CRANFIELD UNIVERSITY

SCHOOL OF ENGINEERING

PhD THESIS
Academic Year 2013-2014

ANDRE MOUTON

On Artefact Reduction, Segmentation and Classification of 3D Computed Tomography Imagery in Baggage Security Screening

Supervised by: Dr Toby Breckon and Dr Carol Armitage
March 2014

This thesis is submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Abstract

This work considers novel image-processing and computer-vision techniques to advance the automated analysis of low-resolution, complex 3D volumetric Computed Tomography (CT) imagery obtained in the aviation-security-screening domain. Novel research is conducted in three key areas: image quality improvement, segmentation and classification.

A sinogram-completion Metal Artefact Reduction (MAR) technique is presented. The presence of multiple metal objects in the scanning Field of View (FoV) is accounted for via a distance-driven weighting scheme. The technique is shown to perform comparably to the state-of-the-art medical MAR techniques in a quantitative and qualitative comparative evaluation.

A materials-based technique is proposed for the segmentation of unknown objects from low-resolution, cluttered volumetric baggage-CT data. Initial coarse segmentations, generated using dual-energy techniques, are refined by partitioning at automatically-detected regions. Partitioning is guided by a novel random-forest-based quality metric (trained to recognise high-quality, single-object segments). A second segmentation-quality measure is presented for quantifying the quality of full segmentations. In a comparative evaluation, the proposed method is shown to produce similar-quality segmentations to the state-of-the-art at reduced processing times.

A codebook model constructed using an Extremely Randomised Clustering (ERC) forest for feature encoding, a dense-feature-sampling strategy and a Support Vector Machine (SVM) classifier is presented. The model is shown to offer improvements in accuracy over the state-of-the-art 3D visual-cortex model at reduced processing times, particularly in the presence of noise and artefacts.

The overall contribution of this work is a novel, fully-automated and efficient framework for the classification of objects in cluttered 3D baggage-CT imagery. It extends the current state-of-the-art by improving classification performance in the presence of noise and artefacts; by automating the previously-manual isolation of objects and by decreasing processing times by several orders of magnitude.

Acknowledgements

I would like to thank Dr Toby Breckon for his optimistic and insightful support and guidance throughout my research and for showing interest beyond the necessary. You made the process all the more enjoyable.

Thank-you to all those who willingly contributed to my research: Dr Carol-Armitage for the last-minute proof reading - your input wasn't ignored; Prof. Johan Nuyts and Dr Katrien Van Slambrouck at KU Leuven for access to their data, simulations and invaluable insight into iterative reconstruction; Dr Carl Crawford for his enthusiastic interest in my work and for access to the ALERT data and Najla and Greg for their input - particularly in the initial stages.

Moving to a new country can be difficult at the best of times, moving to the social hub that is Cranfield is on another level. Thank-you to family, friends (new and old) and everyone else who took the time to care.

And of course, an honorary mention must go to Reinhard, Kenneth and Willard for allowing me the luxury of seven slices of cucumber instead of six (provided it's Tuesday).

Finally, thanks to Joanna for embracing the routine and so much more. I'll see you at 04h30 for a run.

Funding for this project was provided under the Innovative Research Call in Explosives and Weapons Detection (2010), sponsored by the HOSDB, the DfT, the CPNI and the MPS.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The central role of baggage screening in the aviation-security domain has led to an increased interest in the development of automated, software-based solutions to challenging tasks such as the detection and classification of contraband items. This work considers the application of image-processing and computer-vision techniques to advance the automated analysis of low-resolution, complex 3D volumetric baggage Computed Tomography (CT) imagery obtained in the aviation security-screening domain.

1.1 Motivation

Aviation security has traditionally been performed in three sequential stages [Pol94]:

1. **Access denial:** preventing initial access to the civil aviation facility (via police intelligence).
2. **Baggage inspection:** implementing efficient explosives and/or threat detection procedures.
3. **Damage control:** ensuring the installation of sufficient structures and/or systems to minimise aircraft damage and maximise passenger survivability.

Baggage inspection is the principal safeguard against the transportation of illicit and/or dangerous materials and is typically performed using a combination of five approaches [BP02]: 1) manual search; 2) sniffer dogs; 3) Explosive Trace Detection (ETD); 4) 2D X-ray based imaging and 5) Explosive Detection Systems (EDS).

Singh [Sin03] discusses the increasingly important role of image-based automated baggage inspection within the aviation security infrastructure and emphasises two primary objectives: 1) the improvement of image quality to aid visual (i.e. human) inspection and software-based analysis of imagery and 2) the automated detection of explosives. The latter has more recently been extended to consider the

broader task of general threat and/or illicit materials detection [FBM12, CMP13]. The accomplishment of these two objectives requires the development and implementation of efficient software-based techniques for image denoising and artefact reduction, automated segmentation, feature extraction and image classification [Sin03]. These tasks are notoriously challenging in the security-screening domain due to the variability and complexity of security imagery (compared to medical imagery for example) and the demand for high throughput - it has been estimated that the congestion at large international airports such as Heathrow demands baggage inspection times of approximately 6 seconds per item [Spe01]. It has thus been suggested that image-based automated aviation security-screening systems should be characterised by: high-speed detections - to minimise traveller inconvenience; robustness to clutter (i.e. capable of detecting well-hidden objects); low false-positive (false alarms) rates and affordability [Sin03].

Owing to its speed (1200-1500 bags/hour) and relative affordability, X-ray based 2D imaging has traditionally been used for the screening and analysis of baggage items [BP02]. The interpretation of 2D X-ray imagery is however, complicated by variations in object orientation, clutter and density confusion [AZGA06]. Three-dimensional X-ray Computed Tomography (CT), which has enjoyed much success in a broad range of medical applications, has thus been introduced to the security-screening domain in so-called Explosive Detection Systems (EDS), in an attempt to mitigate the limitations of conventional 2D X-ray imagery [ZPA10a].

X-ray CT is based on the same physical principles as conventional radiography. An external X-ray source is used to produce cross-sectional images of the X-ray attenuation properties of the object being scanned. In conventional 2D radiography the attenuation values along the path of each X-ray beam are superimposed resulting in line integrals of the attenuation. In contrast, CT acquires a set of contiguous 2D cross-sectional images (which may be stacked to produce a volumetric image) and then reconstructs the attenuation values in each volumetric element (or voxel) separately. This produces a three-dimensional dataset [Man01]. The attenuation of an X-ray beam is a function of the effective atomic number, density and thickness of the material it traverses. Material-based discrimination is thus possible using the correlations between the effective atomic numbers and densities of materials and has formed the basis of automated explosives detection in security-screening applications. The advent of Dual-Energy Computed Tomography (DECT) [Joh11], whereby objects are scanned at two distinct energies, has provided an effective means for performing such material-based discrimination. Owing to the primary explosives detection-based objective of imaging within the aviation-security domain, DECT machines have thus been the baggage-CT scanners of choice.

The primary, non-object recognition-based objective of typical baggage-CT scanners, coupled with the demand for high throughput, means that 3D baggage-CT imagery typically presents with substantial noise, metal-streaking artefacts and poor voxel resolution (Figures 1.3 and 2.9). Baggage-CT imagery is thus generally of a much poorer quality than medical-CT imagery. In the medical domain, the constraints of throughput and the need for dual-energy materials detection are not forthright.

Currently baggage-CT scanners do not meet the demands of aviation security. Typical baggage-CT scanners process between 150 and 200 bags per hour (which does not meet the 6s inspection window as laid out by Speller [Spe01]) and are characterised by high false-positive rates ($\sim 30\%$) [BP02]. The development of a high-speed, yet accurate CT-based automated baggage screening system is thus an open problem and is addressed in this work through the development and application of efficient techniques for:

1. **Image quality improvement**, whereby denoising and artefact-reduction techniques are implemented to mitigate the detrimental effects of the characteristically high levels of noise and artefacts in baggage-CT imagery (Chapter 4).

2. **Segmentation**, whereby the objects within an image are isolated from one another to allow for accurate labelling (Chapter 7).

3. **Classification**, involving the labelling of images based on their contents (Chapters 5 and 8).

1.2 3D Baggage-CT Imagery

The 3D volumetric baggage-CT data used for the research conducted in this thesis has been obtained from a CT80-DR dual-energy baggage-CT scanner manufactured by Reveal Imaging Inc (Figure 1.1), designed specifically for materials-based explosive detection. A fan-beam geometry was employed with a focus-to-isocentre distance of 550mm, a focus-to-detector distance of 1008.4mm and nominal tube voltages of 160kVp and 80kVp. Raw projection data was rebinned to parallel-beam data [Man01]. Reconstructed 512×512 2D CT images are obtained via Filtered Back-Projection (FBP) [Hsi03] and are represented in Modified Hounsfield Units (MHU), where the CT densities at each pixel fall in the range $[0, 60000]$ with air calibrated to 0 MHU and water calibrated to 10000 MHU. Volumetric data was obtained by *stacking* all axial slices obtained for a given bag. The data is characterised by anisotropic voxel resolutions of $1.56 \times 1.61 \times 5.00$mm.

The complete dataset is composed of 552 scans obtained at the two aforementioned nominal tube voltages. This dataset has been used in a variety of ways

Figure 1.1: Reveal Imaging CT80-DR dual-energy baggage-CT scanner.

(dependent on the particular experimental procedure) throughout this thesis - detailed explanations are provided in the relevant chapters. Throughout this work, references are made to whole volumes, subvolumes and 2D axial slices. A whole-volume scan contains all of the stacked 512×512 axial slices obtained for a given bag. Subvolume scans are generated by cropping out particular regions or items of interest from the whole-volume scans - they retain the same voxel resolutions but have reduced dimensions. Axial slices refer to the individual 512×512 2D FBP reconstructions obtained for each scan. The data-gathering process was performed prior to the commencement of this work and thus not all relevant details are readily available (e.g. precise contents of scans).

The vast majority of CT-based literature is found in the medical domain. It is thus important to emphasise that there exist several significant differences in the nature and quality of typical medical-CT imagery and that encountered in the aviation-security domain. These differences mean that computer-vision techniques, such as segmentation and classification, which have been successfully applied to medical imagery are not guaranteed to be met with the same degree of success when applied to baggage-CT data. The most pertinent of these differences are discussed below.

Image quality: The nature of dual-energy-based baggage-CT scanners and the demands for higher scan speeds in the aviation-security domain (compared to the medical domain), lead to compromises in image quality - both in terms of noise and resolution [Sin03]. Sub-millimetre isotropic resolutions in all three dimensions have become the norm in medical CT scanners - Toshiba and GE Healthcare, for example, have advertised scanners with 0.35mm and 0.23mm isotropic voxel resolutions respectively [tos, geH] (Figure 1.2). In contrast the CT80-DR volumetric data used in this study is characterised by comparatively low anisotropic voxel resolutions of 1.56x1.61x5.00mm (Figure 1.3). Anisotropic voxel resolution and poor resolution in the axial plane in particular are known to compound the effects of image noise and artefacts [KKRH+00]. Consequently, in addition to significantly poorer resolutions, baggage-CT data typically presents with a lower signal-to-noise

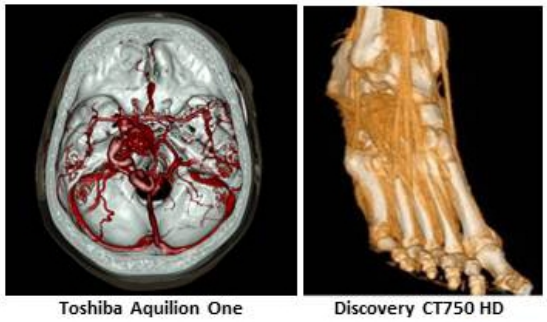


Figure 1.2: Medical-grade CT scans with sub-millimetre isotropic resolution [geH, tos].

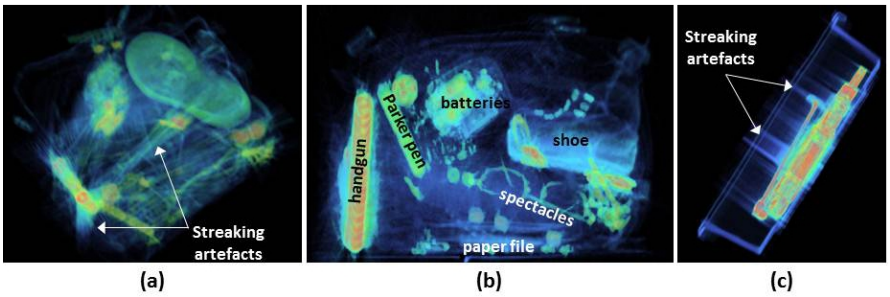


Figure 1.3: Reveal CT80-DR baggage-CT scans illustrating poor image quality, low resolution, artefacts and clutter.

ratio and a greater degree of artefacts. Ibanez *et al.* [ISNC05] make the following emphatic statement regarding the impact of low resolution imagery of this nature on the efficacy of computer-vision techniques: *'... such datasets are close to useless for the purpose of computer assisted image analysis."*. The contributions of this thesis will contrast sharply with this view.

***A priori* information:** In the medical domain, *a priori* knowledge related to the properties and spatial relations of the anatomical structures being scanned exists. It would, for example, be reasonable to assume that a CT scan of the head will be composed of brain matter, bone and air (see Toshiba CT scan in Figure 1.2). Furthermore, theoretical or expected X-ray attenuation properties for most anatomical structures/tissues exist. The exploitation of such *a priori* knowledge allows for the development of algorithms designed or fine-tuned for particular tasks or anatomical structures [KKRH+00]. In contrast, the contents of any given bag are entirely unknown prior to scanning and may exhibit considerable variability in shape, size, material and spatial context (Figure 1.3 (b)), making the fine-tuning of algorithms significantly more challenging.

Image complexity: In addition to the availability of *a priori* knowledge, most medical CT scans exhibit relatively low degrees of complexity and clutter (i.e. they are fairly homogeneous). Checked baggage, on the other hand, is generally tightly packed and thus extremely cluttered/complex (Figure 1.3 (b)) with no *a priori* information available related to the number of objects in any given bag. It is well documented that complexity and clutter have a significant detrimental effect on

both human and computer detection rates [Sin03].

The low-resolution, complex and unpredictable volumetric CT imagery encountered in the aviation-security domain is thus generally of a much poorer quality than that encountered in the medical domain. Throughout this thesis, this data is referred to as low-resolution, complex volumetric imagery to distinguish it from the comparatively high-resolution and uncluttered medical-CT imagery (see Figures 1.2 and 1.3). Algorithms designed for baggage-CT applications are thus required to be independent of the number of objects in an image as well as the composition of these objects, making their development particularly challenging.

1.3 Current State-of-the-Art

Noise and artefact reduction: The majority of denoising and artefact-reduction CT literature is found in the medical domain. The development of novel techniques or the evaluation of existing (medical) techniques in non-medical applications of CT imagery are extremely limited [XZX+09, GSK+12]. Existing comparative studies are limited in the techniques that are compared, the CT domains or applications which are considered and the performance-evaluation techniques that are employed.

Segmentation: Volumetric segmentation techniques (again existing predominantly in the medical literature) are typically fine-tuned for particular anatomical structures and are unlikely to be effective for the segmentation of multiple, unknown objects. The state-of-the-art in the segmentation of unknown objects from cluttered volumetric CT imagery [Gra06] has been developed using high-resolution medical-grade imagery with relatively low levels of noise and metal-streaking artefacts [CMP13]. The segmentation of low-resolution, cluttered volumetric imagery in the presence of multiple metal objects has not been considered previously.

Classification: The current state-of-the-art in 3D object classification in non-medical complex 3D volumetric imagery [FBM12] relies on the manual segmentation of the input data; incurs large computational overhead and suffers a decline in performance in the presence of image noise and/or artefacts. An efficient, fully-automated classification framework that is robust to image noise and artefacts does not currently exist in this domain.

1.4 Contribution to Knowledge

The research conducted in this thesis addresses each of the aforementioned limitations in the state-of-the-art via the following contributions:

- A novel interest-point based quantitative performance measure is presented, extending traditional denoising performance evaluation approaches by eval-

uating the potential benefits of denoising on the application of more complex operations (volume rendering and 3D object classification) within the current imaging context.

- A novel Metal Artefact Reduction (MAR) technique, designed specifically for cluttered baggage-CT imagery containing multiple metal objects, is presented and shown to perform comparably to state-of-the-art medical techniques when applied to cluttered baggage-CT imagery.

- A comprehensive comparative performance evaluation (which has not previously been considered in medical or non-medical CT domains) is conducted for seven image-denoising techniques [ZPA10a, PM90, ROF92, ZG08, CDAO95, BCM05b] and twelve artefact-reduction techniques [WSOV96, KHE87, ZBWW02, BS06, YZB+07, JR09, LBY+10, AAA+10, MRL+10, MRS+11, MRL+12, MMB+13].

- A novel dual-energy-based segmentation technique is presented and shown to provide fast, high-quality segmentations of complex volumetric baggage-CT imagery. Within the proposed framework, four novel contributions have been made: 1) a materials-based coarse segmentation technique; 2) a random-forest-based model for measuring the quality of individual object segments; 3) a random-forest-based model for measuring the quality of entire segmentations and 4) an efficient segmentation-refinement procedure for splitting fused objects. In a comparative performance evaluation, the proposed technique is shown to perform comparably to the state-of-the-art [CMP13, WGW12, Gra06].

- A codebook image classification model constructed using random-forest-based feature encoding, a dense-feature sampling strategy and a Support Vector Machine (SVM) classifier is presented and shown to significantly outperform the current state-of-the-art [FBM12] both in terms of accuracy as well as runtime.

- The culmination of the research conducted in this thesis is a novel, fully-automated and efficient framework for the classification of objects in complex volumetric baggage-CT imagery. The framework is shown to improve on the current state-of-the-art [FBM12] by reducing the detrimental effects of image noise and artefacts; by automating the segmentation process and by improving both runtime as well as accuracy.

Portions of the work presented in this thesis have previously been published in the following peer reviewed publications:

- An Experimental Survey of Metal Artefact Reduction in Computed Tomography (A. Mouton, N. Megherbi, T.P. Breckon, K. Van Slambrouck, J. Nuyts) *Journal of X-ray Science and Technology, IOS Press, Volume 21, No. 2, pp. 193-226, 2013.* Bibliographic reference [MMvS+13].

- A Distance-Driven Method for Metal Artefact Reduction in CT (A. Mouton, N. Megherbi, T.P. Breckon, K. Van Slambrouck, J. Nuyts) *Proceedings of the IEEE International Conference on Image Processing, pp. 2334-2338, 2013.* Bibliographic reference [MMB+13].

- A Novel Intensity Limiting Approach to Metal Artefact Reduction in 3D CT Baggage Imagery (A. Mouton, N. Megherbi, G.T. Flitton, S. Bizot, T.P. Breckon), *Proceedings of the IEEE International Conference on Image Processing, pp. 2057-2060, 2013.* Bibliographic reference [MMFB12]

- An Evaluation of CT Image Denoising Techniques Applied to Baggage Imagery Screening (A. Mouton, G.T. Flitton, S. Bizot, N. Megherbi, T.P. Breckon), *Proceedings of the IEEE International Conference on Industrial Technology, pp. 1063-1068, 2013.* Bibliographic reference [MMFB13].

- 3D Object Classification in Complex Volumes using Randomised Clustering Forests (A. Mouton, T.P. Breckon, G.T. Flitton) *Submitted to IEEE International Conference on Image Processing: under review.*

- A Review of Automated Analysis within 3D Baggage Security Screening Computed Tomography (A. Mouton, T.P. Breckon), *Submitted to Machine Vision and Applications: under review.*

- Materials-Based 3D Segmentation of Unknown Objects from Dual-Energy Computed Tomography Imagery in Baggage Security Screening (A. Mouton, T.P. Breckon), *Submitted to Pattern Recognition: under review.*

- On the Relevance of Denoising and Artefact Reduction in 3D Segmentation and Classification within Complex CT Imagery (A. Mouton, T.P. Breckon), *Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence: under review.*

1.5 Thesis Structure

The reader is introduced to the general topic of X-ray Computed Tomography (CT) in **Chapter 2** through a brief overview of the fundamental principles governing the generation and detection of X-rays; the acquisition of CT data; the reconstruction of CT images and the factors affecting the quality of CT imagery.

The foundation of the research conducted in this thesis is then established via a critical review in **Chapter 3** of the prior literature and current state-of-the-art in noise and artefact reduction, volumetric image segmentation, dual-energy techniques and image classification.

Chapter 4 addresses the topics of noise and artefact reduction in the previously unconsidered context of low-quality, complex volumetric baggage-CT imagery through experimental comparisons and the development of novel dedicated baggage-CT techniques.

The feasibility of codebook-based classification in 3D volumetric baggage-CT imagery is investigated and substantiated in **Chapter 5** via a comparative performance evaluation of five codebook models to the current state-of-the-art 3D visual cortex approach [FBM12].

Thereafter, the potential benefits of incorporating dual-energy CT techniques into an object-classification framework for 3D volumetric baggage-CT imagery are experimentally investigated in **Chapter 6**.

Chapter 7 then addresses the segmentation of unknown objects from cluttered, low-quality volumetric data and presents a novel materials-based approach which is shown to produce fast, high-quality segmentations of baggage-CT images.

The contributions of the preceding chapters are combined in **Chapter 8** to create a fully-automated and efficient classification framework which demonstrates state-of-the-art performance in complex, volumetric baggage-CT imagery.

Finally, an overview of the research conducted in this work and several potential directions for future developments are presented in **Chapter 9**.

Chapter 2

X-Ray Computed Tomography

X-ray Computed Tomography (CT) is a non-destructive imaging modality that produces cross-sectional images representing the X-ray attenuation properties of the substances being scanned. X-ray CT is used for visualising the interior features of solid objects and for obtaining information regarding their three-dimensional geometric characteristics. A single CT slice represents a cross-sectional view of a certain thickness of the object being scanned. A volumetric representation is generated by *stacking* a set of contiguously acquired slices.

The development and evaluation of Metal Artefact Reduction (MAR) techniques (Chapter 4) demands an understanding of the theory of tomographic reconstruction (Section 2.3) as well as knowledge of the causes and effects of noise and artefacts in CT imagery (Section 2.4). The physics governing the interaction of X-rays with matter (Section 2.1.2) is central to the development of dual-energy-based techniques (Chapter 6). A brief overview of the fundamental principles governing the generation and detection of X-rays; the acquisition of CT data; the reconstruction of CT images and the factors affecting the quality of CT imagery thus follows.

2.1 The Physics of X-Rays

Electromagnetic radiation is composed of individual photons. The energy, E, of every individual photon is inversely proportional to its wavelength λ:

$$E = \frac{hc}{\lambda} = hv, \tag{2.1}$$

where $h = 6.6261 * 10^{-20} Js$ is Planck's constant, $c = 3 \times 10^{8} m/s$ is the speed of light and v is the frequency of the wave. X-rays are electromagnetic waves characterised by wavelengths in the Angstrom range ($0.1nm$), or equivalently, by photon energies in the order of 10 keV (1 eV = 1.60210^{-19} J).

2.1.1 X-ray Source

X-rays are generated in an X-ray tube - the main components of which are a vacuum tube, an anode and a cathode. The heated cathode filament releases a stream of electrons, via thermal excitation, which is accelerated across a high voltage towards the anode. The stream of accelerated electrons flowing between the cathode and anode is referred to as the tube current (typically approximately 100mA), while the potential difference between the cathode and anode is referred to as the tube voltage (generally in the range 80kV to 140kV). A vacuum is maintained within the glass envelope of the X-ray tube to prevent the electrons from interacting with gaseous particles [HR03, Man01]. As the high energy electron stream collides with the anode, X-rays are generated by two distinct atomic processes [Lu99]:

1. *The Bremsstrahlung process*: This is the radiation that is released as the incident electrons are decelerated on interaction with nuclei of the anode. The resulting X-rays are characterised by a continuous spectrum and are bounded by the electric charge of the electron q and the tube voltage V:

$$E \leq E_{max} = qV \tag{2.2}$$

   A tube voltage of 80kV thus yields X-rays with a maximum energy of E_{max}= 80keV.

2. *X-ray fluorescence*: High energy electrons may collide with and subsequently eject the orbital electrons of the inner electron shell (e.g. the K-shell). As a result, electrons from a higher energy shell (e.g. the L-shell) fill the gaps left by the ejected orbital electrons and emit X-ray photons with discrete energies, known as the line spectrum (in this case a K-spectrum). More specifically, the energy of a photon emitted in this manner is equal to the difference between the energies of the two electron states [Man01]. The line spectrum is material-dependent and is depicted by characteristic peaks superimposed onto the continuous Bremsstrahlung spectrum and is referred to as the characteristic radiation (Figure 2.1).

The quantity and energy of the emitted photons are controlled by the cathode current (this is not the same as the tube current) and the tube voltage respectively.

2.1.2 The Interaction of X-Rays With Matter

There are four predominant means of interaction between X-ray photons and matter:

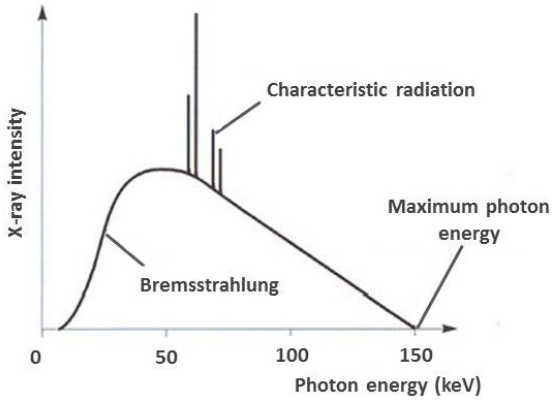


Figure 2.1: Typical X-ray spectrum, characterised by a continuous Bremsstrahlung radiation spectrum with sharp peaks at characteristic energies (caused by X-ray fluorescence).

1. **The photoelectric effect:** this refers to the interaction between an incident X-ray photon and matter in which the photon energy is completely absorbed by an atomic electron that is ejected from the atom.

2. **Coherent scattering:** (also known as Rayleigh scattering) occurs when a very low energy X-ray photon interacts with a strongly bound electron. Since the incident photon energy is insufficient to overcome the binding energy of the electron, the photon is deflected, or *scattered*, from its original path. The interaction involves no transfer of energy to kinetic energy (and hence no change in the wavelength or frequency of the photon) and ionisation does not occur.

3. **Incoherent scattering:** (also known as Compton scattering) this refers to the interaction of an X-ray photon with matter in which the photon is deflected and retains part of its original energy.

4. **Pair production:** occurs when the incident X-ray photon interacts with the nucleus of the target atom, resulting in the creation of a positron-electron pair. Pair production only occurs at very high energy levels (generally greater than 1.022 MeV).

The typical energy range of the X-ray photons generated for use in baggage-security CT, as well as medical-diagnostic CT, fall within the range 20 keV - 200 keV. Pair production only occurs at considerably higher energy ranges and is therefore not of concern in this work. Coherent scattering makes its major contribution at low photon energies (less than 50 keV) for materials with high atomic numbers. Nonetheless, since coherent scattering involves no transfer of energy it is of limited interest in CT. In the context of this work, the interaction of X-ray with matter is thus dominated by the photoelectric effect and incoherent (Compton) scattering.

The manner in which photons interact with matter is probabilistic in nature. The likelihood of a particular type of interaction occurring is defined by the *cross-section* of interaction [Coo98]. In high-energy particle physics, the cross-section is defined as a hypothetical area governing the probability of a particular interaction occurring when small particles collide. In X-ray physics, these cross-sections are material and energy-dependent and may be represented as a function of the atomic number of the target atom (or the effective atomic number for a compound material) and the incident photon energy. The photoelectric cross-section σ_{pe} and the incoherent scattering cross-section σ_{is} may be approximated as follows [AM76]:

$$\sigma_{pe} \simeq K_1 \frac{Z^n}{E^3} \quad (n \approx 4.5) \tag{2.3}$$

$$\sigma_{is} \simeq K_2 Z f_{KN}(E) \tag{2.4}$$

where Z is the atomic number of the target atom, E is the incident photon energy in keV and K_1 and K_2 are constants. The function $1/E^3$ approximates the energy dependence of the photoelectric cross-section, while the energy dependence of the incoherent scattering cross-section is defined by the Klein-Nishina function $f_{KN}(E)$:

$$f_{KN}(\alpha) = \frac{1+\alpha}{\alpha^2}\left[\frac{2(1+\alpha)}{1+2\alpha} - \frac{1}{\alpha}\ln(1+2\alpha)\right] + \frac{1}{2\alpha}\ln(1+2\alpha) - \frac{(1+3\alpha)}{(1+2\alpha)^2} \tag{2.5}$$

where $\alpha = E/510.975\text{keV}$. An important observation regarding these relationships is that σ_{pe} decreases rapidly as the atomic number Z (or the effective atomic number Z_{eff} for compound materials) decreases and as the photon energy E increases. In contrast, σ_{is} decreases much more slowly with increasing energy. Consequently, the photoelectric effect is the dominant effect at low energies, while incoherent scattering becomes the dominant attenuation mechanism at higher photon energies (and especially for organic materials with low atomic numbers). Furthermore, the relationship of σ_{pe} to the atomic number and photon energy ($\sigma_{pe} \propto Z^n$ and $\sigma_{pe} \frac{1}{\propto} E^3$) indicates that lower-energy photons are useful for low-contrast differentiation of materials.

2.1.3 The Principles of X-ray Detection

Detectors are the image receptors that 'capture' the attenuated X-ray beams that have passed through the patient/object and convert them first to an electrical signal and then to a digital (binary-coded) signal for computerised reconstruction. In third-generation CT scanners, the most common detectors are comprised of

either a high-pressure inert gas (usually xenon) or solid-state scintillators coupled with photo-diodes [KKRH+00].

Gas detectors are comprised of a series of thin Tungsten ionisation plates submersed in a high-pressure xenon gas chamber. A single detector cell consists of a low-voltage (cathode) and a high-voltage (anode) Tungsten plate. When the X-ray photons collide with the charged Tungsten plates a photoelectric interaction occurs, resulting in the ionisation of the xenon gas. The ionised xenon nuclei drift towards the cathode, while the free electrons (released in the photoelectric interaction) gather on the anode, producing an electric current. Xenon gas detectors have a characteristically low quantum detection efficiency ($\sim 60\%$) and a high response time ($\sim 700\mu s$). Due to their comparatively low cost, several low-end single-slice CT scanners still use xenon gas detectors.

Solid-state scintillation detectors overcome many of the limitations of xenon detectors and are the most commonly used detectors in state-of-the-art scanners. A scintillation detector consists of a crystal that fluoresces when struck by an X-ray photon, producing a visible photon (light energy). The visible photon impinges on a photodiode coupled to the scintillator and is converted into an electrical signal. Modern solid-state detectors are characterised by low response times (order of 1×10^{-12}s) and a high quantum efficiency ($\geq 98\%$) [Man01].

The net effect of the interactions of an X-ray beam with matter is a gradual decrease in its intensity (due to absorption and scattering). Cooke [Coo98] defines the intensity of an X-ray beam as: *"the rate of flow of photon energy through a unit area lying at right angles to the path of the beam"*. This reduction in intensity is referred to as *attenuation*. The X-ray photons which are neither absorbed nor scattered, pass through the matter via a process referred to as transmission.

The attenuating ability of a material is quantified by its attenuation coefficient. More particularly, the *linear attenuation coefficient* of a material may be defined as the fraction of a parallel X-ray beam that is attenuated per unit distance of the material being traversed [See01]. X-ray images are essentially representations of these linear attenuation coefficients.

The Beer-Lambert Law states that, for monochromatic X-ray beams, the X-ray intensity of the beam passing through a homogeneous material of uniform thickness ds and atomic number Z, falls off exponentially as a function of the product of the path length (i.e. the object thickness) and the linear attenuation coefficient of the object:

$$I = I_0 e^{-\mu ds} \tag{2.6}$$

where I and I_0 are the incident and transmitted X-ray intensities respectively

and μ is the linear attenuation coefficient of the material. The linear attenuation coefficient μ is related to the total cross-section per atom σ_{tot} according to [Dys05]:

$$\mu = \frac{\rho N_A}{uA}\sigma_{tot} \tag{2.7}$$

where $N_A = 6.0225210^{23}$ mol^{-1} is Avogadro's number, u is the unified atomic mass unit (1/12 the mass of an unbound neutral ^{12}C atom) and A is the relative atomic mass of the target element. The total cross-section σ_{tot} may defined as the sum of the contributions of the relevant photon interactions [Dys05]:

$$\sigma_{tot}(Z, E) = \sigma_{pe}(Z, E) + \sigma_{is}(Z, E) \tag{2.8}$$

For non-uniform substances, the Beer-Lambert law in Equation 2.6 may be applied in a cascade fashion [Hsi03]:

$$I(s) = I_0 e^{-\int_0^s \mu(s')ds'} \tag{2.9}$$

Equation 2.9 still assumes a monochromatic X-ray beam. In reality, the X-ray beam emitted by an X-ray tube is polychromatic in nature, covering a broad spectrum (Figure 2.1). Due to the dependence of the linear attenuation coefficient on photon energy, the degree of attenuation experienced by the X-ray beam for each distinct energy in the spectrum is thus different. For polychromatic X-rays and heterogeneous materials, the transmitted intensity is computed as the summation of the intensities over all the energies in the spectrum. Equation 2.9 becomes [Man01, Hsi03]:

$$I(s) = \int_0^{E_{max}} I_0(E) e^{-\int_0^s \mu(s',E)ds'} dE \tag{2.10}$$

where $I_0(E)$ represents the polychromatic spectrum emitted by the X-ray tube and E_{max} is the maximum photon energy (equal to the tube voltage).

When considering the linear attenuation coefficients for different substances at a given energy, the differences are often quite small. The small differences are enhanced via a linear transformation of the linear attenuation coefficients according to the Hounsfield Scale (named after the pioneer of CT, Sir Godfrey Hounsfield). The output of the transformation is the so-called CT number and is measured in Hounsfield Units (HU). The values in reconstructed CT images are

commonly measured in HU. The scale has been designed such that the CT number of distilled water at the Standard Pressure and Temperature (STP) is 0 HU and that of air is -1000 HU:

$$\text{CT Number} = 1000 \cdot \frac{\mu - \mu_{H_2O}}{\mu_{H_2O}} \tag{2.11}$$

where μ is the linear attenuation coefficient of the voxel and μ_{H_2O} is the linear attenuation coefficient of water. The Hounsfield Scale results in CT numbers of -1000 HU, 0 HU and $\pm$1000 HU for air, water and bone respectively. A dynamic range of 2000 HU is beyond what can be represented in grayscale and perceived by the human eye. A gray-level transformation is thus generally applied via a window/level operation such that the window is defined as the total span of a given display interval and the level is defined as the centre of the interval [Man01]:

$$\text{Window} = \text{CT}_{\text{max}} - \text{CT}_{\text{min}} \tag{2.12}$$

$$\text{Level} = \frac{\text{CT}_{\text{max}} + \text{CT}_{\text{min}}}{2} \tag{2.13}$$

where $[\text{CT}_{\text{min}}\ \text{CT}_{\text{max}}]$ is the desired image display interval. Various window/level settings can be applied to view different structures in a given image (e.g. bone settings vs. soft-tissue settings).

2.2 Data Acquisition

The mathematical model typically used in X-ray CT (incorrectly) assumes a monochromatic X-ray beam and ignores the effects of scattering, beam hardening and other physical phenomena. The X-ray beam can thus be approximated as a set of parallel lines. Therefore, we consider the 2D parallel-beam geometry in Figure 2.2. The object being scanned lies along the z-axis (coming out of the page). The distribution of the linear attenuation coefficient of the object in the xy-plane is represented by $\mu(x, y)$. The function $\mu(x, y)$ is assumed to be *compactly supported* - that is to say, it is zero outside a bounded domain. In this case, the bounded domain is defined as a circular region with diameter D_{FoV} known as the scan FoV (the region within the gantry from which the projection data is acquired). A new coordinate system (r, s) is defined by rotating the (x, y) axes through an angle θ such that:

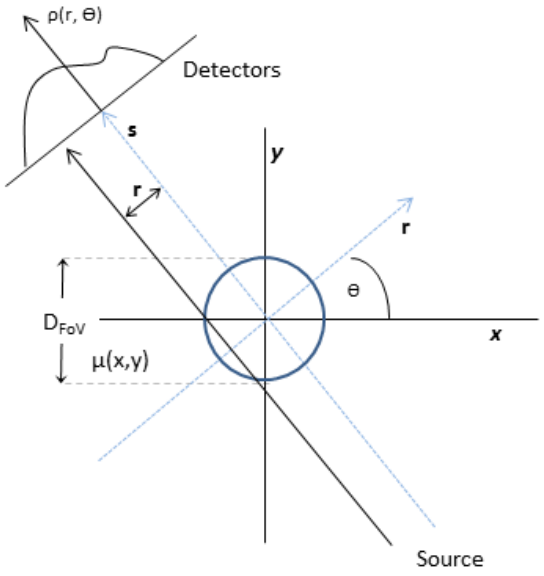


Figure 2.2: A parallel-beam geometry. The X-ray beams make an angle of θ with the y-axis and are at a distance r from the origin.

$$\begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$
$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} \tag{2.14}$$

Assuming that the Beer-Lambert Law (Equation 2.6) is obeyed, the attenuated X-ray intensity profile (as measured by the detector) for some fixed angle θ, is represented as a function of r:

$$I_\theta(r) = I_0 e^{-\int_{L_{r,\theta}} \mu(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta)ds} \tag{2.15}$$

where r is the distance of the X-ray beam from the origin and s is the X-ray path length. $L_{r,\theta}$ is then the line that makes an angle θ with the y-axis at distance r from the origin. Equation 2.15 thus reads: if a monochromatic X-ray beam with intensity I_0 enters an object, and travels along the line $L_{r,\theta}$ it will exit with an intensity $I_\theta(r)$. The measured intensity data $I_\theta(r)$, however, is not typically used directly in computed tomography. Instead, the attenuation profile is obtained by log-converting the ratio of the input to output X-ray intensities:

$$P_\theta(r) = -\ln\frac{I_\theta(r)}{I_0} = \int_{L_{r,\theta}} \mu(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta)ds \tag{2.16}$$

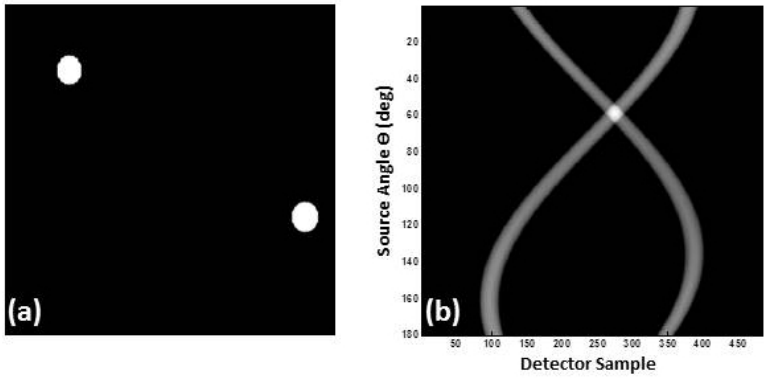


Figure 2.3: A sparse input image (a) composed of two point sources results in a sinusoidally-shaped sinogram (b).

$P_\theta(r)$ is referred to as the *projection* measurement (or ray-sum) of the attenuation function $\mu(x, y)$ and represents the set of line integrals for all parallel lines intersecting the support (the FoV) of μ in a particular direction defined by θ. Note that $P_\theta(r)$ is often denoted as $\rho_\theta(r)$ - an uppercase P is used here to avoid confusion with references made to mass densities in Chapter 6. In the mathematical literature, the integral operator in Equation 2.16 is referred to as the Radon transform of a function $f(x, y)$: $\mathcal{R}\{f(x, y)\}$ (after its inventor Johan Radon). In fact, the Radon transform provides a decomposition of any function $f(x, y)$ (which may represent the linear attenuation distribution of an object) into a set of parallel line projections $P(r, \theta)$:

$$P(r, \theta) = \mathcal{R}\{f(x, y)\} = \int_{-\infty}^{\infty} f(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta)ds \tag{2.17}$$

A collection of projection measurements obtained over a range of angles results in a 2D data set $P(r, \theta)$ referred to as the *singogram*. This nomenclature is used because the Radon transform of the Dirac delta function is sinusoidal in shape (Figure 2.3). The Radon transform of a several small objects (e.g. several points on a 2D grid) will appear as a number of blurred sinusoids with varying amplitudes and phases. Two important properties of the Radon transform are:

1. $P(r, \theta)$ is periodic in θ with period 2π:

$$P(r, \theta) = P(r, \theta + 2\pi) \tag{2.18}$$

2. $P(r, \theta)$ is symmetric in θ with period π:

$$P(r, \theta) = P(-r, \theta \pm \pi) \tag{2.19}$$

Therefore, while projection measurements may be acquired for a full rotation:

$\theta \in [0, 2\pi)$, it is only necessary to consider the subset $\theta \in [0, \pi)$. In *sinogram space*, the horizontal axis represents the distance r from the origin and is defined by the number of detector channels. The vertical axis represents the projection angle θ. A sinogram computed for M views (i.e. M projection angles) and N detector samples can thus be represented as an M x N matrix:

$$P[n, m] = P(n\Delta r, m\Delta\theta) \tag{2.20}$$

To avoid aliasing, the detector spacing Δr and the rotational increment between views $\Delta\theta$ need to be set appropriately.

