LIMITED ANGLE TOMOGRAPHY

BY

MATTIEU DE VILLIERS

M.Sc., University of Cape Town, 2000
B.Sc., University of Cape Town, 1998

THESIS

Presented for the degree of Doctor of Philosophy
in the Department of Electrical Engineering,
University of Cape Town.

Cape Town, June 2004

ABSTRACT

This thesis investigates the limited angle tomography problem where axial reconstructions are produced from few measured projection views covering a 100° angular range. Conventional full angle tomography requires at least a 180° range of projection views of the patient at a fine angular spacing. Inference techniques presented in the literature, such as Bayesian methods, perform inadequately on the information-starved problem of interest.

The concept of transformation maps is developed as a problem solving framework. A transformation map extracts generic prior knowledge quantitatively from independent examples and fuses this knowledge with measured data during a reconstruction process. The success of this method is firstly due to the preciseness in using measured instead of heuristic prior knowledge. Secondly, suboptimal solutions are avoided by catering for the space variant requirements of the prior knowledge.

X-ray projection data of a human head is acquired using the Lodox full body x-ray scanner. This thesis demonstrates how a number of difficulties are overcome in using this trauma based x-ray system to produce successful reconstructions which correspond well enough to conventional tomographic reconstructions to have diagnostic value.

ACKNOWLEDGEMENTS

Thanks to the following people and institutions for their involvement in this project:

Professor Gerhard de Jager and Dr Frederick Nicolls for supervision.

Lodox for the use of their equipment and their support.

Groote Schuur Hospital for use of their equipment.

African Medical Imaging, National Research Foundation and Department of Labour for funding.

TABLE OF CONTENTS

CHAPTER **PAGE**

1 Introduction . . . 1
- 1.1 Background . . . 1
  - 1.1.1 Degenerate classes of tomography . . . 1
  - 1.1.2 Background references . . . 2
  - 1.1.3 Lodox . . . 3
- 1.2 Thesis project . . . 5
  - 1.2.1 Purpose and achievements . . . 5
  - 1.2.2 Scope and limitations . . . 6
  - 1.2.3 Quantitative and qualitative assessment of tomographic results . . . 6
- 1.3 Layout and overview . . . 7
  - 1.3.1 Chapter 2 . . . 7
  - 1.3.2 Chapter 3 . . . 7
  - 1.3.3 Chapter 4 . . . 7
  - 1.3.4 Chapter 5 . . . 7
  - 1.3.5 Chapter 6 . . . 7
  - 1.3.6 Chapter 7 . . . 8

2 Background theory . . . 9
- 2.1 Theory of computerized tomography . . . 9
  - 2.1.1 Radon transform . . . 9
  - 2.1.2 Fourier slice theorem . . . 10
  - 2.1.3 Parallel beam filtered back projection equation . . . 11
  - 2.1.4 Fan beam reconstruction theory . . . 12
- 2.2 X-ray attenuation and detection . . . 14
  - 2.2.1 X-ray attenuation . . . 15
  - 2.2.2 Poisson noise distribution . . . 15
- 2.3 Maximum likelihood solutions . . . 16
  - 2.3.1 Likelihood function . . . 16
  - 2.3.2 Solution by gradient based search methods . . . 18
  - 2.3.3 EM algorithm . . . 19
  - 2.3.4 χ^2 errors . . . 19
- 2.4 Bayes' Rule . . . 20
- 2.5 Gibbs priors . . . 21
  - 2.5.1 Definitions . . . 21
  - 2.5.2 The role of energy functions . . . 22

2.5.3 Clique potentials . . . 22
2.5.4 Maximizing the posterior distribution . . . 24
2.6 Summary . . . 24

3 Image formation process . . . 25
3.1 Data acquisition using the Lodox system . . . 25
3.1.1 Overview . . . 25
3.1.2 Image formation . . . 26
3.1.3 Operator settings and dataset . . . 26
3.1.4 Reliability . . . 27
3.1.5 Image quality . . . 28
3.2 Camera distortion . . . 31
3.2.1 Pixel column position errors . . . 31
3.2.2 Pixel row position errors . . . 32
3.3 Camera overlap artefacts . . . 33
3.3.1 Artefact properties . . . 34
3.3.2 Intensity level based algorithm . . . 35
3.3.3 Adaptive algorithm . . . 36
3.3.4 Examples and performance . . . 38
3.4 Consistency . . . 39
3.5 Projection geometry . . . 41
3.6 Pixel projection . . . 45
3.7 Projection axis resolution . . . 46
3.8 Filtered back-projection reconstruction . . . 47
3.8.1 Results . . . 47
3.8.2 Performance . . . 50
3.9 Summary . . . 51

4 Optimizing reconstruction . . . 55
4.1 Maximum likelihood reconstruction . . . 56
4.1.1 Direct inversion in a single step . . . 56
4.1.2 Iterative reconstruction methods . . . 58
4.1.2.1 Steepest descent . . . 59
4.1.2.2 Conjugate gradients . . . 60
4.1.2.3 Using components of previous search directions . . . 60
4.1.3 Iterative volume reconstruction . . . 61
4.2 Implications of having sparse data . . . 61
4.2.1 Ambiguity and the influence of the starting point . . . 62
4.2.2 Inconsistencies and overfitting . . . 62
4.2.3 Positivity constraint . . . 63
4.3 The properties of medical tomographic images . . . 65
4.3.1 Hounsfield units . . . 68
4.3.2 Noise and artefacts in Lodox CT images . . . 69
4.3.3 Dominant attenuation classes . . . 70
4.3.4 Transformation fields . . . 71
4.4 Regularization by Bayesian methods . . . 72

4.4.1 Summary of prior knowledge . . . 72
4.4.2 Adaptive potential functions and clique potentials . . . 72
4.4.3 Maximizing the posterior . . . 73
4.5 Summary . . . 75

5 Transformation maps . . . 77
5.1 Introduction . . . 77
5.1.1 Neural networks . . . 78
5.1.2 Transformation maps . . . 78
5.2 Examples . . . 78
5.2.1 Signal restoration example . . . 79
5.2.2 Image restoration example . . . 80
5.3 Definitions and implementation details . . . 84
5.4 Application to tomography . . . 86
5.4.1 Randomness of tomographic artefacts . . . 86
5.4.2 Removal of angular aliasing artefacts in single step reconstruction . . . 88
5.4.3 Understanding the transformation map . . . 88
5.4.4 Iterative regularization . . . 90
5.4.5 Populating the transformation map . . . 92
5.5 Map effectiveness . . . 92
5.5.1 Optimal parameters . . . 92
5.5.2 Target accuracy . . . 93
5.6 Summary . . . 93

6 Limited angle tomography . . . 95
6.1 Unregularized limited angle reconstructions . . . 95
6.1.1 Caution on simulations, resolution and data fit . . . 95
6.1.2 Properties of limited angle reconstructions . . . 97
6.1.3 Comparing limited to full angle reconstructions . . . 97
6.2 Reconstructions using partial spatial knowledge . . . 99
6.3 Survey of limited angle tomography methods . . . 102
6.3.1 Direct inversion . . . 102
6.3.2 Regularization . . . 103
6.3.3 Sinogram extrapolation . . . 105
6.3.4 Projection onto convex sets . . . 105
6.3.5 Optimal views . . . 105
6.3.6 Complementary data fusion . . . 106
6.3.7 Shape modeling . . . 106
6.4 Bayesian limited angle reconstructions . . . 108
6.5 Reconstruction using a transformation map . . . 110
6.5.1 Training examples for the transformation map . . . 110
6.5.2 Feature vectors . . . 111
6.5.3 The effect of the transformation map . . . 113
6.5.4 Results . . . 115
6.6 Envisaged limited angle tomographic system . . . 119
6.6.1 Scanning hardware . . . 119

6.6.2 Gathering prior knowledge . . . 119
6.6.3 Speeding up processing time . . . 120
6.7 Summary . . . 120

7 Conclusion . . . 121
7.1 Diagnostic value of reconstructions . . . 121
7.1.1 High resolution full angle reconstructions . . . 121
7.1.2 Low resolution sparse angle reconstructions . . . 121
7.1.3 Coronal views of high resolution limited angle reconstructions . . . 122
7.1.4 Low resolution unregularized limited angle reconstructions . . . 122
7.1.5 Low resolution limited angle reconstructions . . . 122
7.2 Quantitative comparison of results . . . 123
7.3 Summary of developments . . . 124
7.4 Findings . . . 125
7.5 Results . . . 125
7.6 Conclusions . . . 126
7.7 Recommendations . . . 126

REFERENCES . . . 127

LIST OF FIGURES

Figure		Page
1.1	Axial tomographic reconstructions.	2
1.2	Lodox full body x-ray scanner.	3
1.3	Lodox full body scan.	4
2.1	Parallel beam projection.	10
2.2	Fan beam projection.	13
2.3	Discontinuity adaptive functions.	23
3.1	Lodox scans that are part of the dataset.	27
3.2	Faulty Lodox scans.	28
3.3	Image quality measurements.	29
3.4	Image quality results.	30
3.5	Closeup view of steel ruler.	31
3.6	Detected tick marks.	32
3.7	Tick mark spacing and pixel position errors.	33
3.8	High resolution view of one metal bar.	34
3.9	Artefact properties and definitions.	35
3.10	Example of the artefact removal algorithm applied to a chest x-ray.	38
3.11	Results for artefact removal algorithms.	39
3.12	Accumulated row intensity profiles.	40
3.13	Detection of a pin position in a Lodox scan.	42
3.14	Geometry of the Lodox machine.	43
3.15	Contribution of pixel segment to a projection bin.	44
3.16	Accumulative projection of triangle.	45
3.17	High resolution projection profiles of individual pixels.	46
3.18	Filtered back-projection reconstruction slice at $i_z = 225$.	48
3.19	Intensity profiles.	49
3.20	Filtered back-projection reconstruction slice at $i_z = 450$.	52
3.21	High resolution reconstruction.	53
3.22	Filtered back-projection slices.	54
4.1	Single step reconstructions.	56
4.2	Optimal filters for direct inversion of 128×128 reconstructions.	57
4.3	Projection data errors.	58
4.4	Convergence curves for iterative reconstruction methods with $\Delta\theta = 5°$.	59

4.5 Isosurface views of reconstructed volume using the complete dataset. 61
4.6 Iterated reconstructions. 62
4.7 Projection data errors for iterated reconstructions. 63
4.8 Reconstructions using positivity constraint after 100 iterations. 64
4.9 Projection data errors for iterated reconstructions with positivity constraint. . . 65
4.10 High resolution positively constrained iterated reconstructions. 66
4.11 Sagittal and coronal views of a full angle reconstruction. 67
4.12 Full angle reconstruction and attenuation levels in Hounsfield units. 68
4.13 Intensity profiles. 69
4.14 Histograms for iterated $128 \times 128 \times 128$ volumetric reconstructions with positivity constraint. 70
4.15 Transformation fields for $128 \times 128 \times 128$ reconstructions. 71
4.16 Data fit standard deviations for DA-MRF reconstructions. 74
4.17 Regularized reconstructions. 75

5.1 Application of a transformation map to correct the pulse shapes in a signal. . . . 79
5.2 Denoising example. 81
5.3 Transformation map for denoising. 82
5.4 Matlab code for image denoising example. 83
5.5 Removal of angular aliasing artefacts from filtered back-projection reconstructions with no iteration using the transformation field method. 87
5.6 Slices of the 64^3 transformation map used in Figure 5.5 for the removal of tomographic artefacts. 89
5.7 Iterated reconstructions biased using transformation field method with positivity constraint. 91

6.1 Effect of using a high projection axis resolution for limited angle reconstruction. 96
6.2 Unregularized limited angle reconstructions showing the effect of having more available data. 98
6.3 Histogram comparison of limited to full angle reconstructions. 99
6.4 Limited and full angle reconstructions with intensity profiles through the reconstructions. 100
6.5 Closeup views of positively iterated reconstructions. 101
6.6 Sagittal and coronal views of a limited angle reconstructed volume. 107
6.7 Bayesian limited angle reconstructions. 109
6.8 Sample quantities used to determine transformation map. 111
6.9 Steps in building a transformation map. 112
6.10 Iterative limited angle reconstruction cycle. 112
6.11 Enlarged views of the transformation map. 113
6.12 Transformation map for regularizing limited angle reconstructions. 114
6.13 Comparison of transformation map limited angle reconstruction results. 115
6.14 Limited angle reconstructions. 117
6.15 More limited angle reconstructions. 118

7.1 Quantitative comparison of different reconstruction algorithms showing typical values of fit to the ground truth (σ_{1°). 124

CHAPTER 1

Introduction

Computed tomography has widespread application in non-destructive testing, electron microscopy and medicine. The cross-sectional views produced by tomographic reconstructions (see Figure 1.1(a)) exhibit spatial and density information which is not presented by projection views of x-ray imaging and therefore has an added diagnostic capacity.

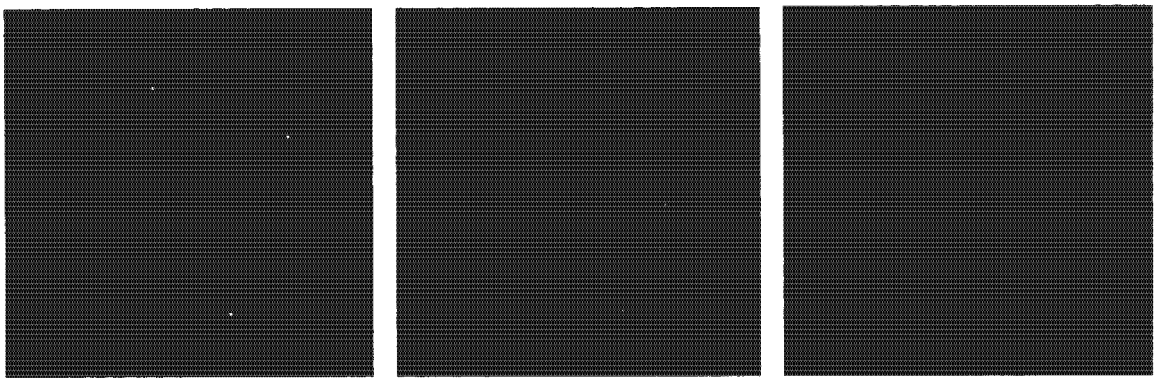
1.1 Background

Conventional computed tomography requires that the patient be x-rayed from all angles in the cross-section, or *axial*, plane. More precisely, an x-ray must pass through every point in that plane from every direction over 180°. This amounts to a substantial radiation dose delivered to a patient and any way of reducing this requirement is helpful.

In some situations of non-destructive testing, the dimensions of the object are such that it cannot practically be x-rayed from all angles, or the density is so great that x-rays cannot penetrate the object from all angles. Likewise in electron microscopy, physical tilt limitations result in a cone of unmeasurable data.

1.1.1 Degenerate classes of tomography

Local tomography refers to the reconstruction from projections that do not expose the entire object to radiation. *Sparse angle tomography* (see Figure 1.1(b)) is performed when data is available at discrete angular steps only. When the angular range of available data is restricted,



(a) Full angle reconstruction using data at $\Delta\theta = 1°$ intervals over an angular range of $180°$, $\sigma_{1^\circ} = 79.13$.

(b) Sparse angle reconstruction using data at $\Delta\theta = 5°$ intervals over an angular range of $180°$, $\sigma_{1^\circ} = 489$.

(c) Limited angle reconstruction using data at $\Delta\theta = 5°$ intervals over an angular range of $90°$, $\sigma_{1^\circ} = 889$.

Figure 1.1 Axial tomographic reconstructions of the same slice are created using a decreasing amount of data from left to right. In comparison to (a), observe angular aliasing artefacts in (b), and a further loss of edge information in the vertical direction in (c). The values for σ_{1° indicate correspondence to the ground truth as discussed in Section 1.2.3 and elsewhere.

limited angle tomography reconstructions are created (see Figure 1.1(c)). This thesis contributes to the field of limited angle tomography.

Any lack of projection data results in artefacts or some form of absence or ambiguity of information in the reconstruction. These artefacts appear as stripes through the image as seen in Figure 1.1(b) due to course angular sampling intervals. The absence of edge information due to a corresponding limited angular range is illustrated in Figure 1.1(c).

In the literature, various methods have been proposed to improve on results when data is incomplete. Better results have been achieved in cases where it is possible to model the object well; but increasingly poor results are obtained the more ill-posed the problem is and the less prior or additional knowledge of the object is available.

1.1.2 Background references

Skilling (Skilling, 1998) outlines the well-known Bayesian framework for dealing with inference problems where prior knowledge is combined with measured data to estimate unknown parameters. This paper provides a broad perspective of the inherent problem faced by this thesis and

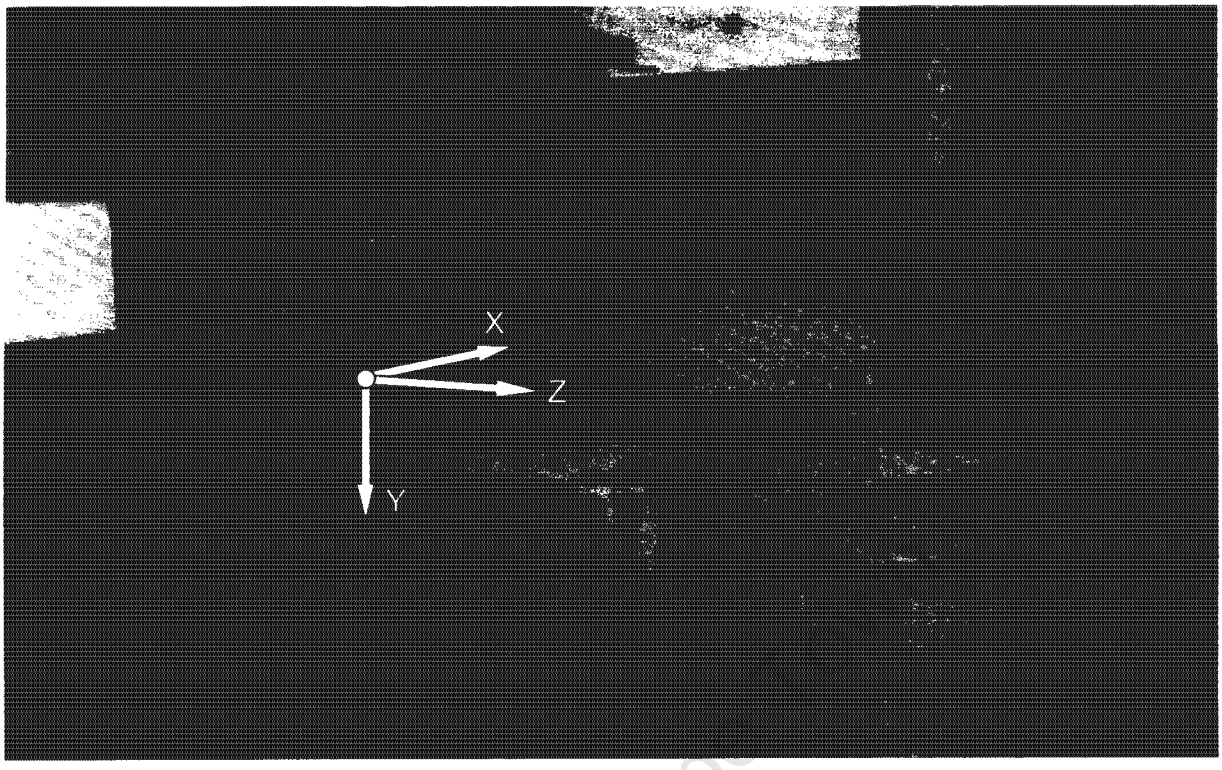


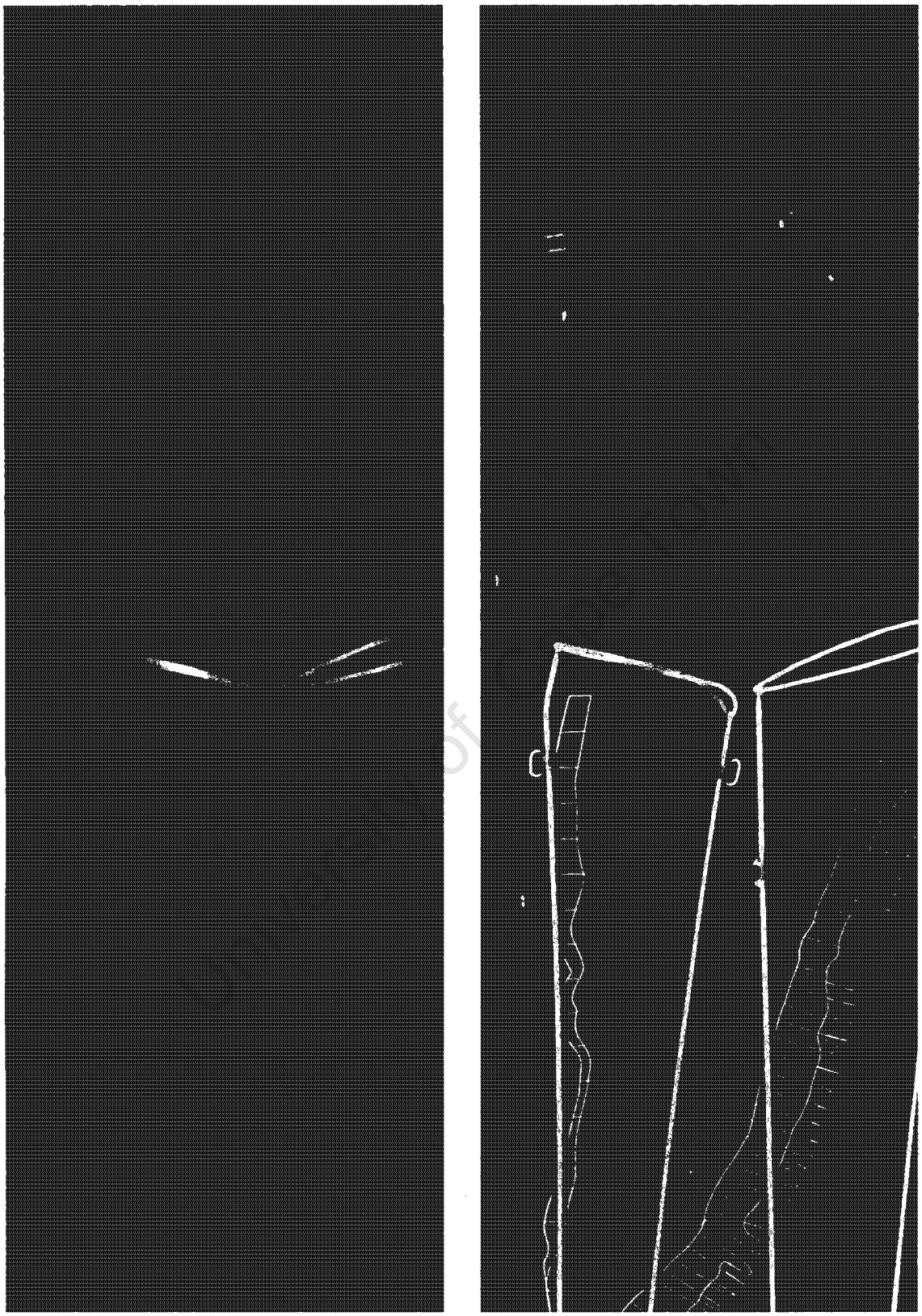
Figure 1.2 Full body x-ray scanner installed in Groote Schuur Hospital used to acquire data for this project.

points out some difficulties experienced in determining solutions. A good overview of sparse and limited angle tomography can be found in (Siltanen *et al.*, 2003).

1.1.3 Lodox

Lodox is a South African company that manufactures full body digital x-ray scanners. Their product StatScan, of which an example is shown in Figure 1.2, is primarily targeted at medical trauma applications. Rapid scanning, high image quality and low dose are the key elements of its design (Beningfield *et al.*, 2003).

A C-arm which houses an x-ray tube on top and a digital detector beneath the table scans a patient from head to toe in the z-direction as indicated in Figure 1.2. This C-arm can also rotate around the patient in the $x \times y$ cross-section plane for lateral imaging. Rotation occurs counter clockwise as in this view from an upright position at 0° to a maximum of 100°.



(a) Unprocessed image.

(b) Enhanced image.

Figure 1.3 Lodox full body scan. Notice fractures in all limbs.

Athough the machine is not designed for this purpose, this thesis proves that it can be used to produce tomographic reconstructions. Compared to a conventional spiral tomography system, this system geometry has the advantage that it offers a high resolution in the scanning direction. The disadvantages are a coarse angular resolution and a limited angular range.

For interest's sake Figure 1.3 shows a classic Lodox full body scan. Lodox employs a proprietary algorithm that improves the sharpness of x-ray images and suppresses the overall dynamic range by enhancing regional contrast as outlined in (Beutel *et al.*, 2000). Such enhancement facilitates the diagnosis of patients since it is unnecessary for radiologists to change the presentation state (contrast and brightness) of images. Note that the enhanced image is only shown for clarity of presentation and no such processing is performed on scans used for tomographic reconstruction.

1.2 Thesis project

The question that this project addresses is: 'Given realistic constraints, can the Lodox machine be used to produce limited angle reconstructions that are of diagnostic value?' This question can be restated in terms of a visual example: 'Can we create a reconstruction that resembles Figure 1.1(a) more closely in a reliable way while only using the measured data that was available to produce Figure 1.1(c)?' And as a less ambitious first step: 'Can we suppress the artefacts in Figure 1.1(b) so that the reconstruction appears more like Figure 1.1(a)?'

1.2.1 Purpose and achievements

The purpose of this thesis is to develop a tomographic capability for the Lodox machine with particular emphasis on limited angle tomography. Using actual measurements and addressing the practical problem specifically for this system supercedes noise free simulation and likewise simplified studies.

There are a number of unavoidable problems that are overcome through the course of this project in pursuit of the end goal which is limited angle tomographic reconstruction. The first portion of this work is concerned with the straightforward yet essential development of tomographic functionality.

Thereafter a novel method is presented which is able to extract generic prior knowledge from examples. This method successfully fuses such recorded knowledge with measured data during a

reconstruction process. Results of this thesis shows that significant improvements are achieved for sparse and particularly limited angle reconstructions.

1.2.2 Scope and limitations

All results are determined from physical noisy measurements. A single complete dataset was acquired from a human head at angular increments of $\Delta\theta = 1°$ over an angular range of 180° at a resolution of 1.66 line pairs per millimeter. Appropriate subsets of this complete dataset (typically at angular increments of $\Delta\theta = 5°$) are used as data for sparse and limited angle tomographic reconstructions.

For limited angle tomography, the scope is restricted to the special case of having projection data available at 5° intervals over a 100° angular range. Only the 128×128 1.8mm per pixel resolution condition is investigated for limited angle reconstruction while higher resolution reconstructions are provided for full angle tomography.

1.2.3 Quantitative and qualitative assessment of tomographic results

The complete dataset can be used as the *ground truth* because it defines the reconstruction unambiguously (as discussed later). In contrast, sparse and limited angle reconstructions are not adequately defined by the limited data available during their reconstruction. Since a complete dataset is acquired after all, a corresponding *complete* set of projections for any sparse or limited angle reconstruction can be compared to actual measured projections to evaluate the effectiveness of the particular reconstruction method in question.

Note that most figures of reconstructions (see Figure 1.1 for example) quote values for $\sigma_{1°}$. This value ($\sigma_{1°}$) is the root mean square error that the projections of the reconstruction make with the complete dataset projections (at angular increments of 1° over an angular range of 180°) regardless of the limitations imposed on the availability of data during reconstruction.

In addition to numerical measures, the opinion of a professional radiologist on the diagnostic quality of the results is quoted in the final chapter. This qualitative assessment complements the trends observed numerically and provides a real world reference frame.

1.3 Layout and overview

As outlined below, more fundamental aspects are discussed first, leading up to limited angle tomography in the final chapters.

1.3.1 Chapter 2

This chapter presents essential background theory on tomographic reconstruction and Bayesian methods. Familiarity with these topics are assumed later in the thesis.

1.3.2 Chapter 3

A chapter on the image formation process describes the peculiarities of the Lodox system and discusses solutions specific to its imaging problems. The physical geometry of the machine is accurately calibrated from the dataset and it is shown how filtered back-projection reconstructions are produced.

1.3.3 Chapter 4

Efficient methods for iterative reconstruction are investigated. Ambiguity due to sparse projection data is demonstrated, which indicates the need for regularization or the incorporation of additional knowledge. Prior knowledge is identified and formulated into a Bayesian framework.

1.3.4 Chapter 5

Transformation maps are introduced as an alternative to Bayesian methods. This machine learning technique is able to capture, as well as apply, prior knowledge of a more natural kind.

1.3.5 Chapter 6

Finally the topic of limited angle tomography is addressed. This chapter includes a survey of existing limited angle tomography methods. After a study of the properties of limited angle reconstructions, the transformation map method is applied to the problem.

1.3.6 Chapter 7

The last chapter provides commentary of a radiologist on the diagnostic value of the reconstructions presented. A summary of the project developments is presented and conclusions are drawn from the findings. The thesis ends with recommendations for further study.

CHAPTER 2

Background theory

The purpose of this chapter is to provide necessary background theory to readers that are inexperienced with tomographic reconstruction and Bayesian methods. The content of this chapter is adapted from (de Villiers, 2000) and does not present any new material.

2.1 Theory of computerized tomography

The following subsections provide a summary of the theory behind the filtered back-projection reconstruction algorithm. A comprehensive overview of tomographic techniques can be found in (Kak and Slaney, 1988) and references found therein (for below). The idea of filtered back-projection was first proposed by Bracewell and Riddle in 1967 and later independently by Ramachandran and Lakshminarayanan in 1971. The superiority of this algorithm over the algebraic techniques, originally used by Hounsfield, was demonstrated by Shepp and Logan in 1974. In 1975 Lakshminarayanan extended the theory to the equispaced collinear detector case with a fan beam geometry similar to that of the Lodox machine.

2.1.1 Radon transform

Figure 2.1 illustrates a parallel beam projection $P_{\theta^{\parallel}}^{\parallel}(\rho^{\parallel})$ of the cross-section function $f(x,y)$ at angle $\theta^{\parallel}$. The value of the projection at $\rho^{\parallel} = \rho_0^{\parallel}$ is the line integral of $f(x,y)$ along ray_0, so

$$P_{\theta^{\parallel}}^{\parallel}(\rho_0^{\parallel}) = \int_{\text{ray}_0} f(x,y)dl. \tag{2.1}$$

The line labeled ray_0 is expressed as

$$x\cos\theta^{\parallel} + y\sin\theta^{\parallel} = \rho_0^{\parallel} \tag{2.2}$$

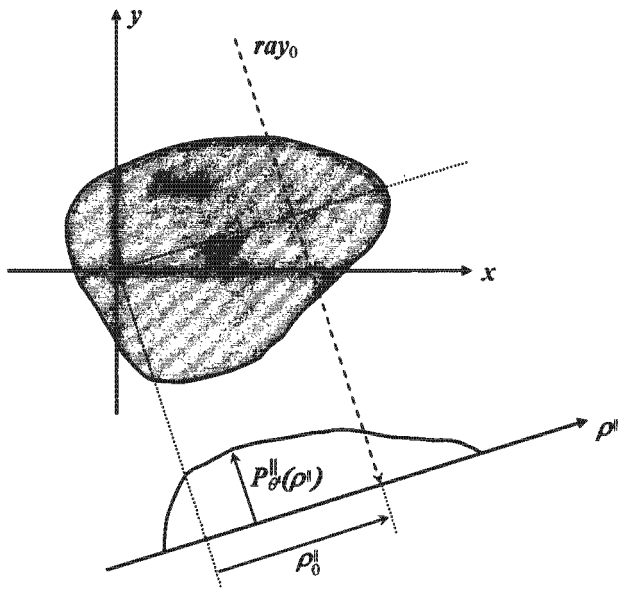


Figure 2.1 An example parallel beam projection geometry.

which can easily be derived from $y = mx + c$ with

$$m = -\frac{1}{\tan\theta^{\|}} \quad \text{and} \quad c = \frac{\rho_0^{\|}}{\sin\theta^{\|}}.$$

Therefore, for different values of $\rho^{\|}$ and $\theta^{\|}$, Equation 2.1 can be rewritten as

$$P_{\theta^{\|}}^{\|}(\rho^{\|}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta^{\|} + y\sin\theta^{\|} - \rho^{\|})dxdy \tag{2.3}$$

which is the *Radon transform* of $f(x,y)$. The forward Radon transform describes mathematically how projections are formed from a cross-section.

2.1.2 Fourier slice theorem

The Fourier slice theorem relates the measured projection data $P_{\theta^{\|}}^{\|}(\rho^{\|})$ to the two-dimensional Fourier transform $F(u,v)$ of the cross-sectional image $f(x,y)$. This theorem forms the basis of the filtered back-projection method which is commonly used to perform the inverse Radon transform when a complete dataset is available.

The Fourier transform of the cross-section function $f(x,y)$ is defined as

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy. \tag{2.4}$$

Along the line in the Fourier domain given by $v = 0$, the Fourier transform of the cross-section is

$$F(u,0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi ux}dxdy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x,y)dy\right]e^{-j2\pi ux}dx. \tag{2.5}$$

However, the parallel beam projection $P^{\|}_{\theta^{\|}}(\rho^{\|})$ at angle $\theta^{\|} = 0$ is

$$P^{\|}_{\theta^{\|}=0}(\rho^{\|}) = \int_{-\infty}^{\infty} f(x,y)dy. \tag{2.6}$$

By substitution, Equation 2.5 can be rewritten as

$$F(u,0) = \int_{-\infty}^{\infty} P^{\|}_{\theta^{\|}=0}(\rho^{\|})e^{-j2\pi ux}dx \tag{2.7}$$

which equals the Fourier transform $S^{\|}_{\theta^{\|}=0}(\omega)$ of the projection $P^{\|}_{\theta^{\|}=0}(\rho^{\|})$ by definition. In other words, the one-dimensional Fourier transform of the projection at angle $\theta^{\|} = 0$ equals the two-dimensional Fourier transform of the cross-sectional image along a line at the same angle. However, due to the rotational invariance property of the two-dimensional Fourier transform this is true for all angles, so

$$S^{\|}_{\theta^{\|}}(\omega) = F_{polar}(\omega,\theta^{\|}) \triangleq F(\omega\cos\theta^{\|}, \omega\sin\theta^{\|}). \tag{2.8}$$

2.1.3 Parallel beam filtered back projection equation

By measuring an infinite number of projections, $F_{polar}(\omega,\theta^{\|})$ and hence $F(u,v)$ will be known over the entire Fourier plane. The cross-sectional image can then be recovered using the inverse Fourier transform in Cartesian coordinates

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}dxdy \tag{2.9}$$

or equivalently, using polar coordinates,

$$f(x,y) = \int_{0}^{\pi}\int_{-\infty}^{\infty} F_{polar}(\omega,\theta^{\|})e^{j2\pi(x\omega\cos\theta^{\|}+y\omega\sin\theta^{\|})}|\omega|d\omega d\theta^{\|} \tag{2.10}$$

where $|\omega|$ is the familiar Jacobian for converting the integrand to polar coordinates. By recalling that $\rho^{\|} = x\cos\theta^{\|} + y\cos\theta^{\|}$ and $S^{\|}_{\theta^{\|}}(\omega) = F_{polar}(\omega,\theta^{\|})$,

$$f(x,y) = \int_{0}^{\pi}\left[\int_{-\infty}^{\infty} S^{\|}_{\theta^{\|}}(\omega)|\omega|e^{j2\pi\omega\rho^{\|}}d\omega\right]d\theta^{\|} = \int_{0}^{\pi} Q^{\|}_{\theta^{\|}}(x\cos\theta^{\|} + y\sin\theta^{\|})d\theta^{\|} \tag{2.11}$$

where

$$Q^{\|}_{\theta^{\|}}(\rho^{\|}) = \int_{-\infty}^{\infty} S^{\|}_{\theta^{\|}}(\omega)|\omega|e^{j2\pi\omega\rho^{\|}}d\omega. \tag{2.12}$$

$Q^{\|}_{\theta\|}(\rho^{\|})$ is regarded as a *filtered* projection since the Fourier transform of the projection $S^{\|}_{\theta\|}(\omega)$ is multiplied by $|\omega|$ in the Fourier domain before the result is inverse Fourier transformed. In the space domain, this operation is equivalent to convolution with the impulse response of the filter $h(\rho^{\|}) = \mathcal{F}^{-1}\{|\omega|\}$, so

$$Q^{\|}_{\theta\|}(\rho^{\|}) = \int_{-\infty}^{\infty} P^{\|}_{\theta\|}(\rho^*)h(\rho^{\|} - \rho^*)d\rho^* \quad \text{where} \quad h(\rho^{\|}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\omega| e^{j2\pi\omega\rho^{\|}} d\omega. \tag{2.13}$$

Finally using this result and exhanging the dummy variable $\rho^* \Leftrightarrow \rho^{\|}$ in Equation 2.13, Equation 2.11 can be rewritten as

$$f(x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} P^{\|}_{\theta\|}(\rho^{\|})h(x\cos\theta^{\|} + y\sin\theta^{\|} - \rho^{\|})d\rho^{\|}d\theta^{\|} \tag{2.14}$$

which is a mathematical representation of the filtered back projection algorithm for parallel beam projections.

2.1.4 Fan beam reconstruction theory

Practical scanners often have a fan beam geometry, as shown in Figure 2.2, where an entire projection is measured at a time using an isotropic x-ray source. Fan beam projections $P_\theta(\rho)$ are distorted when compared to parallel beam projections $P^{\|}_{\theta\|}(\rho^{\|})$ and therefore some modification to the theory is required.

