THE UNIVERSITY OF CALGARY

QUANTITATIVE EVALUATION METHODS FOR MRI POST-PROCESSING ALGORITHMS

by

Jane Xiaojing Zeng

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

JULY, 1996

© Jane Xiaojing Zeng 1996

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "QUANTITATIVE EVALUATION METHODS FOR MRI POST-PROCESSING ALGORITHMS", submitted by Jane Xiaojing Zeng in partial fulfillment of the requirements for the degree of Master of Science.

Dr. M. R. Smith, Supervisor

Dept. of Electrical & Computer Engineering

Dr. S. T. Nichols

Dept. of Electrical & Computer Engineering

Dr. L. Hahn

Dept. of Radiology

Date: 27 September 96

Abstract

In this thesis, we propose several quantitative evaluation methods for magnetic resonance imaging (MRI) post-processing algorithms. The methods include a set of metric measurements that compare a corrected data set or image with a standard data set or image; and a newly designed computer observer receiver operator characteristics (ROC) analysis. We test our evaluation methods with two types of applications. The first is amplitude modulation motion artifact suppression in MRI. The second application is constrained modeling used in MRI image reconstruction, which is previously d eveloped in our lab. The test results suggest that our computer ROC analysis has great potential in the area of quantitative evaluation of imaging algorithms and applications.

Acknowledgement

I'd like to thank my supervisor Dr. M. R. Smith for his help, his patience and his sense of humor during this thesis project. Dr. A. Crawley and Ms. C. Wallace from the magnetic resonance imaging center at the Foothill's Hospital offered a lot of help as well. My sincere gratitude to my parents and all those who are supportive in my life.

Table of Contents

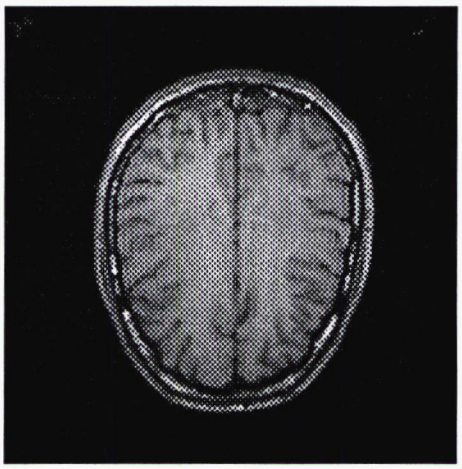
List of Tables

List of Figures

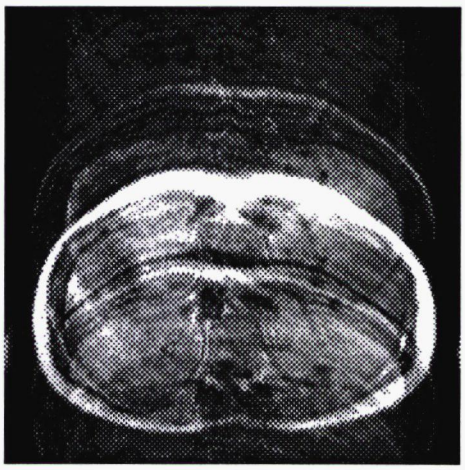
Chapter 1

Introduction

Magnetic resonance imaging (MRI) is a powerful imaging tool for medical applications. One of the shortcomings of MRI is distortion in the final images introduced during data acquisition. These distortions degrade the quality of images and make diagnosis more difficult. The origin of the distortions varies. Two examples are the ringing and blurring caused by data truncation, and the ghosting caused by motion during data acquisition. These distortions can be seen in Figure 1.1.



A brain image with ringing and blurring



An abdomen image with ghosting

Figure 1.1: Sample images with distortions

Over the past years, many computer post-processing algorithms have been developed to reduce MR image distortions. In our laboratory, we investigated the use of autoregressive moving average (ARMA) modeling ([SNHW86] and phase correc-

tion ([MSNC93]) in reducing truncation artifacts. Recently we developed various signal processing algorithms to remove amplitude modulation (AM) motion artifacts ([ZSC94]). One problem we encountered, as shared by many researchers ([BMW86], [FBSM87], [Han88], [Han90], [HY91], [Gra]), is that it is difficult to compare the performance of different post-processing algorithms. There are frequently no suitable standard images and procedures for quantitative performance evaluation.

In this thesis, we attempted to establish standard images and procedures that can be used to evaluate performance of different post-processing algorithms in image processing. Our emphasis has been on evaluating MR image processing techniques. However, the analysis procedures are general enough that they can be applied to other image processing applications. Our test phantoms were generated mathematically in the spatial frequency domain and reconstructed to produce MR images. The reason that it is invalid to generate phantoms in the image domain directly will be explained in this thesis.

There are two broad types of evaluation procedures discussed in this thesis. The first type of procedure uses difference measures between the corrected images and a standard image. The second type attempts to analyze clinical relevancy of the images using a computer observer.

We then evaluated the appropriateness of these procedures for MR image processing in three areas of application – discrete Fourier transform (DFT) reconstruction of MR images, amplitude modulation (AM) motion artifact removal and image reconstruction using modeling. Both mathematical phantom and modified medical data were used for testing.

1.1 Amplitude Modulation Motion Correction in MRI

Compared to other imaging techniques, MRI is superior in its flexibility of imaging plane, resolution and contrast ([WSK88]). The main disadvantages of MRI are its high implementation and operation costs and long imaging time. Long imaging time decreases the throughput of the equipment, and introduces various motion artifacts to the image. Compared to computer tomography (CT), which may be accomplished in milliseconds, a typical MRI spin-echo scan takes 2 to 10 minutes. During the scan, even if the patient stays completely still, periodic motions (cardiac, respiratory) and fluid flow in the blood vessels still exist (see Figure 1.1 right hand side image). These movements produce artifacts in the image that degrade its quality. Reducing the effect of motion remains an important topic in MRI research ([HYR91b], [HY92b], [MPS+93]).

Most of the existing motion correction techniques, such as cardiac gating and gradient moment nulling, require instrumentation monitoring or new RF pulse sequences ([AW90]). Our research interest lies in developing post-processing techniques that utilize only the information contained in the motion degraded image. The main advantage of such post-processing techniques is that they do not require the data acquisition procedures to be altered.

In conventional Fourier MR imaging, the artifacts from periodic motion usually appear as ghosting, i.e., a series of blurred versions of the object are superimposed on the object ([Woo86]). A simplified model of ghosting considers motion degraded images as amplitude modulated original images ([HYR91b], [MPS+93]). Based on this model, we developed an initial series of simple post-processing motion suppression

algorithms.

After obtaining encouraging results during this initial research phase, we tried to evaluate our algorithms relative to those developed by other researchers. It was at this point that we found that there were no quantitative image evaluation methods or standards. Such methods are crucial for evaluating research results.

Meanwhile, the absence of an accurate motion model, and the irrelevancy of developing our own physical motion model, impaired further investigation in the motion suppression area. We therefore decided to devote the remainder of the thesis project to developing quantitative evaluation methods for MR image processing. The various AM motion suppression algorithms we had developed provided an excellent application with which to test our evaluation methods.

1.2 Quantitative Evaluation Methods

Quantitative evaluation methods are crucial for a convincing comparison of algorithms. We did a literature survey of such methods in MRI research and did not find many of them. Most of the performance evaluations of MRI algorithms presented in the literature were subjective visual evaluations. Often corrected images using proposed algorithms were displayed beside the original image. A conclusion was drawn that the algorithm was successful if the corrected images exhibited fewer artifacts and less distortion than the original distorted image.

This subjective approach is sufficient when obvious differences in the images can be perceived. However, when comparing different algorithms, the subtle differences in image quality may be hard to detect visually, especially in reproduced images. The

lack of evaluation standards motivated us to develop methods that can quantitatively assess MR image quality. Such a method should yield a figure of merit (FOM) that could be easily compared across algorithms. Furthermore, it should suggest which algorithm makes clinical diagnosis easiest.

An extended literature search showed that in other fields of image processing, researchers have proposed some quantitative methods to judge the quality of images ([FBSM87], [Han88], [Han90], [HO91], [HY91], [LLG92], [MHN+94], [FHN+94]). Based on this literature search, we attempted to develop two categories of quantitative evaluation methods suitable for MRI.

1. The first category is a set of quantitative image metrics that measure the difference between a corrected image and a standard image to give an FOM. Difference measures can be taken using either image or frequency domain data.

2. The second category is a series of procedures performing clinical image analysis using a computer observer.

These methods were tested using various DFT reconstruction algorithms, the AM motion suppression algorithms developed early in this thesis project, and alternative MRI reconstruction algorithms previously developed in our lab.

1.3 Project Scope

This thesis project consists of developing and testing quantitative image evaluation methods in MRI. The evaluation methods include:

- A set of mathematical measurements indicating the differences between corrected and standard data in frequency and image domain.

- A computer observer approximating the visual process that human observers employ to determine the presence or absence of a feature in an image.

The methods are tested using three MRI applications:

- DFT reconstruction: Discrete Fourier transform is applied to a series of increasingly truncated MR data sets to obtain MR images of the same size. The effect of truncation in the images are evaluated using our methods.

- AM motion artifact suppression: A series of new correction algorithms based on the AM motion model are developed during this thesis project. The motion suppression algorithms are applied on both real MR image and mathematical phantoms with simulated motion and evaluated with our proposed methods.

- ARMA modeling in reconstruction: Image reconstruction is conventionally achieved using a 2D Fourier transformation. With truncated MR data sets, The Fourier transform introduces ringing and blurring in reconstructed images. The use of ARMA modeling ([LBC+92]) in image reconstruction has visually improved the quality of MRI images. We quantitatively evaluate the existing reconstruction algorithms using our criteria.

1.4 Thesis Organization

The thesis is organized into three main parts.

In the first part, Chapters 2 and 3, we present the background of MRI motion suppression and the AM motion suppression algorithms developed during this thesis project.

The next three chapters, Chapters 4, 5 and 6, are devoted to introducing our quantitative evaluation methods. Chapter 4 is a summary of the literature survey we did on existing image evaluation methods. The metric measurements we developed for image evaluation are described in Chapter 5. In Chapter 6, we present a computer receiver operator characteristics (ROC) analysis developed in this thesis project.

The last part of the thesis include results using our proposed quantitative evaluation methods (Chapter 7). We wrap up the thesis with a brief conclusion and suggestions for future work in Chapter 8.

Chapter 2

Motion in MRI

In this chapter, we first present the principles of magnetic resonance imaging (MRI) in a simplified fashion. Some background knowledge of motion in MRI is then introduced. We also include a summary of existing techniques for motion artifact suppression in MRI.

2.1 Fundamental Principles of MRI

MRI is based on nuclear magnetic resonance (NMR). In this section we present the physics of MRI in a greatly simplified fashion, following the articles in [WSK88] and [Reh91].

2.1.1 The nuclear magnetic resonance phenomena

According to Jagannathan ([Reh91]), the protons in the nuclei of atoms possess a magnetic field, the nuclear magnetic dipole. For any individual nucleus, the direction of this dipole is naturally random. In the presence of an external magnetic field B_0, the nuclei tend to precess about the direction of B_0 and the net magnetization produced by the dipoles lies in the direction of B_0. We can represent the state of the ensemble of nuclear dipoles by a net magnetization vector M_0 in the direction of the external magnetic field (Figure 2.1 (a)). This "macro" representation using M_0 makes it easier to visualize NMR signals and will be employed to introduce MRI

principles. A quantum mechanical interpretation can be found in "NMR Imaging in Biomedicine" ([MM82]) and will not be discussed here.

When an electro-magnetic radiation (B_1 in Figure 2.1 (a)) with an appropriate energy (frequency) is applied to the nuclei under the influence of an external magnetic field, the net magnetization M_0 will change direction (Figure 2.1 (a)). The frequency required to excite a nucleus from its ground state is called the resonance frequency, or Larmor frequency, of the nucleus. The resonance frequency can be expressed by the Larmor equation:

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

where ω_0 is the Larmor frequency and B_0 is the external magnetic field strength. γ is a constant called the gyro-magnetic ratio, which is different for different nuclei. The Larmor frequency ω_0 is usually in the MHz or radio frequency (RF) range. The value depends on the external magnetic field strength, the nuclei and their chemical environment.

2.1.2 The NMR signal

When an external magnetic field is present, the initial net magnetization M_0 is parallel to the external magnetic field B_0, along the z-axis in Figure 2.1 (a). An RF magnetic pulse (a strong radio frequency field) produced by an RF coil on the x-axis can tip M_0 away from the z-axis. The duration and power of the RF pulse determines the direction of M_0 after the pulse. If a so called 90^o pulse is applied, the net magnetization M_{xy} lies in xy-plane after excitation (Figure 2.1 (b)). The longitudinal or z-magnetization M_0 is thus transformed into a transversal or xy-magnetization M_{xy}.

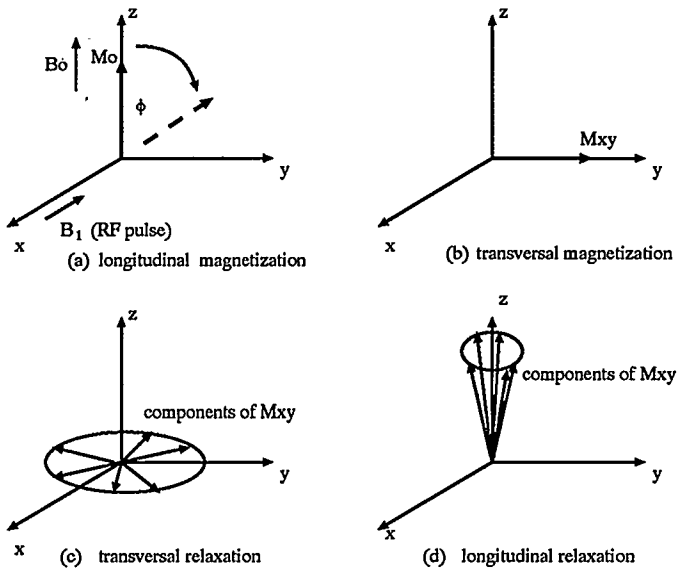


Figure 2.1: Illustration of net magnetization and relaxation

When the RF field is removed, the dipoles tend to return to their initial states (relaxation), causing the net magnetization to go back to M_0. Two simultaneous processes exist during relaxation: transversal relaxation and longitudinal relaxation. Transversal relaxation is characterized by the decay of transversal magnetization M_{xy}. Since the resonance frequencies of nuclei vary due to their local environment, the dipoles precess at different frequencies. As a result, the xy-magnetization M_{xy} initially aligned by the RF pulse splits into magnetizations in different directions (Figure 2.1 (c)), and the resultant M_{xy} diminishes. Meanwhile, longitudinal magnetization increases from 0 to M_0 (Figure 2.1 (d)), because of longitudinal relaxation.

The magnetic dipoles rotating in the xy plane can be made to produce a voltage signal in the receiver coil. Due to transversal relaxation, the magnitude of the voltage signal decays with time, as shown in Figure 2.2. The complex signal in Figure 2.2 is called a free induction decay (FID) signal. The Fourier transform of the time domain

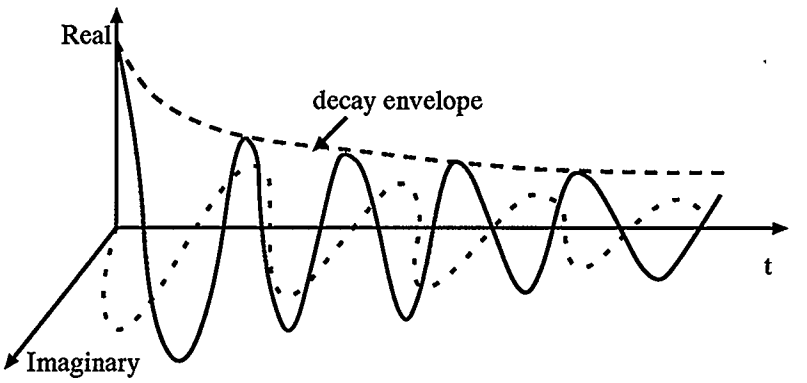


Figure 2.2: The free induction decay signal that can be collected during the relaxation of exited nuclei

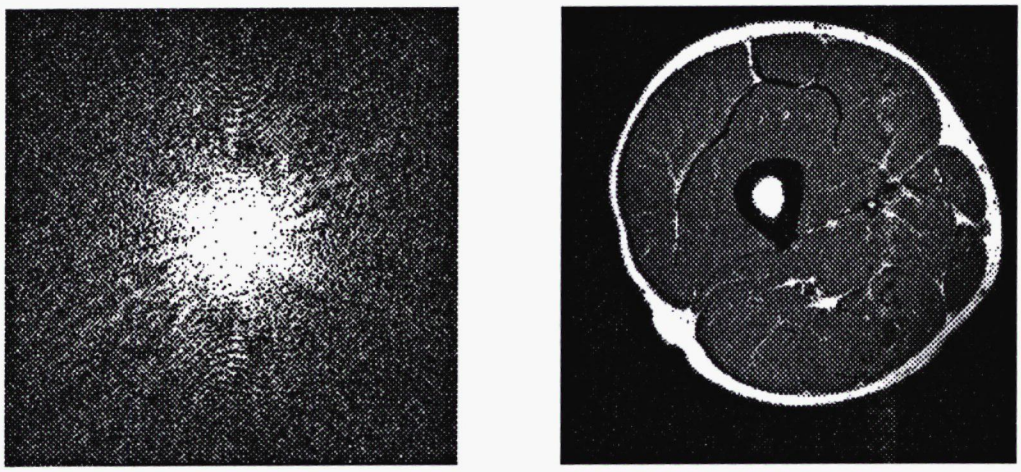
FID yields the NMR signal. An MRI data set is a collection of different FID signals.

2.1.3 The MRI image

To acquire an MRI image, we have to encode spatial information into the NMR signals. Spatial encoding is achieved by using three orthogonal magnetic gradients: x-, y- and z-gradient, denoted by G_x, G_y and G_z. If a gradient magnetic field exists, the Larmor frequency (2.1) will depend on the position of the nuclei along the gradient field. Normally, the z-gradient and the frequency of the RF pulse specifies which image plane to excite. The x-gradient provides frequency encoding of spatial locations along the x-axis of the image. The y-gradient provides phase encoding of spatial locations along the y-axis of the image. The combination of x- and y-gradients allow accurate mapping of the contribution to the NMR signal from each location in the excited slice.

The data acquired in MRI forms an array of time-domain FID signals. The time domain is often called k-space or the frequency domain as the FID signals represent frequency information of the MR signals. A 2D inverse Fourier transform is required

to reconstruct the image from the data acquired. One set of MR data in both frequency and image domain is show in Figure 2.3.



An MR data set in the frequency domain

The data set in the image domain

Figure 2.3: A sample MRI data set of a thigh in both frequency and image domain

In the next section, we'll use the principles of MRI to discuss the production of motion artifacts in MR images.

2.2 AM Motion Artifacts in MRI

To visualize the origin of motion artifacts in MR images, we adopt the following formalism proposed by Xiang *et al.* ([XH93]).

For a 2D MR image, the data set that is collected during data acquisition is a 2D array containing spatial frequency components of the object being imaged. We call the 2D spatial frequency domain *k-space*. The two spatial frequency axes are denoted by k_x and k_y, also called frequency and phase encoding axis, respectively.

Since motion in MRI causes the MR images to change with respect to time, to understand motion, we have to take into consideration the time elapsed during data acquisition. At each instance in time, a 2D spatial frequency component matrix (MR data set) is needed to produce an MR image showing the instantaneous state of the image subject. To completely describe the image subject during data acquisition, we need one 2D MR data set at each time instance. Therefore, the MR data needed to produce a complete representation of an image subject during data acquisition would be a 3D block of spatial frequency components, as illustrated in Figure 2.4. The horizontal axes are k_x and k_y and the vertical axis is time (t).

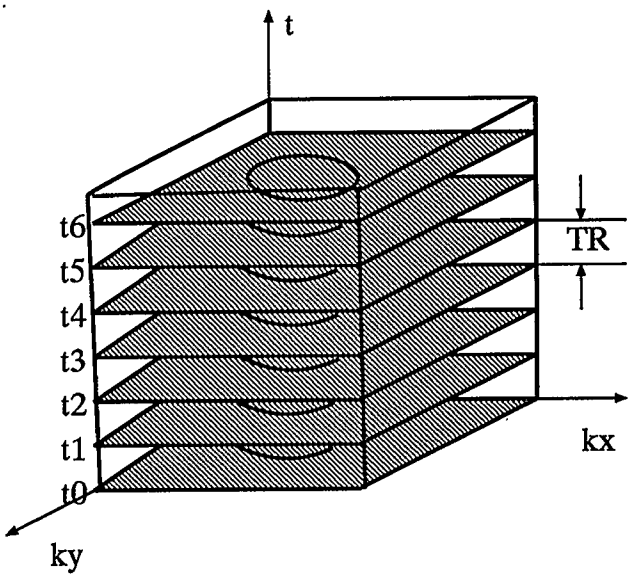


Figure 2.4: An imaginary 3D MR data block required to produce instantaneous MR images at any given time during data acquisition

At any particular time, a data sheet perpendicular to the time axis will represent the 2D spatial frequency components of the image subject at that time. Ideally, the 3D data block should have infinitely many such data sheets in order to generate

images of the subject at all time instances. Such a 3D MR data block is impossible to collect in real life.

Figure 2.4 showed a few discrete 2D data sheets at time intervals t0 to t6, each being taken time TR apart. If we ignore the time required to collect the MR data, we can acquire a series of such discrete 2D data sheets to approximate the 3D data block.

In real life, the time required for data collection normally could not be ignored. During conventional MR data acquisition, at each frequency encoding excitation, one data line along the k_x direction will be collected. The time required to do that is relatively short and can be neglected. Therefore one horizontal line of data is collected at each time instance. However, the phase-encoding (k_y direction) repetition time (TR) is relatively long and the time required to collect this data can't be neglected. As a result, the horizontal k_x data lines move up the time axis with each phase encoding. When the scan is finished, we have collected an oblique data sheet in the 3D data block. One such oblique data sheet is illustrated in Figure 2.5.

If no motion exists during the scan, all the horizontal data sheets in Figure 2.4 are identical. The oblique data sheet, which consists of lines of data from different horizontal sheets, is then identical to all the horizontal data sheets as well. If motion exists, the horizontal data sheets are different at different times and the oblique data sheet is different from any one of the horizontal sheets. When such an oblique data sheet is reconstructed to produce an MR image, motion artifacts will be introduced.

To illustrate the introduction of motion artifacts, we look at two paths to reconstruct an MRI image with motion artifacts (Figure 2.6). The path we use in real life to reconstruct an image is to take a 2D inverse discrete Fourier transform (DFT) of

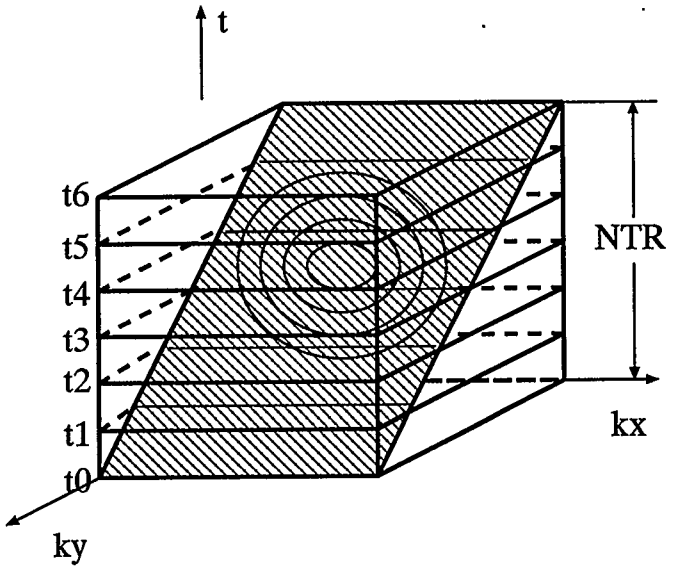


Figure 2.5: A 2D oblique MR data sheet actually acquired in conventional MRI

the oblique data sheet ((a) to (c) in Figure 2.6). The second path of reconstruction is imaginary, used here to explain how motion artifacts emerge. The theoretical background of the second reconstruction path is Fourier's projection-slice theorem, which states that the Fourier transform of a 2D plane passing through the origin is equivalent to the projection of the 3D-Fourier transformed space with the projection direction normal to the 2D plane. This equivalence provides an alternative path for the conventional MRI image reconstruction ((a) to (b) to (d) in Figure 2.7). If a 3D data block exists we can first 3D Fourier transform the block and then take a projection of the transformed block at the right angle to obtain the resulting image.

Figure 2.7 shows an imaginary 3D Fourier transformed data block (as in Figure 2.6 (b)) with motion. The axes are spatial position x and y and temporal frequency ω. The center plane is the DC plane which represents a time-averaged magnetization (image) during data acquisition. The AC planes above and below the DC plane

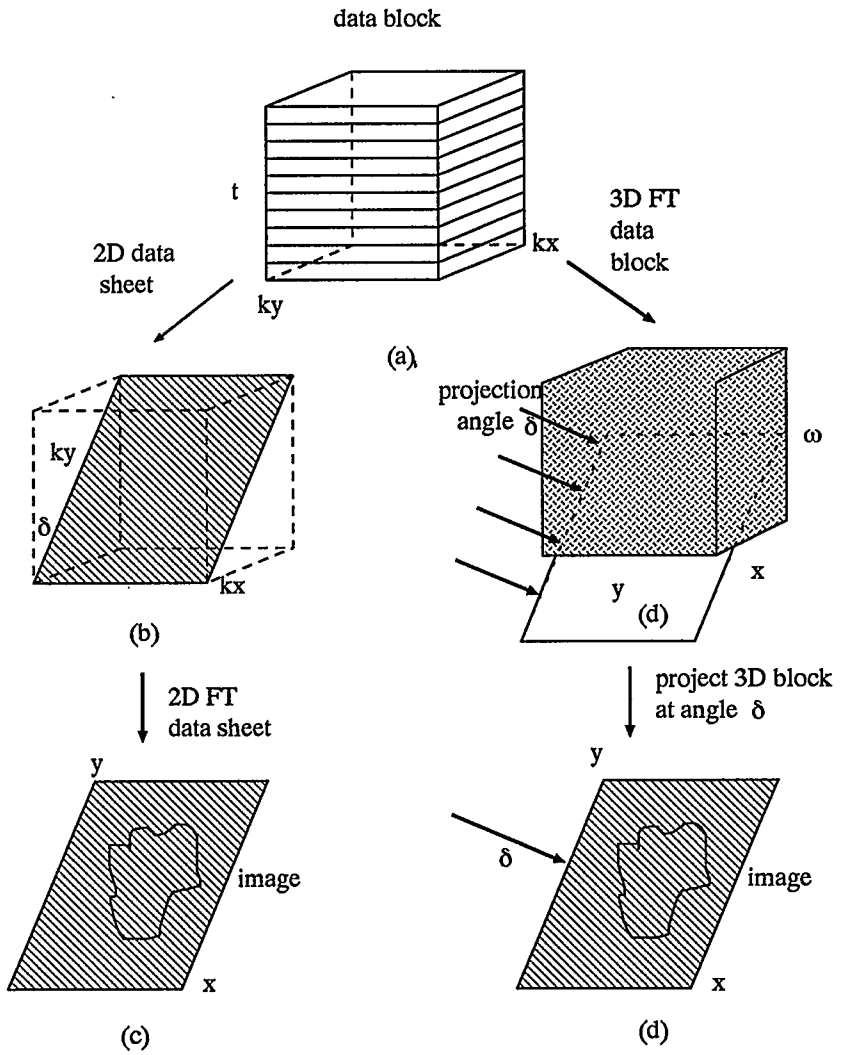


Figure 2.6: Illustration of two MR image reconstruction paths using the oblique MR data sheet

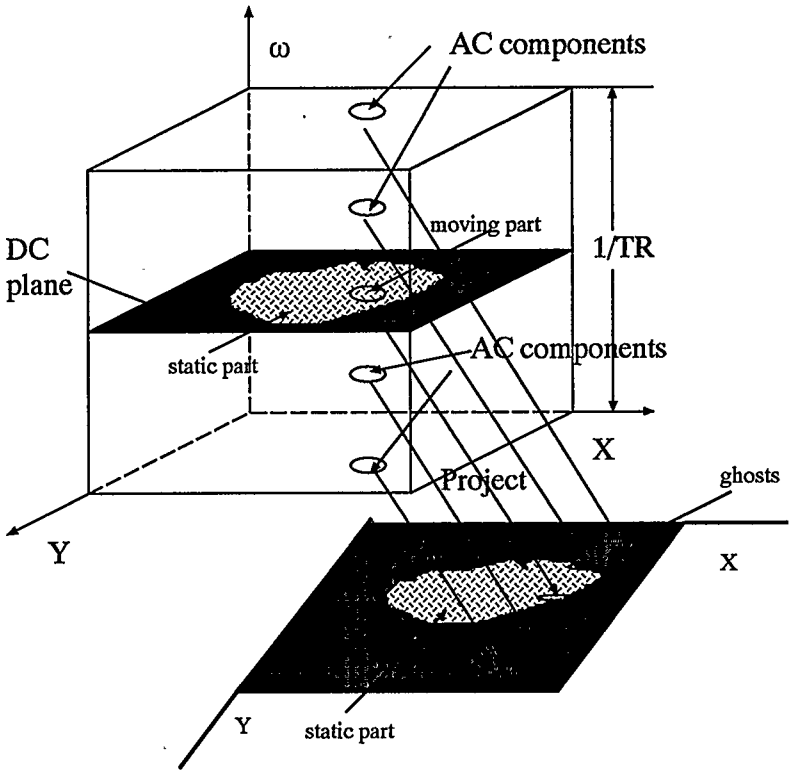


Figure 2.7: The 3D Fourier transformed data block, showing DC and AC components of the image and how they are projected to form the final image

contain information on the temporal variations during data acquisition. If no motion occurs during data acquisition, there will be no signal components outside the DC plane. For pseudo-periodic movements such as cardiac and respiratory motion of the patients, the AC components are located at discrete horizontal levels. The position of the AC components depends on the frequency and harmonics of the motion. When such a 3D space is projected to obtain the 2D image, the discrete AC components will result in ghosting artifacts in the y direction in the reconstructed image.

Figure 2.7 shows the ghosting artifacts of the moving part in the final image. In one simplified model of motion, the ghosts can be interpreted as amplitude modulated side bands of the moving object. In this thesis, we developed our motion suppression algorithms using this AM motion model and obtained good initial results.

The AM model itself is a grossly over-simplified motion model that does not handle the more complicated phase distortions introduced in actual sample motion. We used the AM model to correct ghosting in some real motion images and did not obtain encouraging results. Nevertheless, the motion suppression algorithms provide an excellent test ground for our quantitative evaluation methods developed later in this thesis.

2.3 Existing Motion Suppression Techniques

Motion suppression has been an important research topic in MRI. Various methods have been developed to suppress motion artifacts. The following is a brief list of existing motion suppression algorithms surveyed by Xiang and Henkelman ([XH93]), Runge and Wood ([RW88]), Arvanitis and Watson ([AW90]). Please refer to these

articles for more detailed information.

1. Techniques which restrain the motion of the patients

   Physical restraint and *breath holding* are both effective methods to reduce the amount of artifacts. They are easy to implement. However, they are not always possible with all data acquisition and are not comfortable for the patients.

2. Techniques which modify the pulse sequence applied to acquire the k-space data

   Averaging was one of the earliest and easiest approaches. This technique repeats each phase encoding step more than once and averages the data sets. Averaging reduces the ghosts by phase dispersion ([XH93]).

   If the total time of the repeated acquisition of each phase encoding step equals the period of the motion, the averaged data will have a null ghost ([XH93]). Such a technique is called *pseudo-gating.*

   Gradient moment nulling uses specially designed gradients between the RF pulses to eliminate motion that occurs between these two pulses.

