USC–SIPI REPORT #213

Direct and Statistical Approaches to Tomographic Image Reconstruction

by

Xiao–Hong Yan

August 1992

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 400
Los Angeles, CA 90089-2564 U.S.A.

To my dear parents

Acknowledgements

I would first of all like to express my sincere gratitude to my advisor, Richard Leahy, from whom I have received encouragement and support, excellent guidance and true friendship throughout the duration of my Ph.D. study. These were, not only the most beneficial part of my education, but also the most enjoyable ones. I would like to thank the other members of my committee, Alexander Sawchuk and Zdenek Vorel, for their suggestions and comments on my work. I would also like to thank Rama Chellappa and Jay Kuo for serving as members of my guidance committee and valuable advice on several occasions.

My stay at SIPI has been an enriching experience. I am grateful to the SIPI faculty, staff and students for helping me and supporting me in many different ways. Among them, I wish to express my special thanks to Allen Weber, who has patiently helped me solve computer related problems on numerous occasions. I am grateful to him and to Toy Mayeda for maintaining the SIPI computer systems in an excellent condition. The staff of SIPI has been very helpful and patient. I appreciate the assistance from Linda Varilla, Mayitta Penoliar, Gloria Bullock and Delta Tan.

The graduate students at SIPI have been a warm, kind and wonderful group to share various types of experiences, I feel very fortunate to have them around me. I wish to thank Tom Hebert, Ze'ev Lichtenstein and Zhenyu Wu for listening to my complaints and helping me handle many difficult situations. I would also like to thank the other students and ex-students of SIPI, Stephanie Sandor, John Mosher, Rae Lee, Sam Song, Michael Spencer, Erkan Mumcuoglu, Zhenyu Zhou, Richard Hansen, Ted Broida, B.S.Manjunath, Qinfen Zheng, Tal Simchony, Anand Rangarajan, Yitong Zhou, Chandra Shekhar, Gem-Sun Young, David Feldman, Mike Rude, George Zunich, Ching-Yih Tseng, Jian-Ming Wang,

Contents

List of Figures

List of Tables

Abstract

The problem addressed in this research is three dimensional (3-D) tomographic image reconstruction using direct and statistical approaches. In the former approach, our focus is the reconstruction of a 3-D object from its cone beam projections with potential applications in cone beam 3-D X-ray computed tomography (CT) and single photon emission computed tomography (SPECT). In the latter approach, we restrict our attention to the applications of positron emission tomographic (PET) image reconstruction, although many of the techniques described here can also be applied to SPECT.

A general formula for image reconstruction from cone beam data is derived by modifying a result due to Kirillov. Applying this formula to various cone beam geometries results in a class of filtered backprojection algorithms. This formula is known to lead to exact reconstructions in cases in which the cone vertices form certain unbounded curves. An example of such a curve is an infinite straight line. In the case where the curve is a circle, this formula leads to the well known Feldkamp algorithm, for which the reconstructions are only approximations to the true image. We apply this general formula to the cases where the curve is an ellipse and a spiral and new algorithms are derived. For the approximate inverse, we derive the spatially varying point spread function (PSF), which should be useful in the design of cone beam imaging systems. The properties of these algorithms are investigated through studies of the system point spread function and reconstructions of computer generated phantom data.

The statistical approach to PET image reconstruction described here follows a Bayesian formulation. The PET image is defined on a two dimensional lattice as a collection of random variables representing the mean positron emission rate from the elemental volume (pixel) surrounding each lattice site. Between each

pair of pixels in the image, a binary line process is introduced to model the presence or absence of a discontinuity in the image. The image and its associated line process are jointly modeled as a Markov random field (MRF) with a joint Gibbs distribution chosen to favor the formation of images consisting of locally smooth, connected regions. We describe a maximum *a posteriori* (MAP) estimation algorithm based on the generalized EM algorithm for reconstructing the PET image using the above model. The incorporation of a line process in the image model also provides a useful mechanism for the introduction of strong *a priori* information obtained from high resolution registered anatomical magnetic resonance (MR) images. Through a boundary finding process, we can detect anatomical boundaries from these MR images and introduce them as fixed prior line sites in the PET estimation algorithm. The potential performance of the method is tested using a 3-D brain phantom and patient data.

Chapter 1

Introduction

1.1 Motivation

Tomography is a non-invasive technique through which the internal characteristics of a three dimensional (3-D) object can be studied in a visual way, and is widely used as a diagnostic tool in medicine and industry. Various types of sources, such as acoustic waves, radio waves, visible and invisible light and X-rays are employed in tomography depending on the applications. Our attention in this research will be restricted to X-ray transmission and positron or γ-ray emission tomography for diagnostic medical imaging.

The transmission and emission term used here simply indicate where the source originated. In transmission computed tomography (TCT), a radiation source is located *outside* the subject and transmits through the subject. In emission computed tomography (ECT), a chemical compound with the desired biological activity labeled with a radioisotope is administered into the subject under study, the radioactive decay produces a single photon or photon pair emitting from the *inside* of the subject. Those photons or photon pairs are then collected by a detector or detector pair.

The conventional way to reconstruct the entire 3-D subject is in the stacked 2-D fashion. A 3-D reconstruction is formed by imaging successive cross sections of the subject, using 2-D reconstruction algorithms and then stacking the resulting reconstructions. For each cross section of the subject, a detector strip is placed

around this slice to collect the measurements, which correspond to either parallel or fan shaped line integrals. A more efficient and attractive approach is in the truly 3-D fashion, *i.e.* a generalization of the 2-D parallel or fan beam reconstruction algorithms. A detector plane is used to measure the data, representing 3-D parallel or cone shaped line integrals.

The major emphasis of TCT is anatomical imaging, thus the spatial resolution and sensitivity to different types of tissue are very important. TCT usually uses X-ray sources with high photon flux. The reconstruction approach for such a case is often based on the assumption that the measured data is noiseless, so that the Fourier slice theorem and inverse Radon transform or other inverse formulae play very important roles in the development of this approach.

A cone beam data acquisition system provides great potential to obtain reconstructions with better resolution, sensitivity and faster data collection time than the other systems. Usually it is impossible to handle the information across different transaxial planes in stacked 2-D fashion. Therefore, reconstruction from cone beam data cannot be treated in the same way as the 3-D parallel beam problem, *i.e.* as a set of independent 2-D problems. This is because the theory developed for fan beam and 2-D and 3-D parallel beam does not offer a totally adequate means for inverting cone beam data. New theories or formulae [31] [57] [10] [43] [54] [17] have answered many theoretical questions from different point of views, such as the stability of the inversion and sufficient conditions on the curve of cone beam vertices, and therefore led many advances in cone beam tomography. In practice, closed form inverses can be found for many geometries [9] [66] [6], however accurate numerical implementation of these inverses is computationally expensive. In other words, there are no existing algorithms that are both exact and computationally inexpensive. In addition, in many situations where the inverse is only an approximation, it would be desirable to quantitatively evaluate the performance of the algorithm. Although considerable effort has been directed at developing accurate and numerically efficient inversion techniques, in order to fully realize the potential of cone beam imaging systems, a better understanding of the cone beam inverse problem is required.

ECT is a technique for *in vivo* measurement of a wide variety of functional parameters, for example, local blood flow and local glucose metabolism of an organ. Data acquisition time therefore must be short, which constitutes the great limitation as well as potential for ECT in clinics. Each photon (pair) in ECT must be detected and analyzed. This results in an important difference between TCT and ECT: in general, the statistical variability of measured data in ECT is much higher than that in TCT. Thus, one of the important and difficult tasks in ECT is to improve the quantitative accuracy and temporal resolution of reconstructed images using statistical methodologies and estimation procedures.

A maximum likelihood (ML) reconstruction approach by Shepp and Vardi [49] has pioneered the development of statistical approaches to ECT image reconstruction. Because of the ill-conditioned nature of the ML approach to image reconstruction, a variety of statistical approaches have been developed over the years to overcome this problem. These generally impose additional constraints on the solution, attempt to impose the structure on the solution through a regularizing function or introduce prior information to describe the statistical properties of the solution through the use of Bayes theorem. By far the most popular, and that used in this thesis, is the maximum *a posteriori* (MAP) estimator in which the solution is chosen as that image maximizing the posterior probability. An important motivation for using the MAP estimator is the Markov random field (MRF) model [2], which captures a property common to almost all images, that is, there is a high degree of spatial correlation in image intensity between pixels and their neighbors. But MAP estimators employing MRFs may in many cases produce over-smoothed reconstructions particularly across true boundaries in the image. This problem may be alleviated to some extent by choosing the smoothing term to penalize intensity differences only up to a certain threshold [14]. A more satisfying approach to this problem, is the use of line processes in the MRF model [13]. In this model, a binary line process is introduced between each pair of pixels in the image to model the presence or absence of a discontinuity in the image.

Overall, ECT itself has very limited ability to produce accurate anatomical information due to poor spatial resolution and poor statistics in the measured data. Magnetic resonance (MR) imaging and X-ray CT are capable of producing high

resolution images of human anatomy but little functional information. Therefore, these imaging modalities provide complementary diagnostic information. While they produce images of different parameters, they should share the common underlying anatomical structures and boundaries. Inclusion of *a priori* anatomical information should offer great potential to enhance ECT image quality and is desirable in the regional quantitative analysis of ECT.

According to the above discussions, the goal of this research is to develop models and approaches to accurate, quantitative and efficient reconstruction. This research is directed at the following areas:

1. develop general and efficient cone beam reconstruction algorithms which produce acceptable results for many geometries.

2. develop a methodology to analyse the approximation inherent in these algorithms.

3. develop priors which favor local smoothness as well as discontinuity where appropriate and facilitate numerical computation of MAP reconstructions.

4. develop computationally efficient algorithms for more desirable MAP reconstructions.

5. enhance quantitative accuracy and temporal resolution of PET by incorporating anatomical structures.

1.2   Organization of the Dissertation

This dissertation includes eight chapters and is organized as follows. Chapter 1 states the motivation and goals for this research. This is followed by the organization of this thesis and a statement of original contributions made by this research. In Chapter 2, physics of tomography and observation models are presented. This includes the case of transmission and emission, and maximum likelihood (ML) reconstruction using the EM algorithm for emission computed tomography. Chapter

3 reviews the existing literature dealing with related topics, including the theoretical development and practical inverse techniques for direct 3-D image reconstruction and statistical approaches to image reconstruction and their application to emission computed tomography. Chapters 4 and 5 develop an approximate formula for the reconstruction of a 3-D object from various bounded cone beam geometries and apply this formula to circular, elliptical and spiral geometries resulting in a class of filtered backprojection algorithms. Although the algorithms are only approximate, they are shown to produce acceptable results in many cases. The spatially varying point spread function (PSF) is derived to analyse the approximation inherent in the algorithms. The properties of the algorithms are investigated through studies of the system PSF and reconstructions of simulated data. Chapters 6 and 7 deal with a Bayesian approach to image reconstruction for PET. The image is modeled using a joint Gibbs distribution of emission intensities and line processes, which are introduced to model the presence or absence of a discontinuity in the image. A maximum *a posteriori* (MAP) reconstruction algorithm based on the generalized EM algorithm is developed using the above model, it is shown to produce a (local) optimum solution to this problem. Results are presented from computer simulated phantom, 3-D brain phantom and patient data. The phantom data is used to explore the potential performance of the algorithms in a wide variety of test cases, the real data is necessary to validate the overall approach. Finally, Chapter 8 contains conclusions on the research presented and possible directions for further research.

1.3 Original Contributions

The contributions made by this research are summarized in the following:

1. By modifying a result due to Kirillov, a general inversion formula is derived for the reconstruction of a 3-D object from a set of cone beam projections. This formula is known to lead to exact reconstruction in cases where the cone vertices form certain unbounded curves. Therefore, for any bounded curve, the resulting inverse is only approximation to the true image.

5

2. Using this general formula, it is possible to derive a 1-D filtering and 3-D backprojection algorithm for cone beam tomography for a wide range of imaging configurations. We have investigated the application of this formula to the case of circular, elliptical and spiral orbits, and new algorithms are derived. In the case where the curve is a circle, the algorithms is shown to be essentially equivalent to the well known Feldkamp algorithm. The appeal of this approach is that, although the reconstruction operator is only approximate, the algorithms are of the filtered backprojection type and hence computationally very efficient, and produce acceptable reconstructions in many cases. The derivation given here allows a detailed analysis of the inherent approximation associated with the algorithms.

3. The resulting spatially varying point spread function (PSF) is derived to quantitatively evaluate the performance of the new algorithms. It is shown that in the absence of the bandlimiting, the PSF is singular on the surface formed by the set of all lines that pass through the point source and the curve, and the values of the PSF are identical zero inside the surface. The PSF should be useful in determining an upper bound on the resolution of a proposed cone beam imaging system.

4. An alternative energy function is presented for a Gibbs prior distribution, where in addition to pixels representing image intensities, a binary line process is introduced between each pair of pixels in the image to model the presence or absence of a discontinuity in the image. The energy function is designed to favor the formation of images consisting of locally smooth, connected regions and to facilitate numerical computation of the proposed MAP algorithm. Two types of Gibbs energy functions are proposed for the first order and second order neighborhood to cope with different degrees of image complexity.

5. A maximum *a posteriori* (MAP) estimation algorithm is developed based on the generalized EM algorithm for reconstructing PET image using a joint Gibbs distribution of intensity and line processes. This algorithm, using the suggested energy function, automatically guarantees the non-negativity of

image intensity values. In developing the algorithm, a parameter T is used to allow the values of a binary line site to be on the continuous interval $[0,1]$ but eventually converge to either 0 or 1. Through the use of the parameter T, the rate of convergence of the unknown line process can be controlled relative to the intensity process and hopefully the algorithm then converges to a more desirable local optimum.

6. The use of a line process in the MRF model also provides a useful mechanism for the introduction of strong *a priori* information obtained from high resolution registered anatomical magnetic resonance (MR) images. Through a boundary finding process, we extract anatomical boundaries from these MR images and incorporate them as fixed prior line sites in the PET estimation algorithm. The incorporation of information from other imaging modalities offers great potential to improve the quality of PET images. The overall performance of the proposed approach is tested using patient MR and PET data.

Chapter 2

Observation Models for Tomographic Images

This chapter presents the observation models that will be used in the following chapters. Transmission tomography has its root in X-ray imaging, though other types of sources, for example γ-rays, have been used. Emission tomography includes single photon emission computed tomography (SPECT) and positron emission tomography (PET). The word *single* in the former refers to the product of the radioactive decay, a single photon, while in positron emission tomography the decay produces a single positron, this positron almost immediately annihilates with a nearby electron to form two photons, traveling in almost opposite directions. There are important physical and biological differences between emission tomography and transmission X-ray tomography. The basic physical difference is that in emission tomography the information sought is the intensity distribution of the emitted photons in an attenuating medium, in transmission X-ray tomography the attenuation coefficient is sought. The basic biological difference is that emission tomography is a technique for determining the dynamic changes of biochemical activities but little anatomical information, while X-ray CT produces high resolution of anatomical images with little functional information.
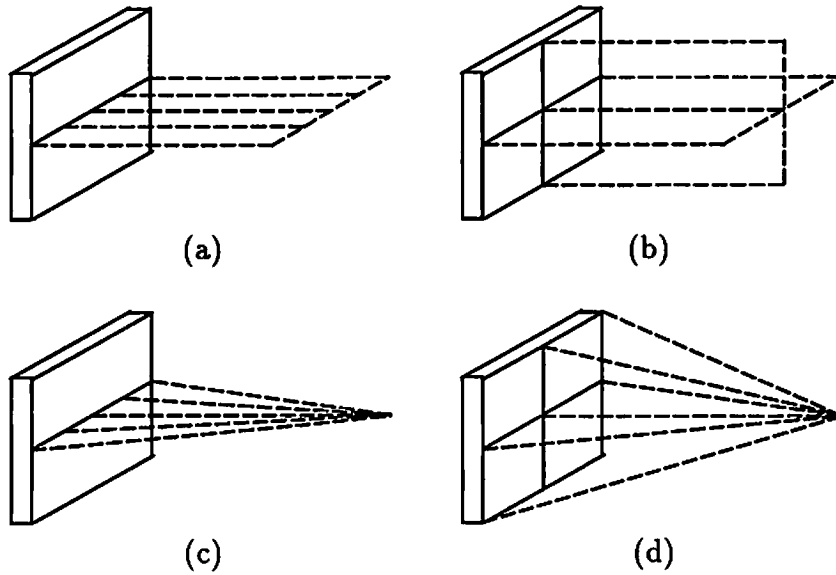
Figure 2.1: Data acquisition geometries for a single 2-D cross section through an object or a 3-D region of interest of an object. (a): 2-D parallel beam. (b): 3-D parallel beam. (c): 2-D fan beam. (d): 3-D cone beam.

2.1 Transmission Tomography

X-ray Computerized Tomography (CT) is an imaging modality producing high resolution images of the human anatomy. In X-ray CT, an attenuation density function of the entire 3-D object or a slice of the object can be reconstructed from the measurements of attenuated X-rays.

In X-ray CT, a X-ray source is placed on one side of a subject, and a detector is located on the other side of the subject. Consider parallel beams of X-ray photons propagate through a cross section of the subject, see Figure 2.1(a). These beams, attenuated due to the photons being absorbed and scattered from their original traveling directions, are measured on a array of detectors located at the opposite side of the source. Then the X-ray source rotates synchronously with the array of detectors to measure another set of data from a different direction. This process is repeated until measurements have been taken from all directions in the plane. The X-ray source can be collimated so that the measurements are collected using either 2-D fan beams or 3-D parallel or cone beams, as illustrated in Figure 2.1.
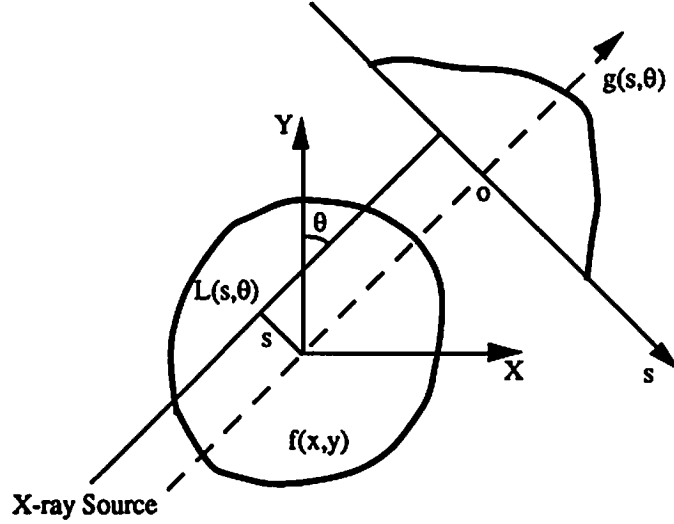
Figure 2.2: Mathematical description for X-ray CT scanning.

For accurate reconstructions of emission data, transmission PET or SPECT is required to correct for the effect of attenuation in emission computed tomography (ECT). For example, during a transmission PET scan, a patient is positioned as for the PET scan, but with no internal positron activity present. An external transmission positron source is placed inside a ring source holder and transmission PET data are collected. Then the *a priori* attenuation information can be used to modify the measured data or can be introduced into an iterative reconstruction algorithm in a systematic fashion.

Mathematically we may describe the problem as follows. Consider a fixed plane through the body (the concepts are straightforward to extend to 3-D), see Figure 2.2. Let $f(x,y)$ denote the attenuation coefficient at a point (x,y) and $L(s,\theta)$ be a line inclined at an angle θ from the Y-axis and at a distance s from the origin. Suppose a thin X-ray beam is directed into this slice along $L(s,\theta)$ and the X-ray beam consists of monoenergy photons, then the intensity of the attenuated X-ray beam measured on a detector has the following form:

$$I = I_0 \exp\left\{ -\int_{L(s,\theta)} f(x,y)du \right\}, \tag{2.1}$$

where I_0 is the intensity of the incident beam and u is the distance along $L(s,\theta)$. Note that when the photon detection is discrete and becomes counting processes,

the variable I thus follows a Poisson distribution. Taking a natural logarithm of the ratio I/I_0 yields the measurements for X-ray CT:

$$g = \ln \frac{I}{I_0},$$

thus from (2.1) we obtain the linear transformation

$$g \equiv g(s, \theta) = \int_{L(s,\theta)} f(x, y) du, \quad \infty < s < \infty, 0 < \theta < \pi, \qquad (2.2)$$

therefore, the image reconstruction problem is to determine $f(x, y)$ from $g(s, \theta)$.

The formula (2.2) is only an approximation even for the X-ray transmission case (a): it ignores statistical fluctuation due to the limited number of photons actually transmitted during each measurement and (b): other sources of error such as the detection of scattered radiation and the assumption of monoenergy X-ray beams.

In the transmission PET or SPECT case, the emitted photons are of monoenergy, however, the major limitation is that the small number of photons is detected. Therefore the statistical fluctuation of the data can no longer be neglected. In this case and X-ray CT case where the photon flux is low and the photon detection becomes discrete, collecting photons becomes counting processes. Thus the variables both I and I_0 in the equation (2.1) follow a Poisson distribution, the discretized version of (2.1) can be written as

$$I_k = I_{0k} \exp \left\{ -\sum_{i,j} P_k(i,j) f(i,j) \right\}, \qquad (2.3)$$

where Poisson random variables I_{0k} and I_k are the incident and detected photons in the k-th beam respectively. $P_k(i,j)$ denotes the length that the k-th beam intersects a pixel at location (i,j). The $f(i,j)$s are the attenuation coefficients to be estimated. In summary, the log likelihood over all X-ray beams can be reduced to

$$L(F|I) = \sum_k \left\{ -\mathbf{E}\{I_{0k}\} e^{-\sum_{i,j} P_k(i,j) f(i,j)} + I_k \ln \mathbf{E}\{I_{0k}\} - \right.$$

$$I_k \sum_{i,j} P_k(i,j) f(i,j) - \ln I_k! \Bigg\},$$

where $\mathbf{E}\{I_{0k}\}$ represents the mean of incident photon beam I_{0k}, F and I are vectors whose components are the $f(i,j)$ and I_k respectively. A ML reconstruction for TCT using EM algorithm is introduced in [34].

The mapping from $f(x,y) \rightarrow g(s,\theta)$ in (2.2) is known as the Radon transform. Radon showed that if $f(x,y)$ is continuous and has compact support, then $f(x,y)$ is uniquely determined by the values of $g(s,\theta)$ for all lines $L(s,\theta)$, *i.e.* the inverse Radon transform:

$$f(x,y) = \left(\frac{1}{2\pi^2}\right) \int_0^\pi \int_{-\infty}^\infty \frac{[(\partial g/\partial s)(s,\theta)]}{x\cos\theta + y\sin\theta - s} ds d\theta.$$

This transform pair plays a very important role in the development of X-ray CT. The projection slice theorem, a fundamental result of Fourier theory, leads to several useful algorithms for obtaining the inverse Radon transform. A conceptually simple method is the Fourier transform based reconstruction, which directly follows from the projection slice theorem. Among the various algorithms developed over the years, the filtered or convolution backprojection method is the most widely used for both parallel and fan beam data.

In theory, one can easily employ the 3-D projection theorem or Radon transform to directly reconstruct a 3-D attenuation density function of an object from a set of cone beam projections. However, the algorithms based on these theorems are not viable in practice, because it is required that measurements be taken from all directions over a sphere. A motivation for the approach discussed in Chapter 4 and 5 is to develop an approximate solution to this problem, which produces fast and acceptable results with the minimum amount of measurements.

2.2 Emission Tomography

2.2.1 The Physics of SPECT
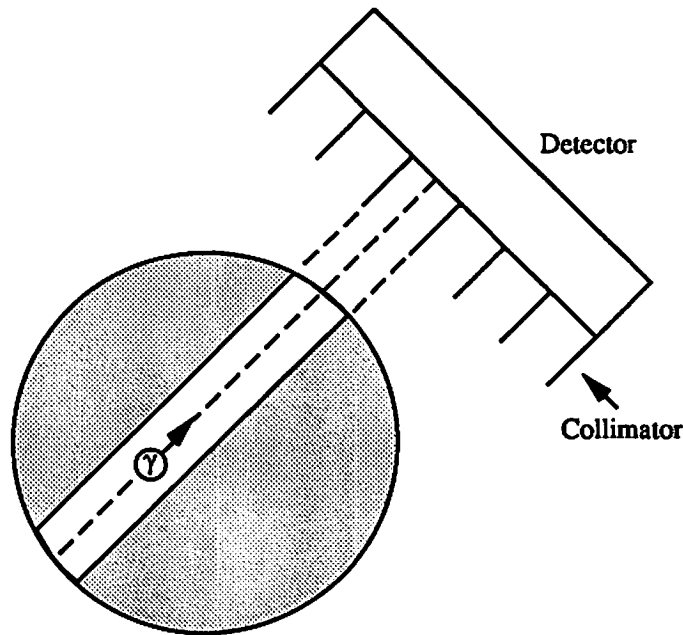
Figure 2.3: Illustration of a cross section of a SPECT imaging system with parallel collimators. The photons detected at each detector are proportional to the summation of the radioactive decays occurred along the strip determined by the collimator.

SPECT is used to determine the spatial distribution of an injected or inhaled chemical compound with some desired biological activity tagged with a γ-ray emitting isotope, called a radiopharmaceutical. The radioactivity decays by emitting a single photon. This photon penetrates the subject and is detected by a strip of scintillation crystals equipped with parallel collimators, see Figure 2.3. The basic purpose of a collimator is to exclude from the detector those photons which are traveling from directions other than a straight line perpendicular to the detector. Due to collimation, when a detection occurs, it is evident that the photon originated somewhere along the line perpendicular to the detector and passing the point of detection. Thus, photons measured at the detector array correspond to a series of parallel line integrals. By altering the direction of collimators into fan shape, the measurements then represent a series of fan shaped line integrals. If a planar detector is used to collect photons, the pattern of the collimators will become either 3-D parallel or cone shaped.
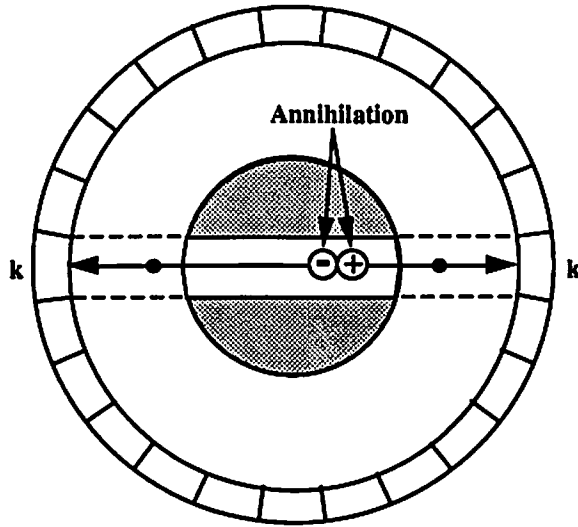
Figure 2.4: Illustration of a PET imaging system with a single ring detectors. Coincidences detected by a detector pair k are proportional to the total number of photons emitted along the strip determined by the detector pair k.

2.2.2 The Physics of PET

Contrasting SPECT, in which a single photon and a collimator are employed to position the location of radioactive decay, is PET. In PET, the chemical compound introduced into the subject is labeled with a positron emitting radioisotope. An emitted positron almost immediately combines with an nearby electron, and the two are annihilated with the emission of two $511kev$ photons. These two photons travel in almost opposite directions, penetrate the surrounding tissue and are detected in coincidence outside the subject by a detector pair, Figure 2.4. From the coincidence detection, one can only determine that an annihilation occurred somewhere along the line defined by the two incident points of photon pair at the detectors

Similar to SPECT, in a single ring detectors of PET, measured coincident photons correspond to approximations to either parallel or fan shaped line integrals. In the case where multiring detectors are employed in PET, coincident photons detected may correspond to 3-D parallel or cone shaped line integrals.

In both SPECT and PET, the measurements represent the radioactivity in the subject viewed from different directions. According to the principle of image

14

reconstruction of computerized tomography, estimation of the radioactivity distribution may be formulated as the problem of tomographic image reconstruction.

2.2.3 Mathematical Model

The following development is suitable for both SPECT and PET, only PET terminology is used. The generation of data in a PET imaging system is the result of radioactive decay. The decay of each individual isotopic nucleus is a random Bernoulli event. Let us examine a time interval I over which a distribution of positron emitting isotope changes very little. Denote p the decay probability of each given individual nucleus in some time interval I. Suppose there is a collection of N atoms in a voxel (small cubic volume), the decay of each nucleus is independent of the decay of the others, thus the number of decays (photon pairs emitted from the voxel) d within time I follows the binomial distribution. Hence the probability P of d decays in time I is equal to

$$P(d) = \left(\begin{array}{c} N \\ d \end{array}\right) p^d (1-p)^{N-d}. \tag{2.4}$$

N is typically very large in the case of radioactive decay, so introducing a new variable $\lambda = Np$ and letting $N \to \infty$ yields

$$\begin{aligned} P(d) &= \lim_{N \to \infty} \frac{N!}{(N-d)!d!} \left(\frac{\lambda}{N}\right)^d \left(1 - \frac{\lambda}{N}\right)^{N-d} \\ &= \frac{\lambda^d e^{-\lambda}}{d!} \lim_{N \to \infty} \left(1 - \frac{\lambda}{N}\right)^{-d} \frac{N!}{N^d(N-d)!}. \end{aligned} \tag{2.5}$$

Since the limit in (2.5) is 1, we obtain that $P(d)$ is the Poisson distribution with the mean λ, *i.e.*

$$P(d) = \frac{\lambda^d e^{-\lambda}}{d!}. \tag{2.6}$$

Hence the radioactive decay follows the Poisson distribution with unknown parameter λ. Since radioactive decays or photon emissions from one voxel are independent of those from any other voxel, the number of photon emissions from the source voxels are well modeled as independent Poisson random variables. Then

the joint distribution of the measured data in emission imaging system can be developed from this model.

For the simplification of notation, we restrict our attention to a single transaxial plane of the subject. Imagine that this plane is divided into $N \times N$ small areas (pixels). We use $d(i,j)$ to denote the number of emissions from pixel (i,j), $i,j = 1, \cdots, N$. The observed data collected on a detector ring around the subject is denoted by Y, whose elements are coincidences y_k detected by detector pair k, $k = 1, \cdots, M$. Assuming there are the total number of D detectors on the ring, then the maximum value of M should be equal to D(D-1)/2. From the previous discussion, it is known that $d(i,j)$s are independent Poisson random variables with unknown mean $\lambda(i,j)$, $i,j = 1, \cdots, N$. However it is not clear that how the detected coincidences y_ks are associated with the number of emissions $d(i,j)$s.