It is convenient to introduce virtual fan beam ρ' as well as parallel beam $\rho^{\|}$ projection axes that pass through the center of rotation. The projection values for the different axes are assumed to be equal where the same ray intersects the axes, i.e.

$$P_\theta(\rho) = P'_\theta(\rho') = P^{\|}_{\theta\|}(\rho^{\|}). \tag{2.15}$$

Therefore parallel beam projection data can be extracted (inconsecutively) from the fan beam data using Equation 2.15 where

$$\begin{aligned} \rho^{\|} &= \rho'\cos\phi = \frac{\rho' t_{cen}}{\sqrt{\rho'^2 + t_{cen}^2}} & \theta^{\|} &= \theta + \phi \\ \rho' &= \rho\frac{t_{cen}}{t_{len}} & \phi &= \tan^{-1}\frac{\rho'}{t_{cen}}. \end{aligned} \tag{2.16}$$

From Equation 2.14, the cross-section function in polar coordinates $f_{polar}(r,\vartheta)$ becomes

$$f_{polar}(r,\vartheta) = \int_0^{\pi}\int_{-\infty}^{\infty} P^{\|}_{\theta\|}(\rho^{\|})h(r\cos(\theta^{\|} - \vartheta) - \rho^{\|})d\rho^{\|}d\theta^{\|} \tag{2.17}$$

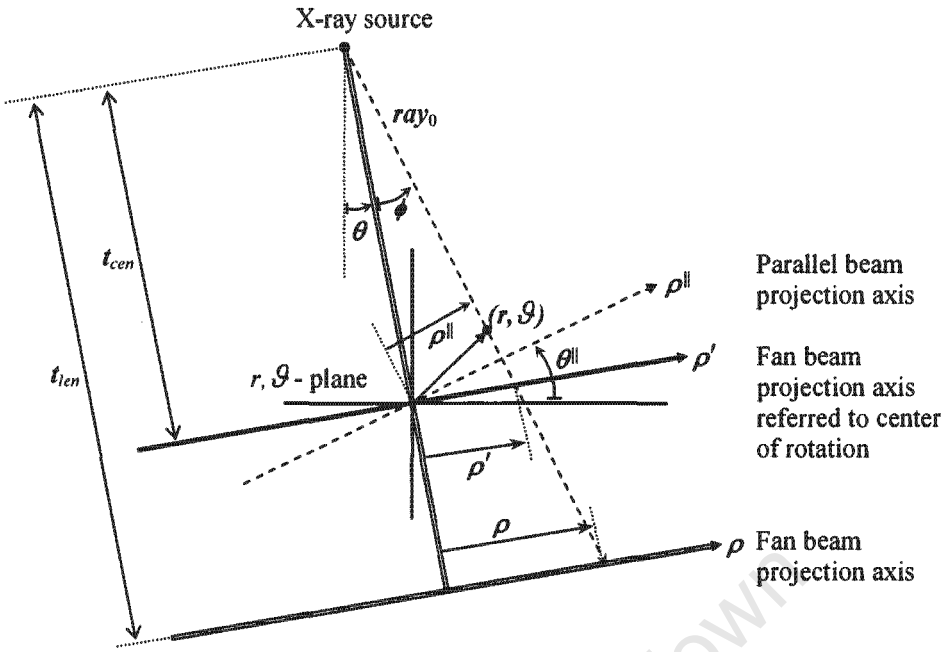


Figure 2.2 A typical fan beam projection geometry.

by substituting $x = r\cos\vartheta$ and $y = r\sin\vartheta$ and using the identity $\cos(A-B) = \cos A\cos B + \sin A\sin B$. Substituting $P'_\theta(\rho') = P^{\|}_{\theta^{\|}}(\rho^{\|})$ and changing the variables of integration using Equation 2.16,

$$f_{polar}(r,\vartheta) = \int_{-\pi/2}^{\pi/2}\int_{-\infty}^{\infty} P'_\theta(\rho')h\left[r\cos(\theta+\tan^{-1}\left(\frac{\rho'}{t_{cen}}\right)-\vartheta)-\frac{\rho' t_{cen}}{\sqrt{\rho'^2+t_{cen}^2}}\right]\frac{t_{cen}^3}{(\rho'^2+t_{cen}^2)^{3/2}}d\rho' d\theta \tag{2.18}$$

where the Jacobian $J = \frac{t_{cen}^3}{(\rho'^2+t_{cen}^2)^{3/2}}$ is calculated from

$$J = \left|\frac{\partial\rho^{\|}}{\partial\rho'}\cdot\frac{\partial\theta^{\|}}{\partial\theta} - \frac{\partial\rho^{\|}}{\partial\theta}\cdot\frac{\partial\theta^{\|}}{\partial\rho'}\right| = \left|\frac{t_{cen}^3}{(\rho'^2+t_{cen}^2)^{3/2}}\cdot 1 - 0\cdot 0\right| \tag{2.19}$$

Using $\cos(A+B) = \cos A\cos B - \sin A\sin B$ and Pythagoras' rule, the argument of h in square brackets of Equation 2.18 can be rewritten as

$$r\cos(\theta+\tan^{-1}\left(\frac{\rho'}{t_{cen}}\right)-\vartheta)-\frac{\rho' t_{cen}}{\sqrt{\rho'^2+t_{cen}^2}}$$

$$
\begin{aligned}
&= r\cos(\theta-\vartheta)\frac{t_{cen}}{\sqrt{\rho'^2+t_{cen}^2}} - (t_{cen}+r\sin(\theta-\vartheta))\frac{\rho'}{\sqrt{\rho'^2+t_{cen}^2}} \\
&= \left(\frac{t_{cen}r\cos(\theta-\vartheta)}{t_{cen}+r\sin(\theta-\vartheta)} - \rho'\right)\frac{t_{cen}+r\sin(\theta-\vartheta)}{\sqrt{\rho'^2+t_{cen}^2}} \\
&= (\rho^*-\rho')\frac{t_{cen}U}{\sqrt{\rho'^2+t_{cen}^2}} \qquad (2.20)
\end{aligned}
$$

if ρ^* and U are defined as

$$
\rho^* = \frac{t_{cen}r\cos(\theta-\vartheta)}{t_{cen}+r\sin(\theta-\vartheta)} \quad \text{and} \quad U = \frac{t_{cen}+r\sin(\theta-\vartheta)}{t_{cen}}. \qquad (2.21)
$$

From the definition of $h(\rho^{\|})$ given in Equation 2.13,

$$
\begin{aligned}
h\left[(\rho^*-\rho')\frac{t_{cen}U}{\sqrt{\rho'^2+t_{cen}^2}}\right] &= \frac{1}{2\pi}\int_{-\infty}^{\infty}|\omega|e^{j2\pi\omega(\rho^*-\rho')t_{cen}U/\sqrt{\rho'^2+t_{cen}^2}}d\omega \\
&= \frac{\rho'^2+t_{cen}^2}{t_{cen}^2U^2}\frac{1}{2\pi}\int_{-\infty}^{\infty}|\omega^*|e^{j2\pi(\rho^*-\rho')\omega^*}d\omega^* \\
&= \frac{\rho'^2+t_{cen}^2}{t_{cen}^2U^2}h(\rho^*-\rho') \qquad (2.22)
\end{aligned}
$$

using the transformation

$$
\omega^* = \omega\frac{t_{cen}U}{\sqrt{\rho'^2+t_{cen}^2}} \qquad (2.23)
$$

for which $d\omega = \frac{\sqrt{\rho'^2+t_{cen}^2}}{t_{cen}U}d\omega^*$ and $|\omega| = \frac{\sqrt{\rho'^2+t_{cen}^2}}{|t_{cen}U|}|\omega^*|$. Finally, by substituting Equation 2.22, Equation 2.18 is rewritten as

$$
f_{polar}(r,\vartheta) = \int_{-\pi/2}^{\pi/2}\frac{1}{U^2}\int_{-\infty}^{\infty}P'_{\theta}(\rho')h(\rho^*-\rho')\frac{t_{cen}}{\sqrt{\rho'^2+t_{cen}^2}}d\rho'd\theta \qquad (2.24)
$$

which can be interpreted as a mathematical formulation of the filtered back projection algorithm for a fan beam geometry.

2.2 X-ray attenuation and detection

This section provides some background on how x-rays are attenuated through an object to produce an image of the object on a digital detector. The statistical nature of the image data is also discussed.

2.2.1 X-ray attenuation

A projection $P(\rho, \theta)$ is the line integral through the density distribution function $f(x, y)$ along the beam path,

$$P(\rho, \theta) = \int_{\text{path}} f(x, y) dl. \tag{2.25}$$

However, due to the Compton scattering and absorption characteristics of x-rays (Kak and Slaney, 1988), the photon count detected is approximated by

$$k'_{i'} \approx e^{\int_{\text{path}} f(x,y) dl}. \tag{2.26}$$

A projection is therefore related to its corresponding photon count by

$$P(\rho, \theta) \approx \ln N_{k'_{i'}} - \ln k'_{i'} \tag{2.27}$$

where $N_{k'_{i'}}$ is the number of photons that would have been counted if there were no loss due to scattering or absorption.

Such logarithmic compensation of measured counts makes density differences appear proportional to intensity differences in x-ray images.

2.2.2 Poisson noise distribution

The number of photons $K'_{i'}$ counted per unit time by an x-ray detector can be modeled as a Poisson random variable (Kak and Slaney, 1988). Equivalently, the time $T'_{i'}$ between the arrival of photons at the detector is exponentially distributed with a mean of $1/\alpha'_{i'}$. The conditional probabilities of the events $K'_{i'} = k'_{i'}$ and $T'_{i'} = t'_{i'}$ are therefore

$$\begin{aligned} P(K'_{i'} = k'_{i'} | f) &= \frac{\alpha'^{k'_{i'}}_{i'}}{k'_{i'}!} e^{-\alpha'_{i'}} \qquad k'_{i'} = 0, 1, \ldots, \quad \alpha'_{i'} > 0 \\ P(T'_{i'} = t'_{i'} | f) &= \alpha'_{i'} e^{-\alpha'_{i'} t'_{i'}} \qquad t'_{i'} \geq 0 \end{aligned}$$

where f represents the three-dimensional density distribution (i.e. the object) that is x-rayed. It can be shown that the means and variances of the random variables $K'_{i'}$ and $T'_{i'}$ are given by

$$\begin{aligned} E[K'_{i'}] &= \alpha'_{i'} \qquad \text{VAR}[K'_{i'}] = \alpha'_{i'} \\ E[T'_{i'}] &= \frac{1}{\alpha'_{i'}} \qquad \text{VAR}[T'_{i'}] = \frac{1}{\alpha'^2_{i'}}. \end{aligned}$$

Due to linear time integration to increase the signal to noise ratio and binning to reduce image resolution, detector measurements are combined additively in the Lodox system. This operation

can be written as

$$k_i = \sum_{i'=(i-1)N_{\text{bin}}+1}^{iN_{\text{bin}}} k'_{i'}. \tag{2.28}$$

Assuming that each of the N_{bin} detector readings which are combined into the ith bin are independent, the resulting probability distribution $P(K_i)$ remains Poisson:

$$P(K_i = k_i|f) = \frac{\alpha_i^{k_i}}{k_i!}e^{-\alpha_i} \quad \text{where} \quad \alpha_i = \sum_{i'=(i-1)N_{\text{bin}}+1}^{iN_{\text{bin}}} \alpha'_{i'}. \tag{2.29}$$

This can easily be proven by considering that the sum of independent random variables correspond to the multiplication of their *probability generating functions* (Leon-Garcia, 1994), i.e.

$$G_{K_i}(z) = \prod_{i'=(i-1)N_{\text{bin}}+1}^{iN_{\text{bin}}} G_{K'_{i'}}(z) = \prod_{i'=(i-1)N_{\text{bin}}+1}^{iN_{\text{bin}}} e^{\alpha'_{i'}(z-1)} = e^{\sum_{i'=(i-1)N_{\text{bin}}+1}^{iN_{\text{bin}}} \alpha'_{i'}(z-1)} = e^{\alpha_i(z-1)} \tag{2.30}$$

where the probability generating function for the Poisson random variable χ with mean μ is

$$G_\chi(z) = E[z^\chi] = e^{\mu(z-1)}. \tag{2.31}$$

This statistical model neglects the effect of additive electrical noise as well as the influence of any remaining artefacts or systematic inaccuracies that occur in imaging systems.

2.3 Maximum likelihood solutions

The following sections describe how tomographic reconstruction can be stated as the solution to an optimization problem. This section starts off by modeling the measurement noise as a probability distribution and discusses how unbiased solutions can be found to inverse problems.

2.3.1 Likelihood function

Assuming that the binned measurements $k = \{k_1, k_2, ..., k_i, ...k_N\}$ can be modeled as N independent Poisson random variables $K = \{K_1, K_2, ..., K_i, ..., K_N\}$, the conditional probability of the joint event $K = k$, i.e. $\{K_1 = k_1, ..., K_N = k_N\}$, is given by

$$P(K|f) = \prod_{i=1}^{N} P(K_i|f) = \prod_{i=1}^{N} \frac{\alpha_i^{k_i}}{k_i!}e^{-\alpha_i} \tag{2.32}$$

where the expected projection data vector $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_i, ...\alpha_N\}$ is calculated from the expected object $f(x, y)$ using

$$\alpha = \mathbf{H}f \quad \text{or} \quad \alpha_i = \sum_{j=1}^{M} \mathbf{H}_{i,j} f_j \tag{2.33}$$

and $\mathbf{H}$ is a matrix that represents the projection process while $f = \{f_1, f_2, ..., f_j, ..f_M\}$ is a vector representation of $f(x, y)$.

The probability density function in Equation 2.32 is commonly known as the *likelihood function* and is conditioned on the object which remains fixed throughout the imaging process. The log-likelihood can be rewritten as follows:

$$\begin{aligned} \ln(P(K|f)) &= \ln \prod_{i=1}^{N} \frac{\alpha_i^{k_i}}{k_i!} e^{-\alpha_i} = \sum_{i=1}^{N} \ln \frac{\alpha_i^{k_i}}{k_i!} e^{-\alpha_i} = \sum_{i=1}^{N} \ln \alpha_i^{k_i} - \ln k_i! + \ln e^{-\alpha} \\ &= \sum_{i=1}^{N} k_i \ln \alpha_i - \alpha_i - \ln k_i! \approx \sum_{i=1}^{N} k_i \ln \alpha_i - \alpha_i - (k_i \ln k_i - k_i) \\ &= \sum_{i=1}^{N} k_i \ln \alpha_i - \alpha_i - k_i \ln k_i + k_i \end{aligned} \tag{2.34}$$

where Stirling's approximation $\ln x! \approx x \ln x - x$ is used.

The maximum likelihood estimate $f^{(\mathrm{ML})}$ of the object corresponds to the peak of the distribution $P(K|f)$ and also $\ln P(K|f)$ since the logarithm is a monotonic function, so

$$f^{(\mathrm{ML})} = \arg_f \max P(K|f) = \arg_f \max \ln P(K|f). \tag{2.35}$$

This optimum is achieved at

$$\frac{\partial \ln P(K|f)}{\partial k} = 0 \tag{2.36}$$

which occurs when

$$\begin{aligned} \frac{\partial \ln P(K_i|f)}{\partial k_i} &= \frac{\partial}{\partial k_i} (k_i \ln \alpha_i - \alpha_i - k_i \ln k_i + k_i) = 0 \\ &= \ln \alpha_i - (1 + \ln k_i) + 1 = \ln \alpha_i - \ln k_i \\ \therefore \quad \alpha_i &= k_i \end{aligned} \tag{2.37}$$

Therefore the maximum likelihood estimator is unbiased.

2.3.2 Solution by gradient based search methods

The solution to Equation 2.35 can be obtained in several ways. The steepest and conjugate gradient ascent algorithms require the gradient $g \equiv \nabla_f \ln P(K|f)$ which can be calculated as

$$\frac{\partial}{\partial f_j} \ln P(K|f) = \frac{\partial}{\partial f_j} \sum_{i=1}^{N} k_i \ln \mathbf{H}_{i,j} f_j - \mathbf{H}_{i,j} f_j - k_i \ln k_i + k_i \tag{2.38}$$

$$= \sum_{i=1}^{N} k_i \frac{1}{\mathbf{H}_{i,j} f_j} \mathbf{H}_{i,j} - \mathbf{H}_{i,j} \tag{2.39}$$

$$= \sum_{i=1}^{N} \mathbf{H}_{i,j} \left(\frac{k_i}{\mathbf{H}_{i,j} f_j} - 1 \right) \tag{2.40}$$

$$\therefore \nabla_f \ln P(K|f) = \mathbf{H}^{\mathsf{T}} \left((k \div (\mathbf{H}f)) - 1 \right) \tag{2.41}$$

where $\div$ represents elementwise division for vectors. In other words, the calculation of the gradient requires both forward and back projection operations.

The steepest ascent method updates an estimate $f^{(n)}$ by performing a line search in the direction of the gradient $g^{(n)} \equiv \nabla_{f^{(n)}} \ln P(K|f)$, so

$$f^{(n+1)} = f^{(n)} + \lambda^{(n)} g^{(n)} \tag{2.42}$$

where each step length $\lambda^{(n)}$ is determined using one-dimensional search methods such that $f^{(n+1)}$ maximizes the objective function with respect to $\lambda^{(n)}$.

Faster convergence can be achieved using the Conjugate Gradient method which is well explained in the literature (Press *et al.*, 1999) in the context of minimization. Instead of searching in the gradient direction $g^{(n)}$, only that component $h^{(n)}$ of this gradient direction which is conjugate to all the previous search directions $(h^{(r)}, r = 1, 2, ..., n-1)$ is used. This conjugacy condition can be expressed as

$$h^{(n)\mathsf{T}} \mathbf{A} h^{(r)} = 0 \tag{2.43}$$

where $\mathbf{A} = \nabla g^{(n)}$ is a Hessian matrix. In addition, the orthogonality conditions

$$g^{(n)\mathsf{T}} g^{(r)} = 0 \quad \text{and} \quad g^{(n)\mathsf{T}} h^{(r)} = 0 \tag{2.44}$$

are also satisfied when the search direction is updated as

$$h^{(n)} = -g^{(n)} + \frac{\left|g^{(n)}\right|^2}{\left|g^{(n-1)}\right|^2} h^{(n-1)} \tag{2.45}$$

which conveniently does not involve the explicit calculation of the Hessian.

2.3.3 EM algorithm

It was found that the line searches in the Conjugate Gradient method are quite time consuming due to the computational demand of the Radon transform. Significant performance improvements can be achieved if no line search is performed.

It is shown in (Rangarajan *et al.*, 1999) by differentiating a likelihood energy function that an iterative update equation can be expressed as

$$f_j^{(n+1)} = \frac{\sum_i k_i \frac{\mathbf{H}_{i,j} f_j^{(n)}}{\sum_{j'} \mathbf{H}_{i,j'} f_{j'}^{(n)}}}{\sum_i \mathbf{H}_{i,j}} \tag{2.46}$$

which maximizes the Poisson likelihood. Such an expectation maximization step involves only one forward and back projection operation.

Equation 2.46 requires that the projection operations have a gain of less than or equal to unity for stability.

2.3.4 χ^2 errors

Many authors, including (Vengrinovich *et al.*, 1998), model x-ray noise with a Gaussian distribution. This is reasonable when the measured counts are logarithmically compensated and the signal to noise ratio is high.

Assuming Gaussian errors, the likelihood function becomes

$$P(p|f) = \prod_{i=1}^{N} P(p_i|f) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\chi^2} \tag{2.47}$$

where p is a vector of logarithmic compensated projection data and the χ^2 error is

$$\chi^2 = \sum_{i=1}^{N} \frac{(p_i - \mu_i)^2}{\sigma_i^2}. \tag{2.48}$$

In Equation 2.48, $\mu_i = p_i^{(n)}$ is the corresponding projections of the object n-th estimate $f^{(n)}$.

Similar to Equation 2.35, the Gaussian maximum likelihood (ML) estimate can be expressed as the optimization problem

$$f^{(\mathrm{ML})} = \arg_f \max P(p|f) = \arg_f \max \ln P(p|f) = \arg_f \min \chi^2. \tag{2.49}$$

A benefit of the χ^2 representation of projection errors is that the objective function is a quadratic. Gradient based search methods, such as the Conjugate Gradient method, assume that the objective function is quadratic and can consequently converge to the optimum in fewer iterations.

2.4 Bayes' Rule

A maximum likelihood estimate maximizes the probability that the data equals the (estimated object's) projections given the object. Bayesian methods argue that it is more sensible to maximize the probability that the object equals the object estimate given the data. This posterior distribution can be calculated using Bayes' formula

$$P(f|p) = \frac{P(f)P(p|f)}{P(p)} \tag{2.50}$$

where the probability of the data $P(p)$ is a constant and normalizes the area of $P(f|p)$. The prior distribution $P(f)$ expresses the *a priori* knowledge of the object.

In its simplest form, the prior could be a uniform distribution of infinite extent. Under this condition $P(f) = P(p)$ and Bayes' rule simplifies so that the maximum *a posteriori* (MAP) estimate

$$f^{(\mathrm{MAP})} = \arg_f \max P(f|p) \tag{2.51}$$

equals the maximum likelihood estimate $f^{(\mathrm{ML})}$.

However, it is known that the density of the object and air is non-negative. It may also be assumed that no density greater than $f_{\max}$ could occur within the object. This prior knowledge can be expressed as

$$P(f) = \prod_{j=1}^{M} P(f_j) \tag{2.52}$$

where

$$P(f_j) = \begin{cases} \frac{1}{f_{\max}} & 0 \leq f_j \leq f_{\max} \\ 0 & \text{otherwise} \end{cases} . \tag{2.53}$$

More sophisticated forms of prior knowledge are discussed in the following section.

2.5 Gibbs priors

In various applications of image restoration (and reconstruction), the target image (or volume) is believed to exhibit certain properties that can be characterized by local site (i.e. pixel or voxel) interactions. Gibbs distributions can be used to model the contextual constraints that sites must satisfy. There are many applications for Gibbs distributions including texture modeling and region labeling (i.e. segmentation). In this section, only low-level Markov random field models expressed as Gibbs distributions will be discussed. A more elaborate treatment of the subject can be found in (Li, 1995).

2.5.1 Definitions

Some notation needs to be defined first. An image (or volume) is denoted $f = \{f_j | j \in \mathcal{S}\}$ where $\mathcal{S} = \{1, 2, .., j, .., M\}$ is a set of sites. A pixel on an image lattice can be indexed by $j \in \mathcal{S}$ and takes on a value f_j in the real interval $\mathcal{L} = [0, f_{\max}]$. In discrete labeling problems, $\mathcal{L}$ is a label set instead, and f is a segmented map.

The spatial relationship of the sites is determined by a neighborhood system $\mathcal{N} = \{\mathcal{N}_j | j \in \mathcal{S}\}$ where $\mathcal{N}_j$ is the set of sites which are regarded as being neighbours to j. A clique c is a single site $c = \{j\}$ or a set of neighbouring sites $c = \{j, j', j'', ...\}$. Very often only pair cliques $c = \{j, j'\}$ are considered.

According to the Markov-Gibbs equivalence (Li, 1995), a Markov random field is a Gibbs distribution

$$P(f) = \frac{1}{Z} e^{-\frac{1}{T}U(f)} \tag{2.54}$$

where the partition function

$$Z = \sum_{\forall f} e^{-\frac{1}{T}U(f)} \tag{2.55}$$

is a normalizing constant. T is a constant referred to as the temperature, while the energy

$$U(f) = \sum_{c \in C} V_c(f) \tag{2.56}$$

is the sum of clique potentials $V_c(f)$ over the set of all possible cliques C. The value of a particular clique potential $V_c(f)$ depends on the local configuration of the clique c.

A set of random variables $F = \{F_1, .., F_M\}$ is called a Gibbs random field on S with respect to $\mathcal{N}$ if its configuration satisfies a Gibbs distribution. Each of these random variables F_j can take on a value $f_j \in \mathcal{L}$.

A Gibbs random field is homogeneous or isotropic if the value of $V_c(f)$ is independent of the position or orientation respectively, of the clique c with respect to S.

2.5.2 The role of energy functions

(Geman and Geman, 1984) draw the analogy between images and statistical mechanical systems. They explain that the maximum *a posteriori* image corresponds to a low energy state of the clique potentials and suggest that this solution be obtained by stochastic relaxation.

Gibbs energy functions are formulated in such a way that a minimum clique potential will be achieved when every clique satisfies the image properties given by the prior knowledge.

(Li, 1995) argues that the role of energy functions are not merely to provide a quantitative cost measure of the image estimate but also to guide the search for the optimal solution. This is essential since the formulation of the objective function in Equation 2.51 is often multimodal resulting in a very difficult optimization problem.

2.5.3 Clique potentials

A simple example of a clique potential represents the prior knowledge of image smoothness,

$$V_c(f) = \frac{1}{2}(f_j - f_{j'})^2 \tag{2.57}$$

where the set of all cliques $\mathcal{C}$ includes only the pair cliques $c = \{j, j'\}$ for which j is an index to the pixel at (x, y) and j' is an index to either of the pixels $(x, y-1)$ or $(x-1, y)$ for all x and y in the image.

Few real life images have this property. Instead, a more popular assumption is that images can be represented by smooth regions separated by discontinuities.

(Li, 1995) discusses the development of the discontinuity adaptive Markov random field model that can represent this image property. An adaptive potential function $g_\gamma(\eta) \equiv V_c(f)$ is defined in terms of an adaptive interaction function $h_\gamma(\eta)$, as

$$g_\gamma(\eta) = 2\int_0^\eta \eta' h_\gamma(\eta') d\eta' \tag{2.58}$$

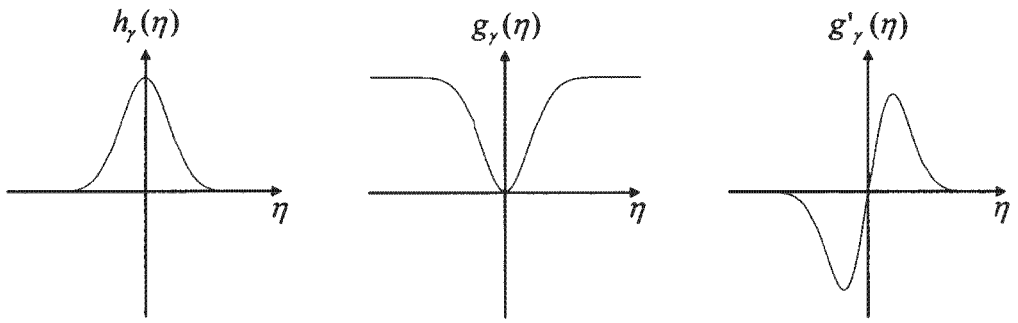


Figure 2.3 Discontinuity adaptive functions.

for which

$$\frac{\partial g_\gamma(\eta)}{\partial \eta} \equiv g'_\gamma(\eta) = 2\eta h_\gamma(\eta) \tag{2.59}$$

and η is an intensity gradient or, in practice, the intensity difference of a pair clique, i.e. $\eta = f_j - f_{j'}$. The adaptive interaction function satisfies the following conditions:

1) $h_\gamma(\eta) = h_\gamma(-\eta)$

2) $h_\gamma(\eta) > 0$

3) $h'_\gamma(\eta) < 0 \quad (\forall \eta > 0)$

4) $\lim_{\eta \to \infty} |\eta h_\gamma(\eta)| = C < \infty$

One example of such discontinuity adaptive functions is illustrated in Figure 2.3 and given by

$$h_\gamma(\eta) = e^{-\frac{\eta^2}{\gamma}}, \quad g_\gamma = -\gamma e^{-\frac{\eta^2}{\gamma}} \tag{2.60}$$

where γ determines the width of $h_\gamma(\eta)$.

The potential function $g_\gamma(\eta)$ has a minimum for a smooth image. Small edges (quantified by γ) are penalized nearly quadratically, however, in comparison to Equation 2.57, the penalty for large edges is much less severe. The result is that this prior potential represents piecewise smoothness.

Another possibility is proposed in the context of tomography. (Vengrinovich *et al.*, 1998) defines a clique potential

$$V_j(\mu) = \sum_{a=1}^{26} \left(\mu'_j - \mu'_{j,a}\right)^2 + \beta {\mu'_j}^2 (\mu'_j - 1)^2, \quad \mu'_j = \frac{\mu_j - m_1}{m_2 - m_1} \tag{2.61}$$

where $\mu \equiv f$ while m_1 and m_2 are two levels of attenuation coefficient in a binary object that is tomographically reconstructed from a limited number of projections. This potential reflects the prior knowledge that the smoothness of the reconstruction should be balanced (using hyper parameter β) by the closeness of the attenuation coefficients to either of two possible attenuation levels.

2.5.4 Maximizing the posterior distribution

Assuming Gaussian noise and a Gibbs distribution prior model, the maximum *a posteriori* solution given by Equation 2.51 can be expressed in logarithmic form and simplified so that

$$f^{(\mathrm{MAP})} = \arg_f \max \ln P(f|p) = \arg_f \min \left[U(f) + \lambda\chi^2\right] \tag{2.62}$$

where λ is a constant. A local optimum of this energy expression can be obtained using a gradient based method. However, $U(f)$ is very often multimodal. (Press *et al.*, 1999) and (Li, 1995) discuss many search methods for local and global optimization.

Although stochastic methods such as simulated annealing and genetic algorithms have a better chance of attaining the true global optimum, the deterministic annealing method of Graduated Non-Convexity (GNC) reportedly finds good solutions at a significantly lower computational expense. The dimensionality of most realistic problems necessitates the use of efficient algorithms.

GNC optimizes an objective function repeatedly while changing the objective function slightly at every major iteration. Initially a hyperparameter, such as γ in the discontinuity adaptive model, is set to a value so that the energy function is essentially convex. After the optimum for this parameter setting is obtained, which is hopefully reasonably close to the desired solution, the parameter value is adjusted so that the energy function becomes less convex. The old estimate is used as the starting point of minimization for the next iteration. In this way GNC attempts to avoid local optima.

2.6 Summary

This chapter introduced background theory. The text continues in a more focussed manner by addressing a series of relevant problems to reach its goal in limited angle tomography. A literature survey of limited angle tomography methods specifically is presented in Section 6.3.

CHAPTER 3

Image formation process

The way in which x-rays travel through the patient onto the detector defines the relationship between the digital image data and the tomographic reconstruction. It is important to know the machine geometry and imaging conditions very precisely to ensure accurate reconstructions and eliminate avoidable artefacts.

An introduction to the scanning hardware is provided in the first section. This is followed by a study of distortions observed in Lodox images. Section 3.3 presents an algorithm that removes camera overlap artefacts. Once the image plane is calibrated, the dataset itself is used to determine the three-dimensional imaging parameters of the system. The final sections of this chapter show how direct tomographic reconstructions are achieved.

3.1 Data acquisition using the Lodox system

The Lodox system used to acquire the dataset was originally developed by Debtech. A brief overview of their technology is given in this section since it provides important background. The conditions under which the machine was used to obtain the dataset are also stated.

3.1.1 Overview

In the Lodox system (refer back to Figure 1.2) both the x-ray tube and detector are attached to a C-arm. The C-arm can rotate and move around a patient that lies on a trolley. Scanning is always done in one direction only at a fixed C-arm angle. The setup resembles that of a third-generation fan beam tomography machine (Kak and Slaney, 1988). Although the 1999

model installed in the Groote Schuur Hospital (which was used to acquire the dataset) can only scan over a 90° range, newer models can scan up to 100°. Mechanical stress on the C-arm bearings limits this angular range in the current design.

3.1.2 Image formation

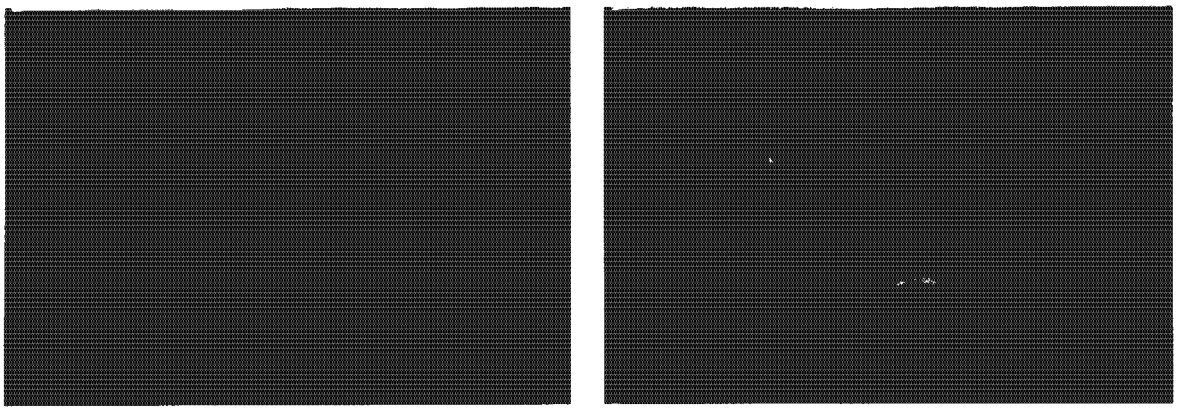
Once the x-rays leave the tube, they are filtered by a 1mm aluminium sheet and collimated by a narrow gap between tungsten plates before passing as a fan beam through the scanned object perpendicular to the scanning direction. At the detector, post-collimation is performed with another pair of closely spaced tungsten plates and the x-rays are converted into visible light by a paper scintillator. At this point an array of fibre-optic tapers demagnify the light rays to 12 different CCD cameras. The tapers are shaped like paralellograms and are tightly spaced so that light is projected redundantly onto adjacent cameras where the tapers overlap. These cameras operate in a linear-time integrating mode in the scan direction while the linear motor drives the C-arm so that rows of the image are produced as the C-arm moves.

The intensity offset, also known as the *dark current* (Beutel *et al.*, 2000), is measured before each scan and is subtracted from the CCD outputs before gain compensation is performed to make the CCD readings uniform. The outputs of the different cameras are combined and overlapped according to the redundancy in the data (which is physically related to the overlap of the fibre-optic tapers and the relative positioning of the tapers on the cameras). Finally these values are binned and logarithmically scaled. The Handbook of Medical Imaging (Beutel *et al.*, 2000) discusses the physics of x-ray imaging in great detail.

The camera gain curves, amount of overlap and positioning parameters are estimated or measured beforehand when the machine is calibrated. Section 3.2 discusses camera alignment and overlaps in more depth.

3.1.3 Operator settings and dataset

There are a number of operator settings. The x-ray tube can function at several different voltage and current settings. Scanning at half and quarter speeds increases exposure. Binning of the fundamental 60μm pixels into 1x1, 2x2, 3x3 and 5x5 bins provides different image resolutions. Other settings such as the C-arm angle, scan start and stop positions, and collimator slit width can be set by the operator using software in newer models of the Lodox system.



(a) Scan at 0°.

(b) Scan at 90°.

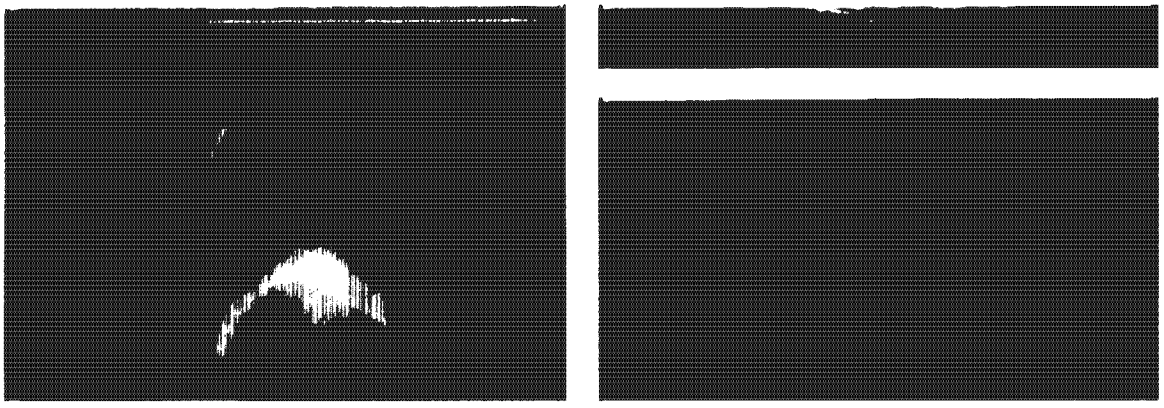
Figure 3.1 Lodox scans that are part of the dataset. Note that the table top (foam with carbon fiber coating) of the trolley is visible in (b). Also note that fan beam magnification occurs in the horizontal direction.

Unless otherwise stated, all the x-ray images used in this project were taken using the typical settings of 100kV 100mA, at fullspeed and 5x5 binning resolution. Each scan image size is approximately 8MB having 1164x2294 pixels of at least 8000 grey intensity levels. Since the C-arm of the Lodox scanner can only be adjusted from 0 to 90°, and at least 180° is needed for full tomographic reconstruction, the trolley was flipped around (reversed in the scan direction) and the object scene was scanned from the other side, totalling 182 scans over a 180° range. Figure 3.1 shows two of the scans in the dataset.

3.1.4 Reliability

Although the Lodox system is constantly being improved and debugged, it behaved somewhat unreliably at the time the dataset was acquired. Out of 200 scans, 18 were considered to be faulty or of too poor quality to be used and were rescanned. See Figure 3.2 for examples of such scans.

Critically speaking, for the early Lodox machines no image produced is perfect. For example, within up to a 100 pixel margin of all the images there appear to be gain compensation problems (see the bright vertical line on the left hand side in Figure 3.1). Also, scanning starts at an unpredictable position. These trivial problems are side-stepped by cropping the regions of interest using markers as reference points.



(a) Saturated scan with stripes.

(b) Truncated and blank scans.

Figure 3.2 Faulty Lodox scans.

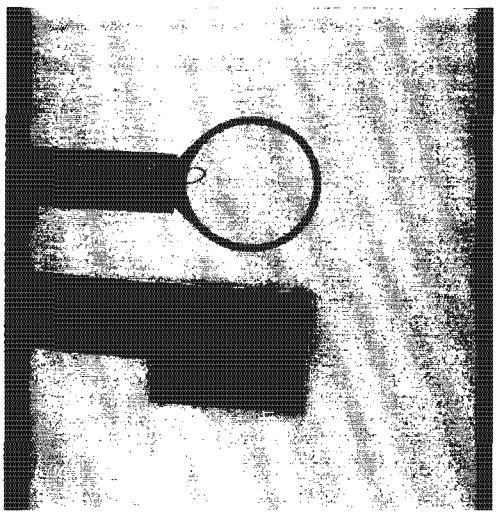
More subtle yet significant distortions in the images are studied in the next two sections. Image correction algorithms were developed in this thesis and have since been integrated into Lodox. Also, Lodox Development has recently improved the effectiveness of their detector significantly by slowing down the readout speeds of the CCDs.

3.1.5 Image quality

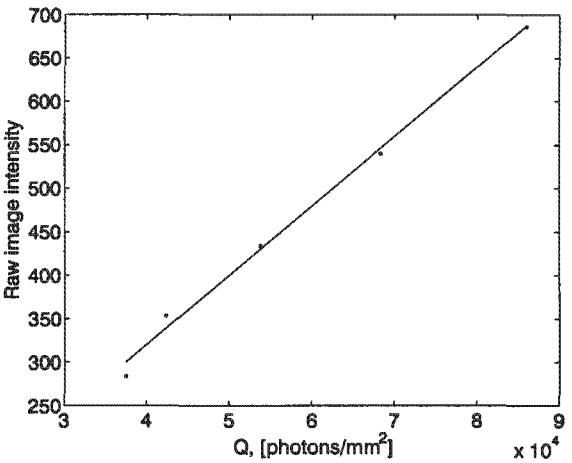
In x-ray systems image quality improves with dose because noise increases linearly with the square of the signal magnitude for Poisson-distributed x-ray quanta (Beutel *et al.*, 2000). Achieving good image quality using a low dose is desired and this property is commonly quantified by means of the DQE, or *detective quantum efficiency*, defined in (IEC 62220-1. Ed 1, 2003). A better DQE also implies that the high voltage generator and x-ray tube need not be heat loaded as much while still producing good images.

The DQE of the Lodox system is measured for the beam quality RQA5 as specified by the International Electrotechnical Commission (IEC 62220-1) and compared to other x-ray imaging systems. Additional information pertaining to the image quality of the Lodox Statscan system can be found in (de Villiers and de Jager, 2003) and (Sheelke *et al.*, 2005).

The detective quantum efficiency describes how much of the signal to noise ratio of the incident radiation is sustained in the resultant digital image. This two-dimensional function of spatial



(a) Raw image of tungsten plate and dosimeter for DQE measurement.



(b) Characteristic curve. Mean image intensities are shown for 5 different exposures by varying the current in the tube.

Figure 3.3 Image quality measurements.

frequency is defined as

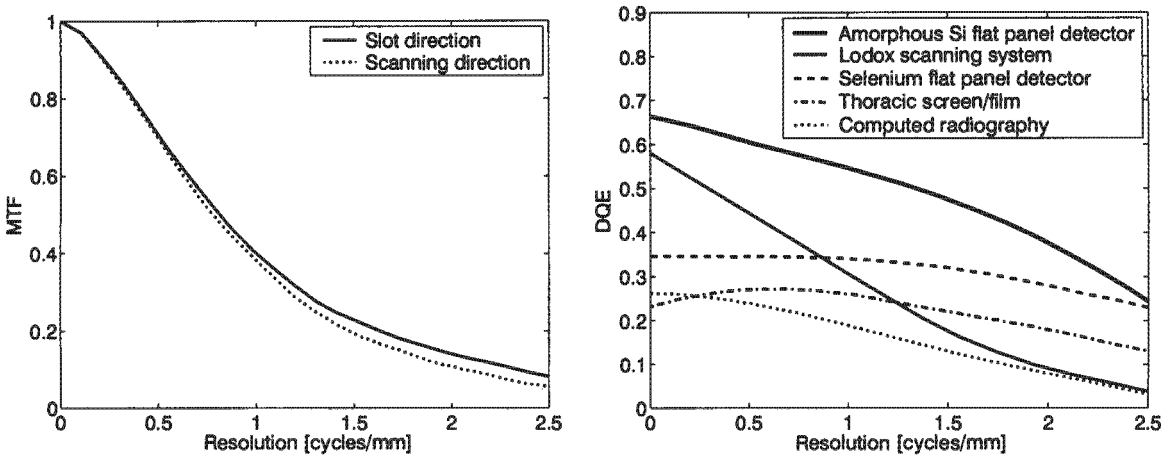
$$DQE(u,v) = \frac{W_{\text{in}}(u,v)}{W_{\text{out,referred to input}}(u,v)} = G^2 MTF^2(u,v)\frac{W_{\text{in}}(u,v)}{W_{\text{out}}(u,v)} \tag{3.1}$$

where $W_{\text{in}}(u,v)$ is the *input noise power spectrum*, $W_{\text{out}}(u,v)$ is the *output noise power spectrum* and G is the *incremental gain*. MTF(u,v) is the *modulation transfer function* which is only measured in the scanning direction and the direction perpendicular to scanning (slot direction), and is therefore actually two one-dimensional functions as described later (hence the parameters u and v are neglected below). This equation can be restated as

$$DQE = G^2 MTF^2 \frac{Q}{W_{\text{out}}(u,v)} \tag{3.2}$$

where Q is the *quantum density* (photons/mm^2 at the detector). The technique for measuring the DQE is well explained in (Stierstorfer and Spahn, 1999). This section gives a brief overview of the measurement process and reports on results. Figure 3.3(a) shows a scan of a dosimeter and a tungsten plate suspended slightly above the detector.

Firstly the incremental gain is determined from a number of exposures by varying the current in the tube. The dose at the detector is converted from μGray to photons/mm^2 using a conversion factor supplied by the IEC for the given beam quality. The slope of a straight line fitted through the plot of quantum density versus average raw pixel values yields G. See Figure 3.3(b).



(a) Modulation transfer function.

(b) Detective quantum efficiency.

Figure 3.4 Image quality results.

The presampled horizontal and vertical MTFs are measured from the scan of a slightly angled tungsten plate with accurately milled square edges as in (Samei *et al.*, 1998). Pixel intensities are projected parallel to the edge direction to produce an edge spread function. The numerical derivative of the edge spread function is the line spread function, and the magnitude of the Fourier transform thereof gives the MTF shown in Figure 3.4(a). A slightly poorer MTF is noted in the scanning direction, especially at faster scanning speeds. This is ascribed to scintillator afterglow.

The noise power spectrum $W_{\text{out}}(u, v)$ at the output is determined by averaging the power spectra of several partly overlapping windows of pixels over a uniformly exposed area. One-dimensional horizontal and vertical cuts through the noise power spectrum are obtained by averaging 15 rows or columns of the two-dimensional spectrum around each axis, while maintaining exact spatial frequencies in terms of radial distance from the origin. The axis itself is omitted. See (IEC 62220-1. Ed 1, 2003) for more details.