   Spatial pre-saturation pulses can be used to saturate the tissue on each side of the imaging slice before phase encoding. Pre-saturation reduces the ambient signals that come into the imaging slice during the motion and thus reduces the artifacts.

   Researchers are also developing some *ultra-fast pulse sequences* that record the spatial frequency components in k-space in a different fashion. With these sequences typical scan time can be reduced, resulting in images that contain

less motion artifacts. However, these pulse sequences give rise to tissue contrast that radiologists are not familiar with. In addition most of these techniques are still under research. Implementing these new pulse sequences on commercial machines will require considerable expertise and training.

3. Techniques which modify the data acquisition rate

   Gating monitors the pattern of the pseudo-periodic motion and acquires data only during a certain part of the motion cycle. Ghosts can be reduced since, for each phase encoding step, the object being imaged is roughly in the same position.

   Ordered phase encoding method also monitors the motion and re-orders the phase encoding steps so that the apparent motion becomes much slower.

4. Post-processing techniques

   These techniques utilize the information contained in the motion degraded image and apply mathematical manipulations to the data to achieve specific goals. The main advantage of post-processing techniques is that they do not require additional time, instrumentation or new RF pulse sequences. We found two types of post-processing techniques in the literature:

   - Mathematically modeling the motion ([MPS+93], [ZFW+93], [AO94]). Using the model, image artifacts are suppressed with standard signal processing methods.
   - Using *a priori* information such as region of support, realness, etc. and applying a least square method to find an optimal solution that supports

the *a priori* information ([HYR91b], [RTS92], [HY92b]).

In this thesis project, we are interested in post-processing techniques utilizing a mathematical model.

2.4 Summary

In this chapter, we first introduced the principles of MRI. Motion in MRI was then explained and illustrated using Xiang's model ([XH93]). A survey of existing MR motion suppression algorithms was given.

In the next chapter, we present several AM motion suppression algorithms and their implementation.

Chapter 3

AM Motion Suppression Algorithms

In this chapter, we present a series of motion suppression algorithms we have developed based on the amplitude modulation (AM) motion model proposed by Mitsa ([MPS+93]). The algorithms include the Mitsa algorithm ([MPS+93]), the ARMA algorithm ([ZSC94]), the Direct algorithm ([ZSC94]) and the Adaptive algorithm. We present both theoretical background and implementation details for each algorithm.

3.1 The Mitsa Algorithm

3.1.1 Background

Mitsa *et al.* ([MPS+93]) proposed a motion suppression algorithm based on an amplitude modulation motion model. The AM model was derived from movement of a point source in the slice selection direction (z-direction). Mathematically, the MR signal acquired from such a moving object can be described as:

$$S(K_x, K_y) = M(K_x, K_y)[1 + \sum_{n=-\infty}^{\infty} \frac{\Delta m_n}{m_0} \cdot cos(\frac{2\pi n N_b K_y}{N_y \Delta K} + \phi_n)] \qquad (3.1)$$

where $S(K_x, K_y)$ and $M(K_x, K_y)$ are the amplitude modulated and raw data in the Fourier domain, respectively. K_x and K_y are k-space indices from -128 to 127. Other variables are related to the characteristics of the motion and MRI data acquisition. (m_0 refers to the intrinsic strength of a stationary point source. Δm_n is the change in mass due to the nth harmonic of the periodic motion. N_b is the number of

movement cycles per entire scan. ΔK are the steps by which K_y is incremented. N_y is the number of phase encoding steps. ϕ_n is the phase of periodic motion with respect to the start of data acquisition for the nth harmonic.)

Equation 3.1 basically states that periodic motion of a single slice in the slice selection direction can be modeled as amplitude modulation of the raw data in the frequency domain with a motion kernel along the phase encoding direction. Mitsa stated that the ghosting artifacts could be modeled as scaled and shifted versions of the moving objects. The effect of this amplitude modulation on the power spectrum of the MR data is that the DC power peak is shifted, scaled, and superimposed on the spectrum in the phase encoding direction. If these extra power peaks could be identified and suppressed in the original data, the amplitude modulation could be corrected. This corrected data can then be used to reconstruct an image with less artifacts.

To identify and then correct for this amplitude modulation, Mitsa's algorithm follows these steps:

1. Project the magnitude of the raw data along the x-direction to form one line of projection data with respect to positions in the y-direction.

2. Take the inverse discrete Fourier transform (DFT) of the projection data to obtain the power spectrum of the projection data.

3. Identify the extra energy peaks in the power spectrum that are related to the motion artifacts. In Mitsa's implementation, this is done picking out the points that are more than two standard deviations above a threshold in the power spectrum.

4. Apply a notch filter to the power spectrum of the projection data to eliminate the identified motion peaks. In Mitsa's implementation, the filter gain is unity for all points except for a four-point window around each peak, where the gain is determined so as to bring the local mean to the neighborhood mean.

5. Fourier transform the filtered data to obtain an estimate of the projection data without motion artifacts.

6. The ratio of projection with motion and without motion yields an estimate of the motion kernel. The motion kernel represents the amplitude modulation that had been applied to the original data.

7. Divide each row of the raw complex valued data by the motion kernel to obtain corrected data.

8. Reconstruct the motion corrected image by applying a 2D DFT on the corrected data.

Figure 3.1 shows a block diagram of the Mitsa algorithm.

3.1.2 Implementation

We will illustrate Mitsa's algorithm by simulating, and then correcting, AM motion artifacts using a mathematical abdominal phantom. The phantom image (with added Gaussian noise) is shown in Figure 3.2. Problems with ensuring proper generation of the phantom will be discussed in Chapter 6.

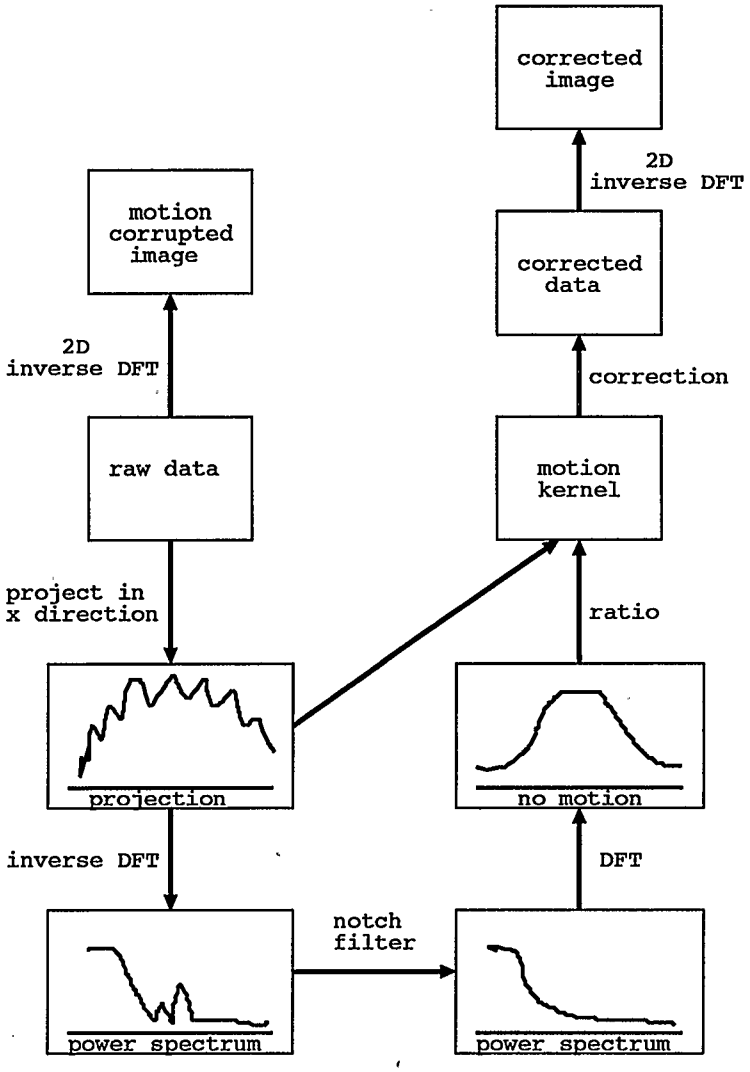


Figure 3.1: A block diagram of the Mitsa algorithm for AM motion suppression

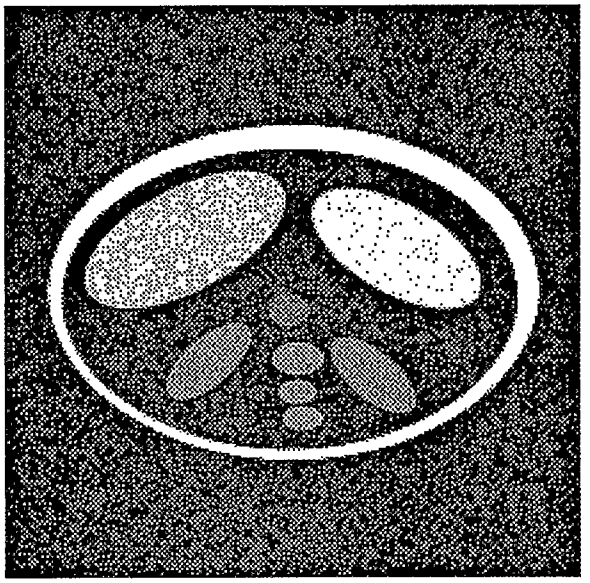


Figure 3.2: An abdomen phantom with added noise to be used for all the AM motion suppression algorithms

Motion Simulation

Following the mathematical model in equation 3.1, we simulated periodic motion using Mitsa's motion kernel:

$$G(K_Y) = 1 + 0.5sin(\frac{2\pi K_y}{12} + 0.785) + 0.15sin(\frac{2\pi K_y}{6} + 1.57) + 0.05sin(\frac{2\pi K_y}{3} + 3.141) \tag{3.2}$$

The amplitude of the sine terms decreases for higher harmonics. Only three terms in equation 3.1 were used: the dominant harmonics. The phase of the periodic motion with respect to the start of data acquisition was varied for different harmonics. The simulated kernel was plotted in Figure 3.3. We ensured that no synchronous sampling ([SMC+93]) were associated with the motion artifact, unlike the motion artifact introduced in the work by Zoroofi *et. al.* ([ZST+95]).

This kernel was then applied on the abdomen phantom without noise to simulate

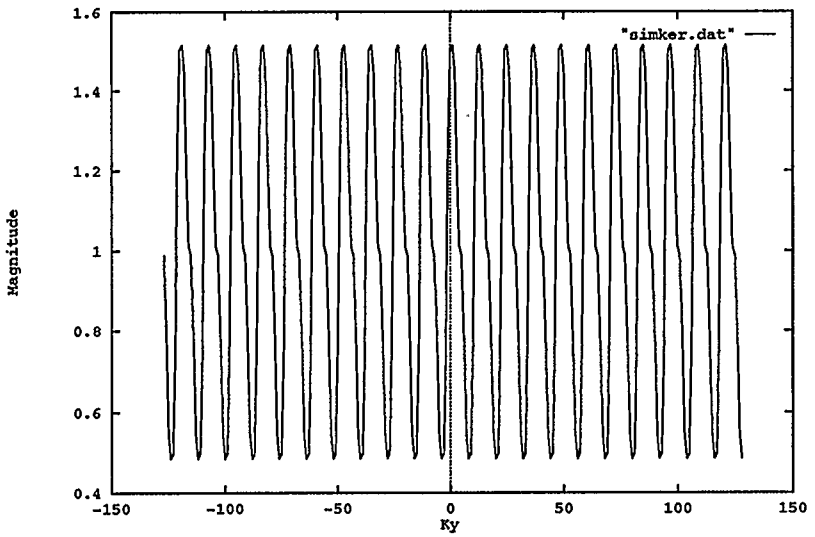


Figure 3.3: Simulated motion kernel used in the Mitsa algorithm

motion. The complex k-space data was multiplied by the motion kernel line by line. Noise was added to the motion corrupted data file. A 2D DFT was then applied to obtain the motion corrupted image (Figure 3.4).

Figure 3.4 exhibits ghosting artifacts. These artifacts are what we would expect to see in real motion corrupted images. The artifacts can blur or hide image details which may be essential for diagnosis.

Projection of the Data

In the AM model, motion results in extra energy peaks in the phase encoding direction (k_y direction). In order to identify these peaks, Mitsa took projection of the magnitude data in the k_x direction. To obtain a sharper DC peak and a better scaling scheme, the middle 31 columns of data with large DC components were removed.

The no motion (original) and motion corrupted projection data are shown in

Figure 3.4: The abdomen phantom with simulated motion, showing ghosting artifacts

Figure 3.5. The projection data is folded where points 0 to 127 correspond to k_x points 0 to 127 in conventional MR data files, points 128 to 255 correspond to -128 to -1 in conventional MR data files. From Figure 3.5 we can see the effect of amplitude modulation distortion. The motion corrupted projection has the same low frequency components as the original projection but also has high frequency terms demonstrated as ripples in the projection.

3.1.3 Power spectrum of the projection

An inverse DFT was performed on the projection data to obtain the power spectrum. The power spectra of the no motion and motion projections are shown in Figure 3.6. Only half (128 points) of the symmetrical power spectrum need to be shown. The three extra peaks in the corrupted data are related to the motion we simulated.

From Figure 3.6, we can clearly identify the extra power peaks associated with

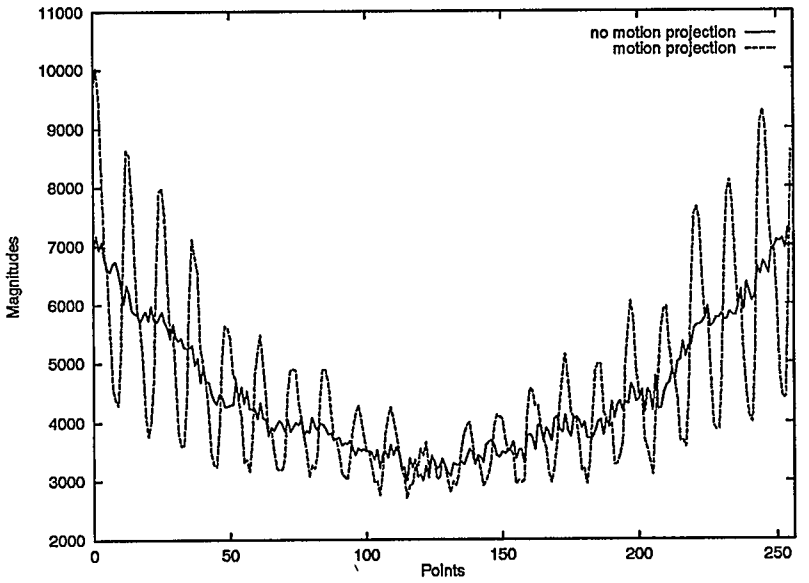


Figure 3.5: Sample projection data of no motion and motion corrupted abdomen phantom

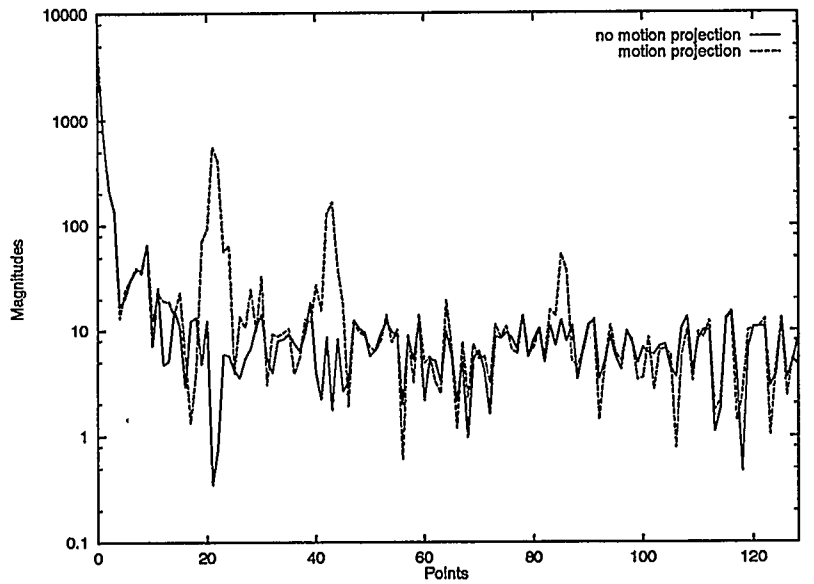


Figure 3.6: A comparison of power spectra of the motion and no motion projection

AM motion. However, locating the position and width of the peaks automatically was not an easy task. In this project we used a thresholding algorithm to detect the center and width of the peaks. Thresholding is more suitable in our case than difference operators which pick out all the peaks, since only those peaks that are well above the base line in the power spectrum can be motion peaks. Our thresholding algorithm first calculates the global mean $Mean$ and standard deviation $Stddev$ of the power spectrum excluding the DC peak. The algorithm then identifies a peak when more than two adjacent points satisfy the criteria:

$$S_i - Mean > M * Stddev \tag{3.3}$$

where S_i is the magnitude of the data and M is an adjustable constant. The algorithm assumes that the point in the peak that has a maximum magnitude is the center of the peak. Changing the constant M will change the strictness by which we define a peak. A large value of M may lead to failure in identifying smaller motion peaks; a smaller M may lead to false motion peak identification. We found that the appropriate value of M needs to be be determined by trial and error with each type of projection data. In this example we used a M value of 40.

Design of the notch filter

After the motion peaks were located, a filtering operation is necessary to remove them. The first notch filter we used was a direct application of Mitsa's notch filter. The notches in the filter are designed to bring the projection power spectrum magnitude at points in the power peaks down to the mean magnitude of the neighboring points. The neighboring points are those points in the projection power spectrum that are adjacent to the power peaks but not in the power peaks. The average mag-

nitude of these points are called the neighborhood mean magnitude. The notches in the filter power spectrum had an odd number of points. For each point we calculated the ratio of neighborhood mean magnitude and the magnitude of that point in the projection power spectrum. The center point of the notch had a gain of that ratio, while other points in the notch had gains half of that ratio. All points outside the notches had a gain of 1.

The notch filter we designed to remove the motion peaks is shown in Figure 3.7. Due to Mitsa's design simplicity, the notches had sharp edges. These sharp edges introduced unnecessary discontinuities into the power spectrum which might lead to distortions in the reconstructed motion kernel ([Har78]).

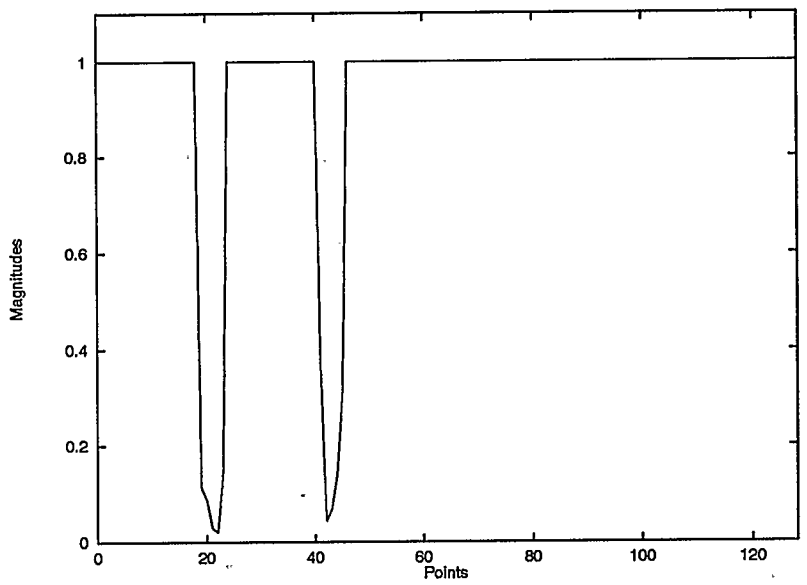


Figure 3.7: A sample notch filter using the Mitsa design for out abdomen phantom

The power spectrum of the filtered signal is shown in Figure 3.8, together with the power spectrum of projection without motion.

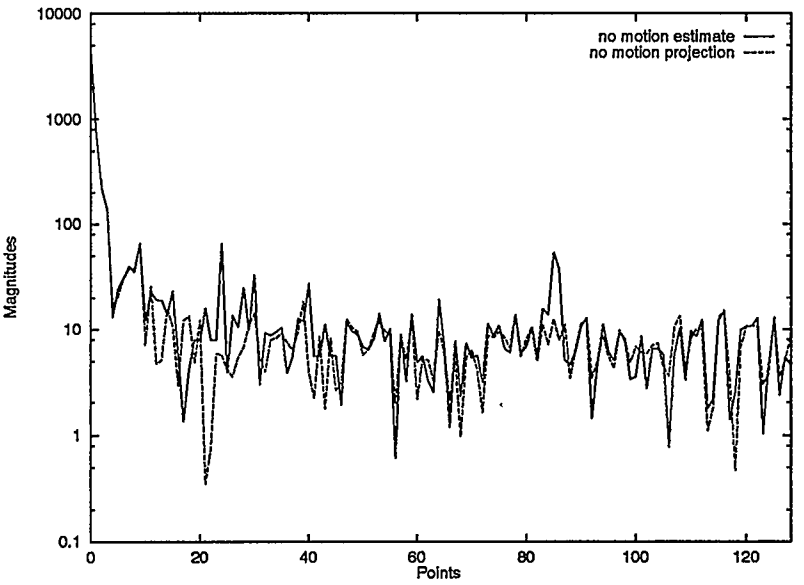


Figure 3.8: The notch-filtered power spectrum as a no motion estimate of the projection power spectrum. The actual no motion projection power spectrum (dashed line) is also shown for comparison

Motion Kernel

Mitsa assumed that the notch-filtered spectra data was an estimate of the projection power spectrum without motion (Figure 3.8). This no motion estimate power spectrum was Fourier transformed to generate an estimate of the projection without motion, which was compared to the actual projection without motion in Figure 3.9.

The motion kernel could then be extracted by dividing the motion corrupted projection by the estimate of the projection without motion. The extracted motion kernel and the original kernel are shown in Figure 3.10.

Data Correction

After the motion kernel was extracted, it could be used to correct the data. Motion corrupted data at a specific K_y value was corrected by dividing by the corresponding

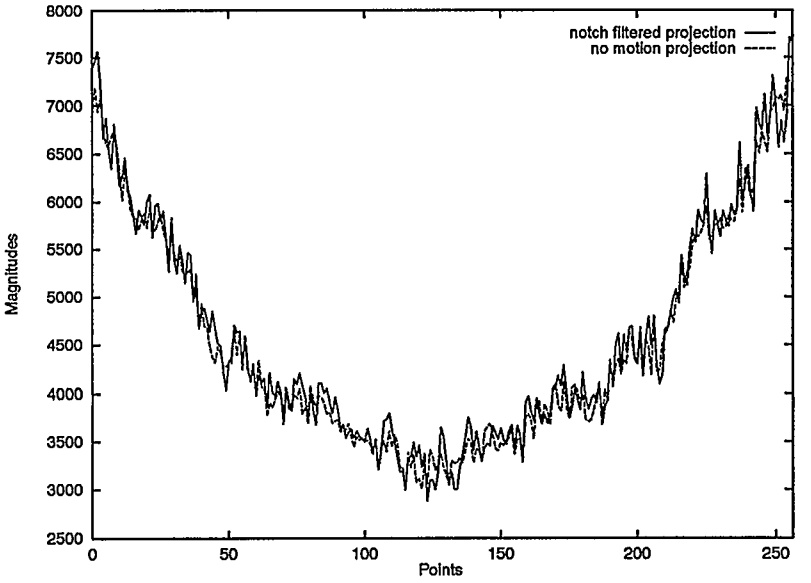


Figure 3.9: The notch filtered projection as an estimate of the no motion projection. The actual no motion projection (dashed line) is also shown for comparison

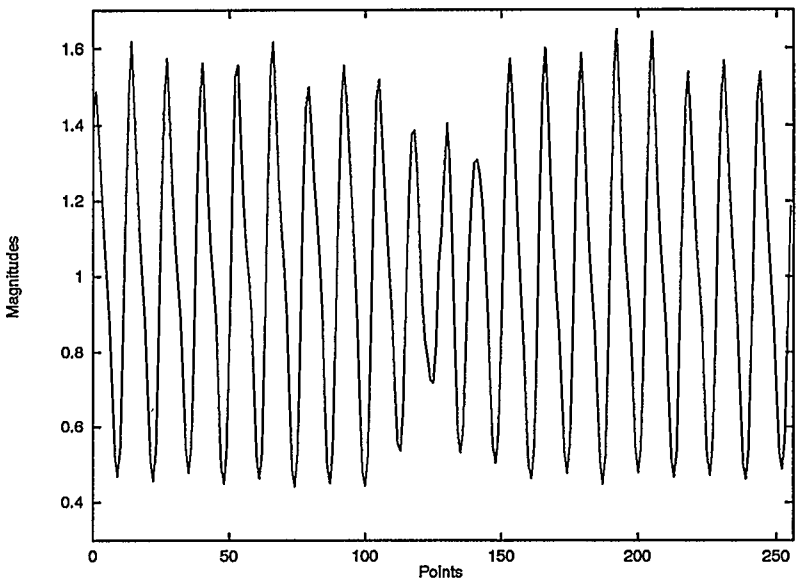


Figure 3.10: A sample motion kernel extracted using the Mitsa algorithm, compared to the original motion kernel (dashed line) used to simulate motion

value of the kernel at the same K_y. The corrected image is shown in Figure 3.11.

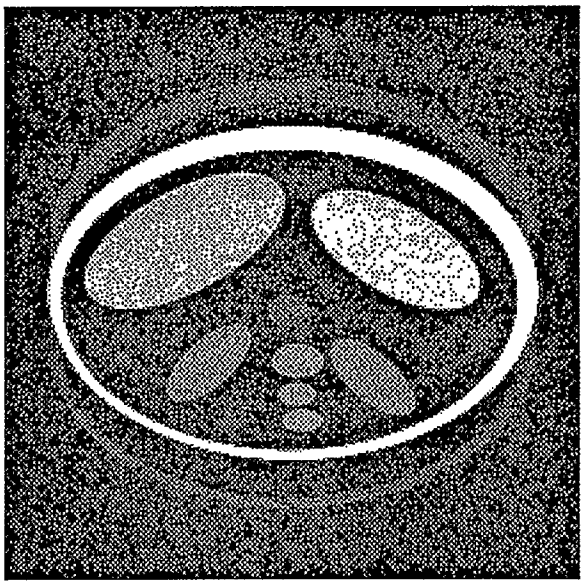


Figure 3.11: The corrected abdomen phantom using the Mitsa algorithm

Most of the noticeable motion artifacts were removed. The image was restored. Nevertheless we observed some residual motion artifacts in Figure 3.11. The residual artifacts came from the difference between the extracted kernel and the actual applied kernel. These differences are very apparent in Figure 3.10.

3.1.4 Problems with Mitsa's method

Overall, Mitsa's method worked well with our mathematical phantom. However, we found some problems with it:

- The motion peaks can be quite wide in some images. If the motion peaks overlap each other as could occur for slow movement, it would be hard for us to identify and thus remove the motion peaks.

- With low signal to noise ratio (SNR) images, some peaks could be buried in noise and not detected.

- The Mitsa notch filter introduced discontinuity into the projection power spectra, which might lead to distortions in the kernel estimate and hence in the corrected data and reconstructed image.

- The simple thresholding peak detection method did not always accurately detect peaks. Better automation of peak detection was needed.

In the following sections, we discuss our proposed AM motion suppression methods to overcome these problems.

3.2 ARMA Modeling for Motion Artifact Removal

To increase the resolution of the projection power spectra for easier identification of the motion peaks in the Mitsa algorithm, Dr. Smith suggested that we use autoregressive moving average (ARMA) modeling to generate the power spectra. We expected that the ARMA generated power spectra would have better resolution, so that more accurate peak identification would be possible.

3.2.1 Background of ARMA modeling

ARMA modeling is a parametric modeling technique. It is widely used in many signal processing fields such as speech analysis, communication spectral analysis, seismic signal analysis, etc. ARMA modeling was first introduced to MR image reconstruction by Smith *et al.* [SNHW86]. Conventionally, MR images are reconstructed using

a 2D DFT on the acquired data. Since the acquired data is a 2D truncated array of spatial frequency components of the object being imaged, it can be viewed as the result of a full spatial frequency component set multiplied by a rectangular window. As a result, the DFT reconstructed image is equivalent to the true image convolved with a *sinc* function ([GW87]). This convolution leads to ringing artifacts in the image and a loss in resolution. Modeling algorithms were proposed to reduce the effect of windowing, thus reducing ringing artifacts and improving resolution. ARMA modeling was one of the successful algorithms.

In essence, after a DFT of the data in the less truncated direction, each row or column of the MRI data set can be considered as a subset of the infinite output of an infinite impulse response (IIR) filter excited by a delta function. In order to describe our digital IIR filter and its input and output, we adopt the Z-transform notation in digital filtering. The Z-transform is a generalization of the Fourier transform which allows us to treat exponential inputs. It is commonly used in digital signal processing.

The Z-transfer function of the filter is a ratio of two polynomials, $B(Z)$ and $A(Z)$:

$$H(Z) = \frac{B(Z)}{A(Z)} \tag{3.4}$$

In signal processing, a moving average filter (MA) has the transfer function of $H_{MA}(Z) = B(Z)$. An autoregressive (AR) filter has a transfer function of $H_{AR}(Z) = 1/A(Z)$. Thus the filter used here is an autoregressive, moving average filter. A simplified block diagram of the process was shown in Figure 3.12.

Smith's algorithm has the following steps [SNHW86]:

1. Each row or column of the data is split into Hermitian and anti-Hermitian

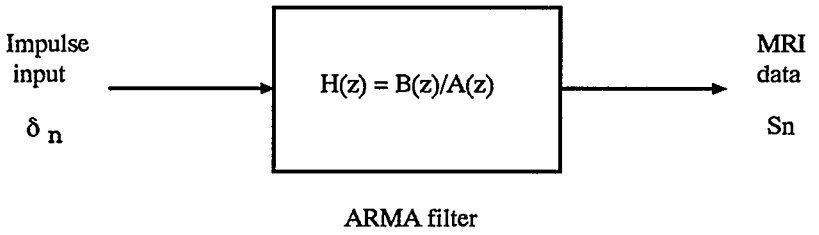


Figure 3.12: An illustration of the ARMA filtering concept

series to account for data symmetry.

2. Each series is modeled as the output of the ARMA filter. The acquired data is used to determine the AR and MA filter coefficients, respectively.

3. The Fourier transform of the infinite data set is estimated from the filter coefficients for the Hermitian and anti-Hermitian series.

A more detailed block diagram of the transient error reconstruction algorithm (TERA) is shown in Figure 3.13.

The AR and MA coefficients are determined separately. For the AR filter, we can write the input and output equation as:

$$x_n = -\sum_{i=1}^{p} a_i x_{n-i} + \epsilon_n \tag{3.5}$$

where a_i is the ith AR coefficient, p is the order of the filter, x_n is the desired output of the filter, and ϵ_n is the forward prediction error. Equation 3.5 represents a classic prediction problem. The goal is to find a set of a_i's that minimize the error using some algorithm, such as a least square algorithm.