2.3 Image Reconstruction

Given a sinogram $P(r, \theta)$, the computational problem in CT is to determine the function $\mu(x, y)$ representing the distribution of linear attenuation coefficients. In other words, a formulation for the inverse Radon transform is required:

$$f(x, y) = \mathcal{R}^{-1}\{P(r, \theta)\} \tag{2.21}$$

2.3.1 Back-Projection

The *back-projection* method (also known as the summation method [Old61] or the linear superposition method [KE63]) is the original and most basic technique for reconstructing a 2D distribution from multiple 1D projections. Back-projection involves taking each view and *smearing* it along the path upon which it was acquired. While offering a rather crude solution to the CT-reconstruction problem, back-projection is a crucial component of the most popular CT reconstruction technique - Filtered Back-Projection (FBP). Mathematically, the back-projection method is represented as:

$$\hat{\mu}(x, y) = \int_0^{\pi} P(x\cos\theta + y\sin\theta, \theta)d\theta \tag{2.22}$$

Therefore, for a given line (or ray) $L_{r,\theta}$ (where $r = x\cos\theta + y\sin\theta$), the back-projection method assigns the projection value $P(r, \theta)$ to all the points (x, y) falling on that line. This is repeated for all $\theta \in [0, 2\pi)$. The final back-projected density at each point (x, y) is thus the sum of all the ray-sums (projection measurements)

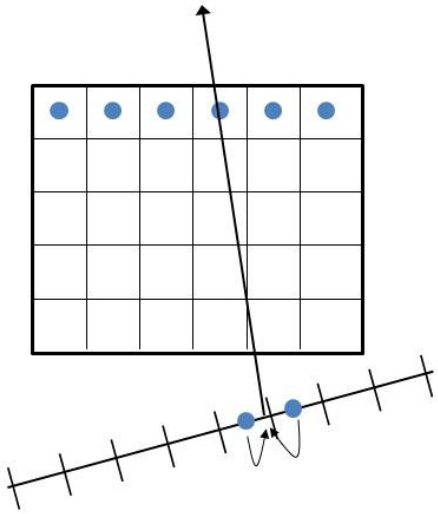


Figure 2.4: Bilinear interpolation process used in discretised back-projection [Man01].

passing through that point. Equation 2.22 is discretised as follows:

$$\hat{\mu}(x_i, y_j) = \sum_{m=1}^{M} P(x_i \cos\theta_m + y_j \sin\theta_m, \theta_m)\Delta\theta \tag{2.23}$$

The discrete implementation of the back-projection algorithm requires an interpolation step (Figure 2.4): for each view, a projection line is drawn through every pixel (x, y). The intersection of this line with the detector array is computed. This intersection (given by $x_i \cos\theta_m + y_j \sin\theta_m$) does not always coincide exactly with the discrete detector positions. Therefore, the corresponding projection value for each intersection is calculated by interpolating between the neighbouring values (i.e. the projection readings at the two nearest discrete detector positions).

Reconstruction by back-projection has several major limitations. Since each ray-sum is applied to all points along that ray and not only to regions of high-density, the reconstructed image will have non-zero values in regions outside of the object of interest. Consequently, the non-zero region in the reconstructed image is larger than the area of increased activity in the object. If a low number of projections is used the rays appear as a star. As the number of projections is increased, the star shape fades into a general blurring of the image [Pre12] (Figure 2.5). These defects are most evident in discrete areas of high density. The Point Spread Function (PSF) of the back-projection operator (the response of the operator to a point source) is circularly symmetric with a magnitude that decreases with the reciprocal of its radius. Mathematically, this can be described as the convolution of the true image with the kernel $1/r$:

$$\hat{f}(x, y) = f(x, y) \star \frac{1}{r} \tag{2.24}$$

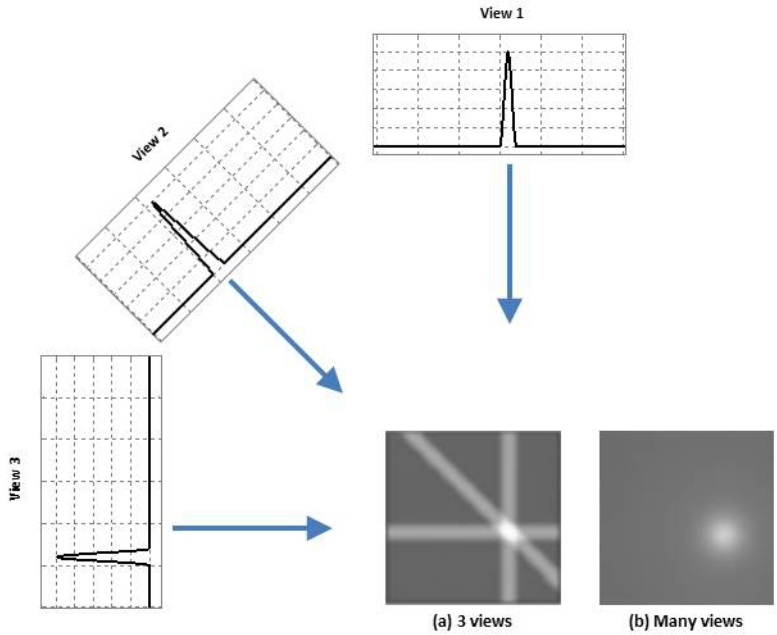


Figure 2.5: Back-projection of a point source. (a) Using only three views results in star-shaped artefacts. (b) Using many views results in a global blurring [Smi03].

2.3.2 The Fourier-Slice Theorem

The Fourier-slice theorem (equivalently the central-slice or projection theorem) is central to tomographic reconstruction. The theorem states that the 1D Fourier transform of a parallel projection of an object $f(x, y)$ obtained at angle θ, is equal to a slice of the 2D Fourier transform of the original object $f(x, y)$ obtained in the same direction θ.

Let $F(k_x, k_y)$ be the 2D Fourier transform of the function $f(x, y)$:

$$F(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi(k_x x + k_y y)i} dx dy \tag{2.25}$$

The inverse Fourier transform is given by:

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(k_x, k_y) e^{-2\pi(k_x x + k_y y)i} dk_x dk_y \tag{2.26}$$

where $i = \sqrt{-1}$ and k_x and k_y are the polar frequency parameters (i.e. the spatial frequencies in the x and y directions respectively):

$$\begin{pmatrix} k_x \\ k_y \end{pmatrix} = k \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \tag{2.27}$$

Let $\mathcal{P}(k,\theta)$ be the 1D Fourier transform of the projection $P(r,\theta)$ with respect to the variable r:

$$\mathcal{P}(k,\theta) = \int_{-\infty}^{\infty} P(r,\theta) e^{-2\pi kir} dr \tag{2.28}$$

The Fourier-slice theorem then states:

$$\mathcal{P}(k,\theta) = F(k_x,k_y) \quad \text{if} \quad \begin{cases} k_x = k\cos\theta \\ k_y = k\sin\theta \end{cases} \tag{2.29}$$

The reader is referred to [Hsi03] for a proof of this theorem. Given sufficient projection data (i.e. obtained over the range $\theta \in [0,\pi)$), the entire 2D Fourier transform of the original object can be obtained. It follows that the original object can be reconstructed by performing a simple 2D inverse Fourier transform (Equation 2.26). The Fourier-slice theorem thus provides a solution to the inverse Radon transform by use of 1D and 2D Fourier transforms. Given a discrete set of projection data $P(r_n,\theta_m)$, the *direct Fourier reconstruction* of the function $f(x,y)$ is performed as follows [Man01]:

1. For all θ_m where $\theta_m \in [0,\pi)$, compute the 1D Discrete Fourier Transform (DFT) of $P(r_n,\theta_m)$ with respect to r_n:

$$\mathcal{F}_1 P(r_n,\theta_m) = \mathcal{P}(k_{n'},\theta_m) \tag{2.30}$$

2. Place $\mathcal{P}(k_{n'},\theta_m)$ on a polar grid $\forall\theta_m$.

3. Resample $\mathcal{P}(k_{n'},\theta_m)$ to a Cartesian grid by bilinear interpolation (to allow for the application of the Fast Fourier Transform (FFT) [Bra90]).

4. Compute the FFT $F(k_{x'_i},k_{y'_j})$ from the resampled $\mathcal{P}(k_{n'},\theta_m)$.

5. Compute the 2D inverse DFT of $F(k_{x'_i},k_{y'_j})$:

$$f(x_i,y_j) = \mathcal{F}_2^{-1}\left\{F(k_{x'},k_{y'})\right\} \tag{2.31}$$

Interpolation errors in the frequency domain (associated with the bilinear interpolation in Step 3 above) are not localised in the spatial domain and thus affect the entire spatial-domain image. The associated artefacts are the main reason that Direct Fourier reconstruction is not generally used in CT.

2.3.3 Filtered Back-Projection

The Filtered Back-Projection (FBP) algorithm is a reconstruction technique that overcomes the data resampling limitations of the direct Fourier method as well as the star-shaped artefacts associated with the characteristic $1/r$ blurring effect of the simple back-projection procedure.

The 2D inverse Fourier transform in Equation 2.26 may be transformed from the Cartesian coordinate system (k_x, k_y) to an equivalent polar coordinate system (k, θ) using the transformation defined in Equation 2.29, thereby avoiding the issue of interpolation. If $|k|$ is the absolute value of the polar coordinate transformation, the 2D inverse Fourier transform is given by:

$$f(x,y) = \int_0^{\pi} \int_{-\infty}^{\infty} \mathcal{P}(k,\theta) e^{-2\pi i k r} |k| dk d\theta \tag{2.32}$$

Letting $\mathcal{P}^*(k,\theta) = \mathcal{P}(k,\theta)|k|$ and $P^*(r,\theta) = \int_{-\infty}^{\infty} \mathcal{P}^*(k,\theta) e^{2\pi i k r} dk$ Equation 2.32 simplifies to:

$$f(x,y) = \int_0^{\pi} P^*(r,\theta) d\theta \tag{2.33}$$

Equation 2.33 requires that the projection data be available for the angular range $[0, \pi]$. Mathematically, $\mathcal{P}^*(k,\theta)$ is obtained by multiplying the Fourier domain sinogram data with the Fourier response of a ramp filter. Since multiplication in the Fourier domain is equivalent to convolution in the spatial domain, $\mathcal{P}^*(k,\theta)$ may also by obtained by convolving the measured sinogram with the impulse response $h(r)$ of the ramp filter (obtained via the 1D inverse Fourier transform of $|k|$):

$$h(r) = \mathcal{F}^{-1}\{|k|\} = \int_{-\infty}^{\infty} |k| e^{2\pi i k r} dk \tag{2.34}$$

Since the ramp filter $|k|$ is a high-pass filter, it propagates high frequency noise. This is typically counteracted by applying a windowing function to the filter (i.e. discretising or band-limiting the filter). The simplest approach is to use a rectangular windowing function (in the Fourier domain), resulting in what is commonly referred to as the Ram-Lak filter (named after its creators Ramachandran and Lakshiminarayanan). Rectangular windowing, however, results in ringing artefacts and aliasing. Windowing functions with smoother roll-offs are thus more effective. Examples include classical signal processing functions (e.g. Hamming, Hanning, Butterworth etc.) as well as dedicated CT reconstruction functions (e.g.

Shepp-Logan).

In summary, reconstruction by Filtered Back-Projection may be achieved as follows:

1. Filter the measured sinogram data with a band-limited filter:

$$P^*(r,\theta) = P(r,\theta) \star h(r) \tag{2.35}$$

2. Back-project the filtered sinogram:

$$f(x,y) = \int_0^{\pi} P^*(r,\theta)d\theta \quad \text{where} \quad r = x\cos\theta + y\sin\theta \tag{2.36}$$

Filtered back-projection remains the most widely implemented reconstruction technique in CT imagery.

2.3.4 Fan-Beam to Parallel-Beam Rebinning

While the aforementioned FBP algorithm assumes parallel-beam sinogram data, the majority of modern CT scanners employ fan-beam geometries. While dedicated fan-beam FBP algorithms do exist [HLN76, HL80], a more popular approach is to resample the fan-beam data to obtain the equivalent parallel-beam data and then to apply the traditional FBP reconstruction approach [Wan77, HL80, PL77].

Consider a fan-beam geometry defined by the angles β, between the focus (fan-apex) and the y-axis and α, between a given ray and the centre line of the fan (Figure 2.6). The parallel-beam coordinates (r,θ) are related to the fan-beam coordinates according to the following transformation:

$$\begin{aligned} \theta &= \alpha + \beta \\ r &= R\sin\alpha \end{aligned} \tag{2.37}$$

where R is the focus-to-isocentre distance. In contrast to the parallel-beam geometry, an angular range of $\beta \in [0,\pi]$ is not sufficient to capture all possible line measurements. Instead, a range of $\beta \in [0, \pi + \Delta\alpha]$ is required, where $\Delta\alpha$ is the total fan-angle. Assuming the fan-beam data has been obtained for lines (α_n, β_m) for $n = 1 \cdots N$ and $m = 1 \cdots M$, the corresponding parallel-beam measurement may be computed as follows [Man01]:

$$P(r_{n'}, \theta_{m'}) = \sum_{i=0}^{1} \sum_{j=0}^{1} c_i d_i P(\alpha_{n'+i}, \beta_{m'+j}) \tag{2.38}$$

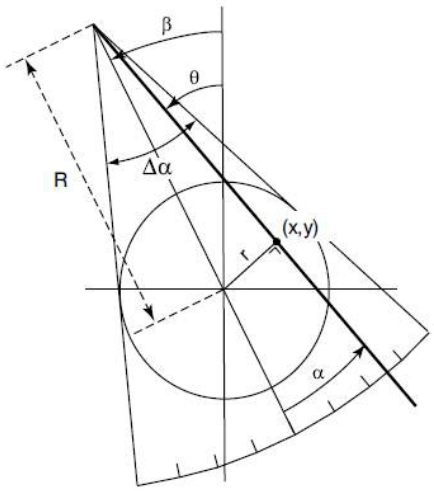


Figure 2.6: Fan-beam geometry: β is the angle between the fan centre line and the y-axis; α is the angle between a given ray and the fan centre line; $\Delta\alpha$ is the total fan-angle and R is the focus-to-isocentre distance [Man01].

where:

$$\begin{aligned} c_0 &= (\alpha_{n'+1} - \alpha)/(\alpha_{n'+1} - \alpha_{n'}) \\ c_1 &= 1 - c_0 \\ d_0 &= (\beta_{m'+1} - \beta)/(\beta_{m'+1} - \beta_{m'}) \\ d_1 &= 1 - d_0 \end{aligned} \tag{2.39}$$

and $\alpha_{n'}$, $\beta_{m'}$ are the largest values satisfying:

$$\begin{aligned} \alpha_{n'} &\leq \alpha = \arcsin \tfrac{r_{n'}}{R} \\ \beta_{m'} &\leq \beta = \theta_{m'} - \alpha \end{aligned} \tag{2.40}$$

2.4 CT Image Quality

The quality of a CT image is typically measured according to resolution (high-contrast spatial resolution, low-contrast spatial resolution and temporal resolution); CT number uniformity; CT number accuracy; image noise and image artefacts [KKRH+00]. A brief overview of each is provided here.

2.4.1 Image Resolution

Spatial resolution measures the degree to which lines can be resolved in an image and is perhaps the most important measurement regarding the clarity of an image. In CT, reference is generally made to high and low-contrast spatial resolution [Man01].

High-contrast spatial resolution refers to the ability of the scanner to resolve closely spaced or small foreground objects [KKRH+00]. Such objects are said to have a high spatial frequency, while large and/or widely spaced objects have a low spatial frequency. Spatial resolution is generally defined in terms of line pairs per millimetre (lp/mm) (i.e. the number of independent pixel values per unit length [PB99]). The in-plane spatial resolution of a CT image is considerably worse than that of a conventional radiograph: the typical limiting resolution of a conventional X-ray screen film is 4-20 lp/mm while the limiting resolution of CT is only 0.5-2 lp/mm [Man01]. The modulation transfer function (MTF) is a plot of spatial frequency as a function of light amplitude (or image sharpness) and is used to measure the response (i.e. performance) of a system to different frequencies. The response of an ideal system would be independent of frequency and thus have a flat MTF curve. Such a system would be able to reproduce all objects perfectly, regardless of their size or proximity to other objects. In reality, however, the magnitude of the frequency response (i.e. the image fidelity) decreases rapidly with increasing frequency [PB99]. In multi-slice CT both the in-plane (i.e. within the imaging plane) and cross-plane (i.e. inter-slice) spatial resolution is of importance.

Factors impacting the in-plane spatial resolution of a CT image include: the physical properties of the scanner; sampling rates; scanning parameters (e.g. slice collimation, tube voltage and tube current) and the choice of reconstruction parameters [KS88, Man01]. The majority of these factors are determined (and hence optimised) by the operator. Image-processing techniques are generally not used to improve the spatial resolution of images.

The introduction of multi-slice CT technologies has increased the relevance of the cross-plane resolution. A major advantage of contemporary medical multi-slice CT scanners is that they offer near isotropic resolution. That is to say, the CT system exhibits equivalent point-spread functions in every spatial direction (i.e. the resolution is uniform across all viewing planes). Isotropic resolution is essential if undistorted visualisation of small structures, independent of the viewing plane, is desired. Radially symmetric 2D isotropy in the $x-y$ (i.e. transverse or axial) plane is virtually guaranteed with current CT reconstruction techniques. With the introduction of multi-slice CT and volumetric imaging, however, viewing is not restricted to this plane. Although much progress has been made in achieving 3D isotropy in medical CT imaging, this is not the case in the security-screening CT domain. While the resolution in the $x-y$ plane is generally isotropic, the resolution in the z-direction usually differs substantially. The effect of this anisotropic cross-plane resolution on the imaging and visualisation of small structures is significant. Beyond the distortion of small structures in the image, noise levels are also impacted by the spatial resolution, leading to directional noise in the

non-isotropic planes. Much research has been conducted in determining the optimal scanning and reconstruction parameters to achieve near-isotropic resolutions [Mah02, Kal95].

The low-contrast resolution of a CT scanner refers to the ability of the scanner to differentiate objects which differ only slightly in intensity from their backgrounds [KS88]. While conventional radiography can discriminate a density difference of approximately 10%, modern CT can detect differences as small as 0.25% [Man01]. This remarkable Low-Contrast Detectability (LCD) is one of the major advantages of CT over conventional radiography. The LCD of a CT scanner is affected by both the size and intensity difference (with respect to its background) of an object. The other major factor affecting LCD is the noise level in the image. While a higher signal-to-noise ratio typically corresponds to an improved LCD, in some scenarios a trade-off exists. Increasing the slice thickness, for example, reduces the noise level in the image while also negatively impacting the visibility of small structures (i.e. degrades the low-contrast resolution). Techniques for improving the LCD are generally operator based, as opposed to software based, and involve selecting optimal scanning parameters.

The temporal resolution of a CT scanner refers to the efficiency of the scanner in producing clear images of moving objects (for example, cardiac imaging in medical CT) [Man01]. The most obvious way of minimising the effects of motion is to increase the scan speed. The majority of modern scanners rely on a high scan speed and the half-scan reconstruction algorithm when imaging moving objects [KKRH+00, Man01].

2.4.2 CT Number Accuracy and Uniformity

CT number accuracy refers to the proximity of a scanner-generated CT number to the theoretical CT number of the material under investigation. Ideally, when scanning a uniform phantom, the CT numbers should remain constant across the entire phantom. The degree of this consistency defines the uniformity of the CT number. In reality, uniformity is negatively affected by numerous factors (e.g. beam hardening, scattered radiation etc.) and can thus only be maintained within a reasonable range (typically 2HU). The chosen reconstruction algorithm has a significant impact on the resulting CT number [KKRH+00].

2.4.3 Noise

In computed tomography there are three predominant factors that contribute to image noise: quantum noise, the inherent physical limitations of the scanner and the chosen reconstruction parameters [Man01].

Quantum (or shot) noise arises due to statistical fluctuations inherent in the

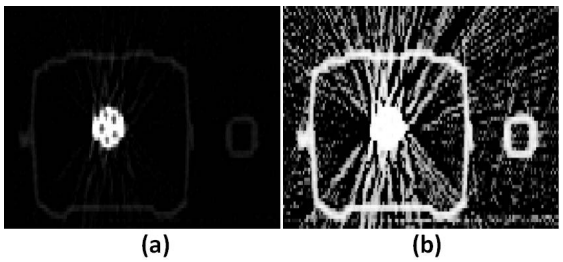


Figure 2.7: Baggage CT image before (a) and after (b) Histogram Equalisation (HE). HE highlights the background noise previously not visible due to the high dynamic range of baggage-CT imagery.

detection of a finite number of photons (or X-ray quanta) at the detector of a CT scanner [Han81]. Quantum noise is most evident when the number of photons emitted by the X-ray tube is sufficiently small such that uncertainties due to the Poisson distribution (describing the occurrence of independent random events) are of significance [HC94]. The only certain way of reducing the effects of statistical quantum noise in a CT image is to increase the number of X-ray photons emitted (by increasing the radiation dose). While quantum noise is an unavoidable statistical occurrence, its quantity is further influenced by the scanner parameters, scanner efficiency, detector efficiency and the physical properties of the object being scanned [Man01].

As with any image acquisition system, the mechanical components of particular scanners pose limitations on the image quality. Factors such as electronic noise, the data acquisition system and scattered radiation all contribute to the level of noise in a CT image. The degree to which these factors impact the noise level varies from scanner to scanner.

The reconstruction procedure further contributes to the noise-level in CT imagery. The majority of noise in digital signals presents as high frequency signals [PB99]. In order to obtain a high-resolution image, the reconstruction kernel thus needs to preserve the high-frequency contents of the sinograms. This has the unfortunate consequence of additionally preserving (or enhancing) the high-frequency noise in the sinograms. A trade-off between noise and resolution thus exists and a suitable balance needs to be determined [KS88].

Owing to the characteristically high dynamic range of baggage-CT imagery, the presence of background noise is not always obvious [ZPA10b]. Image enhancement, using Histogram Equalisation (HE) [SB10], for example, reduces the effective dynamic range of the image and often reveals significant quantities of background projection noise (Figure 2.7). Popular image denoising techniques may be implemented to improve the signal-to-noise ratio in CT imagery - this is explored further in Chapter 4.

2.4.4 Artefacts

Barrett and Keat [BK04] define a CT image artefact as: *"any systematic discrepancy between the CT numbers in the reconstructed image and the true attenuation coefficients of the object."* CT images are inherently more susceptible to artefacts than conventional radiographic images. Each individual 2D CT image is reconstructed from a massive number ($\sim 10^6$) of independent projection readings. Furthermore, the contribution of every projection sample is not limited to a single point in the reconstructed image since the mechanics of the Filtered Back Projection (FBP) process dictate that a single point in the projection profile is mapped to a straight line in the reconstructed image [KS88] (refer also to Section 2.3). In contrast to conventional radiology, projection reading errors are thus not localised, significantly increasing the probability of artefacts in the reconstructed images. These artefacts generally manifest in one of the following ways: streaking, rings, bands or shading [BK04].

Streaking artefacts arise due to errors of isolated projection readings [Man01]. The errors are enhanced during reconstruction and mapped into intense dark and bright lines radiating across the reconstructed image. For an error-free projection, the FBP process maps each individual point in the projection profile to a straight line in the image domain. Positive and negative contributions from neighbouring lines are combined, ensuring that no unwanted straight lines appear in the reconstructed image. If there are inconsistencies in the projection data, however, the positive and negative contributions are not combined correctly, resulting in streaks in the reconstructed image [KS88]. When streaking artefacts appear in large quantities they can significantly degrade the quality of an image (Figure 2.9).

When isolated errors in the projection readings occur over a range of views (i.e. no longer isolated) the back-projection process maps them to a series of straight lines in the image domain. These straight lines occur at a fixed distance from the iso-centre, due to the rotational motion of a third-generation detector. The terminal points of the straight lines thus cancel out, forming ring-like artefacts in the reconstructed image. In a similar fashion to the formation of streaking artefacts, small errors in the projection data are magnified during the filtering stage of reconstruction [Man01].

Shading artefacts are most prominent in the vicinity of high-density objects. They are generally caused by a gradual deviation of a group of projection readings from their true measurements.

The aforementioned CT artefacts are caused by a range of phenomena, most notably: beam hardening, the partial volume effect, partial volume averaging, photon starvation, undersampling (in accordance with the Nyquist sampling theory [PB99]), non-stationary objects, the cone-beam effect and scattered radiation

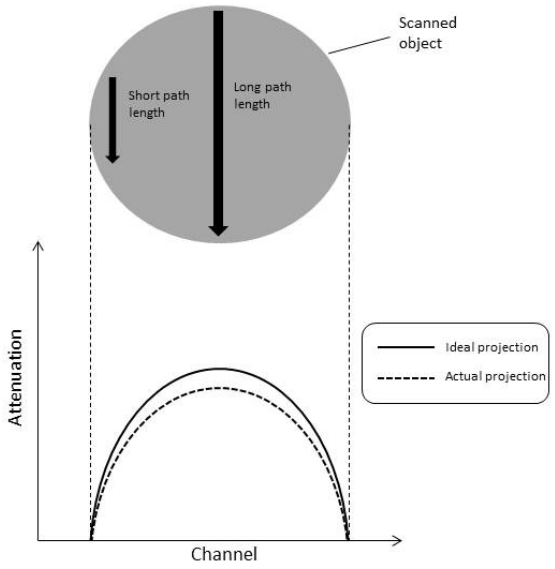


Figure 2.8: The beam hardening effect. Ideal and actual attenuation profiles obtained for an X-ray beam passing through a uniform cylindrical phantom. Adapted from [BK04].

[BK04]. In the presence of high-density objects, such as metals, the effects of noise, beam hardening, scattered radiation, photon starvation and the partial volume effect cause the FBP algorithm to produce reconstructions characterised by streaking and star-shaped artefacts (Figure 2.9). These are the artefacts which are most prominent in the low-quality, complex volumetric imagery obtained from the security-screening domain and used throughout this work. It is not uncommon for the quality of these images to be significantly degraded by such artefacts, making their reliable interpretation extremely challenging (Figure 2.9). The effects and reduction of metal-streaking artefacts in baggage-CT imagery are thus major focal areas of this work and are addressed in greater depth in Chapters 3 and 4. Only a brief overview of the relevant principles (beam hardening, scattered radiation, photon starvation and the partial volume effect) is provided below.

Beam hardening is the process by which the mean energy of a polychromatic X-ray beam increases as it passes through an object. The rate of absorption of the photons of an X-ray beam passing through an object is proportional to the energies of those photons. Subsequently, the mean energy of an X-ray beam increases (i.e. it becomes 'harder') as it traverses an object (Figure 2.8).

As an X-ray beam passes through an object, its photons deviate (or scatter) from their initial straight-line trajectories leading to shading artefacts. Collimators are typically used to 'narrow' (or realign) the deviated X-ray beam, thereby eliminating the scattered photons and reducing (in part) the associated artefacts [KKRH+00].

The partial volume effect (not to be confused with partial volume averaging) leads to shading artefacts in several ways. The most common of which occurs when an X-ray beam partially intersects a dense object. In such scenarios, the object

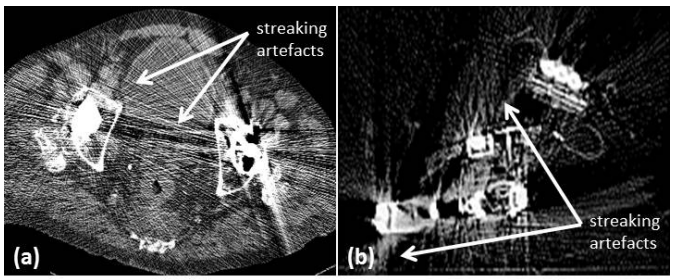


Figure 2.9: Example of streaking artefacts caused by metallic objects. (a) Medical scan containing metallic hip replacements (b) Security screening scan containing metallic handgun

is only correctly scanned at certain tube positions. The partial volume effect is especially problematic in regions where there are large variations (in object shape and/or size) along the $z-$axis and thus a high probability of the X-ray beam encountering off-centre objects. A narrow acquisition section-width is usually an adequate solution, although the corresponding increase in image noise needs to be considered [Man01]. In contrast, partial volume averaging refers to the scenario where a CT number of a given voxel is representative of the average attenuation of the materials constituting that voxel [BK04]. Partial volume averaging, which typically manifests as streaking and bands, is always present to some degree and can never be eliminated entirely [KKRH$^+$00].

Photon starvation occurs when insufficient photons reach the detectors, usually as a consequence of a highly-attenuating material or region in the scanning FoV. The resulting projections are characterised by high levels of noise and/or missing data, leading to streaking artefacts in the reconstructed images. The effects of photon starvation are exacerbated by poor object positioning, incorrect scanning parameters and the inherent physical limitations of the scanner. [KS88].

2.5 Summary

An overview of the fundamental principles of X-ray CT has been presented, with a particular focus on those principles deemed central to the latter chapters of this work.

While the causes of noise and artefacts in CT imagery are far-reaching, virtually all types originate in the data acquisition phase of the CT process, prior to the reconstruction of the CT images. In most cases, in fact, the acquisition errors/inconsistencies are greatly magnified by the widely-used Filtered Back-Projection (FBP) reconstruction algorithm. A sound understanding of the principles of tomographic reconstruction is thus invaluable in the development of effective noise and artefact-reduction techniques (Chapter 4). To this end, the principles of X-ray generation and detection, tomographic reconstruction and the causes and effects of noise and artefacts in CT imagery have been addressed.

Chapter 6 of this work proposes Dual-energy Computed Tomography (DECT) as a means of capturing information related to the chemical characteristics of the CT scans. While the principles of DECT have not been addressed in this chapter, the fundamentals of the physical processes governing the interactions of X-rays with matter (which are central to the understanding and development of DECT-based techniques) have been discussed.

While the theoretical overview presented here is sufficient for the purposes of this work, a more comprehensive and detailed analysis of the theory of X-ray CT may be found in the literature [KS88, Dea93, CDMC90, Hsi03, KBK03].

Chapter 3

Literature Review

This chapter presents an overview of the prior literature relating to and motivating the research presented in this thesis. The broader objective of this work is the development of a fully-automated framework for the classification of objects in low-quality, complex volumetric imagery. From the outset, it is apparent that this framework will require at least the following stages: noise and artefact reduction (based on the theory presented in Chapter 2.4); segmentation (to isolate target objects) and classification. The following review addresses and evaluates the current state-of-the-art in the research topics relevant to these components.

Sections 3.1 and 3.2 investigate prior works addressing the reduction of noise and artefacts in low-resolution, cluttered volumetric CT imagery, with a particular focus on the reduction of metal-streaking artefacts. Thereafter the recent advances in classification and the automated classification of objects in complex volumetric imagery encountered outside of the medical domain in particular, are presented (Section 3.3). The availability of CT imagery captured at different energies allows for the extraction of the chemical characteristics of the materials present in a scan using dual-energy decomposition techniques. Section 3.4 reviews the prior applications and successes of dual-energy-based techniques in an attempt to determine their potential for use in materials-based segmentation and classification. Finally, Section 3.5 presents the recent advances in the automated segmentation of volumetric imagery, drawing predominantly from the medical literature.

Portions of this chapter have previously been published as [MMvS+13, MB14a].

3.1 Denoising

Previous work addressing the reduction of noise in low-resolution, cluttered volumetric CT imagery is limited. Zhou *et al.* [ZPA10b] use image enhancement to remove background noise and improve the resolution of baggage-CT imagery. The approach is comprised of two stages: noise removal and image enhancement. The noise-removal step relies on the notion that much of the projection noise present

in baggage-CT imagery is characterised by very low intensity values relative to the characteristically high dynamic range of such imagery. A simple thresholding-based approach is thus proposed for noise removal. In particular, a given 2D input image is separated into two parts via Alpha-Weighted Mean (AWM) intensity thresholding: 1) an object image (containing the 'valuable' information in the image) and 2) a noise image (which is subsequently discarded). A second threshold is then used to further subdivide the object image into two sub-images: an upper image, containing the brighter regions of the object image and a lower image, containing the darker, yet still informative, regions of the object image. The two intensity thresholds are computed as scalar multiples of the mean intensity of the input image (hence the term 'alpha-weighted mean') and are chosen empirically. The upper and lower sub-images are then enhanced by intensity clipping and Histogram Equalisation (HE) [SB10] respectively. The final, enhanced CT image is computed as the summation of the enhanced sub-images.

The performance of the so-called Alpha-Weighted Mean Separation and Histogram Equalisation (AWMSHE) technique is evaluated using a novel enhancement measure which quantifies the improvement in image contrast using second-order derivatives. In terms of this measure as well as a standard qualitative performance analysis (visual comparisons), the AWMSHE technique is shown to improve the contrast and visual quality of baggage-CT imagery. Performance analysis, however, is focussed on the improvement in image contrast and little mention is made regarding the effectiveness of the denoising stage of the technique (particularly from a quantitative perspective). It is also worth noting that the images used in the study [ZPA10b] are largely free of metal artefacts and the efficacy of the method in terms of metal artefact reduction is thus unclear. Furthermore, performance is found to be sensitive to the chosen thresholds, which need to be manually adjusted on a per-slice basis, rendering the algorithm inefficient when applied to large volumes.

Despite the vast resource of general denoising literature, there do not appear to be any further works concerned specifically with low-quality, complex volumetric CT imagery. An extension of the work of Zhou *et al.* [ZPA10b], by evaluating the performance of popular denoising techniques (obtained from the broader image processing literature) within this previously unconsidered context is a necessary task and is presented in Chapter 4.

3.2 Metal Artefact Reduction (MAR)

Metal artefacts can corrupt CT images such that they become difficult to interpret and of limited diagnostic value. Filtered Back-Projection (FBP) is the CT-reconstruction algorithm that is most widely used in daily clinical practice. The

FBP algorithm, however, is based on an analytical inversion of the Radon transform [Dea93], and only yields satisfactory reconstructions in ideal conditions. In the presence of high-density objects, such as metals, the effects of beam hardening, scattered radiation, photon starvation, noise and the partial volume effect cause the FBP algorithm to produce reconstructions characterised by streaking (Figure 2.9) and star-shaped artefacts. Such streaking can degrade the quality of the image tremendously, often obscuring valuable details and detracting from the usability of the image. It has been shown that for typical abdominal geometries, as little as 1 cm of iron or 2 cm of titanium can be sufficient to produce this effect [KKRH+00]. Efficient strategies to minimise the impact of metal artefacts on the readability of CT images are invaluable.

The problem of Metal Artefact Reduction (MAR) in CT has been widely studied with over 100 publications in the last 10 years. While many of these published techniques claim fairly substantial improvements to previous methods, these claims are often based on rather limited comparisons. For instance, a large portion of publications base their claims solely on qualitative comparisons made with the standard linear-interpolation-based approach [KHE87, GP81] - a technique which is widely accepted to perform poorly in complex (i.e. most real-world) scenarios. Comprehensive comparative studies, where both the qualitative as well as quantitative performance of state-of-the-art methods are compared, are surprisingly limited. Rinkel *et al.* [RDF+08] compared the performance of 3 fairly simple interpolation-based approaches with the primary aim of determining the value of MAR when detecting small features near large metallic objects. The primary objective of this previous study was thus not to determine the optimal available MAR technique. Golden *et al.* [GMB+11] compared the performance of 4 MAR techniques (3 sinogram-completion-based approaches and 1 iterative approach) with the aim of determining their effectiveness in improving the diagnostic quality of medical-CT images (determined by the independent assessments of 3 radiologists). There are presently no reviews or quantitative comparative studies where a broad range of state-of-the-art MAR techniques are considered.

The vast majority of CT-based literature, and MAR-based CT literature in particular, is found in the medical domain. The development of novel MAR techniques or the evaluation of existing, medical MAR techniques in novel applications of CT imagery (outside of medicine) are extremely limited [MMFB12, XZX+09, GSK+12]. The advantages of CT imagery, however, extend beyond the medical domain and its applications in other fields are widespread, ranging from micro CT for non-invasive imaging of wood anatomy to the scanning of baggage for potential threat items in aviation-security settings [vKD05]. While the challenges posed by metal objects extend to all applications of CT imagery, the differences in the nature of medical images and those encountered in other domains (e.g. Section 1.2)

mean that the MAR techniques which have been successfully applied to medical images will not necessarily be successful when applied to non-medical-CT images [MMFB12].

In X-ray CT the most widely implemented reconstruction technique is the analytical Filtered Back-Projection (FBP) (Section 2.3). According to reconstruction theory, FBP yields fast and accurate reconstructions of the attenuation function for ideal (or near ideal) projections which contain a sufficient number of projection samples and low degrees of noise, beam hardening and other imperfections [Hsi03]. In reality, projections are only approximations of the ideal case. This is due to the finite number of projection samples; Poisson noise in the projection data; beam hardening and scattered radiation (Section 2.4). When these approximations are relatively small (as is often the case), FBP still produces satisfactory reconstructions. When the errors become large, however, the reconstructed images become corrupted by artefacts [BK04].

Metal objects in particular cause significant artefacts in CT images [BK04]. In an extensive simulation study, De Man *et al.* [MND+99] cite beam hardening (the preferential attenuation of low-energy photons in a polychromatic X-ray beam [KCWM12]), scattered radiation, photon (projection) noise and the exponential edge-gradient effect (trans-axial non-linear partial volume effect) as the predominant causes of metal-streaking artefacts in high resolution 2D fan-beam CT images. While additional factors contribute to metal artefacts in CT imagery [BK04, KCWM12], the aforementioned factors are considered dominant in the remainder of this work.

The majority of the published Metal Artefact Reduction (MAR) techniques fall into one of four categories: sinogram (or projection) completion methods (Section 3.2.1); iterative methods (Section 3.2.3); hybrid methods (Section 3.2.4) and miscellaneous methods (Section 3.2.4).

3.2.1 Sinogram-Completion Methods

The vast majority of sinogram-completion-based approaches to MAR rely on reconstructions using the FBP approach. Under certain conditions, however, FBP produces reconstructions containing bright and dark streaking artefacts Section 2.4. When streaking artefacts appear in large quantities they can result in an image which is significantly degraded (Figure 2.9).

Sinogram completion methods typically regard these inconsistencies in the projection data as missing data and use various techniques to estimate the correct projection values. The vast majority of sinogram-completion-based approaches adhere to the following framework: metal segmentation, sinogram completion, final image reconstruction.

Metal object segmentation involves isolating the metal objects in the original

CT image and creating a metal-only image. The segmentation results are used to determine which regions of the original sinogram are corrupted and need to be adjusted or replaced. Sinogram completion involves the actual replacement of these corrupted points via a broad range of techniques [KHE87, AAA+10, YZB+07, MRL+10, TMK+06, JR09, ZBWW02, LBY+10, DZX+08]. The final, artefact-reduced image is created by back-projecting the interpolated sinogram and re-inserting the metal-only image into this corrected image.

Segmentation by thresholding, followed by direct interpolation-based sinogram completion (i.e. interpolating the sinogram data on either side of the metal traces) is generally considered the simplest approach to MAR in CT and is used in many recent studies as a benchmark for performance comparisons [AAA+10, AAAZ10, MND+00, DZX+08, JR09]. Although such direct interpolation-based MAR approaches were popular in early studies [KHE87, KKSF90a], many recent studies have highlighted their limitations [MRL+10, AAAZ10, RLP+03, AAA+10, MB09, Man01, ZBWW02].

3.2.1.1 Metal Object Segmentation

The most widely implemented segmentation method employs simple thresholding, whereby a single threshold is used to distinguish the metal from the non-metal objects in the image [KHE87, AAAZ10, AAA+10, DZX+08, JR09, MMFB12]. Thresholding exploits the fact that the CT values of metals are extremely high (due to their high atomic numbers), especially relative to other materials. Despite the simplicity of this approach, thresholding generally produces reasonably accurate results and has been widely implemented, even in some of the most complex MAR techniques [AAA+10, DZX+08]. Nonetheless, several studies have claimed that minor segmentation errors may have significant detrimental effects on the overall performance of the MAR technique [YZB+07, MRL+10, LFN09, LBY+10]. These effects generally manifest as a loss of information from both the structures surrounding the metal objects as well the metal objects themselves, leading to the generation of secondary streaking artefacts in the reconstructed image [MB09, MND+99]. More sophisticated segmentation processes such as the mean-shift technique [YZB+07, CM02] and Mutual Information Maximised Segmentation (MIMS) [LBY+10] claim to better preserve edge and contrast information of the metal objects and their direct surroundings. It is worth noting that the majority of the aforementioned studies involve more sophisticated processes in the other stages of the MAR procedure (e.g. complex sinogram-completion methods, pre-filtering and/or post-filtering etc.), making it unclear if the improved segmentation alone contributes to the improved results.

3.2.1.2 Sinogram Completion

The simplest sinogram-completion techniques employ basic interpolation-based approaches to replace the corrupted data (e.g. linear interpolation [KKSF90b]; spline interpolation [AAAZ10, RLP+03]).

Although simple and convenient to implement, direct interpolation is generally sufficient when only a single, small metal object is present in the Field-of-View (FoV). When larger and/or multiple metal objects are present, however, the reliability of the interpolated values decreases. Muller and Buzug [MB09] have demonstrated that the biggest disadvantage of sinogram correction by direct interpolation is that all edge information lying on the beams passing through the metal objects is lost - in other words, the loss of edge information affects the entire image and is not restricted to the edges in the vicinity of the metal objects. Sinogram correction by this so-called 'naive interpolation' ultimately leads to the generation of secondary streaks in the corrected images [MND+99, MB09]. These secondary streaks may be comparable in severity to the original artefacts. Although the original streaks are usually reduced to some degree with interpolation, they are rarely eliminated entirely [RLP+03]. Variations to the sinogram-completion procedure thus constitute the bulk of MAR-based literature and numerous approaches have been proposed to overcome the aforementioned limitations.

