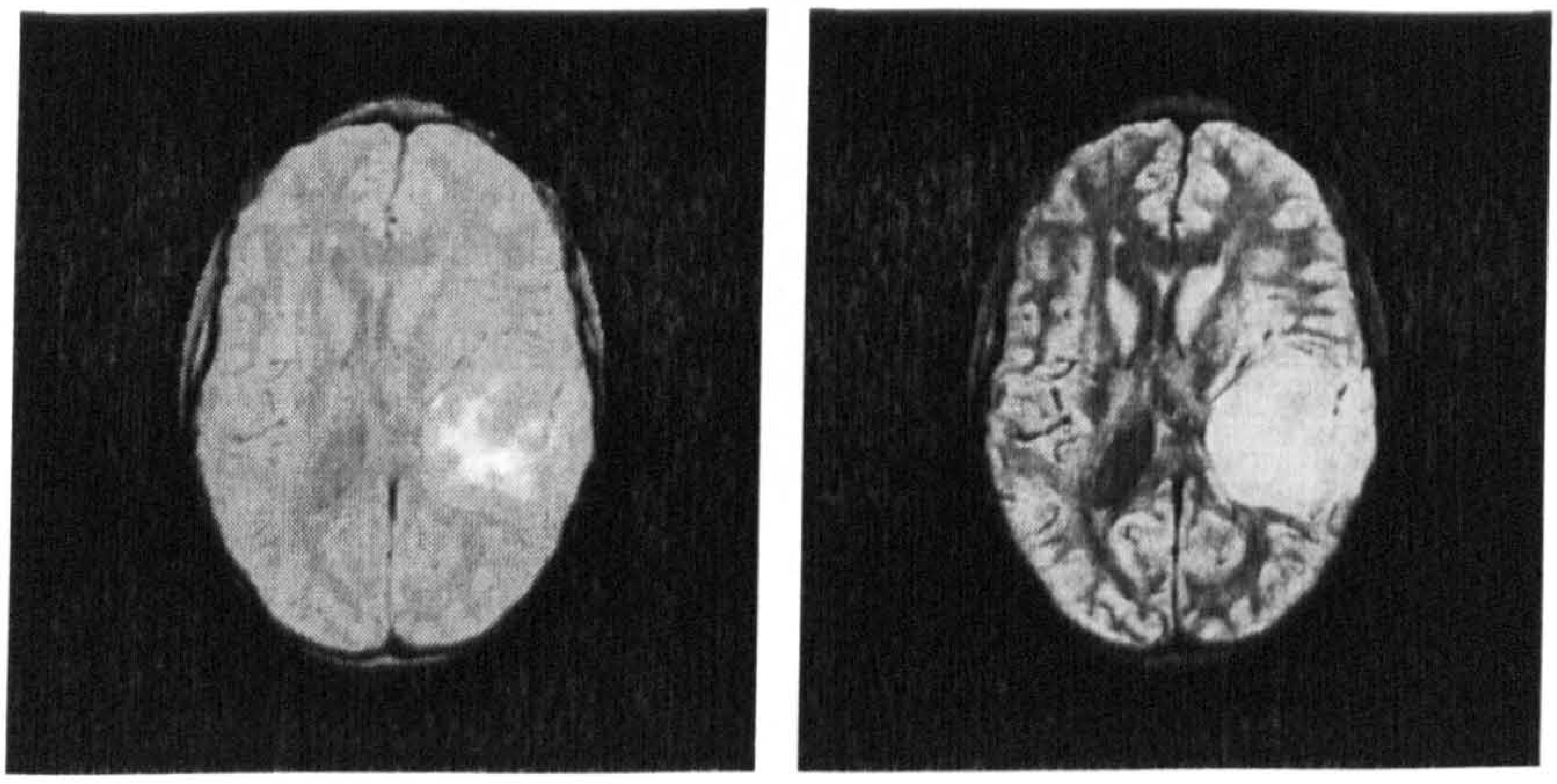
(a) standardized image

(b) binary mask



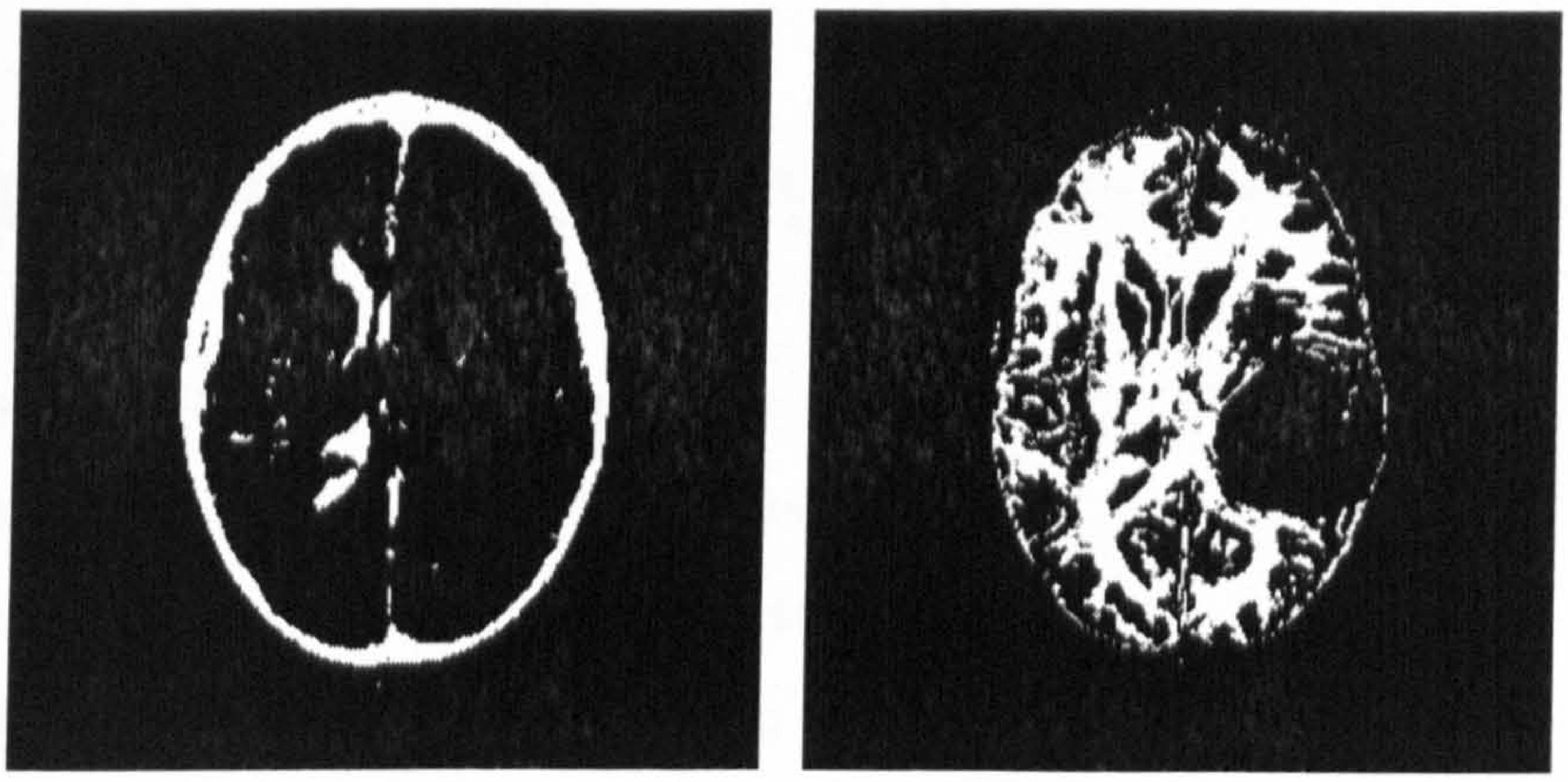
(c) non-brain region after erosion

(d) non-brain region after dilation

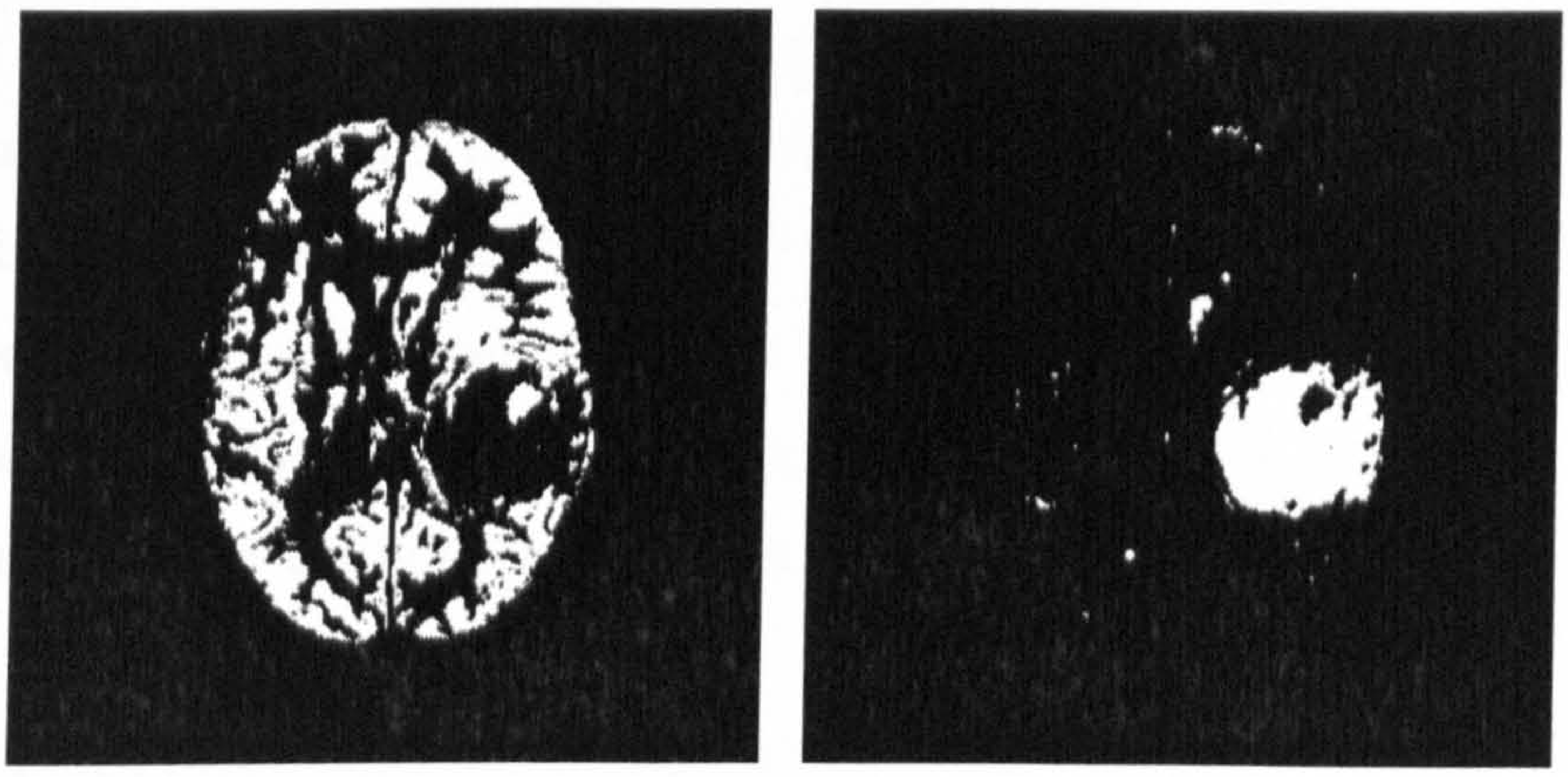


(e) brain region only

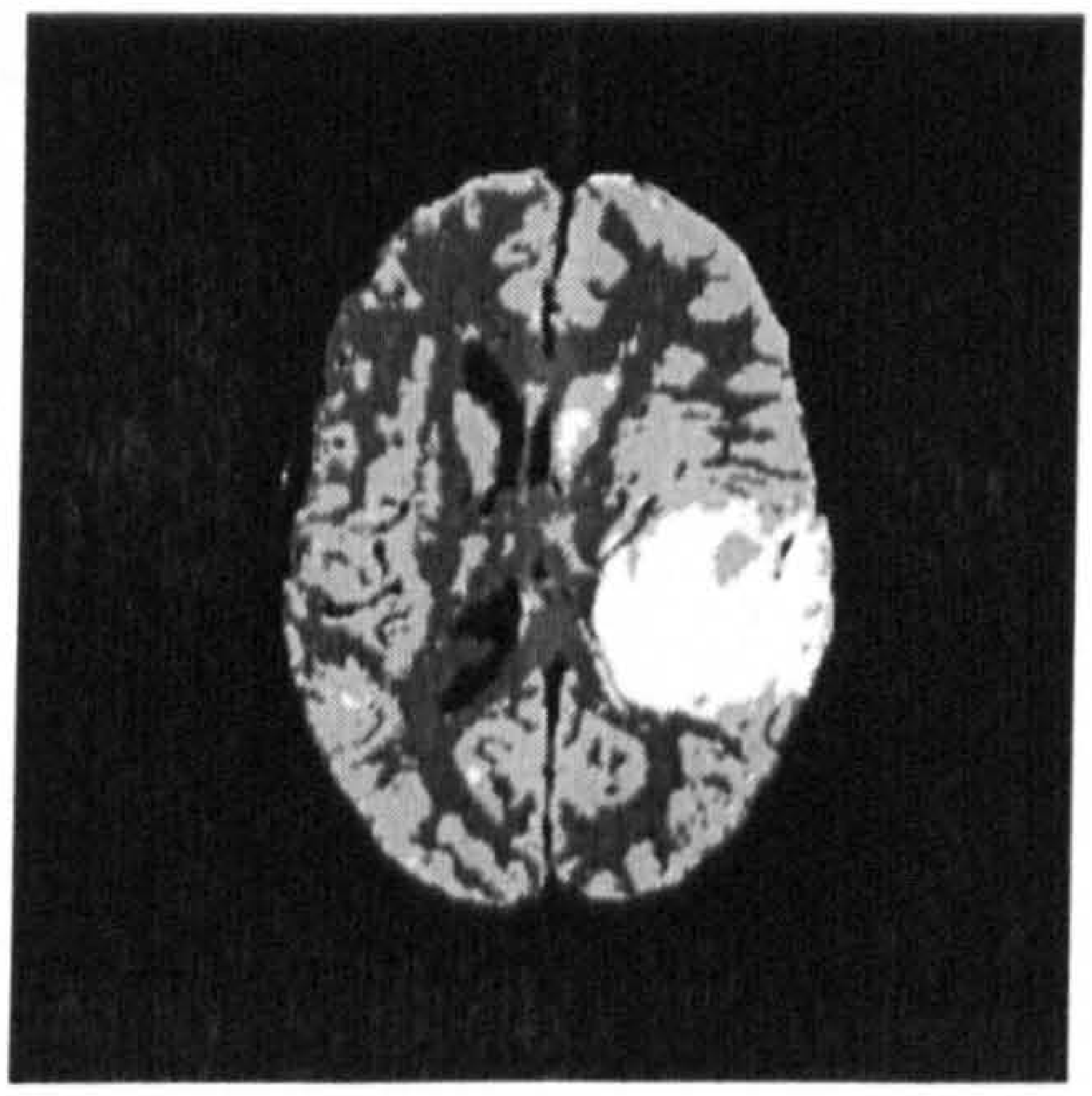
(f) enhanced image



(g) CSF (h) white matter



(i) gray matter (j) brain tumour



(k) segmented image

Figure 6.35 Preprocessing and segmentation of PD-FSE MR image.

6.8 RF INHOMOGENEITY ESTIMATION

Artifacts in MR images can make the conventional intensity-based segmentation methods very difficult, especially spatial intensity inhomogeneity induced by the radio frequency (RF) coil. The RF field sensitivity variation varies slowly across the image. Even advanced techniques such as nonparametric, multi-channel methods cannot solve this problem [Dawant, Zijdenbos and Margolin, 1993]. The most popular method to solve this problem is using the expectation-maximization (EM) algorithm [Wells III *et al.*, 1996; Zhang, Brady and Smith, 2001]. However, there are two main disadvantages of EM approach. First, the EM algorithm is computationally intensive, especially for large problems. Second, the EM algorithm requires a good initial guess for either the bias field or for the clustering estimate. Otherwise the EM algorithm could be easily trapped in a local minimum, resulting in an unsatisfactory solution [Ahmed *et al.*, 1999].