In order to find the joint probability of observed data Y by utilizing the Poisson distribution of a radioactive decay at each location (i,j), we define new random variables, $x_k(i,j), k = 1, \cdots, M; i,j = 1, \cdots, N$, to be the number of emissions from pixel (i,j) which are collected in detector pair k. These variables are intermediate and unobserved. Let $P_k(i,j)$ denote the probability that an emission from source point (i,j) is detected at detector pair corresponding to index k. $P_k(i,j)$ is known and determined by the geometry of the imaging system, the location of source point (i,j), the probability of attenuation of photons and the detector normalization. Rewriting the probability for $x_k(i,j) = m$ as

$$P(x_k(i,j) = m) = \sum_{n=m}^{\infty} P(x_k(i,j) = m|d(i,j) = n)P(d(i,j) = n),$$

where $P(x_k(i,j) = m|d(i,j) = n)$ is a binomial distribution parametrized by $P_k(i,j)$, thus we have

$$P(x_k(i,j) = m) = \sum_{n=m}^{\infty} \binom{n}{m} P_k(i,j)^m (1 - P_k(i,j))^{n-m} \frac{\lambda(i,j)^n e^{-\lambda(i,j)}}{n!},$$

which is simplified to

$$P(x_k(i,j) = m) = \frac{(P_k(i,j)\lambda(i,j))^m e^{-P_k(i,j)\lambda(i,j)}}{m!}.$$

So $x_k(i,j)$ is a Poisson random variable with mean equal to $P_k(i,j)\lambda(i,j)$. Note that the observed data y_k is equal to the summation of all $x_k(i,j)$s over i and j, that is

$$y_k = \sum_{i,j} x_k(i,j),$$

and for a fixed k, $x_k(i,j)$s are independent for different i and j, because all $d(i,j)$s are independent. Therefore, the data y_k is the sum of independent Poisson random variables, thus itself is a Poisson random variable with mean equal to

$$\mathbf{E}\{y_k\} = \sum_{i,j} \mathbf{E}\{x_k(i,j)\} = \sum_{i,j} P_k(i,j)\lambda(i,j),$$

the different values of k represent different locations of detector pairs. Since a direction of an emitted photon pair is uniformly and randomly distributed over the 3-D space, all y_ks are independent Poisson random variables with the joint conditional probability

$$P(Y|\lambda) = \prod_k \frac{(\sum_{i,j} P_k(i,j)\lambda(i,j))^{y_k} e^{-(\sum_{i,j} P_k(i,j)\lambda(i,j))}}{y_k!}. \tag{2.7}$$

2.2.4 Maximum Likelihood Estimation Using EM Algorithm

Consider estimating λ by maximizing $P(Y|\lambda)$ or the log-likelihood function $L(\lambda|Y) = \log P(Y|\lambda)$ from the given set of data Y. The summation in the conditional probability (2.7) implies that it would be difficult to solve this problem directly, because it does not possess a locally dependent property. The EM algorithm presented by Dempster *et al.* [8] is a general approach to iterative computation of maximum likelihood or *a posteriori* estimates when the data can be formulated in complete/incomplete fashion. More precisely, the EM algorithm explicitly makes use of the density $P(X|\lambda)$ to find λ such that

$$\max_\lambda P(Y|\lambda),$$

where $P(X|\lambda)$ is the joint probability of $X = \{x_k(i,j)\}$. As first applied to emission tomography by Shepp and Vardi [49], who treat the actual observed data Y as incomplete version of the unobserved data X, termed as complete data, then the EM algorithm for PET is described as

$$\max_{\lambda} \mathbf{E}_X \{\log P(X|\lambda)|\lambda^n, Y\}, \tag{2.8}$$

where λ^n represents current estimates of emission intensities. This algorithm formularizes a heuristic idea for handling incomplete-data, consisting of essential Expectation (E) and Maximization (M) steps at each iteration stage. The E step is to find the conditional expectation of the complete-data X, given the observed data Y and the current estimated parameter λ^n. Using the fact that the distribution of $x_k(i,j)$, conditioned on λ^n and $y_k = \sum_{i,j} x_k(i,j)$, is binomial $\left(y_k, P_k(i,j)\lambda^n(i,j)/\sum_{i',j'} P_k(i',j')\lambda^n(i',j')\right)$, its mean is equal to

$$\mathbf{E}(x_k(i,j)|\lambda^n, Y) = \lambda^n(i,j)\frac{y_k P_k(i,j)}{\sum_{i',j'} P_k(i',j')\lambda^n(i',j')}. \tag{2.9}$$

In the M step, we first replace the complete data by its expectation (2.9) to yield the result $Q(\lambda|\lambda^n)$ as

$$Q(\lambda|\lambda^n) = \sum_{i,j}\left\{-\sum_k P_k(i,j)\lambda(i,j)+ \right. \tag{2.10}$$
$$\left. \lambda^n(i,j)\sum_k \frac{y_k P_k(i,j)}{\sum_{i',j'} P_k(i',j')\lambda^n(i',j')} \log \lambda(i,j)\right\},$$

where those term which are not dependent on λ are omitted, since they will not affect the maximization result. We then maximize $Q(\lambda|\lambda^n)$ to find updated parameters λ^{n+1} by setting the first derivative of $Q(\lambda|\lambda^n)$ with respect to each $\lambda(i,j)$ equal to zero. After simplification, we obtain the well known EM algorithm for ML PET estimation

$$\lambda^{n+1}(i,j) = \lambda^n(i,j)\frac{\sum_k \frac{y_k P_k(i,j)}{\sum_{i',j'} P_k(i',j')\lambda^n(i',j')}}{\sum_k P_k(i,j)}. \tag{2.11}$$

The images obtained in this manner are generally considered qualitatively better than those from the filtered backprojection algorithms. However it is well noted that after several iterations of the EM algorithm the image quality begins to deteriorate due to the ill-conditioned nature of the ML problem in PET. This has motivated this research to investigate the Bayesian approach to this problem through the use of the MRF model with a line process.

Chapter 3

Review of the Literature

There are many techniques especially developed for tomographic image reconstruction. Some are more targeted at minimizing computational complexity; others are more directed at accurate system modeling. In general, they fall into one of the two types of approaches: direct and iterative or statistical inversion.

Both techniques have strengths and weaknesses. Direct inversion approaches were initially developed for X-ray transmission computed tomography, since the statistical fluctuations in the measurements of X-ray are much less severe than those in the emission data. Therefore, when applied to the cases where relatively large number of photons are measured compared with the number of parameters to be estimated, this type of algorithm often leads to satisfactory reconstructions. In addition, it is computationally efficient and straightforward. In emission tomography, the major factor that limits its performance is the small number of photons detected, which results in reconstructed images with large statistical uncertainty and poor resolution.

Statistical approaches mainly deal with the problem of reconstruction in emission tomography, though some effort has been directed at the problem of transmission tomography, for example in [34]. In the algorithm development for emission tomography, the use of statistical approaches has demonstrated potential advantages over the direct reconstruction approach, especially in cases where the total number of detected photons is relatively small. Under this framework, one can

introduce a prior to describe statistical properties of images or naturally incorporate *a priori* information or structure about the spatial distributions of functional activities in a human body. The images obtained in a statistical manner are generally considered qualitatively better, but computationally more expensive than those from direct approaches.

Due to the present computational ability, it is necessary to study both types of techniques. In fact, the reconstruction of PET or SPECT images in most clinics is based on the well known filtered backprojection type of algorithms. Clearly, the computational cost of statistical approaches can be significantly reduced, if the algorithms are implemented in a hardware board of the type currently used for filtering and backprojection in clinical PET or SPECT systems. By then, the statistical techniques may be widely used in clinical uses.

3.1   Direct Reconstruction of A 3-D Object

The recent development of medical imaging systems which acquire cone beam projection data from 3-D objects has motivated research into direct 3-D image reconstruction algorithms analogous to the well known filtered backprojection methods for 2-D image reconstruction from parallel [48] and fan beam [24] projections. Examples of systems for which such an algorithm would be useful include the dynamic spatial reconstructor (DSR) [59], cone beam collimated SPECT [11], [29] and [28], electronically collimated SPECT [52] and the Philips twin cone beam X-ray CT system [32].

Cone beam tomography has potential advantages over standard parallel and fan beam geometries for both X-ray CT and SPECT. In the case of X-ray CT, a cone beam X-ray source and data acquisition system offers the potential for extremely fast 3-D data acquisition, as in the DSR for example [59]. In SPECT, cone beam collimators have been designed for use in brain and cardiac studies that have improved sensitivity and resolution in comparison to parallel or stacked fan beam collimators [29] [19].

Reconstruction from cone beam data cannot be treated as a set of independent two-dimensional problems and is a challenging problem both in theory and in

practice. This is because the theory developed for fan beam and 2-D and 3-D parallel beam does not offer a totally adequate means for inverting cone beam data. A new theory or formula needs to be developed to answer questions, such as the stability of inverses and sufficient conditions on the curve of cone beam vertices. In practice, one must remember that the computational complexity is one of the major concerns. As far as we know, there is no existing algorithm that is both exact and computationally inexpensive. The influence of others factors, such as detector sensitivity or physical constraints may motivate the use of different cone beam geometries.

Kirillov [31] gives a closed form solution for the reconstruction of an n-dimensional complex valued function from its complex line integrals. Rewriting this work for one-dimensional line integrals of a 3-D object results in a general inversion formula for the reconstruction of a 3-D object from a set of cone beam projections, which requires cone beam curves that are unbounded and satisfy certain other conditions [53] [63]. This formula was first given by Smith [53] who also showed that an example of such a curve is an infinite straight line. A reconstruction algorithm based on this result was proposed in which a 3-D coordinate rotation is employed to reduce the computational complexity of reconstruction. While this result yields an exact reconstruction, the arrangement of the cone vertices on a single line is not practically viable.

The reconstruction formulae in [31] and [53] require that the curve formed by the cone vertices is unbounded. For the case where the object is of finite support, Tuy [57], Smith [54] and Grangeat [17] show that the conditions on the curve may be relaxed. They all point out similar sufficient conditions on a bounded curve of source points from which the image may be exactly recovered, and provide closed form solutions for more general curves satisfying these conditions. The reconstruction formula of Tuy is based on the relation between the processed cone beam data and a filtered form of 3-D inverse Radon transform. Although this formula is known to lead to closed form inverses for many geometries, accurate numerical implementation of these inverses is usually computationally expensive. Smith's derivation [54], unfortunately, contains an error which is pointed out by Defrise et al. [7]. Grangeat [17] has proposed a different, but related approach, in

which a fundamental relation is established between the cone beam projections and the first derivative of the Radon transform. Then the derivation of an inverse formula using this relation becomes straightforward, according to the 3-D inverse Radon transform. One drawback of this approach is that it requires the rebinning of cone beam data beforing using the reconstruction algorithm. For curves not satisfying the conditions stated in [57] [54] [17], Finch [10] describes various inversion procedures for functions of compact support based on a Sobolev space setting. These results do not include a closed form solution and do not appear to lead to a computationally feasible algorithm. An alternative approach was taken in [5] for a cone vertex forming a sphere enclosing the object, rather than a 3-D curve that this research is concerned with. Again these results do not extend to our problem. A detailed discussion of the cone beam reconstruction problem and the question of stability are addressed by Natterer [43], however no algorithm is provided for the problem treated in this research.

The most commonly employed geometry for cone beam tomography is the circular orbit. Unfortunately, an exact inverse does not exist for this arrangement. However, several approximate reconstruction methods have been proposed [41], [9], [54], [17] and [63]. By far the most widely used of these, is the "practical cone beam algorithm" of Feldkamp et al. [9]. This algorithm is derived in a heuristic manner by extending the 2-D fan beam reconstruction approach to 3-D reconstruction. This method is of the filtered backprojection type, and thus very computationally efficient, and produces acceptable reconstructions in many cases. Minerbo [41] has used 3-D inverse Radon transform to derive an approximate solution for the geometry considered here, however his method does not result in an efficient reconstruction algorithm. Three approximation schemes for this configuration are discussed in [54], the first of those is shown in [55] to be equivalent to the algorithm of Feldkamp et al., however the algorithm and its properties are not described in detail. Grangeat's method [17] has been applied to the special case of a circular orbit, where the shadow zone concept is used to describe the approximation associated with the algorithm. This method has the ability to fill the first derivative of the Radon transform by interpolation in this shadow zone. In Chapter 4, a general approximate formula is derived by adapting

a result due to Kirillov [31]. The resulting algorithm for this particular geometry is shown to be equivalent to the well known Feldkamp algorithm. However, the derivation given in this research allows a detailed analysis of the spatially variant point spread function (PSF) implicit in the approximation and other properties of Feldkamp's algorithm.

There are many alternative cone beam geometries and reconstruction algorithms. The influence of other factors, such as detector sensitivity or physical constraints or accuracy of reconstructions motivates the use of either the elliptical or the spiral or more complex configurations. These are addressed in Chapter 5. Elliptical orbits are an attractive geometry for cardiac studies using cone beam collimated SPECT, as they allow the camera to closely follow the chest contour. Gullberg *et al.* [19] have extended the Feldkamp algorithm to this noncircular case. An alternative elliptical cone beam reconstruction algorithm is given in [65] using the general inverse formula described in detail in [63]. Unfortunately, no reconstruction algorithms are exact for this geometry.

Tuy [57] and Grangeat [17] described sufficiency conditions on the curve of cone beam vertices, for which inverse operators do exist and derived the corresponding inverses. In most cases however accurate numerical computation of these inverses is computationally expensive. Fast approximate reconstruction algorithms for geometries for which inverses do exist have been described by Clack *et al.* [6] for the case where the curve is a pair of orthogonal, concentric circles, Zeng and Gullberg [66] for the case where the curve is the union of a circle and a perpendicular line, and Yan and Leahy [65] for a finite spiral curve. The approach of Clack *et al.* can be generally described as: (1) converting the cone beam projections into planar integrals; (2) backprojecting the values obtained from the integrals; (3) three-dimensional filtering. The basic idea in [66] is to consider the redundancy in the cone beam data and discard unwanted projection data using a spatially varying filter implemented in the frequency domain and then employ the algorithm proposed in [56]. The appeal of the approach in [65] is that the algorithms are of the filtered backprojection type and hence computationally very efficient, the approximation inherent in the algorithms are investigated through studies of the system PSF.

3.2 Statistical Approaches to Emission Tomography

Emission computed tomography (ECT) is a unique technique to quantitatively determine dynamic changes in biochemical activity in the body. The potential and utility of ECT in diagnostic medicine has attracted many investigators who have continued to explore potential applications. ECT images have improved significantly over the last few years. The improvements resulted from a better understanding of design principles and trade-offs as well as the introduction of novel concepts. Considerable effort has been directed toward accurate, quantitative reconstruction in three dimensions of spatial distributions of radioactivity.

The ECT imaging problem may be formulated as one of tomographic image reconstruction as similar to that in transmission computed tomography (TCT). One of the important differences between ECT and TCT is the statistics availability. ECT imaging systems collect approximately 10,000 times less data per transverse section image than TCT [22]. Thus the assumption that the projection data is noiseless becomes much less valid than that used in TCT. This is one of the main reasons the reconstruction strategy for ECT involves investigation of statistical approaches in order to take into account the Poisson statistics of data. A statistical approach to ECT image reconstruction offers several potential advantages over the filtered backprojection method currently employed in most clinical ECT systems: (1) the true data formation process may be modeled accurately to include the Poisson nature of the observation process and factors such as attenuation, scatter, detector efficiency and randoms; and (2) an *a priori* statistical model for the image may be employed to model the generally smooth nature of the desired spatial distribution and to include information such as the presence of anatomical boundaries, and hence potential discontinuities, in the image.

PET is the focus of the study of emission tomography in this research for two reasons. The first one is that, although SPECT has the advantage of being less expensive, PET imaging has distinct advantages over SPECT, such as relatively high detection efficiency and relatively good spatial resolution. The second one is that some correction factors, for example, attenuation correction, can be included

in a much easier manner in PET than in SPECT. Thus these advantages has made more efficient to explore the potential performance of proposed algorithms in a variety of test cases, and to validate the overall approach using real data. Consequently, the investigation of patient MR and PET data has pointed out the possible directions for further improving the proposed algorithm.

Reconstruction of PET images in most commercial PET systems is based on the well known filtered backprojection (FBP) algorithm [48] initially developed for transmission X-ray computed tomography. Although this reconstruction algorithm is computationally straight forward, the quality of images is severely limited by the failure of the algorithm to take into account the photon limited nature of the data. The publication by Shepp and Vardi [49] of a new method for PET image reconstruction based on a maximum likelihood (ML) formulation sparked considerable interest in statistical approaches to emission computed tomographic image reconstruction. In their seminal paper, Shepp and Vardi develop a statistical model for the data and apply the expectation maximization (EM) method of Dempster *et al.* [8] to maximize the resulting likelihood function. The images obtained in this manner are generally considered qualitatively better than those from FBP. However, it has been widely noted that after several iterations of the EM algorithm the image quality begins to deteriorate due to the ill-conditioned nature of the ML problem in the reconstruction problem. Methods to overcome this problem include early termination of the EM iterations using a statistical hypothesis test [21], [58] and regularization using the method of sieves [40], smoothing EM [50]. All these techniques have been shown to alleviate the ill-posed problem to some extent. The stopping rule [21] [58] has demonstrated a compromise solution between maximizing the likelihood and minimizing the hypothesis. But it is theoretically unappealing since one begins by searching for an optimal (ML) solution, but then terminates the search before reaching this point. Similarly, the method of smoothing [50] introduces a simple smoothing step at each EM iteration. As far as we know, there is no an appealing direct interpretation of this method in terms of a specified optimization problem. The method of sieves [40] effectively solves the problem by restricting the solution space using a set of regularizing functions. While this method is statistically sound, it often suffers from edge artifacts or

over-smoothing due to the choice of sieve kernel. In the following we will restrict our attention to Bayesian approaches to the image reconstruction problem.

Through the use of Bayes theorem, one can introduce a prior distribution to describe the statistical properties of the unknown image and thus produce a posterior probability distribution for the image conditioned upon the data. Maximizing over the posterior probability yields a maximum *a posteriori* (MAP) estimation in which the prior effectively performs a regularizing function and thus MAP estimation avoids the ill-conditioning inherent in ML. Liang and Hart [38] have developed MAP estimators based on the EM algorithm for Gaussian and Poisson priors [38]. Levitan and Herman have also developed a MAP EM algorithm for an uncorrelated Gaussian prior in [37], using a smoothed FBP reconstruction as an estimated mean. A correlated Gaussian prior is designed to enforce local smoothness in [25].

An important class of image models, the Gibbs distribution or Markov random field (MRF) [2], capture a property common to almost all images *i.e.* that there is a high degree of spatial correlation in image intensity between pixels and their neighbors. The MRF models this local interaction by specifying the conditional probability for each pixel as a function only of those pixels within a local neighborhood of that pixel. It has been shown that the joint distribution for an MRF has a special form termed the Gibbs distribution [2] which facilitates both the formulation of problems involving these models and efficient numerical computation of the resulting estimator. The MRF model was first applied in emission tomographic reconstruction by Geman and McClure [14] and has since been widely studied [20], [18], [30], [4], [64], [35], [15] and [36]. Hebert and Leahy [20] have shown that MAP estimation employing an MRF prior, may be implemented using a modified version of the EM algorithm at little additional cost compared to a ML estimator. The One-Step-Late (OSL) algorithm of Green [18] evaluates the energy function of a Gibbs prior at the *current* estimate to facilitate an MAP-EM type approach. The algorithms presented in [4], [64], [35], [15] and [36] employ a joint Gibbs distribution of image intensities and line processes [13] and incorporate registered *a priori* anatomical information. Research involved in the use of a line process is reviewed in the following.

A problem which arises with the use of MRF models is that in choosing a prior which favors strong local correlations, abrupt variations in image intensity are penalized. The resulting effect is that MAP estimators employing MRFs may produce over smoothed reconstructions, particularly in the vicinity of true boundaries in the image. This problem may be alleviated to some extent by choosing the smoothing term to penalize intensity differences only up to a certain threshold [14], [20]. A more satisfying approach to this problem, and one which allows us to naturely incorporate *a priori* anatomical information, is the use of "line sites" in the MRF model [13]. In this model, in addition to the pixels representing the image intensity, one introduces an additional variable between each pair of neighboring pixels - this variable should take values of either "1" to denote the presence of a boundary between the two pixels or "0" to represent the absence of a boundary. One constructs a Gibbs energy function which is a function of both the pixel intensities and the line sites. Thus when computing the interaction terms between pixels, one only considers those pixel pairs for which the associated line site is "0" (i.e. no boundary); in this way one avoids smoothing across boundaries in the reconstructed image. In addition, a penalty function is added to avoid forming too many boundaries. Finally, one introduces a smoothing term to the line sites themselves, to attempt to obtain closed and connected boundaries. This smoothing term also serves the purpose of discouraging redundant boundaries and very small regions. A similar approach was proposed with some success by Johnson *et al.* [30] for PET imaging. Neighborhood system is expanded in their image model to consider the blurring effect in the image. However, in that case no penalty term was employed to limit the formation of too many boundaries. It can be shown that in the absence of such a penalty term, the algorithm [30] will converge to a ML solution. Chapter 6 presents a MAP estimation algorithm, which is based on the generalized EM algorithm using the MRF model with a joint Gibbs distribution. The Gibbs prior with the second-order neighborhood is designed to encourage the formation of image consisting of locally smooth and connected regions. Hopefully the introduction of the parameter T will make the proposed algorithm converge to a better local maximum solution.

Although the use of line processes has produced encouraging reconstructions [30] [64], It has been observed through reconstructions of simulated data that the estimated line sites are dependent on the total number of coincidence detections and even for a large amount of coincidence detections, are sometimes inaccurately located. This is due to both the finite resolution of the simulated PET system, and the random nature of the data. A MAP reconstruction approach, including line site estimation, has the potential for further enhancing the quality of PET reconstructions.

Magnetic resonance (MR) imaging is capable of producing high resolution images of the human anatomy. An appropriate choice of pulse sequence can produce T1 or T2 weighted images with excellent soft tissue contrast. As discussed in Chapter 1, positron emission tomography (PET) [42] is capable of producing somewhat lower resolution images of functional activity through the use of radiolabeled pharmaceuticals. While PET and MR produce images of different parameters, the underlying anatomical structures and the boundaries between them are common to both images. In other words, the spatial distribution of functional activity is dependent on the underlying anatomical structure in which the activity occurs. Thus if one could extract high resolution structural information from the registered MR image, this could be used as important *a priori* information about potential boundaries between structures in the PET image. In the case where registered MR images are available, edge information may be extracted from these images and used as known *a priori* line sites or modeled as an edge map to reflect the confidence about this information. We note that similar work for performing this task has been reported by Chen *et al.* [4], but there is no details of their approach presented. An alternative algorithm using registered MR boundaries as fixed *a priori* information will be discussed in detail in Chapter 7, based on the use of MRF model with additional line processes [35] [36]. The potential performance of the algorithm was demonstrated using computer simulated phantom and 3-D brain phantom PET data. The overall approach is validated using patient MR and PET data. PET and MR images are registered by scalp surface matching [61], in which scalp contours from both ML reconstructions of PET and MR images are used. Our experiences with real data indicate that the

use of the MRF model with a line process is a promising approach to PET reconstructions, and the incorporation of anatomical MR boundaries should offer the potential to enhance the quality of PET images. However, *a priori* boundary information and line sites in the MRF model should be treated differently such that the *a priori* information is only employed as an extra weighted term either to encourage the formation of line sites or to discourage this formation. A more sound approach to dealing with MR boundaries is presented by Gindi *et al.* [15] who modeled MR boundaries as edge maps whose values are the measurements of confidence believe towards those boundaries, though it is not clear that how the values of the map relate to the *a priori* MR boundaries.

From (4.31), b may be expressed in terms of $\mathbf{x}$, $\mathbf{x_0}$ and θ as:

$$b = x_{03}\frac{R\left(\frac{x_3}{x_{03}} - 1\right) + \left(x_1 - \frac{x_{01}}{x_{03}}x_3\right)\cos\theta + \left(x_2 - \frac{x_{02}}{x_{03}}x_3\right)\sin\theta}{R - x_{01}\cos\theta - x_{02}\sin\theta}, \qquad x_{03} \neq 0. \quad (4.35)$$

The following inequality holds for any $\mathbf{x}$:

$$\left|\left(x_1 - \frac{x_{01}}{x_{03}}x_3\right)\cos\theta + \left(x_2 - \frac{x_{02}}{x_{03}}x_3\right)\sin\theta\right| \qquad (4.36)$$

$$\leq \sqrt{\left(x_1 - \frac{x_{01}}{x_{03}}x_3\right)^2 + \left(x_2 - \frac{x_{02}}{x_{03}}x_3\right)^2}.$$

Therefore, if for some $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$,

$$R\left|\frac{\hat{x}_3}{x_{03}} - 1\right| > \sqrt{\left(\hat{x}_1 - \frac{x_{01}}{x_{03}}\hat{x}_3\right)^2 + \left(\hat{x}_2 - \frac{x_{02}}{x_{03}}\hat{x}_3\right)^2}, \qquad (4.37)$$

then: $|b| > 0$, for any $\theta \in [0, 2\pi)$, and consequently

$$\hat{f}_p(\hat{\mathbf{x}}) = 0, \qquad (4.38)$$

since the argument b of the delta function $\delta(b)$ does not pass through zero as we integrate (4.33) with respect to θ.

Equation (4.37) thus describes a set in $\mathbf{R}^3$ over which $\hat{f}_p(\mathbf{x})$ is identically zero. This inequality describes a distorted hourglass shape, formed by a double cone with vertex $(x_{01}, x_{02}, x_{03})^T$, which cuts the curve $\phi(\theta)$ as depicted in Figure 4.2. The set of points satisfying the inequality (4.37) forms the interior of the cones.

The PSF thus has the following form:

$$\hat{f}_p(\mathbf{x}) = \begin{cases} 0 & R\left|\frac{x_3}{x_{03}} - 1\right| > \sqrt{\left(x_1 - \frac{x_{01}}{x_{03}}x_3\right)^2 + \left(x_2 - \frac{x_{02}}{x_{03}}x_3\right)^2} \\[3mm] \frac{R}{4\pi^2}\int_0^{2\pi} \frac{|\dot{a}_2|}{\dot{a}_{02}^2}\delta(b)\lim_{\epsilon\to 0} E_\epsilon(a)d\theta, & \text{otherwise,} \end{cases} \qquad (4.39)$$

where a and b are defined in (4.34) and $E_\epsilon(\cdot)$ in (4.30).

This result provides some interesting insight into the performance of the approximate reconstruction. Since the PSF is zero inside the cone described above,
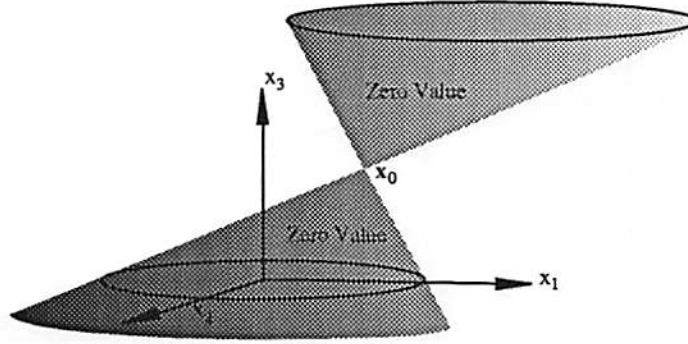
Figure 4.2: Illustration of the double cone with vertex at $\mathbf{x}_0$ whose surface passes through the source curve $\phi(\theta)$. The points in the interior of the cones satisfy the inequality (4.37) and define the set over which the PSF is equal to zero.

the reconstruction should improve as this region becomes larger. From (4.37) it follows that the interior of the cone increases in volume as R is increased. Consequently, one should choose R as large as possible to minimize blurring due to the PSF in (4.39).

The result in (4.39) holds for all $\|\mathbf{x}_0\| < R$ provided $x_{03} \neq 0$. In the case $x_{03} = 0$, it can be shown that $\hat{f}_p(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0)$; this case is discussed in Section 4.3.1. In Section 4.3.2 we return to the PSF and consider the case for a bandlimited data acquisition system

4.3   Approximate Reconstruction from Planar Projection Data

We now assume that the projection data are collected at a planar detector placed outside the circle $\phi(\theta)$. Due to the finite detector resolution, there exists a bandlimiting effect on the data. In this section, we present the reconstruction algorithm for planar data, and consider the effects of finite resolution on the resulting PSF.

4.3.1 Reconstruction by Filtered Backprojection

To conclude the development, the result in (4.29) is expressed in terms of the collected, planar projection data. We can rewrite (4.29) as:

$$\hat{f}(\mathbf{x}) = \frac{R}{4\pi^2} \int_0^{2\pi} \hat{f}_\theta(\mathbf{x}) d\theta, \tag{4.40}$$

where

$$\hat{f}_\theta(\mathbf{x}) = \lim_{\varepsilon \to 0} \left[g_r(\hat{\alpha}, \theta) *_{\hat{\alpha}_1} E_\varepsilon(\hat{\alpha}_1) \right]_{|\hat{\alpha}=M(\mathbf{x}-\phi(\theta))} . \tag{4.41}$$

Assume the projection data $g_1(Z, \theta)$ are collected on the plane $\hat{\alpha}_2 = D$ where D is the distance of the detector from the cone vertex. The arrangement is shown in Figures 4.1 and 4.3. Let $Z = (z_1, z_2)$ denote the planar coordinates, then from the geometry and (4.2):

$$g_1(z_1, z_2, \theta)|_{z_1=\hat{\alpha}_1 D/\hat{\alpha}_2, z_2=\hat{\alpha}_3 D/\hat{\alpha}_2} \equiv g_1(\hat{\alpha}, \theta) = \|\hat{\alpha}\| g_r(\hat{\alpha}, \theta), \tag{4.42}$$

Substituting (4.42) in (4.41) yields

$$\hat{f}_\theta(\mathbf{x}) = \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{g_1(\hat{\alpha}_1' D/\hat{\alpha}_2, \hat{\alpha}_3 D/\hat{\alpha}_2, \theta)}{\sqrt{\hat{\alpha}_1'^2 + \hat{\alpha}_2^2 + \hat{\alpha}_3^2}} E_\varepsilon(\hat{\alpha}_1 - \hat{\alpha}_1') d\hat{\alpha}_1' \Bigg|_{\hat{\alpha}=M(\mathbf{x}-\phi(\theta)).} \tag{4.43}$$

Making the change of variable $z_1' = \hat{\alpha}_1' D/\hat{\alpha}_2$, with $d\hat{\alpha}_1' = |\frac{\hat{\alpha}_2}{D}| dz_1'$, and using the definition of $E_\varepsilon(\hat{\alpha}_1)$ in (4.30):

$$\hat{f}_\theta(\mathbf{x}) = \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{g_1(z_1', z_2, \theta)}{\sqrt{\left(\frac{z_1' \hat{\alpha}_2}{D}\right)^2 + \hat{\alpha}_2^2 + \left(\frac{z_2 \hat{\alpha}_2}{D}\right)^2}} \frac{E_\varepsilon(z_1 - z_1')}{\left(\frac{\hat{\alpha}_2}{D}\right)^2} \left|\frac{\hat{\alpha}_2}{D}\right| dz_1'. \tag{4.44}$$

Noting that $\hat{\alpha}_2 = -x_1 \cos\theta - x_2 \sin\theta + R$ and simplifying (4.44) yields the result:

$$\hat{f}_\theta(\mathbf{x}) = \frac{D^2}{(x_1 \cos\theta + x_2 \sin\theta - R)^2} \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{g_1(z_1', z_2, \theta)}{\sqrt{z_1'^2 + D^2 + z_2^2}} E_\varepsilon(z_1 - z_1') dz_1'. \tag{4.45}$$

Thus $\hat{f}(\mathbf{x})$ is recovered by filtering the weighted projections

$$(z_1^2 + D^2 + z_2^2)^{-1/2} g_1(z_1, z_2, \theta)$$

with $E_\epsilon(z_1)$ and then performing a weighted backprojection.

In summary we have reduced the problem of reconstructing an image from a set of cone beam projections to a two stage procedure:

1. 1-D filtering of weighted planar projection data:

$$\tilde{g}(z_1, z_2, \theta) = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{g_1(z_1', z_2, \theta)}{\sqrt{z_1'^2 + D^2 + z_2^2}} E_\epsilon(z_1 - z_1') \, dz_1'. \qquad (4.46)$$

2. Weighted 3-D backprojection:

$$\hat{f}(\mathbf{x}) = \frac{RD^2}{4\pi^2} \int_0^{2\pi} \frac{\tilde{g}(z_1, z_2, \theta)}{(x_1 \cos\theta + x_2 \sin\theta - R)^2} \, d\theta. \qquad (4.47)$$

We note that (4.46) and (4.47) are very similar to equation (28) in [9] and identical as the allowed bandwidth of the function in equation (28) of [9] tends to infinity. The advantage of the preceding development is that this derivation provides some insight into the inherent approximation associated with the reconstruction algorithm. Furthermore, as shown in the following, it is possible to derive the PSF associated with the approximation for the planar data. As in [9], this result gives exact reconstruction for images confined to the plane of the curve $\phi(\theta)$, i.e. $x_3 = 0$. In this case, reconstruction of plane $x_3 = 0$ only, reduces (4.46) and (4.47) to the (exact) fan beam inversion formula given by equation (49) of [27].