3.2.1.3 Sinogram Completion using Priors

Several approaches have been proposed with the primary objective of better preserving edge and contrast information and thereby reducing secondary artefacts in the corrected images. Many of these methods exploit the predictability of the characteristics (e.g. CT numbers) of the anatomical structures present in medical-CT scans to generate priors which are used to guide the sinogram-completion process. The majority of these methods employ some variation of intensity thresholding segmentation (e.g. k-means clustering [BS06] and multiple thresholding [MRL+10]) to generate priors and then focus on variations in the sinogram-completion phase to improve MAR results. Meyer *et al.* [MRL+10, MRS+11] and Muller and Buzug [MB09] use normalisation schemes based on the ratio of the raw sinograms to the sinograms of the prior images, to increase the homogeneity of the interpolation regions. Interpolation is claimed to be less problematic when applied to relatively homogeneous regions [MRL+10]. Meyer *et al.* [MRL+12] propose frequency-splitting techniques to reduce the characteristic, undesired blurring of edges near to metal objects seen in interpolated images by utilising the high-frequency edge information available in the original FBP reconstructions.

The core of the aforementioned techniques lies in intensity thresholding to generate the prior information. Intensity thresholding however, often leads to poor segmentations and hence inaccurate priors [KCWM12]. While the use of priors is

intended to better preserve edge and contrast information and thereby minimise secondary artefacts [BS06, MRL+10, KCWM12], inaccurate priors can in fact lead to a loss of edge information in the sinogram and hence cause greater degrees of secondary artefacts in the corrected images. Karimi *et al.* [KCWM12] demonstrate that the accuracy of the prior has a greater impact on artefact reduction than the chosen interpolation strategy. They propose focussing on generating an accurate prior (as opposed to improving the interpolation procedure) to improve artefact reduction. In particular, the prior is generated by segmenting regions of the original CT image, and distinguishing between metal artefact regions and anatomical regions. The metal artefact regions are assigned a constant soft-tissue value, while anatomical regions are left unchanged. The sinogram of the resulting prior is used to guide the sinogram-completion phase (performed using standard interpolation techniques). The algorithm successfully reduces metal artefacts and produces fewer secondary artefacts than related (intensity threshold-based) techniques [BS06, MRL+10], even in cases involving multiple metal objects.

While these prior-based techniques have shown impressive results, optimal parameter tuning relies heavily on the predictability of the structures present in the scan. Since they are all intended for use in the medical field, it is appropriate to assume prior knowledge of the likely anatomical structures present in the scans and reliable thresholds and parameters can thus be set. In settings where this prior knowledge or predictability regarding the nature of the scanned objects does not exist, however, the selection of suitable parameters will become significantly more challenging. Prior-based techniques are expected to be less effective in such domains.

3.2.1.4 Sinogram Completion with Multiple Metal Objects

Another major challenge in MAR arises in cases involving multiple metal objects. The presence of multiple (or large) metal objects means that the effective shape of the metal regions will be asymmetric across views resulting in unequal quantities of beam hardening and scatter across views. For example, at a particular tube position the beam may only pass through one of the metal objects and thus be hardened less than at another tube position where it passes through both objects [BK04]. This exacerbates the effects of beam hardening and scatter, usually leading to more prominent (dark) streaks in the regions connecting the metal objects [MND+99, ZBWW02, Man01]. The presence of multiple metal objects also complicates the process of multiclass segmentation, particularly in the regions of and near to the metal objects [KCWM12]. Poor segmentations invariably result in poor priors, characterised by a loss of edge information of the anatomical structures surrounding the metal objects [KCWM12]. As mentioned, Karimi *et al.* [KCWM12] demonstrate that the use of inaccurate priors has a significant detrimental effect

on artefact reduction. Finally, multiple metal objects lead to sinograms with multiple metal traces. Direct interpolation of such data is more likely to result in a loss of edge information in the metal trace of the sinogram which ultimately results in secondary artefacts in the corrected image (as discussed above) [YZB+07]. Several studies have proposed interesting approaches to deal with the challenge of multiple metal objects in particular.

Takahashi *et al.* [TMK+06] propose a very simple approach where the sinogram of the metal-only image is subtracted from that of the original image and the reprojection of the resulting difference image used as the corrected image. Qualitative improvements to the visual quality of the images are, however, limited. Abdoli *et al.* [AAA+10] compute a corrected sinogram as a weighted combination of the spline-interpolated sinogram and the original, uncorrected sinogram, where the optimal weighting scheme is determined using a Genetic Algorithm (GA) [BNKF97]. Yu *et al.* [YZB+07] employ a feedback interpolation strategy whereby interpolation is carried out repeatedly until all interpolated sinogram values are less than the original corrupted values. Jeong and Ra [JR09] employ Total Variation (TV) pre-filtering of the original image to reduce the initial severity of streaking artefacts. Interpolation is then performed only in regions where the sinogram traces of multiple metal objects intersect (elsewhere, the corrupted sinogram data is replaced by the reprojection of the TV-filtered image), thereby reducing the effective size of the interpolated regions. A scaled combination of the original and interpolated sinograms is reconstructed to yield the final image. Zhao *et al.* [ZRW+00, ZBWW02] extend the conventional linear-interpolation-based approach by performing an additional interpolation in the wavelet domains of the original and linearly interpolated sinograms. While effective, the algorithm requires several parameters to be set manually. Kratz *et al.* [KKM+08] treat the problem of CT MAR as a scattered data interpolation problem and perform interpolation based on the 2D Nonequispaced Fast Fourier Transform (NFFT) [Kun06]. The technique is, however, met with limited visual improvements.

3.2.1.5 Sinogram Completion using Image Inpainting

While interpolation is the most common technique for replacing corrupted or missing sinogram data, image inpainting techniques have also been successfully implemented in sinogram-completion schemes [LBY+10, DZX+08, ZPH+11]. Image inpainting is the process of replacing missing or corrupted data in images in a non-detectable way, using local geometric and/or textural information from uncorrupted regions in the same image [RC01]. Li *et al.* [LBY+10] combine Non-Local (NL) inpainting and linear interpolation to replace corrupted sinogram data, while Duan *et al.* [DZX+08] employ Total Variation (TV) inpainting within the conventional sinogram-completion framework. Zhang *et al.* [ZPH+11] present a

fractional-order TV-inpainting approach, where the conditional conductivity coefficient for TV is replaced by a novel fractional-order curvature parameter. The approach is shown to outperform simple linear interpolation as well as the TV-inpainting approach of Duan *et al.* [DZX+08]. The main drawback of using inpainting as opposed to interpolation is the significant increase in computational cost.

3.2.1.6 The Virtual Sinogram

The majority of sinogram-completion methods operate directly on the raw sinogram (projection) data. This approach may pose practical challenges as raw projection data is often vast and stored in a proprietary format. This makes the practical implementation of MAR algorithms difficult and dependent on the scanner manufacturer. Abdoli *et al.* [AAA+10, AAAZ10] introduce the concept of a virtual sinogram (obtained by the forward projection of the CT image) to overcome this challenge. They do assert, however, that CT images obtained by back-projecting virtual sinograms are not of diagnostic quality.

3.2.1.7 Final Reconstruction

The final, corrected image is generally constructed by filtered back-projection of the corrected sinogram, giving a corrected background image (i.e. free of metal objects). The metal-only image is then reinserted into this corrected background image, yielding the final image. A simple addition, however, may result in a loss of edge information for the metal objects. To combat this, Yu *et al.* [YZB+07] propose using an adaptive scale and filter scheme (originally developed by Chen *et al.* [CLSR02]) to compose the final image, whereby a scaled and lowpass filtered version of the metal-only image is added to the background image. Roeske *et al.* [RLP+03] reconstruct the final, corrected image in the usual way and then apply a pixel-by-pixel correction factor to the original image to reduce the severity of metal artefacts.

3.2.2 Pre and Post-Processing Operations

Several studies have implemented interesting pre-processing and post-processing steps which appear to have some beneficial effects. The predominant motivation for pre-filtering operations is to reduce minor streak artefacts and background noise [MB09, LBY+10]. A variety of common 2D image denoising filters have been used for these purposes such as: adaptive filtering [LBY+10, BS06], Total Variation (TV) filtering [JR09] and Non-Local Means (NLM) filtering [MMFB12].

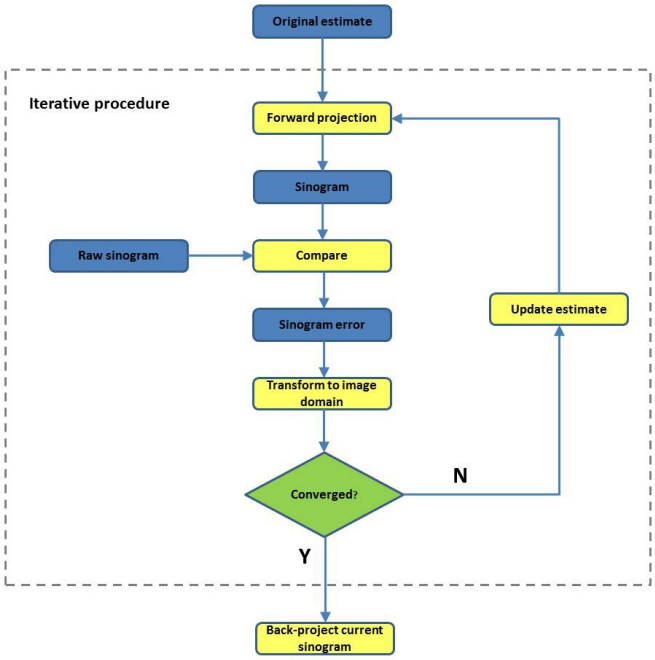


Figure 3.1: General framework for iterative-reconstruction techniques. Starting from the current reconstruction, a sinogram is calculated. A measure for the sinogram error is transformed to the image domain. The reconstruction is updated in a way that reduces the sinogram error.

3.2.3 Iterative Reconstruction Methods

Iterative-reconstruction techniques provide an interesting alternative to the FBP approach with several known advantages (especially in terms of MAR) [LC84, SV82].

The problem of iterative reconstruction is solved by optimising a chosen objective function. Algorithm variations generally occur in the choice of this objective function (e.g. minimum least squares error, maximum likelihood etc.) and the particular optimisation technique used (e.g. steepest ascent, conjugate gradients etc.). The chosen objective function and optimisation technique are used in an iterative optimisation framework composed of the following steps (Figure 3.1): 1) an initial reconstruction is estimated (typically a blank image); 2) the virtual sinogram of the estimate is computed; 3) the error between the virtual sinogram and the raw sinogram is computed; 4) the error is transformed to the image domain and used to update the estimate. Steps 2 to 4 are repeated until the error converges or some predefined termination criteria are met.

The most significant advantage of iterative approaches in general, is the superior performance in reconstructing images from incomplete projection data, as the assumption of uniformly sampled projection data is not required (as is the case for analytical approaches such as FBP).

The two most common iterative methods are algebraic reconstruction and maximum likelihood reconstruction. The Algebraic Reconstruction Technique (ART)

[GBH70, Hou72] is a special case of the Projection Onto Convex Sets (POCS) [BB96] iterative operator. POCS iteratively solves for f in the reconstruction problem:

$$p = A \cdot f \tag{3.1}$$

where p is the log-converted CT data, A is the projection matrix, describing all of the projections lines and f is the unknown reconstruction. The Simultaneous Algebraic Reconstruction Technique (SART) [AK84] and the Simultaneous Iterative-reconstruction technique (SIRT) [Gon72] are variations of ART.

Maximum Likelihood (ML) reconstruction is a statistically-based iterative algorithm that is typically used to find the Maximum Likelihood (ML) estimates of the parameters in a statistical model. The model parameters that yield a distribution giving the observed data the greatest probability are found by maximising the log-likelihood for the observed data [DLR77]. When considering CT reconstruction, the objective of ML is to estimate the reconstructed image that best fits the measured projection data under the assumption that this data obeys the laws of a Poisson distribution. Intuitively, this equates to finding the image which is *most likely* to have produced the measured projection values [Don07]. While transform-based reconstruction techniques (such as FBP) attempt to compensate for the stochastic nature of the projection data by filtering out high frequencies prior to reconstruction, statistical iterative-reconstruction techniques employ statistical models to approximate the projection data [Man01]. Statistical iterative-reconstruction techniques (such as ML approaches) offer the additional benefit of significantly decreased noise levels compared to FBP, provided that the appropriate statistical models are employed. Ultimately, statistical reconstruction leads to a significant reduction in streaking artefacts. These improvements are attributed to the fact that the iterative approach seeks to find the optimal fit to the measured data, while FBP assumes the data to be exact, consistent and complete [WSOV96].

In nuclear medicine applications such as Positron Emission Tomography (PET) and Single Photon Emission Tomography (SPECT), Maximum Likelihood Expectation Maximisation (MLEM) is the common method for image reconstruction [SV82]. The Poisson log-likelihood is optimised by using an Expectation Maximisation (EM) technique. For CT applications, the MLEM algorithm is usually applied to the log-converted data. However, the MLEM algorithms assumes the data are Poisson-distribution which is only correct for the measured data but not for the log-converted projection data.

A dedicated MLEM algorithm for CT was developed by Lange and Carson

[LC84]. Ollinger [Oll94] has, however, demonstrated that the EM algorithm converges extremely slowly for transmission reconstructions. Consequently, several ML [SB93] and Penalised Likelihood (PL) [FFCL97] transmission reconstruction algorithms have adopted the approach of directly maximising the objective function as opposed to relying on the classical EM in order to reduce computational costs. Erdogan and Fessler [EF99] present an alternate simultaneous update algorithm to the transmission EM algorithm of Lange and Carson [LC84]. The so-called Separable Paraboloidal Surrogates (SPS) algorithm is shown to converge considerably faster than the transmission EM algorithm. Other examples of direct-maximisation techniques include the convex algorithm [LF95] and Maximum Likelihood for Transmission (MLTR) [NMD+98].

Hudson and Larkin [HL94], introduced the concept of Ordered Subset Expectation Maximisation (OSEM) to reduce the computational demands of MLEM. In OSEM, the projection views are divided into several Ordered Subsets (OS) which are used to sequentially update the current reconstruction estimate. The same principle can be applied to other ML reconstruction techniques [EF99, BK01]. OS is capable of accelerating the convergence of the ML algorithms by a factor approximately equal to the number of ordered subsets and generally requires only small modifications to the algorithm. The main limitation of the approach is that convergence is only guaranteed if an exact solution exists: if this is not the case, OS results in limit cycles [HL94]. Convergence can, however, be enforced by reducing the step size at each iteration (relaxed update schemes [BdP96]) or by gradually reducing the number of subsets during reconstruction [Ber97]. Relaxed update schemes, however, require the manual specification of relaxation parameters, as there exists no mathematical technique for selecting optimal parameters. Such user input can have a significant detrimental effect on the convergence rate [LAL05].

A major advantage of ML reconstruction algorithms is that *a priori* information about the image to be reconstructed can incorporated into the algorithms. In Maximum *a-posteriori* (MAP) a combined likelihood is used. The first part of MAP is the ML-part, where the likelihood for the reconstructed image with respect to the measurement is given. The second part, the prior, describes the likelihood of the reconstructed image based on *a-priori* information. Alternatively, in Penalised Likelihood (PL), 'less likely' reconstructions are penalised. The *a-priori* information used in both MAP and PL is usually a smoothness constraint.

As has been discussed extensively by de Man [Man01], ML approaches allow for a direct incorporation of mathematical models for various physical limitations of the acquisition process into the system matrix (e.g. noise, beam hardening, partial volume effect etc.) as well as *a priori* information regarding the reconstructed image (e.g. smoothness constraints [MND+00]). The FBP approach

does not account for these limitations, leading to artefacts in the final reconstruction. While accurate modelling of the acquisition process is challenging and has a significant impact on the accuracy of the reconstruction, even techniques using simple acquisition models have been shown to produce better reconstructions (in terms of metal artefacts) than FBP reconstructions. The predominant limitation of iterative-reconstruction techniques, in the context of this study, is the high computational cost.

De Man *et al.* [MND+01] use an effective noise model, a Markov random field smoothness prior, a polychromaticity model (to combat the effects of beam hardening) and increased sampling in the reconstructed image in a transmission maximum likelihood reconstruction (MLTR) framework [NMD+98]. While this Iterative Maximum-Likelihood PolyChromatic Algorithm for CT (IMPACT) is not a dedicated MAR technique, preliminary experimentation does suggest an effective reduction in the effects of metal artefacts. The benefits of incorporating prior information and establishing an accurate acquisition model are clearly demonstrated. As with most iterative approaches, computational cost is an issue - comparing the computational complexity of IMPACT to the MLTR approach yields a ratio of 8:3 [SN12]. Elbakri and Fessler [EF02, EF03] and Menveille *et al.* [MGOS05] describe ML methods which, similarly to the IMPACT algorithm, incorporate a polychromatic acquisition model for multiple materials. Van Slambrouck and Nuyts [SN12] demonstrate that the computational cost of iterative reconstruction using complex reconstruction models (e.g. [MND+01, NMD+98]) can be reduced without a significant decline in performance (in terms of metal artefact reduction) by limiting the use of the complex models for the reconstruction of image regions near to the metal objects. Less complex models can then be used for reconstructing the remainder of the image. The images are automatically subdivided into metal and non-metal regions (patches) and reconstruction models of varying complexity (MLTRC - a simple polychromatic extension to the MLTR model [NMD+98]; IMPACT - a fully polychromatic model [MND+01] with or without increased resolution model) are then applied to these patches depending on the contents of the patch. The study compares the performance of several iterative reconstruction schemes to a number of sinogram-completion-based approaches (linear interpolation [KHE87, GP81], NMAR [MRL+10] and FSMAR [MRL+12]). Applying the MLTRC model in metal-free patches and the IMPACT model in metal patches, while resulting in a considerable reduction in computational cost, is shown to yield reconstructions of a similar quality to those obtained when using the IMPACT model for the entire image. While the sinogram-completion methods are shown to produce reconstructions with less obvious metal artefacts (compared to the iterative approaches) they are also shown to be characterised by a loss of edge and contrast information in the direct vicinity of the metal objects.

Since iterative-reconstruction techniques are inherently better suited to producing metal artefact-free images, the majority of recent work in this area has been concerned with minimising the computational costs of iterative approaches [HL94, LAL05, BF11, BdP96, MND+00, WFV00, YTB+11, CCFT06]. Despite the development of optimised approaches such as Ordered Subset Expectation Maximisation (OSEM) [HL94], the Row-Action Maximum Likelihood Algorithm (RAMLA) [BdP96], Model-Based Iterative Reconstruction (MBIR) approaches [YTB+11], Iterative Coordinate Descent (ICD) optimisation [BS96, TSBH07], Block-Iterative (BI) modifications [Byr97] and numerous hybrid methods [LAL05, BF11, MND+00], high computational cost remains the main factor preventing the universal implementation of such techniques in commercial CT machines. Processing times are still often quoted to be as much as three orders of magnitude higher than corresponding FBP processing times and for this reason iterative techniques have yet to be incorporated into routine clinical practice [Man01].

The work of Wang *et al.* [WSOV96] may perhaps be considered the benchmark for iterative-reconstruction techniques aimed specifically at metal artefact reduction in CT. While several earlier studies [GBH70, Opp77] address the issue of metal artefact reduction using iterative-reconstruction approaches, the majority of earlier work considered only reconstructions from complete projections. Wang *et al.* [WSOV96] present modifications to the Expectation Maximisation (EM) approach [SSO92] and the Simultaneous Iterative-Reconstruction Technique (SIRT) [Gon72] (an algebraic technique) to deal specifically with reconstructing CT images from incomplete projections. In particular, it is shown that two factors set the proposed approaches aside from related works. Firstly, the EM-type algorithm of [WSOV96] adopts simultaneous iterations: while the approach is similar in many ways to the Multiplicative Algebraic Reconstruction Technique (MART) [GBH70], the MART is derived from the Kaczmarz-method [Kac37] for solving a system of linear equations and is sequential in nature. In other words, the correction factors (to be applied to the image estimates at each iteration) are computed and applied based on individual projections profiles. The EM algorithm of Wang *et al.* [WSOV96] computes these correction factors by simultaneously taking into account all of the projection profiles and then updating the image estimates. While sequential approaches generally converge faster than simultaneous approaches, they generally yield poorer reconstructions characterised by noise-induced salt-and-pepper like stripes [AK84]. Furthermore, the convergence of the EM-type iterations has been established under moderate conditions (regardless of whether the projection data is complete or not), while the MART method cannot converge in the data-inconsistent case [WSOV96]. The second factor contributing to the superiority (in terms of MAR) of the methods proposed in [WSOV96] is that the authors adopt a spatially varying relaxation coefficient in each iteration. In previous works, this

coefficient was traditionally constant in each iteration and did not compensate for the non-uniform densities of the re-projecting and back-projecting rays (non-uniformities are especially prominent near metal surfaces). In addition to demonstrating that both iterative approaches (EM-based and ART-based) outperform FBP for incomplete projection data as well as for noisy, but complete projection data, the authors conclude that the EM-type algorithm converges faster than the ART-type algorithm in terms of both the I-divergence [Csi75] and the Euclidean distance between the measured and the reprojected data [WSOV96]. The differences in the computational complexity of the two proposed approaches are shown to be negligible [WSOV96].

The majority of statistically-based iterative reconstruction methods are formulated as unconstrained optimisation models that minimise or maximise some data-dependent cost function [ZWX11]. Zhang *et al.* [ZWX11] consider the metal-affected sinograms as systems with incomplete data and employ a constrained optimisation model to compute the optimal solution. Within the constrained optimisation framework, the data fidelity term (in the optimisation model) becomes an inequality and is used to determine a set of images that satisfy the measured data to within a predefined tolerance. The regularisation term (in the model) becomes the objective function and is used to select the optimal image from the feasible set. The optimisation problem is solved using a combination of the Projection-Onto-Convex-Sets (POCS) iterative operator [BB96] (a combination of the ART updating scheme and the image non-negativity constraint) and the steepest gradient descent of the objective function. The constrained optimisation algorithm is evaluated using a novel Penalised Smoothness (PS) function with an edge-preserving prior to generate an artefact and noise reduced solution. The method is shown to outperform linear interpolation [KHE87], ART [WSOV96] and EM [WSOV96] methods in terms of simultaneous artefact and noise reduction as well as edge and contrast preservation in a series of phantom experiments. While the EM and ART methods yield artefact-reduced images, the authors demonstrate that the proposed approach is able to eliminate artefacts entirely.

3.2.4 Hybrid and Miscellaneous Methods

Several studies have attempted to exploit the advantages of fundamentally differing approaches in hybrid algorithms to improve reconstruction performance.

Watzke and Kalender [WK04] propose merging the outputs of a conventional linear-interpolation-based approach with that of a Multi-dimensional Adaptive Filter (MAF) [KWK01] to combat the common shortcomings of direct interpolation-based approaches. The approach is motivated by the fact that interpolation and MAF compliment one another at varying distances from the metal objects.

Oehler and Buzug [OB06, OB07] present a modified MLEM approach that uses

a weighted MLEM algorithm to reconstruct interpolated sinograms. Different projection lines through the scanned object are weighted such that the influence of the residual inconsistencies of the interpolation procedure are minimised. Although an overall improvement in image quality is demonstrated, high computational expense is again highlighted as a concern.

Lemmens *et al.* [LFN09] show that the success of iterative reconstruction approaches rely heavily on algorithm initialisation and propose a hybrid approach whereby a Maximum *a Posteriori* (MAP) scheme is used to define a constrained image (free of artefacts) which is ultimately used to initialise an MLTR reconstruction. The proposed method involves three separate iterative reconstructions: an initial MLTR reconstruction, a MAP reconstruction and the final MLTR reconstruction and processing times are thus extremely high. To combat this, the authors propose replacing the initial and final MLTR reconstructions with FBP reconstructions and show that the deterioration in performance is minimal.

Based on the principle that CT reconstruction fidelity is proportional to the number of projections used [Hsi03], Bruyant *et al.* [BSM00] present a level line-based interpolation scheme for increasing the number of projections, without increasing the acquisition time, to reduce streaking artefacts. The algorithm is fully automated and eliminates the need for a high-pass filter in the reconstruction process.

Image-domain MAR [NLA+11, NLP+09], whereby all processing is performed in the image domain, has been proposed as an alternative to the virtual sinogram [AAAZ10, AAA+10] (Section 3.2.1.6) in scenarios where raw projection data is unavailable. Naranjo *et al.* [NLA+11, NLP+09] propose two approaches for dealing with situations where raw projection data is not available. The first approach [NLP+09] involves morphological filtering in the polar domain (to exploit the inherent symmetry in FBP reconstructed images). A second, more recent approach [NLA+11] uses a comparison between the current, artefact-containing slice and an adjacent, artefact-free slice to locate corrupted regions in the image, which are ultimately replaced via 2D linear inpainting. Image registration and morphological dilation are used to accurately locate the corrupted regions whilst avoiding anatomical structures. The method requires considerable user input and thus suffers from high processing times.

Finally, it is worth noting that the implementation of all iterative-reconstruction techniques as well as hybrid approaches, combining sinogram-completion and iterative reconstruction, rely on the availability of the raw (original) projection data. When this data is not readily available the iterative approach is not feasible.

3.2.5 MAR Performance Evaluation

Performance evaluation is an important yet challenging task. All of the MAR studies discussed in the preceding sections share a primary objective of improving the diagnostic quality of medical-CT images; in other words making the CT image easier to read for a radiologist. Therefore, in reality, the problem is a qualitative one: a MAR technique is deemed effective if a radiologist (or some other expert human observer) concludes that the image is easier to interpret after applying the MAR algorithm. It is thus not surprising that the majority of studies rely heavily on a subjective analysis of performance. This makes it challenging to reliably compare different techniques. Nonetheless, some studies have presented some form of quantitative analysis.

A common trend is to perform both clinical studies using real-world CT scans, as well as simulated studies using phantoms (objects which are designed to mimic the properties of human tissue and organs). In the medical domain, the use of physical phantoms [MB08, LBY+10, YZB+07, MRL+10, ZBWW02] as well as software-generated phantoms [YZB+07, MND+01, Man01, MND+00, LFN09] have become accepted comparative protocols. The use of phantoms (numerical and physical) allows for the establishment of gold standard images (usually obtained by scanning or simulating the phantom without metal inserts) and hence the implementation of any standard image reconstruction performance measure [SB10].

Meyer *et al.* [MRL+10] compare the projection (sinogram) profiles and image profiles of software generated phantom images (containing artefacts) after applying MAR with the corresponding profiles of a reference image (artefact-free phantom image) to quantify MAR performance. It is claimed that effective MAR will yield profiles that closely resemble those of the reference image. It is important to note, however, that this analysis technique would not be possible using real-world data. The true challenge thus lies in quantifying the performance of an algorithm on real-world data.

A simple reduction in the standard deviation of the reconstructed image is often cited as evidence of successful reduction in streaking [TMK+06]. This approach exploits the notion that streaking results in large and frequent fluctuations in the CT values of the reconstructed image with respect to the underlying CT values. Hence, reducing these fluctuations will result in a more homogeneous image and a lower standard deviation. This approach is somewhat rudimentary however, as it does not account for the unwanted blurring of image details in the artefact-free regions of the images, which would also lead to more homogeneous images and hence reduced standard deviations. Although more sophisticated standard deviation-type measures have been employed [YZB+07], these usually rely on the prior knowledge of the ideal attenuation values of the regions in the direct vicinity

of the metal objects, which is not always available outside of the medical domain [vKD05].

Abdoli *et al.* [AAA+10] perform quantitative performance analysis using Bland-Altman plots [AB83]. Three regions are specified in each of the CT slices: overestimated regions (pixel densities higher than expected due to artefacts), underestimated regions (pixel densities lower than expected due to artefacts) and unaffected regions (regions not affected by streaking). The Bland-Altman plot (or, equivalently, the Tukey mean-difference plot) is used to compare the pixel densities in these regions before and after applying MAR and ultimately to quantify performance. While this approach does rely on the knowledge of the intensity distributions of the images prior to MAR, this extends only to the ability to distinguish between regions in the images which are and are not affected by streaking (a distinction which can be performed by manual inspection of any image). It does not require prior knowledge of the actual characteristics (i.e. ideal CT numbers) of the contents of the scan, as is the case in the more sophisticated standard deviation methods mentioned previously.

Ens *et al.* [EKB10] present a reference-free performance measure termed the Band-Pass filtered Gradient (BPG) measure to quantitatively evaluate the performance of MAR techniques when ground-truth data is not available. The BPG measure is computed as the sum of the pixel values in the gradient image of a given image, where the sum is limited to gradient values falling within a predefined band (hence band-pass filtered). This band is defined by manually selected upper and lower band-limits. The BPG of a given image is expected to be higher in an image corrupted with streaking artefacts.

Kratz *et al.* [KEMB11] also present a reference-free ground-truth metric for quantitatively evaluating the performance of MAR techniques. The authors make the assumption that in the original projection data, the presence of metal objects have no influence on the projection readings outside of the actual metal traces - this is not the case for the virtual sinogram [AAAZ10]. The original projection data, outside of the metal traces, is thus considered to be an inherent ground-truth which is used as a reference for evaluating the degree of image artefacts. Using a standard distance metric, the regions in the original and corrected (virtual) sinograms outside of the metal traces are compared. For a given CT image, the value of this metric should decrease with a reduction in metal streak artefacts. While the quality of the performance measure is shown to be comparable to other reference-free measures, such as the BPG measure [EKB10], the approach is deemed superior in that it is fully automated.

To date, there appear to be no system-level quantitative performance measures (whereby the performance of a given MAR technique is quantified according to its impact on subsequent operations, such as object classification). Furthermore,

existing approaches only consider performance in 2D cross-sectional images (i.e. in $\mathbb{R}^2$) - the performance of MAR techniques in true 3D space has not been considered previously.

3.2.6 MAR Summary

Previous work, where the relative performance of several state-of-the-art MAR techniques are compared, is limited [RDF+08, GMB+11]. There are presently no comprehensive reviews or quantitative comparative studies of the state-of-the-art MAR techniques. Furthermore, all of the MAR studies discussed in this review have been intended for use in the medical-imaging domain only. Many of these rely on the use of priors to guide the sinogram-completion process. While the need for metal artefact reduction in CT imagery extends beyond the medical domain, the performance of the state-of-the-art medical MAR techniques in settings where isolated metal objects occur in non-tissue surroundings (making the generation of accurate priors more challenging), is unclear. An investigation into the performance of these existing, medically-based MAR techniques when applied to non-medical images has not been conducted previously. Existing studies are thus limited in the techniques that are compared, the CT domains or applications which are considered and the performance-evaluation techniques that are employed.

3.3 Classification

The classification of images (or parts thereof) is a core problem in computer vision. Broadly speaking, image classification refers to the labelling of images according to the object categories (i.e. classes) they contain. Virtually all approaches are based on the assumption that the features characterising a given image (e.g. geometric shapes, textures etc.) may be 'matched' in some way to those of the predefined classes [JKS95].

For the most part, supervised image classification (where classification models are inferred from labelled training data) relies on at least the following components/stages [AT13]: 1) feature extraction and description; 2) training and 3) classification. While this section presents a brief review of the recent advances in image classification, the primary focus is on the classification of low-quality, complex volumetric imagery (encountered in non-medical domains). The reader is referred to the literature for a more comprehensive review of image classification [AT13, Jor02, Sze10].

3.3.1 Feature Extraction and Description

Although selecting suitable properties (or features) by which to represent object classes and images is a crucial component of a classification system, it is not the central focus of this review and only a brief overview of the fundamental concepts and popular techniques is provided.

Image representations based on local feature descriptors are widely applied in image-classification and object-recognition frameworks due to their robustness to partial occlusion and variations in object layout and viewpoint. Distinctive features of objects are detected at interest point locations which generally correspond to local maxima of a given saliency measure calculated at each location in an image. The intensity patterns around these interest points are then encoded using a descriptor vector.

Interest-point detection has been performed in a number of ways, depending on the desired characteristics of the detected points/regions. Harris detectors [MS04], for example, respond to corners and highly textured points and are more effective when exact interest points are desired. In contrast, interest-point detectors such the Difference-of-Gaussians (DoG) [Low99] and Laplacian-of-Gaussians (LoG) [MS04]) detectors, respond mainly to image blobs and are preferred when invariant regions are desired. Furthermore, different detectors offer invariance to different scenarios (e.g. scale and/or affine invariance). Lowe [Low99], for example, determined scale invariant interest points by computing local extrema in scale-space pyramids constructed using DoG filters (which is shown to offer a good approximation of the LoG operator at significantly lower computational costs). Tuytelaars and Van Gool [TVG99], on the other hand, compute affine invariant interest points using local edge information computed at 'anchor' points located using the Harris detector. Mikolajczyk and Schmid [MS04] present a detector that results in interests points that are invariant to changes in both scale as well as affine transformations and demonstrate state-of-the-art performance. Particularly, a multi-scale representation of the Harris detector is used to determine candidate points. Scale invariant points are then determined by maximising a local Laplacian measure across all scales. Finally, affine invariance is achieved by estimating the local affine shape of each point by iteratively modifying its location, scale and neighbourhood. The aforementioned techniques are some of the most popular interest-point detection approaches, for a more comprehensive review of existing methods the reader is referred to the literature [SMB00, MLS05, TM08].

Feature description is the process by which the characteristics of the detected interested points are encoded in a vector suitable for classification. To this end, the Scale Invariant Feature Transform (SIFT) [Low04], which encodes the interest point information via localised sets of 3D gradient orientation histograms, has been one of the most successful and widely adopted techniques in

the computer-vision literature. This has been substantiated in a recent comparative evaluation of image features in the context of object class recognition [MLS05]. In addition to evaluating five state-of-the-art interest-point detectors (Harris-Laplace [MS04]; DoG [Low04]; Hessian-Laplace [MTS+05]; salient regions [KB01] and Maximally Stable Extremal Regions (MSER) [MCUP04]) the following five state-of-the-art descriptors are considered: the SIFT descriptor [Low04]; the Gradient Location-Orientation Histogram (GLOH) [MS05b] - an extension of the SIFT descriptor aimed at improving robustness and distinctiveness [MTS+05]; the PCA-SIFT descriptor [KS04] - a (reduced dimensionality) vector of image gradients computed in the x and y directions within a region-of-interest; moment invariants [VGMU96] and a basic cross-correlation descriptor [MTS+05] computed as a set of normalised points sampled at 9×9 pixel locations within a smoothed version of the detected region. Within this extended experimental comparison, it is shown that the extended SIFT descriptor (GLOH [MS05b]) performs best. The SIFT formulation has in fact been met with success in a broad range of computer-vision applications including: object recognition [SEZ05, LMT+07, BD09, FBM10]; segmentation [SFTA08]; registration [BN08, YZY08] and panoramic image stitching [Sze06, BL07]. More recently, Bay *et al.* [BETG08] have presented the Speeded Up Robust Features (SURF) descriptor, which is loosely based on SIFT, but claimed to offer superior performance at a significant reduction in computational cost.

The aforementioned survey [MLS05] considered classification of 2D imagery. In the context of cluttered volumetric baggage imagery, Flitton *et al.* [FBM13] compare the performance of four 3D interest-point descriptors of varying complexities (sampled at SIFT interest points): the Density Histogram (DH) descriptor [FBM13]; the Density Gradient Histogram descriptor (DGH) [FBM13]; the 3D SIFT formulation of [FBM10] and a 3D extension to the Rotation Invariant Feature Transform (RIFT) [LSP03, LSP05]. The study considers the detection of known rigid objects within low-resolution, noisy and complex volumetric CT imagery. Surprisingly, it is shown that the simpler density statistics-based descriptors (DH and DGH descriptors) outperform the more complex 3D descriptors (SIFT and RIFT). The comparatively poor performance of the SIFT and RIFT descriptors are attributed to the low, anisotropic voxel resolution and high level of noise and artefacts characteristic to this type of imagery.

3.3.2 Classifiers

Two predominant classification paradigms exist: generative models and discriminative models. For a given input sample x and a class label y, generative classification models (e.g. Gaussian Mixture Models (GMM) [MP04]; Hidden Markov Models (HMM) [Edd96]; naive Bayes [Ris01]) learn the joint probability $p(x, y)$ and perform a classification by using Bayes' rule [JMF11] to compute the condi-

tional probability $p(y|x)$ and then selecting the most likely label y. Discriminative classification models (e.g. Support Vector Machines (SVM) [BHW10]; boosting [FS95]; random forests [Bre01]) model the conditional posterior probability distribution $p(y|x)$ directly (i.e. a direct mapping from the inputs x to the class labels y is inferred) [Jor02]. In contrast to generative models, discriminative models are inherently supervised. Despite the widely-held belief that discriminative models are better suited to the classification problem, it has recently been shown that while the discriminative model typically has a lower asymptotic error, the generative model often approaches its asymptotic error considerably faster [Jor02]. It is further shown that there exist two scenarios (dependant on the size of the training set) in which each model outperforms the other. Nonetheless (likely due to the aforementioned misconception) much of the recent literature has employed discriminative models for image classification [Cri11, MC11, CRLR12, JWP+12, CMS12].

The Support Vector Machine (SVM) [Vap00] is one of the most widely used classification algorithms. In binary classification tasks (i.e. two classes) SVMs seek the optimal linear separation of the classes by maximising the margin of separation between classes (Figure 3.2). Using this criterion, optimisation results in a separator that can be recovered at any time using only a few data points - namely those lying nearest to the boundary of separation (and hence determining the margin). These data points are aptly named the *support vectors* and can be used to identify the class of a previously unseen observation as lying on one side of the identified hyperplane separator in N-dimensional space. In cases of linearly inseparable data, separation is still attainable via kernel projection. Maximal-margin separation of classes is obtained by projecting the data into this higher-dimensional space using a suitable non-linear kernel function (e.g. Radial Basis Function (RBF) kernel [SSB+97]) (Figure 3.2 (a)). A soft margin may be used, allowing for some degree of misclassification (controlled by an input cost parameter C). The parameter C essentially weights the misclassified examples during training such that the total (training) misclassification error is minimised. The SVM classification algorithm offers guaranteed maximum-margin separation of classes (Figure 3.2 (b)) and good generalisation using only a fraction of the available training data (i.e. the samples near the class boundaries). It has thus enjoyed widespread popularity and success in image-classification and pattern-recognition problems [PV98, BL03, FM04, MC11].

Supervised learning algorithms, such as the aforementioned support vector machine, use labelled training data and search through a given function space for a function that will make good predictions of the labels of new (unseen) data. Suitable predictors, however, are not always easily found. In such scenarios, the advantages of ensemble-based learning, whereby classifiers are constructed by combining or fusing together multiple weak classifiers, are well documented

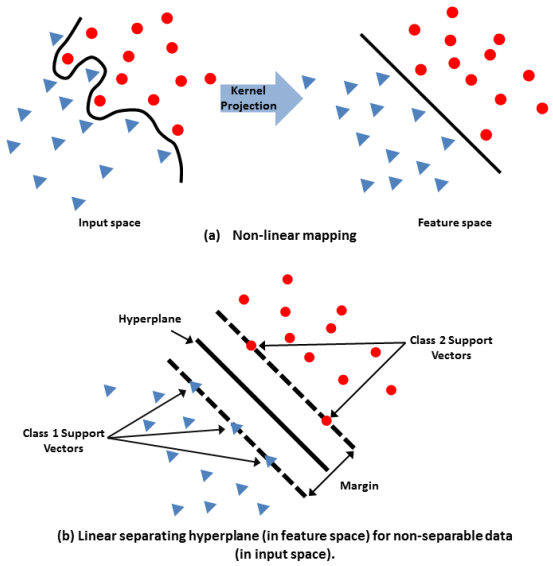


Figure 3.2: Illustration of binary classification using an SVM. The feature space is related to the input space via a non-linear mapping, allowing for linearly inseparable input data to be separated using a linear hyperplane in feature space.

[Fre01, Tu05, MGE11, Cri11]. The random-forest classifier [Bre01] is one such approach which has enjoyed a notable surge in popularity in recent years. Random forests are ensembles of randomly trained decision trees [Bre01]. Decision trees are models composed of sets of *nodes* and *edges* arranged in a hierarchical tree-like manner and employ branching strategies to predict an outcome given some input sample [Qui86] (Figure 3.3).

Decision trees are typically constructed in a supervised manner using a labelled training set and a greedy, top-down approach [Qui86]. Every training instance is completely described by a scalar-valued feature vector and an associated class label. A decision tree is composed of zero or more internal nodes and one or more leaf nodes. In the case of a binary decision tree (as is used in this work), all internal nodes (indexed by j) have exactly two child nodes. Internal nodes are characterised by a binary test (or node split function) defined over the incoming data (feature set) [Cri11]:

$$f(\mathbf{v}, \theta_j) : \mathcal{F} \times \mathcal{T} \rightarrow \{0, 1\} \tag{3.2}$$

where $\mathbf{v}$ is the incoming data point and $\theta_j \in \mathcal{T}$ is the parameter space of the split function. An optimal split function (type and parameters) is that which optimally separates the incoming data into two disjoint subsets - where the optimal separation is determined according to a chosen objective function (e.g. the information gain [Weh96]). Given an incoming training sample S_j at an internal node j, the function that optimally splits S_j into two disjoint subsets S_j^L (left child) and

S_j^R (right child) where $S_j^L \cup S_j^R = S_j$ and $S_j^L \cap S_j^R = \emptyset$ is that which maximises the objective function $I_j = I(S_j, S_j^L, S_j^R, \theta_j)$ [Cri11]:

$$\theta_j^* = \arg\max_{\theta_j \in \mathcal{T}} I_j \tag{3.3}$$

In the context of binary classification, the objective function is typically chosen such that the optimal split is that which produces the purest disjoint subsets of the input sample (a node j is said to be pure if all the training samples at j belong to a single class) [SL91].