The DQE is calculated from Equation 3.2 and the average is displayed in Figure 3.4(b) along with DQEs of a non-scanning systems presented in (Granfors and Aufrichtig, 2000). At low spatial resolutions, the Lodox system performs very well. This is well suited for tomography applications which require relatively low spatial resolution x-ray data.

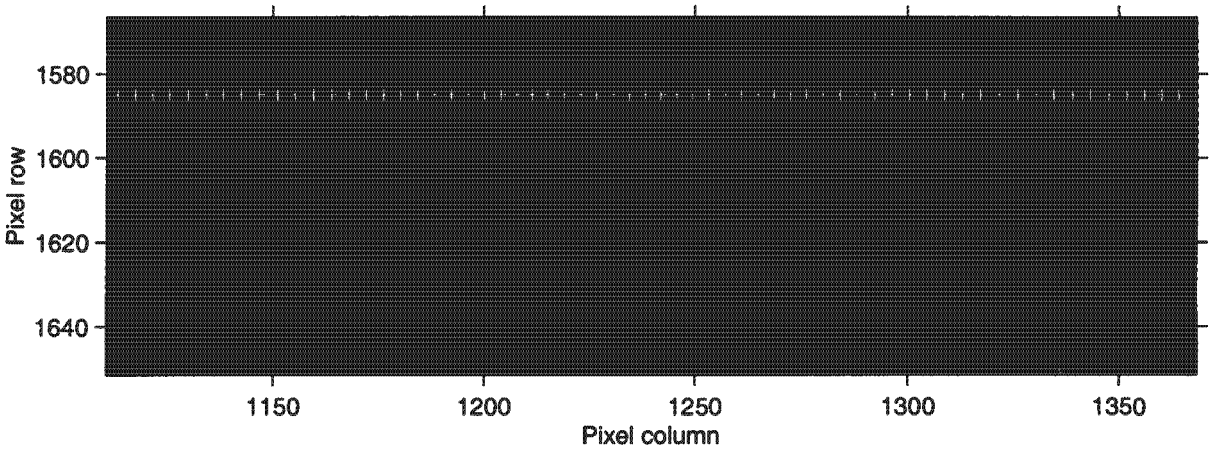


Figure 3.5 Closeup view of a steel ruler with detected tick marks (+) and edge position (solid white line) superimposed.

3.2 Camera distortion

Figure 3.5 shows a closeup view of a steel ruler taken with the Lodox machine. In this image one would expect to see a straight line along the edge of the ruler and also expect to see evenly spaced tick marks. However, there appears to be distortion in both these directions, and at higher resolutions these distortions are more apparent. The cause of the distortions is believed to arise from barrel distortion in the fibreoptic tapers and other inaccuracies in their alignment and orientation. Quantitative knowledge of these distortions is necessary for making image corrections. This is especially important because the geometry of the machine is inferred to a high degree of accuracy in Section 3.5 using sparse data taken from the images.

The overlap artefacts, which can clearly be seen in this figure near pixel columns 1142 and 1334, are investigated in Section 3.3.

3.2.1 Pixel column position errors

As demonstrated in Figure 3.6, the pixel column positions of the tickmarks on the ruler image are detected with subpixel accuracy by fitting sections of sinusoids to the intensity profile of these markings. Firstly, 8 rows of the image are averaged to improve the signal to noise ratio of the markings. The least squares fit performed for each tick mark determines only the intensity

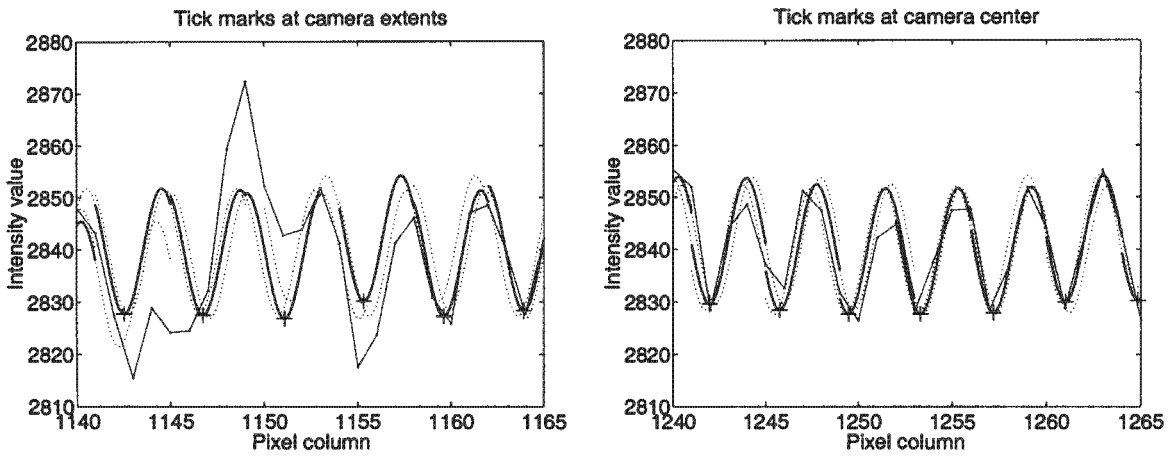


Figure 3.6 Tick marks (+) detected near edge of camera (left) and near center of camera (right). The metal ruler intensity profile (thin solid line) and the sinusoids fitted thereto (dark solid lines extending into dotted lines) are shown.

offset and phase of the localized sinusoid. The average amplitude and frequency is used for all tick marks and the region of interest size is chosen to span 2 tick mark cycles.

The further the tick marks lie from the center of any camera, the further apart they are spaced from one another. Explicit examples of this phenomenon are given in Figure 3.6 while overall results are shown in Figure 3.7.

From these results the tick marks appear to be on average 4 pixels apart, but vary between 3.5 and 4.5 pixels. This average value is bigger than the expected 3.33 pixels per millimeter for 5x5 binning of 60μm pixels due to magnification by the fan beam geometry. The ruler was on the trolley almost 30cm above the detector. Furthermore, absolute position errors of the pixel columns are nearly 3 pixels (or 0.75mm) while relative position errors could be up to 6 pixels in extreme cases.

3.2.2 Pixel row position errors

The edge of the ruler, superimposed at a slight offset in Figure 3.5, is a mean intensity contour of the ruler image. Figure 3.7 shows that the row position error remains mostly below 1 pixel (0.33mm) but can be incorrect by up to 2 pixels. These values are somewhat less than the

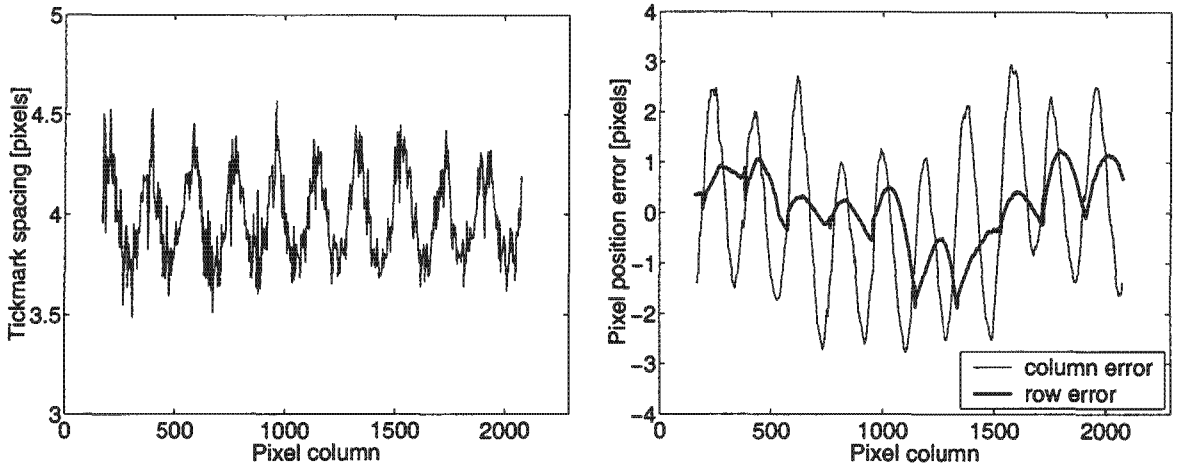


Figure 3.7 Tick mark spacings (left) and pixel position errors (right).

column position errors because there is no fan beam magnification in the scanning direction, and the time delay integration of pixel columns in the CCD also removes some distortions.

Further software developments have since successfully been integrated into the Lodox system during the course of 2004. These algorithms are designed to operate on raw images, remove distortions and perform camera overlapping automatically without any supervision.

3.3 Camera overlap artefacts

The most striking shortcoming of Lodox images is the presence of camera overlap artefacts. These artefacts obscure image details and degrade image quality, but do not destroy underlying image detail. They are believed to exist due to a combination of factors including the collimator slot width, inaccurate manual camera overlapping and variation of the CCD intensity offset, or *dark current*, during a scan.

A lot of research effort has been put into developing a robust standalone algorithm that removes these artefacts well for different Lodox machines and all imaging procedure factors (de Villiers and de Jager, 2001). Two algorithms are discussed in this section namely an intensity level algorithm and an adaptive algorithm. The first algorithm is included not only because it gives some insight into the artefact problem, but because the concept that it presents is fundamental

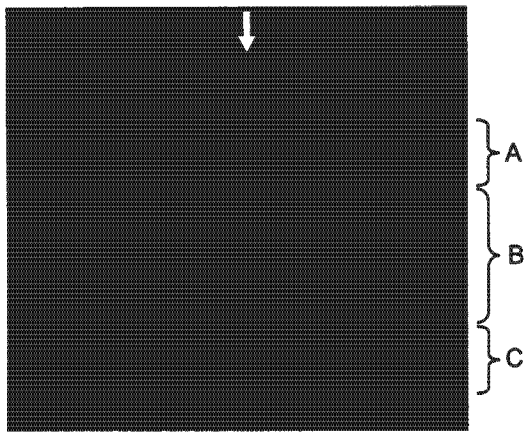


Figure 3.8 High resolution view of one metal bar.

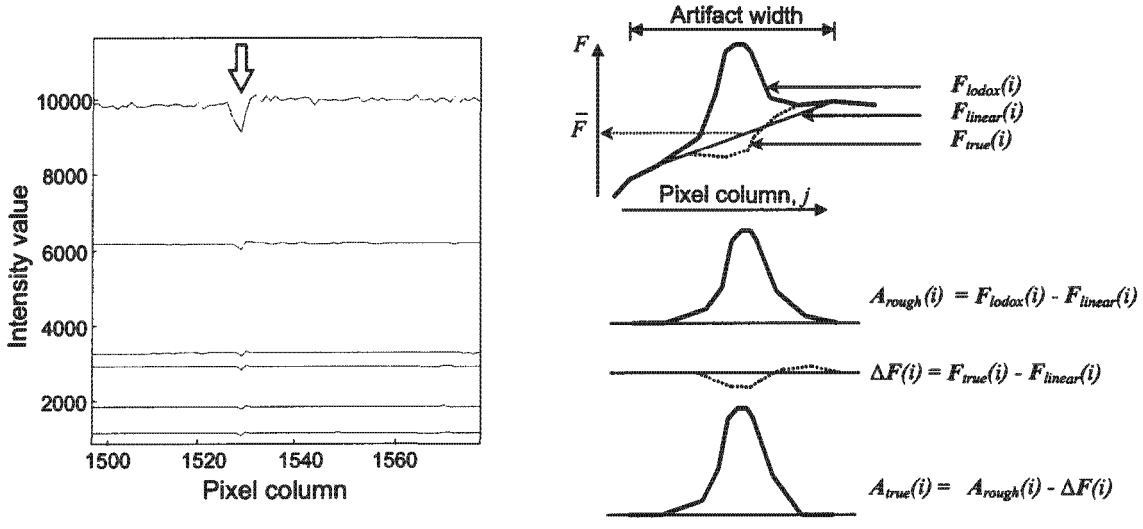
to later chapters in this thesis. Attention is therefore drawn to the understanding of the workings of this algorithm rather than the latter adaptive algorithm.

3.3.1 Artefact properties

Figure 3.8 shows part of a high resolution Lodox image of a metal bar. The artefact, marked with an arrow, affects only a few pixel columns where overlap from one camera to the next occurs. These column indices are the same for all scans and can be identified *a priori* because they are determined by the physical setup of the scanner. The shape of the artefact is significantly different in the three indicated regions. Regions A and C show transient artefacts near the edges of the metal bar while region B shows a steady state artefact.

Image intensity profiles of a Lodox image of six different metal bars are shown in Figure 3.9(a). Since each metal bar is of uniform density, constant valued intensity profiles are expected to match the corresponding x-ray attenuation level for each bar. However, artefacts can be seen that change in shape and size as a function of intensity.

A slight difference in mean intensity for different cameras (on either side of the artefact) is also noticeable in the imaging of high attenuation objects.



(a) Intensity profiles of different metal bars.

(b) Definitions of terms in level algorithm.

Figure 3.9 Artefact properties and definitions.

3.3.2 Intensity level based algorithm

The intensity level based method assumes that the artefact shape is dependent only on the local mean intensity level. Consequently this method fails to remove the artefacts in places where there are abrupt, large changes in image intensity. Nevertheless, a good estimate of the underlying image can be obtained with this algorithm.

Consider the simplest example where a scan is taken of a uniform object resulting in a region of nearly constant intensity (consider Region B only of Figure 3.8). Regard this region as a sequence of row vectors, $F_{lodox}(i)$, where i is the row index. The artefact, $A_{true}(i)$, remains exactly the same for each row. Note that the image detail (labeled $\Delta F(i)$ which in this case is quantum and measurement noise only) changes randomly from one row to the next but the superimposed artefact does not. In this region, an estimate of the artefact (an artefact row vector, $A_{true}(i) \approx < A_{rough}(i) >$) can be obtained by first averaging (in the column direction) all the row vectors comprising the region and then subtracting from it the scalar average intensity of the region. The underlying image with its random fluctuations can be recovered by subtracting the estimated artefact from each original row vector in this region.

In the more general case described below, see Figure 3.9(b) for a visual representation of various terms and note that the column indices, j, are omitted for clarity.

Firstly, linear interpolation is performed horizontally across the artefact (note that a zeroth order or cubic interpolation could be used instead requiring respectively more examples or less examples for this method to estimate the artefact effectively). A rough estimate of the artefact, $A_{rough}(i)$, for each row, i, is obtained by subtracting the relevant columns of the original Lodox image, $F_{lodox}(i)$, from the linear interpolations, $F_{linear}(i)$. Each of these rough estimates $A_{rough}(i)$ is associated with the mean intensity level, $\bar{F} = \bar{F}_{linear}(i)$ of the linear interpolation at that row.

The true shape of the artefact $A_{true}(i)$ at row i can be written as

$$A_{true}(i) = A_{rough}(i) - \Delta F(i) \tag{3.3}$$

where $\Delta F(i)$ is detail of the true image profile that is clipped off by the linear interpolation as defined in Figure 3.9(b). If there are N rows with the same intensity level tag $\bar{F}$, indexed with k, the expected true artefact for this set of rows can now be calculated as

$$\begin{aligned} A_{true}(\bar{F}) \approx < A_{true}(k) > &= < A_{rough}(k) > - < \Delta F(k) > \\ &= < A_{rough}(k) > -0 \\ &= \frac{1}{N} \sum_{k} A_{rough}(k). \end{aligned} \tag{3.4}$$

The expected value of $\Delta F(k)$ is zero because it is assumed that there is no correlation in the clipped off detail amongst the rows in the underlying image. In Figure 3.9(b) the curve for $\Delta F(k)$ will change randomly for all different rows k throughout the image that is associated with the same intensity level tag $\bar{F}$. However, during this variation of $\Delta F(k)$ for different k at the same $\bar{F}$, the artefact remains constant according to the original algorithm assumption. Equation 3.4 shows how these two concepts work together to yield the artefact at $\bar{F}$.

In practice the intensity level tags $\bar{F}(i)$ are quantized and a linear interpolation scheme is used to build a lookup table of artefact shape versus intensity level. This lookup table is then used to predict the artefact shape at an arbitrary intensity level in the image which can then be subtracted out of the image.

3.3.3 Adaptive algorithm

The assumption on which the adaptive algorithm is based is that the artefact changes more slowly than the image detail from one row to the next (in the vertical direction). Therefore,

blurring the image one-dimensionally along the columns (in the vertical direction) will not have much effect on the artefact (nor on low resolution components of the image), but it will smooth out the image details. In this way, blurring is used to separate image details from low resolution components of the image. Afterwards, the incorrect low resolution components of the image, which are corrupted by artefacts, are discarded and replaced by interpolated smooth curves.

The restoration can be written as

$$< F_{true}(i) > \; F_{spline}(i) + F_H(i). \tag{3.5}$$

Here $F_{spline}(i)$ is a cubic spline interpolation over the artefact and the high resolution image detail is

$$F_H(i) = F_{lodox}(i) - F_L(i) \tag{3.6}$$

where the low resolution component $F_L(i)$ of the original $F_{lodox}(i)$ can be calculated efficiently from

$$F_L(i) = \frac{1}{N(i)+1}\left(F_{acc}(i+N(i)/2) - F_{acc}(i-N(i)/2-1)\right) \quad \text{and} \quad F_{acc}(k) = \sum_{i=1}^{k} F_{lodox}(i). \tag{3.7}$$

A bigger blurring kernel size, $N(i)$, is used where the underlying image has small intensity changes, $\Delta_i F_{true}|$, between rows or large intensity changes, $\Delta_j F_{true}|$, between columns. This is necessary to accomodate rapidly changing artefacts and maximizes the extent over which the artefact shape is inferred. These intensity gradients are estimated from restored images as obtained in Section 3.3.2. Thus

$$N(k) = \alpha \sum_{i=k}^{k+\Delta i} \sum_{j=j_0}^{j_1} < |\Delta_j F_{true}(i,j)| > -\beta < \Delta_i F_{true}(i,j)| > \tag{3.8}$$

where summation is done in a neighbourhood Δi of a fixed size that only moves as a function of row index. The size of this neighbourhood Δi equals 50 rows for an image with a resolution of 1.666 line pairs per millimeter and scales in proportion to resolution. Fine tuning of parameters α and β around 5.5 and 0.6 respectively by up to 25% of their nominal values can result in excellent artefact suppression for a particular image or region in an image. The values for these parameters were chosen as quoted above so that good results are obtained for a broad range of images. No single value is optimal for all images.

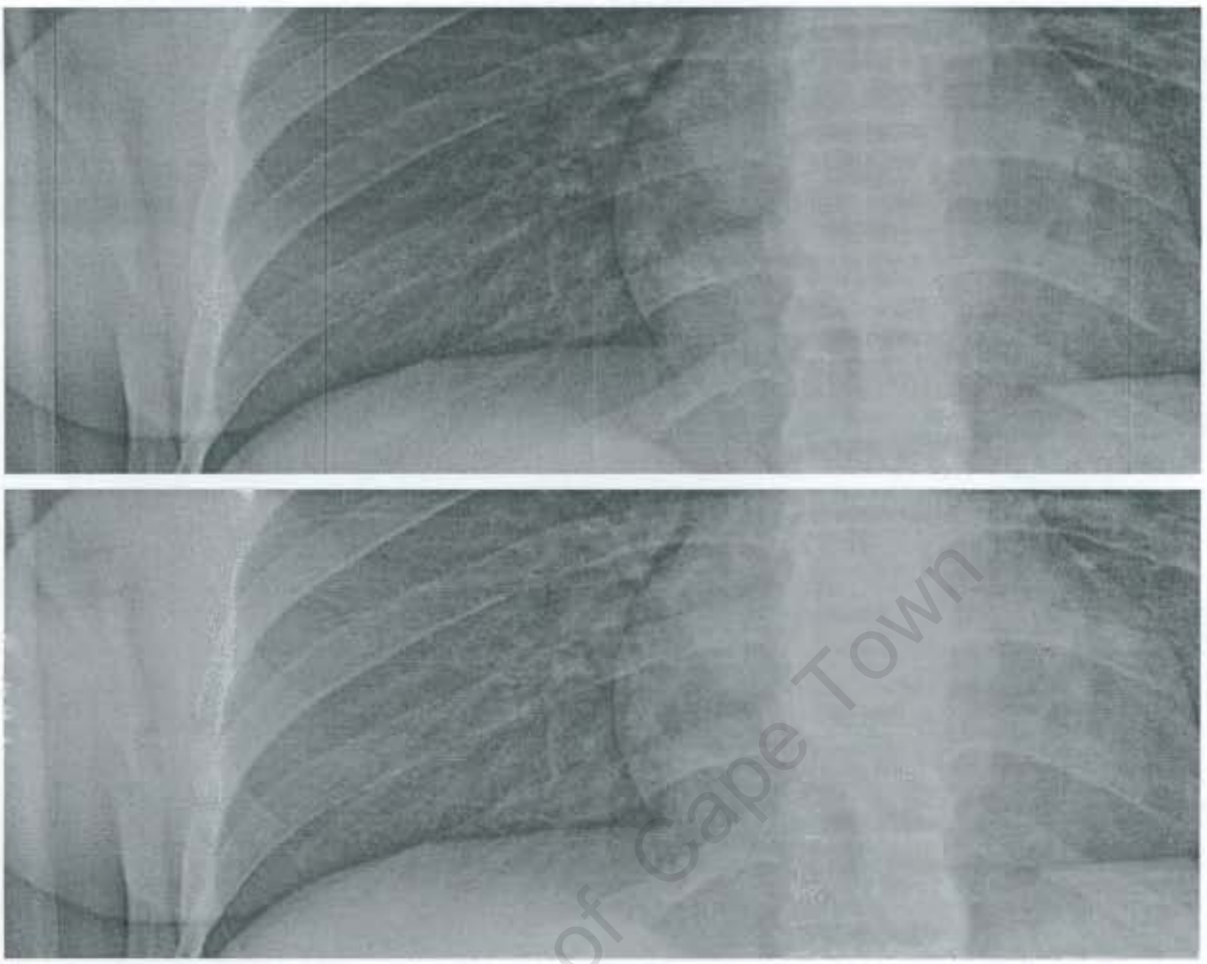


Figure 3.10 An original chest x-ray with artefacts (top) and an image with the artefacts removed (bottom).

3.3.4 Examples and performance

Figure 3.10 shows a typical example of the value that the artefact removal algorithm adds to the image. Note that there are some markings on the spine which are not visible unless the artefact removal algorithm is applied to the image. Both of the algorithms described above produce nearly indistinguishable results in the region of the image shown.

Figure 3.11 shows closeup views of the results of the two different algorithms applied to the same image region (the top of the skull). Evidently the adaptive algorithm is better capable of removing the artefacts near large intensity transitions. There are no instances found where the intensity level algorithm produces superior results to the adaptive algorithm.

(a) Original. (b) Intensity level method result. (c) Adaptive method result.

Figure 3.11 Results for different artefact removal algorithms.

A version of the adaptive algorithm has been used at a number of hospitals around the world for nearly a year with such great approval by its staff that the option of disabling such processing is no longer given to users of Lodox's Statscan systems.

3.4 Consistency

A projection can be considered to be a profile of values that is formed by a set of line integrals through a cross-section plane (see Figure 2.1). The integral of a projection through the same cross-section of an object equals the total mass of that section. It does not matter from what angle the projections are taken, the integral of the projection must still equal the total mass of the cross-section. Such consistency does not exist in the Lodox images used as the dataset, for the following reasons:

- The trolley is only partially imaged in some of the scans because of its width.
- There is some variation in x-ray generation and detection such that the intensity range of images could differ by up to 10% if a scan is repeated under identical operator conditions at different times.
- The image contrast and brightness vary throughout the image and differ abruptly between some adjacent cameras. This is especially noticeable at high attenuation and is affected by the variation in *dark current* and gain in each camera within a single scan. See

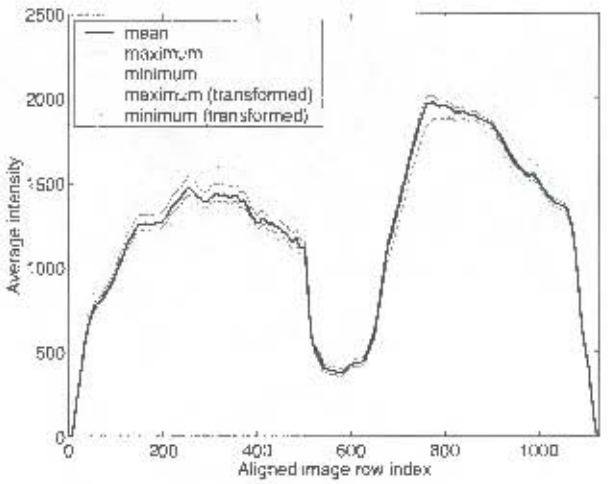


Figure 3.12 Average, maximum and minimum accumulated row intensity profiles in dataset.

Figure 3.9(a). Although these system inconsistencies have been solved in 2004 by slowing down CCD readout speeds, the dataset used in this thesis is subject to such corruption.

- Beam hardening and scattering of the polychromatic x-rays used in the Lodox system prevents the detector from measuring the actual linear attenuation coefficient (Kak and Slaney, 1988).

The contribution of the trolley to the inconsistency is removed by subtracting the mean trolley intensity profile from each row in a scan. This is possible since scanning occurs parallel to the trolley resulting in a nearly equal contribution to all rows. Slight mismatch at trolley edges is removed using the adaptive camera overlap artefact removal algorithm in Section 3.3.3.

Figure 3.12 shows the remaining inconsistency where the dotted lines represent the minimum and maximum mass of the object per row for the dataset (at different angles). This variation is reduced to that of the thin solid lines by optimizing the contrast and brightness (scale) of each scan so that they match the mean mass profile best in terms of mean square error.

Well known techniques in the literature (Kak and Slaney, 1988) for minimizing the effect of beam hardening in reconstructions are not possible in the context of limited angle tomography with Lodox. The dual energy technique involves taking an additional set of x-rays at a secondary energy level. The Lodox system is not designed to do that. Other post-processing techniques require the segmentation of bone out of reconstructions and estimating tissue thickness which is not as straightforward with limited angular data. Therefore, no further attempts are made

to improve the consistency in the dataset and henceforth reconstruction algorithms treat these inconsistencies as imperfections in the dataset along with the measurement noise.

3.5 Projection geometry

It it difficult to determine the system geometry to a high accuracy using conventional measurement instruments. Besides getting hold of suitable equipment and dismantling the system, it would remain a nontrivial task to locate, for example, the center of rotation and the precise position of the x-ray source.

Instead, now that the distortions in the image plane are quantified, the imaging system itself can be used to measure the C-arm geometry. This is done by scanning a reference object from different angles and consolidating markings on the object with their projections visible in the image.

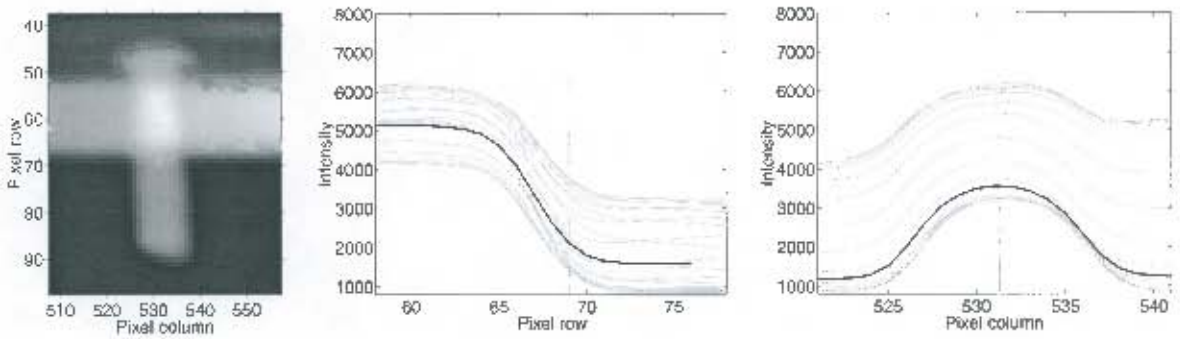
As seen in the upper parts of Figure 3.1, an upright perspex sheet with a regular grid of 9 aluminium pins can be found in every scan of the dataset. For identification purposes, each pin has a different length.

The cross-sectional relative spatial coordinates (x_k, y_k) of these pins are known *a priori* to an accuracy of $\sigma_{pin} = 0.057$mm (maximum fit error: $\epsilon_{pin} = 0.12$mm) from the log likelihood of a set of 36 unique pin to pin distance measurements $r'_{k,\bar{k}}$, i.e.

$$(x_k, y_k, \sigma_{pin}) = \arg\min N \log(\sqrt{2\pi}\sigma_{pin}) + \sum_{k>\bar{k}} \frac{1}{2} \frac{\left(r'_{k,\bar{k}} - \sqrt{(x_k - x_{\bar{k}})^2 + (y_k - y_{\bar{k}})^2}\right)^2}{\sigma_{pin}^2} \tag{3.9}$$

which has the same solution as the minimum variance χ^2 cost function, except that it also estimates the variance. A normally distributed error in the measurements is assumed.

In the Lodox images, the pin positions are manually located to the nearest pixel. Then, as the example in Figure 3.13(b) helps to show, the row index is chosen where the steepest average edge in the neighbourhood, orientated perpendicular to the perspex sheet, reaches 15% elevation. Along this row, the pin column position is determined by fitting a Gaussian curve to the intensity profile shown in Figure 3.13(c). The column positions of pins that appear close together in the images, are chosen manually and are therefore known with less accuracy. This collection of pins, as well as pins near camera overlap artefacts and pins within the image margins, are labeled *bad* pins. Pixel column distortion is not determined in the image margins where excessive gain



(a) Pin image. (b) Pin row position. (c) Pin column position.

Figure 3.13 Detection of a pin position in a Lodox scan.

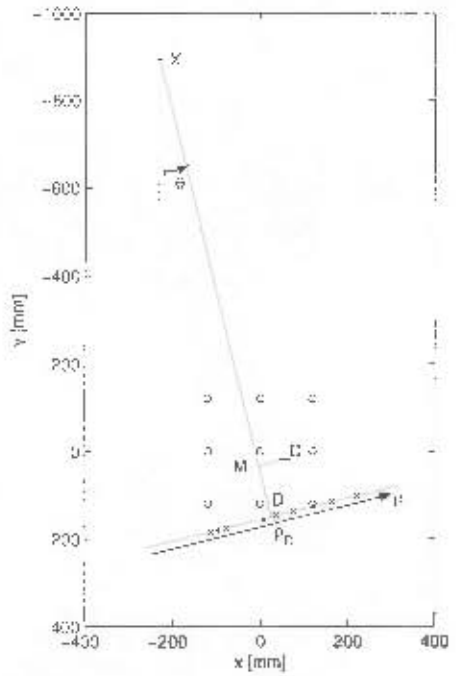
compensation problems occur. In the dataset, 312 out of 1638 pins are labeled *bad*. Although all the pin data is used equally, the results show that the *good* pins have less variance, as expected (see Figure 3.14(c) discussed later).

The pin positions are corrected for pixel column distortion by mapping them onto the phantom ruler represented by the variable ρ (which is approximately 30cm closer to the x-ray source from the physical detector) into millimeter units. Linear interpolation is employed in this mapping and yields the pin's projected position $\rho'_{k,\theta}$ on the ρ axis, where θ is the C-arm angle and k is the pin number.

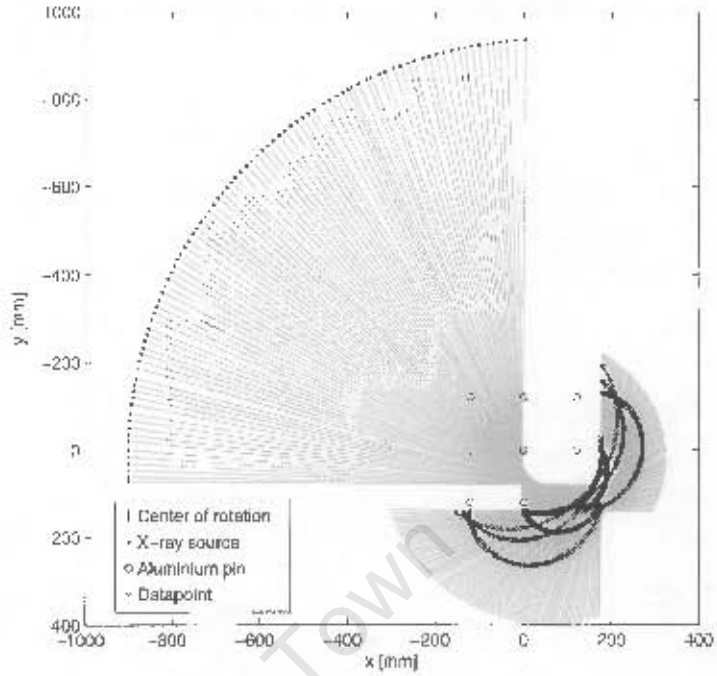
Figures 3.14(a) and 3.14(b) show the geometry of the C-arm and how the pins are projected onto the phantom ruler ρ from their cross-sectional coordinates. This projection is governed by the following equations:

$$\begin{aligned} x_{X,\theta} &= x_C - r_{MC}\cos(\theta) - r_{XM}\sin(\theta) & \phi_{k,\theta} &= \tan^{-1}\left(\frac{x_k - x_{X,\theta}}{y_k - y_{X,\theta}}\right) \\ y_{X,\theta} &= y_C + r_{MC}\sin(\theta) - r_{XM}\cos(\theta) & \rho_{k,\theta} &= r_{XD}\tan(\phi_{k,\theta} - \theta) - \rho_D \end{aligned} \tag{3.10}$$

In these equations, the x-ray source is at point X=$(x_{X,\theta}, y_{X,\theta})$ when the C-arm is at angle θ with respect to the center of rotation at C=(x_C, y_C). Pin k at (x_k, y_k) is projected onto the projection axis at $\rho_{k,\theta}$ where ρ_D is the origin of the projection axis at point D. Further r_{MC}, r_{XM} and r_{XD} are the distances from the points M to C, X to M and X to D respectively as quantified in Figure 3.14. The intermediate quantity $\phi_{k,\theta}$ is used to simplify the calculation of $\rho_{k,\theta}$, and represents the angle to pin k from the x-ray source.



(a) Geometric definitions.



(b) Geometry of dataset for $\theta \geq 0°$.

Label	Value [mm]	Description
σ_ρ	0.14	standard deviation in ruler plane
$\sigma_{\rho,good}$	0.10	standard deviation in ruler plane for good pins
$\sigma_{\rho,bad}$	0.25	standard deviation in ruler plane for bad pins
σ_r	0.12	standard deviation in cross-section plane
$\sigma_{r,good}$	0.09	standard deviation in cross-section plane for good pins
$\sigma_{r,bad}$	0.21	standard deviation in cross-section plane for bad pins
ϵ_ρ	0.64	maximum error in ruler plane
$\epsilon_{\rho,good}$	0.31	maximum error in ruler plane for good pins
ϵ_r	0.54	maximum error in cross-section plane
$\epsilon_{r,good}$	0.28	maximum error in cross-section plane for good pins
x_C	56.01	center of rotation for $\theta \geq 0$
y_C	19.27	center of rotation for $\theta \geq 0$
x_{C-}	-54.51	center of rotation for $\theta \leq 0$
y_{C-}	19.72	center of rotation for $\theta \leq 0$
r_{XD}	1074.53	x-ray source to phantom ruler
r_{XM}	954.55	x-ray source to center of rotation (component)
r_{MC}	60.23	lateral offset
ρ_D	495.39	center value of phantom ruler

(c) Geometry results

Figure 3.14 Geometry of the Lodox machine.

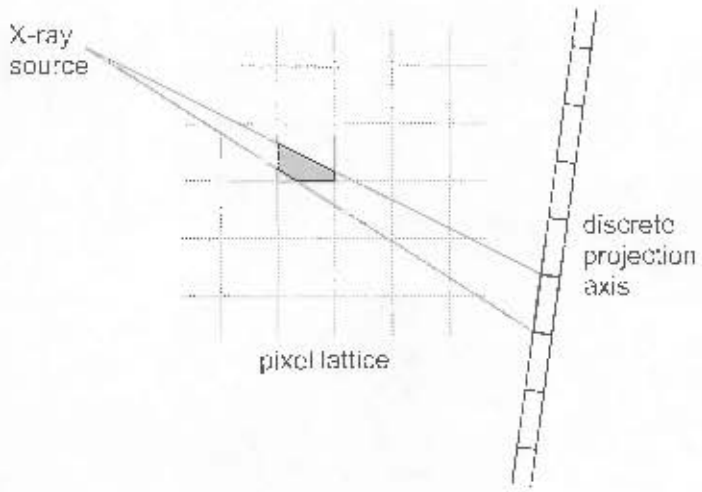


Figure 3.15 Contribution of pixel segment to a projection bin.

The minimum variance data fit is found using a local search algorithm, by minimizing

$$\chi^2 = \sum_{k,\theta} \left(\rho'_{k,\theta} - \rho_{k,\theta}\right)^2 \tag{3.11}$$

which has 190 parameters. Omitting the C-arm angles, θ, the results are given in Figure 3.14(c).

The results suggest that a point in the projection plane can on average be pinpointed to an accuracy of $\sigma_{r,good} = 0.09$mm in the cross-sectional (axial) plane using the image corrections and geometric setup described in this chapter. This limits the resolution to which reconstructions can be performed with no regard to the resolution of the dataset.

Although the geometric parameters could be determined to greater accuracy using more pins in the reference object and higher resolution data, the assumption that the estimated parameters are constants breaks down at some point. Finite element analysis of the C-arm structure by Ken Park (Park, 2002), reveal that deflections of up to $100\mu m$ are expected as a function of C-arm angle due to gravity alone.

Therefore, if a greater order of accuracy is required, then not only should more reference points be used, but each scan's geometric parameters should be considered individually (or at least locally as a function of θ).

However, the accuracy to which the geometry is known at this stage is believed to be well matched to the resolution of the dataset, which is 0.3mm per pixel.

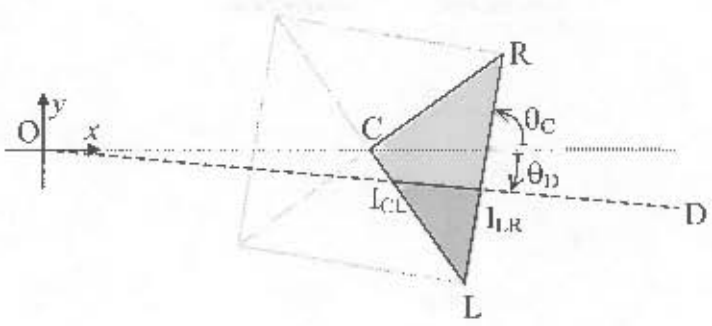


Figure 3.16 Accumulative projection of triangle.

3.6 Pixel projection

Figure 3.15 illustrates the *partial volume effect* (Beutel *et al.*, 2000) where part of a pixel on a regular grid is projected onto a discrete projection axis. In the general case, the contribution that the shaded portion of the pixel makes to the projection bin is an elaborate analytical function of the relative orientation and position of various components in the diagram.

A simplified diagram is shown in Figure 3.16 where a pixel consists of the union of four rotations of a triangle about point C. The pixel orientation is expressed using θ_C.

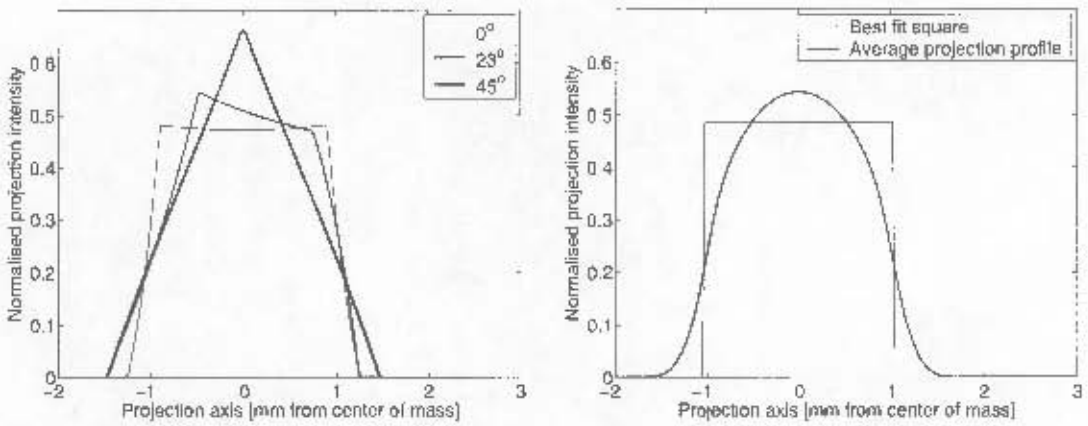
It is useful to define a general function $A_{CLR}(OD)$ which determines that part of the area of a triangle CLR that lies below the line OD. In this simplified example, that function is expressed using the angle of the line OD instead, such that

$$A_{CLR}(\theta_D) = \begin{cases} 0 & \text{if} \quad \theta_D < \tan^{-1}(y_L/x_L) \\ A(I_{CL}, I_{LR}, L) & \text{if} \quad \tan^{-1}(y_L/x_L) < \theta_D \leq 0 \\ A(C, L, R) \quad A(I_{CR}, I_{LR}, R) & \text{if} \quad 0 < \theta_D \leq \tan^{-1}(y_R/x_R) \\ A(C, L, R) & \text{if} \quad \theta_D > \tan^{-1}(y_R/x_R) \end{cases} \tag{3.12}$$

where the area of a triangle with vertices C, L and R can be calculated from

$$A(C, L, R) = \frac{1}{2} \cdot \begin{vmatrix} x_C & y_C & 1 \\ x_L & y_L & 1 \\ x_R & y_R & 1 \end{vmatrix}. \tag{3.13}$$

The area $A_{\theta_C}(\theta_D)$ of the pixel at angle θ_C below the same line is calculated by summing the partial areas of the individual triangles that make up that pixel. Finally, the contribution that



(a) Profiles of pixels at different angles.