4.3.2 Point Spread Function from Bandlimited Data

We have summarized the PSF in the general form (4.33), assuming infinite bandwidth. Probably of more interest is the PSF for bandlimited data. This bandlimited PSF is derived below. The planar projection data for a point source are

derived in the Appendix as:

$$g_1(z_1, z_2, \theta) = \frac{D}{\hat{\alpha}_{02}^2} \sqrt{z_1^2 + D^2 + z_2^2} \ \delta\left(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}} D\right) \delta\left(z_2 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}} D\right). \qquad (4.48)$$

By substituting (4.48) and (4.46) into (4.47), we can write the PSF for the planar data as:

$$\hat{f}_p(\mathbf{x}) = \frac{RD^3}{4\pi^2} \int_0^{2\pi} \qquad (4.49)$$

$$\frac{1}{(x_1 \cos\theta + x_2 \sin\theta - R)^2 \hat{\alpha}_{02}^2} \ \delta\left(z_2 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}} D\right) \lim_{\epsilon \to 0} E_\epsilon\left(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}} D\right) \ d\theta.$$

Finite detector resolution and the use of sampled data impose finite bandwidth on the reconstructed image. To estimate the PSF in this case, we perform a bandlimiting operation on the data by convolving the planar projection data with a low pass filter. Assuming no aliasing and a sample interval of Δ *cm* in z_1 and z_2, the bandwidth obtained from the Nyquist sampling theorem is $(-\pi/\Delta, \pi/\Delta)$ *rad/cm* in both the z_1 and z_2 directions. To obtain the bandlimited data we convolve the data with the ideal low pass filter impulse response:

$$h(z_1, z_2) = \frac{\sin\left(\frac{\pi z_1}{\Delta}\right) \sin\left(\frac{\pi z_2}{\Delta}\right)}{\pi^2 z_1 z_2}. \qquad (4.50)$$

Thus the bandlimited PSF is given by including (4.50) in (4.49) to obtain:

$$\hat{f}_{wp}(\mathbf{x}) = \frac{RD^3}{4\pi^2} \int_0^{2\pi} \frac{1}{(x_1 \cos\theta + x_2 \sin\theta - R)^2 \hat{\alpha}_{02}^2} \qquad (4.51)$$

$$\delta\left(z_2 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}} D\right) \lim_{\epsilon \to 0} E_\epsilon\left(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}} D\right) *_{z_1 z_2} \frac{\sin\left(\frac{\pi}{\Delta} z_1\right)}{\pi z_1} \frac{\sin\left(\frac{\pi}{\Delta} z_2\right)}{\pi z_2} \ d\theta,$$

where the notation $*_{z_1 z_2}$ is a 2-D convolution with respect to z_1 and z_2. The convolution with respect to z_1 is treated in the generalized sense. The result is found in [47]:

$$\lim_{\epsilon \to 0} \left[E_\epsilon(x) *_x \frac{\sin\left(\frac{\pi}{\Delta} x\right)}{\pi x} \right] = \frac{\pi}{2\Delta^2} \left(\frac{\sin\left(\frac{\pi}{\Delta} x\right)}{\frac{\pi}{\Delta} x} \right) - \frac{\pi}{4\Delta^2} \left(\frac{\sin\left(\frac{\pi}{2\Delta} x\right)}{\frac{\pi}{2\Delta} x} \right)^2.$$
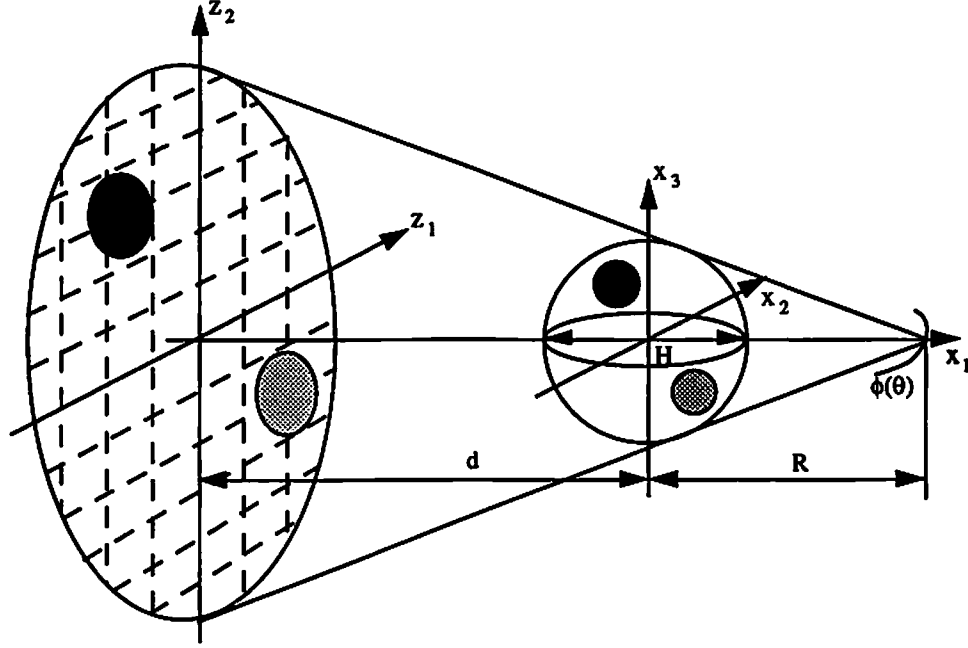
Figure 4.3: Arrangement for the cone beam imaging system used in the computer simulation. The cone vertices form a circle radius R, the detector plane is placed a distance $d = 70cm$ from the origin $\mathbf{x} = 0$ and the object was confined to a sphere of diameter $H = 30cm$. The simulations used values $R = \{20, 40, 60\}cm$.

Finally, $\hat{f}_{wp}(\mathbf{x})$ can be rewritten as

$$\hat{f}_{wp}(\mathbf{x}) = \frac{RD^3}{4\Delta^3} \int_0^{2\pi} \frac{1}{(x_1\cos\theta + x_2\sin\theta - R)^2 \hat{\alpha}_{02}^2} \left(\frac{\sin\left(\frac{\pi}{\Delta}(z_2 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}}D)\right)}{\frac{\pi}{\Delta}(z_2 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}}D)} \right)$$
$$\left[\left(\frac{\sin\left(\frac{\pi}{\Delta}(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}}D)\right)}{\frac{\pi}{\Delta}(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}}D)} \right) - \frac{1}{2} \left(\frac{\sin\left(\frac{\pi}{2\Delta}(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}}D)\right)}{\frac{\pi}{2\Delta}(z_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}}D)} \right)^2 \right] d\theta. \tag{4.52}$$

In the following section, we use the result (4.52) to evaluate the PSF as a function of R.

4.4 Simulation Results

The results from Sections 4.2 and 4.3 indicate that the performance of the filtered backprojection algorithm is dependent on the radius, R, of the circle formed by the

47

cone vertices. In the following we evaluate the PSF associated with the approximate reconstruction, using (4.52), for three different values of $R = \{60, 40, 20\} cm$. We then simulate a small scale 3-D imaging system using (4.46) and (4.47) for each of these values of R and compare the resulting reconstructions. The PSF and phantom reconstructions were computed for the 3-D imaging system illustrated in Figure 4.3. The detector plane was placed a distance of $70cm$ from the origin $\mathbf{x} = \mathbf{0}$ with the origin of the z_1-z_2 plane at position $\hat{\alpha} = (0, D, 0)^T$ shown in Figure 4.1. Thus for $R = 20$, $D = 90$; for $R = 40$, $D = 110$ and for $R = 60$, $D = 130$. To obtain a fair comparison in the following simulations, for each case the number of samples across the projection of a $30cm$ diameter sphere, centered at $\mathbf{x} = \mathbf{0}$, was held constant. Thus the data were collected on the planar array of detectors with sample intervals $\Delta = 1cm$, $\Delta = 1.33cm$ and $\Delta = 3.05cm$ for $R = 60$, $R = 40$ and $R = 20$ respectively. Images were reconstructed and the PSFs, $\hat{f}_{wp}(\mathbf{x})$, were computed from 64 equispaced angles of view for each value of R.

4.4.1 The Point Spread Function

The PSF was evaluated for two point sources $f_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i)$, $i = 0, 1$ where $\mathbf{x}_0 = (0, 0, -7)^T$ and $\mathbf{x}_1 = (0, -7, -7)^T$ for values of $R = \{20, 40, 60\}$ (all values are given in cm). The PSF's $\hat{f}^i_{wp}(\mathbf{x})$ were calculated within a sphere of diameter $30cm$ using (4.52).

The values of the PSFs are shown on selected 2-D planes in Figures 4.4 and 4.5 and along selected lines in Figures 4.6 and 4.7. A number of observations may be made from these results. The 2-D plots in Figures 4.4 and 4.5 clearly show the cone type structure illustrated in Figure 4.2 for the PSF with infinite bandwidth. Although for the bandlimited case, the PSF is no longer zero inside the cone with vertex at $\mathbf{x}$ and passing through the curve $\phi(\theta)$, the general conic shape of the PSF remains. Note also that the shape of the cone changes as the point source is moved.

These results also indicate that the degradation in image quality becomes worse as R is decreased. For example, the profiles in Figures 4.6 and 4.7 clearly
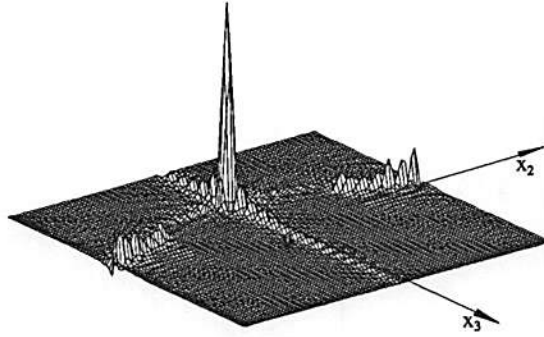
(a): Point source at $(0, 0, -7)^T$ with R=20.
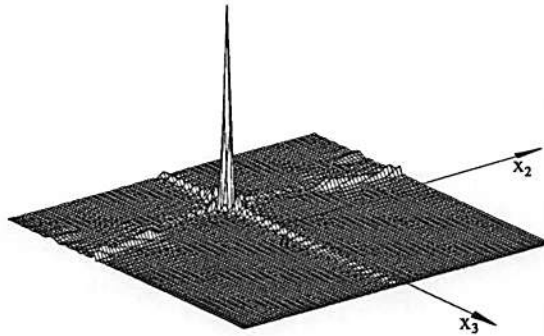
(b): Point source at $(0, 0, -7)^T$ with R=40.
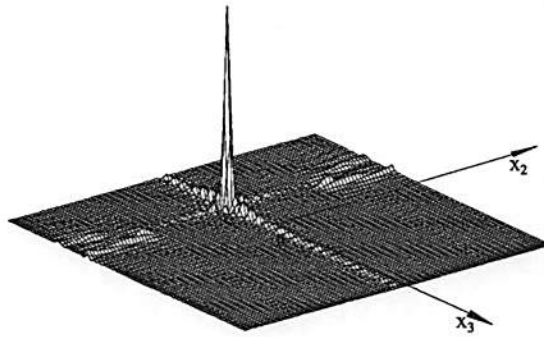
(c): Point source at $(0, 0, -7)^T$ with R=60.

Figure 4.4: The PSFs from (4.52) are computed on the plane $(0, x_2, x_3)^T$ for a point $\mathbf{x}_0 = (0, 0, -7)^T$ and $R = \{20, 40, 60\}$.
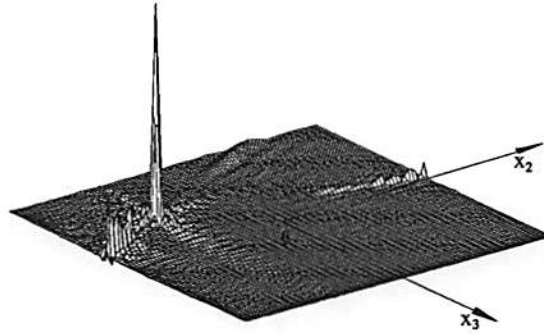
(a): Point source at $(0, -7, -7)^T$ with R=20.
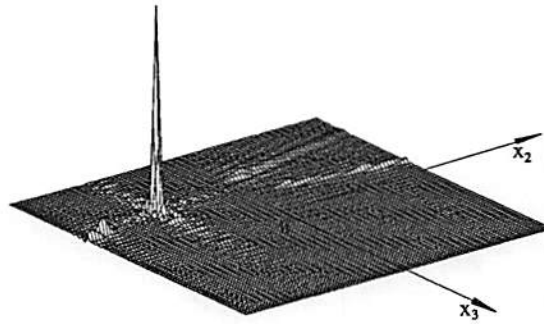
(b): Point source at $(0, -7, -7)^T$ with R=40.
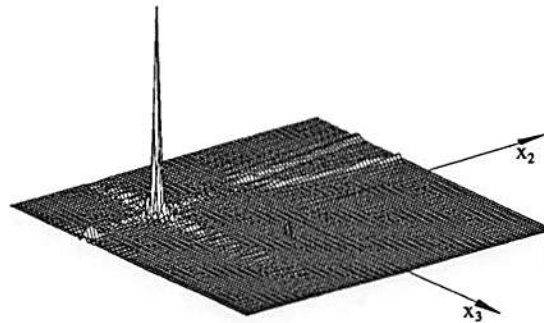
(c): Point source at $(0, -7, -7)^T$ with R=60.

Figure 4.5: The PSFs from (4.52) are computed on the plane $(0, x_2, x_3)^T$ for a point $\mathbf{x}_1 = (0, -7, -7)^T$ and $R = \{20, 40, 60\}$.
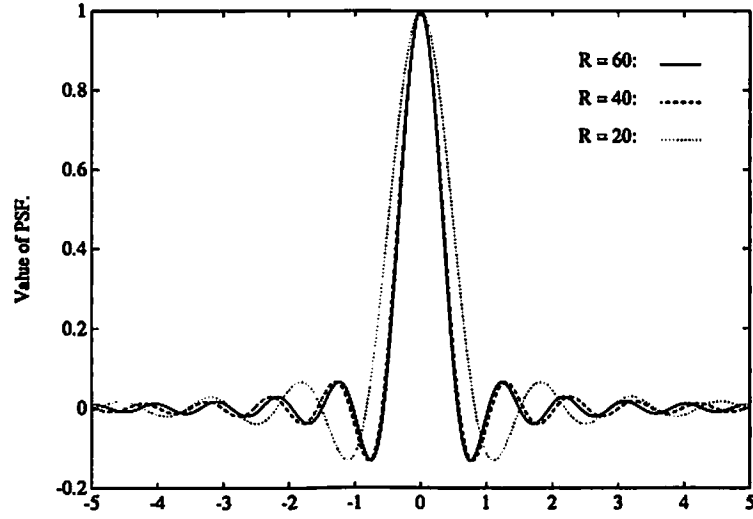
(a): Sampled along $(x_1, 0, -7)^T$.
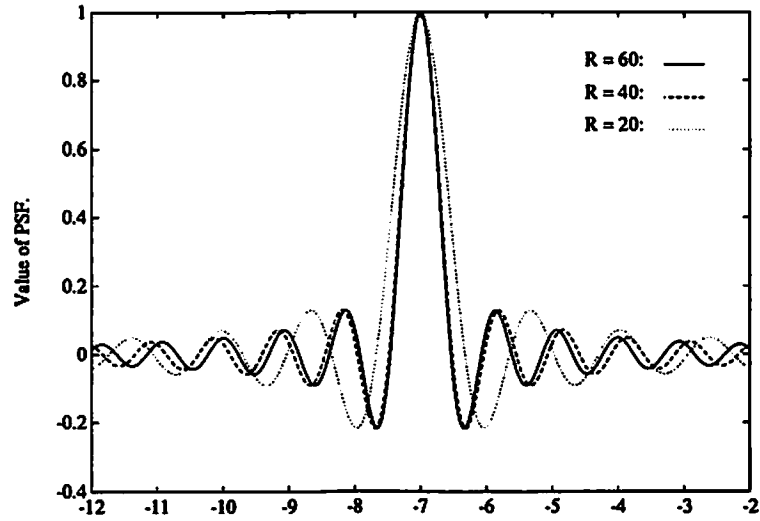
(b): Sampled along $(0, 0, x_3)^T$.

Figure 4.6: The PSFs from (4.52) are computed for $x_0 = (0, 0, -7)^T$ and $R = \{20, 40, 60\}$.
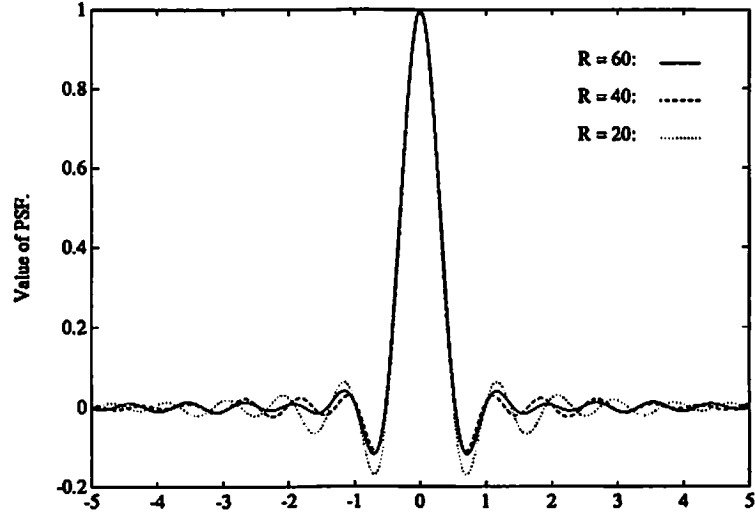
(a): Sampled along $(x_1, -7, -7)^T$.
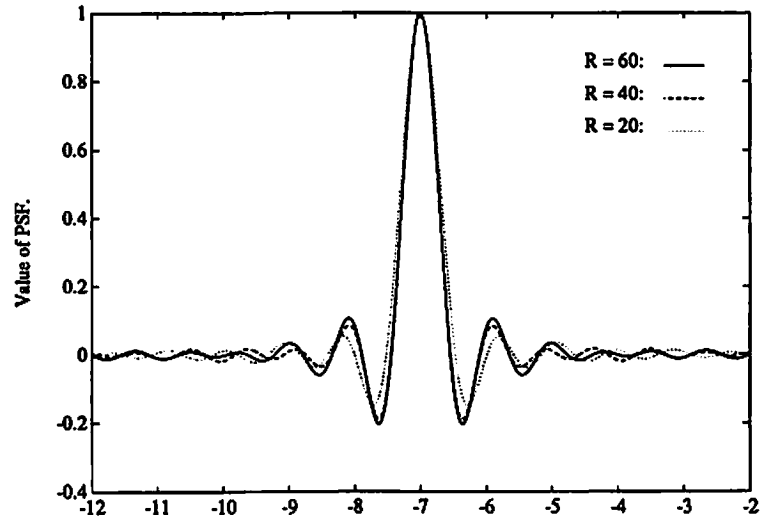
(b): Sampled along $(0, -7, x_3)^T$.

Figure 4.7: The PSFs from (4.52) are computed for $x_1 = (0, -7, -7)^T$ and $R = \{20, 40, 60\}$.

show the mainlobe of the PSF increases in both the x_1 and x_3 directions respectively as R is decreased. Note also that the values of the PSF for both point sources in Figures 4.4(a) and 4.5(a), increase dramatically in the vicinity of the curve, $\phi(\theta)$. This is due to the term $(x_1 \sin\theta + x_2 \cos\theta - R)^{-2}$ in (4.52). Consequently, to minimize reconstruction artifacts, it is important to maintain R as large as possible with respect to the desired imaging volume.

Finally, we note that these observations concur with the conclusions of Feldkamp *et al.* [9], namely that the degradation is worse in the x_3 direction than in the x_1-x_2 plane. Note for example, that the sidelobes in Figures 4.6 and 4.7 are higher in the x_3 direction than in the x_1 direction. However, it is important to recognize the fact, as illustrated in Figures 4.4 and 4.5 that the spatially variant blurring occurs predominantly in the vicinity of the surface of the cone with vertex at the point source x and passing through the curve $\phi(\theta)$.

The utility of these results is that one can determine, for a given proposed system, an upper bound on the possible resolution using the radius R of the cone vertices, the location of the detector plane, the available detector resolution and the maximum extent of the object to be imaged. This information should be useful in the design of this type of imaging system. In addition, it is straightforward to modify equation (4.52) to include different projection filters [23] [46] and therefore examine the effect of these filter functions on the resulting PSF.

4.4.2   Simulated Image Reconstruction

The images were reconstructed using a discretized form of (4.46) and (4.47). The convolution defined in (4.46) can be implemented either in the spatial domain or using a zeropadded FFT [47]. In our simulation, we directly employed a discrete linear convolution with the filter frequency response given by the bandlimited Fourier transform of the function $\lim_{\epsilon \to 0} E_\epsilon(\hat{\alpha}_1)$ defined in (4.30) *i.e.* [47]:

$$H(\omega) = \begin{cases} |\omega| & |\omega| < \pi/\Delta \\ 0 & \text{otherwise,} \end{cases}$$

Region	Center (cm)			Size of Region (cm)			Intensity
	x_1	x_2	x_3	x_1	x_2	x_3	
1	0	0	0	25.0	25.0	25.0	84
2	4.0	4.0	4.0	8.0	8.0	8.0	150
3	-3.0	4.0	-5.5	6.5	9.0	7.0	200
4	-3.5	-4.0	5.0	9.0	5.0	7.0	255
5	-3.5	-3.5	4.5	6.5	6.0	7.5	0

Table 4.1: Parameters of the 4 regions of the phantom image. Regions 2, 3, 4 and 5 are one sphere, two ellipsoids and one cube respectively with the parameters given above embedded in the homogeneous sphere defined as region 1 above.

where Δ is the sample spacing in the planar projection data. Design of appropriate filters for filtered backprojection has been discussed elsewhere [23] [33] and will not be discussed further here. The reconstruction is completed by performing a voxel driven backprojection with linear interpolation according to (4.47) using the trapezoidal method for numerical integration in θ.

The images were reconstructed from 64 equispaced angles of view for each value of R. The parameters of the original object, Figure 4.8, are given in Table 4.1. This object is confined to a sphere of diameter $25cm$ centered at x = 0. The images shown in Figures 4.9, 4.10 and 4.11 were reconstructed on a $30cm$ diameter sphere with a voxel size of $0.5cm^3$. Each figure shows the central 49 slices of the reconstruction for planes $x_3 = (k - 25)/2\ cm, k = 1, 2, ..., 49$. A more quantitative comparison can be made from the line plots, Figures 4.12 and 4.13, which show the values of the original and reconstructed images along the lines $(x_1, -3.5, -3.5)^T$ and $(-3.5, -3.5, x_3)^T$ respectively. These results clearly show deterioration in image quality as R is reduced. It can be seen that the uniform portion of the reconstructed image for $R = 20$ exhibits a decreasing intensity away from the origin. In addition there are a large number of artifacts outside the $25cm$ diameter sphere containing the original object. Although the reconstructions for $R = 40$ and $R = 60$ both show some artifacts, there is a clear improvement over those for $R = 20$, with the case for $R = 60$ slightly better than for $R = 40$.

The above results show that from the view of reconstruction accuracy, it is advantageous to have cone vertices on a circle with the maximum possible radius,
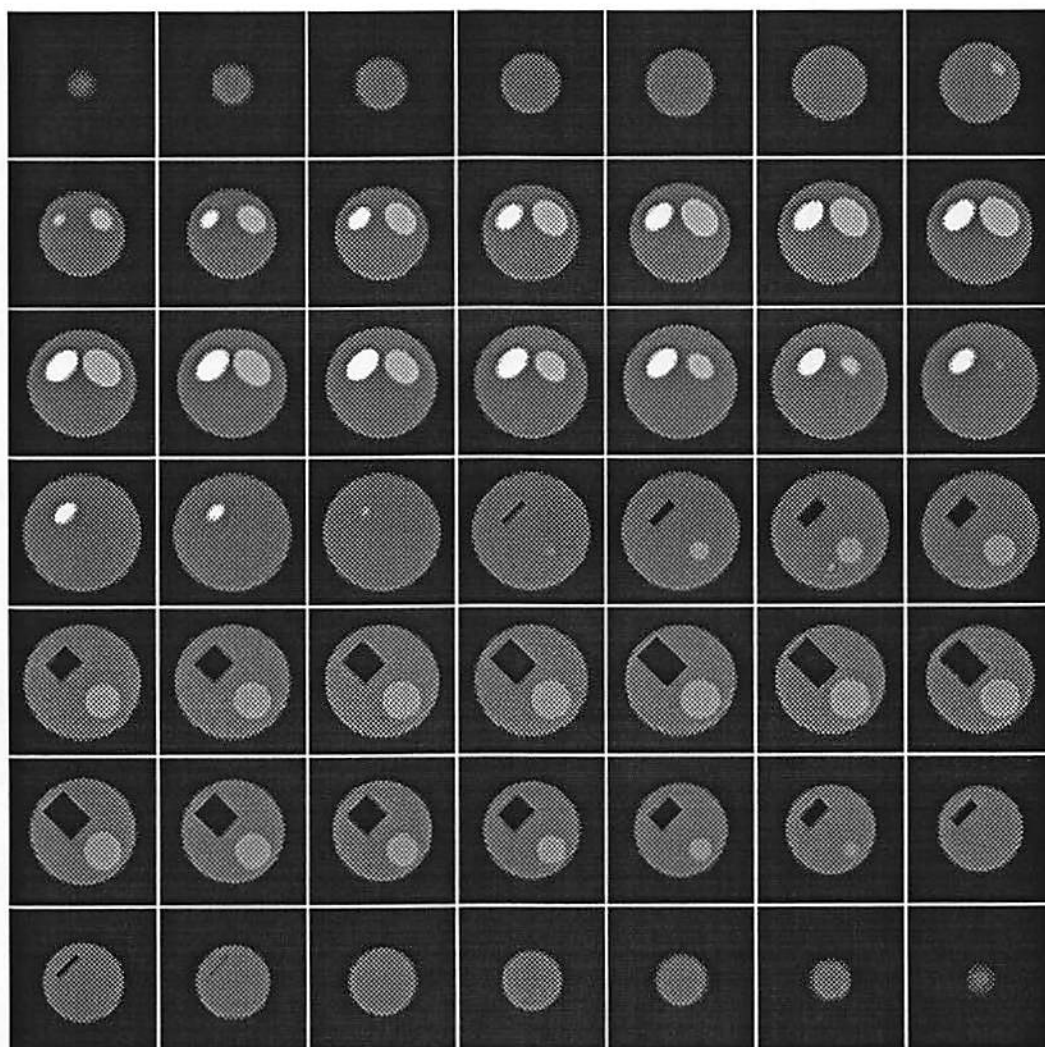
Figure 4.8: Cross sectional planar views of the original phantom, $x_3 = (k - 25)/2$ cm, $k = 1, 2, \cdots, 49$. The same planes of the reconstruction for different values of R are shown in Figures. 4.9, 4.10, 4.11
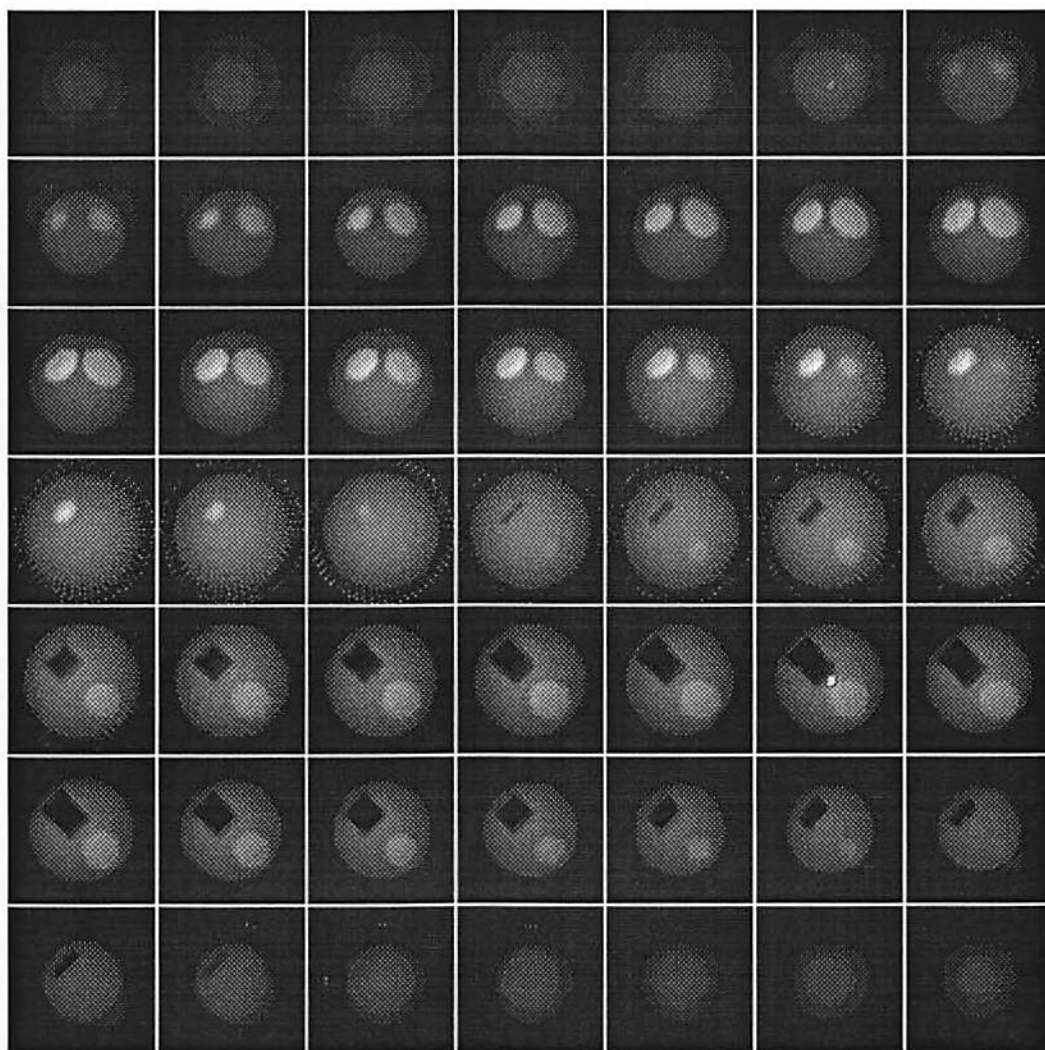
Figure 4.9: Reconstruction of the phantom from 64 equispaced angles of view with $R = 20$.
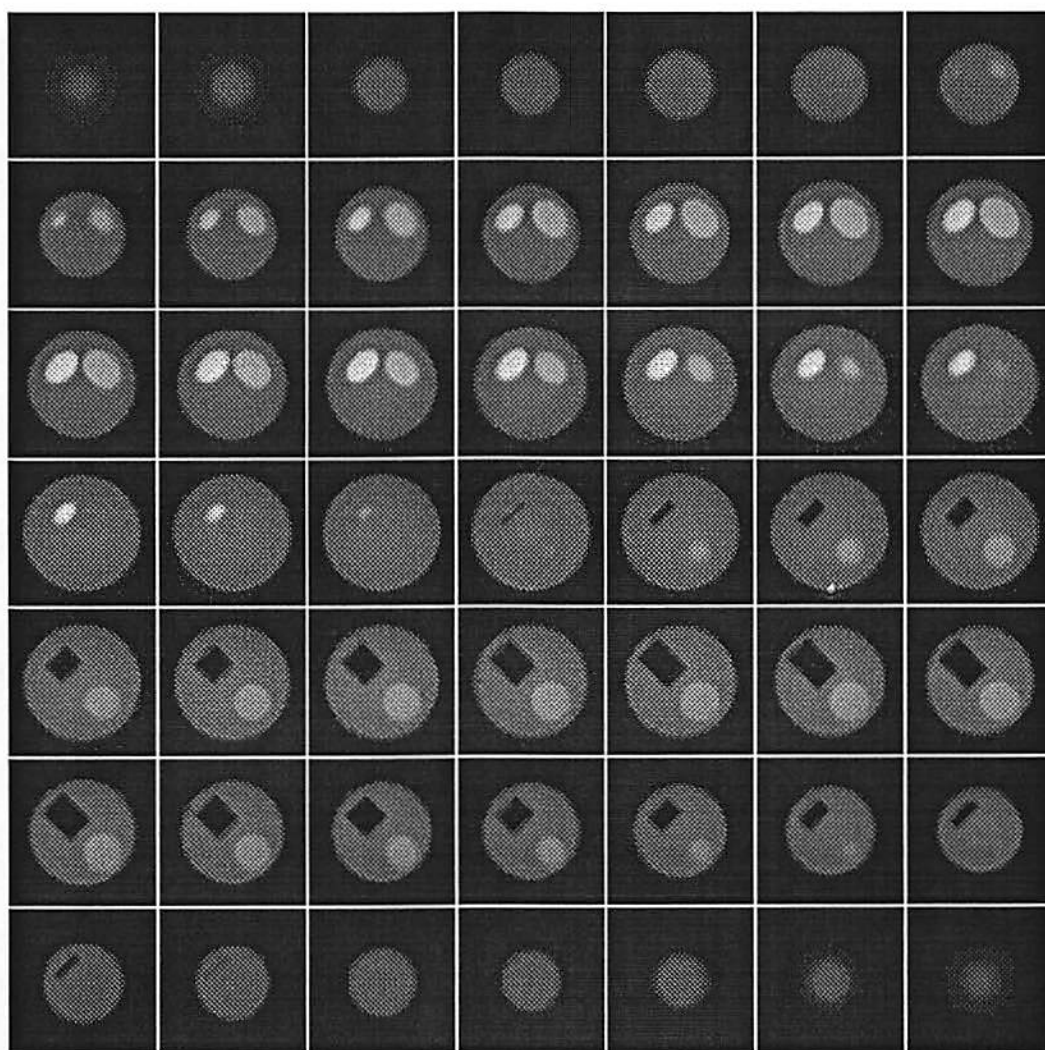
Figure 4.10: Reconstruction of the phantom from 64 equispaced angles of view with $R = 40$.
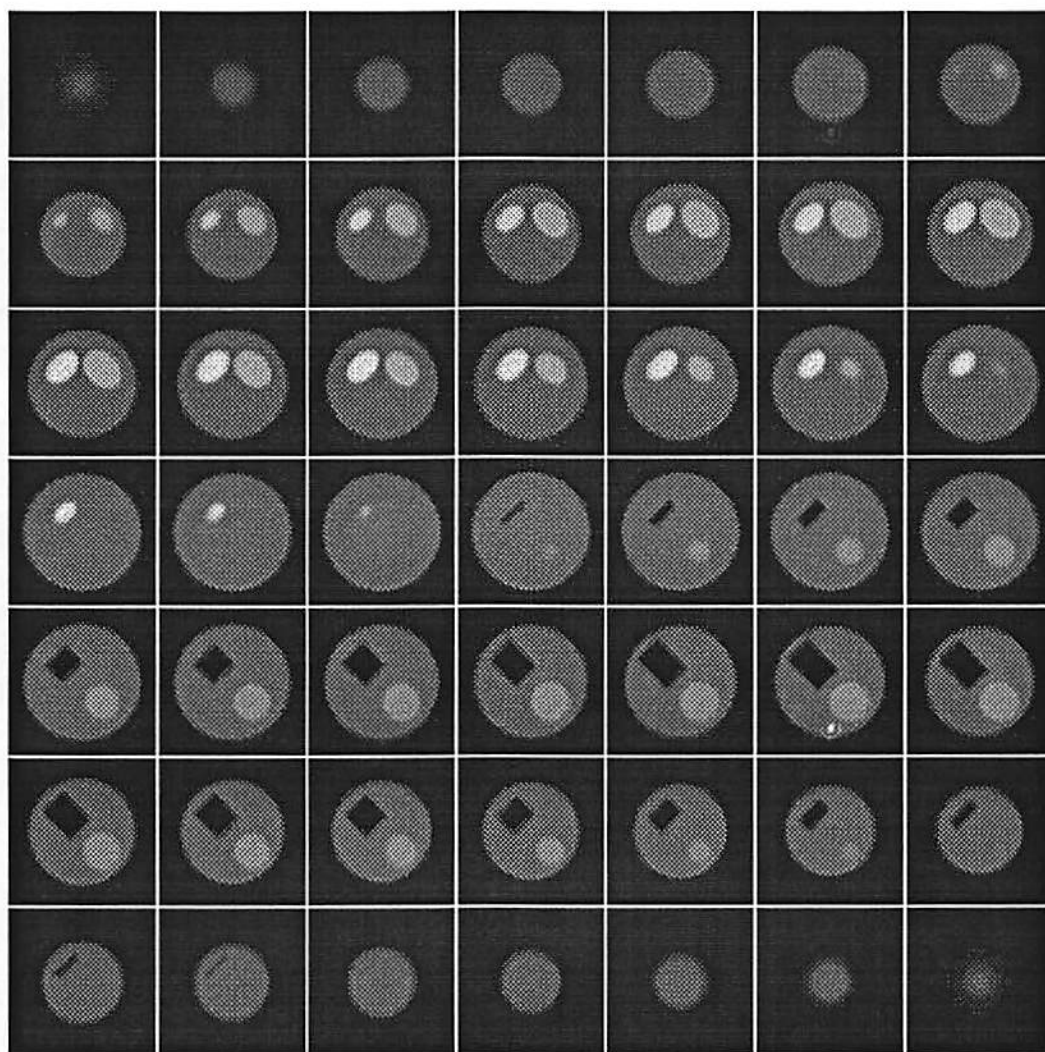
Figure 4.11: Reconstruction of the phantom from 64 equispaced angles of view with $R = 60$.
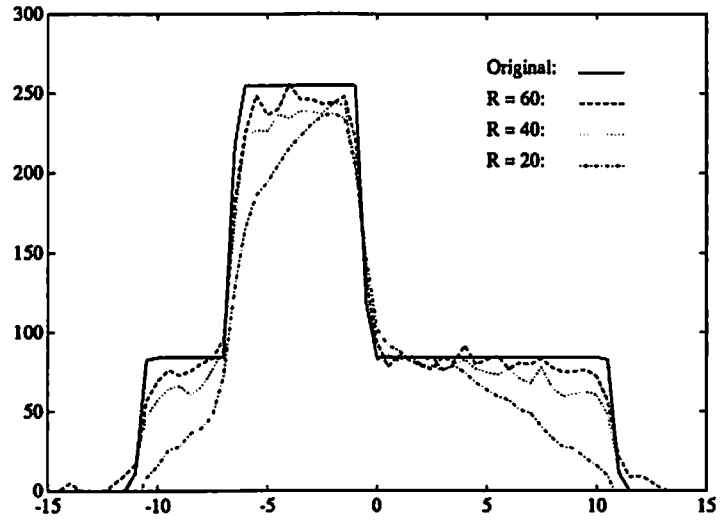
Figure 4.12: The original and reconstructed images for $R = \{20, 40, 60\}cm$ sampled along the line $(x_1, -3.5, -3.5)^T$.
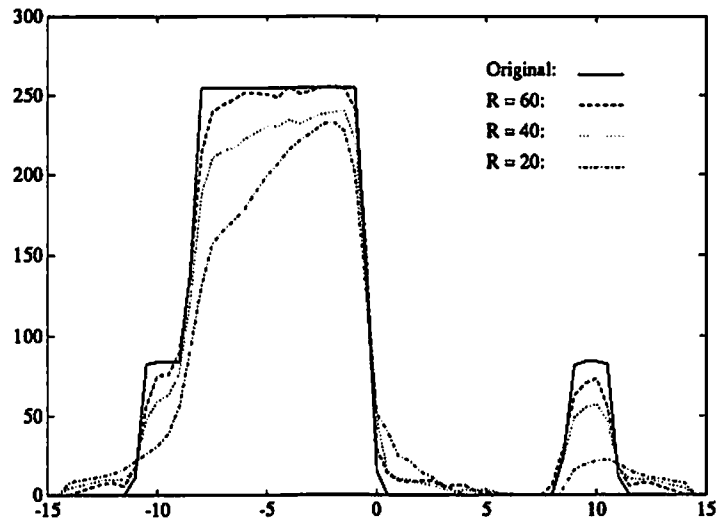
Figure 4.13: The original and reconstructed images for $R = \{20, 40, 60\}cm$ sampled along the line $(-3.5, -3.5, x_3)^T$.