Since the excitation signal is a delta function, the following simple relationship holds between the MA coefficients b_n and the error ϵ_n:

$$b_n = \epsilon_n \tag{3.6}$$

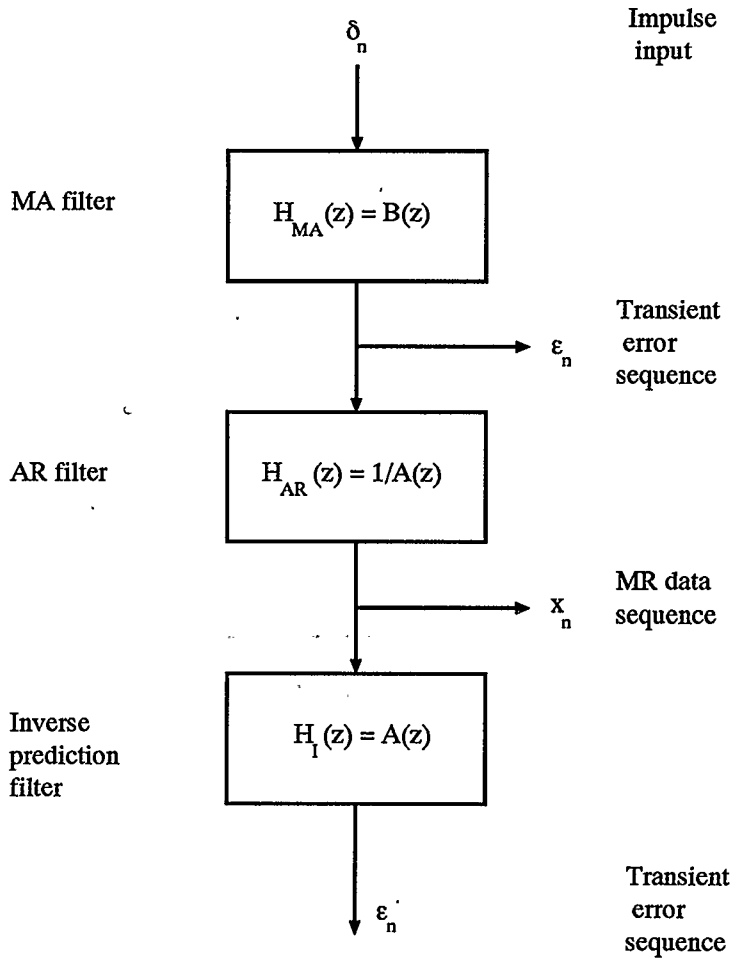


Figure 3.13: A block diagram of the ARMA reconstruction algorithm

Therefore if the error sequence is known, the MA coefficients are known.

Finally, the image line is calculated by:

$$\begin{aligned} FT[x_n] &= (B(Z))/(A(Z))FT(\delta_n) \\ &= (FT[\epsilon_n])/(FT[a_n]) \end{aligned} \tag{3.7}$$

Equation 3.7 shows that the image can be reconstructed using only the ARMA coefficients. Even though the AR and MA coefficients a_n and b_n are of finite lengths, the resulting image line $FT[x_n]$ can have infinitely many terms. As a result, we have implicitly extrapolated the higher frequency data components using modeling.

ARMA modeling has been shown to successfully reduce ringing artifacts and improve resolution of MR images. It is very flexible regarding the kind of data it is used on. It is also computationally efficient ([LBC+92]).

3.2.2 Application of ARMA modeling to AM motion suppression

One problem with the Mitsa algorithm is that the projection power spectrum is generated using DFT. The finite length of the projection data implies a rectangular window applied on the acquired data, leading to widening of the motion peaks and loss of resolution. Widened peaks make accurate motion peak identification difficult.

ARMA modeling's success in reconstruction is due to its ability to extrapolate higher spatial frequency components. Extrapolation leads to better resolution and less ringing artifacts in reconstructed images. If used in generating the projection power spectrum, ARMA modeling can be expected to improve the resolution of the power spectrum, thus sharpening the motion peaks and making them easier for accurate identification ([ZSC94]). The ARMA generated power spectrum of the

projection data is shown in Figure 3.14. The power spectrum is zoomed four times in Figure 3.15 to better show the sharpened motion peak.

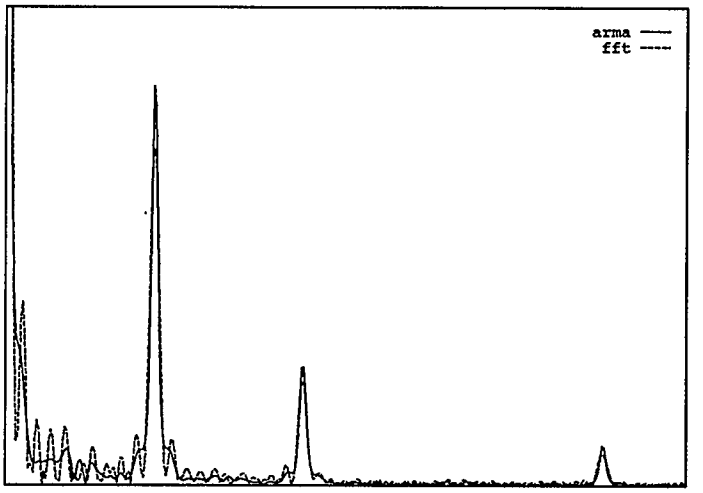


Figure 3.14: ARMA generated power spectrum of the projection data (solid line), compared to DFT generated power spectrum (dashed line)

Figure 3.16 shows the effective extrapolation of modeling on the no motion projection estimate. Instead of the sharp cutoff at high frequency range (dotted line), we can see an extrapolation of the projection data. Such extrapolated projection data yielded a power spectrum with sharper peaks.

3.2.3 BH3 filter applied in the ARMA spectrum

Mitsa's original notch filter was based on neighborhood magnitudes. The filter was not smooth and might introduce discontinuity into the filtered power spectrum. In signal processing we normally want to minimize such distortions ([Har78]). Some low distortion filters have been proposed in the literature to enhance performance in

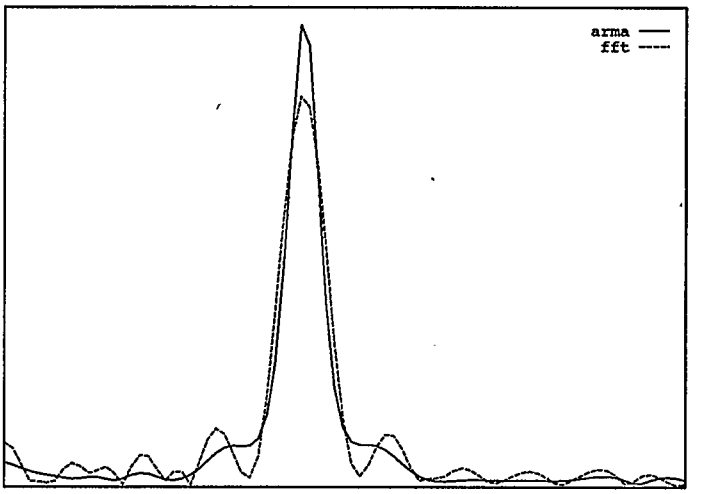


Figure 3.15: A closer look of the middle power peak in the power spectrum. The ARMA generated power peak (solid line) is sharper than the DFT generated one (dashed line)

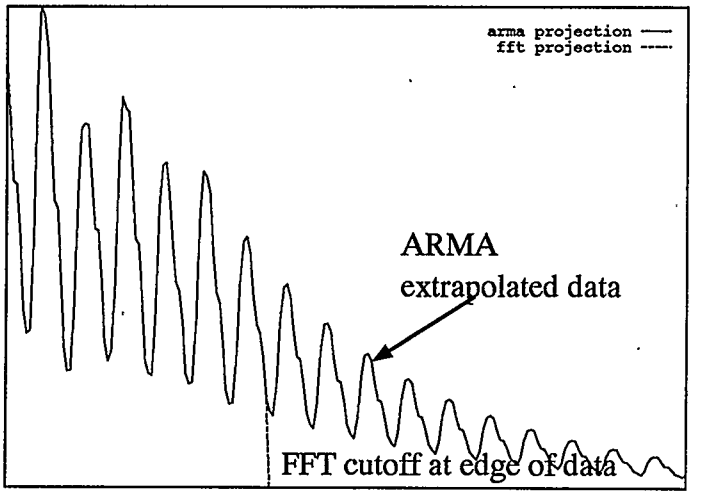


Figure 3.16: Extrapolation of the projection data using ARMA modeling to generate the projection power spectrum (solid line), compared to that using DFT (dashed line)

this respect.

One of the low distortion windows proposed is the three term Blackman and Harris (BH) window. This window has been shown to introduce minimal discontinuity ([Har78]). The mathematical expression of this window is:

$$G(i) = 1 - (A_0 + A_1 * cos(\frac{2\pi i}{2W_h + 1}) + A_2 * cos(\frac{2\pi 2i}{2W_h + 1})) \tag{3.8}$$

where $G(i)$ is the BH filter gain for the ith point inside the window, and A_0, A_1, A_2 are coefficients to control the gain of the notch. W_h is the half-window size, in number of points.

The BH notch filter we used to remove motion peaks is plotted in Figure 3.17. Note that it has smooth stop-bands.

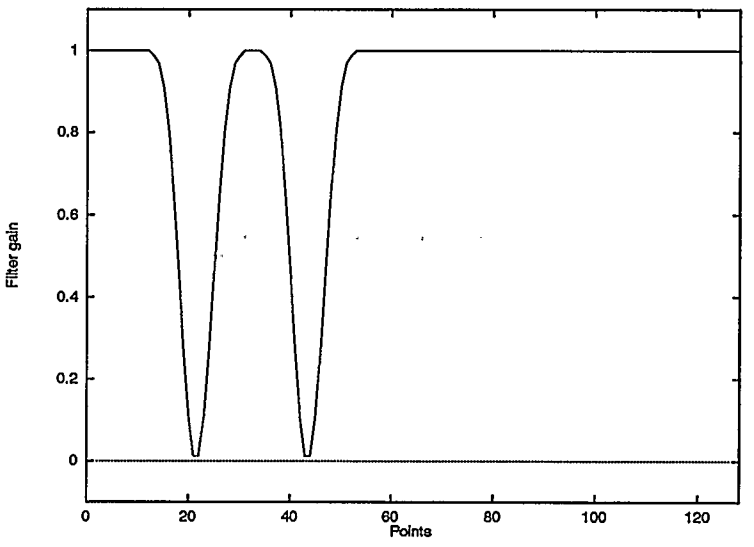


Figure 3.17: A sample Blackman-Harris notch filter we used to suppress the motion peaks in the projection power spectrum

3.3 The Direct method

One of the shortcomings of the Mitsa algorithm is that its performance is sensitive to image details and noise. In an attempt to reduce the effect of image details and noise, we designed the Direct method to extract only the DC and motion peaks. All the data points in the spectrum outside these peaks were set to zero. Thus amplitude modulation was considered while all other image information was ignored. A block diagram of the method is shown in Figure 3.18. We hoped that, by discarding the effect of image details and noise, we would be able to extract better defined motion peaks and obtain a better motion kernel.

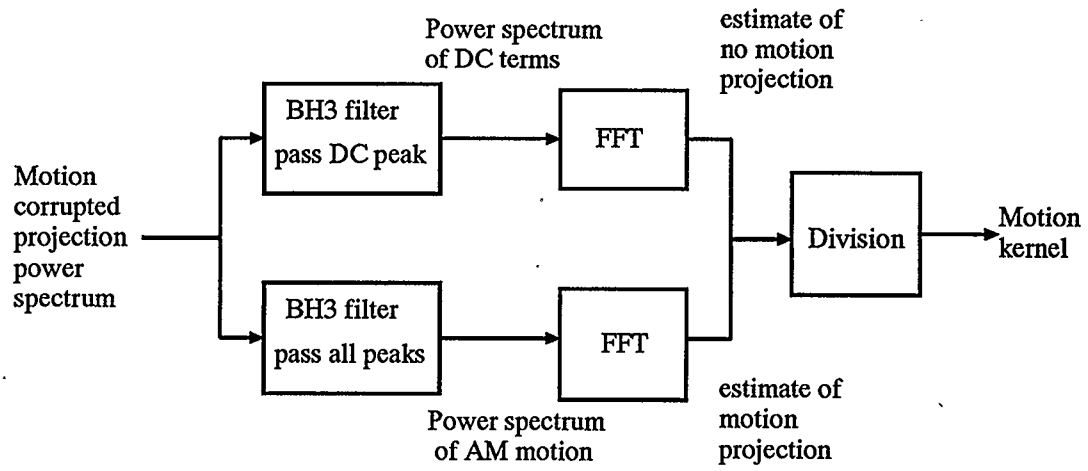


Figure 3.18: Block diagram of the Direct method

To obtain an estimate of the projection data with motion, we first identify the motion peaks in the ARMA generated power spectrum using our thresholding algorithm (Section 3.1.3). A BH3 filter with a series of narrow pass-bands (Figure 3.19, lower) was then used to extract the DC peak and all the motion peaks. The pass-band gain was set to 1, to extract the peaks; The stop band gain was set to 0, to

discard all the frequency components outside the peaks. This assumption is valid only for sharp peaks. We used ARMA modeling to generate the power spectra in order to get appropriate sharper peaks. The filtered power spectrum (Figure 3.20) was then Fourier transformed to obtain motion corrupted projection data without details (Figure 3.21).

The estimate of projection data without motion was obtained by Fourier transforming the DC peak. The DC peak was extracted by applying a BH3 filter with only one narrow passband around DC (Figure 3.19 upper). The filtered power spectrum and projection estimate were shown in Figure 3.20 and Figure 3.21. The motion kernel was then calculated by the ratio of projection estimate with and without motion.

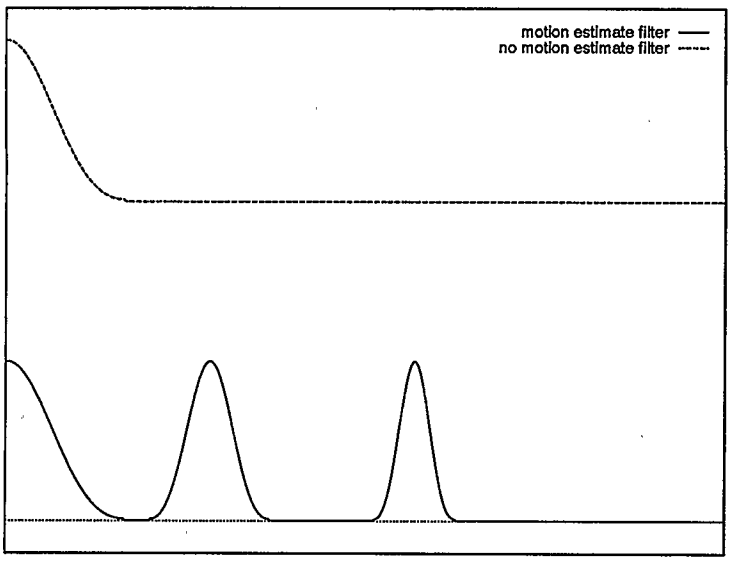


Figure 3.19: Sample BH3 passband filters for the Direct method. The top one is intended to filter the DC peak only; The bottom one is adjusted to filter both the DC and motion peaks

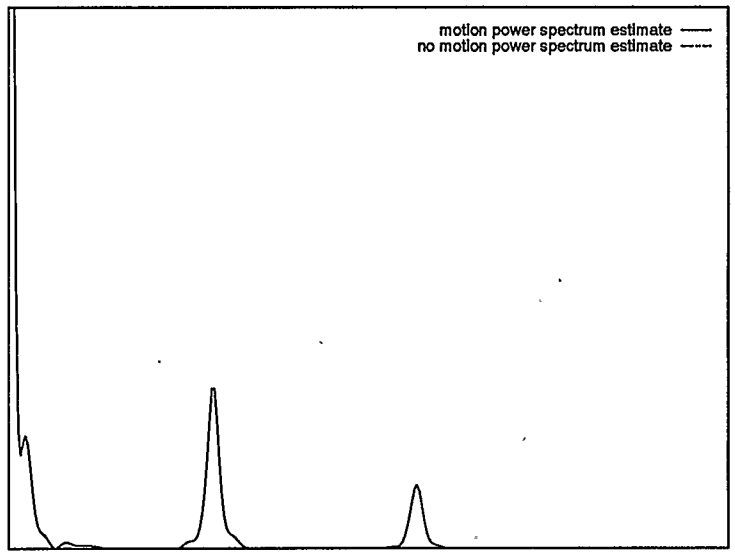


Figure 3.20: Sample filtered power spectra in the Direct method

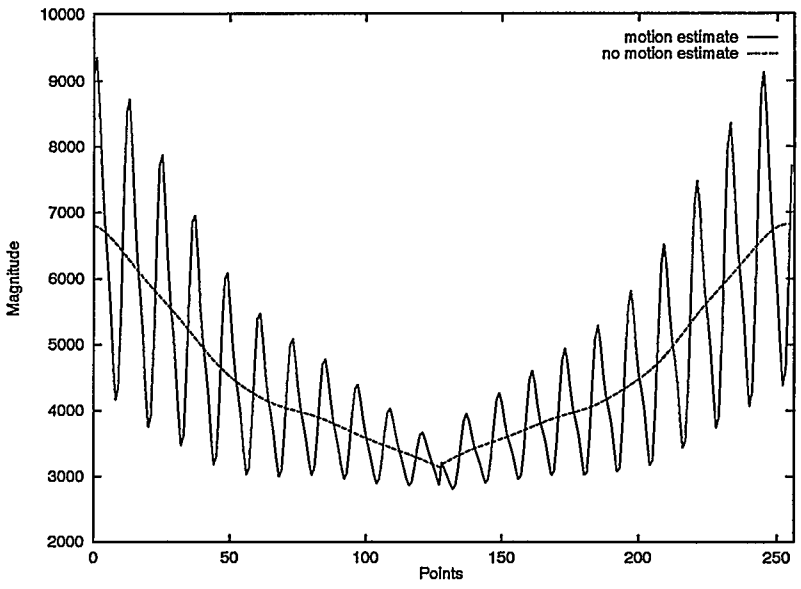


Figure 3.21: Sample projection estimates in the Direct method

3.4 The Adaptive Algorithm

The ARMA algorithm improved the accuracy of peak identification in the Mitsa algorithm and thus improved the quality of motion suppression. However, complete peak detection automation was not realized. The thresholding algorithm for peak detection used in the above algorithms needed tuning by trial and error. Another peak detection algorithm ([Sez90]) was tested and yielded little improvement. Performance of all the algorithms were largely sensitive to the choice of width and center of the power peaks. To completely automate peak detection and reduce the sensitivity of the algorithm to peak identification, we proposed the adaptive algorithm. Adaptivity in this case was exploited to automatically find the peaks and suppress them.

3.4.1 Background

Adaptive filtering techniques have been widely used in communications to solve problems with a changing nature. Most of the applications involve restoring distorted signals (noisy, modulated, dispersed, etc.). Usually a subset of the distorted signal is compared to a known reference signal. Both signals are used to train an adaptive filter so as to minimize the difference or error signal in some way. The filter coefficients are fixed after the training and the filter is used to restore the distorted signal. (See Figure 3.22)

Adaptive filtering algorithms can be categorized into three groups based on the derivation of the algorithms [Hay91]. The third category consists of three different classes of algorithms, depending on the structure used for implementation.

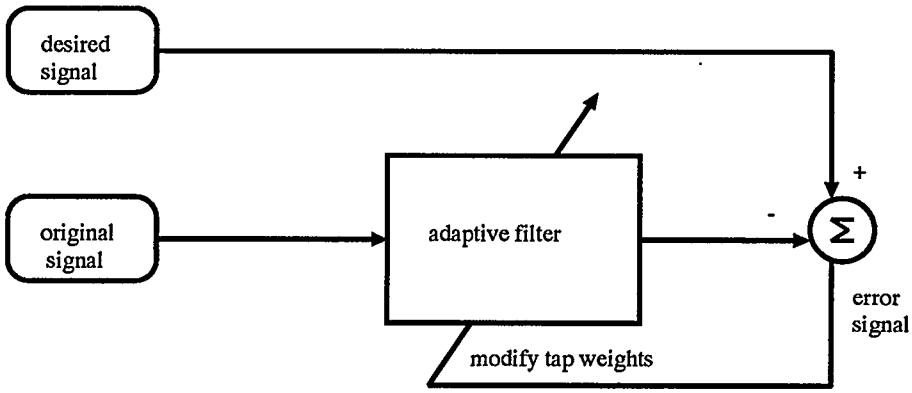


Figure 3.22: Adaptive filtering concept

- Methods based on the Wiener filter theory (LMS algorithm)
- Methods based on the Kalman filter theory
- Methods of Least Square fit:
  - Recursive least-squares (RLS) algorithm
  - Least-squares lattice (LSL) algorithm
  - QR decomposition least-squares algorithm

Based on experience in 1D adaptive signal processing in communications ([Zen94]), the RLS algorithm was chosen as the algorithm to be implemented. The RLS algorithm gives fast convergence rate, low steady state error, and is computationally inexpensive.

3.4.2 The Adaptive algorithm for AM motion suppression

We will first formulate the motion suppression problem in adaptive filtering terms. The projection data with motion is an amplitude modulated data set. The objective

is to remove the amplitude modulation using an adaptive filter. The filter must first be trained using the AM data and some estimate of the desired projection data.

The procedures for our adaptive algorithm can be summarized as follows:

1. Using the k-space motion corrupted data, project the magnitude of each complex pair in the x-direction.

2. Inverse Fourier transform the projection data to obtain the corresponding power spectrum.

3. Low-pass filter the power spectrum to obtain only the DC and low frequency terms. A three term Blackman and Harris filter is used since it introduces minimum distortion.

4. Fourier transform the filtered DC and low frequency data to generate an estimate of the desired projection data.

5. Initialize and train a transversal filter using the RLS algorithm. The input to the filter is the motion corrupted projection data. The estimate of the desired projection data is used as the desired signal.

6. Once the filter tap weights are determined, process the motion corrupted projection data using the filter, to yield the estimated projection data without motion.

7. Use the ratio of the projection data with and without motion to obtain the motion kernel.

8. Correct the motion corrupted raw data using the motion kernel.

9. Reconstruct the image by applying a 2D DFT on the corrected data.

The next section discusses in detail the RLS algorithm we used to implement our adaptive motion suppression algorithm.

3.4.3 The RLS algorithm

The RLS algorithm is a deterministic (as opposed to statistical) adaptive filtering algorithm using a transversal filter structure. Based on the *method of least squares*, we minimize the sum of squared errors, where the error is defined as the difference between a desired signal and the actual filter output. The desired signal is usually a known sequence used as a training signal to determine the filter tap weights.

Minimization of the normal equation

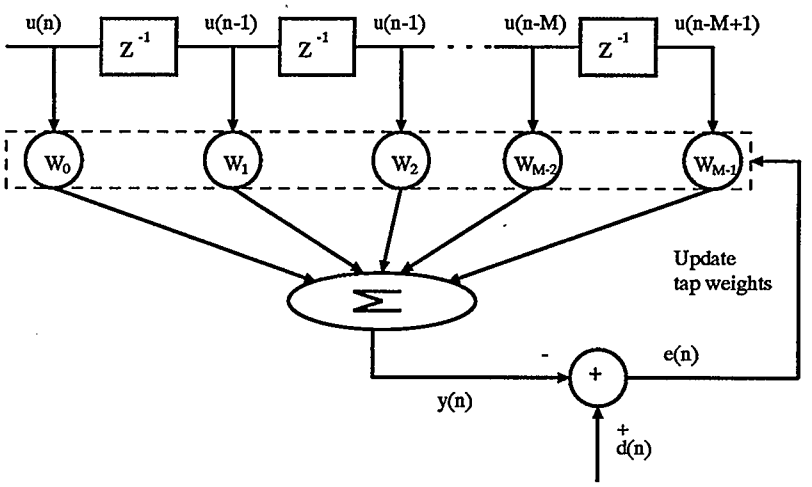


Figure 3.23: A transversal filter with M tap weights

The filter structure, input and output signals are shown in Figure 3.23. At iteration n, $u(n)$ and $y(n)$ are the input and output of the filter, $d(n)$ is the desired signal,

$e(n)$ is the error signal, w_i is the ith filter tap weight, and Z^{-1} denotes a unit time delay in the filter. The output signal y for an M-tap filter is:

$$\begin{aligned} y(i) &= \sum_{k=0}^{M-1} W(k)u(i-k+1) \\ &= \tilde{W}^T\tilde{u}(i) \end{aligned} \tag{3.9}$$

where $\tilde{u}(i)$ is the input vector and $\tilde{W}^T$ is the transpose of the weight vector.

The error signal e is:

$$\begin{aligned} e(i) &= d(i) - y(i) \\ &= d(i) - \tilde{W}^T\tilde{u}(i) \end{aligned} \tag{3.10}$$

The object function J to be minimized is:

$$J = \sum_{i=i1}^{i2} e^2(i) \tag{3.11}$$

where $i1 <= i <= i2$ is the window for the least square estimate. J can be calculated as [Ses94]:

$$\begin{aligned} J &= \sum_{i=i1}^{i2}[d^2(i) - 2\tilde{W}^T\tilde{u}(i)d(i) + \tilde{W}^T\tilde{u}(i)\tilde{u}^T(i)\tilde{W}] \\ &= \sum_{i=i1}^{i2} d^2(i) - 2\tilde{W}^T[\sum_{i=i1}^{i2}\tilde{u}(i)d(i)] + \tilde{W}^T[\sum_{i=i1}^{i2}\tilde{u}(i)\tilde{u}^T(i)]\tilde{W} \end{aligned} \tag{3.12}$$

To simplify the above expression, we define:

$$\begin{aligned} R &= \sum_{i=i1}^{i2}\tilde{u}(i)\tilde{u}^T(i) \\ \tilde{\Phi} &= \sum_{i=i1}^{i2}\tilde{u}(i)d(i) \\ D &= \sum_{i=i1}^{i2} d^2(i) \end{aligned} \tag{3.13}$$

Therefore,

$$J = D - 2\tilde{W}^T\tilde{\Phi} + \tilde{W}^T R\tilde{W} \tag{3.14}$$

To minimize the object function J, we need:

1. A necessary condition

$$\partial J/\partial W = 0 \tag{3.15}$$

2. A sufficient condition

$$\partial^2 J/\partial^2 W > 0 \tag{3.16}$$

To satisfy the necessary condition, we have:

$$\begin{aligned} \nabla J &= \nabla D - 2\nabla(\tilde{W}^T\tilde{\Phi}) + \nabla(\tilde{W}^T R\tilde{W}) \\ &= -2\tilde{\Phi} + 2R\tilde{W} \qquad (3.17) \\ &= 0 \end{aligned}$$

i.e.

$$R\tilde{W} = \tilde{\Phi} \tag{3.18}$$

Equation 3.18 is called the deterministic normal equation.

The sufficient condition can be simplified as follows:

$$\begin{aligned} \nabla(\nabla J) &= \nabla[-2\tilde{\Phi} + 2R\tilde{W}] \\ &= 2R \qquad (3.19) \\ &> 0 \end{aligned}$$

Since the cross-correlation matrix of the input R is positive definite for any real signals, this condition is naturally satisfied. As a result, the normal equation is the one that we must solve.

Recursion

Due to the difficulty of solving the normal equation directly, recursive techniques are used instead. The RLS algorithm is such an recursive algorithm. This algorithm updates the filter tap weights as shown below ([Ses94]):

1. Calculate the input correlation matrix at time n:

$$R(n) = R(n-1) + \tilde{u}(n)\tilde{u}^T(n) \tag{3.20}$$

2. Calculate the cross-correlation vector at time n:

$$\tilde{\Phi}(n) = \tilde{\Phi}(n-1) + d(n)\tilde{u}(n) \tag{3.21}$$

3. Invert $R(n)$ to obtain $R^{-1}(n)$

4. Calculate the filter tap weights at time n:

$$\tilde{W}(n) = R^{-1}(n)\tilde{\Phi}(n) \tag{3.22}$$

For the matrix inversion in step 3 we use the matrix inversion lemma:

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1} \tag{3.23}$$

In our problem, from equation 3.20, we set:

$$\begin{aligned} A &= R(n-1) \\ B &= \tilde{u}(n) \\ C &= 1 \\ D &= \tilde{u}^T(n) \end{aligned} \tag{3.24}$$

and the matrix inversion can be calculated as:

$$\begin{aligned} R^{-1}(n) &= R^{-1}(n-1) - R^{-1}(n-1)\tilde{u}(n)[\tilde{u}^T(n)R^{-1}(n-1)\tilde{u}(n)+1]^{-1}\tilde{u}^T(n)R^{-1}(n-1) \\ &= R^{-1}(n-1) - (R^{-1}(n-1)\tilde{u}(n)\tilde{u}^T(n)R^{-1}(n-1))/(1+\tilde{u}^T(n)R^{-1}(n-1)\tilde{u}(n)) \end{aligned} \tag{3.25}$$

For convenience, we define:

$$\begin{aligned} P(n) &= R^{-1}(n) \\ &= P(n-1) - (P(n-1)\tilde{u}(n)\tilde{u}^T(n)P(n-1))/(1+\tilde{u}^T(n)P(n-1)\tilde{u}(n)) \\ &= P(n-1) - (P^T(n-1)\tilde{u}(n)\tilde{u}^T(n)P(n-1))/(1+\tilde{u}^T(n)P(n-1)\tilde{u}(n)) \end{aligned} \tag{3.26}$$

since, for a symmetric correlation matrix, $P(n-1) = P^T(n-1)$.

To simplify the equation, we define another variable:

$$\tilde{K}(n) = \frac{P(n-1)\tilde{u}(n)}{1+\tilde{u}^T(n)P(n-1)\tilde{u}(n)} \tag{3.27}$$

Using equation 3.27 to further simplify equation 3.26, we get:

$$P(n) = P(n-1) - \tilde{K}(n)\tilde{u}^T(n)P(n-1) \tag{3.28}$$

$K(n)$ can also be calculated from equations 3.27 and 3.28:

$$\tilde{K}(n)[1 + \tilde{u}^T(n)P(n-1)\tilde{u}(n)] = P(n-1)\tilde{u}(n) \tag{3.29}$$

$$\begin{aligned} \tilde{K}(n) &= \tilde{u}(n)[P(n-1) - \tilde{K}(n)\tilde{u}^T(n)P(n-1)] \\ &= P(n)\tilde{u}(n) \end{aligned} \tag{3.30}$$

The solution for the normal equation is:

$$\begin{aligned} \tilde{W} &= R^{-1}(n)\tilde{\Phi}(n) \\ &= P(n)\tilde{\Phi}(n) \\ &= [P(n-1) - \tilde{K}(n)\tilde{u}^T(n)P(n-1)][\tilde{\Phi}(n-1) + \tilde{u}(n)d(n)] \\ &= \tilde{W}(n-1) + \tilde{K}(n)[d(n) - \tilde{u}^T(n)\tilde{W}(n-1)] \\ &= \tilde{W}(n-1) + \tilde{K}(n)\alpha(n) \end{aligned} \tag{3.31}$$

where $\alpha(n) = d(n) - \tilde{u}^T(n)\tilde{W}(n-1)$. $\alpha(n)$ is the error term we use to minimize the mean-squared error.

Summary of the RLS algorithm

The standard RLS algorithm was applied in this project. The following summarizes the algorithm based on the above derivation. The algorithm includes the initialization of variables at iteration $N = 0$ and update of variables at each subsequent iteration $N = n$.

1. At $N = 0$, set up the following initial conditions:

$$P(0) = \delta^{-1} I \tag{3.32}$$

where I =identity matrix, $\delta << 1$;

$$W(0) = 0 \tag{3.33}$$

2. At each step $N = n$, $n \in$ positive integers:

$$\tilde{K}(n) = \frac{P(n-1)\tilde{u}(n)}{1 + \tilde{u}^T(n)P(n-1)\tilde{u}(n)} \tag{3.34}$$

$$P(n) = P(n-1) - \tilde{K}(n)\tilde{u}^T(n)P(n-1) \tag{3.35}$$

$$\alpha(n) = d(n) - \tilde{W}^T(n-1)\tilde{u}(n) \tag{3.36}$$

$$\tilde{W}(n) = \tilde{W}(n-1) - \tilde{K}(n)\alpha(n) \tag{3.37}$$

At each iteration, we adjust the tap weight vector $\tilde{W}(n)$ by adding the scaled error term to $\tilde{W}(n-1)$.