(b) Average and best fit profiles.

Figure 3.17 High resolution projection profiles of individual pixels.

the pixel makes to an interval θ_{D_i} to $\theta_{D_{i+1}}$ on the discrete projection axis is then determined by subtracting the accumulated areas $A_{\theta_C}(\theta_{D_{i+1}}) - A_{\theta_C}(\theta_{D_i})$.

3.7 Projection axis resolution

Given a particular spatial resolution, this section describes how an optimal projection axis resolution is selected.

Figure 3.17(a) shows a few projections of a single square pixel onto a high resolution projection axis at different angles. Clearly the shape of the pixels affects the projections more significantly the higher the projection axis resolution. If the projection axis resolution is too high, the reconstruction is overspecified by the projections because it has structure that either contradicts (real data from continuous origin) or concurs with (simulated data from discrete origin) that which is consistent with the shape of the pixels. As demonstrated in the final chapter of this thesis, a resolution mismatch of this kind introduces avoidable noise in real data iterated reconstructions. On the other hand, if the projection resolution is too low, more pixels are used than necessary to represent the projection data, and reconstructions can be poorly specified.

Figure 3.17(b) shows the average projection profile of all pixels in a 128×128 lattice for typical geometrical conditions used in this thesis. A rectangle of equal area is fitted optimally in mean square error to this curve. The width of this block defines the bin spacing used as the best

choice of projection axis. For a 128×128 lattice this corresponds to a 216 element projection axis in this case.

3.8 Filtered back-projection reconstruction

The filtered back-projection method described in Section 2.1.4 is used to perform tomographic reconstruction. In essence, it requires that the x-ray projections are filtered with $H(\omega) = \alpha|\omega|$ in the frequency domain where α is some constant. The filtered projections are then cumulatively projected back onto a pixel grid. The results of this operation is provided below.

3.8.1 Results

Figures 3.18 and 3.20 to 3.22 show a few typical full angle tomographic reconstruction slices of the object using 182 scans at an angular spacing of $\Delta\theta = 1°$. The head or limbs are the most likely parts of the anatomy to be imaged tomographically in a clinical situation with the Lodox system, because they are relatively easy to keep still for the long periods of time it will take to perform the scans. Acquiring 182 scans of a human head cannot be done faster than 12 minutes if the machine were fully automated for this purpose (manually it took 4 hours).

In contrast to most tomography machines, a scan has to be completed before the x-ray source is rotated. This makes the Lodox system a lot more susceptable to motion-blur problems due to, for example, breathing. The benefit, however, is that the Lodox system offers very high resolution imaging in the scanning direction. For these reasons, perhaps the most suitable application for this study is forensic imaging.

It is possible to develop *single slice* tomographic capabilities into the Lodox system which could offer very high contrast resolution. This requires a special mode where the C-arm rotates while data is acquired. However, this topic falls outside the scope of this thesis.

Figure 3.19 shows an intensity profile through Figure 3.18 which is an enlargement of the slice at scanning index $i_z = 225$ using 1024×1024 0.22mm pixels. An intensity profile of the equivalent low resolution 128×128 1.8mm pixel reconstruction in Figure 4.1(c) is also provided in Figure 3.19.

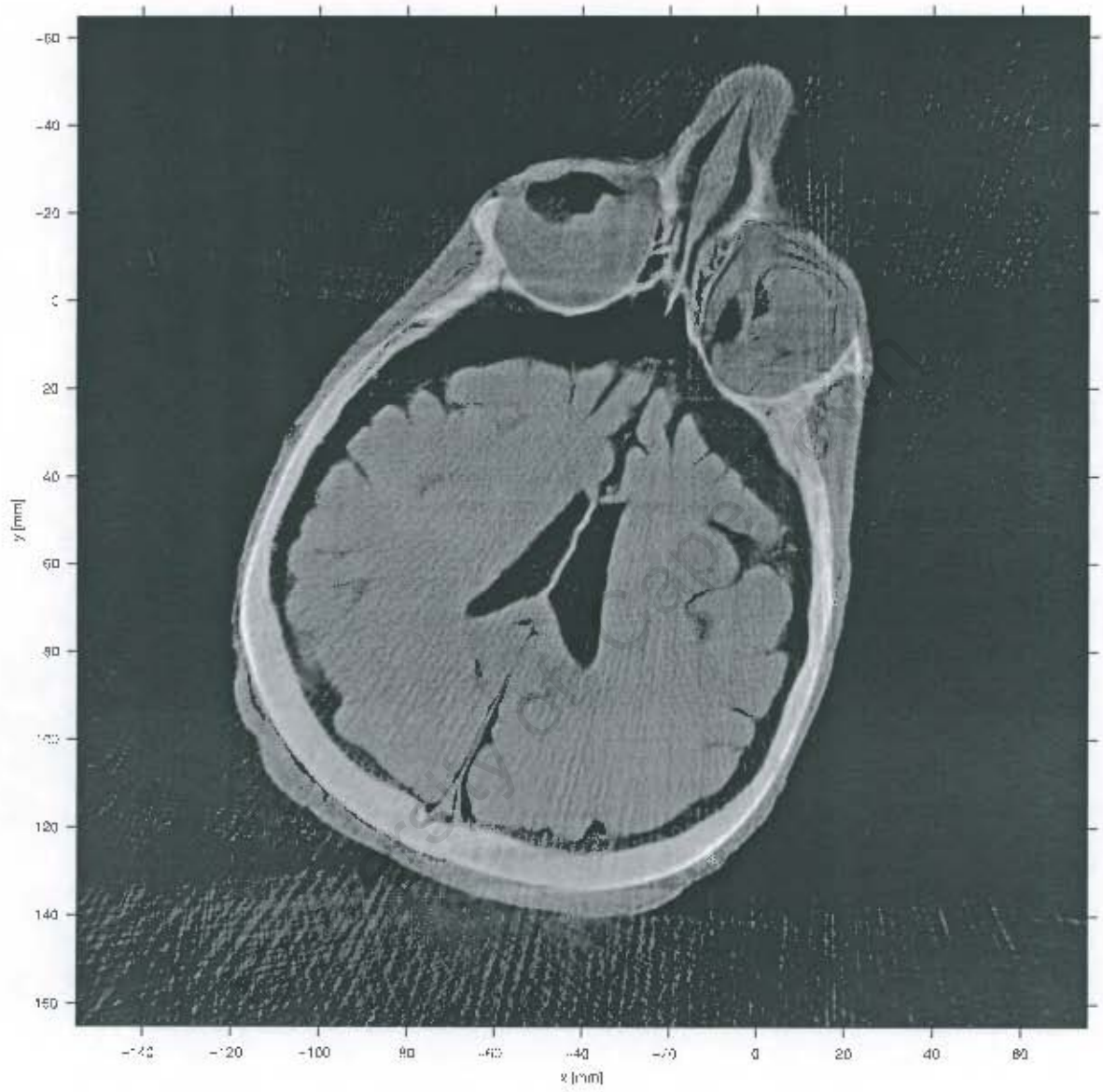


Figure 3.18 Filtered back-projection reconstruction slice at $i_z = 225$.

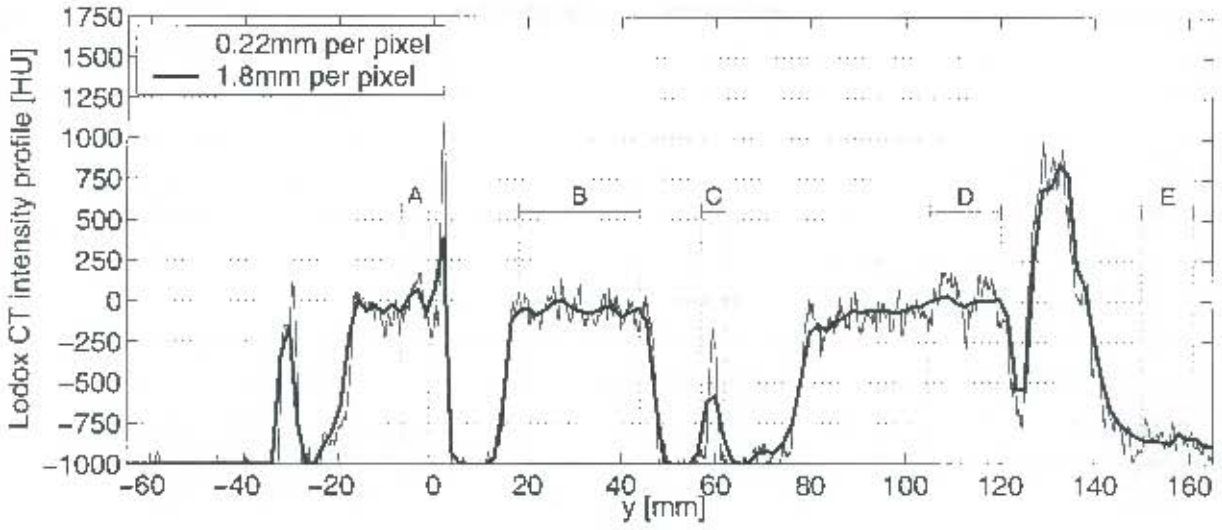


Figure 3.19 Intensity profiles through 128×128 and 1024×1024 filtered back-projection reconstructions at $x = -44.38$mm, $i_z = 225$.

The light streak running horizontally between bright bone sections in the lower part of the left hand side eye is an example of a *beam hardening artefact* (Beutel *et al.*, 2000). This happens because projection data originate from polychromatic instead of monochromatic x-rays and is therefore not an exact measure of the linear absorption coefficient of tissue (Kak and Slaney, 1988). Figure 3.19 indicates this by Region A as an intensity profile in Hounsfield units [HU] (see Figure 4.12(b) for interpreting Hounsfield units).

Noise levels and other fluctuations in the reconstruction of the same homogenous tissue are indicated by Region B. For the 1024×1024 reconstruction, the peak to peak profile variation is nearly 250 HU. For the 128×128 reconstruction in Figure 4.1(c), this variation is reduced to about 100 HU.

Region C illustrates the partial volume effect. At a low spatial resolution, it is not possible to represent detailed structures accurately. The presence of tissue is indicated inconclusively using 1.8mm pixels. Although increasing the slice thickness (not shown) by binning in the scanning direction reduces noise in reconstructions, it does not attenuate the artefacts and it worsens the partial volume effect.

The ripples of fine light and dark streaks seen throughout Figure 3.18 are aliasing artefacts due to the relatively small number of angles at which scans were taken. Region D in Figure 3.19 shows that the variations can be up to 500 HU peak to peak. These ripples are not visible in

the 128 × 128 reconstruction where the angular resolution is more comparable to the display resolution.

Near the bottom of the image, the lower edge of the skull appears to extend a light streak over the scalp tissue to its left. Also notice the slight increase in intensity for both resolutions of over 100 HU below the head which is labeled Region E in Figure 3.19. This happens because the total angular range over which x-rays in the 180° dataset were detected, is less than 180° for points in that region of space. For fan beam tomography, the angular range at which projections must be taken is 180° plus the fan angle to ensure that 180° data exist for all points in the image space. Therefore, the dataset is technically not complete and no reconstructions in this thesis are quite full angle.

Figure 3.20 shows an enlargement of a slice at $i_z = 450$ using 0.22mm pixels while Figure 3.21 displays only the central 40mm×40mm detail region of this slice using 1024 × 1024 pixels. This image pushes the limits of the dataset resolution, as well as the accuracy to which the geometry of each scan is known.

3.8.2 Performance

Reconstruction speed is fairly slow when no compromise is made on the accuracy of the pixel projection process. Whilst determining all intersecting areas of beams and pixels on the fly, back-projection takes approximately 4.87μs per scan per pixel on a 1.8GHz Pentium 4 processor. In other words, the 6 512×512 slices using 182 scans shown in Figure 3.22 took about 23 minutes to compute.

For low resolution cross-sections, the contributions that projected pixels make to the projections can be precomputed and stored in a lookup table. In this way a very large improvement in speed for 128x128 images is possible. A 128 × 128 × 128 cube takes 2 minutes and 30 seconds to reconstruct at 1.17s per slice.

(Coric *et al.*, 2002) reports a speedup for the back-projection process of up to 20 times using dedicated hardware and fixed point arithmetic. Hardware acceleration may be essential to the future commercialization of tomographic functionality for Lodox. However, for the development of new limited angle tomography techniques, no compromise on accuracy is tolerated.

3.9 Summary

In this chapter many real world problems with the system and dataset were investigated and overcome. It was shown that accurate reconstructions are possible and therefore the dataset and the system are well characterized. The encapsulation of this reconstruction ability is a fundamental building block for the rest of this thesis.

The results have shown that although a high spatial resolution is obtainable, relatively poor contrast (or density resolution) is achieved due to the poor angular resolution and artefacts. The next chapter investigates how to obtain better reconstructions from a limited number of projections in the presence of noise and inconsistencies in the data.

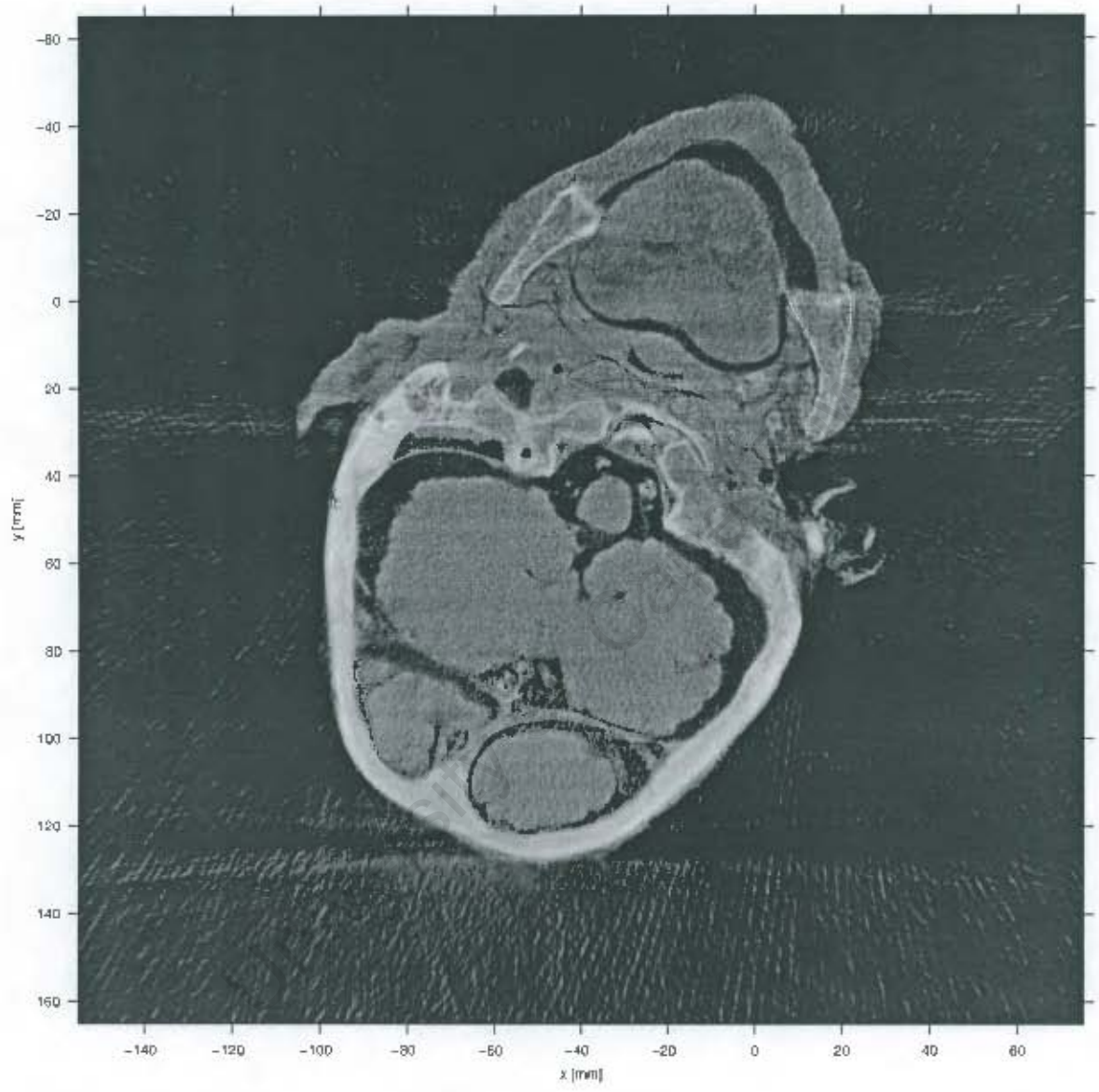


Figure 3.20 Filtered back-projection reconstruction slice at $i_z = 450$.

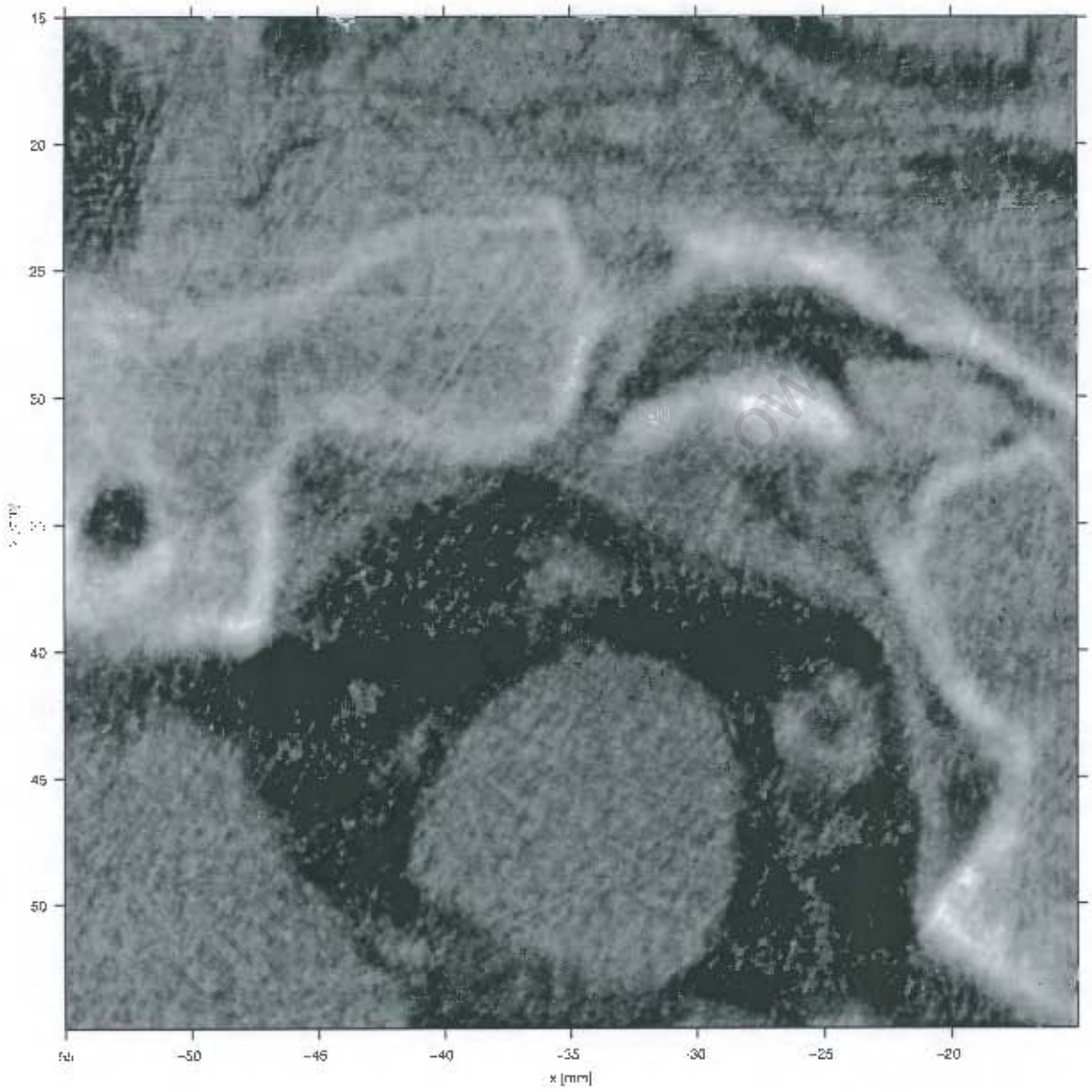


Figure 3.21 High resolution reconstruction using 39μm per pixel at $i_z = 450$.

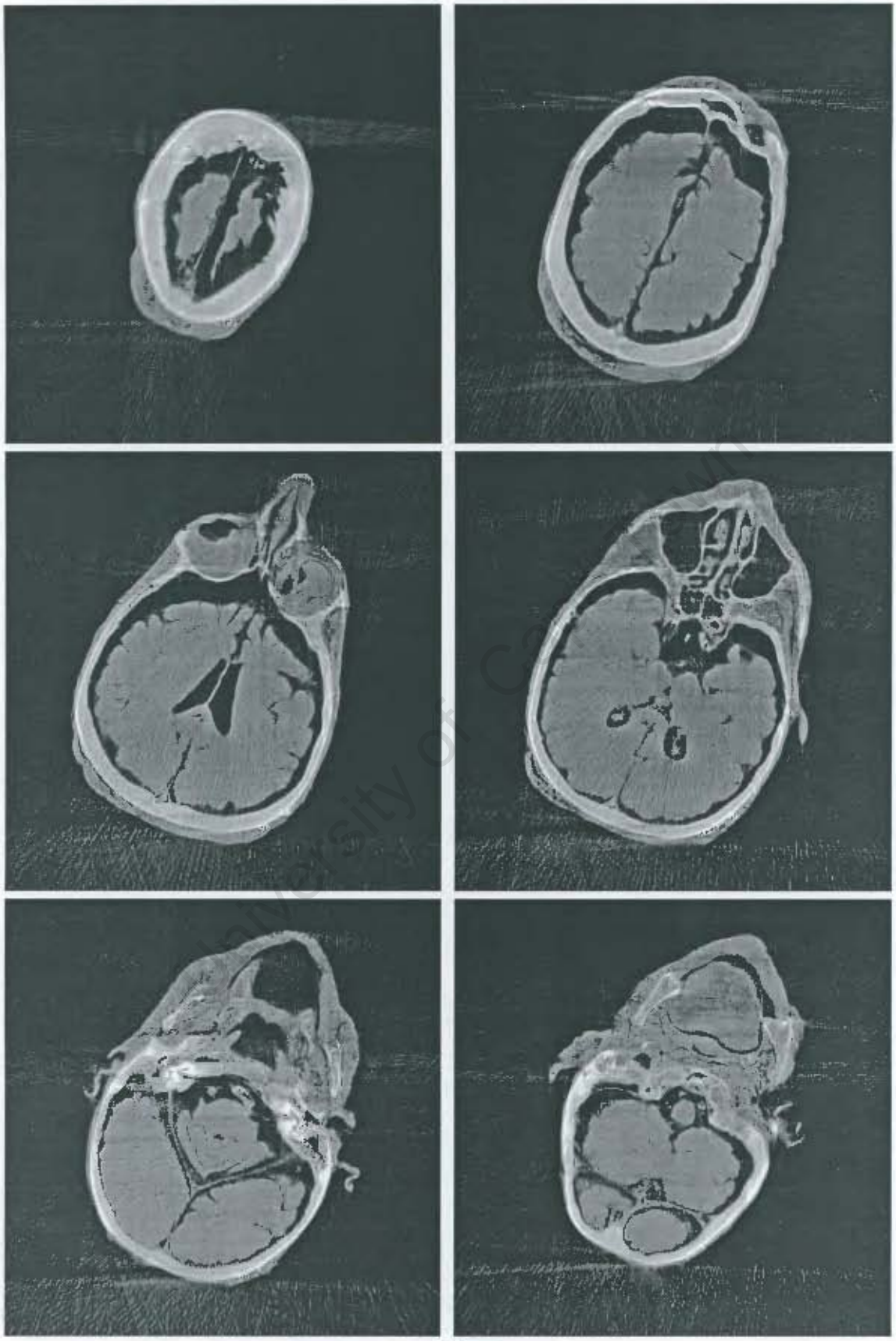


Figure 3.22 Filtered back-projection reconstruction slices at $i_z = \{75, 150; 225, 300; 375, 450\}$.

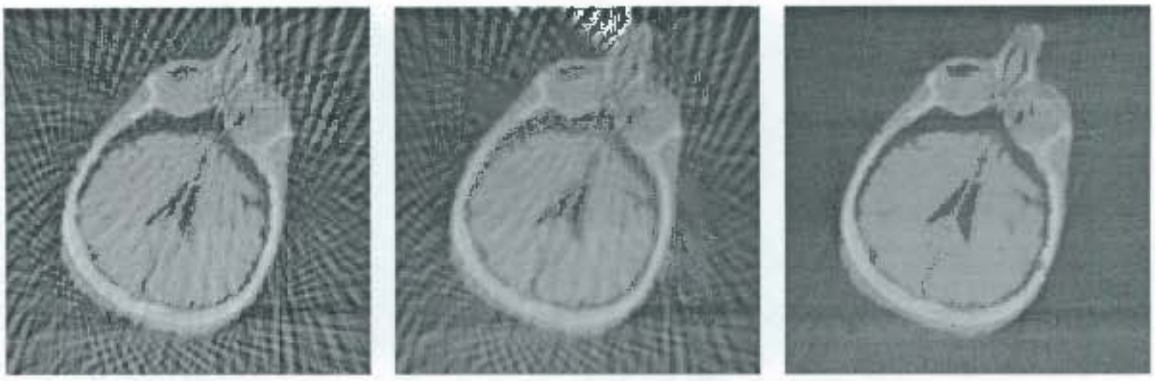
CHAPTER 4

Optimizing reconstruction

Given a complete set of ideal projections, the filtered back-projection algorithm theoretically converts the projection data to an exactly equivalent cross-section with no need for iteration. If data is sparse, subject to measurement errors and beam hardening, this algorithm will however *not* necessarily reconstruct a cross-section such that the projections of that cross-section equal the projection data best in terms of mean square error. For example, in an extreme case where only a single projection is known, an *un*filtered back-projection will clearly result in a reconstruction that is more consistent with the data than a filtered back-projection. This chapter investigates the best way to perform reconstructions from limited datasets so that they are most consistent with measured projection data.

The *criterion of optimality* used for the better part of this chapter is the mean square error, $\sigma^2 \equiv \chi^2$, of the reconstruction projections with respect to the measured projection data. In this thesis, σ_{5° is the root mean square data projection error of projections spaced at $\Delta\theta = 5^\circ$ intervals over a 180° range. Similarly, σ_{1° is the root mean square data projection error of projections spaced at $\Delta\theta = 1^\circ$ intervals over a 180° range. Note importantly that σ_{1°, which is quoted in many figures, represents a quantitative measure of comparison for a reconstruction to the *ground truth*. This parameter is particularly useful for assessing sparse and limited angle reconstructions throughout this thesis.

In the following section it is demonstrated that the filtered back-projection method should be modified when data is sparse to minimize differences of the reconstruction projections with the projection data. These findings are then extended to iterative reconstruction methods. The final sections of this chapter discuss that reconstructions should be regularized to best provide for unmeasured data. Bayesian reconstructions are included.



(a) Ramp filter, $\Delta\theta = 5^\circ$, $\sigma_{5^\circ} = 489$, $\sigma_{1^\circ} = 489$.

(b) Optimal filter, $\Delta\theta = 5^\circ$, $\sigma_{5^\circ} = 304$, $\sigma_{1^\circ} = 326$.

(c) Optimal filter, $\Delta\theta = 1^\circ$, $\sigma_{5^\circ} = 299$, $\sigma_{1^\circ} = 298$.

Figure 4.1 Single step reconstructions ($i_z = 225$).

4.1 Maximum likelihood reconstruction

As discussed in Section 2.3, the maximum likelihood reconstruction is the reconstruction that fits the available projection data best in terms of mean square error. The class of reconstruction techniques advocated below differs from most standard methods in that a search is generally performed in a *subspace* of search directions (Skilling and Bryan, 1984) instead of a *single* search direction at each stage that the reconstruction is updated from $f^{(n)}$ to $f^{(n+1)}$.

It is constructive to point out that *direct inversion* reconstruction methods in Section 4.1.1 employ a self-contained iterative optimization procedure which is used to determine controlling filter parameter values. These are regarded as *single step* methods because they closely resemble a single step of classical gradient based search methods (see Section 2.3.2).

Iterative methods in Section 4.1.2 simply iterates direct inversion steps which automatically recalculates the projection differences and filter parameters due to its new starting position.

4.1.1 Direct inversion in a single step

Figure 4.1 shows direct inversion reconstructions where projection data is known either at $\Delta\theta = 5^\circ$ or 1° intervals as indicated. At a display resolution of 128×128 pixels, the projection data for the 1° case is considered to be *complete* while in the 5° case it is considered to be *sparse*. The poor angular resolution results in aliasing artefacts which are clearly visible in

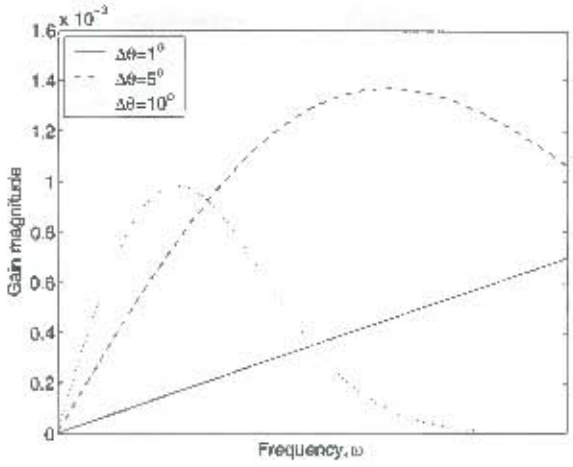


Figure 4.2 Optimal filters for direct inversion of 128×128 reconstructions.

Figures 4.1(a) and (b). Figures 4.1(a) and (c) appear to have visually pleasing sharp edges while Figure 4.1(b) appears slightly blurred.

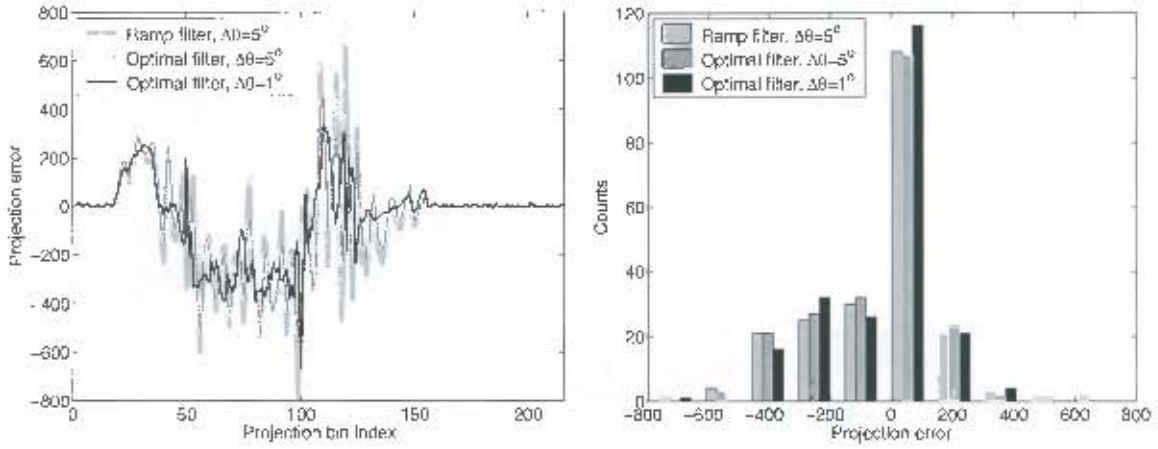
For the reconstruction in Figure 4.1(a) the familiar ramp filter, $H_1(\omega) = h_1 \cdot |\omega|$, was used on the projections prior to back-projecting, while for Figures 4.1(b) and (c) an optimal blurring component is included in the filter of the form:

$$H_3(\omega) = (h_0 + h_1 \cdot |\omega|)e^{-\frac{1}{2}(h_2\omega)^2}. \tag{4.1}$$

Parameter values h_0, h_1 and h_2 define the shape of the filter and are chosen optimally for each reconstruction individually. These values are found by minimizing the σ^2 (equivalent to the familiar χ^2) error between the available data and the corresponding projections of the *resultant* reconstruction using the Nelder Mead simplex minimization algorithm (Press *et al.*, 1999).

As expected, the theoretically correct ramp filter with no blurring component ($h_2 = 0$) was found to be the most optimal filter choice for the case of a complete dataset with 1° data spacing intervals given a blank image as the starting point. For interest's sake, Figure 4.2 shows the optimal filter shapes for a few different angular spacings in projection data. The actual values used are calculated internally in an automatic way, as described above, and vary depending on the starting point, and also vary slightly for different slices.

The reason why a blurring component is included is to reduce the sharpness of the angular aliasing artefacts, which in turn improves the projection error as shown in Figure 4.3(a). In other words, although the reconstruction appears more blurred in Figure 4.1(b), its projections fit the data projections better in mean square error than in Figure 4.1(a).



(a) Projection error profiles.

(b) Projection error histograms.

Figure 4.3 Projection data errors at $\theta = 0°$ for single step reconstructions in Figure 4.1.

Figure 4.3(b) illustrates crudely by means of histograms that the statistical nature of the data fit error is asymmetrical and non-Gaussian. This is due to inconsistencies in the dataset as discussed in Section 3.4.

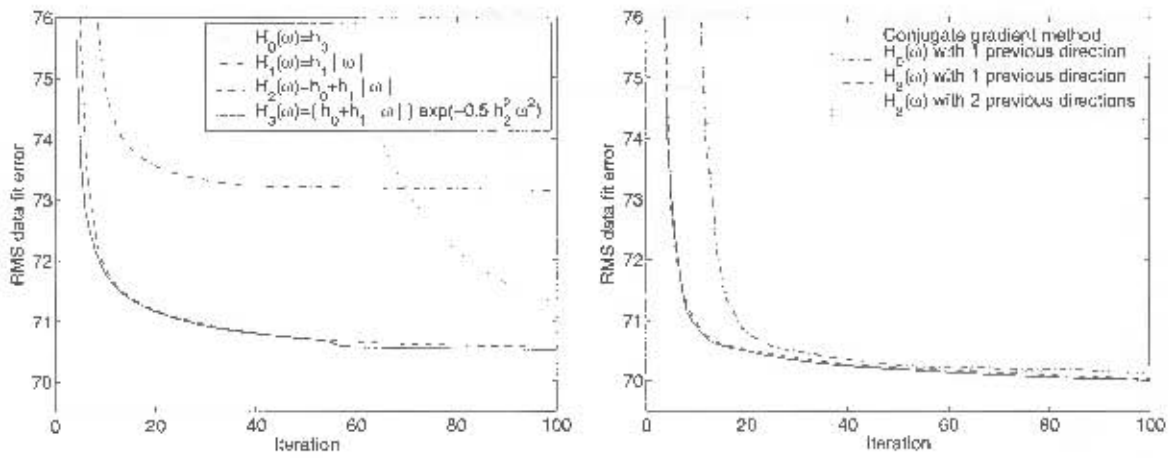
Direct inversion methods are well suited for high resolution reconstructions when a large amount of data is available. For low resolution images it is feasible to use more computationally expensive algorithms in exchange for better reconstruction quality. The rest of this chapter investigates more elaborate iterative techniques.

4.1.2 Iterative reconstruction methods

A much better data fit can be achieved by iterative methods. The aim of these methods is to minimize the σ^2 error that the projections of the reconstruction make with the projection data.

The simplest algebraic reconstruction method (Kak and Slaney, 1988) repeatedly projects back some fixed fraction of the difference between the projections of the current reconstruction $f^{(n)}$ and the data (i.e. the residuals) onto $f^{(n)}$ to produce a new reconstruction $f^{(n+1)}$.

Some improvement is possible if this fraction is determined by minimizing the resultant σ^2 error of $f^{(n+1)}$ at each step as in Section 2.3.2. Evident from the curve for $H_0(\omega)$ in Figure 4.4(a), this alternative strategy still has a poor rate of convergence.



(a) Steepest descent methods.

(b) Methods using previous search directions.

Figure 4.4 Convergence curves for iterative reconstruction methods with $\Delta\theta = 5°$.

Instead of updating the reconstruction in a single direction at each step, the more effective iterative methods discussed below consider multiple directions.

4.1.2.1 Steepest descent

Steepest descent algorithms simply iterate direct inversion steps exactly as in Section 4.1.1. The result of one iteration is used as the starting point for the next iteration. At each iteration step, filter parameters (h_0, h_1, and h_2) are recalculated by the Nelder Mead algorithm so that the σ^2 error decreases maximally. In other words, these values are *automatically* determined and are not constant from one iteration to the next. They are somewhat data-dependent and tend to change dramatically during the first few iterations but still keep changing throughout the reconstruction process.

As shown in Figure 4.4(a), the rate of convergence can be improved by *filtering* the projection difference: the classical steepest gradient as in Section 2.3.2, using $H_0(\omega) = h_0$, converges slowly but steadily while the theoretical filtered back-projection choice, $H_1(\omega) = h_1 \cdot |\omega|$, converges much faster initially but comes to a standstill later on.

The best of both worlds is achieved by adding an offset parameter to the filter, so that $H_2(\omega) = h_0 + h_1 \cdot |\omega|$.

A further yet insignificant development, evident in the first few iterations only, is possible by blurring the projections as in Equation 4.1. The added computation required to determine the blurring parameter does not justify the use of filter $H_3(\omega)$.

4.1.2.2 Conjugate gradients

The conjugate gradient method, as described in Section 2.3.2, is the standard modification to the classical steepest descent algorithm (using $H_0(\omega) = h_0$) aimed at solving quadratic or nearly quadratic optimization problems in the fewest number of iterations possible. In this method, a new search direction is chosen that is conjugate to all previous search directions. A one-dimensional *line search* operation is performed in this conjugate search direction at each step. The conjugate gradient method has been implemented for comparison to the algorithm discussed next.

4.1.2.3 Using components of previous search directions

Previous search directions can easily be stored in a buffer. Instead of searching within the filter parameter (h_0 through h_2) subspace alone for the next iterate, the subspace can be augmented by fractions of previous search directions to be added to the current search direction (in the projection domain). Figure 4.4(b) shows numerically as a function of iteration number that the performance of the conjugate gradient method is almost indistinguishable from the case where the next iterate is searched for in the two directional subspace defined by h_0 and a fraction of the previous search direction.

The figure also shows that the RMS data fit error is significantly less during the first 20 iterations when searching in the three-directional subspace defined (in filter form H_2) by h_0, h_1 and a fraction of the previous search direction. Only a minuscule further improvement is achieved using additional older search directions, possibly because the data fit optimization problem is inherently quadratic only.

It was found (not shown to avoid cluttering) that using the filter form H_3 with previous search directions does not have a noteworthy advantage over filter form H_2 under the same conditions.

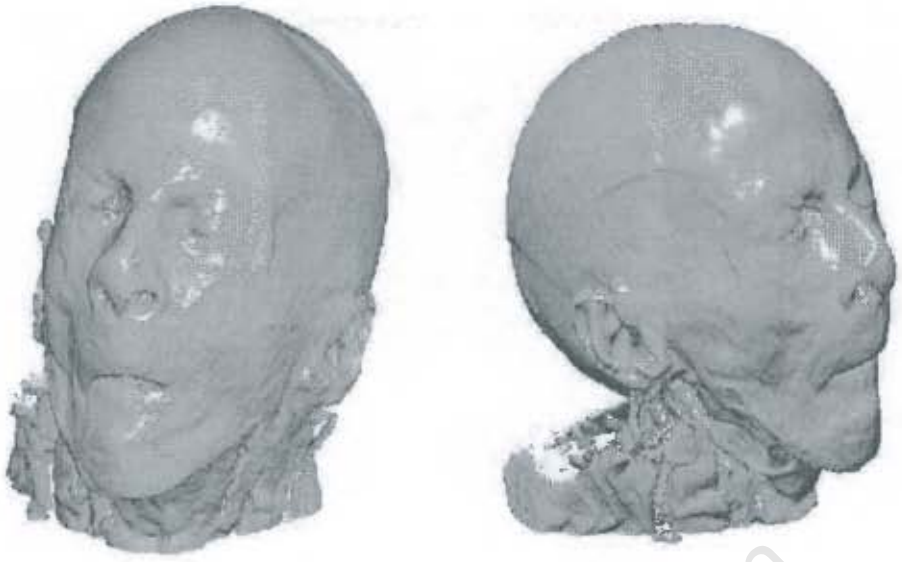


Figure 4.5 Isosurface views of reconstructed volume using the complete dataset.