R, relative to the size of the object to be imaged. As R is reduced, the reconstructions become increasingly sensitive to the errors introduced by the approximate nature of the reconstruction formula, as given in (4.39).

4.A   Appendix

In order to derive the point spread function for the approximate reconstruction algorithm, we must first derive the projection data for a point source. In the following we will assume that the point sources are confined to the interior of a sphere of radius R. Furthermore, we will assume that the imaging volume of interest is also confined to the interior of the sphere of radius R, and thus the PSF need only be compute for values of $\mathbf{x}$, with $\|\mathbf{x}\| < R$.

Consider the point image $f(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0)$ with $\|\mathbf{x}_0\| < R$. Let $\alpha_0 = \mathbf{x}_0 - \phi(\theta)$, then from (4.2)

$$g(\alpha, \theta) = \int_{-\infty}^{\infty} f(\phi(\theta) + t\alpha)dt = \int_{-\infty}^{\infty} \delta(t\alpha - \alpha_0)dt. \qquad (4.53)$$

The rotated projection data, $g_r(\hat{\alpha}, \theta)$ are found using the transformation $\hat{\alpha} = M\alpha$ as:

$$g_r(\hat{\alpha}, \theta) = \int_{-\infty}^{\infty} \delta(M^T(t\hat{\alpha} - \hat{\alpha}_0))dt. \qquad (4.54)$$

Since M is a unitary matrix, $\delta(M^T\hat{\alpha}) = \delta(\hat{\alpha})$, and therefore

$$
\begin{aligned}
g_r(\hat{\alpha}, \theta) &= \int_{-\infty}^{\infty} \delta(t\hat{\alpha} - \hat{\alpha}_0)dt \\
&= \int_{-\infty}^{\infty} \delta(t\hat{\alpha}_1 - \hat{\alpha}_{01})\delta(t\hat{\alpha}_2 - \hat{\alpha}_{02})\delta(t\hat{\alpha}_3 - \hat{\alpha}_{03})dt.
\end{aligned}
\qquad (4.55)
$$

To simplify (4.55) we use the properties [1]:

$$\delta(cx - x_0) = \frac{1}{|c|}\delta\left(x - \frac{x_0}{c}\right) \quad c \neq 0, \qquad (4.56)$$

and

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)dx = f(x_0) \qquad (4.57)$$

60

Chapter 4

Derivation and Analysis of A General Inversion for Cone Beam Tomography

In this chapter, we address the problem of three dimensional (3-D) image reconstruction from cone beam projections. An inversion formula is derived for the case where the cone vertices form an unbounded curve by modifying a result due to Kirillov [31]. To transform the general reconstruction formula to a specific algorithm for a given geometry, one must replace the cone beam projection data to the actual planar projection data. It is shown that the image may be reconstructed by one-dimensional (1-D) filtering of the weighted two-dimensional (2-D) projection data followed by a weighted 3-D backprojection. As discussed in Chapter 3, no exact reconstruction exists for the special case where the cone vertices form a circle. Therefore, an approximate reconstruction formula is developed using this general formula and shown to be essentially equivalent to the 'practical cone-beam algorithm' of Feldkamp et al. [9]. We note that this result was shown independently by Smith [55], i.e. that equation (9.5) in [54] is equivalent to the algorithm given in [9]. The appeal of the approach presented in this chapter is that, although the algorithms for bounded curves are only approximate, they are of the filtered backprojection type, hence computationally very efficient and produce acceptable

results in many cases. The derivation given here allows analysis of the spatially variant point spread function (PSF) implicit in the approximation.

In the following we return to Kirillov's approach. The reconstruction formula is derived for a general curve in Section 4.1 by following a development similar to that in [31] but for real 3-D functionals. The case in which the curve is a single circle enclosing the object is treated in Section 4.2.1. Since this curve violates the conditions in [31], [57] and [54], we introduce an approximate reconstruction formula. The degree of approximation is specifically derived in terms of the resulting PSF in Section 4.2.2. The reconstruction formula for planar projection data is shown in Section 4.3.1. To examine the effect of bandlimited data acquisition due to sampling, we derive the form of the spatially variant PSF for bandlimited projection data by introducing an ideal low pass filter in Section 4.3.2. These results are used to study the effect of the radius of the circle formed by the cone vertices and the sample rate on resulting image quality in Section 4.4.

4.1   The General Reconstruction Formula

In this section we derive an inversion formula for the reconstruction of a 3-D object from a set of cone beam projections whose vertices lie along a continuous curve $\phi(\theta)$ through $\mathbf{R}^3$, parameterized by θ with domain Θ, an interval of the real line. The full derivation is included here since although similar results are given in [54] and [55], the derivation is not readily available in the literature. Furthermore, it is necessary for describing the approximation involved in the case treated in Section 4.2 where $\phi(\theta)$ is a circular curve surrounding the object. Alternative inversion formulae for functions of compact support are given in [57] and [54].

Following the notation in [53], let $f(\mathbf{x})$ denote the object to be reconstructed, where $\mathbf{x} = (x_1, x_2, x_3)^T$. The operation $(\cdot)^T$ denotes the transpose of a vector. The cone beam data may be expressed as

$$g_1(\alpha, \theta) = \int_{-\infty}^{\infty} f(\phi(\theta) + t\frac{\alpha}{\|\alpha\|})dt, \tag{4.1}$$

where

$$\phi(\theta) = (\phi_1(\theta), \phi_2(\theta), \phi_3(\theta))^T$$

is the curve formed by the vertices of the cones,

$$\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$$

and

$$\|\alpha\| = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2}.$$

In the following the modified projection data

$$g(\alpha, \theta) = \frac{1}{\|\alpha\|} g_1(\alpha, \theta) = \int_{-\infty}^{\infty} f(\phi(\theta) + t\alpha) dt \qquad (4.2)$$

is employed. It is easily shown that

$$g(c\alpha, \theta) = \frac{1}{|c|} g(\alpha, \theta)$$

for any scalar c, *i.e.* $g_1(\alpha, \theta)$ is homogeneous in α. Thus we need only know $g(\alpha, \theta)$ for α on the unit sphere in $\mathbf{R}^3$. We maintain the 4-D representation of $g(\alpha, \theta)$ to facilitate derivation of the inversion formula.

Forming the 3-D Fourier transform in α of (4.2) gives

$$G(\beta, \theta) = \int_{\mathbf{R}^3} \int_{-\infty}^{\infty} f(\phi(\theta) + t\alpha) e^{-i(\alpha \cdot \beta)} dt d\alpha, \qquad (4.3)$$

where $(\alpha \cdot \beta)$ denotes the vector inner product and $\beta = (\beta_1, \beta_2, \beta_3)^T$. As noted in [57] and [54], in general the integral in (4.3) does not exist. This problem is overcome by treating $G(\beta, \theta)$ as the generalized Fourier transform [39] of $g(\alpha, \theta)$. In the following, Fourier integrals should be interpreted in the generalized sense where necessary. Let $\mathbf{x} = \phi(\theta) + t\alpha$ in (4.3), then:

$$G(\beta, \theta) = \int_{\mathbf{R}^3} \int_{-\infty}^{\infty} f(\mathbf{x}) e^{-i(\frac{\mathbf{x} - \phi(\theta)}{t} \cdot \beta)} |t|^{-3} dt d\mathbf{x}.$$

A second change of variable $t = 1/\tau$ yields

$$G(\beta, \theta) = \int_{-\infty}^{\infty} \int_{\mathbf{R}^3} f(\mathbf{x}) e^{-i(\mathbf{x} \cdot \tau \beta)} d\mathbf{x} \, e^{i(\phi(\theta) \cdot \tau \beta)} |\tau| d\tau. \qquad (4.4)$$

We now introduce the function from [31]:

$$K(\beta,\omega) = \int_{-\infty}^{\infty} F(\tau\beta)|\tau|e^{i\omega\tau}d\tau. \tag{4.5}$$

Using the Fourier transform relationship

$$F(\tau\beta) = \int_{\mathbf{R}^3} f(\mathbf{x})e^{-i(\mathbf{x}\cdot\tau\beta)}d\mathbf{x}, \tag{4.6}$$

it follows that

$$K(\beta,\omega) = \int_{-\infty}^{\infty}\int_{\mathbf{R}^3} f(\mathbf{x})e^{-i(\mathbf{x}\cdot\tau\beta)}d\mathbf{x}\, e^{i\omega\tau}|\tau|d\tau. \tag{4.7}$$

Comparing (4.4), and (4.7), it is clear that

$$G(\beta,\theta) = K(\beta,\omega)|_{\omega=(\phi(\theta)\cdot\beta)}. \tag{4.8}$$

Furthermore, from (4.6):

$$f(\mathbf{x}) = \frac{1}{(2\pi)^3}\int_{\mathbf{R}^3} F(\tau\beta)e^{i(\mathbf{x}\cdot\tau\beta)}\bigg|_{\tau=1} d\beta. \tag{4.9}$$

Taking the Fourier transform of $K(\beta,\omega)$ defined in (4.5) with respect to ω, we have:

$$F(\tau\beta) = \frac{|\tau|^{-1}}{2\pi}\int_{-\infty}^{\infty} K(\beta,\omega)e^{-i\omega\tau}d\omega. \tag{4.10}$$

Substituting this result for $\tau = 1$ in (4.9) gives:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^4}\int_{\mathbf{R}^3}\int_{-\infty}^{\infty} K(\beta,\omega)e^{-i(\omega-\mathbf{x}\cdot\beta)}d\omega d\beta. \tag{4.11}$$

Since $f(\mathbf{x})$ may be recovered from $K(\beta,\omega)$, according to (4.11), it follows that sufficient conditions for the reconstruction of $f(\mathbf{x})$ from $g(\alpha,\theta)$, or equivalently $G(\beta,\theta)$, are determined by the range of (β,ω) for which the equality in (4.8) holds.

The most restrictive sufficient condition on $\phi(\theta)$ follows directly from (4.8), i.e. $f(\mathbf{x})$ is recoverable from $G(\beta,\theta)$ if for each $\beta \in \mathbf{R}^3$ and each $\omega \in \mathbf{R}$ there exists a point $\theta \in \Theta$, where Θ denotes the domain of $\phi(\theta)$, such that $(\phi(\theta)\cdot\beta) = \omega$.

Equivalently, if for all points $(\beta, \omega) \in (\mathbf{R}^3 \times \mathbf{R})$ the plane $(\mathbf{x} \cdot \beta) = \omega$ intersects the curve $\phi(\theta)$ at least once, then one can reconstruct the object, see [31] and [54]. Suppose for a curve $\phi(\theta)$ there exist a subset $\mathbf{N} \subset (\mathbf{R}^3 \times \mathbf{R})$ with $m(\mathbf{N}) = 0$, where $\mathbf{N} = \{(\beta, \omega) \in (\mathbf{R}^3 \times \mathbf{R}) : \ (\beta \cdot \phi(\theta)) \neq \omega,\ \text{for any}\ \theta \in \Theta\}$ and $m(\cdot)$ is the Lebesgue measure on $\mathbf{R}^4$. The values of $K(\beta, \omega)$ on this subset $\mathbf{N}$ do not affect the integral in (4.11), and therefore a relaxed sufficient condition is as follows: "for almost all $(\beta, \omega) \in (\mathbf{R}^3 \times \mathbf{R})$, there must exist a $\theta \in \Theta$ such that $(\beta \cdot \phi(\theta)) = \omega$". Less restrictive sufficient conditions for bounded source curves and functions of compact support, based on the properties of the function $K(\beta, \omega)$, are given in [57] and [54]. To facilitate the following derivation, we additionally assume that the curve $\phi(\theta)$ is intersected exactly L times by almost all planes [31], however we note that the condition is neither necessary nor easy to satisfy in general.

Using the equivalence of $K(\beta, \omega)$ and $G(\beta, \theta)$ under the above conditions, we substitute $G(\beta, \theta)$ in place of $K(\beta, \omega)$ with $\omega = (\phi(\theta) \cdot \beta)$, and integrate with respect to θ with the Jacobian:

$$J = \left| \frac{d\omega}{d\theta} \right| = \left| (\phi'(\theta) \cdot \beta) \right|,$$

and thus

$$f(\mathbf{x}) = \frac{1}{(2\pi)^4 L} \int_{\mathbf{R}^3} \int_{\Theta} G(\beta, \theta) e^{i(\mathbf{x} - \phi(\theta)) \cdot \beta} sgn(\phi'(\theta) \cdot \beta)(\phi'(\theta) \cdot \beta) d\theta d\beta, \quad (4.12)$$

where $sgn(\cdot)$ denotes the sign of the function. Note that this transformation is only valid under certain restrictions on the curve $\phi(\theta)$; see section VII in [54].

To complete the derivation , we express $g(\mathbf{x} - \phi(\theta), \theta)$ as the inverse Fourier transform of $G(\beta, \theta)$ evaluated at $\alpha = \mathbf{x} - \phi(\theta)$ and take the paritial derivative with respect to θ in the first variable to obtain:

$$(g'_\alpha(\mathbf{x} - \phi(\theta), \theta) \cdot \phi'(\theta)) = \frac{i}{(2\pi)^3} \int_{\mathbf{R}^3} G(\beta, \theta) e^{i((\mathbf{x} - \phi(\theta)) \cdot \beta)}(\phi'(\theta) \cdot \beta) d\beta, \quad (4.13)$$

where $g'_\alpha(\alpha, \theta)$ denotes the gradient of $g(\alpha, \theta)$ with respect to α and $\phi'(\theta)$ the derivative of $\phi(\theta)$ with respect to θ. Comparing (4.12) and (4.13) we may rewrite

(4.12) as

$$f(\mathbf{x}) = \int_{\Theta} [(g'_{\alpha}(\alpha, \theta) \cdot \phi'(\theta)) *_{\alpha} h(\alpha, \theta)]_{|\alpha = \mathbf{x} - \phi(\theta)} d\theta, \qquad (4.14)$$

where $*_{\alpha}$ denotes the 3-D convolution in α and

$$h(\alpha, \theta) = \frac{-i}{(2\pi)^4 L} \int_{\mathbf{R}^3} sgn(\phi'(\theta) \cdot \beta) e^{i(\alpha \cdot \beta)} d\beta \qquad (4.15)$$

is treated as a generalized function. Since $h(\alpha, \theta)$ contains singularities (see (4.24) below), the convolution integral in (4.14) must be treated in the sense of its Cauchy principal value.

Equations (4.12) or (4.14) and (4.15) constitute exact reconstruction formulae for any curve $\phi(\theta)$ obeying the conditions stated above. However, care must be taken in the formulation of the generalized Fourier integrals and the treatment of singular integrals. In addition, direct implementation of (4.14) is computationally expensive.

In [53] Smith proposes a rotation matrix which reduces the 3-D convolution in (4.14) to a 1-D convolution for $\phi(\theta)$ an infinite straight line. In the following section we derive an approximate formula for $\phi(\theta)$ a closed circle and employ the idea of a rotation matrix to arrive at a fast, filtered-backprojection algorithm for cone beam reconstruction.

4.2 Approximate Reconstruction for $\phi(\theta)$ A Circle

Although (4.14) and (4.15) provide an exact reconstruction formula, a source curve $\phi(\theta)$ which satisfies the conditions stated above is not viable in practice. In this section, an approximate reconstruction algorithm is derived for $\phi(\theta)$ a closed circle by making a simple approximation and employing the idea of the rotation matrix. In addition, we analyse the degree of approximation by deriving the spatially varying PSF inherent in this approximation.

4.2.1 Reconstruction by Filtered Backprojection
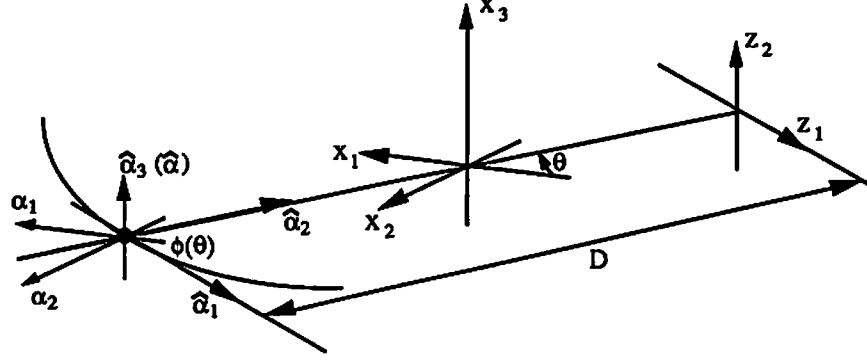
Figure 4.1: Illustration of the cone beam coordinate system. $(x_1, x_2, x_3)^T$ is the fixed object centered coordinate system. The curve $\phi(\theta)$ is a circle with radius R, in the plane $x_3 = 0$ centered at $\mathbf{x} = 0$. Data for the value of θ shown are collected on the planar detector z_1-z_2 and related to the rotated coordinate system $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)^T$ by (4.31).

For a configuration such as the DSR [59] or a cone beam collimated SPECT system [11], an appropriate curve for the cone beam vertices is the circle

$$\phi(\theta) = (R\cos\theta, R\sin\theta, 0)^T, \qquad \theta \in \Theta = [0, 2\pi]. \tag{4.16}$$

The arrangement is illustrated in Figure 4.1. It is clear that this curve violates the conditions stated in Section 4.1. To find an approximate reconstruction formula we return to (4.8) from which we have

$$G(\beta, \theta) = K(\beta, \omega)_{|\omega = (\phi(\theta)\cdot\beta)} \tag{4.17}$$

for all $\theta \in [0, 2\pi]$. Thus for a specific β, $K(\beta, \omega)$ can be recovered from $G(\beta, \theta)$ for all $\omega = (\phi(\theta) \cdot \beta) = \beta_1 R\cos\theta + \beta_2 R\sin\theta$ in the range

$$\omega \in [-R\sqrt{\beta_1^2 + \beta_2^2}, R\sqrt{\beta_1^2 + \beta_2^2}].$$

Furthermore, as θ takes all values in the range $[0, 2\pi]$, the circle $\phi(\theta)$ will intersect almost all planes

$$(\mathbf{x} \cdot \beta) = \omega, \qquad \omega \in [-R\sqrt{\beta_1^2 + \beta_2^2}, R\sqrt{\beta_1^2 + \beta_2^2}]$$

exactly twice, *i.e.* $\phi(\theta)$ has the property: $L = 2$.

We now define the approximate inverse solution as

$$\hat{f}(\mathbf{x}) = \frac{1}{(2\pi)^4} \int_{\mathbb{R}^3} \int_{-R\sqrt{\beta_1^2+\beta_2^2}}^{R\sqrt{\beta_1^2+\beta_2^2}} K(\beta,\omega)e^{-i(\omega-(\mathbf{x}\cdot\beta))}d\omega d\beta. \qquad (4.18)$$

Using (4.17), and repeating steps equivalent to (4.11)–(4.14) with $L = 2$, it then follows that the approximate solution is given by:

$$\hat{f}(\mathbf{x}) = \int_0^{2\pi} [(g'_\alpha(\alpha,\theta) \cdot \phi'(\theta)) *_\alpha h(\alpha,\theta)]_{|\alpha=\mathbf{x}-\phi(\theta)} d\theta \qquad (4.19)$$

where $h(\alpha,\theta)$ is defined in (4.15).

The approximate reconstruction formula (4.19) can now be written in a form more suitable for computation. To achieve this goal we employ a coordinate rotation matrix as suggested by Smith [53].

For $\phi(\theta)$ in (4.16):

$$\phi'(\theta) = (-R\sin\theta, R\cos\theta, 0)^T, \qquad (4.20)$$

and there exists a rotation matrix M such that $M\phi'(\theta) = (\|\phi'(\theta)\|, 0, 0)^T$. In our case the matrix is

$$M = \begin{pmatrix} m_1^T \\ m_2^T \\ m_3^T \end{pmatrix} = \begin{pmatrix} -\sin\theta & \cos\theta & 0 \\ -\cos\theta & -\sin\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (4.21)$$

Note that $det(M) = 1$, $M^T M = I$, and $m_1^T = \phi'(\theta)/\|\phi'(\theta)\|$.

This rotation matrix may be employed to simplify (4.15) by setting $M\beta = \hat{\beta}$ and noting that $d\hat{\beta} = d\beta$ and $L = 2$, we obtain

$$h(\alpha,\theta) = \frac{-i}{2(2\pi)^4} \int_{\mathbb{R}^3} sgn(\phi'(\theta) \cdot M^T\hat{\beta})e^{i(\alpha \cdot M^T\hat{\beta})}d\hat{\beta}. \qquad (4.22)$$

Since $(\alpha \cdot M^T \hat{\beta}) = (M\alpha \cdot \hat{\beta})$ we have

$$sgn(\phi'(\theta) \cdot M^T\hat{\beta}) = sgn(M\phi'(\theta) \cdot \hat{\beta}) = sgn(\|\phi'(\theta)\|\hat{\beta}_1).$$

Setting $\hat{\alpha} = M\alpha$ then gives

$$h(\alpha, \theta) = \frac{-i}{2(2\pi)^4} \int_{\mathbf{R}^3} sgn(\hat{\beta}_1) e^{i(\hat{\alpha}\cdot\hat{\beta})} d\hat{\beta}. \qquad (4.23)$$

Using the result from [16]

$$\int_{-\infty}^{\infty} sgn(x) e^{ixX} dx = \frac{2i}{X},$$

the integral reduces to

$$h(\alpha, \theta) = \frac{1}{4\pi^2} \delta(\hat{\alpha}_2)\delta(\hat{\alpha}_3) \left(\frac{1}{\hat{\alpha}_1}\right). \qquad (4.24)$$

Noting $\hat{\alpha}_i = (m_i \cdot \alpha)$ with m_i defined in (4.21) we can substitute in (4.19), to obtain

$$\hat{f}(\mathbf{x}) = \frac{1}{4\pi^2} \int_0^{2\pi} \qquad (4.25)$$

$$\left[(g'_\alpha(\alpha, \theta) \cdot \phi'(\theta)) *_\alpha \delta(m_2 \cdot \alpha)\delta(m_3 \cdot \alpha)\frac{1}{(\frac{\phi'(\theta)}{\|\phi'(\theta)\|} \cdot \alpha)}\right]_{|\alpha = \mathbf{x} - \phi(\theta)} d\theta.$$

In the case where $\phi(\theta)$ satisfies the conditions stated in Section 4.1, (4.25) is the form of the reconstruction formula given by Smith [53] where M is chosen as the appropriate rotation matrix.

In our case the reconstruction can be further simplified by performing the convolution with respect to the rotated coordinates $\hat{\alpha}$ as follows. Using the same rotation as above, $\hat{\alpha} = M\alpha$, we define the rotated projection measurement as

$$g_r(\hat{\alpha}, \theta) = g(M^T\hat{\alpha}, \theta). \qquad (4.26)$$

Taking partial derivatives with respect to $\hat{\alpha}$:

$$g'_{r_{\alpha}}(\hat{\alpha}, \theta) = M g'_{\alpha}(\alpha, \theta). \qquad (4.27)$$

Thus

$$(g'_{\alpha}(\alpha, \theta) \cdot \phi'(\theta)) = (M^T g'_{r_{\alpha}}(\hat{\alpha}, \theta) \cdot \phi'(\theta)) = (g'_{r_{\alpha}}(\hat{\alpha}, \theta) \cdot M \phi'(\theta)).$$

Since $M \phi'(\theta) = (\|\phi'(\theta)\|, 0, 0)^T$, the inner product reduces to:

$$(g'_{\alpha}(\alpha, \theta) \cdot \phi'(\theta)) = \|\phi'(\theta)\| \frac{\partial g_r(\hat{\alpha}, \theta)}{\partial \hat{\alpha}_1} \Bigg|_{\hat{\alpha} = M\alpha} . \qquad (4.28)$$

Substituting (4.28) in (4.25) we obtain

$$\hat{f}(\mathbf{x}) = \frac{1}{4\pi^2} \int_0^{2\pi} \left[\|\phi'(\theta)\| \frac{\partial g_r(\hat{\alpha}, \theta)}{\partial \hat{\alpha}_1} *_{\hat{\alpha}} \delta(\hat{\alpha}_2) \delta(\hat{\alpha}_3) \left(\frac{1}{\hat{\alpha}_1} \right) \right]_{\hat{\alpha} = M(\mathbf{x} - \phi(\theta))} d\theta.$$

The 3-D convolution now reduces to a one dimensional convolution in $\hat{\alpha}_1$; with $\|\phi'(\theta)\| = R$, then,

$$\hat{f}(\mathbf{x}) = \frac{R}{4\pi^2} \int_0^{2\pi} \left[\frac{\partial g_r(\hat{\alpha}, \theta)}{\partial \hat{\alpha}_1} *_{\hat{\alpha}_1} \left(\frac{1}{\hat{\alpha}_1} \right) \right]_{\hat{\alpha} = M(\mathbf{x} - \phi(\theta))} d\theta$$

where $*_{\hat{\alpha}_1}$ denotes a 1-D convolution with respect to $\hat{\alpha}_1$. Finally, using Cauchy principal value and integrating by parts, see [27], it is straight forward to show the following in the generalized sense:

$$\hat{f}(\mathbf{x}) = \frac{R}{4\pi^2} \int_0^{2\pi} \lim_{\varepsilon \to 0} [g_r(\hat{\alpha}, \theta) *_{\hat{\alpha}_1} E_\varepsilon(\hat{\alpha}_1)]_{\hat{\alpha} = M(\mathbf{x} - \phi(\theta))} d\theta, \qquad (4.29)$$

where

$$E_\varepsilon(\hat{\alpha}_1) = \begin{cases} 1/\varepsilon^2 & |\hat{\alpha}_1| < \varepsilon \\ -1/\hat{\alpha}_1^2 & |\hat{\alpha}_1| \geq \varepsilon. \end{cases} \qquad (4.30)$$

The result (4.29) is now a reconstruction formula for $\hat{f}(\mathbf{x})$ requiring 1-D filtering of the projection data in $\hat{\alpha}_1$ followed by a 3-D backprojection.

4.2.2 Point Spread Function of the Approximate Reconstruction

To reveal the inherent approximation associated with the algorithm and to determine the quantitative effect of this approximation on the reconstruction, we consider the resulting spatially variant PSF. Consider a point image at $\mathbf{x} = \mathbf{x}_0$, i.e. $f(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0)$, where $\delta(\mathbf{x})$ denotes the Dirac delta function. Defining $\boldsymbol{\alpha}_0 = \mathbf{x}_0 - \boldsymbol{\phi}(\theta) = (\alpha_{01}, \alpha_{02}, \alpha_{03})^T$, and $\hat{\boldsymbol{\alpha}}_0 = M(\mathbf{x}_0 - \boldsymbol{\phi}(\theta))$, then:

$$\begin{pmatrix} \hat{\alpha}_{01} \\ \hat{\alpha}_{02} \\ \hat{\alpha}_{03} \end{pmatrix} = \begin{pmatrix} -x_{01}\sin\theta + x_{02}\cos\theta \\ -x_{01}\cos\theta - x_{02}\sin\theta + R \\ x_{03} \end{pmatrix}. \tag{4.31}$$

In deriving the PSF, we will assume that the object of interest is confined to the interior of a sphere centered at $\mathbf{x} = \mathbf{0}$. Consequently, from our definition of $\hat{\boldsymbol{\alpha}}$, it follows that $\hat{\alpha}_{02} \neq 0$. The projections of the point image are derived in the Appendix as (4.59):

$$g_r(\hat{\boldsymbol{\alpha}}, \theta) = \frac{|\hat{\alpha}_2|}{\hat{\alpha}_{02}^2} \delta \left(\hat{\alpha}_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}} \hat{\alpha}_2 \right) \delta \left(\hat{\alpha}_3 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}} \hat{\alpha}_2 \right). \tag{4.32}$$

Substituting (4.32) into (4.29) and applying property (4.57) to the convolution integral, the PSF becomes:

$$\begin{aligned} \hat{f}_p(\mathbf{x}) &= \frac{R}{4\pi^2} \int_0^{2\pi} \\ &\quad \frac{|\hat{\alpha}_2|}{\hat{\alpha}_{02}^2} \delta \left(\hat{\alpha}_3 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}} \hat{\alpha}_2 \right) \lim_{\epsilon \to 0} E_\epsilon \left[\hat{\alpha}_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}} \hat{\alpha}_2 \right]_{|\hat{\boldsymbol{\alpha}} = M(\mathbf{x} - \boldsymbol{\phi}(\theta))} d\theta. \end{aligned} \tag{4.33}$$

Let us define new variables

$$\begin{aligned} a &= \hat{\alpha}_1 - \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{02}} \hat{\alpha}_2 \\ b &= \hat{\alpha}_3 - \frac{\hat{\alpha}_{03}}{\hat{\alpha}_{02}} \hat{\alpha}_2. \end{aligned} \tag{4.34}$$

6.3 A New Algorithm for PET Image Reconstruction

This section presents a MAP estimation algorithm based on the generalized EM algorithm using the MRF model with a joint Gibbs distribution of emission intensities and line processes. The M-step updates both the intensity and line processes, achieved in two stages. The M1-step employs a gradient descent method to update the intensities. The M2-step updates the line sites in a similar manner to the ICA (iterated conditional averages) proposed by Johnson et $al.$ [30].