Although MR images which were affected by artifacts were removed at analysis in this thesis, an extension of the IFCM algorithm is proposed in this section which, can estimate inhomogeneities in MR images. Its capability has been demonstrated on both synthetic non-MR images and MR images.

6.8.1 BIAS FIELD MODEL

MRI intensity inhomogeneity is modeled as a spatially-varying factor called the gain field. Let $X=\{\mathbf{X}_1,\mathbf{X}_2,\ldots\mathbf{X}_j\ldots\mathbf{X}_N\}$ be the observed MRI intensities (with artifacts) and $X^* = \{\mathbf{X}_1^*,\mathbf{X}_2^*,\ldots\mathbf{X}_j^*\ldots\mathbf{X}_N^*\}$ be the ideal intensities (without artifacts). N is the total number of pixels in the image. The gain field is denoted by $B=\{\mathbf{B}_1,\mathbf{B}_2,\ldots\mathbf{B}_j\ldots\mathbf{B}_N\}$. The observed MRI intensity is modeled as a product of the true signal generated by the underlying anatomy and the gain field, expressed as follows:

$$\mathbf{X}_j = \mathbf{X}_j^*\mathbf{B}_j \tag{6.24}$$

The application of a logarithmic transformation to the intensities allows the logarithmic artifact, called bias field to be modeled additive to the ideal intensities. Let $\mathbf{x}_j$ and $\mathbf{x}_j^*$ denote the observed and ideal log-transformed intensities at the j-th pixel. β_j is the bias field at the j-th pixel. So,

$$\begin{aligned} \ln \mathbf{X}_j &= \ln(\mathbf{X}_j^* B_j) \\ \mathbf{x}_j &= \mathbf{x}_j^* + \beta_j \end{aligned} \tag{6.25}$$

Bias field β_j is modeled by an N dimensional zero-mean Gaussian prior probability density, which is one of the most successful models [Wells III *et al.*, 1996]. This model can capture the smoothness of the inhomogeneities. The bias field is defined as follows:

$$\frac{1}{N}\sum_{j=1}^{N}\beta_j = 0 \tag{6.26}$$

6.8.2 BIAS FIELD ESTIMATION

The objective function of both FCM and IFCM algorithms in Eq. (6.1) is therefore extended by considering the definition in Eq. (6.26). A constrained optimization of the objective function is expressed using Lagrange multipliers,

$$J_m(U,V) = \sum_{j=1}^{N}\sum_{i=1}^{C}\mathbf{u}_{ij}^m d^2(\mathbf{x}_j,\mathbf{v}_i) + \eta\left(\frac{1}{N}\sum_{j=1}^{N}\beta_j\right) \tag{6.27}$$

Taking the derivative of J_m with respect to β_j and setting the result to zero regarding the optimization,

$$\frac{\partial J_m}{\partial \beta_j} = \sum_{i=1}^{C}\mathbf{u}_{ij}^m (1-\lambda H_{ij} - \xi F_{ij})^2 \cdot 2\|\mathbf{x}_j - \beta_j - \mathbf{v}_i\| + \frac{\eta}{N} = 0 \tag{6.28}$$

Let $M_{ij} = 1 - \lambda H_{ij} - \xi F_{ij}$, and substitute $\eta = \frac{\eta}{N}$. $\|\cdot\|$ is regarded as Euclidean distance.

$$\sum_{i=1}^{C}\mathbf{u}_{ij}^m M_{ij}^2 \cdot 2\|\mathbf{x}_j - \beta_j - \mathbf{v}_i\| + \eta = 0 \tag{6.29}$$

$$\beta_j = \mathbf{x}_j - \frac{\sum_{i=1}^{C}\mathbf{u}_{ij}^m M_{ij}^2 \mathbf{v}_i}{\sum_{i=1}^{C}\mathbf{u}_{ij}^m M_{ij}^2} + \frac{\eta}{2\sum_{i=1}^{C}\mathbf{u}_{ij}^m M_{ij}^2} \tag{6.30}$$

According to Eq. (6.26),

$$\sum_{j=1}^{N}\beta_j=\sum_{j=1}^{N}\mathbf{x}_j-\sum_{j=1}^{N}\frac{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}\mathbf{v}_i}{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}}+\sum_{j=1}^{N}\frac{\eta}{2\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}}=0$$

$$\eta=\frac{\sum_{j=1}^{N}\left(\frac{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}\mathbf{v}_i}{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}}-\mathbf{x}_j\right)}{\sum_{j=1}^{N}\left(2\cdot\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}\right)^{-1}} \tag{6.31}$$

Substituting Eq. (6.31) into Eq. (6.30), the bias field can be estimated as:

$$\beta_j=\mathbf{x}_j-p_j-\frac{\sum_{j=1}^{N}\left(\mathbf{x}_j-p_j\right)}{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}\cdot\sum_{j=1}^{N}\left(\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}\right)^{-1}} \tag{6.32}$$

where, $$p_j=\frac{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}\mathbf{v}_i}{\sum_{i=1}^{C}\mathbf{u}_{ij}^{m}M_{ij}^{2}}$$

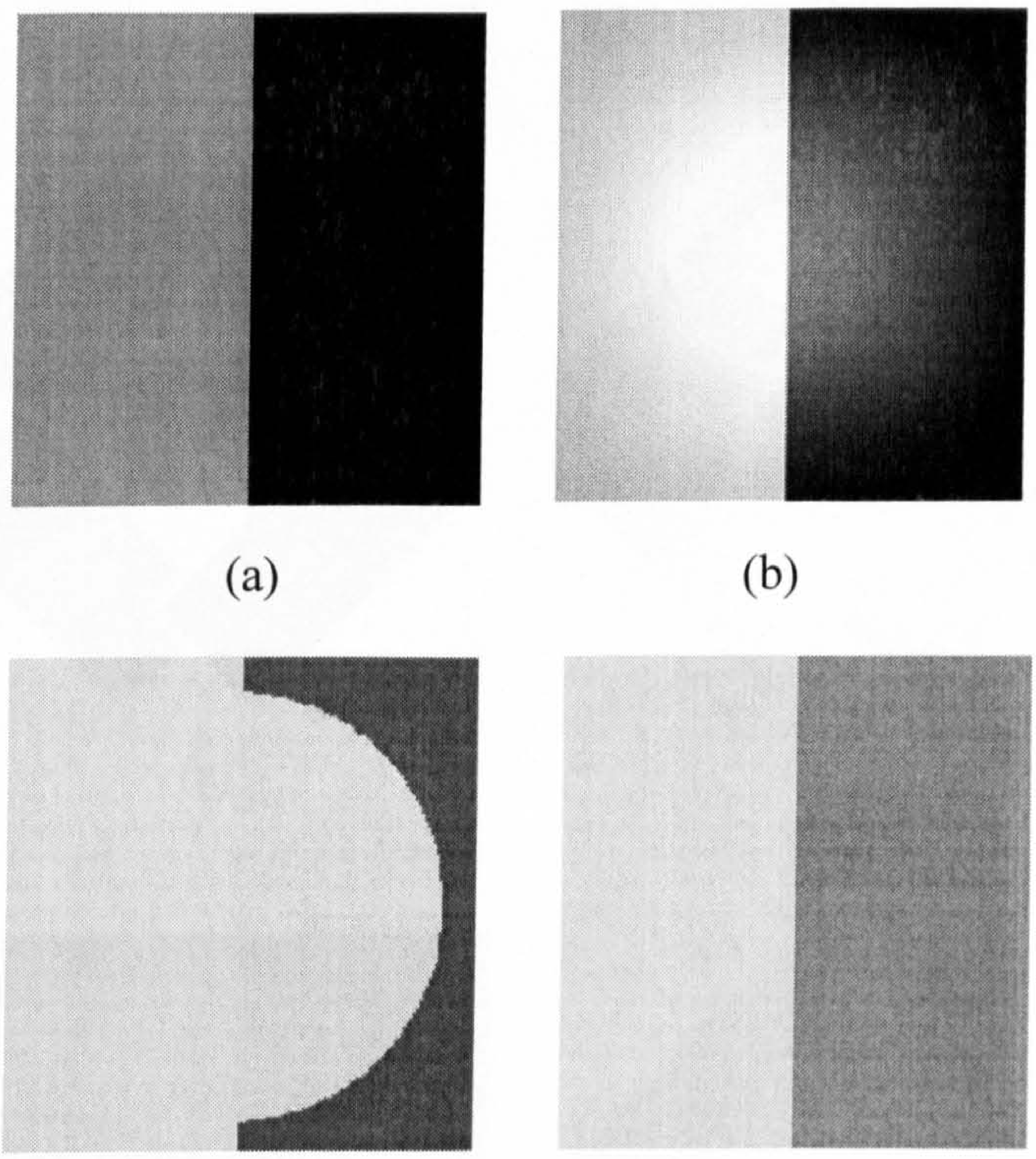
The extension causes no change of membership function but modify the cluster centre as following:

$$\mathbf{v}_i=\frac{\sum_{j=1}^{N}(\mathbf{u}_{ij})^{m}(\mathbf{x}_j-\beta_j)}{\sum_{j=1}^{N}(\mathbf{u}_{ij})^{m}} \tag{6.33}$$

6.8.3 EXAMPLES

The extended IFCM was applied on synthetic images, corrupted multiplicative bias field and also on the simulated MR images. The parameters λ and ξ in Eq. (6.14) were set to be 0.47 and 0.53 by trial and error.

Fig.6.36 (a) shows a two-cluster synthetic image and (b) is the image corrupted by a Gaussian bias field. The segmentation results of both FCM and extended IFCM are shown in Fig.6.36 (c) and (d). Obviously, FCM cannot even segment such a



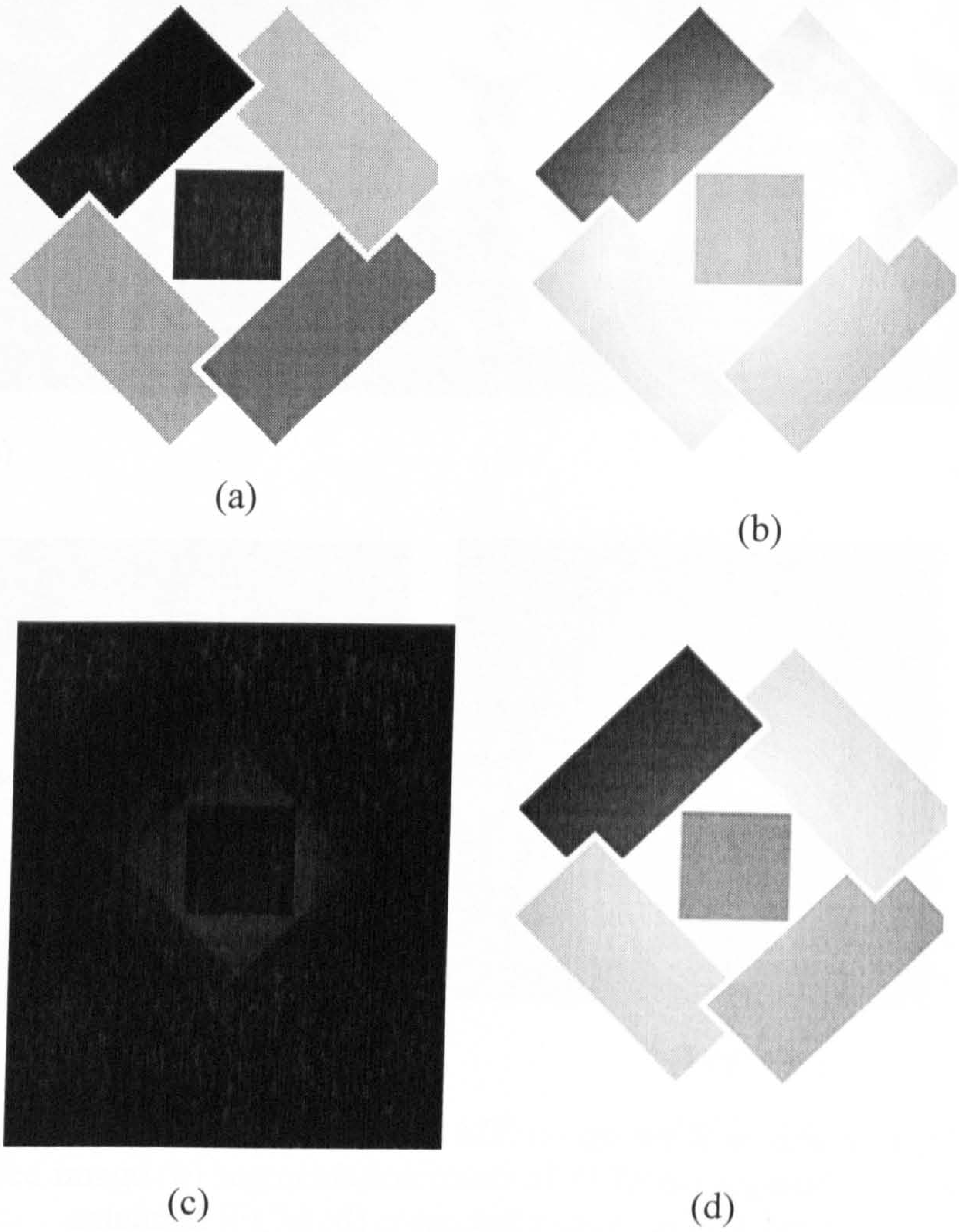
(a) (b)

(c) (d)

Figure 6.36 Synthetic image I for bias field estimation.
(a) original image (b) image corrupted by a Gaussian bias field (c) segmentation result of FCM (d) segmentation result of extended IFCM

simple image corrupted by bias field but the extended IFCM can correct the bias field completely.

Fig.6.37 (a) shows another synthetic image with four clusters and it was also corrupted by a Gaussian bias field. The bias field estimated using extended IFCM algorithm is shown in Fig.6.37 (c) and (d) is its segmentation results.

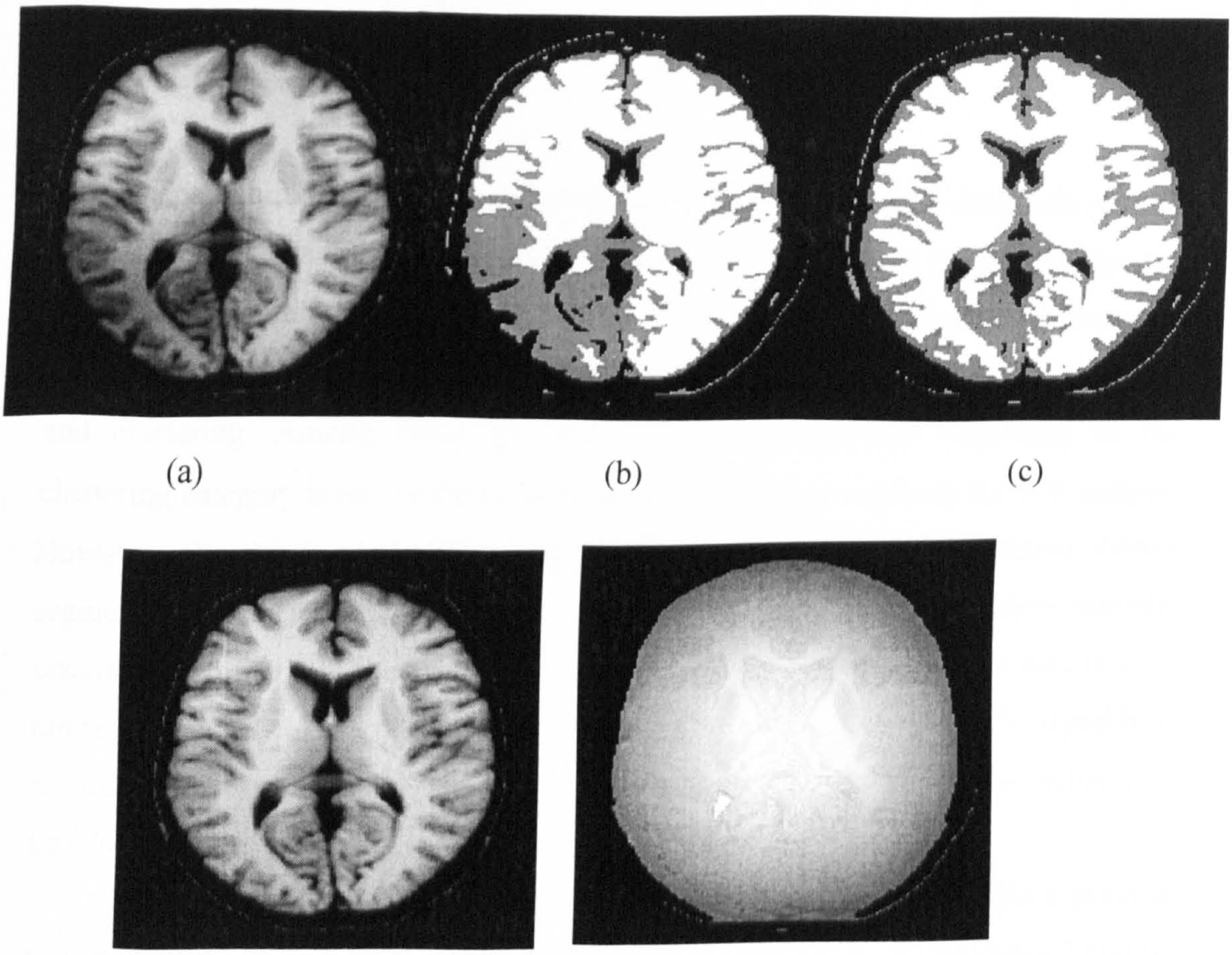


(a) (b)

(c) (d)

Figure 6.37 Synthetic image II for bias field estimation.
(a) original image (b) image corrupted by a Gaussian bias field
(c) bias field estimated (d) segmentation result of extended IFCM

To further demonstrate the ability of extended IFCM, a simulated MR image was downloaded from an MRI simulated brain database [Brainweb, 2003]. It was a T_1-weighted image with 1% noise and 40% non-uniformity. Fig.6.38 (a) shows the original corrupted image and (b) shows the corrected image. The image was segmented into four classes corresponding to background, gray matter, white matter and CSF. The segmentation results of FCM and extended IFCM are given in Fig.6.38 (c) and (d). Similarly, FCM was affected by the bias field while the extended IFCM not only succeeded in segmentation but also estimated the bias field correctly.



