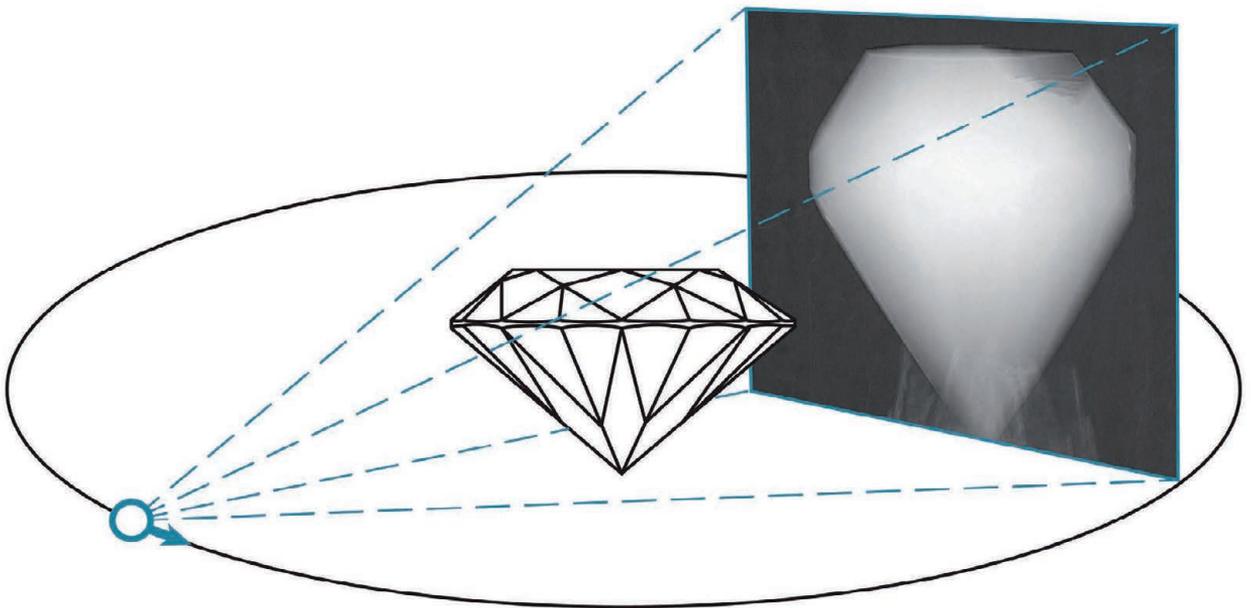


Tomographic segmentation and discrete tomography for quantitative analysis of transmission tomography data

Proefschrift voorgelegd tot het behalen van de
graad van Doctor in de Wetenschappen aan
de Universiteit Antwerpen te verdedigen door

Wim van Aarle



PROMOTOREN

Prof. dr. Jan Sijbers

Prof. dr. K. Joost Batenburg

Faculteit Wetenschappen
Departement Fysica
Antwerpen 2012

 Universiteit
Antwerpen



Faculteit Wetenschappen
Departement Fysica

Tomographic segmentation and discrete tomography for quantitative analysis of transmission tomography data

Tomografische segmentatie en discrete tomografie voor kwantitatieve analyse van transmissie tomografie data

Proefschrift voorgelegd tot het behalen van de graad van

Doctor in de Wetenschappen

aan de Universiteit Antwerpen, te verdedigen door

Wim VAN AARLE

Promotoren:

Prof. dr. Jan Sijbers

Prof. dr. K. Joost Batenburg

Antwerpen, 2012

Doctoral committee:

Prof. dr. Sara Bals
Prof. dr. K. Joost Batenburg
Prof. dr. Michel Defrise
Prof. dr. Jan Sijbers
Prof. dr. Sabine Van Doorslaer
Prof. dr. Luc Van Hoorebeke

Contact information:

✉ **Wim van Aarle**
Vision Lab, Dept. of Physics
University of Antwerp (CDE)
Universiteitsplein 1, Building N1.14
B-2610 Wilrijk, Antwerpen, Belgium

☎ +32 (0) 3 265 28 40

📠 +32 (0) 3 265 22 45

✉ wim.vanaarle@gmail.com

🌐 <http://visielab.ua.ac.be/people/wim-van-aarle>

Summary

Computed Tomography (CT) is a non-destructive imaging technique that allows visualization of interior structures. With CT, an object can be virtually reconstructed based on multiple X-ray projection images, recorded at different directions. Many applications benefit from CT: medical diagnostics, drug trials on small animals, price-optimization of diamonds, materials science, etc.

Conventional reconstruction algorithms (such as FBP and SIRT) are often not suited for accurate analysis as reconstruction artefacts may impede the creation of high quality images. In X-ray CT, only a limited number of projection images can be acquired due to the radiation exposure to the scanned object. Also, if an object to be scanned is too large to fit inside the field of view of the scanner, the projection data is truncated.

Prior to quantitative image analysis, a segmentation algorithm is often performed on the reconstructed images. However, given that this process is posterior to the reconstruction (Fig. 3a), the segmented images are likely to suffer from the same artefacts as the reconstructed images.

In this work, novel approaches are introduced to create high quality segmented images from as little projection data as possible. These approaches are divided into two categories, corresponding to the two major parts in this work. In Part I, several *tomographic segmentation* methods are proposed. These methods exploit available projection data to enhance the segmentation accuracy (Fig. 3b). In Part II, *discrete tomography* methods are discussed. These techniques combine reconstruction and segmentation into a single algorithm (Fig. 3c) and exploit prior knowledge about the scanned objects (e.g. their grey levels) to create highly accurate reconstructions from a limited number of projections.

Part I: Tomographic Segmentation

Global thresholding is a widely applied method to create segmented images. The quality of the resulting segmentations is dependent on the threshold values, the choice of which is commonly based solely on the reconstructed image. In the first part of this work, tomographic segmentation algorithms are presented that make use of the available projection data to estimate the optimal threshold values more accurately than classical methods do.

In Chapter 2, the *Projection Distance Minimization (PDM)* method is discussed. In PDM, the optimal segmentation is assumed to be the one whose forward projection is the closest to the measured projection data. To find this minimum, several optimization strategies are compared.

In Chapter 3, the *Segmentation Inconsistency Minimization (SICM)* method is

SUMMARY

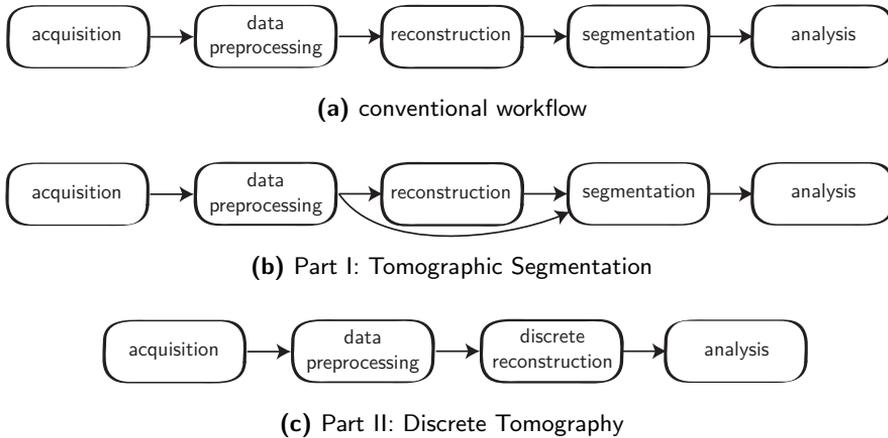


Figure 3: (a) Conventional workflow of a typical tomographic application. (b) Workflow with tomographic segmentation. (c) Workflow with discrete tomography.

proposed. In SICM, a single threshold value is estimated that separates the most dense material from all other materials. The optimal dense object segmentation is the one for which the residual projection data — the forward projection of the part of the image that does not belong to the dense object — is minimally inconsistent. Contrary to PDM, where each image pixel is classified into one of only a few distinct grey levels, SICM can be applied if the non dense objects are non-homogeneous.

Part II: Discrete Tomography

Discrete reconstruction methods for tomography limit the set of possible reconstructions to those that contain only a small number of distinct grey levels. High quality reconstructions can then be computed from substantially fewer projection directions, thus reducing the radiation exposure to the scanned object.

In Chapter 4, an iterative discrete tomography method, called *Discrete Algebraic Reconstruction Technique (DART)*, is discussed. By exploiting prior knowledge about the grey levels of each of the scanned materials, DART often results in high quality reconstructed images even if only a few projection angles are measured, if the angular range is limited, or if the projection data is truncated.

A key problem when applying DART to experimental datasets is its assumption that the set of grey levels in the unknown reconstructed image is known (Fig. 4a). Obtaining such knowledge in practice is non-trivial. Even if the attenuation value of each material is known in advance, accurate hardware calibration is required to translate these values into grey levels to be used in the reconstruction. Such

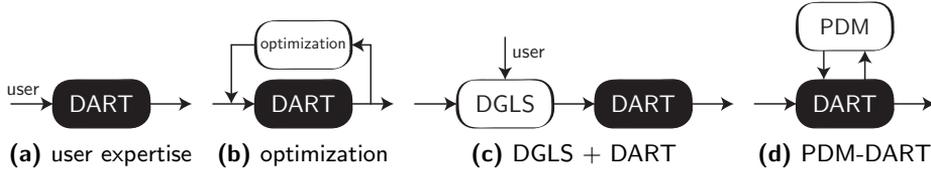


Figure 4: Schematic overview of various methods to estimate the grey level for use in DART. **(a)** *User* expertise is used to set the correct parameters. **(b)** An optimization strategy is used to automatically minimize the Euclidean distance between the forward projection of the DART reconstructed image and the measured projection data. **(c)** The DGLS method is used *prior* to DART to semi-automatically estimate the correct grey levels. **(d)** The PDM segmentation strategy is used *within* each DART iteration to automatically estimate the optimal grey levels and threshold values.

a calibration depends on various properties of the scanning system, which are typically not available to the user.

In Part II of this work, several schemes are presented for automatic grey level estimation.

- In Section 4.4, the Euclidean distance between the forward projection of a DART reconstruction and the measured projection data is proposed as a cost function to evaluate a chosen set of grey levels. A global optimization strategy (e.g. simplex search or adaptive surrogate modelling) can then be applied to find the optimal grey levels (Fig. 4b).
- In Chapter 5, the *Discrete Grey Level Selection (DGLS)* approach is presented for semi-automatic estimation of the grey levels. In DGLS, the user first selects an image region that can be expected to correspond to a homogeneous region in the original object. In certain cases, knowledge of such a constant region allows for reliable estimation of the grey level corresponding to the selected region. The results of DGLS can subsequently be used to run DART (Fig. 4c).
- In Chapter 6, an extension to DART is proposed. *Projection Distance Minimization DART (PDM-DART)* combines discrete reconstruction with fully automatic grey level estimation. PDM-DART adaptively selects the optimal grey levels within each DART iteration such that the Euclidean distance between the forward projection of the intermediate segmented image and the measured projection data is minimal.

In CT, accurate segmentation of structures that are small with respect to the reconstruction pixel size, poses a very complex and difficult problem as recon-

SUMMARY

structured images often lack contrast due to a *partial volume effect (PVE)*. High resolution projection acquisition can provide a solution, but is often not feasible due to X-ray dose limitations, limited scanning time or hardware constraints. The conventional approach to reduce PVEs without increasing the X-ray dose is to upsample the reconstruction voxel grid, allowing for a more accurate representation and potentially improving the overall visualization of small structures. This upsampling is also known as *super-resolution* and typically results in a limited data reconstruction problem: the number of equations (measured projection data) remains the same while the number of unknowns (reconstruction voxels) increases significantly.

In Chapter 7, a novel super-resolution approach is proposed to improve the detection of small structures in low resolution CT acquisitions. By incorporating prior knowledge about the object materials to compensate for the lack of high resolution projection data (i.e. to apply DART), the proposed approach effectively increases the spatial resolution of the tomographic reconstructions.

Part III: Conclusions and Appendices

In Chapter 8, conclusions are drawn. Subsequently, Appendix A provides a concise introduction to the tomographic software framework that was developed at the Vision Lab of the University of Antwerp.

—*Lasciate ogni speranza, voi ch'entrate.*
Dante Alighieri, *Inferno*, 1309.



Introduction

Contents

1.1	A brief history of transmission tomography	2
1.2	The physics of transmission CT	5
1.2.1	X-ray generation	5
1.2.2	X-ray–matter interaction	7
1.2.3	X-ray detection	8
1.3	The mathematics of transmission CT	9
1.3.1	The projection model	9
1.3.2	Analytical reconstruction methods	13
1.3.3	Algebraic reconstruction methods	18
1.4	Typical CT workflow	21
1.4.1	Data acquisition	21
1.4.2	Data preprocessing	24
1.4.3	Reconstruction	27
1.4.4	Segmentation	29
1.4.5	Analysis	29
1.5	Tomography on all scales, application overview	29
1.5.1	Materials science	30
1.5.2	Non-destructive testing	31
1.5.3	Biomedical	32
1.5.4	Medical	32
1.5.5	Large scale	33
	References	34

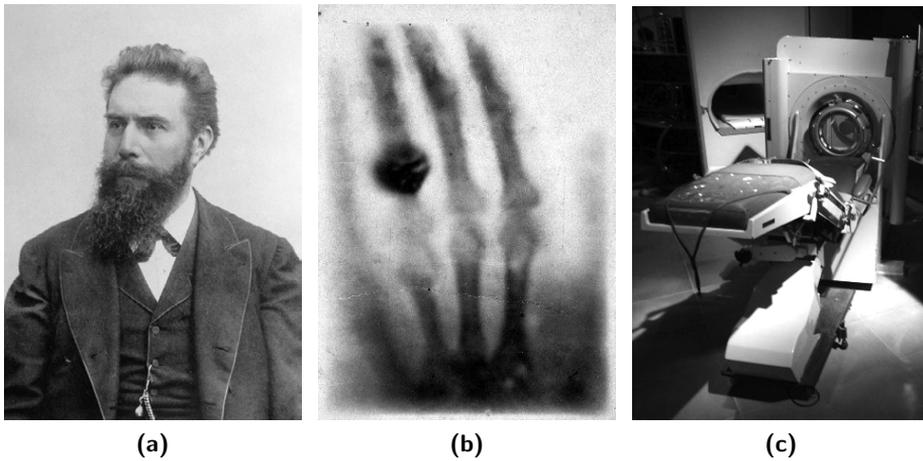


Figure 1.1: (a) Wilhelm Conrad Röntgen (1845-1923) (b) An X-ray image of the hand of Anna Bertha, Röntgen's wife. (c) One of the first CT scanners at the EMI central research laboratories. Designed by Sir Godfrey Hounsfield (1919-2004).

In this chapter, a concise introduction is given to the topic of X-ray physics and transmission Computed Tomography imaging (CT). The term tomography is derived from two Greek words: “*tomos*”, meaning section, and “*graphein*”, meaning to write, i.e. tomography concerns the visualization of slices through certain objects.

A brief history of transmission tomography is given in Section 1.1. In Section 1.2, light is thrown upon the physical background of tomography such as X-ray generation and photon attenuation. Following, the mathematical background and the tomographic reconstruction algorithms are discussed in Section 1.3. In the ensuing Section 1.4, the typical workflow of a tomographic application — from the scanning of an object to its analysis — is illustrated. Special attention is given to what can go wrong, the so-called tomographic *artefacts*. Ultimately, Section 1.5 discusses a selection of the various applications where computed tomography is commonly applied.

1.1 A brief history of transmission tomography

The German physicist Wilhelm Conrad Röntgen (Fig. 1.1a) is generally considered the father of medical radiology imaging. In 1895, while experimenting with various types of vacuum tubes, he discovered X-rays. When, in a fully darkened room, he passed an electrostatic charge through a Hittord-Crookes tube entirely covered

1.1. A BRIEF HISTORY OF TRANSMISSION TOMOGRAPHY

by a black cardboard, he noted a fluorescent effect on a barium platinocyanide screen lying on a bench about a metre away from the vacuum tube. With further experiments, he noted that these X-rays — X for unknown — can penetrate objects where light cannot and that their absorption depends on the density of the object. This led to the advent of radiographies, or X-ray images, which are still commonly used in modern medicine. As radiographies allow visualization of the inside of a patient without the use of a knife, it is generally called a *non-invasive image modality*. One of the first radiographies is shown in Fig. 1.1b, an X-ray image of Röntgen's wife's hand. Röntgen's findings were published in [1], for which he was awarded the very first Nobel Prize in Physics in 1901.

X-ray images are very useful for looking inside opaque objects. However, from a single such image, it is not possible to know the exact position of each internal structure. Consider an X-ray image taken of a human head with a tumour inside. Is the tumour in the centre of the head or is it more near the skull? To answer this question, multiple X-ray images should be taken, each from a different perspective, i.e. with a different location of the X-ray source and detector. However, even then, exact information about the object shape is not available. For that, *transmission Computed Tomography (CT)* had to be invented.

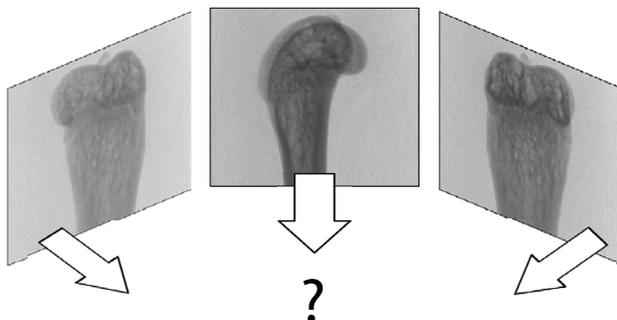


Figure 1.2: Schematic overview of the inverse problem. Given multiple X-ray images of an object, how to reconstruct the interior of the object?

The mathematical foundation for tomography was developed in 1917 by the Austrian mathematician Johann Radon (1887-1956) [2]. He introduced a transformation operator, called the *Radon transform*, that defines a projection of a certain two-dimensional object function, $f(x, y)$, by an infinite series of straight line integrals. He also specified the *inverse Radon transform* and noted that, if the Radon transform is known for each projection direction, $f(x, y)$ can be exactly reconstructed. This is called the *inverse problem* or *reconstruction problem* (Fig. 1.2). The Radon transform and its inverse are further introduced in Section 1.3.1.2.

CHAPTER 1. INTRODUCTION

Due to the computational complexity of the Radon transform, it took another 40 years before solutions to the inverse problem were applied to X-ray images. In 1963, the physicist Allan MacLeod Cormack (1924-1998) laid out the mathematical basics for an implementation of tomographic reconstruction. In 1971, Sir Godfrey Hounsfield (1919-2004), a British engineer working at the EMI research laboratories¹, validated Cormack's work by building a prototype CT scanner (Fig. 1.1c). In such a scanner, an X-ray source and X-ray detector rotate round the object or patient, taking radiographic images under several angles. For their contribution to medicine, both Cormack and Hounsfield were awarded the Nobel Prize in Physiology in 1979.

In the first generation of CT scanners, the spatial resolution was very low and the aperture (the diameter of the opening) was so small that it was only suited for head imaging. Also, they used a translating source and detector, called parallel beam projection (Fig. 1.3a). This resulted in long scanning times and high radiation dose to patients. Subsequent generations of scanners have seen great advances in the spatial resolution (up to sub-millimetre precision) and great reductions in scanning time and radiation dose. The scanning geometry has evolved from the primitive parallel beam projection to fan beam projection (Fig. 1.3b), to circular cone beam projection (Fig. 1.3c), and to helical cone beam projection (Fig. 1.3d). These different scanning geometries (also *projection geometries*) are subject to further introduction in Section 1.3.1.1.

CT has also been applied outside the field of medicine. In the 1950s, tomographic principles were already being used to map the regions of emitted microwave radiation from the disk of the sun. For visualization of small objects, table-top μ CT-scanners have been introduced that reach a very high spatial resolution. They are of use in biomedical research and in materials science. Other imaging systems where tomography can be applied include electron microscopes (even up to atomic resolution [3]) and at synchrotron facilities (high energy X-ray beam lines). A more in-depth overview of various applications that can benefit from the use of CT is given in Section 1.5.

Recent research efforts have focused, amongst others, on dual energy systems, where two X-ray sources with a different energy spectrum are combined to generate reconstructions with enhanced contrast between tissue types [4]; on spectral imaging and photon counting detectors [5] that in the future might eliminate beam hardening artefacts [6]; on phase contrast imaging [7]; on improved reconstruction methods that generate highly accurate reconstructions from very noisy projection data [8]; on dynamic (4D) imaging to visualize cardiac movement; etc.

¹EMI is also well known for its popular record label. It could therefore be argued that the founding fathers of computed tomography were in fact The Beatles as their record sales substantially increased the budget of the EMI company.

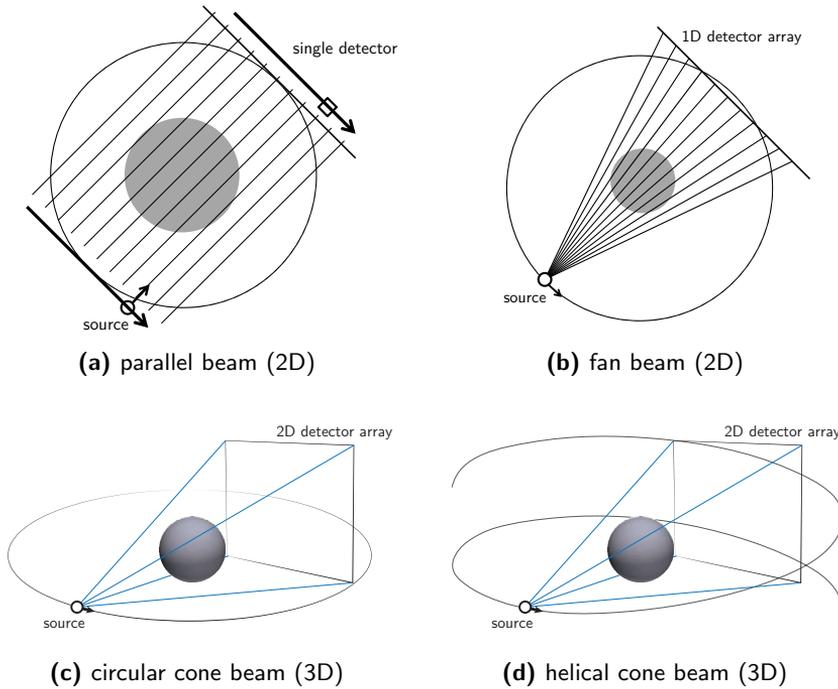


Figure 1.3: Various scanning geometries.

1.2 The physics of transmission CT

This section investigates the physical processes that occur when an X-ray projection image is created. In Fig. 1.4, one can observe three major phases. (1) X-rays are generated, (2) X-rays interact with matter when traversing an object, and (3) X-rays are measured in a detector cell. Note that only the processes that occur in common medical applications are considered here. Other tomographic imaging modalities, e.g synchrotron radiation, are not discussed.

1.2.1 X-ray generation

X-rays are a form of electromagnetic radiation emitted by charged particles (usually electrons) in changing atomic energy levels or in slowing down in a Coulomb force field. They have a wave-length which is smaller than that of UV-light and are typically generated in a vacuum tube in which a cathode is placed at one side and an anode — commonly made of tungsten — at the other (Fig. 1.5).

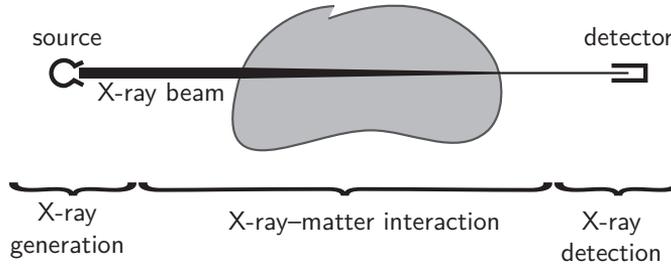


Figure 1.4: Schematic overview of the path of an X-ray beam, from source to detector.

When a high voltage is applied between the cathode and the anode of such an X-ray tube, electrons are accelerated from the cathode to the anode and hit the anode at very high speed. In the anode material the fast electrons undergo multiple interactions with electrons and nuclei and lose their energy, mostly gradually. Most of this energy dissipates as heat, as such the anode has to be well-cooled. However, in the energy loss processes also X-rays can be produced in the following ways:

- If the energy transfer in an interaction with an electron of an anode atom is large enough, ionisation of the anode atom can occur. If an electron is removed from an inner shell, an outer shell electron can fill its gap, thereby generating an X-ray photon. The energy of this photon equals the difference in binding energy of the two shells. These so-called characteristic X-rays give rise to peaks in the emitted X-ray spectrum.
- If a fast electron is deflected and decelerated while interacting with a nucleus, a photon is emitted with an energy equal to the kinetic energy lost by the electron. This effect is called Bremsstrahlung. The energy of the emitted photon can have any value between zero and the initial fast electron energy, depending on the distance between the electron and the nucleus. This effect gives rise to a continuous contribution in the emitted X-ray spectrum, with a maximum photon energy equal to the kinetic energy of the electrons hitting the anode, hence depending on the applied high voltage.

These processes keep on occurring as long as the electric current is applied, thus generating an X-ray beam. Each X-ray beam has a certain intensity, a value proportional to the number of photons. This intensity is not fixed along the path of the X-ray beam. In the next subsection it is shown that the intensity decreases as the beam passes through a material. Let $I(\xi)$ denote the intensity of the X-ray beam at the point ξ along the path of the beam. Define $I(0) = I_0$, the *incident beam intensity*, the intensity of an X-ray beam at its source.

1.2. THE PHYSICS OF TRANSMISSION CT

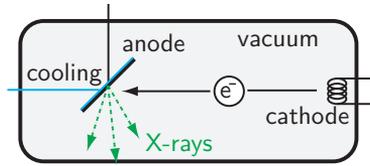


Figure 1.5: Schematic visualization of an X-ray tube.

It should be noted that the generated photons of a beam have a variable energy level. Such an X-ray beam is referred to as a *polychromatic X-ray beam*. This is in contrast to a *monochromatic X-ray beam*, in which all photons have the same energy. For polychromatic X-ray beams, define $I_0(E)$, the incident beam intensity of photons with energy E .

1.2.2 X-ray–matter interaction

After an X-ray photon is generated, it traverses the object in the scanner. Common effects that can occur include:

- The **photoelectric effect**. The photon interacts with an atom and is fully absorbed. As a result a photoelectron is ejected from one of the shells of the atom, whereby the kinetic energy of the electron equals the initial photon energy minus the binding energy of the electron in the atom. As such, the incoming photon energy has to be larger than the electron binding energy.
- **Compton scattering**. The incoming photon interacts with an electron (assumed to be at rest) and changes direction, whereby part of its energy is transferred to the electron. After the process one obtains a scattered photon and a recoil electron. The energy transferred to the electron depends on the scattering angle and the incoming photon energy and has a value between zero and a fraction of the incoming photon energy.
- **Pair production**. In the Coulomb field of a nucleus, a photon with an energy which exceeds twice the electron rest mass energy (i.e. 1022 keV) disappears and an electron-positron pair is created. The excess photon energy above 1022 keV appears as kinetic energy of the electron and positron. The required photon energy for this process is much higher than that being used in medical imaging².

²Note that even though CT is commonly referred to as “non-invasive imaging modality”, it is by no means completely safe as the X-ray photons effectively alter the atoms of the patient. A high radiation dose — due to high energy X-ray photons and/or a long exposure time — therefore severely increases the chance of destroying cells, and thus also of cancer incidence. It

CHAPTER 1. INTRODUCTION

The likelihood of any of these three effects happening for a monochromatic beam at a certain point ξ along its path is correlated to the *attenuation factor*. This value depends on the atomic number and the density of the material at that location. Let $\mu(\xi)$ denote the attenuation factor at point ξ .

Consider a section of the X-ray beam of length $\Delta\xi$ and assume that the attenuation factor of the object is constant in this section. The intensity of the beam at the start of this section, $I(\xi)$, is then decreased by a fraction that depends on that attenuation factor. The intensity at the end of the section, $I(\xi + \Delta\xi)$, thus becomes:

$$I(\xi + \Delta\xi) = I(\xi) - \mu(\xi)I(\xi)\Delta\xi. \quad (1.1)$$

Taking the limit $\Delta\xi \rightarrow 0$ of Eqn. 1.1 yields the total intensity change at point ξ :

$$\frac{dI}{I(\xi)} = -\mu(\xi)d\xi. \quad (1.2)$$

1.2.3 X-ray detection

If an X-ray photon has passed through the object without being absorbed, it is measured by an X-ray detector. Various types of detectors exist: gas detectors, solid-state scintillator detectors, solid-state flat-panel detectors, etc. Rather than describing how these detectors actually work — which is explained in great detail in works such as [9] — this section focusses on what is actually measured, i.e. the intensity of the X-ray beam at the detector point.

Consider a monochromatic beam. Integration of Eqn. 1.2 along the entire X-ray beam yields the measured intensity I :

$$I = I_0 e^{-\int \mu(\xi)d\xi}. \quad (1.3)$$

Eqn. 1.3 is called the law of Beer-Lambert. One should note that, due to the exponential nature of Eqn. 1.3, the measured intensity has a non-linear connection to the attenuation coefficients. For the reconstruction of the attenuation values (Section 1.3), this is very inconvenient. Therefore, the measured intensity data must first be linearized. Define the measured attenuation A , also referred to as *log-corrected* data, by

$$A = -\ln\left(\frac{I}{I_0}\right) = \int \mu(\xi)d\xi. \quad (1.4)$$

Eqn. 1.3 and Eqn. 1.4 assume a monochromatic beam, i.e. beams where each photon has the same energy. While some radiation sources indeed provide

is thus crucial to use algorithms that create accurate images from low dose scans.

1.3. THE MATHEMATICS OF TRANSMISSION CT

monochromatic beams (e.g. synchrotrons and electron microscopes), polychromatic beams are by far more common. For such beams, Eqn. 1.3 must be adjusted accordingly:

$$I = \int I_0(E) e^{-\int \mu(\xi, E) d\xi} dE. \quad (1.5)$$

In Eqn. 1.5, $\mu(\xi, E)$ denotes the attenuation factor at point ξ of a beam with the energy E .

1.3 The mathematics of transmission CT

Once the attenuation values are measured, an image representing the original object can be constructed. To do so, several methods, called *reconstruction algorithms*, are available. Most reconstruction algorithms can be divided into two classes.

- In **analytical reconstruction techniques** various analytical properties of the Radon transform are used to reconstruct the original object function. These methods are computationally efficient and are therefore commonly used. They are described in Section 1.3.2.
- **Algebraic reconstruction techniques** consider the reconstruction problem as the solving of a system of linear equations. They are very powerful but also very computationally demanding. Recently, the rapid increase of available computational power has made algebraic methods feasible. They are portrayed in Section 1.3.3.

In Section 1.3.1, a mathematical model for the projection operation is discussed.

1.3.1 The projection model

Consider the *object function domain*, an orthogonal two-dimensional Cartesian coordinate system (x, y) in which a 2D object function $f(x, y)$ is defined (Fig. 1.6). The values of this function represent the attenuation coefficients μ of the scanned object at the corresponding position.

A single projection of a two-dimension function under the projection direction θ , is a one-dimensional function³. Define this function as $p_\theta(t)$, the *detector function*. Consider the *projection domain*, the coordinate system (θ, t) , in which the detector functions of $f(x, y)$ are stored for $\theta \in [0, 2\pi)$ and $t \in \mathbb{R}$.

³Likewise, for three-dimensional object functions, a single projection image is a two-dimensional function.

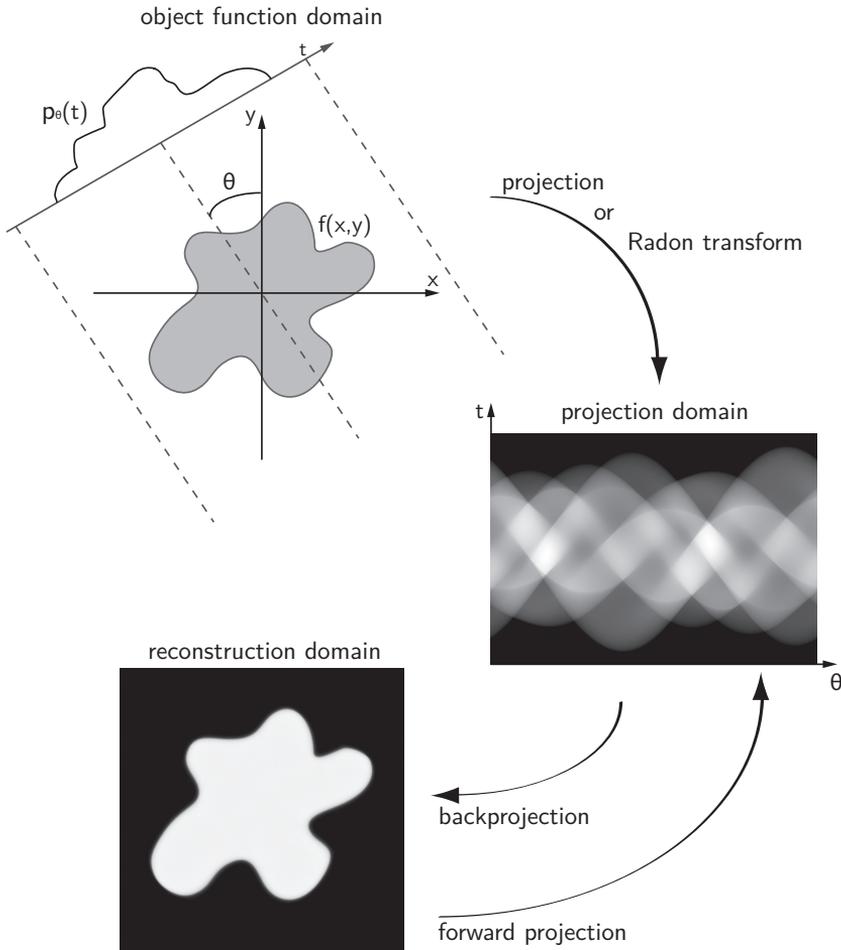


Figure 1.6: Overview of the different domains, and their relation to each other, in which CT data can be visualized.

1.3.1.1 The projection geometry

The exact formula that transforms a function in the object function domain to a function in the projection domain depends on the *projection geometry*, i.e. the positioning and trajectory of the X-ray source and detector in the object function domain. Here, some commonly used projection geometries are introduced. In case of a two-dimensional object function or a three-dimensional object function that is scanned slice-by-slice:

1.3. THE MATHEMATICS OF TRANSMISSION CT

- **Parallel beam projection** (Fig. 1.3a). In each projection direction θ , all projection rays (the rays between the source and a point on the detector function) are parallel to each other. To measure the detector function, the source has to translate on a line perpendicular to the rays. Thereafter, the source and detector rotate around $f(0,0)$, the centre of rotation.
- **Fan beam projection** (Fig. 1.3b). For each projection direction θ , the source is at a fixed position and the projection values are measured on a one-dimensional detector array. The source and detector plate then rotate around the centre of rotation, measuring the detector function at each step. Note that a parallel beam projection geometry is effectively a special case of a fan beam geometry: one where the source is at an infinite distance from the center of rotation. Also note that by interpolation it is possible to convert data from a fan beam projection geometry into data from a parallel beam projection geometry and vice versa. This process is called *rebinning* and is usually performed to apply a reconstruction technique that is suited for only one particular projection geometry.

If the object function, $f(x, y, z)$, is three-dimensional, other geometries are possible:

- **Circular cone beam projection** (Fig. 1.3c). A projection image is measured on a two-dimensional detector array. The source and detector then follow a circular trajectory around the object in the plane $z = 0$. This geometry can be considered as a three-dimensional extension of the previously described fan beam projection geometry.
- **Helical cone beam projection** (Fig. 1.3d). This geometry can be considered as a cone beam geometry in which the rotation is not fixed on the $z = 0$ plane, but instead constantly shifts upwards, i.e. a helical trajectory is followed⁴.

All mentioned geometries are — or have been — used in medical CT scanners. However, as will be explained in Section 1.5, the CT reconstruction approach is widely applicable in applications that make use of an entirely different projection geometry (e.g. tomosynthesis, astrotomography, ...). For the sake of simplicity, only the two-dimensional parallel beam geometry is considered in the remainder of this section.

⁴For practical reasons the source and detector have to remain on the same gantry. A smoothly translating patient is therefore required.

1.3.1.2 The Radon transform

For each angle $\theta \in [0, \pi)$ ⁵ and each detector offset $t \in \mathbb{R}$, the parallel beam projection of the function $f(x, y)$ can be defined by the line integrals of $f(x, y)$ along the paths $L(\theta, t) = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \cos \theta + y \sin \theta = t\}$:

$$(\mathcal{R}f)(\theta, t) = p_\theta(t) = \int_{L(\theta, t)} f(x, y) ds. \quad (1.6)$$

In Eqn. 1.6, \mathcal{R} is the transformation operator that maps the object function $f(x, y)$ onto the complete set of detector functions $p_\theta(t)$. It is commonly referred to as the *Radon transform*. The function $(\mathcal{R}f)(\theta, t)$ is also known as the *sinogram*⁶ of $f(x, y)$.

1.3.1.3 Discretized Radon transform

It is important to note a few key differences between the Radon transform and practically measured attenuated projection data. For one, it is not possible to take an infinite number of projection angles θ . Also, detector elements have a certain width, Δt , and are limited in number. Define $\mathbf{p} = (p_i) \in \mathbb{R}^m$, the vector containing all log-corrected projection data, by:

$$p_i = \int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} p_{\theta_i}(t_i + t') dt', \quad (1.7)$$

where θ_i and t_i are the direction angle and detector offset that correspond to detector value p_i . Let m denote the total number of projection directions multiplied by the number of detectors cells in each projection.

The reconstructed image is commonly represented on a grid and its pixels or voxels have a certain width and height. Consider $\mathbf{v} = (v_j) \in \mathbb{R}^n$, a discretized object function represented on such a grid. Let n denote the total number of pixels on this grid. A projection of \mathbf{v} onto the detector p_i can then be considered as the summation of all pixels, weighted by their contribution to the path of this ray, i.e.:

$$p_i = \sum_{j=1}^n w_{ij} v_j. \quad (1.8)$$

In Eqn. 1.8, w_{ij} represents the contribution of pixel j to the detector cell i . The

⁵For a parallel beam projection geometry, $p_\theta(t) = p_{\theta \pm \pi}(-t)$. Therefore, it is not necessary to have projection data for each angle in $[0, 2\pi)$.

⁶If $f(x, y)$ is a translated Dirac delta function, the projection data represents a sine wave along the θ -axis. As the Radon transform is a linear operation, the projection function can therefore be regarded as a weighted summation of sine waves, hence the term “sinogram”.

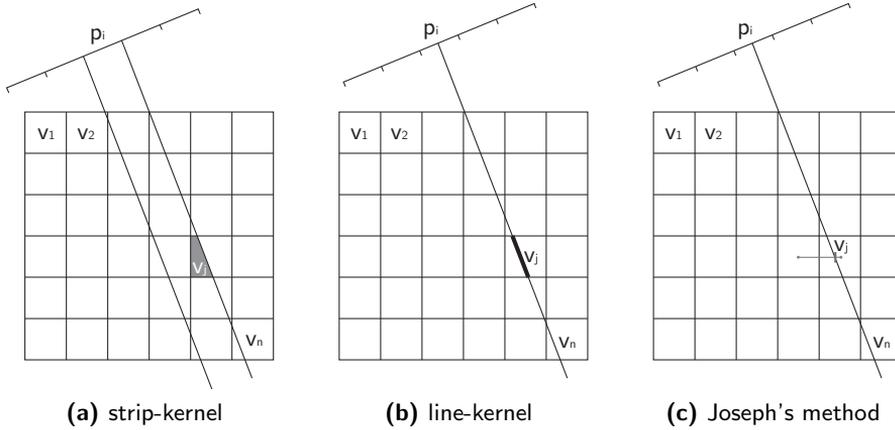


Figure 1.7: Three different approaches for algebraic projection of a certain volume.

projection of \mathbf{v} for each detector pixel can then be written in matrix notation:

$$\mathbf{p} = \mathbf{W}\mathbf{v}. \tag{1.9}$$

The matrix $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{m \times n}$ is called the *projection operator* or *projection matrix* and maps a function in the object or reconstruction domain onto the projection domain. It can be computed or approximated in a variety of ways. Fig. 1.7a shows a strip-kernel projector model where w_{ij} is defined by the size of the intersecting area of the projected strip i and the pixel j . This model resembles the actual X-ray physics, but is not computationally efficient. In Fig. 1.7b, a line-kernel projector model is shown in which w_{ij} is defined by the length of a single detector line i through the pixel j [10]. The reconstruction \mathbf{v} can also be defined as a set of radial basis functions such as the modified Kaiser-Bessel window [11]. As the projection of a radial basis function is equal in each direction, w_{ij} can be computed very efficiently. In Fig. 1.7c, Joseph's projection method is shown [12]. The volume is sampled at each line through the reconstruction volume and interpolation is used to obtain the correct values. This often leads to improved accuracy without sacrificing computational efficiency.

1.3.2 Analytical reconstruction methods

Analytical reconstruction techniques employ analytical properties of the Radon transform to reconstruct the original object function. These methods are computationally efficient and therefore widely utilized in practical applications. Their

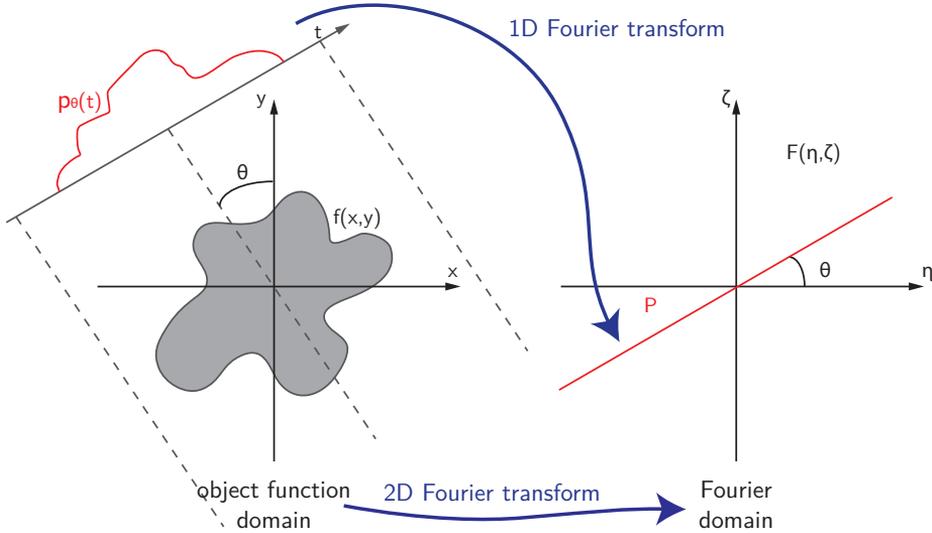


Figure 1.8: Schematic overview of the Fourier slice theorem.

downside is that they are not very flexible. For example, it is complicated to add prior knowledge to the reconstruction process and complicated imaging models cannot be applied. Also, an analytical reconstruction technique can be derived only for a single projection geometry. Other geometries require new derivations.

The *filtered backprojection (FBP)* algorithm, a popular reconstruction technique for parallel beam projection geometry, is derived in Section 1.3.2.2. It is based on the Fourier slice theorem, which is established in Section 1.3.2.1.

1.3.2.1 Fourier slice theorem

The *Fourier slice theorem* states that the one-dimensional Fourier transform of a detector function $p_\theta(t)$, is equal to the slice through the origin at rotation θ of the two-dimensional Fourier space of the object function (Fig. 1.8) [13].

The 2D Fourier transform of $f(x, y)$ is given by:

$$F(\eta, \zeta) = \mathcal{F}\{f(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi(\eta x + \zeta y)} dx dy. \quad (1.10)$$

Furthermore, the 1D Fourier transform of $p_\theta(t)$ is given by:

$$P_\theta(\omega) = \mathcal{F}\{p_\theta(t)\} = \int_{-\infty}^{\infty} p_\theta(t) e^{-i2\pi\omega t} dt. \quad (1.11)$$

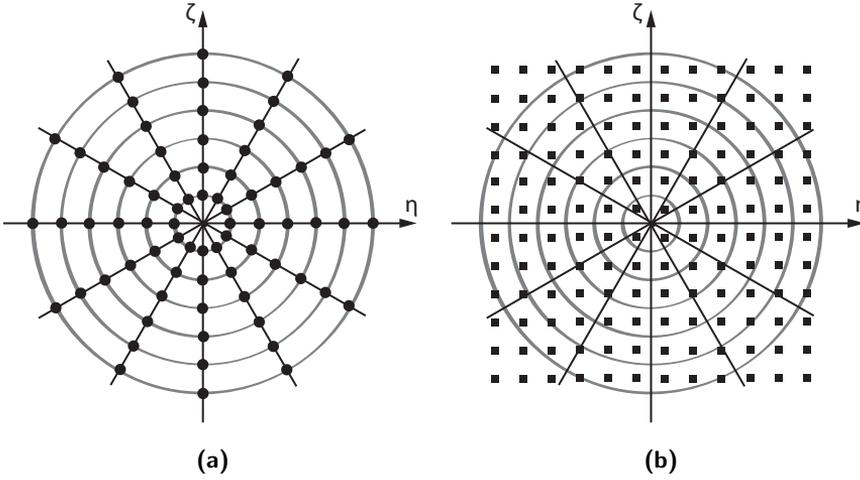


Figure 1.9: (a) Sampling of the Fourier domain with a limited number of projection angles and a non-infinitesimal detector width. Note that the sampling frequency is lower in the high frequency range. (b) Resampling of the Fourier domain.

Substituting Eqn. 1.6 into Eqn. 1.11 results in:

$$P_{\theta}(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi\omega(x \cos \theta + y \sin \theta)} dx dy. \quad (1.12)$$

Therefore, given Eqn. 1.10, it can be concluded that:

$$P_{\theta}(\omega) = F(\omega \cos \theta, \omega \sin \theta), \quad (1.13)$$

proving the Fourier slice theorem.

If projection data is available for each $\theta \in [0, \pi)$ and if the detectors have an infinitesimal width, the Fourier domain of the object function can be fully generated by Fourier transforming each projection. A 2D inverse Fourier transform then yields the original object function:

$$f_{rec}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\eta, \zeta) e^{i2\pi(\eta x + \zeta y)} d\eta d\zeta. \quad (1.14)$$

In practice, the number of projection directions is limited and the detector cells have a non-infinitesimal width. Consequently, it is not possible to sample the entire Fourier domain. Additionally, the sample points do not lie on a rectangular grid (Fig. 1.9a). This obstructs the use of the inverse fast Fourier transform (IFFT)

without an additional interpolation step to resample the data points into a rectangular grid (Fig. 1.9b). This interpolation generally introduces high frequency reconstruction artefacts due to the fact that the higher frequencies are sampled less frequently. For a practical implementation a different, yet related, approach is required: FBP.

1.3.2.2 Filtered backprojection (FBP)

To derive the FBP algorithm, the inverse Fourier transform (Eqn. 1.14) is first expressed in polar coordinates:

$$\begin{aligned}\eta &= \omega \cos \theta, \\ \zeta &= \omega \sin \theta, \\ d\eta d\zeta &= \omega d\omega d\theta.\end{aligned}$$

Eqn. 1.14 then becomes:

$$f_{rec}(x, y) = \int_0^{2\pi} \int_0^\infty F(\omega, \theta) e^{i2\pi\omega(x \cos \theta + y \sin \theta)} \omega d\omega d\theta. \quad (1.15)$$

Using the property $F(\omega, \theta + \pi) = F(-\omega, \theta)$, the above expression can be written as:

$$f_{rec}(x, y) = \int_0^\pi \left[\int_{-\infty}^\infty F(\omega, \theta) |\omega| e^{i2\pi\omega(x \cos \theta + y \sin \theta)} d\omega \right] d\theta. \quad (1.16)$$

As the Fourier slice theorem states (Eqn. 1.13), $F(\omega, \theta)$ can be substituted by the Fourier transform of its corresponding projection, $P_\theta(\omega)$:

$$f_{rec}(x, y) = \int_0^\pi q_\theta(x \cos \theta + y \sin \theta) d\theta, \quad (1.17)$$

with

$$q_\theta(t) = \int_{-\infty}^\infty P_\theta(\omega) |\omega| e^{i2\pi\omega t} d\omega = \mathcal{F}^{-1} \{ P_\theta(\omega) |\omega| \}. \quad (1.18)$$

Eqn. 1.17 and Eqn. 1.18 clearly show the two steps of the algorithm, also present in its name:

1. Each detector function $p_\theta(t)$ is **filtered** by multiplying its Fourier transform with $|\omega|$. This high-pass filter compensates for the varying sampling density of the Fourier domain (Fig. 1.9).
2. The filtered projection data is then **backprojected** — or “smeared out” — onto the reconstruction grid along the lines $t = x \cos \theta + y \sin \theta$.

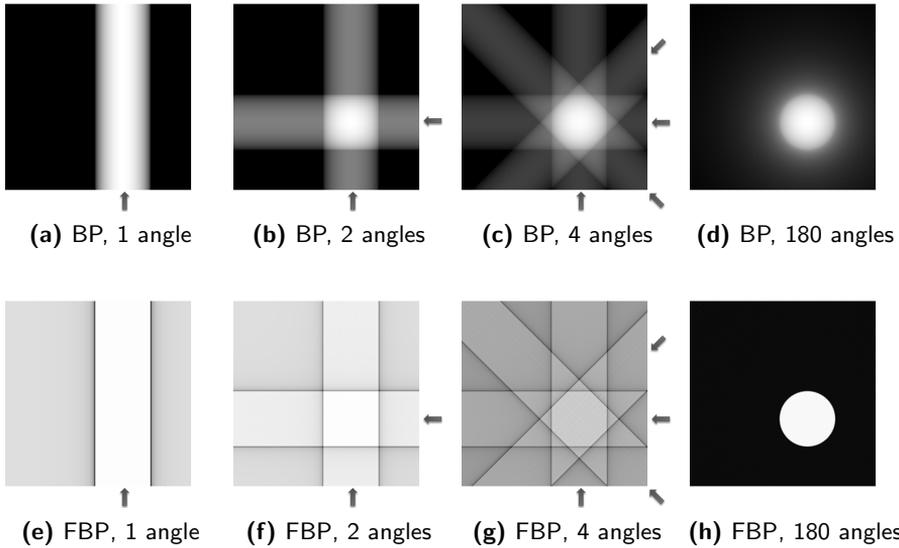


Figure 1.10: Comparison between unfiltered backprojection (BP) and filtered backprojection (FBP) of a homogeneous spheric object. Reconstructions are shown for an increasing number of projection angles. The “smearing out” of the detector functions is clearly visible, as is the sharpness of the FBP reconstruction in comparison to the unfiltered BP.

For practical use, Eqn. 1.17 and Eqn. 1.18 must first be discretized, which does not cause a fundamental change to the algorithm.

In Fig. 1.10, reconstructions of a homogeneous spheric object can be observed with and without the filtering step. Clearly, the sharpness of the reconstruction is acutely higher when the filter is used.

1.3.2.3 Other analytical reconstruction methods

FBP is only applicable in case of a parallel beam projection geometry. For other geometries, a different derivation is required. For the fan beam geometry, such a derivation is given in [13].

To exactly reconstruct the object function, the Radon space, $(\mathcal{R}f)(\theta, t)$, must be completely known. For three-dimensional geometries, the 3D Radon space cannot be filled in a circular trajectory (i.e. with a cone beam geometry). A popular reconstruction method for circular cone beam data has been proposed by *Feldkamp-Davis-Kress (FDK)* [14]. It is an approximation technique that generally results in adequate reconstructions for small cone angles, i.e. for slices close to the central slice $z = 0$. For helical cone beam, the 3D Radon space can be entirely

filled, so exact reconstruction techniques are possible, e.g. Kudo's algorithm [15] or the Katsevich algorithm [16].

1.3.3 Algebraic reconstruction methods

In algebraic reconstruction methods, the reconstruction problem is considered as the solving of the system of linear equations

$$\mathbf{W}\mathbf{v} = \mathbf{p}, \quad (1.19)$$

where \mathbf{W} , \mathbf{v} and \mathbf{p} are defined as in Section 1.3.1.3. Eqn. 1.19 is typically a very large system and cannot be solved directly by inverting \mathbf{W} . Various iterative strategies, such as the *Simultaneous Iterative Reconstruction Technique (SIRT)* [17], are available.

1.3.3.1 Simultaneous Iterative Reconstruction Technique (SIRT)

Put $\mathbf{v}^{(0)}$, an initial reconstruction image (typically a black image, i.e. $\mathbf{v}^{(0)} = \mathbf{0}$). Put $k = 0$, the iteration number. Each iteration of SIRT consists of 3 steps.

1. Compute the forward projection of the current reconstruction image:

$$\mathbf{p}^{(k)} = \mathbf{W}\mathbf{v}^{(k)}. \quad (1.20)$$

2. Compute the current projection difference, or *residual sinogram*:

$$\mathbf{r}^{(k)} = \mathbf{p} - \mathbf{p}^{(k)}. \quad (1.21)$$

3. Update the reconstruction image $\mathbf{v}^{(k)}$ by adding a weighted backprojection of the residual sinogram:

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \mathbf{C}\mathbf{W}^T \mathbf{R}\mathbf{r}^{(k)}. \quad (1.22)$$

In Eqn. 1.22, $\mathbf{R} \in \mathbb{R}^{m \times m}$ is a diagonal matrix of the inverse row sums of \mathbf{W} , i.e. $r_{ii} = 1/\sum_j w_{ij}$. Likewise, $\mathbf{C} \in \mathbb{R}^{n \times n}$ is a diagonal matrix of the inverse column sums of \mathbf{W} , i.e. $c_{jj} = 1/\sum_i w_{ij}$. Increase k by 1 and return to step 1 until a convergence criterion has been met.

The full update for each iteration of SIRT is:

$$v_j^{(k+1)} = v_j^{(k)} + \frac{1}{\sum_{i=1}^n w_{ij}} \sum_{i=1}^m \frac{w_{ij} \left(p_i - \sum_{h=1}^n w_{ih} v_h^{(k)} \right)}{\sum_{h=1}^n w_{ih}}. \quad (1.23)$$

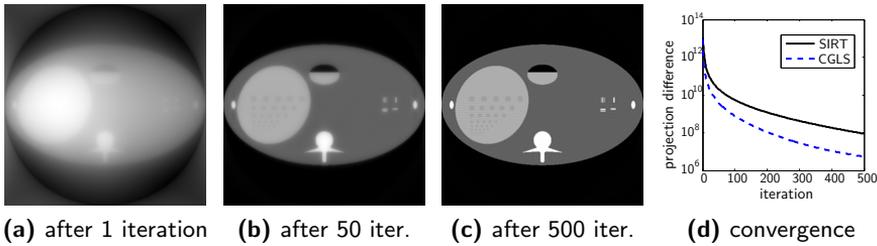


Figure 1.11: SIRT reconstruction of a simulated 512×512 FORBILD abdomen phantom (available from: <http://www.imp.uni-erlangen.de/phantoms/head/head.html>) from 180 equiangular projections. (a-c) Intermediate reconstructions after 1, 50 and 500 iterations. (d) The projection difference as a function of the iteration number for both the SIRT and the CGLS method.