Due to the relatively complex nature of the brain anatomical structure (Figure 7.7(a) in Chapter 7), we consider the second-order neighbors for the intensity, horizontal and vertical and diagonal line fields in the following derivation. In the following, superscripts on the estimates are omitted, because we use both previous and current estimates in intensity and line updating schemes to achieve a faster convergence rate.

A suitable energy function for an eight nearest neighbor interaction model is suggested as follows[1]:

$$
\begin{aligned}
U(\lambda, l) = & \; \textstyle\sum_{i,j} \left\{ \beta \lambda_x^2(i,j)(1 - l_y(i,j)) + \alpha l_y(i,j) \right\} + \\
& \; \textstyle\sum_{i,j} \left\{ \beta \lambda_y^2(i,j)(1 - l_x(i,j)) + \alpha l_x(i,j) \right\} + \\
& \; \textstyle\sum_{i,j} \left\{ (\beta/\sqrt{2}) \lambda_{rx}^2(i,j)(1 - l_{ry}(i,j)) + \alpha l_{ry}(i,j) \right\} + \\
& \; \textstyle\sum_{i,j} \left\{ (\beta/\sqrt{2}) \lambda_{ry}^2(i,j)(1 - l_{rx}(i,j)) + \alpha l_{rx}(i,j) \right\} + \mathcal{H}(l),
\end{aligned}
\tag{6.21}
$$

where $l_x(i,j)$, $l_y(i,j)$, $l_{rx}(i,j)$ and $l_{ry}(i,j)$ represent horizontal, vertical and two diagonal binary line sites respectively. Note that the M2-step of this algorithm cannot be based on gradient descent, more flexibility is therefore allowed in the choice of the potential function. In the following we use the index p to denote any one from $\{x, y, rx, ry\}$, $i.e.$ $p = \{x, y, rx, ry\}$. In each pair of braces, the first term reflects our belief that the image values should not change abruptly except at discontinuities, $i.e.$ when $l_p(i,j) = 1$; the second term penalizes each non-zero

---

[1]For a finite lattice, care must be taken in defining potential functions on cliques near the image boundary. For notational simplicity, in the following we assume that appropriate modifications are made to the potential functions at the image boundary.

line site in the image; and the term $\mathcal{H}(l)$ denotes a measure of interaction between line sites which encourages the formation of connected lines in the reconstructed image and suppresses the formation of isolated or unconnected lines. $\lambda_x(i,j)$, $\lambda_y(i,j)$, $\lambda_{rx}(i,j)$ and $\lambda_{ry}(i,j)$ denote the differences between adjacent horizontal, vertical and diagonal intensities:

$$
\begin{aligned}
\lambda_x(i,j) &= \lambda(i-1,j) - \lambda(i,j) \\
\lambda_y(i,j) &= \lambda(i,j-1) - \lambda(i,j) \\
\lambda_{rx}(i,j) &= \lambda(i-1,j-1) - \lambda(i,j) \\
\lambda_{ry}(i,j) &= \lambda(i+1,j-1) - \lambda(i,j).
\end{aligned} \tag{6.22}
$$

We employ the following potential form as a line interaction term:

$$
\begin{aligned}
\mathcal{H}(l) = -\epsilon\alpha \sum_{i,j} \{ &[l_y(i,j-1)l_{rx}(i-1,j) + l_y(i,j)l_{ry}(i,j) \\
&+ l_y(i+1,j-1)l_{ry}(i+1,j) + l_y(i+1,j)l_{rx}(i,j)] \, l_x(i,j) \\
&- l_x(i-1,j) [l_{rx}(i-1,j) + l_{ry}(i,j) - 1] \, l_x(i,j) \\
&- l_y(i,j-1) [l_{rx}(i-1,j) + l_{ry}(i,j) - 1] \, l_y(i,j) \},
\end{aligned} \tag{6.23}
$$

where $0 < \epsilon < 1$. This type of interaction term is similar to that discussed in [12] [45]. This function serves the purpose of increasing the penalty on a horizontal (vertical) line if either of its horizontal (vertical) neighbors on the same rows (columns) and either of the diagonal neighbors in between are simultaneously "on", and decreasing the penalty on a horizontal (vertical) line if anyone of its four vertical (horizontal) neighbors and the connecting diagonal line are simultaneously "on". It also serves the purpose of increasing the penalty on a diagonal line if both of its horizontal or vertical neighbors are "on" and decreasing the penalty on a diagonal line if its horizontal and vertical neighbors to be connected are simultaneously "on". The parameter ϵ controls the relative penalty associated with this term. We note these functions encourage connected boundaries and smoothed corners and as the number of orientations is increased, the function above can be modified to allow better control of the formation of boundaries.

We now derive the updating procedures for the M1-step and M2-step by following the general recipe (6.18) and (6.20). The Gibbs energy function $U(\lambda, l)$ as defined in (6.21), results in a function $Q(\lambda, l^n | \lambda^n, l^n)$ that is concave with respect

to each $\lambda(i,j) \geq 0$. Hence in the M1-step, we use the concave nature of the function to perform a sequence of one dimensional line searches, updating each of the intensity variables in turn. This simple approach ensures that the function $Q(\lambda^{n+1}, l^n | \lambda^n, l^n)$ is monotonically nondecreasing at each M1-step.

M1-step: Setting the partial derivative of $Q(\lambda, l^n | \lambda^n, l^n)$, defined in (6.17), with respect to $\lambda(i,j)$ equal to zero and simplifying, yields the quadratic equation:

$$C_1 \lambda^2(i,j) - [C_2 - a(i,j)] \lambda(i,j) - b^n(i,j) = 0, \tag{6.24}$$

where C_1 and C_2 are defined as follows:

$$
\begin{aligned}
C_1 =\ & 2\beta\gamma \left[(1 - l_x^n(i,j)) + (1 - l_x^n(i,j+1)) + \right. \\
& \quad (1 - l_y^n(i,j)) + (1 - l_y^n(i+1,j)) \Big] + \\
& \sqrt{2}\beta\gamma \left[(1 - l_{rx}^n(i,j)) + (1 - l_{rx}^n(i-1,j+1)) + \right. \\
& \quad (1 - l_{ry}^n(i,j)) + (1 - l_{ry}^n(i+1,j+1)) \Big] \\
C_2 =\ & 2\beta\gamma \left[\lambda(i,j-1)(1 - l_x^n(i,j)) + \lambda(i,j+1)(1 - l_x^n(i,j+1)) + \right. \\
& \quad \lambda(i-1,j)(1 - l_y^n(i,j)) + \lambda(i+1,j)(1 - l_y^n(i+1,j)) \Big] + \\
& \sqrt{2}\beta\gamma \left[\lambda(i+1,j-1)(1 - l_{rx}^n(i,j)) + \right. \\
& \quad \lambda(i-1,j+1)(1 - l_{rx}^n(i-1,j+1)) + \\
& \quad \lambda(i-1,j-1)(1 - l_{ry}^n(i,j)) + \\
& \quad \lambda(i+1,j+1)(1 - l_{ry}^n(i+1,j+1)) \Big].
\end{aligned}
\tag{6.25}
$$

When computing C_1 and C_2, the most recent estimates of $\lambda(i \pm 1, j \pm 1)$ should be used since the pixel intensities are updated sequentially. Since $\lambda(i,j) \geq 0$, we must choose the non-negative root of (6.24). Assuming C_1 is non-zero, the updating scheme for $\lambda(i,j)$ becomes:

$$
\lambda^{nr}(i,j) = \frac{[C_2 - a(i,j)] + \sqrt{[C_2 - a(i,j)]^2 + 4C_1 b^n(i,j)}}{2C_1}, \quad n_r \Leftarrow n_r + 1;
$$
$$
\lambda^{n+1}(i,j) \Leftarrow \lambda^{nr}(i,j), \quad \text{after } r \text{ iterations.} \tag{6.26}
$$

In this scheme we update each pixel r times. The estimates λ^{nr} are then assigned to λ^{n+1}.

Since this procedure ensures that the gradient of $Q(\lambda, l^n | \lambda^n, l^n)$ with respect to $\lambda(i,j)$ is zero at $\lambda^{n+1}(i,j)$ and the function is concave with respect to $\lambda(i,j) \geq 0$, it follows that the above iteration, when applied to all intensity sites in turn, results in an increase in the function $Q(\lambda, l^n | \lambda^n, l^n)$, i.e. condition (6.18) is satisfied. This result is summarized in the following lemma:

Lemma 6.1 *(i). (6.26) is the only non-negative root of (6.24), since $4C_1 b^n(i,j)$ is always nonnegative. (ii). $Q(\lambda, l^n | \lambda^n, l^n)$ is concave and has a unique maximum solution (6.26) with respect to each $\lambda(i,j) \geq 0$ at each iteration of the M1-step.*

The case where $C_1 = 0$ implies that all line sites involved in C_1 are "on". In this case C_2 is also equal to zero and the M1-step at this site reduces to $\lambda^{nr}(i,j) = b^n(i,j)/a(i,j)$. In other words, one would expect no intensity smoothing between neighbors to occur if all the lines around the pixel are "on".

We now turn to the problem of updating the line sites. Our goal is to obtain a result in which the estimated line sites converge to the values 1 or 0 that, jointly with λ, maximize the posterior density function. We rule out a stochastic search as too computationally expensive and restrict our attention to deterministic algorithms. During the M2-step we wish to update the line processes so as to satisfy (6.19), or equivalently (6.20). The method of iterated conditional mode (ICM) [2] can be shown to produce a sequence with an associated energy function that is nonincreasing and would therefore satisfy (6.20). Combining this ICM approach with the M1-step defined above, we would expect convergence of this algorithm to a local maximum of the posterior distribution. However, when the ICM method is applied to the estimation of line processes, it has been observed that the line estimates converge faster than the image intensities [45] [30]. It is intuitively more appealing to have an updating procedure in which the rate of convergence of the line process can be controlled by allowing the variables $l_p(i,j)$ to assume values on the continuous interval $[0,1]$ but eventually converge to one of the end points. In this case, a zero-one decision about the presence of a boundary is delayed until later in the iteration process. By replacing the conditional mode in the ICM algorithm with a conditional average (ICA), Johnson *et al.* [30] developed an iterative algorithm which does allow the line process to take values in the interval $[0,1]$. However, there is no guarantee in the case of

ICA that the line process converges to the values 0 or 1. Furthermore, is it not proven that ICA necessarily converges to a (local) maximum.

In developing the M2-step below, we adopt a third alternative to ICM and ICA. Recall that the goal of the M2-step is to update the line sites such that $U(\lambda^{n+1}, l^{n+1}) \leq U(\lambda^{n+1}, l^n)$. Here we use the iterated conditional average of a modified energy function with fixed $\lambda = \lambda^{n+1}$: $U_T(\lambda^{n+1}, l) = U(\lambda^{n+1}, l)/T$, where $U(\lambda^{n+1}, l)$ is defined in (6.21). Fixing $T = 1$ would produce an iteration equivalent to the ICA algorithm [30]. However, as $T \to 0$, it can be shown that the variables $l_p(i,j)$ converge to either 0 or 1 and this algorithm would converge to a local minimum of the energy function $U(\lambda^{n+1}, l)$ with respect to l on the set $\{0,1\}^L$. The goal of this procedure is to use the parameter T to control the rate of convergence of the line process relative to the intensity process and hopefully to then converge on a more desirable local maximum of the posterior distribution.

M2-step: Let $N_p(i,j)$ denote the set of intensities and lines which interact with $l_p(i,j)$. The local conditional probability of $l_x(i,j)$, $l_y(i,j)$, $l_{rx}(i,j)$ and $l_{ry}(i,j)$ for the energy function $U_T(\lambda^{n+1}, l)$ is given by

$$P_T(l_x(i,j)|N_x(i,j)) = \frac{e^{-\frac{1}{T}\left\{\beta\lambda_y^{n+1}(i,j)^2(1-l_x(i,j))+\alpha l_x(i,j)-\epsilon\alpha S_x(i,j)l_x(i,j)\right\}}}{e^{-\frac{1}{T}\beta\lambda_y^{n+1}(i,j)^2} + e^{-\frac{1}{T}\{\alpha-\epsilon\alpha S_x(i,j)\}}}$$

$$P_T(l_y(i,j)|N_y(i,j)) = \frac{e^{-\frac{1}{T}\left\{\beta\lambda_x^{n+1}(i,j)^2(1-l_y(i,j))+\alpha l_y(i,j)-\epsilon\alpha S_y(i,j)l_y(i,j)\right\}}}{e^{-\frac{1}{T}\beta\lambda_x^{n+1}(i,j)^2} + e^{-\frac{1}{T}\{\alpha-\epsilon\alpha S_y(i,j)\}}}$$

(6.27)

$$P_T(l_{rx}(i,j)|N_{rx}(i,j)) = \frac{e^{-\frac{1}{T}\left\{\frac{\beta}{\sqrt{2}}\lambda_{ry}^{n+1}(i,j)^2(1-l_{rx}(i,j))+\alpha l_{rx}(i,j)-\epsilon\alpha S_{rx}(i,j)l_{rx}(i,j)\right\}}}{e^{-\frac{1}{T\sqrt{2}}\beta\lambda_{ry}^{n+1}(i,j)^2} + e^{-\frac{1}{T}\{\alpha-\epsilon\alpha S_{rx}(i,j)\}}}$$

$$P_T(l_{ry}(i,j)|N_{ry}(i,j)) = \frac{e^{-\frac{1}{T}\left\{\frac{\beta}{\sqrt{2}}\lambda_{rx}^{n+1}(i,j)^2(1-l_{ry}(i,j))+\alpha l_{ry}(i,j)-\epsilon\alpha S_{ry}(i,j)l_{ry}(i,j)\right\}}}{e^{-\frac{1}{T\sqrt{2}}\beta\lambda_{rx}^{n+1}(i,j)^2} + e^{-\frac{1}{T}\{\alpha-\epsilon\alpha S_{ry}(i,j)\}}}.$$

The terms $\mathcal{S}_x(i,j)$, $\mathcal{S}_y(i,j)$, $\mathcal{S}_{rx}(i,j)$ and $\mathcal{S}_{ry}(i,j)$ are defined as

$$
\begin{aligned}
\mathcal{S}_x(i,j) &= l_y(i,j-1)l_{rx}(i-1,j) + l_y(i,j)l_{ry}(i,j)+ \\
&\quad l_y(i+1,j)l_{rx}(i,j) + l_y(i+1,j-1)l_{ry}(i+1,j)- \\
&\quad l_x(i-1,j)\,[l_{rx}(i-1,j) + l_{ry}(i,j) - 1] - \\
&\quad l_x(i+1,j)\,[l_{rx}(i,j) + l_{ry}(i+1,j) - 1] \\
\mathcal{S}_y(i,j) &= l_x(i-1,j)l_{rx}(i-1,j) + l_x(i,j)l_{ry}(i,j)+ \\
&\quad l_x(i,j+1)l_{rx}(i-1,j+1) + l_x(i-1,j+1)l_{ry}(i,j+1)- \\
&\quad l_y(i,j-1)\,[l_{rx}(i-1,j) + l_{ry}(i,j) - 1] - \\
&\quad l_y(i,j+1)\,[l_{rx}(i-1,j+1) + l_{ry}(i,j+1) - 1] \\
\mathcal{S}_{rx}(i,j) &= l_x(i,j)l_y(i+1,j) + l_x(i+1,j)l_y(i+1,j-1)- \\
&\quad l_x(i,j)l_x(i+1,j) - l_y(i+1,j)l_y(i+1,j-1) \\
\mathcal{S}_{ry}(i,j) &= l_x(i,j)l_y(i,j) + l_x(i-1,j)l_y(i,j-1)- \\
&\quad l_x(i-1,j)l_x(i,j) - l_y(i,j-1)l_y(i,j).
\end{aligned}
\tag{6.28}
$$

The purpose of the interaction term $\mathcal{H}(l)$ defined in (6.23) are further elucidated from the line updating scheme (6.29) and the definition (6.28).

The local conditional mean of the binary random variable $l_p(i,j)$ is then equal to the probability that $l_p(i,j) = 1$. Using the local conditional mean, the updating procedure for $l_p(i,j)$ is defined as

$$
\begin{aligned}
l_x^{n_q}(i,j) &\equiv E[l_x(i,j)|N_x(i,j)] = \frac{1}{1 + e^{\frac{1}{T}[\alpha - \beta\lambda_y^{n+1}(i,j)^2 - \epsilon\alpha\mathcal{S}_x(i,j)]}}, \quad n_q \Leftarrow n_{q+1}; \\
l_y^{n_q}(i,j) &\equiv E[l_y(i,j)|N_y(i,j)] = \frac{1}{1 + e^{\frac{1}{T}[\alpha - \beta\lambda_x^{n+1}(i,j)^2 - \epsilon\alpha\mathcal{S}_y(i,j)]}}, \quad n_q \Leftarrow n_{q+1}; \\
l_{rx}^{n_q}(i,j) &\equiv E[l_{rx}(i,j)|N_{rx}(i,j)] = \frac{1}{1 + e^{\frac{1}{T}[\alpha - \frac{\beta}{\sqrt{2}}\lambda_{ry}^{n+1}(i,j)^2 - \epsilon\alpha\mathcal{S}_{rx}(i,j)]}}, \quad n_q \Leftarrow n_{q+1}; \\
l_{ry}^{n_q}(i,j) &\equiv E[l_{ry}(i,j)|N_{ry}(i,j)] = \frac{1}{1 + e^{\frac{1}{T}[\alpha - \frac{\beta}{\sqrt{2}}\lambda_{rx}^{n+1}(i,j)^2 - \epsilon\alpha\mathcal{S}_{ry}(i,j)]}}, \quad n_q \Leftarrow n_{q+1}; \\
l_p^{n+1}(i,j) &\Leftarrow l_p^{n_q}(i,j), \quad \text{after } q \text{ iterations; reduce } T.
\end{aligned}
\tag{6.29}
$$

When computing (6.29) the most recent estimates are used for the neighboring line sites, $l_p(i \pm 1, j \pm 1)$.

In general, the updating procedure (6.29) defined above does not guarantee that $U_T(\lambda^{n+1}, l)$ increases at each M2-step for a fixed T. In addition, the estimates

from (6.29) vary with the parameter T. This is an motivation of introducing the parameter T, since by changing the values of T, one can avoid line estimates to be trapped at a local minimum too fast. It can be shown that (6.29) converges to a (local) minimum of the function $U(\lambda^{n+1}, l)$ as $T \to 0$ during the M2-step. The advantage of (6.29) is that we are able to control this optimization process through T and we may find a better (local) maxima as T is gradually reduced to zero over several M2-steps.

Lemma 6.2 *(i). The line estimate $l_p^n(i,j)$ obtained using (6.29) satisfies the condition*

$$0 \leq l_p^n(i,j) \leq 1 \quad \text{for any } T, \ n = 0, 1, 2, \cdots. \tag{6.30}$$

(ii). For fixed λ, the iteration (6.29) converges to either 0 or 1 as $T \to 0$. (iii). For fixed λ, as $T \to 0$ the iteration (6.29) becomes the iterated conditional modes (ICM) algorithm with associated energy function $U(\lambda, l)$ and therefore converges to a local minimum of $U(\lambda, l)$ at the M2-step on the set $\{0, 1\}^L$ where L denotes the number of line process variables.

The proof for Lemma 6.2 is given in Appendix 6.A.

6.4 Experimental Results

In order to evaluate and validate the overall performance of the proposed method to PET image reconstructions, we have applied this new algorithm to patient data collected for cerebral glucose metabolism study using [18]FDG and dopamine receptor study using F-DOPA. These patient data, provided by the Division of Nuclear Medicine at UCLA, were obtained using a Siemens ECAT831 brain scanner equipped with a septal collimator. Each detector ring of the scanner has a total of 320 detector crystals. These data were collected with automatic randoms subtractions, where random events were determined through the use of a delayed time window. Each view of projection data was formed by interlacing two sets of adjacent views of projections to increase the detector resolution. So these data were treated as a total of 160 parallel projections with 128 samples per projection.

In all of the reconstructions shown below, negative detections after subtractions are then set to zero. It is followed by rebinning the coincident detections into what they were originally detected, *i.e.* a total of 320 parallel projections with 64 samples per projection. No scatter measurements or corrections were performed in this experiment. Corrections for detector response and attenuation were made and used in all of the reconstructions shown.

Images were reconstructed for two different sets of data, corresponding to patient [18]FDG and F-DOPA study. The [18]FDG data used here were collected for a total of 30 minutes, from which images were reconstructed for two different two-dimensional cross sections through the patient brain. In order to show somewhat dynamic changes of F-DOPA inside a patient brain, the F-DOPA data were used as follows. The first 18 minutes data were discarded and then the remaining collections were sorted into two sets of data corresponding to the second 32 minutes collection (set 1) and the last 90 minutes collection (set 2). A single two-dimensional cross section through the patient brain were then reconstructed from these two sets of F-DOPA data respectively.

For comparison purpose, we have shown reconstructions of corresponding cross sections of patient brains for each set of data using: filtered back-projection, maximum likelihood, maximum *a posteriori* (MAP), in addition to the reconstructions of patient brains using the new algorithm discussed in the previous section. The MAP reconstructions were obtained using the GEM algorithm [20] with the second order neighborhood of the Gibbs energy function proposed by Geman and Mc-Clure [14]. These images clearly demonstrate the advantage of statistical methods to PET reconstruction over the filtered back-projection. Shown in Figures 6.2 and 6.3 are reconstructions of two different cross sections through the patient brain for the data set of [18]FDG study. In Figures 6.4 and 6.5 we show reconstructions of a cross section of brain for the two sets of F-DOPA data. The images in Figure 6.4 may offer some useful information that the spatial distributions during that interval of time are similar to those from [18]FDG studies, while the images in Figure 6.5 are of more interest in the F-DOPA study. These images illustrate dynamic changes occurred during the collection to some extent. The MAP results are a little smoother than the ML images, but overall, the results obtained using
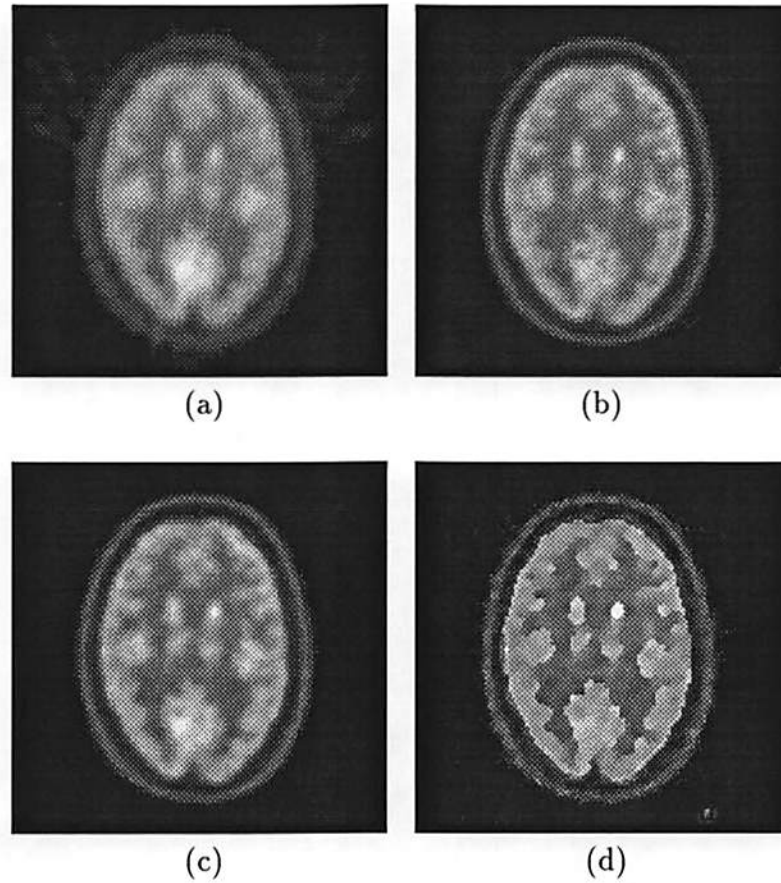
Figure 6.2: Reconstructions for a single two-dimensional cross section through a patient brain after 30 minutes ^{18}FDG data collection. (a): PET reconstruction using filtered backprojection. (b): PET reconstruction using maximum likelihood estimation. (c): PET reconstruction using MAP estimation without a line process. (d): PET reconstruction using the proposed MAP estimation with an estimated line process.
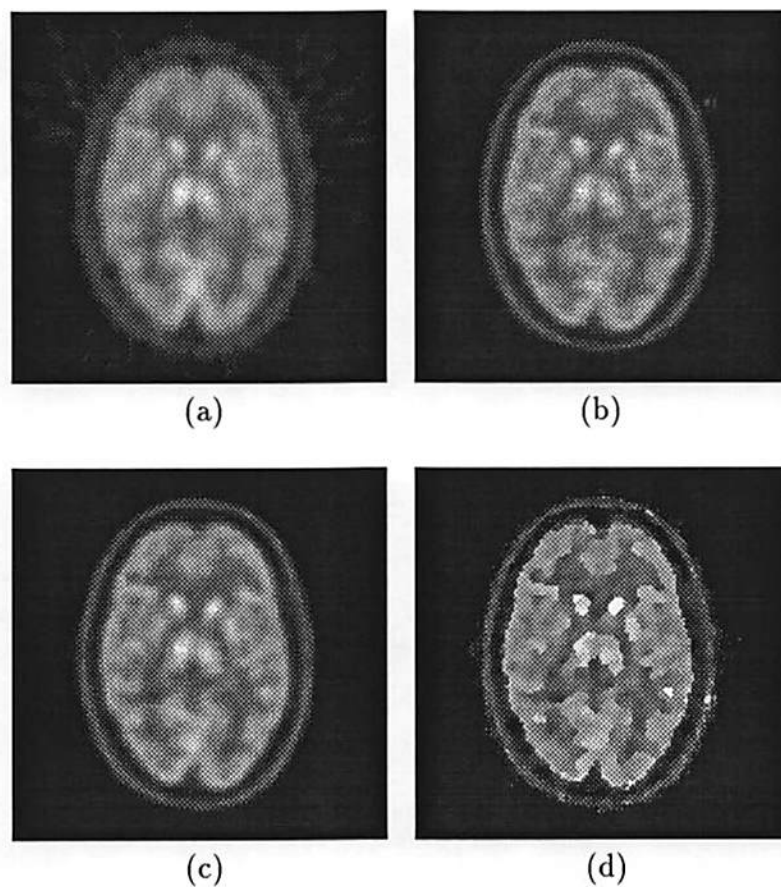
Figure 6.3: Reconstructions for another cross section through the same patient brain as in Figure 6.2 after 30 minutes ^{18}FDG data collection. (a): PET reconstruction using filtered backprojection. (b): PET reconstruction using maximum likelihood estimation. (c): PET reconstruction using MAP estimation without a line process. (d): PET reconstruction using the proposed MAP estimation with an estimated line process.
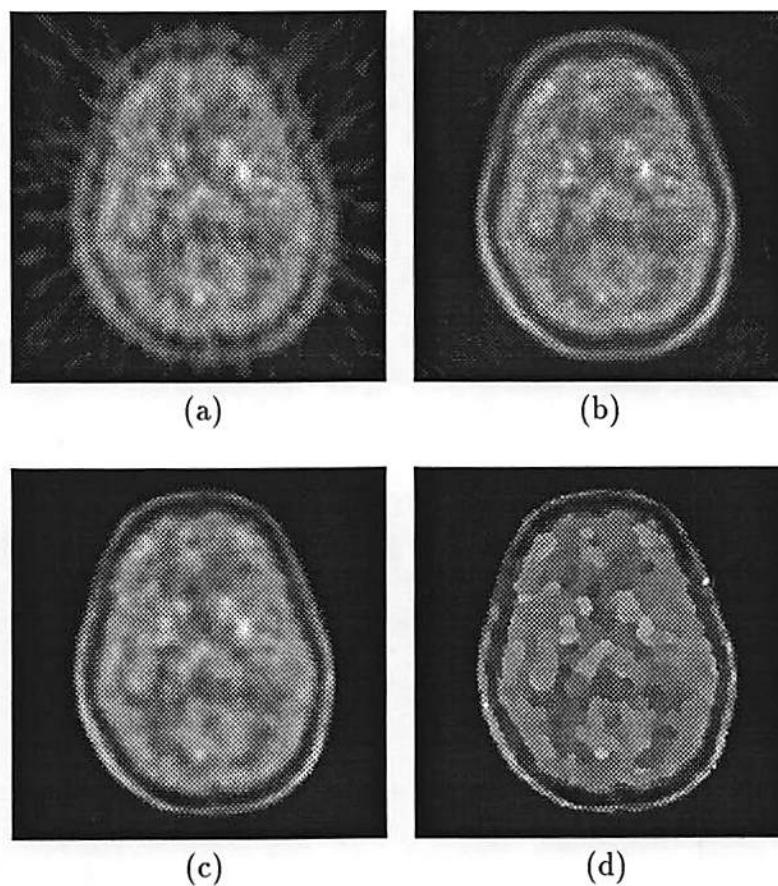
Figure 6.4: Reconstructions for a single two-dimensional cross section through a patient brain after the second period of 32 minutes F-DOPA data collection (set 1). (a): PET reconstruction using filtered backprojection. (b): PET reconstruction using maximum likelihood estimation. (c): PET reconstruction using MAP estimation without a line process. (d): PET reconstruction using the proposed MAP estimation with an estimated line process.

(a)

(b)

(c)

(d)
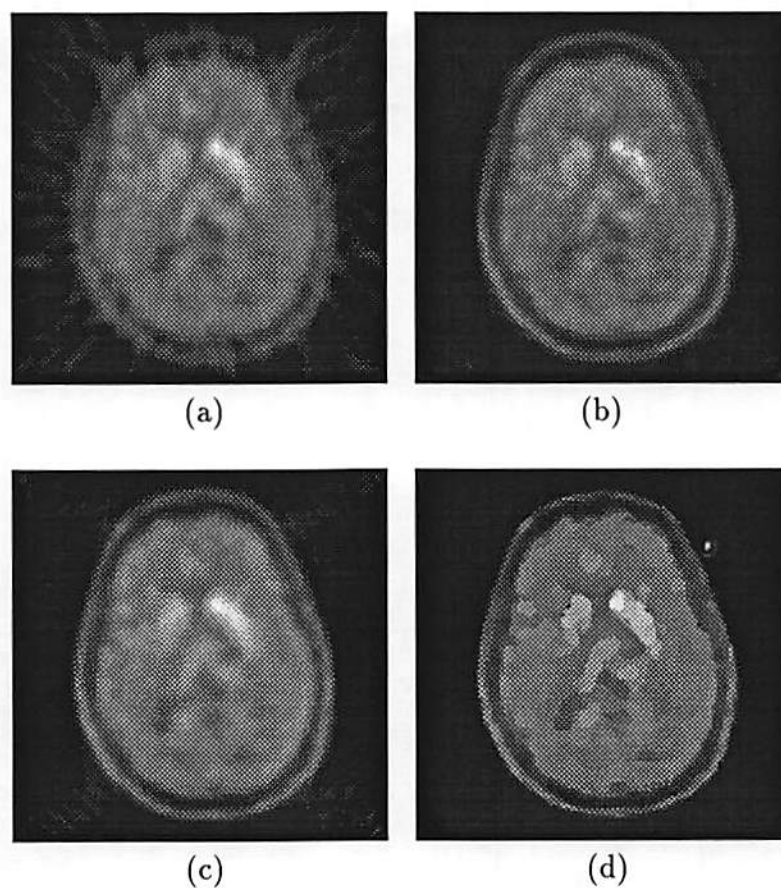
Figure 6.5: Reconstructions for the same cross section as in Figure 6.4 after the last period of 90 minutes F-DOPA data collection (set 2). (a): PET reconstruction using filtered backprojection. (b): PET reconstruction using maximum likelihood estimation. (c): PET reconstruction using MAP estimation without a line process. (d): PET reconstruction using the proposed MAP estimation with an estimated line process.

an MRF with estimated line sites are superior to the previous results with little extra computational cost comparing with that of ML or MAP results. There is now noticeable grey/white matter contrast, the images appear to have more fine details although the shape of some of the regions appears somewhat forced. The control parameter T in this experiment is reduced according to the following rule, similar to that given by Simchony et $al.$ [51]:

$$T_k = \frac{T_0}{\log(1 + 10k)^{1.5}} \tag{6.31}$$

where k is the iteration number and $T_0 = 2.01$.

6.A Appendix

Results in Lemma 6.2 (i) and (ii) are straightforward, the proof for the part (iii) is provided in the following.

Only the estimated horizontal line site $l_x(i,j)$ are considered below, but the same conclusion can be drawn for $l_y(i,j)$, $l_{rx}(i,j)$ and $l_{ry}(i,j)$. We now introduce a new notation Δ_x and let

$$\Delta_x = \alpha - \beta \lambda_y^{n+1}(i,j)^2 - \epsilon \alpha S_x. \tag{6.32}$$

At $T = 0$, the updating procedure (6.29) for $l_x(i,j)$ becomes

$$l_x^{n_q}(i,j) = \begin{cases} 0 & \text{if } \Delta_x \geq 0, \\ 1 & \text{if } \Delta_x < 0. \end{cases} \tag{6.33}$$

By using the notation Δ_x, the term which interacts with $l_x(i,j)$ in the energy function $U(\lambda, l)$ can be written as

$$e^{-\Delta_x l_x(i,j)}. \tag{6.34}$$

Therefore, if one applies the ICM algorithm to estimate each $l_x(i,j)$ with the energy function $U(\lambda, l)$, the solution at n_q-th iteration would be identical to (6.33).