In addition to determining the optimal node split functions, the training process also seeks to optimise the tree structure (i.e. depth, number of leaf nodes etc.). Greedy top-down tree construction [Qui86] is the most widely adopted method for building decision trees. Beginning with the entire dataset and an empty tree, top-down construction involves recursively partitioning the data into meaningful subsets until some termination criteria have been met (e.g. maximum tree depth or completely partitioned data). More particularly, beginning at the root node, $j = 0$, the optimal split function is found as described above. Two child nodes (left and right) are then constructed, receiving as input the two disjoint subsets of the initial training set. For each of the child nodes, the optimal split functions are again determined over the new input samples. This process is repeated for all subsequent child nodes until the termination criteria have been met. A wide variety of stopping criteria have been proposed in the literature (e.g. maximum tree depth; minimum number of samples at a leaf node etc.). The outputs of the tree training process are thus: 1) a tree structure; 2) the optimal split functions for each node and 3) the information about the training samples at each leaf node (e.g. class distribution) [Qui86].

Tree testing: A test sample is classified by passing it down the tree: beginning at the root node, the split function associated with each internal node is applied to the sample and based on the result of the binary test, the sample is sent to the left or right child node. The process is repeated until a leaf node is reached, where an output is assigned to the test sample using the information stored at the leaf node (e.g. class label or posterior probability) (Figure 3.3 (b)).

Leaf predictor model: The data available at a given leaf node may be used in several ways to assign an output to a test point. In the case of classification trees, a widely adopted approach is to estimate the conditional probability $p(c|\mathbf{v})$ that a given test data point $\mathbf{v}$ belongs to the class c, where c is a discrete class label. The distribution is conditional on the specific leaf node reached by the data point [Cri11].

Randomness: Randomness may be incorporated into the tree training pro-

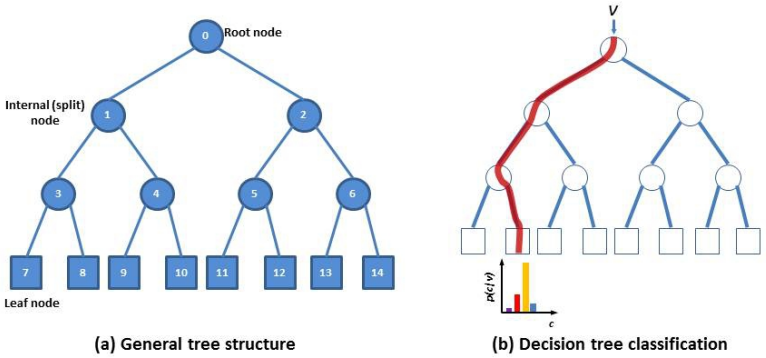


Figure 3.3: Decision tree classification diagram. (a) Generalised tree composed of a set of internal nodes (circles) and leaf nodes (squares). (b) Decision tree classification procedure: internal nodes store binary split functions which are applied to the incoming data sample $\mathbf{v}$. Leaf nodes store specific class information (e.g. posterior probability). Adapted from [Cri11]

.

cess in two ways [BZM07]: 1) random node optimisation and 2) random training set sampling. Random node optimisation uses the entire available training set at each node. Randomness is obtained by considering only a random subset of the available parameter values $\mathcal{T}_j \subset \mathcal{T}$. The degree of randomness is determined by the ratio $|\mathcal{T}_j|/|\mathcal{T}|$. A ratio of 1 indicates no randomness, while a value of $1/\mathcal{T}$ indicates maximum randomness. Random training set sampling (or *bagging*) injects randomness into the tree by considering only a randomly drawn subset of the training sample at each node. Random node optimisation and bagging may be used simultaneously [Cri11].

Random forests: A known limitation of the decision tree model is its tendency to overfit the training data [Bre01]. It has been shown that ensembles of randomly trained decision trees (or random forests) yield superior generalisation and stability relative to individual trees [Bre01]. Furthermore, in an extensive experimental comparison [CKY08], random forests have been shown to outperform a number of popular binary classification algorithms (SVMs [Vap00]; Artificial Neural Networks (ANN) [DOM02]; logistic regression [DOM02]; naive Bayes [Ris01]; k-NN [SB10]; bagged decision trees [Die00] and perceptrons [FS99]), particularly in high-dimensional problems. Several related studies, however, have contradicted these observations, particularly in the bioinformatics domain. Statnikov *et al.* [SWA08] for example, identify methodological biases of prior work comparing random forests and support vector machines. In an extensive comparative study using a broad range of microarray-based diagnostic and prognostic datasets it is shown that SVMs consistently and significantly outperform random forests in the classification of cancer. Similarly, Ogutu *et al.* [OPSS11] demonstrate superior genomic selection results for boosting relative to SVMs and random forests.

The individual trees in a forest are trained independently of one another (using the aforementioned framework), such that a random forest is comprised of an

ensemble of randomly different trees. In this way, where one tree fails or performs poorly on a given task, it is likely that another tree in the forest will perform well. This assumption is dependent on the degree of correlation (similarity) between the individual trees in a given forest. Forests composed of highly decorrelated but individually accurate trees generally lead to improved generalisation and stability [Bre01]. The degree of decorrelation between the individual trees of a forest is determined by the degree of randomness of the individual trees and is also controlled via the ratio $|\mathcal{T}_j|/|\mathcal{T}|$. Given a test data point, the output of a random forest composed of T trees may, for example, be computed as the average of the individual tree predictions or using a majority vote [Bre01].

A random forest is thus characterised by the following parameters [Cri11]: the number of trees in the forest T; the maximum individual tree depth D_T; the type and degree of randomness ($|\mathcal{T}_j|/|\mathcal{T}|$); the node split function(s); the node objective function; the leaf and forest prediction models and the choice of features. The impact of each of the model parameters on system performance is fairly well documented in the literature [Cri11]. In general, increasing the size of the forest T improves the generalisation capabilities of the forest, but also increases computational complexity. While large forests are known to alleviate the problem of overfitting characteristic of individual trees (even for forests composed of individual trees that heavily overfit), growing the trees of a forest to too great a depth D_T has been shown to have a negative impact on generalisation [Cri11]. Fully-grown trees (where each leaf node in the tree contains only a single training point) are therefore ill-advised. Overfitting may be avoided by terminating tree growth when one or more stopping criteria have been met (e.g. maximum tree depth; minimum size of training sample at nodes etc.). Alternatively, overfitting may be alleviated via the *pruning* of fully-grown or very large trees [HTF01]. Pruning involves reducing the depth of a decision tree by recursively removing nodes of the tree (usually bottom-up) until some termination criteria have been met (e.g. a threshold on the information gain). While various methods for tree-pruning exist (e.g. reduced-error pruning [EK01]; minimal cost complexity pruning [PS01]; rule post pruning [NHS01]), they can add significant complexity to the training process.

Extremely randomised forests: Extremely Randomised Trees (ERT) determine the node split functions by randomising both the attribute choices as well as the quantisation thresholds [GEW06, MTJ07]. Ensembles of such trees are referred to as Extremely Randomised Forests (ERF). ERFs typically do not incorporate bagging. With reference to the aforementioned ratio, ERFs use $|\mathcal{T}_j|/|\mathcal{T}| = 1/|\mathcal{T}|$, $\forall j$ - essentially, the degree of randomisation is maximised and node training is eliminated entirely. The degree of randomisation may, however, be controlled [MTJ07]. Although true ERFs may lead to lower overall prediction confidence,

they are highly efficient to train [Cri11].

Random forests have demonstrated state-of-the-art performance (both in terms of classification accuracy as well as runtime) in a broad range of image-classification-related tasks [SLT+03, DUDA06, CEJB+07, GMC+10, BZM07, Cri11]. In contrast to techniques such SVMs and boosting, random forests also extend naturally to multiclass problems [Cri11]. For a more comprehensive review of the role of random forests in computer vision, the reader is referred to [Cri11].

Boosting algorithms [FS95], whereby strong (i.e. good) classifiers are built as collections or ensembles (i.e. linear combinations) of many weak classifiers, have also enjoyed widespread success in the context of image classification. The bases of the majority of boosting algorithms rely on an iterative procedure whereby a weak classifier is learnt at each iteration using a weighted distribution of the training data. Having added the weak classifier to the final classifier, the training distribution is recomputed such the misclassified samples from the previous iteration are assigned heavier weights (i.e. they are 'boosted'), while the weightings of the correctly classified samples are reduced or left unchanged. The next weak classifier is then built using this re-weighted distribution. In this manner, future weak classifiers focus more on the previously misclassified samples. Boosting algorithms (e.g. AdaBoost [FS95]; BrownBoost [Fre01] - which is more robust to noise; linear program boosting (LPBoost) [DBST02]) have demonstrated success in a variety of image-classification [DB03, HES+07], detection [VJ01, VJ04, ZPV05] and recognition [OPFA06] tasks.

Bag-of-Words model: Sivic and Zisserman [SZ03] proposed the original Bag-of-Words (BoW) (or bag-of-visual-words) model for images, whereby local features obtained from images are grouped into a finite number of clusters. The cluster centroids form a codebook which is used to encode the features of images in a vector quantised representation. The cluster centroids are intuitively referred to as *visual words*. Any given image may then be represented by its (orderless) histogram over these visual words. Traditionally, image classification using the BoW model is composed of the following steps [NJT06a] (Figure 3.4): 1) feature detection and description; 2) visual codebook generation and vector quantisation and 3) classification (using any standard classifier).

Although feature detection and description are most commonly performed using interest-point detectors and local descriptors (refer to Section 3.3.1), an extensive investigation into the impact of sampling strategies on the classification performance of BoW models [NJT06b] has demonstrated that dense sampling of the feature space significantly outperforms interest-point detectors, provided the number of sampled points is large enough. The performance gains are attributed to the increase in the amount of information captured by a dense sampling strategy. While performance improves with sampling density, the associated increase

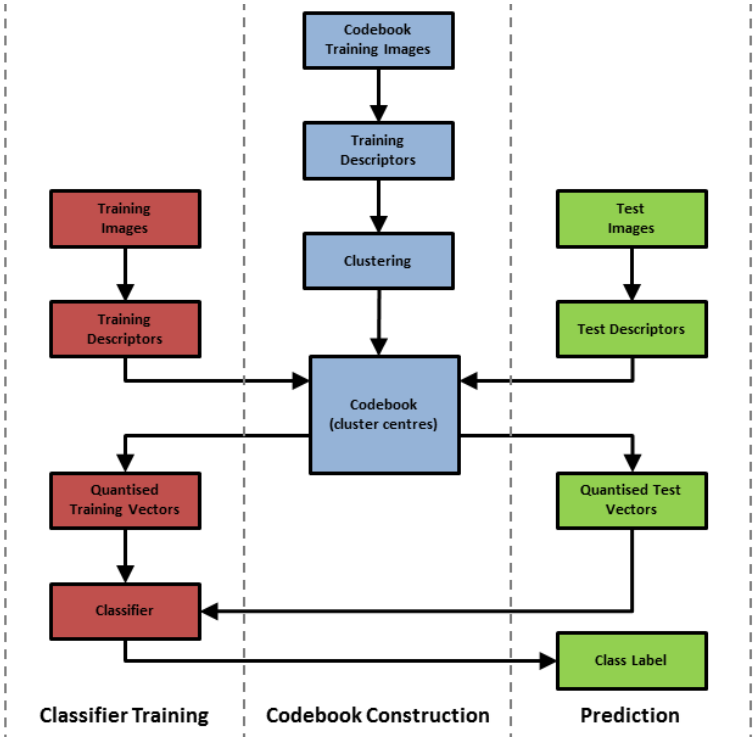


Figure 3.4: The Bag-of-Words (BoW) classification model.

in computational demand needs to be considered.

Feature encoding: Following feature description, an image is represented as an orderless set of descriptors. To allow for the use of standard classification techniques (e.g. SVM), the unordered descriptors are transformed into fixed-sized vectors via vector quantisation (middle column in Figure 3.4). This is achieved by learning a visual *vocabulary* (or codebook) by partitioning a set of training descriptors into clusters. Each cluster centre represents a visual word (or codeword) in the codebook. Provided the set of training descriptors is sufficiently representative of the data, the codebook will be universal. Each descriptor extracted from a new image is then mapped to the *nearest* entry in the codebook (a process known as vector quantisation) and a histogram of visual words is constructed for the image. This histogram represents the BoW representation of the image and is fed to the chosen classifier to determine the image class [MTJ07].

Of the various techniques that have been proposed for creating visual vocabularies the most popular approach is k-means clustering [SZ03]. Given a set of N training descriptors: $x_1, \ldots, x_N \in \mathbb{R}^D$, k-means clustering typically finds k vectors $\mu_1, \ldots, \mu_k \in \mathbb{R}^D$ and a data-to-mean assignment $q_1, \ldots, q_N \in \{1, \ldots, k\}$ that minimises the cumulative error $\sum_{i=1}^{N} ||x_i - \mu_{q_i}||^2$ [CLVZ11]. Despite its general accuracy, the assignment of local descriptors to visual words (usually via some nearest neighbour-based search) performed in both the construction of the codebook as well as in subsequent testing, is computationally expensive. This detracts from the suitability of a dense sampling strategy, where a massive number of descriptors (each requiring quantisation) are extracted from every image.

Several near real-time feature-encoding techniques have been proposed to address the computational limitations of traditional k-means clustering. Moosmann *et al.* [MTJ07] present a random-forest-based clustering algorithm that is considerably faster to train and test, more robust to background clutter and more accurate than traditional clustering methods and has arguably been the most prominent high-speed feature-encoding methodology of late. The proposed algorithm exploits the fact that component-wise decision trees offer logarithmic-time coding ($T(n) = O\log(n)$). This idea has been explored previously by Nister and Stewenius [NS06] who constructed codebooks using a tree-based approach based on hierarchical k-means quantisation. While the technique demonstrated good compromise between improved efficiency and loss of accuracy in image retrieval tasks, Moosmann *et al.* [MTJ07] propose that no single data structure can sufficiently capture the diversity of high-dimensional data, suggesting that k-means clustering is not the optimal approach for descriptor encoding. Instead, an ensemble of decision trees (i.e. a random forest) is proposed - due to its simplicity, speed and accuracy. In particular, codebooks are constructed using Extremely Randomised Forests (ERF) [GEW06], whereby both attribute choices as well as the quantisation thresholds are randomised. While traditional random-forest-based algorithms [Cri11] generate outputs by averaging over the constituent tree outputs, the clustering forests build codebooks by assigning separate codewords to every leaf node in the forest (i.e. a forest containing N leaf nodes, yields a codebook of size N).

Given a new image, every descriptor vector extracted from that image is fed through every tree in the forest and the output of each tree recorded. The result for a single descriptor is thus a set of labels (codewords) - one from each tree. The BoW representation for the entire image is obtained by accumulating the codeword counts after applying the forest to all the descriptors in the image. The resulting histogram of codewords is then used in subsequent classification in the same way as any standard BoW model. Using an SVM classifier, the proposed approach has been shown to outperform both k-means clustering and kd-trees in terms of processing time (training and testing), memory usage, classification accuracy and robustness to background clutter in classical 2D image-classification tasks (using the PASCAL VOC dataset [EVGW+10]) [MTJ07]. The efficacy of incorporating ERC forests into BoW image-classification frameworks has been substantiated further by the near real-time classification framework presented by Uijlings *et al.* [USS09] (tested on the 2007 PASCAL VOC dataset).

Furuya and Ohbuchi [FO09] present a 3D model-retrieval approach using ERC forests to encode densely sampled SIFT feature points. The study demonstrates the benefits of employing a dense-feature sampling strategy as well as the computational advantages of ERC forests for feature encoding. Although the framework is utilised for 3D model retrieval, feature extraction occurs in 2D depth images

and thus does not represent a true 3D application of the clustering forests model for feature encoding.

A similar model to the ERC forest is the class-specific Hough forest proposed for the detection of object classes in 2D imagery [GL13]. Previous works have accomplished this task via the generalised Hough transform-based Implicit Shape Model (ISM) [LLS08]. The ISM for a given object class is essentially a codebook of interest-point descriptors common to that class. Each entry in the codebook casts a vote for the possible positions of the object centroid according to the spatial distribution observed in the training data [LLS08]. The descriptors extracted from test images are then matched against the ISM codebook and votes regarding the position of the object in scale-space are generated. The peaks observed in the resulting Hough image (created by summing the probabilistic votes) are regarded as detection hypotheses [GL13]. Similarly to the traditional BoW model, codebook-based Hough transforms come at a significant computational cost. Class-specific Hough forests utilise random forests to directly map image patch appearances to the probabilistic object position votes. Similarly to the ERC forest, the set of leaf nodes of each tree in the Hough forest represents a discriminative codebook. Hough-forests have demonstrated state-of-the-art performance (as well as reductions in processing times) in a variety 2D object classification and segmentation tasks [GL13, PT13].

To date, it does not appear that clustering forests have been applied to 3D classification tasks. Nonetheless, random-forest-based encoding techniques have been successfully implemented in related tasks, particularly in the medical domain. Zikic *et al.* [ZGC13], for example, have presented a highly efficient Multi-Atlas Label Propagation (MALP) scheme for automatically labelling healthy tissue in 3D Magnetic Resonance (MR) imagery of the human brain, by using a random-forest-based atlas encoding scheme (known as Atlas Forests). State-of-the-art performance is demonstrated at considerably lower computational cost.

3.3.3 Classification of Non-Medical Complex Volumetric Imagery

Prior work related to the automatic classification of objects within complex 3D volumetric imagery is limited. Here it is implied that medical-CT imagery is not complex (or cluttered) in comparison to security-screening imagery (Chapter 1.2). The techniques reviewed here are those which consider complex imagery of the nature described in Chapter 1.2. Chen *et al.* [BCZX08] address the classification of pistols in Dual-Energy CT (DECT) imagery. DECT decomposition is performed using High-Low (HL) energy curves and look-up tables constructed for 28 calibration elements. In this way the chemical characteristics (effective atomic numbers and electron densities) of the scans are determined. For each volumetric image, the

central slice (cross-section) is assumed to contain all the information required for correct classification. Only this central slice is used in the classification procedure, thereby reducing the problem to 2D. Classification is performed by boosting 2D Haar-like features [VJ01]. The technique is evaluated using volumes containing only handguns with no clutter, noise or artefacts. While no experimental results are presented, the data used is not representative of that encountered in the real-world and it is unlikely that the aforementioned simplification to 2D will suffice in cluttered and noisy environments. Further work by the same author [BCZX09] presented a methodology for the detection of planar materials within baggage-CT imagery using a 3D extension to the Hough transform [Bal81].

Megherbi *et al.* [MFB10, MHBF12] present a comparison of classifier-based approaches using volumetric shape characteristics for the classification of pre-segmented objects in cluttered volumetric CT imagery. The performance of combinations of three shaped-based feature descriptors (rotationally invariant 3D Zernike descriptors [NK04]; the Histogram-of-Shape Index (HSI) [DJ95] and a combination of the two) and five classifiers (Support Vector Machines (SVM) [BHW10]; neural networks [Wan90]; decision trees [SL91]; boosted decision trees [CS13] and random forests [CS13]) are considered for the classification of pre-segmented bottles. Although encouraging classification results are presented, particularly for the HSI descriptor used in conjunction with the SVM or random-forest classifier (correct classification rates in excess of 98.0%), only a very limited dataset is considered. The effects of image noise and artefacts are not considered.

Extending upon their earlier work [FBM13], Flitton *et al.* [FMMB13] present an experimental comparison to investigate the suitability of the Bag-of-Words (BoW) model [SZ03] for the detection of threat items in both manually-segmented as well as unsegmented baggage-CT imagery. Combinations of four 3D interest-point descriptors (Density Histograms (DH) [FBM13]; Density Gradient Histograms (DGH) [FBM13]; the 3D Scale-Invariant Feature Transform (SIFT) [FBM10] and the 3D Rotationally-Invariant Feature Transform (RIFT) [LSP05]) and three codebook assignment methodologies ((hard, kernel and uncertainty) are considered. The classification of two classes of threats (handguns and bottles) in manually pre-segmented subvolumes indicates that optimal correct classification rates ($\sim$ 89% for bottles; $\sim$ 97% for handguns) are obtained using an uncertainty assignment protocol [vGVSG10] in conjunction with simple density-based descriptors [FBM13] sampled at 3D SIFT [FBM10] keypoint locations. The impact of the classifier type, the clustering method and the keypoint detection protocol are, however, not considered. Further experimentation, using unsegmented whole volumes, is shown to result in a significant decline in performance (with false-positive rates in excess of 15%). Poor resolution, image noise and metal-streaking artefacts characteristic to baggage-CT imagery, are shown to negatively impact the efficacy of the 3D

descriptors and ultimately the classification performance on both the segmented and unsegmented volumes. Measures to reduce the effects of noise and artefacts are not considered. Despite the decline in performance associated with the unsegmented data, this currently represents the only fully-automated approach to object classification in low-quality, complex volumetric CT imagery.

Finally, Flitton *et al.* [FBM12] have presented what may perhaps be considered the current state-of-the-art in the automated classification of objects in low-quality, complex volumetric imagery. Particularly, a novel 3D extension to the hierarchical visual cortex model for object classification [SWP05] is used for the automated detection of threats in manually segmented 3D baggage-CT imagery. The approach is shown to outperform a traditional BoW approach with correct detection rates in excess of 95% and low false-positive rates. Performance is however, hindered by the presence of noise/artefacts and the high degree of clutter. Furthermore, an extremely high computational cost is associated with the construction of the model. Noise and artefact reduction are again not considered.

3.3.4 Classification Summary

A brief review of the most popular and traditionally successful image-classification techniques has been presented. The individual components of a typical classification framework (feature extraction; feature description and classification) have been discussed. While feature descriptors based on the Scale Invariant Feature Transform (SIFT) [Low04, KS04] are generally considered optimal in 2D classification tasks, this does not appear to be true in complex and noisy volumetric imagery - where simpler density statistics-based descriptors have been shown to outperform the more complex SIFT and RIFT [LSP05] descriptors [FBM13].

Descriptors are typically computed at keypoints detected using a variety of scale and/or affine invariant interest-point detectors. Within the popular Bag-of-Words (BoW) environment, however, it has been shown that classification performance may be significantly improved by adopting a dense-feature-point sampling strategy (whereby interest points are randomly and densely sampled throughout the entire image). The increase in computational demand associated with a dense sampling strategy is proportional to the density of sampling grid. For this reason, the already computationally-demanding k-means clustering vector-quantisation method has not been previously considered within the 3D imaging domain.

Support Vector Machines (SVM) [Vap00] have traditionally been one of the most widely adopted and successful classifiers in the computer-vision literature. Ensemble classifiers (whereby strong classifiers are built as collections of weak classifiers) are known to offer improved classification performance. In particular, random forests-based classifiers [Bre01] have enjoyed a massive increase in popularity in recent years. Owing to their efficiency and good generalisation (par-

ticularly in multiclass and high-dimensional classification tasks [Cri11, CKY08]), random forests have been successfully applied to a broad range of classification and recognition-based tasks.

Randomised clustering forests [MNJ08] offer an efficient alternative to the k-means clustering approach used in the traditional BoW model. This has allowed for the benefits of dense-feature sampling strategies to be exploited without the associated increase in computational demand. Although encouraging results have been demonstrated in classical 2D classification tasks [MTJ07, MNJ08], the concept has, to date, not been considered in 3D.

Finally, the current state-of-the-art in the automated classification of objects in low-quality, complex volumetric CT imagery relies on the manual segmentation of the input data; incurs large computational overhead (in building the cortex model) and suffers a decline in performance in the presence of image noise and artefacts. Although the need for manual segmentations have been eliminated in the fully-automated approach of [FMMB13], the technique is computationally expensive and leads to false-positive classification rates in excess of 15%.

3.4 Dual-Energy Computed Tomography (DECT)

Conventional, single-energy Computed Tomography (CT) systems produce reconstructions representative of the Linear Attenuation Coefficients (LAC) of the object under investigation. That is to say, the greyscale intensity values (i.e. CT numbers, in Hounsfield Units (HU)) in the CT image are dependent on the LAC of the scanned object. Consequently, it becomes challenging and in some cases, impossible, to distinguish between materials that share similar LACs. In contrast, Dual-Energy CT (DECT) techniques, whereby attenuation data is captured using two distinct X-ray spectra, offer a means for characterising the chemical composition (e.g. atomic number and electron density) of the material under investigation based on its response under these different spectral conditions.

Dual-energy computed tomography is not a new concept. In fact Godfrey Hounsfield made mention of it in his pioneering work on computed tomography in 1973 [Hou73]. Despite an early interest in DECT techniques, technological limitations (e.g. unstable CT numbers, insufficient tube currents at low tube voltages and poor separation of energy spectra [Jin11]) have meant that the first commercial clinical DECT system was produced as recently as 2005 by Siemens Healthcare. As a result of recent advances in CT technology, the popularity of dual-energy-based CT imaging has enjoyed a rejuvenation with successful applications in a broad range of medical-imaging tasks [Joh11], industrial applications such as non-destructive material evaluation [NWK+11, MSVGJ03] and illicit material detection in airport security-screening [Sin03, YNC06].

In general, DECT techniques fall into one of three categories [Jin11]: 1) post-reconstruction techniques; 2) pre-reconstruction techniques and 3) iterative-reconstruction techniques.

3.4.1 Post-Reconstruction Techniques

Post-reconstruction (or image-based) DECT techniques are both the most straightforward and the most widely used in the medical-imaging domain. In fact, the most successful clinical application of DECT to date has been the differentiation of iodine (a commonly used contrast medium in CT) [Joh11]. DECT for material differentiation exploits the different absorption characteristics of materials with differing atomic numbers [JKS+07]. Since human tissue is composed predominantly of atoms with low atomic numbers (namely: hydrogen (Z=1), carbon (Z=6), nitrogen (Z=7) and oxygen(Z=8)), they exhibit very similar attenuation characteristics across the clinically acceptable X-ray energy range (i.e. they have similar CT numbers at low and high energies). In contrast, the CT numbers of materials with high atomic numbers vary considerably at differing energies. Therefore, the use of contrast mediums such as iodine (Z=53), which produce higher attenuation at lower tube voltages, allows for the differentiation of materials at different energies by a direct measurement of the ratio of the high and low-energy CT numbers [JKS+07, Joh11]. According to these principles, so-called *three material differentiation* has become a well-established method in medical-CT [JKS+07, LYPM09]. The basic concept is best explained by use of an example. By plotting the high and low-energy CT numbers of three (sufficiently different) materials of known density, a material differentiation may be performed. Figure 3.5 illustrates this plot for material differentiation in the liver [JKS+07]. Hepatic tissue is composed predominantly of soft-tissue, water and fat. These components exhibit an approximately linear relation between attenuation and energy. The addition of the contrast medium iodine to the liver tissue alters the spectral behaviour of the components and causes a displacement in the CT numbers from the (contrast-free) straight line. The change in the spectral behaviour arises due to the strong photoelectric effect of iodine relative to the comparatively weak photoelectric effects of soft-tissue and fat. The shaded area in Figure 3.5 represents the different ratios of soft-tissue, fat and iodine in the liver and is used to determine the material composition ratios for a given low and high-energy CT number [JKS+07].

DECT-based material differentiation has been successfully applied to a variety of tasks including: the characterisation of kidney stones and gallstones in abdominal imaging [GJCM09, CNB+10]; the characterisation, discrimination and monitoring of lesions and nodules in the liver [AHR+11], kidneys [GJCM09, GJH+09], adrenal glands [GHM+10], pancreas [GJCM09, MSK+10] and lungs [SBHC+12];

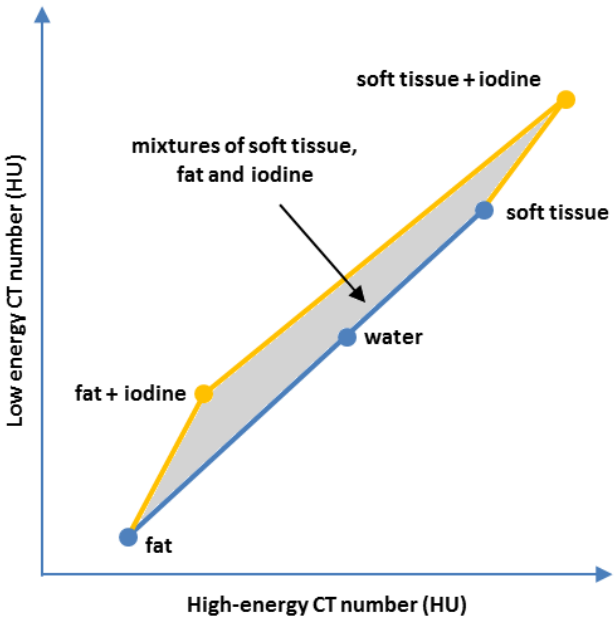


Figure 3.5: Principle of three material differentiation: differentiation of fat, soft-tissue, water (the three main components of liver tissue) and iodine in the liver. The fat, water and soft tissue components are approximately linearly related. The addition of iodine displaces the CT values of the liver tissue from this line [JKS+07].

virtual colonoscopy for the detection of colorectal lesions [NKY12]; lung perfusion imaging [TBH+08]; myocardial perfusion imaging [RLZ+08] and bone and plaque removal for improved quantification of calcified carotid stenoses in head and neck angiography [UWH+09, TKK+10]. The differentiation of tissues without the application of contrast media is considerably more challenging and remains an unsolved problem [Joh11].

A further post-processing application of DECT involves combining the low and high-energy CT images to produce a so-called *mixed* or *fused* image [EHIS+08, Joh11]. In DECT, the low-energy images typically exhibit superior contrast resolution but lower signal-to-noise ratios compared to the high-energy images [EHIS+08]. The objective of a fused image is to optimally combine the low and high-energy images such that the benefits of both are preserved. This fusion is generally achieved via a simple, fixed linear combination of the high and low-energy scans:

$$I_M = w.I_L + (1 - w).I_H \tag{3.4}$$

where I_M, I_L, I_H are the mixed-energy, low-energy and high-energy images and w is a weighting factor (usually fixed at $w = 0.3$). Eusemann *et al.* [EHIS+08] present a comparative study demonstrating that the traditional fixed linear combination of the high and low-energy images does not optimally capture the benefits of the individual scans. Instead, an organ-specific (i.e. tunable) nonlinear, sigmoidal fusion scheme is shown to outperform the linear approach.

The Dual Energy Index (DEI) of a material is a simple technique for quantifying the dual-energy behaviour of scanned materials and can be used as a means for material differentiation [Joh11]. The DEI of a material in air (i.e. not dissolved in water) is given by [Joh11]:

$$DEI = \frac{x_L - x_H}{x_L + x_H + 2000} \tag{3.5}$$

where x_L and x_H are the pixel values (in Hounsfield Units (HU)) for the low and high-energy scans respectively. By definition, the HU value of water is 0, and it remains unchanged at different energy levels of CT imaging. The DEI value of water is thus 0. Materials which have effective atomic numbers less than that of water ($Z_{\text{eff}} < 7.42$) have negative DEI values (since the HU values of such materials decrease with decreasing photon energy). In contrast, materials with atomic numbers greater than that of water have positive DEI values (since HU values increase with decreasing energy) [CZLY13]. The DEI of a mixture of two materials falls between the DEI of the constituent materials [Joh11]. While the DEI is therefore an indicator of the effective atomic number of a material, in contrast to the true effective atomic number (Section 3.4.2), its value does not rely on the photoelectric cross-section characteristics of the material (which are not precisely known) [Joh11]. Furthermore, its computation does not require a calibration procedure or the availability of raw-data. Despite its ease of computation, the DEI has demonstrated potential in material differentiation for a variety of clinical tasks including: the differentiation of air and tagged faecal materials from soft-tissue colonic structures in CT colonography [CZLY13] and the chemical characterisation of urinary stones in abdominal CT imagery [GJCM09]. It is worth noting that the DEI best discriminates between materials with atomic numbers less than 40, beyond which the relationship between DEI and Z becomes approximately uniform (i.e. DEI remains constant for increasing Z) [Joh11].

The predominant limitation of post-reconstruction DECT techniques is their susceptibility to artefacts in the reconstructed images [SM12].

3.4.2 Pre-Reconstruction Techniques

Alvarez and Macovski [AM76] pioneered the so-called pre-reconstruction DECT technique by modelling the total attenuation of X-rays as a linear combination of the photoelectric absorption and Compton scattering coefficients using a non-linear polynomial approximation of the polychromatic measurement models [AM76]. This early work has formed the basis for a broad range of techniques known as *basis material decomposition* methods [CH87, KPVK86, NBC03, YNC06].

The physical basis of DECT imaging relies on the energy dependence of the in-

teraction of X-ray photons with matter (Section 2.1.2). More particularly, within a photon energy range of approximately 30 keV to 200 keV, these interactions are known to be dominated by the photoelectric effect and Compton scattering [AM76]. Alvarez and Macovski [AM76] have demonstrated that, under these circumstances, the total attenuation of an X-ray beam may be modelled as follows:

$$\mu(x,y,E) = \underbrace{a_c(x,y)f_{KN}(E)}_{\text{Compton scatter}} + \underbrace{a_p(x,y)f_p(E)}_{\text{photoelectric effect}} \tag{3.6}$$

where (x,y) represents the coordinates of the material being scanned; E is the incident X-ray energy; $f_p(E)$ represents the energy dependence of the photoelectric effect and $f_{KN}(E)$ is the Klein-Nishina cross section for Compton scattering. Particularly, the photoelectric effect is proportional to $1/E^3$, while the energy dependence of Compton scattering is governed by the Klein-Nishina formula (the functions $f_p(E)$ and $f_{KN}(E)$ have been discussed in further detail in Section 2.1.2). The parameters $a_c(x,y)$ and $a_p(x,y)$ are constants that are dependent on the material composition [NBC03]:

$$a_c(x,y) = \rho_e(x,y) \tag{3.7}$$

$$a_p(x,y) = \rho_e(x,y)BZ^n(x,y) \tag{3.8}$$

where $B = 9.8 \times 10^{-24}$, $n \approx 3$, $Z(x,y)$ is the atomic number and $\rho_e(x,y)$ is the electron density given by:

$$\rho_e(x,y) = N_A \left(\frac{Z(x,y}{uA(x,y)} \right) \tag{3.9}$$

where $N_A = 6.02252 \times 10^{23}$ mol^{-1} is Avogadro's number, u is the unified atomic mass unit (1/12 the mass of an unbound neutral ^{12}C atom) and $A(x,y)$ is the relative atomic mass.

The fundamental principle of dual-energy CT involves acquiring attenuation measurements for an object at two different tube voltages (usually 80kVp and 140kVp for medical applications). This results in two separate attenuation profiles. Considering Equation 3.6 and assuming a polychromatic X-ray beam and the notation outlined in Section 2.2, dual-energy scanning produces two logarithmic projections:

$$P_H(r,\theta) = f_H(A_c, A_p) = -\ln \int S_H(E) e^{-f_{KN}(E)A_c(r,\theta) - f_p(E)A_p(r,\theta)} dE + \ln S_H(E) \tag{3.10}$$

$$P_L(r,\theta) = f_L(A_c, A_p) = -\ln \int S_L(E) e^{-f_{KN}(E)A_c(r,\theta) - f_p(E)A_p(r,\theta)} dE + \ln S_L(E) \tag{3.11}$$

where $S_H(E)$ and $S_L(E)$ are the high and low-energy spectra respectively and $A_c(r,\theta) = \int a_c(x,y)ds$ and $A_p(r,\theta) = \int a_p(x,y)ds$ are the line integrals of the Compton scatter and photoelectric absorption coefficients respectively. These dual-energy projections are typically acquired in one of three ways: 1) through the use of energy-resolving detectors [AST04]; 2) X-ray source spectrum switching [GWI+97] or 3) through the use of sandwich detectors (transmission dependent filtering based on material type or thickness) [RP79]. The dual-energy decomposition problem is to determine the Compton scatter coefficients A_c and the photoelectric absorption coefficients A_p of the material from the measured projections P_H and P_L. Alternatively, it has been shown that the attenuation coefficients for any material may be expressed as a linear combination of the coefficients of two basis materials, provided that the two chosen materials are sufficiently different in their atomic numbers (and hence in their Compton and photoelectric coefficients) [KPVK86].

While Equations 3.10 and 3.11 can be solved by direct approximation [Fen78, BBHM81], the more popular approach is to approximate the relationship between the dual-energy projections P_L and P_H and a set of decomposed projections as polynomial functions. Alvarez and Macovski [AM76] used a non-linear polynomial equation to approximate the integral P_H and P_L by a second order power series in A_c and A_p:

$$P_L = b_0 + b_1 A_c + b_2 A_p + b_3 A_c A_p + b_4 A_c^2 + b_5 A_p^2 \tag{3.12}$$

$$P_H = c_0 + c_1 A_c + c_2 A_p + c_3 A_c A_p + c_4 A_c^2 + c_5 A_p^2 \tag{3.13}$$

A calibration procedure is used to determine the coefficients b_i and c_i. Particularly, the projection values for two known materials of varying thicknesses are measured. Since P_L, P_H, A_c and A_p are known for the chosen materials, the coefficients b_i and c_i are computed using a polynomial least-squares fitting algorithm. Equations 3.12 and 3.13 are then solved using the iterative Newton-Raphson method [SBB+93]. It has been shown, however, that the method is both computationally demanding and unstable, making it susceptible to noise and large approximation

errors. Furthermore, the solutions are sensitive to the coefficients used in the polynomial approximations [CH87].

Methods based on Look-Up Table (LUT) procedures have been proposed and shown to be both faster (compared to Newton-Raphson methods [AM76]) and less sensitive to the numerical procedure, thus producing solutions which are less susceptible to noise [CH87, KPVK86]. Chuang and Huang [CH87] propose a method based on the use of iso-transmission lines and LUTs. For a given logarithmic transmission value, an iso-transmission line is represented by a linear equation in two basis functions:

$$P_L = at_a + bt_p \tag{3.14}$$

$$P_H = dt_a + et_p \tag{3.15}$$

where t_a and t_p are the aluminium and plastic equivalent thicknesses respectively and a, b, d, e are the regression coefficients which are proportional to the total attenuation coefficients of aluminium and plastic. The desired aluminium and plastic thicknesses (t_a, t_p) for a given pair of projection values (P_L, P_H) are determined by solving Equations 3.14 and 3.15 simultaneously (i.e. at the intersection of the two iso-transmission lines - see Figure 3.6). The computation of the regression coefficients require a calibration procedure. A set of predefined projection values are obtained by scanning various combinations of thicknesses of two well-defined materials (most commonly aluminium and plastic). The corresponding regression coefficients (computed from Equations 3.14 and 3.15) are stored in high and low-energy calibration tables. Linear interpolation between two predefined coefficients in a calibration table is then used to compute the regression coefficients for any new projection value. In clinical applications, the aluminium and plastic thicknesses are chosen to mimic the maximum possible equivalent thicknesses of bone and soft tissue in the human body. The acquired plastic and aluminium components may then be transformed into any equivalent set of basis materials or synthesised monoenergetic images using simple linear combinations [LAM+81]. The main drawback of the iso-transmission line method is the intensive calibration procedure required to generate the regression coefficients. It is also evident that for high projection values, the iso-transmission lines become increasingly non-linear. In those cases, the use of the linear equations in Equations 3.14 and 3.15 result in large approximation errors [NBC03].

The majority of dual-energy decomposition algorithms have been motivated by medical applications. There has, however, been a growing interest in the application of similar techniques for explosives detection in baggage-screening systems employed at airport security checkpoints [YNC06, NBC03, SM12]. The fundamen-

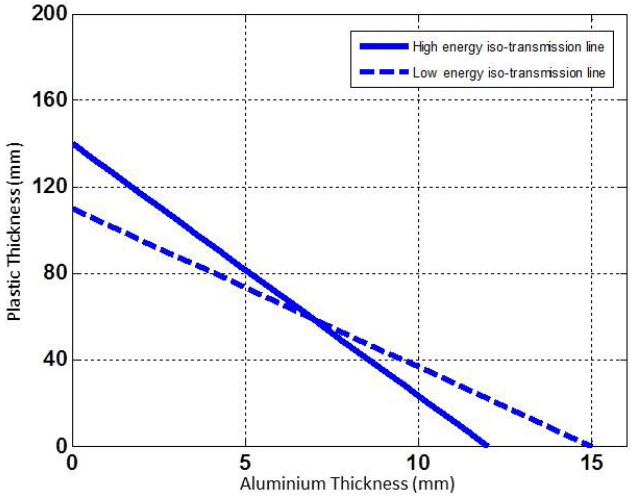


Figure 3.6: Iso-transmission lines for low and high-energy data: intersection gives equivalent plastic and aluminium thicknesses. Plot generated using simulations presented in Chapter 6.

tal objectives of dual-energy decomposition in the medical and security-screening domains differ. In the medical domain, the primary goal is to generate high-quality images to facilitate the diagnostic procedure, while in the security-screening domain, the main objective is the determination of the atomic properties of the objects in a scan to allow for materials-based explosives detection. The nature of baggage-CT data differs substantially from that encountered in the medical domain [MMFB12, MMvS+13]. In particular, the range of possible materials encountered in baggage scans is much broader and more unpredictable. Consequently, dual-energy decomposition using basis materials is more challenging since fairly accurate estimates of the combinations of basis material thicknesses to use in the calibration procedures are required. Ying *et al.* [YNC06] have further highlighted several limitations of traditional medical dual-energy decomposition methods (e.g. [AM76, KPVK86, CH87]) when used for explosives detection. These limitations include: high polynomial approximation errors ($> 200\%$ for [AM76]), caused by the large dynamic range of the photoelectric coefficients (resulting from the broad spectrum of materials encountered in baggage scans); a lack of boundary constraints in dual energy decomposition; image artefacts and X-ray spectral drifts.

Despite these fundamentally differing objectives and the increased complexity of baggage data, the dual-energy techniques designed specifically for baggage screening have been fairly similar to their medical counterparts [YNC06, NBC03]. Naidu *et al.* [NBC03], present a dual-energy decomposition approach for use in baggage-CT using a multi-step fitting procedure. An iso-transmission method based on [CH87], is used to solve the decomposition problem in Equations 3.10 and 3.11. In contrast to [CH87], the iso-transmission method is used to determine the photoelectric and Compton equivalent reconstructions (as opposed to the plastic and aluminium equivalent reconstructions).