4.1.3 Iterative volume reconstruction

Efficient *volume* reconstructions can also be achieved using the techniques described above. At each iteration, the parameters defining the back-projection filter and the fraction of the previous search direction to be added, are optimized for a single slice of the volume only. The slice with the greatest data fit error is selected at each iteration for such optimization. All the other slices make use of the same parameter settings. This technique offers stability, rapid convergence as well as minimal computation. Figure 4.5 shows two isosurface views of a $128 \times 128 \times 128$ volume reconstruction of the complete dataset.

4.2 Implications of having sparse data

Very good maximum likelihood reconstructions are possible when data is complete. Redundancy in the data alleviates noise, artefacts and other errors in the dataset when the σ^2 error is minimized. Of more practical interest, however, is the situation where data is sparse.

This section explores what happens during the reconstruction process when the angular resolution ($\Delta\theta$) of available data is poor.

(a) Checkered start, $\Delta\theta = 5°$, $\sigma_{5°} = 70.05$, $\sigma_{1°} = 289$.

(b) Blank start, $\Delta\theta = 5°$, $\sigma_{5°} = 70.05$, $\sigma_{1°} = 144$.

(c) Checkered start, $\Delta\theta = 1°$, $\sigma_{5°} = 72.43$, $\sigma_{1°} = 69.77$.

Figure 4.6 Iterated reconstructions ($i_z = 225$).

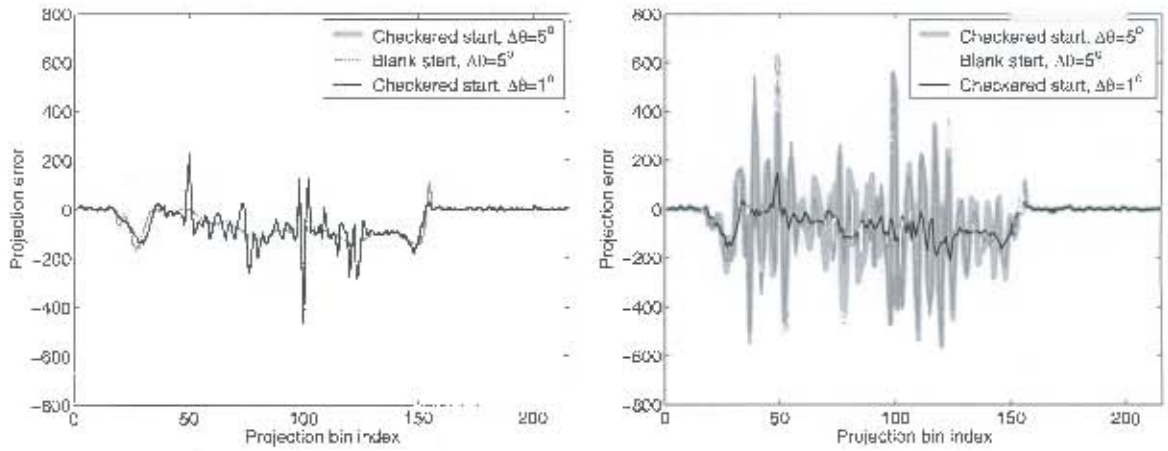
4.2.1 Ambiguity and the influence of the starting point

Figure 4.6 shows iterated reconstructions using the algorithm described in Section 4.1.2.3. The projections of the reconstructions in Figures 4.6(a) and (b) are both fitted to the same projection data with an RMS error of $\sigma_{5°} = 70.05$ equally, yet the reconstructions appear quite different. This illustrates the ambiguity that exists in the space domain due to the incompleteness of data for $\Delta\theta = 5°$ intervals. A much better and definitive reconstruction can be achieved when sufficient data is available. For the reconstruction using $\Delta\theta = 1°$ data intervals, the starting point was found to be irrelevant.

At this stage iterated reconstruction methods do not seem to provide much visual improvement over direct reconstruction methods, despite the immense numerical improvements evident in the data fit.

4.2.2 Inconsistencies and overfitting

The RMS 5° ($\sigma_{5°}$) data errors achieved by the reconstructions in Figure 4.6 (a), (b) and (c) after 100, 89 and 100 iterations, are 70.0486, 70.0498 and 72.43 respectively. All $\sigma_{5°}$ RMS errors are calculated using the same projection geometry at $\Delta\theta = 5°$ intervals, even for reconstructions where projection data was available for $\Delta\theta = 1°$. (The RMS error with respect to the entire dataset at 1° intervals is expressed by $\sigma_{1°}$). These RMS values reflect that a better data fit can be achieved to sparser data in the presence of noise and inconsistencies. Accordingly,



(a) Projection error at $\theta = 0°$.

(b) Projection error at $\theta = 1°$.

Figure 4.7 Projection data errors for iterated reconstructions.

Figure 4.7(a) shows a better error for the $\Delta\theta = 5°$ reconstructions than for $\Delta\theta = 1°$ at the $\theta = 0°$ projection. However, a very poor fit at the $\theta = 1°$ projection is observed for the $\Delta\theta = 5°$ reconstructions because no data was available at that angle. The good fit at $\theta = 0°$ and corresponding poor fit at $\theta = 1°$ is an example of *overfitting* when data is sparse.

Notice that the non-Gaussian nature of the projection errors persists even for the iterated reconstructions. Some method of constraining the reconstruction further seems to be necessary to prevent overfitting as well as to overcome the ambiguity.

4.2.3 Positivity constraint

The main benefit of iterative methods is that additional prior knowledge can be incorporated into the reconstruction. In particular, the knowledge that all densities in the reconstruction must be non-negative was found to be very useful and can be implemented simply by clipping intensities.

Notice how this added knowledge suppresses the checker board pattern in Figure 4.8(a) compared to Figure 4.6(a). A lot of the ambiguity is resolved, but not all of it. The edge sharpness and contrast in Figure 4.8(b) also seems much better than in Figure 4.6(b).

(a) Checkered start, $\Delta\theta = 5°$, $\sigma_{5°} = 77.96$, $\sigma_{1°} = 126$.

(b) Blank start, $\Delta\theta = 5°$, $\sigma_{5°} = 77.96$, $\sigma_{1°} = 101$.

(c) Blank start, $\Delta\theta = 1°$, $\sigma_{5°} = 81.05$, $\sigma_{1°} = 78.34$.

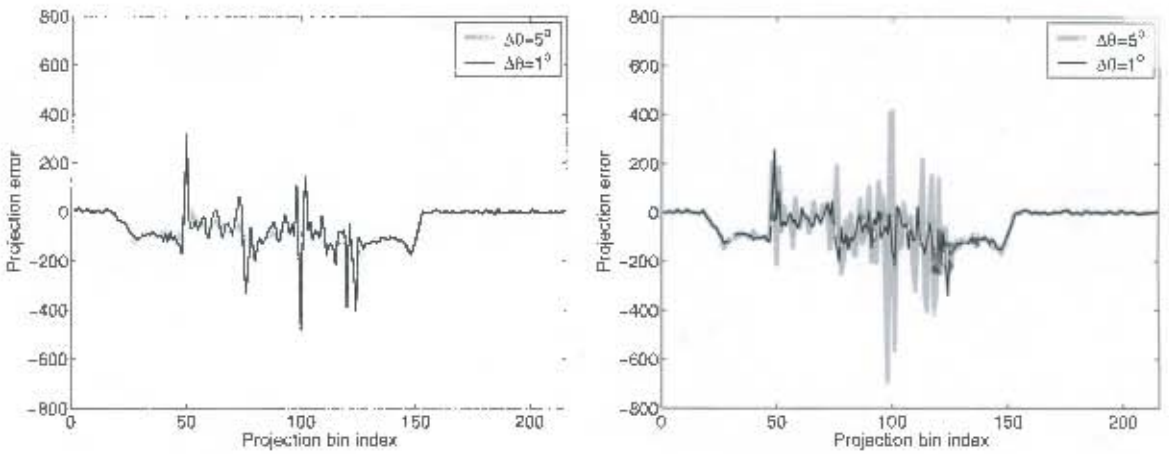
Figure 4.8 Reconstructions using positivity constraint after 100 iterations ($i_z = 225$).

Figure 4.12(a), a few pages further, shows an iterated reconstruction which is terminated prematurely after only 10 iterations using data at $\Delta\theta = 1°$ and applying the positivity constraint. Some small improvement is noticed compared to the unconstrained reconstruction in Figure 4.6(c) at this low resolution.

If iteration is continued to 100 times, as in Figure 4.8(c), the reconstruction appears slightly more noisy again or *overfitted*. This noise is not only due to the data. Negative noise (or artefacts or inconsistencies in the data) that suggests negative densities in the space domain cannot be represented under the positivity constraint. Therefore this noise is transferred to regions of positive density. Another reason for this noisy appearance of the over-iterated reconstruction, which will be addressed in more detail later, is that no data can be represented exactly on a limited resolution pixel grid even if there were no noise, artefacts or inconsistencies.

The projection errors in Figure 4.9 show the same overfitting occurring when the data is sparse as seen before in Figure 4.7. Observe though that the positivity constraint results in a much better data fit for the $\Delta\theta = 5°$ reconstruction at $\theta = 1°$.

For interest's sake, compare the high resolution iterated reconstructions in Figure 4.10 to the filtered back-projection reconstructions in Figure 3.22 and notice how the artefacts are reduced. Figure 4.11 shows sagittal and coronal views for positively iterated reconstructions that will be referred to again later in Chapter 6.



(a) Projection error at $\theta = 0°$.

(b) Projection error at $\theta = 1°$.

Figure 4.9 Projection data errors for iterated reconstructions with positivity constraint.

In conclusion of this section, it is not ideal to achieve a maximum likelihood reconstruction if there are inconsistencies, noise or a lack of data. In fact, there appears to be no sense in achieving a better data fit than $\sigma_{5°} = 81.83$ (of the best 128×128 full data reconstruction in Figure 4.12(a)), otherwise the data will be overfitted one way or another.

Therefore, the best reconstructions according to the ground truth (with minimum $\sigma_{1°}$) are biased somewhat away from the reconstruction with minimum mean square error $\sigma_{5°}$. Although the positivity constraint performs such bias very successfully, it does not seem to be constrictive enough when data is too sparse.

4.3 The properties of medical tomographic images

It is essential to explore the properties of reconstructions extensively so that an intelligent choice can be made as to how the reconstructions should be biased. The aim of this section is to determine the nature of the prior knowledge that can be used to regularize the reconstructions.

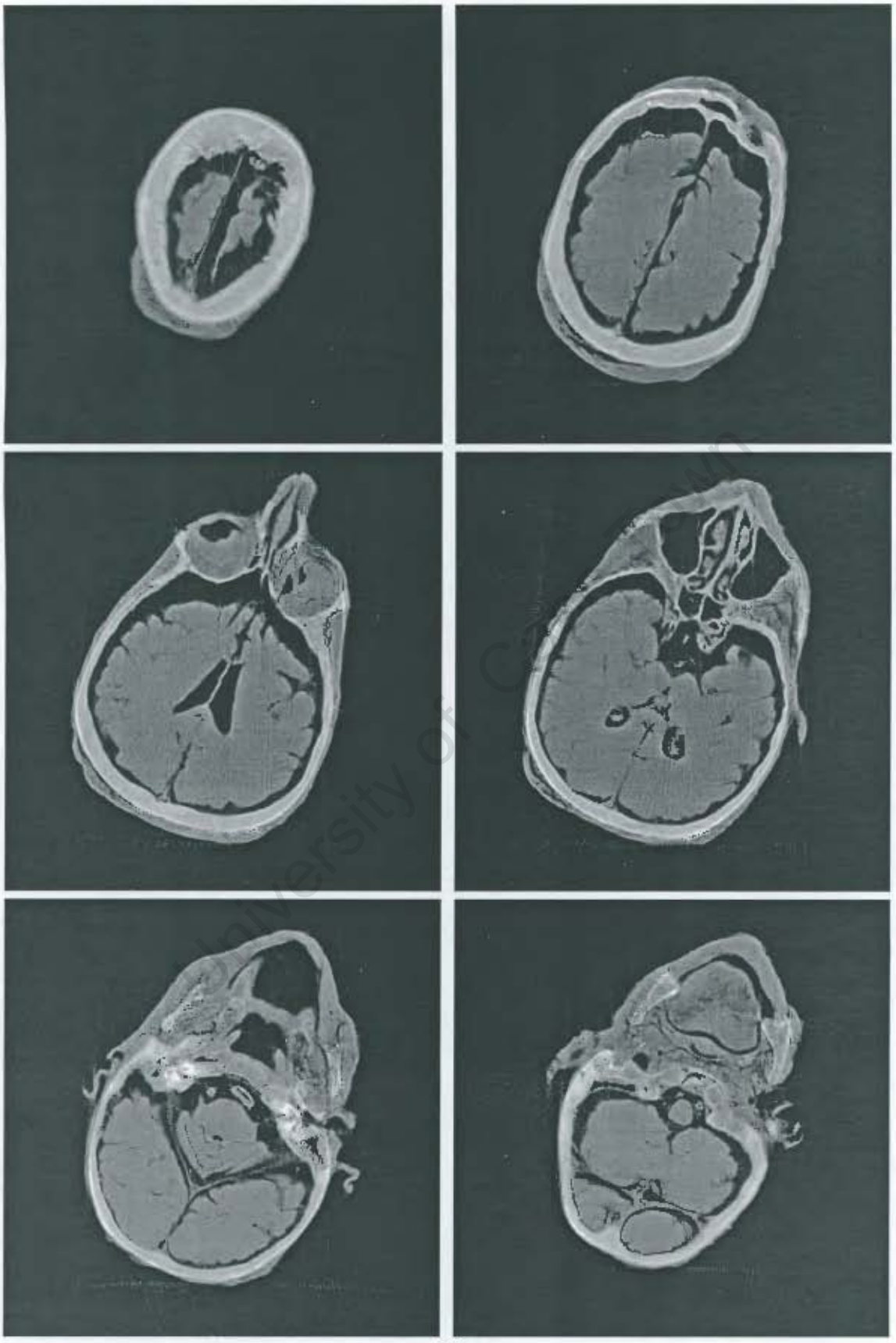


Figure 4.10 High resolution positively constrained iterated reconstructions at $i_z = \{75, 150; 225, 300; 375, 450\}$ using the complete dataset.

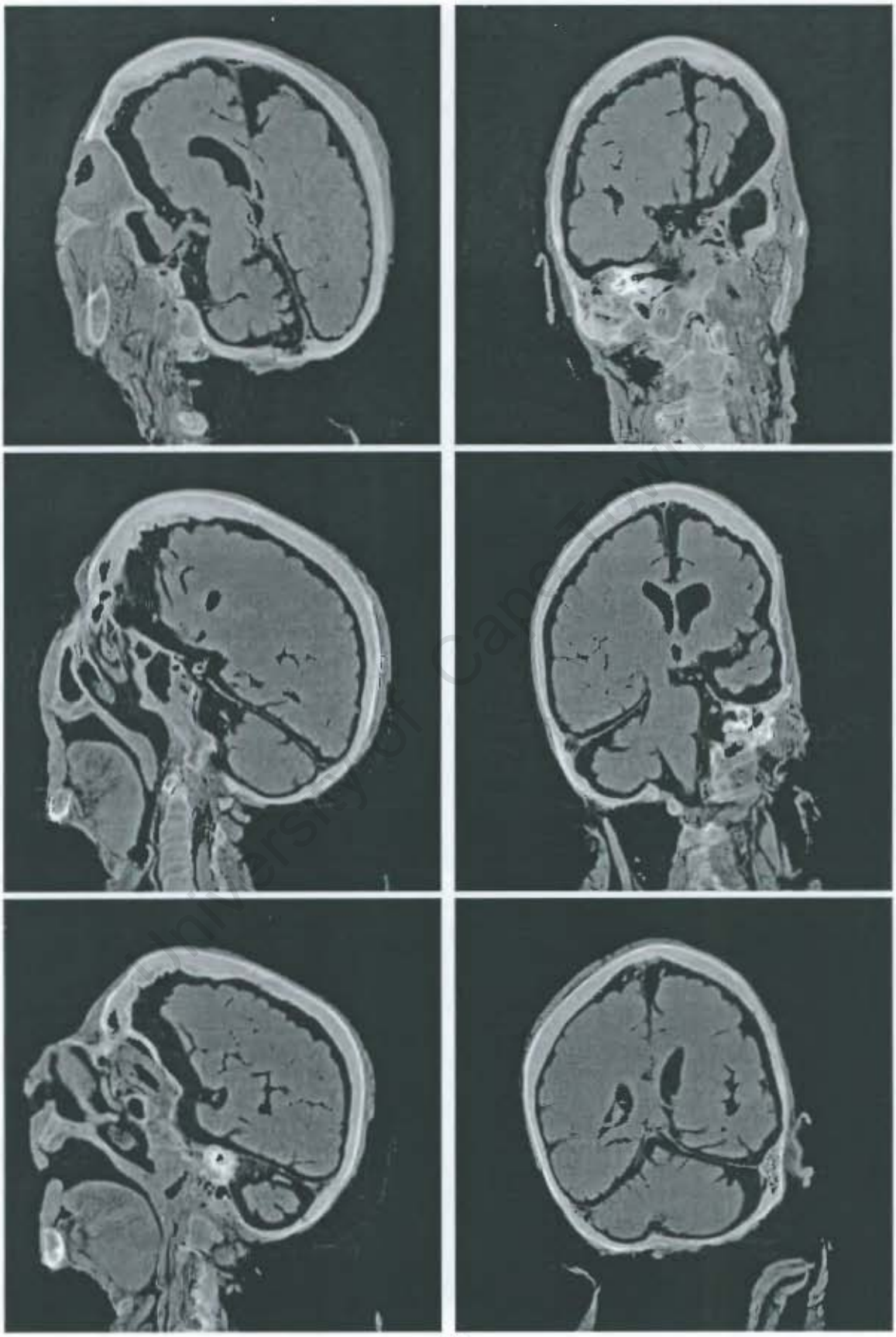


Figure 4.11 A selection of sagittal (left) and coronal (right) views at $256 \times 256 \times 256$.

(a) $\Delta\theta = 1^\circ$, $\sigma_{5^\circ} = 81.83$, $\sigma_{1^\circ} = 79.13$.

Description	Value [HU]
Air	-1000
Fat	-40 to -100
Fluid	0 to 20
Soft tissue	20 to 100
White matter	20 to 35
Grey matter	30 to 40
Acute intracranial hemorrhage	55 to 75
Bone	1000

(b) Table of Hounsfield units.

Figure 4.12 Full angle reconstruction ($i_z = 225$) and attenuation levels in Hounsfield units.

4.3.1 Hounsfield units

The level of attenuation in a tomographic reconstruction is expressed in Hounsfield units. Scaling is done linearly so that air has a value of -1000HU and water 0HU. A more comprehensive list of typical values found in such medical images is given in Figure 4.12(b) (Grossman and Yousem, 1994). Tissue attenuation values may vary slightly for different tomographic systems depending on the x-ray spectra being used (Beutel *et al.*, 2000).

From this table it is evident that discrimination amongst different types of soft tissue and fluids occurs within a very small percentage of the total range of densities. This means that tomographic reconstructions must have very good contrast properties to permit observation of different classes of tissue. Better contrast can be achieved by improving the angular resolution because the signal to noise ratio in the image improves whenever more observations are available.

It is believed that preservation chemicals (Formalin) have unfortunately affected the soft tissue attenuation properties of the human head used in this project in such a way that it becomes very difficult to identify any regions in reconstructions other than the three dominant attenuation classes – namely air, tissue and bone.

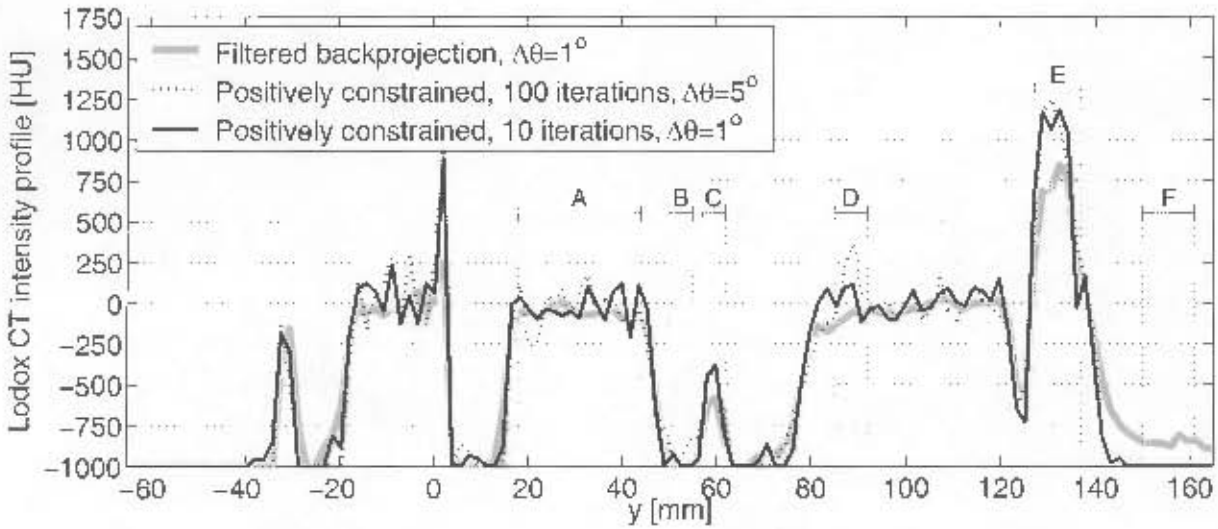
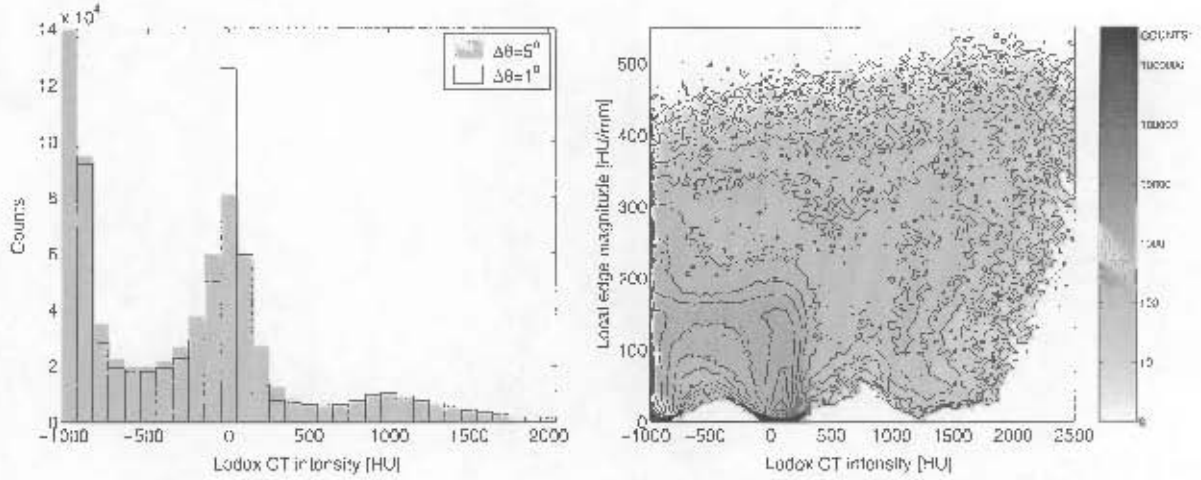


Figure 4.13 Intensity profiles through 128×128 filtered back-projection and positively constrained iterated reconstructions at $x = -44.38$mm, $i_z = 225$.

4.3.2 Noise and artefacts in Lodox CT images

Figure 4.13 compares intensity profiles of positively constrained iterated reconstructions described in Section 4.2.3 to a profile of a filtered back-projection reconstruction. In Region A it is evident that the iterated reconstructions are more noisy. Region B indicates where the $\Delta\theta = 5°$ reconstruction exhibits poorer contrast than the $\Delta\theta = 1°$ reconstruction. The positively constrained reconstructions have sharper edges than the filtered back-projection reconstruction as seen in Region C. The lack of data causes an artefact in Region D for the $\Delta\theta = 5°$ reconstruction. Much more accuracy in the attenuation levels is observed in the iterated reconstructions in Regions E and F than in the filtered back-projection reconstruction. In Region E this is believed to be due to inconsistencies and beam hardening while in Region F it is due to the slight restriction on the angular range of the dataset.

From these intensity profiles it appears that the iterated reconstructions should be constrained to be spatially smoother. It also seems that greater accuracy in the average intensity levels throughout the reconstruction is achieved through iteration.



(a) Sparse and full angle histograms.

(b) Joint histogram for full angle reconstruction.

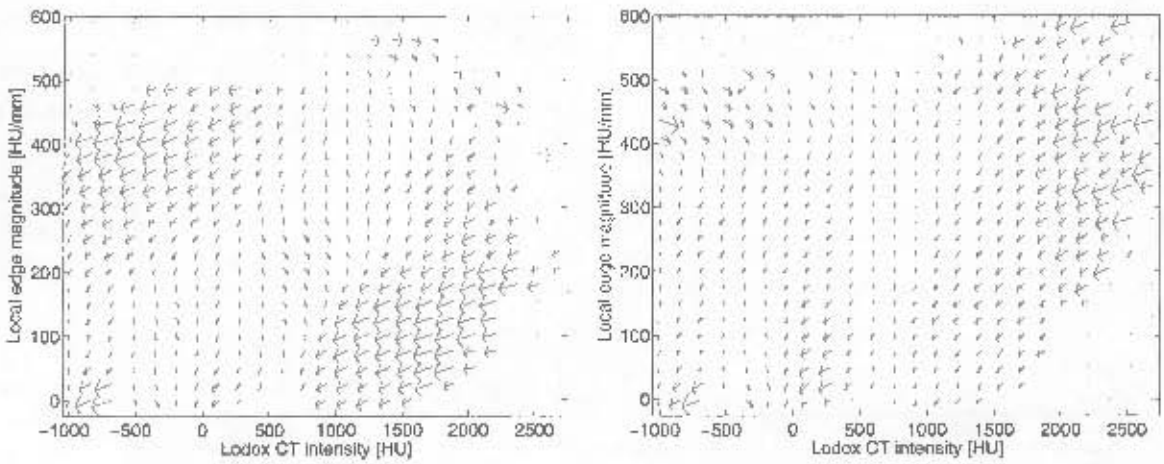
Figure 4.14 Histograms for iterated $128 \times 128 \times 128$ volumetric reconstructions with positivity constraint.

4.3.3 Dominant attenuation classes

Figure 4.14(a) shows histograms of two volumetric reconstructions using different amounts of data. In both cases there appear to be pronounced peaks around -1000, 0 and 1000 HU. When more data is available, at $\Delta\theta$ 1° intervals, the peaks seem to be narrower and more concentrated around these three attenuation levels.

A joint histogram for the $\Delta\theta = 1°$ reconstruction is displayed in Figure 4.14(b) as a contour plot. The number of occurrences is plotted as a function of both intensity and local edge magnitude. The next chapter describes the process used to estimate this edge magnitude.

The joint histogram shows that three clusters occur at -1000, 0 and to a lesser extent at 1000 HU with low local edge magnitude values. This means, as seen intuitively in the reconstruction images themselves, that there are regions having nearly flat spatial attenuation at the three dominant levels. These regions form clusters in the space domain and are separated by edges having high local edge magnitudes with a greater range of attenuation values.



(a) Effect of positivity constraint. Field comparing Figure 4.1(a) → Figure 4.8(b).

(b) Effect of completing the dataset. Field comparing Figure 4.8(b) → Figure 4.12(a).

Figure 4.15 Transformation fields for 128 × 128 × 128 reconstructions.

4.3.4 Transformation fields

A novel way of comparing reconstructions is by examining their joint transformation field (defined later in Section 5.3). Two examples are provided in Figure 4.15. These fields show how the intensities and local edge magnitudes differ between two reconstructions. An arrow is used to represent the mean change in both intensity and local edge magnitude for all voxels of similar intensity and edge magnitude from the poorer reconstruction to the better reconstruction. Although the arrow dimensions are given by the axes shown, they are scaled to avoid clutter in this figure.

Figure 4.15(a) compares a filtered back-projection reconstruction to an iterative reconstruction where the positivity constraint was applied. As shown, the most significant bias that the positivity constraint seems to impose on the reconstruction occurs at low attenuation between regions near sharp edges and also within regions of high attenuation. This is therefore where the greatest visual improvement in the reconstruction is expected to be found.

Figure 4.15(b) compares a sparse angle reconstruction to a full angle reconstruction. Notice the concentration of intensities towards the -1000, 0 and 1000 HU attenuation levels for low local edge magnitude values in both fields shown in Figure 4.15. Generalising, the application of any additional prior (a), or data (b), knowledge seems to result in regions being piecewise smoother around the three dominant attenuation levels.

The significant bias towards central attenuation values observed at high edge magnitudes in Figure 4.15(b) for high attenuation and to a lesser degree for low attenuation too, is attributed to the reduction in noise and less overfitting when more data is available. This is interpreted as the smoothing of sharp noisy edges. Transformation fields and their application to reconstructions are discussed further in the next chapter.

4.4 Regularization by Bayesian methods

Bayesian methods provide a widely used framework for augmenting data with prior knowledge to achieve optimal *a posteriori* estimates. This section describes an implementation of the theory in Section 2.5. See also (Li, 1995) for more details on Markov random field modeling.

Throughout the following discussion, the σ_{1° data fit error is considered to be a quantitative measure of the effectiveness of any regularization method applied during a reconstruction. The data at $\Delta\theta = 1^\circ$ increments, which is not available during the reconstruction, is used for evaluation purposes only and can therefore be named *test data*. As discussed before, comparing respective projections of a reconstruction to this test data yields σ_{1°. The σ_{5° data fit error is simply the square root of the familiar χ^2 ($\equiv \sigma^2$) data error of the reconstruction.

4.4.1 Summary of prior knowledge

The nature of the prior knowledge is twofold. Firstly, regions are expected to be smooth, yet separated by discontinuities. This implies that minor fluctuations in adjacent pixel intensities should be attenuated while abrupt discontinuities should be permitted. Secondly, intensities should be biased closer to the dominant intensity levels. Critically speaking, intensities closer to a dominant level should be biased more towards that level than intensities that are further away from that level (which could perhaps be closer to another dominant level).

4.4.2 Adaptive potential functions and clique potentials

A suitable form of adaptive potential function $g_\gamma(\eta)$ for expressing the first part of the prior knowledge above, as in Equation 2.58, is

$$g_\gamma(\eta) = -\frac{1}{1+\frac{\eta^2}{\gamma^2}} \text{ with } g'_\gamma(\eta) - 2\eta g^2_\gamma(\eta)/\gamma^2 \tag{4.2}$$

where η is the intensity gradient, or rather intensity difference $\eta = f_i - f_{i'}$, of a pair clique in a second order neighbourhood system (8 neighbours per site). The parameter γ controls the permissible η edge magnitude. This form of adaptive potential function is prefered over a Gaussian-based one for its computational advantage. For an isolated cross-section, the *discontinuity adaptive* clique potential is

$$V_{\text{DA-MRF},i}(f) = \sum_{i'=1}^{8} g_\gamma(\eta), \quad \eta = f_i - f_{i'}. \tag{4.3}$$

The second part of the prior knowledge can be expressed in a similar way by redefining the variable η to an intensity level error as $\eta_{\text{I-MRF}} = f_i - \mathrm{f_c}$ where $\mathrm{f_c}$ is one of the three dominant intensity levels, so that

$$V_{\text{I-MRF},i}(f) = \sum_{c=1}^{3} g_\gamma(\eta_{\text{I-MRF}}), \quad \eta_{\text{I-MRF}} = f_i - \mathrm{f_c} \tag{4.4}$$

is the *intensity level* clique potential. In practice a fourth intensity level at $\mathrm{f_4} = 2000\mathrm{HU}$ is introduced to better represent high density bone edges. Note from a coding point of view that only the closest intensity level makes a significant contribution to the potential for a particular pixel element f_i, so summation is not necessary if only the closest level is identified and used.

A third potential function can be defined as

$$V_{\text{DAI-MRF},i}(f) = V_{\text{I-MRF},i}(f) + \alpha V_{\text{DA-MRF},i}(f) \tag{4.5}$$

which is the weighted combination of the *discontinuity adaptive* and *intensity level* (Markov Random Field) clique potentials. The next subsection describes how each of these potential functions are used within the Bayesian framework to produce reconstructions from data projections available at $\Delta\theta = 5^\circ$ intervals only.

4.4.3 Maximizing the posterior

Approximating the noise and inconsistencies in the data to Gaussian statistics, the *a posteriori* solution occurs at

$$f^{(MAP)} = \arg_f \max \ln P(f|p) = \arg_f \min \left[\beta U(f) - \chi^2\right] \tag{4.6}$$

where $U(f) = \sum V_i(f)$ is the energy function which sums all possible clique potentials as described in Section 2.5. $V_i(f)$ is substituted with $V_{\text{DA-MRF},i}(f)$, $V_{\text{I-MRF},i}(f)$ and $V_{\text{DAI-MRF},i}(f)$ to produce DA-MRF, I-MRF and DAI-MRF reconstructions respectively.

This posterior solution is obtained using the conjugate gradient search method starting from a blank image. A range of solutions were determined by varying the parameters β and γ (and also α in the case of DAI-MRF reconstructions).

Figure 4.16 shows the data fit errors for a number of reconstructions using the clique potential form given by Equation 4.3 for different values of β and γ. The σ_{1° errors indicate that the optimal values for β and γ are 0.3 and 150 respectively. Clearly $\beta = 0$ will result in the best σ_{5° error which is not desired. Note that the *ground truth* σ_{1° $\beta \times \gamma$ surface appears somewhat noisy, implying difficulty in its minimization and also some sensitivity to the starting point of reconstruction.

Figure 4.17 shows reconstructions using optimal parameters for three different forms of clique potentials. The DA-MRF reconstruction is biased towards piecewise smoothness by Equation 4.3 while the I-MRF reconstruction (with optimally chosen parameters $\beta = 0.9$ and $\gamma = 150$) is biased towards dominant intensity levels by Equation 4.5. Lastly, the DAI-MRF reconstruction (with optimally chosen parameters $\alpha = 0.3$, $\beta = 0.2$ and $\gamma = 150$) sums both potentials and thereby includes both aspects of the prior knowledge which shows some further improvement in the reconstruction. All these reconstructions appear somewhat unnatural due to the simple nature of the prior knowledge.

When only $\Delta\theta = 5^\circ$ data is available, this form of regularization brings the *ground truth* σ_{1° error down to 91 from 101 for positively constrained reconstructions and from 144 for unconstrained reconstructions. Quite obviously, however, it is much worse than for a complete data set for which $\sigma_{1^\circ} = 79$.

$\gamma\backslash\beta$	0.1	0.2	0.3	0.4	0.5	0.6
70	79.0	80.3	81.6	82.7	84.1	85.4
110	78.9	79.7	80.7	81.7	82.8	83.9
150	78.6	79.3	**80.1**	80.9	81.9	82.8
190	78.8	79.2	79.7	80.6	81.2	82.1
230	79.7	79.3	79.4	80.1	80.6	81.3

(a) σ_{5° fit errors.

$\gamma\backslash\beta$	0.1	0.2	0.3	0.4	0.5	0.6
70	95.7	93.7	92.8	93.2	94.1	95.0
110	96.2	93.3	92.7	92.7	93.4	94.1
150	94.8	92.8	**92.4**	92.4	92.8	93.6
190	98.4	95.1	93.0	93.8	93.5	94.2
230	102	96.9	93.9	93.9	93.8	93.8

(b) σ_{1° fit errors.

Figure 4.16 Data fit standard deviations for various DA-MRF reconstructions showing that parameters $\beta = 0.3$, $\gamma = 150$ are optimal.

(a) DA-MRF, $\sigma_{5^\circ} = 80.12$, $\sigma_{1^\circ} = 92.41$. $\beta = 0.3$, $\gamma = 150$.

(b) I-MRF, $\sigma_{5^\circ} = 79.21$, $\sigma_{1^\circ} = 93.48$. $\beta = 0.9$, $\gamma = 150$.

(c) DAI-MRF, $\sigma_{5^\circ} = 79.52$, $\sigma_{1^\circ} = 91.19$. $\alpha = 0.33$, $\beta = 0.2$, $\gamma = 150$.

Figure 4.17 Regularized reconstructions ($\Delta\theta = 5^\circ$, $d_z = 225$).

4.5 Summary

This chapter presented a method that can be used to update a reconstruction so that its projections fit the data well. The subsequent problem of overfitting calls for additional constraints to supplement (and to some extent even contradict) the projection data.

Prior knowledge is identified by investigating properties of complete dataset reconstructions. Bayesian methods were used to demonstrate that the application of prior knowledge bias improves reconstructions and compensates for ambiguities that exist when less data is available.

The next chapter further explores transformation maps as a method for automatically acquiring and applying more complex prior knowledge.

CHAPTER 5

Transformation maps

In this chapter, transformation maps are presented as an alternative to the Bayesian method for regularizing reconstructions using prior knowledge. The principle advantage of a transformation map is that it provides a way of determining nuances of the prior knowledge quantitatively from examples.

A transformation map defines, in a feature space, how any particular input possibility should be corrected in order to attain a corresponding desired output. If carefully constructed, such a map could translate many different input images to a single regularised output image.

The similarity of neural networks to transformation maps is expressed first. Two examples are provided before the method is introduced formally. Then the transformation map framework is applied to sparse full angle tomography reconstruction.

5.1 Introduction

Neural networks in the context of soft computing (Jang *et al.*, 1996) are capable of solving a wide range of problems. These machine intelligence strategies are mentioned in this introductory section because transformation maps can perform, by approximation, an identical task even though they are implemented entirely differently.

Neural network enthusiasts may for this reason benefit from an alternate tool in situations where transformation maps may be more suitable. Certainly in some cases neural networks may be more suitable. The following overview hopes to put transformation maps into perspective and point out the defining differences of these methods.

5.1.1 Neural networks

Neural networks are commonly used as a *black box* tool for classification and function approximation. There are many types of networks with differing combinations of neurons forming their internal structures. A neuron associates a given stimulus with a given response usually by performing a nonlinear operation, such as thresholding, on the weighted sum of its inputs.

A number of examples, which demonstrate how given inputs are associated with desired outputs, are used during a training phase to regress values for the unspecified weights. Iterative back-propagation, stochastic search methods, or pseudo-matrix inversion in the case of random neural networks, are possible ways of determining the neuron input weights. Training a neural network is a very computationally expensive process for large datasets, and classification time is determined by the complexity of the network's structure.

If the dimensionality of the input space is large (such as an image for the application of face recognition) the number of dimensions is reduced by extracting feature vectors. The feature vectors are then fed as input to the neural network.

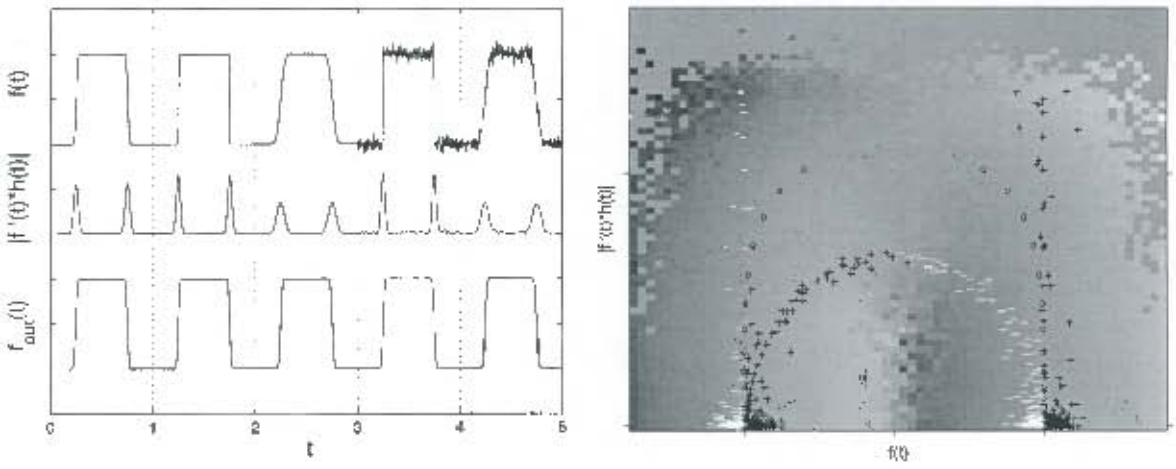
In essence, a neural network performs a mapping from an input space to an output space. It performs this mapping using a model generally defined by unintelligible weights and the architecture of neurons.

5.1.2 Transformation maps

A transformation map can be considered to be simply an explicit illustration of the mappings performed by *implementation technologies* such as neural networks. This implies an exhaustive representation of the mapping of all input possibilities to the desired outputs over the range of allowable inputs. Stated differently, a neural network is a possible mechanism of realizing a transformation map.

5.2 Examples

This section introduces the transformation map method by two examples. Firstly a simple one-dimensional problem is solved, and then an image is denoised. A secondary goal of this presentation is to prove that transformation maps have the ability to store edge properties, since this capability is exploited in the next chapter.



(a) Signals.

(b) Transformation map.

Figure 5.1 Application of a transformation map to correct the pulse shapes in a signal.