(a) (b) (c)

(c) (d)

Figure 6.38 Simulated T_1-weighted MR image for bias field estimation.
(a) corrupted image (b) segmentation result of FCM (c) segmentation result of extended IFCM (d) corrected image (e) bias field

However, using the extended IFCM for bias estimation may lead to loss of some image details. Further work will be needed to solve the problem.

6.9 CONCLUSIONS

This chapter introduced the second major part of a brain tumour diagnosis system, MR image segmentation. Accurate segmentation of brain tissues can result in proper diagnosis of brain tumours. A variety of image segmentation methods in the literature have been reviewed, such as thresholding, region growing, edge-based segmentation and clustering. Among these, the FCM clustering algorithm belonging to the clustering category is one of the most popular segmentation methods for MR images. However, the drawback of FCM algorithm, as with many other intensity-based segmentation methods, is its sensitivity to noise. Since medical images contain uncertain information, noise and even artifacts, the intensity-sensitive methods may not segment medical images correctly. On the basis of the traditional FCM algorithm, an improved segmentation algorithm called IFCM was developed to solve this problem.

Generally, a pile of pixels may represent details of partial image but a pixel is too small to do so. If a pixel has a totally different intensity to its surrounded pixels, it may be regarded as noise. As a result, attractions among nearby pixels appear to draw others towards their own features. On the basis of this theory, a neighbourhood attraction including two factors was considered in our IFCM algorithm. One is called feature attraction. The feature in MR images represents to pixel intensity. The other is distance attraction. The former indicates that the attraction emerges depending on the intensity difference among pixels. The lower the intensity difference is, the stronger the feature attraction. The latter indicates that the attraction also emerges depending on the spatial distance among pixels. The closer the pixels are, the stronger the distance attraction. By considering neighbourhood attraction, segmentation is therefore determined not only by the pixel intensity itself, but also by the intensities of neighbouring pixels and the structure of its neighbourhood. This greatly reduces the sensitivity of segmentation to noise. A neural network model was presented to search the optimized degree for both feature attraction and distance attraction. Different images may require various values for the two attractions.

Other approaches also based on FCM, such as NCM, PCM and RFCM were described and compared to IFCM. Synthetic square images and Lena facial image with different noise were used to test all methods for segmentation. Besides the better visual effect of IFCM than other methods, five parameters were defined to confirm the superiority of IFCM algorithm quantitatively. Due to no ground truth available to MR images, simulated brain MR images whose anatomic structure was known were downloaded from Brainweb (2003). Typical images from T_1-weighted, T_2-weighted and PD-weighted categories with different noise levels were segmented by IFCM along with other methods. The better performance and higher accuracy of IFCM were suggested by visual segmentation results and evaluation parameters obtained. After going through all the other tests successfully, real brain MR images could finally be segmented using IFCM algorithm. For each MRI modality, only one typical image was shown in this chapter. Each brain image was segmented into white matter, gray matter, CSF and brain tumour, while some T_1-weighted image displayed no brain tumour. The average intensity of each part in each image was then easily calculated and saved in a database for the purpose of further diagnosis.

RF inhomogeneity is one the most prevalent artifacts in MR images. The last section of this chapter described how the IFCM algorithm can be extended to estimate this inhomogeneity. Its capability has been demonstrated on both synthetic images and simulated MR images.

Chapter 7

BRAIN TUMOUR DIAGNOSIS USING FUZZY LOGIC BASED GENETIC PROGRAMMING

7.1 INTRODUCTION

Brain tumour diagnosis is the last major part of the diagnosis system in this thesis. From the previous two parts, image preprocessing and image segmentation, a database was built which comprises intensities of white matter, gray matter, CSF and brain tumour from each MR images. The size of the database has no limits but its actual size depends upon the number of images and the number of slices of each image. It could contain a huge amount of data making it difficult to read all of them. How to dig deeply into this data and discover diagnosis rules for brain tumours is our main concern. As stated above, data mining techniques emerged especially useful for excavating interesting knowledge from a large number of data; therefore they are the best choice dealing with this kind of problem. On the other hand, in terms of the number of tumour types, the diagnosis problem could be regarded as a multi-classification problem. The next task is therefore to choose an appropriate method from the data mining techniques for multi-class classification. It is worthy of note that medical data has its own characteristics which should be considered in the processing.

In this chapter, the literature review of data mining methods in multi-class classification is described. Most methods have to convert a multi-class classification problem into several two-class classification problems. A fuzzy logic based GP is proposed here which can deal with multi-class classification directly. Classification rules for brain tumour types are discovered with acceptable accuracies.

7.2 DATA MINING METHODS IN MULTI-CLASS CLASSIFICATION

Multi-class classification is a popular problem with a number of applications in a variety of areas, such as computer vision, bioinformatics, speech recognition, robotic control, marketing, drug discovery and medical diagnosis. Usually the classification problems have a common feature in that a large amount of data is involved. Due to the excellence of data mining in dealing with volumes of data, many attempts have been made on classification by means of data mining methods.

Medical diagnosis has been paid a great attention among all these applications of multi-class classification using data mining techniques. An effective classification can automatically diagnose the diseases or analyze the conditions of patients, which may assist or enhance the clinical decision of doctors. It may also save a great deal of time in an urgent case and save lives. For example, chest pain is a symptom related to several diseases such as cardiovascular, pulmonary, esophageal, and psychogenic. In terms of the report of the World Health Organization, cardiovascular diseases accounts for about 25% of death rate in the whole world, especially in developed countries. A critical problem faced by doctors particularly in emergency room is how to discriminate the life threatening diseases from other less serious pathologies quickly and effectively, which all have the similar symptom of chest pain. Bojarczuk *et al.* (2000) proposed a paradigm to classify 12 diseases (classes) related to chest pain. Many contributions to this problem were also presented by such as Assanelli *et al.* (1993), Mair (1995). Other applications for medical diagnosis using classification methods include differentiating the patients with the coronary artery disease from healthy people (two-class classification) by Cios *et al.* (2002), prognostic predicting of bilharziasis-related bladder cancer by Wei *et al.* (2003), diagnosing whether the solitary pulmonary nodule is cancerous or benign by Kusiak *et al.* (2000) and determining the malignancy of breast cancer by Kovalerchuk, Vityaev and Ruiz (2000).

For solving classification problems, many effective and efficient methods in data mining have been delivered in the literature. Most of them are based on mathematic models or theories [Chien, Lin and Hong, 2002]. For example, the statistic classifiers were built on the Bayesian theory which provides a probability model to assign data to a certain class with the highest probability. The major limitation is that effective classification is highly related to the knowledge users have of data properties. Neural network is another popular classification method in which a multi-layered network is trained with the given training dataset. The drawback of a neural network is the opaque representation of the results and the inefficiency of the training process. Classification problems are also dealt with traditionally by the maximum likelihood classifier (MLC). It measures distances among input data and classifies data into a class with the least distance. However, a normal distribution is generally assumed for the input data, which is not usually true, because it leads to minimum classification error.

Genetic programming (GP) is gaining attention in multi-class classification area due to its ability to discover the underlying data relationships and express them clearly. GP no longer has to encode classifier to sequences of strings as in genetic algorithm. It begins with a population of randomly created computer programs which represents a potential solution. The primary considerations in applying GP to multi-class classification are as following [Kishore *et al.*, 2000]:

(1) Require no particular data distribution, i.e., no a prior knowledge is needed about the statistical distribution of the data as in MLC

(2) Detect underlying but unknown relationships that exists among data

(3) Discover the most important discriminative features of a class

(4) Operate directly on the data in their original form

However, GP is usually applied to distinguish between just two classes of objects [Koza, 1992] as in many other methods. A discriminate function is optimized such that for values larger than a certain threshold value, the object is classified as class A, and otherwise to class B. Thus, the most common approach in practice is to convert the multi-class classification (e.g. N class) problem to binary classification problems and iterate the classifier N times until each class is classified [Kishore, 2000]. Obviously, it increases the computing time and reduces the efficiency of GP incredibly.

For the sake of solving the problem, a fuzzy logic based GP is proposed in this thesis to deal with the multi-class classification problems. It does not request to convert N-class classification into N two-class classification, but classifies all classes simultaneously.

7.3 FUZZY LOGIC BASED GP

The combination of fuzzy logic and GP has previously been presented by different researchers. Edmonds, Burkhardt and Adjei (1995) published a combination method which can produce a powerful methodology for the generation of fuzzy production rules that are effective and intelligible. Chien, Lin and Hong (2002) introduced another combined method for data classification which can increase the accuracy. However, none of them tried to combine fuzzy logic and GP for direct multi-class classification.

GP implements a classification task as described in Chapter 4.3. Initially, a population of individuals is generated randomly. Each individual in the population of GP represents a classifier for a specified class. Generally for GP, the classifier can be accomplished in either two ways: classification rules [Fretas, 1997] or classification functions [Kishore *et al.*, 2000]. The latter, also called discrimination functions, is more concise and efficient than the former. However, the discrimination functions are not easily understood and interpreted directly due to their non-linearity. Due to the need for comprehensible results, classification rules are chosen as the modality of the classification results, which have the form as:

IF conditions *THEN* classification

The logical combination of conditions on the values of predicting attributes composed of the 'IF' part. The 'THEN' part contains the predicted class.

In a two-class classification problem, the classification rules are to predict whether or not the data belongs to a prior specified class. Therefore, the individuals of GP encode the 'IF' part of the rules; whereas the 'THEN' part is the prior specified class. During the evolution of GP, only the 'IF' part (individual) in classification rule is operated by genetic operators but the 'THEN' part is ignored. The classification results can be explained as: if the individual is true, the subject is classified to the specified class; on the contrary, it belongs to the other class. The conventional GP for multi-class classification executes the complete GP process the number of times equal to the number of classes. One class is predicted each time. The diagnosis of the chest pain related diseases by Bojarczuk, Lopes and Freitas (2000) predicted 12 diseases, so the GP proceeded 12 times.

In our fuzzy logic based GP, the individual still stands for the 'IF' part of the classification rule, the 'THEN' part is not specified in advance but determined in the first stage of each generation. Generally, the individual consists of the functional set (FS) and the terminal set (TS), which are variable in different applications. Fig.7.1 shows the basic structure of classification rules.

The fitness function in GP evaluates the quality of each individual, or the capability to solve the problem. The definitions of fitness function are different depending on the features of the problem. The most popular fitness function in classification approach consists of following aspects:

True positive (*tp*): the number of subjects that the individual predicts belongs to a certain class and they do belong to the class.

False positive (*fp*): the number of subjects that the individual predicts belongs to a certain class but they do not belong to the class.

True negative (*tn*): the number of subjects that the individual predicts that does not belong to a certain class and they indeed do not belong to the class.

False negative (*fn*): the number of subjects that the individual predicts does not belong to a certain class but they do belong to the class.

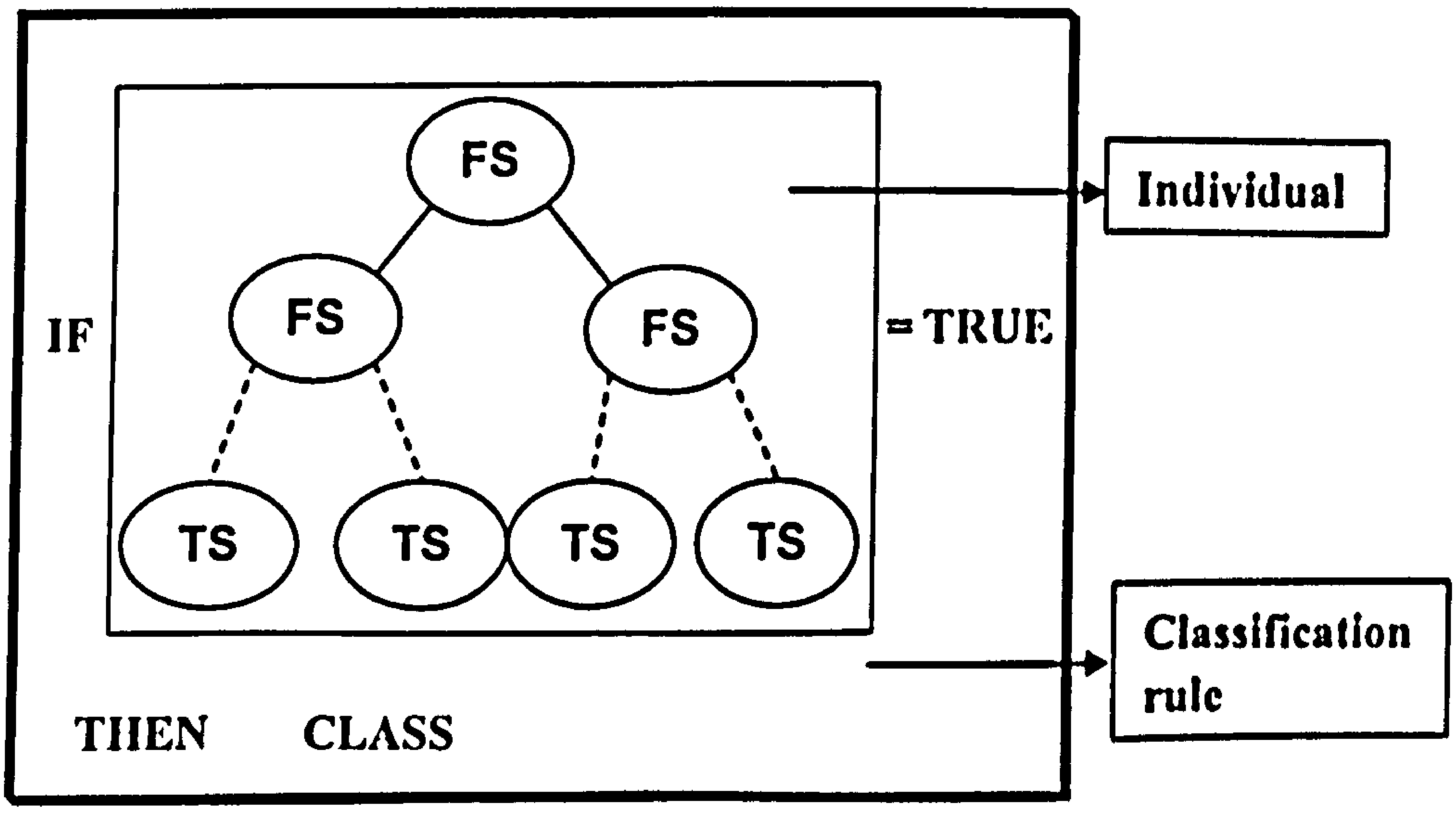


Figure 7.1 Structure of classification.

Two parameters are then defined:

$$Se = \frac{tp}{tp + fn} \tag{7.1}$$

$$Sp = \frac{tn}{tn + fp} \tag{7.2}$$

Where Se is named sensitivity, and Sp is named specificity.