3.4.4 Implementation

The filter training signals

The image we used in this project was the abdomen phantom shown in Figure 3.2, on page 26. The same motion kernel as in the Mitsa algorithm was used to simulate motion. A projection operation was then performed on the motion corrupted abdomen phantom. The projection data was used as the input signal for transversal filter training.

We used the DC and low frequency components of the motion corrupted projection data as an estimate of the desired projection data. These frequency components contain information of the shape of the images without details. Our simulated AM motion corruption has little effect in this frequency range. A three term Blackman-Harris window was used to filter out the first 15 terms in the power spectrum. The filtered power spectrum is then Fourier transformed to produce an estimate of the no motion projection. We use this estimate as the desired signal to train the adaptive transversal filter.

Both the desired and input signal for filter training are plotted in Figure 3.24.

Filter training

The filter structure used in our algorithm was a 9-tap transversal filter as shown in Figure 3.23. The length of training depended on the application. In this project, the RLS algorithm was efficiently implemented and was computationally inexpensive. As a result, we trained our 9-tap transversal filter with the whole data set. The frequency response of the trained filter is shown in Figure 3.25. The valleys in the frequency spectrum corresponded to the location of extra energy peaks in the motion corrupted

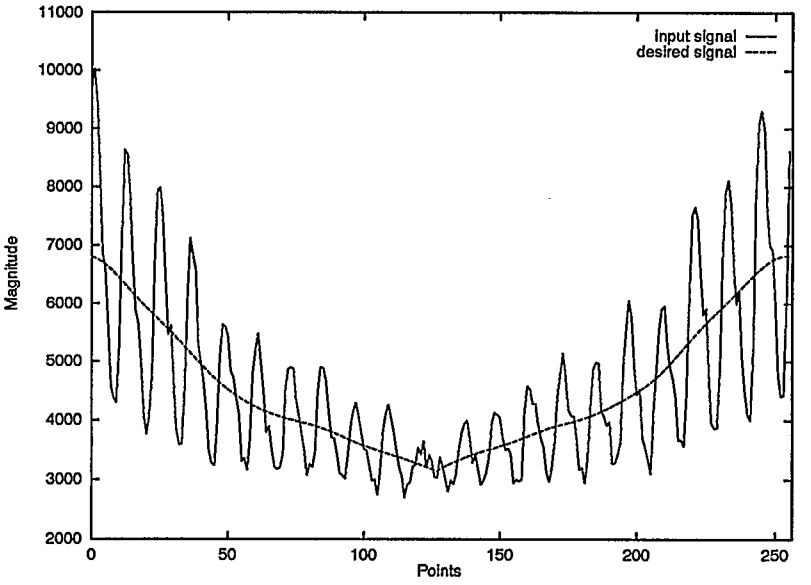


Figure 3.24: The desired (dashed line) and input (solid line) signals to the adaptive transversal filter for training

projection power spectrum. The trained adaptive filter automatically identified all three motion peaks. The adaptive filter also identified a non-motion peak in the high frequency range of the projection data. The effect of suppressing this extra peak is not clear to us.

Motion correction

Once the filter tap weights were determined, the motion corrupted data was sent through the filter and the output was used as a projection without motion. The ratio of projection data with and without motion was used for the estimated motion kernel. The motion kernel was then used on the original motion corrupted data to correct for the motion.

3.5 Performance

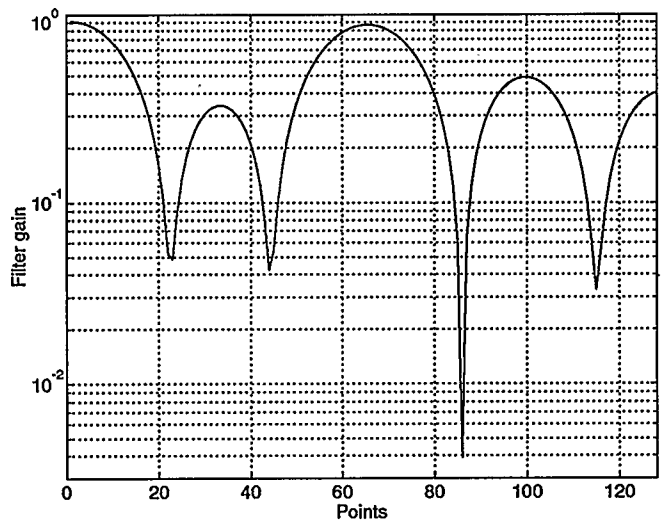
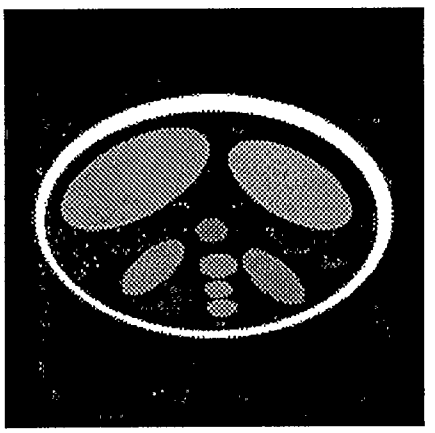


Figure 3.25: Frequency spectrum of the RLS filter after training to suppress the motion peaks

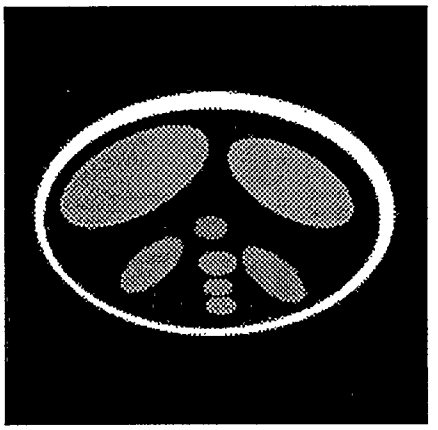
We present the motion corrected images produced with different algorithms in Figure 3.26. To the untrained eye there appears to be little visual differences between the corrected images. However, if we compare the motion kernels determined by the algorithms (Figure 3.27), there is considerable differences in the algorithms.

How do we compare the performance of different algorithms? Is there a convincing way to demonstrate which algorithm would perform the best under clinical conditions, for example when looking for lesions? The lack of convincing evaluation methods has been a long standing problem in image processing. Without proper evaluation methods, improvement of algorithms is hard to prove.

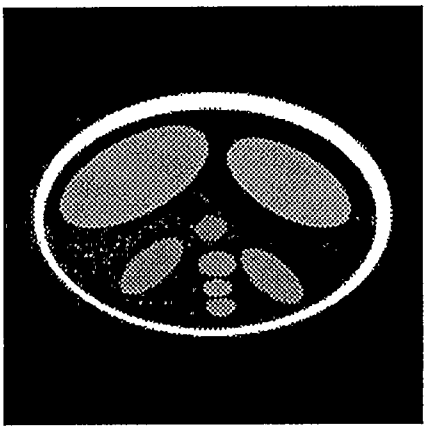
Meanwhile, a few difficulties blocked me from further investigating motion suppression in MRI. The AM model we used was over-simplified. When we applied our algorithms based on this model on a real motion corrupted abdomen image, the



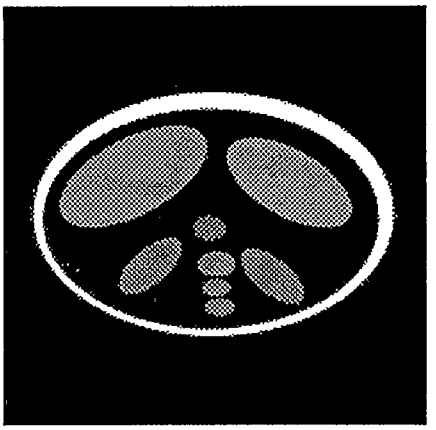
Mitsa



ARMA



DIRECT



ADAPTIVE

Figure 3.26: Motion corrected images

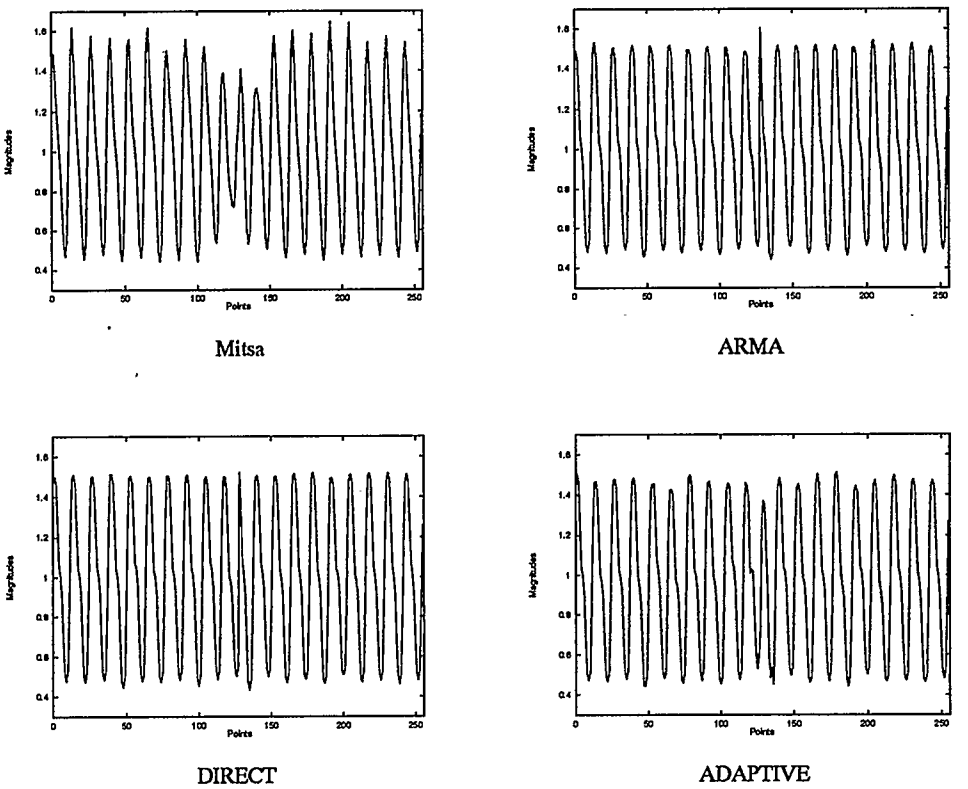


Figure 3.27: Motion kernels extracted using the different motion suppression algorithms

results were disappointing (Figure 3.28), but not unexpected. Figure 3.28 shows the motion corrupted abdomen image on the left hand side and the motion corrected image using the Adaptive algorithm on the right hand side. The ghosting artifacts were not removed. We need better motion models to handle real motion in MRI. Such models were not available from my literature survey. I thought of using the 3D motion model presented in Chapter 2 for 3D data processing. However, we did not have access to an MRI machine and acquiring 3D data was impossible. Developing our own motion model was also impossible due to my lack of physics and medical background. It was at this point that we decided to devote the rest of the thesis to developing quantitative evaluation methods.

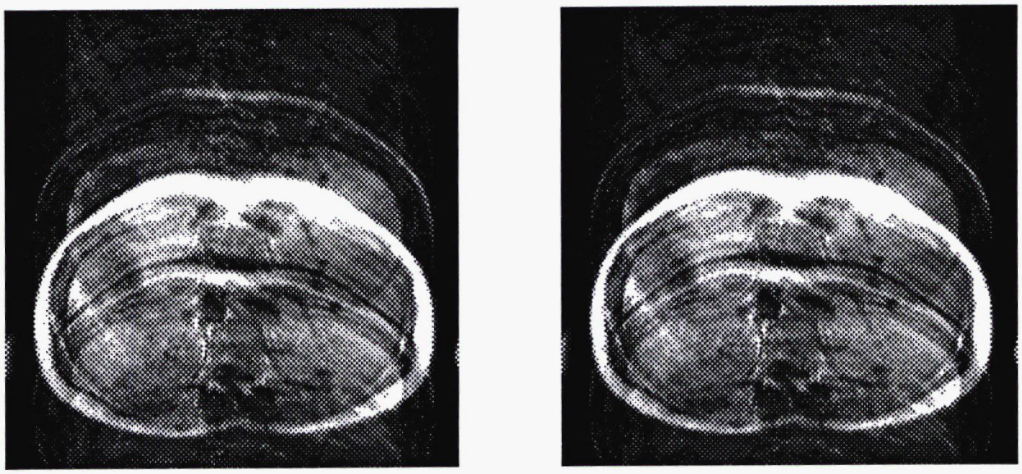


Figure 3.28: A real motion corrupted MRI abdomen image (left) and the abdomen image after motion suppression (right) using the Adaptive algorithm based on the simplified AM model

In the next part of the thesis, we set out to investigate quantitative evaluation methods that would yield convincing algorithm performance results. We hope that

our effort will help to set up an integrated environment for developing and evaluating MRI post-processing algorithms.

3.6 Summary

We presented four motion suppression algorithms and their implementation in this chapter. The algorithms are: the Mitsa algorithm, the ARMA algorithm, the Direct algorithm and the Adaptive algorithm. The last three algorithms were proposed to improve the performance of the Mitsa algorithm. However, we need better evaluation methods to be able to examine the differences in performance of these algorithms, and to determine if the differences are numerical or clinical.

Chapter 4

The Image Quality Evaluation Problem

Quantitative evaluation of image quality has been a long-standing problem. Ideally, we'd like to obtain a figure of merit (FOM) to indicate how good an image is for a given clinical task. Two approaches have been commonly used in the medical society([Wag86]). One involves physical characterization of the equipment and images, which includes measuring the modulation transfer function (MTF), signal to noise ratio (SNR), etc. Usually, these measurements can be taken quite easily. However, the clinical relevance of such measurements is difficult to determine. The other approach is clinical receiver operator characteristics (ROC) analysis based on signal detection theory (SDT). ROC analysis yields clinically relevant results for image evaluation. However, a large database of clinical cases is required in order to perform this analysis. Such a database is hard to acquire. In addition, one such costly test is usually valid for a narrow range of clinical cases only. An alternative to clinical evaluation is phantom ROC analysis. Phantom study can be meaningful when proper procedures are followed.

In this chapter, we present a survey of existing approaches for quantitative evaluation of algorithms and techniques in image processing.

4.1 Some Standard Metrics

The intuitive approach to image quality evaluation is physical measurement. Physical measurements include such factors as contrast, resolution, signal to noise ratio (SNR), etc. When distortions are assessed, measures based on the mean squared error (MSE) between the original image and the distorted or corrected image can be used. Usually, the SNR and resolution are determined during data acquisition and are not affected much by post-processing algorithms. However, contrast and MSE-based metrics can be used to quantify distortions in the images and evaluate algorithm performance.

4.1.1 Contrast

Contrast is an important measure for medical images since it affects clinical diagnosis directly. It can be defined as:

$$C = \frac{I_o - I_b}{I_o + I_b} \tag{4.1}$$

where I_o is the average intensity of an object and I_b is that of the background. Contrast can be used to compare different algorithms for lesion detection. Higher contrast normally leads to more efficient diagnosis and so it is preferable.

4.1.2 MSE based metrics

The MSE for a M by N image can be defined as:

$$MSE = \sum_{x \in M} \sum_{y \in N} (f(x,y) - f_0(x,y))^2/(MN) \tag{4.2}$$

where f and f_0 are some pixel measurements (magnitude, energy, etc.) of the two images to be compared ([Hal81]).

MSE can be normalized with respect to the total energy in one reference image as:

$$NMSE = \frac{\sum_{x \in M} \sum_{y \in N} (f(x,y) - f_0(x,y))^2}{\sum_{x \in M} \sum_{y \in N} (f_0(x,y))^2} \tag{4.3}$$

By using the absolute values rather than the squared ones, we also have the normalized error:

$$NE = \frac{\sum_{x \in M} \sum_{y \in N} |f(x,y) - f_0(x,y)|}{\sum_{x \in M} \sum_{y \in N} |f_0(x,y)|} \tag{4.4}$$

The MSE based metrics can be used in various measurements as long as both the image to be measured and a reference image exist. They are easy to calculate and yield a FOM that can be used to compare different images.

4.1.3 Diffenergy

The *diffenergy* method was first proposed by Smith *et al.* ([MSNC93], [Yan93]) to measure the success of modeling in MRI reconstruction. It was used on the frequency domain data files. Two data files were compared on a point by point basis. Diffenergy is a FOM indicating the total difference between the two data files. For algorithm evaluation, we can compare the corrected data file with the reference (uncorrupted) data file. A smaller diffenergy indicates closer resemblance of the corrected data to the reference data.

Diffenergy was designed to overcome problems image domain measurements may

have when comparing reconstruction algorithms. Image domain measurements calculate the differences in reconstructed images. Usually the reference or standard image is the DFT reconstructed image from a full MR data set. However, the "full" data set is windowed during data acquisition, and the standard image reconstructed from it may have artifacts. Thus when a modeled image is compared to the standard image the results may be inaccurate, as the modeled image may be better than the standard one. Diffenergy compares the data in the frequency domain to avoid this problem.

Yang ([Yan93]) defined a *global normalized diffenergy* (GDF) when comparing modeling reconstruction algorithms:

$$GDF = \frac{\sum_{x \in M} \sum_{y \in N} (f_{model}[x][y] - f_{ref}[x][y])^2}{\sum_{x \in M} \sum_{y \in N} (f_{trunc}[x][y] - f_{ref}[x][y])^2} \tag{4.5}$$

where $f_{model}[x][y]$ and $f_{trunc}[x][y]$ are the real or imaginary parts of a point in the modeled and truncated data files, respectively, and $f_{ref}[x][y]$ represents reference file. A smaller GDF indicates better recovery of high frequency components in image reconstruction from short data sets. However, a large GDF is obtained when a large localized error exists, even though most of the data appeared be well modeled. Yang further defined a *localized diffenergy* (LDF) which localized the error:

$$LDF = \sum_{m \in M} \frac{\sum_{y \in N} (f_{model}[x][y] - f_{ref}[x][y])^2}{\sum_{y \in N} (f_{trunc}[x][y] - f_{ref}[x][y])^2} / M \tag{4.6}$$

where M is the minimum number of common rows in all data files. If the reconstructed data has a large GDF and a small LDF, it is likely that a localized area of the data was not successfully reconstructed.

4.2 Receiver Operator Characteristics (ROC)

The ROC methodology is based on signal detection theory (SDT). An ROC analysis attempts to answer the question, "how effective is a particular imaging procedure?" The results of an ROC analysis provide an index of accuracy of the imaging system and an estimate of the probabilities of observer decisions. ROC analysis incorporates the human factor in the imaging procedure. It is more meaningful than pure physical measures since the ultimate goal in imaging is to assist in human diagnosis.

In this section, we first lay out the basics of signal detection theory. ROC analysis is then introduced and its limitations discussed.

4.2.1 Signal Detection Theory

Derived from war-time development of radar systems, signal detection theory uses mathematical and statistical approaches to solve some problems in psychophysics. One of the most popular approaches models an ideal-observer. In this scenario, a specific task is defined – often to detect a known, low-contrast signal in a noisy environment. Data from different imaging systems is collected and analyzed to determine how well an ideal observer with knowledge of the signal parameters can perform the stated task. The performance of the observer can be summarized by ROC curves or some metrics, e.g., the area under the ROC curve. The metrics are taken as a figure of merit (FOM) and the highest FOM corresponds to the best performing imaging system.

To further illustrate SDT with the above ideal-observer example, we present some mathematical terms. In a detection problem with only two events and two responses,

there are four event-response pairs. Let T and F stand for true and false responses. Let s stand for conditions in which a signal exists. Let n stand for conditions in which a signal does not exist. We have:

Term	Symbol	Description
True Positive	$T\|s$	A hit – the signal existed, and was detected
False Negative	$F\|s$	A miss – the signal existed, but was not detected
False Positive	$T\|n$	A false alarm – the signal did **not** exist, but was detected
True Negative	$F\|n$	A correct rejection – the signal was absent and was not detected

Let $P()$ denote the probability of an event. In statistical terms, if a signal exists, we have:

$$P(T|s) + P(F|s) = 1 \tag{4.7}$$

If a signal does not exist, we have:

$$P(T|n) + P(F|n) = 1 \tag{4.8}$$

We can calculate $P(T|s)$ from $P(F|s)$ since their sum is 1, and vice versa. The same holds for $P(T|n)$ and $P(F|n)$. Therefore, to completely characterize a system, we need to know only one of the following two probability pairs:

1. True Positive Fraction (TPF) $P(T|s)$ and False Positive Fraction (FPF) $P(T|n)$ (hit and false alarm)

2. False Negative Fraction (FNF) $P(F|s)$ and True Negative Fraction (TNF) $P(F|n)$ (miss and correct rejection)

In clinical applications, the first pair is typically used. TPF represents the *sensitivity* of the system [Met86]: the fraction of patients who have the disease in question and who are correctly diagnosed as 'positive'. FPF represents the *specificity*: the fraction of patients without the disease who are incorrectly diagnosed as 'positive'. The sensitivity and specificity pair meaningfully describes diagnostic performance. The pair has been used in clinical ROC analysis ([Ega75], [Met86]).

4.2.2 ROC analysis

Metz ([Met86]) listed some requirements for a meaningful ROC analysis:

- **Diagnostic truth:** the diagnostic positive and negative states have to be properly identified.
- **Sampling issues:** a proper sample of patients and observers should be used for the analysis to be meaningful.
- **Diagnostic alternatives:** a traditional ROC analysis deals with cases where only two diagnostic states, positive and negative, exist.

These requirements should be satisfied before we proceed with the analysis.

A key result in ROC analysis is an ROC curve. One typical ROC curve was shown in Figure 4.1. Details on generating an ROC curve will be presented in the following paragraphs. The ROC curve indicates clinical performance by plotting TPF (sensitivity) against FPF (specificity). Each TPF and FPF pair represents the hit ratio at a given false alarm rate. Both TPF and FPF range from 0 to 1. A higher ROC curve or larger area under the curve would indicate better discrimination

capacity. The area under the curve may also be used as an FOM to summarize the result of an ROC analysis.

Figure 4.2 illustrates the process required to generate an ROC curve. The horizontal axis represents a detectability index that radiologists use for diagnosis. This index can be contrast, size or location of certain features in the image, etc. A threshold of the detectability index is the criterion above (or below) which positive diagnosis is issued. The two curves shown are the probability density functions of the detectability index for negative and positive patients, also called frequency curves. For ideal discrimination of the patients, there should be no overlapping of the two curves. However, in real life, a positive diagnosis based on any detectability index can arise from either a positive patient or a negative patient. A better algorithm or technique should result in a larger hit ratio (TPF) at any given false alarm rate (FPF).

To generate an ROC curve we need multiple values of TPF and FPF. The threshold detectability index is varied and multiple TPF and FPF are calculated one pair at a time. At each threshold setting, TPF is calculated as the fractional area under the positive patient frequency curve above (or below) the threshold. FPF is the fractional area under the negative patient frequency curve above (or below) the threshold. By varying the threshold, we can gather multiple pairs of TPF and FPF and plot an ROC curve.

In practice, ROC data gathering requires a large database of medical cases and human observers. Two different experimental approaches can be used:

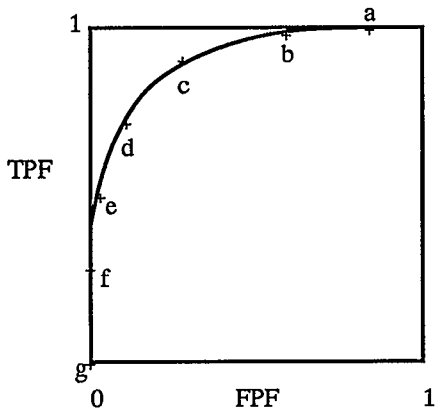


Figure 4.1: A typical ROC curve

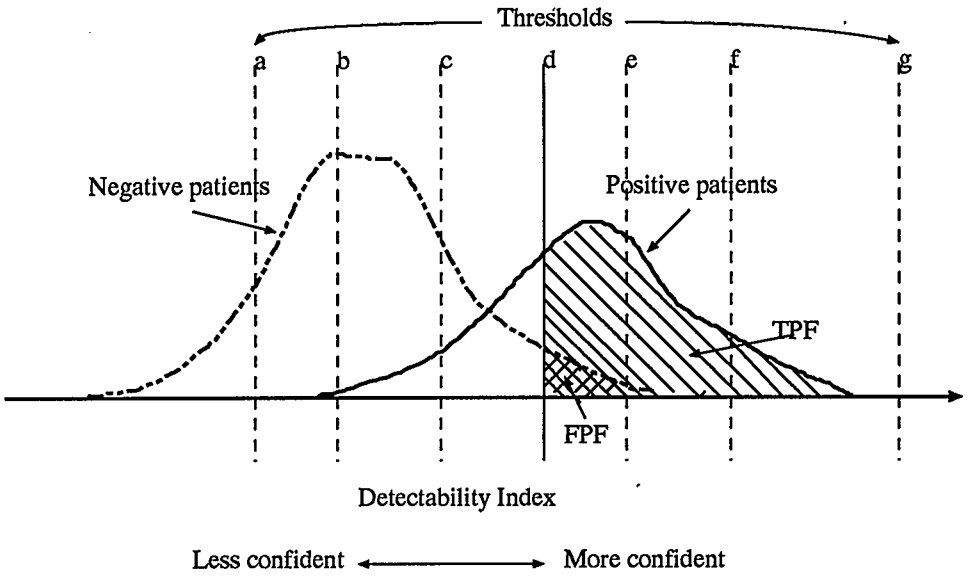


Figure 4.2: Sample frequency curves and thresholding methods used to calculate TPF and FPF pairs for the ROC curve. TPF and FPF are shown for threshold d.

1. **yes/no method.** The observer views a series of images sequentially at a certain threshold and is required to give a yes/no response for each image. One pair of TPF and FPF is obtained after a sequence. The process is repeated using different thresholds, obtaining more data points for the ROC curve.

2. **rating method.** The observer is required to select one of the several confidence categories (definite, possible, etc.). One sequence will yield multiple pairs of TPF and FPF. Apparently this is the more efficient method ([Met86]).

After a sufficient number of data points are obtained, a curve fitting algorithm can be employed to generate a smooth ROC curve.

ROC analysis is widely used in evaluating medical imaging methods. Most papers in the literature are concerned with radiologic imaging, such as radiography, mammography, CT, ultrasonography and SPECT.

4.2.3 Limitations

The major limitation of signal detection theory based ROC analysis is that it oversimplifies the complex perceptual tasks that humans perform. The task in ROC analysis is usually detection, with known signals and simple noise models. In real life, the tasks could be estimation or classification, with completely unknown signals and noise. With SDT we are normally dealing with the simplest task, signal and noise models. The simplicity of the test limits the applicable range of the performance evaluation. Nonetheless, the ROC analysis is statistically representative of clinical decision making. However, it requires a large database of patients and observers, which may lead to high costs and long analysis time.

4.3 Phantom Study

To increase the efficiency of ROC analysis, phantoms can be used instead of real patient images ([Wag86]). This approach applies when an observer is given an image of a simple test phantom and must discriminate where certain shapes appear in the image. Two steps should be carefully monitored in the design of phantom studies:

1. **Phantom generation.** Realistic phantoms, based on the specific imaging method, are needed instead of simplistic mathematical images. Realistic phantoms are especially important in MRI, since the data is collected in k-space. Proper phantom generation in MRI will be discussed in detail in Chapter 6.

2. **Statistical distribution.** Phantom studies have to fulfill certain requirements to be statistically sound. Designers should take into consideration the sample size, the distribution of normal and abnormal samples, the distribution of lesion intensity, etc..

Phantom studies can be carried out by either human observers or computer observers. Phantom study by human observer is usually used to generate ROC curves. Computer observers are relatively new in image evaluation and some of the computer observer based techniques will be discussed in the next sections.

4.4 Computer Observers

To reduce the time and money required to carry out ROC analysis, computer observers were introduced. Ideally, we would like to apply some artificial intelligence

(AI) so that the computer can emulate the performance of a human observer. However, the complexity of AI makes this type of application prohibitively complicated at the present time. A more realistic approach is to use a certain metric which has already been proven to have medical relevance. Two types of approaches can be found in the literature:

- Development of metrics that have a high correlation to a comparative human observer ROC analysis.
- Designing a new metric or figure of merit (FOM), based on a specific task, to evaluate the performance of different algorithms. Repeat the experiment on a statistically significant set of phantoms and evaluate the algorithms based on the resulting FOMs.

The following sections present a few examples from the literature that are relevant to this project.

4.4.1 A perceptual image quality measure

Hall ([Hal81]) proposed a perceptual image quality measure, perceptual MSE (PMSE), which is an MSE computed in a transformed space. The preprocessor used to obtain the data for PMSE is related to the human vision system (HVS). This measure has been reported to have close correlation to a comparative ROC analysis, indicating that it may be suitable as a metric for computer observers. We implemented this measure and the details are shown in Chapter 5.

4.4.2 Evaluation method based on task performance

Hanson ([Han90]) developed a methodology to evaluate image reconstruction algorithms based on task performance. His method attempted to judge an algorithm by how well one could perform a defined visual task using reconstructed images. The method simulated the entire imaging process. Defined representative scenes and corresponding data sets were generated randomly. A specific task was then performed using the reconstructed images. Finally, the accuracy of the task performance was evaluated. The basic steps of this method are:

1. Define the class of scenes to be imaged.
2. Define the geometry of the measurements, including noise.
3. Define the task to be performed on the images.
4. Define the method of task performance, including the computer observer or human observer.
5. Create a representative scene and the corresponding data by a Monte Carlo simulation (pseudo-random selection of the uncertain and variable parameters).
6. Reconstruct the scene with the algorithm being tested.
7. Perform the specified task, using the reconstructed images.
8. Repeat the above three steps a sufficient number of times to obtain results.
9. Evaluate the task performance, using an ROC curve or other metrics.

Hanson demonstrated the methodology in an evaluation of algebraic reconstruction technique (ART). The task was simplified to detecting a possible disk at known locations of the background. The mean magnitude of the reconstruction over the area of the disk was used as a decision variable. Frequency graphs of the decision variable evaluated at positions where a low-contrast disk was known to exist and where none existed were plotted. An ROC curve was generated from the frequency graphs by setting different decision thresholds and calculating the TPFs and FPFs. The area under the ROC curve was an effective index of performance ([Met86]).

Other computer observer techniques include those proposed in [FBSM87], [HO91], [HY91], [LLG92], [FHN+94].

4.5 Summary

We presented a literature survey of the existing quantitative evaluation methods in medical imaging. Some standard metrics were found to be used to generate quantitative results when evaluating different algorithms and techniques. ROC analyses yield more clinically relevant results. Computer observers have also been used to reduce evaluation costs.

Chapter 5

Performance Evaluation Metrics

We now present a set of metrics we developed to quantitatively evaluate the performance of different MRI motion suppression and reconstruction algorithms. In most cases, a *reference* (image or data) is needed to calculated the metrics. In the case of motion suppression, our reference is the original phantom data/image before motion simulation. In the case of MR image reconstruction, our reference is a full (not truncated) MR data set or image reconstructed from the full data set.

5.1 Contrast

For MRI images, the magnitude image is normally used by radiologists for diagnosis. Contrast in an MRI magnitude image can be defined as:

$$Contrast = \frac{S_o - S_b}{S_o + S_b} \tag{5.1}$$

where S_o is the mean magnitude of an object and S_b is that of the background. We defined the background as the area contained in a box surrounding an object excluding the object. (See detailed contrast calculation in Figure 6.2 in chapter 6.)

In our project, we measured the contrast of simulated lesions in the phantom before and after applying our post-processing algorithms. A larger contrast value indicates easier recognition of features in the images.

5.2 Kernel Error

Kernel error is designed for post-processing algorithms using a mathematical model. Researchers often use a mathematical model to simulate distortions in the images and then apply the model to correct for simulated distortions. In our case, we used a motion kernel as our model. All our algorithms extracted the motion kernel and used it to correct the motion. The success of the correction depended on the accuracy with which the extracted kernel approximated the kernel we used to simulate motion. Ideally the extracted motion kernel should be identical to the motion kernel used to simulate motion. Accordingly, a measure of the difference between the extracted and original kernels could be a measurement of algorithm performance. A smaller kernel error value indicates better performance.