In [17], it is proven that SIRT is a linear operation, and that $\mathbf{v}^{(k)}$ converges to the weighted least squares solution \mathbf{v}_{rec} :

$$\mathbf{v}_{rec} = \operatorname{argmin}_{\mathbf{v} \in \mathbb{R}^n} \|\mathbf{W}\mathbf{v} - \mathbf{p}\|_{\mathbf{R}}^2, \quad (1.24)$$

where $\|\mathbf{W}\mathbf{v} - \mathbf{p}\|_{\mathbf{R}}^2 = (\mathbf{W}\mathbf{v} - \mathbf{p})^T \mathbf{R} (\mathbf{W}\mathbf{v} - \mathbf{p})$. If multiple solutions exist — which is likely to occur if $m \ll n$ — SIRT converges to the solution that is the closest to the initial image. In Fig. 1.11a-c, SIRT reconstructions of a phantom image are shown with an increasing number of iterations.

In comparison to the analytical reconstruction methods, SIRT is much more flexible. For example, a positivity constraint can be easily enforced by setting the negative values of $\mathbf{v}^{(k)}$ to 0 after the backprojection in step 3. Also, SIRT can be readily applied for any other projection geometry, provided that its projection matrix is known or can be computed.

A key problem with iterative reconstruction techniques is their computational requirements. This is mostly due to the constant forward- and backprojection operations. Furthermore, as \mathbf{W} is typically too large to store in memory⁷, its values must be recomputed every time they are needed and discarded immediately afterwards. Fortunately, SIRT lends itself perfectly for acceleration with parallel computing [18]. Especially the recent increase in computational power of modern GPU hardware has led to rapid accelerations and has made iterative algorithms feasible for everyday use [19].

⁷Consider a relatively small 512×512 reconstruction grid and 360 projection directions with 512 detector values in each projection. The size of \mathbf{W} is then 184320×262144 . A single precision representation of \mathbf{W} then requires 180 gigabytes of memory. Even with a sparse representation, several gigabytes would be required.

1.3.3.2 Other iterative reconstruction methods

Here, a selection of other commonly used iterative reconstruction methods is given.

- In SIRT, each iteration contains the forward- and backprojection along each detector value. In contrast, each iteration of the *Algebraic Reconstruction Technique (ART)* only uses a single detector value to update the reconstruction. The update formula is then:

$$v_j^{(k+1)} = v_j^{(k)} + \frac{p_i - \sum_{h=1}^n w_{ih} v_h^{(k)}}{\sum_{h=1}^n w_{ih}^2} w_{ij}, \quad (1.25)$$

where i is the index of the detector that is used in iteration k . ART reaches a solution more quickly than SIRT, but does not have stable convergence if Eqn. 1.19 is inconsistent (e.g. due to noise).

- In the *Simultaneous Algebraic Reconstruction Technique (SART)*, each iteration updates all detector values that belong to a single projection direction. It combines the increased reconstruction stability of SIRT with the increased convergence speed of Algebraic Reconstruction Technique (ART). Its update formula is:

$$v_j^{(k+1)} = v_j^{(k)} + \frac{\sum_{p_i \in P_\theta} \left(\frac{p_i - \sum_{h=1}^n w_{ih} v_h^{(k)}}{\sum_{h=1}^n w_{ih}} \right) w_{ij}}{\sum_{p_i \in P_\theta} w_{ij}}, \quad (1.26)$$

where P_θ denotes the set of all detector values p_i that belong to the projection direction θ . The order in which the projection directions are chosen is essential for optimal reconstruction quality. In Chapter 3 of [20], a weighted distance scheme is proposed to maximally increase the convergence speed and accuracy. If more than one projection direction is used in each iteration, the algorithm is called *ordered subsets SIRT (OS-SIRT)* [21].

- ART, SART and SIRT are additive iterative algorithms as each update step adds an update term to the current reconstruction image. In *Multiplicative Algebraic Reconstruction Technique (MART)*, however, the current reconstruction image is multiplied by the update step:

$$v_j^{(k+1)} = v_j^{(k)} \frac{p_i}{\sum_{h=1}^n w_{ih} v_h^{(k)}}. \quad (1.27)$$

Provided that $v_j^{(0)} > 0$, MART has a built-in non-negative constraint. Moreover, if $v_j^{(0)} < 0$, that pixel value can never be updated to a positive value.

- The *Conjugate Gradient Least Squares (CGLS)* method utilizes a different reconstruction approach. It applies the iterative Conjugate Gradient (CG) method to solve the least squares normalization of Eqn. 1.19 [22]:

$$\mathbf{W}^T \mathbf{W} \mathbf{v}_{rec} = \mathbf{W}^T \mathbf{p}. \quad (1.28)$$

With $\mathbf{r}^{(0)} = \mathbf{p}$, $\mathbf{z}^{(0)} = \mathbf{W}^T \mathbf{p}$ and $\mathbf{v}^{(0)} = \mathbf{z}^{(0)}$, the update step of CGLS is the following:

$$\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \frac{\|\mathbf{z}^{(k)}\|_2}{\|\mathbf{W} \mathbf{v}^{(k)}\|_2} \mathbf{W} \mathbf{v}^{(k)} \quad (1.29)$$

$$\mathbf{z}^{(k+1)} = \mathbf{W}^T \mathbf{r}^{(k+1)} \quad (1.30)$$

$$\mathbf{v}^{(k+1)} = \mathbf{z}^{(k+1)} + \frac{\mathbf{v}^{(k)}}{\|\mathbf{z}^{(k)}\|_2} \quad (1.31)$$

CGLS converges to a solution much faster than SIRT (Fig. 1.11d).

1.4 Typical CT workflow

In this section, an overview is given of a typical workflow in CT imaging (Fig. 1.12). It starts with data acquisition (Section 1.4.1), continues to data preprocessing (Section 1.4.2), to image reconstruction (Section 1.4.3), to segmentation (Section 1.4.4), and finally to analysis (Section 1.4.5). Common problems and pitfalls are outlined for each step.

1.4.1 Data acquisition

Initially, a sample is placed into the *field of view (FOV)*⁸ of the scanner and intensity projection images (Eqn. 1.3) are acquired for a set of projection angles. Ideally, each measured image is an exact projection of the scanned object under the exact geometrical parameters. In that case, the projection data is said to be *consistent*. In [23, 24] consistency conditions are given for parallel beam projection data. Unfortunately, there are a variety of reasons that can make different projection images inconsistent with each other.

- **Noise.** The total number of photons at the source of a certain X-ray beam, I_0 , is Poisson distributed. Consequently, high signal-to-noise ratios (SNR) can only be obtained if I_0 — and thus also the radiation dose — is sufficiently

⁸The FOV is the area of the reconstruction domain that is projected onto the detector in every projection direction.



Figure 1.12: Typical CT workflow.

high. Fig. 1.13 shows FBP reconstructions of a phantom image from 360 equiangular parallel beam projections with simulated Poisson noise. It is clear that the contrast of the reconstructions decreases as also I_0 decreases.

- **Detector quality.** Not all detector cells have the same sensitivity, an effect likely to increase as detector plates wear out. Fig. 1.14a shows a sinogram where each detector has a random sensitivity error. This results in visible *ring artefacts* in the reconstructed image (Fig. 1.14b).

To counter this problem, the projection data I can be normalized by also measuring a dark field image, D (an image acquired when the X-ray source is turned off), and a bright field image B (an image acquired when the X-ray source is turned on but without the sample):

$$I_{norm} = \frac{I - D}{B - D}. \quad (1.32)$$

- **Scattering.** As described in Section 1.2.2, when an X-ray photon traverses an object, there is a chance of it being scattered, i.e. of it continuing — with a lower energy — on a refracted path. It is commonly assumed that such a refracted photon will not reach the detector. However, as is sketched in Fig. 1.15a, scattered photons might still be measured. The assumption that each detector cell only measures non-attenuated X-ray photons is thus violated. Instead, each measured value is a combination of primary photons and scattered photons. Note that the contribution of scattered rays is not proportional to the contribution of primary rays (Fig. 1.15b). The SNR of the measured signal then effectively decreases if high attenuating objects are on the path of the primary ray.
- **Photon starvation.** As an X-ray beam traverses an object, the total number of photons in that beam decreases. If the object contains a very dense material, such as a metal, and if the energy of the X-ray beam is not sufficiently high, the entire X-ray beam might be attenuated. Information about the attenuation before and after the dense materials is then lost, resulting in streaking artefacts or *metal artefacts* [25]. This can be observed in Fig. 1.16a, where an FBP reconstruction is shown of a plexiglass hardware

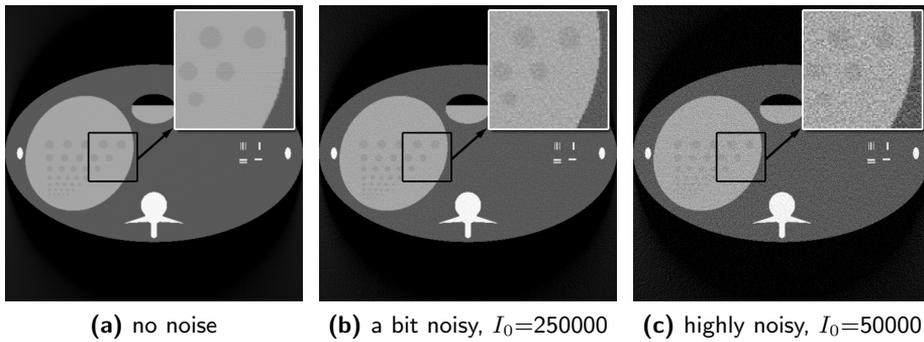


Figure 1.13: Three FBP reconstructions of a simulated 512×512 FORBILD abdomen phantom from 360 equiangular projection angles with different levels of Poisson noise applied to the projection data. The contrast visibly decreases as more noise was present.

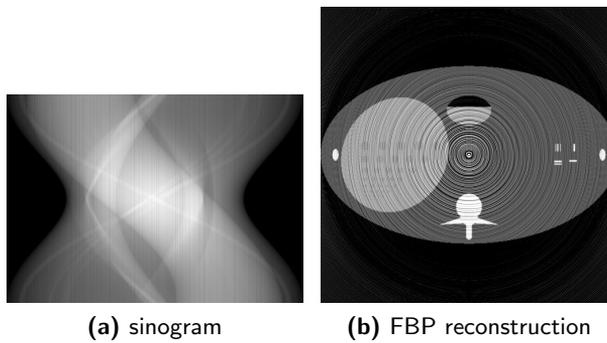


Figure 1.14: (a) A simulated sinogram of a 512×512 FORBILD abdomen phantom from 360 equiangular projection angles where each detector has a random, Gaussian distributed, error in its sensitivity. (b) On the FBP reconstruction, ring artefacts are clearly visible.

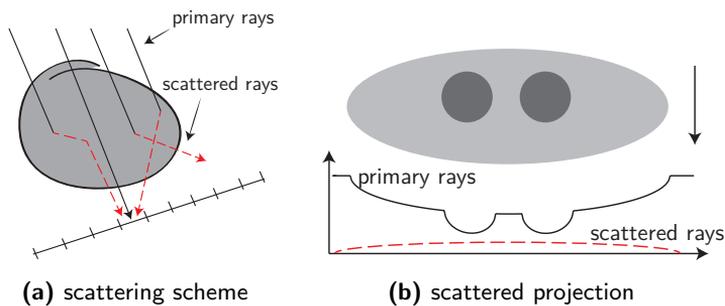


Figure 1.15: Photon scattering can lead to inconsistent projection data.

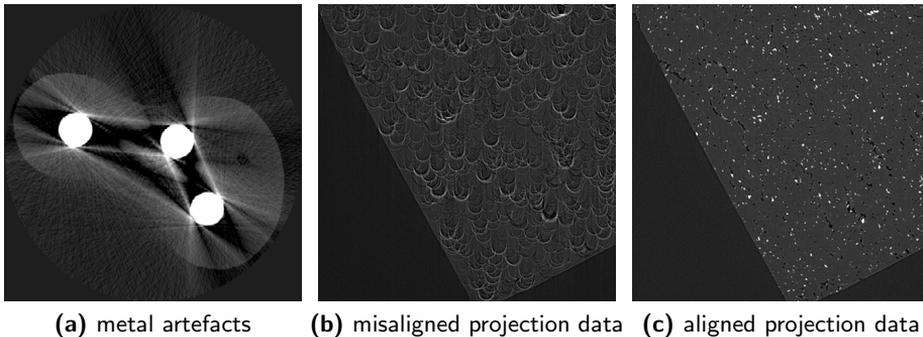


Figure 1.16: (a) FBP reconstruction of the barbabapa phantom, scanned in a SkyScan 1172 μ CT system with 320 projection directions. Metal streaking artefacts are clearly visible. (b) FBP reconstruction from misaligned projection data of a salt cube with various copper particles. Banana-shaped misalignment artefacts are clearly visible. (c) FBP reconstruction of the same dataset where this misalignment was corrected for.

phantom containing very dense steel rods, scanned in a SkyScan 1172 μ CT system.

- **Misalignment.** Projection data is also inconsistent if the geometric parameters of a certain projection (i.e. the position of the source and detector-array with respect to the origin $(0,0)$) are not what is assumed. This can occur if the gantry on which the source and detector are mounted is not stable during its rotation. In Fig. 1.16b, an FBP reconstruction is shown of a salt cube with various copper particles, scanned at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. In the projection data, the centre of rotation was off by the width of several detector cells, resulting in banana-shaped *misalignment artefacts*. Fig. 1.16c shows the FBP reconstruction with correctly aligned projection data.
- **Motion.** It is assumed that the scanned sample does not move or change shape during the entire acquisition phase. For certain applications, such as cardiac imaging, this is however not possible. To minimize this effect, modern medical CT scanners have a very high rotation speed.

1.4.2 Data preprocessing

After the projection data is acquired, a preprocessing step is applied to prepare the data for use in a reconstruction algorithm. This step includes the computation of the logarithmic transform that turns intensity values of the projection images into

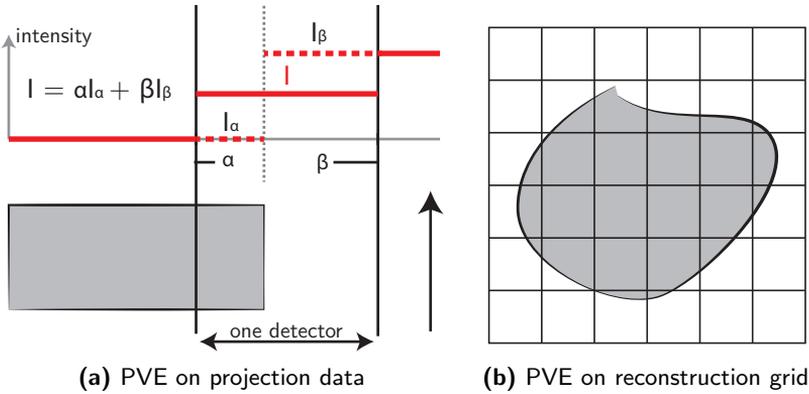


Figure 1.17: (a) The partial volume effect (PVE) on the projection data. Detector cells have a certain width. The total measured intensity in the detector is a weighted summation of contributions. This linearity does not hold for the log-corrected values, resulting in reconstruction artefacts. (b) The partial volume effect (PVE) on the reconstruction grid. Pixels have a certain width and height and objects are not aligned with the pixel edges. The value of many pixels will thus not represent the attenuation coefficient of any of the materials of the object.

log-corrected attenuation values (Eqn. 1.4). Other common operations include rebinning (converting the data to another projection geometry), and correcting the data for one or more inconsistencies that were introduced during the previous phase (e.g. attempting to correct projection misalignments). The preprocessed data often contains the following problems.

- Any **data inconsistency** of the acquired projection images that was not corrected for.
- A **partial volume effect (PVE)** on the projection data. Each detector cell has a certain width and object edges are typically not aligned with an edge of a detector. Consider the case presented in Fig. 1.17a. In one part of length α of the detector the intensity, I_α , is low as the photon beam has passed through the object. In the other part of length β the intensity, I_β , is high. The total measured intensity, I , is then a weighted summation of both contributions, i.e. $I = \alpha I_\alpha + \beta I_\beta$. However, due to the logarithmic nature of Eqn. 1.4, this linearity does not hold for the log-corrected values:

$$\ln(\alpha I_\alpha + \beta I_\beta) \neq \ln(\alpha I_\alpha) + \ln(\beta I_\beta). \quad (1.33)$$

This means that the log-corrected values do not exactly correspond to the line integrals, leading to streaking artefacts in the reconstructed images.

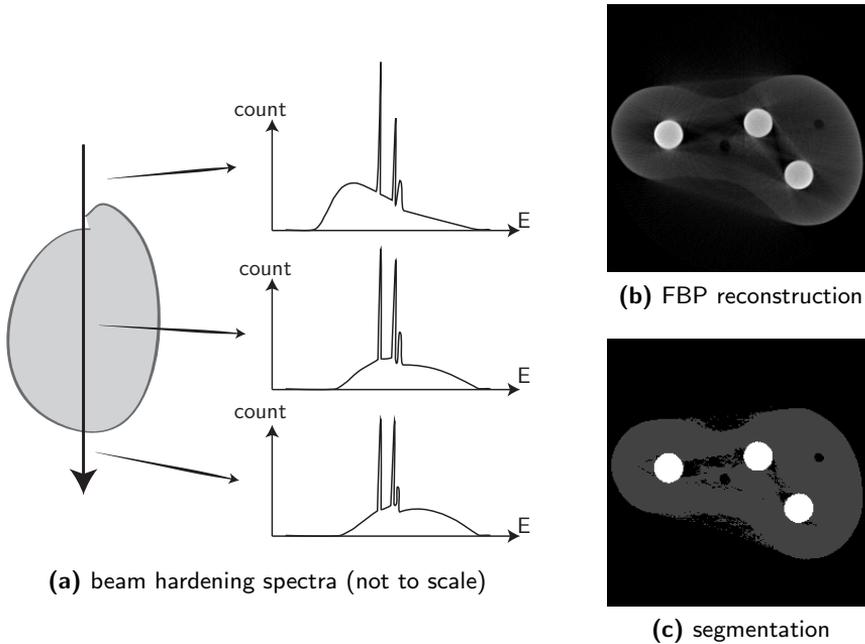


Figure 1.18: (a) Schematic overview of beam hardening. As an X-ray beam traverses an object, its energy spectrum constantly changes. Given that low energy photons are absorbed more strongly than high energy photons, this results in an effective “hardening” of the X-ray beam. (b) An FBP reconstruction of the barbabapa phantom, scanned in a SkyScan 1172 μ CT system with 600 projection angles (data at the courtesy of Gert Van Gompel[26]). The cupping and streaking artefacts are clearly visible. (c) Segmentation of the FBP reconstruction.

- **Beam hardening.** For polychromatic X-ray sources, the measured intensity value does not exclusively depend on the ray length through the object but also on the energy spectrum of the X-ray beam (Eqn. 1.5). Seeing that low energy photons are commonly absorbed more strongly than high energy photons, the shape of the energy spectrum will change (“harden”) as the beam passes through an object (Fig. 1.18a). This eradicates the linear relationship between the ray length and the attenuation, resulting in inconsistent log-corrected projection data that will culminate into cupping and streaking reconstruction artefacts, as can be recognized in Fig. 1.18b.

In the literature, an assortment of beam hardening reduction strategies have been proposed. For example, a hardware filter that narrows the energy spectrum can be placed after the X-ray source. While this reduces the effect

of beam hardening, it also significantly reduces the SNR of the projection data. Attempts have been made to linearize the projection data — effectively turning polychromatic data into monochromatic data — by exploiting prior knowledge about the object materials and energy spectra [27], or by using advanced iterative methods to solve linearization cost functions [28, 29]. Also, the polychromatic model can be included in statistical reconstruction algorithms, such as maximum likelihood (ML) [30].

1.4.3 Reconstruction

In the next phase the reconstruction is performed. Section 1.3 introduced several of such algorithms. Often, the reconstructed images contain pixels whose values do not correspond to the attenuation factors of the original object. They are referred to as reconstruction errors or *reconstruction artefacts* and can arise due to a variety of reasons.

- Any **data inconsistency** of the preprocessed data that was not corrected for. Fig. 1.14b shows an FBP reconstruction polluted with ring artefacts after the projection data was not normalized. Fig. 1.16a shows how photon starvation has led to metal artefacts, Fig. 1.16b shows misalignment artefacts and Fig. 1.18b shows cupping and streaking artefacts that arise when beam hardening is neglected.
- The **partial volume effect (PVE) of the reconstruction grid**. The pixels in the reconstruction grid have a finite width and height, Δs . The original function $f(x, y)$ mapped onto the reconstruction grid, $\mathbf{v} = (v_j) \in \mathbb{R}^n$, is then:

$$v_j = \frac{1}{\Delta s^2} \int_{-\frac{\Delta s}{2}}^{-\frac{\Delta s}{2}} \int_{-\frac{\Delta s}{2}}^{-\frac{\Delta s}{2}} f(x_j + x', y_j + y') dx' dy', \quad (1.34)$$

where x_j and y_j are the coordinates of the centre of the pixel v_j . The value v_j depends on an entire area of values of the real object function. If the object has an edge running through its area (as seen in Fig. 1.17b), or if the object is not homogeneous inside the pixel boundaries, v_j does not represent the attenuation coefficient of any of the materials of the object. This effect is studied in detail in Chapter 7 of this work.

Reconstruction artefacts also occur when the projection data is **incomplete** and no unique solution to Eqn. 1.19 exists. Instead, various optima exist for

$$\operatorname{argmin}_{\mathbf{v} \in \mathbb{R}} \|\mathbf{W}\mathbf{v} - \mathbf{p}\|_{\mathbf{R}}^2, \quad (1.35)$$

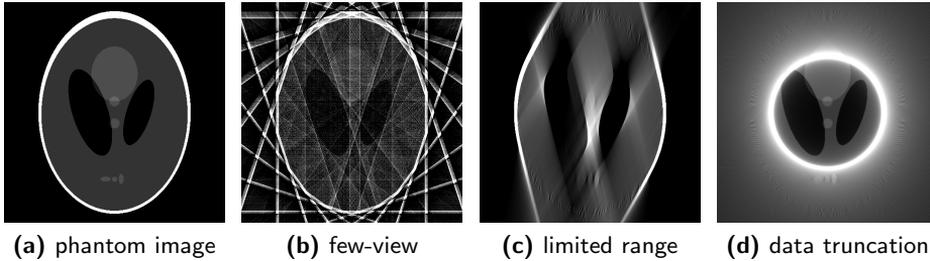


Figure 1.19: Several types of incomplete projection data. **(a)** A 512×512 Shepp-Logan phantom image. **(b)** FBP reconstruction of **(a)** from only 10 equiangularly spaced projections. **(c)** FBP reconstruction of **(a)** from 360 projections in the limited range $[-\frac{\pi}{3}, \frac{\pi}{3}]$. **(d)** FBP reconstruction of **(a)** from truncated data.

developing reconstructed images that do not correspond to the reality. Various types of incomplete data can be noted.

- **A small number of projection directions**, also referred to as *few-view tomography*. To reduce the scanning time or to restrict the radiation dose without lowering the SNR of the projection images, the number of projection directions is often limited. Fig. 1.19b shows an FBP reconstruction of the Shepp-Logan phantom (Fig. 1.19a) from only 10 equiangularly spaced parallel beam projection directions.
- **A limited angular range**. Typically, all projection directions θ are equiangularly distributed in the $[0, \pi)$ ⁹. However, sometimes it is technically impossible to create a projection image under a certain angle. This is for example true in electron microscopes, where the sample is placed on a structure that can only rotate in an angular range such as $[-\frac{\pi}{3}, \frac{\pi}{3}]$. It is also true for laminography, where projection directions parallel to the object surface result in photon starvation and metal artefacts. Fig. 1.19c shows an FBP reconstruction of the Shepp-Logan phantom from 360 projection directions in the limited range $[-\frac{\pi}{3}, \frac{\pi}{3}]$. *Missing wedge* artefacts are clearly visible.
- **Truncated projections**. If the detector array is not sufficiently wide to capture the entire object, the projection data is said to be *truncated*. This can be due to technical constraints (i.e. objects that are too large to fit inside the FOV) or due to an attempt to reduce the radiation dose outside the *region of interest (ROI)* by blocking the X-ray beams to the outside detector

⁹This is only in the case of a parallel beam projection geometry. For a fan beam geometry an even larger range is required.

1.5. TOMOGRAPHY ON ALL SCALES, APPLICATION OVERVIEW

cells. This type of incomplete data obstructs accurate image reconstruction outside the FOV. Furthermore, given that common reconstruction techniques are non-local (i.e. it is not possible to exactly reconstruct only a subset of the reconstruction domain), it is also not possible to accurately reconstruct inside the FOV. This can be observed by the cupping artefact in Fig. 1.19d.

Proposed solutions to this problem include the completion of projection data by extrapolation with a certain smoothing function [31, 32] or by combining multiple scans with a shifted sample [33] or with a lower radiation dose on a full detector array [34]. Approaches have also been introduced that use local reconstruction techniques such as Differentiated Backprojection (DBP) [35].

1.4.4 Segmentation

Given the collection of possible reconstruction artefacts, the reconstructed image may not be suitable for analysis without an extra segmentation step. An overview of various segmentation techniques is given in Chapter 2. Important to note here is that all common segmentation techniques are effectively post-processing steps, i.e. they are solely based on the finished reconstructed image. However, from this section and from Fig. 1.12, it is clear that the typical tomographic workflow is a very sequential process and that errors propagate into each subsequent step. This can be observed in Fig. 1.18c, where the streaking artefacts of the FBP reconstruction have a clear effect on the accuracy of the segmentation.

Consequently, an ideal segmentation technique should not only be based on the reconstructed image, but should also use information of the previous steps in the workflow, i.e. the projection data [36]. This observation is in fact the major thesis of this work. In Part I, segmentation techniques are proposed that use projection data to improve the segmentation quality, the so called *tomographic segmentation* techniques. In Part II, segmentation and reconstruction are combined into a single step, referred to as *discrete tomography*.

1.4.5 Analysis

At the final step in the workflow, the data is ready for analysis. This step is fully dependent on the application, which is discussed in the next section.

1.5 Tomography on all scales, application overview

While CT is frequently referred to as a “medical imaging technique”, its applicability is by no means limited to that field. Indeed, many are the applications that have prospered by utilizing tomographic principles. In this section, a short

CHAPTER 1. INTRODUCTION

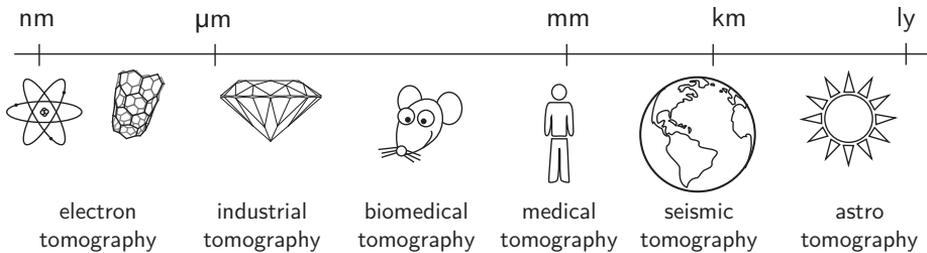


Figure 1.20: Overview of different tomographic applications at a variety of scales.

overview is given. It will become clear that tomography need not be restricted to imaging modalities of a single scale or resolution (Fig. 1.20).

1.5.1 Materials science

Modern electron microscopes (Fig. 1.21a) can be used to visualize nano structures up to atomic resolution [37]. They do not make use of an X-ray beam — the wavelength of an X-ray beam is simply too long for such a resolution — but instead utilize a beam of electrons to create a bright-field projection image of the scanned sample. By tilting the plane on which this sample is placed, different projection images can be recorded and *electron tomography* can be applied, resulting in a 3D reconstruction. This is called *transmission electron microscopy (TEM)*. Next to TEM, electron microscopes can also be used to measure other information. With *high-angle annular dark field scanning TEM (HAADF STEM)*, the contrast scales with the atomic number of the material. With *energy-filtered TEM (EFTEM)*, energy loss information is measured to map the composition of the sample [38]. These images can also be used in combination with tomographic reconstruction algorithms, leading to a variety of interesting applications: the study of biological structures such as viruses; metallurgy, the search for durable and strong super-alloys that requires a highly optimized structure at a nano scale; electronics, improving the performance of ultra small semiconductors; etc. Given that the electron gun is in a fixed position and the sample is placed on a supporting structure that is then tilted, it is usually not possible to acquire projection data for the full angular range, leading to common missing wedge artefacts. Additionally, due to the small samples and the often inaccurate tilting, the projection data is often severely misaligned.

A second imaging modality that is often used in materials science is a synchrotron radiation facility (Fig. 1.21b). In these facilities, charged particles are accelerated to near the speed of light in large storage rings. This typically results in very intense monochromatic bundles that are able to create very high resolution

1.5. TOMOGRAPHY ON ALL SCALES, APPLICATION OVERVIEW

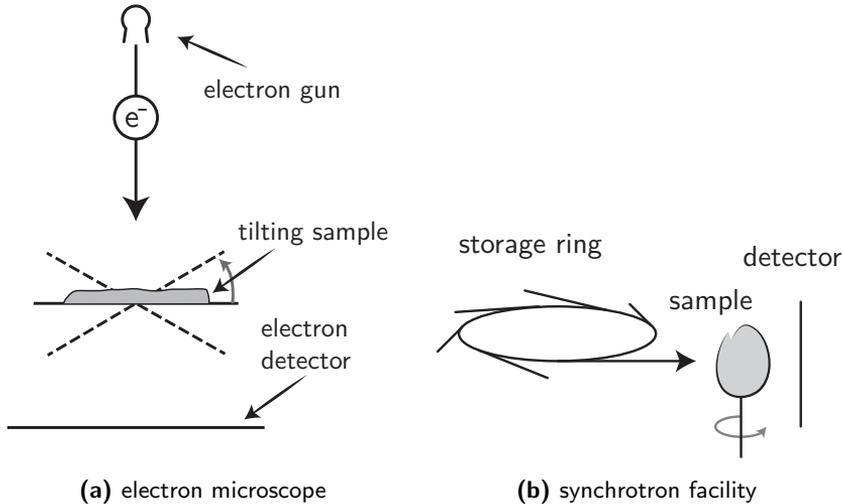


Figure 1.21: Schematic overview of image modalities that are commonly used in materials science.

projection images and thus also very high resolution reconstructed images. Next to classic tomography, many applications are made possible by synchrotron radiation, e.g. the change in X-ray phase can be measured to increase the contrast between two materials that have a similar attenuation value. In *3D X-ray diffraction (3DXRD)*, the diffraction of a certain homogeneous object is measured under several projection angles. Tomographic principles can then be used to reconstruct the orientation of each crystal of the object [39].

More information on electron and synchrotron tomography for use in materials science can be found in [40].

1.5.2 Non-destructive testing

In industry, tomography is often used for non-destructive testing of objects. By visualizing the interior of manufactured objects their required material properties, e.g. its structural integrity, can be confirmed.

Related, tomography can also be applied for planning an industrial process. Such an application can be found in the diamond industry. A diamond is a collection of carbon atoms linked together in a very strong lattice. They form under high pressure and high temperature and are therefore very rare and valuable. In nature, they are found in their raw form and prior to being used as a jewel, they have to be cut into an aesthetically pleasing form. The value of this diamond

depends on what is known as “the four C’s”: Clarity (the absence of flaws and inclusions), Colour (colourless diamonds are preferred as they allow the prettiest refraction of light), Cut (its shape) and Carat (its weight, 1 carat = 200 mg). The choice of how and where to cut the diamond — such that its eventual value is maximal — is thus of great importance and is typically made by trained specialists. Tomography, however, provides a cheaper and more objective way to do so. With a μ CT system, a high quality virtual model of the raw diamond can be constructed. This is then used by a clever optimization technique to automatically locate the optimal cut diamond inside this model. As a small increase in carat can lead to a big increase in value, the segmentation step in the tomographic workflow is crucial. Fig. 1.22b shows a virtual model of a raw diamond stone as computed with CT. One can also observe the optimal location for the cut diamond.

1.5.3 Biomedical

In biomedical research and in preclinical trials, small animals (mice or rats) are used for research into various diseases or to evaluate the effects of new drugs. As an example, consider the study of a new drug designed to counter osteoporosis, a disease which degenerates the structural integrity of bones. With an in-vivo μ CT scanner the temporal effect of these drugs can be investigated by scanning the femur of lab animals (Fig. 1.22c) at various stages in the trial [41]. The bone structures of interest are the cortical bone and trabecular bone, a network of very small structures. Accurate segmentation and determination of the morphometric bone parameters [42] therefore requires a high *spatial resolution*. Also, to get the most information, the time between two scans of the same animal should be as low as possible. This is referred to as the *longitudinal resolution*. However, it is crucial to refrain from subjecting a high radiation dose to the animal. For one, they might succumb to radiation poisoning. But also, measuring with radiation is effectively also interacting: it may enlarge or shrink the trabecular structures, preventing accurate analysis.

A trade-off has to be made between the various requirements. If the spatial resolution is high, then the animals cannot be scanned very often. If the longitudinal resolution is high, scans are of a low spatial resolution and analysis is difficult. In Chapter 7 of this work, a super-resolution approach is proposed to compute high resolution reconstructions from low resolution, low dose projection data.

1.5.4 Medical

The most well-known application of CT is its use in medical imaging. CT allows a view inside a patient’s body without doing invasive surgery. For decades, it has been one of the standard diagnostic tools at the aid of physicians. It is typically

1.5. TOMOGRAPHY ON ALL SCALES, APPLICATION OVERVIEW

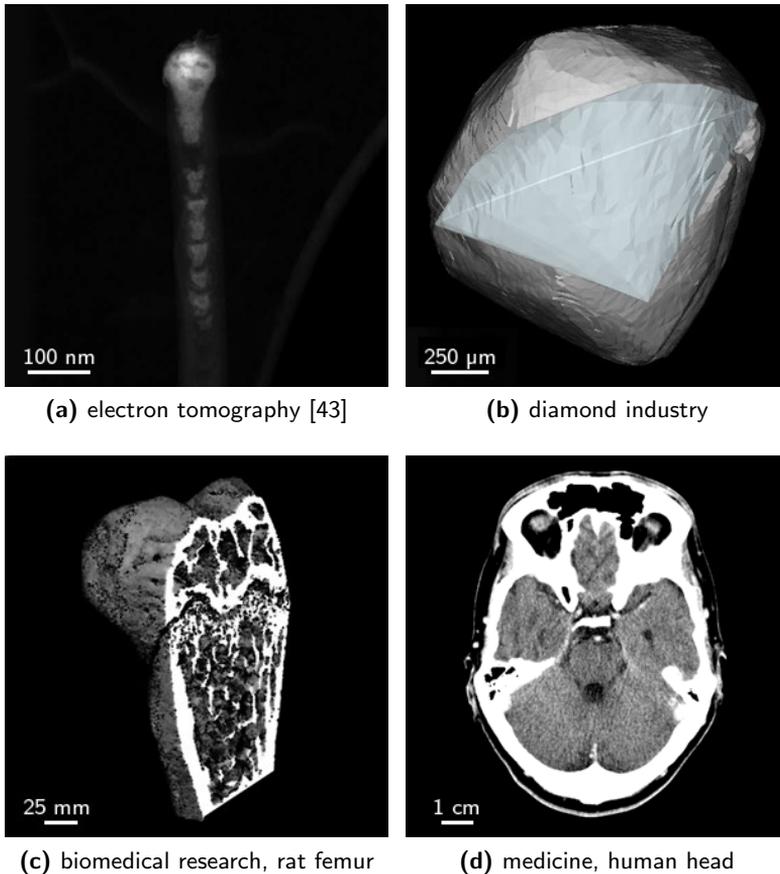


Figure 1.22: Various applications in which tomography is often used.

used to detect and follow up on tumours, calcifications, brain trauma, . . . Its use is not restricted to head imaging (Fig. 1.22d), but also extends to lungs, pulmonary arteries, the abdominal area, . . . Modern scanners have a high rotation speed and also support accurate cardiac imaging. Contrast fluids, typically iodine-base, can be applied to increase the contrast of certain internal structures, such as blood vessels.

1.5.5 Large scale

In seismic tomography, earthquakes are measured at various stations throughout the globe. Common reconstruction algorithms such as SIRT are used to reconstruct

REFERENCES

the P-waves and S-waves of the earthquakes as they have propagated through the earth's crust. These techniques have been used to obtain valuable insights into the density and rigidity of the earth [44] and to measure the alignment of motorways in tunnels [45]. The resolution of such reconstructed images is typically about a few kilometres. Common issues in seismic tomography include limited data (due to the low number of measuring stations), misalignment (due to the fact that the hypocentre of the earthquake must be exactly known), and a very complex projector model which should include scattering and surface refractions.

On an even larger scale, tomography is used in astronomy, for example for a three-dimensional study of the corona of the sun. Given that projection images have to be taken over a period of at least several days — only a few satellites are available to take the pictures — the dynamics of the sun must be taken into account during the reconstruction process. This can be modelled using techniques such as Markov random processes [46].

This concludes the basic introduction to the world of tomography. For a more in-depth overview of the field, the reader is referred to detailed works by Kak&Slaney [13], Natterer [47], and Buzug [9].

References

- [1] W. C. Röntgen, “Ueber eine neue art von strahlen,” *Sonderdruck aus den Sitzungsberichten der Wuzburger Physik.-medic. Gesellschaft*, 1896.
- [2] J. Radon, “Ueber die bestimmung von funktionen durch ihre integralwaerte laengs gewisser maennigfaltigkeiten,” *Ber. Saechs. Akad.*, vol. 69, pp. 262–277, 1917.
- [3] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, “Three-dimensional atomic imaging of crystalline nanoparticles,” *Nature*, vol. 470, pp. 374–377, 2011.
- [4] P. Sukovic and N. H. Clinthorne, “Penalized weighted least-squares image reconstruction for dual energy x-ray transmission tomography,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 11, pp. 1075–1081, 2000.
- [5] J. S. Iwaczyk, E. Nygard, O. Meirav, J. Arenson, W. C. Barber, N. E. Hartsough, N. Malakhov, and J. C. Wessel, “Photon counting energy dispersive detector arrays for x-ray imaging,” *IEEE Transactions on Nuclear Science*, vol. 56, pp. 535–542, June 2009.
- [6] M. G. Bisogni, A. Del Guerra, N. Lanconelli, A. Lauria, G. Mettivier, M. C. Montesi, D. Panetta, R. Pani, M. G. Quattrocchi, P. Randaccio, V. Rosso, and P. Russo,

-
- “Experimental study of beam hardening artifacts in photon counting breast computed tomography,” in *Proceedings of the 11th International Vienna Conference on Instrumentation*, vol. 581, pp. 94–98, 2007.
- [7] A. V. Bronnikov, “Theory of quantitative phase-contrast computed tomography,” *Journal of the Optical Society of America*, vol. 19, no. 3, pp. 472–480, 2002.
- [8] L. Tianfang, L. Xiang, W. Jing, W. Junhai, L. Hongbing, H. Jiang, and L. Zhen-gong, “Nonlinear sinogram smoothing for low-dose x-ray ct,” *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2505–2513, 2004.
- [9] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer-Verlag Berlin Heidelberg, 2007.
- [10] R. L. Siddon, “Fast calculation of the exact radiological path for a three-dimensional ct array,” *Medical Physics*, vol. 12, no. 2, pp. 252–255, 1985.
- [11] R. M. Lewitt, “Alternatives to voxels for image representation in iterative reconstruction algorithms,” *Physics in Medicine and Biology*, vol. 37, no. 3, pp. 705–716, 1992.
- [12] P. M. Joseph, “An improved algorithm for reprojecting rays through pixel images,” *IEEE Transactions on Medical Imaging*, vol. 1, no. 3, pp. 192–196, 1982.
- [13] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.
- [14] L. Feldkamp, L. Davis, and J. Kress, “Practical cone-beam algorithm,” *Journal of the Optical Society of America*, vol. 1, pp. 612–619, 1984.
- [15] H. Kudo, T. Rodet, F. Noo, and M. Defrise, “Exact and approximate algorithms for helical cone-beam ct,” *Physics in Medicine and Biology*, vol. 49, no. 13, pp. 2913–1397, 2004.
- [16] A. Katsevich, “Analysis of an exact inversion algorithm for spiral cone-beam ct,” *Physics in Medicine and Biology*, vol. 43, no. 15, pp. 2583–2597, 2002.
- [17] J. Gregor and T. Benson, “computational analysis and improvement of SIRT,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 918–924, 2008.
- [18] J. Gregor, M. Lenox, P. Bingham, and L. Arrowood, “multi-core cluster implementation of sirt with application to cone beam micro-ct,” in *Nuclear Science Symposium Conference Record*, pp. 4120–4125, 2009.
- [19] W. Xu, F. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard, and K. Mueller, “High-performance iterative electron tomography reconstruction with long-object compensation using graphics processing units (gpus),” *Journal of Structural Biology*, vol. 171, no. 2, pp. 142–153, 2010.
-

REFERENCES

- [20] K. Mueller, *Fast and accurate three-dimensional reconstruction from cone-beam projection data using algebraic methods*. PhD thesis, The Ohio State University, 1998.
- [21] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 601–609, 1994.
- [22] C. Paige and M. Saunders, "Lsqr: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software*, vol. 8, pp. 43–71, 1982.
- [23] D. Ludwig, "The Radon transform on Euclidean space," *Communications on Pure and Applied Mathematics*, vol. 19, pp. 49–81, 1966.
- [24] S. Helgason, *The Radon Transform*. Boston, MA: Birkhauser, 1980.
- [25] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Metal streak artifacts in x-ray computed tomography: a simulation study," *IEEE Transactions on Nuclear Science*, vol. 46, no. 3, pp. 691–696, 1999.
- [26] G. Van Gompel, *Toward accurate image reconstruction from truncated X-ray CT projections*. PhD thesis, University of Antwerp, 2009.
- [27] P. M. Joseph and C. Ruth, "A method for simultaneous correction of spectrum hardening artifacts in ct images containing both bone and iodine," *Medical Physics*, vol. 24, pp. 1629–1634, 1997.
- [28] M. Krumm, S. Kasperl, and M. Franz, "Reducing non-linear artifacts of multi-material objects in industrial 3d computed tomography," *Nondestructive Testing and Evaluation*, vol. 41, pp. 242–251, 2008.
- [29] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. J. Batenburg, J. de Mey, J. Sibjers, and N. J., "Iterative correction of beam hardening artifacts in ct," *Medical Physics*, vol. 38, 2011.
- [30] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Seutens, "An iterative maximum-likelihood polychromatic algorithm for ct," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 999–1008, 2001.
- [31] R. M. Lewitt and R. H. T. Bates, "Image reconstruction from projections iii: projection completion methods (theory)," *Optik*, vol. 50, pp. 189–204, 1977.
- [32] J. L. Prince and A. Willsky, "Constrained sinogram restoration for limited-angle tomography," *Optical Engineering*, vol. 29, pp. 535–543, 1990.
- [33] P. S. Cho, A. D. Rudd, and R. H. Johnson, "Cone-beam ct from width-truncated projections," *Computerized Medical Imaging and Graphics*, vol. 20, no. 1, pp. 49–57, 1996.

-
- [34] I. K. Chun, M. H. Cho, S. C. Lee, M. Cho, and S. Y. Lee, "X-ray micro-tomography system for small-animal imaging with zoom-in imaging capability," *Physics in Medicine and Biology*, vol. 49, pp. 3889–3902, 2004.
- [35] F. Noo, R. Clackdoyle, and J. D. Pack, "A two-step hilbert transform method for 2d image reconstruction," *Physics in Medicine and Biology*, vol. 49, no. 17, pp. 3903–3923, 2004.
- [36] J. Fessler, "Segmented attenuation correction for pet," in *Nuclear Science Symposium and Medical Imaging Conference*, pp. 1182–1184, 1992.
- [37] P. A. Midgley and R. E. Dunin-Borkowski, "Electron tomography and holography in materials science," *Nature Materials*, vol. 8, pp. 271–280, 2009.
- [38] P. A. Midgley and M. M. Weyland, "3D electron microscopy in the physical sciences: the development of z-contrast and efem tomography," *Ultramicroscopy*, vol. 96, pp. 413–431, 2003.
- [39] X. Fu, E. Knudsen, H. F. Poulsen, G. T. Herman, B. M. Carvalho, and H. Y. Liao, "Optimized algebraic reconstruction technique for generation of grain maps based on three-dimensional x-ray diffraction (3dxrd)," *Optical Engineering*, vol. 45, 2006.
- [40] J. Banhart, *Advanced Tomographic Methods in Materials Research and Engineering*. Oxford University Press, 2008.
- [41] E. Perilli, V. Le, B. Ma, P. Salmon, K. Reynolds, and N. L. Fazzalari, "Detecting early bone changes using in vivo micro-ct in ovariectomized, zoledronic acid-treated, and sham-operated rats," *Osteoporosis International*, vol. 21, pp. 1371–1382, 2010.
- [42] A. Parfitt, M. Drezner, F. Glorieux, J. Kanis, H. Malluche, P. Meunier, S. Ott, and R. Recker, "Bone Histomorphometry: standardization of nomenclature, symbols and units," *Journal of Bone and Mineral Research*, vol. 2, no. 6, pp. 595–610, 1987.
- [43] S. Bals, K. J. Batenburg, J. Verbeeck, J. Sijbers, and G. Van Tendeloo, "Quantitative three-dimensional reconstruction of catalyst particles for bamboo-like carbon nanotubes," *Nano Letters*, vol. 7, no. 12, pp. 3669–3674, 2007.
- [44] G. Nolet, *A Breviary of Seismic Tomography: Imaging the Earth and Sun*. Cambridge University Press, 2008.
- [45] J. Roser, R. Stopar, and A. Gosar, "Application of seismic tomography in investigations of the motorway alignment in the sentvid tunnel area," *Materials and geoenvironment*, vol. 55, no. 1, pp. 67–84, 2008.
- [46] M. D. Butala, R. J. Hewett, R. A. Frazin, and F. Kamalabadi, "Dynamic three-dimensional tomography of the solar corona," *Solar Image Processing and Analysis*, vol. 262, no. 2, pp. 495–509, 2010.
- [47] F. Natterer and F. Wubberling, *Mathematical Methods in Image Reconstruction*. SIAM: Society for Industrial and Applied Mathematics, 2001.
-

Part I

Tomographic Segmentation

—If you don't eat yer meat, you can't have any pudding!
How can you have any pudding if you don't eat yer meat?!

Roger Waters, Another Brick in the Wall, 1979.

2

Segmentation of tomographic images using Projection Distance Minimization

Contents

2.1	Introduction	41
2.2	Concept and notation	44
2.3	Implementation	46
2.3.1	Grey level optimization	47
2.3.2	Optimal threshold selection	48
2.4	PDM with local thresholding	51
2.5	Experiments	52
2.5.1	Simulation studies	52
2.5.2	Experimental studies	57
2.6	Conclusions	58
	References	58

2.1 Introduction

To extract quantitative information from a tomographic reconstruction, it is often segmented first. Segmentation refers to the act of classifying image pixels into a few distinct classes based on a certain characteristic, typically their grey level. Practically speaking, it answers the question: “Which pixels of the reconstructed image belong to which structure?”

CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION

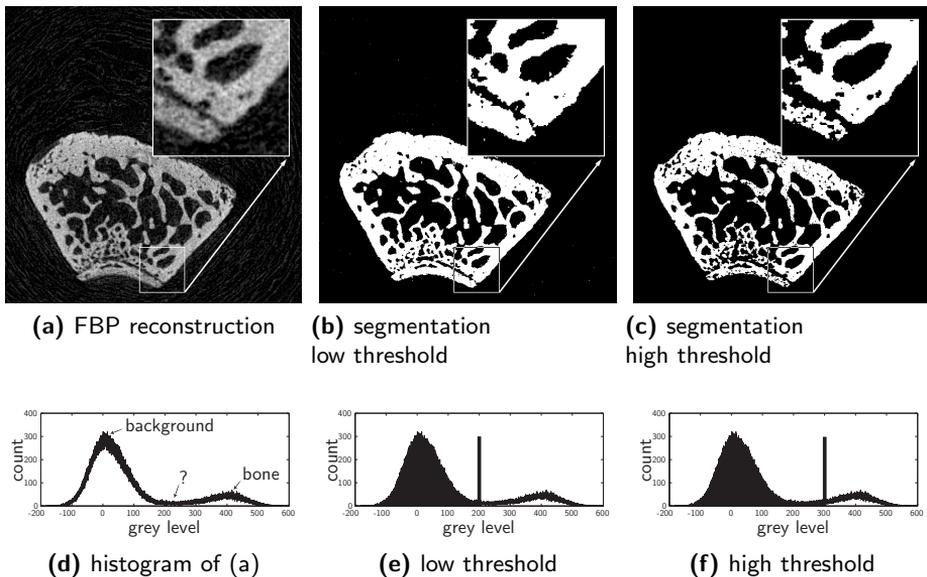


Figure 2.1: (a) An FBP reconstruction of a rat femur. The scan was made in a SkyScan 1172 μ CT scanner with a camera pixel size of $11.73\mu\text{m}$. 368 projection angles were used. (b-c) Segmentations of (a) with a low and high threshold value. (d) Histogram of (a) (e-f) threshold values of (b) and (c) visualized on the histogram.

Consider Fig. 2.1a, an FBP reconstruction of a rat femur from 368 equiangular cone beam projections, scanned with a SkyScan 1172 μ CT scanner. The reconstructed image has a continuous grey scale¹ and while creating a segmentation of this image may look trivial at first, a close-up view reveals several pixels — especially those at the border of the object — that cannot be easily attributed to either the foreground or the background. This is also visible in Fig. 2.1d, the histogram of Fig. 2.1a. Two peaks clearly represent the pixels of the background and of the bone, but there are also many pixels whose grey level lies in the middle. It is especially for these pixels that an accurate segmentation technique is required.

In the literature, a multitude of segmentation algorithms have been described. Some methods are region-based, e.g. watershed segmentation [1] and region growing [2]. In other methods, object structures are modelled with active contours or snakes [3, 4]. Presumptively the simplest, yet most widely used technique, is

¹This is technically not true as the set of possible values in reconstructed images is limited by what can be represented by computational hardware. However, the rounding errors thereby introduced have a negligible effect on the adequacy of the reconstruction and can therefore be ignored.

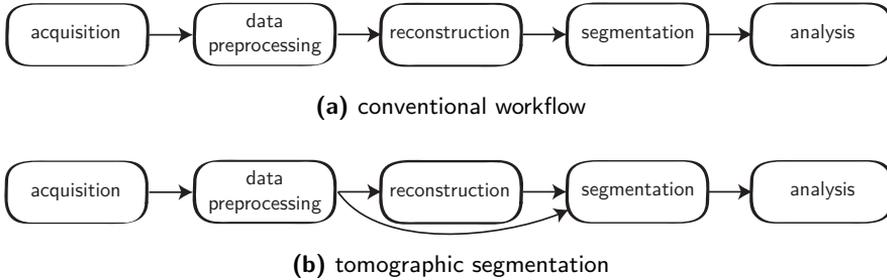


Figure 2.2: (a) Conventional workflow of a typical tomographic applications. (b) Workflow with tomographic segmentation.

thresholding. However, with thresholding, the accuracy of the segmented images is determined by the chosen threshold values [5]. In Fig. 2.1b and Fig. 2.1c, two different threshold values have been applied to Fig. 2.1a (the threshold values are also visualized in the histograms shown in Fig. 2.1e and Fig. 2.1f). This shows that even a seemingly small change in threshold choice can have a large effect on the outcome.

To select the optimal threshold, several algorithms have been proposed [6]. *Histogram shape-based methods* consider the shape of the histogram and use this information to locate its peaks (the grey levels of the classes) and valleys (the threshold values) [7, 8]. *Clustering-based methods* divide the histogram into several clusters (one for each class). This can be done by numerous methods. With Expectation Maximization (EM) the histogram is modelled with parametric basis functions such as Gaussian Mixture Models (GMM) [9]. With k-means clustering, each pixel is attributed to the class whose nearest mean it is closest to [10]. With the iterative Otsu’s clustering method, the weighted sum of intra-class variances is minimized [11]. *Entropy-based methods* do not make use of the histogram. Instead, they attempt to maximize the entropy between the foreground and background pixels [12] or to minimize the cross-entropy between the reconstruction and the segmentation [13].

Most thresholding techniques provide satisfactory results only if the contrast between two materials is sufficiently large. If this is not the case, the grey level distributions of each class overlap too much. Also, small objects (e.g. microcalcifications in medical scans), are easily ignored as their contribution in the histogram is barely noticeable.

For tomographic applications, advanced schemes are typically used that combine classical segmentation methods with additional post-processing steps. As an example, in [14], segmentation of lung tissue is proposed as a three step algorithm

CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION

containing (1) thresholding and a sequence of morphological operations to extract the lungs from the scan, (2) a region-growing method to detect the large airways, and (3) a dynamic programming step to separate the left and right lung. In [15], lung lobe segmentation is performed with an atlas-driven approach.

None of the methods here described contain an inherent measure to assess the accuracy of the segmented image. Furthermore, they are all solely based on the reconstructed image (Fig. 2.2a). As seen in Section 1.4, such a reconstructed image is prone to several types of reconstruction artefacts that might obstruct accurate threshold selection. In [16], a threshold selection method, called *Projection Distance Minimization (PDM)*, was introduced that uses the projection data to determine the optimal threshold value. Such a method is referred to as an *tomographic segmentation* technique. Note that this method effectively transforms the tomographic workflow to Fig. 2.2b.

In Section 2.2, the general concept of PDM is explained and notation is introduced. In Section 2.3, two optimization approaches for PDM are discussed. Section 2.4 extends the PDM approach to a local thresholding scheme. Section 2.5 then validates the method on several simulated and experimental datasets. The method is compared to conventional segmentation methods such as Otsu's method and k-means clustering. Ultimately, Section 2.6 concludes this chapter.

2.2 Concept and notation

Consider $\mathbf{v} = (v_j) \in \mathbb{R}^n$, a reconstructed image that approximately satisfies

$$\mathbf{W}\mathbf{v} = \mathbf{p}. \quad (2.1)$$

In Eqn. 2.1, $\mathbf{p} = (p_i) \in \mathbb{R}^m$ is the projection data and $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{m \times n}$ is the projection matrix that maps \mathbf{v} onto \mathbf{p} (as introduced in Section 1.3.1.3).