Hence (6.29) reduces to the ICM and converges to a (local) minimum of $U(\lambda, l)$ as $T \rightarrow 0$.

Chapter 6

Statistical Models and Methods for PET Image Reconstruction

In this chapter, we concentrate on a statistical reconstruction approach to PET. As mentioned previously, one important thing that distinguished PET reconstruction problem from that arising from X-ray CT is that the measurements tend to be much more noisy. Therefore it is desirable to develop an estimation approach that considers the statistical nature of data and to accurately model the data so that correction factors, such as attenuation, scatter, point or line spread function can be systematically included. A iterative or statistical approach appears to to be the top candidate for fulfilling these desires. In addition, the computation required in PET reconstruction is far from trivial because there are a large amount of information involved: the algorithms are in general iterative; the minimum dimension of this problem, even for brain scanning, is $15 \times 128 \times 128$ based on 15×320 projections, each consisting of a 64 array of photon counts. For this reason, it is essential to develop a fast and efficient reconstruction algorithm that incorporates the statistical properties of data.

We follow a Bayesian formulation introduced by Besag [2] and Geman and Geman [13]. Based on the belief that image values should not change abruptly except at discontinuities, the image is model using a joint Gibbs distribution of emission intensities and line processes. The line process is used to avoid smoothing cross discontinuities, which commonly occurs in Bayesian image estimation where

a line process is not included. We describe a MAP (maximum *a posteriori*) estimation algorithm using the above model. This algorithm employs the generalized EM (GEM) algorithm to avoid direct optimization over the posterior distribution which does not share the Markovian property of the prior. The M-step in the MAP estimation procedure has the interesting property that the function to be maximized at this step is with the same neighborhood as the prior. Thus the optimization in the M-step requires only local computations on the image lattice. This M-step is achieved in two stages. In the M1-step a gradient descent method is used to update intensities with the non-negativity guarantee, while holding the line process constant. An alternative M2-step based on the local conditional mean is described to update line sites, with the intensity process held constant. The parameter T is introduced in the M2-step so that the rate of convergence of the unknown line sites can be controlled and we may fine a better (local) maxima as T is gradually reduced to zero over several M2-steps.

In the following sections, we develop a statistical model for the PET imaging system. We then develop MAP estimation schemes for MRF priors within the framework of the generalized EM algorithm. We conclude this chapter with an illustration of the performance of these estimators as applied to a set of patient data collected using a Siemens ECAT831 brain scanner.

6.1   PET System Modeling

An important requirement for quantitative PET imaging is that one should have an accurate statistical description for the data. Let y_k, the k-th element of the observed data Y, denote the number of coincidences detected between the k-th detector pair. Let λ denote an $N \times N$ array of mean emission rates from the source, thus $\lambda(i,j)$ represents the mean emission rate from pixel (i,j) of the source image.

Consider the number of coincidences y_k detected between detector pair k. This number is the sum of three components: t_k – the number of true coincidence events detected; s_k – the number of events detected after one or both of the primary photons has undergone scattering; and r_k – the number of random coincidences, *i.e.* two independent photons arriving at the detector pair within the coincidence

window and counted as a true coincidence. Then:

$$y_k = t_k + s_k + r_k. \tag{6.1}$$

Each of the three components in (6.1) is well modeled as a Poisson process. The true unscattered coincidences, t_k, form the desired data, $i.e.$ the tomographic projection data or "sinogram". The latter two components introduce additive Poisson noise into the data. Let us consider each of the three components in (6.1) in more detail.

True coincidences: the probability, $P_k(i,j)$ of detecting a true coincidence (photon pair) at detector pair k due to a positron emission from pixel (i,j) is dependent on a number of factors. The most important of these are: the probability of attenuation, the response of the detectors, and the position of pixel (i,j) relative to the detector pair k. We can model t_k as a Poisson random variable with mean:

$$\mathbf{E}\{t_k\} = \sum_{i,j} P_k(i,j)\lambda(i,j). \tag{6.2}$$

We can factor the detection probability as follows: $P_k(i,j) = a_k g_k P_k^g(i,j)$ where $P_k^g(i,j)$ denotes the probability that a positron emitted from site (i,j) produces a photon pair that, in the absence of an attenuating medium, strike detector pair k; g_k denotes the probability that a photon pair striking detector pair k is detected; and a_k denotes the probability of attenuation of photons that would otherwise reach detector pair k. The first of these factors can be determined to reasonable accuracy from simple geometrical considerations. The probabilities of detection and attenuation are determined experimentally. Inclusion of these factors in either the standard filtered backprojection algorithm or in statistical models is straightforward, and they are included in the experimental studies presented below. There are other factors which affect the detection probability such as detector dead time and positron range and angular separation. Consideration of these factors is beyond the scope of this thesis.

Scattered coincidences: these are coincidence detections which occur after one or both of the primary photons produced by a positron/electron annihilation is scattered within the body. If one does not consider the possibility of scatter in

developing the data model, then the annihilation is considered to have occurred along the strip joining the two detectors - as would be the case for unscattered coincidences. In practice the scattered component of the data is usually smoothly varying between detector pairs. In clinical applications, a scatter fraction is often estimated and subtracted from the data before image reconstruction. In many cases, the presence of scatter is simply ignored. Due to the smoothness of the scatter component, the practical effect of this is to produce a positive bias in the reconstructed image. For quantitative imaging, the scatter must be included in the formulation. In our statistical model, we could exploit the smoothness of the scatter function to model the scatter component across a single parallel projection parametrically. The parameters of this model could then be estimated in conjunction with the image intensities within the EM framework. A more sophisticated model would attempt to relate the scatter component to the source activity, however this is probably infeasible due to the highly complex nature of the scattering mechanism and the variability of the scatter function between subjects.

Random coincidences: these coincidence detections arise from the simultaneous detection of two photons produced by two different positron annihilations. They differ from true and scattered coincidences in that the randoms rate scales as approximately the square of the source activity as opposed to linearly [42]. In most PET scanners randoms are corrected for by measuring a randoms rate using a delayed timing window [26] and subtracting this from the measured data. This correction is strictly inappropriate for our statistical model since although both the data and the measured randoms are Poisson processes, their difference is not. One way in which randoms may be included in a statistical formulation is to maintain the off-timing randoms counts r_k as an additional set of Poisson data with unknown mean, q_k. The parameters q_k could then be estimated along with the image intensities as derived in Section 6.2.3, although clearly this results in a huge increase in the number of parameters. Alternatively, it may be possible to use a lower dimensional parametric model, similar to that proposed above for modeling scatter.

From the system modeling point of view, the new algorithm developed for PET image reconstruction is in the normal fashion, *i.e.* we simply ignore the presence of scatter, and subtract the randoms counts from the observed coincidence data and assume the resulting measurements are Poisson, although this clearly introduces some degree of modeling error into the reconstruction process.

6.2 Statistical Approaches in PET

In this section we develop several estimation schemes for PET image reconstruction that are based on the generalized EM algorithm. We begin in section 6.2.1 with a definition of the generalized EM algorithm for parameter estimation.

6.2.1 The GEM Algorithm

The MAP estimate of a parameter vector A given a data vector B maximizes the posterior probability $P(A|B)$, i.e.

$$\max_A P(A|B) = \max_A \left\{ \frac{P(B|A)P(A)}{P(B)} \right\},$$

where $P(B|A)$ denotes the conditional probability for the data given A and $P(A)$ denotes a prior probability distribution on A. The EM algorithm is a general method for solving this problem by choosing an intermediate set of unobserved "complete" data C such that there is a known many-to-one mapping from C to B. The problem may then be solved by choosing some feasible initial estimate A^0 of A and applying the iteration:

$$
\begin{aligned}
&\text{E-step:} \quad \mathbf{E}_C \left\{ \log P(C|A) | A^n, B \right\} \\
&\text{M-step:} \quad A^{n+1} = \arg \left\{ \max_A Q(A|A^n) \right\}, \quad n \Leftarrow n + 1,
\end{aligned}
\tag{6.3}
$$

where the function $Q(A|A^n)$ is defined as

$$Q(A|A^n) = \mathbf{E}_C \left\{ \log P(C|A) | A^n, B \right\} + \log P(A).$$

In general, a closed form solution to the M-step optimization problem is not available and searching for a minimum of $Q(A|A^n)$ at each iteration is computationally prohibitive. A more general form of this algorithm, the generalized EM (GEM) algorithm, replaces the M-step with an updating procedure for A which satisfies the following inequality:

$$Q(A^{n+1}|A^n) \geq Q(A^n|A^n).$$

As shown in [8], the M-step only maximizes a portion of the posterior distribution while ensuring that the remaining portion is increased, but not maximized. Therefore at best, this approach only guarantees convergence of the sequence to a stationary point of the posterior distribution [60].

6.2.2 Maximum Likelihood via EM

For PET image reconstruction, the parameters to be estimated are the emission intensities, i.e. $A = \lambda$ where λ denotes the set of pixel intensities $\lambda(i,j)$, $i,j = 1, \cdots, N$. The observed data are the set of coincident detections denoted by $B = Y$, with elements y_k $k = 1, \cdots, M$. Let $P_k(i,j)$ denote the known probability that an emission from source point (i,j) is detected at the detector pair corresponding to index k. The unobserved complete data C, as defined in [49], are denoted $x_k(i,j)$ and are defined to be the number of emissions from pixel (i,j) which are collected in detector pair k. The variables $x_k(i,j)$ are modeled as conditionally independent Poisson random variables with mean $P_k(i,j)\lambda(i,j)$. The incomplete data y_k are related to the complete data $x_k(i,j)$ as $y_k = \sum_{i,j} x_k(i,j)$ and hence are also conditionally independent Poisson random variables with mean $\sum_{i,j} P_k(i,j)\lambda(i,j)$.

Applying the EM algorithm to the ML estimation problem as discussed in Chapter 2 we rewrite (2.10) in terms of the result from E-step as

$$\mathbf{E}_X \left\{ log P(X|\lambda)|\lambda^n, Y \right\} = \sum_{i,j} \left\{ -a(i,j)\lambda(i,j) + b^n(i,j) \log \lambda(i,j) \right\} + \quad (6.4)$$
$$constant,$$

where

$$a(i,j) = \sum_k P_k(i,j)$$
$$b^n(i,j) = \lambda^n(i,j) \sum_k \left\{ y_k P_k(i,j) / \sum_{i',j'} P_k(i',j') \lambda^n(i',j') \right\}. \tag{6.5}$$

Substituting this result into the M-step and optimizing over λ, we obtain the well known EM algorithm for ML estimation:

$$\lambda^{n+1}(i,j) = \lambda^n(i,j) \frac{\sum_k \frac{y_k P_k(i,j)}{\sum_{i',j'} P_k(i',j') \lambda^n(i',j')}}{\sum_k P_k(i,j)}. \tag{6.6}$$

6.2.3 An EM Algorithm for PET with Independent Random Measurements

The Siemens ECAT831 brain scanner collects random coincidences occurring in a delayed time window to correct for randoms. In the case where the random measurements can be treated as an additional data set one can modify the above EM algorithm by choosing the data as $B = \{Y, R\}$, where R denotes the set of random measurements, whose elements are $r_k, k = 1, \cdots, M$. The unknown parameters to be estimated as $A = \{\lambda, Q\}$, where Q denotes the mean of the randoms process, including $q_k, k = 1, \cdots, M$. The observed data are the sum of true coincidences and randoms, i.e.

$$y_k = C_k^T + C_k^R,$$

where C_k^T denotes the true coincidences including scatters, and C_k^R denotes the collected randoms at the detector pair k during the period of PET data collection.

The ML estimate of this problem now becomes to find a $\lambda^* = \{\lambda^*(i,j)\}$ and a $Q^* = \{q_k^*\}$ such that the conditional probability $P(Y, R | \lambda, Q)$ is maximized, i.e.

$$\max_{\lambda, Q} P(Y, R | \lambda, Q).$$

In order to apply the EM algorithm to solve the problem, we must form a set of

complete data first. Let the complete data be

$$C = \left\{X, C^R, R\right\} = \left\{x_k(i,j), C_k^R, r_k; i,j = 1, \cdots, N; k = 1, \cdots, M\right\}. \qquad (6.7)$$

Then the conditional probability of complete data C given the parameters λ and Q follows

$$P(X, C^R, R|\lambda, Q) = \prod_{i,j,k} \frac{e^{-P_k(i,j)\lambda(i,j)} \left(P_k(i,j)\lambda(i,j)\right)^{x_k(i,j)}}{x_k(i,j)!} \prod_k \frac{e^{-q_k} q_k^{C_k^R}}{C_k^R!} \prod_k \frac{e^{-q_k} q_k^{r_k}}{r_k!},$$

so the likelihood of the above probability is proportional to

$$\sum_{i,j,k} \left\{-P_k(i,j)\lambda(i,j) + x_k(i,j)\log(P_k(i,j)\lambda(i,j))\right\} + \qquad (6.8)$$

$$\sum_k \left\{-2q_k + [C_k^R + r_k]\log q_k\right\}.$$

The E-step of the EM algorithm replaces $x_k(i,j)$ and C_k^R by their conditional mean given $\{\lambda^n, Q^n, Y, R\}$. Since $x_k(i,j)$ and C_k^R, $i,j = 1, \cdots, N$ are mutually independent Poisson random variables and $y_k = \sum_{i,j} x_k(i,j) + C_k^R$. Therefore, the probability of $x_k(i,j)$ given $\{\lambda^n, Q^n, Y, R\}$ is binomial with mean

$$\mathbf{E}\left\{x_k(i,j)|\lambda^n, Q^n, Y, R\right\} = y_k \frac{P_k(i,j)\lambda(i,j)}{\sum_{i,j} P_k(i,j)\lambda^n(i,j) + q_k^n}, \qquad (6.9)$$

and so is C_k^R with mean

$$\mathbf{E}\left\{C_k^R|\lambda^n, Q^n, Y, R\right\} = y_k \frac{q_k^n}{\sum_{i,j} P_k(i,j)\lambda^n(i,j) + q_k^n}. \qquad (6.10)$$

Substituting the results (6.9) and (6.10) into (6.8) and taking the partial derivative with respect to $\lambda(i,j)$ and q_k respectively and letting the results be equal to zero yields

$$q_k^{n+1} = \frac{1}{2}\left[r_k + \frac{y_k q_k^n}{\sum_{i,j} P_k(i,j)\lambda^n(i,j) + q_k^n}\right], \qquad (6.11)$$

$$\lambda^{n+1}(i,j) = \frac{\lambda^n(i,j)}{\sum_k P_k(i,j)} \sum_k \frac{y_k P_k(i,j)}{\sum_{i',j'} P_k(i',j')\lambda^n(i',j') + q_k^n}. \qquad (6.12)$$

One chooses feasible initial estimates, λ^0 and Q^0 of λ and Q and performs the above iterative algorithm to obtain a ML estimate. We expect the above iteration to suffer from the same ill-conditioning which usually occurs in conventional ML PET reconstruction. This problem arises due to the high dimensionality of the estimated parameters relative to the data. To overcome this problem we could employ MAP estimators similar to those described below, but with an additional prior distribution to perform spatial smoothing of the randoms process, Q.

6.2.4 MAP Using MRFs with Intensity Process Only

In [20] a GEM algorithm is described for MAP estimation where the prior distribution for λ is a Markov random field:

$$P(\lambda) = \frac{1}{Z} e^{-\gamma U(\lambda)}, \tag{6.13}$$

where Z is a normalizing constant, γ is a prior parameter and $U(\lambda)$ denotes a Gibbs energy function defined on the set

$$\Omega = \{\lambda(i,j) \in \mathcal{R} \mid \lambda(i,j) \geq 0; i,j = 1, \cdots, N\}. \tag{6.14}$$

Applying the GEM algorithm to this problem one obtains the following procedure:

E-step: $\mathbf{E}_X \{\log P(X|\lambda)|\lambda^n, Y\} = \sum_{i,j} \{-a(i,j)\lambda(i,j) + b^n(i,j) \log \lambda(i,j)\} +$

$$constant \tag{6.15}$$

M-step: $\max_{\lambda} Q(\lambda|\lambda^n) = \mathbf{E}_X \{\log P(X|\lambda)|\lambda^n, Y\} - \gamma U(\lambda).$

The E-step remains the same as for the ML-EM algorithm. A suitable M-step is described in [20]. In this case the definitions of the parameters, A, and data, B, remain the same as for the original form of the ML-EM algorithm. Although with a suitable choice of $U(\lambda)$, this method avoids the ill-posedness of the ML approach, the algorithm tends to over-smooth the reconstructed image particularly in the vicinity of true boundaries. To overcome this problem, we now consider the use of MRFs with an additional line process.
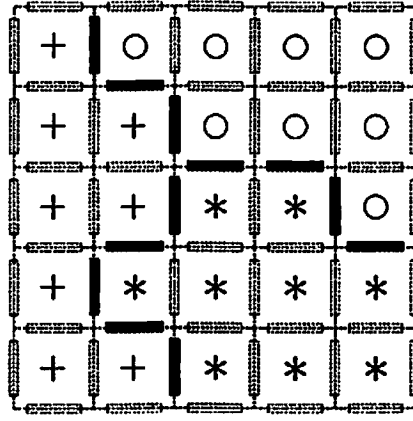
Figure 6.1: Line sites are introduced between each pair of neighboring pixels, and take the value "1" to indicate the presence of a boundary between adjacent pixels and otherwise take the value "0". Three different regions are represented using '+', '*' and 'o'. The line sites are equal to "1" (solid bars) on the boundaries of these regions.

6.2.5   MAP Using MRFs with Intensity and Line Process

The line process is used to model discontinuity-like edges in images [13], and can be regarded as an indicator variable which is set to zero wherever neighboring pixels lie in the same region (*i.e.* have similar values) and is otherwise set to one. Let $\{l(i,j)\}$ represent line sites lying between each pair of neighboring intensities, as illustrated in Figure 6.1. For the purposes of computation, each $l(i,j)$ may be either binary valued $\{0,1\}$ or take any real value on the interval $[0,1]$. The parameters to be estimated in the GEM algorithm are $A = \{\lambda, l\}$. The joint Gibbs distribution of λ and l may be written in the form [13]:

$$P(\lambda, l) = \frac{1}{Z} e^{-\gamma U(\lambda, l)}, \qquad (6.16)$$

where γ is a prior parameter. The Gibbs energy function $U(\lambda, l)$ is defined on the set $\Omega \times \mathcal{I}^L$, where Ω is defined in (6.14). The notation $\mathcal{I}$, for the case where the line processes $l(i,j)$ are binary random variables, represents the set $\{0,1\}$.

The GEM algorithm for the MAP problem in this case becomes

E-step: $\mathbf{E}_X \{ log P(X|\lambda)|\lambda^n, Y \} = \sum_{i,j} \{ -a(i,j)\lambda(i,j) + b^n(i,j)\log \lambda(i,j) \} +$

$$constant \qquad (6.17)$$

M-step: $\max_{\lambda,l} Q(\lambda, l|\lambda^n, l^n) = \mathbf{E}_X \{ \log P(X|\lambda)|\lambda^n, Y \} - \gamma U(\lambda, l).$

Note that in this case, the E-step remains identical to that in (6.4), only the M-step is changed. The parameters to be estimated in the M-step are the intensities $\lambda(i,j)$ and the lines $l(i,j)$. The M-step in the GEM algorithm is selected as an updating procedure for both λ and l that satisfies the following condition at each iteration

$$Q \left(\lambda^{n+1}, l^{n+1}|\lambda^n, l^n \right) \geq Q \left(\lambda^n, l^n|\lambda^n, l^n \right).$$

This is achieved in two steps:

- M1-step: find estimates of intensities λ^{n+1} satisfying

$$Q \left(\lambda^{n+1}, l^n|\lambda^n, l^n \right) \geq Q \left(\lambda^n, l^n|\lambda^n, l^n \right); \qquad (6.18)$$

- M2-step: find estimates of lines l^{n+1} satisfying

$$Q \left(\lambda^{n+1}, l^{n+1}|\lambda^n, l^n \right) \geq Q \left(\lambda^{n+1}, l^n|\lambda^n, l^n \right). \qquad (6.19)$$

Note that $\mathbf{E}_X \{ \log P(X|\lambda)|\lambda^n, Y \}$ is a constant with respect to l^{n+1} for the M2-step and therefore (6.19) is equivalent to the following:

$$-U(\lambda^{n+1}, l^{n+1}) \geq -U(\lambda^{n+1}, l^n). \qquad (6.20)$$

MR images share the same underlying anatomical boundaries. Thus if one could extract high resolution structural information from the MR image, this could be used as important *a priori* information about potential boundaries between structures in the PET image. In this chapter we describe a method by which this information, extracted from MR images, may be incorporated into the PET image reconstruction process.

7.1 Incorporation of Prior Anatomical Information

This section presents a MAP reconstruction algorithm for incorporation of anatomical MR boundaries. This algorithm is a modified version of the algorithm developed in Chapter 6 for the case where *a priori* edge information is partially available. With the aid of *a priori* edge information and interaction measurements between neighboring edges, one would expect that the task of estimating unknown edges becomes easier than the case considered in the previous chapter. Hence, in order to reduce the complexity of implementation while producing a similar image quality, We only consider the first-order neighbors for intensity and horizontal and vertical line fields in the following derivation. The model can be easily extended to higher-order neighborhoods.

Let us now assume that the line sites, l, consist of two parts: 1) l_a: the set of known edges obtained through some edge detection process from an MR image; and 2) $l_F = \{l - l_a\}$. This second set may contain additional edges due to the presence of discontinuities in the PET image which do not correspond to anatomical boundaries. It is important to include and estimate the set l_F since it is possible to obtain additional edges in the PET image due to sharp changes in function within a single anatomical region. Following the notation in Chapter 6, then for this problem, the parameters to be estimated are $A = \{\lambda, l_F\}$, the observed data are $B = \{Y\}$ and l_a are treated as known constants. The GEM

algorithm in this case can then be written as:

E-step: $\mathbf{E}_X \{\log P(X|\lambda)|\lambda^n, Y\} = \sum_{i,j} \{-a(i,j)\lambda(i,j) + b^n(i,j)\log \lambda(i,j)\} +$

$$\text{constant} \qquad (7.1)$$

M-step: $\max_{\lambda, l_F} Q(\lambda, l_F|\lambda^n, l_F^n) = \mathbf{E}_X \{\log P(X|\lambda)|\lambda^n, Y\} - \gamma U(\lambda, l_F, l_a).$

We denote the line process by l_s, where $s = \{F, a\}$. Let l_a denote the known locations of edges extracted from the registered and segmented MR image and l_F the unknown functional boundaries to be estimated. Where appropriate we add a second index to l_s to denote the orientation of the line site: $l_{p,s}$ where $p = x$ denotes a horizontal line site and $p = y$ denotes a vertical line site. For notational simplicity, in the following we assume that appropriate modifications are made to the potential functions at the image boundary, a suitable energy function for first order interactions is then

$$U(\lambda, l_s) = \sum_{i,j}\{\beta\lambda_x^2(i,j)(1-l_{y,s}(i,j))+\alpha l_{y,s}(i,j)+\mathcal{H}_y(l_s)\}+ \qquad (7.2)$$
$$\sum_{i,j}\{\beta\lambda_y^2(i,j)(1-l_{x,s}(i,j))+\alpha l_{x,s}(i,j)+\mathcal{H}_x(l_s)\}.$$

The line sites $l_{p,s}(i,j)$ are binary random variables. In each pair of braces, The first two terms function in a similar manner to those discussed in Section 6.3, except that a measure of interaction between line sites, $\mathcal{H}_p(l_s)$, is defined separately in the horizontal and vertical direction. In addition to encouraging the estimated edges to be connected, $\mathcal{H}_p(l_s)$ also suppresses redundant line sites. $\lambda_x(i,j)$ and $\lambda_y(i,j)$ are the differences between adjacent horizontal and vertical intensities respectively, defined as

$$\begin{aligned} \lambda_x(i,j) &= \lambda(i-1,j) - \lambda(i,j) \\ \lambda_y(i,j) &= \lambda(i,j-1) - \lambda(i,j). \end{aligned} \qquad (7.3)$$

A simple choice of line interaction term would be the functions:

$$
\begin{aligned}
\mathcal{H}_x(l_s) &= -\epsilon\alpha[l_{x,s}(i-1,j) - l_{x,s}(i,j-1)]l_{x,s}(i,j) \\
\mathcal{H}_y(l_s) &= -\epsilon\alpha[l_{y,s}(i,j-1) - l_{y,s}(i-1,j)]l_{y,s}(i,j),
\end{aligned}
\tag{7.4}
$$

which indicate that any two line sites from different orientations are independent to each other. These functions serve the purpose of increasing the penalty on a horizontal (vertical) line if either of its neighbors on adjacent rows (columns) is "on" and decreasing the penalty on a horizontal (vertical) line if either of its horizontal (vertical) neighbors on the same row (column) is "on". This can be seen clearly later in the line updating scheme. Again, the parameter ϵ controls the relative penalty associated with this term. In the case where the term $\mathcal{H}_p(l_s)$ includes an *a priori* edge information, using a relatively large ϵ would be beneficial. We note that as the neighborhood size is increased, the function above can be modified to allow better control of the formation of boundaries.

We now describe an updating procedure for both λ and l_F that is based on the GEM method described in Section 6.2.5. Here two types of edge processes are included: the known prior edges $l_{p,a}$ obtained from the MR image, and the unknown functional edges, $l_{p,F}$, to be estimated. The known anatomical boundaries $l_{p,a}$ are unchanged throughout the updating process.

In our simulation we found that the overall computational cost was reduced by updating each intensity (line) site more than once at each M1-step (M2-step). We also found that the line processes converge faster than the intensity processes and are therefore updated less often. We thus restate the M1-step and M2-step in the following way:

M1-step: Condition (6.18) in Chapter 6 for this case can be rewritten as:

$$
Q\left(\lambda^{n+1}, l_F^m | \lambda^n, l_F^m\right) \geq Q\left(\lambda^n, l_F^m | \lambda^n, l_F^m\right).
\tag{7.5}
$$

Here λ^{n+1} denotes the estimates from this step, which are obtained by updating each intensity site an arbitrary number of times, $r \geq 1$.

M2-step: Based on the result λ^{n+1} from the M1-step, the estimates l_F^{n+1} should

110

satisfy condition (6.20) discussed in Chapter 6 rewritten as:

$$Q\left(\lambda^{n+1}, l_F^{n+1} | \lambda^n, l_F^n\right) \geq Q\left(\lambda^{n+1}, l_F^n | \lambda^n, l_F^n\right),$$

or equivalently

$$-U(\lambda^{n+1}, l_F^{n+1}, l_a) \geq -U(\lambda^n, l_F^n, l_a). \tag{7.6}$$

where l_F^{n+1} are the line estimates obtained by updating each line site an arbitrary number of times, $q \geq 0$.

The Gibbs energy function $U(\lambda, l_s)$ results in a function $Q(\lambda, l_F^n | \lambda^n, l_F^n)$ that is concave with respect to each $\lambda(i,j) \geq 0$. Hence in the M1-step, we follow a similar derivation to that described in Section 6.3

M1-step: Each intensity site is updated in turn using a one dimensional line search, since $Q(\lambda, l_F^n | \lambda^n, l_F^n)$ is concave with respect to each $\lambda(i,j) \geq 0$. Thus setting the derivative with respect to $\lambda(i,j)$ equal to zero, and taking the positive root of the resulting quadratic equation, yields the unique maximum with respect to $\lambda(i,j)$. The updating equation is identical to equation (6.26) in Section 6.3, i.e.

$$\lambda^{nr}(i,j) = \frac{[C_2 - a(i,j)] + \sqrt{[C_2 - a(i,j)]^2 + 4C_1 b^n(i,j)}}{2C_1}, \quad n_r \Leftarrow n_r + 1;$$

$$\lambda^{n+1} \Leftarrow \lambda^{nr}, \quad \text{after } r \text{ iterations,} \tag{7.7}$$

but with simpler constants C_1 and C_2 given by

$$
\begin{aligned}
C_1 &= 2\beta\gamma[(1 - l_{x,s}^n(i,j)) + (1 - l_{y,s}^n(i,j)) + \\
&\quad (1 - l_{y,s}^n(i+1,j)) + (1 - l_{x,s}^n(i,j+1))] \\
C_2 &= 2\beta\gamma[\lambda(i-1,j)(1 - l_{y,s}^n(i,j)) + \lambda(i,j-1)(1 - l_{x,s}^n(i,j)) + \quad (7.8) \\
&\quad \lambda(i,j+1)(1 - l_{x,s}^n(i,j+1)) + \lambda(i+1,j)(1 - l_{y,s}^n(i+1,j))].
\end{aligned}
$$

When computing C_1 and C_2, the most recent estimates of $\lambda(i \pm 1, j \pm 1)$ should be used for fast convergence. In the scheme (7.7) we also update each pixel r times. The estimates λ^{nr} are then assigned to λ^{n+1}.

The above procedure, when applied to all intensity sites in turn, results in an increase in the function $Q(\lambda, l_F^n | \lambda^n, l_F^n)$, i.e. condition (7.5) is satisfied. The case where $C_1 = 0$ implies that all lines involved in C_1 are "on". In this case C_2 is also equal to zero and the M1-step at this site reduces to $\lambda^{nr}(i,j) = b^n(i,j)/a(i,j)$.

We now turn to the problem of updating the unknown functional boundaries $l_{p,F}$ while keeping the known prior anatomical boundaries $l_{p,a}$ fixed. In developing the M2-step below, we adopt the ICA approach developed for the algorithm in Section 6.3. The goal of the M2-step is to update the unknown line sites satisfying $U(\lambda^{n+1}, l_F^{n+1}, l_a) \leq U(\lambda^{n+1}, l_F^n, l_a)$. In this case, we again use the iterated conditional average of a modified energy function: $U_T(\lambda^{n+1}, l_F, l_a) = U(\lambda^{n+1}, l_F, l_a)/T$, where $U(\lambda^{n+1}, l_F, l_a)$ is defined in (7.2). Fixing $T = 1$ would produce an iteration equivalent to the ICA algorithm of [30]. However, as $T \to 0$, the variables $l_{p,F}(i,j)$ converge to either 0 or 1 and this algorithm would converge to a local minimum of the energy function $U(\lambda^{n+1}, l_F, l_a)$ with respect to l_F on the set $\{0,1\}^L$. Through the use of the control parameter T, we hope that the algorithm converges to a more desirable local maximum of the posterior distribution.

M2-step: Let $N_{p,F}(i,j)$ denote the set of intensities and lines which interact with $l_{p,F}(i,j)$. The local conditional probability of $l_{x,F}(i,j)$ and $l_{y,F}(i,j)$ for the energy function $U_T(\lambda^{n+1}, l_F, l_a)$ is given by

$$P_T(l_{x,F}(i,j) | N_{x,F}(i,j)) =$$
$$\frac{e^{-\frac{1}{T}\{\beta\lambda_y^{n+1}(i,j)^2(1-l_{x,F})+\alpha l_{x,F}-\epsilon\alpha[l_{x,s}(i-1,j)+l_{x,s}(i+1,j)-l_{x,s}(i,j-1)-l_{x,s}(i,j+1)]l_{x,F}\}}}{e^{-\frac{1}{T}\beta\lambda_y^{n+1}(i,j)^2} + e^{-\frac{1}{T}\{\alpha-\epsilon\alpha[l_{x,s}(i-1,j)+l_{x,s}(i+1,j)-l_{x,s}(i,j-1)-l_{x,s}(i,j+1)]\}}}$$
(7.9)

$$P_T(l_{y,F}(i,j) | N_{y,F}(i,j)) =$$
$$\frac{e^{-\frac{1}{T}\{\beta\lambda_x^{n+1}(i,j)^2(1-l_{y,F})+\alpha l_{y,F}-\epsilon\alpha[l_{y,s}(i,j-1)+l_{y,s}(i,j+1)-l_{y,s}(i-1,j)-l_{y,s}(i+1,j)]l_{y,F}\}}}{e^{-\frac{1}{T}\beta\lambda_x^{n+1}(i,j)^2} + e^{-\frac{1}{T}\{\alpha-\epsilon\alpha[l_{y,s}(i,j-1)+l_{y,s}(i,j+1)-l_{y,s}(i-1,j)-l_{y,s}(i+1,j)]\}}}.$$

The local conditional mean of the binary random variable $l_{p,F}(i,j)$ is then equal to the probability that $l_{p,F}(i,j) = 1$. Using the posterior conditional mean, the updating procedure for $l_{p,F}(i,j)$ is defined as

$$l_{x,F}^{n_q}(i,j) = \{1 + \exp\{\frac{1}{T}[\alpha - \beta\lambda_y^{n+1}(i,j)^2 - \epsilon\alpha(l_{x,s}(i-1,j) +$$

$$l_{x,s}(i+1,j) - l_{x,s}(i,j-1) - l_{x,s}(i,j+1))]\}\}^{-1},$$

$n_q \Leftarrow n_q + 1$ and reduce T,

$l_{x,F}^{n+1}(i,j) \Leftarrow l_{x,F}^{n_q}(i,j)$,   after q iterations;

$$(7.10)$$

$$l_{y,F}^{n_q}(i,j) = \{1 + \exp\{\frac{1}{T}[\alpha - \beta\lambda_x^{n+1}(i,j)^2 - \epsilon\alpha(l_{y,s}(i,j-1)+$$
$$l_{y,s}(i,j+1) - l_{y,s}(i-1,j) - l_{y,s}(i+1,j))]\}\}^{-1},$$

$n_q \Leftarrow n_q + 1$ and reduce T,

$l_{y,F}^{n+1}(i,j) \Leftarrow l_{y,F}^{n_q}(i,j)$,   after q iterations.