5.2.1 Kernel mean squared error – KMSE

Since the kernel inverse was used to multiply the motion corrupted data in the correction, the mean squared error of the kernel inverse (KMSE) is defined as :

$$KMSE = \frac{1}{M} \sum_{i=0}^{M-1} (1/ker_{extr}[i] - 1/ker_{orig}[i])^2 \tag{5.2}$$

where M is the number of points in the kernel, and $1/ker[i]$ is the value of the kernel inverse (1/kernel) at the ith point. The original kernel used to simulate motion (ker_{orig}) is used as the reference here. ker_{extr} is the kernel extracted using different algorithms.

For a particular reference kernel, a smaller KMSE indicates better performance of the algorithm used to extract the kernel. In order to compare results from different

kernels, we need to normalize KMSE by the total energy of the original kernel inverse. We denote this normalized KMSE as KNMSE:

$$KNMSE = \frac{\sum_{i=0}^{M-1}(1/ker_{extr}[i] - 1/ker_{orig}[i])^2}{\sum_{i=0}^{M-1}(1/ker_{orig}[i])^2} \tag{5.3}$$

KNMSE yields the kernel error as a fraction of the total reference kernel energy. KNMSE excludes the effect of reference kernel energy on KMSE. KNMSE should be used when comparing the performance of algorithms tested with different reference kernels.

5.2.2 Weighted mean squared error – KWMSE

Due to the nature of the MRI data, the amplitude of the low frequency components in k-space is always greater than that of the higher frequency components. We found that a kernel error which appeared in the low frequency region had greater effect than one which appeared in the higher frequency region. In order to reflect this characteristic of MRI data, we defined a weighted MSE. First, a set of weights were calculated from the M by N motion corrupted data file:

$$w(i)|_{i=0}^{M-1} = \frac{\sum_{j=0}^{N-1}(s_{mo}(i,j))^2}{\sum_{k=0}^{M-1}\sum_{j=0}^{N-1}(s_{mo}(k,j))^2} \tag{5.4}$$

where $s_{mo}[i][j]$ is the magnitude of the point (i,j) in the data file. The numerator is a sum of the energy in a row of data. The denominator is the total energy of the data. Each $w[i]$ represents the fractional energy of the ith row in the data file. The weights w form an M by 1 vector.

A weighted KMSE (KWMSE) can then be calculated:

$$KWMSE = \sum_{i=0}^{M-1} (w[i] * (1/ker_{extr}[i] - 1/ker_{orig}[i])^2) \tag{5.5}$$

This metric incorporates the effect of energy distribution in the MRI data set. The distortion in an energy concentrated area is emphasized while distortion in a low energy area is deemphasized.

5.3 Diffimage

The *Diffimage* (DI) measures the difference between the corrected image and a reference image. The reference image is assumed to have no artifacts. A *Diffimage* metric is calculated by the difference between the image produced by first distorting and then correcting the reference image, and the original reference image. The mean energy of the difference images were then obtained. A smaller value indicates closer resemblance of the corrected image to the reference image. When the reference image contains distortion and artifacts itself, a large DI could mean that the corrected image has less artifacts than the reference image. Under such circumstances the Diffimage measure is not reliable.

For an M by N image, *Diffimage* is defined as:

$$DI = \frac{\sum_{x \in M} \sum_{y \in N} (I_{corr}[x][y] - I_{ref}[x][y])^2}{M * N} \tag{5.6}$$

where $I_{corr}[x][y]$ is the magnitude of the point (x, y) in the distorted and corrected image, and $I_{ref}[x][y]$ is a point in the reference image.

5.4 Diffdata

Diffdata (DD) is based on the *diffenergy* measurement, which was first proposed by Smith *et al.* ([MSNC93], [Yan93]) to measure the success of modeling in MRI reconstruction. It is used on the frequency domain data files of the corrected and reference MRI images:

$$DD = \frac{\sum_{x \in M} \sum_{y \in N} (s_{corr}(x,y) - s_{ref}(x,y))^2}{M * N} \tag{5.7}$$

where $s_{corr}(x,y)$ and $s_{ref}(x,y)$ are the real or imaginary components of the point (x,y) in the corrected and reference data files, respectively.

Diffdata can be normalized by the mean diffenergy between the corrupted and reference data (NDD). After normalization, 100% will indicate no improvement. A smaller value indicates better performance.

$$NDD = \frac{\sum_{x \in M} \sum_{y \in N} (s_{corr}(x,y) - s_{ref}(x,y))^2}{\sum_{x \in M} \sum_{y \in N} (s_{dstr}(x,y) - s_{ref}(x,y))^2} \times 100\% \tag{5.8}$$

where $s_{corr}(x,y)$, $s_{ref}(x,y)$ and $s_{dstr}(x,y)$ are the magnitudes of the point (x,y) in the corrected, reference and distorted data files, respectively.

Usually, *Diffdata* and *Diffimage* are related for the same images, since the frequency domain data is the Fourier transform of the image. However, when the reference image contains artifacts, the *Diffimage* measure may be inaccurate. On the other hand, if an image has "super-resolution," the Fourier transformed frequency domain data set may contain data aliasing. The *Diffdata* measurement on aliased frequency domain data sets may not be reliable. Under these circumstances the two measures may yield different results.

5.5 A perceptual image quality measure

This measure is based on the perceptual mean squared error (PMSE) measure proposed by Hall ([HH77] [Hal80], [Hal81]). It is simply an MSE computed on images processed by a filter. The filter used to obtain the data for PMSE is related to the human vision system (HVS). A conceptual block diagram of the PMSE measure is shown in Figure 5.1.

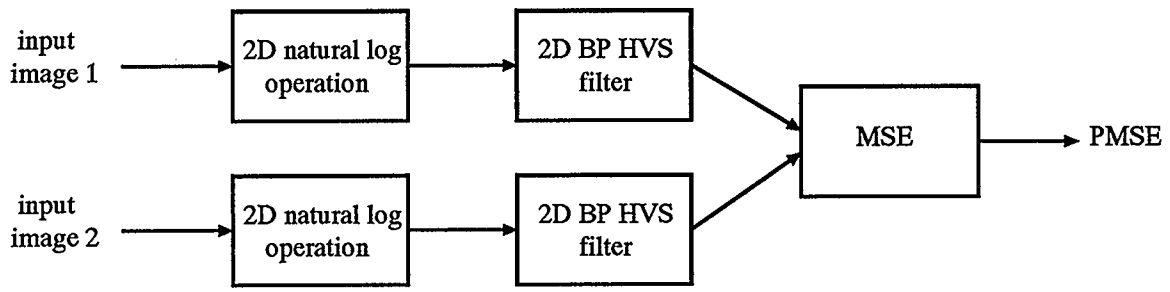


Figure 5.1: Block diagram of the PMSE measurement

Mathematically,

$$PMSE = \frac{\sum_{x \in M} \sum_{y \in N} (z(x,y) - z_{ref}(x,y))^2}{\sum_{x \in M} \sum_{y \in N} (z_{ref}(x,y))^2} \tag{5.9}$$

where $z(x,y)$ and $z_{ref}(x,y)$ are intermediate variables calculated according to:

$$z(x,y) = \ln(f(x,y)) \otimes h_{bp}(x,y)$$

and

$$z_{ref}(x,y) = \ln(f_{ref}(x,y)) \otimes h_{bp}(x,y)$$

where $f(x,y)$ and $f_{ref}(x,y)$ are magnitudes of the images (corrected and reference) at points (x,y), ln denotes the natural log operation, $\otimes$ denotes a convolution operation, and $h_{bp}(x,y)$ is a point spread function of the HVS.

This metric has been applied on non-medical images ([Hal81]). Here we introduce

it to MRI images. The procedures for obtaining the PMSE are:

- Perform a natural log operation on a processed magnitude image and a reference magnitude image.
- Fourier transform the images into k-space and multiply the k-space data with the frequency response of a 2D HVS filter.
- Inverse Fourier transform the k-space filtered data to generate filtered images.
- Calculate the PMSE using the filtered images.

5.5.1 The HVS model

The main difficulty in implementing the PMSE measure is to define a 2D HVS filter. We searched the literature for a HVS model which has been proven to match experimental measurements, and is both clearly defined and documented.

The model we found and used in our project was based on that developed by Barten [Bar92]. It was a comprehensive model accounting for effects of various parameters on the contrast sensitivity of the eye, including optical modulation transfer function (MTF) of the eye, internal noise (photon noise and neural noise) and external noise. The model had been shown to fit various physical measurements if its parameters are properly chosen. For a comprehensive explanation of the model, please refer to [Bar92]. The following provides a brief outline of the model:

The overall contrast sensitivity function of the HVS system is:

$$\frac{1}{M_t} = \frac{1}{k}\sqrt{\frac{T}{2}}[\frac{1}{\eta PI} + \frac{\Phi_0}{(1-F(u))^2}\frac{1}{X_0^2} + \frac{1}{X_e^2} + (\frac{u}{N_e})^2]^{-0.5} M_{opt}(u) \qquad (5.10)$$

Some of the variables in equation 5.10 need to be further calculated:

$$I = \frac{\pi d^2 L}{4} \tag{5.11}$$

$$d = 4.6 - 2.8 \tanh(0.4 \log \frac{L}{1.6}) \tag{5.12}$$

$$F(u) = 1 - [1 - \exp(\frac{-u^2}{u_0^2})]^{0.5} \tag{5.13}$$

$$M_{opt} = \exp(-\pi^2 \sigma^2 u^2) \tag{5.14}$$

$$\sigma = \sqrt{\sigma_0^2 + (C_{sph} d^3)^2} \tag{5.15}$$

The variables are defined in Table 5.1.

In our project, we tried to describe a typical viewing environment for MRI images. The values we used in our project are shown in Table 5.1. Most constants were fixed by Barten([Bar92]) using average values in typical cases. Only the following three parameters need to be adjusted according to different viewing conditions.

- **angular spatial frequency**

  All the recorded frequency responses of HVS use cycles/degree as the unit for spatial frequency. Barten's model uses cycles/degree as well. Cycles/degree is the reciprocal of the optical viewing angle. In the literature, cycles/degree is only clearly defined in a sine-wave grating test where each test uses bar images at one specific spatial frequency. There is no clear description of how cycles/degree can be calculated or measured for the Fourier transform of an image. In our project, we made two assumptions before calculation so that we can use Barten's HVS model properly. The first one was that the eye always

Table 5.1: Description of variables

Var	Description	Value	Units
k	constant	3.0	
T	temporal dimension of the image	0.1	s
η	total quantum efficiency	2	%
P	constant	350	
I	illuminance of the eye		cd/m^2
d	diameter of the eye pupil		
L	illuminance of object		
Φ_0	noise density at high spatial frequency	3e-8	$sec * deg^2$
$F(u)$	MTF of the low pass filter		
u	angular spatial frequency		
u_0	spatial frequency below which the attenuation of the contrast sensitivity takes place	8	$cycles/deg$
X_0	angular size of the picture		deg
X_e	maximum angular size of the picture	12	deg
N_e	maximum number of cycles over which the eye can integrate the information	15	$cycles$
M_{opt}	optical MTF of the eye		
σ	radial standard deviation of the optical point-spread function		
σ_0	value of σ at small pupil sizes	0.75	$arcmin$
C_{sph}	constant describing the spherical aberration effect	0.006	$arcmin/mm^3$

points toward the center of the image. The second was that the highest spatial frequency term in k-space after DFT of the image equals to the reciprocal of the separation angle between two adjacent points in the digital image (angular resolution). The second assumption was based on sampling theory. With the first assumption, we were able to determine the angular size of the image and angles between two adjacent points. With the second assumption, we labeled our frequency domain data with the unit cycles/degree.

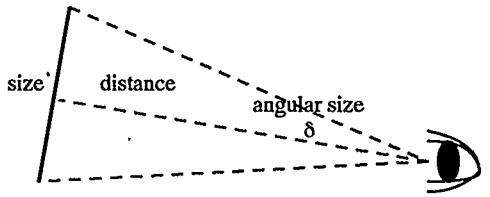


Figure 5.2: A sample viewing condition

Figure 5.2 shows an example calculation. Suppose we were viewing an s cm image from d cm away. With the eye pointing at the center of the image, the angular size of the image will be

$$\delta(indegree) = \tan^{-1}(\frac{s/2}{d}) \tag{5.16}$$

If the image has M by M pixels, then the highest frequency term of the image should be:

$$f_{max}(incycles/degree) = 1/(\tan^{-1}(\frac{s/M}{d})) \tag{5.17}$$

- **angular size of the observed image**

  As described in the last item, it is calculated as:

$$angularSize = \tan^{-1}(\frac{size/2}{distance}) \tag{5.18}$$

Figure 5.3 showed frequency responses with different distances between the observer and the image, assuming each image was exactly the same size.

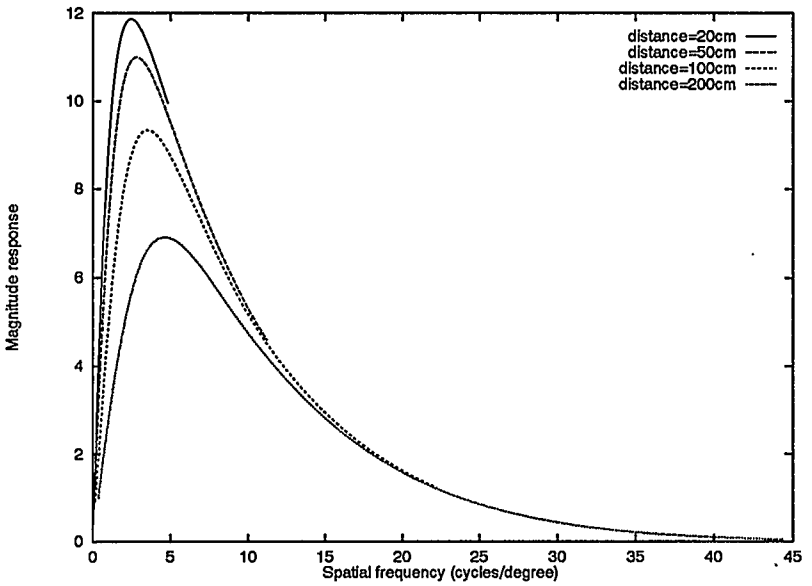


Figure 5.3: Spatial frequency response of Barten's HVS model at different viewing distance

- **illuminance**

  Under normal viewing conditions, illuminance can range from 0.00001 to 1000 cd/m^2. Figure 5.4 shows the same curve with only illuminance varying.

From Figures 5.3 and 5.4, we can see that the setting of parameters affect the shape of the HVS bandpass filter. We would expect the PMSE results to be different under different viewing conditions as well. In this project, we are not interested in investigating what effect different viewing conditions have on the HVS model. We will assume a typical viewing condition and use a fixed HVS filter in all our evaluations.

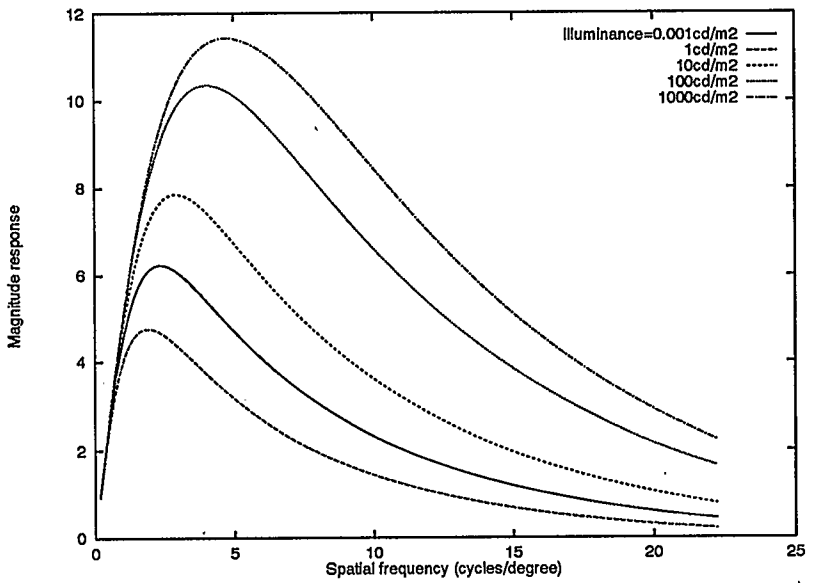


Figure 5.4: Spatial frequency response of HVS with varying illuminance

With an average angular size of 6 degrees and illuminance of 34 cd/m^2 (the median of what was used in all the HVS experimental measures we found in the literature) the frequency response of the HVS system in this project is shown in Figure 5.5.

The PMSE metric is essentially a metric of the image. For MRI images, we need to first reconstruct an image from the acquired data using appropriate reconstruction methods. After we obtain a reconstructed image with the best possible quality, we can calculate the PMSE using the procedures listed in Section 5.5. Figure 5.6 shows one line of MRI data with the corresponding HVS filter. Since the HVS filter is not defined at DC (where spatial frequency is 0 cycles/degree), we do not filter the DC value. Since a 2D Fourier transform on the image generates symmetrical positive and negative frequency components, both components are filtered by the same HVS filter.

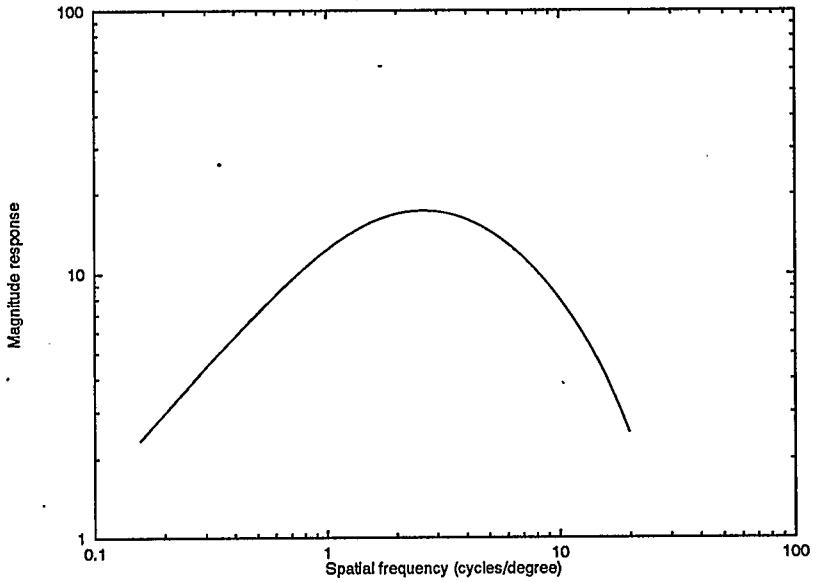


Figure 5.5: Spatial frequency response of HVS in this project

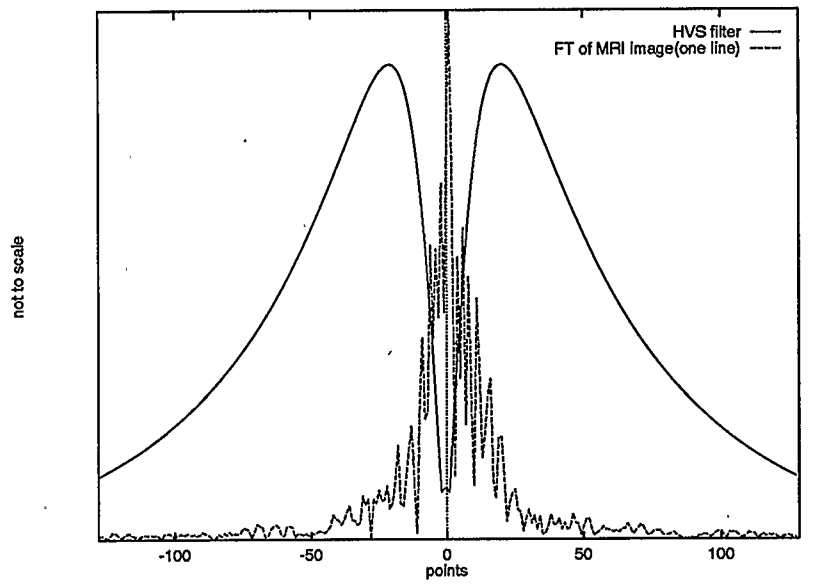


Figure 5.6: One line of MRI data and the HVS filter which will be used on this line of data

To extend the 1D HVS filter to 2D, we assume that the frequency response of the eye is invariant in all directions ([Hal80]). Such a 2D HVS filter is shown in Figure 5.7.

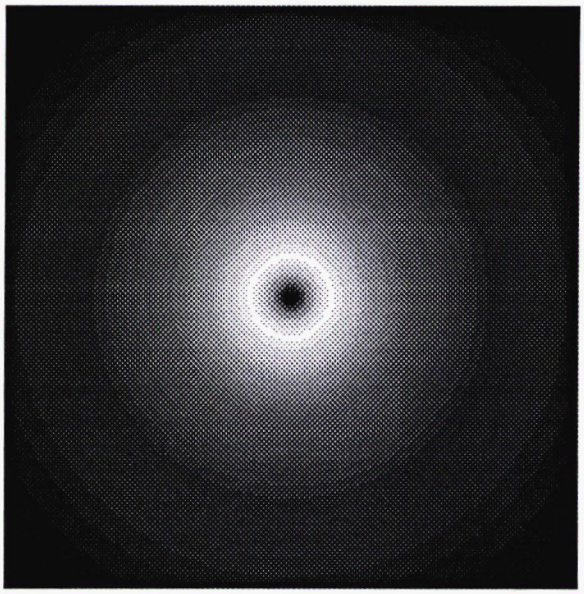


Figure 5.7: The 2D HVS filter used in this project

5.6 Summary

In this chapter, we listed all the metric measurements we assembled for MR image quality evaluation. The metric measurements included: contrast, kernel error, Diffimage, Diffdata and a perceptual image quality measure. Kernel error is suited for evaluating motion suppression algorithms when a motion kernel is extracted. Diffimage and Diffdata compare the differences between a processed data set and a standard data set in the image and frequency domain, respectively. The perceptual image quality measure incorporates the effect of the human vision system in

algorithm evaluation. These measurements will be used in Chapter 7 for algorithm evaluation.

Chapter 6

A Novel Computer Observer ROC Analysis

Our proposed method attempts to create a computer observer that would yield receiver operator characteristics (ROC) analysis results using mathematical phantoms. We took the following steps to set up the computer observer.

- Define the image scene to be used in the computer ROC analysis.
- Define the task to be performed by the computer observer.
- Define the variations used to obtain statistically significant results.
- Define the approach to be employed by the computer observer to generate ROC curves

We will discuss the key steps for this novel computer observer ROC analysis in this chapter.

6.1 Phantom Generation

Our computer ROC analysis is designed to be a general paradigm for computer automated quantitative evaluation of algorithms and techniques. The evaluation process starts from generating phantoms using the computer. We would like the phantoms to be realistic and representative of typical MR images.

Throughout the literature, phantoms are mathematically generated in the image domain ([HY92b], [HYR91a], [TOS+94], [RTS92]). However, for MRI applications

this may not be appropriate. To understand this, we use a 1D example to illustrate k-space data aliasing caused by an image domain defined phantom.

In Figure 6.1, we assume that (a) is an image domain generated 1D phantom and (b) is its corresponding k-space data set. If (a) has mathematically defined sharp edges, its corresponding data set (b) has an infinite number of frequency components.

In digital image processing, images are represented as arrays of discrete pixels. Therefore the continuous function shown in (a) is discretized, or sampled, to generate the digital phantom. Sampling functions in the image domain and k-space are illustrated in Figure 6.1 (c) and (d). δx is the sampling rate. The higher the sampling rate (smaller δx in (c)), the larger the separation between the delta functions in (d). (e) and (f) show the sampled signals. In the image domain, the sampled signal is the product of (a) and (c). In k-space, the sampled signal results from convolving (b) and (d). After convolution, the k-space data in (f) exhibit overlapping parts. This overlapping is called *aliasing*. In this case, the aliasing is on data in k-space and we call it data aliasing.

If there are infinitely many frequency components in k-space, we cannot sample the image without k-space (data) aliasing. However, increasing the sampling rate in the image domain may lessen data aliasing, since the frequency data normally tends to zero as frequency increases. A higher sampling rate means less overlapping. A grossly over-sampled image domain generated phantom may exhibit very little data aliasing.

The effect of data aliasing can be seen in the comparison of data generated in the frequency domain and image domain (Figure 6.2). The aliased data set is not realistic and may skew the ranking of image processing algorithms.

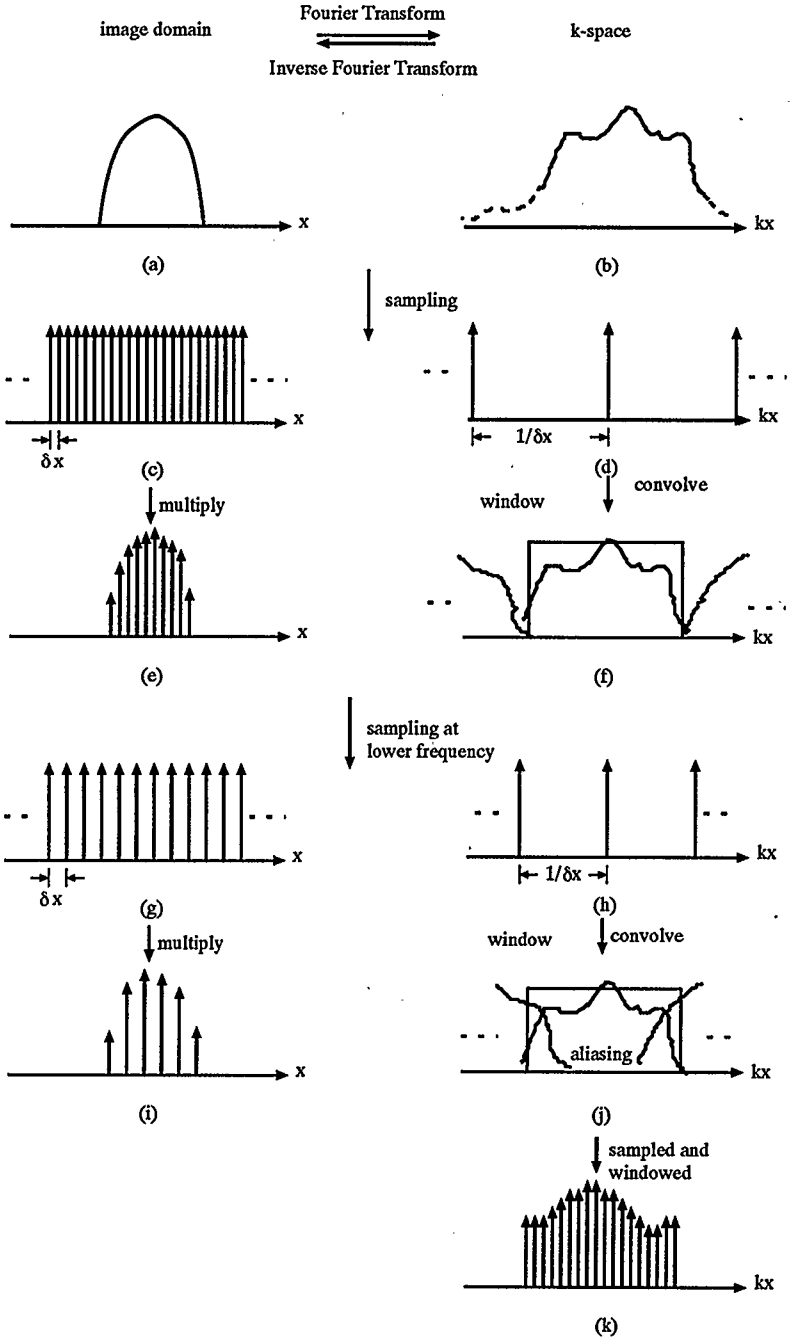
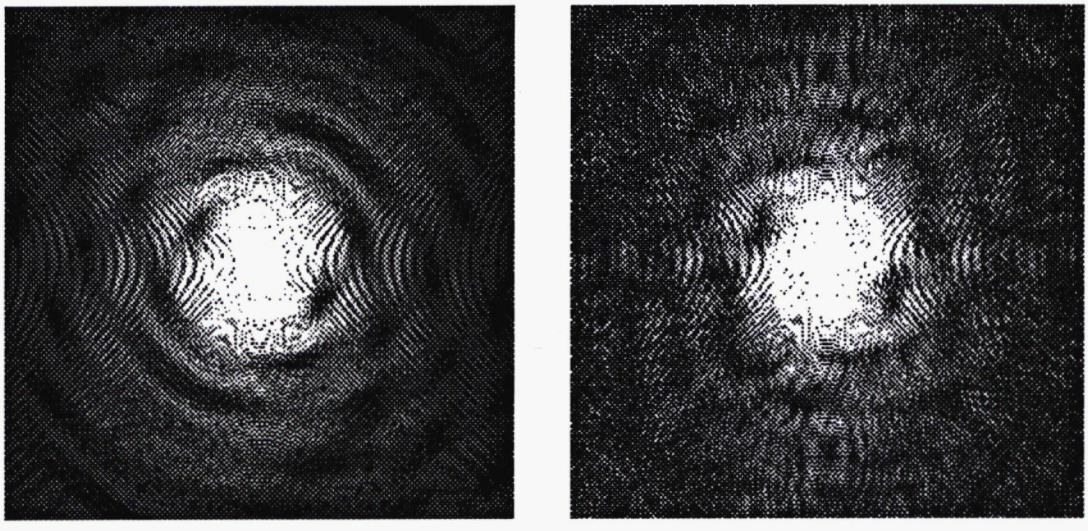


Figure 6.1: Data aliasing in image domain generated phantom



Frequency domain generated phantom

Image domain generated phantom with aliasing

Figure 6.2: Data aliasing in image domain generated phantom

In our project, the phantoms were generated in k-space to avoid data aliasing. The phantom program accepts image domain descriptions of the phantom, calculates its FT, and then generates a truncated set of data in k-space. Due to the difficulty of calculating a close-form expression for the FT of an arbitrary shape, our phantom program only generates ellipses that do not partially overlap one another. The resulting phantoms lack realness when compared to clinical MR images. However, the phantoms are still representative of typical medical MR images.

The abdomen phantom used throughout this thesis results in a 256 by 256 pixel image. All the parameters required to create the phantom are shown in Table 6.1. We gave each ellipse in the image a descriptive name to identify them. We assume the image has $(0,0)$ in the center and extends from -128 to 127 along the x- and y-axes. The *intensity* column indicates the intensity of each ellipse. Angle, X-center, Y-center, X-radius and Y-radius describe the geometry of each ellipse.

intensity	X-center	Y-center	X-rad.	Y-rad.	Angle	Desc.
0.319	4.192	-56.596	8.909	5.240	0.000	spine 3
0.319	2.096	-45.067	8.385	5.240	0.000	spine 2
0.319	2.306	-30.394	11.529	7.337	0.000	spine 1
0.129	-1.887	-8.385	9.433	7.337	0.000	aorta
0.262	35.635	-37.731	23.058	10.481	140.000	right kidney
0.262	-37.731	-31.442	23.058	10.481	40.000	left kidney
0.432	-48.212	23.058	49.260	23.058	28.000	liver
0.494	46.115	18.865	41.923	19.913	150.000	stomach
-0.987	-1.677	-3.144	101.663	67.077	0.000	anti-fat
1.004	0.000	0.000	109.000	74.413	0.000	fat

Table 6.1: AbdomenT1 phantom parameters

The reconstructed phantom image is shown in Figure 3.2 in Chapter 3 on page 26. The phantom had no data aliasing.