Similarly to [CH87], a calibration procedure is performed to generate the LUT. Interestingly, the calibration data is generated using simulated low and high-energy spectra $S_L(E)$ and $S_H(E)$ (as opposed to the true, measured spectra). The literature indicates that this is a more commonly adopted approach in the security-screening domain [NBC03, YNC06, YNSC07]. It is worth noting, however, that although simulated spectra are used, CT images of known materials (termed Image Quality Phantoms or IQPs), obtained on the CT scanner under investigation, are generally used to calibrate the simulated spectra [YNC06].

Ying *et al.* [YNC06], propose a pre-reconstruction basis material decomposition method, whereby the photoelectric and Compton sinograms are obtained by solving a constrained least squares minimisation problem:

$$(A_c, A_p) = \arg\min_{A_c, A_p} \underbrace{[P_L(A_c, A_p) - P_L]^2 + [P_H(A_c, A_p) - P_H]^2}_{Q(A_c, A_p)} \tag{3.16}$$

subject to the constraint: $A_c, A_p \geq 0$ and where $P_L(Ac, A_p)$ and $P_H(A_c, A_p)$ are the measured (or simulated) low and high-energy projections respectively (Equations 3.10 and 3.11) and $Q(A_c, A_p)$ is the cost function. Additionally, techniques for adaptive scatter correction based on the work of Glover *et al.* [Glo82], destreaking by nonlinear filtering of the decomposed projections and real-time image-based correction for X-ray spectral drifts are incorporated into the proposed framework. The resulting approach, termed the Constrained Decomposition Method (CDM), is shown to yield numerically stable and physically meaningful solutions to A_c and A_p. Furthermore, the solutions A_c and A_p are continuous functions of P_L and P_H, eliminating the artefacts caused by discontinuities in A_c and A_p. The CDM is also shown to yield a significant reduction in the approximation and boundary constraint errors common to earlier methods - a comparison between the CDM approach and the approach of Alvarez and Macovski (AM) [AM76], for example, resulted in improvements in the approximation errors for the A_c and A_p projections respectively from 1.50% and 238.25% for the AM method to 0.00008% and 0.0002% for the CDM method. The large approximation error for A_p in the AM method leads to physically meaningless values (e.g. negative atomic numbers) preventing the direct application of such methods to the baggage-CT problem. The high-fidelity recovery of the photoelectric coefficient is, in fact, a common challenge in the majority of dual-energy decomposition methods. This is largely due to the domination of Compton scattering relative to the photoelectric effect in the photon energy range of interest for X-ray CT applications (Section 2.1.2). The photoelectric component typically presents with much higher degrees of noise (compared to the Compton coefficients), making the stable recovery of the photoelectric coefficients more challenging [SM12].

3.4.3 Iterative-Reconstruction Techniques

Iterative-reconstruction techniques [NH85, SV82], whereby CT images are reconstructed iteratively using a statistical model and a chosen objective function, are known to outperform traditional analytical approaches such as FBP in the reconstruction of images from incomplete projection data. The improved performance generally comes at a considerable increase in computation demand [MMvS+13]. The ever-increasing computational power of modern hardware, however, has seen an increased interest in iterative techniques.

Semerci and Miller [SM12] present a polychromatic DECT algorithm, tailored particularly for the detection of objects in unknown, cluttered environments (as typically encountered in baggage-CT images). The availability of some degree of *a priori* information regarding the Compton scatter and photoelectric absorption coefficients of the objects of interest is assumed. This prior information is incorporated (as a series of constraints) into a variational framework, using the Levenberg-Marquardt algorithm [Mar63] for minimisation. The photoelectric and Compton scattering parameters are then modelled as the superposition of a parametrically defined object of interest and a non-parametric background. The object model contains a geometric component (equal in the photoelectric and Compton images) and a contrast component (specific to the photoelectric and Compton images) and is based on a parametric level-set representation of the characteristic function of the object (via radial basis functions). The proposed approach provides simultaneous solutions to the problems of object detection and background reconstruction. Tested on simulated data, the algorithm is shown to successfully detect, locate and determine the geometric characteristics of objects of interest, while simultaneously producing reasonable background reconstructions.

3.4.4 Computation of the Effective Atomic Number

Intuitively, the effective atomic number of a material is an estimate of the equivalent, hypothetical element that will result in the same X-ray attenuation of the given material. The formal computation of the effective atomic number of a material requires precise knowledge of the composition of the material [WdB69]:

$$Z_{\text{eff}} = \left(\sum_i \frac{Z_i/A_i}{\sum_j Z_j/A_j} Z_i^n \right)^{\frac{1}{n}} \tag{3.17}$$

where i and j are indices referencing the individual elements composing the material; Z_i and A_i are the atomic number and relative atomic weight respectively of each individual element and n is a constant (traditionally $n \approx 2.94$ [Spi46]).

In the baggage-screening context, the predominant application of DECT has

been the determination of effective atomic numbers and densities for materials-based detection of explosives. [Sin03]. In its simplest form, the detection of explosives in baggage scans is based on two fundamental assumptions [Rod79]: 1) the majority of explosives may be characterised as organic substances with effective atomic numbers of approximately $Z_{\text{eff}} = 8$ and densities of $1.15 \leq \rho \leq 1.85$ g/cm^3 and 2) the majority of (non-metallic) innocuous items typically found in packed luggage (e.g. clothing, toiletries, books etc.) have densities of $\rho < 1.0$ g/cm^3. Figure 3.7, for example, illustrates Z_{eff} as a function of density for common substances (and several illicit materials) found in packed luggage [EK93]. Innocuous materials include organic substances (e.g. books, sausages, alcohol, leather, cotton etc.); inorganic substances (e.g. salt, PVC, plastic) and metals (e.g. iron, copper). The illicit drugs plotted are heroine and cocaine, while the explosives include C4, TNT, Semtex and Detasheet [EK93]. Such plots have traditionally been used by the US Federal Aviation Association (FAA) to evaluate the detection capabilities of a given scanner. Most importantly, the plot demonstrates that typical explosive materials (as well as illegal narcotics) are easily clustered and hence distinguished from other innocuous organic and/or inorganic materials. Based on these observations, it is then theoretically possible to distinguish between illicit and innocuous items by computing the effective atomic numbers and densities of the materials in a scan [Rod79]. Traditionally, this has been achieved via a simple calibration and interpolation procedure [SMK11].

A set of reference materials, with known chemical characteristics and whose effective atomic numbers and densities span the expected range of the materials of interest, are chosen. For explosives detection (where the materials of interest are typically organic with low Z_{eff}) a reasonable range is typically $5 \leq Z_{\text{eff}} \leq 14$ [Rod79, SMK11]. The low and high-energy Linear Attenuation Coefficients (LACs) for each of the reference materials are then measured on the scanner under investigation. The relationship between the known Z_{eff} and the measured LAC ratio (μ_H/μ_L) is approximated by an interpolating polynomial. Finally, the Z_{eff} of any unknown material may then be interpolated from the measured LAC ratio and the approximation polynomial. The procedure is summarised by the curve in Figure 3.8, which was generated using the following reference materials: ethanol, Delrin, water, Teflon, neon, sodium, magnesium and aluminium. The chemical properties of these materials are summarised in Table 3.1. The effective atomic numbers for ethanol, Delrin, water and Teflon were computed according to the classical equation (Equation 3.17). The low and high LACs were approximated using simulated X-ray energy spectra (Figure 6.1), the energy-dependent photon interaction cross-sections for each reference material (interpolated from published databases [BH87, CHK97]) and Equation 3.6.

It is important to note, however, that this fundamental approach assumes a

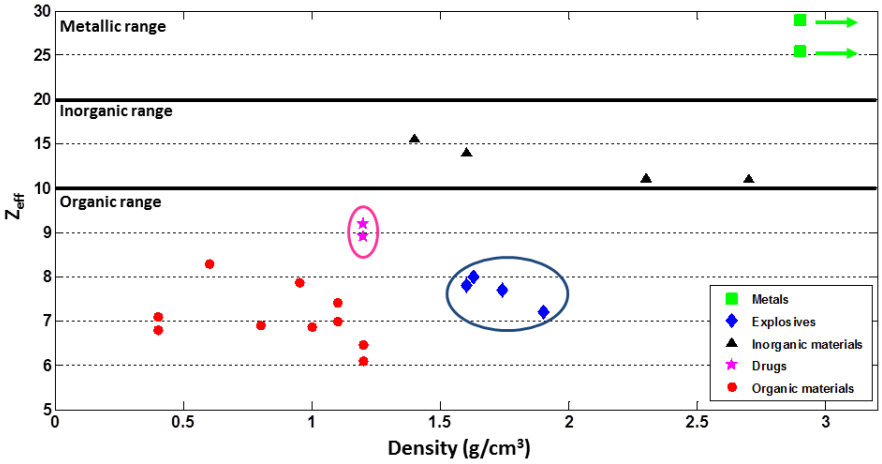


Figure 3.7: Z_{eff} as a function of density for common innocuous and illicit materials found in packed luggage [EK93].

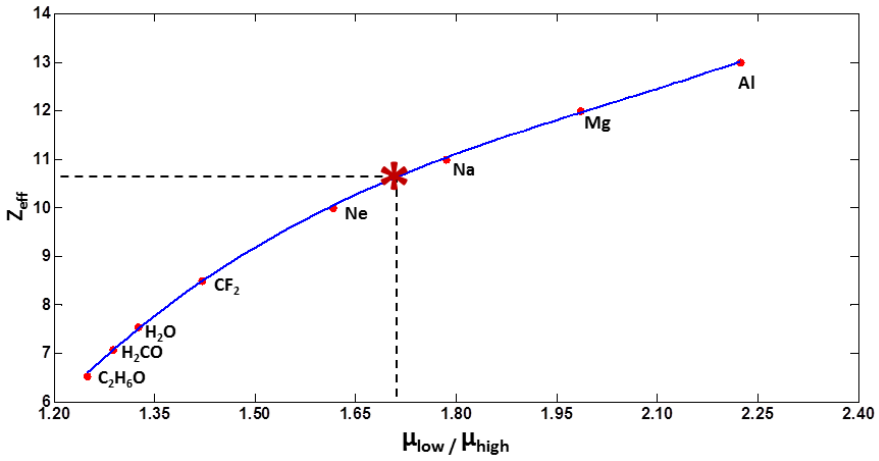


Figure 3.8: Determination of Z_{eff} by interpolation of approximating polynomial - approximated using the ratio of the measured (or simulated) low and high-energy LACs for a range of known reference materials [SMK11].

perfect measurement and observation of the constituents of the illicit and innocuous substances [Lu99]. Naturally, such ideal conditions are not encountered in reality. Furthermore, it is worth noting that various innocuous substances that have very similar chemical characteristics to common explosives (e.g. honey and chocolate), are typically not included in the evaluation of the detection capabilities of a system (as they would fall within very similar regions as the explosives in Z_{eff} vs. density plots) [Lu99, EK93]. Therefore, even under the assumption of ideal conditions, the discrimination of explosives by the interpolation of the Z_{eff} vs. density curve is at best a crude approximation.

Several more robust DECT-based techniques exist for computing the effective atomic number [AM76, LAM+81, NBC03, ZCZC06, YNC06, SMK11, SM12]. The majority of these rely on the decomposition of the low and high-energy data into equivalent Compton scatter and photoelectric absorption coefficients (using the methods described in Section 3.4.2). The most widely adopted approach for

Material	**Molecular Formula**	Z_{eff}
Ethanol	C_2H_6O	6.35
Delrin	H_2CO	6.95
Water	H_2O	7.42
Teflon	CF_2	8.43
Neon	Ne	10
Sodium	Na	11
Magnesium	Mg	12
Aluminium	Al	13

Table 3.1: Reference materials used for computation of Z_{eff} interpolating polynomial [SMK11].

computing an estimate of Z_{eff}, given the Compton scatter coefficient a_c and the photoelectric absorption coefficient a_p, is formulated as follows [AM76]:

$$Z_{\text{eff}} = K' \left(\frac{a_p}{a_c} \right)^{\frac{1}{n}} \tag{3.18}$$

where K' and n are constants. To obtain the photoelectric and Compton coefficients requires two separate reconstructions (one for the photoelectric image and one for the Compton image). The determination of Z_{eff} in this manner is thus computationally demanding. Ying *et al.* [YNC06] propose an alternative formulation for computing the effective atomic number:

$$Z_{\text{eff}} = K \left(\frac{a_p}{a_{hct}} \right)^{\frac{1}{n}} \tag{3.19}$$

where K and n are constants and a_{hct} is the CT number of the scanned materials (obtained from the high-energy CT image). This approach eliminates the need to compute the Compton reconstruction image, resulting in a significant reduction in computational demand. Furthermore, the division of the photoelectric image by the high-energy CT image, results in an elimination of the partial volume effect in the Z_{eff} image.

Equation 3.17 may be modified to compute the effective atomic number image given a decomposition of the dual energy data into two basis material images $I_1(x, y)$ and $I_2(x, y)$ [ZS12]:

$$Z_{\text{eff}}(x, y) = \left[\frac{\rho_1 I_1(x, y) Z_1^n + \rho_2 I_2(x, y) Z_2^n}{\rho_1 I_1(x, y) + \rho_2 I_2(x, y)} \right]^{\frac{1}{n}} \tag{3.20}$$

where ρ_1 and ρ_2 are the densities of the two basis materials and n is a constant typically in the range of 3-4 [YNC06].

3.4.5 DECT Summary

This section has provided a review of various Dual-Energy Computed Tomography (DECT) techniques. For the most part, DECT techniques may be grouped into one of three categories: post-reconstruction techniques; pre-reconstruction techniques and iterative-reconstruction techniques. Post-reconstruction techniques operate directly on the low and high-energy scans and are the most straightforward and computationally efficient approaches. The literature does however, indicate that the effectiveness of post-reconstruction techniques are limited by artefacts and noise and provide comparatively little discriminative power (compared to more advanced techniques). Pre-reconstruction DECT techniques are the most widely implemented techniques, particularly in the security-screening domain, where DECT decomposition and subsequent effective atomic number computations have been successfully used for materials-based explosives detection [Rod79, Sin03, SMK11, YNC06, NBC03]. Such techniques operate in the projection domain and typically decompose the low and high-energy scans into equivalent Compton and photoelectric parts. Using the fundamental laws of DECT, effective atomic number reconstructions may then be computed - providing an indication of the chemical makeup of the objects present in a scan. Theoretically, different object classes are distinguishable based on this information. Similarly to the post-reconstruction techniques, reconstruction artefacts pose a significant challenge to these techniques. Furthermore, decompositions typically require two FBP reconstruction per image, making pre-reconstruction techniques computationally demanding.

Similarly to single-energy CT, DECT based on iterative-reconstruction techniques provides superior performance, particularly considering the reduction of artefacts. Improved performance does however, come at an increase in computational demand. Nonetheless, such techniques are gaining popularity with the ever increasing computational power of modern hardware.

3.5 Segmentation

Image segmentation is a fundamental task in the field of computer vision and is a critical component of a large variety of image analysis algorithms. In statistics, the equivalent problem is generally referred to as *cluster analysis* and has been the source of a wealth of literature [NJW+02, FR98]. The prior literature addressing the problem of image segmentation is similarly extensive. Different algorithms are

often classified according to their underlying methodology, for example: thresholding techniques [S+04]; region-growing techniques [PBLL11, GVR+10, LW10]; split-and-merge techniques [Che91]; clustering techniques [CA79]; partial differential equation-based methods (e.g. level sets [VC02, HJWL13]); deformable model-based methods (e.g. active contours [KWT88] and Active Shape Model (ASM) [CHTH94, CTCG95]); graph-based methods [BJ01]; watersheds [VS91] etc. A comprehensive analysis of the current state-of-the-art in segmentation is not readily feasible. For the purposes of this study, the review of the current state-of-the-art is thus restricted to the topic of volumetric-image segmentation and CT imagery in particular [SSW88, SZD06, HvGS+09]. The reader is urged to refer to Appendix A which is presented in parallel with this review and provides a more comprehensive discussion of image segmentation and related concepts.

3.5.1 The Segmentation of Medical Imagery

In most medical applications, the segmentation task is concerned with a single anatomical structure and therefore different segmentation techniques have been developed specifically for particular structures or organs. Nonetheless, two common trends exist: 1) the exploitation of prior knowledge of the properties and characteristics of the relevant human anatomy to develop effective models and 2) the use of very simple techniques (e.g. greylevel thresholding) to provide coarse, initial segmentations, which are subsequently refined or completed. The literature addressing the problem of medical-image segmentation is vast (refer to [PXP00]) and only a brief overview of the most popular techniques is presented here.

Thresholding techniques are effective when the target structures have contrasting intensities or some other quantifiable features [PXP00]. Surprisingly, such a simple approach is often sufficient to achieve the desired segmentation in medical applications, where the target structures or organs are typically characterised by distinctive features such as image intensity or gradient magnitude [MTJM10]. Healthy lung tissue, for example, is characterised by a lower attenuation value relative to that of the surrounding anatomy. The majority of related segmentation techniques thus employ simple intensity-based operations to obtain a coarse segmentation of the healthy lung fields. These estimations are then refined (if necessary) using, for example, connected-component analysis [HHR01, UR05, SPvG05, vRdHV+09]. Basic thresholding, however, produces sub-optimal results in images with low signal-to-noise ratios and cases where the foreground is not easily distinguishable from the background (based on intensity values alone). A more comprehensive survey of thresholding techniques is provided in [SSW88].

Region-growing techniques extract image regions based on predefined connectivity/similarity criteria (e.g. intensity or edge information) [SZD06, BGMG+00].

Region growing, however, typically relies on the selection of an accurate seed-point (starting position) from which the region-growing kernel expands, extracting all pixels connected to the initial seed (according to the predefined criteria). In medical applications, seed-points are usually easily determined using prior-knowledge of the surrounding anatomy. A seed-point located in the trachea (which is easily located using a simple tube detection procedure) for example, generally yields satisfactory results for the segmentation of the lung fields. Several methods for eliminating the dependence of region-growing algorithms on accurate seed-points have been proposed and typically incorporate prior knowledge and statistical information [DYC03, PT01]. A more comprehensive review of segmentation by region-growing is presented in Appendix A.2.

Supervised machine learning based methods employ pattern recognition techniques to partition input images according to a model inferred from a set of training data [Sze10]. These techniques again rely on target structures being characterised by distinctive features (e.g. texture or shape). A large variety of approaches have been proposed, including: Multiple Sclerosis (MS) lesion segmentation using a kNN classifier [AVV08]; liver segmentation in CT imagery using artificial neural network classifiers [TT94]; segmentation of ultrasound images of the liver using support vector machines [KP03] and multi-organ segmentation in CT imagery using random forests [CSB09a, MSW+11b, GPKC12].

Atlas-based techniques (which employ machine-learning) are particularly popular in medical-image segmentation. An *atlas* is constructed by compiling information related to the (approximately constant) locations and shapes of the anatomical structures to be segmented. This may be performed manually or using information from existing segmentations. An unseen image is segmented via image registration techniques [MV98], whereby the input image is mapped to the coordinate space of the atlas. Atlas-based segmentation algorithms have been widely implemented - particularly in the segmentation of the structures of the brain within MR imagery [CPB+04, AHH+09, LWK+10].

Clustering-based techniques have enjoyed much success in the medical-imaging domain. Classical hard-clustering methods [Bab98] require that a given data point (e.g. voxel) either belongs to a particular cluster or it does not. The data is thus partitioned into a pre-specified number of mutually exclusive clusters. While hard-clustering techniques have been used for the segmentation of medical images [OSE05], fuzzy-clustering techniques are much more popular. In contrast to hard-clustering, fuzzy-clustering methods, as in fuzzy logic, allow data points to belong to several clusters simultaneously. The degree to which a given data point belongs to each cluster is defined by a value between 0 and 1, such that the sum of its membership degrees across all clusters is 1 [Bab98]. The vast majority of fuzzy-clustering algorithms are based on the optimisation of the basic c-means objective

function (or some modification thereof) [Bez81, Bab98], leading to what is commonly termed the Fuzzy C-Means (FCM) clustering algorithm. FCM clustering-based segmentation has been used for the segmentation of: neurological structures within MR images of the brain [SBSA03, JSX11, ASB+12]; breast lesions in dynamic contrast-enhanced MR images [CGB06]; tissues in ophthalmic-MR imagery [YHLL02]; the carotid artery in ultrasound images [ADES07]; pulmonary lesions within Positron Emission Tomography (PET) images [BZ10]; colonic polyps within CT colonography [YMFS04] and structures within abdominal-CT images [Tab07, WP08].

It is worth noting that several parameters need to be specified or initialised prior to using the FCM algorithm, namely [Bab98]: the number of clusters; the fuzziness exponent (determining the degree of fuzziness of the partition); the termination criterion; the norm-inducing matrix (determining the shape of the resulting clusters) and the initial fuzzy partition matrix (defining the initial cluster centroids). Of these, the number of clusters has the most significant impact on the final segmentation. Medical applications allow for relatively accurate estimates of the number of clusters to be selected - significantly increasing the likelihood of accurate segmentations. Furthermore, while the fuzzy partitioning matrix is typically initialised randomly (i.e. random initial cluster centroids are assigned), medical applications usually exploit knowledge of typical tissue distributions to more accurately initialise this matrix - improving accuracy and the rate of convergence [LY03].

Deformable statistical models are attractive options for a wide range of anatomical segmentation tasks due to the combination of ease of incorporating prior knowledge into the models and the predictable nature (e.g. shape and appearance) of the target structures or organs. The active contour model (or *snakes*), originally proposed by Kass *et al.* [KWT88], is a technique for detecting the boundary of an object in an image and was the first deformable model applied to the task of medical-image segmentation. Kass *et al.* [KWT88] define a snake (or contour) as an *"energy-minimising spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges"*. More particularly, an initial, parametrised contour is iteratively deformed (or evolved) by minimising an energy function designed to be locally minimal at the object boundary. The model is, however, sensitive to the initial placement of the contour and requires prior knowledge regarding the position and shape of the target object. Alternate constrained deformable models that have been met with success in the domain of medical-image segmentation include: Active Shape Models (ASM) [CHTH94, CTCG95, VGFS+02]; Active Appearance Models [CET01, BBLS05]; Geometric (or Geodesic) Active Contours (GAC) using level-sets [LGF00, Par02] and active contours without edges [CV01]. The general

technique of matching deformable models to image data has been extensively applied to the task of medical-image segmentation - for a more complete analysis of such techniques, the reader is referred to one of many relevant surveys (e.g. [MT96, HM09, MTJM10]. The predominant characteristic of such approaches, which make them ill-suited to the segment-all task in the baggage imagery domain, is their reliance on the consistency or predictability in the properties of the object being segmented (especially regarding shape).

Several medical-segmentation tasks present specific challenges (e.g. the segmentation of abnormal or pathological anatomy or the segmentation of complex structures such as the pulmonary vessel trees). In such cases, highly specified algorithms are typically developed with a particular abnormality or target structure in mind [vRvG13]. Consider, for example, the segmentation of the complete pulmonary vessel trees. *Vesselness* filters [FNVV98], based on the eigenvalues of the Hessian matrix or the eigenvalues of the structure tensor, exploit the greyscale curvature characteristics of tube-like structures against a dark background such that the vessels are enhanced and the surrounding anatomical structures are suppressed [FNVV98, WBB05]. Vesselness filters are often used as the core segmentation tool [FNVV98, ZCS+07, AAIW05] but have also been used to provide starting points for tree growing or tracking techniques [SHS04, ZCK+12]. Similar approaches have been successfully applied to the segmentation of the lung fissures [WBB05].

Approaches for the segmentation of pathological anatomy are often initiated with a coarse segmentation obtained via simple greylevel thresholding and then refined or completed using some case-specific technique. Various approaches have been proposed to this end including: probabilistic atlas-based segmentation [SPvG05, ASM+11, DFC+11, vRAvG07]; textural classification using statistical features [KKK+08, WLL09, RK96]; knowledge-based methods exploiting knowledge of the surrounding anatomy [SWB+11] and statistical shape-based learning approaches [SWB+11, SBB12].

The simultaneous segmentation of multiple anatomical structures is more akin to the task of segmenting complex baggage imagery. This problem has received notably less attention in the literature. The most significant contributions in this field address the issue of multi-organ segmentation in varied CT imagery [CSB09a, MSW+11b, GPKC12]. The task is approached as a voxel classification problem and is addressed via modifications to the popular random forest classifier [Bre01] - thus falling in the category of machine-learning based segmentation (see above). Random forests are feature-based classifiers and thus require the extraction of informative image features. As is the trend in medical-imaging applications (referred to throughout this study), the availability of prior knowledge is exploited. Particularly, anatomical context is captured via context-rich features which describe the relative position of visual patterns in the local anatomy. These

features are then used to build (or learn) a random-forest-based spatial-context model. An important advantage of such an approach is its low computational overhead [Bre01, CSB09a]. In fact, random forests have enjoyed increasing popularity in medical-image segmentation in general and have been successfully applied to the segmentation of adrenal gland abnormalities in CT imagery [SCS+13]; synaptic contacts in electron microscopy images [KSS+11]; foetal brain structures in ultrasound images [YNI+12]; the myocardium in real-time 3D echocardiography [LVNB09] and a range of structures in MR imagery (e.g. multiple sclerosis [ABGG+09, GCM+11]; high-grade gliomas [ZGK+12]; left ventricle [MGCA12], neurological structures [YCSB09, MBC+11]).

Overall, while this has not been a comprehensive review of the medical-imaging segmentation literature (refer to [PXP00, WK08, HvGS+09, MTJM10, GOM+12, vRvG13]), it has allowed for the emphasis of an important trend (in the context of this work) - namely, the dependency of the majority of the current techniques on the availability of prior anatomical information. While this is not a criticism of the aforementioned techniques, it does make them ill-suited to the task of segmenting *unknown* objects from complex 3D imagery.

3.5.2 Automatic Segmentation of Non-Medical Volumetric Imagery

Complex volumetric imagery acquired in the security-screening domain is typically characterised by low, anisotropic voxel resolutions; a high level of noise and artefacts; clutter (i.e. potentially large number of objects to segment) and a lack of prior knowledge regarding the contents of the scan [MMFB13]. Consequently, the segmentation of such data is a challenging task. Techniques that have been successfully applied to medical-CT imagery (Section 3.5.1), where the segmentation objectives are typically well defined, are thus unlikely to be equally effective in this domain.

This hypothesis has been verified by Megherbi *et al.* [MBFM13], who investigated the effectiveness of classical medical-segmentation techniques when applied directly to low-quality baggage-CT scans. In particular, four methods were evaluated on the task of segmenting bottles and handguns from complex baggage-CT imagery: 1) confidence connected region growing [PB99]; 2) fuzzy connectedness [US96]; 3) the watershed transform [VS91] and 4) fast marching [Set99]. It is found that careful parameter tuning is required on a per-case basis to obtain meaningful results and even then, the effects of image noise (despite pre-filtering), image complexity and the lack of prior knowledge regarding the target objects lead to suboptimal segmentation results which are notably poorer than those observed in the medical domain.

The bulk of the prior literature addressing baggage-CT segmentation in partic-

ular has its origins in a recent collaborative initiative between the US Department for Homeland Security (DHS) and the Awareness and Localization of Explosives-Related Threats (ALERT) Center of Excellence [CMP13]. The initiative (tasked with promoting academic and third party research in security-screening) led to the development of the following five baggage-CT-segmentation algorithms (two of which have appeared in peer-reviewed publications [WGW12, GSK+12]).

Wiley *et al.* [WGW12] present a 3D region-growing method based on the Stratovan Tumbler medical-segmentation technology [Wil09]. The technique is composed of five stages: 1) definition of a 3D kernel; 2) determination of the kernel movement criteria; 3) seed initialisation; 4) flood-fill and 5) splitting and merging. Optimal results are obtained using a spherical kernel, provided the size of the kernel (determined automatically, based on the amount of local clutter) is smaller than the object being segmented and larger than any expected holes in its boundary. The movement criteria for a given kernel are determined automatically using a training procedure, whereby initial criteria are matured by manually improving errant segmentations and adding each improvement to a central training file. A polynomial is fitted to the training points and used to determine the future movement criteria at any voxel. Seed-points are determined according to a voxel ordering method which ensures that large kernel sizes, high intensity voxels and voxels in the centres of objects are considered first. The 3D kernel traverses a volume in a flood-fill manner provided the traversed voxels satisfy the movement criteria. Composite objects are represented by hierarchical tree-like models. In particular, objects are initially segmented into multiple parts and pairs of segmented parts are merged if their degree of overlap exceeds a threshold. The study demonstrates high-quality segmentations for homogeneous objects and results in good separation of touching objects. Performance deteriorates for low-contrast objects, thin objects and in the presence of artefacts. It is also indicated that high-quality segmentations rely on near isotropic voxel resolutions in all three dimensions. The technique is presented in further detail in Appendix A.2.2.

Song *et al.* (TeleSecurity Sciences, Inc.) [CMP13] present a sequential approach composed of three stages: 1) pre-processing (by 2D bilateral filtering [TM98]; 2) object segmentation and 3) post-processing. Object segmentation is achieved using a sequential 'Segment-and-Carve' (SC) approach, operating on the principal that *easy* objects should be segmented first. The objects segmented in each stage are *carved* out of the image before proceeding to the next stage. Segmentation is performed using the Symmetric Region-Growing (SymRG) technique [WH03] (Appendix A.2.1) - a seedless (i.e. unsupervised) region-growing technique based on a symmetric function and which is invariant to starting conditions. In total, five SC stages are proposed, each targeting objects with different characteristics: 1) homogeneous, bulk objects; 2) homogeneous, medium thickness

objects; 3) homogeneous, sheet-like objects; 4) homogeneous, metal objects and 5) heterogeneous objects. Each stage is composed of five steps: 1) binary mask generation by thresholding; 2) mask pre-processing; 3) segmentation by SymRG; 4) boundary correction and 5) object carving. The five steps each require parameter tuning, with parameters differing for each stage. On completion of the five-stage sequential SC procedure, the segmented objects from each stage are subjected to extensive post-processing operations to correct for over and under-segmentations. Particularly, object-splitting is performed in four stages: 1) splitting by histogram analysis; 2) splitting by RANSAC; 3) splitting by recursive k-means clustering and 4) splitting by morphological opening. Object-merging is performed based on three thresholds: 1) spatial proximity; 2) mean intensity and 3) object type. While the study demonstrates high-quality segmentations for selected objects, the results for complete scans are not presented. The approach is extremely convoluted (with a large parameter set) and optimal performance requires careful parameter tuning.

Grady *et al.* [GSK+12] present a graph-based segmentation technique composed of three stages: 1) foreground identification; 2) candidate splitting and 3) candidate refinement. Foreground identification is performed by applying a Mumford-Shah functional [GA09] to artefact-reduced volumes (obtained by linear interpolation-based MAR [Tuy93]), producing labelled volumes (voxels labelled as foreground or background). Connected component analysis is applied to the labelled volumes. Each of the connected components in the foreground is recursively partitioned into candidate segments using the Isoperimetric Distance-Tree (IDT) algorithm [Gra06] (Appendix A.1). Recursions are driven by a novel Automated QUality Assessment (AQUA) metric, which automatically computes the quality of a segmentation without *a priori* knowledge of the object being segmented. Computational expense is optimised by performing coarse-to-fine segmentation (i.e. the segmentation from the previous level is used as the initial mask for further splitting). High-quality segmentations are demonstrated for challenging cases. Manageable run-times of approximately four minutes per volume (on an Intel Core 2 Duo 2.8 GHz machine) are presented. Low-density and sheet-like objects present the greatest challenges and it is suggested that superior MAR would be beneficial.

Harvey *et al.* (University of East Anglia) [CMP13], present a technique based on the multi-scale sieves class of algorithms [BCPL96, BHHC98]. Sieves function by filtering input signals to remove the intensity extrema at specific scales. In the context of image segmentation, *semantically meaningful* objects are removed at specific (typically higher) scales. The proposed approach is composed of four steps: 1) sieve the input volume to four logarithmically-spaced scales; 2) compute four *channel* volumes; 3) label the channel volumes by connected component analysis and 4) merge the labelled channel volumes into a single labelled volume. Merging

is performed by determining the similarities between the density histograms for each labelled object in each channel volume using the Kolmogorov-Smirnov (K-S) test [LHT10], which computes the probability that the histograms have been drawn from the same distribution. The specific strengths and weaknesses of the approach are not addressed in any great detail [CMP13]. An interesting observation is that since sieves segments all objects at all scales, at least one channel will always contain a segmentation of an object. It is thus proposed that a more useful approach (compared to channel merging) would be to pass the channel volumes directly into some artificial intelligence system (e.g. a classifier, object detector or salient region detector). The decision to merge the channels was dictated by the specifications of the ALERT initiative [CMP13]. The computational complexity of sieves is approximately $N \log p$, where p is image dependent and is proportional to the number of *flat-zones* (the largest connected components where the signal is constant) in the image.

Feng *et al.* (Marquette University) [CMP13] present a true 3D (as opposed to per-slice) technique which, although not explicitly specified, draws significantly from the automatic segmentation and merging technique of Ugarriza *et al.* [USV+09]. The approach is composed of three stages: 1) seed map generation; 2) adaptive region-growing and 3) merging. Seed maps are generated by locating sufficiently large homogeneous regions in the input volume. Homogeneous regions are determined by thresholding of the Sobel gradient map of the volume [SB10], while region size is determined by connected component analysis. Seed regions are grown by dynamic region-growing [USV+09], where the region-growing threshold is not constant. To compensate for the variation of intensities within objects (due to CT artefacts), the region-growing threshold is modelled as a non-linear function of the mean intensity of the region. On completion of the region growing, pairs of touching objects (i.e. those sharing a common edge) are merged based on their similarity in a 2D feature space (characterising mean texture and intensity). This merging heuristic is applied recursively. The technique is shown to be sensitive to parameter tuning and susceptible to under-segmentations (occurring in approximately 15% of cases).

Each of the five aforementioned baggage-CT segmentation techniques were developed and evaluated using a fully labelled volumetric baggage-CT data set captured on a single-energy medical-CT scanner with a resolution of $0.98 \times 0.98 \times 1.29$ mm. This data is not representative of the current benchmark in baggage screening, where data is typically captured on dual-energy scanners and are characterised by considerably poorer voxel resolutions. The development of segmentation algorithms for such data has not been considered previously.

3.5.3 Segmentation Evaluation Metrics

An accurate quantitative analysis of a segmentation produced by an algorithm is vital for the reliable comparison of different segmentation algorithms and to demonstrate performance improvements of novel approaches. Segmentation evaluation techniques may be divided into two broad categories: *subjective* and *objective* evaluation [Zha96]. In subjective evaluation segmentations are rated by human observers (e.g. five-level rating [AIGM98]). Although these are considered the gold-standard in clinical practice (if the rating is performed by expert radiologists), such evaluation is both costly and time consuming and not guaranteed to produce repeatable results (hence subjective) [HvGS+09]. The objective evaluation category presents a much richer array of techniques. In a comprehensive survey, Zhang *et al.* [Zha96] present an informative hierarchy of segmentation evaluation methods. Particularly, the objective methods are divided into two categories: *system-level techniques*, quantifying the impact of the segmentation on the larger system/application (e.g. impact on object recognition results) and *direct techniques*, which directly quantify the performance of a segmentation method. The direct evaluation techniques are then subdivided further into *analytical techniques*, evaluating the method itself and *empirical techniques*, evaluating the results of a given method. Finally, the empirical methods are classed as either *supervised techniques*, which employ ground-truth segmentations, or *unsupervised techniques*, which do not require ground-truth images. It should be noted that supervised evaluation techniques generally are not truly objective when the ground-truth images are manually created.

The most common approach in the medical literature may be classified as a supervised-empirical technique, whereby the algorithm-generated segmentations are compared with expertly delineated ground-truth segmentations using any number of discrepancy measures [Zha96, NBVV00] or combinations thereof [GJC01, DZS+07]. Such measures are typically based on volumetric overlap or surface distances [HvGS+09]. A list of the most commonly used metrics may be found in [GJC01]. It is worth noting again, however, that the manual generation of ground-truth data is a time consuming task (especially for 3D data) and is not guaranteed to provide true ground-truth segmentations (due to its intrinsic subjectivity) [BMFU+07].

Quantification of the segmentation error is straightforward when ground-truth data is available. In many real-world scenarios (i.e. beyond the algorithm development stage), online evaluations of segmentations may be required for a number of purposes (e.g. flagging poor segmentations; selecting optimal segmentations from a candidate set etc.). Naturally, ground-truth segmentations are not available in such scenarios (hence the requirement for a segmentation in the first place), demanding reliable unsupervised evaluation techniques. A number of methods have

been proposed to perform unsupervised evaluation of segmentations. Several such techniques are reviewed here - for a more comprehensive review, the reader is referred to the survey of Zhang *et al.* [Zha96].

Warfield *et al.* [WZW04] propose an Expectation-Maximization (EM) based algorithm for Simultaneous Truth And Performance Level Estimation (STAPLE). Essentially, the performance of different algorithms on a given segmentation task are estimated by their common agreement. In particular, the algorithm considers a collection of segmentations (produced by different segmentation algorithms) and computes a probabilistic estimate of the true (reference) segmentation as a weighted combination of these segmentations. The reference standard estimate may then be used for the evaluation of each algorithm using any standard discrepancy measure. Bouix *et al.* [BMFU+07] successfully employed this method to compare the performance of brain-tissue classifiers. An obvious limitation of STAPLE, however, is the danger of good segmentation scores arising from poor segmentations which are characterised by similar errors.

Kohlberger *et al.* [KSA+12] present an approach for estimating the segmentation error in the absence of ground-truth segmentations. Particularly, a generic learning approach, based on a set of novel segmentation features, is adopted to predict the volumetric overlap error [HvGS+09] and Dice coefficient [Dic45] of any given segmentation. A novel set of 42 shape and appearance features is proposed to characterise each segmentation. The proposed features are based on the objective functions used in popular energy-based and graph-based segmentation algorithms and are grouped into five categories: 1) unweighted geometric features (quantifying the size and regularity of the segmentation); 2) weighted geometric features (locally emphasising the geometric features when intensity values are similar to each other); 3) intensity features (measuring absolute intensity and intensity distributions within segmentations); 4) gradient features and 5) ratio features (computed as ratios of previously computed features). These features are then used to train a Support Vector Machine (SVM) classifier using the segmentation errors measured against a known ground-truth. The trained SVM can then be used to predict a given segmentation error using the features extracted from a segmentation without the need for a ground-truth segmentation. Having trained the classifier using the segmentations of eight organs, the proposed approach produces strong correlations between the predicted and true errors when applied to an unseen test set. The proposed approach is shown to produce considerably stronger correlations than the responses of Probabilistic Boosting Classifiers [Tu05] trained on the ground-truth segmentation boundaries.

Grady *et al.* [GSK+12] present an Automated QUality Assessment (AQUA) measure that provides a novel confidence measure that automatically computes the quality of a segmentation without *a priori* knowledge of the object being seg-

mented. The confidence measure is obtained via a data-driven approach for model learning. Particularly, 92 *good* object segments are identified, from which a set of 42 features (the same as those used in [KSA+12]) is extracted. A Gaussian Mixture Model (GMM), using 8 Gaussians, is then fitted to the data via Expectation Maximisation (EM) to create a model of a *high-quality* object. The trained GMM is then used to compute the AQUA measure for subsequent segmented objects. A high value for AQUA would indicate a high probability that the segmented object is a single, high-quality object. It is further suggested that the AQUA measure may be used to evaluate the overall segmentation of an image by averaging the AQUA scores of each individual object. In this way, different segmentation algorithms may be quantitatively compared in the absence of ground-truth. It is worth noting, however, that such an evaluation may fail in scenarios where an incorrect number of objects are segmented from a given image. Given an input image composed of three objects, for example - if algorithm A correctly segments the image into three components with AQUAs $(0.8, 0.7, 0.9)$, while algorithm B incorrectly segments the image into a single object (i.e. missing two objects) but with an AQUA of 0.9, then the overall AQUA for algorithm A would be 0.8 while that of algorithm B would be 0.9. This incorrectly suggests that the segmentation result of B is superior to that of A.

In conclusion, different segmentation tasks (and even different stages within a given segmentation algorithm) require different performance measures. Considering, for example, the broader objective of this study (i.e. the automatic subdivision of baggage volumes for object recognition) the ultimate performance of a particular segmentation technique would perhaps best be quantified via a system-level evaluation technique [Zha96] - whereby the impact of the segmentation on the overall classification results are compared. In order to determine the optimal results for a given segmentation method, however, an online empirical evaluation technique [Zha96] is more appropriate.

3.5.4 Segmentation Summary

This review has considered the segmentation of 3D volumetric imagery obtained from two imaging domains: the medical-imaging domain and the security-screening domain.

Based on the medical-imaging literature, several important observations (in the context of this study) have been made. The majority of medical-segmentation techniques exploit the availability of prior knowledge and are thus highly specified to particular anatomical structures. Comparatively few methods address the issue of the simultaneous segmentation of multiple anatomical structures - those that do, still rely on prior, contextual information [CSRK11, CSB09b, MSW+11a]. The dependence of medical-segmentation techniques on prior knowledge is an indica-

tion that such techniques will yield suboptimal results when the segmentation of multiple, unknown objects is required. This hypothesis has been substantiated by the comparative work of Megherbi *el al.* [MBFM13].