So, the fitness function is the product of both Se and Sp. The higher the fitness is, the better the individual in this case.

$$Fitness = Se \times Sp \tag{7.3}$$

However, the fitness function does not fit for many classification problems, especially in the medical diagnosis domain. Suppose several types of diseases are to be classified, the sensitivity can be described as the capability that the classification rule diagnoses the patient with the disease they actually have; the specificity can be described as the capability that the classification rule correctly diagnoses the patient without the disease. Obviously, the former is more important than the latter. It is not enough just simply increase the weighted value of the sensitivity to improve the performance of the fitness function, but more factors need to be considered.

In our approach, two fitness functions are employed rather than one to evaluate the classification rules more comprehensively. The fuzzy logic technique is also combined in fitness function determination, which can solve the multi-class classification problem directly without converting to binary classification.

Assume a population of one generation is obtained; the membership of each individual to each class is defined as follows:

$$u_{ij} = \frac{C_{ij}}{N_i} \tag{7.4}$$

Where $i=1...n$ (class number), $j=1...L$ (rule number). C_{ij} represents the number of subjects that the j-th individual can classify into the i-th class correctly. N_i is the number of subjects belonging to the i-th class. The higher the membership u_{ij} is, the better the ability of the j-th individual to classify the subjects into the i-th class.

The first fitness function is therefore expressed as:

$$[fitness1_j, Class_j] = \max(u_{1j} \ldots u_{ij} \ldots u_{nj}) \tag{7.5}$$

Where $fitness1_j$ indicates the degree of correct classification ($fitness1_j \in [0,1]$), and $Class_j$ labels the class each individual can classify the best. If u_{kj} is the maximum of the memberships, $fitness1_j = u_{kj}$ and $Class_j = k$. It denotes that the j-th individual has the best ability to classify the k-th class with $fitness1_j$.

Previously, the j-th individual is like 'IF j-th conditions' without the 'THEN' part. Now the j-th classification rule corresponding to the j-th individual has been completed as: 'IF j-th conditions, THEN k-th class'.

A little modification has been made on the definitions of tp, fp, tn, fn stated above and specified several new parameters for this particular approach.

f_{1j} denotes the percentage that the classification rule classifies the subjects into the k-th class and they do belong to the k-th class.

f_{2j} denotes the percentage that the classification rule classifies the subjects into the k-th class and they do not belong to the k-th class.

f_{3j} denotes the percentage that the classification rule does not classify the subjects into the k-th class and they do not belong to the k-th class.

f_{4j} denotes the percentage that the classification rule does not classify the subjects into the k-th class and they do belong to the k-th class.

So, the second fitness function is described as following:

$$fitness2_j = \frac{\alpha f_{1j} + \beta f_{3j}}{1 + (\gamma f_{2j} + \varphi f_{4j})} \tag{7.6}$$

Where, $\alpha, \beta, \gamma, \varphi \in [0,1]$, and they must satisfy the relations below:

$$\begin{cases} \alpha + \beta = 1 \\ \gamma + \varphi = 1 \end{cases} \tag{7.7}$$

From Eq. (7.6), the higher the numerator and the lower the denominator is, the more optimum the *fitness*2. Ideally, if all subjects are classified correctly either positively or negatively, then: $f_{1j} = f_{3j} = 1, f_{2j} = f_{4j} = 0$. The optimized value of the second fitness is

$$fitness2_j = \frac{\alpha + \beta}{1 + 0} = 1$$

In the worst situation, $f_{1j} = f_{3j} = 0, f_{2j} = f_{4j} = 1$ and $fitness2_j = 0$. So, $fitness2_j \in [0,1]$.

Similarly as stated above for the sensitivity and specificity, f_{1j} is more important than f_{3j} because the former classifies positively and the latter classifies negatively. Especially in medical diagnosis, it is more crucial to decide if the patient has a certain disease than decide if a healthy person does not have this disease. On the contrary, f_{2j} describes mis-classification and f_{4j} represents unclassified cases. Diagnosing a patient with a disease, rather than the one he really has, is clearly much worse than being unable to diagnose. Because if the type of disease is known incorrectly; then a totally different therapy may be carried which might lead to fatal results. But for the type of disease not being able to determined, a further diagnosis can be applied. As a result, usually $\alpha > \beta$ and $\gamma > \varphi$. In summary, each individual is decided its class with respect to its *fitness1* and evaluated its classification ability with *fitness2*.

Suppose the number of population is 100 in each generation and the maximum generation is 100. Only one out of 100×100 individuals is useful in the conventional GP, because just one classification rule for one class can be discovered at each time of GP implementation. For *N* class classification, *N*×100×100 classification rules are created but only *N* rules are useful. In our fuzzy based genetic programming, *N*-class classification can be completed in one GP implementation. One individual may be no fit to a class but be very fit to another class; whereas the conventional GP discards all other rules. So, *N* out of 100×100 created rules are valuable. It is a simple mathematic question which indicates the significance of the proposed method over the conventional one.

Fig.7.2 shows the configuration of the fuzzy logic based genetic programming for multi-class classification problems. The detailed implementation is as following:

(1) Select the number of population, the maximum generation or termination criteria

(2) Determine functional set and terminal set

(3) Generate an initial population at Generation=0 by the ramped half-and-half method

(4) Compute membership functions of each individual using Eq. (7.4)

(5) Decide *fitness*1 and corresponding class to each individual using Eq. (7.5) and complete classification rules

(6) Decide *fitness*2 of each classification rules using Eq. (7.6)

(7) Apply genetic operations (reproduction, crossover and mutation etc) to the individuals chosen by a specified selection method and create a new

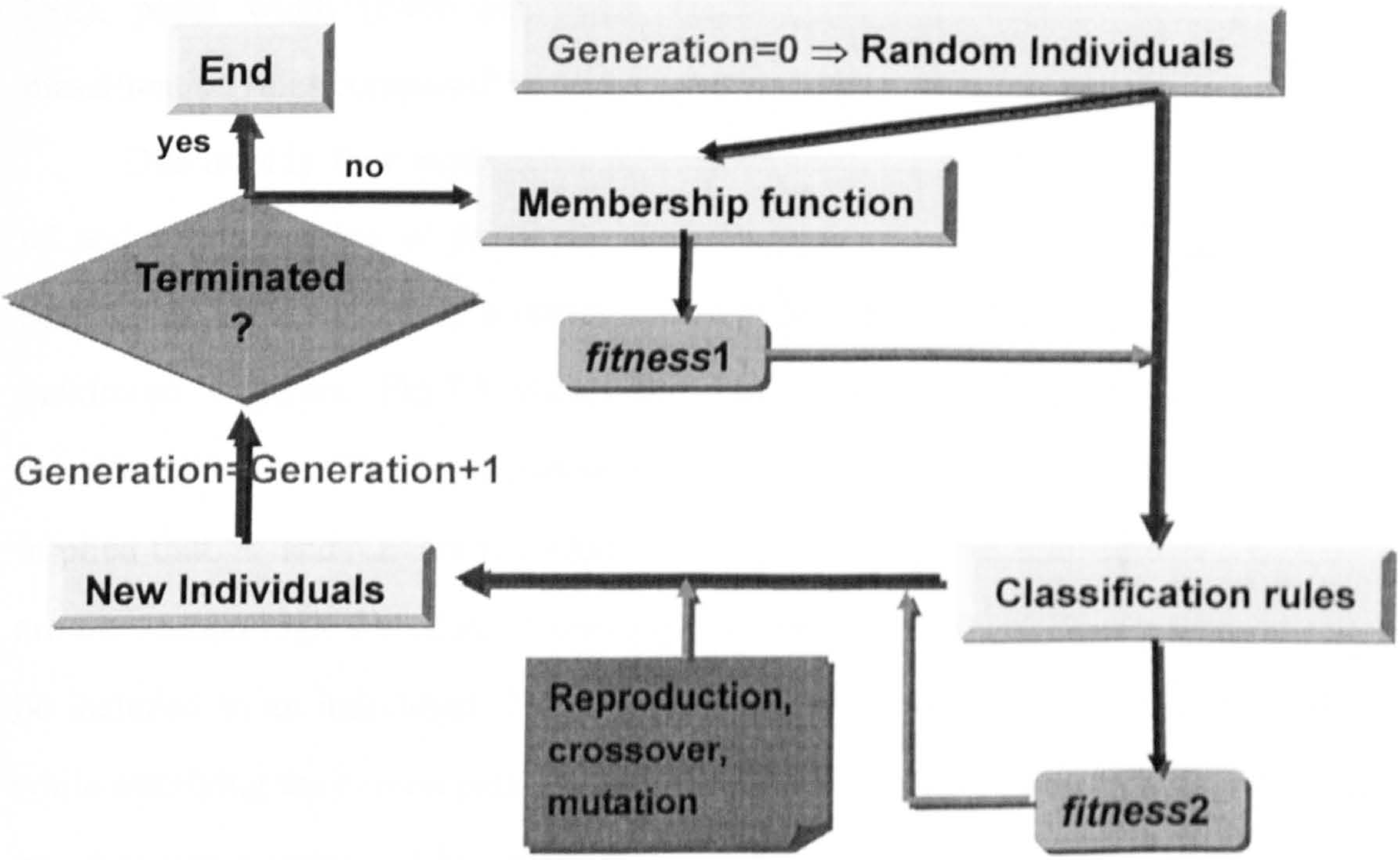


Figure 7.2 A brief configuration of fuzzy logic based GP.

population for the next generation

(8) Repeat from step (4) until the termination criteria are satisfied

7.4 METHODOLGY EVALUATION

After the description of the fuzzy logic based GP, two classification examples are utilized to evaluate the accuracy and efficiency of the method.

7.4.1 Fisher's Iris Dataset

Fisher's Iris dataset is a famous dataset to evaluate the performance of classification methods [Fisher, 1936]. The dataset consist of three species of Iris flower, setosa, versicolour and virginica. Each species includes 50 subjects and in total 150 subjects are in the dataset. The data report four characteristics: sepal width (SW), sepal length (SL), pedal width (PW) and pedal length (PL). The task is to discover the classification rules composed by four characteristics for the three species.

Due to only four attributes involved in the individuals, the maximum number of nodes (the number of functions, terminals and constants) in an individual was selected as 15. If expressed a rule as a binary tree, 15-node represents a tree with a maximum 4 layers. Fig.7.3 shows the structure of a binary tree representing classification rules. The functional set was $\{\leq, \geq, +, -, \times, /\}$. The grammar of the rules implied that '≤' and ' ≥' are regarded as the first logic functions, and '+', '-', 'x', '/' are the second logic functions. Generally, one and only one first logic function must be included in an individual. The number of second logic function can be variable while satisfying the correct grammar and the number limit of nodes. Fig.7.4 gives the tree structure constructed by grammar. The terminal set was composed of {SW, SL, PW, PL}. The probabilities of reproduction, crossover and mutation were 0.1, 0.7 and 0.2, respectively. The population size per generation was 500, and the maximum

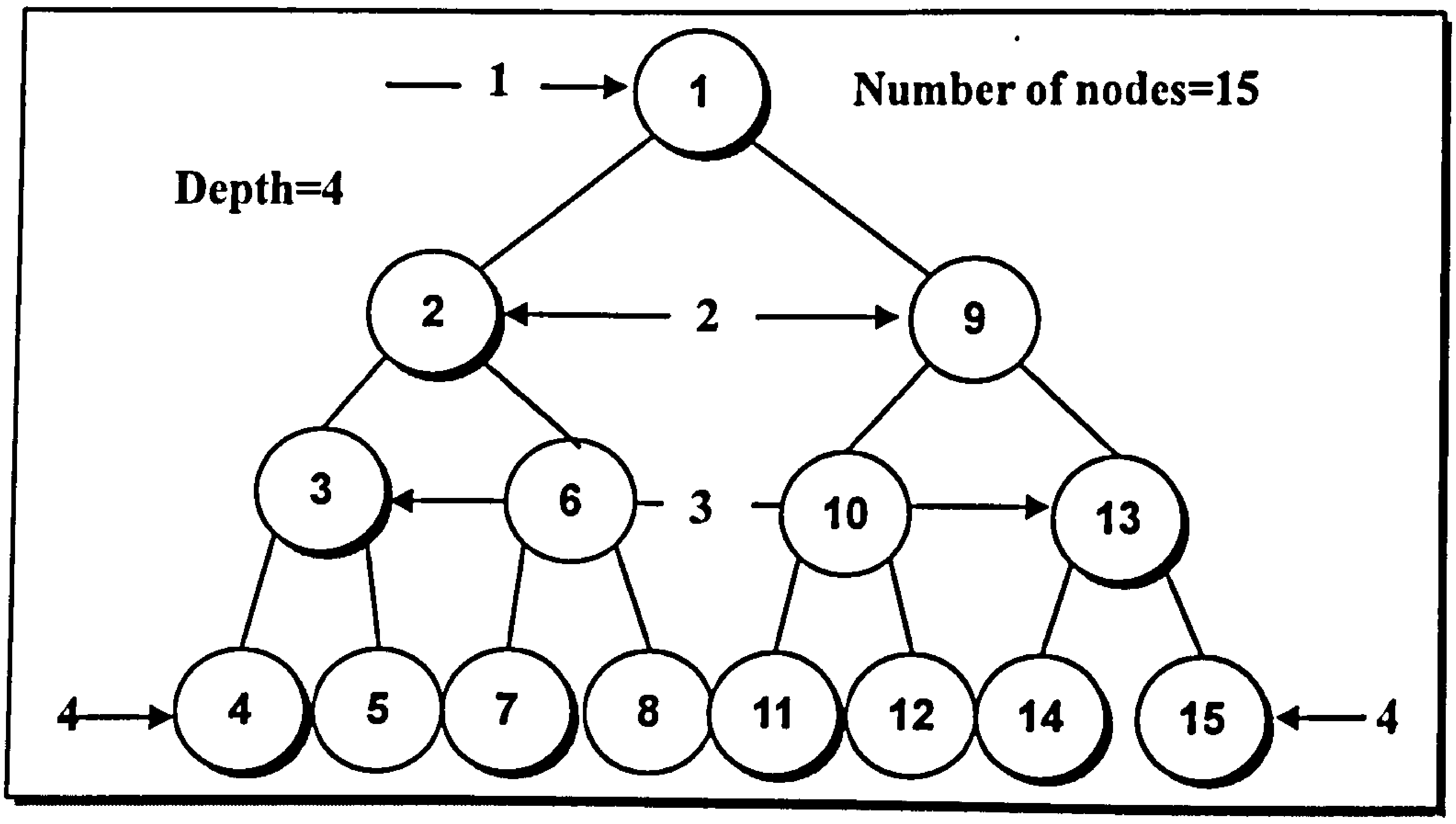


Figure 7.3 Structure of binary tree representing classification rules.