6.2 Lesion Detection

We defined the task to be performed in this computer observer experiment as detecting low contrast lesions in the abdomen phantom. Based on the advice of a pathologist (Carla Wallace, MRI Center of the Foothills Hospital), we selected 10 locations where lesions were most likely to occur. The locations were scattered among the simulated liver, kidneys and spine. In a real MR image with similar tissue contrast to our phantom, the intensity of a lesion would be lower than that of healthy tissues. We created negative intensity ellipses at the known locations to represent lesions and superimposed them on the phantom image. The parameters for the lesions are shown in Table 6.2. In order to simulate both the positive and negative conditions in an ROC analysis, we used a computer generated random number to

X-center	Y-center	X-rad.	Y-rad.	angle	desc.
-39.827	18.865	9.0	7.0	10.0	lesion 1
-50.308	33.538	6.0	4.0	70.0	lesion 2
-16.769	44.019	5.0	3.0	0.0	lesion 3
-81.750	8.385	4.0	7.0	30.0	lesion 4
-62.885	14.673	6.0	2.0	40.0	lesion 5
31.442	-35.635	5.0	6.0	130.0	lesion 6
46.115	-40.019	4.0	3.0	10.0	lesion 7
-33.538	-37.731	7.0	3.0	22.0	lesion 8
-40.019	-36.683	5.0	4.0	12.0	lesion 9
3.144	-31.442	2.0	3.0	50.0	lesion 10

Table 6.2: Lesion parameters in AbdomenT1 phantom

determine whether or not there should be a lesion at each known location.

The intensity of the lesions was randomly selected so that it could be any number between 0 and -0.0735, not including 0. The lesions were placed so they overlap with organs. The actual intensity of the lesions was calculated as the sum of organ and lesion pixel intensity. A phantom with four random lesions is shown in Figure 6.3.

After applying different reconstruction or correction algorithms to phantoms with lesions, we use the computer observer to perform lesion detection. The detectability index we chose is the contrast at known locations where lesions might occur.

$$Contrast = \frac{M_l - M_b}{M_l + M_b} \tag{6.1}$$

where M_l is the mean intensity of a known lesion location and M_b is that of the background. The contrast was calculated according to the algorithm shown in Figure 6.4. The algorithm starts by creating a square surrounding the known location of a possible lesion. For each pixel in the square, we calculate to see if it's inside

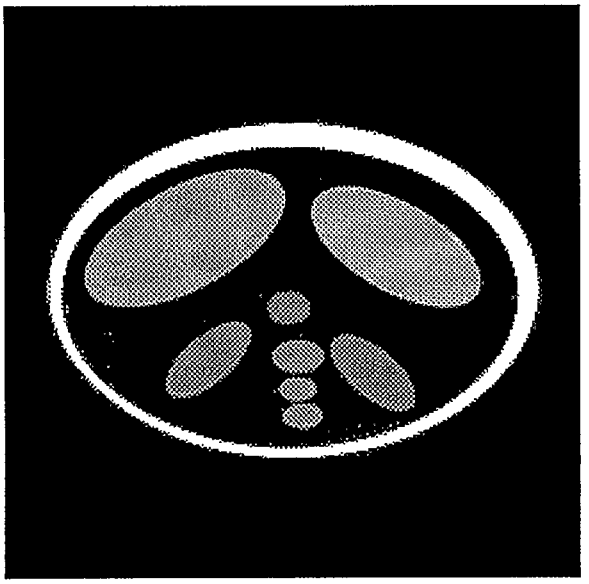


Figure 6.3: A sample abdomen phantom with lesions

the lesion or outside. Based on its position, we add the pixel to the inside group or the outside group. After all the pixels in the box are grouped, we calculate the mean intensity of both inside pixels and outside pixels and thus the contrast. If the contrast is above a threshold, the computer observer detects a lesion.

6.3 Variations

The ROC analysis requires variations when collecting data to plot the frequency and ROC curves. In AM motion suppression, we vary the following factors:

1. **Motion kernel:** The motion kernel can be described as:

$$K(y) = 1 + \sum_{n=-p}^{p} M(n) \cdot cos(2\pi N(n) K_y + \phi_n)] \tag{6.2}$$

The number of terms p, the amplitude of the modulation term M, the frequency of motion N and the phase ϕ can all be varied. In our motion suppression

Inputs: The image file
Outputs: Contrast at all possible lesion locations (known)

Procedure CalcLesionContrast {
    allocate and load image array from file
    for(*each possible lesion location*)
        create a square surrounding the lesion
            $sideOfSquare = 2 + max$(X-rad, Y-rad)
        $totalInsideIntensity = totalOutsideIntensity = 0.0$
        $totalInsidePixels = totalOutsidePixels = 0$
        for(*each pixel in the square*)
            set pixel coordinates (x, y) relative to X-center and Y-center
            rotate x,y about (X-center,Y-center) the same angle as the lesion
            calculate $e = (x_{rotated}/\text{X-rad})^2 + (y_{rotated}/\text{Y-rad})^2$
            if ($e < 1.0$)
                $totalInsideIntensity += intensity(x, y)$
                $totalInsidePixelsincrement$
            if ($e >= 1.0$)
                $totalOutsideIntensity += intensity(x, y)$
                $totalOutsidePixelsincrement$
        next *pixel*
        $meanInsideIntensity = totalInsideIntensity/totalInsidePixels$
        $meanOutsideIntensity = totalOutsideIntensity/totalOutsidePixels$
        $contrast = (meanInsideIntensity - meanOutsideIntensity)/$
                $(meanInsideIntensity + meanOutsideIntensity)$
    next *possible lesion location*
    }

Figure 6.4: The algorithm to calculate contrast at all possible lesion locations

algorithm evaluation, we vary the frequency of motion, N.

2. **Lesion contrast:** The contrast of the simulated lesions are set at different levels.

Data can be collected by varying one or more of the above factors. The resulting detectability index data is used to generate the ROC curve. In image reconstruction, only the lesion contrast was varied.

One of the problems associated with motion kernel variation is synchronous sampling, such as found in references [TOS+94] and [ZST+95]. Synchronous sampling occurs when data at the two ends of the finite data set are continuous. Synchronous sampling is usually due to improper placement of the data window. Since fast Fourier transform (FFT), the most commonly used DFT algorithm, assumes that all data sets are periodic with the finite windowed data having a length of one period, a synchronously sampled data set produces sharper power spectrum peaks that are easier to identify and suppress ([SC95]). However, real MRI data sets are discontinuous at the two ends and the peaks in the power spectra are wider. The simulated motion kernel should contain the discontinuities at the two ends of the data to represent realistic motion. In this project, we made sure that all variations of motion frequency are not synchronously sampled.

6.4 ROC Curve

After we generated phantoms with our variations, applied our algorithms and obtained the lesion detection results from the computer observer, we were ready to plot ROC curves. The ROC curves were produced in two steps:

1. Plotting frequency curves.

   The computer records for each known lesion location whether or not a lesion was generated and the contrast at that location. Contrast was used as the detectability index. The frequency graphs were plotted with frequency (number of occurrences) as the y-axis and detectability index (contrast) as the x-axis. The frequency graph for each algorithm contains two curves: a frequency curve when there was a lesion at the known location and one when there was no lesion. In order to count the number of occurrences at a certain contrast, we digitized contrasts to discrete values at small intervals. The actual contrast recorded during computer observer lesion detection was rounded to the nearest discrete contrast value. The computer algorithm used to plot the frequency curves is shown in Figure 6.5.

   The frequency vs. contrast data files were then used to plot the frequency curves for each algorithm. A sample frequency graph is shown in Figure 6.6. In this project, the lesion area showed lower intensity than the surrounding healthy tissues. As a result, the frequency curve with lesion showed contrast mostly lower than zero. The frequency curve without lesions showed a peak centered around zero contrast, as expected.

2. Plotting the ROC curves.

   To obtain an ROC curve, we set up multiple thresholds in the frequency graphs to calculate multiple pairs of TPF and FPF. For each thresholding contrast, TPF is calculated as the area (total frequency counts) to the left of the threshold and under the frequency curve with lesions. FPF is calculated in the same

Inputs: Computer observer lesion detection results
Outputs: Data file to plot frequency curves

```
Procedure CalcFreq {
    initialize all the counters to be zero
    for( each algorithm )
        for( each possible lesion location )
            read flag (is there a lesion?)
            read contrast
            round contrast to nearest discrete contrast value
            if ( flag == TRUE )
                TrueCount[Alg][contrast] increment
            if ( flag == FALSE )
                FalseCount[Alg][contrast] increment
            next lesion location
        next algorithm
    for( each algorithm )
        write to a file TrueCount and FalseCount vs. contrast
    next algorithm
    }
```

Figure 6.5: The algorithm to plot frequency curves

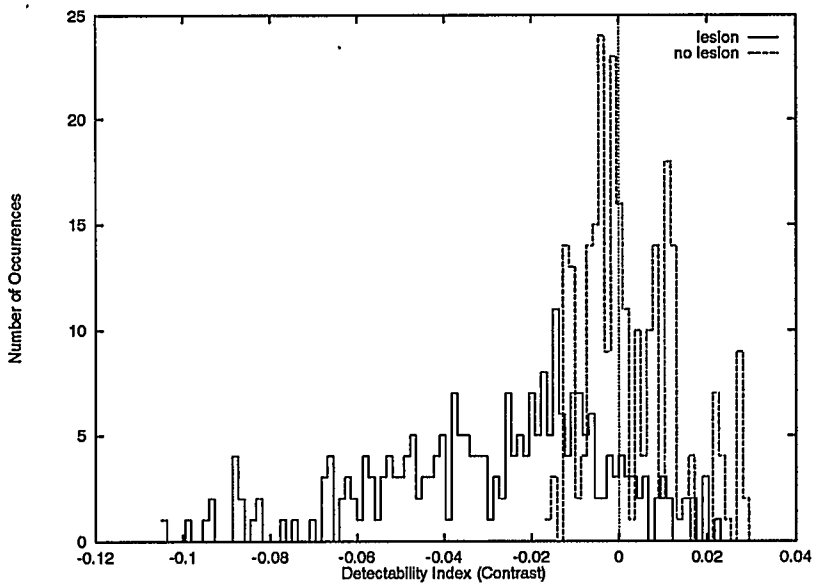


Figure 6.6: A sample computer generated frequency graph

way using the frequency curve without lesions. The TPF and FPF pairs are then used to plot the ROC curve.

6.5 Summary

In this chapter we proposed a novel computer ROC analysis for quantitative evaluation of algorithms and techniques. The steps in setting up such an ROC analysis were discussed. We summarize our computer observer ROC analysis for algorithm evaluation as the following:

1. **Test phantom generation.** A computer generated abdomen phantom is used as the background for lesion detection. Lesions at 10 known locations are created randomly. The intensity of a lesion is also randomly determined to generate different contrasts.

2. **Algorithm application.** The algorithms to be evaluated are applied on the phantom. Final images after algorithms have been applied are obtained.

3. **Contrast calculation.** Contrasts are calculated at all the known lesion locations for each algorithm.

4. **Repetition.** The above steps were repeated and for each know lesion location, we record whether or not a lesion is present and the contrast at that location.

5. **Frequency curve plotting.** The lesion contrast data are used to generate frequency plots for each algorithm.

6. **ROC curve generation.** Different thresholds are used to obtain multiple pairs of TPF and FPF's from the frequency plots. One ROC curves for each algorithm is plotted.

Results were obtained using this methodology with AM motion suppression and ARMA reconstruction algorithms and are presented in Chapter 7.

Chapter 7

Results

In this chapter, we present some of our results and discuss the implications of those results.

In the first section we provide some background information on why and how we performed our quantitative evaluation. Next we present some evaluation results obtained during testing of our computer ROC analysis. Statistical variations are studied and discussed. This section aims to answer some of the questions we had regarding the protocol used for our computer ROC analysis to compare different MR reconstruction algorithms.

The third part contains detailed quantitative evaluations of both the AM motion suppression algorithms and the TERA modeling reconstruction algorithms. All the evaluation methods discussed in Chapters 5 and 6, the metrics and the computer ROC analysis, were applied on both computer generated phantoms and real medical images.

7.1 Background

In this chapter, we use several different evaluation methods to quantitatively compare a number of different MR algorithms. Our purpose is to identify the method that offers the most realistic evaluation of the algorithms. The evaluation approaches we apply are the following:

- a computer observer receiver operator characteristics (ROC) analysis
- Mean-Squared-Error (MSE) based measures, including
  - Diffdata (DD) – the MSE between the reconstructed and standard MR data measured in the frequency domain
  - Diffimage (DI) – the MSE between the reconstructed and standard MR image measured in the image domain
  - Perceptual MSE (PMSE) – the MSE between the reconstructed and standard MR images that are preprocessed by human vision system filters
  - Kernel MSE – the various MSE measures between reconstructed and actual motion kernels, applicable only to motion suppression algorithms

The computer observer ROC analysis is designed to draw clinical rather than numerical conclusions about the quality of MR images. However, this analysis is computationally expensive. On the other hand, the various MSE based measures are quick to yield numerical results with no apparent medical implications. We would also like to investigate whether the simple, fast MSE measures are closely correlated to the slow, clinically valid, ROC method.

7.1.1 Protocol for the computer observer ROC analysis

To obtain reliable results, we establish a protocol to use our new computer ROC analysis for algorithm evaluation. The protocol must be justified in terms of statistical validity and the computer observer's ability to distinguish, in a similar way to a human observer. We evaluate the protocol using the DFT reconstructed MR images

of a series of increasingly truncated MRI data sets. We expect the computer ROC analysis to show the deterioration in image qualities as the truncation increases. Once the computer ROC analysis has been validated, we can have some confidence in using this technique to evaluate AM motion suppression algorithms and TERA reconstruction algorithms.

After some initial tests, we designed a protocol to conduct our computer ROC analysis. The protocol includes the following steps:

1. Generate a series of MR phantom data sets in the frequency domain with simulated lesions. The lesions randomly exist at known locations and have intensities randomly distributed between two thresholds.

2. Apply the MR reconstruction algorithms to be evaluated to the data sets and calculate the contrast at known lesion locations in the images (lesion contrast data point).

3. Collect 2000 lesion contrast data points for each reconstruction algorithm to be evaluated, using common MR data sets.

4. Divide the 2000 lesion contrast data points to form 7 sets of partially overlapping 500 points lesion contrast data.

5. Generate ROC frequency plots for each of the 7 lesion contrast data sets.

6. Plot ROC curves using the frequency plots and calculate the area under the ROC curves as a figure of merit (FOM) for each data set.

7. Calculate the mean and standard deviation of our FOM, the area under the ROC curves, for all the reconstruction algorithms using the 7 data sets.

8. Compare the reconstruction algorithms based on the mean and standard deviation of their FOM.

In the next section, we present some of the initial test results to explain why we adopted the above protocol.

7.2 Initial Tests for the Computer ROC Analysis Protocol

Since ROC analysis is essentially a statistical study on signal detection, statistical characteristics, such as sample size, mean and standard deviations, need to be examined to justify the meaning of the results. Merely presenting the results without providing any statistical perspectives is incomplete and the results can not be convincible. In an attempt to properly set up our computer ROC analysis protocol, we did a series of tests and the following are the results of the tests.

7.2.1 Number of samples

Our proposed computer ROC analysis generates simulated lesions at known locations using a random number generator. The intensity of a lesion could be any value between two thresholds. The analysis uses contrast at the lesion locations as the detectability index for plotting frequency curves and ROC curves.

If an ROC data set contains predominantly simulated lesions with very low intensity, it will be harder for the computer observer to detect the lesions, which results in a lower ROC curve and a smaller area under the ROC curve. To obtain unbiased ROC analysis results, we need our lesion intensities to cover the range of possible lesion intensities reasonably uniformly. This means we need a large number of samples

to establish the uniformity. However, using a very large sample to reduce standard deviation can be very time consuming. What is the optimal sample size needed?

We did a literature survey on signal detection theory and ROC analysis ([Ega75], [Wha71], [GS66], [McN72], [Met86]) and found no analytical answer to this question. While writing up the thesis, we found a paper by Constable *et. al.* on ROC analysis for evaluating functional brain MR imaging and post-processing protocols [CSG95]. Since the ROC method in that paper differs greatly from our method, we could not find useful hints in this paper either. A brief comment on the Constable *et. al.* paper will be included at the end of this chapter.

In this project, we tried to answer this question by using different sample sizes for our ROC analysis and comparing the resulting FOMs and their standard deviation. We used two different reconstruction algorithms: DFT reconstruction on a truncated 256×128 data set, and DFT reconstruction on a severely truncated 256×64 data set.

We collected 4000 lesion contrast data points for each reconstruction using our computer observer ROC method. We then generated ROC curves using different number of lesion contrast data points (50, 100, 200, 300, 400, 500 and 1000). From the 4000 lesion contrast data points we obtained 80 sets of 50 points lesion contrast data, 40 sets of 100 points lesion contrast data, etc.. Our FOM, the area under ROC curves were calculated for each sample size. The mean and standard deviation of these FOM are shown in Table 7.1. We also plot the standard deviation of the FOM for the 256×128 and 256×64 point MR reconstruction in Figure 7.1.

The standard deviation decreases, as expected, with larger sample size for both reconstruction of 256×128 and 256×64 MR data sets. The change in standard

Table 7.1: Mean and standard deviation of area under the ROC curve for different ROC sample size

		128	128	64	64
data size	number of data sets	mean area under ROC	standard deviation	mean area under ROC	standard deviation
50	80	0.917	0.042	0.868	0.045
100	40	0.917	0.032	0.869	0.029
200	20	0.917	0.028	0.869	0.021
300	13	0.918	0.019	0.871	0.019
400	10	0.917	0.017	0.870	0.016
500	8	0.918	0.014	0.870	0.010
1000	4	0.918	0.011	0.870	0.004

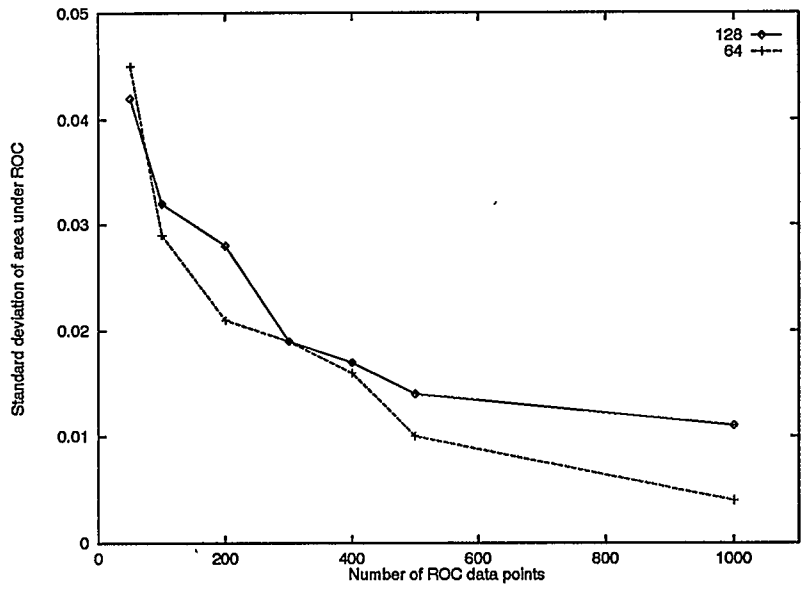


Figure 7.1: Decrease of standard deviation in FOM with increased ROC data size

deviation becomes less significant above 500 lesion contrast data points.

Figure 7.1 shows a tradeoff between ROC sample size and the standard deviation of the ROC analysis. A larger ROC sample size takes longer to collect, but has a slightly smaller standard deviation. Usually, when comparing algorithms, the difference in FOM between algorithms should be greater than the standard deviations in order to say that the algorithms produce different results. Greatly increasing the ROC sample size can decrease the standard deviation, thus perhaps making two algorithms with little difference in FOM distinguishable in theory. However, the difference in algorithms would be too small to detect visually by human observers from a single measurement, making the two algorithms indistinguishable in reality. Therefore, increasing the ROC sample size to a very large number simply to distinguish the minute difference between algorithms makes no clinical sense.

In our project, we choose 500 points lesion contrast data as our sample size. It has a small standard deviation ($< 2\%$) and is much more computationally inexpensive than analyzing 1000 or more lesion contrast data points. We performed another 3 sets of 8X500 ROC analysis for 256×128 DFT reconstruction to verify that the standard deviation is stable. The results are shown in Table 7.2.

Table 7.2: Mean and standard deviation of area under the ROC curve for repeated 8X500 points ROC analysis

Test No.	mean area under ROC	standard deviation
1	0.918	0.014
2	0.914	0.008
3	0.910	0.010
4	0.926	0.008

7.2.2 Overlapping of data sets

In order to shorten the time required to do the ROC analysis, we thought of using partially overlapping data sets, as illustrated in Figure 7.2, to calculate the mean and standard deviation of the FOM.

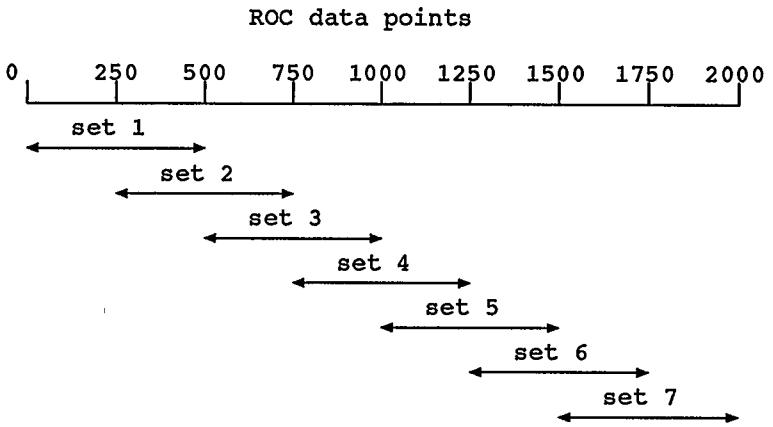


Figure 7.2: Overlap of data sets to calculate mean and standard deviation in a shorter period of time

Using 2000 points lesion contrast data in total, we can only generate 4 groups of 500 points lesion contrast data without overlapping. With overlapping we can generate 7 groups of 500 points data, giving a more realistic measure of the FOM and its standard deviation. However, there must be little correlation between the partially overlapping data sets for this overlapping approach to be valid. Usually, a decrease in standard deviation is an indication of correlation. If there is no correlation, the standard deviation calculated using overlapping data sets should be equivalent to that using non-overlapping data sets, but determined in a shorter period of time.

We tested our assumptions with 256×128 and 256×64 DFT reconstruction and the results are shown in Table 7.3. The standard deviation did not change significantly with overlapping, which suggests that overlapping can be used to shorten

analysis time.

Table 7.3: Standard deviation of area under the ROC curve for overlapping and non-overlapping 500 points ROC analysis

		128	128	64	64
	No. of data sets	mean area under ROC	standard deviation	mean area under ROC	standard deviation
non-overlapping	8	0.918	0.014	0.870	0.010
overlapping	15	0.917	0.017	0.869	0.014
overlapping	8	0.918	0.014	0.869	0.009

7.2.3 DFT reconstruction using MR data of different sizes

Computer ROC analysis

Would the computer ROC analysis of different reconstruction algorithms give similar results to the equivalent analysis by human observers? We did a computer ROC analysis with DFT reconstruction on increasingly truncated MR data. We truncated the original 256×256 data file to generate a series of 256×192, 256×128, 256×96, 256×64 and 256×32 MR data sets. The reconstructed images all contain 256×256 pixels.

We know that larger MR data sets are reconstructed, using the DFT algorithm, to give images with better resolution and fewer truncation artifacts. ([HB87]). We would expect the computer ROC analysis to show this trend. The ROC analysis should also generate meaningful distinction in terms of the mean and standard deviation of the FOM for the various algorithms. We performed the ROC analysis using 500 lesion contrast data points. The mean and standard deviation are calculated using 7 groups of partially overlapping data sets.

The ROC curves for all the reconstruction algorithms are shown in Figure 7.3. The mean and standard deviation of the area under the ROC curves are presented in Table 7.4. We also plotted the mean with standard deviation as error bars versus the data truncation in Figure 7.4.

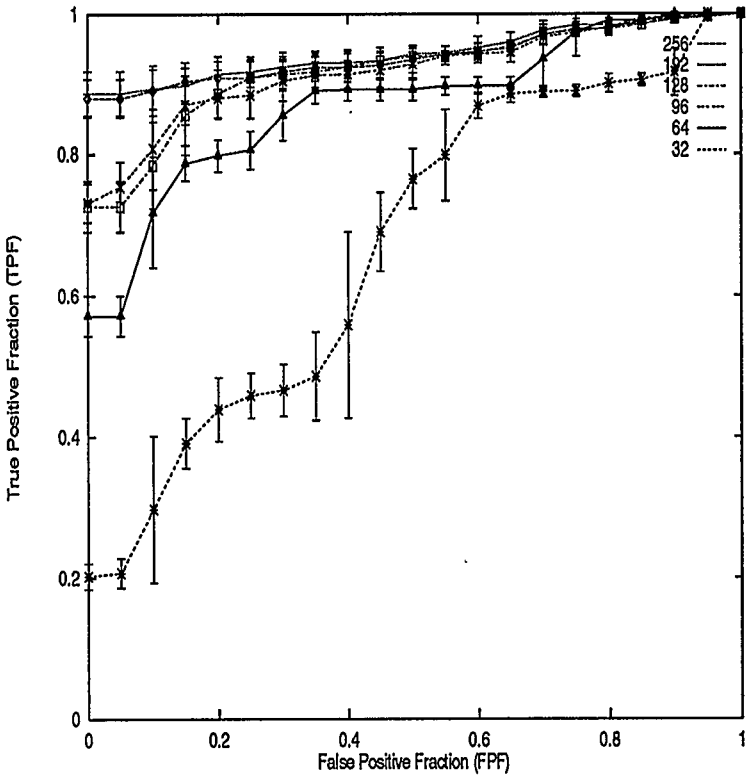


Figure 7.3: ROC curves of DFT reconstruction on increasingly truncated MR data

From Figure 7.4, we can say that when we reconstructed 256×256 images from decreasingly truncated MR data sets, the reconstructed image quality, indicated by the area under the ROC curves, improved as we expected. The computer ROC analysis recognized a large difference between images reconstructed from 256×32 and 256×64 MR data sets. The differences among images reconstructed from MR data of more than 256×128 are much smaller. The results comply with our knowledge

Table 7.4: Mean and standard deviation of area under the ROC curves for DFT reconstruction using MR data of different sizes

number of rows	mean area under ROC	standard deviation
32	0.675	0.014
64	0.879	0.014
96	0.915	0.012
128	0.916	0.011
192	0.937	0.009
256	0.933	0.010

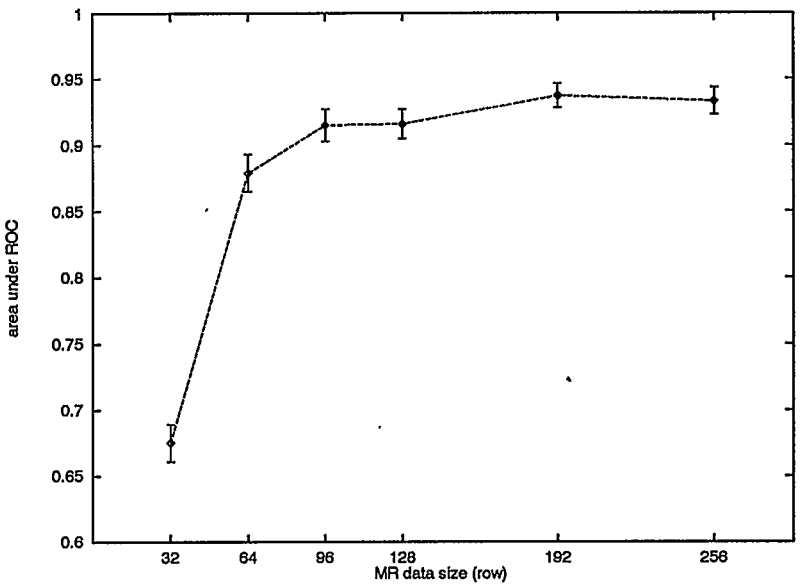


Figure 7.4: Area under the ROC curves for DFT reconstruction using MR data of different sizes

and reinforced our confidence in the computer ROC analysis.

Metric measurements

We also calculated the applicable metric measurements, Diffdata (DD), Diffimage (DI) and perceptual mean squared error (PMSE), for the same DFT reconstructions. The measurements are taken using 10 phantom images with randomly added lesions. The results are shown as the mean value of the 10 measurements $\pm$ the standard deviation in Table 7.5. All three measurements are calculated against the full (256×256) data/image sets.

Table 7.5: Metric measurements of DFT reconstruction on increasingly truncated data files

	DI(e-3)	DD	PMSE
192	0.505 ± 0.0001	33.126 ± 0.004	0.1570 ± 0.0001
128	1.522 ± 0.0003	99.733 ± 0.009	0.226185 ± 0.000002
96	2.334 ± 0.00002	152.99 ± 0.02	0.315923 ± 0.000004
64	4.628 ± 0.001	303.29 ± 0.04	0.46942 ± 0.00004
32	11.078 ± 0.004	726.0 ± 0.3	0.5451 ± 0.0003

As the truncation increases all the MSE measures increase as expected. Diffimage and Diffdata show a similar change pattern among reconstructions as the ROC analysis, i.e. larger difference between 256×32 and 256×64 reconstruction than the differences between reconstructions of the less truncated data sets. The PMSE measure shows a different change pattern than the ROC analysis. This is due to the effects of the human vision system (HVS) filter used in the PMSE measure. The HVS filter emphasizes high spatial frequency components. The result of using this filter before calculating the MSE is that the change in image quality in the higher

frequency range is emphasized. This result is shown by larger difference between the less truncated reconstructions (256×192 and 256×128) than the more truncated reconstructions (256×64 and 256×32). We do not know which change pattern is more realistic at this point.

An observation

One interesting observation is that while we can not detect the differences visually in images reconstructed from 256×192 and 256×256 data files, the computer observer seems to consistently rank the 256×192 point reconstruction slightly higher than the 256×256 point one on common data sets. This is shown in Table 7.6 for four separate tests. Although the standard deviations for the 256×192 and 256×256 reconstruction indicates that the difference in area under ROC is not significant, the consistency of the analysis results suggests that the computer observer is detecting something when the algorithms are used on a common data set.

Table 7.6: Repetition of area under the ROC curves for DFT reconstruction using MR data of 192 and 256 rows

	192	192	256	256
test No.	mean area under ROC	standard deviation	mean area under ROC	standard deviation
1	0.937	0.009	0.933	0.010
2	0.952	0.005	0.950	0.007
3	0.933	0.004	0.928	0.005
4	0.944	0.015	0.941	0.015

One possible hypothesis to explain why 256×192 point reconstruction might be better than 256×256 point reconstruction is that the high frequency components of MR data sets typically contain more noise than useful information. By truncating

the high frequency components to produce the 256 × 192 reconstruction, we may have reduced noise but not significantly reduced the information contained in the MR data. Perhaps the ROC analysis is sensitive enough to pick up this.

We did another set of ROC analysis on common data to test our hypothesis. The reconstruction algorithms are:

- 256 × 256 point DFT reconstruction
- 256 × 192 point DFT reconstruction, where the data is zero padded to generate a 256 × 256 image
- 256 × 192 point DFT reconstruction, where the data is noise padded to generate a 256 × 256 image

We expect that the zero padded 256 × 192 reconstruction would perform the best, since it has the least amount of noise in the high frequency regions. The noise padded 256 × 192 reconstruction would have a ROC value closer to the the 256 × 256 reconstruction. The computer ROC analysis seems to agree with our hypothesis(Table 7.7). If the ROC analysis is capable of detecting this effect we can place additional confidence in the approach.