5.2.1 Signal restoration example

Consider the problem of restoring the shape of an *input* pulse, $f(t)$, which may have been distorted slightly by noise, blurring or edge sharpening. The leftmost pulse in Figure 5.1(a), $f_{\text{target}}(t) = f(t)$ for $0 < t < 1$, illustrates the desired *output* pulse shape and the pulses $f(t)$ for $1 < t < 2$ (sharpened), $2 < t < 3$ (blurred), $3 < t < 4$ (sharpened with noise) and $4 < t < 5$ (blurred with noise) are examples of possible input pulses.

A derived signal, or *feature*, $\mathcal{H}_1 \equiv |f'(t) * h(t)|$ is determined which indicates the local mean edge magnitude. Here $*h(t)$ represents convolution with a Gaussian blurring kernel and the prime denotes derivative. This signal is shown in Figure 5.1(a).

The error difference, $f_{\text{target}}(t) - f(t)$, between the (appropriately time-shifted) ideal pulse shape and the input signal is calculated for each value of t and is plotted in Figure 5.1(b) on a plane invariant to t as follows: for each value of t at $\mathcal{H}_1 \equiv |f'(t) * h(t)|$ and $\mathcal{H}_2 \equiv f(t)$, a $+$, $-$ or 0 is plotted depending on whether the input signal should be increased, decreased or remain unchanged respectively at that value of t to recover $f_{\text{target}}(t)$ from $f(t)$.

The plane, or map, can be populated by generating many more examples varying the amount of sharpening, blurring and noise levels. This yields the underlying intensity surface shown in

the figure. The average correction to the input signal is represented by a pixel intensity in this transformation map which is independent of t.

Figure 5.1(a) also shows the result $f_{\text{out}}(t)$ of restoring the input signal $f(t)$. The correction value is simply read off the map indexed by $(i, j) = (\mathcal{H}_1, \mathcal{H}_2)$ and added to $f(t)$.

5.2.2 Image restoration example

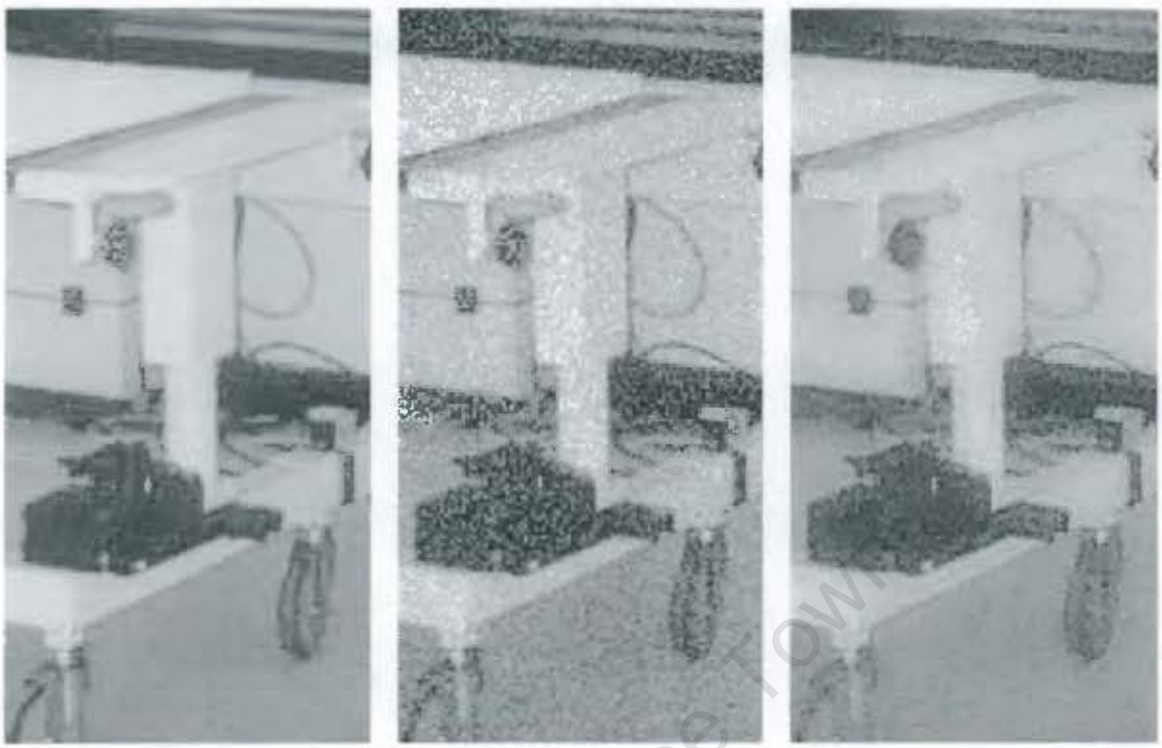
Figure 5.2 presents an example where an input image $f(x, y)$, shown in Figure 5.2(b), is denoised while preserving edges using the transformation map method. The input image is obtained by adding normally distributed noise with a standard deviation of $\sigma = 15$ intensity units (out of 256 intensity levels) to the original image $f_{\text{target}}(x, y)$, shown in Figure 5.2(a). After denoising, the standard deviation of the difference between the output image $f_{\text{out}}(x, y)$, shown in Figure 5.2(c), and the original is $\sigma = 8.6$ intensity units.

Figure 5.2(a) is actually cropped from Figure 1.2. Other regions, which exclude the cropped regions in Figure 5.2, are used in generating the transformation map which is shown in Figure 5.3. However, results did not change significantly when the entire image (Figure 1.2) was used to generate the transformation map.

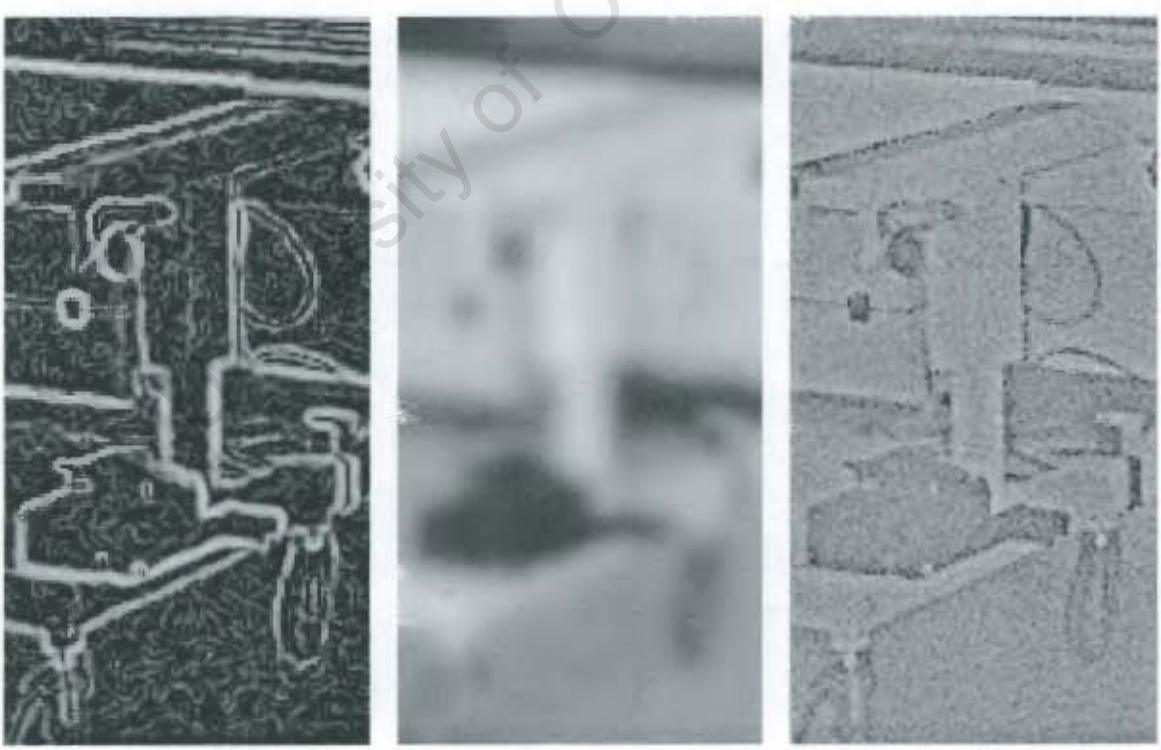
The first transformation map dimension or *feature*, illustrated in Figure 5.2(d), is the local edge magnitude given by $\mathcal{H}_1 \equiv |f'(x, y) * h(x, y)|$ where $*h(x, y)$ represents convolution by a Gaussian bivariate blurring kernel (with a standard deviation of 0.5, discussed later within this subsection) and $f'(x, y)$ is the intensity gradient at (x, y).

These calculations are performed by multiplication with the appropriate filters in the frequency domain as discussed in the next section (see also lines 51-71 of the Matlab code in Figure 5.4). Choosing local edge magnitude as a dimension makes sense because regions near edges are expected to be treated differently from regions with a low intensity gradient magnitude.

As a second dimension $\mathcal{H}_2 \equiv f(x, y) - f(x, y) * h_2(x, y)$, illustrated in Figure 5.2(f), is chosen where $*h_2(x, y)$ indicates convolution with another Gaussian bivariate blurring kernel (with a standard deviation of 1.5). For a given edge-free region, this measure closely resembles the explicit noise which is to be subtracted out of the degraded image $f(x, y)$. Hence a near-linear increase in intensity is observed in Figure 5.3 for low $|f'(x, y) * h(x, y)|$ (bottom) as $f(x, y) - f(x, y) * h_2(x, y)$ increases around zero (middle) from left to right.



(a) Original $f_{\text{target}}(x, y)$, $\sigma = 0$ (b) Input $f(x, y)$, $\sigma = 15.0$ (c) Output $f_{\text{out}}(x, y)$, $\sigma = 8.6$



(d) $|f'(x, y) * h(x, y)|$ (e) $f(x, y) * h_2(x, y)$ (f) $f(x, y) - f(x, y) * h_2(x, y)$

Figure 5.2 Denoising example.

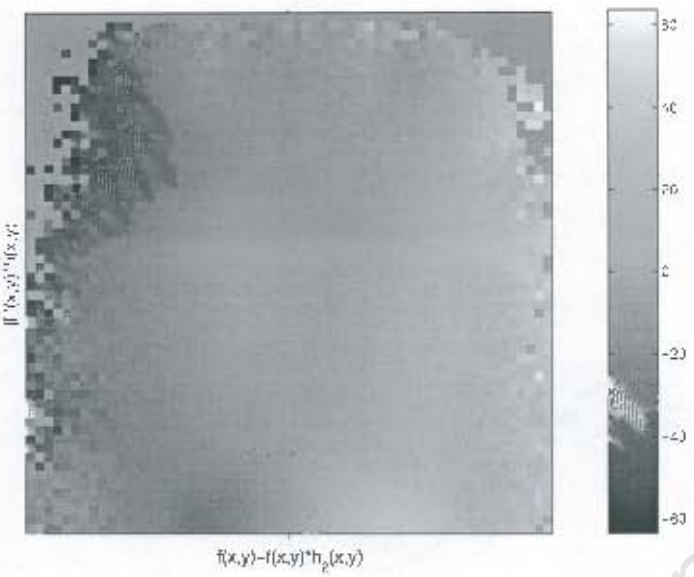


Figure 5.3 Transformation map for denoising.

The intensity of a pixel at cell $(i, j) = (\mathcal{H}_1, \mathcal{H}_2)$ in Figure 5.3 is the average intensity correction, $\langle f_{\text{target}}(x, y) - f(x, y) \rangle$, of several occurrences of pixels in many different noisy images compared to the original or target image all corresponding to the same value for (i, j). (Indices are quantized in this implementation.) Each cell (i, j) not only has an associated mean but also a variance due to conflict in intensity correction values $f_{\text{target}}(x, y) - f(x, y)$ for different examples corresponding to the same cell in the feature space $\mathcal{H}_1 \times \mathcal{H}_2$.

The Gaussian blurring kernel standard deviations used in $h(x, y)$ and $h_2(x, y)$ for this example are chosen experimentally to produce a subjectively good-looking result with a $\sigma = 8.6$ fit to the ground truth $f_{\text{target}}(x, y)$. However, if values are chosen systematically as described in a later section of this chapter, so that the collective variance of the conflicting examples in the map is minimized, then a restoration result with $\sigma = 6.9$ can be achieved. Although the result is numerically superior, it is not necessarily better subjectively. Nevertheless, this technique provides reasonable guidance in choosing values for unspecified parameters.

For interest's sake, on the same restoration problem a 3×3 median filter produces an image with a somewhat 'blotchy' appearance even though its standard deviation is $\sigma = 7.8$ with respect to $f_{\text{target}}(x, y)$. Also, by simple Gaussian blurring, a result with $\sigma = 7.2$ can be achieved by choosing an optimal kernel size (that minimizes σ).

```
randn('state',0);                                          %initialize random number generator
img=double(imread('lodoxmachine.bmp','bmp'));              %load image
img512=img(1:512,1:512);                                   %crop image
sz=64;                                                     %transformation map size
tx=zeros(sz,sz);txv=tx;txn=tx;                             %initialize transformation map
dblfact=0.5;                                               %hyper parameter 1
bblfact=1.5;                                               %hyper parameter 2
dscale=max(max(abs(GetDerivative(img512,dblfact))))*1.2;   %fixed scale for first feature dimension
bscale=max(max(abs(GetBlur(img512,bblfact))))*1.2;         %fixed scale for second feature dimension
for nnoise=1:100                                           %generate transformation map
  noisimg=img512+15*randn(512,512);                        %creates noisy source image example
  dimg=abs(GetDerivative(noisimg,dblfact))/dscale;         %get first feature dimension
  bimg=GetBlur(noisimg,bblfact);                           %get blurred image
  beimg=((noisimg-bimg))/bscale+0.5;                       %determine second feature dimension
  eimg=noisimg-img512;                                     %determine error image
  for iy=1:512
    for ix=1:256                                           %uses only left hand side of image
      id=floor(dimg(iy,ix)*(sz-1))+1;                      %index first feature dimension
      ib=floor(beimg(iy,ix)*(sz-1))+1;                     %index second feature dimension
      if (id>=1&&id<=sz&&ib>=1&&ib<=sz)                    %populate transformation map, variance and counter
        tx(id,ib)=tx(id,ib)+eimg(iy,ix);
        txv(id,ib)=txv(id,ib)+eimg(iy,ix).^2;
        txn(id,ib)=txn(id,ib)+1;
      end
    end
  end
end
for id=1:sz                                                %normalize transformation map
  for ib=1:sz
    if(txn(id,ib)>0)
      tx(id,ib)=tx(id,ib)/txn(id,ib);
      txv(id,ib)=txv(id,ib)/txn(id,ib);
    end
  end
end
noisimg=img512+15*randn(512,512);                          %generate example to denoise
dimg=abs(GetDerivative(noisimg,dblfact))/dscale;           %get first feature dimension
bimg=GetBlur(noisimg,bblfact);                             %get blurred image
beimg=((noisimg-bimg))/bscale+0.5;                         %get second feature dimension
nimg=noisimg;
for iy=1:512
  for ix=1:512                                             %denoise image using transformation map
    id=floor(dimg(iy,lx)*(sz-1))+1;
    ib=floor(beimg(iy,ix)*(sz-1))+1;
    if (id>=1&&id<=sz&&ib>=1&&ib<=sz)
      nimg(iy,ix)=noisimg(iy,lx) -tx(id,ib);
    end
  end
end
std(reshape(noisimg-img512,512*512,1))                     %display RMS error of noisy image
std(reshape(nimg-img512,512*512,1))                        %display RMS error of transformation map
std(reshape(medfilt2(noisimg,[3 3])-img512,512*512,1))     %display RMS error of median filter result
sqrt(mean(mean(txv-tx.*tx)))                               %display transformation map standard deviation
```

```
function nim=GetDerivative(img,var)
nx=size(img,1);
nx2=nx/2;
r2=zeros(nx,nx);
ang=r2;
for x=1:nx2
  for y=1:nx2
    r2(x,y)=((x-1).^2+(y-1).^2);
    r2(nx-x+1,nx-y+1)=(x.^2+y.^2);
    r2(x,nx-y+1)=((x-1).^2+y.^2);
    r2(nx-x+1,y)=(x.^2+(y-1).^2);
    ang(x,y)=atan2(y-1,x-1);
    ang(nx-x+1,nx-y+1)=atan2(-y,-x);
    ang(x,nx-y+1)=atan2(-y,x-1);
    ang(nx-x+1,y)=atan2(y-1,-x);
  end
end
derfilt=sqrt(r2).*exp(j.*ang);
nim=abs(ifft2(fft2(img).*derfilt.*exp(-0.5*r2/nx/pi.*(var.^2))));
```

```
function nim=GetBlur(img,var)
nx=size(img,1);
nx2=nx/2;
mask=zeros(nx,nx);
for x=1:nx2
  for y=1:nx2
    r2(x,y)=((x-1).^2+(y-1).^2);
    r2(nx-x+1,nx-y+1)=(x.^2+y.^2);
    r2(x,nx-y+1)=((x-1).^2+y.^2);
    r2(nx-x+1,y)=(x.^2+(y-1).^2);
  end
end
nim=real(ifft2(fft2(img).*exp(-0.5*r2/nx/pi.*(var.^2))));
```

Figure 5.4 Matlab code for image denoising example.

Figure 5.4 provides the Matlab source code for implementing the image restoration problem described in this subsection. As indicated, the map is populated first using many noisy examples, then the map is normalized by dividing by the number of examples that occur for every cell in the map. Finally the map is used to denoise a particular input image.

5.3 Definitions and implementation details

Blurring, indicated with $*h(x,y)$, is performed by multiplication in the two-dimensional frequency domain with the Fourier transform of the Gaussian kernel

$$h(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}} \tag{5.1}$$

that has some standard deviation σ. Derivative action, $f'(x,y)$, on image $f(x,y)$ is calculated in conjunction with blurring in the Fourier domain (simultaneously) by multiplication with

$$j \cdot (u + j \cdot v) \tag{5.2}$$

where u and v are the frequency domain variables and $j = \sqrt{-1}$. This is a two-dimensional equivalent of the familiar differentiation by multiplication with $j\omega$ in the one-dimensional case, where ω is the frequency variable.

In this thesis, a *transformation field*, $\vec{\mathcal{T}} \equiv \vec{\mathcal{T}}(f_{\text{target}}(\mathbf{x}), f(\mathbf{x}))$, is a vector function that is defined in a *feature space* $\mathcal{H}$ as the average change in respective feature space dimensions $\mathcal{H}_1$, $\mathcal{H}_2$, ..., $\mathcal{H}_\text{N}$ between all source $f \equiv f(\mathbf{x})$ and target $f_{\text{target}} \equiv f_{\text{target}}(\mathbf{x})$ example pairs:

$$\vec{\mathcal{T}} = (\langle \mathcal{H}_1(f_{\text{target}}) - \mathcal{H}_1(f) \rangle, \langle \mathcal{H}_2(f_{\text{target}}) - \mathcal{H}_2(f) \rangle, ..., \langle \mathcal{H}_\text{N}(f_{\text{target}}) - \mathcal{H}_\text{N}(f) \rangle). \tag{5.3}$$

The angled brackets indicate averaging over examples that coincide in the feature space domain. Input f and target f_{target} functions are co-aligned by means of the space (or time) independent variable $\mathbf{x}$ (in the two-dimensional case $\mathbf{x} \equiv (x,y)$). The N-dimensional feature space $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times ... \times \mathcal{H}_\text{N}$ is constructed by functionals $\mathcal{H}_1 \equiv \mathcal{H}_1(f(\mathbf{x}))$ through $\mathcal{H}_\text{N} \equiv \mathcal{H}_\text{N}(f(\mathbf{x}))$ which serve as the domain of the feature space. This domain is not directly dependent on $f_{\text{target}}(\mathbf{x})$ nor on $\mathbf{x}$.

A *transformation map* $\mathcal{T}$ is a scalar function of all example pairs that exist in the feature space $\mathcal{H}(f(\mathbf{x}))$ which is derived from $f(\mathbf{x})$ but not $f_{\text{target}}(\mathbf{x})$ and is not dependent on $\mathbf{x}$:

$$\mathcal{T}(f_{\text{target}}, f) = \langle f_{\text{target}}(\mathbf{x}) - f(\mathbf{x}) \rangle \tag{5.4}$$

and the *map variance* is defined as

$$\sigma^2_{\text{map}} = \left\langle (f_{\text{target}}(\mathbf{x}) - f(\mathbf{x}))^2 \right\rangle - \mathcal{T}^2 \tag{5.5}$$

in the feature space. The *total map variance* $\sigma^2_{\text{totalmap}} = \overline{\sigma^2_{\text{map}}}$ is a scalar quantity which measures the degree of conflict in example pairs. These definitions will be elaborated on later in this chapter.

A transformation map is typically implemented using a multidimensional array. In order to create a transformation map from examples, three of these arrays exist in memory. One is used to accumulate all the examples $f_{\text{target}}(\mathbf{x}) - f(\mathbf{x})$ at their corresponding array indices in the array. These indices are calculated from appropriately quantized and scaled feature dimension values $\mathcal{H}_1(f(\mathbf{x}))$ to $\mathcal{H}_N(f(\mathbf{x}))$. A second array is used to tally the number of examples accumulated in each cell of the first array. A third array keeps record of the squared quantity $(f_{\text{target}}(\mathbf{x}) - f(\mathbf{x}))^2$ for the calculation of the map variance.

The transformation map is determined by elementwise division of the accumulated example array by the example count array. Note, importantly, that it is possible to populate empty cells in the transformation map by blurring each of these arrays (identically) prior to this division.

After its creation, the *transformation map bias* $\mathcal{T}(f)$ is additively applied to a source image $f(\mathbf{x})$ to produce a resultant output image $f_{\text{out}}(\mathbf{x})$, i.e.

$$f_{\text{out}} = f + \lambda \mathcal{T}(f) \tag{5.6}$$

where $\lambda = 1$ for non-iterative methods ($\lambda < 1$ typically for iterative methods which are discussed later). The bias $\mathcal{T}(f)$ is simply read off the transformation map, indexed using the appropriately quantized and scaled feature dimension values $\mathcal{H}_1(f(\mathbf{x}))$ to $\mathcal{H}_N(f(\mathbf{x}))$. In other words, each image voxel has its own set of feature indices (one per feature dimension) to index the transformation map, and these indices must be (re)calculated from features of $f(\mathbf{x})$ before the transformation map can be indexed.

The method of scaling and quantization of the feature indices must be identical to when the map was created. The scaling must obviously be such that the feature space which the transformation map occupies captures all relevant examples. This can be done by finding minimum and maximum values of feature dimensions over all typical examples. The quantization and hence size of the map could easily be limited by physical memory and is chosen to be either 512^2, 64^3 or 32^4.

5.4 Application to tomography

Tomography is a very suitable problem to be explored in conjunction with transformation maps as a method for extracting and applying prior knowledge. Both iterative and non-iterative reconstruction techniques are discussed.

5.4.1 Randomness of tomographic artefacts

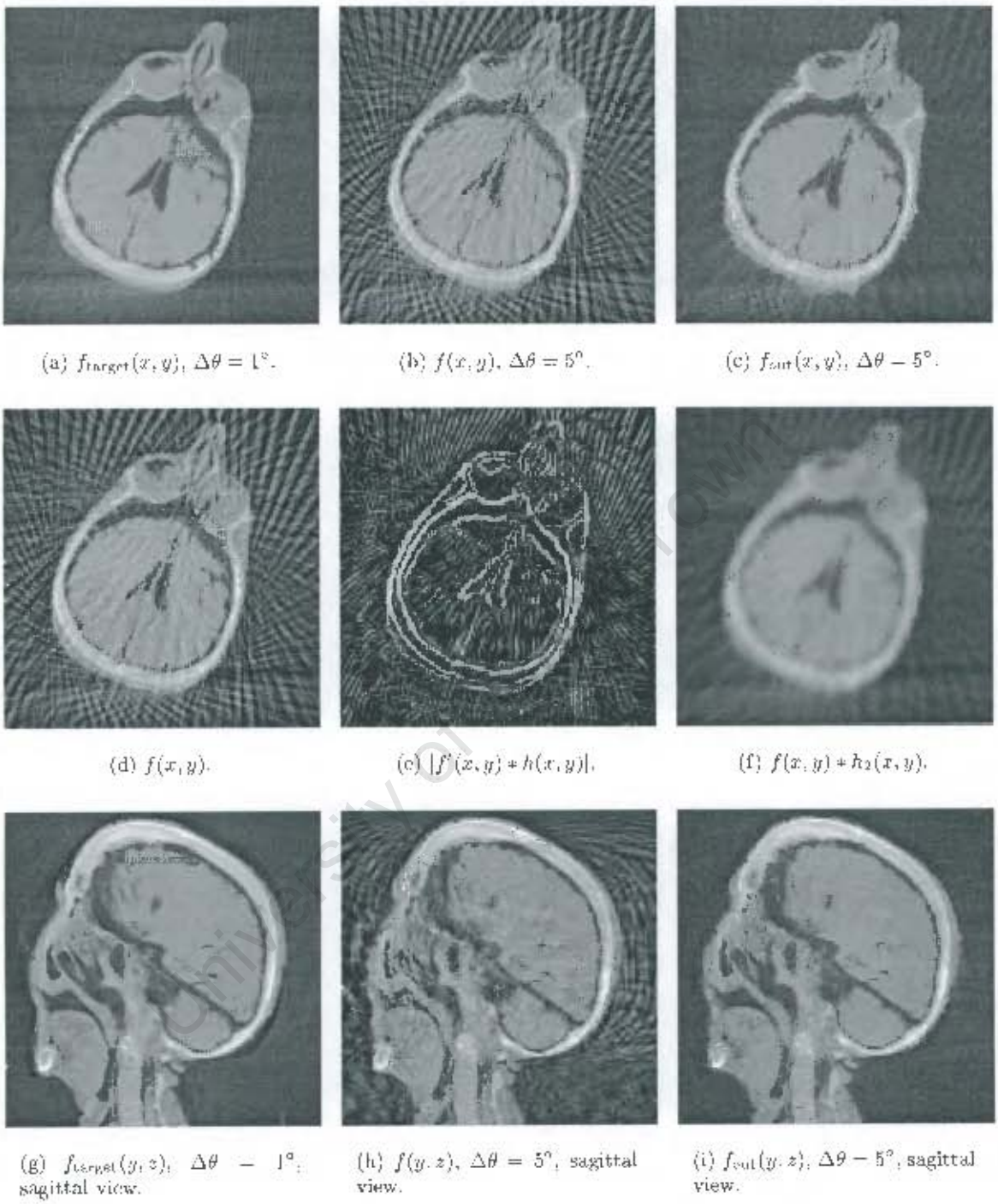
Tomographic angular aliasing artefacts (the steaks seen in Figure 5.5(b) compared to Figure 5.5(a), and in a sagittal view Figure 5.5(h) compared to Figure 5.5(g)) are known to contain a lot of information and are not random (Kak and Slaney, 1988). In fact, it is shown in (de Villiers, 1999) that samples of these artefacts along the reconstruction image border alone can be used to reconstruct sparse angle tomographic images surprisingly well if the geometry of the projections is known *a priori*. Sparse angle tomographic artefacts are substantially interchangeable with the measured data that created them. Likewise, being able to remove artefacts correcly is equivalent to knowing *un*measured projections exactly.

Even though these artefacts are not of a random origin, intensities in reconstructions are effected in a seemingly random manner if only local information (a neighbourhood of intensity values) is available. For this reason any denoising filter (for example: Gaussian kernel blurring, the median filter, or the more effective Bayesian method described in Section 4.4) can be used to suppress these artefacts to some extent.

Instead of trying to predict artefacts or noise directly, most methods (including the transformation map method) use knowledge or assumptions about the underlying image to estimate a better image given a corrupted image. If the artefacts are overwhelmingly destructive such that no features of the underlying image can be recognized in the corrupted image for some region, then it is not (directly) possible to predict anything reliably in that region (as discussed in the next chapter).

In other words, it is only possible to suppress artefacts that are small relative to (or are distinguishable from) underlying image features so that the underlying image can still be recognized from its prevailing features in the corrupted image. The features shown in Figures 5.5(d), 5.5(e) and 5.5(f) are used in Section 5.4.2 to predict Figure 5.5(c) from Figure 5.5(b).

Refer to Section 4.3 for a discussion on properties of medical tomographic reconstructions since artefacts are simply any deviation from these properties.



(a) $f_{target}(x, y)$, $\Delta\theta = 1^\circ$.

(b) $f(x, y)$, $\Delta\theta = 5^\circ$.

(c) $f_{out}(x, y)$, $\Delta\theta = 5^\circ$.

(d) $f(x, y)$.

(e) $|f(x, y) * h(x, y)|$.

(f) $f(x, y) * h_2(x, y)$.

(g) $f_{target}(y, z)$, $\Delta\theta = 1^\circ$, sagittal view.

(h) $f(y, z)$, $\Delta\theta = 5^\circ$, sagittal view.

(i) $f_{out}(y, z)$, $\Delta\theta = 5^\circ$, sagittal view.

Figure 5.5 Removal of angular aliasing artefacts from filtered back-projection reconstructions with no iteration using the transformation field method.

5.4.2 Removal of angular aliasing artefacts in single step reconstruction

Direct inversion algorithms, such as the filtered back-projection method, are capable of reconstructing high resolution cross-sections rapidly without providing any means of applying prior knowledge. The transformation map method can, however, be used to apply prior knowledge subsequent to reconstruction.

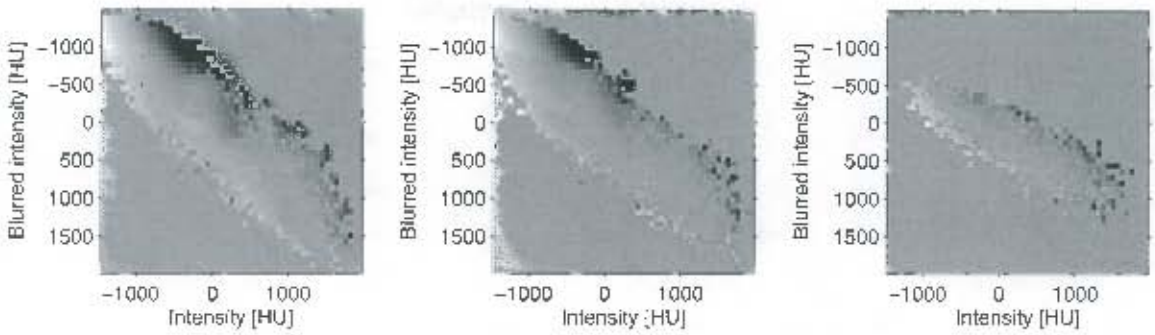
Figure 5.5 shows how the aliasing artefacts caused by a sparse angular resolution can be reduced using a three-dimensional transformation map. Axial and sagittal views of filtered back-projection reconstructed volumes are shown where data is available at either $\Delta\theta = 1°$ or $\Delta\theta = 5°$ ($f_{\text{target}}(x, y)$ or $f(x, y)$ respectively). The space variable z is omitted from the notation when convenient since the volumes are treated as sets of individual slices.

Restored reconstructions, i.e. those with prior knowledge applied by means of a transformation map, are labeled $f_{\text{out}}(x, y)$ and shown in Figures 5.5(c) and 5.5(i). The derived measures that constitute the dimensions of the three-dimensional (64^3) transformation map, in this case chosen to be $\mathcal{H}_1 \equiv f(x, y)$, $\mathcal{H}_2 \equiv |f'(x, y) * h(x, y)|$ and $\mathcal{H}_3 \equiv f(x, y) * h_2(x, y)$, are also shown in Figures 5.5(d), 5.5(e) and 5.5(f) respectively. Standard deviations for $h(x, y)$ and $h_2(x, y)$ are 0.7 and 1.5 respectively in this example. The feature $f(x, y) - f(x, y) * h_2(x, y)$ used for denoising in Section 5.2.2, is broken up into two separate features $\mathcal{H}_1$ and $\mathcal{H}_3$ in this subsection. This modification adds to the map the capability to function differently depending on the intensity $f(x, y)$. Denoising in Section 5.2.2 is ignorant of intensity $f(x, y)$ as it should be.

In this example it was found that generating the transformation map (discussed thoroughly in the next subsection) using examples comparing the entire volume $f(x, y, z)$ to $f_{\text{target}}(x, y, z)$ achieves very similar results to using only the top half of the volumes, where $z < z_{\max}/2$. This is because there are ample examples in both situations that agree in the transformation map. Results reflect that artefacts are greatly reduced – without blurring or oversharpening edges. A natural look seems to be retained. Details in the target image that are not already present in the input image are not recovered in the output image.

5.4.3 Understanding the transformation map

It is important to be able to interpret the transformation map visually to understand how it is able to reduce artefacts. The following continues the discussion from the previous section on the effect of the transformation map on the reconstruction with the aim of removing the artefacts. Figure 5.6 partly illustrates the transformation map used.



(a) Low local edge magnitude.

(b) Medium local edge magnitude.

(c) High local edge magnitude.

Figure 5.6 Slices of the 64^3 transformation map used in Figure 5.5 for the removal of tomographic artefacts.

Regard the transformation map method as a type of nonlinear edge preserving filter such as anisotropic diffusion or the bilateral filter. See (Perona and Malik, 1990), (Tomasi and Manduchi, 1998) and in particular (Delaney and Bresler, 1995) on removing tomographic artefacts with edge preserving filters. Any edge preserving denoising filter could be used to reduce tomographic artefacts if their tunable parameters are chosen accordingly. The difference is that the transformation map method is more likely to produce natural looking results because, unlike classical regularization methods, its characteristics are not modeled with simple mathematical functions. Instead, the transformation map is created from realistic examples and leaves no tunable parameters.

Nonlinear filters are able to reduce artefacts in cases where the artefacts are not as prominent as edge features in the image. It is the nonlinearity in the algorithm that 'decides' if an image characteristic is an artefact or not and biases it accordingly. For example, Figure 5.6(a) illustrates the behaviour of the transformation map on image characteristics with low local edge magnitudes, i.e. $\mathcal{H}_2 \approx 0$. In this figure darker pixels indicate that the transformation map bias is negative and will result in decreased intensities while brighter pixels result in increased intensities. Grey implies no change and equals zero.

Consider points in this map near $(\mathcal{H}_1, \mathcal{H}_3) \approx (0, 0)$, for which the intensity ($\mathcal{H}_1$ or horizontal axis) as well as the mean neighbouring intensities ($\mathcal{H}_3$ or vertical axis) are close to the density of soft tissue. The bias is as follows: if the intensity of a voxel is greater than 0 while the mean is still around 0, that voxel intensity will be biased down towards 0. Likewise, if the intensity

is less than 0 it will be biased up towards 0. Therefore, if there were a *slight* artefactual streak through a big region of tissue in the image, its intensity will be biased towards the intensity of the tissue. The same applies to air at $(\mathcal{H}_1, \mathcal{H}_3) \approx (-1000, -1000)$.

If the artefactual streak were a bit more prominent in the hypothetical region of soft tissue, or if we are considering a smaller region of tissue (or perhaps the border of the region) then Figure 5.6(b) will be relevant instead because the local edge magnitude value $\mathcal{H}_2$ will be larger. The intensities of the points in the map near $(\mathcal{H}_1, \mathcal{H}_3) \approx (0, 0)$ and also (-1000,-1000) still ramps down from left to right as before, yet the intensity variation is less pronounced. In other words, the map bias will behave very similarly for these local image characteristics, yet will be slightly smaller in magnitude. This makes sense because firstly, if an artefact is more prominent, then there is less certainty of it being an artefact rather than an edge. Secondly, if it were an edge, it should not be biased much because then the edge may disappear or be sharpened unjustifiably. Similar logic dictates classical edge preserving filters (see Section 4.4.1).

For high local edge magnitudes as in Figure 5.6(c), the map bias magnitudes seem comparatively lower, and not only flatter around the diagonal $\mathcal{H}_1 = \mathcal{H}_3$ but also more uniform (there are no more distinct peaks and troughs around (0,0), for example). This means that prominent edges are generally not much affected (due to the central flatness) although outliers are indeed smoothed (light left extreme and dark right extreme entries in the map bias intensities towards the center). Note that the uniformity along the diagonal $\mathcal{H}_1 = \mathcal{H}_3$ (or specifically that intensity profiles of this figure along $\mathcal{H}_1$ for different $\mathcal{H}_3$ are nearly identical except that they are shifted proportionally to $\mathcal{H}_3$) implies invariance to intensity.

Of course, each of the nonzero entries in the 64^3 transformation map has some specific meaning that defines the map's behaviour for a given image characteristic. It is this unsimplified complexity that gives transformation maps the advantage in delivering natural-appearing results compared to methods that parameterize image characteristics in mathematical functions as in the previous chapter.

5.4.4 Iterative regularization

In this subsection, the transformation map models the difference between a volume originating from $\Delta\theta = 5°$ data and a volume originating from $\Delta\theta = 1°$ data. The transformation map is calculated in the same manner as above using the same derived dimensions. The only difference is that positively constrained iterated reconstructed slices are used to generate the transformation map, instead of filtered back-projection slices. Further, the values for the stan-

(a) Biased afterwards, $\Delta\theta = 5°$, $\sigma_{5°} = 90.02$, $\sigma_{1°} = 102.6$.

(b) Fixed fraction, $\Delta\theta = 5°$, $\sigma_{5°} = 78.52$, $\sigma_{1°} = 90.82$.

(c) More populated map, $\Delta\theta = 5°$, $\sigma_{5°} = 78.98$, $\sigma_{1°} = 90.67$.

Figure 5.7 Iterated reconstructions biased using transformation field method with positivity constraint.

dard deviations for $h(x, y)$ and $h_2(x, y)$ are both equal to 1. A systematic way of choosing these parameter values are discussed in Section 5.5.

Similar to the previous section, a positively constrained iterated reconstruction can be biased in a single operation after the iterated reconstruction is completed. Performing this operation results in a poorer data fit $\sigma_{5°}$ as indicated in Figure 5.7(a) with $\sigma_{5°} = 90.02$. However, the artefacts appear much reduced compared to the positively constrained iterated reconstruction shown in Figure 4.8(b) with $\sigma_{5°} = 77.96$.

A better data fit can be achieved if such a bias correction operation is performed at each step during reconstruction. Figure 5.7(b) with $\sigma_{5°} = 78.52$ shows the result using a *fixed fraction* (in this case $\lambda = 0.1$ in Equation 5.6) of the bias direction which is calculated from the transformation map at each iteration step (in this case 100 iterations). Conceivably, this fraction can be chosen to achieve a desired level of data fit (so that $\sigma_{5°} = \sigma_{5°,\text{aim}}$), as in regularization with Lagrange multipliers.

In other words, following every projection data error correction step during iterative reconstruction, the transformation map correction value (indexed by the feature dimensions which is recalculated at each step) is determined for each voxel. Instead of adding this correction value to the voxel intensity as is done for non-iterative regularization, only a fraction of this value is added to the voxel intensity before the cycle iterates.

As shown in Figure 5.7 compared to Figure 4.17, the ground truth fitness value for this reconstruction ($\sigma_{1^\circ} = 90.67$) improves on an equivalent result obtained in the previous chapter using Bayesian methods ($\sigma_{1^\circ} = 91.19$). The Bayesian reconstructions also seem to have an artificial look by comparison due to the simplicity of its mathematical approximations.

5.4.5 Populating the transformation map

The option of populating the transformation map by blurring has been mentioned earlier. It is preferable to have as many examples as possible to extract a well defined transformation map which is smooth and noise free. Instead of a single reconstruction volume pair, a whole sequence of reconstruction volumes can be used as *training* examples.

The volumes corresponding to iterate numbers 10 to 15 for a positively constrained iterated reconstruction, with a projection data spacing of $\Delta\theta = 5^\circ$, are used as source examples. One target volume reconstructed from $\Delta\theta = 1^\circ$ interval projection data is paired with the six source volumes separately to generate the transformation map.

Figure 5.7(c) shows that this new transformation map performs slightly better numerically. No visual advantage is noticed in this case. The smallness in magnitude of improvement is attributed to the fact that the original transformation map is already well populated.

5.5 Map effectiveness

In grey scale image processing, an element at a particular location in a transformation map indicates a *mean* change in intensity for a large number of example pairs. The effectiveness of a transformation map is determined by its *variance*. In other words, a large degree of variation in the intensity change over all the examples at a particular location in the transformation map implies that the map is ineffective in that region and suggests that more suitable map dimensions should be chosen. In other words, the map needs to be made to depend on additional image features in this case.

5.5.1 Optimal parameters

So far the dimension features have been chosen successfully in the examples presented. It is conceivable that the task of choosing features may be more difficult for some problems. Given

a poor choice of feature dimensions, the map variance could be used as a tool for identifying problem areas within the map. This may assist in finding more suitable dimensions that resolve ambiguities or conflict in the map more effectively.

For the same reason it was found that a sensible way to optimize unspecified parameters in the feature dimensions (such as the blurring kernel sizes used in $h_1(x, y)$ and $h_2(x, y)$ of Section 5.2.2) is to choose values that minimize the total transformation map variance.

5.5.2 Target accuracy

Target accuracy can be defined as the *correctness* of $f_{\text{target}}(\mathbf{x})$. An ideal example might not be available and therefore target examples must sometimes be estimated. The transformation map variance is also dependent on the target accuracy or correctness and does not exclusively represent the map effectiveness. If target examples are not available but are estimated instead, then the collective source to target differences could result in a high variance transformation map, even though the map mean remains effective. The example in the following paragraph illustrates this point.