When computing (7.10) the prior anatomical edges or the most recent estimates are used for the neighboring line sites, $l_{p,s}(i \pm 1, j \pm 1)$.

It is clear that this line updating scheme shares all properties stated in Section 6.3. As $T \rightarrow 0$, the updating procedure (7.10) defined above guarantee that $U_T(\lambda^{n+1}, l_F, l_a)$ decreases at each iteration of the M2-step.

7.2   Experimental Results

To illustrate the potential performance of the method described above we have conducted some preliminary studies in a wide variety of test cases, using both a computer generated phantom and the real three dimensional Hoffman brain phantom. Patient data is also used to validate the overall approach. The results from patient data validated the underlying structure between PET and MR is correlated, if not identical. They have also pointed out the possible direction for the further improvement of the algorithms.

7.2.1   Simulated Phantom: Case 1

We began with a real MR image of a patient with a large white matter lesion, Figure 7.1(a). Using the segmentation method described in [62] we segmented this image into four distinct tissue types: white matter, grey matter, lesion and ventricles. The resulting segmentation is shown in Figure 7.1(b) using the mean

intensity for each class; the boundaries between different regions are shown in Figure 7.1(c). The result of this segmentation was used as a template for generating a phantom PET image, *i.e.* we assume that a radiopharmaceutical is used in which there will be different amounts of activity in each of the four anatomical regions. The PET phantom is shown in Figure 7.1(d). In addition to the variations in activity in different anatomical regions, we added a hot spot in the left side of the white matter region to demonstrate the ability of the new algorithm to detect functional boundaries not present in the MR (anatomical) image. We will refer to this as a functional lesion below. To generate a more realistic image, the intensity within each region varies according to a first order Gauss Markov random field model [3].

The computer generated PET data were based on a simplified model of a single ring of the Siemens ECAT system. The source was assumed to be of maximum dimension 22cm diameter, the detector ring of diameter 76cm with 384 detectors. Data were generated as a set of fan beam projections obtained by pairing each of the 384 detectors with sufficient of the remaining detectors to cover the required 22cm diameter field of view. Factors such as randoms, scatter, attenuation and detector efficiency were not considered here. However, since the comparisons are all based on the same data, the comparison should indicate the relative potential performance of the various reconstruction methods. After generating the means of the projection data, a pseudo-random Poisson generator was used to generate Poisson data for each data sample with a mean of 200 counts per sample. The results are shown in Figures 7.1(e) though 7.1(i). Figure 7.1(e) is a filtered backprojection (FBP) reconstruction and shows the typical noisy image obtained with low count rates. Note that there is virtually no distinction between the mean activity in the grey and white matter and the additional functional lesion is barely visible. Figures 7.1(f) and 7.1(g) show ML and MAP reconstructions from the same data - the MAP reconstruction does not include line sites. There is in both cases a qualitative improvement over the FBP method with the MAP result smoother than the ML image, but neither of the results shows good white/grey matter contrast.
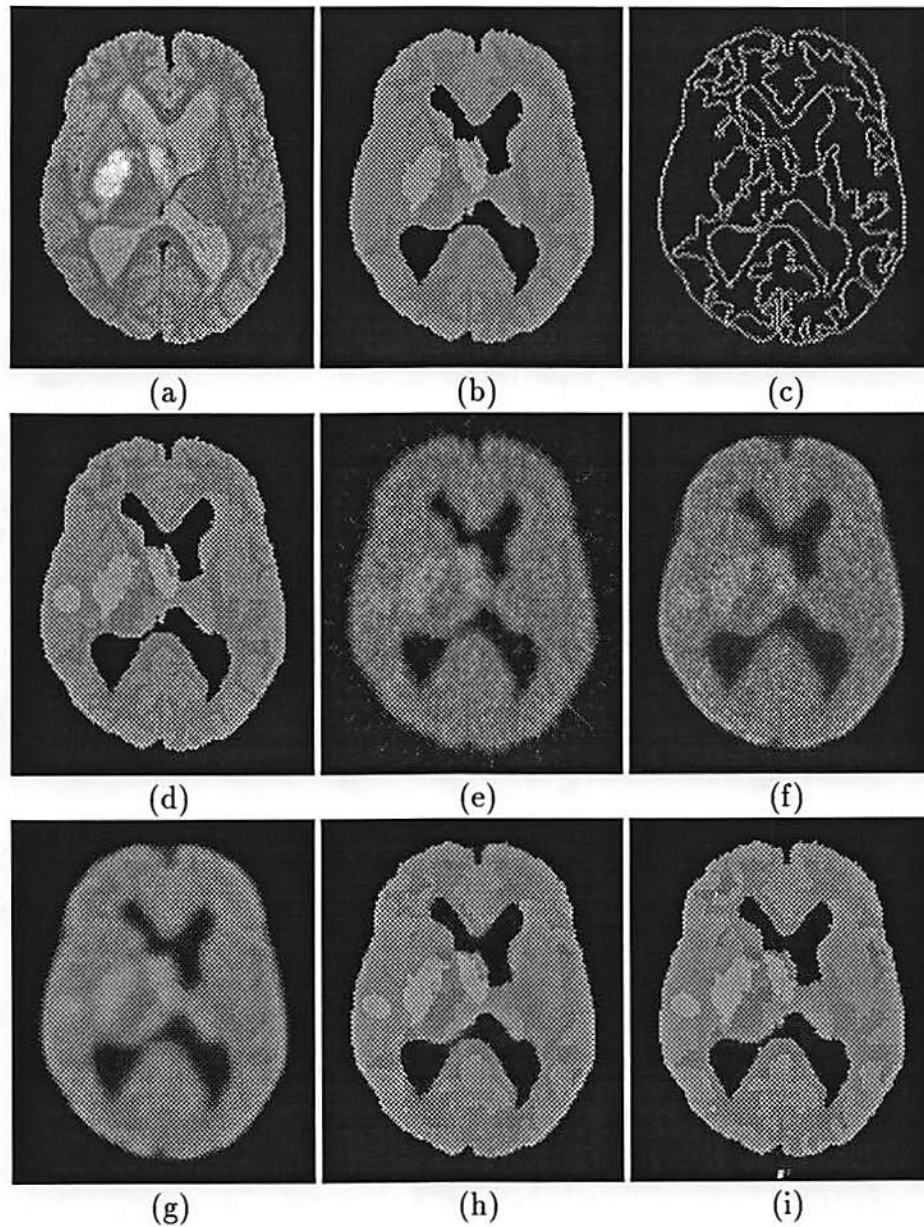
Figure 7.1: (a): Original MR brain image. (b): Segmentation of (a) into four tissue types. (c): Anatomical boundaries extracted from (b). (d): Computed PET phantom generated from MR template (b). (e): PET reconstruction using filtered backprojection. (f): PET reconstruction using maximum likelihood estimation. (g): PET reconstruction using MAP estimation - no lines. (h): PET reconstruction using this proposed algorithm - estimated lines. (i): PET reconstruction using this proposed algorithm - some line sites from MR image.

115

Figure 7.1(h) shows a reconstruction in which all line sites are estimated. Figure 7.1(i) shows the MAP reconstruction using line sites in which the anatomical line sites, Figure 7.1(c), are included, the other line sites are then estimated. These results indicate that there is a clear benefit to the use of line sites in MAP PET reconstruction in comparison to the other methods described above. Even when a priori edge information is not available, this methods performs surprisingly well, there is now noticeable grey/white matter contrast and the functional lesion is clearly visible. When the MR information is included the situation becomes even better. Note also, that although the functional lesion, in the left middle portion of the PET phantom, does not correspond to any anatomical boundaries, the MAP estimator, with a line process, is able to clearly detect the presence of this lesion. In this case the shape of the lesion is somewhat distorted, however, this is to be expected as it arises from the finite resolution of the simulated PET system. The importance of this observation is that the use of prior anatomical information does not preclude the presence of additional structure in the reconstructed PET image.

7.2.2 Simulated Phantom: Case 2

This experiment involves a simple phantom study. Its purpose is to examine that the *a priori* edge information does not impose discontinuities in PET reconstructions if this does not correspond to functional edges. Figure 7.2(a) is the disk phantom with an uniform intensity in each region. The simulated PET data is a set of parallel beam data generated in a similar manner to that described in the section 7.2.1. The total amount of Poisson data is equivalent to the collection time of 5 minutes for the Hoffman phantom FDG study. The configuration of PET imaging system used here is the same as a Siemens ECAT831 brain scanner, explained in details in Chapter 6. Factors such as randoms, scatters and *etc.* are not included. Shown in Figure 7.2(b) is a boundary image extracted from a cross section of Hoffman brain phantom. This image will be used as anatomical boundary information in the Hoffman brain phantom study. Figure 7.2(c) is the reconstruction using the proposed algorithm with *a priori* edges in Figure 7.2(b). We also showed the reconstruction using this algorithm with estimated edges in

Figure 7.2(d). This test case clearly demonstrates that the presence of a boundary in the anatomical image does not force a corresponding boundary in the reconstructed PET image, if the data shows no evidence for such a boundary, then it will not be present in the reconstructed PET image. This is because the function of the anatomical boundary information is to avoid smoothing across anatomical boundaries rather than to force the formation of corresponding boundaries in the PET image.

7.2.3   Real Hoffman Phantom

Images were reconstructed for a single two dimensional cross section through the three dimensional Hoffman brain phantom. A template for this section is shown in Figure 7.3(a). The image intensity in Figure 7.3(a) illustrates the activity in the phantom according to the ratio $4(white) : 1(grey) : 0(black)$. Figure 7.3(b) displays the boundaries extracted directly from the template (Figure 7.3(a)) from which the 3-D Hoffamn phantom was constructed. In the case where the anatomical boundaries were used in PET reconstructions, these boundaries were incorporated in the proposed algorithm, rather than those from an MR scan of the phantom. However we expect a high resolution MR scan should achieve a similar resolution to that of the digitized template shown. This template was rescaled to match PET reconstructions, therefore the boundaries were overlaid onto the top of a MAP PET reconstruction to illustrate the accuracy of the match between them in Figure 7.3(c).

The phantom was scanned using a Siemens ECAT831 brain scanner, which also scanned patient data discussed in Chapter 6. The data were collected for a total of 30 minutes and sorted into three data sets corresponding to collection times of 30 minutes (set 1), 5 minutes (set 2) and 1 minute (set 3). The three sets of data are preprocessed the exactly same as the patient data, including rebinning and setting negative values of coincidences to zero. Again, no scatter measurements or corrections were performed in this experiment. Detector response and attenuation correction measurements were made and used in all of the reconstructions shown below.

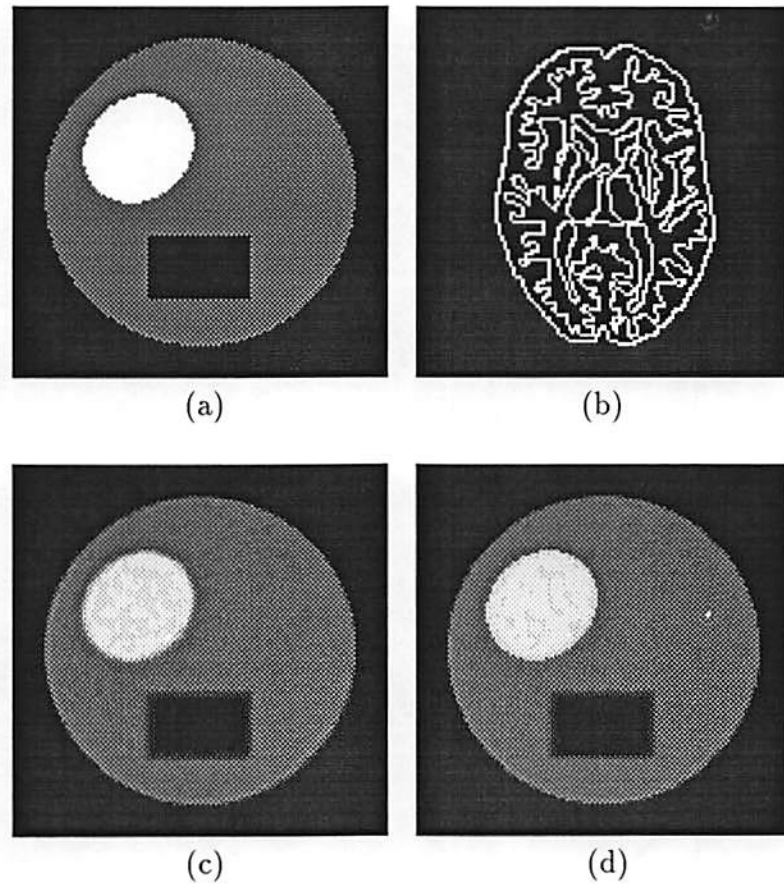Figure 7.2: Study of a simple computer generated phantom. (a): PET phantom to be reconstructed. (b): Boundaries extracted from a single cross section of the 3-D Hoffman brain phantom, which is also employed as anatomical MR boundaries in section 7.2.3. (c): Reconstruction using the proposed MAP algorithm with the incorporation of the boundaries in (b). (d): Reconstruction using the same algorithm but with an estimated line process.

In this experiment, we have shown reconstructions of the PET phantom using filtered backprojection for data sets 1, 2 and 3 in Figures7.4(b), 7.5(b) and 7.6(b) respectively. These images clearly show the deterioration in image quality as the total number of coincidence detections is decreased. Even for the case of 30 minutes worth of data, the boundaries in the image tend to be blurred and much of the fine detail evident in the phantom template is lost. Also shown in Figures 7.4, 7.5 and 7.6 are the reconstructions of the phantom from data sets 1, 2 and 3 using: maximum likelihood estimate and MAP estimate for an MRF prior without line process. The MRF without a line process uses the second order neighborhood of the Gibbs energy function proposed by Geman and McClure [14].

In Figures 7.4, 7.5 and 7.6 we show reconstructions of the PET phantom from data sets 1, 2 and 3 using the proposed algorithms with both estimated and *a priori* line processes. The most surprising aspect of these results is the performance of the MRF prior in the case where the line process is estimated. Comparison of the boundaries estimated in Figure 7.4(e) with the true boundaries in Figure 7.4(a) shows a close correspondence. The reconstructions that include *a priori* line processes employ the boundaries in Figure 7.3(b), directly obtained from the template.

While the performance clearly deteriorates in all cases, the proposed approach is able to reconstruct a reasonable image in cases where the filtered back projection algorithm yields virtually no useful information. It is clear that the advantage of using a line process, especially *a priori* line process in the MRF model becomes significant over the other methods presented here, as the number of detected coincidences is decreased. In this example we used only horizontal and vertical line processes, the effect of this can be seen in Figure 7.1(e), i.e. the vertical and horizontal boundaries in the phantom are found more accurately than boundaries at other orientations. This effect was not visible in our experiment with patient data, in which diagonal line sites were employed. So we hope that by modifying the model to also include a diagonal line process, the algorithm becomes more isotropic.

(a)

(b)   (c)
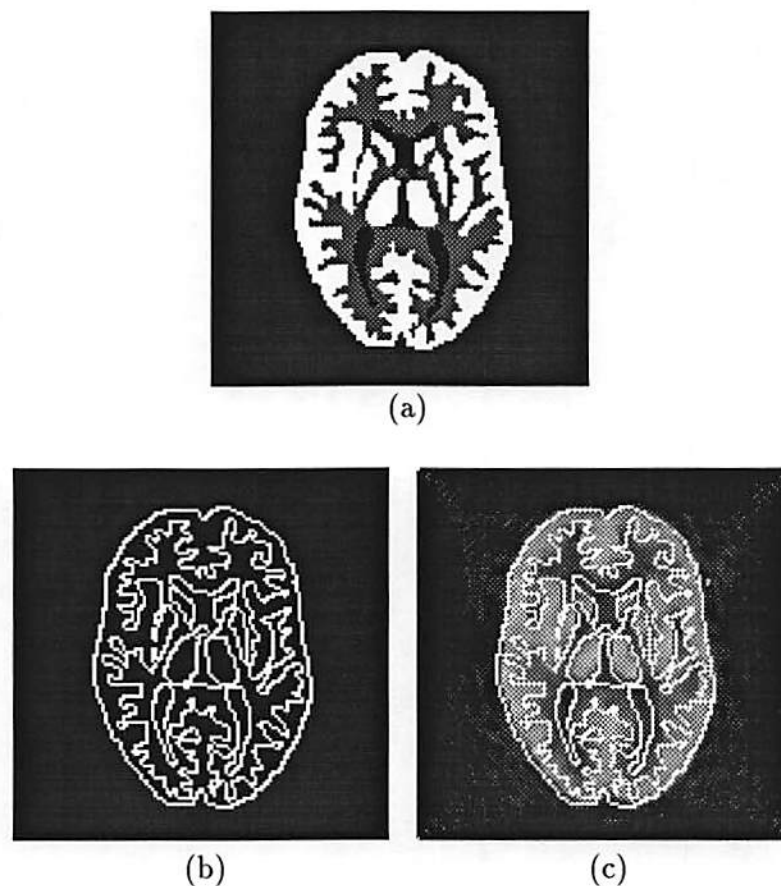
Figure 7.3: (a): Hoffman brain phantom template to be reconstructed. (b): Boundaries extracted from (a) to be used as anatomical MR boundaries in the following reconstructions. (c): An image obtained by overlaying the boundaries in (b) onto the corresponding slice of an MAP reconstruction.
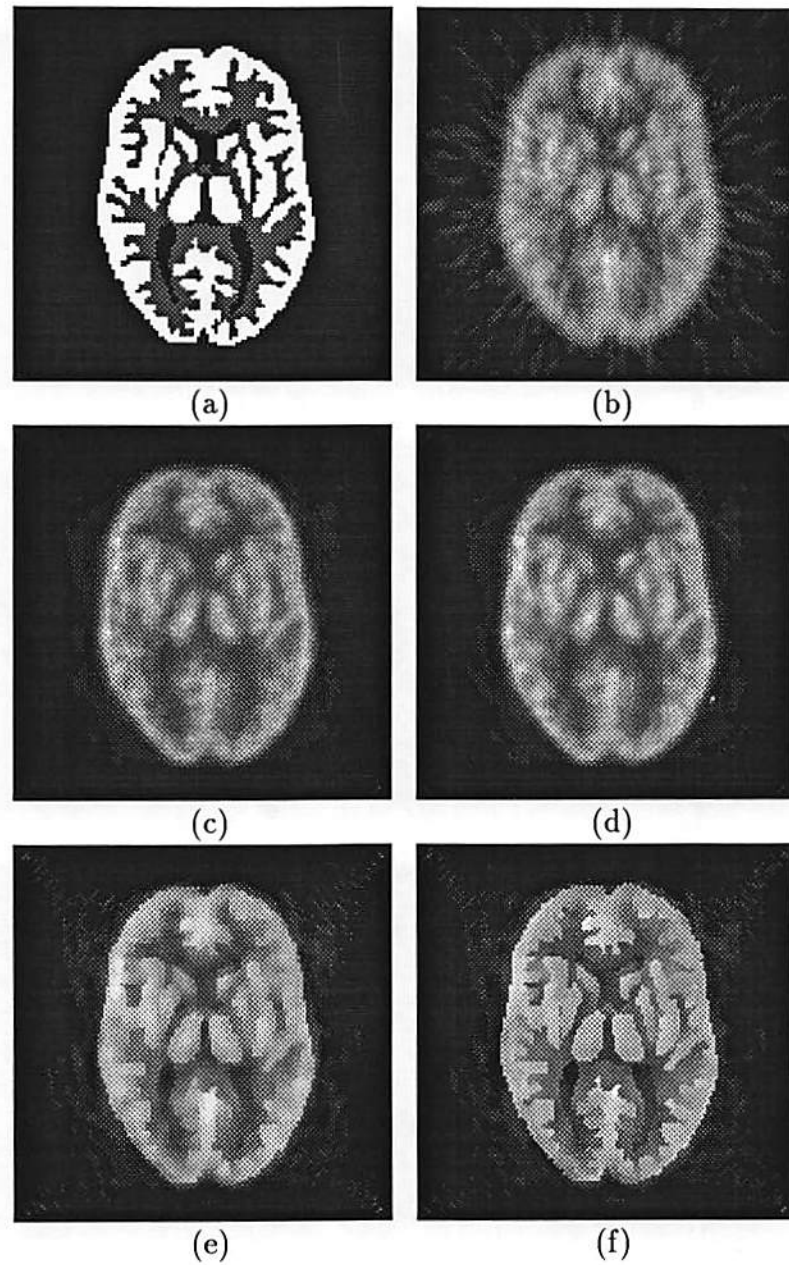
Figure 7.4: Hoffman brain phantom and reconstructions after 30 minutes data collection (3,339,279 coincidences). (a): Brain phantom template. (b): Reconstruction using filtered backprojection. (c): Reconstruction using maximum likelihood estimation. (d): Reconstruction using MAP estimation without a line process. (e): reconstruction using the proposed algorithm with an estimated line process. (f): Reconstruction using the proposed algorithm with *a priori* boundaries in Figure 7.3(b).
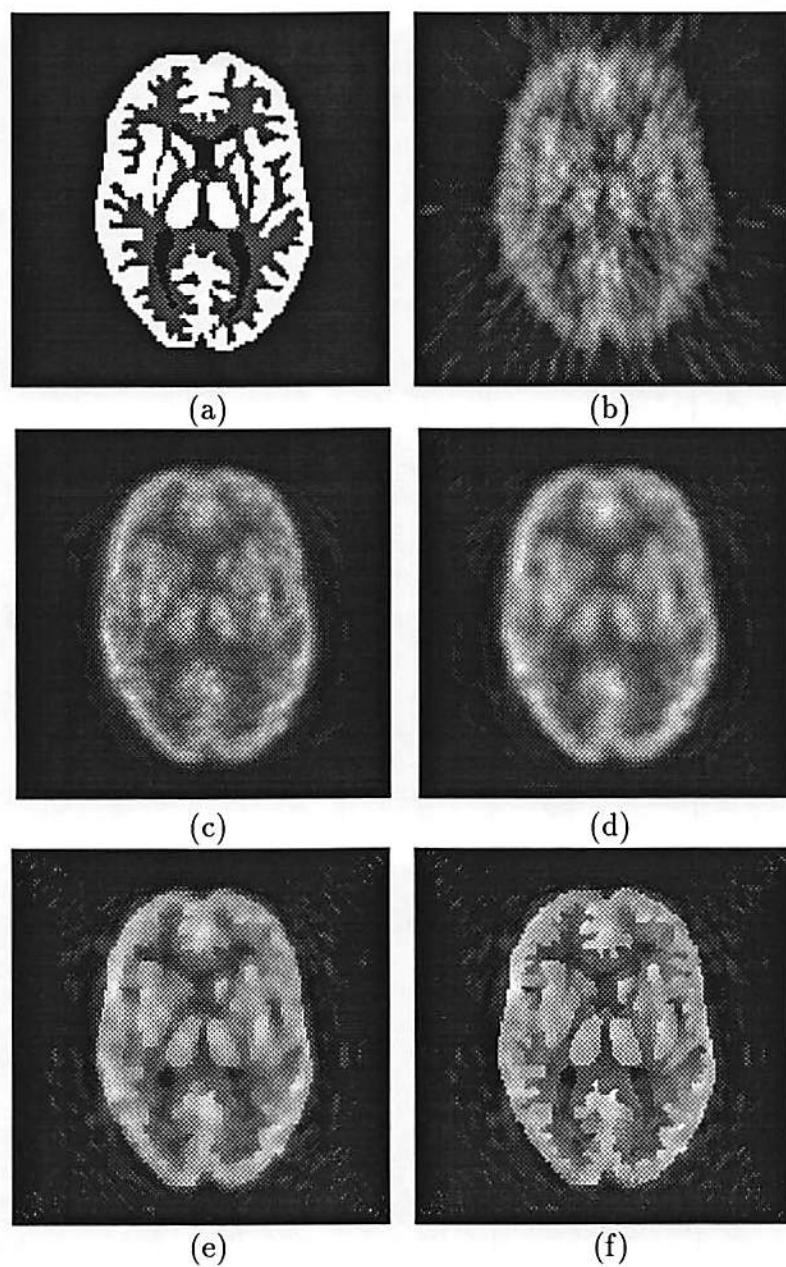
Figure 7.5: Hoffman brain phantom and reconstructions after 5 minutes data collection (580,021 coincidences). (a): Brain phantom template. (b): Reconstruction using filtered backprojection. (c): Reconstruction using maximum likelihood estimation. (d): Reconstruction using MAP estimation without a line process. (e): reconstruction using the proposed algorithm with an estimated line process. (f): Reconstruction using the proposed algorithm with *a priori* boundaries in Figure 7.3(b).
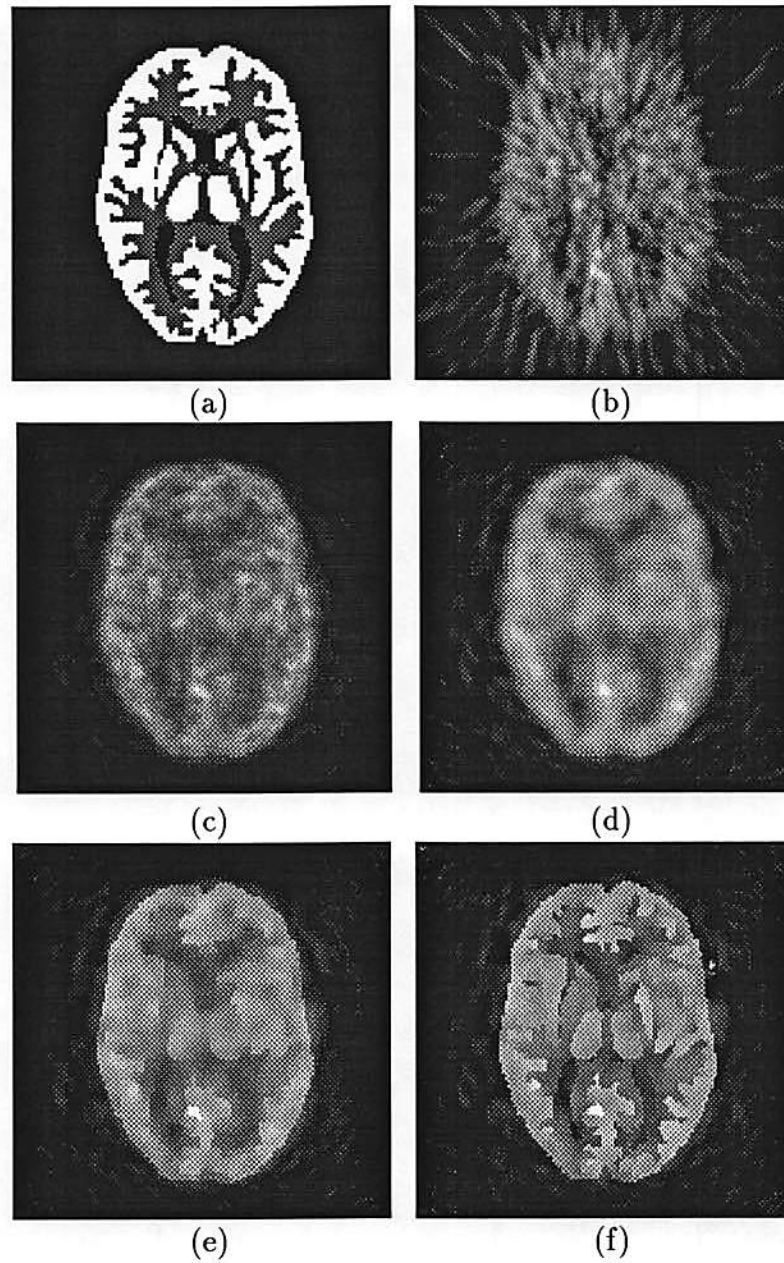
Figure 7.6: Hoffman brain phantom and reconstructions after 1 minute data collection (120,612 coincidences). (a): Brain phantom template. (b): Reconstruction using filtered backprojection. (c): Reconstruction using maximum likelihood estimation. (d): Reconstruction using MAP estimation without a line process. (e): reconstruction using the proposed algorithm with an estimated line process. (f): Reconstruction using the proposed algorithm with *a priori* boundaries in Figure 7.3(b).

7.2.4 Patient Data

The MAP algorithm developed in this chapter was applied to the MR scan and PET [18]FDG data of a patient brain. In order to incorporate *a priori* anatomical MR boundary information into this algorithm, one must first register both 3-D PET and MR images to a common framework. The registration technique used in this research is developed in [61], termed scalp surface matching, where scalp contours are extracted from both an ML reconstruction of PET and MR images and used for this registration. The registration technique searches for a transformation involving scaling, rotation and translation such that the total distance from the transformed set of points on PET scalp contours to the B-spline surface of MR images is minimized. The points-to-surface matching is similar to that described in [44] in many ways, however it employs a B-spline surface representation for MR contours to provide a smoother scalp surface. The average error, measured by a distance from a point on a PET contour to a B-spline surface of MR, is about $2mm$. This is acceptable comparing to the resolution of PET images.

Figure 7.7(a) is the reinterpolated and registered 2-D cross section of a 3-D MR scan. The brighter contour is the scalp contour of corresponding slice of an ML reconstruction of the PET image. This overlaying image demonstrates accuracy of this matching to some extent. Using the segmentation approach described in [62], we segmented this image into five tissue types and displayed the resulting segmentation in Figure 7.7(b) using the mean intensity value for each region. The anatomical boundaries between different regions were extracted from Figure 7.7(b) and shown in Figure 7.7(c). We then overlaid the boundaries onto the corresponding cross section of an ML reconstruction of PET, Figure 7.7(d) to show the matchness between anatomical and functional structures. There is a close correspondence between them. The image shown in Figure 7.7(e) is the reconstruction using the MAP algorithm developed in Chapter 6, which used the second-order neighborhood for intensities and four directions for line sites. In Figure 7.7(f), we showed the reconstruction using the proposed algorithm with both estimated and *a priori* anatomical boundaries in Figure 7.7(c). Again, it can be seen that estimated line sites tend to be horizontally or vertically oriented,

however, this effect is not visible in the image, Figure 7.7(e). This indicates that using larger size of neighborhoods is necessary. Mismodeling is also present in the case where *a priori* anatomical structure is included. This algorithm treats *a priori* anatomical boundaries as fixed and exact boundaries of a PET image to be reconstructed. This is not the case, because many factors and errors, such as the motion of a head, registration and segmentation error, would make this assumption less valid. The reconstruction is potentially distorted as a result.

It is expected that the overall performance of the proposed approach is better in the Hoffman phantom study. This is because that the phantom almost perfectly fits the assumption of MRF model: sharp boundaries between different regions, and homogeneous within each region. And also the *a priori* boundaries used correspond to the exact boundaries in PET reconstructions.
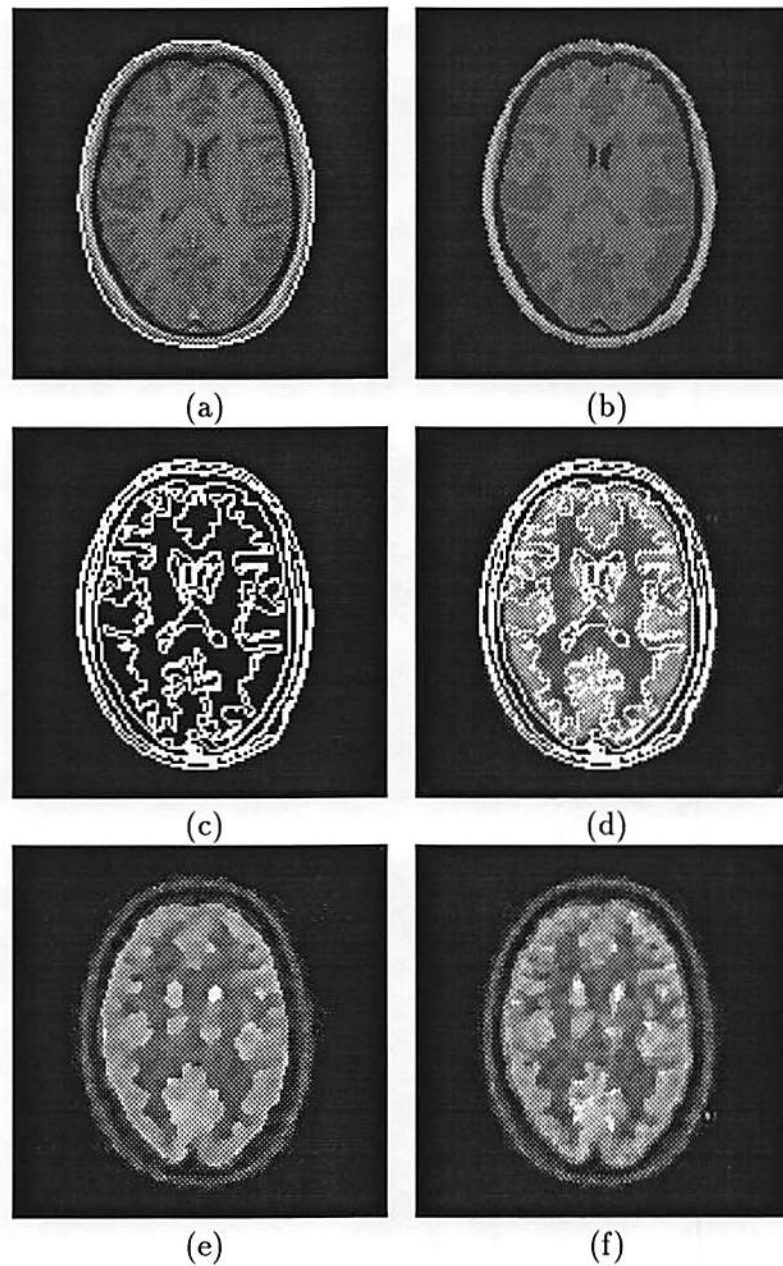
Figure 7.7: (a): A single cross section of a PET registered MR scan. The white contour is the scalp contour from the corresponding slice of an ML reconstruction of PET. (b): Segmentation of MR scan (a) into five tissue types. (c): Anatomical boundaries from (b). (d): An image obtained by overlaying the anatomical boundaries of (c) onto the corresponding slice of the ML reconstruction of PET. (e): Reconstruction using the algorithm presented in Chapter 6 with four directions of estimated line sites. (f): Reconstruction using the algorithm described in this chapter with *a priori* anatomical boundaries of (c) and estimated line sites.