Secondly, a review of segmentation methods developed for the security-screening domain has been presented. Of the five methods reviewed, three are based on region-growing algorithms [WGW12, CMP13]. These techniques are shown to suffer from region leakage (where regions grow beyond the true object boundaries) when structures have poor contrast at their edges or when structures of similar intensities are adjacent to one another. As a consequence, a common post-processing step in the three region-growing approaches is to apply some form of additional splitting and/or merging operations. Such additional processing, however, expands the parameter space (and hence the degree of user interaction) and increases computational expense. Grady *et al.* [GSK+12] propose an optimisation to the isoperimetric graph partitioning method [GS06a] to address the issue of leakage through bottlenecks. The proposed isoperimetric distance tree algorithm (a graph-based method) produces high-quality segmentations, with relatively few cases of over-segmentation. The segmentation of thin, sheet-like objects presents difficulties for each of the five methods and it has been suggested that separate approaches be developed specifically for such cases. Furthermore, low-intensity objects are often missed and incorrectly labelled as background.

Finally, several methods for the quantitative evaluation and comparison of segmentation algorithms have been presented. The most suitable approach is shown to be dependent on the particular objectives of the segmentation task.

3.6 Summary

A review of the prior literature relevant to the research conducted in this thesis has been presented. The following research areas have been addressed: the reduction of noise and artefacts in CT imagery (particularly in the presence of metallic objects); dual-energy-based decomposition techniques; unsupervised 3D object segmentation and 3D object classification within low-resolution, complex volumetric imagery.

While the key findings and observations related to each have been summarised at the ends of their respective sections, it is worth highlighting those particular areas where the current state-of-the-art is lacking and where opportunities for novel contributions exist.

Noise and artefact reduction: The vast majority of denoising and MAR-based CT literature is found in the medical domain. The development of novel techniques or the evaluation of existing (medical) techniques in novel applications of CT imagery (outside of medicine) are extremely limited [MMFB12, XZX+09,

GSK+12]. The differences in the nature of medical images and those encountered in other domains (particularly regarding the lack of *a priori* information) mean that the state-of-the-art techniques from the medical literature are not guaranteed to be successful when applied to non-medical-CT imagery. Existing studies are limited in the techniques that are compared, the CT domains or applications which are considered and the performance-evaluation techniques that are employed.

Classification: The current state-of-the-art in 3D object classification in non-medical complex 3D volumetric imagery (such as that obtained in the security-screening domain) [FBM12] relies on the manual segmentation of the input data; incurs large computational overhead and suffers a decline in performance in the presence of image noise and artefacts. An efficient, fully-automated classification framework that is robust to image noise and artefacts does not currently exist in this domain. Furthermore, despite the proven benefits (in terms of classification accuracy) of densely sampled feature points, such strategies have not previously been considered in this domain.

Segmentation: Volumetric segmentation techniques (existing predominantly in the medical literature) are typically designed with a particular target object in mind and are unlikely to be effective for the segmentation of multiple, unknown objects. The blind segmentation of unknown objects in cluttered volumetric imagery is considerably more challenging and comparatively few solutions exist. The state-of-the-art in this domain [Gra06] has been developed using high-resolution medical-grade imagery with relatively low levels of noise and metal-streaking artefacts [CMP13]. Such data is not representative of that encountered in the current security-screening domain. The segmentation of low, anisotropic resolution volumetric imagery in the presence of multiple metal objects has not been considered previously.

The remainder of this work strives to address each of the aforementioned limitations via the development of novel methodologies, with the ultimate objective of producing a fully-automated 3D object classification framework.

Chapter 4

Noise and Artefact Reduction

The origins and effects of image noise and artefacts in volumetric-CT imagery have been discussed in the preceding chapters of this work.

Although the topic of digital-image denoising has been studied extensively, resulting in a vast resource of literature, the denoising of complex volumetric-CT imagery, of the nature considered in this work, has received comparatively little attention [ZPA10b, ZPA10a]. There is evidence in the medical literature however, that simple denoising operations (which were not necessarily designed for transmission imagery) can improve the quality of CT images and benefit the implementation of subsequent, more complex operations [Hsi03, See01, YZB+07, DZX+08]. The value of such techniques when applied to low-resolution, complex CT imagery has however, not been considered previously.

Similarly, the vast majority of the literature addressing the reduction of metal-streaking artefacts in CT imagery is found in the medical domain. It has been shown (Sections 2.4.4 and 3.2) that the effects of metal-streaking artefacts are particularly pronounced in cluttered, low-resolution imagery. The differences in the nature of medical imagery and that encountered in other domains (e.g. the security-screening domain) mean that the Metal Artefact Reduction (MAR) techniques which have been successfully applied to medical images are not guaranteed to be met with the same degree of success when applied to non-medical-CT imagery.

The review of the literature presented in Chapter 3 has revealed that in the context of this work, the existing denoising and artefact-reduction literature is limited in the following areas:

1. The potential benefit of popular denoising techniques when applied to low-resolution complex volumetric-CT imagery has not been investigated.

2. The efficacy of state-of-the-art MAR techniques from the medical-CT literature has not been evaluated in non-medical domains.

3. To date, there do not exist any MAR techniques designed particularly for

the reduction of artefacts in low-resolution, cluttered baggage-CT imagery containing multiple metal objects.

4. Performance analysis is predominantly qualitative in nature. Comprehensive quantitative analyses of both denoising as well MAR techniques are rare [KEMB11].

These limitations/shortcomings are addressed by the following contributions presented in this chapter:

1. A comparative performance evaluation is conducted for six popular 2D image denoising techniques (based on the study of Buades *et al.* [BCM05b]) and the baggage-CT-enhancement technique of Zhou *et al.* [ZPA10b, ZPA10a] when applied to low-resolution, cluttered volumetric-CT imagery (Section 4.4.1).

2. A novel quantitative performance measure is presented, extending traditional performance evaluation approaches by evaluating the potential benefits of denoising on the application of more complex operations (volume rendering and 3D object classification) within the current imaging context (Section 4.1).

3. A comprehensive comparative performance evaluation (considering both qualitative as well as quantitative measures) is conducted for eleven state-of-the-art MAR techniques from the medical literature when applied to low-resolution, cluttered volumetric-CT imagery (Section 4.4.2).

4. A novel MAR technique designed specifically for cluttered baggage-CT imagery containing multiple metal objects is presented (and included in the above comparison) (Section 4.2).

The research presented in this chapter has been previously published as [MMFB12, MMFB13, MMvS+13, MMB+13].

4.1 3D SIFT-Based Performance Measure

In the context of this work, a predominant motivation for effective denoising is to aid the implementation of subsequent automated 3D object classification. The performance measure proposed here is developed with this objective in mind. Flitton *et al.* [FBM10] have investigated the implementation of object recognition in complex volumetric-CT imagery using 3D SIFT features. It is shown that the presence of noise and artefacts is the predominant factor that negatively impacts the quantity of valuable 3D-SIFT interest points detected and ultimately leads to

a decline in classification performance. With this in mind, it is proposed that a denoising algorithm that leads to an increase in the number of high-quality 3D SIFT points will benefit subsequent classification.

Interest points are detected using a 3D extension [FBM10] to Lowe's SIFT algorithm [Low99, Low04]. Similarly to the traditional 2D formulation, an initial candidate set of keypoints is taken as the local extrema of multi-scale Difference of Gaussian (DoG) volumes, where the DoG volumes are created by convolving the input volume $I(x, y, z)$ with 3D Gaussian filters $G(x, y, z, k\sigma)$ at different scales:

$$DoG(x, y, z, k) = I(x, y, z) \star G(x, y, z, k\sigma_s) - I(x, y, z) \star G(x, y, z, (k-1)\sigma_s) \quad (4.1)$$

where k is an integer representing the scale index. A voxel is then considered a local extrema if it is a minimum or maximum in its local 3 x 3 x 3 (i.e. 26 voxels) neighbourhood at the current scale k as well as in the 27-voxel neighbourhoods in the two adjacent scale space DoG volumes (i.e. at scales $(k+1)$ and $(k-1)$). This initial candidate set of keypoints is refined by discarding unstable keypoints caused by poor contrast if their densities are below a given threshold τ_c. The candidate set is refined further by discarding the keypoints related to poor localisation on edges - determined by a second threshold τ_e related to the Trace and Determinant of the 3×3 Hessian matrix of the DoG volume [FBM10].

This 3D SIFT point detector is applied to a given volume before and after denoising and the number of object and noise SIFT points are recorded (Figure 4.1). An object feature point is identified as one located on an object of interest within the CT image whilst a noise feature point is considered as one which is not on the primary object within the CT image (i.e. assumed to be caused by noise and/or artefacts). The ratio of the object feature points to total feature points (object + noise) is used as an indication of the performance of the given technique. It is assumed that an increase in this ratio will ultimately correspond to improved object recognition results. It is important to note that the method is best suited to volumes containing single isolated target objects, such that object and noise feature points are easily distinguished (e.g. Figure 4.1).

4.2 Distance-Driven Metal Artefact Reduction

The majority of sinogram-completion-based MAR techniques adhere to the following framework: metal segmentation, sinogram completion, final image reconstruction. The technique proposed here employs the concept of the virtual sinogram [AAAZ10] and follows the same general framework. Additionally, a

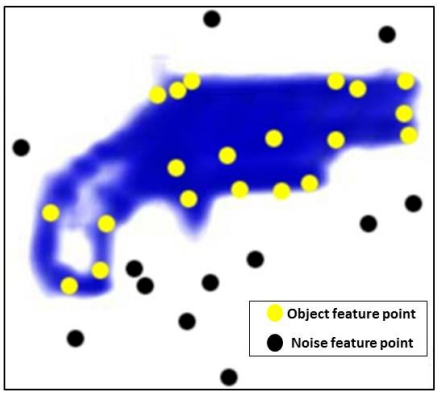


Figure 4.1: Object (yellow) and noise (black) 3D-SIFT interest points detected for an isolated handgun.

novel post-reconstruction refinement step is proposed to address the tendency of interpolation-based sinogram completion to introduce secondary streaking artefacts in the interpolated images [MB09] (refer to Section 3.2.1). The components of the proposed technique are discussed below (Figure 4.2).

4.2.1 Metal Segmentation

Metallic objects present in the original reconstructed image are segmented by binary thresholding, yielding a 'metal-only' image. Thresholding exploits the fact that the CT values of metals are extremely high, especially relative to other materials. A metal-free image is then constructed by assigning a constant pixel value to the metallic regions in the original, reconstructed image (the mean value of the background (i.e. non-metallic) region of the image is used). The metal-free image is then filtered with the edge preserving Non-Local Means (NLM) filter [BCM05b] (see Section 4.3) to reduce weak streaking artefacts and background noise while preserving the non-metallic regions of the image.

4.2.2 Reprojection and Sinogram Completion

The metal-only image and the filtered, metal-free image are forward projected using the Radon transform [KKRH+00], yielding the corresponding virtual sinograms. The metal-only sinogram is used as a mask to reference the corrupted/missing bins in the metal-free sinogram. The affected bins in the metal-free sinogram are then replaced by interpolated estimates from adjacent bins using spline interpolation.

4.2.3 Reconstruction

The interpolated sinogram is reconstructed to obtain the corrected, metal-free image. Reconstruction is based on the FBP algorithm [KKRH+00] (Section 2.3).

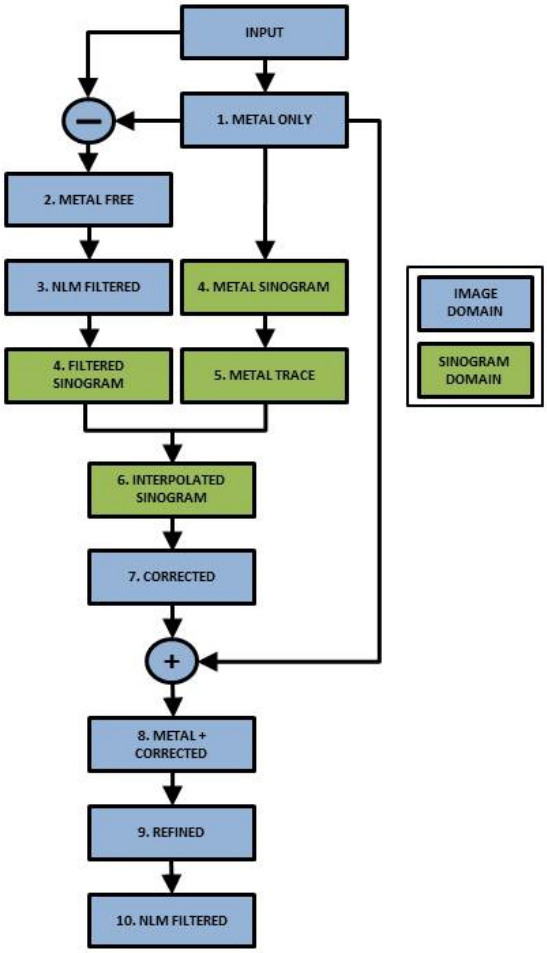


Figure 4.2: Distance-driven MAR flow chart. 1) Metal segmentation. 2) Metal removal 3) NLM filtering of metal free image. 4) Forward projection of metal only and filtered metal free images. 5) Metal trace identification. 6) Sinogram completion. 7) Image reconstruction 8) Reinsertion of metal objects. 9) Image refinement. 9) NLM filtering.

The metal objects are reinserted into the interpolated image, yielding the corrected image.

4.2.4 Image Refinement

The image refinement step is motivated by the fact that the secondary streaking artefacts introduced by the interpolation procedure generally manifest as intense, bright lines affecting the entire image [KCWM12]. Regions previously unaffected by streaking, but exhibiting secondary streaks, are thus characterised by higher intensities (see white arrows in Figures 4.28 and 4.29). It is likely that much of this secondary streaking may be removed by simply imposing an upper limit on the corrected pixel values, such that the intensities of the corrected pixels are less than or equal to the corresponding pixels in the original image. It is worth noting, however, that when considering images containing multiple metallic objects, a common manifestation of unequal beam-hardening across views [MND+99], is the appearance of dark bands (underestimated attenuation values) in the straight-line

regions connecting the metal objects (Figure 4.3). While the sinogram-completion approach generally yields adequate correction of such regions, the correction is likely to be undone by the aforementioned intensity-limiting procedure.

A distance-weighted refinement procedure, whereby the degree of intensity limiting is dependent on the location of the pixels relative to the metal objects, is proposed to address this limitation. Pixels falling within the straight-line regions connecting two metal objects are subjected to less intensive intensity refinement. For every pair of metal objects, a set of 'refinement weights' are computed in the following way (illustrated in Figure 4.3):

1. The centroid of each metal object is determined (red 'x' in Figure 4.3).

2. For the smaller of the two metal objects (Metal 1 in Figure 4.3), an ellipse having the same second-moments as the metal object is determined.

3. The angle that the line passing through the centroids makes with the horizontal is computed and compared to the angles that the major and minor axes of the ellipse make with the horizontal.

4. The width of the weight-mask is set to the length of the ellipse axis which is nearest in orientation to the normal of the line connecting the centroids.

5. The weights (in the range [0, 1]) are computed based on the Euclidean distance from the mask pixel to the nearest of the two metal objects such that pixels nearer to metal objects have higher weights. For a pixel $\mathbf{p} = (p_i, p_j)$ in the mask located at (i, j) on a rectangular image grid, the corresponding weight $w(i, j)$ is computed as follows:

$$w(i,j) = \frac{|D(i,j) - D_{max}|}{D_{max}} \tag{4.2}$$

$$D(i,j) = \min\{D(\mathbf{p}, L_1), D(\mathbf{p}, L_2)\} \tag{4.3}$$

where $D(\mathbf{p}, L)$ is the perpendicular Euclidean distance between the pixel $\mathbf{p}$ at image location (i, j) and the straight line L; $L_{1,2}$ are the straight lines passing through the centroids of the two metal objects, perpendicular to the line connecting the two centroids; D_{max} is the distance by which the mask extends beyond each metal object and is a tunable parameter (Figure 4.3). For every pixel outside of the mask, $w(i, j)$ is set to zero.

6. For a pixel $\mathbf{p} = (p_i, p_j)$ the refined intensity $I(i, j)$ is then computed as follows:

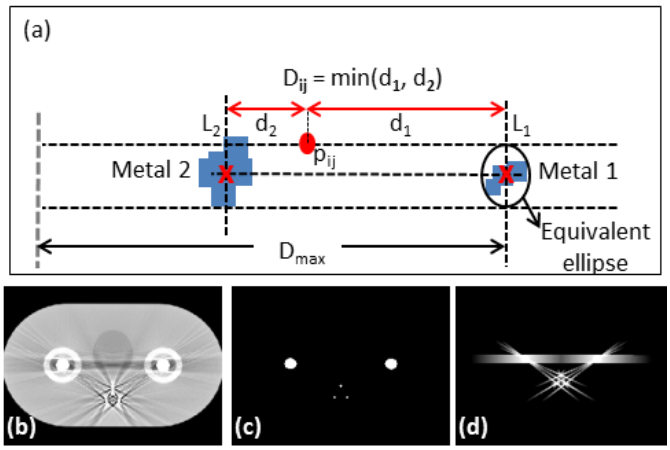


Figure 4.3: Distance-driven MAR weight computation. (a) Illustration of the weighted mask generation. The intensity refinement of pixel p_{ij} is dependent on the distance to the nearest metal object. The width of the mask is equal to the width of the ellipse surrounding the smaller of the two objects (Metal 1). (b) Example input image (c) Multiple metal objects (d) Resulting weighted mask.

$$I(i,j) = [1 - w(i,j)] \cdot I_0 + w(i,j) \cdot I_c \tag{4.4}$$

where I_c is the corrected intensity value of the pixel (i.e. after Step 3. in the aforementioned algorithm) and I_0 is the original (uncorrected) pixel intensity.

Finally, in order to eliminate the remaining weak streaking artefacts, the corrected image is again filtered with the NLM filter [BCM05b].

4.3 Experimental Methodologies

An explanation of the proposed experimental methodology is presented below. In particular, a brief overview of the selected denoising and metal artefact reduction techniques is provided together with an explanation of the proposed comparative methodologies (quantitative and qualitative performance-evaluation techniques) employed in the respective studies.

4.3.1 Denoising Techniques Compared

In addition to the Alpha-Weighted Mean Separation and Histogram Equalisation (AWMSHE) approach (a dedicated baggage-CT technique) of Zhou *et al.* [ZPA10b], the following six popular denoising techniques (based on the recent denoising survey of Buades *et al.* [BCM05b]) are compared: anisotropic diffusion [PM90]; Total Variation (TV) denoising [ROF92]; bilateral filtering [ZG08];

translation invariant wavelet shrinkage [CDAO95] and Non-Local Means (NLM) filtering [BCM05b]. While a brief mathematical basis for each is provided below, for a more detailed explanation the reader is referred to the relevant literature (as cited above).

In the following descriptions, $I(x, y)$ denotes the input (unfiltered) image and $I^*(x, y)$ the denoised (or restored) image. All the filters described here are applied in $\mathbb{R}^2$ (i.e. to the individual slices of a CT volume).

Anisotropic diffusion is a shape-adaptive filtering technique introduced by Perona and Malik [PM90]. The basis of the Perona-Malik algorithm is to evolve an image under an edge-controlled diffusion operator where the orientation of the filter is determined by the local gradient in the image. Details such as edges and lines are thereby preserved (or enhanced), while regions within edges are smoothed. The generalised Perona-Malik diffusion equation is given by [PM90]:

$$g(x, y) = \frac{\delta}{\delta t} I(x, y, t) = div(c(x, y, t)\nabla I(x, y, t)) \tag{4.5}$$

$$I(x, y, 0) = I(x, y) \tag{4.6}$$

where $\nabla(I) = \frac{\delta I}{\delta x}\hat{x} + \frac{\delta I}{\delta x}\hat{y}$ denotes the image gradient, $div(.)$ is the divergence operator and $c(x, y, t)$ is the diffusivity function, controlling the rate of diffusion. Edge information is preserved by modelling $c(x, y, t)$ as a function of the local image gradient [PM90]:

$$c(x, y, t) = 1/(1 + \frac{|\nabla I|^2}{K^2}) \tag{4.7}$$

where K is a so-called contrast parameter and is determined automatically using the noise estimator described by Canny [Can86]. Anisotropic diffusion has been successfully applied to a variety of medical-imaging applications including: the denoising of MRI data [GKKJ92]; pre-reconstruction sinogram restoration in X-ray CT imagery [HLC12] and the enhancement of tubular structures to aid the segmentation of vessels in 3D imagery [KMA97].

Total Variation (TV) denoising relies on the prinicple that reducing the Total Variation (TV) [ROF92] of an image, while maintaining a close match to the original image, removes image noise while preserving important details such as edges [BT09, GSZ03, Cha04]. Rudin *et al.* [ROF92] define the total variation of an image, $I(x, y)$, as:

$$J(I(x,y)) = \sum_{1 \leq x,y \leq N} |(\nabla I)| \tag{4.8}$$

where $|I| := \sqrt{I_1^2 + I_2^2}$ for every $I = (I_1, I_2) \in \mathbb{R}^2$. The total variation denoising problem is then formulated as [ROF92]:

$$\min_{I \in X} \frac{\|I - I^*\|^2}{2\lambda} + J(I) \tag{4.9}$$

where $\|.\|$ is the Euclidean norm and $\lambda > 0$ is a regularisation parameter. The iterative TV minimisation approach developed by Chambolle [Cha04] is used in this work to solve the minimisation problem in Equation 4.9.

Bilateral filtering is an edge-preserving smoothing filter defined by a Gaussian-weighted average of the pixels in a predefined local neighbourhood [ZG08, Gun10, TM98]. The technique is based on the principle that two pixels are similar not only if they are close to one another spatially but also if they exhibit some similarity in their photometric range or intensity [PKTD07]. The filtered pixel is computed as a weighted combination of its neighbouring pixels according to [PKTD07]:

$$I^*(x,y) = \frac{\sum_{k,l} I(k,l) w(x,y,k,l)}{\sum_{k,l} w(x,y,k,l)} \tag{4.10}$$

where the weighting coefficient $w(x,y,k,l)$ is computed as the product of a *domain* kernel $d(x,y,k,l)$ and a *range* kernel $r(x,y,k,l)$:

$$d(x,y,k,l) = e^{-\frac{(x-k)^2+(y-l)^2}{2\sigma_d^2}} \tag{4.11}$$

$$r(x,y,k,l) = e^{-\frac{\|I(x,y)-I(k,l)\|^2}{2\sigma_r^2}} \tag{4.12}$$

Translation-Invariant Wavelet Shrinkage: Image denoising in the wavelet domain has garnered considerable popularity in recent years. Wavelet shrinkage [DJ94, Don95], whereby a hard or soft threshold is applied to the wavelet coefficients of a noisy image, is the most straightforward approach to denoising in the wavelet domain. It has, however, been shown that traditional wavelet shrinkage leads to the generation of artefacts in the denoised image as the wavelet basis is not invariant under translations [CDAO95]. These artefacts may be reduced by averaging out the translation dependence of the wavelet basis using a technique termed *cycle spinning* [CDAO95].

Cycle spinning involves shifting (translating) the image in the spatial domain, applying the traditional wavelet-denoising procedure to the shifted image (i.e. in the wavelet domain) and then 'unshifting' the denoised image. This is repeated for a range of shifts. The final denoised image is obtained by averaging the results over all the shifts. This technique is used in this work in conjunction with the Symlet mother wavelet [GC11] with eight vanishing moments and a hard, VisuShrink threshold [CDAO95]:

$$\tau = \sigma\sqrt{2log(N)} \tag{4.13}$$

where N is the number of pixels in the image and σ is the standard deviation of the noise.

Non-Local Means (NLM) filtering [BCM05a, BCM05b] computes the mean of the values of all points whose Gaussian neighbourhood is similar to the neighbourhood of the current pixel. The estimated value for a pixel at coordinates (x, y) is computed as a weighted average of all the pixels in the image:

$$I^*(x,y) = \frac{\sum_{k,l} w(x,y,k,l)I(k,l)}{\sum_{k,l} w(x,y,k,l)} \tag{4.14}$$

where the weights $w(x, y, k, l)$ are computed based on the similarity of pixels $I(x, y)$ and $I(k, l)$:

$$w(x,y,k,l) = e^{-\frac{\|I(x,y)-I(k,l)\|_a^2}{h^2}} \tag{4.15}$$

where $\|I(x, y) - I(k, l)\|_a^2$ is a Gaussian-weighted Euclidean norm and a is the standard deviation of the Gaussian kernel. The parameter h is a constant proportional to the estimated noise in the input image. The NLM filter is widely accepted to be a powerful denoising technique which often yields superior results compared to other popular denoising techniques [BCM05b]. The NLM filtering algorithm in its original form is computationally demanding, making it ill-suited for practical applications [MS05a, BCM10]. Several recent studies have presented optimised implementations of the NLM algorithm, several of which use some form of block pre-classification to reduce the number of weighted-average computations required for denoising each pixel in an image [MS05a, WGY+06, BV08, OEW08, AGDL09]. Of these, the accelerated NLM implementation of Mahmoudi and Sapiro [MS05a] is used in this work. The algorithm is accelerated by pre-classifying neighbourhoods using the mean neighbourhood intensities and local gradients as measures

of similarity (i.e. by computing their responses to a mean intensity-based filter and a local gradient-based filter). The technique is shown to reduce the quadratic complexity of the original NLM algorithm to a linear complexity.

Alpha-Weighted Mean Separation and Histogram Equalisation: Zhou *et al.* [ZPA10b] present an image-enhancement algorithm that combines alpha-weighted mean separation and histogram equalisation to remove background noise and improve the contrast in CT-baggage imagery. The proposed algorithm is comprised of two stages: noise removal and image enhancement. The noise-removal step exploits the fact that much of the projection noise present in CT-baggage imagery is characterised by very low pixel values relative to the high dynamic range of the image, allowing for denoising using simple thresholding. An initial 2D CT image $I(x, y)$ is separated into an object image $I_O(x, y)$ (containing the valuable information in the image) and a noise image $I_N(x, y)$ (comprised of only noise) via Alpha-Weighted Mean (AWM) thresholding:

$$I_O(x, y) = I(x, y) \quad \text{for} \quad I(x, y) \geq \tau_1 \tag{4.16}$$

$$I_N(x, y) = I(x, y) \quad \text{for} \quad I(x, y) < \tau_1 \tag{4.17}$$

where the noise threshold $\tau_1 = \alpha_1 \overline{I}$ and $\overline{I}$ is the mean intensity of I. The noise image $I_N(x, y)$ is subsequently discarded, while the object image $I_O(x, y)$ is subdivided further into an upper and lower sub-image ($I_U(x, y)$ and $I_L(x, y)$ respectively) by applying a second threshold $\tau_2 = \alpha_2 \overline{I_O}$:

$$I_U(x, y) = I_O(x, y) \quad \text{for} \quad I_O(x, y) \geq \tau_2 \tag{4.18}$$

$$I_L(x, y) = I_O(x, y) \quad \text{for} \quad I_O(x, y) < \tau_2 \tag{4.19}$$

The upper image $I_U(x, y)$ contains the brighter regions of the object image, while the lower image $I_L(x, y)$ contains the darker (yet still informative) regions. The lower sub-image $I_L(x, y)$ is enhanced via Histogram Equalisation (HE) [SB10] yielding the enhanced image $E_L(x, y)$. The upper image $I_U(x, y)$ is clipped to the maximum value of $I_L(x, y)$ (to compress the data range without introducing new artefacts) yielding the upper enhanced image $E_U(x, y)$. The final image $I^*(x, y)$ is computed as the summation of $E_U(x, y)$ and $E_L(x, y)$:

$$I^*(x, y) = E_U(x, y) + E_L(x, y) \tag{4.20}$$

4.3.2 MAR Techniques Compared

Based on the review in Section 3.2, those methods claiming substantial performance gains are experimentally reviewed here. Techniques which show little to no improvement (qualitative or quantitative) when compared to simple interpolation-based approaches, as well as techniques which are characterised by extremely high processing times have been excluded from this comparative study. Furthermore, the literature shows that many of the more recent iterative reconstruction-based and hybrid MAR approaches rely on the efficacy of initial iterative approaches such as Maximum Likelihood Expectation Maximisation (ML-EM). Therefore, for the sake of simplicity, only the EM-based approach of Wang *et al.* [WSOV96] is considered here. It is expected, however, that the performance of the simple EM-based approach will be inferior to more recent iterative approaches that employ sophisticated priors and acquisition models. As is common practice in MAR-based literature, a linear-interpolation-based approach [KHE87] has also been included in the comparison.

In summary, eleven sinogram-completion-based approaches [KHE87, ZBWW02, BS06, YZB+07, JR09, LBY+10, AAA+10, MRL+10, MRS+11, MRL+12, MMB+13] and one iterative reconstruction approach [WSOV96] are compared. Hereafter, these twelve techniques are referred to using the following descriptors: Kalender [KHE87]; Wang [WSOV96]; Zhao [ZBWW02]; Bal [BS06]; Yu [YZB+07]; Jeong [JR09]; Li [LBY+10]; Abdoli [AAA+10]; Meyer1 [MRL+10]; Meyer2 [MRS+11]; Meyer3 [MRL+12]; DDMar [MMB+13]. Note that the DDMar method [MMB+13] refers to the distance-driven MAR approach presented in Section 4.2 which was denoted as Mou_{β} in the corresponding publication [MMB+13].

The MAR techniques were implemented according to the details available in the original publications. There is thus a possibility that the implementations evaluated here differ to some degree to those in the original works. Several of the selected techniques are dependent on several input parameters - these were determined empirically.

4.3.3 Denoising Performance Evaluation

Three images obtained on the Reveal CT80-DR baggage scanner are used in the evaluation of the selected denoising techniques. 1) A whole-volume scan composed of 71 512×512 axial slices and containing a single handgun in a container with no background clutter. This volume is used in the qualitative analyses of the selected denoising techniques. 2) A whole-volume scan composed of 102 512×512 axial slices and containing a handgun and a variety of background clutter objects (dumbbell, binoculars, pliers, batteries etc.). This volume is used in the volume-rendering experiments described below. 3) A subvolume scan composed of 34

60×40 axial slices and containing a single handgun only - used for 3D SIFT point evaluations. The subvolume scan was generated by manually cropping a handgun from a whole-volume scan containing a single handgun in an empty container. The densities in all three images are in Modified Hounsfield Units (MHU) (Section 1.2).

Denoising performance is evaluated in three ways (two qualitative and one quantitative), with a particular focus on the potential impact of denoising on subsequent operations. Firstly, a standard qualitative visual comparison of the volumes before and after denoising is performed. This is done for both the original volumes as well as the original volumes corrupted with Gaussian noise of known standard deviation. It should be noted that CT noise is generally assumed to be correlated Poisson noise (Section 2.4). Although it has been argued that this noise may be approximated by a Gaussian distribution with a nonlinear signal-dependent variance [LLHL02], this addition of Gaussian noise does not represent a model for reality and is instead intended to demonstrate the noise-removal and edge-preservation capabilities of each of the denoising techniques. Secondly, a volume-rendering technique, which uses a combination of Alpha-Compositing ray-tracing and Marching Cubes surface rendering (denoted ACMC) [LCNC98], is applied to a cluttered volume before and after filtering and the visual quality of the resulting volumes compared (thereby giving an indication of the impact of denoising on volume rendering as well as the efficacy of each technique in the presence of clutter). Finally, the 3D SIFT-based measure presented in Section 4.1 is used to provide a quantitative measure of performance.

4.3.4 MAR Evaluation Data

The CT data used in the evaluation of the selected MAR techniques is described below and summarised in Table 4.1. It is worth noting that, with the exception of the distance-driven MAR technique presented in Section 4.2, each of the MAR techniques included in the comparison have been designed for application in $\mathbb{R}^2$. Performance evaluation is restricted to this domain in order to remain consistent with the majority of MAR-based literature. Performance is evaluated under three scenarios: 1) a simulated medical environment; 2) a true medical environment and 3) a true, non-medical (security-screening) environment.

Prior work (Section 3.2) shows that it is accepted practice in the literature to use numerical simulations and mathematical phantoms to measure the performance of MAR techniques, as this allows for reliable quantitative performance analysis [YZB+07, MND+01, Man01, MND+00, LFN09, SN12].

A 2D phantom is thus employed. The phantom is composed of two large circular iron inserts (diameter 2 cm) surrounded by circles of cancellous (soft) bone; three small isolated circular iron inserts (diameter 0.4 cm) and a region of fatty

Experiment	Type	Size	Units	Num. images
Medical phantom	Axial slice	512 × 512	HU	1
Clinical (medical)	Axial slice	512 × 512	HU	1
Clutter-free bag	Axial slice	512 × 512	MHU	1
Cluttered bag	Axial slice	512 × 512	MHU	1
Bland-Altman (medical)	Axial slice	512 × 512	HU	42
Bland-Altman (bag)	Axial slice	512 × 512	MHU	72
Clutter-free volumetric	Whole volume	512×512×77	MHU	1
Cluttered volumetric	Whole volume	512×512×99	MHU	1

Table 4.1: Breakdown of test data used in MAR evaluation. The phantom, clinical, clutter-free bag and cluttered bag experiments refer to the evaluation of all 12 selected MAR techniques. The two Bland-Altman experiments refer to the comparative evaluation between the DDMar (Section 4.2) and Mou [MMFB12] techniques. The clutter-free volumetric and cluttered volumetric experiments refer to the evaluation of the DDMar technique when applied to all axial slices in the specified whole volumes.

tissue. The remainder of the phantom is water. The artefact free phantom is shown in Figure 4.5 (a). The numerical simulation of the phantom was performed using a simulator based on that presented by De Man *et al.* [MND+99], extended with a distance driven projector [MB04]. This 2D simulator has been used extensively in previous MAR studies [MND+00, LFN09, MND+01, MND+99, SN12]. The simulation models the effects of beam hardening (due to the polychromatic nature of X-ray spectra), scattered radiation, projection noise and the trans-axial non-linear partial volume effect (or the exponential edge-gradient effect (EEGE)). De Man *et al.* [MND+99] have cited these as the predominant causes of streaking in medical-CT images. The simulation includes a 10-times subsampling of the detector elements, a 5-times subsampling of the source (using a source width of 1mm) and a 5-times subsampling of the projection views (to model the continuous rotation of the gantry). The effects of afterglow and detector-crosstalk are not considered. Scatter is simulated according to the following formula:

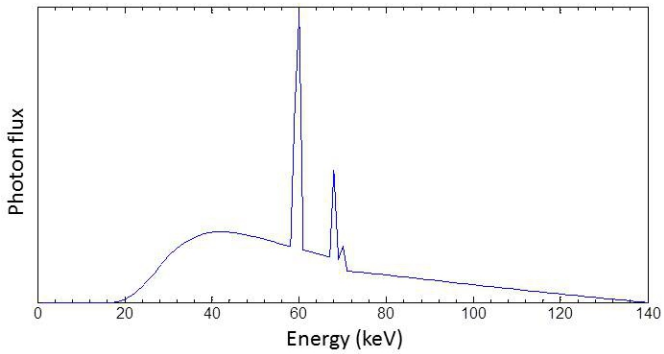


Figure 4.4: Simulated spectrum at a nominal tube voltage of 140kV used in polychromatic simulations. Generated using the Xcomp5 software [NH85].

$$s_i = F_0 \cdot C_c \cdot y_i \cdot m_i \tag{4.21}$$

where s_i is the scatter value at position i in the sinogram and i indicates both the angle and position (within the detector array) of the projection line; F_0 is the fraction of photons scattered forward (at an angle of 0°); C_c is the fraction of the attenuation resulting from Compton scatter (i.e. electron-photon interactions); y_i is the transmission simulation value at i and m_i is the log-converted sinogram:

$$m_i = \ln(b_i/y_i) \tag{4.22}$$

where b_i is the blank scan value at position i (the detected number of photons in sinogram pixel i in the absence of an absorber).

A fan-beam acquisition was simulated using (Table 4.2): 672 detectors and 1160 views per rotation (360°); a focus-to-isocentre distance of 570mm; a focus-to-detector distance of 1040mm; a Field of View (FoV) diameter of 50 cm and a detector angular aperture of 0.0741° (giving a fan-angle of approximately 52°). These geometric parameters approximate the Siemens SOMATOM Sensation 64 CT scanner. A simulated X-ray spectrum was generated using the Xcomp5r software [NH85] at a nominal tube voltage of 140kV (see Figure 4.4). After simulation, the fan-beam sinograms were rebinned to parallel-beam data using the methods described in Section 2.3.4. Filtered Back Projection (FBP) was used to create the reconstructed image with 512×512 pixels. The final, artefact corrupted simulation is illustrated in Figure 4.5 (c). A reference image (Figure 4.5 (a)) was generated using a monochromatic simulation at 70 keV (which approximates the mean energy of the polychromatic spectrum in Figure 4.4) - this image is used as a reference in the qualitative (visual comparison) analyses. To allow for quantitative

Parameter	Description
Beam geometry	Fan-beam
Focus-to-isocentre distance	570mm
Focus-to-detector distance	1040mm
FoV diameter	50 cm
Number of detectors	672
Detector angular aperture	0.0741°
Fan-angle	52°
Number of views/rotation	1160/360°

Table 4.2: Scanner geometry and reconstruction parameters used to generate simulated CT data using the simulator of [MND+99].

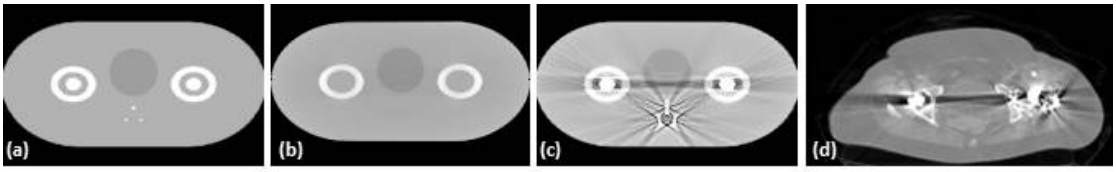


Figure 4.5: Simulated and clinical medical data. a) Monochromatic FBP reconstruction of software simulated phantom image (no artefacts). b) Polychromatic FBP reconstruction of metal-free phantom (used as a reference image in quantitative analysis). c) Polychromatic FBP reconstruction of phantom with metal inserts (test image). d) FBP reconstruction of patient scan with double hip prosthesis.

analysis, a second reference image was generated by an identical polychromatic simulation but without metal inserts (Figure 4.5 (b)).

4.3.5 Performance Evaluation

In order to evaluate the MAR methods on clinical data, a true CT scan of a patient with a double hip prosthesis is included. The spiral CT data was acquired on a Siemens Sensation 16 system as part of a Biograph16 PET/CT scanner (Siemens Medical Solutions, Knoxville, TN) at a nominal tube voltage of 120kVp, a Computed Tomography Dose Index (CTDI) of 11.9, using a collimation of 16 x 1.5mm and a reconstruction slice thickness of 3mm. Prior to reconstruction, the spiral data is rebinned to parallel beam data. The 512×512 FBP-reconstructed slice used in this study is shown in Figure 4.5 (d).

Furthermore, two baggage-CT scans obtained using the CT-80DR baggage scanner, have been used to provide insight into the performance of the predominantly medically-based MAR techniques when applied to novel CT applications (in this case aviation security). The scanning configuration employed has been discussed in Section 1.2. The fan-beam projection data was rebinned to parallel beam data prior to reconstruction. Individual 512×512 axial slices obtained from the volumes generated by the CT-80DR scanner (Figure 4.6 (c) and (d)) are used in the comparisons. The first of the two real-world security scans contains two

metallic objects (handguns) in a container with no background clutter (Figure 4.6 (a)). The second real-world scan is of a cluttered passenger bag containing multiple metallic objects (handgun, belt buckles, metallic zipper etc.) and a variety of commonly encountered objects of varying density (e.g. clothing, bottles etc.) (Figure 4.6 (b)). In the context of this comparative experimentation, all non-metallic objects in the scan are regarded as clutter.

In addition to standard qualitative evaluations (visual comparisons), quantitative performance analysis for the phantom study is performed by computing the Normalised Root Mean Squared Error (NRMSE) [BDM10, Her09] of the difference between the FBP-reconstructed image without metal inserts (Figure 4.5 (b)) and the MAR-corrected images:

$$\text{NRMSE} = \sqrt{\frac{\sum_{j=1}^{\tilde{N}}(f_j - f_j^{ref})^2}{\sum_{j=1}^{\tilde{N}}(f_j^{ref} - \mu)^2}} \tag{4.23}$$

where f_j is the corrected image; f_j^{ref} is the reference image; μ is the mean of all the reference image intensities and $\tilde{N}$ is the (reduced) number of pixels in the image (as the regions corresponding to the metal inserts are not considered). A large difference (between f_j and f_j^{ref}) in a few pixels results in a high NRMSE. An NRMSE value of 1 would correspond to a uniformly-dense corrected image with an intensity value equal to μ [Her09].

For the patient and baggage-CT data (where no ground-truth is available), quantitative performance analysis is performed using the reference-free ground-truth metric of Kratz *et al.* [KEMB11], which utilises the raw projection data outside of the metal trace as ground-truth data. The technique has been described in further detail in Section 3.2.5. The Normalised Reference-Free Errors (denoted as NRFE) are represented as factors of the unprocessed (FBP) error (a value of 1 would correspond to no improvement). In order to determine the veracity of the reference-free metric, the NRFE and the NRMSE for the phantom data are computed and compared.

While absolute computational times are presented, it is emphasised that little attention has been paid to optimisation in the implementation of each of the compared techniques. It is therefore acknowledged that the computational performance results may be misleading in some cases. The relationship between the error and processing time is quantified by computing a normalised product of the error and processing time for each of the methods. The value of this product falls in the range $[0, 1]$ with a value of 1 being the worst possible value (i.e. highest error and highest processing time).