Table 7.7: Mean and standard deviation of area under the ROC curve for 256 and 192 point reconstructions

reconstruction description	mean area under ROC	standard deviation
192 point, zero padded DFT	0.938	0.011
192 point, noise padded DFT	0.936	0.011
256 point DFT	0.933	0.012

7.2.4 Summary

In this section we have investigated and set up a protocol to use our computer ROC analysis. We determined initially:

- the sample size of our ROC analysis.
- the method to group the lesion contrast data and calculate the mean and standard deviation of the FOM.
- the validity of the computer ROC analysis.

We have also taken a set of metric measurements for comparison to the ROC analysis results. The measurements rank the algorithms in the same way as the computer ROC analysis, but does not show the same change pattern.

We can now evaluate the AM motion suppression algorithms and TERA reconstruction algorithms to see whether the computer ROC analysis can detect differences between algorithms. We also want to investigate the correlation between our computer ROC analysis and the fast MSE based measures.

7.3 Detailed Quantitative Evaluation Examples

We will now use our quantitative evaluation methods on two examples: AM motion suppression and TERA reconstruction. We present the metrics measurements as well as the ROC analysis.

7.3.1 AM motion suppression

In this section we quantitatively evaluate the AM motion suppression algorithms discussed in Chapter 3. Four algorithms are being compared: the Mitsa algorithm, the ARMA algorithm, the Direct algorithm and the Adaptive algorithm. We evaluate the performance of the algorithms on both low noise and high noise phantoms. For each set of phantoms, we first present the computer ROC analysis results, followed by the metric measurements.

Low noise phantoms

Mathematical phantoms are usually generated with zero background noise. In real MR images, there is always a certain amount of noise acquired during data collection. To make the phantoms as realistic as possible, we add Gaussian noise to the phantom. The amount of noise added is such that the signal to noise ratio of the phantom is similar to that of a real medical image. For the low noise phantoms, we measured the signal to noise ratio of a typical low noise MR image and added noise to the phantom accordingly.

The computer observer ROC analysis

Using the low noise abdomen phantom, we obtained 2000 lesion contrast data points and processed them in 7 groups of 500 point for the ROC analysis. To illustrate the procedures of our ROC analysis, we present the frequency graphs in one of 500 point ROC analysis in Figures 7.5 to 7.10 for the different reconstruction algorithms. The x-axis of the frequency graphs are broken into small ranges of contrast. The y-axis of these graphs indicates the number of occurrences, or frequency at a certain contrast range.

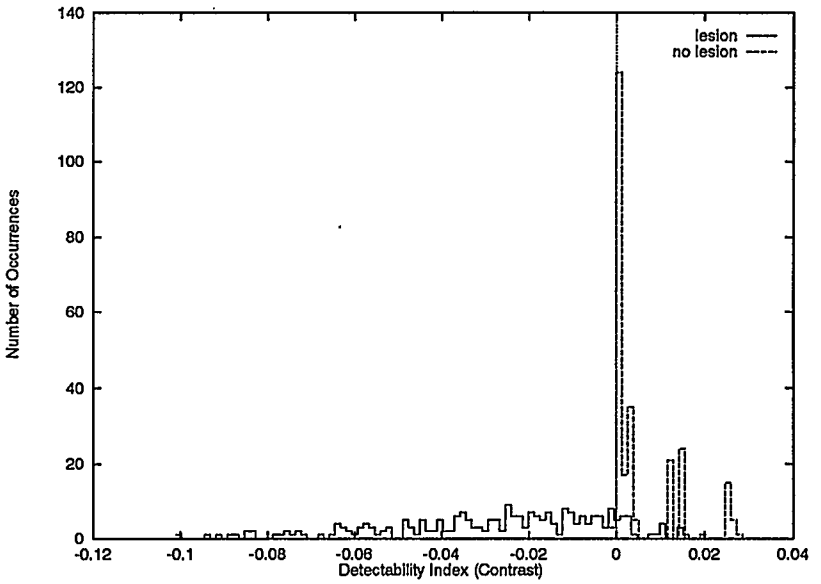


Figure 7.5: Frequency graph for the no motion low noise phantom

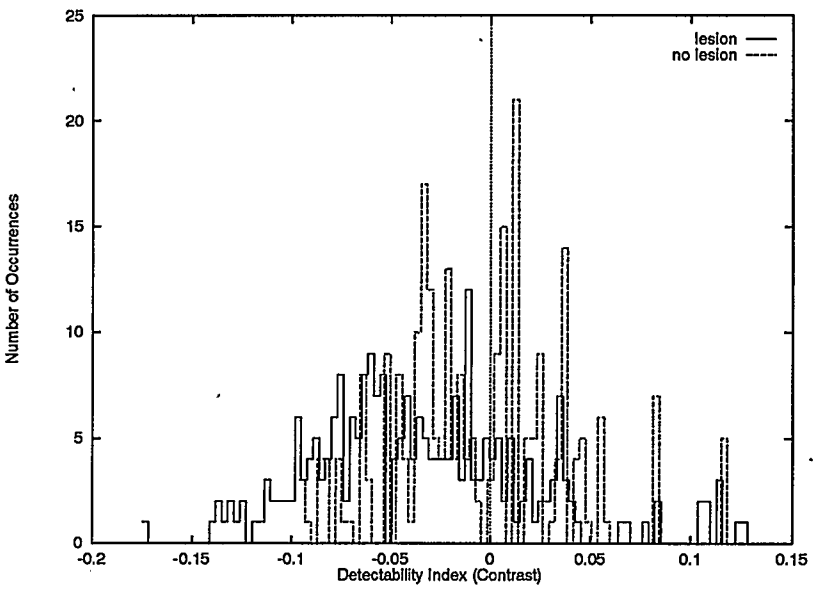


Figure 7.6: Frequency graph for the motion corrupted low noise phantom

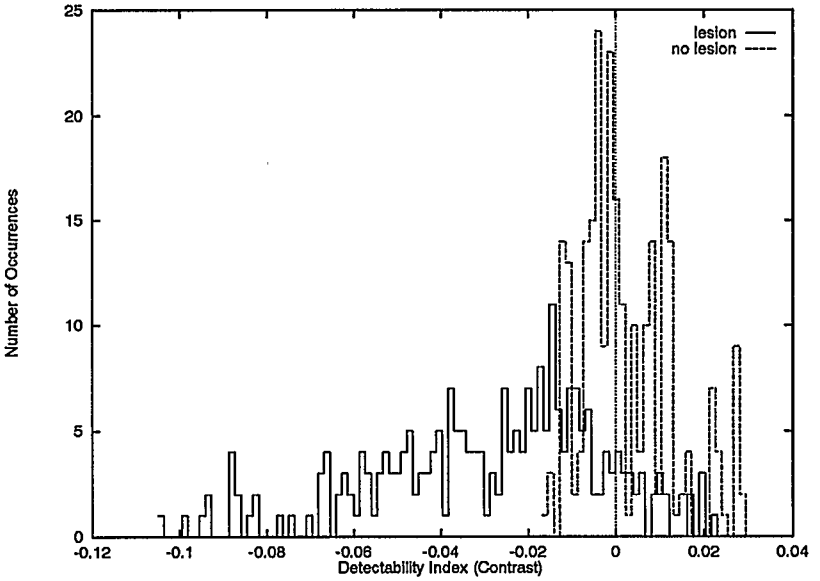


Figure 7.7: Frequency graph for Mitsa corrected low noise phantom

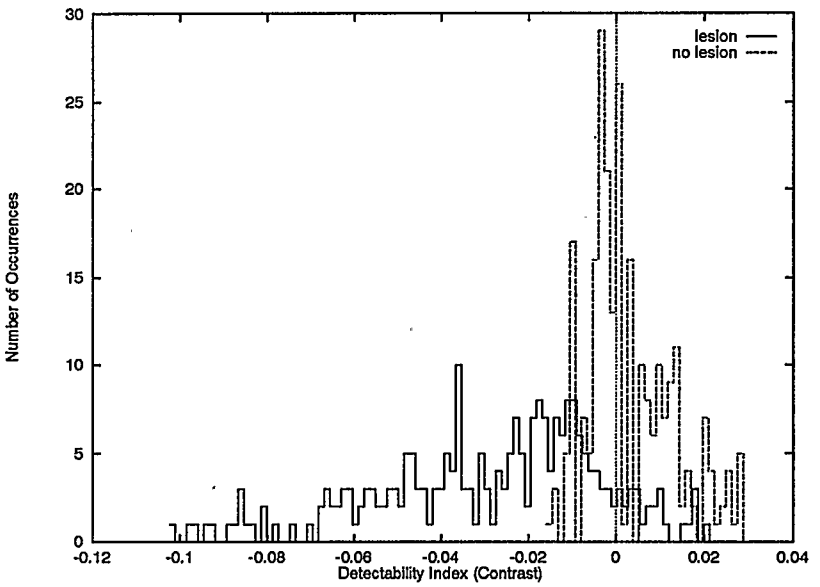


Figure 7.8: Frequency graph for ARMA corrected low noise phantom

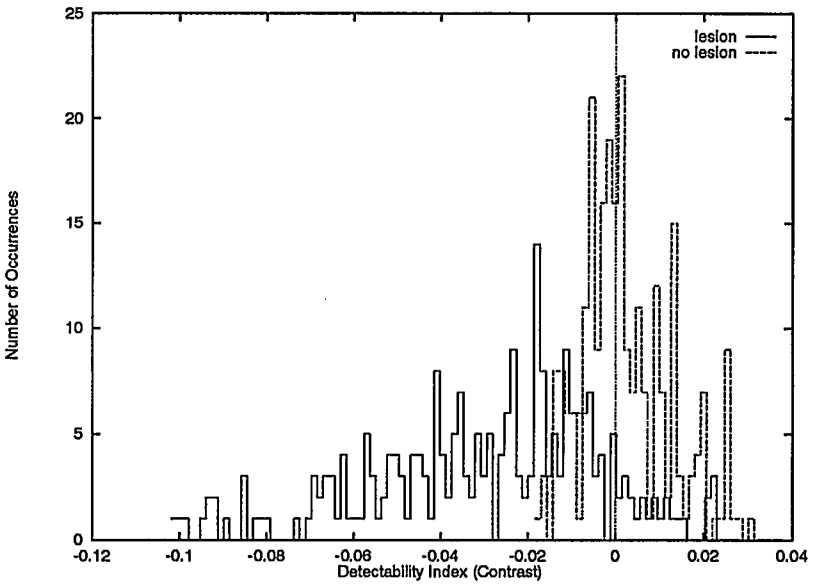


Figure 7.9: Frequency graph for Direct corrected low noise phantom

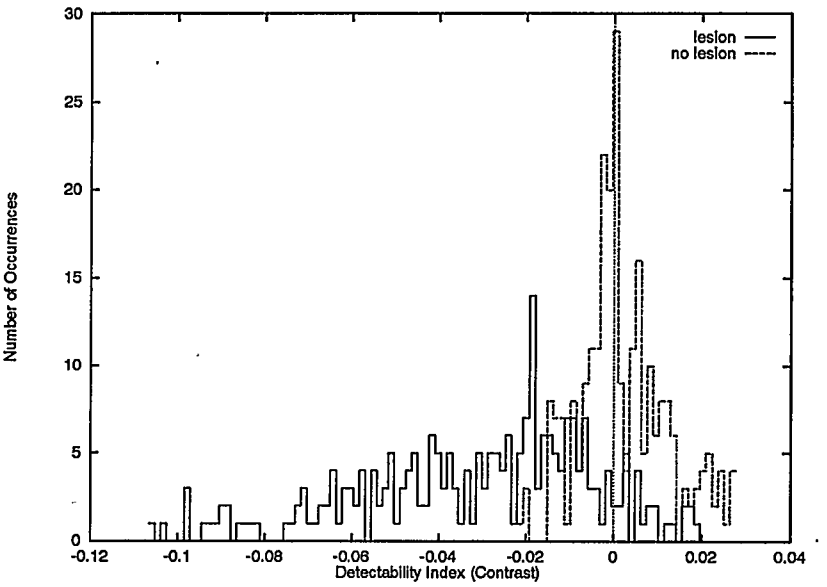


Figure 7.10: Frequency graph for Adaptive corrected low noise phantom

Comparing the frequency graphs in Figure 7.5 and Figure 7.6, we notice that in Figure 7.5 (without motion) there is less overlap between the no-lesion (negative) and lesion (positive) curves. Better separation of the no lesion and lesion frequency curves leads to better lesion detection. All the motion suppression algorithms reduced the overlapping of curves compared to that in Figure 7.6.

The resulting mean ROC curves from all 7 sets of ROC analysis are shown in Figure 7.11. The error bars show the standard deviation among 7 sets of ROC analysis.

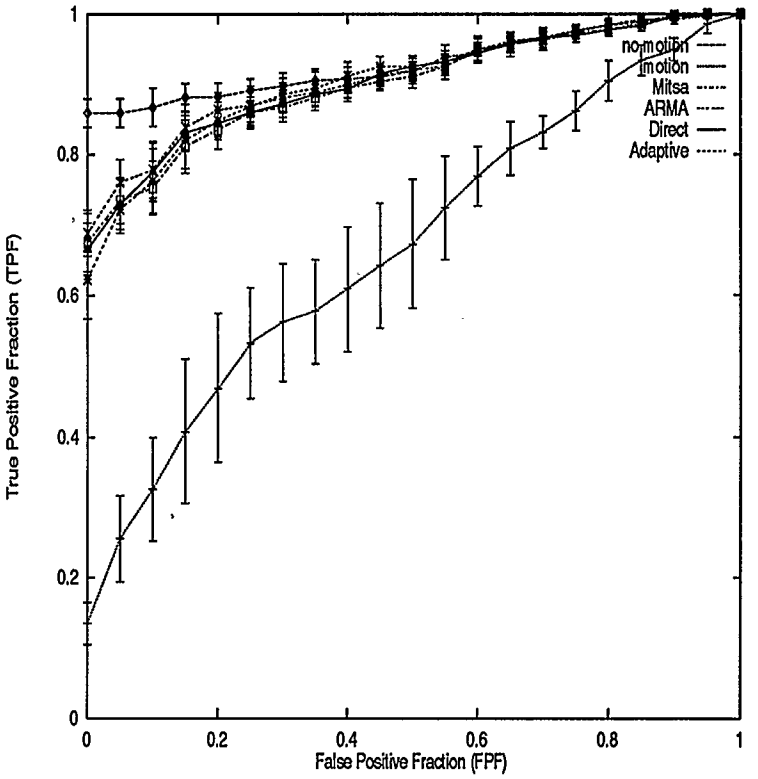


Figure 7.11: ROC curve for no motion low noise phantom, motion corrupted low noise phantom, and motion corrected low noise phantom using our motion suppression algorithms

As we can see, there is a large difference in area under the ROC curve between the

corrected phantoms and the motion corrupted phantom. However, the difference in performance of AM motion suppression algorithms is harder to detect. We calculated the area under the ROC curves and listed the results in Table 7.8. The differences between algorithms are smaller than 1%. We think that if all the algorithms are compared using common data with low noise, the ARMA algorithm probably outperforms the other algorithms. However, in terms of the standard deviations there is a borderline difference in performance of all the motion suppression algorithms in this case, as indicated by the ROC analysis.

Table 7.8: Area under the ROC curves for motion suppression on low noise phantoms

	motion	no motion	Adaptive	ARMA	Direct	Mitsa
mean area under ROC	0.670	0.931	0.909	0.911	0.905	0.902
standard deviation	0.043	0.012	0.015	0.013	0.013	0.014

Metrics

The metrics that can be applied to the AM motion suppression algorithms include:

- Diffimage (DI) - the mean squared error (MSE) between the reconstructed and standard MR images measured in the image domain.
- Diffdata (DD) - the MSE between the reconstructed and standard MR data sets in the frequency domain.
- Kernel mean squared error (KMSE) - the MSE between the reconstructed motion kernel and the original motion kernel.

- Kernel weighted mean squared error (KWMSE) - the weighted MSE between the reconstructed and original motion kernels. The weights represent the energy distribution of the MR data in the frequency domain.

- Perceptual mean squared error (PMSE) - the MSE between reconstructed and standard MR images that are preprocessed by human vision system (HVS) filters.

Using our set of metrics, we evaluated and ranked the AM motion suppression algorithms we developed. The test phantoms are the abdomen phantom with low Gaussian noise added. We generated 10 phantoms with randomly added lesions at know locations. The measurements in Table 7.9 show the mean value $\pm$ standard deviation over the resulting 10 test images.

The ranking from the metric measurements are quite consistent. All the metric measurements favor the Adaptive algorithm. The Mitsa algorithm has consistently been ranked the worst.

High noise phantoms

The high noise phantoms are produced by adding Gaussian noise to the noise free mathematical phantom until the signal to noise ratio is comparable to that of a real high noise MR image. The noise added is approximately 5 times the noise added to produce the low noise phantoms.

The computer observer ROC analysis

The resulting ROC curves from our ROC analysis are shown in Figure 7.12. The area under the ROC curves are shown in Table 7.10.

Table 7.9: Metric measurements of motion suppression on low noise phantoms

	motion	Adaptive	ARMA	Direct	Mitsa
DI(e-4)	137	2.2	3.6	4.4	4.9
	± 1	± 0.1	± 0.4	± 0.2	± 0.9
rank		*1*	*2*	*3*	*4*
DD	895	14.6	24	29	32
	± 7	± 0.7	± 2	± 1	± 6
rank		*1*	*2*	*3*	*4*
KMSE(e-2)		2.6	2.8	2.8	3.9
		± 0.2	± 0.1	± 0.1	± 0.4
rank		*1*	*2*	*2*	*4*
KWMSE(e-3)		1.27	2.5	3.1	3.6
		± 0.09	± 0.3	± 0.2	± 0.8
rank		*1*	*2*	*3*	*4*
PMSE(e-2)	57.5	8.4	8.9	9.0	11
	± 0.6	± 0.2	± 0.6	± 0.7	± 1
rank		*1*	*2*	*2*	*4*

Table 7.10: Area under the ROC curves for motion suppression on high noise phantoms

	motion	no motion	Adaptive	ARMA	Direct	Mitsa
mean area under ROC	0.640	0.856	0.699	0.668	0.669	0.668
standard deviation	0.024	0.023	0.019	0.020	0.021	0.021

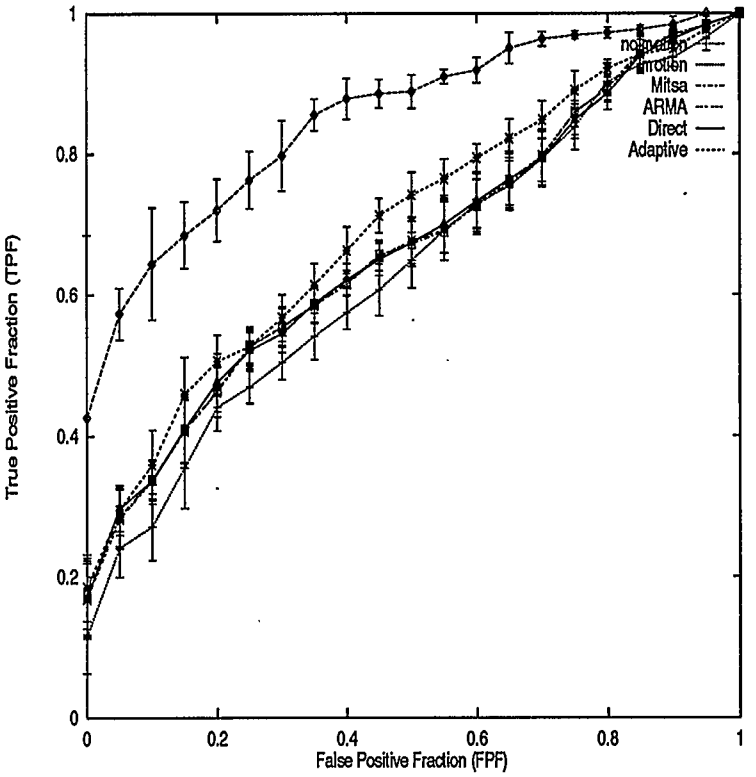


Figure 7.12: ROC curve for no motion high noise phantom, motion corrupted high noise phantom, and motion corrected high noise phantom using our motion suppression algorithms

With large amount of noise added, all the motion suppression algorithms perform less successfully, as expected. There is no difference in area under ROC for ARMA, Direct and Mitsa algorithms. All three algorithms made small improvement in terms of ROC results compared to the motion corrupted data. However, the Adaptive algorithm outperforms all the other algorithms.

One interesting phenomena is that the adaptive algorithm emerges above all the other algorithms. One explanation for the better performance of the adaptive algorithm is that noise corrupt the motion peaks and in turn the relationship (amplitude modulation) between the motion peaks and the DC peak. The algorithms that work directly on extracting this relationship are affected most and their performance deteriorate most. The adaptive algorithm uses the DC and low frequency components of the corrupted signal as an estimate of the desired signal and filters the corrupted signal to produce the desired signal. Since the DC and low frequency components are less affected by noise, the adaptation not only suppressed the motion peaks, but also reduced the noise effect. This can be seen in Figure 7.13. As a result, the adaptive algorithm performs better than the other motion suppression algorithms in the presence of large noise.

Metrics

The metric measurements for high noise phantoms are shown in Table 7.11.

With the high noise phantoms all the measurements rank the Adaptive algorithm the best. This favor for the Adaptive algorithm confirms what was shown by the computer ROC analysis. The ranking of the other three algorithms are rather consistent among the metric measurements.

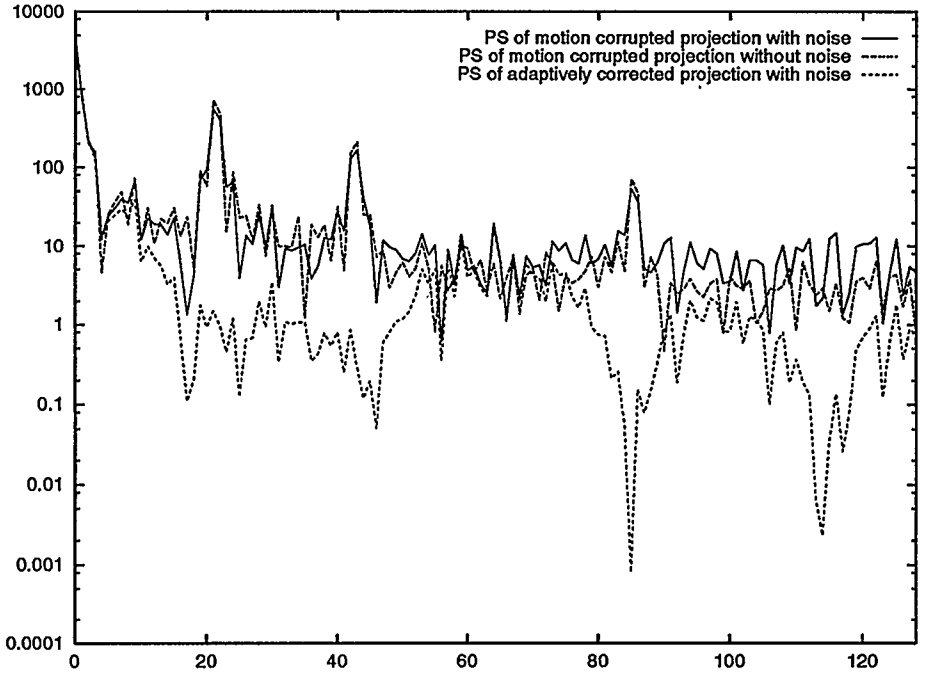


Figure 7.13: The noise suppression effect of the adaptive algorithm

Table 7.11: Metric measurements of motion suppression on high noise phantoms

	motion	Adaptive	ARMA	Direct	Mitsa
DI(e-3)	13.68	3.7	5.9	7.6	5.9
	± 0.01	± 0.2	± 0.2	± 0.5	± 0.2
rank		*1*	*2*	*4*	*2*
DD	898	242	385	495	384
	± 8	± 13	± 16	± 30	± 11
rank		*1*	*2*	*4*	*2*
KMSE(e-1)		1.94	2.04	2.8	2.08
		± 0.03	± 0.02	± 0.8	± 0.02
rank		*1*	*2*	*4*	*3*
KWMSE(e-2)		4.4	6.1	7.8	6.1
		± 0.2	± 0.2	± 0.7	± 0.1
rank		*1*	*2*	*4*	*2*
PMSE(e-1)	4.4	1.56	2.31	3.7	2.29
	± 0.2	± 0.05	± 0.03	± 0.5	± 0.06
rank		*1*	*2*	*4*	*2*

7.3.2 TERA reconstruction algorithms

In this section, we tested our evaluation methods with TERA reconstruction algorithms. The T1-weighted abdomen phantom was used. The data file was truncated from 256×256 to 256×128. We first did an ROC analysis using different model orders with TERA reconstruction and selected an optimum model order. TERA algorithm was then applied with the pre-selected model order. The resulting modeled image was zoomed to 256×256. This image was compared to DFT reconstructed images of 256×256 (long) and 256×128 (short).

The constrained TERA (CTERA) algorithm was also tested. This algorithm uses Sigma-filtered data as the input to the TERA algorithm. It has been shown to improve the performance of TERA algorithm [Yan93].

The low noise and high noise phantoms are generated using the same approach as in motion suppression.

ARMA model order

We did this computer ROC analysis to see if the computer observer can tell the differences between TERA reconstruction algorithms using different model orders. We used 5, 10, 15, 20, 25 and 30 as our TERA model order on the low noise phantom. The resulting ROC curves are shown in Figure 7.14. The mean and standard deviation of area under the ROC curves are shown in Table 7.12.

The results in Table 7.12 show that TERA reconstruction using model order 15, 20 and 25 performed better than using model order 5, 10 and 30. Model order 20 is optimal in this analysis. These results are consistent with our experience in determining model orders. It agrees with the results Yang obtained using Normalized

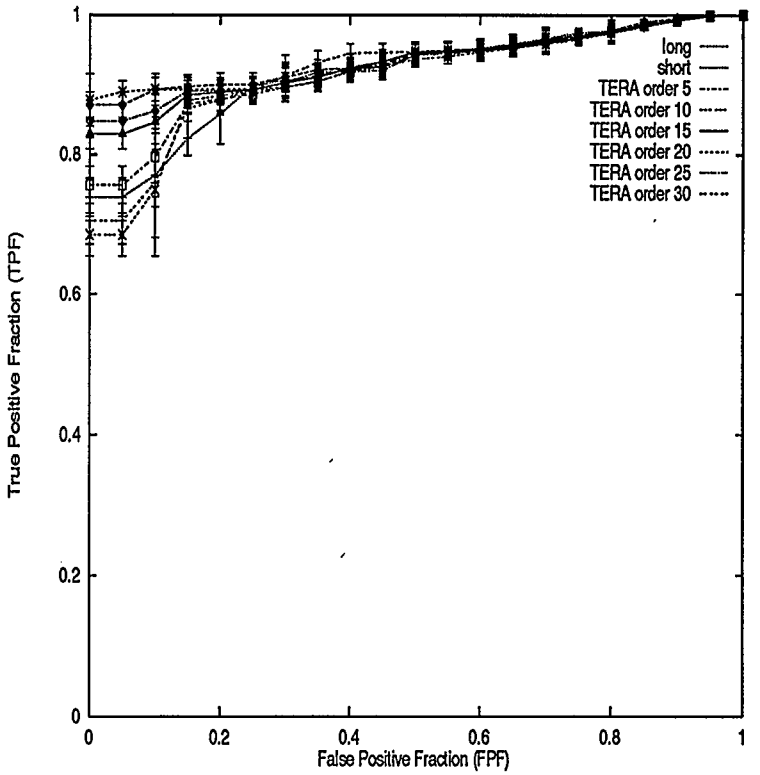


Figure 7.14: ROC curves of TERA reconstruction algorithms using different model orders on low noise phantoms

Table 7.12: Area under the ROC curves for TERA reconstruction using different model orders on low noise phantoms

TERA model order	mean area under ROC	standard deviation
long	0.939	0.011
short	0.917	0.013
5	0.920	0.012
10	0.912	0.013
15	0.932	0.014
20	0.939	0.011
25	0.932	0.011
30	0.919	0.014

Diffenergy measurement ([Yan93]). Our computer ROC analysis in this case managed to track the little differences ARMA modeling order can make in MR image reconstruction using common data sets.

In the following sections, TERA reconstruction algorithm will be applied with model order 20.

Low noise phantoms

Computer Observer ROC Test

With the low noise phantoms, the area under the ROC curves are shown in Table 7.13. We also show the mean ROC curves of all 7 500 point ROC analysis in Figure 7.15. "short" denotes DFT reconstruction of 256 $\times$ 128 point data file. "long" denotes DFT reconstruction of 256 $\times$ 256 point data file. "TERA" and "CTERA" denote modeling reconstruction from 256 $\times$ 128 point data file.

Table 7.13: Area under the ROC curves for TERA reconstruction on low noise phantoms

	short	long	TERA	CTERA
mean area under ROC	0.884	0.929	0.929	0.930
standard deviation	0.018	0.007	0.006	0.007

From figures in Table 7.13, we see that TERA and CTERA constructed images from short data sets are significantly better than DFT-reconstructed images from short data sets, and are as good as DFT-reconstructed images from long data sets.

Metrics

The metrics applicable to this application are DD, DI and PMSE. Since the reconstruction algorithms do not extract a kernel, the kernel related MSE metrics are not

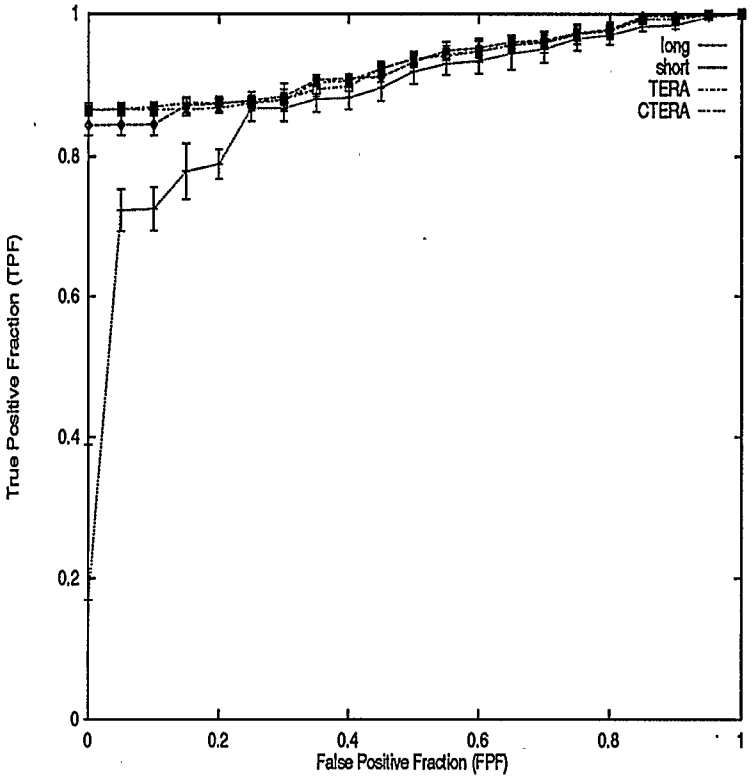


Figure 7.15: ROC curve for all reconstructions on low noise phantoms

applicable. The results on low noise phantoms are shown in Table 7.14.

Table 7.14: Metric measurements of TERA and CTERA reconstruction on low noise phantoms

	Short	TERA	CTERA
DI(e-3)	1.522 ± 0.0001	1.054 ± 0.009	1.005 ± 0.003
DD	99.731 ± 0.006	69.1 ± 0.6	65.9 ± 0.2
PMSE	0.2262 ± 0.0001	0.1942 ± 0.0005	0.187 ± 0.001

From the metrics, we notice that TERA and CTERA reconstructed images from short data sets are closer to the images reconstructed from the long data set than DFT-reconstructed images from short data sets. However, since we are comparing to the images DFT-reconstructed from the long data sets, we will not be able to tell if TERA or CTERA reconstruct better images from short data sets than DFT from the long data sets. All three measurements rank CTERA higher than the TERA algorithm for this phantom, which is the same as what we get from the computer ROC analysis.