A segmentation of \mathbf{v} can be considered as a partition of its pixels into l classes. Each pixel v_j is contained by exactly one class, denoted by $s_j \in \{1, \dots, l\}$. The vector $\mathbf{s} = (s_j) \in \{1, \dots, l\}$ is referred to as the *classification mask*. Define $\boldsymbol{\rho} = (\rho_1, \dots, \rho_l)^T$, the vector containing all grey levels of the segmented image, one for each class. The *segmentation function* or *classification function* is then:

$$\mathcal{C}(\mathbf{s}, \boldsymbol{\rho}) = (\rho_{s_1}, \dots, \rho_{s_n})^T. \quad (2.2)$$

The *Projection Distance Minimization (PDM)* method, as introduced in [16], provides an objective measure to assess the accuracy of the chosen parameters \mathbf{s} and $\boldsymbol{\rho}$. As its name implies, it states that the optimal parameters are those for which the Euclidean distance between the projection of the segmentation and the

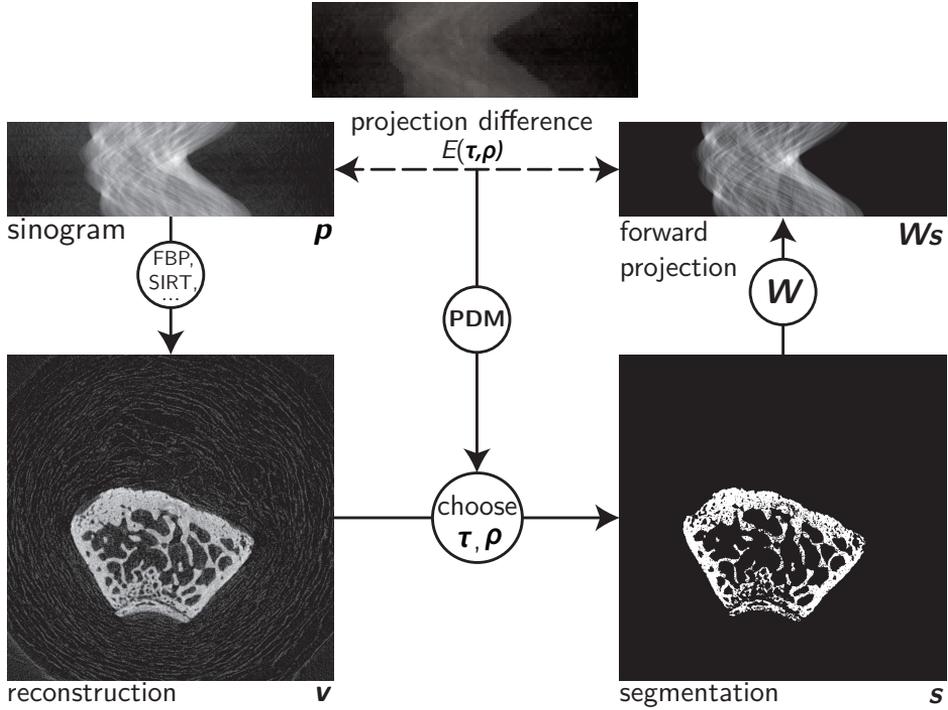


Figure 2.3: Schematic overview of the Projection Distance Minimization (PDM) concept.

measured projection data is minimal, i.e.:

$$s_{opt}, \rho_{opt} = \operatorname{argmin}_{s \in \mathbb{R}^n, \rho \in \mathbb{R}^l} \|WC(s, \rho) - p\|_2. \quad (2.3)$$

The search space of the optimization problem in Eqn. 2.3 has an extremely high dimensionality ($n+l$), rendering the optimum hard to locate. Therefore, segmented images are commonly obtained by application of a global thresholding scheme on a certain, priorly reconstructed image $\bar{v} \in \mathbb{R}^n$. Define the *threshold function* $\mathcal{I}_\tau(\bar{v}) : \mathbb{R}^n \rightarrow \{1, \dots, l\}^n$ as the function that assigns a value in $\{1, \dots, l\}$ to each pixel of the reconstructed image according to threshold values $\tau = (\tau_1, \dots, \tau_{l-1})^T$:

$$\mathcal{I}_\tau(\bar{v}_j) = \begin{cases} 1 & \bar{v}_j < \tau_1 \\ 2 & \tau_1 \leq \bar{v}_j < \tau_2 \\ \vdots & \\ l & \tau_{l-1} \leq \bar{v}_j \end{cases}, \quad j = 1, \dots, n. \quad (2.4)$$

CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION

Also define the *segmentation function* $\mathcal{C}_{\tau,\rho}(\mathbf{v}) : \mathbb{R}^n \rightarrow \{\rho_1, \dots, \rho_l\}^n$ as the function that assigns one of the grey levels in the vector $\boldsymbol{\rho} = (\rho_1, \dots, \rho_l)^T$ to each pixel of the reconstructed image according to the threshold function:

$$\mathcal{C}_{\tau,\rho}(\bar{\mathbf{v}}) = \{\rho_{\mathcal{I}_{\tau}(\bar{v}_1)}, \dots, \rho_{\mathcal{I}_{\tau}(\bar{v}_n)}\}. \quad (2.5)$$

The optimization problem of Eqn. 2.3 can then be translated into:

$$\boldsymbol{\tau}_{opt}, \boldsymbol{\rho}_{opt} = \underset{\boldsymbol{\tau} \in \mathbb{R}^{l-1}, \boldsymbol{\rho} \in \mathbb{R}^l}{\operatorname{argmin}} \|\mathbf{W}\mathcal{C}_{\boldsymbol{\tau},\boldsymbol{\rho}}(\mathbf{v}) - \mathbf{p}\|_2. \quad (2.6)$$

Evaluating Eqn. 2.6 is an optimization problem of dimension $2l - 1$.

A schematic overview of PDM segmentation is shown in Fig. 2.3. At first, a reconstruction is created using any available reconstruction method. Values for $\boldsymbol{\tau}$ and $\boldsymbol{\rho}$ are then chosen and a segmented image is created. This segmentation is then forward projected into the projection domain and its Euclidean distance from the original projection data is computed. An optimization strategy then updates $\boldsymbol{\tau}$ and $\boldsymbol{\rho}$ to improve the accuracy of the segmentation. This optimization can be implemented in multiple ways, discussed in the next section. Note that there is no restriction on the kind of projection geometry that is used as long as it can be expressed in a projection matrix \mathbf{W} .

2.3 Implementation

For computational efficiency, Eqn. 2.6 can be split into two smaller optimization problems.

- Inner optimization: Given certain threshold values $\bar{\boldsymbol{\tau}}$, determine the optimal grey levels, $\boldsymbol{\rho}_{opt}$:

$$\boldsymbol{\rho}_{opt} = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \|\mathbf{W}\mathcal{C}_{\bar{\boldsymbol{\tau}},\boldsymbol{\rho}}(\mathbf{v}) - \mathbf{p}\|_2. \quad (2.7)$$

This optimization problem is discussed in Section 2.3.1.

- Outer optimization: Optimize $\boldsymbol{\tau}$ by solving:

$$\boldsymbol{\tau}_{opt} = \underset{\boldsymbol{\tau}}{\operatorname{argmin}} \|\mathbf{W}\mathcal{C}_{\boldsymbol{\tau},\boldsymbol{\rho}_{opt}}(\mathbf{v}) - \mathbf{p}\|_2. \quad (2.8)$$

In each function evaluation, $\boldsymbol{\rho}_{opt}$ is given by the inner optimization. This optimization problem is discussed in Section 2.3.2, where two different implementation approaches are compared.

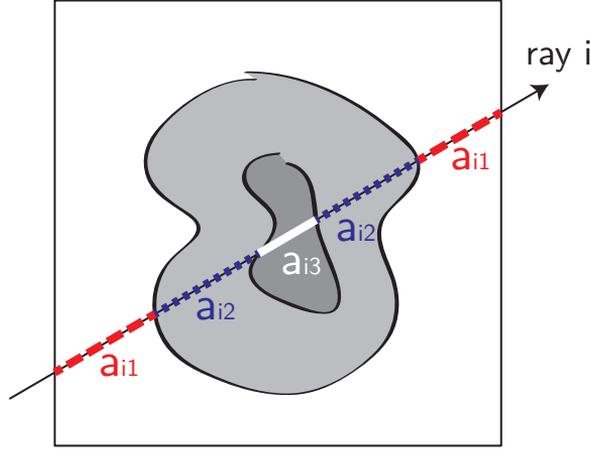


Figure 2.4: Each column of \mathbf{A} contains the length of each ray through the partition i .

2.3.1 Grey level optimization

Let $\mathbf{v} = (v_j) \in \mathbb{R}^n$ be an image vector and let $\bar{\tau}$ be a given vector of threshold values, for each $t \in \{1, \dots, l\}$, let $\bar{\mathbf{s}}_t \in \{0, 1\}^n$ denote the *classification mask of class t* :

$$\bar{s}_j^{(t)} = \begin{cases} 1 & \text{if } \mathcal{I}_{\bar{\tau}}(v_j) = t \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, \dots, n. \quad (2.9)$$

Let $\mathbf{A} = (a_{it}) \in \mathbb{R}^{m \times l}$ denote an $m \times l$ matrix where column t contains the forward projection of $\bar{\mathbf{s}}_t$. The value a_{it} is thus defined by the total weight contribution to ray i of all pixels that belong to partition t . This equals the length that each ray traverses through partition t . This is visualized in Fig. 2.4.

$$a_{it} = \sum_{j=1}^n w_{ij} \bar{s}_{ti} = \sum_{j: \bar{s}_{tj}=1} w_{ij} \quad (2.10)$$

For any vector of grey levels $\boldsymbol{\rho}$, the forward projection of $\mathcal{C}_{\bar{\tau}, \boldsymbol{\rho}}(\mathbf{v})$ can then be written as

$$\mathbf{W} \mathcal{C}_{\bar{\tau}, \boldsymbol{\rho}}(\mathbf{v}) = \sum_{t=1}^l \mathbf{W} \bar{\mathbf{s}}_t \rho_t = \mathbf{A} \boldsymbol{\rho}. \quad (2.11)$$

Let \mathbf{a}_j denote the j^{th} row of \mathbf{A} and let $\mathbf{c}_j = -2p_j \mathbf{a}_j$, $\bar{\mathbf{c}} = \sum_{j=1}^m \mathbf{c}_j$, $\mathbf{Q}_j = \mathbf{a}_j \mathbf{a}_j^T$, $\bar{\mathbf{Q}} = \sum_{j=1}^m \mathbf{Q}_j$. The projection difference of the segmented image, from here on

noted by $\mathcal{E}(\boldsymbol{\tau}, \boldsymbol{\rho})$, then becomes:

$$\mathcal{E}(\boldsymbol{\tau}, \boldsymbol{\rho}) = \|\mathbf{W}\mathcal{C}_{\bar{\boldsymbol{\tau}}, \boldsymbol{\rho}}(\mathbf{v}) - \mathbf{p}\|_2 = \|\mathbf{A}\boldsymbol{\rho} - \mathbf{p}\|_2 \quad (2.12)$$

$$= \sum_{j=1}^m (\mathbf{a}_j^T \boldsymbol{\rho} - p_j)^2 \quad (2.13)$$

$$= \sum_{j=1}^m (\boldsymbol{\rho}^T \mathbf{Q}_j \boldsymbol{\rho} - \mathbf{c}_j^T \boldsymbol{\rho} + p_j^2) \quad (2.14)$$

$$= \boldsymbol{\rho}^T \bar{\mathbf{Q}} \boldsymbol{\rho} - \bar{\mathbf{c}}^T \boldsymbol{\rho} + |\mathbf{p}|^2. \quad (2.15)$$

Given that the projection difference is a quadratic polynomial in $\boldsymbol{\rho}$ which cannot take negative values, the optimal grey levels, $\boldsymbol{\rho}_{opt}$, can be computed by setting the derivatives of Eqn. 2.15 with respect to ρ_1, \dots, ρ_l to zero, obtaining the system Eqn. 2.16 which is easy to solve, as the number of grey levels is typically very small.

$$2\bar{\mathbf{Q}}\boldsymbol{\rho}_{opt} = -\bar{\mathbf{c}}. \quad (2.16)$$

2.3.2 Optimal threshold selection

To obtain the optimal threshold values, the following function must be minimized with respect to $\boldsymbol{\tau}$:

$$\|\mathbf{W}\mathcal{C}_{\boldsymbol{\tau}, \boldsymbol{\rho}_{opt}}(\mathbf{v}) - \mathbf{p}\|_2. \quad (2.17)$$

Each evaluation of this function requires the computation of $\boldsymbol{\rho}_{opt}$. Note that this function is not differentiable to $\boldsymbol{\tau}$. In the remainder of this section, two optimization strategies to find the minimum of Eqn. 2.17 w.r.t. $\boldsymbol{\tau}$, are investigated.

2.3.2.1 Simple approach

A simple optimization strategy to solve Eqn. 2.8 involves classical global optimization techniques such as the Nelder-Mead simplex search [17].

In Fig. 2.5, the pseudo code of this approach is given. The largest computational burden lies with the l forward projection operations that are to be performed per function evaluation. One should consider, however, that these operations are very suited for implementation on a modern *graphical processing unit (GPU)*, greatly increasing the applicability of this approach.

2.3.2.2 Advanced approach

An advanced optimization strategy tailored to the specific problem, was proposed in [16]. One should observe that the projection difference equation (Eqn. 2.15) is

```

function  $\mathcal{E}(\bar{\tau}, \rho_{opt}), \rho_{opt} = \text{optimize\_grey\_level}(\bar{\tau}, \mathbf{v}, \mathbf{W}, \mathbf{p})$ 
for  $t = 1:l$ 
     $\bar{s}_j^{(t)} = \begin{cases} 1 & \text{if } \mathcal{I}_{\bar{\tau}}(v_j) = t \\ 0 & \text{otherwise} \end{cases}, j = 1, \dots, n.$ 
    column  $t$  of  $\mathbf{A} = \mathbf{W}\bar{\mathbf{s}}^{(t)}$ ;
end

 $\rho_{opt} = \text{solution of } 2\bar{\mathbf{Q}}\rho = -\bar{\mathbf{c}} \text{ w.r.t. } \rho;$ 
 $\mathcal{E}(\bar{\tau}, \rho_{opt}) = \|\mathbf{A}\rho_{opt} - \mathbf{p}\|_2;$ 
end

function  $\tau_{opt} = \text{PDM\_simple}(\mathbf{v}, \mathbf{W}, \mathbf{p})$ 
 $\tau_{opt} = \text{argmin}_{\tau} [\text{optimize\_grey\_level}(\tau, \mathbf{v}, \mathbf{W}, \mathbf{p})]$ 
end
    
```

Figure 2.5: Pseudo code for the simple PDM approach.

written in such a way that it can be easily updated if only a few pixels switch to a different class. That way, an optimization technique can be introduced that finds the optimal thresholds without the need for constant forward projection.

First, assume \mathbf{A} , $\bar{\mathbf{Q}}$ and $\bar{\mathbf{c}}$ are computed for a certain set of threshold values τ . Now, consider the alteration of one threshold value τ_t by a very small amount. This means that only a few pixels v_j switch into a different partition. Define $B \in \{1, \dots, m\}$ as the set of projection rays i that pass through any switched pixel, i.e. B contains all i 's for which $w_{ij} \neq 0$ and for which the classification of v_j is switched by the threshold update. The updated projection mask matrix, \mathbf{A}' , is then identical to \mathbf{A} except for the rows $i \in B$. The updated vector $\bar{\mathbf{c}}'$ and matrix $\bar{\mathbf{Q}}'$ can then be computed as follows:

$$\bar{\mathbf{c}}' = \bar{\mathbf{c}} + \sum_{i \in B} (\mathbf{c}'_i - \mathbf{c}_i), \quad (2.18)$$

$$\bar{\mathbf{Q}}' = \bar{\mathbf{Q}} + \sum_{i \in B} (\mathbf{Q}'_i - \mathbf{Q}_i). \quad (2.19)$$

Given that the projection of a single pixel is typically non-zero in only a few detectors, these update steps can be exploited to create an efficient optimization technique. Suppose τ_0 is an initial estimate of the threshold values. The matrices \mathbf{A} , $\bar{\mathbf{c}}$ and $\bar{\mathbf{Q}}$ are then fully constructed, requiring one full forward projection for each class. Also, ρ_{opt} is determined using Eqn. 2.16 and the projection difference is computed using Eqn. 2.15. Next, a small update step is applied to any threshold.

CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION

```

function  $\tau_{opt} = \text{PDM\_advanced}(\mathbf{v}, \mathbf{p}, \tau_0)$ 

     $\tau = \tau_0$ ;
     $\bar{\mathbf{s}}^{(1)} = \mathbf{v} < \tau$ ,  $\bar{\mathbf{s}}^{(2)} = \mathbf{v} \geq \tau$ ;
     $\mathbf{A}_1 = \mathbf{W}\bar{\mathbf{s}}^{(1)}$ ,  $\mathbf{A}_2 = \mathbf{W}\bar{\mathbf{s}}^{(2)}$ ;

    for  $i = 1:m$ ,  $\mathbf{c}_i = -2p_i\mathbf{a}_i$ ,  $\mathbf{Q}_i = \mathbf{a}_i\mathbf{a}_i^T$ , end
     $\bar{\mathbf{c}} = \sum_{i=1}^m \mathbf{c}_i$ ,  $\bar{\mathbf{Q}} = \sum_{i=1}^m \mathbf{Q}_i$ ;

    stepsize = initial_step_size;
    while stepsize  $\geq$  min_step_size

         $\rho_{opt}$  = solution of  $2\bar{\mathbf{Q}}\rho = -\bar{\mathbf{c}}$  w.r.t.  $\rho$ ;
         $\mathcal{E} = \bar{\mathbf{c}}^T \rho_{opt} + \rho_{opt}^T \bar{\mathbf{Q}} \rho_{opt} + \mathbf{p}^2$ ;

        if  $\mathcal{E} < \mathcal{E}_{opt}$ ;
             $\tau_{opt} = \tau$ ;
             $\mathcal{E}_{opt} = \mathcal{E}$ ;
             $\tau = \tau + \text{stepsize}$ ;
        else
            stepsize = stepsize/2;
             $\tau = \tau - \text{stepsize}$ ;
        end

        for  $j = 1:n$ 
            if  $\mathbf{v} < \tau$ ,  $\bar{\mathbf{s}}'^{(1)} = 1$ ,  $\bar{\mathbf{s}}'^{(2)} = 0$ ;
            else,  $\bar{\mathbf{s}}'^{(1)} = 0$ ,  $\bar{\mathbf{s}}'^{(2)} = 1$ , end
        end
         $B = \{i : \exists j : \bar{\mathbf{s}}_j'^{(1)} \neq \bar{\mathbf{s}}_j^{(1)} \text{ and } w_{ij} \neq 0\}$ ;
         $\bar{\mathbf{c}}' = \bar{\mathbf{c}} + \sum_{i \in B} (\mathbf{c}'_i - \mathbf{c}_i)$ ;
         $\bar{\mathbf{Q}}' = \bar{\mathbf{Q}} + \sum_{i \in B} (\mathbf{Q}'_i - \mathbf{Q}_i)$ ;

         $\bar{\mathbf{s}}^{(1)} = \bar{\mathbf{s}}'^{(1)}$ ,  $\bar{\mathbf{s}}^{(2)} = \bar{\mathbf{s}}'^{(2)}$ ;
         $\bar{\mathbf{c}} = \bar{\mathbf{c}}'$ ,  $\bar{\mathbf{Q}} = \bar{\mathbf{Q}}'$ ;
    end
end

```

Figure 2.6: Simplified pseudo code for the advanced PDM approach. For brevity, $l = 2$ and it is assumed that τ_0 is an underestimation, i.e. $\tau_0 \leq \tau_{opt}$. The method can be easily extended for $l > 2$ and fitted with a more robust optimization strategy.

\mathbf{A} , \bar{c} and \bar{Q} are then updated (Eqn. 2.19), requiring a forward projection of only a few pixels. The new ρ_{opt} and projection difference are recomputed very efficiently. These steps are repeated until the projection difference does not improve. Fig. 2.6 shows simplified pseudo code for such an optimization strategy in the case where $l = 2$.

2.4 PDM with local thresholding

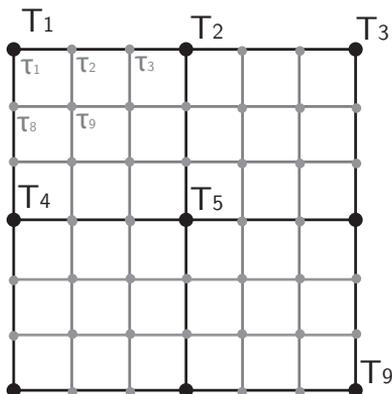


Figure 2.7: Local threshold field. Here $l = 2$ and $r = 3^2$.

In the previous section, a global thresholding scheme was used to find the partition mask with the minimal projection difference. However, the achievable accuracy of this method is limited by the accuracy of the reconstruction. If the reconstructed image displays variations in the intensity of certain image features, global thresholding can never lead to accurate segmentation. For example, small trabecular structures of bone tend to be represented less bright than large structures, even if they consist of the same material. To counter this problem, a scheme can be applied where different thresholds are used for different regions in the image, so-called *local thresholding* or *adaptive thresholding*.

Local threshold values can be selected based on different criteria. Many approaches compare the grey level of a certain pixel to those of its neighbours to determine the local threshold for that pixel [18, 19]. In [20] a small window is slid over the image, computing a different clustering segmentation method for each pixel. In [21], the decision is based on local entropy information.

As with the global thresholding techniques discussed in Section 2.1, these local thresholding methods do not take tomographic information into account. Additionally, adaptive techniques may perform even worse as they are sensitive to local

CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION

variations due to several possible image artefacts. In [22], the previously discussed PDM optimization method was extended to a local thresholding scheme.

Let $\mathbf{T} = (T_j) \in \mathbb{R}^{r \times (l-1)}$ denote the *coarse local threshold field*, the threshold values for certain points on the reconstruction grid. Define $\boldsymbol{\tau}_{local} \in \mathbb{R}^{n \times (l-1)}$ as the *fine local threshold field*, the threshold values for each point on the reconstruction grid. The values of $\boldsymbol{\tau}_{local}$ are defined by bilinear interpolation of \mathbf{T} (Fig. 2.7). The value r defines the number of local threshold values that are to be estimated. Generally, the higher r , the higher the quality of the resulting segmentations, but the more difficult the optimization problem becomes. If $r = 1$, there is no difference between global and local thresholding. If $r = n$, each pixel is considered separately and segmentation effectively becomes a, hard to solve, discrete tomography problem. This is investigated further in Chapter 4.

The values T_j are initialized with the optimal threshold values found using a global PDM approach. A single point on the coarse threshold field is then selected for further optimization. This selection can be done at random or guided by the current segmentation. Ideally, a point close to where the segmentation errors occur is chosen. This information can be gathered by computing a reconstruction from the current projection difference $\|\mathbf{W}\mathbf{s} - \mathbf{p}\|_2$. Areas where segmentation errors occur typically have higher values in this reconstructed image. Once a point is chosen, its threshold values are optimized using one of the two PDM approaches described in the previous section. This process is repeated until no further improvements can be made. At that point, the grid size of the coarse threshold field can optionally be increased to improve the result even further.

2.5 Experiments

To validate the PDM method, this section describes a series of experiments that was performed on simulated (Section 2.5.1) as well as experimental μ CT data (Section 2.5.2). PDM was compared to two classic threshold selection techniques: Otsu's method [11] and k-means clustering [10].

2.5.1 Simulation studies

Three 512×512 simulated phantom images (Fig. 2.8) with grey levels in the range $[0, 255]$ were considered. To assess the quality of the threshold selection, the *relative number of misclassified pixels (rNMP)* of the segmented images was calculated. This is the total number of pixels that are classified in a different partition than in the original phantom image, divided by the total number of non-zero pixels in the phantom image.

In a first experiment, CGLS reconstructions were created of Fig. 2.8a and

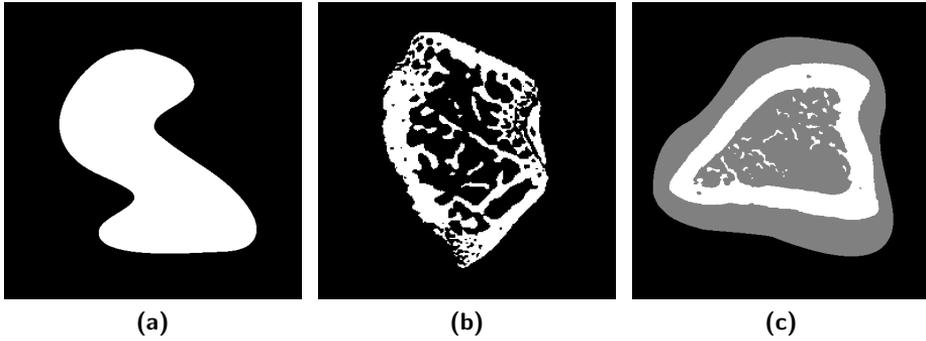


Figure 2.8: Three simulated 512×512 phantom images.

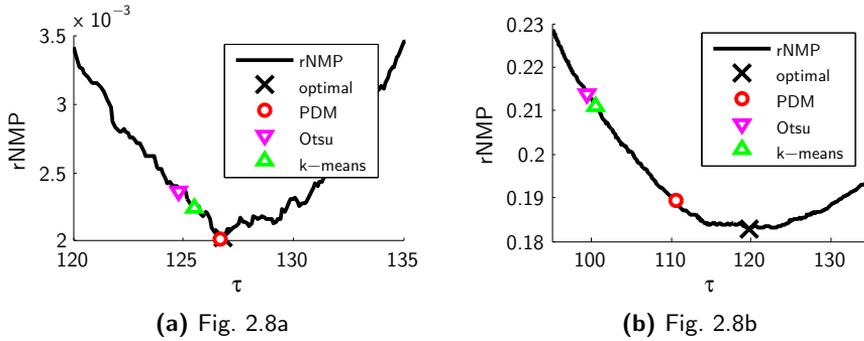


Figure 2.9: The rNMP as a function of the threshold value, τ . The optimal threshold values according to the different threshold selection techniques are also plotted.

Fig. 2.8b from 30 equiangular parallel beam projections. On the simulated projection data of Fig. 2.8b, Poisson noise was applied with incident beam intensity $I_0 = 50000$. Segmented images were created for a range of threshold levels τ and the corresponding rNMP was computed for each. PDM, Otsu and k-means were also applied on the reconstructed images. As visible in Fig. 2.9, in both cases the threshold selected by the PDM technique was clearly the best approximation of the optimal threshold (the one for which rNMP value was minimal).

In a second experiment, the computation time of the two implementation approaches described in Section 2.3.2 was investigated. To that extent, a CGLS reconstruction of Fig. 2.8b was computed from 90 equiangular parallel beam projections. PDM with a simplex-search optimization strategy (Section 2.3.2.1) was performed with a CPU and a GPU (NVIDIA CUDA) implementation. For PDM

CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION

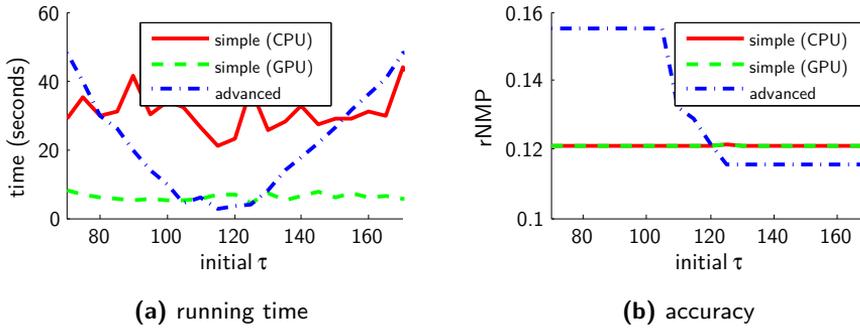


Figure 2.10: Comparison of the two different approaches for global PDM on a reconstructed image of Fig. 2.8b. The simple approach was often faster and its accuracy was less dependent on the initial threshold choice.

with the advanced optimization strategy (Section 2.3.2.2) only a CPU based implementation was used as this approach is not suited for GPU acceleration. Given that only small updates can be taken in the advanced method, its computation time is dependent on the initial choice for τ . Each approach was therefore evaluated for a range of initial values.

In Fig. 2.10a, the running time of each implementation is plotted as a function of the initial threshold. For the advanced method, the initial threshold value indeed had a large influence on the overall running time. As it required more forward projections than the advanced approach, the CPU implementation of the simple approach was generally much slower than the advanced method. Surprisingly however, when accelerated with GPU hardware, the simple approach proved to be faster than the advanced approach.

In Fig. 2.10b, the corresponding rNMP is visualized. It can be seen that with the advanced method also the accuracy depended on the initial choice. While the simple method did not always result in the lowest rNMP, it was much more consistent. In the experiments described in the remainder of this section, the simple approach was therefore used.

Next, the accuracy gain that can be achieved using local thresholding instead of global thresholding was investigated. Parallel beam projection data with 15 projection angles was generated of Fig. 2.8a. Then, a global PDM segmentation was created of a CGLS reconstruction (Fig. 2.11a). This result was subsequently further refined using local PDM optimization with $r = 64^2$ and $r = 16^2$ (Fig. 2.11b and Fig. 2.11c). In Fig. 2.11d, the rNMP is plotted for each step of the local optimization process. As expected, with local thresholding it was possible to achieve segmented images that were more accurate than the segmented image with the

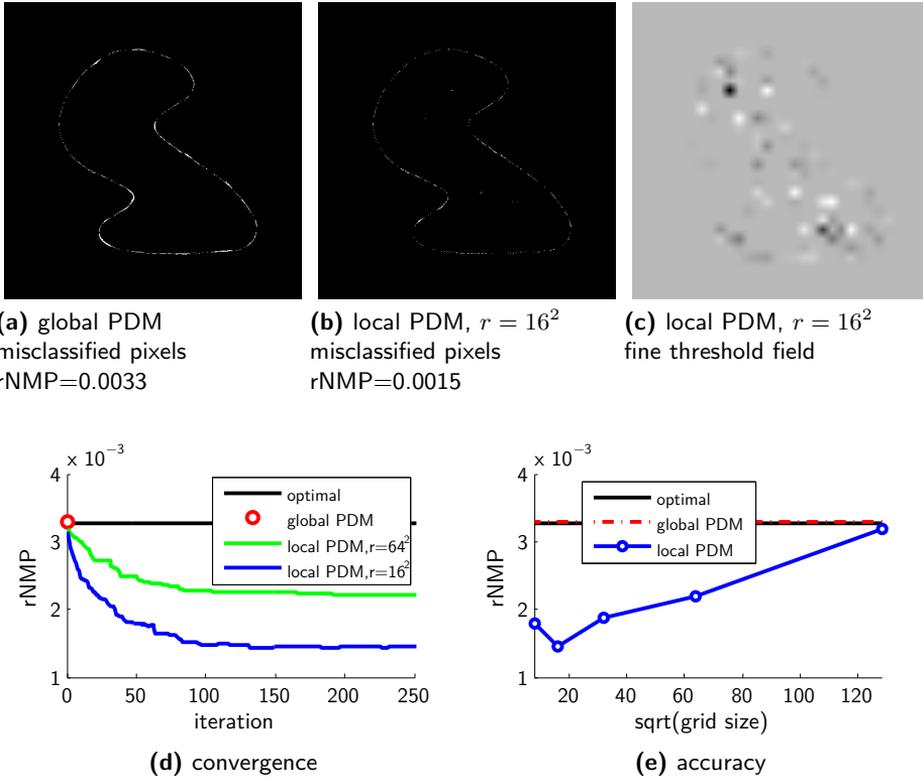
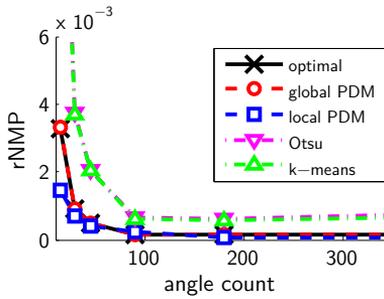


Figure 2.11: Comparison of global PDM and local PDM on Fig. 2.8a. **(a,b)** Misclassified pixels after global and local PDM segmentation respectively. **(c)** The fine local threshold field after local PDM converged. **(d)** The rNMP as a function of the iteration. **(e)** The rNMP as a function of the grid size r .

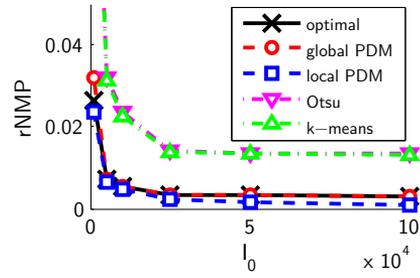
optimal global threshold. However, the eventual result greatly depended on the size of the local thresholding grid. In Fig. 2.11e, the rNMP of global and local PDM is plotted as a function of the gridsize. For these simulated images, $r = 16^2$ generally resulted in the most accurate segmentation. All remaining experiments were therefore ran with this value.

In a final set of simulation experiments, the performance of the segmentation techniques was measured with respect to the number of projection directions. As before, Poisson noise was applied to the projection data ($I_0 = 50000$) and CGLS reconstructions were computed. From Fig. 2.12a-c, one can note that from all global thresholding techniques, the PDM method provided the best approximation of the optimal threshold. This effect was the most profound if the number of

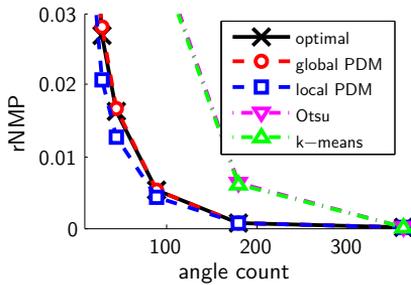
CHAPTER 2. SEGMENTATION OF TOMOGRAPHIC IMAGES USING PROJECTION DISTANCE MINIMIZATION



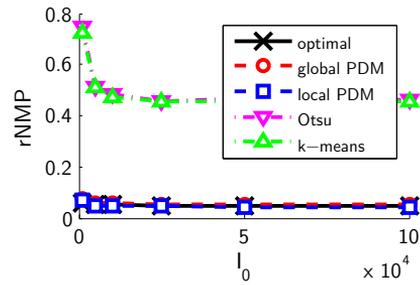
(a) Fig. 2.8a



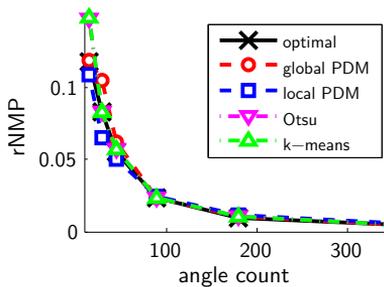
(d) Fig. 2.8a



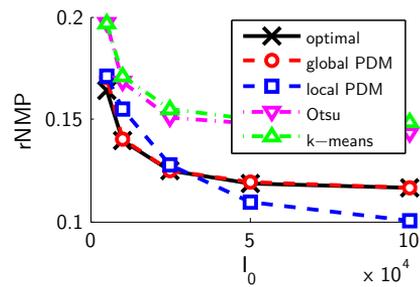
(b) Fig. 2.8b



(e) Fig. 2.8b



(c) Fig. 2.8c



(f) Fig. 2.8c

Figure 2.12: (a-c) The rNMP for all methods as a function of the number of projection directions. (d-f) The rNMP for all methods as a function of the number of counts per detector pixel. Lower counts imply more Poisson noise on the projection data.

projection angles was low. If the local thresholding scheme was applied, results were obtained that were more accurate than with the optimal global threshold. The performance was also evaluated with respect to the noise level, expressed by I_0 . 15 equiangular projection angles were used, effectively simulating low dose scans. In Fig. 2.12d-f, it can be observed that even with noisy projection data, segmented images created using the PDM approach provided the best results.

2.5.2 Experimental studies

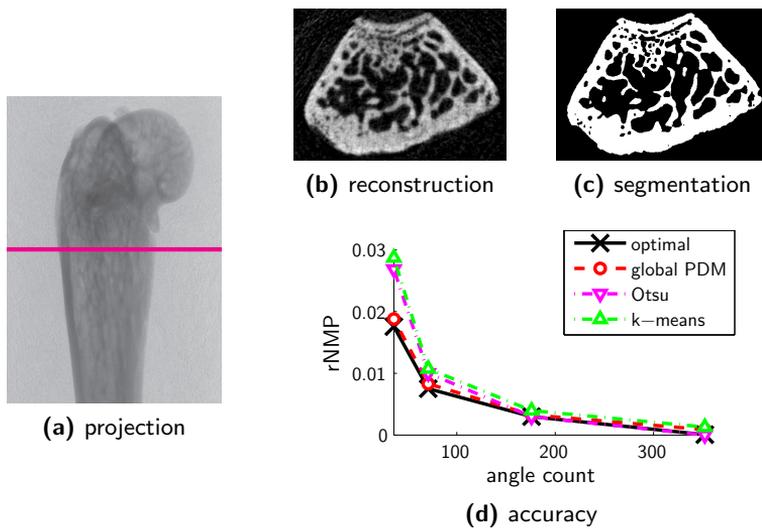


Figure 2.13: Experimental data experiment. **(a)** Projection image of a rat femur, scanned in a SkyScan 1172 μ CT scanner. The central slice is highlighted. **(b)** SIRT reconstruction of the central slice of **(a)** from 368 projections. **(c)** Otsu segmentation of **(b)**. Used as ground truth. **(d)** The rNMP as a function of the number of projections.

The segmentation methods discussed in this chapter were also applied on experimental μ CT data. Fig. 2.13a shows a single projection image of a rat femur, acquired with a SkyScan 1172 μ CT scanner with a camera pixel size of $11.73\mu\text{m}$ and a cone beam projection geometry. In total, 368 of such images were recorded equiangularly. This data was corrected for ring- and beam hardening artefacts using the SkyScan NRecon software. Fig. 2.13b shows a reconstruction of the central slice created using 500 SIRT iterations. This reconstruction was then segmented using Otsu’s commonly used clustering method. This segmentation, shown in Fig. 2.13c, was then taken as the ground truth. The number of projection angles was then reduced, reconstructed images were created and segmented images were

REFERENCES

computed and validated with respect to the ground truth. The results shown in Fig. 2.13d demonstrate the same effects as found in the simulation experiments. PDM segmentation provided the best approximation of the optimal global segmentation. This effect was the most profound when the number of projection directions was very low.

2.6 Conclusions

Prior to image analysis, a segmented image must first be created. This is commonly obtained by simply thresholding the reconstructed image. However, for accurate segmentation it is crucial to find adequate threshold values. A multitude of solutions to this problem have been proposed in the literature. Whereas these techniques attempt to find thresholds that conform with the image and its histogram, they do not search for thresholds that conform to the measured projection data.

In this chapter, the *Projection Distance Minimization (PDM)* algorithm, introduced in [16, 22], was discussed and investigated. The algorithm involves the forward projection of a segmented image that is defined by its segmentation parameters: the threshold values and the grey levels for each class. Its Euclidean distance from the measured projection data is then used to assess the quality of the segmentation parameters. As PDM uses the available projection data to guide the threshold selection, it is called an *tomographic segmentation* technique.

Applying a global threshold to an image can only create sufficiently accurate segmentation if the underlying reconstructed image is an accurate representation of the actual object. This is often not a valid assumption to make as multiple reconstruction artefacts might be present. To counter this, other segmentation schemes should be applied. Here, the use of the PDM metric on a local thresholding technique was investigated.

Experiments on simulated and μ CT data have shown that using PDM optimization of the global threshold values results in threshold values that are a very good approximation of the optimal global threshold. This is also true if the number of projection directions is low or the projection data is noisy. In those circumstances, PDM segmentation outperforms classical threshold selection techniques such as Otsu and k-means.

References

- [1] S. Beucher and F. Meyer, *The morphological approach to segmentation: the watershed transformation*. Mathematical Morphology in Image Processing, New York: Dekker, M., Dougherty E. R. ed., 1993.

-
- [2] X. Yu and J. Ylaaski, "A new algorithm for image segmentation based in region growing and edge detection," in *Proc. Int. Symp. Circuits and Systems*, vol. 1, pp. 516–519, 1991.
- [3] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.
- [4] A. J. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 199–209, 1997.
- [5] M. J. Eichler, C. H. Kim, R. Müller, and X. E. Guo, "Impact of thresholding techniques on micro-CT image based computational models of trabecular bone," in *ASME Advances in Bioengineering*, vol. 48, pp. 215–216, September 2000.
- [6] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–168, 2004.
- [7] M. I. Sezan, "A peak detection algorithm and its application to histogram-based image data reduction," *Computer Vision Graphics and Image Processing*, vol. 49, pp. 36–51, 1990.
- [8] A. Rosenfeld and P. Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, pp. 231–235, 1983.
- [9] S. Cho, R. Haralick, and S. Yi, "Improvement of kittler and illingworth's minimum error thresholding," *Pattern Recognition*, vol. 22, no. 5, pp. 609–617, 1989.
- [10] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, July 2002.
- [11] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, March 1979.
- [12] T. Pun, "A new method for grey-level picture thresholding using the entropy of the histogram," *Signal Processing*, vol. 2, pp. 223–237, 1980.
- [13] C. Li and P. Tam, "An iterative algorithm for minimum cross entropy thresholding," *Pattern Recognition Letters*, vol. 19, no. 8, pp. 771–776, 1998.
- [14] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 490–498, 2001.
-

REFERENCES

- [15] L. Zhang, E. A. Hoffman, and J. M. Reinhardt, "Atlas-driven lung lobe segmentation in volumetric x-ray ct images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 1, pp. 1–16, 2006.
- [16] K. J. Batenburg and J. Sijbers, "Optimal threshold selection for tomogram segmentation by projection distance minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 5, pp. 676–686, 2009.
- [17] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [18] J. M. White and G. D. G. D. Rohrer, "Image thresholding for optical character recognition and other applications requiring character image extraction," *IBM Journal of Research and Development*, vol. 27, pp. 400–411, 1983.
- [19] W. Niblack, *An introduction to digital image processing*. Strandberg Publishing Company, 1986.
- [20] L. Eikvil, T. Taxt, and K. Moen, "A fast adaptive method for binarization of document images," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 435–443, 1991.
- [21] S. A. Abutaleb, "Automatic thresholding of gray-level pictures using 2D entropy," *Computer Vision Graphics and Image Processing*, vol. 47, no. 1, pp. 22–32, 1989.
- [22] K. J. Batenburg and J. Sijbers, "Adaptive thresholding of tomograms by projection distance minimization," *Pattern Recognition*, vol. 42, no. 10, pp. 2297–2305, 2009.

—Weather forecast for tonight: Dark. Continued dark
overnight, with widely scattered light by morning.

George Carlin, 1978

3

Optimal threshold selection for segmentation of dense, homogeneous objects

Contents

3.1	Introduction	62
3.2	Concept and notation	64
3.2.1	SIRT	64
3.2.2	Sinogram inconsistency	65
3.3	Computational approach	66
3.3.1	Segmentation inconsistency	66
3.3.2	Grey level estimation	69
3.3.3	The SICM algorithm	71
3.4	Experiments	71
3.4.1	Simulation studies	71
3.4.2	Experimental studies	77
3.5	Conclusions	79
	References	80

This chapter has been published as:

Wim van Aarle, K. Joost Batenburg, Jan Sijbers, “*Optimal Threshold Selection for Segmentation of Dense Homogeneous Objects in Tomographic Reconstructions*”, IEEE Transactions on Medical Imaging, vol.30, no.4, pp.980–989, April 2011.

3.1 Introduction

This chapter deals with the segmentation of dense objects in tomographic images. It is assumed that the objects to be segmented have a constant density that is higher than that of the surrounding materials. In medical imaging, dense object segmentation is required in many applications. It can be used to suppress streak or beam hardening artefacts caused by metal implants [1, 2, 3]. Object identification and motion estimation (e.g. for co-registration) also often require detection of implanted markers on a device or guide wire [4, 5]. Furthermore, accurate localization of individual cochlear implant electrodes within the inner ear is important to model the electrical field of the cochlea [4, 6]. Also the extraction of the trabecular bone features (such as the cortical thickness or the cortical area) from the surrounding marrow spaces involves segmentation of the dense bone with respect to the background [7, 8].

Without noise or artefacts in the reconstructed image, segmentation of dense homogeneous objects would be trivial. However, in practice, accurate separation of such objects from the surroundings within a tomographic reconstruction is a non-trivial task for several reasons (in Section 1.4, a more in-depth overview is given).

- **Few-view tomography.** In many cases, the number of available projections is not sufficient to guarantee a unique reconstruction. Therefore, the computed reconstruction most likely does not correspond entirely to the underlying, unknown object.
- Experimental projection data is inevitably polluted by **noise and other data inconsistencies**, leading to inaccuracies in the reconstruction.
- **Approximations in the reconstruction algorithm.** Common reconstruction algorithms typically do not compute an exact inverse of the Radon transform, resulting in discrepancies between the reconstruction and the original object.

As a consequence, dense objects do not always show up as clear peaks in the histogram of a tomographic reconstruction.

For dense object segmentation, a common approach is to set a global threshold somewhere between the grey level of the pixels belonging to the object and those of the maximum value of the other pixels, which is referred to as the background in the remainder of this chapter. Typically, this threshold is selected based on the histogram of the reconstructed image [9]. If only a few materials are present and each of these correspond to a distinct grey level peak in the histogram, it is possible to accurately determine appropriate thresholds, for example by analysing

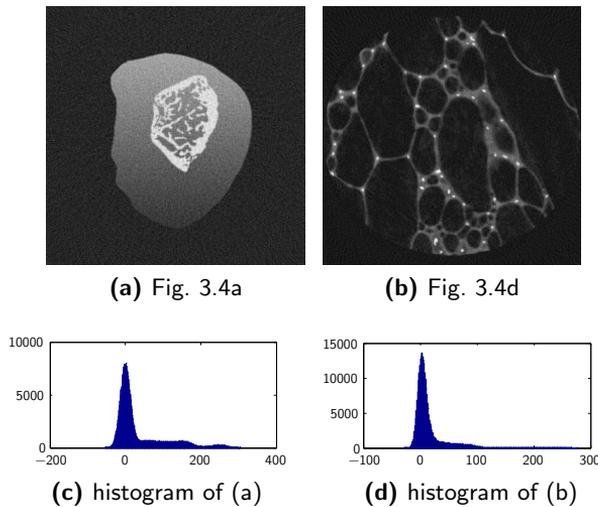


Figure 3.1: (a,b) SIRT reconstructions of Fig. 3.4a and Fig. 3.4d. (c) Histogram of (a). The peak representing the dense object partly overlaps that of the background. (d) Histogram of (b). Here, the contribution of the dense particles is not visible.

the concavity points on the convex hull of the histogram [10] or by modelling the histogram as a mixture of a series of Gaussian distributions [11]. The most popular global threshold selection method is the clustering method of Otsu [12]. It minimizes the weighted sum of intra-class variances of the different segmentation partitions.

A key issue with histogram-based methods in the context of segmenting a homogeneous object in a continuous grey level image, however, is that there are no guaranteed histogram peaks representing the continuous background. In that case, histogram-based methods are particularly inadequate if the object of interest is only slightly more dense than the surrounding materials. This is visible in Fig. 3.1.

Different approaches to segmentation of dense objects also exist, e.g. region-based algorithms such as region growing [13] and watershed segmentation [14]. These methods, however, are also solely based on the reconstructed image and are therefore very susceptible to reconstruction artefacts.

Ideally, reconstruction algorithms for tomography should be “invertible”, so that computed projections of the reconstructed image would equal the measured projection data. For FBP (Section 1.3.2.2), the most common reconstruction algorithm used in practice, this assumption does not hold, mainly due to various interpolation steps involved in the algorithm. Iterative algebraic methods (e.g. SIRT,

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

Section 1.3.3.1) only satisfy the invertibility assumption for the case of noiseless projection data and an infinite number of iterations. As the reconstructed image does not correspond accurately with the measured projections, using the projection data for the segmentation can potentially result in a segmentation that is more faithful to the original measurements.

In Chapter 2, a method called *Projection Distance Minimization (PDM)* was discussed. PDM exploits available projection data to optimally select the threshold values. This approach assumes that the scanned object contains a small number of different densities, each corresponding to a constant grey level in the reconstruction. By segmenting the reconstructed image, this property is restored in the reconstructed image. To measure the quality of this segmentation, projections of the segmentation are computed and compared to the measured projection data. An optimal segmentation results in maximal correspondence between the simulated projections and the measured dataset. However, this approach requires that the scanned object contains only a few different densities, and does not allow for segmentation of objects with a constant grey level in a continuously varying surrounding.

In this chapter, a new global threshold selection method for dense object segmentation is introduced. The *Segmentation Inconsistency Minimization (SICM)* method employs similar concepts as the PDM methods, but also allows parts of the image to vary freely. For each candidate segmentation, the projections of the segmented object are subtracted from the measured projection data, after which the remaining part of the image is reconstructed and checked for consistency with the residual projections. The threshold for which minimal inconsistency is obtained is selected for the segmentation. The only assumptions that are made, is that the density of the object is constant and that it is higher than all remaining densities in the scanned object.

The chapter is structured as follows. In Section 3.2, the tomographic notation is introduced. Section 3.3 describes the SICM threshold selection algorithm in detail. Experimental results are presented in Section 3.4. Section 3.5 concludes the chapter.

3.2 Concept and notation

3.2.1 SIRT

Consider $\mathbf{p} = (p_i) \in \mathbb{R}^m$ the vector of projection data and $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{m \times n}$ the projection operator. As described in Section 1.3.3.1, the iterative SIRT algorithm finds a solution to:

$$\mathbf{W}\mathbf{v} = \mathbf{p}, \tag{3.1}$$

by solving:

$$\tilde{\mathbf{v}} = \operatorname{argmin}_{\mathbf{v} \in \mathbb{R}^n} \|\mathbf{W}\mathbf{v} - \mathbf{p}\|_{\mathbf{R}}^2, \quad (3.2)$$

where $\tilde{\mathbf{v}} \in \mathbb{R}^n$ is the reconstructed image and $\mathbf{R} \in \mathbb{R}^{m \times m}$ is a diagonal matrix of the inverse row sums of \mathbf{W} , i.e. $r_{ii} = 1/\sum_j w_{ij}$. The norm $\|\cdot\|_{\mathbf{R}}$ is defined by $\|\mathbf{W}\mathbf{v} - \mathbf{p}\|_{\mathbf{R}}^2 = (\mathbf{W}\mathbf{v} - \mathbf{p})^T \mathbf{R}(\mathbf{W}\mathbf{v} - \mathbf{p})$.

For the case $\tilde{\mathbf{v}}^{(0)} = \mathbf{0}$, the SIRT algorithm is a linear algorithm in the sense that a reconstructed image $\tilde{\mathbf{v}} \in \mathbb{R}^n$ is formed by applying a linear transformation to the input vector $\mathbf{p} \in \mathbb{R}^m$ of projection data. Let $\mathcal{S} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denote the linear operator that corresponds with a certain, fixed number of iterations of the SIRT algorithm, starting with $\tilde{\mathbf{v}}^{(0)} = \mathbf{0}$:

$$\tilde{\mathbf{v}} = \mathcal{S}\mathbf{p}. \quad (3.3)$$

The SIRT algorithm can also be performed on a subset $A \subset \{1, \dots, n\}$ of the image pixels by removing the columns of \mathbf{W} that are not in A . In this way, a reconstruction in the set of all reconstruction images that are zero outside of A , can be computed for which the projection difference is minimal. Consider the case where $A = \{1, \dots, n\} \setminus \{j\}$. The reconstruction equation then becomes:

$$\begin{pmatrix} \vdots & & \vdots & \vdots & & \vdots \\ w_{i,1} & \dots & w_{i,j-1} & w_{i,j+1} & \dots & w_{i,n} \\ \vdots & & \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_{j-1} \\ v_{j+1} \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix}. \quad (3.4)$$

Let $\mathcal{S}_A : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denote the linear operator that corresponds with a certain number of iterations of the SIRT algorithm, restricted to a set of pixels $A \subset \{1, \dots, n\}$, starting with $\tilde{\mathbf{v}}^{(0)} = \mathbf{0}$:

$$\tilde{\mathbf{v}} = \mathcal{S}_A \mathbf{p}. \quad (3.5)$$

3.2.2 Sinogram inconsistency

Not all vectors $\mathbf{p} \in \mathbb{R}^m$ are valid sinograms. The set of all valid continuous sinograms has been characterized by Ludwig and Helgason in [15, 16]. They describe a set of conditions that must be satisfied by all sinograms, known as *consistency conditions*. In a discretized setting, where projection data is available only for a limited set of angles, a measured sinogram is called *consistent* if there exists a solution to Eqn. 3.1, i.e. $\mathbf{p} \in \operatorname{span}\{\mathbf{w}_a : 0 < a \leq n\}$. In practice, a sinogram is rarely

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

consistent due to noise, discretization, partial volume effects, etc. Therefore, also the *inconsistency* of a sinogram \mathbf{p} is introduced:

$$\min_{\mathbf{v} \in \mathbb{R}^n} \|\mathbf{W}\mathbf{v} - \mathbf{p}\|, \quad (3.6)$$

with $\|\cdot\|$ a vector norm defined by the reconstruction algorithm. In other words, the inconsistency of \mathbf{p} is the distance between \mathbf{p} and the nearest consistent sinogram.

An important role in the proposed algorithm is played by sets of sinograms that correspond to images where certain pixels are known to be zero. A sinogram \mathbf{p} is called *A-consistent* if $\mathbf{p} \in \text{span}\{\mathbf{w}_a : a \in A\}$. Thus, for each *A*-consistent sinogram \mathbf{p} there exists a reconstructed image $\tilde{\mathbf{v}}$ with $\mathbf{W}\tilde{\mathbf{v}} = \mathbf{p}$ and $\tilde{v}_j = 0$ for each $j \notin A$. Line in Eqn. 3.6, the distance between \mathbf{p} and the nearest *A*-consistent vector is called the *A-inconsistency* of \mathbf{p} . Note that when $A = \{1, \dots, n\}$, the concepts of inconsistency and *A*-inconsistency are equivalent.

In practice, it is not possible to compute the *A*-inconsistency directly. To approximate the *A*-inconsistency of a vector \mathbf{p} with respect to the norm $\|\cdot\|_{\mathbb{R}}^2$, one can compute a SIRT reconstruction restricted to *A* from \mathbf{p} with a fixed number of iterations, compute the forward projection of this reconstruction and compare it to the vector \mathbf{p} of measured projections. Define the *A-pseudo-inconsistency* of \mathbf{p} by:

$$IC_A(\mathbf{p}) = \|\mathbf{W}\mathcal{S}_A\mathbf{p} - \mathbf{p}\|_{\mathbb{R}}^2. \quad (3.7)$$

3.3 Computational approach

In this section, a computational approach to the dense object segmentation problem is described. Section 3.3.1 introduces the *Segmentation Inconsistency (SIC)* measure which can be used to assess the quality of a thresholded segmentation of the dense objects in the reconstructed image. This metric is explained first by arguing that if a threshold is chosen too low, the segmentation inconsistency is high, and then by experimentally showing that if the threshold is chosen too high, the measured segmentation inconsistency also increases. Note that this is possible only if the grey level of the dense object is known a priori. In Section 3.3.2, an automatic grey level estimation method is discussed. Subsequently, the *Segmentation Inconsistency Minimization (SICM)* algorithm is presented in Section 3.3.3.

3.3.1 Segmentation inconsistency

Although measured sinograms are typically polluted by noise and other errors, here it is assumed, for the sake of clarity, that a “perfect” sinogram \mathbf{p} has been measured of an unknown image \mathbf{v} , i.e. $\mathbf{W}\mathbf{v} = \mathbf{p}$. All that is known of \mathbf{v} is that it contains one or more dense objects with a constant, maximal grey level $\rho \in \mathbb{R}$

3.3. COMPUTATIONAL APPROACH

and thus satisfies the prerequisites of the suggested algorithm. Let $\tilde{\mathbf{v}}$ be an image such that $\mathbf{W}\tilde{\mathbf{v}} = \mathbf{p}$ is approximately satisfied. This image may have been obtained with any reconstruction algorithm (e.g. FBP or SIRT).