Chapter 7

Incorporation of Anatomical MR Data for Improving PET Quality

Our experiences with patient data indicate that the use of MRF with a line process is a promising approach to PET image reconstructions and the approach proposed in Chapter 6 has produced encouraging reconstructions. However, we have observed through reconstructions of patient data that the locations of estimated line sites are dependent on the total number of coincident detections and, even for a large amount of coincident detections, are sometimes inaccurate. The MAP reconstruction algorithm proposed in Chapter 6 has the potential for solving this problem and further enhancing the quality of PET reconstructions if accurate line locations can be extracted from images obtained from a registered anatomical imaging modality (MR or CT) and incorporated in the PET algorithm.

Magnetic Resonance (MR) imaging is capable of producing high resolution images of the human anatomy. An appropriate choice of pulse sequence can produce T1 or T2 weighted images with excellent soft tissue contrast. Positron emission tomography (PET) [42] is capable of producing somewhat lower resolution images of functional activity through the use of radiolabeled pharmaceuticals. While PET and MR produce images of different parameters, one would not expect smooth variations in PET images across MR anatomical boundaries. In other words, the spatial distribution of functional activity is dependent on the underlying anatomical structure in which the activity occurs. This implies that we assume PET and

Chapter 8

Conclusions and Directions for Future Work

8.1 Conclusions

The goal of the research stated in this dissertation was to develop models and approaches to accurate, quantitative and efficient reconstructions. This goal has been achieved in several aspects.

We began by reconstructing a 3-D object from a set of cone beam projections. The algorithms developed here should also work for any other cone beam imaging systems, such as 3-D X-ray CT, SPECT and 3-D PET machines. The motivating factor in this approach is as follows. Knowing whether an inverse exists for a given cone beam geometry is clearly of importance, however algorithms for approximately inverting cone beam data in case where the inverse does not exist are also definitely needed. The computational time required is costly because of the dimensionality: the minimum scale used for a 3-D imaging system requires reconstruction of $64 \times 64 \times 64$ grids based on 64 projections. Each projection consists of a 64×64 array. Therefore, it may be more desirable to use an efficient algorithm that produces reasonably good reconstructions than an exact algorithm with exorbitant computational cost.

For this reason, we have developed a general inversion formula and investigated the application of this formula to the case of circular, elliptical and spiral

orbits. For a wide range of configurations, this approach will in general lead to a class of filtered back-projection type of algorithms, but does not produce exact inverses and we have therefore included a study of the resultant PSF for the three geometries. In assessing the performance of the new algorithms using both a system PSF and computer generated phantom, it is clear that it is desirable to make the curve of the largest possible radius relative to the object to be imaged. It also appears, from comparing these results from the three geometries, that using a circular orbit, with $R = 60cm$, produces less significant artifacts than any other configurations tested. Thus if one were free to use any configuration, then a large circular orbit would be preferred when using this class of algorithm. However, the influence of other factors, such as detector sensitivity or physical constraints may motivate the use of either the elliptical or spiral configuration.

Our experience with different sizes of configurations indicates that as long as the size of a configuration is large relative to the object being imaged, the quality of reconstructed images does not appear to be sensitive to the change of sizes. In other words, the error introduced by the effect of bounded curves can be neglected for a relatively large curve. The major limitation of this approach is felt to be the requirement that the curve $\phi(\theta)$ must be intersected exactly L times by almost all planes [31]. In order to satisfy the condition, one must reorganize the measured data so that the algorithm uses the data from each ray exactly the same number of times. This requirement may inspire a possible approach to the improvement of the current algorithms. For example, by pursuing this approach for the spiral geometry, but modifying the general form (4.29) to account for redundancy in the data, it may be possible to derive a more accurate, and yet still fast, reconstruction algorithm. One should expect a great improvement in the image quality for this case compared with those shown in the thesis.

We then formulated the PET or SPECT image reconstruction problem in the complete/incomplete data framework and derived a generalized EM (GEM) algorithm for MAP estimation based on a joint Gibbs prior of intensities and line sites. We confirmed that the use of the MRF model with a line process is a promising approach to PET image reconstruction and the need to incorporate

strong *a priori* anatomical information or structure in the algorithms, particularly in situations where the total number of detected events is low such as the measurement of changes in regional blood flow in activation studies, or in cases where the dynamics of a radiopharmaceutical are of interest.

In the 3-D phantom experiment, we know exactly the template to be reconstructed and the MRF model with a line process used fits the 3-D brain phantom almost perfectly. One attempts to make the claim from the phantom study that the proposed algorithms provide more accurate and quantitative reconstructions. However, in the patient experiment, we make no attempt to claim whether the reconstructions are quantitatively accurate, since it is not certain whether this MRF model is one of the best ways to describe a human brain. The images appear to be better than those obtained from the other methods.

Although the reconstruction algorithms are more complex than filtered back projection, the use of these methods with currently available computers is feasible. One complete iteration of the GEM algorithm, including updating the line process, requires about 40 seconds of CPU time on a SUN Sparcstation330. The number of iterations required for effective convergence varies depending on the quality of measured data. This per-iteration time is only slightly longer than the time required to perform one iteration of the ML-EM algorithm. Clearly in both cases the time could be significantly reduced using a vector array processor board of the type currently used for filtering and backprojection in clinical PET systems. The major reason that the computational cost for ML and MAP are approximately equal is that the dominant operation is the calculation of forward and back projections during the E-step. This is identical for both ML and MAP. The optimization in the M-step of the proposed algorithms requires only local computations on the image lattice, and consequently the computational cost is relatively small in comparison to forward and back projection operations which are global.

These results are preliminary and many problems remain to be addressed. Principal among these are the treatment of scatter and randoms, modifications in the procedure to provide robustness against registration errors between the PET and MR images, the development of some procedure to quantify the properties

of the estimator, and incorporation of a scheme to select the parameters of the prior.

8.2   Directions for Future Work

The research presented in this dissertation is seen as only a start in a longer-term effort to quantitatively improve the quality of reconstructions from X-ray CT and PET or SPECT, and to develop fast and accurate image reconstruction algorithms for the routine clinical diagnostic use. The applications are so far restricted to fundamental and clinical research. There are various problems that must be addressed and solved, before the research can reach the stage of clinical use.

Different areas of this research could be extended and improved, we are now working on the following two areas:

- An alternative cone beam reconstruction algorithm for two orthogonal and concentric circles.

- Modeling the inconsistence between the functional and *a priori* anatomical boundaries.

Reconstruction from cone beam data: Recall that in deriving the general formula (4.19), besides the assumption that almost all planes must intersect the curve $\phi(\theta)$ at least once, we also assumed that the curve $\phi(\theta)$ is intersected exactly L times by almost all planes [31]. As discussed in the previous section, this condition introduces the major error to reconstructions when applying to large 3-D curves relative to the object to be imaged.

Here we employ the function $L(\beta, \phi(\theta) \cdot \beta)$ in [56] to denote the number of times that a source curve is intersected by a plane defined by $\mathbf{x} \cdot \beta = \phi(\theta) \cdot \beta$. Thus, there are $L(\beta, \phi(\theta) \cdot \beta)$ values of $G(\beta, \theta)$ corresponding to each value of $K(\beta, \omega)$ in the domain Θ. For any bounded curve, a better approximation to the equation (4.12) becomes

$$\hat{f}(\mathbf{x}) = \frac{1}{(2\pi)^4} \int_{\mathbf{R}^3} \int_{\Theta} \frac{G(\beta, \theta)e^{i((\mathbf{x} - \phi(\theta)) \cdot \beta}}{L(\beta, \phi(\theta) \cdot \beta)} sgn(\phi'(\theta) \cdot \beta)(\phi'(\theta) \cdot \beta)d\theta d\beta. \quad (8.1)$$

Following a similar derivation to that in Chapter 4, we may rewrite (8.1) as

$$\hat{f}(\mathbf{x}) = \int_{\Theta} [(g'_{\alpha}(\alpha, \theta) \cdot \phi'(\theta)) *_{\alpha} h(\alpha, \theta) *_{\alpha} h_w(\alpha, \theta)]_{|\alpha = \mathbf{x} - \phi(\theta)} d\theta, \qquad (8.2)$$

where $*_{\alpha}$ denotes the 3-D convolution in α and $h(\alpha, \theta)$ is defined in (4.15) without the constant L. $h_w(\alpha, \theta)$ is defined as

$$h_w(\alpha, \theta) = \frac{1}{(2\pi)^3} \int_{\mathbf{R}^3} \frac{1}{L(\beta, \phi(\theta) \cdot \beta)} e^{i(\alpha \cdot \beta)} d\beta. \qquad (8.3)$$

Again, $h_w(\alpha, \theta)$ is treated as a generalized function. The integer function $L(\beta, \phi(\theta) \cdot \beta)$ is piecewise constant, whose value depends on both a source location and an orientation of a plane passing the source point. However, determining the exact analytical form of $L(\beta, \phi(\theta) \cdot \beta)$ would be a difficult task in many cases. In the following, we will restrict our attention to the curve consisting of two orthogonal and concentric circles. It is known that this curve satisfies the sufficient conditions stated in [57] [17] and there exists a closed form for this geometry. But it is difficult to implement as an algorithm. A fast algorithm for this geometry was developed by Clack *at el.* [6]. The ideas we will present can be considered as an alternative approach to this problem by following the general formula described in this dissertation.

First of all one must derive an analytical expression for the function $h_w(\alpha, \theta)$. Consider the case where the cone vertices are on the horizontal circle $\phi_{x_3=0}(\theta) = (R \cos \theta, R \sin \theta, 0)^T$, see Figure 8.1. The discussion and derivation is straight forward for the case where the cone vertices are on the other circle $x_1^2 + x_3^2 = R^2$.

It is clear that the circle $x_1^2 + x_3^2 = R^2$ and a source point $\phi_{x_3=0}(\theta) = (R \cos \theta, R \sin \theta, 0)^T$ uniquely determine a cone with the vertex at $\phi_{x_3=0}(\theta)$. We claim that the curve will be intersected 4 times by any plane that passes through the source point and intersects the cone mentioned above; otherwise, the curve will be intersected 2 times. Let us denote $\mathbf{C}_1$ to be the region that for any $\beta \in \mathbf{C}_1$, the plane $\mathbf{x} \cdot \beta = \phi_{x_3=0}(\theta) \cdot \beta$ intersects the curve $\phi(\theta)$ 4 times, and use $\mathbf{C}_2$ to denote the region that the plane intersects the curve 2 times. Note that $\mathbf{C}_1 + \mathbf{C}_2 = \mathbf{R}^3$.
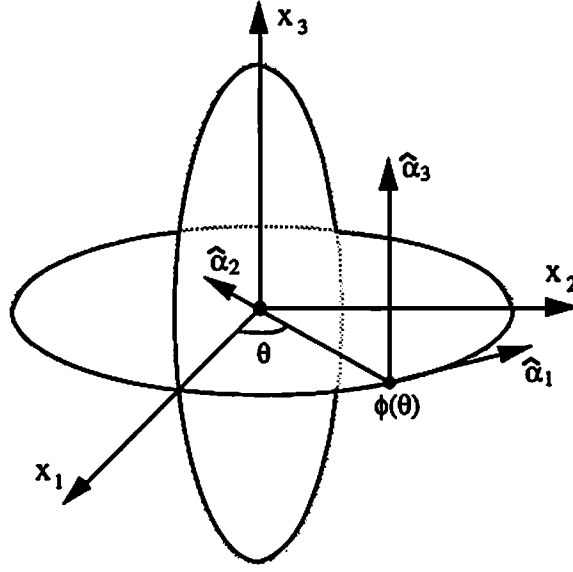
Figure 8.1: A cone beam data acquisition system with a source curve consisting of two orthogonal and concentric circles, where $\hat{\alpha} = M(\mathbf{x} - \phi_{x_3=0}(\theta))$ and $\phi_{x_3=0}(\theta) = (R\cos\theta, R\sin\theta, 0)^T$.

Thus $h_w(\alpha, \theta)$ in (8.3) can be written as

$$h_w(\alpha, \theta) = \frac{1}{4(2\pi)^3} \int_{C_1} e^{i(\alpha \cdot \beta)} d\beta + \frac{1}{2(2\pi)^3} \int_{C_2} e^{i(\alpha \cdot \beta)} d\beta. \qquad (8.4)$$

Next, we need to find the mathematical expression for the region C_1 and C_2, it is then possible to derive a closed form for the function $h_w(\alpha, \theta)$ in terms of either $\hat{\alpha}$ or α. Substituting this result into (8.2) hopefully leads to a fast algorithm.

Modeling the inconsistence: the MAP estimation algorithm described in this dissertation treats *a priori* boundaries from registered MR images as the fixed and known boundaries in PET reconstructions. Our experience with real PET and registered MR data indicates that this is not the case, it is essential to take this error between them into account. A approach similar to [15], but not strict MAP approach, is presented as follows.

Consider the first order neighborhood of image intensities and horizontal and vertical line sites. We introduce new notations $a_x(i, j)$ and $a_y(i, j)$ to represent a blurred version of the horizontal and vertical *a priori* anatomical boundaries

respectively. This blurring step is to provide somewhat robustness against registration error between the PET and MR images. It may be, for example, implemented by convolving a Gaussian distribution with the horizontal and vertical binary boundaries respectively. Then the value of $a_x(i,j)$ or $a_y(i,j)$ is on the interval $[0,1]$. The value close to "1" indicates the present of an *a priori* boundary at the location (i,j). The value close to "0" denotes the absence of an *a priori* boundary at such a location. Here we use $l_x(i,j)$ and $l_y(i,j)$ to denote the functional boundaries in PET images. The modified MAP algorithm will update each $l_x(i,j)$ and $l_y(i,j)$ at each M2-step with the aid of additional information obtained from $a_x(i,j)$ and $a_y(i,j)$. The basic criterion is that the PET reconstructions should mainly depend on measurements if *a priori* information is inconsistent with the measurements.

We may, for example, define an alternative energy function as

$$U(\lambda, l) = \sum_{i,j} \left\{ \beta \lambda_x^2(i,j)(1 - l_y(i,j)) + \alpha_y^n(i,j) l_y(i,j) \right\} + \\ \sum_{i,j} \left\{ \beta \lambda_y^2(i,j)(1 - l_x(i,j)) + \alpha_x^n(i,j) l_x(i,j) \right\} + \mathcal{H}(l), \quad (8.5)$$

where $\alpha_x^n(i,j)$ and $\alpha_y^n(i,j)$ are suggested as

$$\alpha_x^n(i,j) = \alpha[1 + (-1)^{Int(a_x(i,j)+0.5)} \epsilon_x^n(i,j)]$$
$$\alpha_y^n(i,j) = \alpha[1 + (-1)^{Int(a_y(i,j)+0.5)} \epsilon_y^n(i,j)]. \quad (8.6)$$

The function $Int(C)$ represents a operation truncating a real value C to its nearest integer and α is a constant parameter used in the previous chapters. A suitable choice for $\epsilon_x^n(i,j)$ and $\epsilon_y^n(i,j)$ may be given as

$$\epsilon_x^n(i,j) = \exp \left\{ -[a_x(i,j) - l_x^m(i,j)]^2 \right. \\ -[a_x(i-1,j) - l_x^m(i-1,j)]^2 - [a_x(i+1,j) - l_x^m(i+1,j)]^2 \\ -[a_x(i,j-1) - l_x^m(i,j-1)]^2 - [a_x(i,j+1) - l_x^m(i,j+1)]^2 \\ -[a_y(i,j) - l_y^m(i,j)]^2 - [a_y(i+1,j) - l_y^m(i+1,j)]^2 \\ \left. -[a_y(i,j-1) - l_y^m(i,j-1)]^2 \right.$$
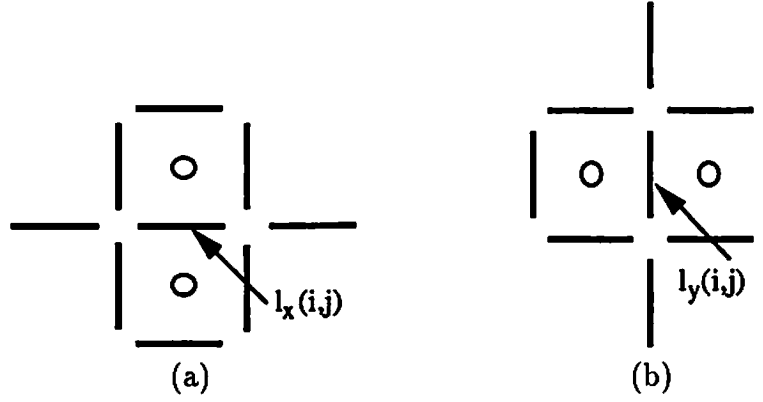
Figure 8.2: The neighborhood of line sites considered. (a): A horizontal line site $l_x(i,j)$ and its neighbors. (b): A vertical line site $l_y(i,j)$ and its neighbors.

$$-[a_y(i+1,j-1) - l_y^m(i+1,j-1)]^2\Big\}$$

$$\text{(8.7)}$$

$$\begin{aligned}
\epsilon_y^n(i,j) &= \exp\Big\{-[a_y(i,j) - l_y^m(i,j)]^2 \\
&\quad -[a_y(i,j-1) - l_y^m(i,j-1)]^2 - [a_y(i,j+1) - l_y^m(i,j+1)]^2 \\
&\quad -[a_y(i-1,j) - l_y^m(i-1,j)]^2 - [a_y(i+1,j) - l_y^m(i+1,j)]^2 \\
&\quad -[a_x(i,j) - l_x^m(i,j)]^2 - [a_x(i,j+1) - l_x^m(i,j+1)]^2 \\
&\quad -[a_x(i-1,j) - l_x^m(i-1,j)]^2 \\
&\quad -[a_x(i-1,j+1) - l_x^m(i-1,j+1)]^2\Big\}.
\end{aligned}$$

This definition involves the neighborhood of line sites as shown in Figure 8.2. Note that the penalty to each non-zero line estimate depends on the fitness between *a priori* and estimated boundaries at the location and its neighbors. Therefore the estimation based on the energy function (8.5) is not strictly MAP. For notational simplicity, we again use the index p to denote either the horizontal or vertical direction of boundaries. Let us now examine the case where $a_p(i,j) \geq 0.5$. Using (8.6), then $\alpha_p^n(i,j) = \alpha[1 - \epsilon_p^n(i,j)]$. If the values of estimated lines are close to the values of the *a priori* boundaries at each location involved for (i,j), see Figure 8.2, $\alpha_p^n(i,j)$ is small such that a "on" line site at (i,j) is encouraged. When

134

the estimated and *a priori* boundaries do not match very well at the locations, $\alpha_p^n(i,j)$ tends to the constant α, which implies that the data will dominate at this situation. The similar explanation can also apply for the case where $a_p(i,j) < 0.5$, and $\alpha_p^n(i,j) = \alpha[1 + \epsilon_p^n(i,j)]$.

One may choose a potential interaction term as

$$
\begin{aligned}
\mathcal{H}(l) \;=\; -\epsilon\alpha\sum_{i,j}\{&[l_x(i-1,j) - l_x(i,j+1)]l_x(i,j)+ \\
&[l_y(i,j-1) - l_y(i-1,j)]l_y(i,j) - \\
&l_x(i,j)l_x(i,j+1)l_y(i,j)l_y(i+1,j)\}\,.
\end{aligned}
\tag{8.8}
$$

The first two terms in the braces serve the exactly same purpose as those described in Chapter 7. The third term discourages isolated and one-pixel wide regions. Solving the problem along this direction, we hope that the modified algorithm will produce PET reconstructions with better quality.

References

[1] H. H. Barrett and W. Swindell, *Radiological Imaging*, vol. 1, Academic Press, 1981.

[2] J. E. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of Royal Statistical Society, Series B*, Vol. 48, pp. 259–302, 1986.

[3] R. Chellappa and R. Kashypa, "Texture Synthesis Using 2-D Noncausal Autoregressive Models," *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. ASSP-33, pp. 194–203, February 1985.

[4] C. T. Chen, X. Ouyang, *et al.*, "Sensor Fusion in Image Reconstruction," *IEEE Nuc. Sci. Symp.*, October 1990.

[5] Z. H. Cho, J. B. Ra, and S. K. Hilal, "True Three-Dimensional Reconstruction (TTR) - Application of Algorithm Toward Full Utilization of Oblique Rays," *IEEE Transactions on Medical Imaging*, Vol. MI-2, pp. 6–18, March 1983.

[6] R. Clack, G. L. Zeng, Y. Weng, P. E. Christian, and G. T. Gullberg, "Cone Beam Single Photon Emission Computed Tomography Using Two Orbits," *XIIth International Conference on Information Processing in Medical Imaging*, 1991.

[7] M. Defrise, R. Clack, and R. Leahy, "A Note on Smith's Reconstruction Algorithm for Cone Beam Tomography," *Submitted the Correspondence of IEEE Trans. Med. Imag.*, 1992.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical Society, Series B*, Vol. 39, No. 1, pp. 1–38, 1977.

[9] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical Cone-beam Algorithm," *J. Opt. Soc. Am.*, Vol. 1, pp. 612–619, June 1984.

[10] D. V. Finch, "Cone Beam Reconstruction With Sources On A Curve," *SIAM Journal of Applied Mathematics*, Vol. 45, pp. 665–673, August 1985.

[11] C. E. Floyd, R. J. Jaszczak, *et al.*, "Cone Beam Collimation for SPECT: Simulation and Reconstruction," *IEEE Transactions on Nuclear Science*, Vol. NS-33, pp. 511–514, February 1986.

[12] D. Geiger and A. Yuille, "A Common Framework for Image Segmentation," *MIT Technical Report*, No. 89-7, 1989.

[13] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, pp. 721–741, November 1984.

[14] S. Geman and D. E. McClure, "Bayesian Image Analysis: An Application to Single Photon Emission Tomography," *Proceedings of the American Statistical Association*, pp. 12–18, 1985.

[15] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian Reconstruction of Functional Images Using Registered Anatomical Images as Priors," *In: Information Processing in Medical Imaging, XIIth IPMI International Conference, Wye, UK.*, pp. 121–131, July 1991.

[16] I. K. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, Academic Press, Inc., 1980.

[17] P. Grangeat, "Mathematical Framework of Cone Beam 3D Reconstruction Via The First Derivative of The Radon Transform," *To appear in Mathematical Methods in Tomography, Lecture Notes in Mathematics, Springer Verlag*, 1990.

[18] P. J. Green, "Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm," *IEEE Transactions on Medical Imaging*, Vol. 9, pp. 84–93, March 1990.

[19] G. T. Gullberg, G. L. Zeng, P. E. Chrintian, B. M. Tsui, and H. T. Morgan, "Single Photon Emission Computed Tomography of the Heart Using Cone Beam Geometry and Noncircular Detector Rotation," *In: Information Processing in Medical Imaging, XIth IPMI Int. Conf.*, June 1990.

[20] T. Hebert and R. Leahy, "A Generalized EM Algorithm For 3D Bayesian Reconstruction From Poisson Data Using Gibbs Priors," *IEEE Transactions on Medical Imaging*, Vol. 8, pp. 194–202, June 1989.

[21] T. Hebert, R. Leahy, and M. Singh, "Fast MLE for SPECT Using an Intermediate Polar Representation and a Stopping Criterion," *IEEE Transactions on Nuclear Science*, Vol. 34, pp. 615–619, February 1988.

[22] G. T. Herman, *Image Reconstruction From Projections: Implementation and Applications*, Springer-Verlag, 1979.

[23] G. T. Herman, *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography*, New York: Academic Press, 1980.

[24] G. T. Herman, A. V. Lakshminarayanan, and A. Naparstek, "Convolution Reconstruction Techniques For Divergent Beams," *Comp. Biol. Med.*, Vol. 6, pp. 259–271, 1976.

[25] G. T. Herman and D. Odhner, "Performance Evaluation of an Iterative Image Reconstruction Algorithm for Positron Emission Tomography," *IEEE Transactions on Medical Imaging*, Vol. 10, pp. 336–346, September 1991.

[26] T. Holmes, D. Ficke, and D. Snyder, "Modeling of Accidental Coincidences in both Conventional and TOF PET," *IEEE Transactions on Nuclear Science*, Vol. 31, pp. 627–631, 1984.

[27] B. K. P. Horn, "Fan-Beam Reconstruction Methods," *Proceeding of the IEEE*, Vol. 67, pp. 1616–1623, December 1979.

[28] R. J. Jaszczak, K. L. Greer, and R. E. Coleman, "SPECT Using a Specially Designed Cone Beam Collimator," *The Journal of Nuclear Medicine*, Vol. 29, pp. 1398–1405, August 1988.

[29] R. J. Jaszczak, K. L. Greer, C. E. Floyd, S. H. Manglos, and R. E. Coleman, "Imaging Characteristics of a High Resolution Cone Beam Collimator," *IEEE Transactions on Nuclear Science*, Vol. 35, pp. 644–648, February 1988.

[30] V. E. Johnson, W. H. Wong, X. Hu, and C.-T. Chen, "Bayesian Restoration of PET Images Using Gibbs Priors," *XIth International Conference on Information Processing in Medical Imaging*, 1990.

[31] A. A. Kirillov, "On A Problem of I. M. Gel'fand," *Soviet Math. Dokl.*, Vol. 2, pp. 268–269, 1961.

[32] G. Kowalski, "Multislice Reconstruction from Twin-Cone Beam Scanning," *IEEE Transactions on Nuclear Science*, Vol. NS-26, pp. 2895–2903, April 1979.

[33] Y. S. Kwoh, I. S. Reed, and T. K. Truong, "A Generalized $|W|$-Filter For 3-D reconstruction," *IEEE Transactions on Nuclear Science*, Vol. NS-24, No. 5, pp. 1990–1998, 1977.

[34] K. Lange and R. Carson, "EM Reconstruction Algorithms for Emission and Transmission Tomography," *Journal of Computer Assisted Tomography*, Vol. 8, pp. 306–316, April 1984.

[35] R. Leahy and X. Yan, "Incorporation of Anatomical MR Data For Improved Functional Imaging with PET," *In: Information Processing in Medical Imaging, XIIth IPMI International Conference, Wye, UK.*, July 1991.

[36] R. Leahy and X. Yan, "Statistical Models and Methods for PET Image Reconstruction," *In Proceedings of American Statistical Association*, August 1991.

[37] E. Levitan and G. T. Herman, "A Maximum A Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography," *IEEE Transactions on Medical Imaging*, Vol. MI-6, pp. 185–192, September 1987.

[38] E. Liang and H. Hart, "Bayesian Image Processing of Data from Constrained Source Distributions I. Non-valued, Uncorrelated and Correlated Constraints," *Bull. Math. Biol.*, Vol. 49, pp. 51–74, 1987.

[39] F. M. J. Lighthill, *Introduction to Fourier Analysis and Generalised Functions*, Cambridge University Press, 1958.

[40] M. Miller, D. Snyder, and S. Moore, "An Evaluation of the Use of Sieves for Producing Estimates of Radioactivity Distribution with the EM Algorithm for PET," *IEEE Transactions on Nuclear Science*, Vol. 33, 1986.

[41] G. N. Minerbo, "Convolutional Reconstruction From Cone-beam Projection Data," *IEEE Transactions on Nuclear Science*, Vol. NS-26, pp. 2682–2684, April 1979.

[42] G. Muehllehner and J. G. Colsher, *Instrumentation in Computed Emission Tomography*, New York, 1982.

[43] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley and Sons, 1986.

[44] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C.-T. Chen, "Accurate Three-Dimensional Registration of CT, PET and/or MR Images of the Brain," *Journal of Computer Assisted Tomography*, Vol. 1, No. 13, pp. 20–26, 1989.

[45] A. Rangarajan and R. Chellappa, "Generalized Graduated Non-Convexity Algorithm For Maximum A Posteriori Image Estimation," *Proceedings of The Tenth International Conference on Pattern Recognition*, June 1990.

[46] R. Roberts and C. Mullis, *Digital Signal Processing*, Addison-Wesley, 1987.

[47] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, vol. 1, Academic Press, 2 ed., 1982.

[48] L. A. Shepp and B. F. Logan, "The Fourier Transform of a Head Section," *IEEE Transactions on Nuclear Science*, Vol. NS-21, pp. 21–43, 1979.

[49] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Transactions on Medical Imaging*, Vol. MI-1, pp. 113–122, October 1982.

[50] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka, "A Smoothed EM Approach to Indirect Estimation Problems, with Particular Reference to Stereology and Emission Tomography," *Journal of Royal Statistical Society, Series B*, Vol. 52, No. 2, pp. 271–324, 1990.

[51] T. Simchony, R. Chellappa, and Z. Lichtenstein, "Relaxation Algorithms for MAP Estimation of Gray-Level Images with Multiplicative Noise," *IEEE Transactions on Information Theory*, Vol. 36, No. 3, pp. 608–613, 1990.

[52] M. Singh, "An Electronically Collimated Gamma Camera for Single Photon Emission Computed Tomography. Part I: Theoretical Considerations and Design Criteria," *Medical Physics*, Vol. 10, pp. 421–427, July/August 1983.

[53] B. D. Smith, "Cone-beam Convolution Formula," *Comput. Biol. Med.*, Vol. 13, No. 2, pp. 81–87, 1983.

[54] B. D. Smith, "Image Reconstruction From Cone-beam Projections: Necessary and Sufficient Conditions and Reconstruction Methods," *IEEE Transactions on Medical Imaging*, Vol. MI-4, pp. 14–25, March 1985.

[55] B. D. Smith, *Computer-Aided Tomographic Imaging From Cone-Beam Data*, Ph.D. dissertation, University of Rhode Island, 1987.

[56] B. D. Smith, "Cone-Beam Tomography: Recent Advances And A Tutorial Review," *Optical Engineering*, Vol. 29, pp. 524–534, May 1990.

[57] H. K. Tuy, "An Inversion Formula For Cone-beam Reconstruction," *SIAM Journal of Applied Mathematics*, Vol. 43, No. 3, pp. 546–552, 1983.

[58] E. Veklerov and J. Llacer, "Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing," *IEEE Transactions on Medical Imaging*, Vol. MI-6, December 1987.

[59] E. H. Wood, J. H. Kinsey, R. A. Robb, B. K. Gilbert, L. D. Harris, and E. L. Ritman, *Image Reconstruction From Projection: Implementation and Applications*, Vol. 32 of Topic in Applied Physics, Berlin: Springer-Verlag, 1979. pp. 247-279.

[60] C. F. J. Wu, "On the Convergence Properties of the EM Algorithm," *Annual of Statistics*, Vol. 11, No. 1, pp. 95–103, 1982.

[61] Z. Wu, "PET and MR Image Registration By Scalp Surface Matching," *Private Communication*, 1992.

[62] Z. Wu and R. Leahy, "A Graph Theoretic Approach to Tissue Classification in MR Images," *In Proceedings of SPIE*, Vol. SPIE-1450, Feb. 1991.

[63] X. Yan and R. Leahy, "Derivation and Analysis of A Filtered Backprojection Algorithm for Cone Beam Projection Data," *IEEE Transactions on Medical Imaging*, Vol. 10, pp. 462–472, September 1991.

[64] X. Yan and R. Leahy, "MAP Image Reconstruction Using Intensity and Line Processes for Emission Tomography Data," *In Proceedings of the SPIE*, Vol. 1452, February 1991.

[65] X. Yan and R. Leahy, "Cone Beam Tomography with Circular, Elliptical and Spiral Orbits," *Physics in Medicine and Biology*, Vol. 37, pp. 493–506, March 1992.

[66] G. L. Zeng and G. T. Gullberg, "Implementation of Cone Beam Tomography with Perpendicular Orbits," *Physics in Medicine and Biology*, Vol. 37, March 1992.