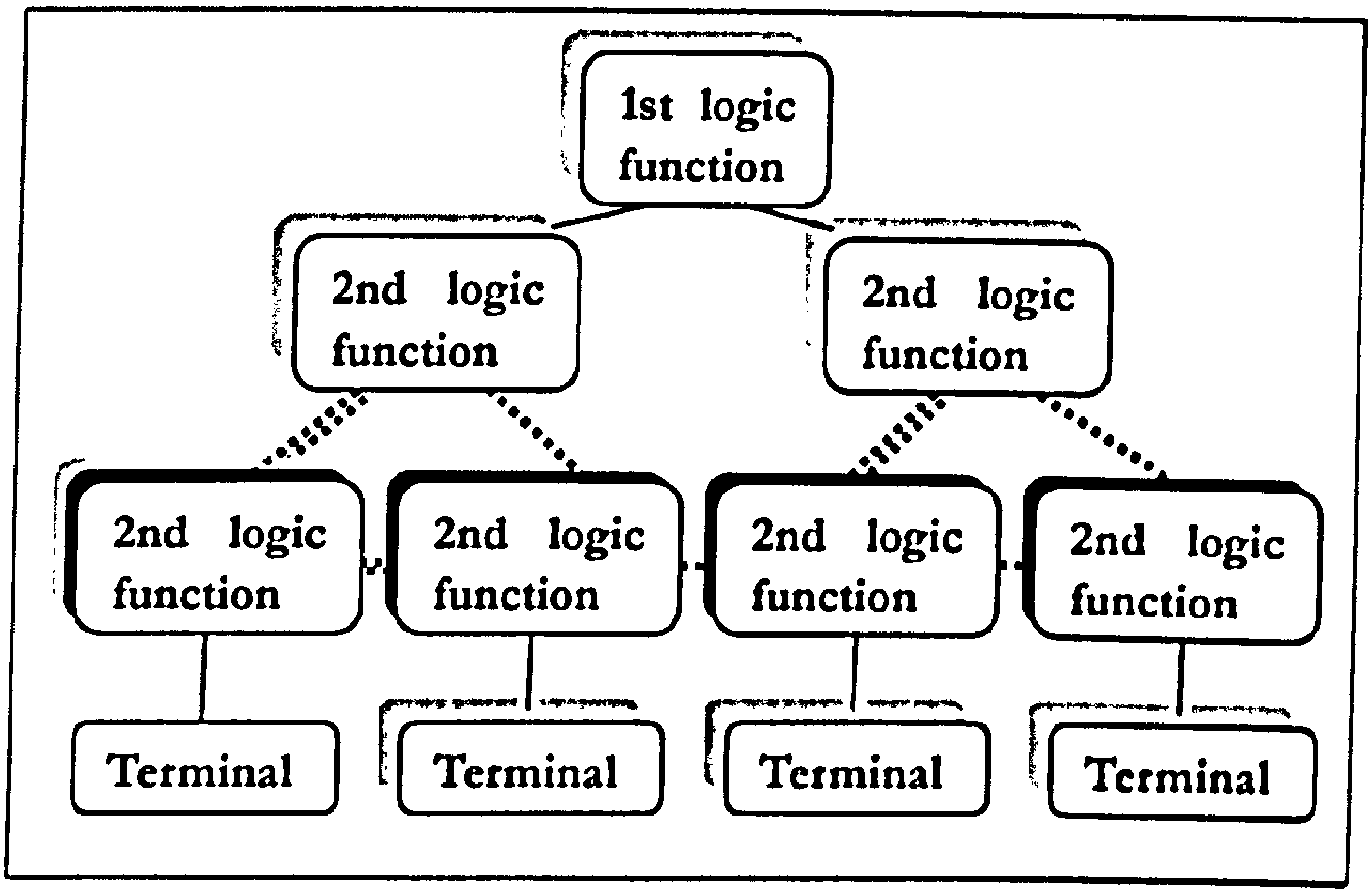


Figure 7.4 Tree structure with first and second logic functions.

generation was 500.

Initially, 150 subjects were randomly divided into two sets with equal number of subjects: training set and validation set. The fuzzy logic based GP was applied to the training set, where, $\alpha = 0.6, \beta = 0.4, \gamma = 0.8, \eta = 0.2$. One classification rule for each class was therefore discovered. The expressions of the rules and their accuracies tested on the validation set are listed in Table 7.1.

Conventional fitness, sensitivity and specificity in Eq. (7.1-3) of the classification rules were also computed, so that they could be more comparable to other classification methods. Further comparison can be done by defining two global indexes, average accuracy and overall accuracy, which are utilized by many classifiers in the literature. They can be expressed as follows:

$$\text{Average accuracy}=\frac{1}{n}\sum_{i=1}^{n}\frac{tp_i}{N_i} \tag{7.8}$$

$$\text{Overall accuracy}=\frac{\sum_{i=1}^{n}tp_i}{\sum_{i=1}^{n}N_i} \tag{7.9}$$

Where n is the total class number and N_i is the number of subjects in each class.

Table 7.1 Classification rules for Iris data.

Classification rules	Proposed fitness	Traditional fitness
IF $\frac{PL\times SW}{4.5009\times PL}\ge (SL\times PW)-(SW-0.30409)$ THEN Setosa	$fitness2$=100%; f_1=100%, f_2=0%, f_3=100%, f_4=0%;	tp=25; Se=100%; Sp=100%; $Fitness$=100%;
IF $\frac{PW-PL}{PL-4.8652}\ge\frac{SW+SL}{PW\times 7.8654}$ THEN Versicolor	$fitness2$=94.78%; f_1=96.15%, f_2=3.85%, f_3=100%,f_4=0%;	tp=25; Se=100%; Sp=98%; $Fitness$=98%;
IF $4.6444\le PW\times SL-SL$ THEN Virginica	$fitness2$=97.71%; f_1=100%, f_2=0%, f_3=96.15%, f_4=3.85%;	tp=23; Se=92%; Sp=100%; $Fitness$=92%

Table 7.2 lists the accuracies of different classifiers for the Iris data, referring to [Hong and Tseng, 1997; Lee *et al.* 2001; Lin and Chen 2001; Wang *et al.* 1999; Chien, Lin and Hong 2002]. Since the Iris data has equal number of subjects in each class, its average accuracy and overall accuracy are equal to each other. It is evident that our fuzzy logic based GP has one of the highest accuracy among other methods.

7.4.2 Wisconsin Breast Cancer Dataset

Wisconsin breast cancer dataset is another famous dataset for classification evaluation. It was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg [Mangasarian and Wolberg, 1990; Wolberg and Mangasarian, 1990; Mangasarian, Setiono, and Wolberg, 1990; Bennett and Mangasarian, 1992]. The data set includes 699 instances with breast cancer which are either benign (458 instances) or malignant (241 instances). Instances arrived periodically, thus the data set therefore reflects the chronological grouping of the data.

Table 7.2 Classification accuracy comparison of Iris data.

Classifiers	**Accuracy (%)**
GVS [Hong, Tseng, 1999]	96.0
FEBFC with 4 features [Lee *et al.*, 2001]	96.7
FRG with GA [Lin and Chen, 2001]	96.9
FEBFC with 2 selected features [Lee *et al.*, 2001]	97.1
FIL [Wang *et al.*, 1999]	97.3
Fuzzy attributes GP [Chien, Lin and Hong, 2002]	97.3
Fuzzy logic based GP	97.3

Group 1: 367 instances (January 1989)
Group 2: 70 instances (October 1989)
Group 3: 31 instances (February 1990)
Group 4: 17 instances (April 1990)
Group 5: 48 instances (August 1990)
Group 6: 49 instances (Updated January 1991)
Group 7: 31 instances (June 1991)
Group 8: 86 instances (November 1991)

Total: 699 instances (as of the donated database on 15 July 1992)

Nine attributes compose of the data set listed as follows:

1. Clump Thickness
2. Uniformity of Cell Size
3. Uniformity of Cell Shape
4. Marginal Adhesion
5. Single Epithelial Cell Size
6. Bare Nuclei
7. Bland Chromatin
8. Normal Nucleoli

Although it only involves a two-class classification problem, it can also be applied to evaluate and compare the capability of our fuzzy based GP since many test results of different classification methods are available in the literature.

The maximum number of nodes in the classification rules was 32, which is five layers of a binary tree, since more attributes are included. The functional set was similar as in the previous example: {≤, ≥, +, -, ×, /}, so were the related parameters. The terminal set was the nine attributes indicated using their abbreviations: {CT, UCSize, UCShape, MA, SECSize, BN, BC, NN, Mit}. The 699 instances were randomly divided into the training set and the validation set, thus two classification rules were discovered as shown in Table 7.3.

Table 7.3 Classification rules for Wisconsin breast cancer dataset.

IF $((UCShape+BC)-(10-NN))\times(BC\times Mit)\le(2\times Mit-UCShape)$ THEN Benign Proposed fitness: $fitness2$=89.79%; f_1=95.26%, f_2=4.74%, f_3=93.16%, f_4=6.84%; Traditional fitness: Se=96.51%, Sp=90.83%; $Fitness$=87.66%
IF $NN\ge\frac{10}{UCShape+BN}$ THEN Malignant Proposed fitness: $fitness2$=89.86%; f_1=92.31%, f_2=7.69%, f_3=100%, f_4=0%; Traditional fitness: Se=100%, Sp=95.63%; $Fitness$=95.63%

Wei *et al.* (2003) emphasized the importance of traditional sensitivity for classification performance measurement and compared the sensitivity and specificity of three classification methods, such as logistic, MLPNN and ANFIS, applied on Wisconsin data. Table 7.4 quotes the results in Wei *et al.* (2003) along with that of our method.

Table 7.4 Sensitivity and Specificity of Wisconsin breast data.

Method	Sensitivity (%)	Specificity (%)
Logistic	95.0	97.3
MLPNN	97.5	97.3
ANFIS	98.3	96.4
Fuzzy logic based GP	98.2	93.3

Although the specificity obtained in the fuzzy logic based GP is lower than others, the sensitivity is comparable to other methods. It is also worth noting that its classification rules may not have the highest sensitivity and specificity among the whole generations, because the selection of the rules depends on *fitness1* and *fitness2*.

7.5 BRAIN TUMOUR DIAGNOSIS

After processed by the first two parts of the brain tumour diagnosis system, 763 instances from 45 brain tumour patients were stored in our database. An image from a pituitary tumour patient was discarded. The images with high artifacts and with resection cavity and post-operative fluid were also removed. The data contains five types of MRI modalities, FLAIR-FSE, T_1-SE, T_2-FSE, PD-FSE and T_1-contrast. Three types of brain tumours were finally included, meningioma, (high and low) glioma and medulloblastoma. The number of attributes was ten which consist of the intensities of white matter, gray matter, CSF, brain tumour, and the intensity difference between each of them. The terminal set was therefore composed of {IW, IG, IC, IT, IWT, IWG, IWC, IGC, IGT, ICT} as in Table 7.5, where 'I' refers to the intensity, 'W' is white matter, 'G' is gray matter, 'C' is CSF, 'T' is tumour and 'IWT' is the intensity difference between white matter and tumour, and so on. The number of slices for each type of MRI is listed in Table 7.6

Table 7.5 Terminal set of brain tumour data.

IT			
IC	ICT		
IG	IGC	IGT	
IW	IWG	IWC	IWT

Table 7.6 Number of slices in each MRI modality.

Scan type→ Tumour types↓	Flair-FSE	T_1-SE	T_2-FSE	PD-FSE	T_1-contrast
meningioma	63	12	84	11	14
high glioma	60	60	198	32	12
low glioma	24	58	71	21	10
medullo-blastoma	8	8	9	0	8

The functional set in this case was $\{And, Or, Not, \geq, \leq\}$. The terminating criterion was reaching the maximum number of generations (100). The population size of each generation was set to 100. The maximum depth of individual was 5 and totally 32 nodes in each individual. $\alpha = 0.6, \beta = 0.4, \gamma = 0.8, \eta = 0.2$.

As stated previously, for each tumour class, meningioma, glioma (low grade and high grade) and medulloblastoma, *fitness1* divided the individuals into the classes they classified and completed the classification rules, and *fitness2* selected the best individual for each class. The best classification rules are as follows (all values have been divided by 1000 to make the rules more manageable):

FLAIR-FSE:

$IF\ (IGC \geq 23.3775) And (IGM \geq 30.9807) And (Not(ITC \geq 12.4049))$
$THEN$ Meningioma

fitness2=66.97%; f_1=80%, f_2=20%, f_3=80.67%, f_4=19.33%;

$IF\ (ITW \leq 17.8935) And (ITI \leq 44.6204) And (ITC \geq IGM) And (ITC \leq 34.2346)$
$THEN$ High grade glioma

$fitness2=87.33\%; f_1=100\%, f_2=0\%, f_3=77.95\%, f4=22.05\%;$

$IF\ (Not(IWG \geq ITI)Or(IWC \leq 21.1934))$
$THEN$ Medulla blastoma

$fitness2=89.69\%; f_1=100\%, f_2=0\%, f_3=82.20\%, f4=17.80\%;$

$IF\ (IGM \leq 32.0818)And(Not(IWM \leq 31.87081))$
$THEN$ Low grade glioma

$fitness2=88.94\%; f_1=100\%, f_2=0\%, f_3=80.87\%, f4=19.13\%;$

T_1-SE:

$IF\ (Not(IWC \geq 28.1206))And((IWM \geq 33.4338)And(IGC \leq 7.5124))$
$THEN$ Meningioma

$fitness2=62.88\%; f_1=66.67\%, f_2=33.33\%, f_3=99.32\%, f4=0.68\%;$

$IF\ (Not(ITC \geq 7.6964))And((IWC \leq 8.1291)And(ITG \geq 29.9638))$
$THEN$ High grade glioma

$fitness2=91.22\%; f_1=100\%, f_2=0\%, f_3=84.93\%, f4=15.07\%;$

$IF\ ((ITG \geq 26.901)And(IWC \geq 18.2416))And(Not(IWC \leq 10.8082))$
$THEN$ Medulla blastoma

$fitness2=98.57\%; f_1=100\%, f_2=0\%, f_3=97.60\%, f4=2.40\%;$

$IF\ ((IGM \leq 30.8435)And(IGC \leq 10.4578))And(ICSF \leq 33.4338)$
$THEN$ Low grade glioma

$fitness2=71.37\%; f_1=78.50\%, f_2=21.50\%, f_3=93.62\%, f4=6.38\%;$

T_2-FSE:

$IF\ (Not(IWM \geq 24.3506))And(IGM \leq ITG)$
$THEN$ Meningioma

$fitness2=92.04\%; f_1=100\%, f_2=0\%, f_3=86.38\%, f4=13.62\%;$

$IF\ ((IGC \leq 13.9397)And(IWG \leq 8.0231))And(IWM \geq 26.21)$
$THEN$ High grade glioma

$fitness2$=65.42%;f_1=84.62%,f_2=15.38% ,f_3=67.41%,$f4$=32.59%;

$IF\ ((ITG \geq 33.6057) And (IGM \leq 34.118)) And (Not (IWC \geq 25.377))$
THEN Medulla blastoma

$fitness2$=87.36%;f_1=93.75%,f_2=6.25% ,f_3=92.13%,$f4$=7.87%;

$IF\ (Not (IWM \leq 25.2864)) And ((IWC \geq 25.7411) And (ITG \geq 20.2199))$
THEN Low grade glioma

$fitness2$=71.23%;f_1=91.67%,f_2=8.33% ,f_3=64.94%,$f4$=35.06%;

PD-FSE:

$IF\ (Not ((ITG \leq 34.2609) And (ICSF \leq 23.9588)))$
THEN Meningioma

$fitness2$=96.44%;f_1=100%,f_2=0% ,f_3=94%,$f4$=6%;

$IF\ (Not (ICSF \geq 19.9868))$
THEN High grade glioma

$fitness2$=93.97%;f_1=100%,f_2=0% ,f_3=89.74%,$f4$=10.26%;

$IF\ (Not ((IGC \geq 15.7559) Or (IGM \leq 34.3851)))$
THEN Low grade glioma

$fitness2$=70.43%;f_1=84.38%,f_2=15.63% ,f_3=78.95%,$f4$=21.05%;

T_1-Constrast:

$IF\ (Not (ICSF \leq 20.6497)) And (Not (ICSF \leq 8.8551))$
THEN Meningioma

$fitness2$=95.78%;f_1=100%,f_2=0% ,f_3=92.86%,$f4$=7.14%;

$IF\ (Not (Not (ITI \geq 17.0341))$
THEN High grade glioma

$fitness2$=97.87%;f_1=100%,f_2=0% ,f_3=96.43%,$f4$=3.57%;

$IF\ ((IWG \leq 5.8745) And (ICSF \leq 15.0941))$
THEN Medulla blastoma

fitness2=93.22%;f_1=100%,f_2=0% ,f_3=87.10%,f_4=12.90%;

IF (*Not*(*IWG* $\leq$ 5.8745))
THEN Low grade glioma

fitness2=90.15%;f_1=100%,f_2=0% ,f_3=83.02%,f_4=16.98%;

The recognition rate of each MRI modality is summarized in Table 7.6. Although certain types of tumour had relatively low accuracy in classification, the classification rate in all was quite promising. It suggests that the idea about brain tumour diagnosis based on MR images is not only feasible but very effective.

Table 7.7 Accuracy of brain tumour diagnosis.

(%)	FLAIR-FSE	T_1-SE	T_2-FSE	PD-FSE	T_1-contrast
Average Accuracy	81.50	91.23	78.27	89.54	91.37
Overall Accuracy	80.24	86.41	72.61	89.31	92.26

7.6 CONCLUSIONS

In this chapter, the final part of brain tumour diagnosis system has been presented. The database, including brain tissue intensities obtained from previous two stages, was used to provide useful information about diagnosis. Since there was a great number of data available, data mining and knowledge discovery techniques were employed to complete the job. After a brief review in data mining category, GP was appealing due to its popular application, especially in medical domain.

Brain tumour diagnosis was regarded as a multi-class classification problem. However, most classifiers, including GP, cannot deal with the multi-class classification problem directly but convert it into several binary-class classification problems. It sacrifices the computing time dramatically.