The performance of the DDMar technique (Section 4.2) is further quantified

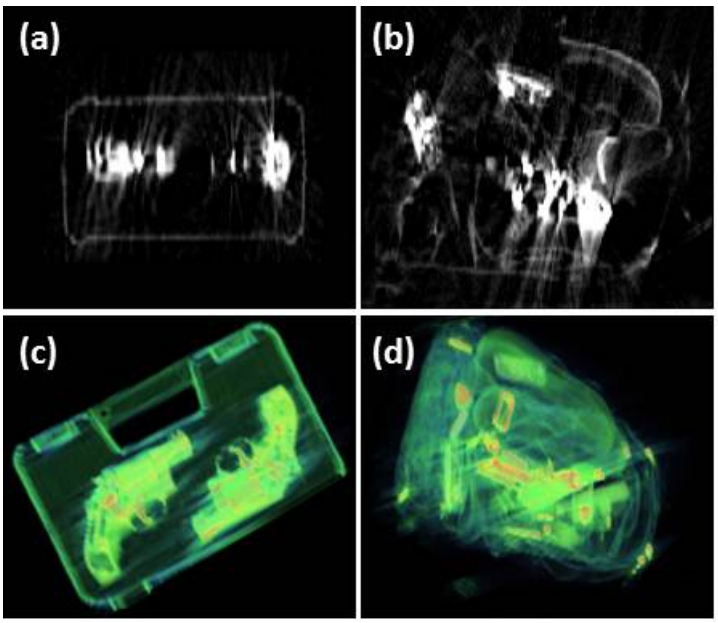


Figure 4.6: Real-world security-screening CT data. a) Axial slice of two handguns in clutter-free environment. b) Axial slice of multiple metal objects in cluttered environment. c) Volumetric rendering of clutter-free bag from which slice in (a) was obtained. d) Volumetric rendering of cluttered bag from which slice in (b) was obtained.

using Bland-Altman plots [AB83] (Section 3.2.5). The Bland-Altman plot considers the mean intensities in three types of image regions before and after MAR: 1) overestimated regions (pixel densities higher than expected due to artefacts); 2) underestimated regions (pixel densities lower than expected due to artefacts) and 3) unaffected regions (pixels unaffected by streaking). In each individual 2D axial CT slice obtained from the clinical data as well as selected axial slices obtained from the 12 volumetric baggage scans, 5 ROIs are manually specified (2 overestimated, 2 underestimated and 1 unaffected). Two measurements are made for every ROI: 1) the mean intensity of the ROI before MAR and 2) the mean intensity of the ROI after MAR. The Bland-Altman plot then plots the difference of the two ROI measurements as a function of their mean. Successful MAR should yield a decrease in the mean intensity of the overestimated regions, an increase in the underestimated regions and little/no change in the unaffected regions. The Bland-Altman plots are generated for the DDMar method as well as the global intensity-limiting approach presented in [MMFB12], which imposes an upper limit on the intensities of all the pixels in the corrected images (regardless of their locations relative to the metal objects). In the results presented in Section 4.4.2, the method of [MMFB12] is denoted *Mou*. This additional quantitative evaluation is included to demonstrate the impact on performance of the distance-driven refinement (compared to a global intensity limiting - see Section 4.2). Finally, a qualitative performance analysis of the DDMar technique applied to volumetric imagery (as opposed to individual slices) is presented.

4.4 Results

All the software for this study was developed according to the information available in the original publications. Where relevant, optimal parameters were determined empirically. All experiments were performed on an Intel Core i5 machine running a 2.30GHz processor with 6GB of RAM.

4.4.1 Denoising Results

Figure 4.7 shows the results of applying the denoising algorithms to a scan of a container containing a single handgun. The scalar opacity mappings and colours have been chosen to provide the clearest visualisation of the relevant features in each volume (i.e. noise and edges). Figure 4.7 (a) displays the original, unfiltered volume (with notable streaking artefacts emanating from the handgun) while images (b) - (g) display the results of each of the denoising techniques. The TV filter [Cha04] (Figure 4.7 (d)), wavelet shrinkage [CDAO95] (Figure 4.7 (e)) and NLM filtering [BCM05a, MS05a] (Figure 4.7 (f)) resulted in the most significant improvements in image quality. Although the streaking artefacts were considerably reduced for each of these methods, they were not removed entirely. Anisotropic diffusion [PM90] (Figure 4.7 (b)) and bilateral filtering [PKTD07] (Figure 4.7 (c)) resulted in less of an improvement in image quality, characterised by a noticeable blurring of the artefacts. The AWMSHE [ZPA10b] (Figure 4.7 (g)), yielded a virtually artefact-free image but additionally resulted in a noticeable loss of edge and contrast information (particularly evident in the outline of the container).

Figure 4.8 displays the denoising results for the same scan (in a different orientation), corrupted with Gaussian noise of standard deviation of 15. Figure 4.9 illustrates a single slice from each of the volumes in Figure 4.8. The NLM filter and wavelet-shrinkage techniques resulted in the greatest degree of noise reduction, with the majority of the noise removed and a clear image of the handgun and container remaining. The edge preservation in the NLM image was marginally superior to that of the wavelet-shrinkage image. Anisotropic diffusion, bilateral filtering, TV filtering and the AWMSHE technique removed comparatively low degrees of noise (particularly evident in the axial slices in Figure 4.9).

Figure 4.10 displays the results of the ACMC volume rendering algorithm [LCNC98] pre-denoising and post-denoising. Wavelet shrinkage and NLM filtering again yielded the most satisfactory results, with notable reductions in the spurious structures in the vicinity of the pliers. The anisotropic diffusion, bilateral filtering, TV filtering and AWMSHE techniques performed comparatively poorly and yielded renderings similar in quality to the unprocessed rendering in Figure 4.10 (a). To illustrate the effects of denoising on the rendering results more clearly, Figure 4.11 shows a magnified region of the rendered volumes before and

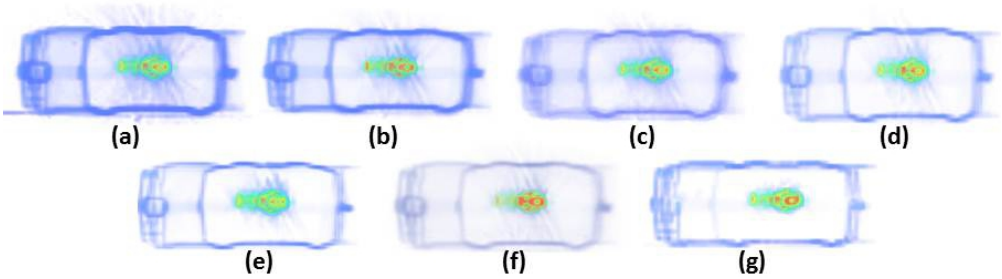


Figure 4.7: Volumetric renderings illustrating denoising results: (a) Original (b) Anisotropic diffusion (c) Bilateral filter (d) TV filter (e) Wavelet thresholding (f) NLM filter (g) AWMSHE

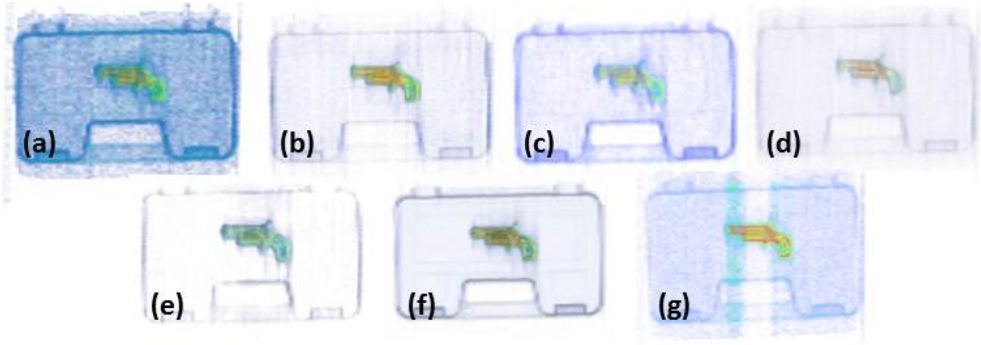


Figure 4.8: Volumetric renderings illustrating denoising results with added Gaussian noise (standard deviation = 15). (a) Original image with Gaussian noise, $\sigma = 15$ (b) Anisotropic diffusion (c) Bilateral filter (d) TV filter (e) Wavelet thresholding (f) NLM filter (g) AWMSHE

after applying the NLM filter, which produced the optimal visual results. Denoising resulted in a considerably cleaner result, as is indicated by the demarcated regions.

Table 4.3 and Figure 4.12 display the results of the 3D SIFT-based quantitative analysis. As mentioned, the SIFT-point-detection algorithm includes a refinement procedure whereby candidate SIFT points are rejected due to poor contrast and/or poor localisation on edges [FBM10]. These rejections are governed by two thresholds which were set according to the optimal values recommended by Flitton *et al.* [FBM10] (where the same data was used). The numbers of object and noise SIFT points were manually recorded across three scale-space levels. The results in Table 4.3 indicate that there was no significant variation in the number of object feature points detected for each of the volumes. For the unfiltered volume a total of 19 noise feature points was detected, yielding a ratio of 0.66. In every case, excluding TV filtering, denoising resulted in significantly fewer noise feature points and subsequently much higher ratios. Wavelet thresholding (indicated in bold in Table 4.3) yielded the optimal results with 0 noise feature points and thus a perfect ratio. The bilateral filter (2 noise feature points and ratio = 0.94) and NLM filter (1 noise feature point and ratio = 0.97) also returned significant improvements. The TV filter resulted in a deterioration in image quality represented by a reduction in

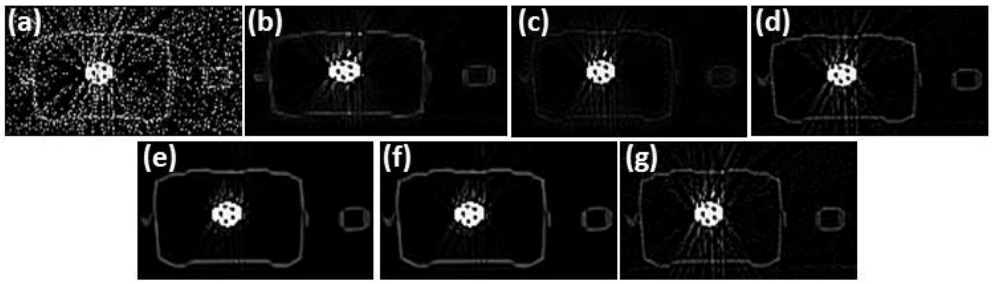


Figure 4.9: Denoising results for a single axial slice with Gaussian noise corruption. (a) Original image with Gaussian noise, $\sigma = 15$ (b) Anisotropic diffusion (c) Bilateral filter (d) TV filter (e) Wavelet thresholding (f) NLM filter (g) AWMSHE

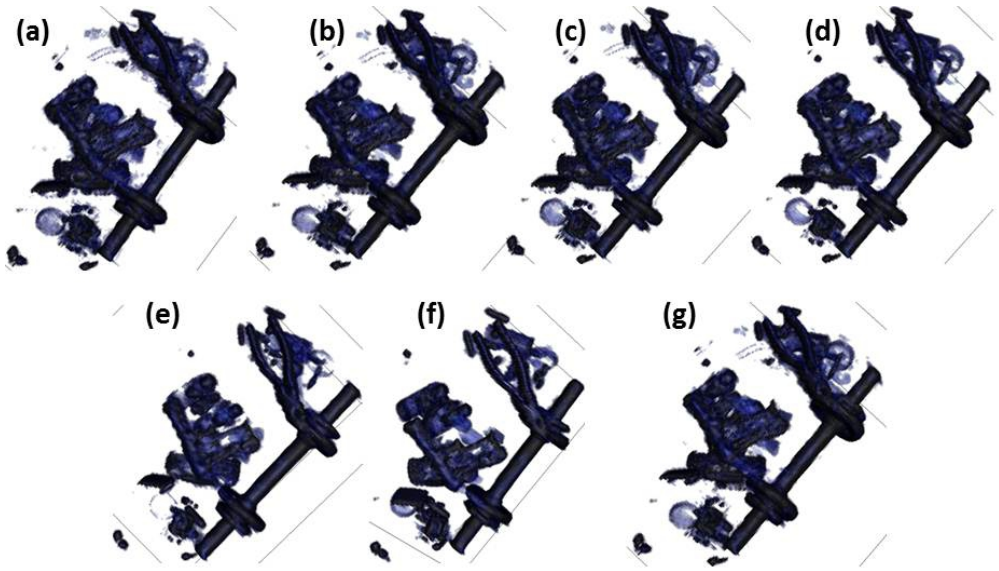


Figure 4.10: Denoised volume visualisations using ACMC volume rendering technique: (a) Original (b) Anisotropic diffusion (c) Bilateral filter (d) TV filter (e) Wavelet thresholding (f) NLM filter (g) AWMSHE

the SIFT ratio relative to the unprocessed volume (0.64 vs. 0.66). Anisotropic diffusion (0.74) and AWMSHE (0.76) yielded only minor improvements in the SIFT ratio and resulted in a significant decrease in the number of object feature points (27 and 26 respectively).

For illustrative purposes, the SIFT interest-point locations at the first scale-space level on the volumes before and after applying each of the denoising techniques are shown in Figure 4.12. These images illustrate keypoint locations at the first scale-space level only and so the numbers of object and noise feature points do not correspond directly with those in Table 4.3, which represent the numbers of keypoints across all three scale-space levels. With the exception of TV filtering (Figure 4.12 (d)), each of the denoising techniques resulted in a clear reduction in the number noise feature points (black dots).

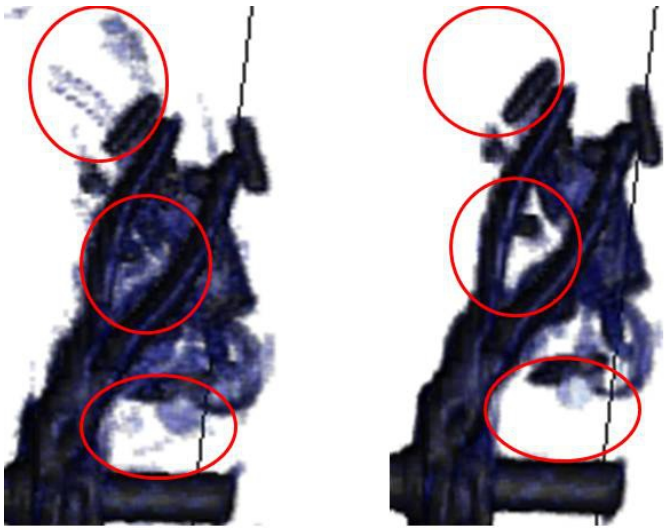


Figure 4.11: Magnified region of ACMC rendered volumes from Figure 4.10 for NLM filtering with regions of interest marked: (a) Original volume (b) NLM filtered volume

Method	Object points	Noise points	Ratio
Unfiltered	37	19	0.66
Anisotropic	27	8	0.77
Bilateral	33	2	0.94
TV	36	20	0.64
Wavelets	**35**	**0**	**1.00**
NLM	33	1	0.97
AWMSHE	26	8	0.76

Table 4.3: Quantitative analysis results using the 3D SIFT-based performance measure. Optimal performing method indicated in bold.

4.4.2 Metal Artefact Reduction Results

The results for the medical-CT data (simulated and patient) as well as the real-world security-screening CT data are presented below.

4.4.2.1 Medical-CT Results

Tables 4.4 and 4.5 and Figures 4.13 - 4.16 show the results of the phantom data experiments. Both the quantitative and qualitative results are, for the most part, in agreement with the observations made in the literature. The iterative approach of Wang *et al.* [WSOV96] yielded the lowest error (NRMSE = 0.174) and produced an image with a substantial reduction in artefacts (compared to FBP) and good preservation of edge and contrast information. The interpolation-based approaches (especially the Kalender [KHE87], Meyer1 (NMAR) [MRL+10], Meyer2 (ANMAR)

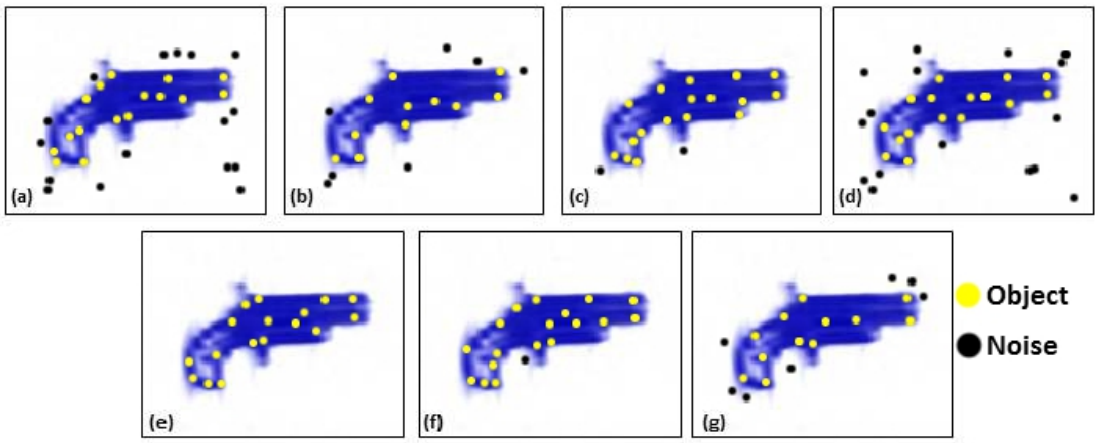


Figure 4.12: SIFT point locations at the first scale space level: (a) Original (b) Anisotropic diffusion (c) Bilateral filter (d) TV filter (e) Wavelet thresholding (f) NLM filter (g) AWMSHE

[MRS+11], and to a lesser degree, DDMar (Section 4.2)) yielded images which, at first glance, appear smooth and apparently free of artefacts. Upon closer inspection however (Figure 4.15), the loss and/or deformation of image detail is evident. This is reflected in the higher errors for these methods when compared to the iterative approach. The Meyer3 (FSMAR) [MRL+12] approach resulted in superior preservation of edge and contrast information (compared to the Kalender, Meyer1 and Meyer2 images) but also reintroduced much of the original streaking from the FBP image. Of the sinogram-completion-based approaches, the Zhao [ZBWW02], Abdoli [AAA+10] and DDMar approaches yielded the most satisfactory images in terms of artefact reduction and edge and contrast preservation. While the Zhao (0.189), Bal (0.398), Yu (0.242), Abdoli (0.192), Meyer1 (0.243), Meyer2 (0.239), Meyer3 (0.336) and DDMar (0.191) approaches all produced significant improvements in the NRMSE, compared to the FBP (0.872) image, the Jeong (0.726) and Li (0.478) approaches performed notably poorer than these methods - both in terms of error and visual quality. In fact,both of these approaches yielded higher errors than the original linear-interpolation-based approach of Kalender (0.401) [KHE87].

To emphasise the difference in image quality, the image regions surrounding and connecting the metallic objects have been magnified and are shown in Figure 4.15. The Wang, Zhao, Abdoli and DDMar images show significant reductions in streaking, while the Meyer1, Meyer2, Bal and Kalender images are notably blurred. The Jeong image, although showing a reduction in the original streaks, contains significant amounts of secondary streaking, especially in the regions connecting the metal objects.

The NRFE and NRMSE for the phantom data are compared in Table 4.5 and Figure 4.16. To allow for a direct comparison, the errors are represented as a percentage of the unprocessed (FBP) error. While smaller variations in the

Method	NRMSE	Time (seconds)	Normalised Product
FBP	0.872	0.18	0.002
Kalender	0.401	1.62	0.006
Wang	0.174	123.30	0.200
Zhao	0.189	11.65	0.021
Bal	0.398	48.95	0.181
Yu	0.242	9.88	0.022
Jeong	0.726	16.59	0.112
Li	0.478	54.72	0.243
Abdoli	0.192	6.42	0.012
Meyer1	0.243	6.67	0.015
Meyer2	0.239	6.61	0.015
Meyer3	0.336	14.81	0.046
DDMar	0.191	10.48	0.019

Table 4.4: MAR quantitative analysis results for simulated scan (Figure 4.5 (c)).

reference-free metric are observed between the different methods, the two metrics resulted in a very similar performance ranking.

Table 4.6 and Figure 4.17 and 4.18 show the results of the patient experiments. Quantitative error analysis was performed using the NRFE metric of Kratz *et al.* [KEMB11]. For the most part, the results are similar to the phantom data experiments and are again largely in agreement with the observations made in the literature. The Wang (0.172), DDMar (0.180) and Zhao (0.196) approaches yielded the lowest errors, while the Jeong approach (0.842) was again the worst-performing method, performing considerably worse than the linear interpolation approach of Kalender (0.369) and only marginally better than standard FBP reconstruction (1.0) (Table 4.6). The remaining approaches all resulted in significant reductions in the NRFE (compared to the FBP reconstruction). With reference to the qualitative results in Figure 4.18, the interpolation-based approaches (Kalender, Meyer1, Meyer2, Bal and DDMar) produced smooth images, with apparently few artefacts but a notable loss in edge and contrast information. Although the Meyer3 approach better preserved edge and contrast information, a greater degree of the original streaking artefacts remained. The improvement in image quality yielded by the Jeong approach is limited and despite noticeable secondary streaking (see Figure 4.19), the approach still yielded a reasonable reduction in the NRFE (Table 4.6) as the original streaking was reduced. This highlights the importance of considering both quantitative and qualitative results in performance evaluation.

As expected, the iterative reconstruction approach (Wang) was the most computationally intensive in both medical experiments (Tables 4.4 and 4.6). Although all the techniques yielded higher processing times compared to the Kalender approach, the majority of these times are still considerably lower than the Wang

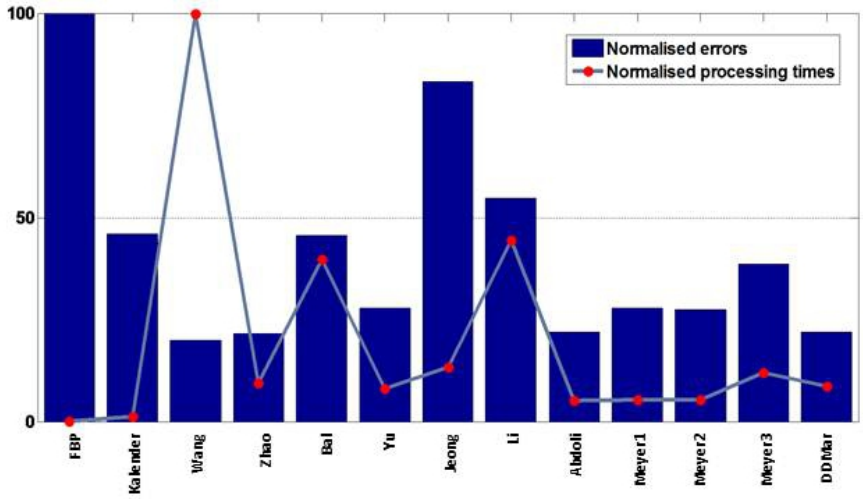


Figure 4.13: Graphical comparison of MAR quantitative analysis results for phantom scan (Figure 4.5 (c)). The plot illustrates the tradeoff between processing times and error.

approach and within a manageable range. Of the sinogram-completion-based approaches, the Bal [BS06] and Li [LBY+10] approaches yielded times which were considerably higher than the other methods. The significant computational expense of the Bal approach can most likely be attributed to the adaptive prefiltering stage of the algorithm. The plots in Figures 4.13 and 4.17 indicate that performance (in terms of error) is not necessarily correlated with computational expense. The Li approach, for example, yielded the highest processing time, but also the second highest error. To quantify this relationship between the error and processing time, the products of the normalised errors and processing times for each of the methods are shown in the third column of Tables 4.4 and 4.17. Judging performance based on these products alone, emphasises the drawback of the high computational cost associated with iterative reconstruction (Wang approach).

Figure 4.19 shows magnified regions in the patient images. The reduction in streaking achieved by the best-performing techniques (Wang, DDMar and Zhao) is perhaps less than that observed in the phantom experiments. In fact, the Wang and DDMar images appear to contain greater degrees of streaking than the Zhao image, despite lower errors. The Zhao image shows some blurring of the metal edges and regions surrounding the metal objects. The Jeong image shows little, if any, reduction in the original streaks and again contains significant amounts of secondary streaking. Similarly to the phantom experiments, the sinogram-completion-based approaches resulted in image blurring.

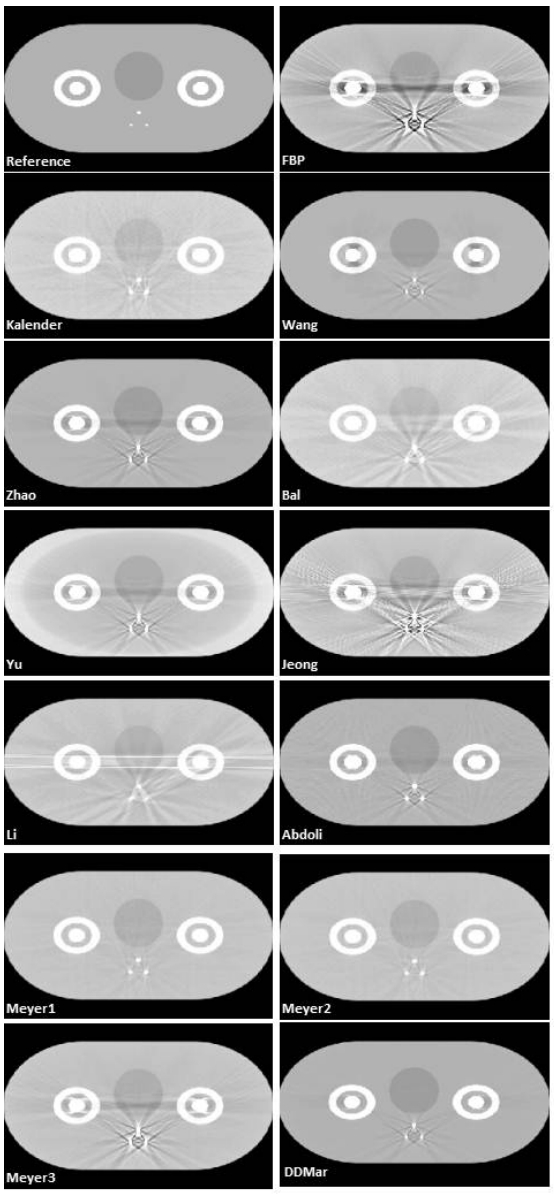


Figure 4.14: MAR results for phantom scan in Figure 4.5 (c). (Window = 800 HU, Centre = 0 HU)

4.4.2.2 Security-Screening Results

The quantitative analysis results for the security-screening CT scans are shown in Tables 4.7 and 4.8 and Figures 4.20 and 4.21. For those MAR techniques that utilise priors, the optimal input parameters were determined empirically, based on preliminary experimentation using a large set of 2D slices (80 slices) randomly selected from 20 real-world volumetric baggage scans with varying contents.

The results follow a similar trend for the two scenarios (clutter-free and cluttered). While all the methods yielded some reduction in error, the Wang (errors: 0.089 for clutter-free, 0.189% for cluttered), Zhao (0.087 and 0.231), DDMar (0.087 and 0.227) and Abdoli (0.151 and 0.292) approaches showed the most significant improvements. Notably, the Zhao and DDMar approaches outperformed (albeit marginally) the Wang approach for the clutter-free scenario. As was the case in

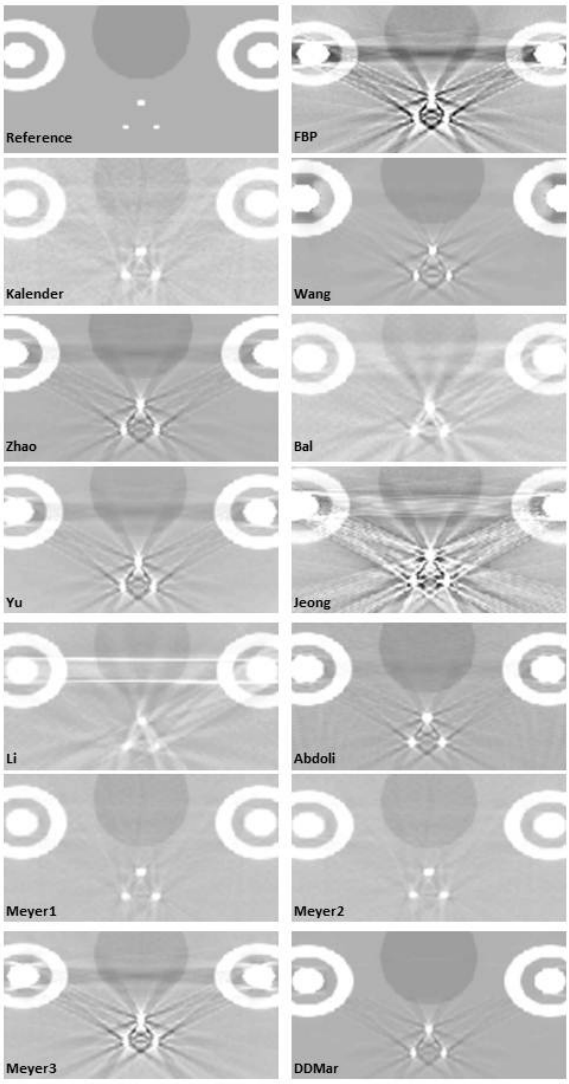


Figure 4.15: Magnification of phantom images in Figure 4.14 illustrating the performance each technique in the regions in and around the metallic inserts (Window = 800 HU, Centre = 0 HU)

the phantom experiments, the Jeong (0.420 and 0.890) and Li (0.353 and 0.653) approaches were the two worst-performing methods - yielding significantly higher errors than the Kalender approach (0.201 and 0.447). Perhaps the most interesting observation from these results is that several of the MAR methods that are considered state-of-the-art in the medical domain yielded minimal performance gains over the simple linear-interpolation-based approach of Kalender. This is particularly evident for the methods presented by Meyer *et al.* [MRL+10, MRS+11, MRL+12] (Meyer1 (0.243 and 0.444), Meyer2 (0.224 and 0.440) and Meyer3 (0.348 and 0.428)) and Bal and Spies [BS06] (0.197 and 0.336) which use multiclass segmentation to generate priors. The small performance gains are emphasised further by the normalised products in the third column of Tables 4.7 and 4.8, which represent a combined performance measure, assuming the error metrics and computational times to be of equal performance. At the very least, judging from these error metrics alone, the performance gains of the more complex methods over the simple

Method	NRMSE (% original error)	NRFE (% original error)
FBP	100	100
Kalender	45.99	26.31
Wang	19.95	11.34
Zhao	21.67	12.90
Bal	45.64	27.21
Yu	27.75	26.40
Jeong	83.26	42.13
Li	54.82	36.69
Abdoli	22.02	22.34
Meyer1	27.87	23.42
Meyer2	27.41	22.39
Meyer3	38.53	26.33
DDMar	25.34	17.68

Table 4.5: Comparison of NRMSE and NRFE [KEMB11] for phantom image in Figure 4.5. The correlation between the two measures verifies the feasibility of the reference-free NRFE metric. Errors represented as percentage of FBP error.

linear interpolation approach were not as significant as indicated in the medical literature.

The images in Figures 4.22 to 4.25 show the qualitative results for the real-world scans. While the Wang, Zhao and DDMar approaches yielded significant reductions in streaking for both the clutter-free (Figure 4.22) and cluttered (Figure 4.23) scenarios, all of the sinogram-completion-based approaches led to a loss of and/or distortion in image details. This is especially evident in the cluttered scenario where, despite the notable reductions in streaking, the Zhao, Bal, Abdoli, Meyer1, Meyer2 and DDMar images are characterised by a noticeable blurring of the low-density regions of the image (Figure 4.23). In both the clutter-free and cluttered scenarios, the Jeong approach produced images characterised by intense secondary streaking with little, if any, improvement in image quality. As expected, performance on a whole was poorer in the cluttered scenario (Figure 4.23), where even the best-performing methods (Wang, Zhao and DDMar) contained noticeable streaking and/or loss of image detail in the non-metal regions.

Figure 4.24 and Figure 4.25 show magnifications of the clutter-free and cluttered scenarios respectively. For the clutter-free case (Figure 4.24), although the best-performing methods in terms of error (Zhao and DDMar) produced considerably cleaner images, a noticeable loss of edge and contrast information is again evident (in the form of blurring). The Jeong image is again heavily corrupted by secondary streaking. For the cluttered case (Figure 4.25), the Wang and Zhao images again exhibit a significant reduction in streaking relative to the FBP image, but image detail is compromised (e.g. the outline of the bag and low-density objects in the direct vicinity of the metal objects). This is especially noticeable in

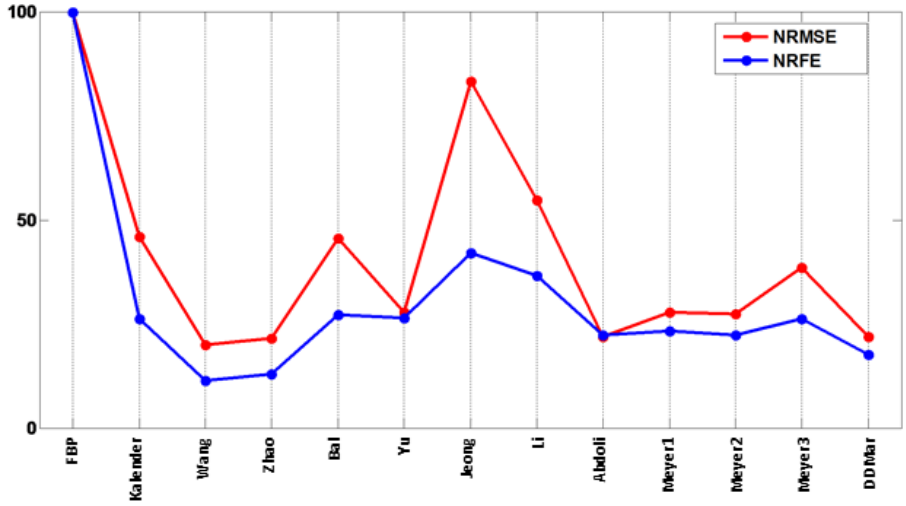


Figure 4.16: Graphical comparison of NRMSE and NRFE [KEMB11] for phantom image in Figure 4.5. As desired, the NRFE follows a similar trend to the NRMSE. Errors represented as % of FBP error.

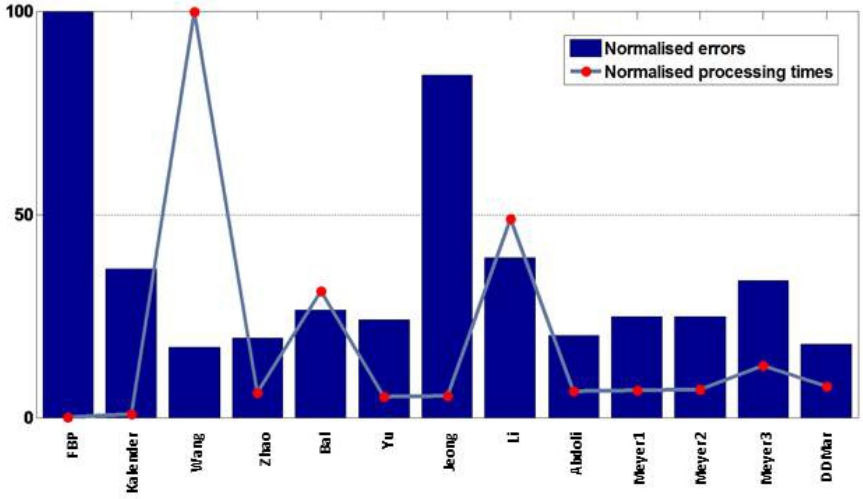


Figure 4.17: Graphical comparison of MAR quantitative analysis results for patient scan (Figure 4.5 (d)). The plot illustrates the tradeoff between processing times and error.

the Zhao image, where the MAR process has resulted in a blurring of the regions surrounding the metal objects. Again, the Jeong image shows little, if any, improvement in image quality. While the differences in the errors for the Wang, Zhao and DDMar approaches are negligible, the Wang images appear to be of a slightly superior visual quality. For the most part, however, the degree of improvement in image quality for all the methods does seem to correlate fairly well with the error analyses in Tables 4.7 and 4.8, thereby further verifying the authenticity of the reference-free error metric.

Judging from the shapes of the curves in Figure 4.20 and 4.21 as well as the readings in Tables 4.7 and 4.8, the processing times in both real-world experiments followed a very similar pattern to that seen in the phantom experiment (Table 4.4 and Figure 4.13).

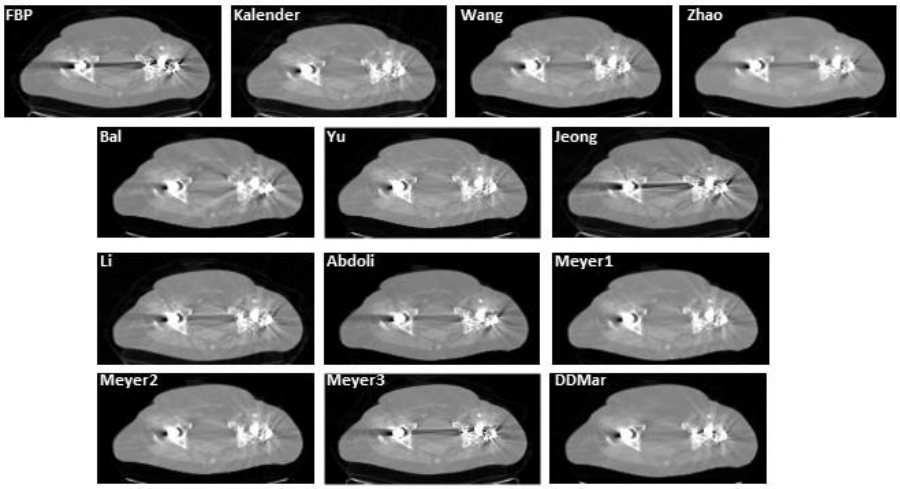


Figure 4.18: MAR results for patient scan in Figure 4.5 (d). (Window = 500 HU, Centre = 0 HU)

Method	NRFE	Time (seconds)	Normalised Product
FBP	1.00	0.21	0.001
Kalender	0.369	1.32	0.003
Wang	0.172	149.80	0.172
Zhao	0.196	9.22	0.012
Bal	0.264	46.81	0.083
Yu	0.242	7.77	0.013
Jeong	0.842	8.00	0.045
Li	0.393	73.31	0.192
Abdoli	0.202	9.83	0.013
Meyer1	0.250	10.10	0.017
Meyer2	0.249	10.39	0.017
Meyer3	0.337	19.07	0.043
DDMar	0.180	11.65	0.012

Table 4.6: MAR quantitative analysis results for patient scan (Figure 4.5 (d)). Errors computed using the NRFE metric of [KEMB11].

Figures 4.26 and 4.27 show the Bland-Altman plots generated from the uncorrected and corrected CT data for the patient and baggage data sets respectively. The overestimated (OE), underestimated (UE), and unaffected (UA) regions in each data set are represented using different coloured markers. The plots are included to investigate the necessity of the distance-weighted intensity limiting employed by the DDMar approach (versus a global intensity limit - refer to Section 4.2). The graphs for both experiments confirm the desired modifications to the image intensities in the overestimated (yellow markers) and unaffected regions (red markers) after MAR. Since the DDMar approach only modifies the behaviour of the Mou approach [MMFB12] in the regions connecting metal objects, the two approaches give the same results for the overestimated and unaffected regions. As desired, for both the patient (Figure 4.26) and baggage data (Figure 4.27), the

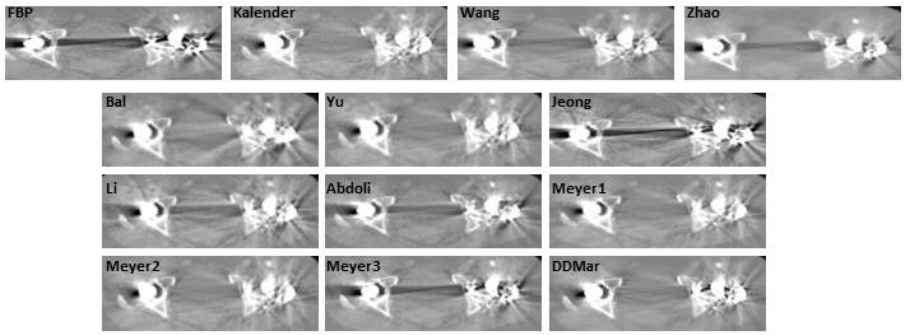


Figure 4.19: Magnification of patient images in Figure 4.18 illustrating the performance each technique in the regions in and around the metallic inserts (Window = 500 HU, Centre = 0 HU).

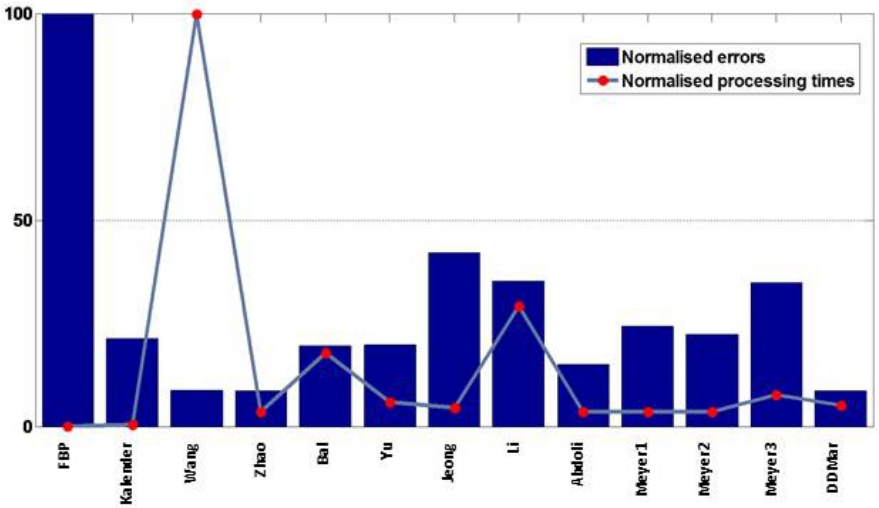


Figure 4.20: Graphical comparison of MAR quantitative analysis results for clutter-free baggage scan (Figure 4.6 (a)). The plot illustrates the tradeoff between processing times and error.

readings in the overestimated regions are consistently in the positive portions of the graphs (along the vertical axes), indicating a reduction in the overestimated intensities after MAR. Furthermore, for both data sets, the readings in the unaffected regions are closely clustered around the horizontal axes, indicating little to no change in the image intensities after MAR. The green and blue markers display the behaviour of the Mou and DDMar approaches respectively in the underestimated regions and clearly demonstrate the improvement achieved by the proposed modifications to the Mou approach. While the Mou approach yielded readings clustered around the horizontal axes (indicating little/no change in the intensities after MAR), the DDMar readings in both experiments consistently occurred in the negative vertical portions of the graphs - indicating a successful increase in the underestimated intensities.