Looking back at Section 3.3.2, it is apparent that the intensity level based artefact removal algorithm falls within the framework of transformation maps where the map has one feature dimension (mean intensity) and several output dimensions (one output dimension for each pixel column as a function of the feature dimension). In this case no ideal target examples are available but they are crudely estimated by means of linear interpolation. Averaging many diverse artefact examples recovers effective mean artefact profiles for each given intensity level despite that linear interpolation is a poor estimator.

In this example, the poor estimates of the target examples result in a large total map variance, even though the map is effective.

5.6 Summary

This chapter introduced a problem solving framework which is closely related to neural networks. Besides its comparative simplicity and implementation efficiency, the advantage of transformation maps over neural networks is that they permit the explicit investigation of the input space. Conflict in the input space can be assessed, which assists with the choice of feature dimensions.

The major difference to neural networks is that transformation maps use significantly more training examples and such additional examples are typically generated by adding noise artificially. This method is best suited to problems having a relatively low number of feature dimensions as it suffers from the so-called *curse of dimensionality*.

It was shown that transformation maps have applicability to a number of different problems. Of particular interest is that transformation maps provide a means of extracting prior knowledge from, and regularizing, tomographic reconstructions. It was shown that more natural-looking results than with Bayesian methods were obtained for sparse angle reconstructions. In the next chapter, limited angle tomography is investigated using transformation maps.

CHAPTER 6

Limited angle tomography

This chapter is concerned with the reconstruction of fan beam x-ray data at a 5° angular spacing covering a 100° angular range. Such a data limitation severely underspecifies the reconstruction as discussed in (Tam and Perez-Mendez, 1981). For the sake of giving perspective, a 90° range of missing data implies that two quadrants of the Fourier transform of a reconstruction remain undetermined and are left to be recovered from prior knowledge.

First the qualities of limited angle tomographic reconstructions are illustrated. After a presentation of techniques found in the literature, an example of a Bayesian limited angle reconstruction is provided for reference. Towards the end of this chapter it is shown how the transformation map method is applied to the reconstruction problem to achieve a functional solution.

6.1 Unregularized limited angle reconstructions

Unregularized reconstructions are produced using projection data only while no additional assumptions are made. However, in some cases the positivity constraint is applied, which causes a noteworthy reduction in artefacts. This section discusses the results of using the most straightforward reconstruction methods possible.

6.1.1 Caution on simulations, resolution and data fit

All the reconstructions in Figure 6.1 are produced from projection data of an identical angular range (90°) and spacing (5°). These limited angle reconstructions (at $i_z = 225$) correspond to the full angle reconstruction in Figure 4.12(a).

(a) Real data reconstruction, 1760 bins per projection.

(b) Real data reconstruction, 216 bins per projection.

(c) Simulated data reconstruction, 1760 bins per projection.

Figure 6.1 Limited angle $\Delta\theta = 5°$ over $90°$ reconstructions after 5000 iterations showing the effect of using a high projection axis resolution. Notice definitive edges appearing in (c).

Figure 6.1(a) results from real measured projection data at the maximum available resolution of 1760 bins per projection. In the reconstruction process, a fraction of the error between the projections of the reconstruction with the available data is projected back onto the reconstruction 5000 times repeatedly. At each step the reconstruction intensities are clipped to be positive.

However, following the discussion in Section 3.7, the ideal number of projection bins is only 216 for the given geometry. Figure 6.1(b) shows great noise reduction with no loss in sharpness under this condition. It has been verified that Figure 6.1(a) has a worse projection fit to the 216 binned projection data than Figure 6.1(b) while Figure 6.1(a) has a better projection fit to the 1760 binned projection data than Figure 6.1(b).

Figure 6.1(c) is produced in exactly the same way using simulated instead of real measured data at 1760 bins per projection. The simulated data are the corresponding projections of the full angle reconstruction in Figure 4.12(a). For such a high projection axis resolution it is understandable that the shape of the pixels plays a significant role in the overall projection. This pixel shape information is responsible for the significant difference between Figures 6.1(a) and 6.1(c).

For any simulations regarding limited angle tomography, it is advisable to use a phantom of a higher resolution than that of the reconstruction because real data is of continuous origin. It should be noted that doing this makes a zero error data fit unattainable because it is not possible to represent exactly a high resolution object at a lower resolution. For a discrete reconstruction,

the best possible representation of a phantom will not have a zero data error, even if there were no measurement noise. This is true also using a lower projection axis resolution, although the effect is mitigated.

6.1.2 Properties of limited angle reconstructions

Figure 6.2 shows a number of unregularized limited angle reconstructions created from data covering a selection of angular ranges and spacings. It is apparent how the diagnostic value of reconstructions steadily deteriorates as the angular range of data decreases. Edges in the vertical direction disappear in extreme cases while edges in the horizontal direction remain very much the same.

Filtered back-projection reconstructions at $\Delta\theta = 5^\circ$ data intervals exhibit streaks due to the poor angular resolution. Applying a positivity constraint in an iterative reconstruction procedure seems to improve results significantly: better contrast and a reduction in artefacts are observed comparing Figure 6.2 (a) to (d), for example, despite no more data being used.

At a higher angular resolution, as in Figure 6.2(g), the streak or ripple artefacts are much reduced while edge definition in the vertical direction remains poor. It is concluded that restricting the angular range of data results in a loss of edge information and contrast in a corresponding angular range of the reconstruction.

6.1.3 Comparing limited to full angle reconstructions

The diagrams in Figure 6.3(a) relate the limited angle to full angle reconstructions shown in Figure 6.4. By comparing the histograms in Figure 4.14(a) and Figure 6.3(a) it becomes apparent that the intensities concentrate around the dominant levels (air, soft tissue and bone) even less for limited angle than for sparse full angle reconstructions. This implies that the limited angle reconstructions could benefit significantly from using piecewise smoothness as prior knowledge.

The intensity profiles in Figure 6.4(d) demonstrate how inadequately the limited data affirms the existence of edges in the vertical direction. Edges are much better defined in the horizontal direction. Figure 6.3(b) shows explicitly by means of a transformation field that full angle reconstructions have stronger edges in the vertical direction than corresponding limited angle

(a) Filtered back-projection, $\Delta\theta = 5°$ over 90°.

(b) Filtered back-projection, $\Delta\theta = 5°$ over 120°.

(c) Filtered back-projection, $\Delta\theta = 5°$ over 150°.

(d) Positively iterated, $\Delta\theta = 5°$ over 90°.

(e) Positively iterated, $\Delta\theta = 5°$ over 120°.

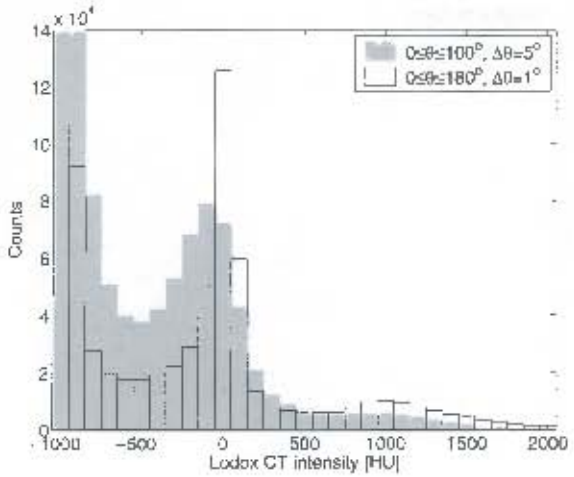
(f) Positively iterated, $\Delta\theta = 5°$ over 150°.

(g) Positively iterated, $\Delta\theta = 1°$ over 100°.

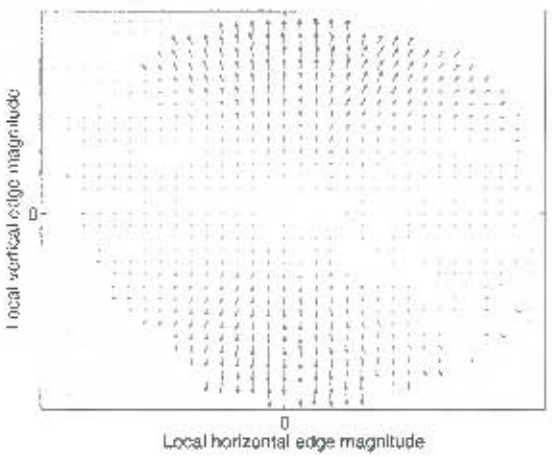
(h) Positively iterated, $\Delta\theta = 1°$ over 130°.

(i) Positively iterated, $\Delta\theta = 1°$ over 160°.

Figure 6.2 Unregularized limited angle reconstructions showing the effect of having more available data.



(a) Histograms showing how the distribution of intensity differ between the reconstructions.



(b) Transformation field showing how edges differ between the reconstructions.

Figure 6.3 Comparison of limited angle 100° in Figure 6.4(a) to full angle 180° in Figure 6.4(c) $128 \times 128 \times 128$ volume reconstructions.

reconstructions. It is evident from examining Figures 6.4(d) and 6.4(e) that the intensity levels for the limited angle reconstruction are more correct for the intensity profiles in the horizontal than in the vertical direction.

Note that the $\sigma_{5^\circ}^{100^\circ}$ data fit errors in Figure 6.4 refer to the projection errors with respect to the limited angle data subset, while the σ_{1° data fit errors refer to the projection errors of the reconstruction with respect to the entire dataset (which is not used as data during limited angle reconstruction).

6.2 Reconstructions using partial spatial knowledge

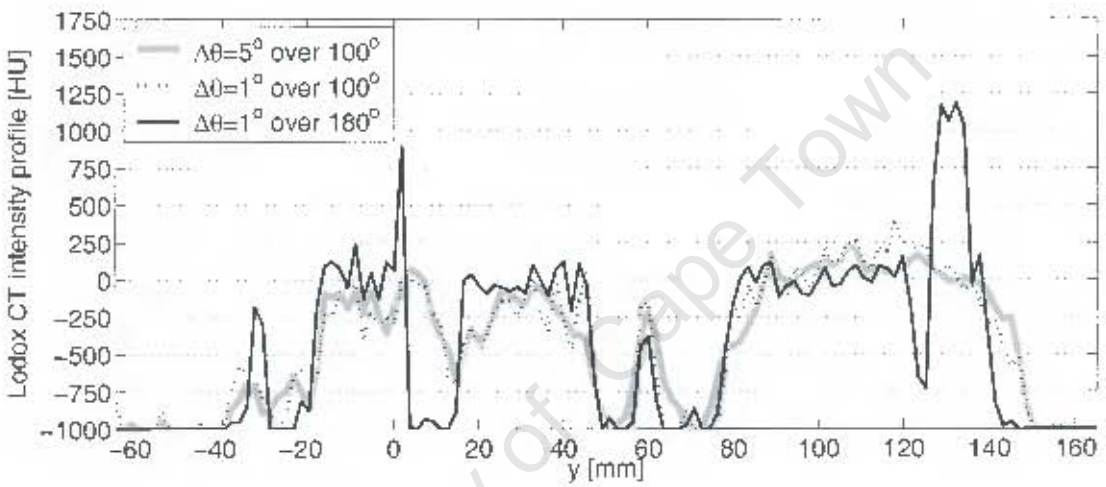
In Figure 6.5 the left hand column of images are closeups of limited angle reconstructions in Figure 6.4. The right hand column of images are corresponding views of full angle reconstructions. For the center column, the same limited angle data constraint is applied as for the left hand column, while the intensities in the spatial region outside the dotted circle are forced to equal the full angle reconstruction on the right at all times during iterative reconstruction.

These closeups reveal that there exist some regions in limited angle reconstructions such as (g) that resemble full angle reconstructions such as (i) well.

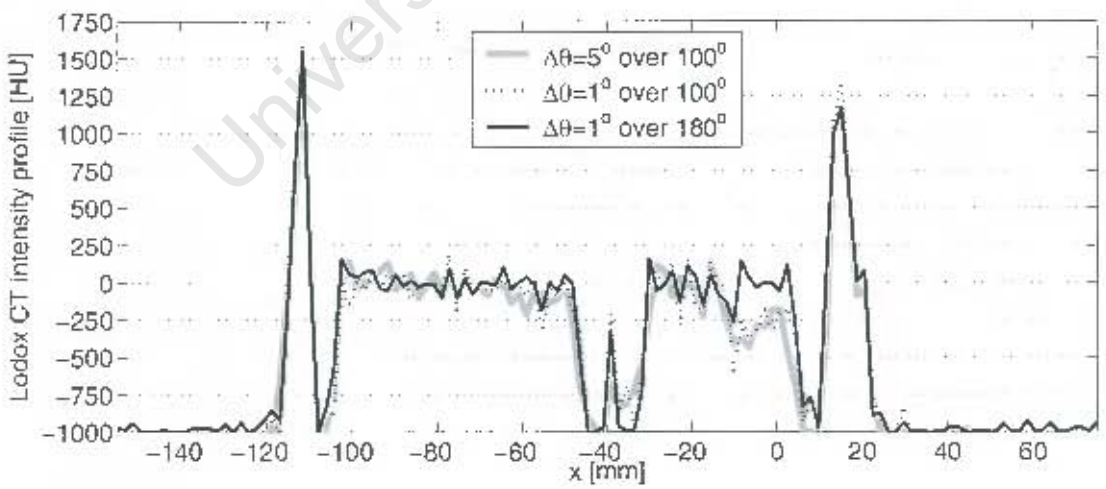
(a) $\Delta\theta = 5^\circ$ over 100°, $\sigma_{5^\circ}^{100^\circ} = 50.2$, $\sigma_{1^\circ} = 286$.

(b) $\Delta\theta = 1^\circ$ over 100°, $\sigma_{5^\circ}^{100^\circ} = 60.2$, $\sigma_{1^\circ} = 269$.

(c) $\Delta\theta = 1^\circ$ over 180°, $\sigma_{5^\circ}^{100^\circ} = 69.2$, $\sigma_{1^\circ} = 79.1$.



(d) Vertical profiles through center with poor correspondence of edges at $x = -44.38$mm.



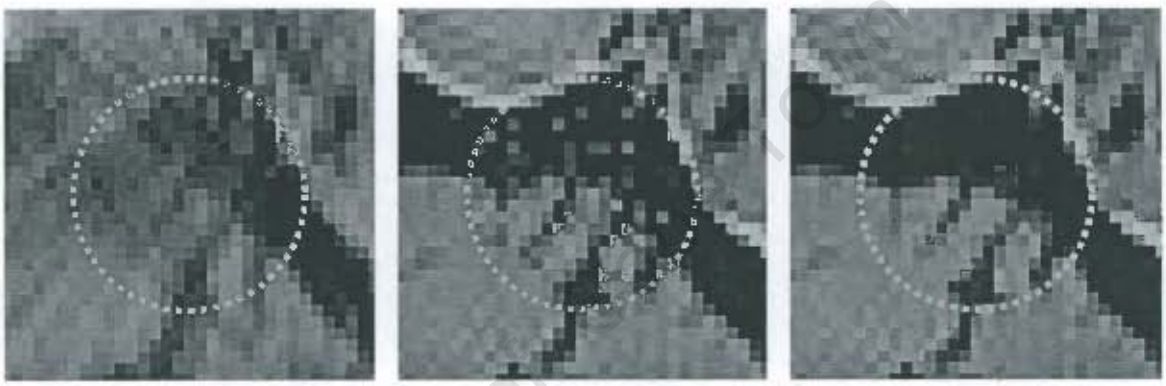
(e) Horizontal profiles through center with good correspondence of edges at $y = -49.09$mm.

Figure 6.4 Limited and full angle reconstructions with intensity profiles through the reconstructions, $i_2 = 225$.

(a) Unconstrained, $\Delta\theta = 5°$ over 100°.

(b) Spatially constrained, $\Delta\theta = 5°$ over 100°.

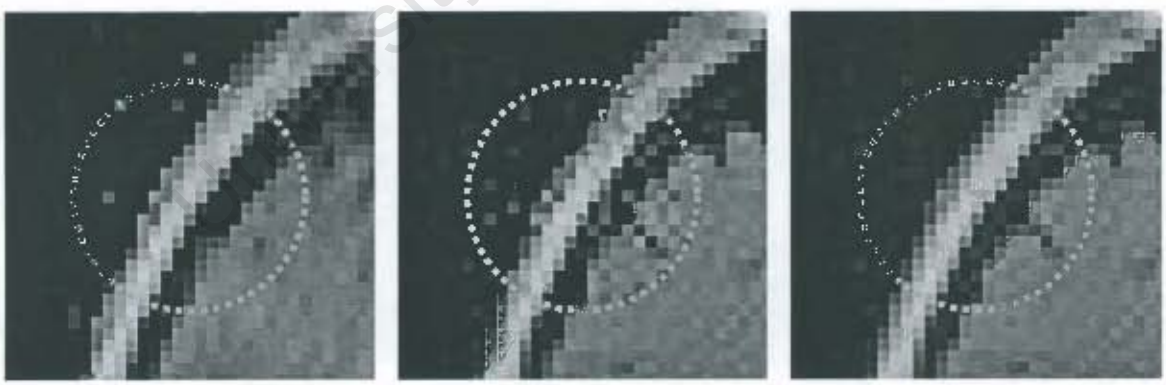
(c) Unconstrained, $\Delta\theta = 1°$ over 180°.



(d) Unconstrained, $\Delta\theta = 5°$ over 100°.

(e) Spatially constrained, $\Delta\theta = 5°$ over 100°.

(f) Unconstrained, $\Delta\theta = 1°$ over 180°.



(g) Unconstrained, $\Delta\theta = 5°$ over 100°.

(h) Spatially constrained, $\Delta\theta = 5°$ over 100°.

(i) Unconstrained, $\Delta\theta = 1°$ over 180°.

Figure 6.5 Closeup views of positively iterated reconstructions in Figure 6.4.

Regions corresponding to the complementary angular range of edges in the limited angle reconstruction such as (a) resemble the full angle reconstruction such as (c) very poorly.

Great improvement within the dotted circles is observed when correct spatial knowledge is applied outside the dotted circles (in the center column of Figure 6.5). This is especially true where edges in the vertical direction are lacking, as in this case. The important observed quality of limited angle tomographic reconstruction is that a correcting bias in some region results in an intensity contrast improvement in another region.

The conclusion that is drawn from this qualitative demonstration is as follows: if some region can be regularized well, then the improvement resulting from such bias is not restricted to the region to which it was applied. The complementary region is also improved indirectly by maintaining the data constraint which bonds the regions together.

As demonstrated earlier with negative noise being constrained to positivity, the converse is also true: the application of any incorrect bias will propagate elsewhere in a reconstruction. Accordingly, if a region is biased incorrectly, then not only that region but complementary regions will also be biased incorrectly through the data constraint.

6.3 Survey of limited angle tomography methods

Now that limited angle tomography has been introduced, this chapter proceeds by presenting a selection of strategies proposed in the literature for dealing with the problem. The less effective methods are discussed first. No methods found adequately solve the problem in a realistic ill-posed situation. An overview of state-of-the-art techniques is provided in (Siltanen *et al.*, 2003).

6.3.1 Direct inversion

There are a number of ways to perform reconstructions in a single step. It is possible to set up a linear system of equations $p = Hf$ which expresses measured projections p as raysums through a discrete cross-section f. Having more unknowns than linearly independent equations for the limited angle tomography case implies that the matrix H is singular. Because high frequency components are poorly transferred in tomography, H has been shown to be singular even in full angle tomography (Wilson, 1998). A solution with the smallest $|f|^2 = \sum_j f_j^2$ can be obtained by singular value decomposition, but is infeasible for realistically sized problems due

to computational expense and memory requirements (Boyd and Little, 1994). More efficient implementations of this inversion in the frequency domain have been explored by (Tam and Perez-Mendez, 1981) and (Jaffe, 1990). Reconstructions obtained through direct inversion are very much equivalent to results obtained by the filtered back-projection method.

Squashing is a technique introduced by (Reeds and Shepp, 1987) to perform an affine transform on limited angle projection data. (Olson and Jaffe, 1990) explains that squashing does little more than converting a limited angle data set into a suitable form to be applied directly to a full angle tomography reconstruction algorithm.

(Yagle, 1998) expresses a closed form solution to limited angle discrete tomography of finite support objects. In this paper, the phantom is expressed as impulses on a regular lattice and projections as well as phantom densities must be integer values. It is reported that totally incorrect reconstructions could result from noise that exceeds integer roundoff errors. Very large extrapolation coefficients need to be calculated for larger images. Success of equation-based solutions to limited angle tomography that does not incorporate prior knowledge other than geometric conditions, must be considered in the light of caution expressed in Section 6.1.1.

6.3.2 Regularization

Ill-conditioned inverse problems are unstable in the sense that small changes in data can cause large changes in the solution (Kwee *et al.*, 1997). Such problems can be stabilized by regularization so that the solution has some desired property such as smoothness. This is done by the minimization of an expression of the form

$$f_{reg} = \arg_f \min \left[U(f) + \lambda \chi^2 \right] \tag{6.1}$$

where $U(f)$ has a minimum when the solution f has the desired property and λ can be tuned so that the level of data fit χ^2 matches the measurement error in the data. Many methods use this framework for achieving a solution.

The Laplacian smoothness condition,

$$U(f) = \sum_j \left[\nabla f^2 \right]^2 \tag{6.2}$$

is used by (Tam and Perez-Mendez, 1981) to regularize the limited angle inverse problem. The maximum entropy method used by (Lawrence, 1989) for limited angle tomography maximizes

the entropy $S(f) = -U(f)$ of an image f, with $f_j > 0$, where

$$S(f) = -\sum_j p_j \ln p_j, \quad p_j = \frac{f_j}{\sum_j f_j}. \tag{6.3}$$

(Tam and Perez-Mendez, 1981) concludes that maximum entropy does not produce better results than any other simple limited angle tomography method. This is understandable since the results of the maximum entropy method show the least amount of configurational structure (which is least informative) yet are consistent with the measured data (Bontekoe *et al.*, 1994). The maximum entropy method is better suited to problems such as gas and plasma diagnostics (Denisova, 1998) where the observed quantity obeys Boltzmann statistics, or in astronomical imaging (Narayan and Nityananda, 1986) where entropy describes the properties of the image intensity distribution well.

Correlation amongst pixels can be specified using minimum cross-entropy, defined by (Skilling and Bryan, 1984) as

$$S(f) \quad \sum_j f_j - f'_j - f_j \ln(f_j / f'_j) \tag{6.4}$$

where f' is the default or expected image. (Fielden *et al.*, 1991) concludes in a full angle magnetic resonance imaging application that truncation artefacts cannot be suppressed using minimum cross-entropy without some loss of resolution.

(Hanson and Wecksung, 1983) demonstrates that good Bayesian limited angle reconstructions can be achieved in a simple example if the prior knowledge is specific enough.

Good results are presented by (Delaney and Bresler, 1995) using piecewise smooth regularization with an objective function of the form

$$\chi^2 + \lambda'_1 \mathcal{K} f| + \lambda \sum_j \phi(|\nabla f|_j) \tag{6.5}$$

where $\mathcal{K}$ penalizes negative values of f and ϕ is a nonconvex functional that operates on the intensity gradient of f. This objective function is equivalent to discontinuity adaptive priors discussed extensively in (Li, 1995) within the Bayesian framework. The results, however, deteriorate significantly as the range of available projections is decreased.

Deterministic annealing is used by (Hsiao *et al.*, 2001) in a simulated Bayesian sparse angle segmentation example to overcome problems due to sensitivity to initial conditions.

6.3.3 Sinogram extrapolation

A linear neural network approach is taken by (Wong and Yau, 1998) to extrapolate 12.5% missing projections with good accuracy in a simulated example. The missing projections are expressed as a function of the projection data and undetermined parameters. These parameters are initialized using a sinogram of a similar phantom, and iterated a few times. The exterior of the object is assumed to be known.

(Andia *et al.*, 2002) obtains comparable results to Bayesian methods reconstructing SPECT data non-iteratively by performing nonlinear operations on the sinogram before back-projection. Training is performed using a similar dataset to the projections of the phantom.

In (Prince and Willsky, 1993) the phantom's mass, center of mass and boundary of its sinogram is estimated from available projections. Together with parameters that specify smoothness constraints which are assumed to be known *a priori*, a complete sinogram is extrapolated.

The success of all these algorithms is dependent on how representative the *a priori* parameters or training examples are of the phantom to be reconstructed.

6.3.4 Projection onto convex sets

(Sezan and Stark, 1984) shows that the iterative application of a set of constraints which can be expressed as projection operators onto closed convex sets converges to a solution that satisfies all the constraints.

(Oskoui-Fard and Stark, 1989) explains that expressing raysum data constraints using this method is equivalent to the algebraic reconstruction technique used in (Andersen, 1989). However, these projection operators may include positivity constraints, amplitude limits, energy limits, spatial support or known Fourier coefficients. Although this technique is effective, the range of prior knowledge that can be applied using this method is somewhat limited.

6.3.5 Optimal views

Conventional spiral tomography machines provide poor resolution in the direction normal to the cross-section plane. Cross-section slices are typically printed as a montage onto film to be viewed on a light box.

(McCauley *et al.*, 2000) displays limited angle reconstructions that appear very similar to corresponding full angle reconstructions. These are, however, not axial views. (Kolehmainen *et al.*, 2003) provides such views for dental tomography. Due to its high resolution in the scanning direction, the Lodox system can also provide such views.

Figure 6.6 shows a good correspondence of limited angle reconstructions to full angle reconstructions in Figure 4.11 for coronal views but not for sagittal views. All of these are simply positively iterated reconstructions. Note that even for coronal views there is some interference visible from out-of-plane structures as predicted by (Starke and de Jager, 1998).

6.3.6 Complementary data fusion

(Boyd and Little, 1994) uses laser range detection and ultrasonic measurements to determine the exterior and outer layer thickness of an aircraft wing in a limited angle tomography problem. *A priori* knowledge of x-ray attenuation coefficients for certain structures in the wing is also available. Projection onto convex sets is the recommended method to combine the various constraints for a realistically sized problem. Good results are presented. In another paper (Tam *et al.*, 1990), a similar technique is used to reconstruct turbine blades by combining wall thickness measurements.

In positron emission tomography (PET), which is a sparse data rather than limited angle problem, registered magnetic resonance images (MRI) are used to enhance the reconstruction quality. (Chen *et al.*, 1991) shows how Gibbs energies are applied to achieve improved resolution where anatomical and functional boundaries match. Alternatively, after performing an intensity transform to the MR image, (Oakley *et al.*, 2000) biases the PET reconstruction towards this reference.

6.3.7 Shape modeling

It is possible to convert an ill-posed inverse problem into a well-defined problem by modeling the solution. However, it is difficult to express unforseen special cases into any model especially if the model describes a physical shape instead of a mere quality which is always true.

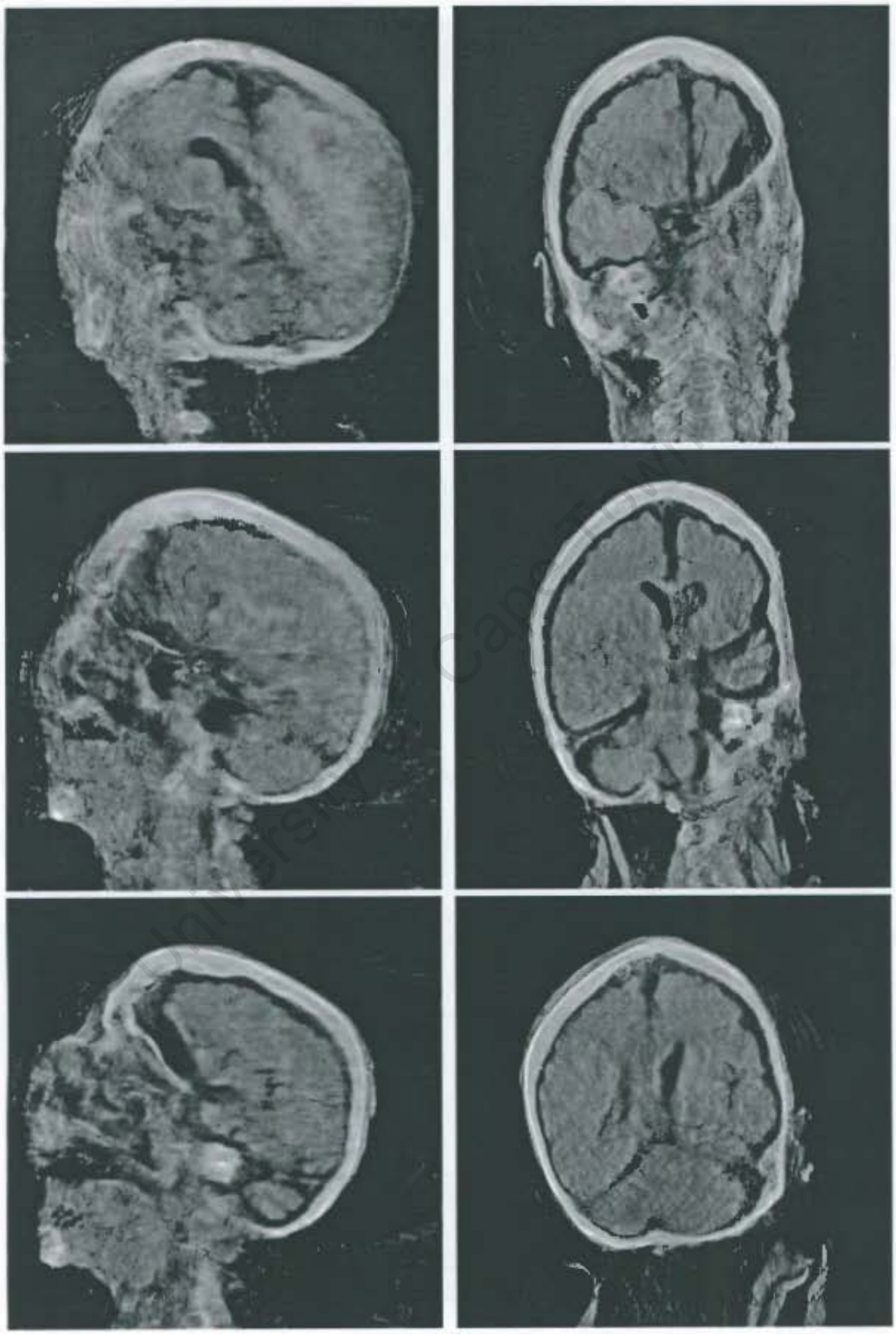


Figure 6.6 A selection of limited angle sagittal (left) and coronal (right) views at $256 \times 256 \times 256$.

(Gureyev and Evans, 1998) modeled hardwood samples and successfully reconstructed cross-sections in simulations from very few projections. From regularity assumptions, the three-dimensional problem is converted into a quasi-one-dimensional reconstruction problem. It is argued that the shape of vessels (voids) or parenchyma (nonfibrous tissue) can be determined by optical methods in practice and therefore be known *a priori* in their simulations. Their results show some tolerance to irregularities.

In atmospheric tomography, (Semeter and Mendillo, 1997) uses so-called *Chapman* functions to model the atmosphere which constrains the vertical profile initially. This constraint is removed later when entropy is maximized instead.

Boundary modeling using snakes (deformable polygons with curvature constraints) is investigated by (Mohammad-Djafari *et al.*, 1997) in the context of sparse angle tomographic reconstruction of simple binary blob-like phantoms.

In this thesis little attention is given to the modeling of shapes since there is so much physical shape variablity in expected reconstructions. Instead, effort is concentrated on the modeling of local image properties.

6.4 Bayesian limited angle reconstructions

Figure 6.7 shows Bayesian limited angle tomographic reconstructions using projection data at 5° spacing over a 100° range. Tunable parameters in the potential functions (defined in Section 4.4.2 with values as quoted in Figure 6.7) are chosen such that the ground truth σ_{1° data fit is minimized as in Section 4.4.3. Such *ad hoc* tuning produces the best possible reconstructions for the given forms of potential functions.

Although the Bayesian reconstructions have a noticeably better ground truth data fit (e.g. $\sigma_{1^\circ} = 249$) than the corresponding unregularized reconstruction in Figure 6.4(a) (with $\sigma_{1^\circ} = 286$), they appear quite unnatural and still resemble the best possible reconstruction in Figure 6.4(c) (with $\sigma_{1^\circ} = 79.1$) poorly. Edge information in the missing angular range is not recovered correctly despite the attainment of piecewise smoothness as suggested by the prior knowledge. It is noted that the reconstructions are correctly biased by the prior knowledge in some but not in other regions.

In the somewhat related example of sensitivity encoding magnetic resonance imaging, noise levels are not uniform throughout reconstructions. (Samsonov and Johnson, 2002) proposes that

(a) DA-MRF, $\sigma_{5^\circ}^{100^\circ} = 50.2$, $\sigma_{1^\circ} = 275$. $\beta = 0.2$, $\gamma = 360$.

(b) I-MRF, $\sigma_{5^\circ}^{100^\circ} = 50.8$, $\sigma_{1^\circ} = 254$ $\beta = 2.4$, $\gamma = 470$.

(c) DAI-MRF, $\sigma_{5^\circ}^{100^\circ} = 53.3$, $\sigma_{1^\circ} = 249$. $\alpha = 0.33$, $\beta = 1.125$, $\gamma = 210$.

Figure 6.7 Bayesian limited angle reconstructions ($\Delta\theta = 5^\circ$ over 100°).

space-varying anisotropic diffusion is superior to space-invariant methods of edge-preserving denoising in that situation.

As explained more analytically in (de Villiers, 2000), the reason why space-invariant methods found in the literature, as demonstrated in Figure 6.7, fail to converge to the correct reconstruction, is because the methods bias intensities incorrectly in regions where edges are poorly defined. By applying prior knowledge bias only to those regions that closely resemble the final reconstruction, one stands a much better chance of avoiding local optima in maximizing the posterior distribution. Although it may be possible to represent a space-varying form of some clique potential to deal with this inherent convergence problem, it is not clear how to derive such a prior knowledge model, nor to find values of any tunable parameters with reasonable justification.

The contribution made by the masters thesis (de Villiers, 2000) which may be difficult to get hold of, is presented in a more accessible and simplified document (de Villiers, 2004).

In the following section, the transformation map method is applied to the problem. The ability of the method to extract prior knowledge from examples overcomes the difficulties experienced in modeling using Bayesian methods.

6.5 Reconstruction using a transformation map

A good choice of feature vectors for the regularization of sparse angle tomographic reconstructions in the previous chapter has been intensity, blurred intensity and local edge magnitude. These three-dimensions provide the transformation map with the ability to smooth out noise without blurring or sharpening edges beyond the curvature indicated by the training examples (refer to Section 5.2.1). Of course, these dimensions also provide the map with the ability to bias mean intensities towards any of the dominant classes of intensity levels.

In limited angle reconstructions, however, little or no bias should occur where the reconstruction is degenerate, otherwise it will be biased unjustifiably, resulting in convergence to a local optimum. This is what happened in the Bayesian reconstructions of the previous section.

To achieve this goal, keeping in mind that the better defined regions occur near well defined edges in the horizontal direction and poorly defined regions exist where edges in the vertical direction prevail, the edge magnitude dimension is split up into vertical and horizontal local edge magnitude dimensions.

6.5.1 Training examples for the transformation map

In this chapter the transformation map is calculated by comparing full angle to limited angle reconstructions for slice index i_z . 150 only (see Figure 4.10), unless specified otherwise. The full angle reconstruction (at $i_z = 150$) is rotated by multiples of 90° and mirrored, providing 8 permutations. Limited angle reconstructions are produced from artificially generated projection data for each of these 8 cases. An example of such a limited angle and full angle pair is shown in Figures 6.8(a) and 6.8(b). The limited angle reconstructions are produced in 500 iterations using the positivity constraint in accordance to the absolute geometry specified by $\Delta\theta = 5°$ over a 100° range.

Not only the final 8 limited angle reconstructions, but also intermediate reconstructions at intervals of 50 iterations are used in calculating the transformation map. More examples (as in Figure 6.8(c)) are generated from these by adding Gaussian noise with standard deviations ranging from zero to 100 in 20 steps. Still more examples are created by performing edge sharpening (by standard unsharp masking) of strengths ranging from 0 to 1 in 5 steps. These details are not important since they can be changed vastly with little effect to the transformation map. The idea is simply that many representative examples are created.

(a) Target $f_{target}(x, y)$, rotated full angle reconstruction.

(b) Limited angle reconstruction of target image.

(c) Source $f(x, y)$, sharpened reconstruction with noise added.

(d) Blurred source.

(e) Local horizontal edge magnitude of source.

(f) Local vertical edge magnitude of source.

Figure 6.8 Sample quantities used to determine transformation map.

See Figure 6.9 for an outline of the operations involved in creating the transformation map. Also, note that the limited angle reconstruction cycle, illustrated in Figure 6.10, is identical to what is described previously for sparse angle tomography.

6.5.2 Feature vectors

The four-dimensional transformation map being calculated is indexed by intensity ($\mathcal{H}_1 \equiv f(x, y)$), blurred intensity ($\mathcal{H}_2 \equiv f(x, y) * h(x, y)$), local horizontal edge magnitude ($\mathcal{H}_3 \equiv \frac{df(x,y)}{dx} * h_2(x, y)$) and local vertical edge magnitude ($\mathcal{H}_4 \equiv \frac{df(x,y)}{dy} * h_2(x, y)$) as illustrated in Figures 6.8(c), 6.8(d), 6.8(e) and 6.8(f) respectively for the example. As before, $*h(x, y)$ and

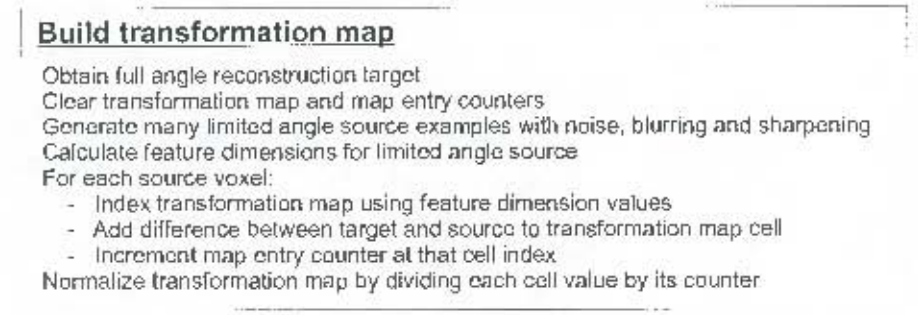


Figure 6.9 Steps in building a transformation map.

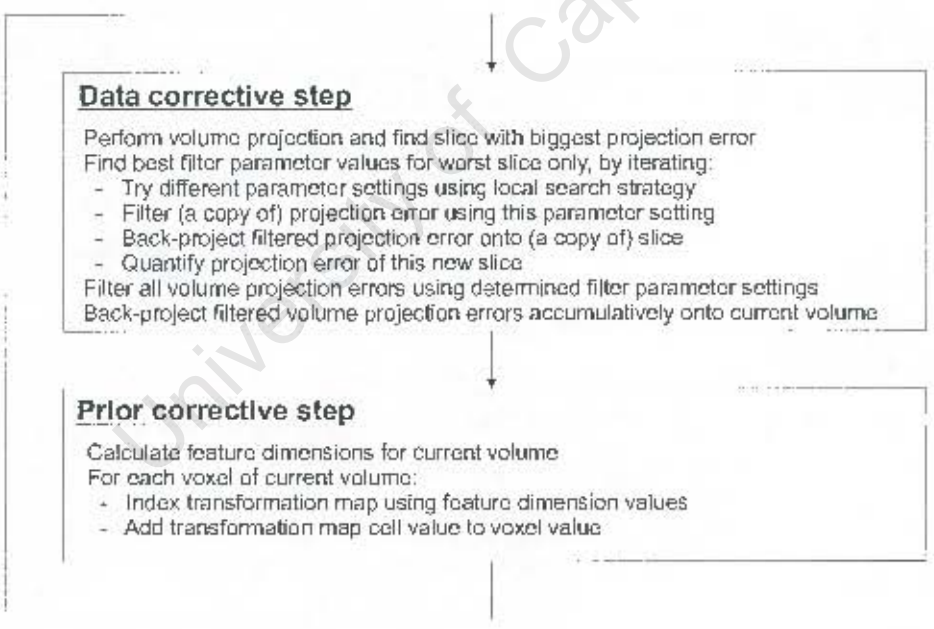
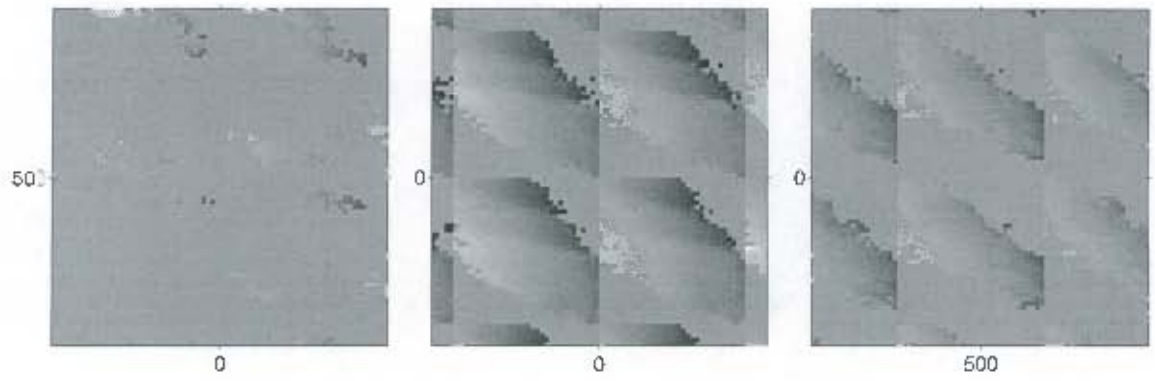


Figure 6.10 Iterative limited angle reconstruction cycle.