A new classification method which combined two typical data mining techniques, fuzzy logic and GP was developed for this thesis. It can classify multi-class directly without any transformation. Fig.7.5 shows the classification strategies of both conventional GP and the proposed fuzzy logic based GP. Clearly, the conventional GP creates a rule (individual) set only for one class each time and finds a classification rule for this class, and requires iterations of GP to classify other classes. The fuzzy logic based GP only creates one rule (individual) set during whole

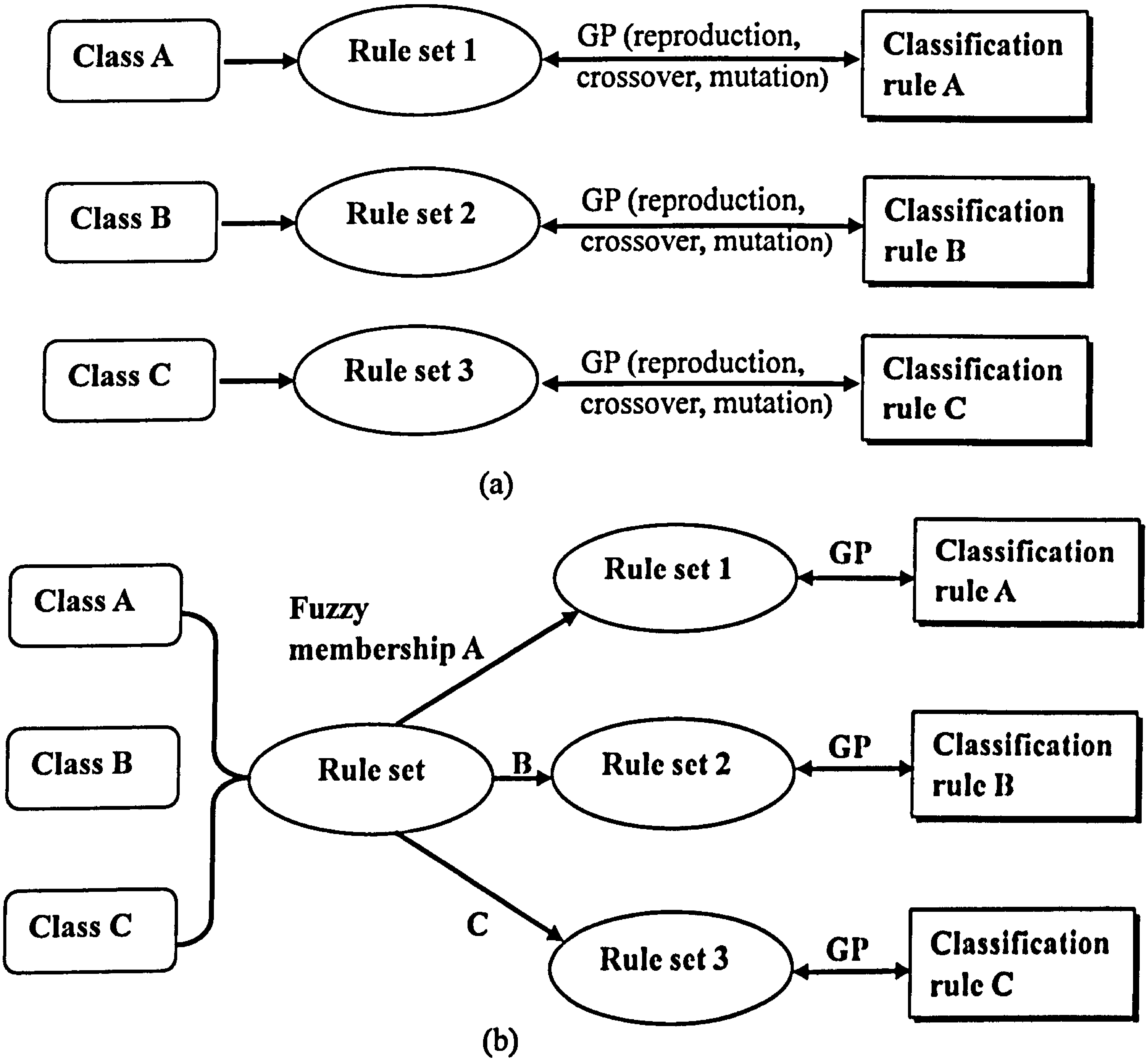


Figure 7.5 Classification strategies. (a) conventional GP (b) fuzzy logic based GP

classification. A fuzzy membership was defined to measure the classification ability of each individual to each class. The individuals are divided into subgroups in terms of their highest fuzzy memberships. The highest membership refers to the class that the individual should classify. The number of groups is equal to the number of classes. One classification rule from each subgroup can be discovered by GP operations simultaneously.

The characteristics of medical data were considered in the fuzzy logic based GP. Instead of using the traditional fitness function, two fitness functions were introduced. The first fitness was acquired directly from the highest fuzzy membership of each individual. The second fitness contained four factors which consider that to diagnose the patient has a disease is more important than to diagnose a healthy person having no this disease; to diagnose the incorrect disease is much worse than being unable to diagnose a disease. Therefore, more appropriate classification rules could be found in medical domain.

Fisher's Iris data and Wisconsin breast cancer dataset were both employed to test the new method. The classification accuracies achieved for both datasets were very comparable or better than those of other classification methods. It indicated the good classification capability of fuzzy logic based GP and therefore it should be applied to the database for brain tumour diagnosis. Classification rules for each MRI modalities and each type of brain tumour were discovered. Three types of brain tumours, glioma, meningioma and medulloblastoma were diagnosed and the grade of glioma was also diagnosed. The accuracies of the classification rules were quite promising. Ultimately, the feasibility of brain tumour diagnosis non-invasively but based on MR image has been significantly demonstrated. Although the classification rules obtained cannot apply in clinics at this stage since more details should be considered, their further effectiveness can be highly expected.

Chapter 8

CONCLUSIONS AND FUTURE WORK

8.1 CONCLUSIONS

As we know, a brain tumour is a very serious disease with a very low survival rate. Thousands upon thousands people are struck by this disease each year. Brain tumours are typically categorized by the tissue of origin. The most common brain tumour type is glioma which makes up approximately 50% of all primary brain tumours. Following are meningioma comprising approximately 25%, pituitary tumour making up 10% and acoustic neuromas comprising 7.5%. Other rare tumour types occupy the remaining 7.5%. Detecting and diagnosing brain tumours quickly and accurately is essential to carry out effective treatments. The prevalent diagnosis process in clinics is composed of three steps: neurological exam, brain image analysis and biopsy. A biopsy is a surgical procedure to take a sample tissue from the suspected tumour to determine the exact diagnosis such as malignance and tumour types. However, this procedure can cause a great deal of pain to the patients and is also difficult for clinicians to do. Some brain tumours cannot have a biopsy, such as brain stem glioma, because removing any healthy tissue from the brain stem can affect vital functions. A non-invasive diagnosis method is anyhow a better way to both brain tumour patients and clinicians.

With the development of imaging techniques, it greatly extends the range of human vision into realms that would otherwise be inaccessible, such as the anatomic structure inside human body. MRI is currently an indispensable diagnostic imaging technique for the early detection of any abnormal changes in tissues and organs, due to its fairly good resolution for different tissues. The significance of the technique indicates that the intensities in MR images may contain more important information than we ever expect. It may not only distinguish brain tumours from normal tissues, but also distinguish the tumour types. Siromoney *et al.* (2000) has firstly discriminated two types of brain tumours, meningioma and astrocytoma merely based on MR images. It substantiated the possibility of using MR image based diagnosis of brain tumours.

In this thesis, a brain tumour diagnosis system based on MR images was proposed. The first chapter explained the reasoning behind developing the diagnosis system and the organization of the thesis. The second chapter briefly introduced the characteristics of MRI. The third chapter reviewed brain tumour types and their general diagnosis methods.

Since a great number of MR images are created everyday in hospitals and added to those images already stored, it is impossible to extract diagnosis knowledge manually. Even so, this would most likely be incomplete. Data mining and knowledge discovery techniques were described in the fourth chapter, which are ideal methods for excavating useful information from huge amounts of data. Typical methods of data mining and the process of knowledge discovery were introduced.

The non-invasive brain tumour diagnosis system was constructed by three major parts. Each part was presented in following three chapters. The fifth chapter addressed the first part, image preprocessing. It was applied to the original MR images from the hospital. Generally, MR images have no uniformed intensity scales especially on images from different patients. The nonuniformity of intensity scales leads to the difficulty of automatic implementation in any image processing stages and also the incomparability of tissue intensities in different images. Therefore, the principal step of image preprocessing is standardizing the intensity scales by means of statistical analysis. A novel and simple standardization method based on the statistic features of intensities was proposed. The distribution of intensities in brain MR images was described using a sub-normal distribution. This distribution has a standard deviation four times that of normal distribution. Its mean also sometimes has a linear displacement to that of normal distribution. Depending on the characteristics of MR intensities, all MR images acquired for this thesis were transformed into new images, which had a predetermined mean and standard deviation, 32500 and 10000, respectively. These specific values of mean and standard deviation not only scaled the intensities of MR images into similar distribution shapes but also unified the images displayed as unsigned 16-bit format with full resolution. The standardization did not require a prior or post template as in other standardization methods and also greatly retained the major features of intensities while transforming them into a similar range.

Following standardization, the non-brain removal was proposed which was the second step of image preprocessing. The existence of non-brain regions can extremely influence the diagnosis because they have similar intensity values to brain tumours or CSF in some MRI modalities. Morphological operations were employed to complete this task. Since morphological operations always require binary images, a threshold corresponding to the intensity value having the most number of pixels was chosen to generate the binary images from the standardized images. The binary images were then eroded until the non-brain regions could be labeled and separated from the brain. Dilation was applied to thicken the non-brain regions that were thinned down through erosion, followed by a processing to mask the whole non-brain regions out of the original MR images. The occasional failure using this method was compensated by the supplementary method.

The third step of image processing was the enhancement of tissue contrast using histogram equalization. The contrast between white matter and gray matter is always quite close which may cause difficulty in further segmentation. Only the intensities lying in a specified range referring to white matter and gray matter were treated without changing other pixels in the image. A typical image from each MRI modality used in this thesis was presented to demonstrate the great effects of each step in image preprocessing stated above.

The second major part of the diagnosis system, image segmentation was presented in chapter six. It was applied to preprocessed MR images from the first part. Accurate segmentation of brain tissues can result in proper diagnosis of brain tumours. A variety of image segmentation methods in the literature have been reviewed, such as thresholding, region growing, edge-based segmentation and clustering. The FCM clustering algorithm, belonging to the clustering category, is one of the most popular segmentation methods for MR images. However, the drawback of FCM algorithm, as with many other intensity-based segmentation methods, is its sensitivity to noise. Since medical images contain uncertain information, noise and even artifacts, the intensity-sensitive methods may not segment medical images correctly. On the basis of the traditional FCM algorithm, an improved segmentation algorithm called IFCM was developed to solve this problem. A neighbourhood attraction including two factors was considered in the IFCM algorithm. One is called feature attraction. The feature in MR images represents pixel intensity. The other is distance attraction. The former indicates that the attraction emerges depending on the intensity difference among pixels. The lower the intensity difference is, the stronger the feature attraction. The latter indicates that the attraction also emerges depending on the spatial distance among pixels. The closer the pixels are, the stronger the distance attraction. By considering neighbourhood attraction, segmentation is therefore not only determined by pixel intensity itself but also by the intensities of neighbour pixels and the structure of its neighbourhood. It greatly reduces the sensitivity of segmentation to noise. A neural network model was presented to search the optimized degree for both feature attraction and distance attraction. Different images may require various extents of the two attractions. Other approaches also based on FCM, such as NC, PCM and RFCM were described and compared to IFCM. Synthetic square images and Lena facial image with different noise were used to test all methods for segmentation. Besides the better visual effect of IFCM than other methods, five parameters were defined to confirm the superiority of IFCM algorithm quantitatively. Due to no ground truth being available to MR images, simulated brain MR images whose anatomic structure is known were downloaded from Brainweb. Typical stimulated images from T_1-weighted, T_2-weighted and PD-weighted categories with different noise levels were segmented using IFCM along with other methods. The best performance and highest accuracy of IFCM were suggested by visual segmentation results and the evaluation parameters obtained. After completing all tests successfully, the real brain MR images could finally be segmented using IFCM algorithm. A typical image of each MRI modality from the images acquired was shown. Each brain image was segmented into white matter, gray matter, CSF and brain tumour, while some T_1-weighted images displayed no brain tumour. The average intensity of each part in each image was then easily calculated and saved in a database for further diagnosis. The last section of this chapter described how the IFCM was extended to estimate the most popular artifacts, RF inhomogeneity in MR images. Its capability has been demonstrated on both synthetic images and simulated MR images corrupted by artifacts.

The seventh chapter outlined the last major part, brain tumour diagnosis. The database including brain tissue intensities obtained from the previous two stages was used to extract useful information about diagnosis. Since there was a great number of data available, data mining and knowledge discovery techniques were employed to complete the job. After brief review in data mining category, GP appealed due to its popular application especially in medical domain. Brain tumour diagnosis was regarded as a multi-class classification problem. However, most classifiers including GP cannot deal with multi-class classification problem directly but convert it into several binary-class classification problems. It sacrifices the computing time dramatically. A new classification method combining two typical data mining techniques, fuzzy logic and GP, was developed in this chapter. It can classify multi-class directly without any transformation. Conventional GP creates a rule (individual) set only for one class each time and finds a classification rule for this class, and requires iterations of GP to classify other classes. The fuzzy logic based GP only creates one rule (individual) set during whole classification. A fuzzy membership was defined to measure the classification ability of each individual to each class. The individuals are then divided into subgroups in terms of their highest fuzzy memberships. The highest membership refers to the class that the individual should classify as. The number of groups is equal to the number of classes. One classification rule from each subgroup was discovered by GP operations. The characteristics of medical data were considered in the fuzzy logic based GP. Instead of using traditional fitness function, two fitness functions were defined. The first fitness is acquired directly from highest fuzzy membership of each individual. The second fitness contains four factors which consider that diagnosing the patient has a disease is more important than diagnosing a healthy person has no this disease; and diagnosing a disease as another one is much worse than being unable to diagnose a disease. Therefore, more appropriate classification rules can be found in medical applications.

Fisher's Iris data and Wisconsin breast cancer dataset were both employed to test the new method. The classification accuracies achieved for both datasets were comparative or better than those of other classification methods. It indicated the good classification capability of fuzzy logic based GP and it therefore should be applied to the database for brain tumour diagnosis. Classification rules for each MRI modalities and each type of brain tumour were discovered. Their accuracies were quite promising. Ultimately, the feasibility of brain tumour diagnosis based on MR image has been significantly suggested. Although the classification rules obtained cannot be applied in clinics at this stage since more details should be considered, their further effectiveness can be highly expected.

In summary, the primary achievement of this thesis was constructing a three-part non-invasive brain tumour diagnosis system. Novel methods were developed in each part of the system whose advantages over other methods were demonstrated respectively. The diagnosis results greatly enhanced the superiority of this diagnosis system with promising accuracies.

8.2 FUTURE WORK

Although the accuracies of brain tumour diagnosis were quite encouraging in this thesis, there are still wide gaps to clinical applications. The precision of diagnosis is so crucial and any mistakes can lead to serious consequences. Due to complex and diversified structure of different brains, more details may be considered to manipulate all sorts of circumstances that could happen. Future work may be addressed to following aspects:

(1) More detailed anatomy of brain should be concerned, instead of only considering white matter, gray matter, CSF and brain tumours. Perhaps some other parts of brain are the sticking points for brain tumour diagnosis.

(2) The volume and shape of brain tumours may be taken into account. They could also imply the tumour types.

(3) The algorithms in this thesis were applied on 2D independent slices. The more practical implementation should be applying the algorithms to 3D images. Manually segmentation of 3D MR images by expert may be acquired to evaluate the performance of IFCM.

(4) If different MRI modalities can be combined, higher diagnosis accuracy may be achieved. In the data source for the thesis, every tumour patient only had two or three types of MRI scan. These two or three types were different between some patients. Therefore, each classification rule of brain tumour was discovered just from one type of MRI modality. Suppose each patient has all MRI types involved, the classification rules would incorporate information from different MRI modalities. They would therefore be more comprehensive and meaningful.

(5) Only transverse images were used in this thesis. Coronal and sagittal images should be combined to improve the diagnosis.

Another addition would be to apply the extended IFCM algorithm on real MR images for bias field estimation. MR images affected by artifacts were excluded in the processing, but after corrected by the extended IFCM, these images may then be of use.

A similar application of the diagnosis system would be to distinguish benign and malignant brain tumours. It is as important as the determination of tumour types, because the property of tumours indicate whether the tumours spread out of the brain or not. Their severities are not equal. The only difference in applying this system is to collect MR images grouped into benign and malignant tumours. Beyond all doubt, it would be a great approach to conduct using the diagnosis system proposed. The classification of tumour or not (edema, signs of ageing) may be also an application.