Put $B = \{b : v_b = \rho\}$, the set of pixels belonging to the dense object, i.e. the set of pixels to be found by the algorithm. One seeks to approximate B by applying a threshold operation to \mathbf{v} . Let $\tilde{\tau}$ be a candidate threshold. Define $\tilde{B} = \{b : \tilde{v}_b \geq \tilde{\tau}\}$, the estimated set of pixels of the homogenous, dense object. Define $\tilde{A} = \{a : \tilde{v}_a < \tilde{\tau}\}$, complementary to \tilde{B} , the estimated set of pixels of the background. Let $\tilde{\mathbf{s}} = (\tilde{s}_j) \in \mathbb{R}^n$ denote the *dense object segmentation* based on the estimated segmentation \tilde{B} :

$$\tilde{s}_j = \begin{cases} 0, & \text{if } j \in \tilde{A} \\ \rho, & \text{if } j \in \tilde{B} \end{cases}, \quad j = 1, \dots, n. \quad (3.8)$$

Based on $\tilde{\tau}$, \mathbf{p} can be divided into two parts: a part that belongs to the dense object: $\mathbf{W}\tilde{\mathbf{s}}$, and a part that belongs to the background of image \mathbf{v} . The second part is called the *residual sinogram* of the region \tilde{A} and is given by:

$$\mathbf{p}_{\tilde{A}} := \mathbf{p} - \mathbf{W}\tilde{\mathbf{s}}. \quad (3.9)$$

The *Segmentation Inconsistency* of any segmented image $\tilde{\mathbf{s}}$ with only two grey levels (0 for background pixels and ρ for foreground pixels), is defined as:

$$SIC(\tilde{\mathbf{s}}) = IC_{\tilde{A}}(\mathbf{p}_{\tilde{A}}) := \|\mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p}_{\tilde{A}} - \mathbf{p}_{\tilde{A}}\|_{\mathbf{R}}^2. \quad (3.10)$$

In the following, it is argued and demonstrated that $SIC(\tilde{\mathbf{s}})$ is a useful measure for determining the quality of a segmentation $\tilde{\mathbf{s}}$.

- Suppose that the reconstructed set \tilde{B} is an *overestimation* of B , i.e. $B \subset \tilde{B}$. This typically occurs if $\tilde{\tau}$ is chosen too low. Put $\tilde{\mathbf{e}} := \tilde{\mathbf{s}} - \mathbf{v}_{\tilde{B}}$, a vector that contains the exact overestimation of the dense object. This vector is non-zero only for pixels in $\tilde{B} \setminus B$, and is strictly positive in this region because $\forall j \in \tilde{B} \setminus B : v_j < \rho$. The residual sinogram can then also be computed as follows:

$$\mathbf{p}_{\tilde{A}} := \mathbf{W}\mathbf{v}_{\tilde{A}} - \mathbf{W}\tilde{\mathbf{e}}. \quad (3.11)$$

As the set of \tilde{A} -consistent sinograms is a linear subspace of the set of all sinograms, $\mathbf{p}_{\tilde{A}}$ is \tilde{A} -consistent if and only if $\mathbf{W}\tilde{\mathbf{e}}$ is \tilde{A} -consistent, i.e. there exists an image $\tilde{\mathbf{x}} \in \mathbb{R}^n$ such that $\mathbf{W}\tilde{\mathbf{x}} = \mathbf{W}\tilde{\mathbf{e}}$ and $x_j = 0$ for all $j \notin \tilde{A}$.

For any image $\bar{\mathbf{v}} \in \mathbb{R}^n$ and any set $\bar{A} \subset \{1, \dots, n\}$, let $\bar{\mathbf{v}}_{\bar{A}}$ be an image that is equal to $\bar{\mathbf{v}}$ for all pixels $j \in \bar{A}$ and 0 for all pixels $j \notin \bar{A}$.

If both the number of projections and the region $\tilde{B} \setminus B$ are small, it may

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

occur that such an image \tilde{x} exists, resulting in two different segmentations for which the residual sinogram is consistent. However, as the number of projections increases and the difference between \tilde{B} and B becomes larger, it becomes highly unlikely that the residual sinogram is still consistent. In fact, the \tilde{A} -inconsistency typically increases as \tilde{B} is made larger. This is demonstrated in Section 3.4.

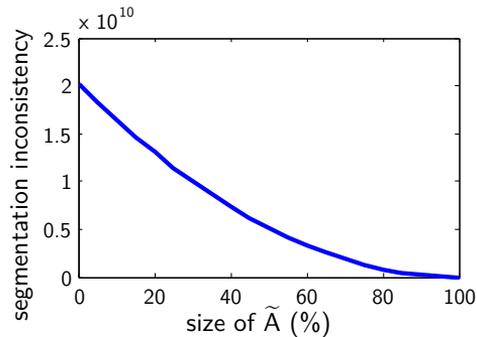


Figure 3.2: The relation between the number of fixed pixels and the SIC after 10 iterations, for a Shepp-Logan phantom of size 256×256 and 100 equiangular parallel beam projections.

- Now suppose that the reconstructed set \tilde{B} is an *underestimation* of B , i.e. $\tilde{B} \subset B$. This typically occurs if $\tilde{\tau}$ is chosen too high. In that case, the residual sinogram $p_{\tilde{A}}$ will be \tilde{A} -consistent. Here the experimental convergence properties of the SIRT algorithm can be used effectively. As the segmented dense object becomes smaller (i.e. the threshold $\tilde{\tau}$ is increased), the background \tilde{A} becomes larger, resulting in slower convergence for the iterative SIRT algorithm that is applied to those pixels. If the reconstruction is terminated after a fixed number of iterations, the computed $SIC(\tilde{s})$ will therefore generally increase along with the threshold $\tilde{\tau}$. Fig. 3.2 shows an experimental confirmation of this algorithm property. For a Shepp-Logan phantom image of 256×256 pixels using 100 projections, an increasingly large random subset of pixels was kept fixed at their true values, while computing the $SIC(\tilde{s})$ for the remaining pixels. A strictly decreasing relation between the size of \tilde{A} and the computed segmentation inconsistency can be observed.
- As a result of reconstruction errors and artefacts, there may be a threshold interval where neither $B \subset \tilde{B}$ nor $\tilde{B} \subset B$. In that case, there is an increase of segmentation inconsistency due to the false positive pixels, the effect proven in the first item of this list. However, the effect explained in the second item of this list will be somewhat cancelled out due to a mixture of false negatives and false positives, where the size of the segmented region is about

the same as the original dense object. Experimental results in Section 3.4 suggest that the segmentation inconsistency measure can still be used as an effective measure for the segmentation quality within this range as well.

In summary, to determine the optimal dense segmentation threshold, two properties of the pseudo-inconsistency measure are exploited. If the size of the dense object is overestimated, the residual sinogram is typically \tilde{A} -inconsistent, which is detected by the pseudo-inconsistency measure. If, on the other hand, the dense object is underestimated, the residual sinogram is always \tilde{A} -consistent, yet convergence properties of the SIRT algorithm favours a larger segmented region. It can therefore be expected that the SIC is minimal when \tilde{B} and B are equal.

3.3.2 Grey level estimation

The SIC concept of the previous subsection is based on the assumption that the grey level ρ of the dense object is known. In practice, this assumption is generally not valid. In this section, the SIC concept is extended by including the estimation of ρ .

Define $\tilde{\mathbf{t}} = (\tilde{t}_j) \in \mathbb{R}^n$, the binary image corresponding to the segmentation \tilde{B} :

$$\tilde{t}_j = \begin{cases} 0, & \text{if } j \in \tilde{A} \\ 1, & \text{if } j \in \tilde{B} \end{cases}. \quad (3.12)$$

Note that, $\mathbf{W}\tilde{\mathbf{s}} = \rho\mathbf{W}\tilde{\mathbf{t}}$. Let $\tilde{\rho}$ be a candidate grey level used to compute the residual sinogram. Assume that $\tilde{B} \subset B$. The residual sinogram is then given by

$$\mathbf{p}_{\tilde{A}} = \mathbf{p} - \tilde{\rho}\mathbf{W}\tilde{\mathbf{t}} = \mathbf{p} - \rho\mathbf{W}\tilde{\mathbf{t}} - (\tilde{\rho} - \rho)\mathbf{W}\tilde{\mathbf{t}} = \mathbf{W}\mathbf{v}_{\tilde{A}} - (\tilde{\rho} - \rho)\mathbf{W}\tilde{\mathbf{t}}. \quad (3.13)$$

In Eqn. 3.13, $\mathbf{W}\mathbf{v}_{\tilde{A}}$ is \tilde{A} -consistent by definition. Therefore, for the residual sinogram to be \tilde{A} -consistent, $\mathbf{W}\tilde{\mathbf{t}}$ must be \tilde{A} -consistent. That is, there must exist an image $\tilde{\mathbf{x}} \in \mathbb{R}^n$ such that $\mathbf{W}\tilde{\mathbf{x}} = \mathbf{W}\tilde{\mathbf{t}}$ and $x_j = 0$ for all $j \in \tilde{B}$. Again, if the number of projections and the region \tilde{B} are both small, it may occur that such an image $\tilde{\mathbf{x}}$ exists, resulting in two different grey levels for which the residual sinogram is consistent. However, as the number of projections is increased and the region \tilde{B} becomes larger, it becomes highly unlikely that the residual sinogram is consistent. This means that if $\tilde{B} \subset B$, the SIC concept defined in the previous subsection can also be used to estimate the grey level ρ , by computing the SIC over all grey levels $\tilde{\rho}$, and choosing the grey level for which it is minimal. If $\tilde{B} \not\subset B$, characterizing the SIC as above becomes more complex. Still, also in this case, the SIC measure can be employed for threshold selection with unknown ρ , as demonstrated in the experimental results of Section 3.4.

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

The grey level ρ for which the SIC is minimal can be computed efficiently by exploiting the linearity of the SIRT algorithm:

$$\begin{aligned}
 SIC(\tilde{\rho}\tilde{\mathbf{t}}) &= \|\mathbf{W}\mathcal{S}_{\tilde{A}}(\mathbf{p} - \tilde{\rho}\mathbf{W}\tilde{\mathbf{t}}) - (\mathbf{p} - \tilde{\rho}\mathbf{W}\tilde{\mathbf{t}})\|_{\mathbf{R}}^2. \\
 &= \|(\mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p} - \mathbf{p}) - \tilde{\rho}(\mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\tilde{\mathbf{t}})\|_{\mathbf{R}}^2 \\
 &= (\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}})\mathbf{R}(\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}})^T \tilde{\rho}^2 \\
 &\quad + (\mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p} - \mathbf{p})\mathbf{R}(\mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p} - \mathbf{p})^T \\
 &\quad + 2(\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}})\mathbf{R}(\mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p} - \mathbf{p})^T \tilde{\rho}.
 \end{aligned} \tag{3.14}$$

The optimal value $\tilde{\rho}_{opt}$ is found where the first derivative of Eqn. 3.14 vanishes. As $SIC(\tilde{\rho}\tilde{\mathbf{t}})$ is by definition non-negative, the minimum of this quadratic polynomial in $\tilde{\rho}$ can be found where the derivative is zero:

$$\tilde{\rho}_{opt} = \frac{(\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}})^T (\mathbf{p} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p})}{(\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}})^T (\mathbf{W}\tilde{\mathbf{t}} - \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}})}. \tag{3.15}$$

In Fig. 3.3, a pseudo code description of a single SIC evaluation is given.

Input: reconstructed image $\tilde{\mathbf{v}} = \mathcal{S}\mathbf{p}$ such that $\mathbf{W}\tilde{\mathbf{v}} \approx \mathbf{p}$

```

 $\tilde{A} = \{a : \tilde{v}_a < \tilde{\tau}\};$ 
for  $j = 1 : n$ 
  if  $j \in \tilde{A}$ 
     $\tilde{t}_j = 0$  ;
  else
     $\tilde{t}_j = 1$ ;
  end
end

 $\mathbf{p}_{\tilde{t}} = \mathbf{W}\tilde{\mathbf{t}};$ 
 $\mathbf{p}_{\tilde{A}} = \mathbf{p} - \mathbf{p}_{\tilde{t}};$ 
 $\mathbf{q} = \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p}, \quad \mathbf{r} = \mathbf{W}\mathcal{S}_{\tilde{A}}\mathbf{p}_{\tilde{t}};$ 
 $\tilde{\rho} = \frac{(\mathbf{p}_{\tilde{t}} - \mathbf{r})(\mathbf{p} - \mathbf{q})^T}{(\mathbf{p}_{\tilde{t}} - \mathbf{r})(\mathbf{p}_{\tilde{t}} - \mathbf{r})^T};$ 
 $\mathbf{p}'_{\tilde{A}} = \mathbf{p}_{\tilde{A}} - \tilde{\rho}\mathbf{r};$ 
 $\tilde{\mathbf{s}} = \tilde{\rho}\tilde{\mathbf{t}};$ 
 $SIC(\tilde{\mathbf{s}}) = \|\mathbf{p}_{\tilde{A}} - \mathbf{p}'_{\tilde{A}}\|_{\mathbf{R}}^2;$ 

```

Figure 3.3: Pseudo code for the computation of a single SIC measure.

3.3.3 The SICM algorithm

The *Segmentation Inconsistency Minimization (SICM)* algorithm combines grey level estimation with segmentation inconsistency computation. It uses the segmentation inconsistency found after a fixed number of SIRT iterations as a quantitative measure for the quality of the selected threshold. To find the threshold τ_{opt} (the one where $SIC(\tilde{s})$ is minimal), the Nelder-Mead simplex search method [17], an unconstrained derivative-free optimization method, is used.

Fig. 3.5 shows a flowchart of the calculation of the segmentation inconsistency for a certain threshold $\tilde{\tau}$. It should be noted that for each threshold evaluation three SIRT reconstructions are required. Two for the calculation of $\tilde{\rho}$: $\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}}$ and $\mathcal{S}_{\tilde{A}}\mathbf{p}$, and one for the calculation of $SIC(\tilde{s})$: $\mathcal{S}_{\tilde{A}}\mathbf{p}_{\tilde{A}}$. However, because $\mathbf{p}_{\tilde{A}} = \mathbf{p} - \tilde{\rho}\mathbf{W}\tilde{\mathbf{t}}$, one can easily compute $\mathcal{S}_{\tilde{A}}\mathbf{p}_{\tilde{A}}$ by subtracting $\tilde{\rho}\mathcal{S}_{\tilde{A}}\mathbf{W}\tilde{\mathbf{t}}$ from $\mathcal{S}_{\tilde{A}}\mathbf{p}$, thereby reducing the number of SIRT reconstructions per threshold evaluation to two.

3.4 Experiments

3.4.1 Simulation studies

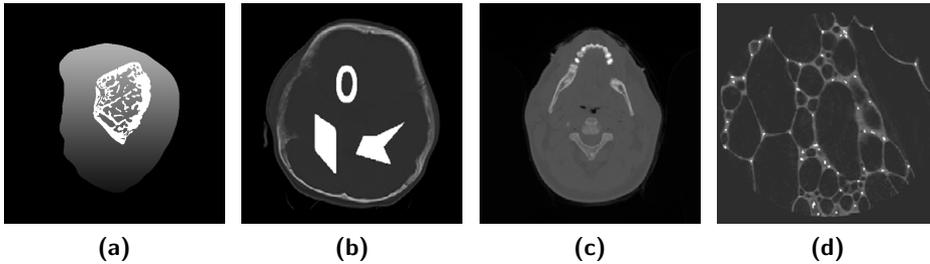


Figure 3.4: Simulated phantom images of size 512×512 . (a) Femur of a rat in a surrounding of various densities (b) Slice of human head with differently shaped objects (c) Slice of human head with a few small dental fillings (d) Foam object with metal marker particles that are often used for image registration.

Simulation experiments were performed based on four simulated phantom images of size 512×512 (Fig. 3.4). All phantom images contain at least one area with a constant, maximal grey level and a background with a continuous set of lower grey levels. Fig. 3.4a represents a slice of a rat femur in a gradient-filled surrounding object. The femur contains large as well as small trabecular structures. Fig. 3.4b represents a slice of a human head filled with 3 differently shaped large objects representing implants. Fig. 3.4c represents a different slice of a human

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

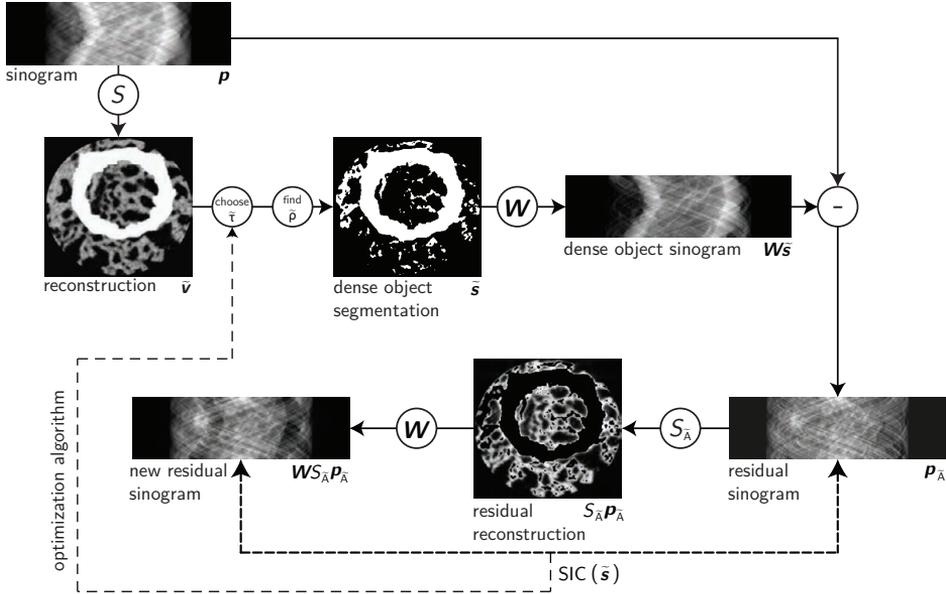


Figure 3.5: Schematic overview of the Segmentation Inconsistency Minimization (SICM) algorithm. In the shown example, the threshold $\tilde{\tau}$ is chosen too low. It can be observed that the residual sinogram $p_{\bar{A}}$ is quite different from $WS_{\bar{A}}p_{\bar{A}}$, indicating a large segmentation inconsistency.

head. It has a few very small dense objects (dental implants). Finally, Fig. 3.4d represents a slice of a foam-like structure with a large amount of small metal particles. Each phantom image thus presents a different application and indeed a different segmentation problem (small objects versus large objects, many objects versus only a few, different type of background, ...).

First, the *SIC* and the *relative number of misclassified pixels* (rNMP) were evaluated as a function of the threshold value. The rNMP is defined as the total number of misclassified pixels divided by the total number of pixels belonging to the dense objects. To be a useful measure for threshold selection, the minimum of the *SIC* should correspond well with the minimum of the rNMP. For all phantoms of Fig. 3.4, parallel beam sinograms were simulated using 180 equally spaced projection angles in $[0, \pi)$. Grey level reconstructions were computed using 300 iterations of the SIRT algorithm described in [18]. Then, for a range of global threshold values, segmentations were created and the rNMP and *SIC* values were computed. For the SIRT reconstructions of the *SIC* measurements, 300 iterations were used. Also computed were the rNMP with the Otsu's clustering method [12], k-means clustering [19] and Expectation Maximization (EM) [20] segmentation

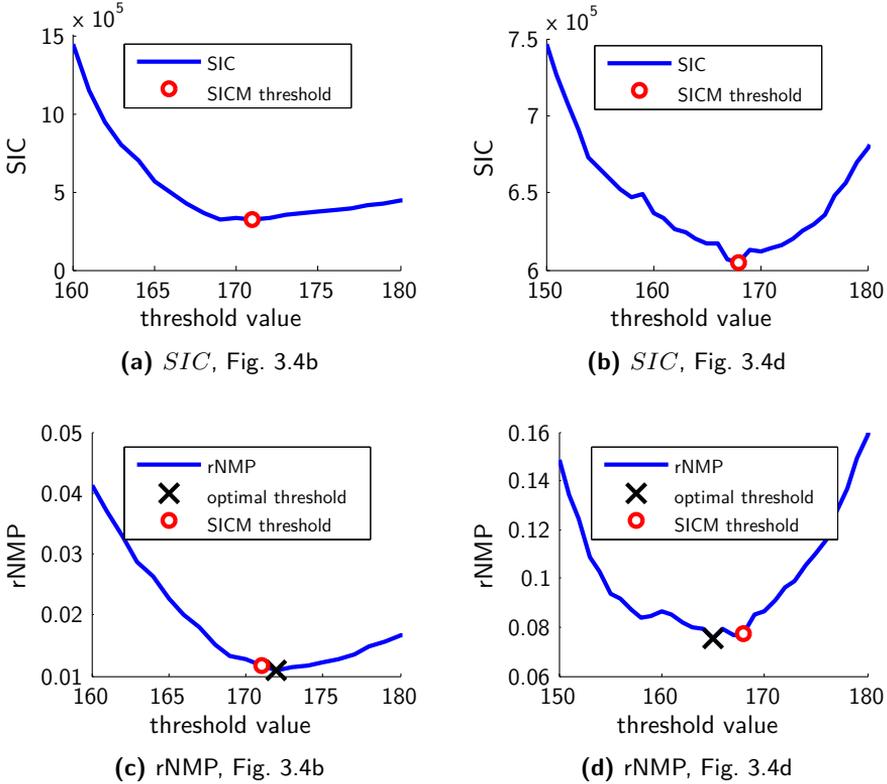


Figure 3.6: (a,b) *SIC*-scores as a function of the threshold values (c,d) *rNMP* as a function of the threshold values.

phantom	threshold value		rNMP				
	optimal	SICM	optimal	SICM	Otsu	k-means	EM
Fig. 3.4a	190.38	195.22	0.041	0.046	0.069	0.069	0.072
Fig. 3.4b	172.31	171.80	0.010	0.012	0.016	0.018	0.174
Fig. 3.4c	208.72	205.30	0.102	0.123	0.898	0.112	0.107
Fig. 3.4d	164.80	167.40	0.076	0.076	0.988	0.091	0.224

Figure 3.7: Numerical results for the first simulation experiment.

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

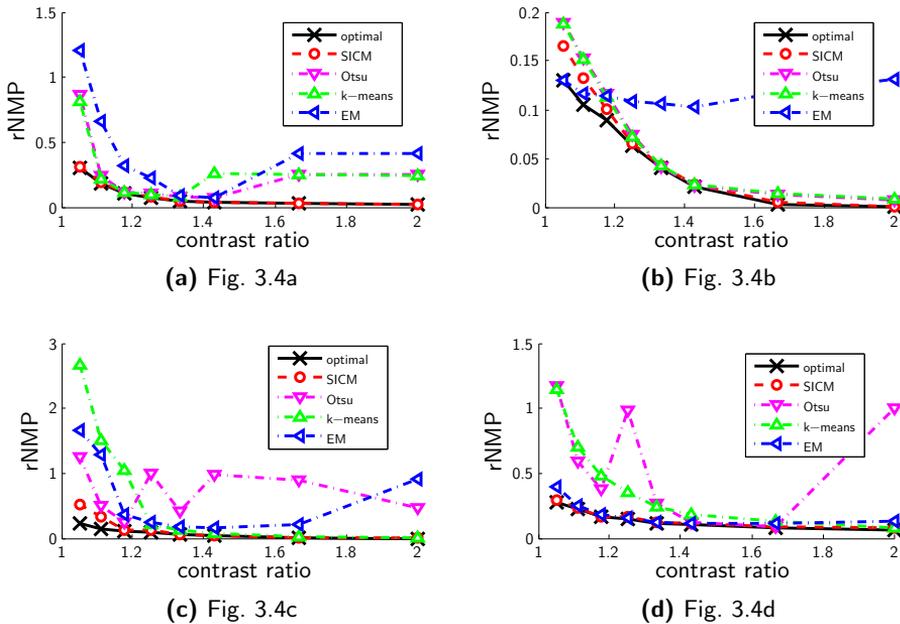


Figure 3.8: The rNMP as a function of the contrast between the background and the dense object in the original image phantom.

methods. All these methods require that the user specifies the number of segmentation partitions. This additional prior knowledge is not always available and depends on the scanned object. In these experiments, the number of partitions was different for each phantom image and was chosen such that the segmentation methods from the literature, in general, generated the lowest rNMP. 6 partitions were used for phantom image Fig. 3.4a; 4 for Fig. 3.4b; 8 for Fig. 3.4c and 7 for Fig. 3.4d.

Fig. 3.6 shows the *SIC*-score and rNMP graphs for phantom images Fig. 3.4b and Fig. 3.4d. It can be observed that the distances from the minima of the *SIC*-curves to the minima of the rNMP-curves (the squares and circles in Fig. 3.6, respectively) were very small. In Fig. 3.7, the results are shown numerically. The thresholds suggested by the SICM method were accurate approximations of the optimal thresholds, i.e. the thresholds for which the rNMP is minimal (found with an exhaustive search of the search space)¹. Furthermore, the accuracy of the SICM segmentation was close to the optimal scenario. For phantoms Fig. 3.4a,b,d, the

¹When using any global threshold technique, a lower number is not possible.

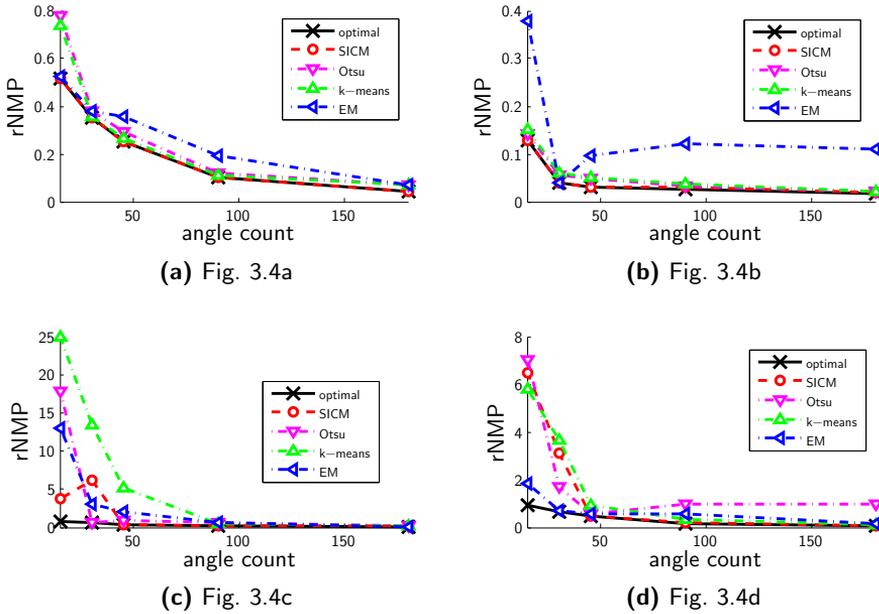


Figure 3.9: The rNMP as a function of the number of projection angles.

other segmentation methods resulted in a substantially higher rNMP. For Fig. 3.4c the results of SICM were comparable to the best performing alternatives.

In the following three series of experiments the effect of various tomographic conditions on the segmentation performance of the SICM algorithm is investigated. The resulting rNMP was compared with that of the optimal threshold and those of the other previously used segmentation methods.

- Firstly, the **contrast** between the maximal density of the background and the density of the continuous dense object was varied. This contrast is defined as $\frac{\rho}{\max_{j \in A} v_j}$. As this contrast decreases, the grey level distribution of the dense object in the tomographic reconstruction overlaps more and more with that of the background. One can therefore expect that the accuracy of the segmentations will also decrease.

Fig. 3.8 shows the rNMP as a function of the phantom contrast for the optimal global threshold, the SICM thresholding method and the other clustering algorithms. In general, the SICM curve accurately approximated the curve of the optimal threshold, whereas the other methods were much less stable and often did not provide a decent estimation.

CHAPTER 3. OPTIMAL THRESHOLD SELECTION FOR SEGMENTATION OF DENSE, HOMOGENEOUS OBJECTS

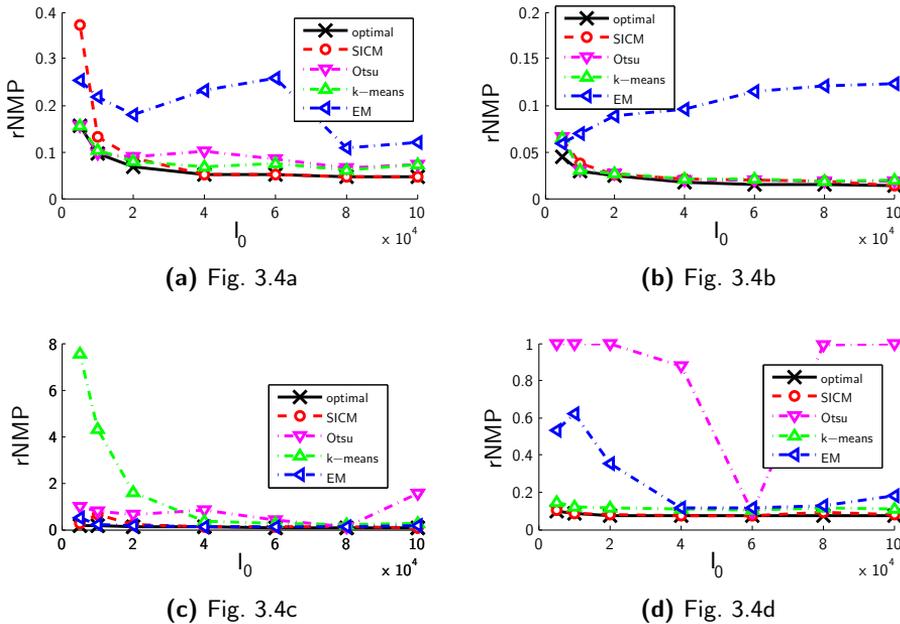


Figure 3.10: The rNMP as a function of the incident beam intensity, I_0 .

- Secondly, the **number of projection angles** was reduced. In practice, such a reduction decreases both scan time and radiation dose. However, it is expected that for a low angle count the results are unreliable as the system will become more and more underdetermined. The results, shown in Fig. 3.9, indeed confirm that the rNMP increased drastically as the number of projection angles was decreased beyond a certain minimum number — a number that is specific for each phantom. However, the SICM method still lead to significantly more accurate segmentations than other methods.
- Finally, low radiation dose scans were simulated by applying **Poisson noise with a varying source intensity** to the sinograms. The intensity of the noise is related to the incident beam intensity, I_0 . The results visible in Fig. 3.10 indicate that the presence of noise on the projection images had little effect on the ability of the SICM algorithm to estimate accurate global threshold values.

It should be noted that in all reported experiments, both on simulated and on experimental data, 300 iterations were computed for each SIRT reconstruction that is part of the *SIC* computations. This number is based on empirical findings.

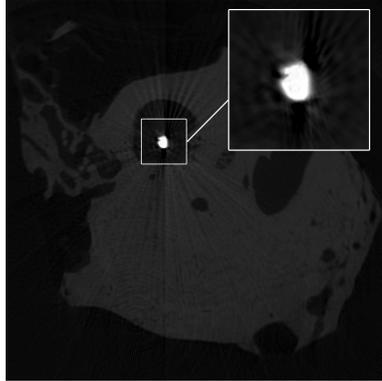


Figure 3.11: SIRT reconstruction of a cochlear implant in surrounding tissue.

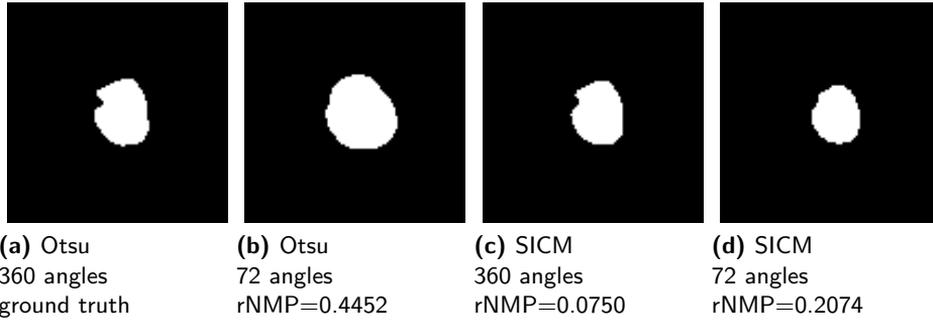


Figure 3.12: Dense object segmentations for the cochlear implant experiment.

If the number of iterations is too low, the pseudo-inconsistency (computed by a fixed number of SIRT iterations) does not match well with the true inconsistency. On the other hand, if the number of iterations is too high, one can no longer make use of the SIRT reconstruction technique to determine if $\tilde{B} \subset B$.

3.4.2 Experimental studies

The proposed algorithm was also applied on experimental μ CT data. Fig. 3.11 shows a reconstructed image of a cochlear implant in surrounding tissue, acquired with a SkyScan 1076 μ CT scanner using 360 projection angles at a detector resolution of $12\mu\text{m}$. The standard SkyScan NRecon software package was used to correct for ring- and beam hardening artefacts.

The goal is to accurately locate the small cochlear implant. However, validating the quality of these segmentations is difficult. Notice that in Fig. 3.12 the

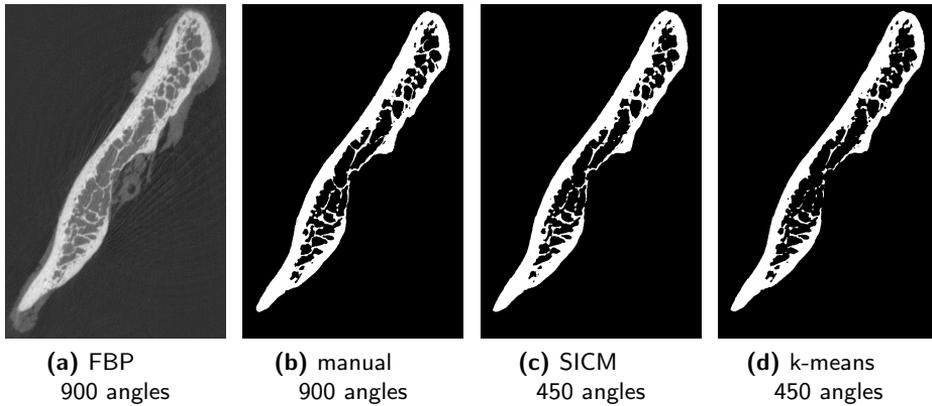


Figure 3.13: Reconstructions of the experimental mandible dataset.

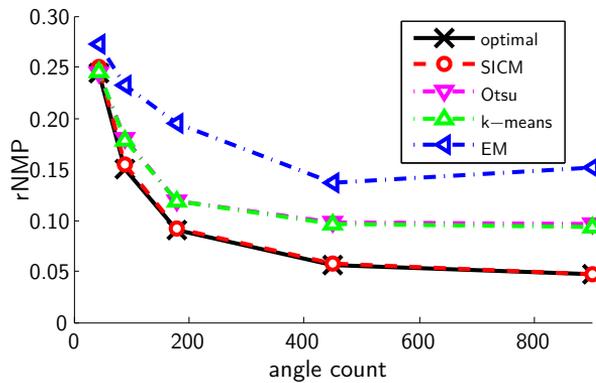


Figure 3.14: The rNMP as a function of the angle count for the mandible experiment.

segmentations provided by Otsu’s method and SICM with 360 projection angles are very similar, indicating that for this high number of angles the segmentation result does not depend strongly on the particular segmentation method. The Otsu’s segmentation with 360 projection angles was therefore considered as the ground truth image to which the SICM and Otsu’s segmentation were compared with using only 72 projection angles. In Fig. 3.12, it is clear that if the number of projection angles was lowered, the SICM segmentations were much more accurate than common Otsu segmentations.

Fig. 3.13a shows an FBP reconstruction of a slice through a human mandible. This image was recorded using a SkyScan 1173 μ CT scanner using 900 projection

angles at a detector resolution of $50\ \mu\text{m}$. Important in this application is the accurate segmentation of the cortical and trabecular bone. The SICM algorithm and the other standard methods from the literature were again applied, comparing the rNMP with respect to a manually segmented image, shown in Fig. 3.13b. The experiments were repeated using a decreasing number of projection angles, simulating low dose scans. Fig. 3.13 clearly shows that the SICM method outperformed the other methods. Fig. 3.13c and Fig. 3.13d respectively show the SICM and k-means segmentation using 450 projection angles. The SICM segmentation is an accurate approximation of the manual segmentation and indeed of the reconstructed FBP image as almost every detail of the trabeculae was present.

3.5 Conclusions

Accurate image segmentation is a difficult problem, in particular if one needs to segment objects that are very detailed or very small with respect to the image resolution. This is especially true in tomography if the image on which the segmentation is based is polluted by reconstruction errors or artefacts. Segmentation errors generally occur at the edges of the object where a hard edge is usually blurred in the reconstruction image. Popular methods, such as global clustering methods (Otsu, k-means, EM) or local neighbourhood methods (region growing, watershed segmentation) all have different strategies to define where exactly the object edges are located, making it difficult to measure the accuracy of these segmentations in practice. Fortunately, tomography inherently provides a way to counter this problem as projection data is available that can be used to improve or optimize standard segmentation techniques.

In this chapter, a novel method for finding a global threshold to accurately locate dense objects in a continuous surrounding in a tomographic reconstruction has been presented. Contrary to existing methods, the *Segmentation Inconsistency Minimization (SICM)* method is not only based on the reconstructed image, but also on the available projection data. For an optimal segmentation, the residual sinogram (i.e. the sinogram of the part of the image that does not belong to the dense object) is consistent. The inconsistency can be measured by applying a linear iterative reconstruction technique (such as SIRT) to the pixels not belonging to the dense object, and by comparing the forward projection of this reconstruction to the residual sinogram.

Experiments were performed on both simulated and experimental data. From them, it can be concluded that the SICM algorithm generally finds a good approximation of the optimal global threshold. In a large majority of the experiments performed, the SICM method outperformed other thresholding methods.

One downside of the proposed method is its computational cost. Whereas

REFERENCES

methods that work only on the reconstruction image are typically very fast, the SICM method has an optimization routine where each function evaluation requires two SIRT reconstructions. For a full optimization of a 512×512 image with 180 projection angles, on a system with a modern 3 GHz Intel CPU, the typical computation time for a single SIC evaluation was 10 minutes, and 221 minutes for a full optimization. To drastically lower the computation time, one can resort to GPU programming using the CUDA programming language. The experiments of Section 3.4 were all performed on an NVIDIA GeForce GTX 480 GPU, on which the average computation time for a single SIC evaluation was 13 seconds, and for a full optimization up to 4.5 minutes.

References

- [1] T. M. Link, W. Berning, S. Scherf, U. Joosten, A. Joist, K. Engelke, and H. E. Daldrup-Link, "CT of metal implants : Reduction of artifacts using an extended CT scale technique," *Journal of Computer Assisted Tomography*, vol. 24, no. 1, pp. 165–172, 2000.
- [2] C. Lemmens, D. Faul, J. Hamill, S. Stroobants, and J. Nuyts, "Suppression of metal streak artifacts in CT using a MAP reconstruction procedure," in *IEEE Nuclear Science Symposium Conference Record*, vol. 6, pp. 3431–3437, October 2006.
- [3] W. Veldkamp, R. Joemai, and J. Geleijns, "Automated segmentation and interpolation in sinograms for metal artifact suppression in CT," *Medical Physics*, vol. 36, pp. 2443–2443, June 2009.
- [4] B. R. Whiting, K. T. Bae, and M. W. Skinner, "Cochlear implants: Three-dimensional localization by means of coregistration of ct and conventional radiographs," *Radiology*, vol. 221, pp. 543–549, 2001.
- [5] G. Schoonenberg, R. Florent, P. Lelong, O. Wink, D. Ruijters, J. Carroll, and B. ter Haar Romeny, "Projection-based motion compensation and reconstruction of coronary segments and cardiac implantable devices using rotational X-ray angiography," *Medical Image Analysis*, vol. 13, pp. 785–792, 2009.
- [6] Q. Yang, G. Wang, M. W. Skinner, J. T. Rubinstein, and M. W. Vannier, "Localization of cochlear implant electrodes in radiographs," *Medical Physics*, vol. 27, no. 4, pp. 775–777, 2000.
- [7] Y. An and R. A. Draughn, eds., *Mechanical testing of bone and the bone-implant interface*. CRC Press, 2000.
- [8] W. L. Roque, A. C. A. de Souza, and D. X. Barbieri, "The euler-poincaré characteristic applied to identify low bone density from vertebral tomographic images," *Revista Brasileira de Reumatologia*, vol. 49, no. 2, 1983.

-
- [9] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–168, 2004.
- [10] A. Rosenfeld and P. Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, pp. 231–235, 1983.
- [11] S. Cho, R. Haralick, and S. Yi, "Improvement of kittler and illingworth's minimum error thresholding," *Pattern Recognition*, vol. 22, no. 5, pp. 609–617, 1989.
- [12] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, March 1979.
- [13] X. Yu and J. Ylaaski, "A new algorithm for image segmentation based in region growing and edge detection," in *Proc. Int. Symp. Circuits and Systems*, vol. 1, pp. 516–519, 1991.
- [14] S. Beucher and F. Meyer, *The morphological approach to segmentation: the watershed transformation*. Mathematical Morphology in Image Processing, New York: Dekker, M., dougherty e. r. ed., 1993.
- [15] D. Ludwig, "The Radon transform on Euclidean space," *Communications on Pure and Applied Mathematics*, vol. 19, pp. 49–81, 1966.
- [16] S. Helgason, *The Radon Transform*. Boston, MA: Birkhauser, 1980.
- [17] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [18] J. Gregor and T. Benson, "computational analysis and improvement of SIRT," *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 918–924, 2008.
- [19] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, July 2002.
- [20] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1026–1038, Aug. 2002.

Part II

Discrete Tomography

—There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.

Douglas Adams, *The Restaurant at the End of the Universe*, 1980.

4

The Discrete Algebraic Reconstruction Technique (DART)

Contents

4.1	Introduction	86
4.2	Discrete Algebraic Reconstruction Technique	88
4.2.1	Notation	88
4.2.2	Concept	89
4.2.3	In-depth view	91
4.3	Experiments	94
4.3.1	Simulation studies	95
4.3.1.1	Few-view tomography	96
4.3.1.2	Limited angular range	98
4.3.1.3	Projection truncation	98
4.3.1.4	Noisy projection data	102
4.3.2	Experimental studies	102
4.4	Automatic parameter optimization	107
4.4.1	Adaptive surrogate modelling	107
4.4.2	Simulation studies	108
4.5	Conclusions	110
	References	112

4.1 Introduction

If the amount of available projection data is limited (e.g. with few-view tomography, a limited angular range or data truncation) the reconstruction problem is ill-posed and conventional algebraic reconstruction methods tend to find solutions that do not correspond to the measured object. It then makes sense to exploit prior knowledge about this object. Such prior knowledge can come in many forms.

- If only a limited number of pixels do not represent air, the image is said to be sparse with respect to the pixel basis. Research in the field of compressed sensing has shown that sparse images can be reconstructed from a small number of measurements, provided that the measurements satisfy certain randomization properties [1, 2, 3].
- Structures often have a homogeneous density and are very large with respect to the pixel size. In that case the gradient image is very sparse. Reconstruction is then referred to as a total variation minimization problem [4].
- Noise statistics can be taken into account with Expectation Maximization (EM) algorithms.
- If a scanned object contains known components such as screws or implants, their shapes can be modelled into the reconstruction process using likelihood-based methods [5, 6].
- Attenuation values are by definition always positive.
- Scanned objects — or parts of it — are usually piecewise constant and homogeneous.
- The exact attenuation factor of the materials present in the object — and thus also the exact grey level in the reconstructed image — can be known in advance.

In this chapter, an algorithm is considered that exploits the last two items of this list. Consider an object that consists of a single homogeneous material. Its reconstructed image should then ideally contain only two grey levels: one for the background and one for the object. Conventionally, this constraint is enforced posteriori to the reconstruction process (Fig. 4.1a). In Part I, segmentation techniques were discussed that searched for the global threshold values such that the resulting segmented image adhered the most to the measured data (Fig. 4.1b). While it was shown that this approach indeed resulted in more accurate images than by using conventional segmentation techniques, reconstruction artefacts (e.g. due to limited data) still defined a lower bound on the achievable accuracy.

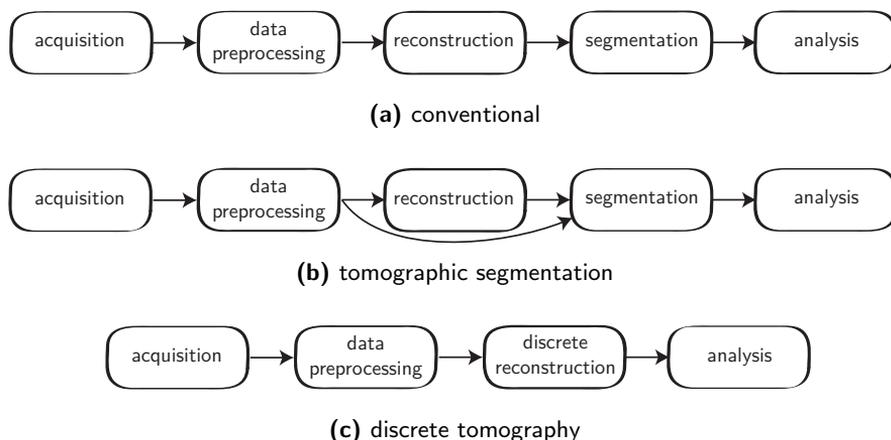


Figure 4.1: (a) Conventional workflow of a typical tomographic application. (b) Workflow with tomographic segmentation. (c) Workflow with discrete tomography.

In Section 2.4, a local thresholding scheme was discussed in which threshold values could differ in various regions of the reconstruction grid. It was also observed that, while this method indeed increased the achievable accuracy, it vastly increased the size of the search space and made the optimization problem much more difficult to solve. The use of additional prior knowledge is thus recommended. Note that, if each pixel in the reconstruction domain has its own threshold value, this local thresholding scheme does, in theory, not require an accurate reconstructed image. The reconstructed image can even be a black image. A threshold can then be set to a negative value for a pixel belonging to the foreground and to a positive value for a pixel belonging to the background. This method is then no longer just a segmentation technique, but in effect also a reconstruction technique (Fig. 4.1c). Such a solution is called *discrete tomography*. Discrete tomography concerns the accurate reconstruction of discrete-valued functions from multiple projections¹.

Originally, discrete tomography was considered a combinatorial problem in which binary matrices were to be determined by its row and column sums [7]. Recent advances have resulted in a number of discrete tomography techniques ready for use in practical applications [8]. In [9], a smoothness prior and a primal-dual subgradient algorithm are applied to provably converge to a spatially coherent binary solution that fits the projection data. In [10], a multilevel approach is proposed where Gibbs priors are used to model the distribution of the segmented

¹It should be noted that the literature supports conflicting definitions of the term “discrete tomography”. For example, it has also been used for the reconstruction of binary objects on a discrete domain.

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

reconstructions.

In [11], an iterative discrete tomography algorithm was introduced. The *Discrete Algebraic Reconstruction Technique (DART)* alternates between regular algebraic reconstruction iterations and a discretization step in which it is assumed that the grey levels corresponding with each material are known. This approach is not limited to binary images, but also works with objects that contain several grey levels². Although DART is a heuristic algorithm without guaranteed convergence properties, it has demonstrated the ability to compute highly accurate reconstructions from limited data in a range of applications, such as electron tomography [12, 13] and to reconstruct grain maps from X-ray diffraction data [14].

The chapter is structured as follows. In Section 4.2, the concept and mathematical formulation of DART is given. Section 4.3 then presents an experimental validation of DART on a variety of simulated and experimental datasets. In Section 4.4, the projection difference is proposed as a simple cost function that can be used by an optimization technique to automatically estimate the optimal parameters. Finally, Section 4.5 concludes this chapter.

4.2 Discrete Algebraic Reconstruction Technique

Here, DART is explained in detail. Section 4.2.1 introduces the required notation, Section 4.2.2 discusses the general concept of DART, and Section 4.2.3 presents the mathematical foundation of the algorithm.

4.2.1 Notation

As in the previous chapters, define $\mathbf{p} = (p_i) \in \mathbb{R}^m$, the measured projection data, $\mathbf{v} = (v_j) \in \mathbb{R}^n$, the object function, and $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{m \times n}$, the projection matrix such that:

$$\mathbf{W}\mathbf{v} = \mathbf{p}. \quad (4.1)$$

Let $\mathcal{S}^{(t)} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denote the linear operator that corresponds with t iterations of a non-specified algebraic reconstruction technique (e.g. SIRT or CGLS), starting with $\bar{\mathbf{v}}^{(0)} = \mathbf{0}$:

$$\bar{\mathbf{v}} = \mathcal{S}^{(t)}\bar{\mathbf{p}}. \quad (4.2)$$

An algebraic reconstruction algorithm can also be performed on a certain subset $A \subset \{1, \dots, n\}$ of the reconstruction grid by removing the columns of \mathbf{W} that are not in A . Consider the case where $A = \{1, \dots, n\} \setminus \{j\}$, with $j \in \{1, \dots, n\}$.

²There is no real upper limit on the number of materials that can be used in a DART reconstruction. However, the more grey levels are present, the lower the achievable accuracy will be. Depending on the complexity of the scanned structures, up to 4 or 5 grey levels are feasible.

4.2. DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE

Eqn. 4.1 then becomes:

$$\begin{pmatrix} \vdots & & \vdots & \vdots & & \vdots \\ w_{i,1} & \cdots & w_{i,j-1} & w_{i,j+1} & \cdots & w_{i,n} \\ \vdots & & \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_{j-1} \\ v_{j+1} \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix}. \quad (4.3)$$

Let $\mathcal{S}_A^{(t)} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denote the linear operator that corresponds with t iterations of a non-specified algebraic reconstruction technique, restricted to a set of pixels $A \subset \{1, \dots, n\}$, starting with $\bar{\mathbf{v}}^{(0)} = \mathbf{0}$:

$$\bar{\mathbf{v}} = \mathcal{S}_A^{(t)} \bar{\mathbf{p}}. \quad (4.4)$$

A global thresholding scheme on a certain image $\bar{\mathbf{v}} \in \mathbb{R}^n$ is defined as follows. Let l denote the number of distinct grey levels in the thresholded image. Define the *segmentation function* $\mathcal{C}_{\tau, \rho}(\bar{\mathbf{v}}) : \mathbb{R}^n \rightarrow \{\rho_1, \dots, \rho_l\}^n$ as the function that assigns a value in the vector of grey levels $\boldsymbol{\rho} = (\rho_1, \dots, \rho_l)^T$ to each pixel of the reconstructed image, according to threshold values $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{l-1})^T$:

$$\mathcal{C}_{\tau, \rho}(\bar{v}_j) = \begin{cases} \rho_1 & \bar{v}_j < \tau_1 \\ \rho_2 & \tau_1 \leq \bar{v}_j < \tau_2 \\ \vdots & \\ \rho_l & \tau_{l-1} \leq \bar{v}_j \end{cases}, \quad j = 1, \dots, n. \quad (4.5)$$

4.2.2 Concept

Consider the case presented in Fig. 4.2, where a binary 512×512 phantom image (Fig. 4.2a) is reconstructed from only 5 parallel beam projections. The reconstruction problem is then very ill-posed and multiple solutions are consistent with the projection data. In that case, the SIRT reconstruction (Fig. 4.2b) converges to the solution that is closest to the initial image [15]. A segmentation of the reconstruction is shown in Fig. 4.2c. One should note that, while this segmentation is by no means accurate, the original phantom shape is somewhat approximated. Indeed, as is visible in Fig. 4.2d, all misclassified pixels are near the object border.

This observation lies at the core of the DART algorithm. Combined with prior knowledge of the grey levels, the reconstruction problem can then be significantly reduced in size. From a certain estimation (Fig. 4.3a) the boundary pixels are determined (Fig. 4.3b). The contribution of the non-boundary pixels — those

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

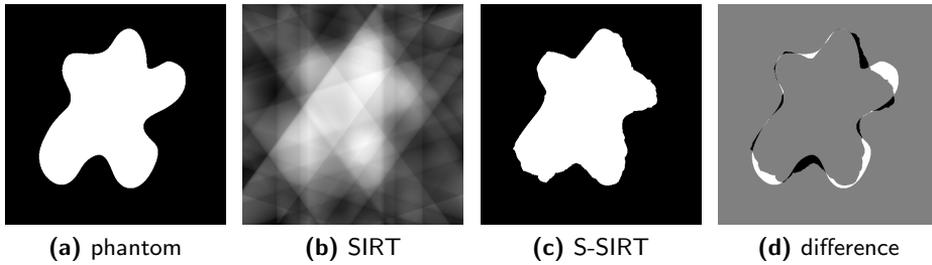


Figure 4.2: (a) A binary 512×512 phantom image. (b) Classical SIRT reconstruction from 5 equiangular parallel beam projection directions. (c) Segmentation of (b). (d) Difference between (a) and (c). Misclassified pixels mostly occur near the edges of the object.

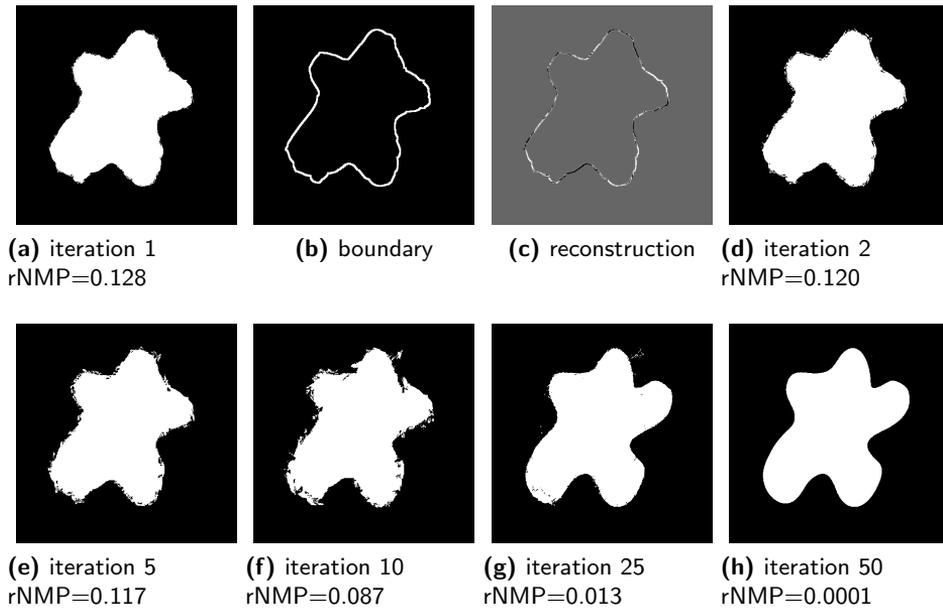


Figure 4.3: DART reconstructions of Fig. 4.2a from 5 equiangular parallel beam projections.

which are assumed to have the correct grey level — is then subtracted from the original projection data. This *residual sinogram* is reconstructed in the boundary pixels only (Fig. 4.3c), a reconstruction problem that is less ill-posed as before: the number of unknown values has been significantly reduced whereas the number of projection values has remained the same. The new reconstruction is subsequently segmented and used to update the boundary pixels of the original segmentation

4.2. DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE

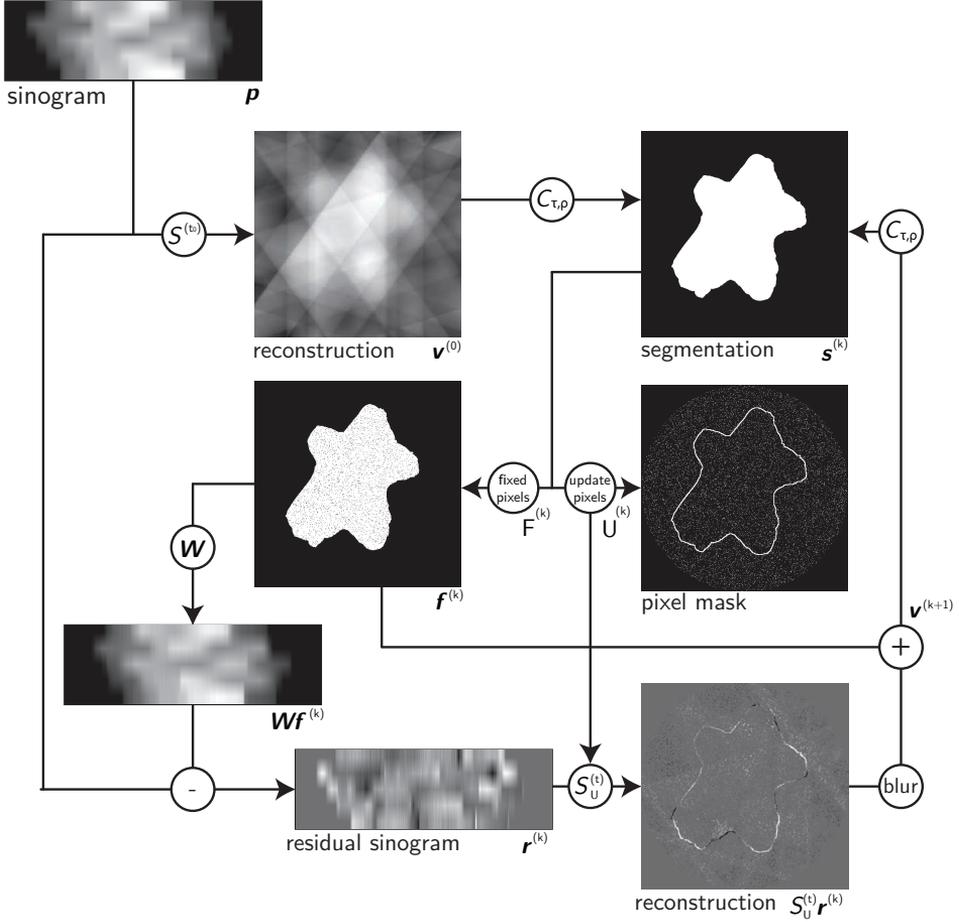


Figure 4.4: Schematic overview of the Discrete Algebraic Reconstruction Technique (DART) method.