(a) High local vertical edge magnitude.

(b) Low local edge magnitude.

(c) High local horizontal edge magnitude.

Figure 6.11 Enlarged views of transformation map in Figure 6.12.

$*h_2(x, y)$ indicates convolution with Gaussian blurring kernels, both having standard deviations of 1 in this instance. The transformation map holds the average difference in intensity from the source $f(x, y)$ to the target $f_{\text{target}}(x, y)$ (Figures 6.8(c) and 6.8(a) respectively) at the indexed feature coordinates.

6.5.3 The effect of the transformation map

Figure 6.11 shows selected closeup views of the resultant $32 \times 32 \times 32 \times 32$ transformation map displayed as a montage in Figure 6.12. The small subimages in Figure 6.12 are slices in the *intensity* $\times$ *blurred intensity* plane. These subimages are ordered from left to right in increasing *local horizontal edge magnitude*, and are ordered from top to bottom in increasing *local vertical edge magnitude*. It is clear that the transformation map is not well populated in the case of large local vertical edge magnitude and would therefore not have much of a biasing effect for that condition.

The ill-defined condition of large local vertical edge magnitude is shown in Figure 6.11(a). In this region most of the transformation map values are zero.

For low edge magnitudes, Figure 6.11(b) shows the biasing effect of the map towards dominant intensity levels. If an intensity is higher than a dominant level, the map indicates (with darker pixels) a negative bias back down towards it. If the intensity is lower, it is biased (indicated with brighter pixels) up towards the closest dominant level.

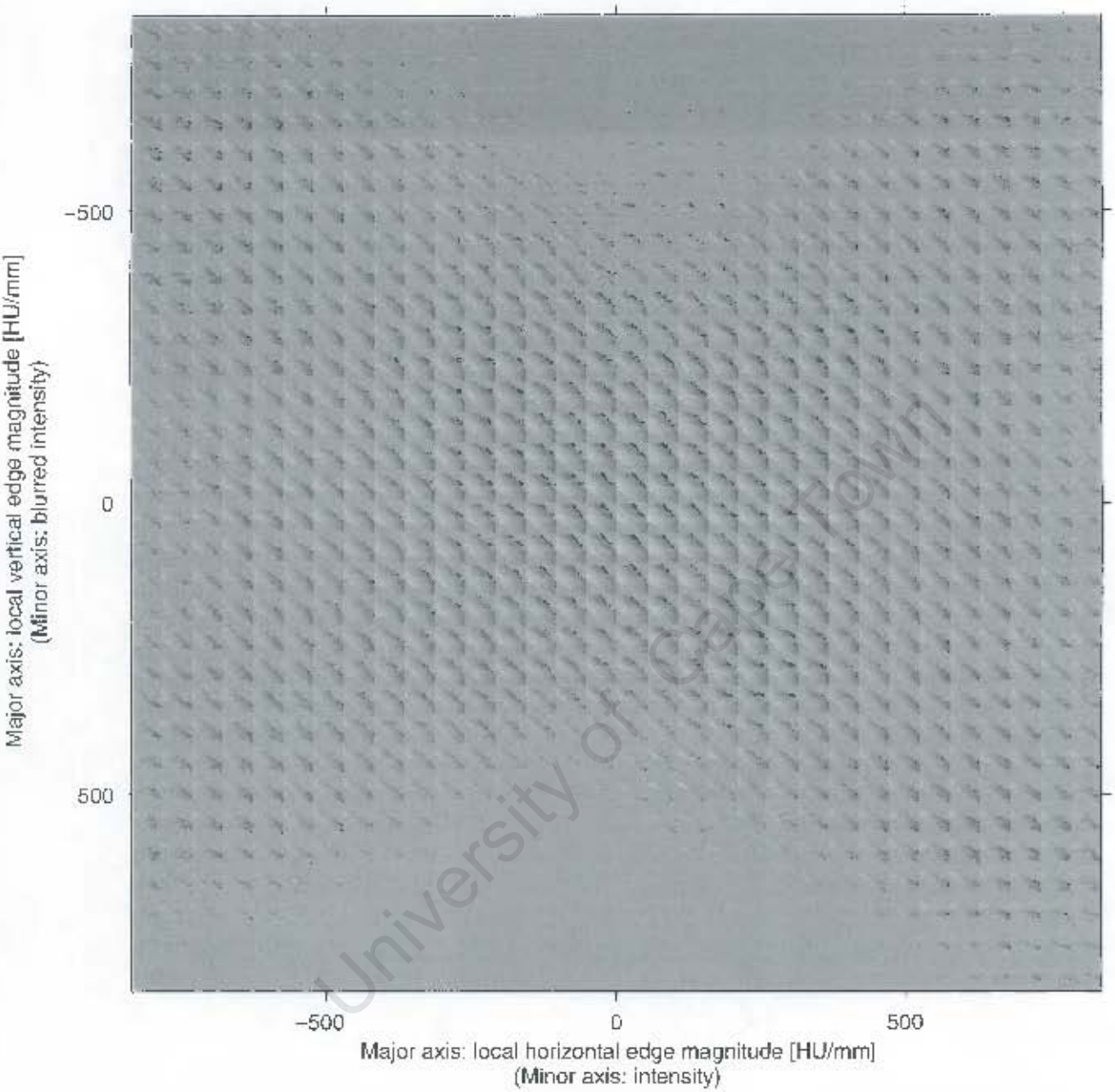
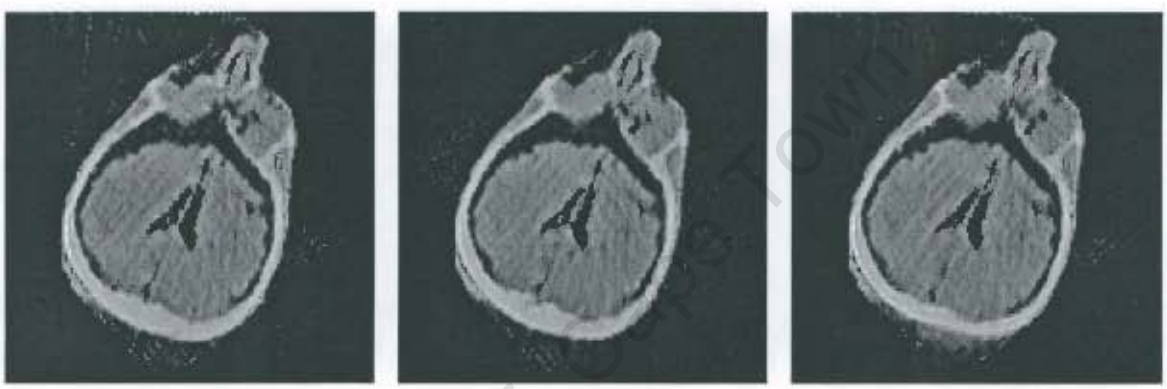


Figure 6.12 Four-dimensional transformation map for limited angle reconstruction displayed as a montage. The *intensity* × *blurred intensity* planes are ordered according to the horizontal and vertical edge magnitude dimensions leftwards and downwards respectively. Enlargements are provided in Figure 6.11.

(a) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100^6} = 69.2$, $\sigma_{1°} = 79.1$.

(b) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100^\circ} = 50.2$, $\sigma_{1°} = 286$.

(c) DAI-MRF, $\sigma_{5°}^{100^0} = 53.3$, $\sigma_{1°} = 249$.



(d) Using transformation map calculated from slice $i_z = 150$, $\sigma_{5°}^{100^6} = 49.5$, $\sigma_{1°} = 188.0$.

(e) As for Figure 6.13(d) except using simulated data, $\sigma_{5°}^{100^a} = 70.5$, $\sigma_{1°} = 183.5$.

(f) Using transformation map calculated using all slices, $\sigma_{5°}^{110^6} = 49.3$, $\sigma_{1°} = 172.8$.

Figure 6.13 Comparison of transformation map limited angle reconstruction results.

Figure 6.11(c) shows the map for high local horizontal edge magnitudes. In such regions smoothing occurs to some extent. If intensities are higher than the blurred intensity, the intensity is biased down towards the mean and if intensities are less than the blurred intensity, the intensity is biased up towards the mean. The extent to which this happens depends on the other features as well.

6.5.4 Results

Figure 6.13 shows limited angle reconstruction results using the transformation map method. Figure 6.13(d) is produced using the transformation map in Figure 6.12. The reconstruction

cycle illustrated in Figure 6.10 was iterated 5000 times starting from blank to produce this result. For each iteration after every data correcting step, the intermediate reconstruction is updated by a fixed fraction λ of the direction indicated by the transformation map following the discussion in Section 5.4.4. Specifically, each pixel is biased by this fraction λ (=0.0035 in this case) of the value in the transformation map which is indexed by the feature vector for that pixel. λ controls the goodness of fit to the $\Delta\theta = 5°$ projection data used.

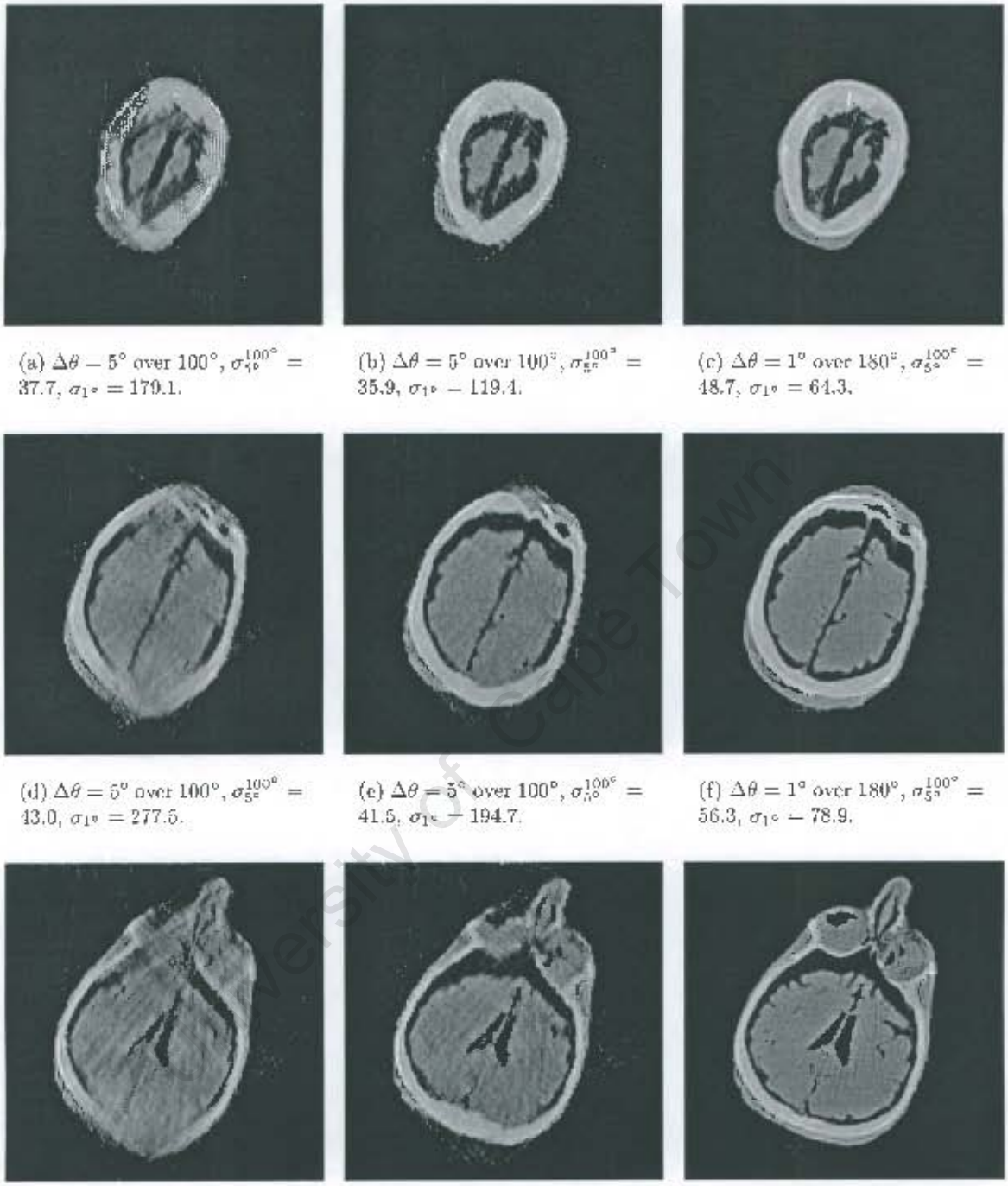
A ground truth full angle reconstruction with $\sigma_{1°} = 79.1$, as well as limited angle reconstructions using the positivity constraint ($\sigma_{1°} = 286$) and Bayesian methods ($\sigma_{1°} = 249$) are also provided for comparison. The transformation map reconstructions with $173 \leq \sigma_{1°} \leq 188$ appear significantly better than the limited angle reconstructions shown for the other methods, both visually and in terms of ground truth $\sigma_{1°}$ data error.

For interest's sake, Figure 6.13(e) shows a similar reconstruction where simulated data projections were used instead of measured projections. Noise was added to the projections whilst the projection geometry should be perfect. Little difference in the reconstruction quality for these results is observed.

A different transformation map is used for the reconstruction in Figure 6.13(f) giving a slighly better result ($\sigma_{1°} = 172.8$). In this particular case, the transformation map is calculated using 128 slices instead of only 1 at $i_z = 150$. The reason that only a single slice is used to determine the transformation map in general is to prove independence of prior knowledge in the sense that the prior knowledge extracted here can be applied equally successfully to another dataset under similar imaging conditions.

Lastly, Figures 6.14 and 6.15 present results for a number of different slices. The transformation map used for these reconstructions is shown in Figure 6.12. In all the examples, a significant improvement is observed in the transformation map reconstructions over straightforward positively iterated reconstructions. A natural look is retained in the results and edges come into appearance that exist in the full angle reconstructions.

Having a complete dataset remains the superior option. However, if the measurements are not available, transformation maps seem to provide a successful alternative.



(a) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 37.7$, $\sigma_{1°} = 179.1$.

(b) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 35.9$, $\sigma_{1°} = 119.4$.

(c) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100°} = 48.7$, $\sigma_{1°} = 64.3$.

(d) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 43.0$, $\sigma_{1°} = 277.5$.

(e) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 41.5$, $\sigma_{1°} = 194.7$.

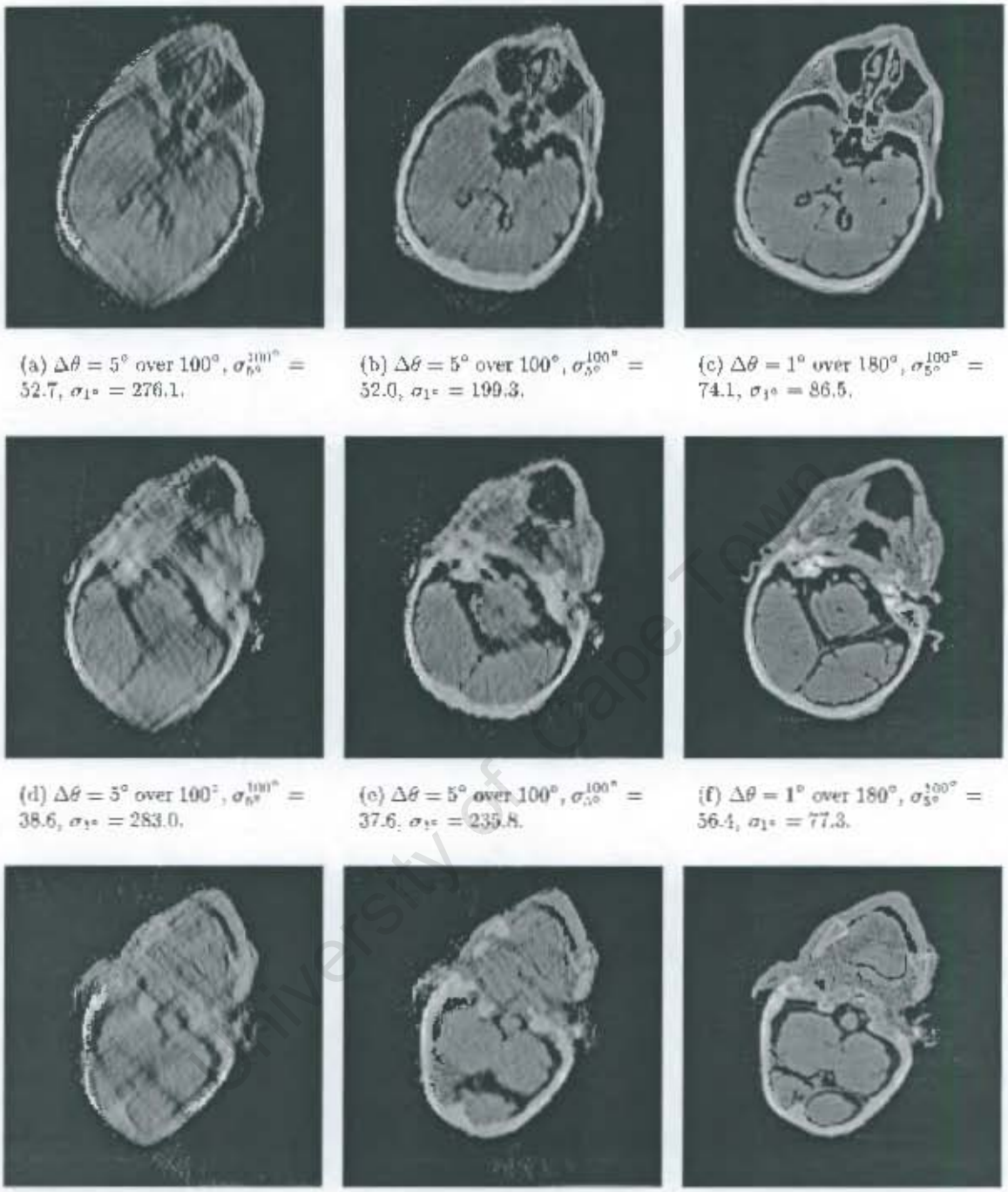
(f) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100°} = 56.3$, $\sigma_{1°} = 78.9$.

(g) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 50.5$, $\sigma_{1°} = 287.6$.

(h) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 49.5$, $\sigma_{1°} = 188.0$.

(i) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100°} = 68.3$, $\sigma_{1°} = 82.7$.

Figure 6.14 Limited angle reconstructions at $i_s = \{75, 150, 225\}$ using transformation map regularization (center) compared to positively iterated limited angle (left) and full angle (right) reconstructions.



(a) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 52.7$, $\sigma_{1°} = 276.1$.

(b) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 52.0$, $\sigma_{1°} = 199.3$.

(c) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100°} = 74.1$, $\sigma_{1°} = 86.5$.

(d) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 38.6$, $\sigma_{1°} = 283.0$.

(e) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 37.6$, $\sigma_{1°} = 235.8$.

(f) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100°} = 56.4$, $\sigma_{1°} = 77.3$.

(g) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 53.8$, $\sigma_{1°} = 240.6$.

(h) $\Delta\theta = 5°$ over $100°$, $\sigma_{5°}^{100°} = 53.1$, $\sigma_{1°} = 208.3$.

(i) $\Delta\theta = 1°$ over $180°$, $\sigma_{5°}^{100°} = 66.9$, $\sigma_{1°} = 86.4$.

Figure 6.15 Limited angle reconstructions at $i_z = \{300, 375, 450\}$ using transformation map regularization (center) compared to positively iterated limited angle (left) and full angle (right) reconstructions.

6.6 Envisaged limited angle tomographic system

This section gives an overview of what is practically required to build limited angle tomographic functionality into the Lodox system. Some further development work is essential for achieving results as presented above in a commercial system, as opposed to in an offline development system.

6.6.1 Scanning hardware

As mentioned before, in existing Lodox systems, scanning does not start accurately enough at any requested offset (in the z direction). It is also difficult to rotate the C-arm to a precise angular position. These fundamental problems need to be addressed. Although software techniques could be used to overcome these issues, it is a waste of valuable processing time which could much more productively be used on the reconstruction.

It would be useful, but not be essential, to change the detector hardware so that scanning can also be performed in the reverse, $-z$, direction. This would halve acquisition time although it will heat load the x-ray tube and generator more. Allowing faster scanning speeds than the current maximum of 140mm/s is also advised. Such developments are important considerations because the patient needs to remain immobile for the duration of at least 20 scans.

Furthurmore, the determination of the C-arm geometry must be automated by means of a three-dimensional calibration tool, since different machines (and machine model versions) may have slightly different geometries. Projection lookup tables that relate voxels to projection data can then be precomputed for predefined C-arm angles at which the patient will be scanned.

6.6.2 Gathering prior knowledge

Transformation maps extract prior knowledge from examples. Full angle and limited angle reconstruction example pairs must be available to generate the transformation map which will subsequently be used in a commercial system. It will be necessary to obtain a complete high resolution dataset of a subject that has not been treated with formalin, as in this thesis, since such treatment may have a significant effect on the attenuation properties of tissue.

Although it is not certain at this point, even more datasets may be required for different kV settings that change the x-ray spectra and ultimately, to some extent, the Hounsfield units in the

final reconstruction. Likewise, any technology change, such as the removal of fibre optic tapers or additional filtration, may have a significant effect which may necessitate the acquisition of another complete dataset. The sensitivity of the transformation map reconstruction method to such variability to its input has not yet been investigated.

Of course the limited angle reconstruction source example must correspond well to the C-arm geometrical conditions under which the system will be used. The prior knowledge has an angular dependency which relates to the C-arm geometry. Similarly, the spatial resolution of the examples used to generate the transformation map must also correspond to the resolution of subsequent reconstructions that employ that transformation map.

It is conceivable that superior results could be obtained by using a transformation map tailored to the region of interest. In other words, if a head is scanned, a transformation map trained using head examples would be recommended, while if a torso is scanned, a transformation map trained from torso examples would be preferred instead. These recommendations are not requirements.

6.6.3 Speeding up processing time

Around the order of 5000 data correcting and prior knowledge correcting cycles are iterated for a single transformation map based 128×128 limited angle reconstruction. This is a substantial amount of computation. For such a limited angle tomographic system to produce results in reasonable time to be commerically viable, hardware acceleration is essential.

Fortunately, many of the required operations can be performed in parallel. It seems sensible to dedicate one processor per scan angle to speed up the projection and back-projection processes, which constitute the bulk of the computation.

6.7 Summary

This chapter investigated the properties of limited angle reconstruction. Methods used in the literature are reported. Bayesian limited angle reconstructions were performed for reference. Thereafter the transformation map method is applied to the problem. It is shown that, using a transformation map, much improved results are achieved compared to other methods. Results resemble the full data reconstructions much more closely. Even though the results are not perfect they give insight into the true underlying reconstruction when the data is not available.

CHAPTER 7

Conclusion

This final chapter summarizes the most important features of this project and provides some closure to the topic of limited angle tomography for the Lodox system. The diagnostic value of various reconstructions is qualified first to give added perspective. Then a summary of the project developments, findings, conclusions and recommendations follows.

7.1 Diagnostic value of reconstructions

Professor S. J. Beningfield, Head of the Department of Radiation Medicine, at Groote Schuur Hospital, is quoted below on the diagnostic value of results presented in this thesis.

7.1.1 High resolution full angle reconstructions

Refer to the $\Delta\theta = 1°$ over 180° positively iterated axial, sagittal and coronal full angle reconstructions in Figures 4.10 and 4.11 on pages 66 and 67 respectively:

> *These are excellent quality; the coronals are reminiscent of the multiplanar reconstructions now obtainable with multi-slice CT with 0.5mm isotropic voxel data.*

7.1.2 Low resolution sparse angle reconstructions

Refer to the $\Delta\theta = 5°$ over 180° positively iterated (in Figure 4.8(b) on page 64), Bayesian (in Figure 4.17(c) on page 75) and transformation map (in Figure 5.7(c) on page 91) sparse angle cross-sections:

Figure 4.8(b) is not diagnostic, other than for gross pathology.
Figure 5.7(c) and Figure 4.17(c) are very similar, apart from coarser pixels in Figure 4.17(c) and are acceptable for diagnosis.

7.1.3 Coronal views of high resolution limited angle reconstructions

Refer to the $\Delta\theta = 5°$ over $100°$ positively iterated sagittal and coronal reconstructions in Figure 6.6 on page 107:

Sagittal views - blurring compromises diagnostic quality significantly, especially the top image. The bottom image verges on acceptability for moderate size pathology. The lower coronal image is good and diagnostically acceptable, while there is blurring in the right lower quadrant of the top coronal image.

7.1.4 Low resolution unregularized limited angle reconstructions

Refer to the filtered back-projection and positively iterated limited angle reconstructions in Figure 6.2 on page 98 showing the effect of having more data available:

(a) unusable
(b) better but still not usable
(c) the first barely acceptable image
(d) not useful
(e) barely acceptable for gross pathology
(f) just usable
(g) smoother image but not usable in practice
(h) barely acceptable for gross pathology
(i) reasonable for moderate pathology

7.1.5 Low resolution limited angle reconstructions

On page 115, refer to the $\Delta\theta = 1°$ over $180°$ low resolution full angle reconstruction in Figure 6.13(a), and the $\Delta\theta = 5°$ over $100°$ positively iterated (b), Bayesian (c) and transformation map, (d) to (f), limited angle reconstructions:

(a) very good

(b) unacceptable for any diagnostic purpose

(c) also not usable, limited greyscale and artefactual errors

(d) barely acceptable - about the quality of first-generation 80 pixel matrix CT

(e) slightly better than (d)

(f) minimal improvement on (e), just acceptable

Refer to Figures 6.14 and 6.15 on pages 117 and 118:

Left column: poor - minimal diagnostic value

Middle column: some moderate diagnostic value

Right column: good value

To summarize, the transformation map limited angle reconstructions have functional diagnostic value, while limited angle reconstructions using other methods do not. When the full dataset is available, the quality of the results obtained is comparable to that of leading commercial tomographic systems. When data is sparse, the transformation map method generates reconstructions that are of diagnostic quality.

7.2 Quantitative comparison of results

Throughout this thesis reconstructions are compared to the ground truth with the parameter σ_{1°. This parameter represents the root mean square error of the projections of a particular reconstruction, no matter how it was created, with the complete dataset of measured projections over a 180° range at an angular spacing of $\Delta\theta = 1°$. It is believed that this is the best way of evaluating or comparing reconstructions quantitatively, even though it gives no indication of how natural-looking the reconstuction may be.

Of particular interest is an evaluation of the transformation map method compared to other methods on the low resolution sparse angle and limited angle tomography problems. Figure 7.1 shows a table of the ground truth fitness parameter σ_{1° for the positively constrained iterative method, the Bayesian method and the transformation map method.

This table of results reflects firstly that the quality of limited angle reconstructions are in general significantly worse than of sparse and full angle reconstructions. Secondly, the results using a

Method\Problem	Sparse angle tomography	Limited angle tomography
Positivity constraint	around 101	around 286
Bayesian	91.2 to 92.4	249 to 275
Transformation map	90.7 to 90.8	173 to 188
Ground truth best	78.3 to 79.1	78.3 to 79.1

Figure 7.1 Quantitative comparison of different reconstruction algorithms showing typical values of fit to the ground truth (σ_{1°).

transformation map is numerically superior to the other techniques implemented, especially in the case of limited angle tomography.

7.3 Summary of developments

This thesis addresses and overcomes a number of difficulties in achieving a high degree of accuracy in measurements. The first set of problems solved in this thesis, such as correcting distortion and artefacts in scans, and inferring the three-dimensional geometry of the imaging system, is peculiar to the Lodox machine.

Later developments, such as the strategy devised for effectively constraining a cross-section or volumetric reconstruction to measured projections, have more general applicability.

Of particular emphasis is the chapter on transformation maps. Transformation maps were shown to be a powerful problem solving framework. Its service to image artefact removal and the regularization of tomographic reconstructions is demonstrated. It was in fact the unnatural looking results due to Bayesian methods that prompted research into transformation maps which characterise prior knowledge more precisely.

Overcoming all these diverse problems was instrumental to proceeding with the investigation into limited angle tomography.

A study of the properties of limited angle reconstructions guided in the choice of feature dimensions used in the transformation map. Ultimately in this thesis, the transformation map method is used to solve the limited angle tomographic problem.

7.4 Findings

It was found that for medical cross-sections, there are a limited number of dominant density classes and that the distribution of these densities have a wider spread in limited and sparse angle reconstructions than in full angle reconstructions. Full angle reconstructions are also less noisy and have particular edge sharpnesses at interfaces between different tissue types.

Both for limited and sparse angle tomographic reconstruction, it was found that a transformation map can successfully extract such knowledge quantitatively from examples. It was also found that transformation maps can effectively apply this knowledge during reconstruction to achieve superior results: other regularization methods seem to produce artificial-looking results.

For limited angle reconstructions, it is shown that biasing a reconstruction with correct knowledge in one region results in a corresponding improvement in the reconstruction for a complementary region. Using this property, which is due to the nature of the data constraint, the transformation map biases predictable regions so that improvements occur indirectly in less predictable regions. The iteration of this principle leads to a final reconstruction.

7.5 Results

In terms of full and sparse angle tomography, this thesis has illustrated that the Lodox system can achieve good quality high resolution medical reconstructions. For unregularized limited angle reconstructions, high resolution coronal views are produced which resemble those of full angle reconstructions well.

It is shown that reconstruction quality improves significantly the bigger the angular range of available projection data. However, this thesis is only concerned with the special case of having projections at 5° intervals over a 100° range. For this situation, considerable numerical as well as visual improvements in reconstruction quality over other methods are demonstrated using transformation maps.

7.6 Conclusions

In response to the research question, it is concluded that the Lodox machine can be used to create limited angle reconstructions that have diagnostic value. However, these reconstructions are not as precise and therefore do not have as much diagnostic value as full angle reconstructions.

At some point in reducing the amount of available data, a reconstruction becomes medically of no diagnostic value. This thesis presented a technique that produces useful results from severely restricted measurements when other methods do not.

7.7 Recommendations

This study may have provoked many questions such as 'how does the algorithm perform at higher resolutions?' or 'what about a bigger or even smaller angular range?'. It would be interesting to know what the minimal data requirements are in terms of angular resolution and range to achieve diagnosable reconstructions. These questions are not answered in this thesis and warrant further investigation.

It seems reasonable to presume that more elaborate forms of prior knowledge which specify properties of interfaces between tissue types, particularly at higher reconstruction resolutions, could advance reconstruction. Also, volumetric limited angle reconstructions (as opposed to individual cross-sections), could benefit from information of neighbouring cross-sections and improve results. Perhaps it would be wise to add a local edge magnitude dimension in the scanning direction to the transformation map in that case.

Given the time required to aquire scans of the patient, the best medical application for sparse, full and limited angle tomography using the Lodox system is perhaps forensic medicine. In a linear scanning system such as Lodox, patient movement would have more severe consequences to reconstruction precision than in spiral computerized tomography where slices are scanned sequentially.

Since medical imaging inherently places the utmost demands on imaging quality and scanning speeds, it is advisable to investigate limited angle tomography using the technology developed in this thesis for certain industrial non-destructive testing applications or electron microscopy where limited angle tomography is the only alternative.

REFERENCES

Andersen, A. H. (1989). Algebraic reconstruction in CT from limited views. *IEEE Trans. Medical Imaging* **8**(1), 50–55.

Andia, B. I., K. D. Sauer and C. A. Bouman (2002). Nonlinear backprojection for tomographic reconstruction. *IEEE Trans. on Nuclear Science* **49**(1), 61–68.

Beningfield, S., H. Potgieter, A. Nicol, S.van As, G. Bowie, E. Hering and E. Latti (2003). Report on a new type of trauma full-body digital X-ray machine. *Emergency Radiology* **10**, 23–29.

Beutel, J., H. Kundel and R. Van Metter (2000). *Handbook of Medical Imaging.* Vol. 1. SPIE Press.

Bontekoe, T., E. Koper and D. Kester (1994). Pyramid maximum entropy images of IRAS survey data. *Astron. Astrophys.* **284**, 1037–1053.

Boyd, J. E. and J. J. Little (1994). Complementary data fusion for limited-angle tomography. *Proceedings of Computer Vision and Pattern Recognition* **94**, 288–294.

Chen, C. et al. (1991). Incorporation of structural CT and MR images in PET image reconstruction. *SPIE Image Processing* **1445**, 222–225.

Coric, S., M. Leeser and E. Miller (2002). Parallel-beam backprojection: An FPGA implementation optimized for medical imaging. *FPGA* pp. 217–226.

de Villiers, M. S. (1999). Information content of artifacts in tomography images. In: *Proceedings of Pattern Recognition Association of South Africa.* Vol. 10.

de Villiers, M. S. (2004). Method of reconstructing tomographic images. United States Patent US6,788,758B2.

de Villiers, M. S. and G. de Jager (2001). Removal of camera overlap artifacts in the Lodox system. In: *Proceedings of Pattern Recognition Association of South Africa.* Vol. 12. pp. 95–98.

de Villiers, M. S. and G. de Jager (2003). Detective quantum efficiency of the Lodox system. In: *SPIE Medical Imaging.* Vol. 5030. pp. 955–960.

de Villiers, M.S. (2000). Limited Angle Tomography. Electrical engineering MSc thesis. University of Cape Town.

Delaney, A. H. and Y. Bresler (1995). Efficient edge-preserving regularization for limited-angle tomography. *In Proc. Internat. Conf. Image Processing* **3**, 176–179.

Denisova, N. V. (1998). Maximum-entropy-based tomography for gas and plasma diagnostics. *J. Phys. D: Appl. Phys.* **31**, 1888–1895.

Fielden, J., H. Kwong and J. Wilbrink (1991). Reconstructing MR images from incomplete Fourier data using the maximum entropy method. *SPIE Image Processing* **1445**, 145–154.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(6), 721–741.

Granfors, P. R. and R. Aufrichtig (2000). Performance of 41 x 41 cm^2 amorphous silicon flat panel x-ray detector for radiographic imaging applications. *Medical Physics* **6**, 1324–1331.

Grossman, R. I. and D.M. Yousem (1994). *Neuroradiology. The Requisites.*. Mosby.

Gureyev, T. E. and R. Evans (1998). An extension of quasi-one-dimensional tomography. *Applied Optics* **37**(13), 2628–2636.

Hanson, K. and G. Wecksung (1983). Bayesian approach to limited-angle reconstruction in computed tomography. *Optical Society of America* **73**, 1501–1509.

Hsiao, I., A. Rangarajan and G. Gindi (2001). Bayesian image reconstruction for transmission tomography using mixture model priors and deterministic annealing algorithms. In: *SPIE Medical Imaging*. Vol. 4322. pp. 899–908.

IEC 62220-1. Ed 1 (2003). Medical electrical equipment - characteristics of digital x-ray imaging devices - part 1: Determination of the detective quantum efficiency. *International Electrotechnical Commission.*

Jaffe, J. S. (1990). Limited angle reconstruction using stabilized algorithms. *IEEE Trans. Medical Imaging* **9**(3), 338–344.

Jang, J., C. Sun and E. Mizutani (1996). *Neuro-Fuzzy and Soft Computing: a computational approach to learning and machine intelligence.* Prentice Hall.

Kak, A. C. and M. Slaney (1988). *Principles of Computerized Tomographic Imaging.* IEEE Press.

Kolehmainen, V., S. Siltanen, S. Jarvenpaa, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttila and E. Somersalo (2003). Statistical inversion for medical x-ray tomography with few radiographs: II. application to dental radiology. *Phys. Med. Biol.* **48**, 1465–1490.

Kwee, I. W., Y. Tanikawa, S. Proskurin, S.R. Arridge, D. T. Delpy and Y. Yamada (1997). Performance of a nullspace-MAP image reconstruction algorithm. *Proceedings of the SPIE - International Society for Optical Engineering* **2979**, 185–196.

Lawrence, M. C. (1989). The application of the maximum entropy method to electron microscopic tomography. *Ultramicroscopy* **31**, 285–302.

Leon-Garcia, A. (1994). *Probability and Random Processes for Electrical Engineering.* Addison-Welsey.

Li, S. Z. (1995). *Markov Random Field Modeling in Computer Vision.* Springer.

McCauley, T. G., A. Stewart, M. Stanton, T. Wu and W. Phillips (2000). Three-dimensional breast image reconstruction from a limited number of views. In: *International Conference on Medical Imaging.*

Mohammad-Djafari, A., K. Sauer, Y. Khayi and E. Cano (1997). Reconstruction of the shape of a compact object from few projections. *IEEE Comput. Soc.* **1**, 165–168.

Narayan, R. and R. Nityananda (1986). Maximum entropy image restoration in astronomy. *Ann. Rev. Astron. Astrophys.* **24**, 127–170.

Oakley, J., J. Missimer and G. Szekely (2000). Better anatomical priors for PET reconstruction. In: *Medical Imaging Understanding Analysis.* Vol. MI-3.

Olson, T. and J. S. Jaffe (1990). An explanation of the effects of squashing in limited angle tomography. *IEEE Trans. Medical Imaging* **9**(3), 242–246.

Oskoui-Fard, P. and H. Stark (1989). A comparitive study of three reconstruction methods for a limited-view computer tomography problem. *IEEE Trans. Medical Imaging* **8**(1), 43–49.

Park, K. (2002). Finite Element Analysis of Lodox System. Biomedical engineering MSc thesis. University of Cape Town.

Perona, P. and J. Malik (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Mahine Intelligance* **12**(7), 629–639.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1999). *Numerical recipes in C: The art of scientific computing, second edition.* Cambridge University Press.

Prince, J. L. and A. S. Willsky (1993). Hierarchical reconstruction using geometry and sinogram restoration. *IEEE Trans. Image Proc.* **2**(3), 401–416.

Rangarajan, A., I. T. Hsiao and G. Gindi (1999). A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. *Journal of Mathematical Imaging and Vision.*

Reeds, J. A. and L. A. Shepp (1987). Limited angle reconstruction in tomography via squashing. *IEEE Trans. Medical Imaging* **MI-6**(2), 89–97.

Samei, E., M.J. Flynn and D.A. Reimann (1998). A method for measuring the presampled mtf of digital radiographic systems using an edge test device. *Medical Physics* **25**, 102–113.

Samsonov, A. and C. Johnson (2002). Noise-adaptive anisotropic diffusion filtering of MRI images reconstructed by SENSE (sensitivity encoding) method. In: *IEEE International Symposium on Biomedical Imaging: Macro to Nano.* pp. 701–704.

Semeter, J. and M. Mendillo (1997). A nonlinear optimization technique for ground-based atmospheric emission tomography. *IEEE Trans. Geoscience and Remote Sensing* **35**(5), 1105–1116.

Sezan, M. I. and H. Stark (1984). Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion. *IEEE Trans. Medical Imaging* **MI-3**(2), 91–98.

Sheelke, M., J.H. Potgieter and M. S. de Villiers (2005). System characterization of the STATSCAN full body slit scanning radiography machine: Theory and experiment. In: *SPIE Medical Imaging.*

Siltanen, S., V. Kolehmainen, S. Jarvenpaa, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttila and E. Somersalo (2003). Statistical inversion for medical x-ray tomography with few radiographs: I. general theory. *Phys. Med. Biol.* **48**, 1437–1463.

Skilling, J. (1998). Probabilistic data analysis: An introductory guide. *Journal of Microscopy* **190**, 28–36.

Skilling, J. and R. K. Bryan (1984). Maximum entropy image reconstruction: general algorithm. *Mon. Not. R. Astr. Soc.* **211**, 111–124.

Starke, P. L. and G. de Jager (1998). Further work on limited angle tomography with Lodox. Technical report. Digital Image Processing Laboratory, University of Cape Town.

Stierstorfer, K. and M. Spahn (1999). Self-normalizing method to measure the detective quantum efficiency of a wide range of x-ray detectors. *Medical Physics* **26**, 1312–1319.

Tam, K. C. and V. Perez-Mendez (1981). Tomographical imaging with limited-angle input. *J. Opt. Soc. Am.* **71**(5), 582–592.

Tam, K. C., W. Eberhard and K. W. Mitchell (1990). Incomplete-data CT image reconstructions in industrial applications. *IEEE Trans. Nuclear Science* **37**(3), 1490–1499.

Tomasi, C. and R. Manduchi (1998). Bilateral filtering for gray and color images. In: *Proceedings of the 1998 IEEE International Conference on Computer Vision.*

Vengrinovich, V., Y. Denkevich and G. Tillack (1998). Limited projections and views Bayesian 3D reconstruction using Gibbs priors. In: *Proceedings of the 7th European Conference on Non-Destructive Testing.*

Wilson, D. W. (1998). Decomposition of images and objects into measurement and null components. *Optics Express* **2**(6), 254–260.

Wong, S. H. and S. F. Yau (1998). Linear neural network for the solution of limited angle problems in computer-aided tomography. *J. Electronic Imaging* **7**(1), 70–78.

Yagle, A. E. (1998). An explicit closed-form solution to the limited angle discrete tomography problem for finite-support objects. *International Journal of Imaging Systems and Technology* **9**(2-3), 174–180.