All algorithms were implemented in MATLAB. As MATLAB is only powerful in easy mathematical coding, it needs rather long computing time. Programmes may be converted to C or C^{++} for speeding up.

Owing to schedules, high workload and limits of data, the brain tumour diagnosis system is not fully automatic, but works in parts. The work involved is mainly a feasibility study rather than the development of practical software. The most expected development would be to produce an automatic diagnosis system to participate in clinical diagnosis.

REFERENCES

AccessMed Health Information Library (2002), "Brain biopsy," Retrieved 28th May 2003. http://www.ehendrick.org/healthy/002033.htm

Acton S. T. and Mukherjee D. P. (2000), "Scale space classification using area morphology," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 623-635.

Ahmed M. N., Yamany S. M., Frag A. A., and Moriarty T. (1999), "Bias field estimation and adaptive segmentation of MRI data using a modified fuzzy c-means algorithm," *Computer Vision and Pattern Recognition*, vol. 1, pp. 1250-1255.

Alvira M. (1997), "Case 123-frontal lobe tumour," *Neuropathology*, Retrieved 19th May 2004, http://path.upmc.edu/cases/case123.html

American Accreditation HealthCare Commission (A. M. A. D) Inc (2002), "Brain tumors: primary," *Well connected series*, Retrieved 27th May 2003. http://www.healthandage.com

Andre D. (1994), "Learning and upgrading rules for an OCR system using genetic programming," *Proceedings of 1994 IEEE World Congress on Computational Intelligence*, pp. 462-467.

Andre D. Bennett III F. H. and Koza J. R. (1996), "Discovery by genetic programming of a cellular automata rule that is better than any known rule for the majority classification problem," *Proceedings of 1st International Conference on*

Genetic Programming, pp. 3-11.

Andrews M. and Prager R, (1994), "Genetic programming and emergent intelligence," In Kinnear, K. E. (Eds.), *Advances in Genetic Programming*, pp. 75-98, MIT Press.

Apte C. and Hong S. J. (1996), "Predicting equity returns from securities data with minimal rule generation," *Advances in knowledge discovery and data mining*, AAAI press.

Assanelli D, Cazzamalli L, and Stambini M. (1993), "Correct diagnosis of chest pain by an integrated expert system," *Proceedings of Computers in Cardiology*, pp. 759-762.

Aziz A and Uetani M. (2002), "MRI of the foot," Retrieved 23rd May 2003 from Nagasaki University School of Medicine, Department of Radiology website: *http://www.med.nagasaki-u.ac.jp/radiolgy*

Ballinger J. R. (1996), "Introduction to MRI," Retrieved 21 May 2003 from MRI Tutor Website and Radiology Teaching files website: *http://www.mritutor.org/mritutor/index.html*

Bauer P., Nouak S. and Winker. R (1996), "A brief course in fuzzy logic and fuzzy control," Retrieved 4th June 2003 from: *http://www.mines.inpl-nancy.fr/~gueniffe/CoursEMN/I31 /fcourse/fuzzy.html*

Beardsley S. (2004), Sandy's homepage, Website, Accessed 19th May 2004 *http://www.glaciergraphics.net/sandy/sandys_mri.html*

Bennett A. (2000), "Modeling evolving populations," Retrieved 3rd June 2003 from University of Sourthampton, Department of Electronics and Computing Science website: *http://www.isis.ecs.soton.ac.uk/isystems/evolutionary/notes/evol/*

Bennett K. P. and Mangasarian O. L. (1992), "Robust linear programming discrimination of two linearly inseparable sets,", *Optimization Methods and Software*, Gordon & Breach Science Publishers, pp. 23-34.

Bergadano F., Giordana A. and Saitta L. (1991), *Machine Learning - an integrated framework and its applications*, Ellis Horwood Limited, England.

Beyer H. G. Brucherseifer E., Jakob W., Pohlheim H., Sendhoff B. and To T. B. (2002), "Evolutionary algorithm glossary," Retrieved 3rd June 2003 from University of Dortmund, Department of Computer Science website: http://ls11-www.cs.uni-dortmund.de/people/beyer/EA-glossary/def-engl-html.html

Bezdek J. C. (1981), *Pattern Recognition with fuzzy object function algorithms*, Plenum Press.

Bojarczuk C., Lopes H. and Freitas A. (1999), "Discovering comprehensible classification rules using genetic programming: a case study in a medical domain," *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-99)*, pp. 953-958.

Bojarczuk C., Lopes H. and Freitas A. (2000), "Genetic programming for knowledge discovery in chest-pain diagnosis. *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, pp. 38-43.

BrainWeb (2003), www.bic.mni.mcgill.ca/brainweb/, website. Accessed 22nd Sep. 2003.

Breunig M. M. (1995), "Location independent pattern recognition using genetic programming," *Genetic Algorithm at Stanford*, Stanford Bookstore.

Brill E. (1994), "Some advances in transformation-based part of speech tagging," *American Association for Artificial Intelligence*, vol. 1, pp. 722-727.

Burgess C. (1999), "Evolutionary computing," Retrieved 3rd June 2003 from University of Bristol, Department of Computer Science website: http://www.cs.bris.ac.uk/%7Ecolin/evollect1/evollect0/index.htm

Bushong S. C. (1995), *Magnetic Resonance Imaging*, Mosby-Year Book InC.

Cannon R. L., Dave J.V. and Bezdek J. C. (1986), "Efficient implementation of the fuzzy c-means clustering algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 2, pp. 248-255.

Cheriet M., Said H. N. and Suen C.Y. (1998), "A recursive thresholding technique for image segmentation', *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 918-920.

Chien B. C., Lin J. Y. And Hong T. P. (2002), "Learning discriminant functions with fuzzy attributes for classification using genetic programming," *Expert Systems with Applications*, vol. 23, pp. 31-37.

Cho Z. H and Ahn I. S. (1975), "Computer algorithm for the tomographic image reconstruction with x-ray transmission scans," *Computers and Biomedical Research*, vol. 8, pp. 8-25.

Cios K., Pedrycz W. and Swiniarski R. (1998), *Data Mining Methods for Knowledge Discovery*. Kulwer Academic Publishers, Boston.

Cios K., Teresinska A., Konieczna S., Potocka J. and Sharma S. (2000), "A knowledge discovery approach to diagnosing myocardial perfusion," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, pp.17-25.

Clark P. and Boswell R. (1989), "The CN2 induction algorithm," *Machine Learning*, vol. 3, no. 4, pp. 261–283.

Cocosco C. A., Killokian V., Kwan R. K., and Evans A. C., (1997) "Brainweb: Online Interface to a 3D MRI Simulated Brain Database," *NeuroImage*, vol. 5, no.4, part 2/4, S425.

Coleman G. B. and Andrews H. C. (1979), "Image segmentation by clustering," *Proceedings of IEEE*, vol. 67, no. 5, pp. 773-785.

Collins D. L., Zijdenbos A. P., Kollokian V., Sled J. G., Kabani N. J., Holmes C. J., and Evans A. C. (1998), "Design and Construction of a Realistic Digital Brain Phantom," *IEEE Transactions on Medical Imaging*, vol.17, no.3, pp.463-468.

Dagan I., Karov Y., and Roth D. (1997), "Mistake-driven learning in text categorization," *Proceedings of The Second Conference on Empirical Methods in Natural Language Processing*, pp. 55-63.

Dave R. N. and Krishnapuram R. (1997), "Robust clustering methods: a unified view," *IEEE Transactions on Fuzzy System*, vol. 5, no.2, pp. 270-293.

Dave R. N. (1991), "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, pp. 657-664.

David C. and Munson J. R. (1996), "A note on lena," *IEEE Transactions on Image Processing*, vol.5, no. 1.

Dawant B. W., Zijdenbos A. P., and Margolin R. A. (1993), "Correction of intensity variations in MR Images for computer-aided tissue classification", *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp.770-781.

Douglas T. S. (1998), *Imaging and Image Processing in Trans-femoral Prosthetics*, Ph.D thesis, University of Strathclyde, UK.

Dutta, S. (1993). *Knowledge Processing & Applied Artificial Intelligence.* Butterworth-Heinemann, England

Edmonds A. N., Burkhardt D. and Adjei O. (1995), "Genetic programming of fuzzy logic production rules," *Proceedings of IEEE Conference on Evolutionary Computation,* Australia, vol. 2, pp. 765-770.

Ellard D. (2003), "History of MRI," Poster Presentation, Retrieved 19 May 2003 from University of Manchester, Imaging Science and Biomedical Engineering website: *http://www.isbe.man.ac.uk/personal/dellard/dje/history_mri/*

EMRF Foundation (2003), "A short history of magnetic resonance imaging from European point of view," Presentation, EMRF Online. Retrieved 19th May 2003. *http://www.emrf.org*

Fayyad U., Shapiro P. G. and Smyth P. (1996a), "From data mining to knowledge discovery in databases," *American Association for Artificial Intelligence,* pp. 37-54.

Fayyad U. Shapiro P. G., Smyth P. and Uthuruasmy R. (1996b), *Advances in Knowledge Discovery and Data Mining.* AAAI Press/ The MIT Press, London.

Feelders A., Daniels H. and Holsheimer H. (2000), "Methodological and practical aspects of data mining," *Information & Management,* vol. 37, pp. 271-281.

Fisher, R. A. (1936), "The use of multiple measurements in taxonomic problems," *Annals of Eugenics,* vol. 7, pp. 179-188.

Fonar Corporation (2003), "MRI glossary," Retrieved 21st May 2003. *http://fonar.com/glossary.htm*

Freitas A. A. (1997), "A genetic programming framework for two data mining tasks:

classification and generalized rule induction," *Proceedings of 2nd Annual Conference of Genetic programming*, San Francisco, California, pp. 96-101.

Gaborski R. S. (2001), "Binary morphological operation," Retrieved 24th June from Rochester Institute of Technology, Department of Computer Science website: *http://www.cs.rit.edu/~rsg/Morphological Processing.pdf*

Galit K. F. (2001), Medical Encyclopedia, Retrieved 27th May 2003 from United States National Library of Medicine website: *http://www.nlm.nih.gov/medlineplus/ency*

Gibbs P., Buckely D. L., Blackband S. J. and Horsman A. (1996), "Tumour volume determination from MR images by morphological segmentation," *Physics in Medicine and Biology*, vol. 41, pp. 2437-2446.

GoH, and Huang K. (1994), *Wavelet transform based image and video coding*, Ph.D. thesis, University of Strathclyde, UK

Gonzalez R.C. and Wintz P. (1977), *Digital Image Processing*, Addison-Wesley Publishing Company.

Hall L. O., Bensaid A. M., Clarke L. P., Velthuizen R. P., Silbiger M. S., and Bezdek J. C. (1992), "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain", *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 672-682.

Handley S. (1993), "Automatic learning of a detector for alpha-helices in protein sequences via genetic programming," *Proceedings of 5th International Conference on Genetic Algorithms*, pp. 271-278.

Heckerman D. (1997), "Bayesian networks for data mining," *Data Mining and Knowledge Discovery*, Kluwer Academic Publisher, pp. 79-119. Retrieved 2nd June

2003 from: *http://www.cs.unr.edu/~sushil/class/ml/papers/local/bayes.pdf*

Hesselink J. R. (2003), "Basic principles of MR imaging," Retrieved 21st May 2003 from University of California, Department of Radiology website: *http://spinwarp.ucsd.edu/NeuroWeb/*

Hiker H. and Beasley D. (2000), "Guide to evolutionary computing," Retrieved 30 May 2003 from University of Birmingham, Department of Computer Science website: *http://www.cs.bham.ac.uk/Mirrors/ftp.de.uu.net/EC/clife/www/*

Holland J. H. (1975), *Adaptation in Natural and Artificial Systems*, University of Michigan Press.

Hornak J. P. (1996), "The basic of MRI," Retrieved 21st May 2003 from Rochester Institute of Technology, Center for Imaging Science website: *http://www.cis.rit.edu/htbooks/mri/index.htm*

Jagielska I., Mattews C. and Whitfort T. (1999), "An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems," *Neurocomputing* vol. 24, pp. 37-54.

Jelinek F. (1998). *Statistical Methods for Speech Recognition*. The MIT Press, Cambridge, Massachusetts.

Kenneth E. and Kinnear J. (1994), "Advances in Genetic Programming," The MIT Press, London.

King M. (2003), "Basic principles of MRI," Retrieved 19th May 2003 from erads.com website: *http://www.erads.com*

Kishore J. K., Patnaik L. M., Mani V. and Agrawal V. K. (2000), "Application of genetic programming for multicategory pattern classification," *IEEE Transactions on Evolutionary Computation*, vol. 4, no. 3, pp. 242-258.

Kovalerchuk B., Vityaev E. and Ruiz J. (2000), "Consistent knowledge discovery in medical diagnosis," *IEEE Engineering in Medicine and Biology Magazine*, vol.19, pp. 26-37.

Koza J. (1992), *Genetic Programming: On the Programming of Computers by Natural Selection*, MIT Press, Cambridge.

Koza J. (1994), *Genetic Programming II: Automatic Discovery of Reusable Programs*, MIT Press, Cambridge.

Koza J., Bennett III F. H., Andre D., Keane M. A. and Dunlap F. (1997), "Automated synthesis of analog electrical circuits by means of genetic programming," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 2 pp. 109-128.

Krantz B. (1999), "A 'crisp' introduction to fuzzy logic," Retrieved 2nd June 2003 from University of Colorado at Boulder, Computer Science Educational Labs website: http://www-ugrad.cs.colorado.edu/~cs3202/papers/Brigette_Krantz.html

Krishnapuram R. and Keller J. M. (1993), "A Positivistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98-110.

Kukar M., Kononenko I., Groselj C., Kralj K. and Fettich J. (1999), "Analysing and improving the diagnosis of ischaemic heart disease with machine learning," *Artificial Intelligence in Medicine*, vol. 16, pp. 25-50.

Kurz L. and Benteftifa M. H. (1997), *Analysis of Variance in Statistical Image Processing*, Cambridge University Press.

Kusiak A. Kernstine K. H., Kern J. A., McLaughlin K. A. and Tseng T. L. (2000), "Data mining: medical and engineering case studies," *Proceedings of 2000 Industrial Engineering Research Conference*, Ohio, pp. 1-7.

Kwan R. K. S., Evans A. C., Pike G. B. (1999): "MRI simulation-based evaluation of image-processing and classification methods." *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1085-1097.

Langdon, W. B. and Poli, R. (1997), "Genetic programming in Europe," Technical report, the European Network of Excellence in Evolutionary Computing, Retrieved 2nd June 2003 from NEC Research Institute CiteSeer website: *http://citeseer.nj.nec.com/langdon97genetic.html*

Larsen R. (2000), "Mathematic morphology," Lecture notes, Retrieved 24th June 2003 from Technical University of Denmark, Department of Informatics and Mathematic Modelling website: *http://www.imm.dtu.dk/courses/02503/*

Lavrac N. (1999), "Selected techniques for data mining in medicine," *Artificial Intelligence in Medicine*, vol. 16, pp. 3-23.

Law T. Y. and Heng P. A. (2000), "Automated extraction of bronchus from 3D CT images of lung based on genetic algorithm and 3D region growing", *Proceedings of SPIE- Medical Imaging*. vol. 3979, pp. 906-916.

Lee D. and Seung H. (1997), "Unsupervised learning by convex and conic coding," In Mozer M. C., Jordan M. I. and Petsche T. (Eds), *Advances in Neural Information Processing Systems*, vol. 9, pp515. The MIT press.

Lee W. P., Hallam J and Lund H. H. (1997), "Applying genetic programming to evolve behaviour primitives and arbitrators for mobile robots," *Proceedings of IEEE 4th International Conference on Evolutionary Computation*, vol. 1, pp. 501-506. IEEE Press.

Leemput K. V., Maes F., Vandermeulen D. and Suetens P. (1999), "Automated model-based tissue classification of MR images of the brain", *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897-908.

Lemieux L., Hagemann G., Krakow K., and Woermann F. G. (1999), "Fast, accurate, and reproducible automatic segmentation of the brian in T_1-weighted vlume MRI data," *Magnetic Resonance in Medicine*, vol. 42, pp. 127-135.

Li C. L., Goldgof D. B. and Hall L. O. (1993), "Knowledge-based classification and tissue labelling of MR images of human brain", *IEEE Transactions on Medical Imaging,* vol.12, no.4, pp. 740-750.