(Fig. 4.3d). This process is then iteratively repeated (Fig. 4.3e-h). In Fig. 4.4, DART is schematically visualized and in Fig. 4.5, pseudo code is provided.

4.2.3 In-depth view

Initially, an approximate reconstruction image, $v^{(0)}$, is computed using any algebraic reconstruction technique with a certain number of iterations, t_0 :

$$v^{(0)} = S^{(t_0)} p. \quad (4.6)$$

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

Subsequently, DART follows an iterative scheme with the following steps. Put $k = 0$ the iteration number of DART.

1. The current reconstructed image, $\mathbf{v}^{(k)}$, is segmented to obtain an image that contains only grey levels from the vector $\boldsymbol{\rho} = (\rho_1, \dots, \rho_l)^T$. A simple global thresholding scheme with threshold values $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{l-1})^T$ is used. Let $\mathbf{s}^{(k)} = (s_j^{(k)}) \in \{\rho_1, \dots, \rho_l\}^n$ denote that segmented image:

$$\mathbf{s}^{(k)} = \mathcal{C}_{\boldsymbol{\tau}, \boldsymbol{\rho}}(\mathbf{v}^{(k)}). \quad (4.7)$$

More advanced segmentation methods, such as the PDM method of Chapter 2, may lead to improved accuracy. This is covered in Chapter 6.

2. The set, $U^{(k)} \subset \{1, \dots, n\}$, of *update pixels* is determined. This set contains pixels that are likely to be misclassified in the current segmented image. The pixels in the complementary set, $F^{(k)} = \{1, \dots, n\} \setminus U^{(k)}$, of *fixed pixels* are those that will be removed from the reconstruction equation. $U^{(k)}$ contains two types of pixels.

- As demonstrated in Fig. 4.2d, most errors occur near the edges of the object. All boundary pixels of the current segmented image are therefore added to $U^{(k)}$. A pixel is considered to be a boundary pixel if its value is different from at least one of its neighbours:

$$U^{(k)} = \left\{ j : \exists h \in N(j) \text{ such that } s_j^{(k)} \neq s_h^{(k)} \right\}. \quad (4.8)$$

In Eqn. 4.8, $N(j) \subset \{1, \dots, n\}$ denotes a certain neighbourhood window of pixel j . Various neighbourhood definitions can be used, e.g. 8-connectivity for 2D images. In [16], a variation of DART, called *Adaptive DART (ADART)*, was introduced. In ADART, the required number of pixels from $h \in N(j)$ for which $s_j^{(k)} \neq s_h^{(k)}$ is increased each time a certain convergence criterion is met.

- Each non-boundary pixel is added to $U^{(k)}$ with a certain probability $0 \leq r \leq 1$. This increases the accuracy of DART reconstruction in case of noisy projection data.
3. To restrict the set of unknowns in Eqn. 4.1 to those in the set of update pixels, the forward projection of the fixed pixels must first be subtracted from the measured projection data. Let $\mathbf{f}^{(k)} \in \mathbb{R}^n$ denote the current segmented

4.2. DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE

```

v =  $\mathcal{S}^{(t_0)}\mathbf{p}$ ;
while stop criterion is not met

    s =  $\mathcal{C}_{\tau,\rho}(\mathbf{v})$ ;
    U = {};
    for j = 1 : n
        if  $\exists h \in N(j) : s_j \neq s_h$  or rand() < r
            fj = 0;
            U = U  $\cup$  j;
        else
            fj = sj;
        end
    end
    r =  $\mathbf{p} - \mathbf{W}\mathbf{f}$ ;
    v' =  $\mathbf{f} + \mathcal{S}_U^{(t)}\mathbf{r}$ ;

    for j = 1 : n
        vj =  $(1-b)v'_j + (b/\bar{n})\sum_{h \in N(j)} v'_h$ ;
    end
end

```

Figure 4.5: Pseudo code for the Discrete Algebraic Reconstruction Technique (DART).

image in which all update pixels are set to zero, i.e.:

$$\mathbf{f}_j^{(k)} = \begin{cases} s_j^{(k)} & j \notin U^{(k)} \\ 0 & j \in U^{(k)}. \end{cases} \quad (4.9)$$

The *residual sinogram*, $\mathbf{r}^{(k)} \in \mathbb{R}^m$, is then given by:

$$\mathbf{r}^{(k)} = \mathbf{p} - \mathbf{W}\mathbf{f}^{(k)}. \quad (4.10)$$

4. The current reconstructed image is updated by the reconstruction of the residual sinogram, restricted to the set of update pixels:

$$\mathbf{v}^{(k+1)} = \mathbf{f}^{(k)} + \mathcal{S}_{U^{(k)}}^{(t)}\mathbf{r}^{(k)}. \quad (4.11)$$

In Eqn. 4.11, t denotes the number of iterations that are performed to reconstruct the residual sinogram. If certain pixels are misclassified in step 1, the residual sinogram is inconsistent. In that case, there does not exist

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

an exact solution to Eqn. 4.11, leading to erroneous values in the updated reconstruction. To counter this, a smoothing filter of intensity $b \in [0, 1]$ is applied to $\mathbf{v}^{(k+1)}$ as a means of regularization:

$$v_j^{(k+1)} := (1 - b)v_j^{(k+1)} + \frac{b}{\bar{n}} \sum_{h \in N(j)} v_h^{(k+1)}. \quad (4.12)$$

In Eqn. 4.12, $N(j) \subset \{1, \dots, n\}$ denotes the neighbours of pixels j . The size of N is denoted by \bar{n} .

5. Increase k by 1 and return to step 1 until some termination criterion is reached.

4.3 Experiments

In this section, a range of experiments is described that was carried out to evaluate DART. In Section 4.3.1, results are presented on simulated projection data. In Section 4.3.2, experimental datasets were used.

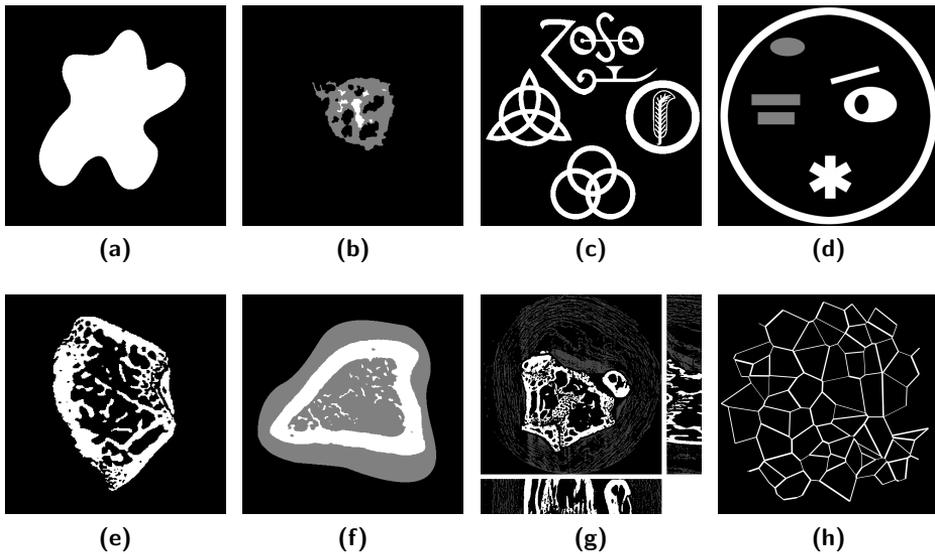


Figure 4.6: Simulated phantom images that were used to validate DART.

4.3.1 Simulation studies

In the simulation studies, eight phantom images were considered.

- (a) Fig. 4.6a: A 512×512 binary phantom image of a concave blob-like object.
- (b) Fig. 4.6b: A 512×512 phantom image representing an electron tomography experiment of a bamboo-like carbon nanotube that was formed around a Copper catalyst particle. This data is based on the experiments reported in [12].
- (c) Fig. 4.6c: A 512×512 binary phantom image of several complex shapes.
- (d) Fig. 4.6d: A phantom containing a few different shapes and three grey levels. This is an analytical phantom, i.e. its projection data was generated on 512 detector cells with the well-known projection formulas of squares and ellipsoids.
- (e) Fig. 4.6e: A 512×512 binary phantom image of a slice through a rat femur.
- (f) Fig. 4.6f: A 512×512 phantom image of a slice through a rat femur. A third grey level, representing the surrounding tissue, was added. This value was exactly half that of the bone grey level.
- (g) Fig. 4.6g: A three-dimensional $512 \times 512 \times 100$ phantom image of a rat femur. This dataset was based on an actual ex-vivo μ CT scan. It contained a third grey level, representing paper that was used to hold the bone in place. This value was one fourth of the bone grey level.
- (h) Fig. 4.6h: A phantom representing a polyurethane foam. The thickness of the cell edges varied from 2 to 5 times the detector width. Projection data (for 512 detector cells) was created analytically. That way, a partial volume effect is simulated into the projection data.

Reconstructions were computed with FBP, SIRT (500 iterations) and DART (200 iterations). The FBP and SIRT reconstructions were subsequently segmented using Otsu's clustering method [17]. These segmented images are hereafter referred to as S-FBP and S-SIRT. For DART, the parameters were set to the following values: $r = 0.1$, $b = 0.1$, $t = 10$, $t_0 = 500$, $\mathcal{S}=\text{SIRT}$. The grey levels ρ were set to the levels measured in the phantom images and the threshold values τ were chosen in the middle of each pair of consecutive grey levels.

To assess the quality of the segmented images, the *relative Number of Misclassified Pixels (rNMP)* was calculated. This is the ratio of the number of pixels that is classified into an incorrect partition to the total number of non-zero pixels in the phantom images. For the analytical phantom images Fig. 4.6d and Fig. 4.6h, the rNMP was approximated by resizing the 512×512 reconstructed images to 8192×8192 and comparing them to high resolution rasterized versions of the phantom images.

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

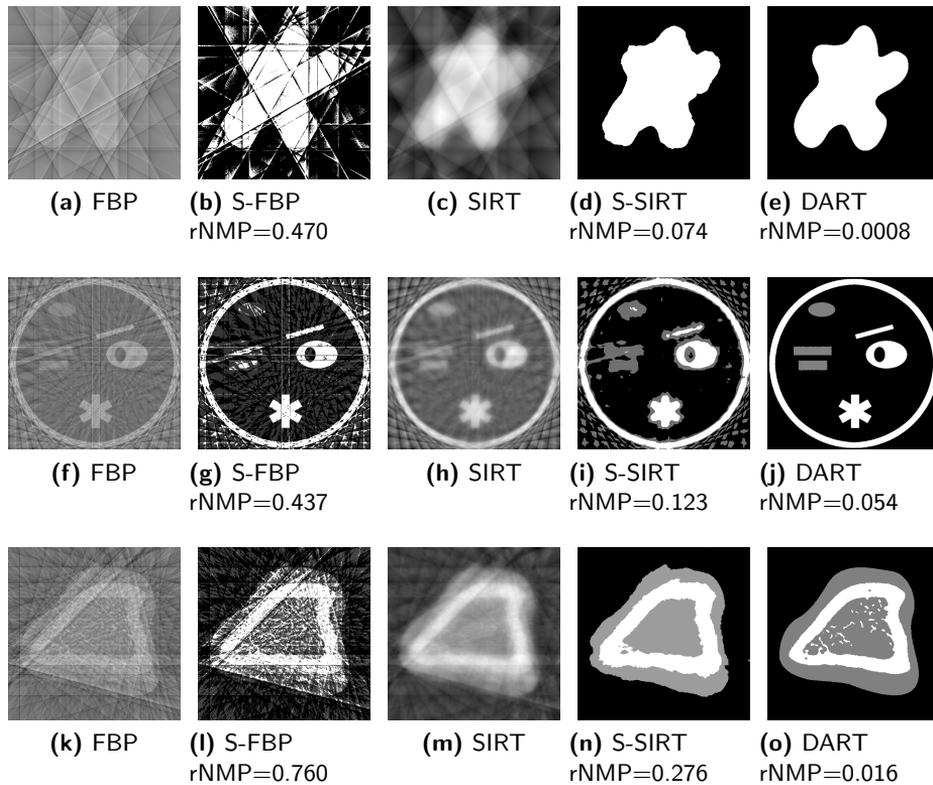


Figure 4.7: (a-e) Reconstructions of Fig. 4.6a from 6 projection angles. (f-j) Reconstructions of Fig. 4.6d from 20 projection angles. (k-o) Reconstructions of Fig. 4.6f from 12 projection angles.

4.3.1.1 Few-view tomography

In a first set of experiments, reconstructions were made of all of phantom images in Fig. 4.6 from a varying number of equiangular projection angles. Fig. 4.7 shows a selection of reconstructed and segmented images of three different phantom images with a low number of projection angles. Clearly, DART provided segmented images that were of a far higher quality than those of the conventional techniques.

This can also be seen in Fig. 4.8, where the rNMP is plotted for all phantom images as a function of the number of projection angles. If there were many projection directions, the difference between the reconstruction techniques was negligible. However, as the number of projection angles was reduced, S-SIRT generally started outperforming S-FBP and DART generally started outperforming S-SIRT. It can

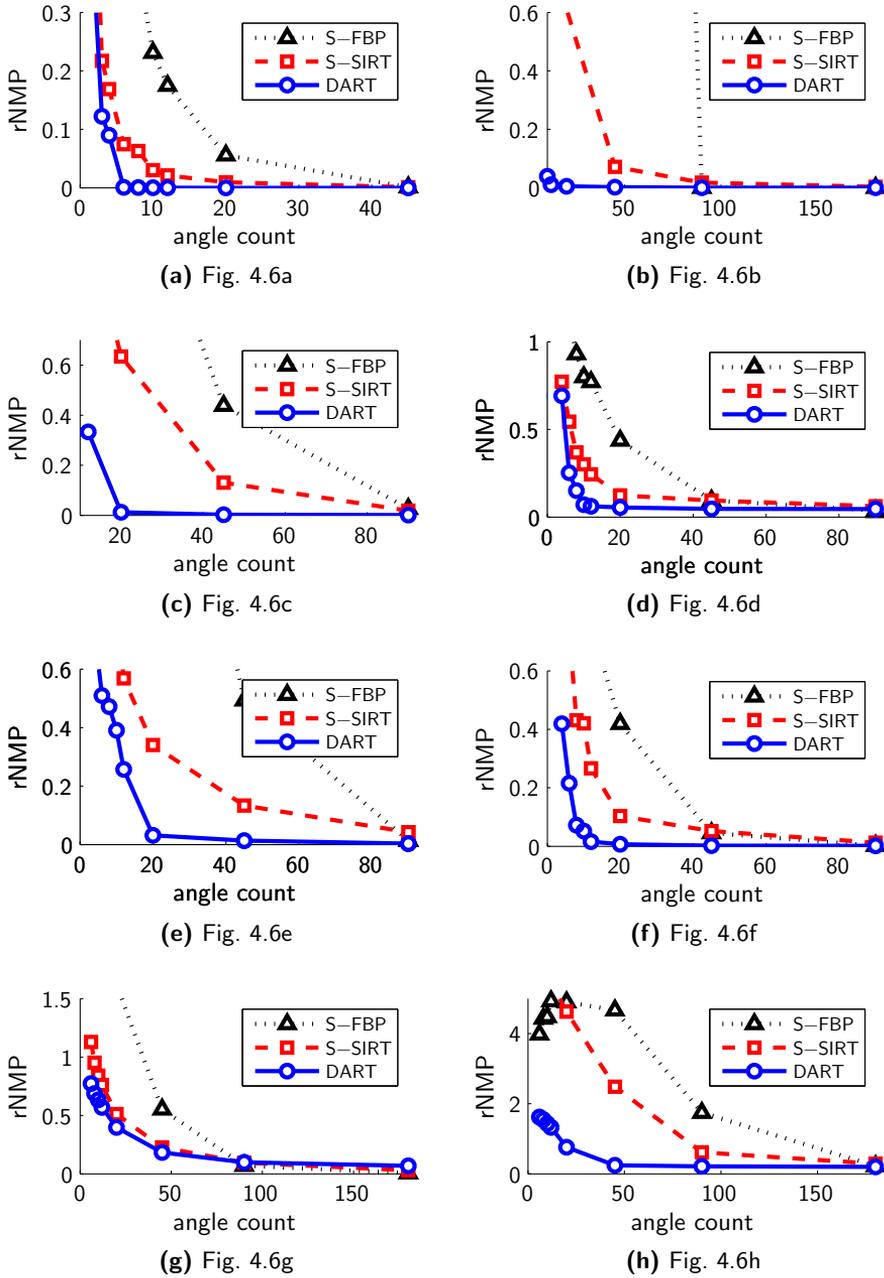


Figure 4.8: The rNMP of reconstructions of Fig. 4.6 is plotted as a function of the number of projection directions.

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

be observed that, for S-SIRT and S-FBP, the accuracy decreased gradually as the number of projections was reduced. The DART reconstructions, however, tended to remain more or less equally accurate up to a certain point where the projection data was so limited that even the exploitation of prior knowledge no longer sufficed to obtain high quality approximations. After that point, reconstruction errors increased rapidly. It should be noted that the location of this point depends on the “complexity” of the dataset. Large piecewise constant shapes such as in Fig. 4.6a required substantially fewer directions than more complex structures, such as the trabecular structures in the bone phantoms.

For phantom images Fig. 4.6g and Fig. 4.6h, the quality of the reconstructions (even for DART) became rapidly insufficient for accurate analysis. This is due to the fact that the structures in these phantoms were very small with respect to the detector and pixel size. The set of pixels to update, $U^{(k)}$, was then too large to sufficiently reduce the reconstruction problem in size. To overcome this problem, a super-resolution approach is investigated in Chapter 7.

4.3.1.2 Limited angular range

In a second set of experiments, projection data was generated with an increasingly limited angular range. This type of incomplete data often occurs in electron tomography (Section 1.5.1). In Fig. 4.9a-e, reconstructions are shown of Fig. 4.6b from 180 parallel beam projection angles with a full angular range ($[0^\circ, 180^\circ]$). Fig. 4.9f-j show reconstructions of the same image from an angular range limited to $[25^\circ, 155^\circ]$. With the conventional reconstruction techniques, the missing angular range resulted in clearly visible *missing wedge* artefacts. Only DART managed to obtain high quality reconstructions. Similar results can be observed in Fig. 4.9k-n, where for Fig. 4.6a,b,c,e, the rNMP is plotted as a function of the angular range.

4.3.1.3 Projection truncation

Incomplete projection data problems also occur when the scanned object is too large to fit in the *field of view (FOV)* of the scanner. The projection data is then said to be *truncated*. An example is given in Fig. 4.10a and Fig. 4.10b, where projection data of phantom image Fig. 4.6a is visualized with a sufficiently large detector array (512 pixels wide) and an insufficiently large detector array (192 pixels wide).

In [18], it is argued that also in these cases DART can provide an accurate solution. Fig. 4.10d-g shows reconstructed images for the projection data pictured in Fig. 4.10b. The field of view can be identified by the blue ring. Unlike the conventional techniques, DART clearly resulted in very accurate reconstructed images,

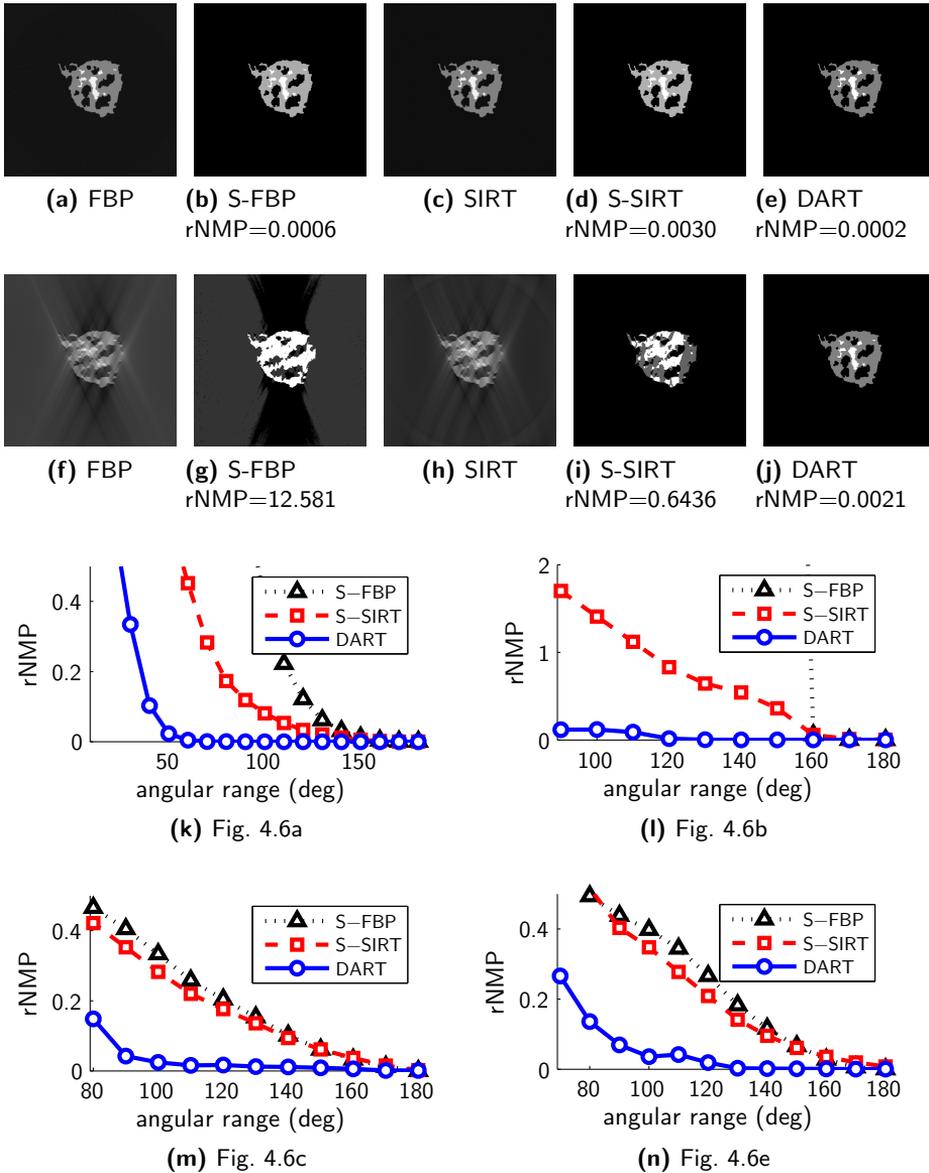


Figure 4.9: (a-e) Reconstructions of Fig. 4.6b from 180 projections in a full (180°) angular range. (f-j) Reconstructions of Fig. 4.6b from 180 equiangular projections in a 130° angular range. (k-n) For reconstructions of four phantom images, the rNMP is plotted as a function of the angular range (in degrees).

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

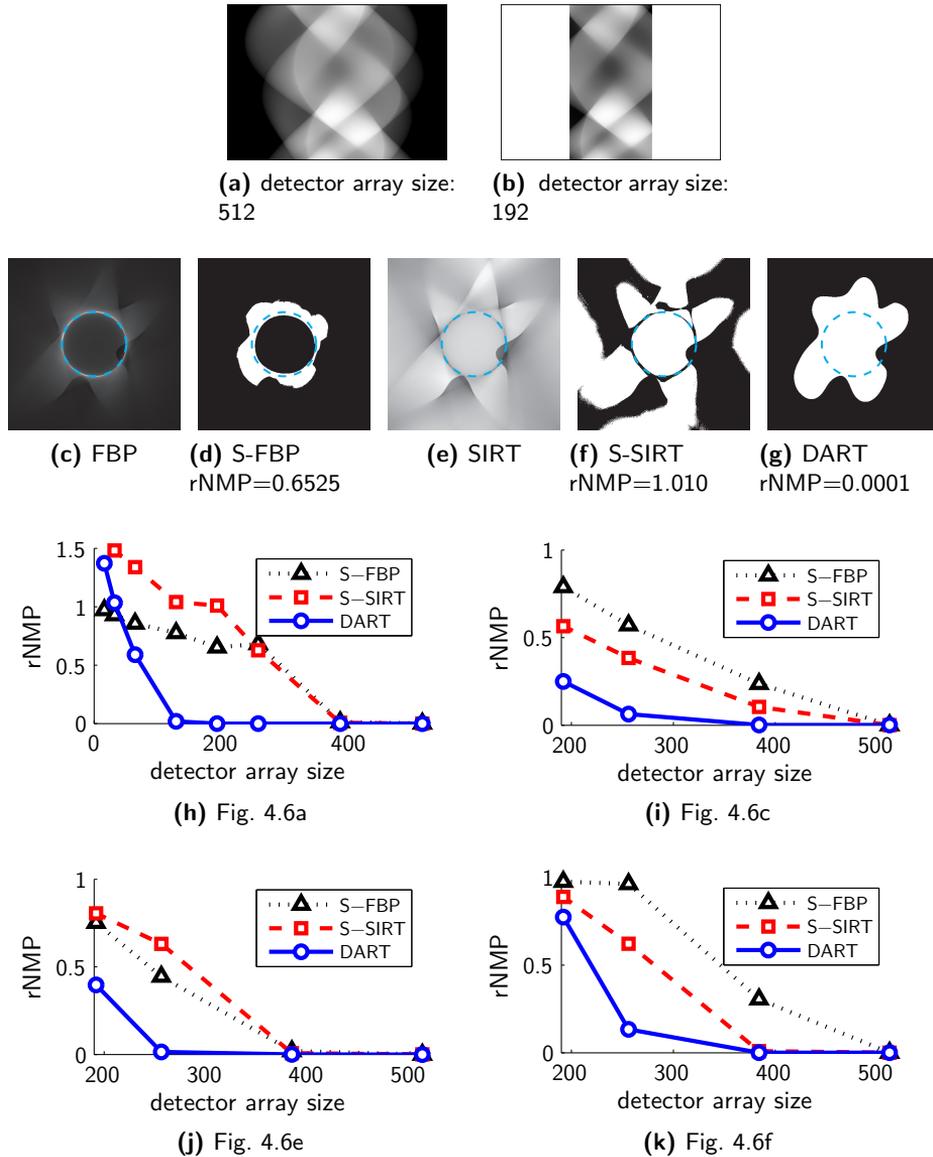


Figure 4.10: (a) Simulated projection data of Fig. 4.6a with 180 non-truncated equiangular projections. (b) Simulated projection data of phantom image Fig. 4.6a with 180 truncated equiangular projections. (c-g) Reconstructions of Fig. 4.6a from the projection data shown in (b). The FOV is depicted by the blue ring. (h-k) For reconstructions of four phantom images, the rNMP is plotted as a function of the total detector size.

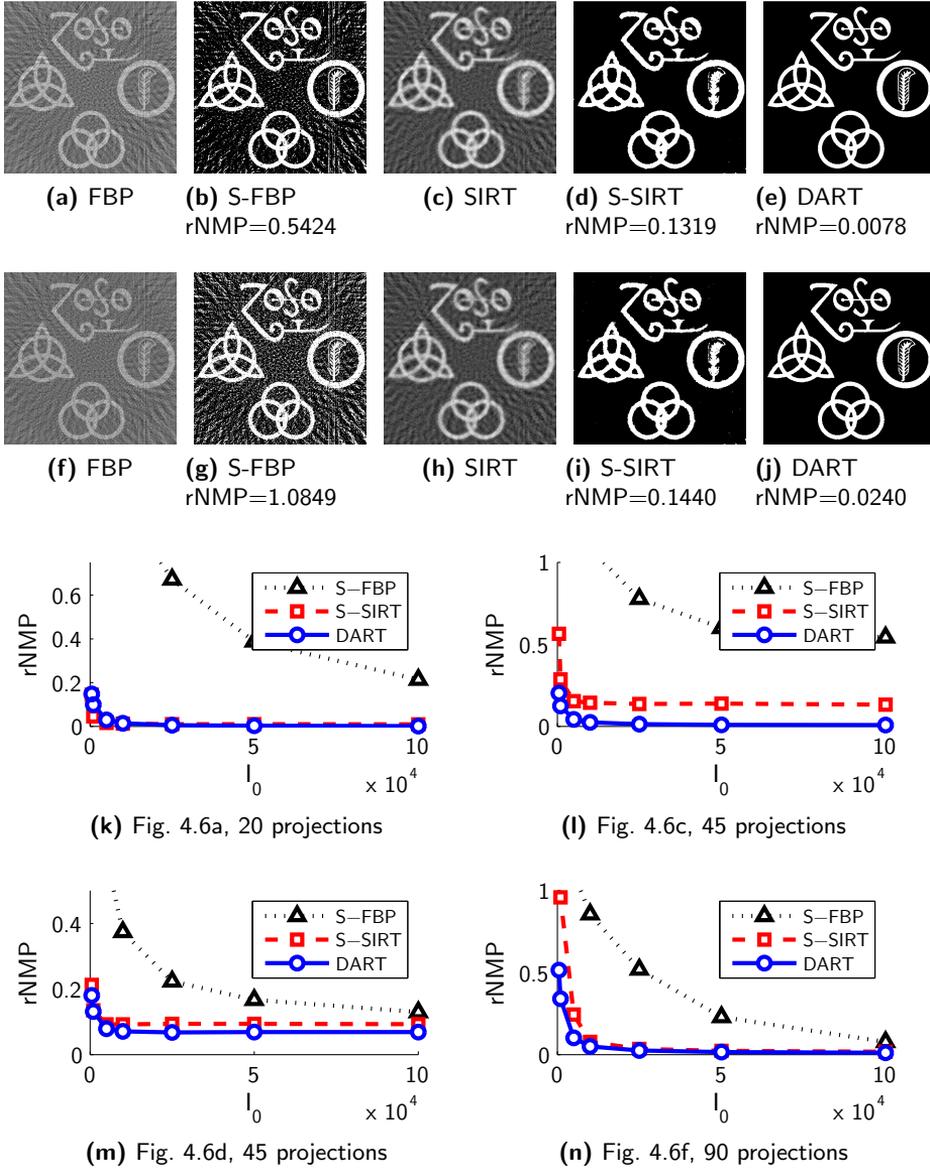


Figure 4.11: (a-e) Reconstructions of Fig. 4.6c from 45 equiangular projections with $I_0=100000$. (f-j) Reconstructions of Fig. 4.6c from 45 equiangular projections with $I_0=10000$. (k-n) For reconstructions of four phantom images, the rNMP is plotted as a function of the incident beam intensity, I_0 .

even for a high level of truncation. These findings are confirmed in Fig. 4.10h-k, where for reconstructions of Fig. 4.6a,c,e,f, the rNMP is plotted as a function of the detector array size.

4.3.1.4 Noisy projection data

For the final set of simulation experiments, Poisson noise was applied and the incident beam intensity, I_0 , was lowered. In Fig. 4.11a-e, reconstructions are shown of phantom image Fig. 4.6c from 45 equiangular, high SNR ($I_0 = 100000$) projection images. Fig. 4.11f-j show reconstructions of the same image from the same projection directions, but with a low SNR ($I_0 = 10000$). The presence of Poisson noise had a clear negative effect on the quality of the DART reconstructions. However, the accuracy of the DART reconstructions was still superior to that of the other methods. In Fig. 4.10k-n, the rNMP is plotted as a function of I_0 for reconstructions of Fig. 4.6a (with 20 projections), Fig. 4.6c (with 45 projections), Fig. 4.6d (with 45 projections) and Fig. 4.6f (with 90 projections).

4.3.2 Experimental studies

DART was also applied on three experimental datasets.

- (a) A raw diamond stone (Fig. 4.12a). This scan was performed at Diamcad, a Belgian diamond processing company that uses computed tomography to perform a detailed study of raw stones. Diamonds are an ideal case for discrete tomography as they consist of only one material and they have a fairly simple shape. The stone in question was scanned in a Scanco Medical μ CT 40 with a circular cone beam geometry with a 1024×56 pixel detector. 250 equiangular projections were recorded in the interval $[0, \pi)$. Using the manufacturer's software, the data was linearized and rebinned to a parallel beam projection geometry. Here, only a single slice of the projection data is considered (depicted by the full pink line in Fig. 4.12a).
- (b) An aluminium foam (Fig. 4.12b), recorded with a SkyScan 1172 μ CT scanner with 481 equiangular cone beam projections. The detector resolution was $9.7\mu\text{m}$. Prior to reconstruction, the data was corrected for ring artefacts and beam hardening with the SkyScan NRecon software package. Only the central slice was considered. Thus, the projection data had a fan beam projection geometry.
- (c) A rat femur (Fig. 4.12c), recorded with a SkyScan 1172 μ CT scanner with 368 equiangular cone beam projections. The detector resolution was $4.5\mu\text{m}$. The SkyScan NRecon software package was again used to preprocess the data.

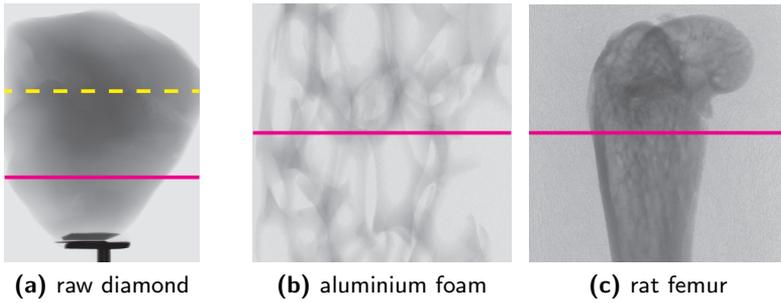


Figure 4.12: Projection images of three different experimental datasets. The slices that are considered are highlighted with a full pink line. For dataset (a), a truncated slice is highlighted with a striped yellow line.

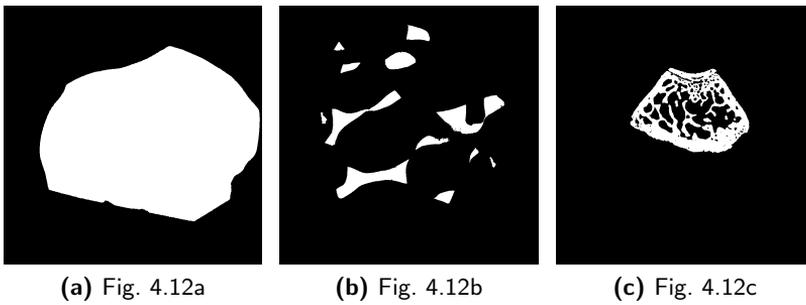


Figure 4.13: Manually created segmented image of the pink slices depicted in Fig. 4.12. These images are used as ground truth for validation.

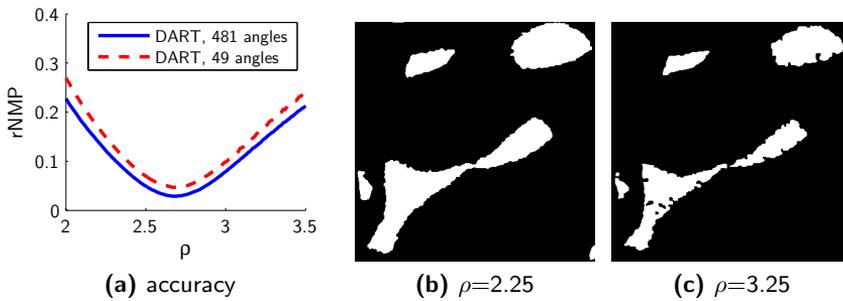


Figure 4.14: (a) The rNMP of the DART reconstructions of dataset Fig. 4.12b as a function of the chosen grey level, for two different angle counts. (b,c) DART reconstructions of the same dataset (with 49 projection angles) with two different choices for ρ .

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

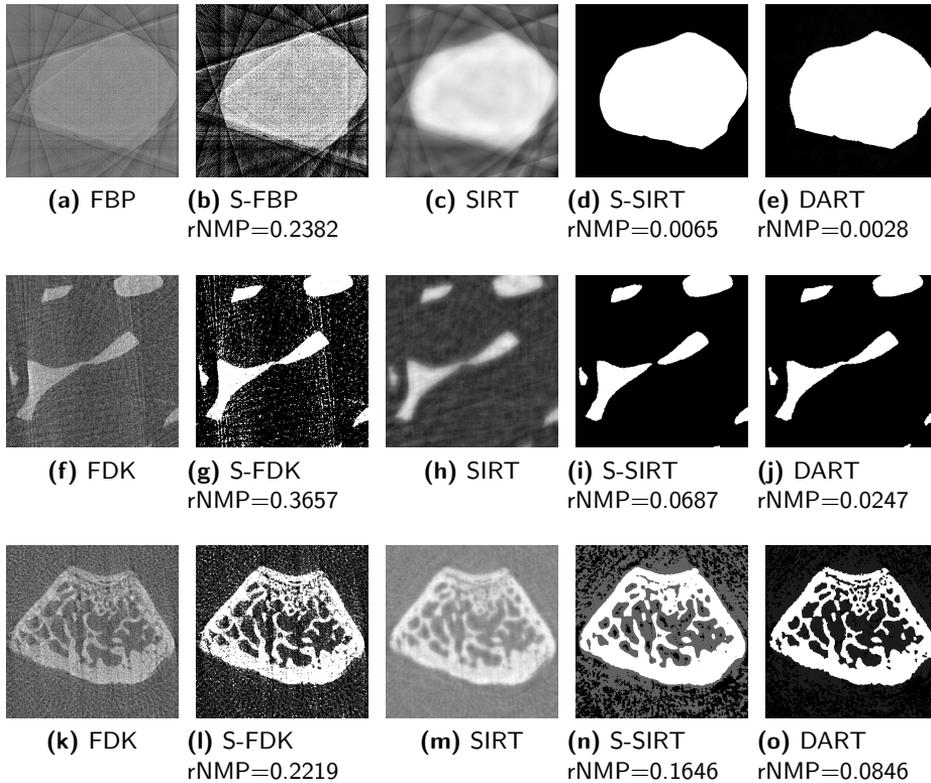


Figure 4.15: (a-e) Reconstructions of Fig. 4.12a with 10 projection angles. (f-j) Reconstructions of Fig. 4.12b with only 49 projection angles. (k-o) Reconstructions of Fig. 4.12c with only 62 projection angles.

Also here, only the central slice was considered. Note that the sample was held in place with paper. This means that a third grey level had to be added in DART and in the segmentation algorithms.

For validation, ground truth images were created by manually segmenting SIRT reconstructions (Fig. 4.12d-f).

In DART, it is assumed that the grey levels are prior knowledge. While this assumption was easy to satisfy for the simulation experiments, grey level estimation for experimental datasets is non-trivial. Furthermore, the quality of the reconstruction greatly depends on the accuracy of the estimation. To demonstrate this, DART reconstructions of Fig. 4.12b were created with a range of different grey level choices. This was done once considering all 481 projection angles and once

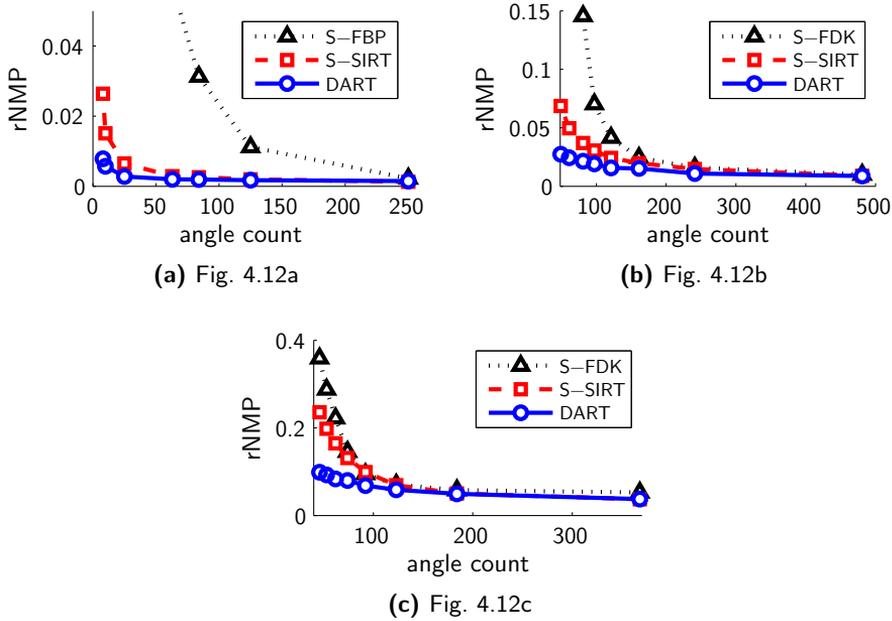


Figure 4.16: For all experimental datasets of Fig. 4.12, the rNMP as a function of the number of projection angles.

considering only 49 angles. In Fig. 4.14a, the corresponding rNMP is plotted as a function of the grey level. Wrongly estimating the grey level had a significant impact on the reconstruction quality. One can also observe this in Fig. 4.14b and Fig. 4.14c, where reconstructions are shown for grey levels that were too low and too high, respectively. If the grey level was too low, the reconstructed objects were too thick, if the grey level was too high, holes appeared inside the homogeneous structures. For the remainder of all experiments described in this section, the grey levels were manually chosen on a trial-and-error base. In Section 4.4, Chapter 5 and Chapter 6, methods are proposed to automate these estimations.

The number of projection angles in each sinogram was gradually reduced and reconstructions were created with FBP, SIRT and DART. Otsu's method was applied to obtain segmented images. Note, however, that the projection data of Fig. 4.12b and Fig. 4.12c had a fan beam projection geometry. This means that conventional FBP could not be used. Instead, the *Feldkamp-Davis-Kress* (FDK) method was applied. Its segmentation is referred to by *S-FDK*. In Fig. 4.15 reconstructions are shown for all datasets with a significantly reduced angle count (10 projections instead of 250 for Fig. 4.12a, 49 instead of 481 for Fig. 4.12b and

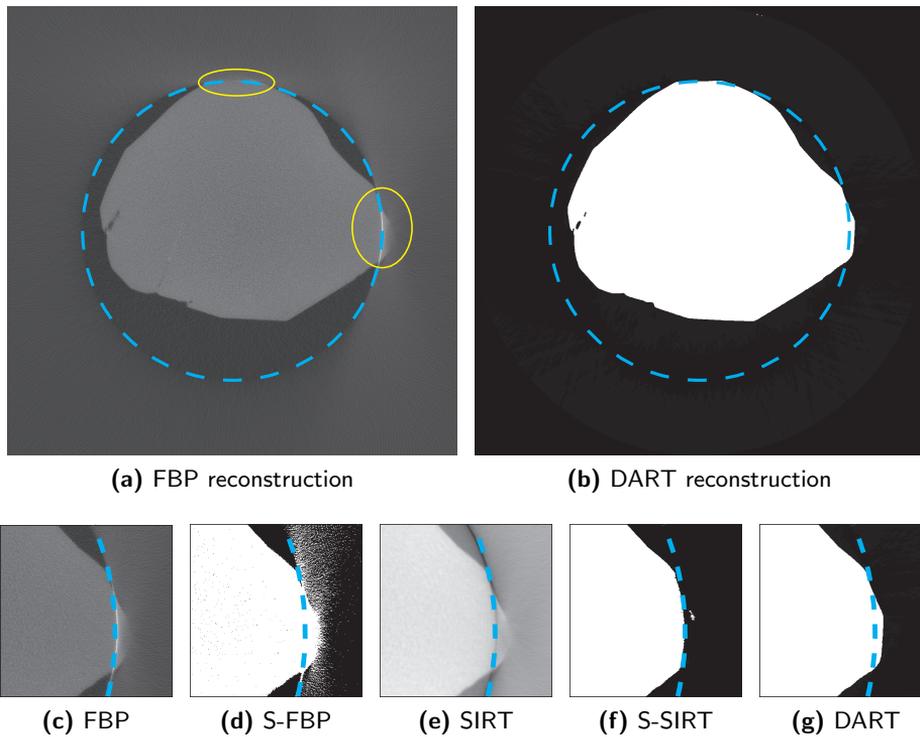


Figure 4.17: (a) FBP reconstruction of the truncated slice of Fig. 4.12a. The FOV is highlighted by the striped blue line. The areas in which the diamond stone transcended the FOV are marked by the yellow circles. (b) DART reconstruction of the same object. (c-g) Zoomed in reconstructed images with various reconstruction methods. Clearly, only DART provided an accurate estimation of the diamond border outside the FOV.

62 instead of 368 for Fig. 4.12c). In Fig. 4.16, the rNMP for all experiments is plotted as a function of the angle count. The results are consistent with those of the simulation study in Section 4.3.1.1: DART outperformed the conventional reconstruction techniques, substantially so for a very low angle count.

In the projection image Fig. 4.12a, it can be noted that, at the striped line, the data was truncated. The diamond stone was simply too large to fit into the FOV of the scanner. In Fig. 4.17a an FBP reconstruction is shown. The two areas in which the diamond stone transcended the FOV are clearly marked. FBP, SIRT and DART were applied on the truncated slice (Fig. 4.17c-g). Clearly, only DART succeeded in accurately estimating the boundary of the object outside the FOV. While numerical validation is impossible (a ground truth image could not be created), the DART reconstruction corresponded very well with the actual stone.

4.4 Automatic parameter optimization

This section has been published as:

Wim van Aarle, Karel Crombecq, Ivo Couckuyt, K. Joost Batenburg, Jan Sijbers, “Efficient parameter estimation for discrete tomography using adaptive modelling”, Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, pp.229–232, July 2011.

From the algorithm description in Fig. 4.2c, it is clear that there are various parameters to DART that influence the accurate of the reconstructed images. Notably the grey levels ρ must be known exactly, but also the threshold values τ must be accurately estimated. Additionally, the percentage of additional random pixels, r , and the intensity of the blurring filter, b , also affect the outcome of the DART algorithm.

As DART is an algorithm that typically requires a long time to compute, the estimation of the algorithm parameters by trial-and-error can be a very daunting task. Fortunately, iterative tomography inherently provides an easy and objective method to automatically evaluate each reconstruction. The Euclidean distance between a forward projection of the reconstructed image and the measured projection data can be used to score the corresponding algorithm parameters. The following cost function can therefore be introduced:

$$\Phi(\rho, \tau, b, r) = \|\mathbf{W} \mathbf{v}_{(\rho, \tau, b, r)}^{(k)} - \mathbf{p}\|_2. \quad (4.13)$$

In Eqn. 4.13, $\mathbf{v}_{(\rho, \tau, b, r)}^{(k)}$ is defined by the reconstructed DART image after k iterations created with the parameter values ρ , τ , b and r .

Eqn. 4.13 is non-differentiable and the search space is likely to contain multiple local optima. Furthermore, due to the computational cost of DART, the number of function evaluations required for minimizing the cost function should be as low as possible. Popular optimization routines, such as *Nelder-Mead simplex search* [19], are able to solve complex optimization problems, but typically do so by using a large number of function evaluations. To deal with these issues, recent advances in the field of adaptive surrogate modelling can be used. A concise introduction to this field is given in Section 4.4.1. In Section 4.4.2, various optimization strategies are applied to optimize the cost function Eqn. 4.13.

4.4.1 Adaptive surrogate modelling

Adaptive modelling is a technique in which a surrogate or approximation model of the original, expensive cost function is built by evaluating this function in a number of points, and training a mathematical model on this data. This model is

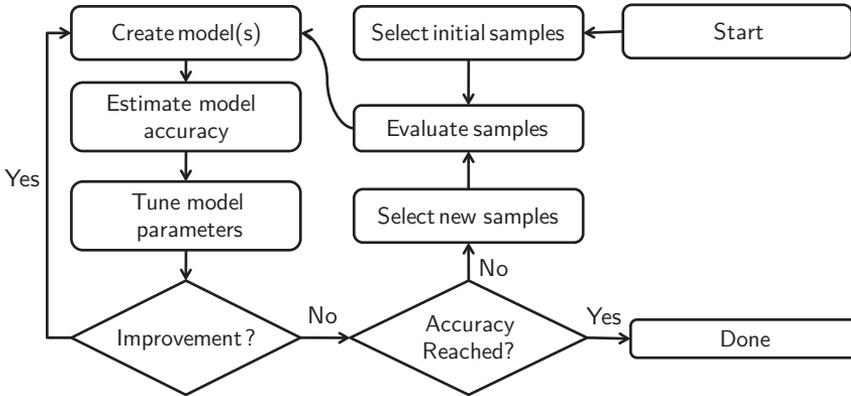


Figure 4.18: Flowchart of an adaptive modelling routine.

then iteratively refined by selecting additional points in interesting locations, and rebuilding the model with this new data [20]. The algorithm that selects these additional points is called the *sampling strategy*. The resulting approximation model can then be used as a drop-in replacement for the original cost function that can be evaluated much faster. A schematic overview of the adaptive modelling process is given in Fig. 4.18.

A popular choice for the mathematical model is *Kriging* [21]. In addition to predicting the behaviour of the cost function on the entire domain, it also provides a measure of uncertainty, indicating where in the domain the model is the most uncertain about its predictions. This information can be used to determine where additional points should be located. A sampling algorithm that exploits this property is the *Efficient Global Optimization (EGO)* algorithm [22]. The EGO method combines the prediction and prediction variance (uncertainty) of the Kriging model to search for the optimum and refining the surrogate model in unexplored locations.

The main advantage of using approximation models is that it can reduce the number of function evaluations required to find the optimum. Instead of searching for the optimum by evaluating the cost function directly, the Kriging approximation model is thoroughly explored, and new points are only selected when it is deemed absolutely necessary, thus saving a lot of expensive function evaluations.

4.4.2 Simulation studies

To compare various optimization strategies to optimize the cost-function Eqn. 4.13, simulated parallel beam projection images were computed of Fig. 4.19a (with 10 equiangular projections), Fig. 4.19b and Fig. 4.19c (both with 60 equiangular

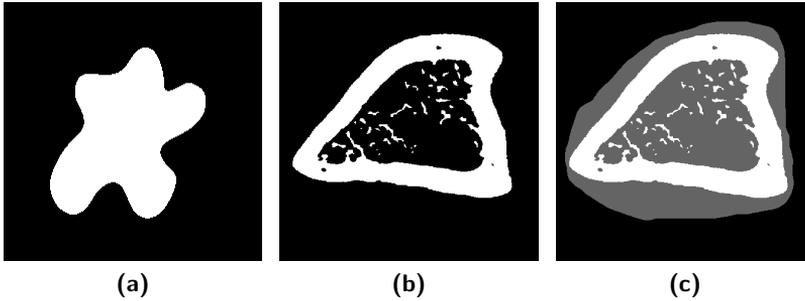


Figure 4.19: Simulated phantom images that were used to demonstrate various optimization strategies for automatic parameter estimation.

projections). To reduce the high dimensionality of the search space, only a few algorithm parameters were optimized at a time.

Firstly, the proposed cost function was compared to the rNMP space. In Fig. 4.20, this is shown for $\Phi(\rho_1, \rho_2)$ on phantom image Fig. 4.19a and for $\Phi(r, \rho_2)$ on phantom image Fig. 4.19b. Even though these spaces were not an exact match, their minima (white dots) were close together.

Secondly, the following optimization strategies were compared.

- The Nelder-Mead simplex optimization as described in [19] and as implemented in `fminsearch` of the MATLAB global optimization toolbox.
- A pattern search optimization method [23] implemented in `patternsearch` of the MATLAB global optimization toolbox.
- An adaptive optimization method as discussed in Section 4.4.1. The SUMO (SURrogate MOdeling) Toolbox, a free MATLAB toolbox designed for adaptive surrogate modelling and sampling, was used [20].

These optimization techniques were run for on a variety of parameter combinations until a certain target for the cost function was reached. For the simplex and pattern search optimization, the number of evaluations depends greatly on the initial parameter choice. These experiments were therefore performed 20 times using randomly selected initial parameters. Also, the number of optimization attempts where the target value was never reached due to stranding in a local optima were counted.

Fig. 4.21 shows, for each experiment, the number of function evaluations that was required to reach the accuracy target. Overall, the adaptive modelling optimization required fewer function evaluations when compared to the simplex method, which in turn required substantially fewer function evaluations than a

CHAPTER 4. THE DISCRETE ALGEBRAIC RECONSTRUCTION TECHNIQUE (DART)

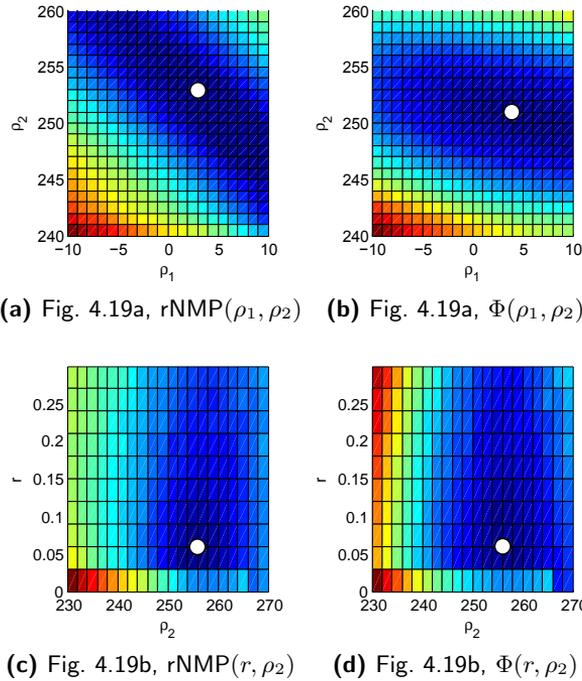


Figure 4.20: Comparison of cost function Φ to the rNMP space. The minima are marked by white dots.

pattern search method. Pattern search was more robust than a simplex search with respect to local optima. During this set of experiments the adaptive method never ran into a local optimum.