High noise phantoms

Table 7.15: Area under the ROC curves for TERA reconstruction on high noise phantoms

	short	long	TERA	CTERA
mean area under ROC	0.845	0.866	0.858	0.867
standard deviation	0.008	0.009	0.010	0.010

We present the mean ROC curves of all 7 500 point ROC analyses and the area under ROC curves in Figure 7.15 and Table 7.13, respectively. The TERA and CTERA reconstructed images from the short data sets again have similar quality to

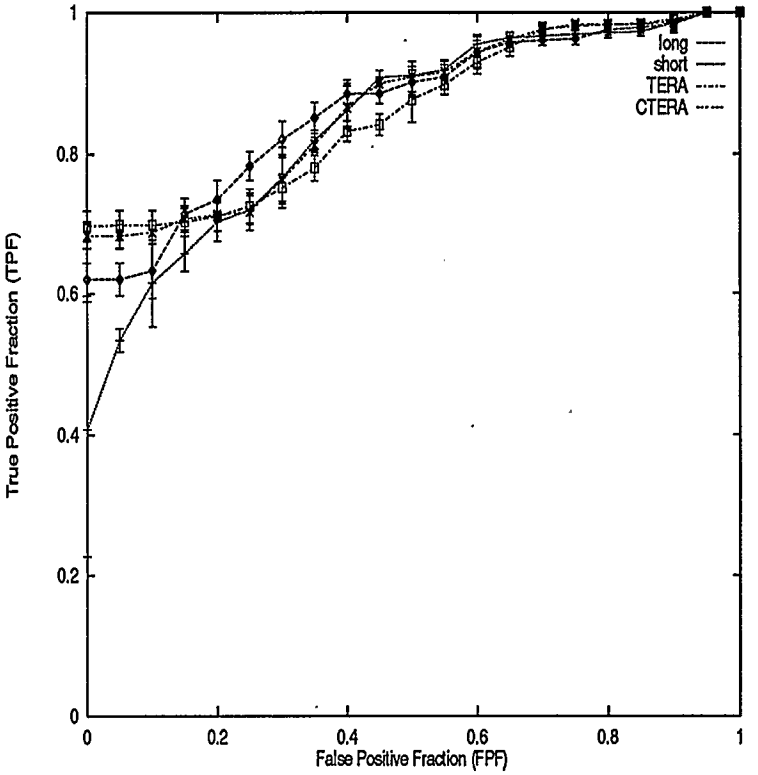


Figure 7.16: ROC curve for all reconstructions on high noise phantoms

the DFT-reconstructed images from the long MR data sets. With high noise, the CTERA reconstruction seem to perform better than TERA on common data sets.

Metrics

Table 7.16: Metric measurements of TERA and CTERA reconstruction on high noise phantoms

	Short	TERA	CTERA
DI(e-3)	7.388 ± 0.000	8.708 ± 0.006	8.697 ± 0.009
DD	484.15 ± 0.02	570.7 ± 0.4	570.0 ± 0.6
PMSE	0.2429 ± 0.0002	0.2486 ± 0.0001	0.2472 ± 0.0004

If we only look at the metric measures, we might draw the conclusion that TERA and CTERA reconstructed images have worse quality than DFT reconstructed images using short MR data sets. However, if we look at the ROC analysis results this is not true clinically. In this sense the computer ROC analysis is superior, able to generate an clinical indication of image quality.

7.4 Computer ROC Analysis Using a Real Medical Image

We also tested our motion suppression algorithms on a real medical image with a simulated lesion added to it. The background image was a section of a thigh. The lesion was added to a fixed location, as shown in Figure 7.17. As in other ROC analyses, the lesion was generated in the frequency domain. It was then added to the frequency domain data of the thigh. A random variable was used to decide whether the lesion should be added or not. The intensity of the lesion was also randomly determined between two thresholds. Since we are only using this analysis as an

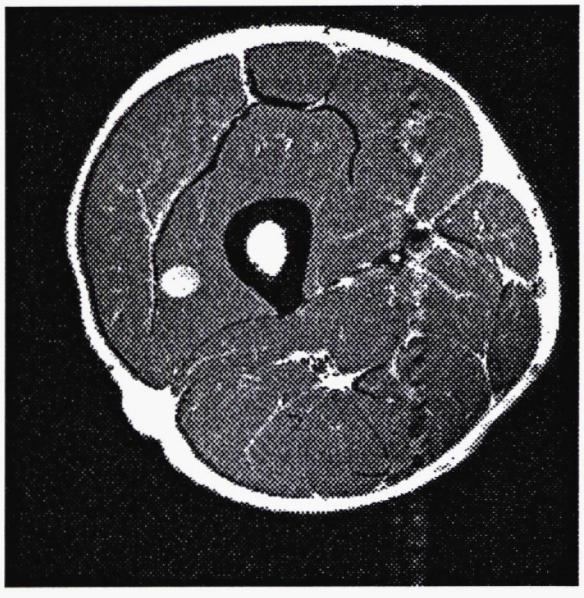


Figure 7.17: The thigh image with simulated lesion

quick example on how to use the computer ROC analysis on real medical images, we did not consult a radiologist on the lesion simulation process. The lesion is simply superimposed on the image and retains much of the structure of the underlying tissue without the lesion. For a practical computer ROC analysis using medical images, the lesion simulation process should be carefully designed so that the lesions can be as realistic as possible.

All the AM motion suppression algorithms were then applied on the thigh image with lesion. Contrasts at the lesion location were calculated and recorded. 2000 lesion contrast data points are collected and devided into 7 sets of overlapping 500 lesion contrast data points. The mean ROC curves for all 7 sets of data are shown in Figure 7.18. The mean and standard deviation of area under the ROC curves was calculated and shown in Table 7.17. The metric measurement results from 10 thigh images with a randomly added lesion are shown in Table 7.18.

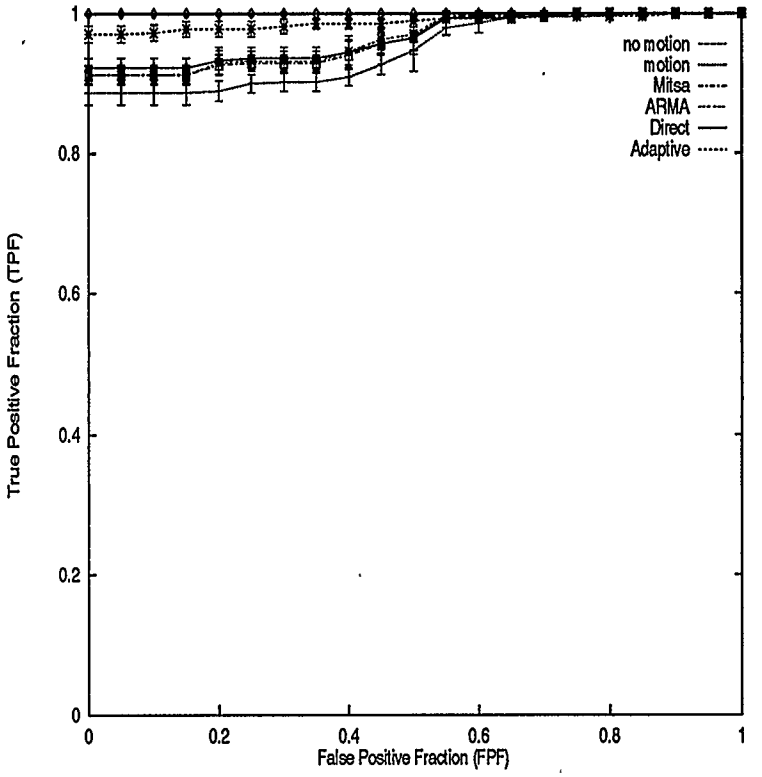


Figure 7.18: ROC curves of AM motion suppression algorithms using the thigh image with a simulated lesion

Table 7.17: Area under the ROC curves for motion suppression on a real medical image with a simulated lesion

	motion	no motion	Adaptive	ARMA	Direct	Mitsa
mean area under ROC	0.946	1.000	0.988	0.964	0.965	0.962
standard deviation	0.007	0.000	0.006	0.007	0.008	0.006

Table 7.18: Metric measurements of motion suppression on medical thigh image with a simulated lesion

	motion	Adaptive	ARMA	Direct	Mitsa
DI(e-4)	171 ± 8	2.4 ± 2.7	3.1 ± 1.2	3.0 ± 1.0	5.6 ± 2.1
DD	1126 ± 55	16 ± 17	20 ± 8	20 ± 7	36 ± 14
KMSE(e-2)		5.8 ± 2.2	3.1 ± 1.1	1.0 ± 0.1	1.5 ± 0.4
KWMSE(e-3)		1.2 ± 1.3	1.6 ± 0.6	1.6 ± 0.5	3.0 ± 1.1
PMSE(e-2)	32 ± 3	2.5 ± 0.5	1.4 ± 0.3	1.3 ± 0.3	1.9 ± 0.5

The metric measurements have relatively large standard deviations. It is difficult to rank the algorithms based on these metric measurement results.

The ROC analysis results favor the Adaptive algorithm. This ROC analysis extends the area of application of our computer ROC analysis method. Algorithms can be compared using real medical images as the background and simulated lesions can be added to them. However, the process of lesion simulation needs to be carefully designed so that the images with lesions are realistic.

7.5 Comments on other computer ROC analysis using in MRI

While writing up this thesis, a paper on ROC analysis for evaluating functional brain MR imaging and post-processing protocols was published [CSG95]. In this paper, Constable *et. al.* obtained multiple real MR images and superimposed on the images

activation signals of various intensity. Different computer post-processing algorithms were applied to detect the activation signals pixel by pixel. The "strictness" of the detection criterion was varied and the true positive fraction (TPF) and false positive fraction (FPF) of activated pixels were calculated to generate an ROC curve for each post-processing algorithm. The ROC results from different types of images were averaged and shown with error bars. From the ROC curves the post-processing algorithms were ranked.

The pixel based activation detection in [CSG95] is specific to functional MR imaging and is quite different from our location based lesion detection. We think that our lesion detection task is more general and could be applied to a wide range of MR images and post-processing application evaluations. The ROC curves in [CSG95] also have different meanings than our ROC curves. As shown in Figure 7.19, the error bars represent repeatability of results, since the standard deviations are calculated using images of different subjects. Our standard deviation indicates variations in results from test to test using the same phantom with added lesions. With major differences between our ROC analysis approach and that of Constable *et. al.*, we can not compare the ROC results. We think Constable's ROC analysis is specific to functional MR imaging while our ROC analysis is more general.

Although Constable *et. al.* used the ROC approach to generate quantitative evaluation results, they did not show how the ROC results may change when the number of images and the number of types of images are varied. Thus we think the results would only be valid for the specific case in this paper and could not be generalized. In addition, if activation signals were generated in the image domain and superimposed on the real MR images, the resulting images are not realistic ([SC94],

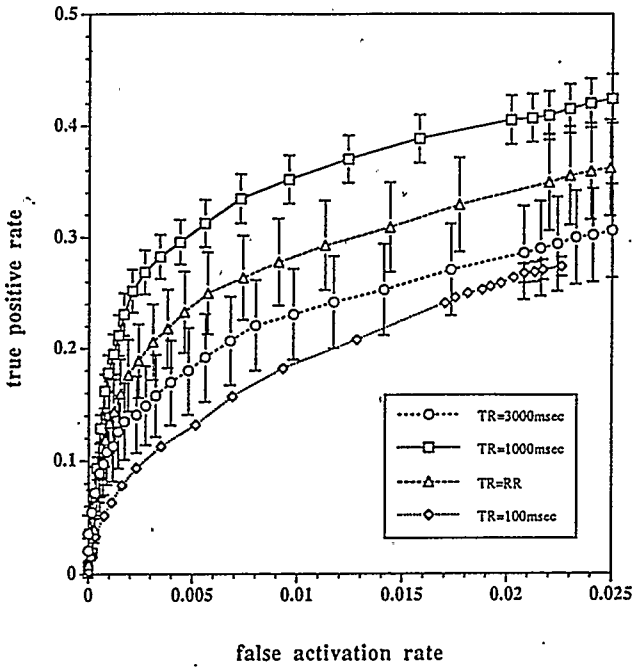


Figure 7.19: A sample ROC plot from Constable's paper ([CSG95])

Page 6.1 of this thesis). In real MR images we can not have shapes with infinitely sharp edges. Such unrealistic MR images may generate unreliable ROC results.

Chapter 8

Conclusions

8.1 Project Summary

During this thesis project, two main areas were explored: quantitative evaluation methods and AM motion suppression.

8.1.1 Quantitative evaluation methods

In magnetic resonance imaging and other medical imaging fields, image processing algorithms are developed to enhance the quality of images, or fulfill other medical requirements. Currently it is difficult to obtain an objective comparison of how well algorithms satisfy medical requirements. In this project, we attempted to provide researchers with a variety of quantitative algorithm evaluation methods, including:

- Mean-Squared-Error (MSE) based measures, including
  - Diffdata (DD) – the MSE between the reconstructed and standard MR data measured in the frequency domain
  - Diffimage (DI) – the MSE between the reconstructed and standard MR image measured in the image domain
  - Perceptual MSE (PMSE) – the MSE between the reconstructed and standard MR images that are preprocessed by human vision system filters

- Kernel MSE – the various MSE measures between reconstructed and actual motion kernels, applicable only to motion suppression algorithms

- a computer observer receiver operator characteristics (ROC) analysis

We first compiled the set of MSE based measures to compare algorithm performance using a standard data/image set. The metrics provide numerical results that lack medical relevancy. We then designed a computer observer ROC analysis to give quantitative comparisons of medical performance.

Our computer ROC analysis provides a paradigm for algorithm comparison. Using our computer ROC analysis method, researchers can design their own phantom with lesions added at known locations, apply the image processing algorithms to be evaluated, and obtain ROC curves for each algorithm. They can also add lesions to actual medical images, rather than to synthetic images.

We tested our ROC method using different motion suppression and reconstruction algorithms in MRI. The results show a consistent ranking of algorithm performance by the computer ROC analysis. We included some problematic observations made during our experiments to illustrate that various factors can affect the result of the ROC analysis.

Used properly, the computer ROC analysis can be a powerful and efficient tool for quantitative algorithm evaluation in medical imaging. It is flexible in selection of the imaging scene. The computer ROC method is much cheaper and faster than human ROC analysis. It gives clinically meaningful results. With continued development, including the use of more complex signal models and better machine recognition, it may be used to replace human ROC analysis.

8.1.2 AM motion suppression

A simplified motion model was used and a series of motion suppression algorithms were developed based on that model. All the algorithms effectively removed the simulated motion. The introduction of ARMA modeling in generating the power spectrum increased the resolution of the power spectrum and made peak detection more accurate. The adaptive algorithm was flexible, completely automated and robust in the presence of noise. However, the simplified motion model could not describe real life motion in MRI. Better mathematical motion models are needed. Nevertheless, our motion suppression exercise indicates that, with an accurate mathematical model of an artifact, computer post-processing algorithms can efficiently correct artifacts and enhance medical image quality.

8.2 Future Work

8.2.1 Motion suppression

In the area of motion suppression, the amplitude modulation model has been investigated extensively. We have shown that the simplified AM model could not account for real life motion in MRI. To solve the overall motion problem in MRI using post-processing, we need a generalized motion model that gives an accurate and complete description of MRI motion. With such a model, post-processing may be the easiest motion suppression technique to implement, requiring no change in instrumentation or pulse sequences.

Since motion in MRI is inherently a 3D problem (k-space and time), we can also try to suppress motion artifacts using 3D signal processing. The conventional MRI

data collection techniques, such as those using the spin-echo pulse sequence, collapse the 3D information into 2D. To obtain 3D information, we have to design new data collection methods that acquire a series of 2D data sets. The data collection method suggested in reference [LR93] acquires a series of short 2D data sets. This method can be combined with modeling to generate a larger 3D data set. 3D signal filters could then be used on the 3D MR data acquired to suppress motion. This 3D approach will involve new pulse sequence design and design of multi-dimensional filters.

8.2.2 Quantitative evaluation methods

The computer observer used in this project has excellent potential in the area of quantitative evaluation for both research algorithms and clinical applications. It has the advantages of low cost and fast evaluation. Future work can be carried out in the following areas:

- Generation of more realistic medical phantoms. The current model only supports non-interceptive ellipses.. More shapes and fewer restrictions would better represent real medical images.

- More sophisticated analysis to generate FOMs using the ROC curves. The total area under the ROC curve is not a sufficient FOM under all circumstances. Sometimes certain segments of the ROC curve mean more clinically than others. The ROC curve analysis should take into account cost factors associated with treating falsely identified patients and missing out unidentified patients. This was brought to our attention during the thesis defense. After the thesis defense we did some preliminary ROC curve analysis to evaluate algorithms at

a small false alarm rate. We measured the mean area under the ROC curves in Figure 7.3, on page 114 at FPF of 5%, 10% and 20% and compared that with the total area under the ROC curves (FPF=100%). The results are shown in Table 8.1.

Table 8.1: Mean and standard deviation of partial and total areas under the ROC curves for DFT reconstruction using MR data of different sizes

No. of rows	FPF=5%	FPF=10%	FPF=20%	FPF=100%
32	0.011 ± 0.001	0.025 ± 0.003	0.063 ± 0.005	0.675 ± 0.014
64	0.031 ± 0.002	0.064 ± 0.004	0.141 ± 0.007	0.879 ± 0.014
96	0.038 ± 0.001	0.077 ± 0.002	0.162 ± 0.005	0.915 ± 0.012
128	0.037 ± 0.001	0.075 ± 0.002	0.158 ± 0.004	0.916 ± 0.011
192	0.043 ± 0.001	0.086 ± 0.002	0.175 ± 0.003	0.937 ± 0.009
256	0.043 ± 0.001	0.086 ± 0.002	0.174 ± 0.004	0.933 ± 0.010

This preliminary ROC curve analysis show that based on parial or total areas under the ROC curves, the ranking of algorithms are not much different. This results show that more in-depth ROC curve analysis is needed in the future studies to interpret the ROC curves better.

- Further tests on the computer observer. If possible, correlation between the computer observer and human observers should be calculated. We may need to carry out extensive ROC analysis by both human and computer observers using the same images.

- A more sophisticated computer observer should be developed. Our computer observer performs the simplest type of tasks in the signal detection scenario. A future computer observer could be developed using:

- A complex signal model. The signals in our project are the lesions in the images. Here, the lesions appear at known locations. More realistically, we need to recognize shapes and be able to classify them. The development of signal models may involve image segmentation and classification.
- A complex decision-maker. Here, the decision of whether or not a lesion exists is based only on contrast. However, human decision-makers undertake complex visual and psychological processes before reaching a decision. Better understanding of these processes will add artificial intelligence to our computer observer and make it more convincing.

With the development of artificial intelligence and computer vision, we hope that one day, our computer observer will be trusted like a human observer. Until that day, extensive research needs to be carried out to better understand human intelligence.

Bibliography

[AO94] E. Atalar and L. Onural. A respiratory motion artifact reduction method in magnetic resonance imaging of the chest. *IEEE Transactions on Medical Imaging*, 10, No. 1, March 1994.

[AW90] T. N. Arvanitis and D. Watson. Overcoming respiratory motion artifacts in MRI. In *1990 IEEE*, 1990.

[Bar92] P. G. J. Barten. Physical model for the contrast sensitivity of the human eye. *SPIE Proc.*, 1666, 1992.

[BMW86] H. H. Barrett, K. J. Myers, and R. F. Wagner. Beyond signal detection theory. *SPIE Proc.*, 626, 1986.

[CP91] R. Cromartie and S. M. Pizer. Structure-sensitive adaptive contrast enhancement methods and their evaluation. *IEEE Transactions on Medical Imaging*, 10, No. 3, September 1991.

[CSG95] R. T. Constable, P. Skudlarski, and J. C. Gore. An ROC approach for evaluating functional brain MR imaging and post-processing protocols. *Magnetic Resonance in Medicine*, 34, 1995.

[Ega75] J. P. Egan. *Signal Detection Theory and ROC Analysis.* Academic Press, New York, 1975.

[FBSM87] R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers. Hotelling trace criterion and its correlation with human-observer performance. *Journal*

of Optical Society of America, 4, No. 5, May 1987.

[FHN+94] S. S. Furuie, G. T. Herman, T. K. Narayan, P. E. Kinahan, J. S. Karp, and R. M. Lewitt. A Methodology for testing for statistically significant differences between fully 3D PET reconstruction algorithms. *Physics of Medical Biology*, 39, 1994.

[GA93] Gmitro and Alexander. Use of a projection reconstruction method to decrease motion sensitivity in diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 29, 1993.

[GP92] Glover and Pauly. Projection reconstruction techniques for reduction of motion effects in MRI. *Magnetic Resonance in Medicine*, 28, 1992.

[Gra] E. M. Granger. A summary measure for image quality. *?*, ? ?

[GS66] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics.* John Wiley and Sons Inc., New York, 1966.

[GW87] R. C. Gonzalez and P. Wintz. *Digital Image Processing.* Addison-Wesley Publishing Company, 1987.

[Hal80] C. F. Hall. Perceptual coding in the Fourier transform domain. *NTC80*, 36.1, 1980.

[Hal81] C. F. Hall. Subjective evaluation of a perceptual quality metric. *SPIE Proc.*, 310, 1981.

[Han88] K. M. Hanson. Method to evaluate image-recovery algorithms based on task performance. *SPIE Proc.*, 914, 1988.

[Han90] K. M. Hanson. Method of evaluating image-recovery algorithms based on task performance. *Journal of Optical Society of America*, 7, No. 7, July 1990.

[Har78] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66, No. 1, January 1978.

[Hay91] S. Haykin. *Adaptive Filter Theory.* Prentice-Hall, 1991.

[HB87] R. M. Henkelman and M. J. Bronskill. Artifacts in magnetic resonance imaging. *Magnetic Resonance in Medicine*, 2, No. 1, 1987.

[HH77] C. F. Hall and E. L. Hall. A nonlinear model for the spatial characteristics of the human vision system. *IEEE Transactions on System, Man and Cybernetics*, 7, No. 3, March 1977.

[HO91] G. T. Herman and D. Odhner. Performance evaluation of an iterative image reconstruction algorithm for Positron Emission Tomography. *IEEE Transactions on Medical Imaging*, 10, No. 3, September 1991.

[How89] Howe. Implementation and evaluation of data compression of MR images. *Magnetic Resonance Imaging*, 8, 1989.

[HY91] G. T. Herman and K. T. Yeung. Evaluator of image reconstruction algorithms. *International Journal of Imaging Systems and Technology*, 1, 1991.

[HY92a] M. Hedley and H. Yan. Restoration of MR images corrupted with translational motion. In *Proceedings of ICIP-92*, 1992.

[HY92b] M. Hedley and H. Yan. Suppression of slice selection axis motion artifacts in MRI. *IEEE Transactions on Medical Imaging*, 11, No. 2, June 1992.

[HYR91a] M. Hedley, H. Yan, and D. Rosenfeld. An improved algorithm for 2-D translational motion artifact correction. *IEEE Transactions on Medical Imaging*, 10, No. 4, December 1991.

[HYR91b] M. Hedley, H. Yan, and D. Rosenfeld. Motion artifact correction in MRI using generalized projections. *IEEE Transactions on Medical Imaging*, 10, No. 1, March 1991.

[IH93] M. F. Insana and T. J. Hall. Methods for estimating the efficiency of human and computer observers in ultrasonography. In *1993 IEEE*, 1993.

[JS86] P. F. Judy and R. G. Swensson. Size discrimination of feature on CT images. *SPIE Proc.*, 626, 1986.

[Jud92] M. D. Judd. A simple, low-computation peak detection algorithm for the angle-of-arrival spectrum for signal subspace methods. *IEEE Transactions on Aerospace and Electronic Systems*, 28, No. 4, October 1992.

[LBC+92] Z. Liang, F. E. Boada, R. T. Constable, E. M. Haacke, P. C. Lauterbur, and M. R. Smith. Constrained reconstruction methods in MR Imaging. *Reviews of Magnetic Resonance in Medicine*, 4, 1992.

[Ler93] R. A. Lerski. Trial of modifications to Eurospin MRI test objects. *Magnetic Resonance Imaging*, 11, 1993.

[LLG92] H. Lopez, M. H. Loew, and D. J. Goodenough. Objective analysis of ultrasound images by use of a computational observer. *IEEE Transactions on Medical Imaging*, 11, No. 4, December 1992.

[LR93] M. L. Lauzon and B. K. Rutt. Generalized k-space analysis and correction of motion effects in MR images. *Magnetic Resonance in Medicine*, 30, 1993.

[MB91] J. P. Mugler and J. R. Brookman. Evaluation of a simple method for reconstructing asymmetrically sampled echo data. *Journal of Magnetic Resonance Imaging*, 1, No. 4, July/August 1991.

[McG91] G. McGibney. *Phase sensitive reconstruction of MR images.* MSc. thesis, University of Calgary, 1991.

[McN72] D. McNicol. *A Primer of Signal Detection Theory.* George Allen & Unwin Ltd., Great Britain, 1972.

[Met86] C. E. Metz. ROC methodology in radiologic imaging. *Investigative Radiology*, September 1986.

[MHN+94] S. Matej, G. T. Herman, T. K. Narayan, R. M. Lewitt, and P. E. Kinahan. Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms. *Physics of Medical Biology*, 39, 1994.

[MM82] P. Mansfield and P. G. Morris. *NMR Imaging in Biomedicine.* Academic Press, 1982.

[MML92] T. R. McCauley, S. McCarthy, and R. Lange. Pelvic phased array coil: image quality assessment for spin-echo MR imaging. *Magnetic Resonance Imaging*, 10, No. 4, 1992.

[MPS+93] T. Mitsa, K. J. Parker, W. E. Smith, A. M. Tekalp, and J. Szumowski. Correction of periodic motion artifacts along the slice selection axis in MRI. *IEEE trans. on Medical Imaging*, 9/3, 1993.

[MSNC93] G. McGibney, M. R. Smith, S. T. Nichols, and A. Crawley. Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. *Magnetic Resonance in Medicine*, 30, No. 1, January 1993.

[PCG93] J. S. Petersson, J. O. Christoffersson, and K. Golman. MRI simulation using the k-space formalism. *Magnetic Resonance Imaging*, 11, 1993.

[PY92] D. C. Popescu and H. Yan. MR image compression using iterated function systems. *Magnetic Resonance Imaging*, 11, 1992.

[Reh91] M. M. Rehani. *Physics of Medical Imaging.* MacMillan India Ltd., 1991.

[Rie94] J. Riek. *Modeling, detection and suppression of motion artifacts in magnetic resonance imaging.* PhD thesis, University of Rochester, 1994.

[RTS92] J. Riek, A. M. Tekalp, and W. E. Smith. Modeling and correction of artifacts due to z-motion in 2D MRI. In *1992 IEEE*, 1992.

[RW88] V. M. Runge and M. L. Wood. Fast imaging and other motion artifact reduction schemes: a pictorial overview. *Magnetic Resonance Imaging*, 6, 1988.

[SC94] M. R. Smith and A. Crawley. Evaluation methods for MRI post-processing techniques. submitted to IEEE Transactions on Medical Imaging, 1994.

[SC95] M. R. Smith and A. Crawley. Protocols for MRI post-processing techniques to avoid unintentional evaluation bias. submitted to IEEE Transactions on Medical Imaging, 1995.

[Ses94] A. Sesay. Adaptive filtering theory. Class notes, 1994.

[Sez90] M. I. Sezan. A peak detection algorithm and its application to histogram-based image data reduction. *Computer Vision, Graphics, and Image Processing*, 49, 1990.

[SL90] P. Santago and K. M. Link. Restoration of cardiac magnetic resonance images. In *1990 IEEE*, 1990.

[SMC+93] M. R. Smith, G. McGibney, A. Crawley, Y. Hui, and S. T. Nichols. Systematic errors associated with phase correction algorithms used in MRI. Submitted to Magnetic Resonance in Medicine, 1993.

[Smi85] M. R. Smith. Interpolation, differentiation, data smoothing, and least squares fit to data with decreased computational overhead. *IEEE Transactions on Industrial Electronics*, 32, No. 2, May 1985.

[SN83] M. R. Smith and S. T. Nichols. Improved resolution in the analysis of multicomponent exponential signals. *Nuclear Instruments and Methods*, 205, 1983.

[SNC91] M. R. Smith, S. T. Nichols, and A. Crawley. The subjective effect of noise on MRI images reconstructed from short data sets. Submitted to IEEE Transactions on Medical Imaging, 1991.

[SNHW86] M. R. Smith, S. T. Nichols, R. M. Henkelman, and M. L. Wood. Application of autoregressive moving average parametric modeling in magnetic resonance image reconstruction. *IEEE Transactions on Medical Imaging*, MI-5, No. 3, September 1986.

[SZC94] M. Smith, J. Zeng, and A. Crawley. A moving target Evaluation algorithms for removing MRI motion artifacts. In *Proceedings of ICIP-94*, 1994.

[TOS+94] L. Tang, M. Ohya, Y. Sato, S. Tamura, H. Naito, K. Harada, and T. Kozuka. MRI artifact cancelation algorithms for 2-D translational motion. In *1994 IEEE*, 1994.

[Wag86] R. F. Wagner. Fundamentals and applications of signal detection theory in medical imaging. *SPIE Proc.*, 626, 1986.

[Wha71] A. D. Whalen. *Detection of Signals in Noise.* Academic Press, New York, 1971.

[WMB+89] R. F. Wagner, K. J. Myers, D. G. Brown, M. J. Tapiovaara, and A. E. Burgess. Higher-order tasks: Human vs. machine performance. *SPIE Proc.*, 1090, 1989.

[Woo86] M. L. Wood. *Respiratory motion artifacts in magnetic resonance imaging.* PhD thesis, University of Toronto, 1986.

[WRBY92] H. Wang, D. Rosenfeld, M. Braun, and H. Yan. Compression and reconstruction of MRI images using 2D DCT. *Magnetic Resonance Imaging*, 10, 1992.

[WSK88] Wehrli, Shaw, and Kneeland. *Biomedical Magnetic Resonance Imaging.* VHS Publishers Inc., 1988.

[WWJH89] V. J. Wedeen, R. E. Wendt, and M. Jerosch-Herold. Motion phase artifacts in Fourier transform MRI. *Magnetic Resonance in Medicine*, 11, 1989.

[XH93] Q. Xiang and R. M. Henkelman. K-space description of MR imaging of dynamic objects. *Magnetic Resonance in Medicine*, 29, 1993.

[Yan93] J. Yang. *Alternative MRI reconstruction techniques using modeling.* MSc. thesis, University of Calgary, 1993.

[YS96] J. Yang and M. R. Smith. Constrained and adaptive ARMA modeling as an alternative to the DFT – with application to MRI. *Control and Dynamic Systems*, 77, 1996.

[Zak90] A. Zakhor. Ghost cancellation algorithms for MRI images. *IEEE Transactions on Medical Imaging*, 9, No. 3, September 1990.

[Zen94] J. Zeng. Adaptive equalization using LMS, RLS and LSL algorithms. Project report, 1994.

[ZFW+93] L. H. Zang, J. Fielden, J. Wilbrink, A. Takane, and H. Koizumi. Correction of translational motion artifacts in multi-slice spin-echo imaging using self-calibration. *Magnetic Resonance in Medicine*, 29, 1993.

[ZSC94] J. Zeng, M. Smith, and A. Crawley. Motion artifact removal using ARMA modeling. In *Proceedings of SMR-94*, 1994.

[ZST+95] R. A. Zoroofi, Y. Sato, S. Tamura, H. Naito, , and L. Tang. An improved method for MRI artifact correction due to translational motion in the imaging plane. *IEEE Transactions on Medical Imaging*, 14, No. 3, September 1995.