Li J. and Gray R.M. (2000), *Image Segmentation and Compression Using Hidden Markov Models*, Kluwer Academic Publishers.

Liang Z. P. and Lauterbur. P. (2000), *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*, IEEE Press Series in Biomedical Enginnering.

Lin Y., Tian J. and He. H. G. (2002), "Image segmentation via fuzzy object extraction and edge detection and its medical application," *Journals of X-ray Science and Technology*, vol. 10, pp. 95-106.

Louis D. N., Posner J., Jacobs T., Kaplan R., Ho. P. T. C., Nichols C., Liang B. C., Packer R., Martuza R. L., Pardridge W. M., Mazziotta J., Shrieve D. C., Meyers C. A., Strausberg R. L., Miller R. H., Weiner S. L., Neuberg D. (2000). *Report of the brain tumour progress review group.* Retrieved 30th May 2003. *http://osp.nci.nih.gov/Prg_assess/PRG/BTPRG*

Mackiewich B. (1995), *Automatic segmentation of the brain in MRI*, MSc thesis, Simon Fraser University, Retrieved 20th May 2003. *http://www.cs.sfu.ca/~stella/grads.html*

Mair J, Smidt J, Lechleitner P, Dienstl F. and Puschendorf B. (1995), "A decision tree for the early diagnosis of acute myocardial infarction in non-traumatic chest pain patients at hospital admission," *Chest*, vol. 108, no. 6, pp. 1502-1509.

Mangasarian O. L. and Wolberg W. H. (1990), "Cancer diagnosis via linear programming,", *SIAM News*, vol. 23, no. 5, pp. 1-18.

Mangasarian O. L., Setiono R., and Wolberg W.H.(1990), "Pattern recognition via linear programming: Theory and application to medical diagnosis," In Coleman T. F. and Li Y. Y (Eds.), *Large-scale numerical optimization*, SIAM Publications, pp. 22-30.

Medica.de (2004), Website, Accessed 19th May 2004, *http:www11.medica.de/../ category./content.1*

McRobbie D. W., Moore E. A., Graves M. J. and Prince M. R. (2003), *MRI from picture to proton,* Cambridge Press.

Miettinen K., Makela M., Meittaanmaki P. and Periaux J. (1999), *Evolutionary Algorithms in Engineering and Computer Science.* John Wiley & Sons Ltd, England.

Mitchell T. (1999), "Machine Learning and Data Mining," *Communications of the ACM*, vol. 42, no.11, pp. 31-36.

MRI CHUQ (2003), A Teaching file, Retrieved 23rd May 2003 from Laval University, Department of Radiology website: *http://www.mri-chuq.com*

Musella Foundation (2003), *Clinical Trials and Noteworthy Treatments for Brain Tumors*, Retrieved 27th May 200.*http://virtualtrials.com/faq*

National Cancer Institute (2003), "What you need to know about brain tumors," Retrieved 27th May 2003. *http://www.cancer.gov/cancerinfo/wyntk/brain*

Nyúl, L.G. and Udupa J. K. (1999), "On standardizing the MR image intensity scales," *Magnetic Resonance Imaging*, vol. 42, pp. 1072-1081.

Obitko M. (1998), "Genetic algorithms," Retrieved 3rd June 2003 from The Czech Technical University in Prague, Department of Computer Science and Engineering website: *http://cs.felk.cvut.cz/~xobitko/ga/*

Pal N. R. and Pal S. K. (1993), "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, pp. 1277-1294.

Patola W. and Coulter B. (1997), "MRI artefacts," Retrieved 23rd May 2003 from University of Brithish Columbia, Department of Radiology website: *http://www1.stpaulshosp.bc.ca/stpaulsstuff/MRartefacts.html*

Pennstate Children's Hospital (2003), *Health information*, Retrieved 27th May 2003. *http://www.hmc.psu.edu/childrens/healthinfo/b/braintumors.htm*

Pham D. L. and Prince J. L. (1999), "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 737-752.

Pham D. L. (2001), "Spatial Models for Fuzzy Clustering," *Computer Vision and Image Understanding*, vol. 84, pp. 285-297.

Pohle R. and Toennies K. D. (2001), "Segmentation of medical images using adaptive region growing", *Proceedings of SPIE- Medical Imaging*, vol. 4322.

Poli R. (1996), "Genetic programming for image analysis," *Proceedings of 1st International Conference on Genetic Programming*, pp. 363-368.

Powers D. (2000), "An introduction to the EM algorithm," Retrieved 7th July 2003 from University of Texas at Austin, Liberal Arts Computer Lab website: *http://www.la.utexas.edu/course-materials/sociology/soc386L/EM1.pdf*

Pudi K. (2003), "Neural networks," Retrieved 2nd June 2003 from Indian Institute of Science website: *http://dsl.serc.iisc.ernet.in/~vikram/nn_intro.html*

Pulo K. (1999), "Fuzzy logic and machine learning- an investigation," Retrieved 2nd June 2003 from Kevin Pulo Personal website: *http://www.kev.pulo.com.au/ai/fuzzyml_report/*

Quinlan J. R. (1986), "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106.

Quinlan J. R. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann.

Ramirez J., Cook D., Peterson L., Peterson D. (2000), "Temporal pattern discovery in course-of-disease data," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, pp. 63-71.

Rich B. and Lasley E. (1998), *The Dana brain daybook: what's new in neuroscience*, Retrieved 28th May 2003 from Dana Foundation website: *http://www.dana.org/articles/dbk_0498.cfm*

Robb A. R. (1985), *Three-Dimensional Biomedical Imaging*, CRC Press.

Robb A. R. (2000), *Biomedical Imaging, Visualization, and Analysis*, Wiley-Liss, Inc.

Rokahr I., Engels S. A, Svanberg S. Ekseth K. and Backlund E. (1996), "Laser-Induced Autofluorescence in Malignant Brain Tumours and Normal Brain Studied During Stereotactic Biopsy Procedures," Retrieved 28th May 2003 from Lund University Medical Centre website: *http://www.mlc.lu.se/Prog9395/p30.htm*

Sabbatini R. (1997), "The PET scan: a new window into the brain," Retrieved 27th May 2003 from e*Pub website: http://www.epub.org.br/cm/n01/pet/pet.htm

Savoy R. and Jovicich J. (2001), "MRI glossary," Retrieved 23rd May 2003 from Massachusetts Institute of Technology website: http://web.mit.edu/hst.583/www/course2001/LECTURES

Siegel F. and Morgan C. J. (1996), *Statistics and Data Analysis- An Introduction*, 2nd ed., John Wiley & Sons, Inc.

Siemens AG (2001), "MR glossary," Retrieved 23rd May 2003 from Siemens Medical Solutions website: http://www.siemensmedical.com/siemens/en_INT/gg_mr_FBAs/files/brochures/MR_Glossar_en.pdf

Siromoney A., Raghuram L., Siromoney A., Korah I. and Prasad G.N.S., "Inductive Logic Programming for Knowledge Discovery from MRI Data," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, pp. 72-77, 2000

Shaw M., Subramaniam C., Tan GW. and Welge M. (2001), "Knowledge management and data mining for marketing," *Decision Support System*, vol. 31, pp. 127-137.

Sled J. G, Zijdenbos B. and Evans A.C. (1998), "A nonparametric method for automatic correction of intensity nonuniformity in MRI Data", *IEEE Transactions on Medical Imaging*, vol. 17, pp. 89-97.

Smith J. (2002), "Genetic algorithm I," Retrieved 3rd June 2003 from University of West England, Faculty of Computing, Engineering and Mathematical Sciences website: http://www.cems.uwe.ac.uk/~jsmith/ec

Soille P. (1999), *Morphological Image Analysis- Principles and Applications*, Springer.

Spector L., Langdon W. B., O'Reilly U. and Crosa P. (1999), *Advances in Genetic Programming*, vol. 3, MIT Press.

Srdanovic M. (1998), *Jumping Jack Icon/Glyph - Scientific Visualization Project*, Retrieved 28th May 2003 from University of Massachusetts Lowell, Department of Computer Science website: http://www.cs.uml.edu/~msrdanov/viz/

Siromoney A., Raghuram L., Siromoney A., Korah I. and Prasad G. N. S. (2000), "Inductive logical programming for knowledge discovery from MRI data," *IEEE Engineering in Medicine and Biology Magazine*, pp. 72-77.

Slevin M. and Ryan P. (1988), "Understanding brain tumours," *British Association of Cancer United Patients.*

Stark D. and Bradley W. G. (1999), *Magnetic Resonance Imaging*, Mosby: St Louis.

Stergiou C. and Siganos D. (1996), "Neural networks," Retrieved 2nd June 2003 from Imperial College London, Department of Computing website: http://www.doc.ic.ac.uk/~nd/surprise_96

Suzuki H. and Toriwaki J. (1991) "Automatic segmentation of head MRI images by knowledge guided thresholding," *Computerized Medical Imaging and Graphics*, vol.15, no.4, pp. 233-240.

Taguchi G, Chowdury S. and Wu Y. (2000), *The Mahalanobis-Taguchi System*, McGraw-Hill Professional.

The Brain Tumor foundation (2003), "Brain tumor information," Retrieved 27th May 2003. http://www.braintumorfoundation.org/tumors/index.htm

The Brain Tumor Society (2002). Website. Accessed 14th May 2003. *http://www.tbts.org/biabt.htm*

The Central Brain Tumor Registry of the United States (2000). Website. Accessed 14th May 2003. *http://www.cbtrus.org*

The Hyman-Newman Institute for Neurology and Neurosurgery (2002), "Pediatric brain tumors," Accessed 27th May 2003. *http://nyneurosurgery.org/*

The Wallace-Kettering Neuroscience Institute (2003), Website. Accessed 14th May 2003. *http://www.wkni.org/brains.htm*

ThirdAge (2003). Website. Accessed 27th May 2003. *www.thirdage.com/health/adam/imagepage/8893.htm*

Tolias Y. A. and Panas S. M. (1998a), "On applying spatial constraints in fuzzy image clustering using a fuzzy rule-based system," *IEEE Signal Processing Letters*, vol. 5, no. 10, pp. 245-247.

Tolias Y. A. and Panas S. M. (1998b), "Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions," *IEEE Transactions on System, Man and Cybernet*, Part A, vol. 28, no. 3, pp. 359-369.

Two Crows Corporation (1999), Introduction to Data Mining and Knowledge Discovery, Third Edition, Retrieved 3rd June 2003. *www.twocrows.com*

Walker M. (2001), "Introduction to genetic programming," Lecture notes, Retrieved 3rd June 2003 from Massey University website: *http://www.massey.ac.nz/~mgwalker/gp/*

Wei J., Naguib R. N. G., Macall J., Petrovic D., Gaura E. and Ghoneim M. (2003), "Prognostic prediction of bilharziasis-related bladder cancer by neuro-fuzzy classifier," *Proceedings of 4th Annual IEEE Conference on Information Technology Applications in Biomedicine, UK*, pp. 181-183.

Weiss S. I. and Kulikowski C. (1991), *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems*, Morgan Kaufmann.

Wells, III W. M., Grimson W. E. L., Kikinis R. and Jolesz F. A. (1996), "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429-442.

Westbrook C. and Kaut C. (1994), *Handbook of MRI Technique*, Blackwell Science Ltd.

Westbrook C. (2002), *MRI at a Glance*, Blackwell Publishing Ltd.

Westbrook C. and Kaut C. (2002), *MRI in Practice*, Second edition, Blackwell Publishing Ltd.

Wilson Memorial Hospital (2003), Website, Accessed 16th April 2004. http://www.wilsonhospital.com.

Wolberg W. H. and Mangasarian O. L. (1990), "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 87, pp. 9193-9196.

Your Neurosurgical Associates (2003), "Stereotactic brain biopsy," Retrieved 28th May 2003. *http://www.yna.org/SBBWebL.pdf.*

Zaden L. A. (1965), "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353.

Zhang Y. J. (1997), "Evaluation and comparison of different segmentation algorithms," *Pattern Recognition Letters*, vol. 18, pp. 963-974.

Zhang Y. Y., Brady M. and Smith S. (2001), "Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol.20, no.1.

APPENDIX I: List of figures

APPENDIX II: List of tables

APPENDIX III: COMPUTING IMPLEMENTATION

Hardware: Pentium IV 2.0G

Platform: Windows XP

Implementation software: MATLAB 6.1

Approximate number of lines of coding: 5000

Approximate running time:

- Image preprocessing; 10 seconds
- Image segmentation: 3 minutes/per slice
- Brain tumour classification: 20 minutes (including training and validation)

An Example of codes:

%1 Read and open original images

```
Filename=input('Please input the filename of image:','s');
fid=fopen(Filename,'r');   % Open the image file.
[imagedata,count]=fread(fid,inf,'uint16');   % Read the original unsigned 16 bit data.
Slice=input('Which slice do you need?');   % Choose the slice.
slicedata=imagedata(1+(Slice-1)*256*256:256*256*Slice,1);  %Store one slice of image
                                                                  data.
```

%2 Order swap

```
%The scanning system which obtained the data uses a different order of the elementary bytes
%in binary image data to that of windows, the byte order is needed to be swapped.
```

```
low=mod(slicedata,256);
high=floor(slicedata/256);
swapdata=low*256+high; %swapdata is unsigned 16 bit.
clear low high;
```

%3 Standardize the intensities of the images.

%As a great number of pixels in an image is the background, these background pixels occupy a big portion of the histogram. Their existence may dominate standardization and affect the mean and the standard deviation. Therefore they should be excluded initially (Fig.5.2 pp.63).

```
i=1; thrpercent=0;
number1=0;
while isempty(number1)==0 | thrpercent<=0.1,
    thrpercent=0.01+(i-1)*0.01;
    threshold1=thrpercent*max(swapdata);
    [number2,b]=find(swapdata>=threshold1);
    Newhist=hist(swapdata(number2),max(swapdata(number2)));
    [number1,b]=find(Newhist(1:10)>=100);
    i=i+1;
end;
clear number1 b;

%thrpercent: Threshold percentage
%threshold1: Threshold for background pixels and foreground pixels. A pixel with intensity
%higher than the threshold is a foreground pixel.
%number1: the number of intensity levels within the lowest ten intensity levels in the new
%histogram which include over 100 pixels.
%number2: the number of foreground pixels.

BW=double(im2bw(swapdata/threshold1,1)); %Obtain the binary mask which only contains
                                          %foreground pixels.
StandardData=10000*((BW.*swapdata)-mean(swapdata(number2)))/std(swapdata(number2)
)+32500; %Standardize the image to a new image with a standard deviation Of 10000 and a
```

```
mean 32500.

[a,b,s]=find(StandardData<0); %Because of standardizing, the intensities of background
%pixels became negative.

%The foreground pixels keep the original intensities and the background pixels are set to be
%zero.

S=full(sparse(a,b,s,256*256,1));
StandardData=StandardData.*(1-S);
Standardimage=(reshape(StandardData,256,256))';
figure
imshow(uint16(Standardimage)) %Display standardized image
```

APPENDIX IV: PUBLICATIONS

S. Shen, W. A. Sandham, and M. H. Granat, "Preprocessing and segmentation of magnetic resonance imaging", Proc. 4th IEEE Annual International Conf Information Technology Applications in Biomedicine, ITAB.03, UK, pp149-152, 2003.

S. Shen, W. A. Sandham, and M. H. Granat, "Discovering multi-class classification rules using evolutionary computing and fuzzy logic", Proc. PREP2003, UK, pp206-207, 2003.

S. Shen, W. A. Sandham, M. H. Grant, J. Patterson and M. F. Dempsey, "A new approach to brain tumour diagnosis using fuzzy logical based genetic programming", Proc. IEEE EMBS 25th Annual International Conf. pp. 870-873, 2003.

S. Shen, W. A. Sandham, M. H. Grant, J. Patterson and M. F. Dempsey, "Fuzzy clustering based applications to medical image processing", Proc. IEEE EMBS 25th Annual International Conf, pp.747-750 2003.

S. Shen, W. A. Sandham, M. H. Grant, J. Patterson and M. F. Dempsey, "Bias field estimation for magnetic resonance images using an improved fuzzy clustering algorithm", Proc. IEEE EMBSS UK & ROI Postgraduate Conf. in Biomedical Engineering and Medical Physics, 2003.

S. Shen, W. A. Sandham, M. H. Granat, J. Patterson and M. F. Dempsey, "Application of data mining and knowledge discovery to brain tumour diagnosis", Proc. IEEE EMBSS UK & ROI Postgraduate Conf. in Biomedical Engineering and Medical Physics, pp25, 2002.