The rNMP of the DART reconstructions with the optimized parameters is shown in Fig. 4.22. For the simplex and pattern search optimization, the parameters of the median run, in terms of function evaluations, were used. Given that Eqn. 4.13 was optimized until a sufficiently low value of the cost function has been reached, the rNMP was comparably low for each of the optimization routine.

4.5 Conclusions

The *Discrete Algebraic Reconstruction Technique (DART)* can be used if the scanned object is known to consist of only a few different homogeneous materials, each corresponding to a priori known grey level. While DART is a heuristic algorithm and its convergence has not been proven, it has been thoroughly vali-

4.5. CONCLUSIONS

	<i>simplex</i>				<i>pattern search</i>				<i>adaptive</i>
	min	max	med	failed	min	max	med	failed	
Fig. 4.19a									
b, ρ_2	8	28	17	0	15	74	52	2	13
r, ρ_2	8	23	16	0	27	91	46	0	18
ρ_1, ρ_2	30	49	38	3	32	99	43	6	16
ρ_1, τ, ρ_2	46	120	75	7	80	129	147	0	56
Fig. 4.19b									
b, ρ_2	1	17	10	0	1	49	24	0	10
r, ρ_2	5	22	11	6	5	78	42	0	10
ρ_1, ρ_2	1	39	32	1	19	89	58	1	11
ρ_1, τ, ρ_2	31	59	49	8	63	113	96	0	20
Fig. 4.19c									
b, ρ_2, ρ_3	10	36	28	5	37	138	104	2	23
r, ρ_2, ρ_3	14	49	29	3	25	239	128	1	14
ρ_1, ρ_2, ρ_3	15	84	50	4	99	240	147	0	30
$\tau_1, \rho_2, \tau_2, \rho_3$	37	50	43	6	97	253	153	0	39
τ_1, τ_2	3	16	7	0	1	40	17	0	11

Figure 4.21: For the three optimization methods, the number of function evaluations required to reach a certain accuracy.

Fig. 4.19a	<i>simplex</i>	<i>pattern search</i>	<i>adaptive</i>
b, ρ_2	0%	0%	0%
r, ρ_2	0%	0%	0%
ρ_1, ρ_2	0%	0%	0%
ρ_1, τ, ρ_2	0.01%	0.02%	0%
Fig. 4.19b	<i>simplex</i>	<i>pattern search</i>	<i>adaptive</i>
b, ρ_2	0.36%	0.46%	0.20%
r, ρ_2	0.39%	0.33%	0.32%
ρ_1, ρ_2	0.49%	0.47%	0.56%
ρ_1, τ, ρ_2	0.47%	0.55%	0.56%
Fig. 4.19c	<i>simplex</i>	<i>pattern search</i>	<i>adaptive</i>
b, ρ_2, ρ_3	0.58%	0.59%	0.77%
r, ρ_2, ρ_3	0.84%	0.61%	0.70%
ρ_1, ρ_2, ρ_3	0.75%	0.71%	0.68%
$\tau_1, \rho_2, \tau_2, \rho_3$	0.89%	0.96%	0.74%
τ_1, τ_2	1.04%	0.71%	0.89%

Figure 4.22: The rNMP of the reconstructions with the estimated parameter values.

REFERENCES

dated with simulation experiments and has been applied on various experimental datasets.

A key problem of DART is its assumption that the set of grey levels in the unknown reconstructed image is known a priori. In practice, obtaining such knowledge is non-trivial.

In Section 4.4, the Euclidean distance from the forward projection of the reconstruction to the measured data was proposed as a cost function to score parameter values. By minimizing this function, optimal grey levels and algorithm parameters can be automatically approximated. However, given the computational requirements of DART, such an optimization can take a long time, even with an advanced optimization strategy such as adaptive modelling optimization.

In the ensuing two chapters, two computationally efficient techniques are proposed that estimate optimal values for the grey levels without the need to perform multiple costly DART reconstructions. In Chapter 5, a semi-automatic method is proposed that estimates the optimal grey levels *prior* to the DART reconstruction. It requires the user to select a region of the reconstruction grid in which he is sure the object is homogeneous. In Chapter 6, the Projection Distance Minimization method (Chapter 2) is used to automatically estimate the grey levels and the optimal threshold values *during* the reconstruction.

References

- [1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Transactions on Signal Processing*, vol. 25, no. 2, pp. 21–30, 2008.
- [3] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial ct scanners still employ traditional, filtered back-projection for image reconstruction?," *Inverse Problems*, vol. 25, no. 12, pp. 1–36, 2009.
- [4] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, no. 17, pp. 4777–4807, 2008.
- [5] J. W. Stayman, Y. Otake, A. Uneri, J. L. Prince, and J. H. Siewerdsen, "Likelihood-based ct reconstruction of objects containing known components," in *11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2011.
- [6] D. L. Snyder, J. A. O'Sullivan, B. R. Whiting, R. J. Murphy, J. Benac, J. A. Cataldo, D. G. Politte, and J. F. Williamson, "Deblurring subject to nonnegativity

- constraints when known functions are present with application to object-constrained computerized tomography,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 10, pp. 1009–1017, 2001.
- [7] G. T. Herman and A. Kuba, eds., *Discrete Tomography: Foundations, Algorithms, and Applications*. Applied and Numerical Harmonic Analysis, Birkhuser Boston, 1999.
- [8] G. T. Herman and A. Kuba, eds., *Advances in Discrete Tomography and Its Applications*. Applied and Numerical Harmonic Analysis, Birkhuser Boston, 2007.
- [9] T. Schule, C. Schnorr, S. Weber, and J. Hornegger, “Discrete tomography by convexconcave regularization and d.c. programming,” *Discrete Applied Mathematics*, vol. 151, pp. 229–243, 2005.
- [10] H. Y. Liao and G. T. Herman, “A coordinate ascent approach to tomographic reconstruction of label images from a few projections,” *Discrete Applied Mathematics*, vol. 151, pp. 184–197, 2005.
- [11] K. J. Batenburg and J. Sijbers, “DART: A practical reconstruction algorithm for discrete tomography,” *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2542–2553, 2011.
- [12] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, C. E. A., and G. Van Tendeloo, “3D imaging of nanomaterials by discrete tomography,” *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.
- [13] S. Bals, K. J. Batenburg, D. Liang, O. Lebedev, G. Van Tendeloo, A. Aerts, J. A. Martens, and K. C. E., “Quantitative three-dimensional modeling of zeolite through discrete electron tomography,” *Journal of the American Chemical Society*, vol. 131, no. 13, pp. 4769–4773, 2009.
- [14] K. J. Batenburg, J. Sijbers, H. F. Poulsen, and E. Knudsen, “DART: a robust algorithm for fast reconstruction of three-dimensional grain maps,” *Journal of Applied Crystallography*, vol. 43, no. 6, pp. 1464–1473, 2010.
- [15] J. Gregor and T. Benson, “computational analysis and improvement of SIRT,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 918–924, 2008.
- [16] F. J. Maestre-Deusto, G. Scavello, J. Pizarro, and P. L. Galindo, “Adart: An adaptive algebraic reconstruction algorithm for discrete tomography,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2146–2152, 2011.
- [17] N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, March 1979.
- [18] G. Van Gompel, *Toward accurate image reconstruction from truncated X-ray CT projections*. PhD thesis, University of Antwerp, 2009.

REFERENCES

- [19] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the nelder-mead simplex method in low dimensions,” *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [20] D. Gorissen, K. Crombecq, I. Couckuyt, T. Dhaene, and P. Demeester, “A surrogate modeling and adaptive sampling toolbox for computer based design,” *Journal of Machine Learning Research*, vol. 11, pp. 2051–2055, 2010.
- [21] J. Sacks, W. J. Welch, T. Mitchell, and H. P. Wynn, “Design and analysis of computer experiments,” *Statistical science*, vol. 4, pp. 409–435, 1989.
- [22] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient global optimization of expensive black-box functions,” *Global Optimization*, vol. 13, pp. 455–492, 1998.
- [23] V. Torczon, “On the convergence of pattern search algorithms,” *SIAM Journal of Optimization*, vol. 7, pp. 1–25, January 1997.

—Yes, we scan.

Dirk De Nef, 2011.

5

A semi-automatic algorithm for grey level estimation in tomography

Contents

5.1	Introduction	116
5.2	Concept and notation	117
5.2.1	Estimating grey levels	117
5.2.2	Switching components	117
5.2.3	Discretization	118
5.2.4	Grey level penalty function	119
5.3	Discrete Grey Level Selection algorithm	120
5.4	Experiments	122
5.4.1	Grey level estimation	122
5.4.2	DGLS + DART	129
5.5	Discussion	131
5.6	Conclusions	132
	References	133

This chapter has been published as:

K. Joost Batenburg, *Wim van Aarle*, Jan Sijbers, “A semi-automatic algorithm for grey level estimation in tomography”, *Pattern Recognition Letters*, vol.32, no.9, pp.1395–1405, July 2011.

5.1 Introduction

A common assumption in discrete tomography reconstruction algorithms, is that the set of grey levels in the unknown original image is known before the reconstruction. In this chapter, the grey level estimation problem is considered. In practical applications, this estimation is often very difficult. Even when the materials and their densities are known in advance, calibration is required to translate the material properties into a reconstructed grey level. The actual grey levels depend on these parameters, and they may vary among different scanner devices. Also, imaging properties of the scanning system may change over time. For example, the X-ray source of a CT scanner heats up while scanning, which might change its operational characteristics. While this may have a negligible effect on the resulting grey levels during a single scan, grey levels in scans acquired at different times may show significant differences.

In many applications of discrete tomography, the number and orientation of the projections that can be acquired is restricted by experimental limitations. In materials science, for example, discrete tomography is used to reconstruct three-dimensional nano-structures from a series of projection images acquired by an electron microscope [1]. Due to the structure of the sample holder, projections can only be obtained for a limited range of angles, resulting in severe reconstruction artefacts for conventional algorithms.

When several similar objects are scanned as a single batch, it is sometimes possible to obtain a high quality reconstruction of one of these objects, based on a large number of projections. This reconstruction can then be used to estimate the set of grey levels for the remaining objects, which can then be reconstructed from few projections. Even when only a limited series of projections is available, one can still compute a reconstruction by classical, non-discrete reconstruction algorithms. The resulting image will be degraded by artefacts, yet it may still be possible for an expert user to delineate certain areas that are likely to have a constant composition in the original object.

In this chapter, a simpler version of the grey level estimation problem is proposed. The user first selects an image region that can be expected to correspond to a homogeneous region in the original object, based on an initial reconstructed image, obtained by a classical, non-discrete reconstruction algorithm. In certain cases, knowledge of such a constant region allows for reliable estimation of the grey level corresponding to the selected region.

The outline of this chapter is as follows. In Section 5.2, the problem of grey level estimation is introduced, along with formal notation. In Section 5.3, a semi-automatic approach is presented for estimating the grey levels when for each grey level a subset of corresponding pixels is given. Experimental results for a range

of simulation experiments are described in Section 5.4. Section 5.5 discusses the possibilities for obtaining more accurate estimates, compared to the approach of this chapter. Section 5.6 concludes the chapter.

5.2 Concept and notation

5.2.1 Estimating grey levels

The reconstruction problem in discrete tomography concerns the recovery of an object function $f(x, y) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ from its Radon transform $\mathcal{R}f$, where the set of unique values in f is denoted by $\boldsymbol{\rho} = \{\rho_1, \dots, \rho_l\}$. The *grey level estimation problem* is stated as follows.

Problem 1. Given $\mathcal{R}f$, estimate the values ρ_1, \dots, ρ_l .

The grey level estimation problem does not have a unique solution in general. When only a small number of projections is available and f contains many grey levels, it is straightforward to find examples where the grey levels cannot be determined uniquely. To deal with this non-uniqueness problem, additional prior knowledge must be incorporated in the grey level estimation problem. Suppose that for each grey level $\rho_t \in \boldsymbol{\rho}$, a region $A_t \subset \mathbb{R} \times \mathbb{R}$ is given on which $f(x, y)$ is known to have a constant grey level. This prior knowledge can be obtained by computing an initial grey level reconstruction using a standard tomography algorithm, and letting an expert user select regions in this reconstruction that are likely to be constant in the original scanned object. In some cases, this additional constraint will be sufficient to determine the grey levels for the selected regions uniquely from the projection data.

5.2.2 Switching components

Unfortunately, even specifying an entire region where the image is known to be constant is not sufficient to guarantee that the grey levels can be estimated accurately. In [2], a procedure is introduced to generate so-called *switching components*, non-zero images that have a zero projection in a given set of projection angles. Fig. 5.1 shows switching components in the horizontal, vertical, and diagonal direction. For each new direction, a negated copy of the switching component is added and translated in the given direction. A similar construction can be applied for any finite set of projection directions, leading to the following proposition regarding the non-uniqueness of a result.

Proposition 1. Let $\Theta = \{\theta_1, \dots, \theta_d\}$ be a given set of projection angles. Let $\boldsymbol{\rho} = \{\rho_1, \dots, \rho_l\}$, $A \subset \mathbb{R} \times \mathbb{R}$ and $\rho \in \boldsymbol{\rho}$. Let $f(x, y) = \rho$ for all $(x, y) \in A$. Then for each grey level $\tilde{\rho} \in \mathbb{R}$, there is an image \tilde{f} such that $\tilde{f}(x, y) = \tilde{\rho}$ for all $(x, y) \in A$

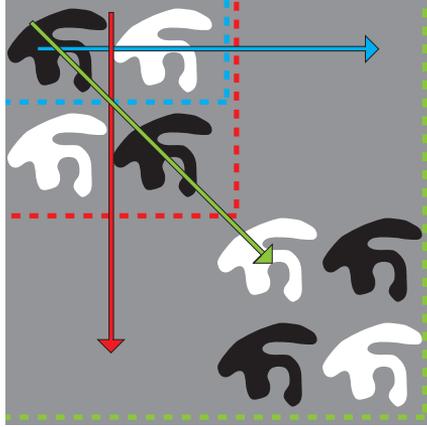


Figure 5.1: Example of a switching component for three directions, which is constant and non-zero on each of the black and white regions.

and $\mathcal{R}f(t, \theta_i) = \mathcal{R}\tilde{f}(t, \theta_i)$ for $i = 1, \dots, l$, $t \in \mathbb{R}$. Moreover, there is such an image \tilde{f} that has at most $3k$ grey levels.

Proof (sketch) By the construction depicted in Fig. 5.1, an image can be created that has a constant value of 1 on A , has constant zero projections in all given directions and only contains grey levels from $\{-1, 0, 1\}$. Let $\bar{\rho} = \tilde{\rho} - \rho$. By adding a multiple of $\bar{\rho}$ times the switching component to f , an image \tilde{f} that conforms to the proposition can be created. This image will have at most $3k$ grey levels, included in the set $\{\rho_1 - \bar{\rho}, \rho_1, \rho_1 + \bar{\rho}, \rho_2 - \bar{\rho}, \rho_2, \rho_2 + \bar{\rho}, \dots, \rho_l - \bar{\rho}, \rho_l, \rho_l + \bar{\rho}\}$. Note that some of these grey levels may be negative, even if the image f has only non-negative grey levels.

5.2.3 Discretization

Experimental tomography data consists of a vector of measured line integrals, obtained at a finite set of detector positions. Denote the total number of measured detector values, for all angles, by m . Denote the measured projection data by $\mathbf{p} = (p_i) \in \mathbb{R}^m$. The images $f(x, y)$ can be discretized and represented on a grid. Let n be the number of pixels in the image. It is assumed that the image is zero outside the rectangle covered by this grid. Let $\mathbf{v} = (v_j) \in \mathbb{R}^n$ denote the discretized image of the object, where all pixel values are stored as consecutive elements of a vector. The Radon transform for a finite set of angles is modelled as a linear operator \mathbf{W} , called the *projection operator*, that maps the image \mathbf{v} to the projection data \mathbf{p} :

$$\mathbf{W}\mathbf{v} = \mathbf{p}. \quad (5.1)$$

In Eqn. 5.1, \mathbf{W} represents an $m \times n$ matrix $\mathbf{W} = (w_{ij})$. The vector \mathbf{p} is called the *forward projection* or *sinogram* of \mathbf{v} . With this notation the key problem that is investigated in this chapter can be defined.

Problem 2. Let $\mathbf{v} \in \mathbb{R}^n$ be an unknown image and let $\rho \in \mathbb{R}$ be an unknown grey level. Suppose that $\mathbf{p} = \mathbf{W}\mathbf{v}$ is given, and that a set $A \subset \{1, \dots, n\}$ is given such that $v_j = \rho$ for all $j \in A$. Find ρ .

5.2.4 Grey level penalty function

Let $A \subset \{1, \dots, n\}$ and $\rho \in \mathbb{R}$, as in Problem 2. Reorder the pixels $\{1, \dots, n\}$ and the corresponding columns of \mathbf{W} , such that:

$$\mathbf{W} = (\mathbf{W}_A \mathbf{W}_B), \tag{5.2}$$

where \mathbf{W}_A contains the columns of \mathbf{W} corresponding to the pixels in A , and \mathbf{W}_B contains the columns corresponding to B , the complement of A . As all elements of A must have an unknown, but constant grey level ρ , it follows that:

$$\mathbf{W}\mathbf{v} = (\mathbf{W}_A \mathbf{W}_B) \begin{pmatrix} \bar{\rho} \\ \mathbf{v}_B \end{pmatrix} = \mathbf{p}, \tag{5.3}$$

where $\bar{\rho}$ denotes a constant vector for which all entries are equal to ρ . Furthermore, \mathbf{v}_B refers to the vector containing all v_j for which $j \in B$. This leads to:

$$\mathbf{W}_B \mathbf{v}_B = \mathbf{p} - \mathbf{W}_A \bar{\rho}, \tag{5.4}$$

which provides a necessary condition for a grey level estimate to be correct: Eqn. 5.4 must have a solution. Although this condition is not always sufficient to estimate ρ , it will often be sufficient if the size of A is a relatively large fraction of n and the number of projections is not too low.

Even if ρ is determined uniquely, it may not be possible to solve Eqn. 5.4 exactly, e.g. due to noise and discretization errors. Given A , one can measure the inconsistency of a grey level ρ with respect to the projection data \mathbf{p} using the following *grey level penalty function*:

$$P(\rho) = \min_{\mathbf{v}_B} \|\mathbf{p} - \mathbf{W}_A \bar{\rho} - \mathbf{W}_B \mathbf{v}_B\|, \tag{5.5}$$

where $\|\cdot\|$ denotes a certain norm, to be defined below.

Note that for each grey level $\tilde{\rho}$, there may be multiple vectors \mathbf{v}_B for which the minimum penalty is attained. This does not have to be a problem for grey level estimation, as long as the grey level for which the penalty is minimal *is* uniquely determined by the projection data. According to Proposition 1, it may occur that

the grey level cannot be determined uniquely, in which case one should seek to find at least one of the grey levels for which the penalty is minimal.

A range of iterative methods are available for evaluating Eqn. 5.5. In this chapter the SIRT algorithm is applied. As discussed in Section 1.3.3.1, the SIRT algorithm converges to a solution of Eqn. 5.5 where the norm to be minimized is a weighted sum of squares [3].

To minimize the grey level penalty function Eqn. 5.5, it must typically be evaluated in several points, resulting in a series of SIRT reconstructions, corresponding to different grey level estimates. Depending on the particular type of tomography problem (e.g. X-ray transmission tomography, electron tomography), an additional constraint may be incorporated in some cases, restricting the admissible grey levels to non-negative values. However, solving Eqn. 5.5 for *non-negative* \mathbf{v}_B requires significantly more computation effort compared to the unconstrained variant. As an alternative, a heuristic adaptation of the SIRT algorithm can be used, where non-negative entries of \mathbf{v}_B are set to 0 after each iteration. This approach was already suggested in [4], and in many cases results in more accurate reconstructions than the unconstrained version. This constrained version of SIRT is denoted by *SIRT-P*.

5.3 Discrete Grey Level Selection algorithm

In this section, the *Discrete Grey Level Selection* approach (DGLS) is introduced for estimating grey levels from projection data. Fig. 5.2 shows a schematic overview of the steps involved in estimating one or more grey levels.

The first part of the procedure requires user interaction. Based on an initial reconstruction, obtained by a conventional, non-discrete, reconstruction algorithm, the user selects a region, hereafter referred to as the *user-selected part (USP)*. This USP is expected to belong to a single grey level in the ground truth image, i.e. it should be constant. The user, who typically has substantial implicit prior knowledge of the particular object under investigation, is responsible for selecting a proper region. In this phase, a significant amount of implicit prior knowledge can be input by the user. For example, if the object is known to contain no holes, it will be safe to select a USP that is clearly within the interior of the object. It can be difficult to select a proper region if the number of projections is small. The selected region should not be too small, but should certainly not contain any pixels that correspond to a different grey level. If the original image contains several grey levels, a region must be selected for each grey level that needs to be estimated. The regions A_t , along with the projection data, now form the input of the estimation algorithm.

After selecting the USP of a single grey level, an optimization algorithm is used

5.3. DISCRETE GREY LEVEL SELECTION ALGORITHM

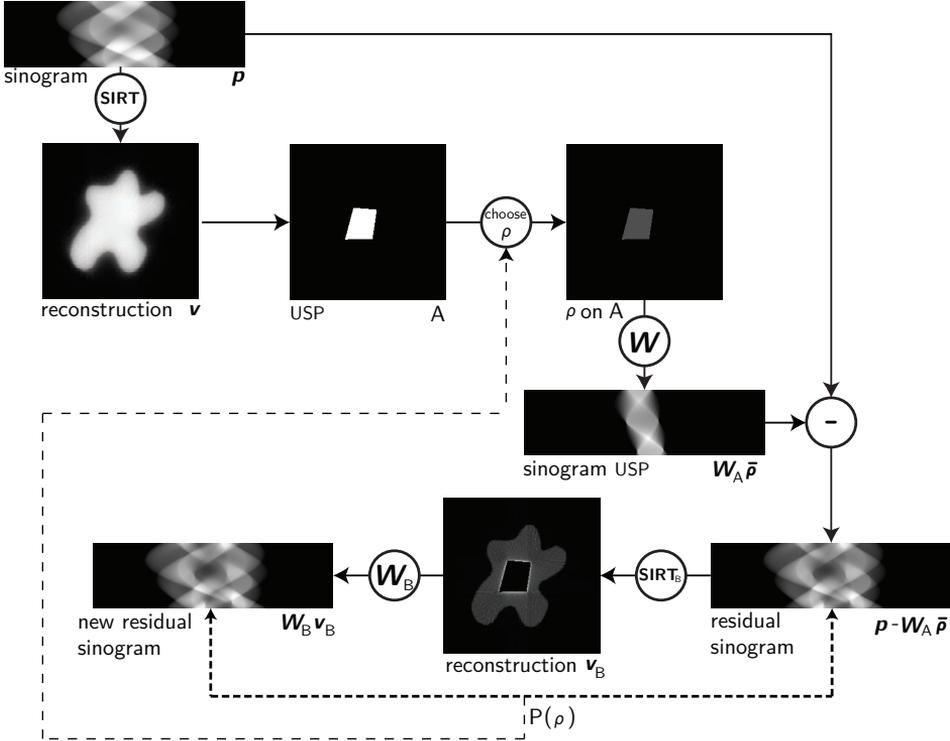


Figure 5.2: Schematic overview of the Discrete Grey Level Selection (DGLS) procedure.

to minimize the grey level penalty function (Eqn. 5.5). To evaluate the penalty function for a given grey level ρ , the sinogram $W_A \bar{\rho}$ of the USP is first subtracted from the complete sinogram p , forming the right-hand side of Eqn. 5.4.

Subsequently, inconsistency of the remaining part of the image with respect to the remaining projection data is determined by evaluating the penalty function (using SIRT or SIRT-P). Based on the value of the penalty function, the grey level estimate is updated iteratively, until the penalty function is minimized. Evaluating the grey level penalty function can be computationally expensive. To find the minimum of this function using a small number of evaluations, Brent's method is used, as described in Chapter 5 of [5]. This optimization strategy can be used for both the SIRT and SIRT-P variants of DGLS and it can be easily extended to incorporate other reconstruction algorithms. Note that for the SIRT-variant of DGLS, the minimization problem to be solved corresponds to Eqn. 5.5 and is in fact a quadratic problem in the unknown grey level ρ and the unknown image v_B . In this case, solving for both ρ and v_B in a combined optimization algorithm, can

yield the same results in much less computation time.

The proposed algorithm estimates a single grey level at a time. If there are multiple grey levels, the USPs for each grey level are selected based on the same initial reconstruction, and the grey level estimations are performed independently for each grey level. A combined approach, that maintains a constant USP for each grey level simultaneously, could also be used. However, this would lead to errors in all estimated grey levels if just one of the USPs is not properly selected. For the experiments in this chapter, focus is put on independent estimation, as it is likely to be more robust in practice.

5.4 Experiments

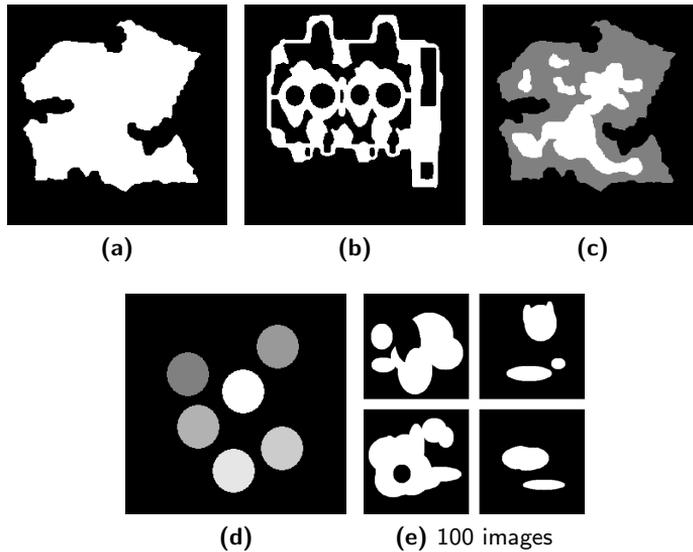


Figure 5.3: Phantom images used for the simulation experiments.

5.4.1 Grey level estimation

Simulation experiments have been performed on multiple 256×256 simulated phantoms. Some were binary images (Fig. 5.3a,b,e), one image had 3 grey levels (Fig. 5.3c), and one and 7 distinct grey levels (Fig. 5.3d). Fig. 5.3e actually represents a family of 100 similar phantom images that were automatically generated. Each contained between 3 and 12 ellipses, randomly generated with radii between

10 and 50. For 20% of these ellipses, the value of the corresponding pixels were set to zero, allowing for the creation of gaps inside ellipses.

To avoid the subjectiveness of an actual user in the experimental results, the USPs for the experiments were automatically generated, using knowledge of the phantom. For each grey level, the USP for that grey level was computed by iteratively applying a binary erosion operation on the phantom region for that grey level, until a certain fraction of the pixels was left. The USP that resulted from this erosion operation was typically similar to a USP selection made by an actual user based on an initial grey level reconstruction.

Using these USPs, the grey levels were estimated with four methods: DGLS with SIRT, DGLS with SIRT-P, the median value (MED) of the USP pixels in a SIRT reconstruction, and the median value of the USP pixels in a SIRT-P reconstruction. The grey level of the background was assumed to be zero and was therefore not estimated. For numerical validation, the absolute difference with respect to the true grey level was computed for each estimation. In case multiple grey levels had to be estimated (such as for phantom image Fig. 5.3c and Fig. 5.3d), the absolute differences were summed. For the family of phantom images in Fig. 5.3e, the average absolute differences and their confidence intervals were computed. The following series of experiments were run.

- The **number of projections** from which the image was reconstructed was varied from 180 down to 15. The projections were equiangularly selected in the interval $[0^\circ, 180^\circ)$. For each grey level, the USP contained 25% of the pixels for that grey level. Fig. 5.4a and Fig. 5.4b show reconstructions of phantom image Fig. 5.3c for 180 down to 15 projection angles, respectively. The results, plotted in Fig. 5.7, show that DGLS for both SIRT and SIRT-P generally yielded much more accurate estimations than the MED estimations. Only when there were very few projection angles, a significant error was visible. This is likely related to the non-uniqueness issues of Proposition 1. In all subsequent experiments, only 30 projection angles were used as it can be observed that the addition of more projection angles did not significantly improve the estimation accuracy.
- The **angular range** of the projections was varied from 180° down to 100° . Fig. 5.4c shows that reducing the angular range had a degrading effect on reconstructions. Fig. 5.8 shows that for both SIRT and SIRT-P, the corresponding DGLS variants yielded far more accurate estimates than the respective MED results.
- The **noise level** of the projections was varied. In this experiment, Poisson noise was applied to the projection data, for a varying source intensity. The

CHAPTER 5. A SEMI-AUTOMATIC ALGORITHM FOR GREY LEVEL ESTIMATION IN TOMOGRAPHY

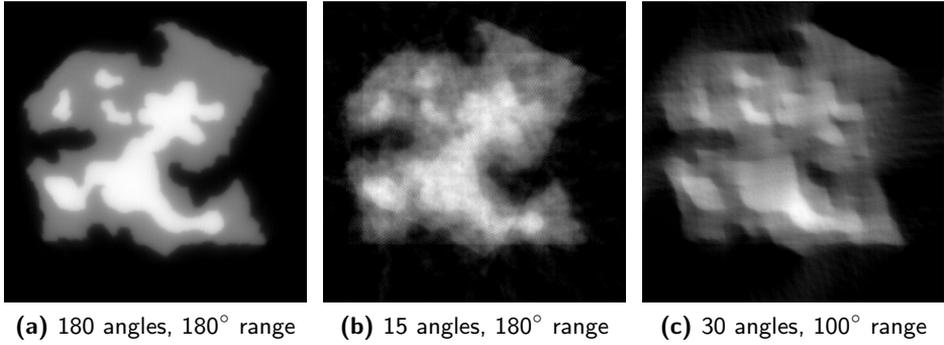


Figure 5.4: SIRT reconstructions of Fig. 5.3c from projection data with a varying angle count and angular range.

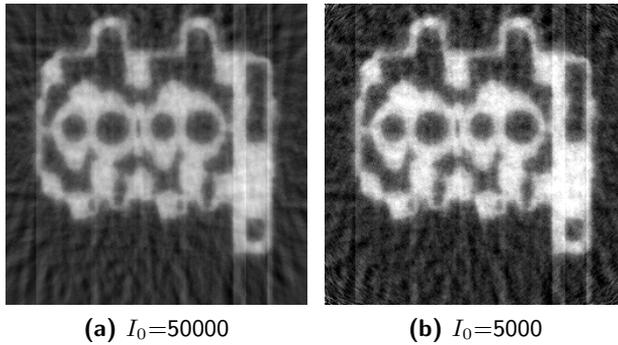


Figure 5.5: SIRT reconstructions of Fig. 5.3b from projection data with varying levels of applied Poisson noise.

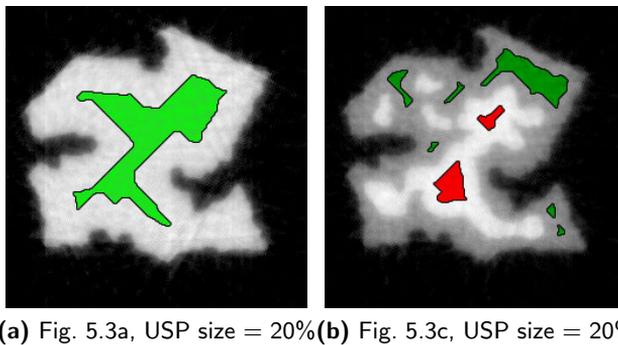


Figure 5.6: The USP as selected for the presented experiments.

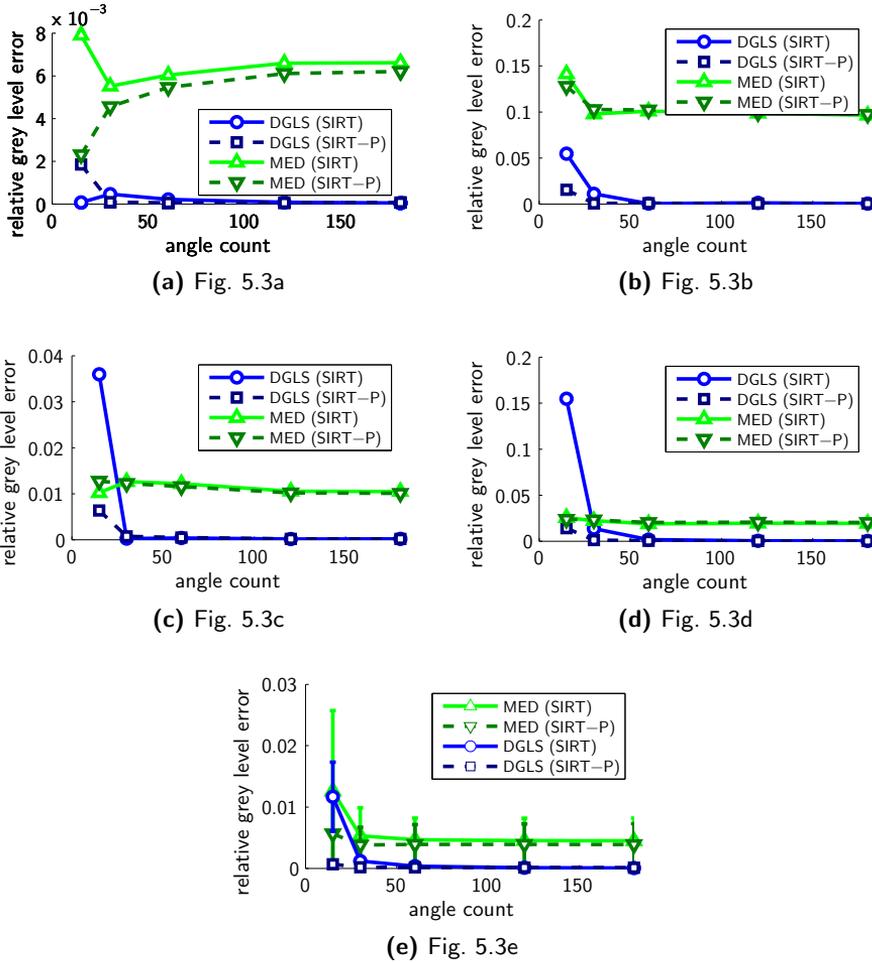


Figure 5.7: The relative grey level estimation error as a function of the angle count.

incident beam intensity, I_0 , refers to the measured detector count when there is no object blocking the path from source to detector. The higher this count, the higher the SNR. The effect of Poisson noise on a reconstruction is visible in Fig. 5.5a and Fig. 5.5b. Fig. 5.9 shows that for the investigated noise range, the noise level did not have a major impact on the estimation error.

- The **size of the USP**, as a percentage of the total region for each grey level, was varied by iteratively applying an erosion filter as discussed above.

CHAPTER 5. A SEMI-AUTOMATIC ALGORITHM FOR GREY LEVEL ESTIMATION IN TOMOGRAPHY

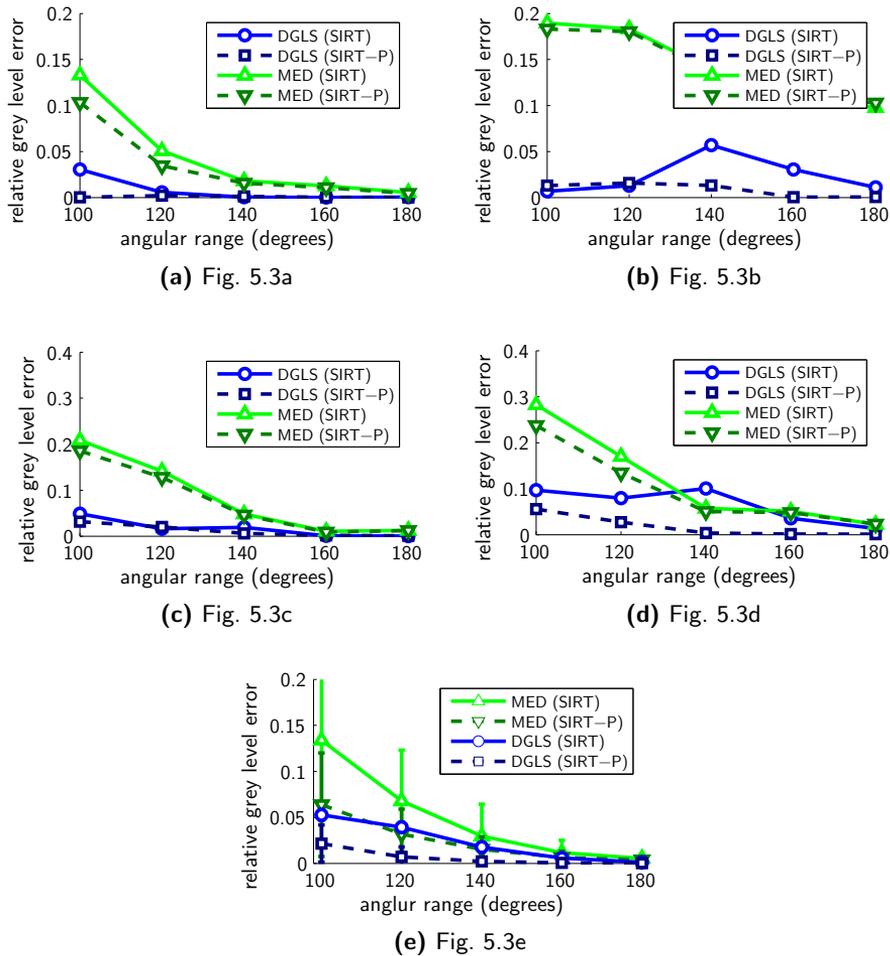
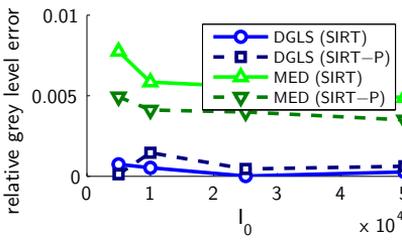


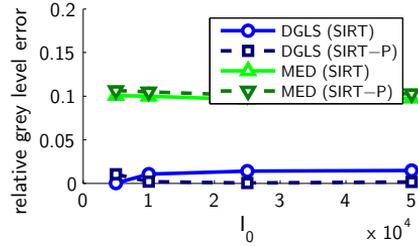
Figure 5.8: The relative grey level estimation error as a function of the projection angle range in degrees.

Fig. 5.10 shows the user-selected part for phantom image Fig. 5.3a and Fig. 5.3c, 20% the size of the original phantom. Fig. 5.10 shows that accurate grey level estimation was possible even if the USP is relatively small, down to 10% of the object size.

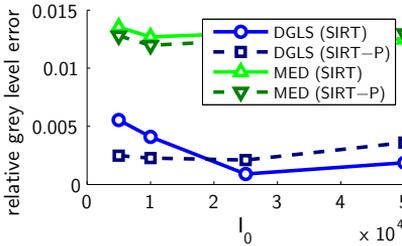
Overall, the results suggest that the DGLS method is robust with respect to the number of projection angles, the angular range, the level of noise and the



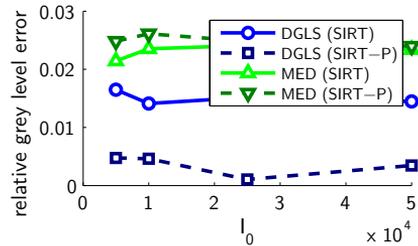
(a) Fig. 5.3a



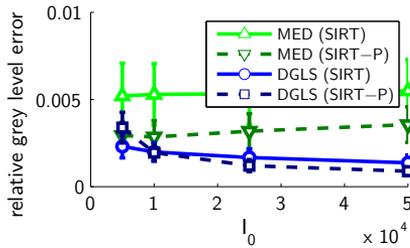
(b) Fig. 5.3b



(c) Fig. 5.3c



(d) Fig. 5.3d



(e) Fig. 5.3e

Figure 5.9: The relative grey level estimation error as a function of I_0 .

size of the user-selected part. Moreover, DGLS typically yielded a significantly more accurate estimate compared to computing the median value of the USP. The results also demonstrated that more accurate estimation can be achieved by using the heuristic SIRT-P method as a scoring function instead of SIRT.

In all experiments described above, the USP for each grey level was selected as a proper subset of the actual regions in the phantom image. In practice, these regions must be selected by the user, based on an initial non-discrete reconstruction

CHAPTER 5. A SEMI-AUTOMATIC ALGORITHM FOR GREY LEVEL ESTIMATION IN TOMOGRAPHY

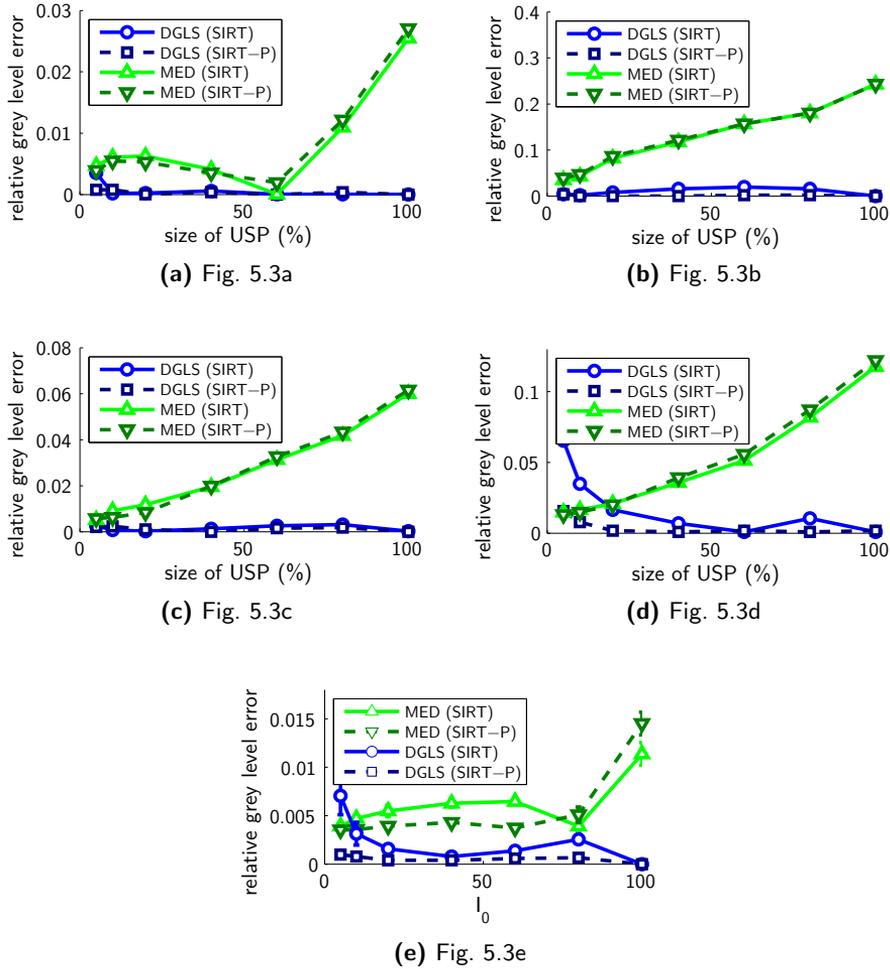


Figure 5.10: The relative grey level estimation error as a function of the size of the user-selected part.

that may be inaccurate and difficult to interpret. If the user selects a region that partially corresponds to a different grey level, the DGLS approach may fail. To investigate this, the following experiment was performed. For phantom image Fig. 5.3a, the USP was selected as a circle of varying radius, centred in a fixed point (see Fig. 5.11a and Fig. 5.11b). The DGLS methods, based on SIRT and on SIRT-P, were then applied based on 30 simulated parallel beam projections. Fig. 5.11c shows the estimated grey level as a function of the radius of the circular

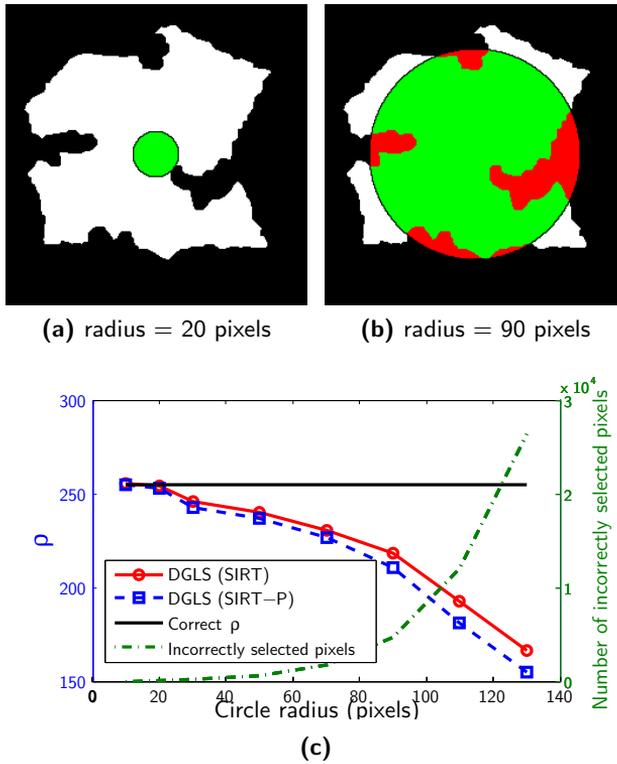


Figure 5.11: (a,b) Circular USP on phantom image Fig. 5.3a. Green pixels in the USP are part of the object, red pixels are not. (c) The accuracy of the DGLS method as a function of the circle radius. Also the corresponding number of incorrectly selected pixels is plotted.

region. It can be observed that the grey level estimate became highly unreliable when a part of the USP extended beyond the actual grey level region. Therefore, it can be concluded that proper selection of the USP is crucial to obtaining reliable grey level estimation.

5.4.2 DGLS + DART

The key motivation for the presented grey level estimation algorithm is that it allows for subsequent application of discrete tomography reconstruction algorithms such as the *Discrete Algebraic Reconstruction Technique (DART)* (Chapter 4). Here, the effect of the DGLS estimation on the DART reconstruction is investigated. Experiments were performed for phantom image Fig. 5.3b, based on 30

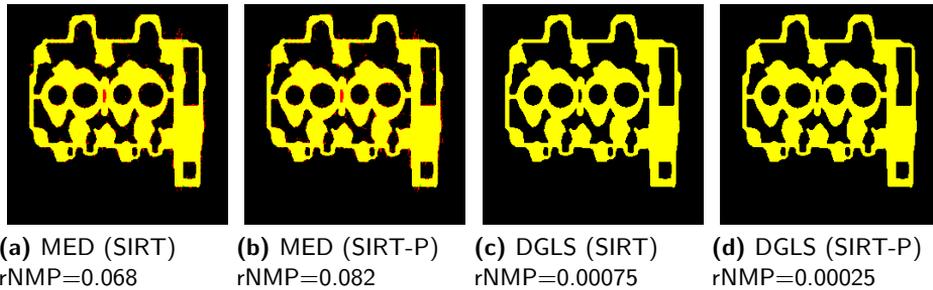


Figure 5.12: DART reconstructions of phantom image Fig. 5.3b with various grey level estimation methods. Yellow pixels are correctly classified, red pixels are erroneously classified to the foreground, green pixels are erroneously classified to the background.

parallel beam projections. After estimating the grey levels with both the MED and DGLS approach, DART reconstructions were performed. Fig. 5.12 shows the misclassified pixels in the DART reconstruction for the various estimation approaches. The overlap between the phantom image and the reconstruction is coloured yellow. The differences are indicated in red (false positive) and green (false negative). The number of misclassified pixels relative to the number of non-background pixels is denoted by $rNMP$ in the figure captions.

Experiments were also performed on a simulated electron tomography experiment using a phantom based on an experimental dataset of a bamboo-like carbon nanotube that was formed around a Copper catalyst particle, similar to the experiments reported in [1]. Fig. 5.13a represents a cross-section of the catalyst nanoparticle that contains both Copper (Cu, white) and Copper-Oxide (CuO, grey), as well as several voids. Projection data was simulated for an angular range of $[-77^\circ, +77^\circ]$. Experimental instability in the alignment of the projections was simulated by shifting each projection by a random distance in the interval $[-1, +1]$. Poisson noise was incorporated in the projection simulation. Fig. 5.13b shows the resulting SIRT reconstruction. To mimic a real experiment, the USP was selected manually (Fig. 5.13c). Fig. 5.13d shows the reconstruction image of Fig. 5.13b segmented with the well-known Otsu’s segmentation method [6]. Fig. 5.13e-i show the segmented DART reconstructions based on perfect prior knowledge of the grey levels (Fig. 5.13e), MED (SIRT) (Fig. 5.13f), MED (SIRT-P) (Fig. 5.13g), DGLS (SIRT) (Fig. 5.13h) and DGLS (SIRT-P) (Fig. 5.13i).

In Fig. 5.14, the $rNMP$ of the different reconstructions is shown alongside the amount of CuO and Cu present in these reconstructions (relative to the phantom image). The DGLS approach resulted in the most accurate segmentation of the object, even slightly better than with the correct grey levels.

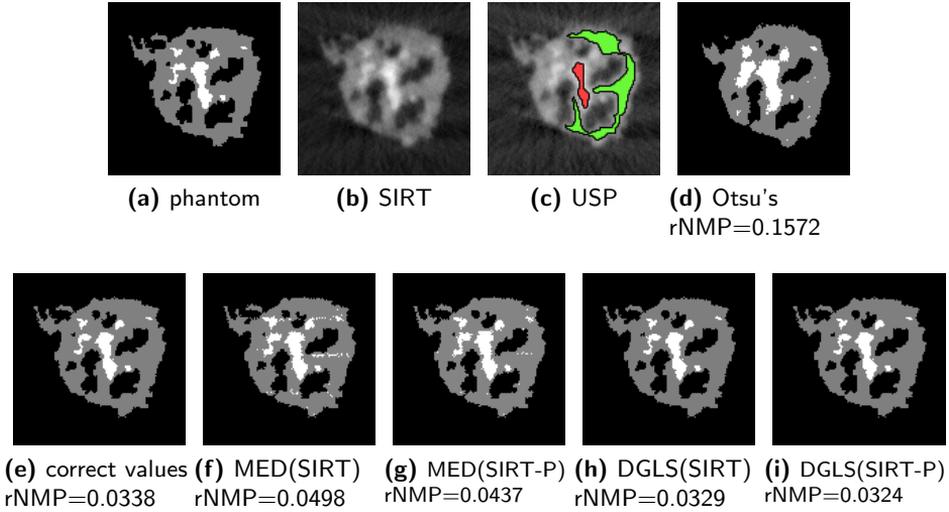


Figure 5.13: DART reconstructions with various grey level estimation techniques.

method	rNMP	CuO	Cu
phantom image	0	100.00%	100.00%
SIRT + Otsu's	0.1572	98.85%	167.22%
correct values + DART	0.0338	100.16%	98.57%
MED(SIRT) + DART	0.0498	99.53%	134.68%
MED(SIRT-P) + DART	0.0437	100.08%	123.04%
DGLS(SIRT) + DART	0.0329	100.34%	97.86%
DGLS(SIRT-P) + DART	0.0324	100.08%	98.82%

Figure 5.14: The rNMP and the amount of the Copper-Oxide (CuO) and Copper (Cu) in the different segmented reconstructions.

5.5 Discussion

Although the experimental results in Section 5.4 have demonstrated that accurate grey level estimation can be achieved by the DGLS approach in a range of scenarios, the grey level estimation problem does not guarantee a unique solution in general. The experimental results show that even for a moderately large number of 15 projections, additional constraints may be necessary to obtain an accurate estimate of the grey levels using our proposed approach. To improve the accuracy of the

resulting estimate, various techniques could be employed.

- **Incorporating additional constraints in the grey level penalty.** The projection distance, as computed using SIRT, does not incorporate non-negativity constraints. Adding a non-negativity constraint to the reconstruction algorithm seems to result in improved accuracy of the grey level estimate, but does not yield the non-negative least square solution, which would be preferable. A scoring function that incorporates both minimization of the projection distance and non-negativity constraints (based on the non-negative least squares problem) could result in more accurate scoring, at the expense of an increase of the running times.
- **Simultaneous estimation of several grey levels.** In the approach presented in this chapter, each of the grey levels is estimated independently. Simultaneous estimation of all grey levels, where the grey level in all user-selected parts is required to be constant, would provide more constraints for the estimation problem. However, this would also make the approach less robust, as an error in the choice of one USP will lead to errors in all estimated grey levels.

Restricting the reconstruction outside the user-selected part to non-negative values still does not capture the full set of available constraints. To incorporate the fact that the entire image should contain only a small, discrete set of grey levels, it seems necessary to incorporate the grey level estimation step *within* a reconstruction algorithm for discrete tomography. Such an approach is presented in Chapter 6.

As demonstrated in Section 5.4.1, proper selection of the USP is crucial to the accuracy of the resulting DGLS estimate, which could be a problem in cases where selecting a proper USP is challenging.

5.6 Conclusions

Grey level estimation is a necessary step before applying discrete tomography algorithms, as these algorithms typically assume the set of admissible grey levels to be known a priori. In this chapter, the DGLS approach was presented for estimating grey levels, where the user first selects regions that are likely to be homogeneous in the original image, and the grey levels are subsequently estimated based on this selection.

The proposed algorithm, which minimizes a penalty function while varying the grey level of the user-selected part, was shown to yield more accurate grey level estimates compared to direct estimation based on a continuous reconstruction. In

particular, the grey level penalty variant that incorporates a positivity constraint on the reconstruction yields accurate estimates even from a small number of projection images, or a small angular range.

References

- [1] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, C. E. A., and G. Van Tendeloo, “3D imaging of nanomaterials by discrete tomography,” *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.
- [2] G. T. Herman and A. Kuba, eds., *Discrete Tomography: Foundations, Algorithms, and Applications*. Applied and Numerical Harmonic Analysis, Birkhuser Boston, 1999.
- [3] J. Gregor and T. Benson, “computational analysis and improvement of SIRT,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 918–924, 2008.
- [4] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections,” *Journal of Theoretical Biology*, vol. 36, pp. 105–117, 1972.
- [5] R. P. Brent, *Algorithms for minimization without derivatives*. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.
- [6] N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, March 1979.

—Random fact: During the course of this work, roughly 650GB worth of data was generated. No animals were harmed however.



Automatic parameter estimation for DART

Contents

6.1	Introduction	135
6.2	PDM-DART	137
6.3	Experiments	140
6.3.1	Comparison of different estimation methods	141
6.3.2	PDM-DART convergence	143
6.3.3	Possible speedup of PDM-DART	144
6.3.4	Limited angles and noisy data	145
6.3.5	Experimental data	148
6.4	Discussion	148
6.5	Conclusions	150
	References	150

At the time of writing, this chapter is submitted for publication:

Wim van Aarle, K. Joost Batenburg, Jan Sijbers, “Automatic parameter estimation for the Discrete Algebraic Reconstruction Technique (DART)”, IEEE Transactions on Image Processing, submitted, 2011.

6.1 Introduction

A key problem when applying DART to experimental datasets is its assumption that the set of grey levels in the unknown reconstructed image is known a priori. While this is easy to satisfy in simulation experiments, obtaining such knowledge in practice is non-trivial.

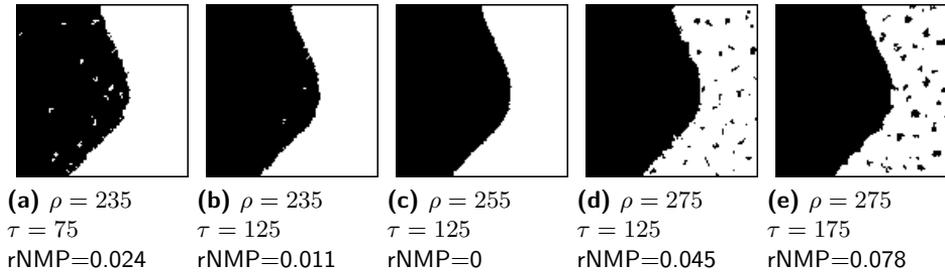


Figure 6.1: Parts of DART reconstructions of Fig. 6.5a for different values of ρ and τ (with 30 equiangular projections). For each (ρ, τ) combination, the relative Number of Misclassified Pixels (rNMP) is given. The rNMP is defined by the number of erroneously classified pixels divided by the amount of non-zero pixels in the phantom image.

The knowledge of the reconstruction’s grey levels is of crucial importance to obtain high quality DART reconstructions. In Fig. 6.1, DART reconstructions are shown of the binary phantom image Fig. 6.5a. Each reconstruction was performed adopting a different grey level for the foreground, ρ . Clearly, a sufficiently accurate reconstruction can be achieved only with a correct estimation of ρ . Note that the estimation of the threshold value, τ , used in the segmentation phase of each DART iteration, also influences the reconstruction accuracy. The combination of grey levels and threshold values are hereafter referred to as the *segmentation parameters* of DART.

To estimate the grey levels of an image composed of only a few grey levels, the semi-automatic *Discrete Grey Level Selection (DGLS)* method was introduced in Chapter 5. For each grey level, the user selects an area in the reconstruction volume that is assumed to be homogeneous. These user-selected parts are set to a candidate grey level, after which their forward projection is subtracted from the measured projection data. The optimal grey levels are those for which this residual sinogram is the least inconsistent. However, using DGLS, the optimal threshold values cannot be estimated and still have to be set manually. Furthermore, it is not always possible to manually select a homogeneous region. For example, in bone or foam-like objects, there are no sufficiently large homogeneous areas.

In Section 4.4, the projection difference was proposed as a cost function to optimize the segmentation parameters of DART. Optimization strategies, such as the Nelder-Mead simplex search [1] or the more advanced adaptive surrogate modelling optimization [2], can then be employed to obtain accurate values (Fig. 6.2c). Given that such an estimation approach requires a full DART reconstruction for each function evaluation in the optimization process, this approach is very computationally inefficient.

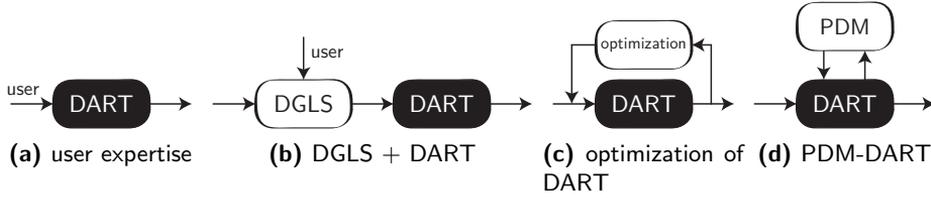


Figure 6.2: Schematic overview of various methods to estimate the segmentation parameters of DART. **(a)** *User* expertise is used to set the correct parameters. **(b)** The DGLS method is used *prior* to DART to semi-automatically estimate the correct grey levels. **(c)** An optimization strategy is used to automatically minimize the projection difference between the DART reconstructed images and the measured projection data. **(d)** The PDM segmentation strategy is used *within* each DART iteration to automatically estimate the optimal grey levels and threshold values.

In this chapter, the DART method is extended with an advanced segmentation algorithm. In the original DART description [3], intermediate reconstructions are segmented using a global thresholding approach with fixed, a priori specified, segmentation parameters. Here, the more flexible segmentation method called *Projection Distance Minimization (PDM)*, as described in Chapter 2, is applied in each iteration of DART. This method is hereafter referred to as *PDM-DART*. In PDM-DART, the threshold values and grey levels are automatically optimized such that the Euclidean distance between the projection of the intermediate segmented image and the measured projection data is minimal. By applying PDM-DART, results are obtained that are comparable to those obtained by conventional DART, without prior knowledge about the segmentation parameters. This eliminates the need for labour intensive user interaction.

This chapter is structured as follows. In Section 6.2 a detailed description of PDM-DART is given. In the subsequent Section 6.3, the proposed method is experimentally validated on simulated as well as experimental data. These results are discussed in Section 6.4. Finally, in Section 6.5, conclusions are drawn.

6.2 PDM-DART

In this section, the DART method is extended with an automatic parameter estimation method. The same notation is used as in Chapter 4.

In the first step of each DART iteration (Section 4.2.3), a segmented image is computed from the current reconstruction $\mathbf{v}^{(k)} \in \mathbb{R}^n$ by application a global thresholding scheme with fixed, priorly specified, grey levels $\boldsymbol{\rho} = (\rho_1, \dots, \rho_l)^T$ and

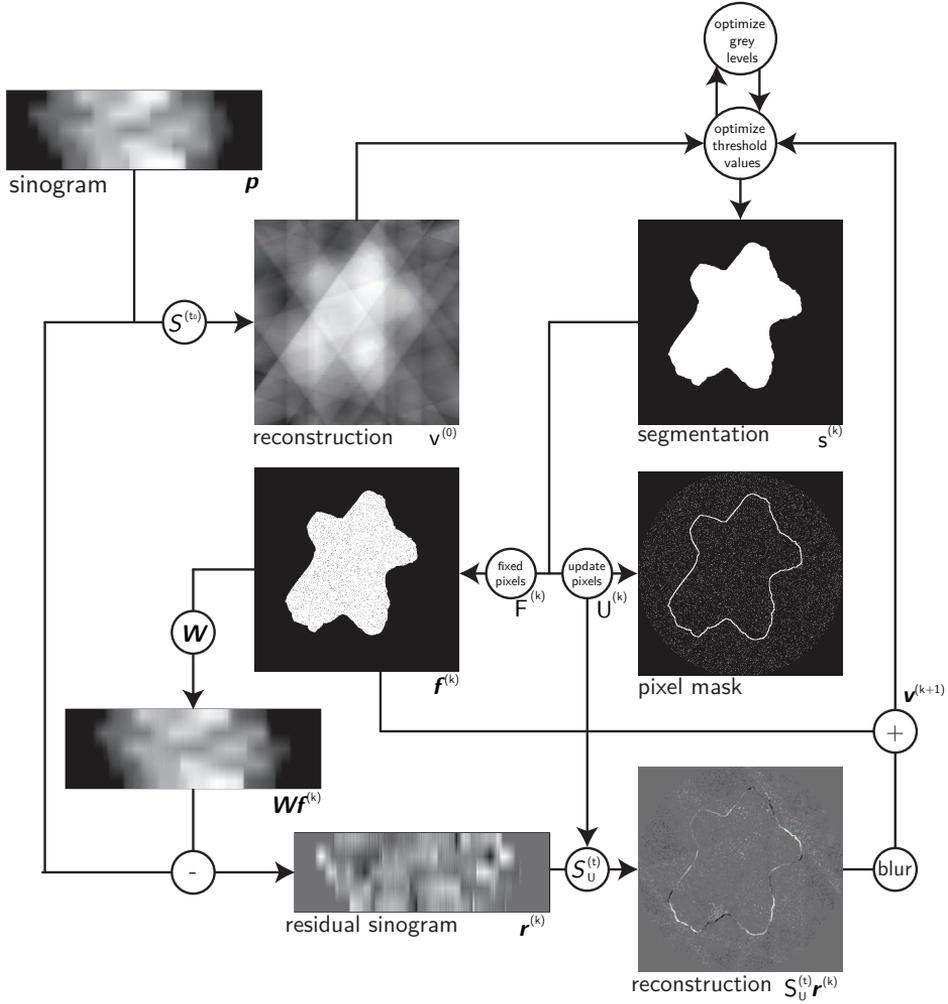


Figure 6.3: Schematic overview of the Projection Distance Minimization Discrete Algebraic Reconstruction Technique (PDM-DART).

threshold values $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{l-1})^T$:

$$s_j^{(k)} = \mathcal{C}_{\boldsymbol{\tau}, \boldsymbol{\rho}}(v_j^{(k)}) = \begin{cases} \rho_1 & v_j^{(k)} < \tau_1 \\ \rho_2 & \tau_1 \leq v_j^{(k)} < \tau_2 \\ \vdots & \\ \rho_l & \tau_{l-1} \leq v_j^{(k)} \end{cases}, j = 1, \dots, n. \quad (6.1)$$

```

v =  $\mathcal{S}^{(t_0)}\mathbf{p}$ ;
 $\tau$  = initial threshold estimate;
while stop criterion is not met
    [ $\tau, \rho$ ] = estimate_tau_rho(v,  $\tau$ );
    s =  $\mathcal{C}_{\tau, \rho}(\mathbf{v})$ ;
    U = {};
    for j = 1:n
        if ( $\exists h \in N(j) : s_j \neq s_h$ ) || rand() < r
            fj = 0;
            U = U  $\cup$  j;
        else
            fj = sj;
        end
    end
    r = p - Wf;
    v = f +  $\mathcal{S}_U^{(t)}\mathbf{r}$ ;
    v = v * Gaussian smoothing filter;
end

function [ $\tau_{opt}, \rho_{opt}$ ] = estimate_tau_rho(v,  $\tau_{init}$ )
    [ $\tau_{opt}, \rho_{opt}$ ] = simplex search of @compute_distance
    with  $\tau_{init}$  as initial estimate;
end

function [distance,  $\rho_{opt}$ ] = compute_distance( $\tau$ )
    compute A;
    compute  $\bar{Q}$  and  $\bar{c}$ ;
     $\rho_{opt}$  = solution of  $2\bar{Q}\rho = -\bar{c}$  w.r.t  $\rho$ ;
    distance =  $\|\mathbf{W}\mathcal{C}_{\tau, \rho_{opt}}(\mathbf{v}) - \mathbf{p}\|_2$ ;
end

```

Figure 6.4: Pseudo code for the Projection Distance Minimization Discrete Algebraic Reconstruction Technique (PDM-DART).

If the chosen values for the segmentation parameters are not correct, $\mathbf{W}\mathbf{f}^{(k)}$ represents an incorrect forward projection of all free pixels, leading to an incorrect residual sinogram, $\mathbf{r}^{(k)}$.

Here, it is proposed to apply the PDM method, as discussed in Chapter 2, to automatically estimate the segmentation parameters *within* each DART iteration:

$$\tau^{(k)}, \rho^{(k)} = \operatorname{argmin}_{\tau, \rho} \|\mathbf{W}\mathbf{s}^{(k)} - \mathbf{p}\|_2. \quad (6.2)$$

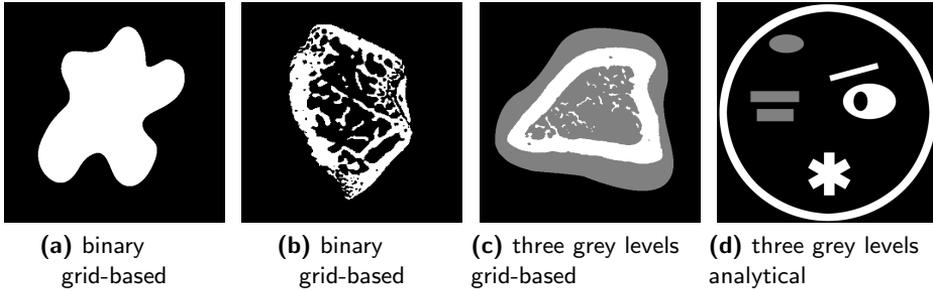


Figure 6.5: Simulated phantoms that are used to validate PDM-DART.

This method is called *Projection Distance Minimization DART (PDM-DART)*. PDM-DART is schematically visualized in Fig. 6.3. In Fig. 6.4, simplified pseudo code is listed.

For low values of k , intermediate reconstructions, $\mathbf{v}^{(k)}$, are typically very inaccurate. The application of PDM is then likely to result in inaccurate segmentation parameters. However, as the experimental results in Section 6.3 demonstrate, even in that case the next intermediate reconstruction, $\mathbf{v}^{(k+1)}$, is very likely to improve. This then makes the subsequently optimized segmentation parameters more accurate, ultimately converging to high quality parameter estimations and reconstructed images.

Note that classical threshold selection techniques, such as Otsu’s clustering methods [4] or k-means clustering [5], are not suited for use within DART. Incorrect grey levels of an intermediate segmentation will negatively affect the selection of new segmentation parameters in the subsequent iteration. In PDM, however, the optimal grey levels are based on the measured projection data, not exclusively on the values obtained in the reconstructions.

The segmentation parameter estimation process does present an increased computational cost to each iteration of the DART algorithm. To reduce this cost, one can opt to perform this optimization only once every few iterations. The optimized segmentation parameters are then stored and reused until the next optimization takes place.

6.3 Experiments

In this section, a series of experiments are described that were carried out to evaluate PDM-DART. Both simulation data and experimental μ CT data are included.

Four simulated phantom images were considered (Fig. 6.5). Whereas Fig. 6.5a

and Fig. 6.5b are binary images, Fig. 6.5c and Fig. 6.5d contain three grey levels. Fig. 6.5a-c were defined on a 512×512 grid while Fig. 6.5d, and its projection data, were analytically defined. All reconstructions were performed on a 512×512 grid. Only parallel beam projection geometries with equiangular projection angles have been considered. However, the proposed method can easily be applied on other projection geometries as well.

To validate the accuracy of the segmented images, the *relative Number of Misclassified Pixels (rNMP)* was computed. This is the total number of erroneously classified pixels with respect to the phantom, divided by the total number of non-zero pixels in the phantom image.

All experiments were performed on a 2.4GHz Intel system, using a single core and 8GB of memory. All tomographic operations, such as DART, SIRT and the forward projection, were accelerated with NVIDIA CUDA and were run on an NVIDIA GeForce GTX285.

6.3.1 Comparison of different estimation methods

In a first set of experiments, projection data was simulated of all phantom images depicted in Fig. 6.5. The number of projection directions was varied and a number of different segmentation parameter estimation methods were applied. That way, for each method, the number of projection angles that is required to compute high quality reconstructions can be compared.

- **Exhaustive parameter search.** The parameters resulting in the most accurate reconstructions were found by performing an exhaustive search. This method can only be applied for simulation experiments, as knowledge of the ground truth is required.
- **Manual selection** (Fig. 6.2a). The grey levels were measured in the phantom images. The threshold values were set in the middle of each pair of consecutive grey levels.
- **DGLS** (Fig. 6.2b). The user-selected regions were manually chosen based on an initial SIRT reconstruction. The threshold values were again set in the middle of each pair of consecutive grey levels. A positivity constraint was included in the DGLS optimization process, as described in [6], as this was observed to result in more accurate estimations.
- **Optimization** (Fig. 6.2c). The Nelder-Mead simplex search optimization strategy was performed on Eqn. 6.2 to estimate all grey levels and threshold values.

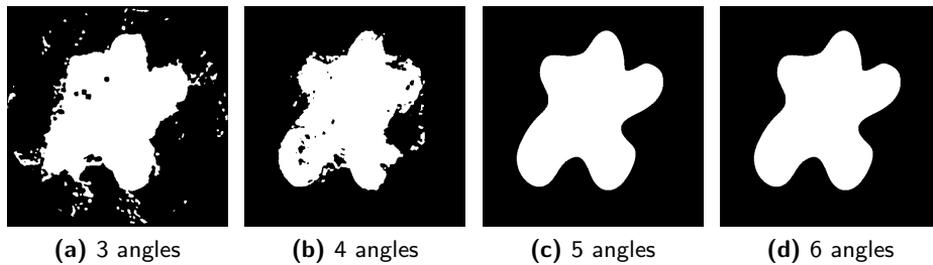


Figure 6.6: PDM-DART reconstructions of Fig. 6.5a from an increasing number of projection directions.

angles	search	manual	DGLS	simplex	PDM-DART
3	0.0947	0.1280	0.2068	0.1208	0.2117
4	0.0575	0.0798	0.1019	0.0911	0.1057
5	0.0007	0.0009	0.0022	0.0014	0.0008
6	0.0006	0.0009	0.0017	0.0009	0.0008
7	0.0004	0.0004	0.0009	0.0012	0.0004
8	0.0003	0.0004	0.0004	0.0003	0.0003
9	0.0002	0.0002	0.0003	0.0003	0.0002
10	0.0001	0.0001	0.0001	0.0001	0.0001

Figure 6.7: The reconstruction accuracy (rNMP) for the phantom in Fig. 6.5a, for a variety of different segmentation parameter estimation methods.

phantom	rNMP goal	search	manual	DGLS	simplex	PDM-DART
Fig. 6.5a	0.001	5	5	5	5	5
Fig. 6.5b	0.050	38	39	44	40	43
Fig. 6.5c	0.010	28	33	36	32	33
Fig. 6.5d	0.100	16	21	34	17	16

Figure 6.8: For each image of Fig. 6.5, the lowest number of projection directions with which it is possible to reach a certain rNMP goal.

- **PDM-DART** (Fig. 6.2d). The PDM-DART algorithm was used to estimate all grey levels and threshold values in each iteration of DART.

In Fig. 6.6a-d, PDM-DART reconstructions of Fig. 6.5a are shown from an increasing number of projection directions. The rNMP of these reconstructions is listed in Fig. 6.7 for all previously mentioned estimation methods. Both manual estimation and PDM-DART provided the best approximation of the optimal values

discovered by the exhaustive parameter search. Simplex search was slightly less accurate. This is likely due to various local optima in the search space. The DGLS approach was also slightly less accurate, especially for very low angle counts.

It can be clearly noted that, for the phantom image in Fig. 6.5a, there exists a certain lower bound to the number of projection angles that is required to create highly accurate reconstructions (5 in this example). For all phantom images and estimation methods, it was determined how many projection directions were needed in order to reach a certain level of accuracy (in rNMP). These results are shown in Fig. 6.8. Overall, PDM-DART did not require substantially more projection directions to reach a certain accuracy than with the optimal segmentation parameters. The same experiments were also performed with ADART [7] instead of DART, achieving similar results.

6.3.2 PDM-DART convergence

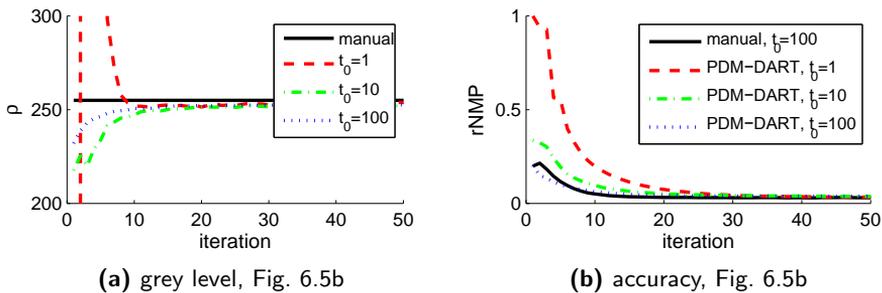


Figure 6.9: To study how PDM-DART performs when initial reconstructions do not allow for accurate tomographic segmentation, the algorithm was run with different values for t_0 , the number of SIRT iterations to create $\mathbf{v}^{(0)}$. **(a)** The estimated grey level of the foreground of phantom image Fig. 6.5b as a function of the iteration number. **(b)** The corresponding accuracy in terms of rNMP.

It may be expected that, if the parameter estimation in PDM-DART is very inaccurate in one iteration, it can still be improved in subsequent iterations.

To investigate the dependence of inaccurate parameter estimation in one iteration on the eventual result, projection data of each phantom image was generated for 30 projections to which Poisson noise was applied. The intensity of this noise is defined by the *incident beam intensity*, I_0 , i.e. the photon count in the incident X-ray beam. Here, I_0 was 50000. PDM-DART reconstructions were performed using different values for t_0 , the number of SIRT iterations that were performed to compute $\mathbf{v}^{(0)}$. For low values of t_0 , the intermediate reconstructions in the

first few iterations are typically severely inaccurate, leading to highly inaccurate segmentation parameters.

In Fig. 6.9a, the estimated grey level of the foreground of phantom image Fig. 6.5b is visualized as a function of the iteration number. The correct grey level is also visualized. It can be seen that, even if the estimated grey level was initially inaccurate, after a few iterations the correct grey level was still adequately approximated. The accuracy of the corresponding reconstructed images is plotted in Fig. 6.9b. Even for very low values of t_0 , the accuracy of the images was comparable to those computed with a manual approach. However, for low values of t_0 , PDM-DART did require more computationally demanding iterations.

6.3.3 Possible speedup of PDM-DART

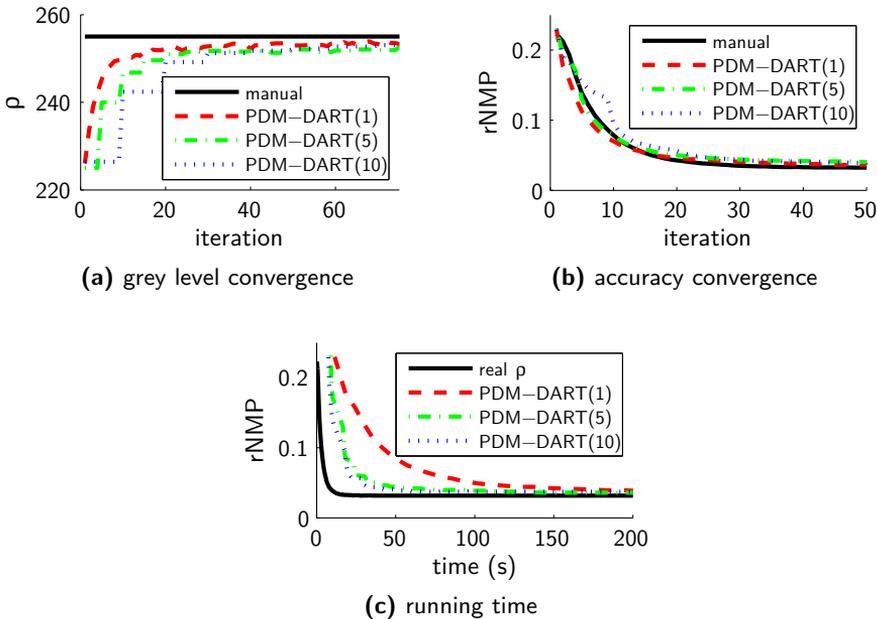


Figure 6.10: (a) The chosen grey level of the foreground of Fig. 6.5b as a function of the iteration number. (b) The corresponding accuracy in terms of rNMP. (c) The total running time of the algorithm as a function of the running time.

In Section 6.2, it was suggested that the additional computational load of PDM-DART over regular DART can be limited by performing the parameter optimization only once per block of iterations. This was experimentally validated on

all phantom images, for a parallel beam projection geometry with 30 projection angles and incident beam intensity $I_0 = 50000$.

In Fig. 6.10a and Fig. 6.10b, it can be seen that for Fig. 6.5b, even if the optimization was performed only once every ten iterations, accurate parameter estimation and reconstruction were still obtained. While the convergence speed towards these results was slower and more iterations were thus demanded to reach a certain level of accuracy, the overall runtime was still substantially decreased. This can be seen in Fig. 6.10c.

6.3.4 Limited angles and noisy data

Next, the effect of limited angles (Fig. 6.11a,c) and noise on the projection data (Fig. 6.11a,b) was studied by performing experiments on all phantom images of Fig. 6.5. The same parameter estimation methods were performed as in Section 6.3.1, with the exception of the exhaustive parameter search. To reduce computation time for phantom images Fig. 6.5c and Fig. 6.5d, only the grey levels were estimated in the simplex search approach. Threshold values were then always put in the middle of two grey levels. For all PDM-DART experiments, parameter estimation was performed only once every 10 iterations.

In Fig. 6.12, the rNMP values of the different parameter estimation schemes are plotted as a function of the number of projection angles used for the reconstruction. In these experiments, the incident beam intensity I_0 was set to 50000. Even in the presence of noise, the PDM-DART algorithm did not require substantially more projection directions than the manual approach with optimal grey levels. DGLS, however, generally required more projection directions to provide accurate grey level estimations. This was especially true for objects for which it was not possible to select large homogeneous parts in an initial reconstruction, an example of which is shown in Fig. 6.11d.

In Fig. 6.13, the rNMP values are plotted as a function of the beam intensity, using 45 projection angles. The achievable accuracy clearly decreased as the noise level increased. If the level of the noise was very high, the simplex search method was often not able to find the global optimum.

In Fig. 6.14, the runtimes of the various estimation methods are plotted for each phantom image of Fig. 6.5 (with $I_0 = 50000$ and 45 projection angles). Given that these runtimes are dependent on the implementation, they only offer a rough overview of the computational efficiency. The fastest reconstructions occurred when the grey levels and threshold values were manually supplied and no further estimation was required. The addition of the DGLS method as an extra preprocessing step slightly increased the reconstruction time. Note, however, that these timings do not take into account the extra time that it took to select the homogeneous part. The PDM-DART scheme required more computation time

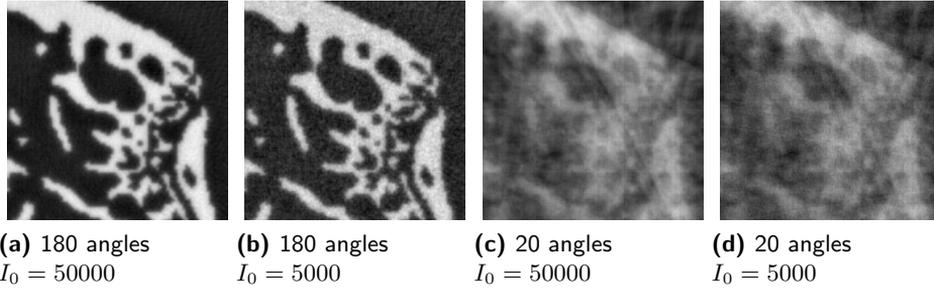


Figure 6.11: Close-up of SIRT reconstructions of Fig. 6.5b from 180 and 20 equiangular parallel beam projection and $I_0 = 50000$ and $I_0 = 5000$. Note how, based on these images, it becomes more difficult to select a homogeneous region as the number of projections decreases.

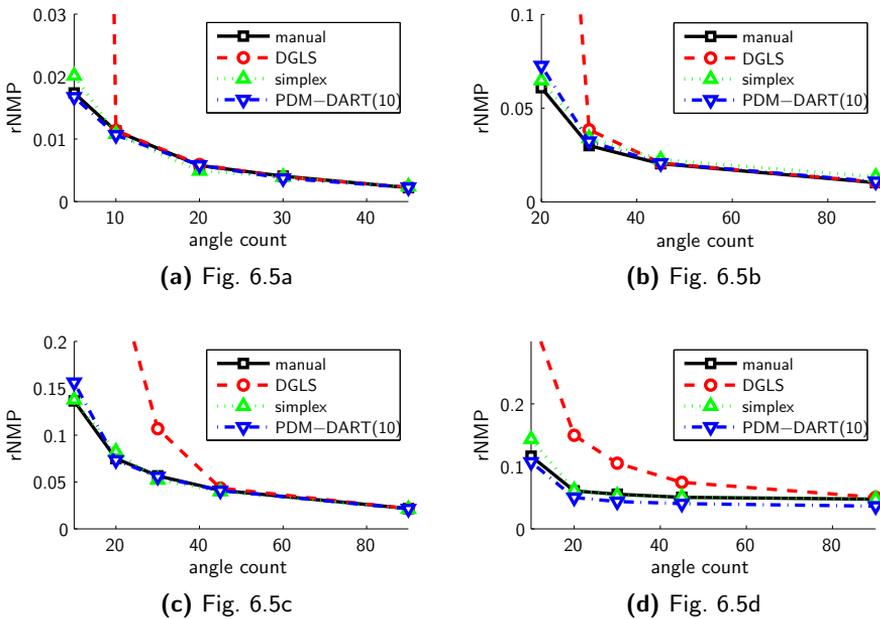


Figure 6.12: The rNMP values for reconstructed images of each phantom image of Fig. 6.5 as a function of the number of projection angles, for four different parameter estimation methods. The projection data was generated for a parallel beam projection geometry and Poisson noise with intensity $I_0 = 50000$ was applied.

6.3. EXPERIMENTS

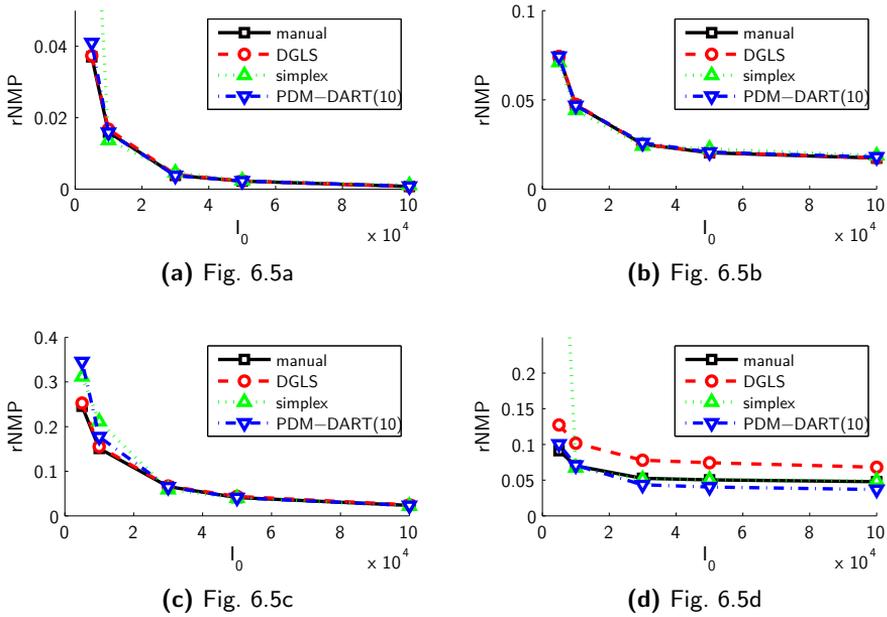


Figure 6.13: The rNMP values for reconstructed images of each phantom image of Fig. 6.5 as a function of the intensity of the applied Poisson noise, for four different parameter estimation methods. The projection data was generated for a parallel beam projection geometry with 45 equiangular projection angles.

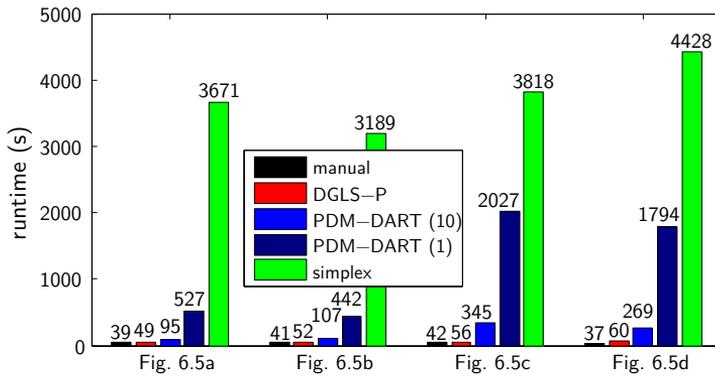


Figure 6.14: The running time of each estimation method for each of the phantom images of Fig. 6.5 with simulated parallel beam projection data with 45 angles and $I_0 = 50000$.

than DGLS, especially for high-dimensional optimization problems (e.g. Fig. 6.5c and Fig. 6.5d). However, the runtime was reduced by excluding the estimation of the segmentation parameters. The simplex search scheme required by far the most computation time.

6.3.5 Experimental data

The proposed method was also applied on experimental μ CT data. Fig. 6.15a shows an FBP reconstruction of a slice through an aluminium foam, which was recorded with a SkyScan 1172 μ CT scanner with 481 equiangular projection angles in the interval $[0, \pi)$. The detector resolution was $9.7\mu\text{m}$. Prior to reconstruction, the data was corrected for ring artefacts and beam hardening with the standard SkyScan NRecon software package.

In Fig. 6.15b, the segmentation of Fig. 6.15a, performed with Otsu's clustering method [4], is shown. This image was used as the ground truth. Next, the number of projection angles in the sinogram was gradually reduced and reconstructions were performed with five different techniques: (1) S-SIRT (a SIRT reconstruction segmented using Otsu's method); (2) a manual approach in which the median value of a user-selected part was used to estimate the grey levels; (3) the DGLS approach; (4) the simplex search method; and (5) PDM-DART. Fig. 6.15c and Fig. 6.15d show the S-SIRT and PDM-DART reconstructions from only 50 projection images respectively. In Fig. 6.15e, all computed rNMP values are plotted as a function of the angle count. In Fig. 6.15f, runtimes are shown. Note that only for a very small number of projections, the slow simplex search method slightly outperforms the PDM-DART approach, which in turn slightly outperforms the DGLS approach.

6.4 Discussion

In the previous section, a series of experiments was described in which the novel PDM-DART approach was compared to other parameter estimation techniques. Each strategy has advantages as well as drawbacks.

- In simulation experiments, **manual estimation** (Fig. 6.2a) generally provides the highest quality reconstructed images with the least amount of time. However, this method is often not feasible in practical applications as various reconstruction artefacts can prevent accurate estimation.
- The **DGLS** technique (Fig. 6.2b) can generate accurate grey level estimations prior to the DART reconstruction and does not present a significant computational overhead. However, it cannot be used to optimally select the threshold values and it can only be applied if the user-based selection of

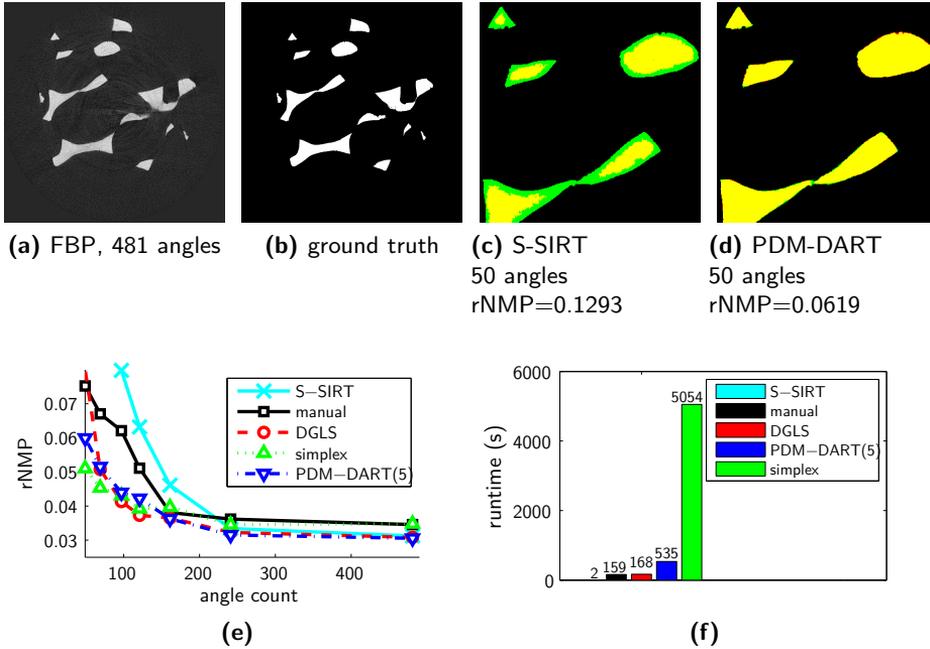


Figure 6.15: Results of the experimental μ CT dataset of an aluminium foam. **(a)** FBP reconstruction with 481 projection angles. **(b)** The segmentation of **(a)** is used as the ground truth image. **(c,d)** Part of PDM-DART reconstructions with 481 and 50 projection angles. The ground truth image is displayed in red and the reconstructions in green. Where both images overlap, i.e. where the segmentation is correct, the corresponding pixel is yellow. **(e)** The rNMP w.r.t to **(b)** of various estimation methods, in function of the number of projection directions.

homogeneous areas is easy. It also tends to be inaccurate if the number of projection angles is very low. It is only *semi-automatic*.

- Optimization with the commonly used Nelder-Mead **simplex search strategy** (Fig. 6.2c) can approximate all optimal segmentation parameters. The search space for this problem may contain many local optima. However, experiments have shown that very accurate estimations can still be made. Furthermore, the method is *fully automatic* and can also be used to optimize other algorithm parameters such as the number of additional random pixels in $U^{(k)}$ in step 2 of each DART iteration or the intensity of the smoothing filter in step 4 of each DART iteration. Estimation with this method is computationally intensive.

REFERENCES

- The *fully automatic PDM-DART* (Fig. 6.2d) generally results in accurate reconstructions even in cases where the DGLS method typically fails. While its computational overhead is substantially lower than the simplex estimation scheme, it is drastically larger than DGLS or manual estimation, especially for high dimensional estimation problems, i.e. if there are many unique grey levels present in the image. This dimensionality can be reduced by removing the grey level of the background, which should always be 0, or by excluding the estimation of the threshold values, fixing them at the middle of two grey levels. With a minimal loss of accuracy, a large speedup can also be achieved if the estimation of segmentation parameters is performed only once per block of DART iterations.

6.5 Conclusions

In this chapter, a novel method for estimating the segmentation parameters within DART was proposed. PDM-DART combines discrete reconstruction with estimation of segmentation parameters. In contrast to the original DART algorithm, where these parameters remain fixed throughout the entire reconstruction process, PDM-DART adaptively selects the optimal segmentation parameters *within* each DART iteration.

Experiments have been performed on a range of different simulation images as well as on experiment μ CT data. They have confirmed that with the PDM-DART approach, high quality reconstructions can be made even for a low number of projection angles. Furthermore, it was demonstrated that the use of PDM-DART does not require more projection directions than other estimation methods. This paves the way for fully automatic DT in a range of application areas in medicine, industry and science, where discrete tomography has thus far remained a highly labour intensive procedure.

References

- [1] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the nelder-mead simplex method in low dimensions,” *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [2] D. Gorissen, K. Crombecq, I. Couckuyt, T. Dhaene, and P. Demeester, “A surrogate modeling and adaptive sampling toolbox for computer based design,” *Journal of Machine Learning Research*, vol. 11, pp. 2051–2055, 2010.
- [3] K. J. Batenburg and J. Sijbers, “DART: A practical reconstruction algorithm for dis-

- crete tomography,” *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2542–2553, 2011.
- [4] N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, March 1979.
- [5] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithm: analysis and implementation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, July 2002.
- [6] K. J. Batenburg, W. van Aarle, and J. Sijbers, “A semi-automatic algorithm for grey level estimation in tomography,” *Pattern Recognition Letters*, vol. 32, no. 9, pp. 1395–1405, 2011.
- [7] F. J. Maestre-Deusto, G. Scavello, J. Pizarro, and P. L. Galindo, “Adart: An adaptive algebraic reconstruction algorithm for discrete tomography,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2146–2152, 2011.

—I was standing in the park the other day, wondering why a Frisbee appears larger the closer it gets. Then it hit me.

Stewart Francis, Live at the Apollo, 2010

7

Super-resolution for computed tomography based on discrete tomography

Contents

7.1	Introduction	154
7.2	Notation	156
	7.2.1 Partial volume effect	156
	7.2.2 Algebraic tomography	157
7.3	Super-resolution	158
	7.3.1 Sinogram upsampling	158
	7.3.2 Detector supersampling	159
	7.3.3 Limited data problem	160
7.4	Experiments	161
	7.4.1 Simulation studies	161
	7.4.2 Experimental studies	167
7.5	Discussion and conclusions	168
	References	169

At the time of writing, this chapter is being prepared for publication:

Wim van Aarle, K. Joost Batenburg, Gert Van Gompel, Elke Van de Casteele, Jan Sijbers, “*Super-resolution for computed tomography based on discrete tomography*”, Journal of X-Ray Science and Technology, in preparation, 2012.

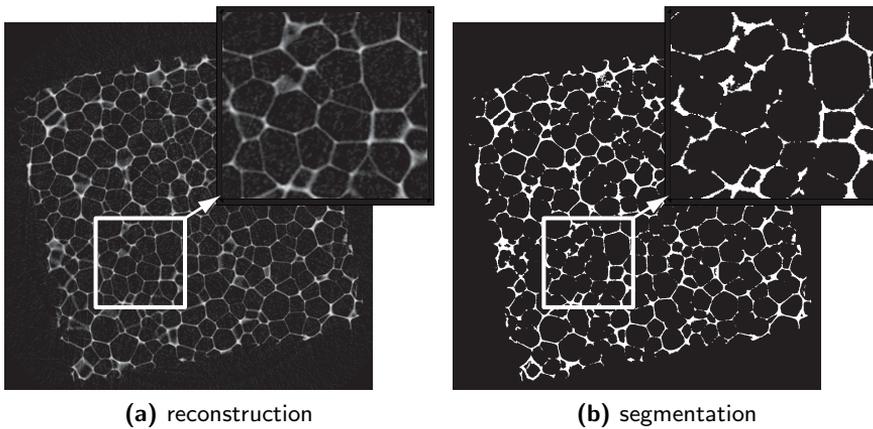


Figure 7.1: (a) Reconstruction of a polyurethane foam, scanned in a SkyScan 1172 μ CT scanner at a pixel resolution of $17\mu\text{m}$. (b) Otsu's segmentation of the reconstruction. Many cell walls remain undetected in the segmentation while other structures are overestimated.

7.1 Introduction

In X-ray computed tomography, images are typically reconstructed on a voxel grid. Given that each voxel is represented by a constant grey level, it is typically assumed that the material within such a voxel is homogeneous. It is clear, however, that a voxel representation cannot properly represent structures that have a varying density within a voxel. Thus, each voxel in the images could contain more than one material or tissue type. This phenomenon is referred to as the *partial volume effect (PVE)*. PVEs cause object boundaries to be smeared out across the boundary voxels. Also, if a feature of the scanned object is small relative to the nominal voxel size, PVEs reduce the contrast between the structure of interest and its background signal. Consequently, it is difficult to achieve the intrinsic resolution of the detector. Fig. 7.1a shows an FBP reconstruction of a polyurethane foam for which the widths of the edges of the pores are comparable to the detector size. A globally thresholded segmentation of Fig. 7.1a, created with the commonly used clustering method of Otsu, is shown in Fig. 7.1b. Clearly, many thin structures remain undetected, whereas the thickness for some larger structures is overestimated.

To reduce PVEs — and hence to obtain sufficient contrast — a high resolution scan can be acquired. This, however, requires a high radiation dose and a long scanning time [1]. In Fig. 7.2, an FBP reconstruction with a spatial resolution of $35\mu\text{m}$ is shown of a rat femur along with an FBP reconstruction with a spatial

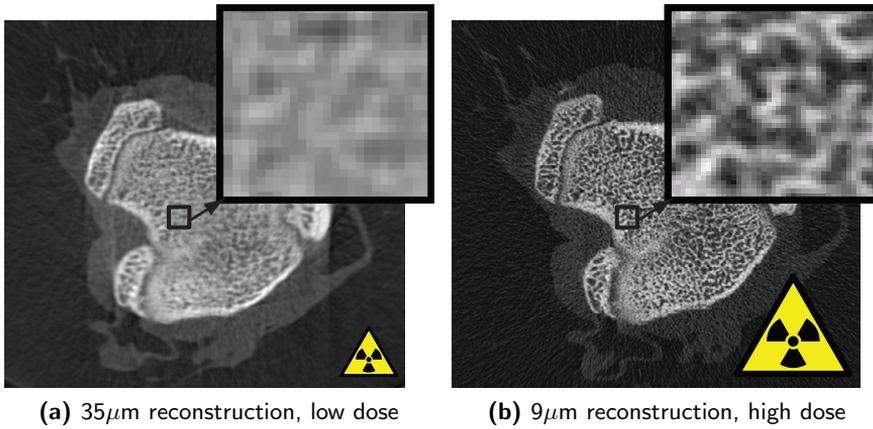


Figure 7.2: FBP reconstructions of the epiphyseal plate of a rat femur taken at two different resolutions in a SkyScan 1172 μ CT scanner. The high dose reconstruction (b) is clearly much easier to segment. Note that as both slices were taken from different scans, the object was slightly displaced between the acquisition of both datasets. Even though image registration was performed, there is still a residual difference.

resolution of $9\mu\text{m}$ of the same femur. It is clear that the contrast in Fig. 7.2b is significantly better than that in Fig. 7.2a. Fig. 7.2b is therefore much better suited for accurate segmentation and estimation of the morphometric bone parameters [2].

The conventional approach to reduce PVEs without increasing the X-ray dose is to upsample the reconstruction voxel grid, allowing for a more accurate representation and potentially improving the overall visualization of small structures. This upsampling is also known as *super-resolution* [3]. Conventional reconstruction upsampling typically results in a limited data reconstruction problem: the number of equations (measured projection data) remains the same while the number of unknowns (reconstruction voxels) increases significantly. This problem can be overcome by combining the information from multiple low resolution CT images into a high resolution CT image [4, 5, 6], but this would again result in an increased scan time and radiation dose.

In this chapter, a super-resolution approach for CT is proposed that effectively solves the limited data problem by incorporating some form of prior knowledge about the unknown object. In CT, such prior knowledge comes in many forms, e.g. sparsity of the reconstructed image [7] or its gradient [8, 9]. Here, the novel super-resolution scheme is based on the *Discrete Algebraic Reconstruction Technique (DART)*, as presented in Chapter 4. It is shown that by upsampling the

CHAPTER 7. SUPER-RESOLUTION FOR COMPUTED TOMOGRAPHY BASED ON DISCRETE TOMOGRAPHY

reconstruction grid and incorporating available prior knowledge to compensate for the lack of high resolution projection data, the proposed approach effectively increases the spatial resolution of the tomographic reconstructions [10, 11].

This chapter is organized as follows. Section 7.2 introduces a model for the PVE. The new super-resolution approach is subsequently introduced in Section 7.3. In Section 7.4, experiments are described that were performed to evaluate the reconstruction accuracy for the proposed super-resolution approach. Results are presented for both simulated data and experimental μ CT data. Finally, Section 7.5 concludes this chapter.

7.2 Notation

In this section, relevant tomographic notation is introduced and a tomographic model for the PVE is described. For simplification, a monochromatic X-ray beam will be assumed. Note, however, that this does not preclude application of the method to polychromatic X-ray imaging since preprocessing methods can be applied to compute monochromatic from polychromatic projections [12, 13, 14, 15].

For clarity, all concepts will be presented on a two-dimensional parallel beam projection geometry, but the proposed methods can be easily generalized to any acquisition geometry.

7.2.1 Partial volume effect

Let $f(x, y) \in \mathbb{R} \times \mathbb{R}$ represent the two-dimensional object function of a certain object. A parallel beam projection geometry defines the tomographic projection of f as the line integrals of f along the lines $l_{\theta, t} = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \cos \theta + y \sin \theta = t\}$, where $\theta \in [0, \pi)$ represents the angle between the line and the y -axis and where $t \in \mathbb{R}$ represents the coordinate along the projection axis. For a finite set of lines $l_{\theta, t}$, the X-ray beam intensity at the detectors, $I(\theta, t)$, is measured by:

$$I(\theta, t) = I_0 e^{-\int_{l_{\theta, t}} f(x, y) ds}, \quad (7.1)$$

with I_0 the incident beam intensity. Define the *attenuation projection function* or *sinogram* $p(\theta, t)$ as follows:

$$p(\theta, t) = -\ln \left(\frac{I(\theta, t)}{I_0} \right) = \int_{l_{\theta, t}} f(x, y) ds. \quad (7.2)$$

In practice, a projection is measured at a set of projection angles and at a set of detector elements with a width Δt . Let $\mathbf{I} = (I_i) \in \mathbb{R}^m$ denote the measured intensity data, with m the number of detector values multiplied by the number of

projection angles. For $i \in 1, \dots, m$, I_i can then be modelled as:

$$I_i = \int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} I_0 e^{-p(\theta_i, t_i + t')} dt', \quad (7.3)$$

with t_i and θ_i the detector coordinate and projection angle of the measured detector value I_i , respectively. The attenuation projection data $\mathbf{p} = (p_i) \in \mathbb{R}^m$ can then be defined as follows:

$$p_i = -\ln \left(\frac{I_i}{I_0} \right). \quad (7.4)$$

Note that, due to the logarithmic operation in Eqn. 7.4, the contribution of a pixel to the measured projection values does not only depend on the average value of that pixel, but also on the *distribution* of the attenuation *within that pixel*.

Tomography deals with the reconstruction of $f(x, y)$ based on \mathbf{p} . This reconstructed function is represented by an image, a grid of square pixels with a finite width and height, Δs . Let $\mathbf{v} = (v_j) \in \mathbb{R}^n$ denote a discretized square image of the function $f(x, y)$, where n denotes the number of pixels. The image value v_j can then be modelled as the total value of f , taken over the square pixel:

$$v_j = \frac{1}{\Delta s^2} \int_{-\frac{\Delta s}{2}}^{\frac{\Delta s}{2}} \int_{-\frac{\Delta s}{2}}^{\frac{\Delta s}{2}} f(x_j + x', y_j + y') dx' dy', \quad (7.5)$$

with x_j and y_j the coordinates of the centre point of pixel v_j .

The value of a certain pixel v_j thus depends on an entire area of values of the real object function. If the object has an edge running through the area of pixel v_j , or if the object is not homogeneous inside the pixel boundaries, the value of v_j does not represent the attenuation coefficient of any of the materials of the object, but instead represents an average of all attenuation coefficients. This is called the *partial volume effect (PVE)*. Note that for object functions that consist of piece-wise constant regions, the fraction of pixels for which PVEs occur is directly related to the size of Δs .

7.2.2 Algebraic tomography

Using the discretized definitions of projection data (Eqn. 7.4) and reconstructed image (Eqn. 7.5), a computational model — approximating the mathematical projection model — can be constructed. The forward projection of the object for a finite set of angles is modelled as a linear operator \mathbf{W} , called the *projection operator*, which maps the image \mathbf{v} to the projection data \mathbf{q} :

$$\mathbf{q} := \mathbf{W}\mathbf{v}. \quad (7.6)$$

CHAPTER 7. SUPER-RESOLUTION FOR COMPUTED TOMOGRAPHY BASED ON DISCRETE TOMOGRAPHY

In Eqn. 7.6, $\mathbf{W} = (w_{ij})$ is an $m \times n$ matrix where w_{ij} represents the contribution of image pixel j to detector value i . The vector \mathbf{q} is called the *forward projection* of \mathbf{v} . The reconstruction problem in CT can then be modelled as the recovery of \mathbf{v} from a given vector \mathbf{p} of projection data, such that:

$$\mathbf{W}\mathbf{v} = \mathbf{p}. \quad (7.7)$$

Throughout this chapter, SIRT (Section 1.3.3.1) is used to solve Eqn. 7.7 without any constraints on \mathbf{v} .

If m is much smaller than n (e.g. when the number of projection directions is very low or the data is truncated), Eqn. 7.7 is an underdetermined system of linear equations, leading to a *limited data reconstruction problem*. As was demonstrated in Chapter 4, DART is very suited for solving such problems. It exploits knowledge about the discrete grey levels to iteratively solve Eqn. 7.7 under the constraint that v_j can only take values that are elements of a vector $\boldsymbol{\rho} = (\rho_1, \dots, \rho_l)^T$.

7.3 Super-resolution

To counter the PVE, the reconstruction grid can be upsampled (Fig. 7.3b). Let a be the upsampling factor in each dimension. Each pixel of width Δs is then subdivided into a^2 pixels of width $\frac{\Delta s}{a}$. Denote the upsampled reconstruction image by $\mathbf{v}' \in \mathbb{R}^{a^2 n}$.

Note that, typically, $\frac{\Delta s}{a}$ is different from Δt , the width of the detector cell. If the projection weights w_{ij} are computed by intersection of a single ray with the upsampled image (Fig. 7.3a), some pixels will not have a ray going through them for each projection angle and the projection data will not be computed correctly. Two methods are investigated to overcome this problem: *sinogram upsampling* (Fig. 7.3c, Section 7.3.1) and *detector supersampling* (Fig. 7.3d, Section 7.3.2).

7.3.1 Sinogram upsampling

With *sinogram upsampling* (SU) — sometimes also referred to by *sinogram stretching* [16] —, the number of detector cells is artificially increased by subdividing each detector of size Δt into a detectors of size $\frac{\Delta t}{a}$. Fig. 7.3c shows a schematic overview of this geometry. The value of each detector point is determined by one-dimensional linear interpolation of \mathbf{p} . Note that this interpolation assumes a certain smoothness in the projection data.

Let $\mathbf{p}' \in \mathbb{R}^{am}$ be the upsampled sinogram and let $\mathbf{W}_{SU} \in \mathbb{R}^{am \times a^2 n}$ be the corresponding projection operator. The reconstruction equation then becomes:

$$\mathbf{W}_{SU}\mathbf{v}' = \mathbf{p}'. \quad (7.8)$$

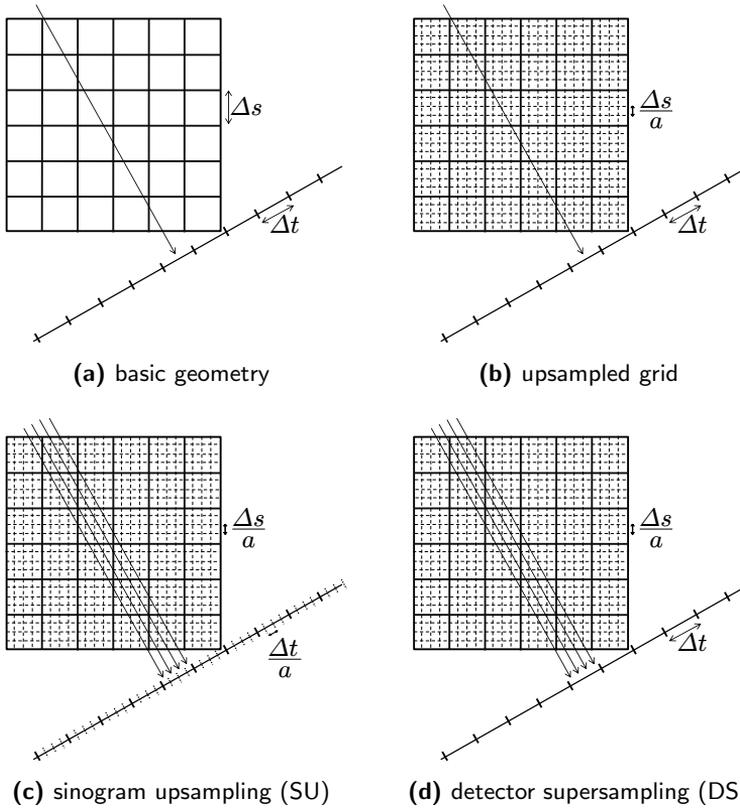


Figure 7.3: (a) Basic projection geometry. Each detector cell corresponds to a single ray. (b) Upsampled reconstruction grid. Certain pixels are not hit by a ray. (c) Sinogram upsampling, each detector is subdivided into multiple detectors with interpolated values. (d) Detector supersampling, each detector corresponds to multiple rays. The contribution of each ray is summed.

7.3.2 Detector supersampling

With *detector supersampling (DS)*, the sinogram \mathbf{p} remains unaltered. However, the number of virtual rays targeting each detector cell is increased by a factor a , each $\frac{\Delta t}{a}$ apart. Fig. 7.3d shows a schematic overview of this geometry. The reconstruction equation is:

$$\mathbf{W}_{DS}\mathbf{v}' = \mathbf{p}, \tag{7.9}$$

where each row in the projection operator $\mathbf{W}_{DS} \in \mathbb{R}^{m \times a^2 n}$ is the summation of the a corresponding rows of \mathbf{W}_{SU} .

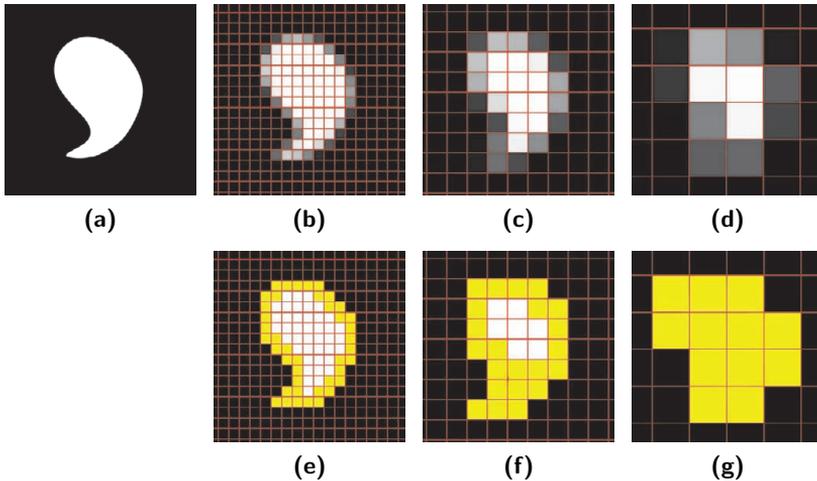


Figure 7.4: (a) An object that cannot be correctly represented on a pixel grid (b-d) The representation of (a) on reconstruction grids with various pixel sizes. The red lines denote the pixel borders. (e-g) Highlighted in yellow are the pixels of the reconstruction grids of (b-d) that would be updated during the a DART iteration.

7.3.3 Limited data problem

In Eqn. 7.8 and Eqn. 7.9, the number of unknowns has been increased by a factor a^2 while the number of equations has been increased by a factor a (Eqn. 7.8) and remained unaltered (Eqn. 7.9), respectively. Solving the reconstruction equation is then a limited data problem. As was noted in Chapter 4, prior knowledge about the scanned objects can be used to solve Eqn. 7.8 and Eqn. 7.9 by using DART.

By design, DART is especially suited for structures that are large with respect to Δs . If the object to be reconstructed consists of many small structures, such as foams or trabecular bone, the PVE breaks the assumption of DART that the object can be represented with a constant grey level. This is visualized in Fig. 7.4. Fig. 7.4b-d shows the PVE on the structure visible in Fig. 7.4a with three different pixel sizes. Furthermore, for a large Δs , the number of pixels that is updated in the subsequent DART iteration (referred to as the set A) is still large, limiting the possible improvements of that DART iteration. In Fig. 7.4e-g, the boundary pixels of the object are shown. In Fig. 7.4g, the pixels in A even cover the entire structure, this might lead to it not being segmented in step 3 of DART.

Note that there is a non-linear relationship between the measured projection data \mathbf{I} and the actual attenuation projection data \mathbf{p} (Eqn. 7.3 and Eqn. 7.4). As DART uses a linear projection model, the proposed super-resolution approaches do not accurately model the PVE. In the next section, however, it will be ex-

perimentally demonstrated that even with this limited model, super-resolution on piecewise homogeneous objects with known attenuation coefficients can indeed be achieved, leading to significant improvements in reconstruction accuracy.

7.4 Experiments

In this section, the proposed super-resolution method is demonstrated and its effectiveness is evaluated on various simulated images (Section 7.4.1) and on experimental datasets (Section 7.4.2).

7.4.1 Simulation studies

Experiments were performed on five simulation phantoms (Fig. 7.5a and Fig. 7.6). Some phantoms were generated analytically (Fig. 7.5a and Fig. 7.6c), while other phantoms were generated based on high resolution rasterized images.

7.4.1.1 Analytical rings phantom

In the first experiment, the efficacy of a discrete super-resolution technique was examined as a function of the size of a structure with respect to Δt . To this end, a simulated analytical binary phantom containing 11 rings with a varying width, Δq , was created. A rasterized rendering of this phantom is depicted in Fig. 7.5a. As a measure of the magnitude of the PVE, the notion of *relative projection resolution*, R_p , is introduced. It is defined as the ratio of the object width, i.e the thickness of the ring, to the detector width:

$$R_p = \frac{\Delta q}{\Delta t}. \quad (7.10)$$

For the phantom in Fig. 7.5a, the R_p of the outer three rings was 10, 5 and 3. The R_p of the fourth ring was 1 and could thus be used to measure if the intrinsic detector resolution was achieved. The seven most inner rings had an R_p of $\frac{1}{2}$ to $\frac{1}{8}$.

Projection data was analytically generated (using Eqn. 7.4) for a parallel beam geometry with 60 equiangular projection angles and 256 detector pixels. Reconstructions were computed for both S-SIRT (a SIRT reconstruction segmented with Otsu's method) and DART and with both the sinogram upsampling approach and the detector supersampling approach, with increasing levels of super-resolution: $a = 1, 2, 4$ and 10.

As indicated by the appearance of the rings in the center in Fig. 7.5b-d, it is clear that by increasing a (combined with DS and DART), the spatial resolution drastically improved. This effect was less pronounced if sinogram upsampling was

CHAPTER 7. SUPER-RESOLUTION FOR COMPUTED TOMOGRAPHY
 BASED ON DISCRETE TOMOGRAPHY

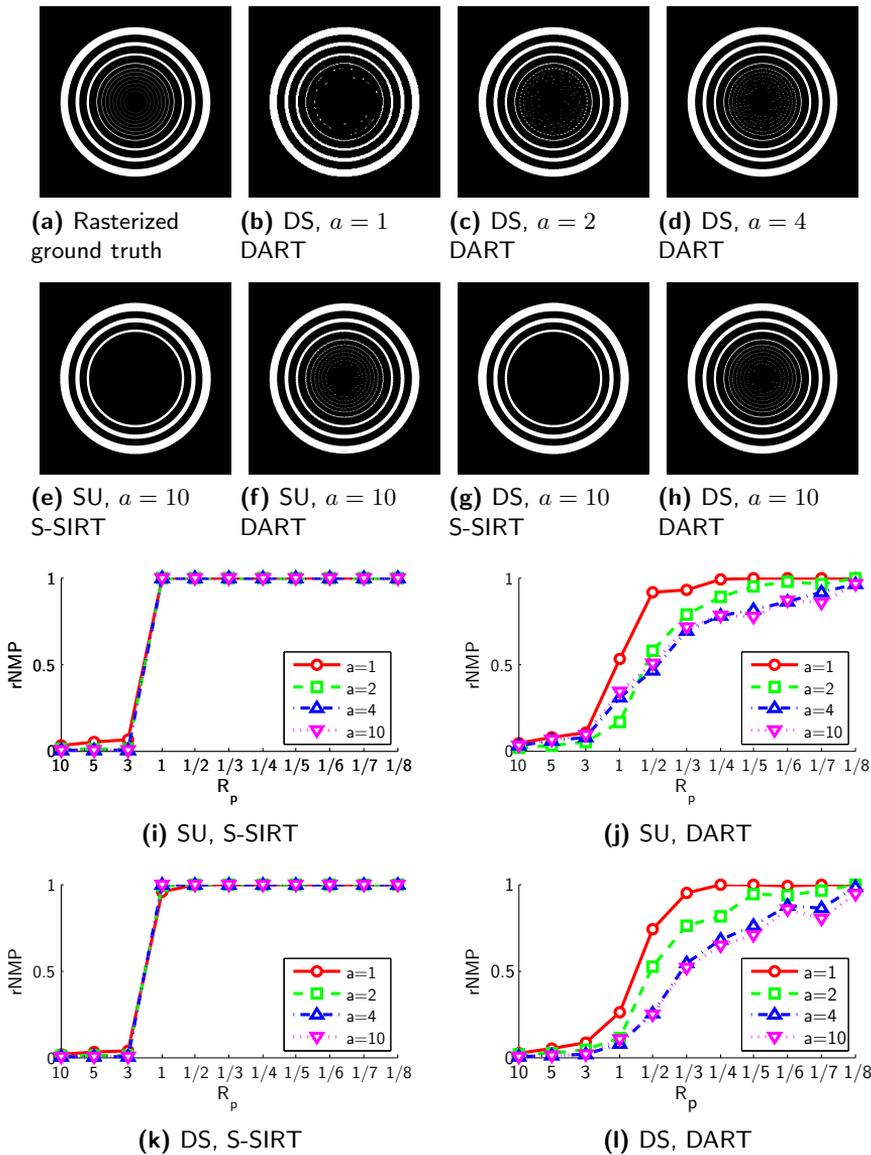


Figure 7.5: Experimental results from a simulated analytical phantom containing 11 rings of varying width. **(a)** High resolution rendering of the phantom image, used as ground truth image. **(b-d)** The inner rings become more visible as a increases. **(e-h)** Reconstructions of each proposed super-resolution approach, with (DART) and without (S-SIRT) prior knowledge. **(i-l)** The rNMP as a function of the widths of each ring, R_p .

used (Fig. 7.5f). Furthermore, the thin rings could not be seen at all if no prior knowledge was included in the reconstruction (Fig. 7.5e and Fig. 7.5g). These results can also be observed in Fig. 7.5i-l, where for each a the *relative Number of Misclassified Pixels (rNMP)* of each ring is plotted. The rNMP measures the total number of pixels that are classified in a wrong partition (false negatives as well as false positives) with respect to the total number of pixels of that object. For analytical phantom images, this rNMP value was approximated by comparing the reconstruction to a very high resolution rasterization of the phantom image. For this experiment, the rNMP was computed for each ring separately. The false negatives of each ring could be easily counted, but counting false positive pixels was more difficult as it was not clear to which ring such pixel belonged. In the results shown in Fig. 7.5, each false positive pixel was accounted to the ring that it was closest to.

7.4.1.2 Other simulated phantom images

Experiments were also performed on the simulated datasets presented in Fig. 7.6. Fig. 7.6a,b,d show three 1024×1024 pixel phantoms based on actual reconstructions of rat femurs and where $\Delta s = \frac{\Delta t}{4}$. Fig. 7.6c represents a set of 20 randomly generated, analytically defined polyurethane foam phantom images. The width of each cell wall was chosen randomly in the interval $[\frac{\Delta t}{2}, \frac{5\Delta t}{2}]$.

For each dataset, projection data was generated on a parallel beam projection geometry with 180 equiangular projection angles and 256 detector cells with $\Delta t = 1$. For the pixel based phantoms (Fig. 7.6a,b,d), the PVE was induced by simulating high resolution projection data (with 1024 detector cells with $\Delta t = \frac{1}{4}$) in the intensity domain (i.e. \mathbf{I} , Eqn. 7.3). The detector bins were then summed 4 by 4 after which the resulting data was converted to the attenuation domain (i.e. to \mathbf{p} , Eqn. 7.4). For the analytical phantoms (Fig. 7.6c), projection data was computed analytically, inherently modelling the PVE. For every dataset, Poisson noise was applied; the intensity of which is defined by the incident beam intensity, I_0 . In these experiments, $I_0 = 20000$.

To quantify the segmentation accuracy, the rNMP was computed. As the experiments were performed at varying pixel or voxel sizes, the reconstructions were first rescaled to the size of the original, high resolution ground truth images. For the analytical phantoms, high resolution rasterizations were used as the ground truth. For phantoms Fig. 7.6b and Fig. 7.6d the rNMP was computed with respect to the most dense partition, i.e. the bone structures.

Several reconstruction methods were evaluated: S-SIRT (visualized for phantom Fig. 7.6a in Fig. 7.8a,c,e) versus DART (Fig. 7.8b,d,f); $a = 1$ versus $a > 1$; and sinogram upsampling (Fig. 7.8c,d) versus detector supersampling (Fig. 7.8e,f). In Fig. 7.7 the rNMP for all phantoms is shown.

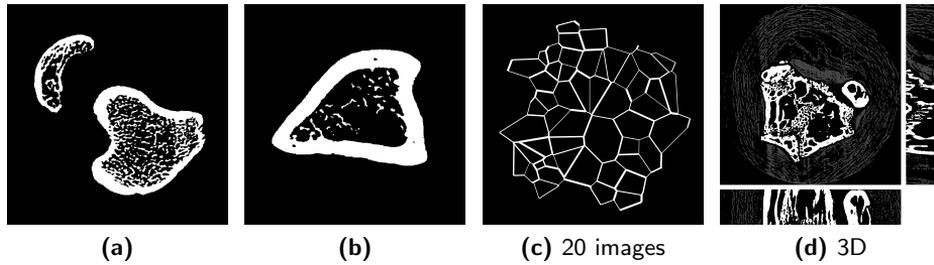


Figure 7.6: Simulated phantom images. Projection data of (a), (b) and (d) was generated from high resolution pixel images, based on actual reconstructions of rat femurs. The set of phantoms (c) were analytically defined and their projection data were also calculated analytically.

phantom	$a > 1$							
	$a = 1$		a	SU		DS		
	S-SIRT	DART		S-SIRT	DART	S-SIRT	DART	
Fig. 7.6a	0.388	0.460	4	0.249	0.302	0.208	0.086	
Fig. 7.6b	0.519	0.036	4	0.519	0.026	0.519	0.003	
Fig. 7.6c	0.468	0.347	8	0.463	0.381	0.389	0.107	
Fig. 7.6d	0.168	0.172	4	0.112	0.090	0.091	0.068	

Figure 7.7: Numerical results (rNMP) for all phantom experiments of Fig. 7.6. For the set of phantom Fig. 7.6(c), the average rNMP is given. A parallel beam geometry with 180 projection direction and $I_0 = 20000$ was used.

For phantom Fig. 7.6b and Fig. 7.6c, the advantage of using DART and detector supersampling can be seen in Fig. 7.9a-d. Small trabecular structures were properly segmented only on an upsampled reconstruction grid. Similar results can be seen for the foam segmentation, where it is clear that especially the thinnest cell edges benefited the most from the proposed super-resolution approach.

Fig. 7.10 shows the improvement of detector supersampling on 3D DART reconstructions of phantom Fig. 7.6d for two orthogonal viewing directions. It can be seen that by applying super-resolution, the small three-dimensional trabecular structures were segmented much more accurately, also through the X-axis.

7.4.1.3 Few-view tomography

To reduce the radiation dose, the number of projection angles can be lowered. However, this leads even further to limited data reconstruction problems. To

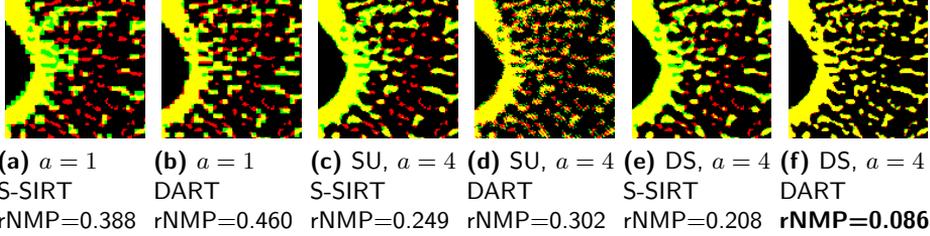


Figure 7.8: Region of the various reconstructions of Fig. 7.6a from 180 parallel beam projections with $I_0 = 20000$. The ground truth image is displayed in red and the reconstructions in green. Where both images overlap, i.e. where the segmentation is correct, the corresponding pixel is yellow.

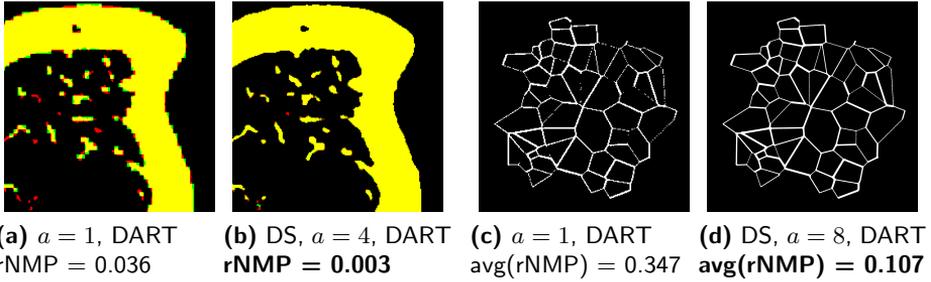


Figure 7.9: (a,b) The use of detector supersampling for phantom Fig. 7.6b. (c,d) The use of detector supersampling for phantom Fig. 7.6c.

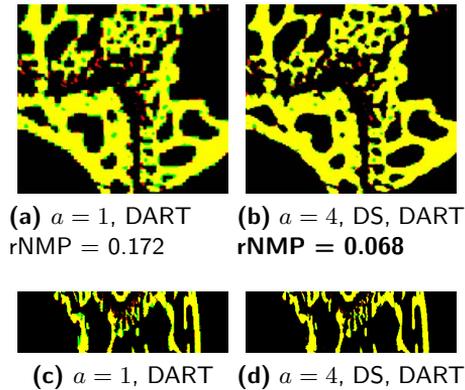


Figure 7.10: The improvement of detector supersampling on 3D-DART reconstructions of phantom Fig. 7.6d is clearly visible from slices through the Z-axis (a,b) and the X-axis (c,d).

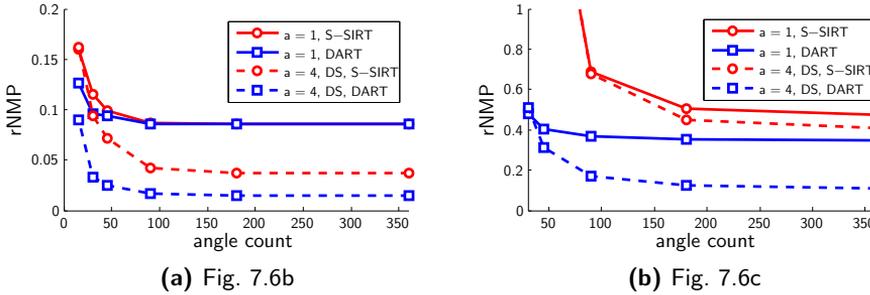


Figure 7.11: The rNMP as a function of the number of projection angles for phantoms Fig. 7.6b and Fig. 7.6c and for $a = 1$ and $a = 4$.

demonstrate the effect of a low angle count on the proposed super-resolution, projection data of Fig. 7.6b and Fig. 7.6c was generated with a decreasing number of projection angles, effectively simulating scans with a reduced radiation dose. The same downsampling strategy and I_0 was used as before in Section 7.4.1.2. For each set of projection data, DART and S-SIRT reconstructions were created with and without the detector supersampling approach. The rNMP values are plotted in Fig. 7.11a and Fig. 7.11b. One can conclude that, even with a drastically lowered number of projection angles, the combination of detector supersampling with the exploitation of prior knowledge resulted in reconstructed images that were more accurate than conventional S-SIRT reconstructions without a super-resolution approach and with a high number of projection angles.

7.4.1.4 Robustness of assumed principles

In DART — and thus also in the proposed super-resolution approach — it is assumed that the object has a homogeneous density and that this density is known in advance. A study was performed to investigate what happens if one of these assumptions are only approximately satisfied.

To demonstrate the robustness of the algorithm with respect to deviations from the first assumption, each pixel of phantom Fig. 7.6b was multiplied with a normally distributed random number with mean = 1. This was done multiple times for an increasing standard deviation. For each such image, projection data was generated with 30 projection angles, downsampled by a factor 4 — as explained before — and with $I_0 = 20000$. Reconstructions were made with S-SIRT and DART and with detector supersampling ($a = 1$ and $a = 4$). In Fig. 7.12a, the rNMP is plotted as a function of the standard deviation of the applied noise. While the rNMP of DART with the super-resolution approach indeed increased as the

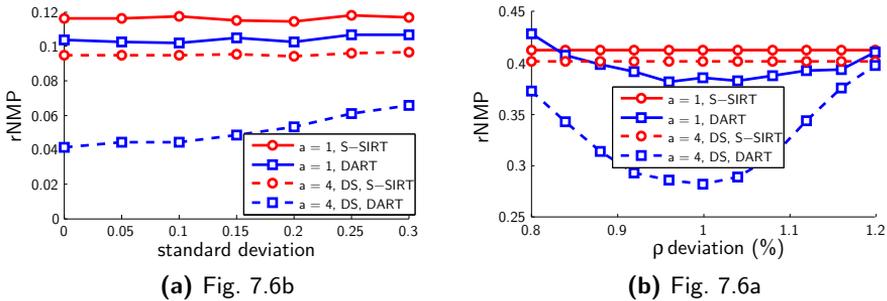


Figure 7.12: (a) The rNMP as a function of the standard deviation of the normally distributed noise that was multiplied with the phantom image Fig. 7.6b prior to simulating the projection data. (b) The rNMP as a function of the deviation on the correct grey level ρ during DART reconstructions of Fig. 7.6c.

objects grew more and more inhomogeneous, improvements over the conventional methods were still achieved.

For phantom Fig. 7.6a, projection data was created with 30 projection angles, again downsampled by a factor 4 and with $I_0 = 20000$. Multiple DART reconstructions were created where the assumed grey level was varied between 0.8 and 1.2 times the correct grey level. Fig. 7.12b plots the rNMP values for these DART reconstruction with detector supersampling ($a = 1$ and $a = 4$) and for S-SIRT. It can be seen that the rNMP of the DART reconstructions indeed increased as the assumed grey level was incorrect. However, drastic improvements over the conventional S-SIRT method without super-resolution could be achieved even if the chosen grey level was just an approximation of the correct grey level.

7.4.2 Experimental studies

The proposed method was also applied on experimental μ CT data. Fig. 7.13a shows an FBP reconstruction of a slice through a human mandible, which was recorded using a SkyScan 1173 μ CT scanner with 900 equiangular projection angles in the interval $[0, \pi)$. The detector resolution was $50\mu\text{m}$. The data was corrected for ring artefacts and beam hardening with the standard SkyScan NRecon software package. Only 100 projection angles were used in the experiments and the projection data was downsampled by summing the detector bins 4 by 4 in the intensity domain (Eqn. 7.3), such that many of the smaller structures were relatively small compared to the new detector sizes.

The discrete grey levels ρ were manually approximated, guided by the grey levels present in the initial SIRT reconstruction. From the FBP reconstruction

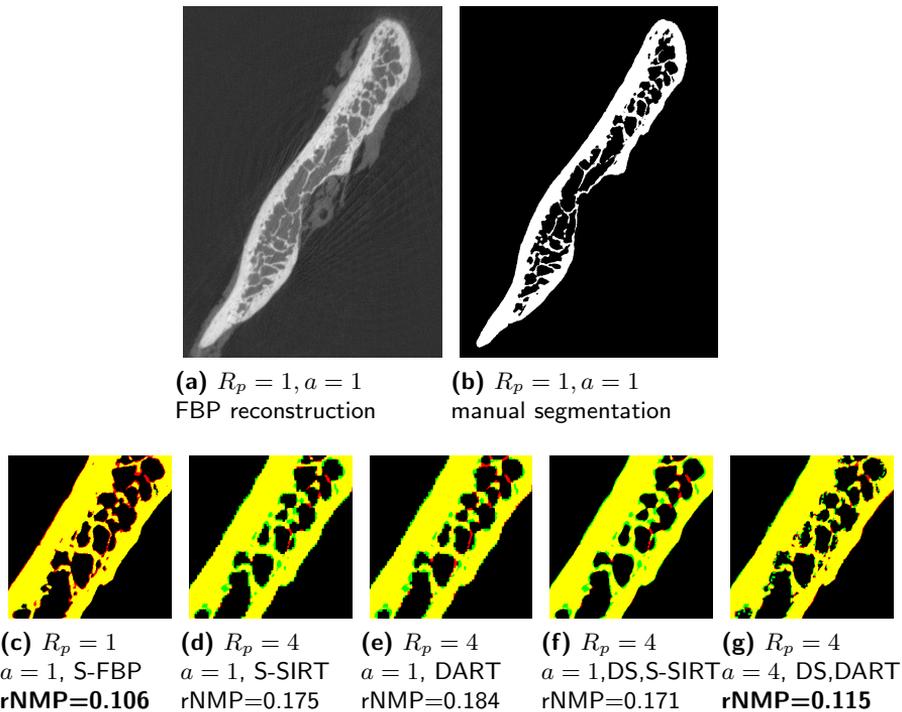


Figure 7.13: Results for an experimental μ CT dataset of a slice through a human mandible.

(Fig. 7.13a), computed with the non-downsampled and full set of projection data, a segmentation was manually created (Fig. 7.13b). This segmentation was used as a ground truth image.. This FBP reconstruction was also segmented using Otsu's clustering method (Fig. 7.13c). Fig. 7.13d-g indicate that also for experimental datasets, the addition of a super-resolution and DART was beneficial for the accuracy of the segmentation. When comparing Fig. 7.13c with Fig. 7.13g, it can be seen that with the proposed super-resolution approach, segmentations of low resolution projection data could be obtained that were of comparable quality to that of high resolution, high dose scans.

7.5 Discussion and conclusions

Accurate segmentation of structures that are small with respect to the reconstruction pixel size, poses a very complex and difficult problem as reconstructed images often lack contrast due to a partial volume effect. This often means that even the

intrinsic detector resolution cannot be achieved. High resolution reconstructions can provide a solution, but are often not feasible due to X-ray dose limitations, limited scanning time or hardware constraints.

To improve the detection of small structures in low resolution CT acquisitions, the use of a super-resolution approach has been proposed in which reconstructed images are computed on an upsampled reconstruction grid. Two geometrical methods for achieving super-resolution have been investigated: sinogram upsampling, where the projection data is upsampled by linear interpolation, and detector supersampling, where multiple rays per detector element are cast through the reconstruction grid. To counter limited data problems, prior knowledge about the object materials was exploited.

Experiments were performed on simulated as well as experimental data of objects containing small structures. Without using a super-resolution technique on objects containing small structures, the addition of prior knowledge (DART) sometimes resulted in less accurate segmentations when compared to the conventional S-SIRT algorithm. This effect was predicted in Section 7.3.2, where it was noted that DART is only suited for objects that are large with respect to the pixel size. However, if a super-resolution technique was applied, the use of prior knowledge with the DART method clearly resulted in more accurate reconstructions than the conventional S-SIRT approach. This effect was observed to be generally more profound if detector supersampling was chosen over sinogram upsampling.

In conclusion, the use of the detector supersampling super-resolution technique in which prior knowledge about the object density is exploited, can effectively increase the spatial resolution of a reconstructed image. In that way, small structures can be segmented more accurately even in cases with a low number of projection angles, and therefore with a low radiation dose.

References

- [1] P. L. Salmon and A. Y. Sasov, *Advanced Bioimaging Technologies in Assessment of the Quality of Bone and Scaffold Materials*. Springer Berlin Heidelberg, 2007.
- [2] A. Parfitt, M. Drezner, F. Glorieux, J. Kanis, H. Malluche, P. Meunier, S. Ott, and R. Recker, "Bone Histomorphometry: standardization of nomenclature, symbols and units," *Journal of Bone and Mineral Research*, vol. 2, no. 6, pp. 595–610, 1987.
- [3] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Transactions on Signal Processing*, vol. 20, pp. 21–36, May 2003.
- [4] J. A. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari, "Super-

REFERENCES

- resolution in PET imaging,” *IEEE Transactions on Medical Imaging*, vol. 25, pp. 137–147, 2006.
- [5] G. E. Marai, D. H. Laidlaw, and J. J. Crisco, “Super-resolution registration using tissue-classified distance fields,” *IEEE Transactions on Medical Imaging*, vol. 25, pp. 177–187, 2006.
- [6] A. A. Patil, R. Singhai, and J. Singhai, “Curvelet transform based super-resolution using sub-pixel image registration,” in *Computer Science and Electronic Engineering Conference (CEECE)*, pp. 1–6, 2010.
- [7] M. Li, H. Yang, and H. Kudo, “An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections,” *Physics in Medicine and Biology*, vol. 47, pp. 2599–2609, 2010.
- [8] E. Sidky and X. Pan, “Image reconstruction by constrained, total-variation minimization,” *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, 2008.
- [9] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, Y. L. Prince, C. A. Pelizzari, and X. Pan, “Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT,” *Physics in Medicine and Biology*, vol. 55, pp. 6575–6599, 2010.
- [10] W. van Aarle, G. Van Gompel, K. J. Batenburg, E. Van de Castele, and J. Sijbers, “A 3-dimensional discrete tomography approach for superresolution micro-ct images: application to foams,” in *The first international conference on image formation in X-ray computed tomography*, (Salt Lake City), pp. 45–48, 2010.
- [11] G. Van Gompel, K. J. Batenburg, E. Van de Castele, W. van Aarle, and J. Sijbers, “A discrete tomography approach for superresolution micro-CT images: application to bone,” in *Biomedical Imaging: From Nano to Macro, IEEE International Symposium on*, pp. 816–819, 2010.
- [12] P. M. Joseph and C. Ruth, “A method for simultaneous correction of spectrum hardening artifacts in ct images containing both bone and iodine,” *Medical Physics*, vol. 24, pp. 1629–1634, 1997.
- [13] E. Van de Castele, D. Van Dyck, J. Sijbers, and E. Raman, “A model-based correction method for beam hardening artefacts in x-ray microtomography,” *Journal of X-ray science and technology*, vol. 12, no. 1, pp. 53–57, 2004.
- [14] M. Krumm, S. Kasperl, and M. Franz, “Reducing non-linear artifacts of multi-material objects in industrial 3d computed tomography,” *Nondestructive Testing and Evaluation*, vol. 41, pp. 242–251, 2008.
- [15] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. J. Batenburg, J. de Mey, J. Sijbers, and N. J., “Iterative correction of beam hardening artifacts in ct,” *Medical Physics*, vol. 38, 2011.
- [16] T. M. Peters, “Algorithms for fast back- and re-projection in computed tomography,” *IEEE Transactions on Nuclear Science*, vol. 28, pp. 3641–3647, 1981.

Part III

Conclusions and appendices

—Um, can you repeat the part of the stuff
where you said all about the things?

Homer Simpson

8

Conclusions

Conventional segmentation techniques attempt to find segmented images that adhere to the reconstructed image and its histogram, but not necessarily to the measured projection data. In case the reconstruction is degraded by artefacts, for example due to a low number of projections, this is likely to lead to inaccurate images — and therefore also to inaccurate quantitative analysis.

In this work, multiple novel segmentation algorithms have been proposed for use in tomography. By design, these methods take the following principles into account.

- Compared to a reconstructed image, available projection data often contains additional information about the scanned object. This is exploited as much as possible. Optimal segmentations are those which maximally adhere to this projection data.
- User interaction is tedious and often error-prone. If possible, the algorithm parameter settings are estimated automatically.
- As tomography can be used in several application fields (Section 1.5), the proposed algorithms were developed without a specific application in mind.
- The radiation exposure can be reduced by reducing the number of projections or by reducing the field of view. High quality reconstructions should be provided even under these difficult circumstances.

All methods discussed in this work can be divided into two categories: *tomographic segmentation* and *discrete tomography*.

Part I: Tomographic Segmentation

In the first part of this work, tomographic segmentation methods were presented. These algorithms are to be executed as a post-processing step to the reconstruction.

At first, it was assumed that the scanned object contains only a few distinct homogeneous materials. The *Projection Distance Minimization (PDM)* method, discussed in Chapter 2, can then be applied. In PDM, the grey levels and optimal global threshold values are automatically estimated such that the Euclidean distance between the forward projection of the segmented image to the measured projection data is minimal. To find these optimal values, several optimization strategies were compared. Extensive experimental validation has shown that a common simplex search approach provides adequately accurate results with relatively little computational load.

Subsequently, the assumption that the object can only contain homogeneous materials was weakened. If only the densest material of the scanned object is assumed to be homogeneous, the *Segmentation Inconsistency Minimization (SICM)*, discussed in Chapter 3, can be applied to automatically approximate the optimal global threshold value for dense object segmentation. In SICM, the optimal segmentation is the one such that the residual sinogram (i.e. the sinogram of the part of the image that does not belong to the dense object) is minimally inconsistent. SICM also automatically estimates the grey level of the dense material.

Part II: Discrete Tomography

The second part of this work concerned discrete tomography methods that combine reconstruction and segmentation into a single algorithm.

In Chapter 4, the *Discrete Algebraic Reconstruction Technique (DART)* was discussed. It can be used if the scanned object is known to consist of a only a few different homogeneous materials, each corresponding to a grey level that is assumed to be known a priori. DART has been thoroughly validated with simulation experiments and has been applied to various experimental datasets.

The assumption that the set of grey levels must be known prior to the DART reconstruction forms a key problem when applying DART on an experimental dataset. If these grey levels are not accurately estimated, high quality reconstructions cannot be obtained. Several schemes for automatic grey level estimating were presented.

- In Section 4.4, the Euclidean distance between the forward projection of the DART reconstructed image and the measured projection data was proposed as a cost function to evaluate a chosen grey levels. A global optimization strategy, such as simplex search or adaptive surrogate modelling, can then be applied to find the optimal grey levels. While the typical search space contains many local optima, experiments have shown that this approach generally results in high quality estimations. This approach is *fully automatic* and can also be used to optimize other algorithm parameters. As a downside, one should note its extreme computational cost.

-
- In Chapter 5, the *Discrete Grey Level Selection (DGLS)* approach was presented for *semi-automatic* estimation of the grey levels. In DGLS, a user must select an image region that can be expected to correspond to a homogeneous region in the original object. Experiments have demonstrated that the DGLS approximations generally results in highly accurate DART reconstructions, without a large computational overhead. However, it cannot be used to estimate the optimal threshold values required in DART. Furthermore, it can only be applied if the user-based selection of homogeneous areas is easy, which in some cases (e.g. foams) is not true.
 - In Chapter 6, *PDM-DART* was proposed. PDM-DART combines discrete reconstruction with estimation of segmentation parameters. In contrast to normal DART, where these parameters remain fixed throughout the entire reconstruction process, PDM-DART adaptively selects the optimal segmentation parameters *within* each DART iteration. That way, accurate reconstructions can be computed even in cases where the DGLS method typically fails. While its computational overhead is substantially lower than the global optimization estimation scheme presented in Section 4.4, it is drastically larger than DGLS or manual estimation, especially for high dimensional estimation problems (i.e. if there are many unique grey levels present in the image). It should be noted that PDM-DART is *fully automatic*.

Another key problem in both conventional as well as discrete tomography is the accurate reconstruction of structures that are small with respect to the reconstruction pixel size. The partial volume effect then reduces the contrast of the reconstructed images. High resolution projection acquisition can provide a solution, but is often not feasible due to X-ray dose limitations, limited scanning time or hardware constraints. The conventional approach to reduce PVEs without increasing the X-ray dose is to upsample the reconstruction voxel grid, allowing for a more accurate representation and potentially improving the overall visualization of small structures. This upsampling is also known as *super-resolution* and typically results in a limited data reconstruction problem: the number of equations (measured projection data) remains the same while the number of unknowns (reconstruction voxels) increases significantly.

In Chapter 7, a novel super-resolution approach is proposed that improves the detection of small structures in low resolution CT acquisitions. By incorporating available prior knowledge (DART) to compensate for the lack of high resolution projection data, the proposed approach effectively increases the spatial resolution of the tomographic reconstructions.

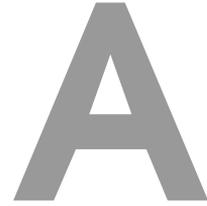
Overall, the advances made in this work, pave the way for fully automatic discrete tomography in a wide range of applications, (e.g. materials science, biomed-

CHAPTER 8. CONCLUSIONS

ical research, industrial tomography, ...), which was thus far a highly labour intensive procedure.

—I am completely operational, and all my
circuits are functioning perfectly.

HAL 9000, 2001



The ASTRA-toolbox

Contents

A.1	Motivation	177
A.2	Features	180
A.3	Examples	181
A.3.1	Simple reconstruction	181
A.3.2	Algorithm prototyping	183

A.1 Motivation

The *All Scale Tomographic Reconstruction Antwerp (ASTRA) toolbox* is a high performance framework that incorporates many common CT algorithms while supporting easy modification and extension, allowing fast prototyping and the accommodation of specific application requirements.

Tomographic reconstruction problems across various application fields share many common features and can often be solved using similar methods. In spite of this, new research is typically focussed solely on a single problem and on a single application, depending largely on the background of the researcher. The philosophy of the ASTRA research group, however, is that new techniques should not only benefit the application it was designed for, but that it should also be tested on other applications that have demonstrated the same problem or weakness. The ASTRA framework greatly facilitates the porting of tomographic techniques. Furthermore, it has the added benefit that various techniques can be easily compared to each other and that a technique to solve a specific problem can be rigorously tested in combination with techniques that solve other problems. The major design goals of the toolbox are the following.



Figure A.1: Logo of the All Scale Tomographic Reconstruction Antwerp (ASTRA) toolbox.

Modular, flexible and extensible. Since there is no one-size-fits-all solution for tomography, the toolbox has been designed to be fully modular and extensible using standard object oriented design. In Fig. A.3, a schematic overview is shown of the different components in the ASTRA toolbox. For each component, multiple implementations can exist, all with the same overall goal, but with a different approach that optionally takes application specifics into account. In the ASTRA framework, various algorithms, geometries or data-structures can be plugged in, compared or replaced by a custom implementation. In this way, a common platform is provided to easily test and benchmark different tomographic methods. Furthermore, each item can be considered as a building block, supporting the construction of complicated tomographic algorithms without tedious implementation work. This allows for easy prototyping of new algorithms or new application fields.

High performance. Computation time is an important issue in CT. The framework was therefore written and optimized in the C++ programming language. Additionally, given the fact that many tomographic algorithm are highly parallelizable, some methods are also implemented in the NVIDIA CUDA programming language. That way, the vast amounts of cheap computation power of modern day *graphical processing units (GPU's)* can be harvested (see the box on Page 179).

Easy to use. While an optimized C++ implementation guarantees efficiency, it is not an easy-to-use language when it comes to data visualization or fast pro-

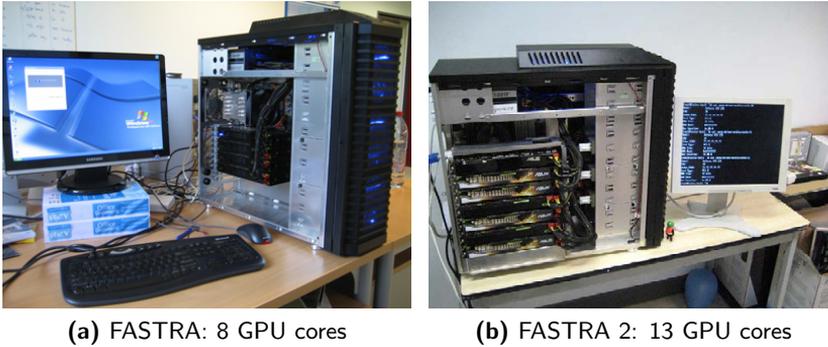


Figure A.2: Two very cheap, yet extremely powerful desktop supercomputers.

Intermezzo: the FASTRA's

In 2007, when scientific computing on GPU's was still in its infancy, a powerful desktop supercomputer was developed at the Vision Lab research group at the University of Antwerp. This computer, named the *FASTRA* (Fig. A.2a), was built with regular consumer hardware at a total cost of 4000EUR. The system contained 4 dual core gaming GPU's, resulting in a total computation power of roughly 3 terraflops. Given that tomography is highly suited for parallel computing, this system proved invaluable for the development of advanced tomographic algorithms, which were otherwise not feasible due to their computation cost.

In 2010, a successor was built (Fig. A.2b). This time, a total of 13 GPU cores were placed into a single desktop system, totalling 12 terraflops for less than 6000EUR. Under full load, this system consumes a relatively low 1200 Watt. As a comparison, typical supercomputers consume more than 100000 Watt and cost millions of euros.

At their release, both systems became a viral hit on the internet. More information about the FASTRA projects can be found at <http://fastra2.ua.ac.be>.

otyping. Therefore, an intuitive MATLAB wrapping layer was developed around the entire framework. That way, all C++ objects can be created, configured, applied and destroyed from within the MATLAB environment. Extensive documentation and examples are provided, two of which are presented in Section A.3.

Platform independence. The software has been designed to work on both Windows and Linux machines.

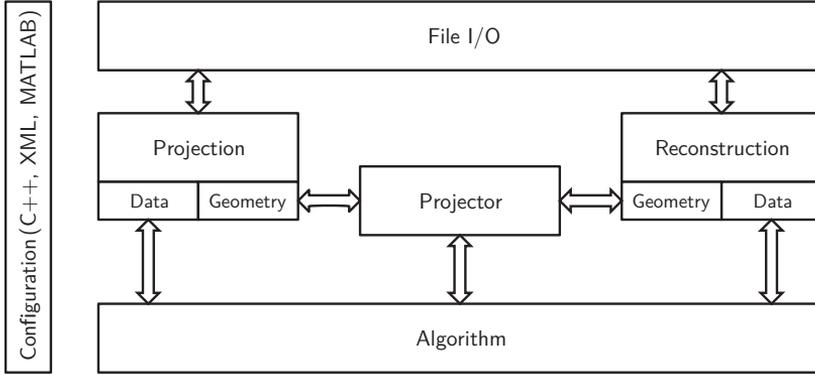


Figure A.3: A schematic overview of the different modules in the ASTRA-toolbox.

A.2 Features

In this section, a short overview of the most important features of the ASTRA-toolbox is presented.

Data structures Projection or reconstruction data is stored into two- or three-dimensional data structures. This data can be read from and written to file using specialized file I/O routines. Each data object is linked to a certain geometry. Projection geometries that are currently available are: parallel beam, fan beam, circular cone beam and cone beam with a trajectory specified by the user.

Projectors Central to the toolbox workflow are projector objects. These objects provide implementations to compute or approximate the values of the projector matrices (\mathbf{W} in Eqn. 1.19). They create a link between the projection geometry and the geometry of the reconstruction grid. Currently implemented: projection with a strip-based kernel (Fig. 1.7a), projection with a line-based kernel (Fig. 1.7b), Joseph’s method (Fig. 1.7c), and projection using any sparse matrix provided by the user.

It should be noted that projector objects are only available for CPU-based implementations. For the CUDA accelerated algorithms, the projection code is integrated in the actual algorithm code, sacrificing some flexibility for a significant speedup.

Algorithms Algorithm classes provide implementations for various tomographic operations. They are applied on projection and/or reconstruction data and

typically make use of a projector object. Currently, the following algorithms are implemented:

- On CPU and GPU: Simple forward projection, simple backprojection, filtered backprojection (FBP), SIRT, CGLS.
- On CPU only: ART, SART.
- On GPU only: Feldkamp-Davis-Kress (FDK), Expectation Maximization (EM).

In the future, many other algorithms will be added.

Matlab interface For better usability, bindings to the MATLAB scripting language are available. With the supplied MEX-files, all functionality of the C++ classes can be accessed from the MATLAB command line. Internally, the XML language is used to transfer configuration data to and from the C++ objects.

A.3 Examples

In this section, two examples are provided of how the toolbox can be operated from MATLAB. In Section A.3.1, a SIRT reconstruction is created of a phantom image. In Section A.3.2, the advanced prototyping capabilities of the toolbox are demonstrated with a simplified DART implementation.

A.3.1 Simple reconstruction

In this example, projection data is simulated of a 512×512 image that is stored in “blob.png”. This data is then used to create a SIRT reconstruction.

1. `Im = imread('blob.png');`

A 512×512 image is loaded and stored into a 2D matrix.

2. `proj_geom = astra_create_proj_geom('parallel', 1, 512, linspace(0,pi,180));`

The geometry of the projection data is specified as a parallel beam geometry with 512 detector cells of width 1.0. There are 180 projection angles, spaced equiangularly in $[0, \pi]$.

3. `vol_geom = astra_create_vol_geom(512,512);`

The geometry of the reconstruction data is defined as a simple 512×512 grid.

APPENDIX A. THE ASTRA-TOOLBOX

4. `proj_id = astra_create_projector('line', proj_geom, vol_geom);`

A line-kernel based projector object is created and stored in the memory. An identifier to this object is returned. This can subsequently be used as a handle to the object.

5. `sinogram_id = astra_create_sino(lm, proj_id);`

A sinogram is created of the 2D phantom image. Only an identifier of this data is returned. That way, the data is only stored in memory once.

6. `reconstruction_id = astra_mex_data2d('create', '-vol', vol_geom, 0);`

A new data object is created. This object will store the reconstructed image. It is initialized as a black image (it contains only zeros).

7. `cfg = astra_struct('SIRT');
cfg.ProjectorId = proj_id;
cfg.ProjectionDataId = sinogram_id;
cfg.ReconstructionDataId = reconstruction_id;
cfg.option.UseMinConstraint = 'yes';`

A MATLAB structure is created that contains all configuration settings for the reconstruction algorithm. Here, SIRT (CPU-based) will be used. The projector and data objects are specified and a positivity constraint is enabled that will force all values of the reconstructed image to remain positive.

8. `alg_id = astra_mex_algorithm('create', cfg);`

The algorithm object is created and initialized. As with the projector and data objects, an identifier is returned.

9. `astra_mex_algorithm('iterate', alg_id, 100);`

100 iterations of the algorithm are performed.

10. `reconstruction = astra_mex_data2d('get', reconstruction_id);`

The data contained in the reconstruction data object is fetched and placed into the MATLAB workspace.

11. `astra_mex_data2d('delete', sinogram_id, reconstruction_id);
astra_mex_algorithm('delete', alg_id);
astra_mex_projector('delete', proj_id);`

All objects are deleted from the memory.

A.3.2 Algorithm prototyping

With a few simple building blocks, complicated algorithms can be constructed. To demonstrate, a simplified implementation of the DART algorithm (Chapter 4) is presented.

```

function reconstruction = create_reconstruction ...
    (proj_id, sinogram_id, vol_geom, mask_id, iterations)

    reconstruction_id = astra_mex_data2d('create', '-vol', vol_geom, 0);
    cfg = astra_struct('SIRT');
    cfg.ProjectorId = proj_id;
    cfg.ProjectionDataId = sinogram_id;
1.  cfg.ReconstructionDataId = reconstruction_id;
    cfg.option.ReconstructionMaskId = mask_id;
    alg_id = astra_mex_algorithm('create', cfg);
    astra_mex_algorithm('iterate', alg_id, 100);
    reconstruction = astra_mex_data2d('get', reconstruction_id);
    astra_mex_algorithm('delete', alg_id);
    astra_mex_data2d('delete', reconstruction_id);
end

```

First, a function is defined that creates a SIRT reconstruction, restricted on a certain set of pixels, from an object containing projection data. The `ReconstructionMaskId` option ensures that the reconstruction is only applied on the pixels that have a value larger than zero in the data object to which `mask_id` points.

```

2.  lm = imread('blob.png');
    proj_geom = astra_create_proj_geom('parallel', 1, 512, linspace(0,pi,180));
    vol_geom = astra_create_vol_geom(512,512);
    proj_id = astra_create_projector('line', proj_geom, vol_geom);
    [sinogram_id, sinogram] = astra_create_sino(lm, proj_id);

```

As in the first example, projection data is simulated of a 512×512 binary phantom image. Note that in the last line, the sinogram data is also returned as a 2D matrix.

```

3.  mask_id = astra_mex_data2d('create', '-vol', vol_geom, 1);
    reconstruction = create_reconstruction( ...
        proj_id, sinogram_id, vol_geom, mask_id, 100);

```

An initial reconstruction is created using 100 iterations. The reconstruction mask contains only ones.

APPENDIX A. THE ASTRA-TOOLBOX

4. `for iteration = 1:100`

Next, DART steps into an iterative scheme.

5. `segmentation = (reconstruction > 128) * 255;`

The current reconstruction is segmented by application of a threshold. Each value in the segmentation is assigned the grey level corresponding to the prior knowledge. Here, the grey levels are 0 and 255 and the threshold value is 128. Note that, for brevity, only binary images can be reconstructed with this implementation.

6. `mask = ...
segmentation_fixed = segmentation .* ~mask;`

A new mask image is created. This image defines which pixels will be updated in the next iteration. Each boundary pixel is assigned to 1, all other pixels are assigned to 0. The actual implementation can be done in numerous ways and is left for the reader. Subsequently, the segmentation is multiplied with the inverse of the mask image.

7. `[sino_fixed_id, sino_fixed] = astra_create_sino(segmentation_fixed, proj_id);
astra_mex_data2d('delete', sino_fixed_id);
astra_mex_data2d('store', sinogram_id, sinogram - sino_fixed);`

A forward projection of the fixed pixels is then created. Subsequently, the residual sinogram is computed and stored.

8. `reconstruction = update_reconstruction(...
proj_id, sinogram_id, vol_geom, mask_id, 10);
reconstruction = reconstruction + segmentation_fixed;`

10 SIRT iterations are applied on the residual sinogram. To this reconstruction, the segmentation containing only the fixed pixels are added.

9. `end`

End of the iterative scheme.

From these examples, it is clear that the MATLAB layer around the ASTRA-toolbox provides an easy and intuitive approach to tomography. Given that changing any parameter (e.g. geometry parameter, projector type, reconstruction algorithm, algorithm options, ...) is a mere triviality, the toolbox is highly useful for experimenting.

—Look at me, I'm a train on a track
I'm a train, I'm a train, I'm a chucka train, yeah
Look at me, got a load on my back
I'm a train, I'm a train, I'm a chucka train, yeah.
Albert Hammond, I'm a train, 1973.



List of common abbreviations and symbols

Contents

Common abbreviations	185
Common symbols	186

Common abbreviations

- ART** Algebraic Reconstruction Technique.
- ASTRA** All Scale Tomographic Reconstruction Antwerp.
- CGLS** Conjugate Gradient Least Squares.
- CT** Computed Tomography.
- DART** Discrete Algebraic Reconstruction Technique.
- DGLS** Discrete Grey Level Selection.
- FBP** Filtered Backprojection.
- FDK** Feldkamp-Davis-Kress.
- FOV** Field of View.
- GPU** Graphical Processing Unit.

Common abbreviations

PDM Projection Distance Minimization.

PDM-DART Projection Distance Minimization DART.

PVE Partial Volume Effect.

rNMP relative Number of Misclassified Pixels.

ROI Region of Interest.

S-FBP Segmented Filtered Backprojection.

S-FDK Segmented Feldkamp-Davis-Kress.

S-SIRT Segmented Simultaneous Iterative Reconstruction Technique.

SART Simultaneous Algebraic Reconstruction Technique.

SICM Segmentation Inconsistency Minimization.

SIRT Simultaneous Iterative Reconstruction Technique.

SNR Signal-to-Noise Ratio.

Common symbols

a	$\in \mathbb{N}$	Upsampling factor for super-resolution.
i	$\in \{1, \dots, m\}$	Indexing variable for data in the projection domain.
j	$\in \{1, \dots, n\}$	Indexing variable for data in the reconstruction domain.
k	$\in \mathbb{N}$	Iteration number.
l	$\in \mathbb{N}$	Number of distinct grey levels.
m	$\in \mathbb{N}$	Number of detector cells multiplied by the number of projection angles.
n	$\in \mathbb{N}$	Number of pixels in the reconstruction grid.
\mathbf{p}	$= (p_i) \in \mathbb{R}^m$	Vector representing projection data.
\mathbf{s}	$= (s_j) \in \mathbb{R}^n$	Vector representing segmented image.
t	$\in \mathbb{R}$	Detector offset.
\mathbf{v}	$= (v_j) \in \mathbb{R}^n$	Vector representing reconstructed image.
\mathbf{W}	$= (w_{ij}) \in \mathbb{R}^{m \times n}$	The projector matrix that maps a function in the object or reconstruction domain onto the projection domain.

$\mathcal{C}(\mathbf{v})$	$:\mathbb{R}^n \rightarrow \{\rho_1, \dots, \rho_l\}^n$	Segmentation function.
$\mathcal{I}(\mathbf{v})$	$:\mathbb{R}^n \rightarrow \{1, \dots, l\}^n$	Threshold function.
\mathcal{R}	$:\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$	Radon transform.
\mathcal{S}	$:\mathbb{R}^m \rightarrow \mathbb{R}^n$	Linear operator that corresponds to a certain number of iterations of a linear algebraic reconstruction technique.
\mathcal{S}_A	$:\mathbb{R}^m \rightarrow \mathbb{R}^n$	Linear operator that corresponds to a certain number of iterations of a linear algebraic reconstruction technique, restricted to the pixels in set $A \subset \{1, \dots, n\}$.
$\boldsymbol{\rho}$	$= (\rho_1, \dots, \rho_l)^T$	Vector containing all distinct grey levels.
$\boldsymbol{\tau}$	$= (\tau_1, \dots, \tau_{l-1})^T$	Vector containing all threshold levels.
Δs	$\in \mathbb{R}^+$	The length and width of a pixel. In this work only square pixels are considered.
Δt	$\in \mathbb{R}^+$	The width of a detector cell.

—And if you listen very hard, the tune will come to you at last.
When all are one and one is all. To be a rock and not to roll.
Jimmy Page and Robert Plant, Stairway to Heaven, 1971.



Scientific contribution

Journal articles

- Wim van Aarle, K. Joost Batenburg, Jan Sijbers, “*Optimal Threshold Selection for Segmentation of Dense Homogeneous Objects in Tomographic Reconstructions*”, IEEE Transactions on Medical Imaging, vol.30, no.4, pp.980–989, April 2011.
- K. Joost Batenburg, Wim van Aarle, Jan Sijbers, “*A semi-automatic algorithm for grey level estimation in tomography*”, Pattern Recognition Letters, vol.32, no.9, pp.1395–1405, July 2011.
- Wim van Aarle, K. Joost Batenburg, Jan Sijbers, “*Automatic parameter estimation for the Discrete Algebraic Reconstruction Technique (DART)*”, IEEE Transactions on Image Processing, submitted, 2011.
- Wim van Aarle, K. Joost Batenburg, Gert Van Gompel, Elke Van de Castele, Jan Sijbers, “*Super-resolution for computed tomography based on discrete tomography*”, Journal of X-Ray Science and Technology, in preparation, 2012.

Conference proceedings (full paper)

- Wim van Aarle, K. Joost Batenburg, and Jan Sijbers, “*Threshold Selection for Segmentation of Dense Objects in Tomograms*”, International Symposium on Visual Computing, vol. 5358, Berlin Heidelberg, Springer-Verlag, pp. 700–709, December, 2008.

Conference proceedings (extended abstract)

- K. Joost Batenburg, Wim van Aarle, and Jan Sijbers, “*Grey Level Estimation for Discrete Tomography*”, *Discrete Geometry for Computer Imaging*, vol. 5810: Springer, pp. 517–529, September, 2009.
- Gert Van Gompel, K. Joost Batenburg, Elke Van de Castele, Wim van Aarle, and Jan Sijbers, “*A discrete tomography approach for superresolution micro-CT images*”, *IEEE International Symposium on Biomedical Imaging*, Rotterdam, pp. 816–819, April, 2010.
- Wim van Aarle, Gert Van Gompel, K. Joost Batenburg, Elke Van de Castele, and Jan Sijbers, “*A 3-dimensional discrete tomography approach for superresolution micro-CT images: application to foams*”, *The first international conference on image formation in X-ray computed tomography*, pp. 45–48, June, 2010.
- Wim van Aarle, Karel Crombecq, Ivo Couckuyt, K. Joost Batenburg, Jan Sijbers, “*Efficient parameter estimation for discrete tomography using adaptive modelling*”, *Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pp.229–232, July 2011.