Finally, Figures 4.28 and 4.29 show the volumetric artefact reduction results of the DDMar technique applied to two volumes. Similarly to the 2D experiments,

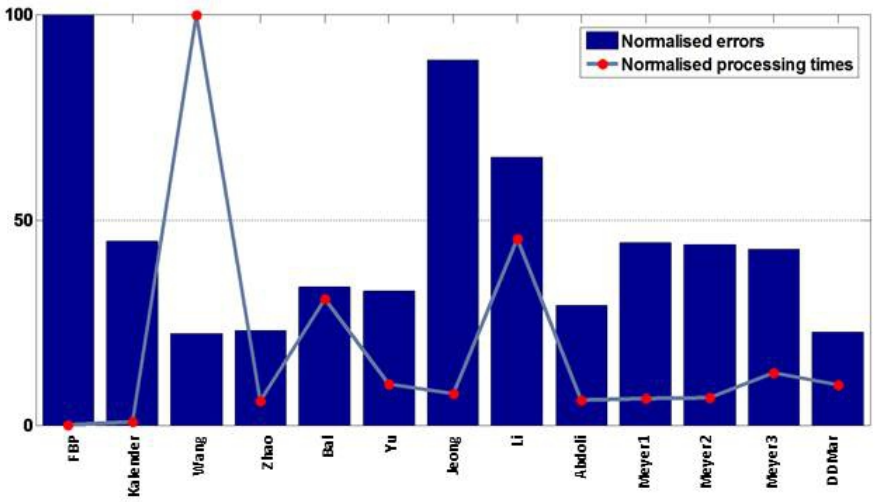


Figure 4.21: Graphical comparison of MAR quantitative analysis results for cluttered baggage scan (Figure 4.6 (b)). The plot illustrates the tradeoff between processing times and error.

one of the test images contained two handguns with no background clutter (Figure 4.28), while the other contained multiple metal/high-density objects in a cluttered environment (Figure 4.29). Each input and artefact-reduced volume is shown in three different orientations to better illustrate the presence and subsequent reduction of artefacts. In both the clutter-free (Figure 4.28) and cluttered (Figure 4.29) volumes the application of DDMar resulted in significant reduction in streaking artefacts, particularly evident in the direct surroundings of the high-density objects (white arrows in Figures 4.28 and 4.29). In addition to the removal of streaking artefacts, however, an obvious blurring is evident in both volumes (indicated in yellow).

4.5 Discussion

Several observations related to the aforementioned denoising and metal artefact reduction results are worth noting and elaborating.

4.5.1 Denoising

With reference to the images in Figure 4.7, even for the two best-performing techniques (wavelet shrinkage and NLM filtering), a notable degree of artefacts remain. The obvious limitation of denoising is that there exists a tradeoff between the quantity of noise/artefacts removed and the fidelity of the edge and contrast information in the image. That is to say, a greater degree of filtering is likely to remove more of the streaking while simultaneously compromising valuable image information. This scenario is particularly evident in the AWMSHE [ZPA10b] image (Figure 4.7 (g)): although virtually all of the streaking has been removed,

Method	NRFE	Time (seconds)	Normalised Product
FBP	1.00	0.13	0.001
Kalender	0.201	0.73	0.001
Wang	0.089	143.94	0.089
Zhao	0.087	5.10	0.003
Bal	0.197	25.81	0.035
Yu	0.198	8.55	0.012
Jeong	0.420	6.56	0.019
Li	0.353	42.14	0.103
Abdoli	0.151	5.33	0.006
Meyer1	0.243	5.24	0.009
Meyer2	0.224	5.23	0.008
Meyer3	0.348	11.18	0.027
DDMar	0.087	7.49	0.005

Table 4.7: MAR quantitative analysis results for clutter-free baggage scan in Figure 4.6 (a). Errors computed using NRFE.

the edges of the container have been almost entirely eliminated. The AWMSHE approach relies on the assumption that the noise in an image is characterised by significantly lower grey values (intensities) than the 'valuable' image regions. Threshold-based denoising approaches (e.g. AWMSHE) are ill-suited to removal of streak-like artefacts (which are by definition characterised by high-density lines) and high-frequency noise as the required threshold is likely to additionally eliminate edge information. The aforementioned observations indicate that denoising alone is not sufficient for the reduction of metal-streaking artefacts.

The variations in performance of the six denoising techniques are more pronounced in Figures 4.8 and 4.9 where synthetic Gaussian noise ($\sigma = 15$) has been added to the input image. While this degree of noise corruption is of course unlikely in reality, it illustrates the efficacy of the denoising algorithms well. The limitations of simple denoising in terms of artefact removal are also illustrated clearer here: while the background noise is reduced, considerable streaking artefacts remain - even for the two most effective methods (NLM filtering and wavelet shrinkage). Interestingly, despite the fact that the AWMSHE approach is the only dedicated baggage-CT-denoising technique, it results in very little (if any) improvement in image quality (Figure 4.8 (g) and Figure 4.9 (g)) and performs notably worse than the five standard denoising techniques.

Several similar trends are revealed in the quantitative performance analysis of the denoising techniques. The AWMSHE is again outperformed by the majority of the standard denoising techniques. It is worth noting that the AWMSHE technique was initially developed using images which were not representative of those encountered in practical security-screening settings and contained comparatively low levels of artefacts and noise [ZPA10b]. The results presented here indicate that

Method	Normalised Error	Time (seconds)	Normalised Product
FBP	1.00	0.23	0.001
Kalender	0.447	1.46	0.004
Wang	0.224	167.80	0.224
Zhao	0.231	10.12	0.014
Bal	0.336	51.68	0.104
Yu	0.328	16.74	0.033
Jeong	0.890	13.01	0.069
Li	0.653	76.12	0.296
Abdoli	0.292	10.33	0.018
Meyer1	0.444	11.04	0.029
Meyer2	0.440	11.23	0.029
Meyer3	0.428	21.38	0.055
DDMar	0.227	16.45	0.022

Table 4.8: MAR quantitative analysis results for cluttered baggage scan Figure 4.6 (b). Errors computed using the NRFE.

the technique is not well-suited to environments characterised by high degrees of artefacts and noise.

The improvements in visual quality and the improvements in the 3D SIFT interest-point ratios are indications that standard denoising techniques (particularly NLM filtering [BCM05a, MS05a] and translation-invariant wavelet shrinkage [CDAO95]) will benefit the implementation of subsequent operations such as the object-classification techniques presented in [FBM10, FBM12] and the volume-rendering techniques presented in [LCNC98].

4.5.2 Metal Artefact Reduction

While the interpolation-based approaches (especially the Kalender [KHE87], Meyer1 [MRL+10], Meyer2 [MRS+11], Bal [BS06] and DDMar (Section 4.2)) successfully remove streaking, a common shortcoming is the loss of and/or distortion in edge and contrast information. The Meyer3 approach (FSMAR) [MRL+12] compensates for this by utilising high-frequency information from the initial FBP reconstruction. While this results in better preservation of image details, an unfortunate byproduct is that much of the original streaking is reintroduced into the corrected image. Consequently, the Meyer3 images in both the phantom and real-world studies present with greater degrees of streaking than the comparative methods.

The results of the four experiments reveal several trends. The Jeong [JR09] and Li [LBY+10] techniques consistently produce the poorest results - quantitatively as well as qualitatively. Coupled with their high computational costs, these two techniques appear to be of little comparative value to the other approaches considered in this study. The iterative reconstruction approach (Wang [WSOV96]),

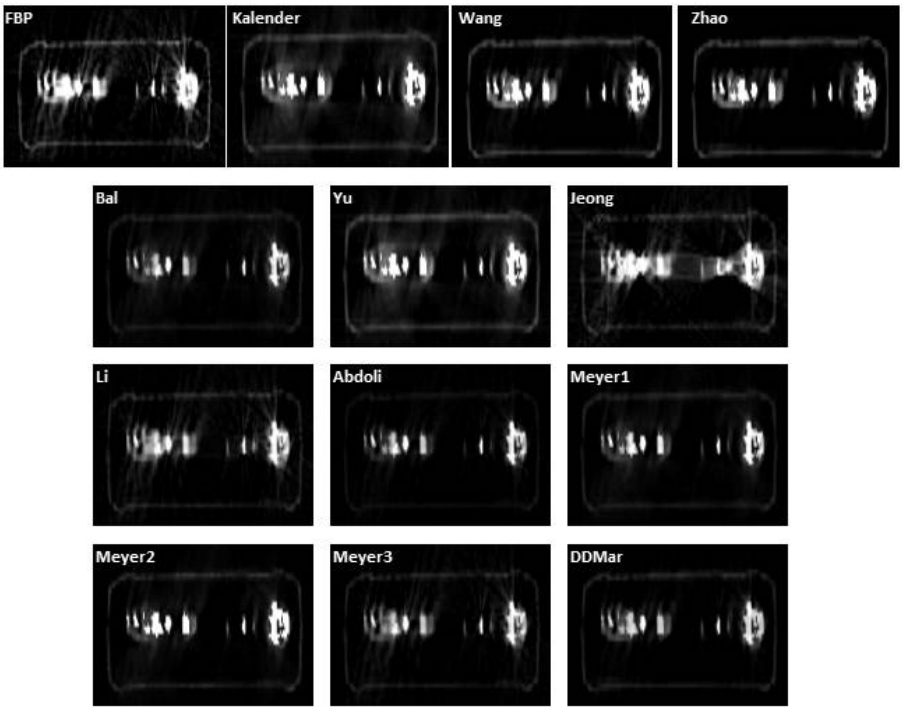


Figure 4.22: MAR results for clutter-free baggage scan in Figure 4.6 (a).

despite not employing a complex prior model, consistently yields low errors in all four experiments but is also consistently the most computationally intensive. It is likely that incorporating more sophisticated modelling processes into the iterative approach (e.g. [OB07, MND+01, SN12]) will lead to further reductions in the errors. The Zhao [ZBWW02], Abdoli [AAA+10] and DDMar approaches also yield significant quantitative and qualitative improvements across all four experiments - with significant and noticeable reductions in streaking and relatively good preservation of details. Of these, the Zhao and DDMar approaches yield the lowest errors and in some cases perform comparably to the iterative approach of Wang [WSOV96]. The processing times of the majority of the sinogram-completion-based approaches remain manageable and considerably lower than the iterative approach. It is worth emphasising, however, that little attention was paid to computational optimisation of the algorithms.

Perhaps the most interesting observation is that the performance gains of the state-of-the-art methods over the simple linear interpolation approach of Kalender *et al.* [KHE87] are not as significant in the security-screening domain compared to the medical domain. This observation is supported further by quantifying the relationship between error and computational cost. When considering performance based on this measure alone, simple linear interpolation comfortably produces the best results in all four experiments. Performing a direct scalar multiplication of the two values does, however, assume that the errors and processing times are of equal importance in terms of performance. In reality the relative importance of the two measures is dependent on the application. In security-screening appli-

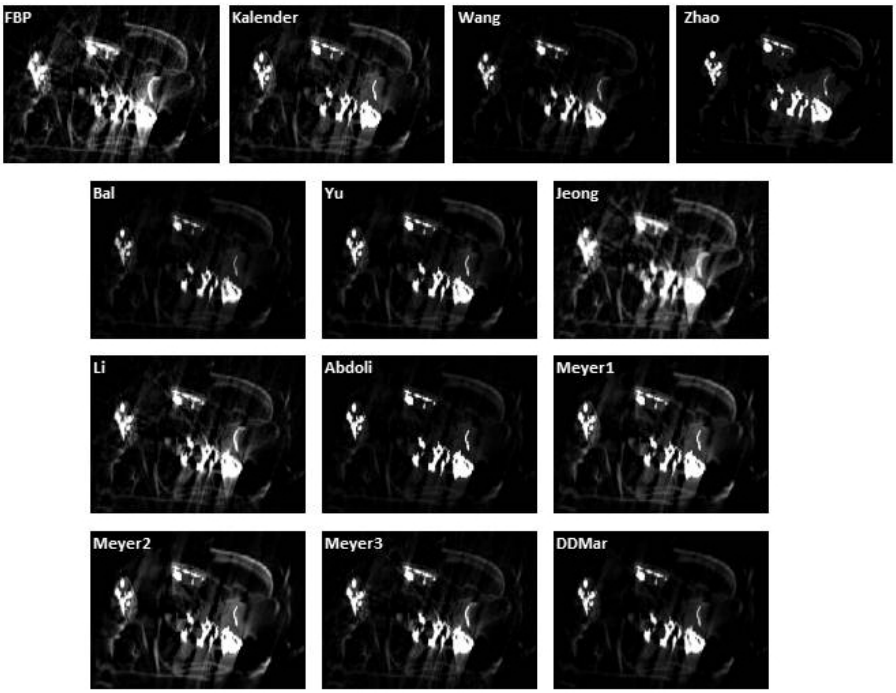


Figure 4.23: MAR results for cluttered baggage scan in Figure 4.6 (b)

cations processing times are of greater relative importance compared to medical applications, where accurate diagnoses and hence image quality, are of the highest importance. This highlights the potential benefit of establishing a suitable tradeoff between the degree of artefact reduction and computational cost and performing a weighted multiplication accordingly. Naturally, this tradeoff would be application dependent.

The fact that the more complex methods perform only marginally better, or worse in some instances (in terms of error) than linear interpolation in the security-CT experiments can be attributed to the fact that many of the state-of-the-art methods employ priors to guide the sinogram-correction process [ZBWW02, BS06, MRL+10, MRS+11, MRL+12]. While medical-CT scans are consistent enough in their appearance to allow for priors to be reliably generated based on known tissue characteristics, the variability and unpredictability in the contents of baggage-CT data, makes the generation of such priors more challenging. These observations indicate that a poor choice of prior may lead to poorer results than not using any prior information at all (i.e. the Kalender approach). This is not entirely surprising and is in concurrence with the conclusions of Karimi *et al.* [KCWM12] that an inaccurate prior has a significant detrimental effect on the performance of a MAR algorithm. Fine tuning of the parameters that influence the computation of the priors would most likely result in improved performances and values that better support the results obtained on medical images. Such tuning, however, can be a laborious, empirical process and would detract from the efficiency of the method. An exhaustive optimisation of the parameter space is left as an area for future

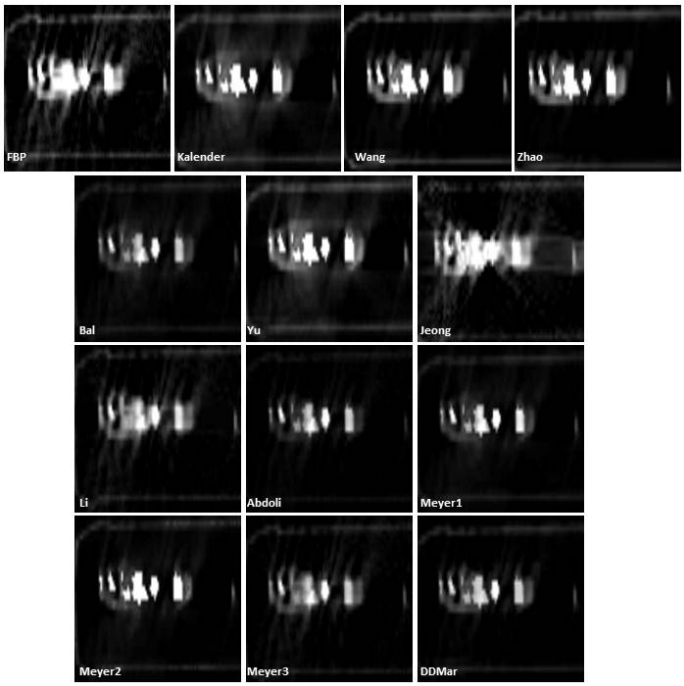


Figure 4.24: Magnification of MAR results for clutter-free baggage scan in Figure 4.22.

work. While these results unfortunately indicate that the MAR techniques that employ prior information are less well-suited to settings where the generation of such priors is difficult (e.g. the security domain), they do not detract from the claims made in the original publications regarding the success of the methods in the medical domain.

As expected, the introduction of clutter complicates the MAR process further, especially in terms of generating accurate priors. This is reflected in the universal decrease in the overall performance of all of the methods. While streaking artefacts are reduced to some degree in most cases, the overall improvements in visual quality of the images are in most cases minimal. In several cases, the negative impact of the secondary artefacts and the corruption of important image details (especially in the vicinity of metal objects) outweighs the positive impact of the reduction of the original streaks.

The similar performance rankings produced in the phantom experiments by the reference-free error metric (NRFE) [KEMB11] and the NRMSE, as well as the correlation between image quality and error in the real-world studies, gives credence to the authenticity of the reference-free metric. Despite the fact that a reduction in error generally corresponds to some improvement in image quality, it is important to emphasise that considering the results of either performance measure (quantitative or qualitative) alone, is likely to be misleading. Even when reference data is available, reliable quantitative performance analysis is challenging and at the very least needs to be presented in conjunction with qualitative results

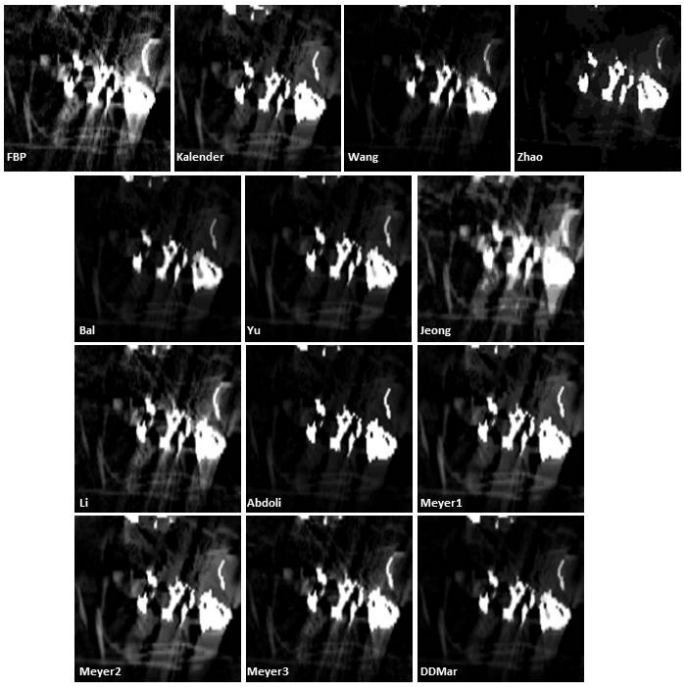


Figure 4.25: Magnification of MAR results for cluttered baggage scan in Figure 4.23.

- this is a factor which has been neglected to a large degree in the literature and has possibly contributed to exaggerated performance claims.

The additional Bland-Altman performance analysis of the DDMar and Mou approaches has highlighted the necessity of the distance-based weighting of the image refinement procedure (Section 4.2). As predicted, the Mou approach is shown to perform poorly in image regions corrupted by dark bands (i.e. characterised by underestimated intensity values). This phenomenon typically occurs in the regions connecting the multiple metal objects and is appropriately handled by the distance-weighting scheme. Nonetheless, the volumetric performance analysis of the DDMar approach reveals that, although a notable degree of artefacts are removed, image blurring (particularly within the neighbourhoods of metal objects) is an obvious concern. The impact of such blurring on further processing (e.g. classification and/or segmentation) is addressed in Chapters 5 to 8 of this work.

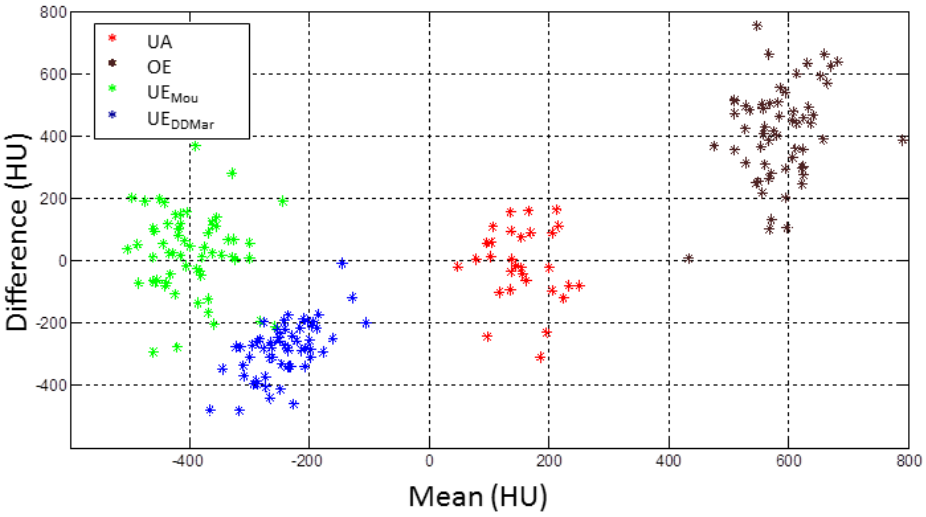


Figure 4.26: Bland-Altman plot for patient data. The plot illustrates the change in intensity values for different regions in the scan after applying MAR. Yellow = overestimated; red = unaffected; green = underestimated (Mou); blue = underestimated (DDMar)

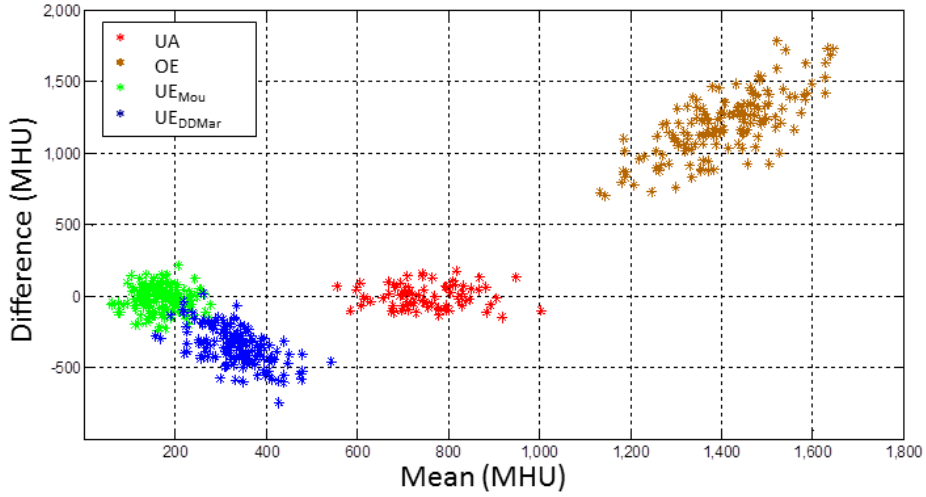


Figure 4.27: Bland-Altman plot for baggage data using Modified Hounsfield Units (MHU). The plot illustrates the change in intensity values for different regions in the scan after applying MAR. Yellow = overestimated; red = unaffected; green = underestimated (Mou); blue = underestimated (DDMar)

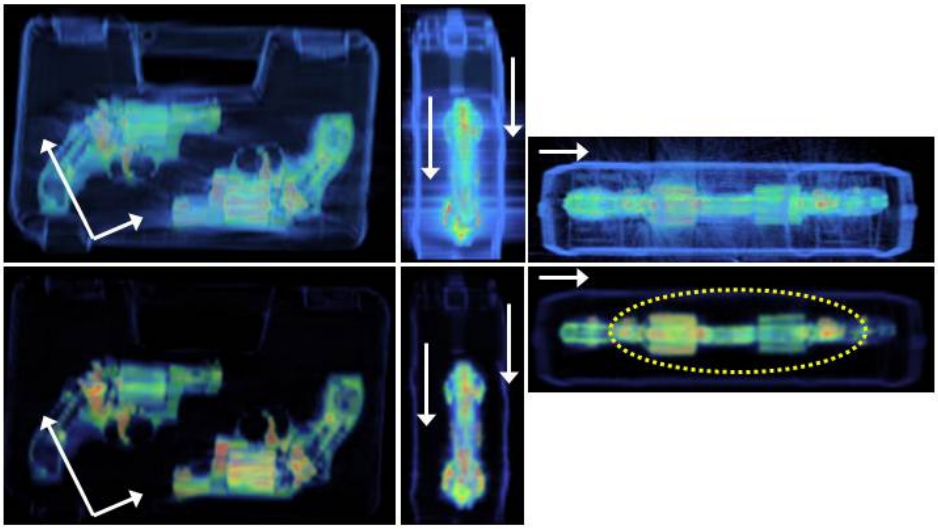


Figure 4.28: Distance-driven MAR results showing volumetric visualisations of clutter-free baggage scan in three different orientations. Top row: Visualisation of input. Bottom row: Visualisation post-DDMar. Regions illustrating significant metal artefact reduction (white arrows) and undesired blurring (yellow circle) indicated.

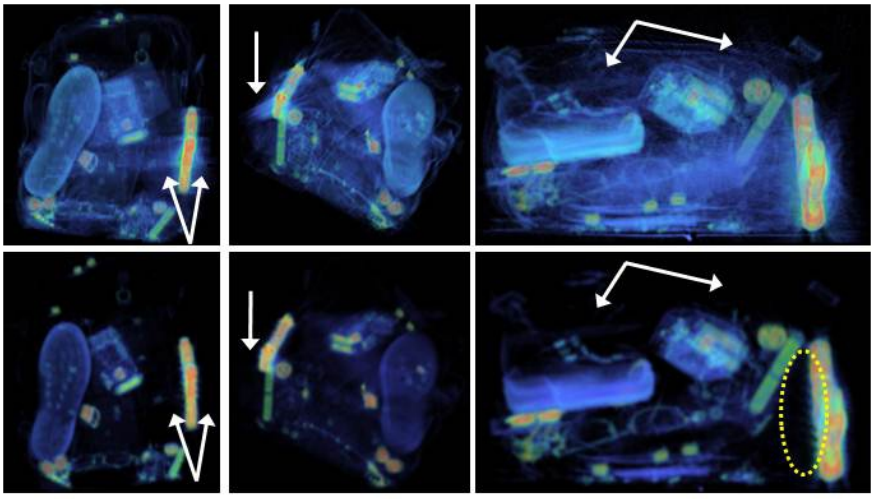


Figure 4.29: Distance-driven MAR results showing volumetric visualisations of the cluttered baggage scan at three different orientations. Top row: Visualisation of input. Bottom row: Visualisation post-DDMar. Regions illustrating significant metal artefact reduction (white arrows) and undesired blurring (yellow circle) indicated.

4.6 Conclusion

This chapter has presented a comparison of denoising and Metal Artefact Reduction (MAR) techniques in the previously unconsidered context of low-resolution, complex volumetric-CT imagery. Previous studies considering the denoising of such imagery are limited to the work of Zhou *et al.* [ZPA10b, ZPA10a], where images with comparatively low degrees of noise and artefacts are considered and no comparative analysis is performed.

Qualitative performance analysis indicated that, although all of the standard 2D denoising techniques yield improvements in the visual quality of the volumes, the most significant improvements are offered by wavelet shrinkage [CDAO95] and the Non-Local Means (NLM) filter [BCM05a, MS05a], both of which significantly outperform the dedicated CT-baggage-denoising approach of Zhou *et al.* [ZPA10b].

A quantitative performance analysis using a novel performance metric (based on the ratio of object to noise 3D-SIFT interest points) was used to quantify the potential impact of denoising on subsequent feature-based automated classification. Performance evaluation using this technique demonstrated the positive impact of denoising, particularly for wavelet shrinkage. The improvements observed in the quantitative analysis and improved volume-rendering results for the NLM filter and wavelet shrinkage, is an indication that these standard 2D denoising techniques will benefit the application of complex computer-vision techniques to low-resolution, cluttered volumetric-CT imagery [FBM10, MFB10, FBM12].

The contributions of the denoising component of this chapter have been the extension of the previous works of Zhou *et al.* [ZPA10b, ZPA10a] by considering low-resolution, cluttered volumetric-CT imagery which is more representative of that encountered in industry; by comparing the performance of a variety of simple, yet popular denoising algorithms, which have been met with success in other areas of image processing and by considering the impact of denoising on subsequent feature-based automated object classification within this environment via the development of a novel 3D SIFT-based performance measure.

Current MAR literature is restricted almost entirely to the medical domain, where CT imagery is typically of a much higher quality with comparatively low degrees of artefacts and clutter and where *a priori* knowledge of the contents and characteristics of the data exists. The majority of the state-of-the-art MAR techniques have exploited these characteristics and their applicability to other domains is thus unclear. A comprehensive evaluation of MAR in non-medical-imaging domains has not been considered previously.

This chapter has presented an experimental comparison grounded in an evaluation based on a standard scientific comparison protocol for MAR methods, using a software generated medical phantom image. The experimental comparison has

been extended beyond the medical-imaging domain by considering novel applications of CT imagery where the MAR literature is limited. In particular, CT images obtained from the aviation security-screening domain, which consist of metal objects with no surrounding tissue in both isolated and cluttered environments, have been considered. A quantitative analysis (of eleven state-of-the-art techniques and one novel technique) has been performed by considering both image quality as well as computational cost and has demonstrated the importance (especially in non-medical applications) of considering the tradeoff between the two measures when determining overall performance.

While the performance trends observed on the medical data (simulated and clinical) are fairly similar to those predicted by the literature, two important observations are made: 1) the medical MAR techniques that employ prior information, are less well-suited to settings where the generation of such priors is difficult (i.e. the security-screening domain); 2) sinogram-completion-based approaches are generally sensitive to input parameters, require manual tuning and result in a characteristic blurring of the regions surrounding high-density objects. The experimental analysis has additionally supported the claims of Kratz *et al.* [KEMB11] regarding the reliability and effectiveness of their reference-free quantitative performance measure. A comparative study of this nature, that considers both medical and non-medical applications, has not been conducted previously.

The limitations in the existing MAR literature have been addressed further through the development of a novel Distance-Driven MAR (DDMar) technique designed particularly for images containing multiple metal objects in complex, non-tissue surroundings. A qualitative as well as quantitative analysis of the technique (in both 2D and 3D) has indicated that high-intensity streaking as well as dark bands (typically occurring between metal objects) are effectively reduced and image quality is improved. Nonetheless, as is a common trend with sinogram-completion-based approaches, the corrected images contain undesired blurring.

The impact of denoising and MAR on the performance of subsequent automated object classification is directly investigated and quantified in Chapter 8 of this work. The importance of determining an appropriate tradeoff between image quality and computational cost and the development of techniques to automatically determine optimal MAR algorithm parameters are highlighted and left as areas for future work. The superiority of iterative reconstruction (in terms of artefact removal), the ever increasing computational power of modern hardware and the sensitivity of sinogram-completion-based approaches to parameter settings, indicates that iterative reconstruction optimisation techniques may be a more fruitful avenue to pursue in future work, as opposed to attempting to develop novel sinogram-completion-based approaches.

Chapter 5

Classification of Subvolumes

Prior literature addressing the task of 3D object classification in low-resolution cluttered volumetric CT imagery (such as that obtained in the security-screening domain) is limited (Section 3.3). The majority of related studies are found in the medical domain, where techniques such as Support Vector Machines (SVM) [MMBB+05, MMK+09], boosting [EBPP12] and random forests [Cri11] have enjoyed success in a broad range of classification tasks.

The Bag-of-Words (BoW) model [SZ03] (Section 3.3.2) has enjoyed success in various object recognition and image classification tasks. Although the BoW model, constructed using 3D SIFT keypoints and descriptors [FBM10], has demonstrated reasonable performance in the classification of threats in cluttered volumetric baggage-CT imagery [FBM13, FBM12], it is shown to suffer a significant decline in performance in the presence of noise and artefacts and does not offer the same level of accuracy as the current state-of-the-art 3D visual cortex classification model [FBM12]. Due to the characteristically high computational cost of traditional clustering techniques used for feature encoding in the BoW model (e.g. k-means clustering [JT05]) the known advantages of dense-feature sampling strategies have not previously been exploited in the baggage-CT domain.

Feature encoding and codebook generation using techniques such as Extremely Randomised Clustering (ERC) forests [MTJ07] have been shown to offer significant gains in terms of classification performance as well as runtime. The significant reduction in computational cost associated with such techniques has allowed for the incorporation of dense-feature sampling strategies into high-dimensional domains. Although clustering forests have not previously been applied directly to the task of 3D object classification in complex volumetric CT imagery, similar techniques have demonstrated success in related 3D problems such as Multi-Atlas Label Propagation (MALP) for the labelling of healthy brain tissue in MR imagery [ZGC13].

The review of the literature in Section 3.3 has shown that prior works addressing the classification of objects in low-resolution complex volumetric baggage-CT

imagery are limited in the following ways:

1. Classification is performed on manually segmented subvolumes.
2. The state-of-the-art visual cortex approach [FBM12] incurs large computational overhead.
3. Classification performance declines in the presence of noise and artefacts.
4. Codebook approaches using salient keypoints produce suboptimal performance (especially in the presence of noise and artefacts).
5. High-speed feature-encoding techniques (e.g. ERC forests [MTJ07]) have not been considered.
6. Dense feature sampling strategies have not been considered.

In an attempt to address these limitations in the state-of-the-art, the performance of five codebook-classification models are compared to the baseline 3D visual cortex approach [FBM12]. The codebook models are constructed using various combinations of sampling strategies (salient and dense), feature-encoding techniques (k-means clustering and ERC forests) and classifiers (Support Vector Machines (SVM) and random forests). A codebook model constructed using ERC forests, a dense-feature sampling strategy and an SVM is shown to yield correct classification rates in excess of 98% and false-positive rates of less than 1% in the classification of handguns and bottles, representing an improvement over the state-of-the-art [FBM12]. These improvements, in conjunction with a significant reduction in computational cost, make the proposed approach an attractive option for the classification of threats in 3D baggage-CT imagery.

The BoW model and random forests form the bases of the research presented in this chapter. An overview of the most relevant concepts related to each has been presented in Section 3.3.2. The remainder of this chapter presents a detailed description of the proposed classification methodologies (Section 5.1) followed by a presentation and discussion of the results of the experimental comparisons (Sections 5.2 - 5.4).

Portions of this chapter have been submitted for publication and are currently under review [MBF14].

5.1 Methods

Sivic and Zisserman [SZ03] proposed the original BoW model (or bag-of-visual-words) for images, whereby local features obtained from images are grouped into a finite number of clusters. The BoW model has since enjoyed success in a broad

range of computer-vision tasks (Section 3.3.2). The traditional BoW-classification framework is adopted here [NJT06a]: 1) feature detection (sampling strategies); 2) feature description; 3) visual codebook generation and vector quantisation and 4) classification.

5.1.1 Interest Point Sampling Strategies

The performance of two feature point sampling strategies are evaluated and compared: 1) a sparse sampling strategy (using the 3D SIFT interest-point detector [FBM10] and 2) a dense sampling strategy (described below).

The benefits (in terms of classification performance) of dense-feature-point sampling strategies are well documented [NJT06b] (Section 3.3.1). Prior works in object classification in baggage-CT imagery [FBM12, FBM13, FMMB13] have adopted a sparse sampling strategy using the 3D SIFT keypoint detector [FBM10]. In accordance with the observations made in Section 3.3.1, the performance of a dense-feature-point sampling strategy, whereby interest points are sampled uniformly and randomly, is evaluated against the SIFT keypoint approach (implemented according to [FBM12]). An invariance to uniform changes in image scale is obtained by sampling interest points from three image scales [Lin94] (as per [FBM10]). At each of the scales a limit of $\tau_N = 0.006N$ on the number of randomly sampled points is enforced (where N is the number of voxels in the Gaussian scale-space image and τ_N is determined empirically)). For the volumes used in this study ($N \sim 3 \times 10^5$), the proposed sampling strategy typically leads to an increase of two orders of magnitude in the number of sampled points compared to the 3D SIFT keypoint detection approach of Flitton *et al.* [FBM10] (making conventional k-means unsuitable).

5.1.2 Feature Description

Flitton *et al.* [FBM13] have shown that simple density statistics-based descriptors outperform more complex 3D descriptors (SIFT [FBM10] and RIFT [LSP03]) in object detection within low-resolution, complex volumetric CT imagery (Section 3.3.3). In accordance with these findings, the Density Histogram (DH) descriptor [FBM13] is used here.

This DH descriptor defines the local density variation at a given interest-point as an N-bin histogram defined over a continuous density range. With reference to Figure 5.1, for a given interest-point, P, every voxel k in the local neighbourhood of that point contributes to a single histogram bin as follows. The active histogram bin (determined by the density I_k of k) is incremented by an amount $w(d_k, \sigma)$, where d_k is the distance (in voxels) from I to k and $w(d, \sigma)$ is a Gaussian window function:

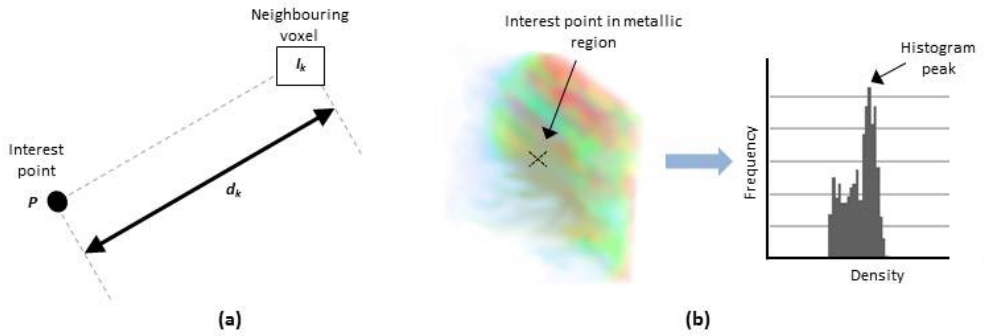


Figure 5.1: Density Histogram Descriptor [FBM13]. (a) Local neighbourhood of interest-point P. (b) Example density histogram descriptor for interest-point in metallic region.

$$w(d,\sigma) = \exp\left[-\left(\frac{d}{\sigma}\right)^2\right] \tag{5.1}$$

It should be noted that, given the definition of distance in voxels, this window remains consistent with the resolution of the volume being examined. On completion, the resulting histogram is normalised to unit area. The descriptor is parametrised by the width of the Gaussian window function σ and the histogram bin-width N_{dh}. Figure 5.1 (b) illustrates an example of the density histogram resulting from an interest-point located within a predominantly metallic region. The histogram is shown to have a corresponding peak resulting from the high concentration of metal within the neighbourhood of the interest-point.

5.1.3 Visual Codebook Generation

The performance of two feature-encoding techniques are evaluated and compared: 1) k-means clustering (using a sparse-feature sampling strategy) and 2) Extremely Randomised Clustering (ERC) forests [MTJ07] (using both sparse and dense sampling strategies). A description of the particular ERC forests procedure adopted is presented below.

Traditional k-means clustering is computationally expensive, limiting its suitability for use with a dense-feature sampling strategy [NJT06b] (Section 3.3.2). To allow for the incorporation of a dense sampling strategy, codebooks are constructed using Extremely Randomised Clustering (ERC) forests [GEW06, MTJ07], which assign separate codewords to every leaf node in a given forest (a full description of the technique is provided in Section 3.3.2). The BoW representation for a given image is obtained by accumulating the codeword counts after applying the forest to all the descriptors in the image. The resulting histogram of codewords is then used in subsequent classification in the same way as any standard BoW model. In contrast to k-means clustering, ERC forests are supervised. Trees are trained in

a top-down recursive fashion [Bre01] using a set of labelled training descriptors, where the labels are obtained from global image annotations (i.e. all descriptors from a given image share the same label). A simple thresholding function is used as the node split function for all internal nodes of the forest:

$$f(v_i, \theta_j) = \begin{cases} 0 & v_i < \theta_j \\ 1 & \text{otherwise} \end{cases} \tag{5.2}$$

where $v_i, i = 1, \ldots, D$ is a single feature attribute selected from a D-dimensional descriptor vector $\mathbf{v} \in \mathbb{R}^D$ and θ_j is a scalar valued threshold ($D = 60$). The optimality criterion used for node splitting is the classical Information Gain (IG) [Cri11]:

$$IG_j = H(S_j) - \sum_{i \in \{L,R\}} \frac{|S_j^i|}{|S_j|} H(S_j^i) \tag{5.3}$$

where S_j is the input sample at node j; i indexes the left (L) and right (R) child nodes and $H(S)$ denotes the Shannon entropy of a set of points:

$$H(S) = -\sum_{c \in C} p(c) \log p(c) \tag{5.4}$$

where $p(c)$ is the normalised class-label distribution in S and c indicates the class label (C being the set of all possible classes). A high value for IG_j indicates good class separation. Randomness is injected into the trees via random node optimisation, whereby a random subset of the available node test parameter values $\mathcal{T}_j \subset \mathcal{T}$ is considered at each node (i.e. performing $|\mathcal{T}_j|$ tests at each node) and selecting the test that produces the highest gain. For each test a new randomly selected threshold is considered and this threshold is tested for all feature attributes. The attribute resulting in the best split is accepted for that threshold. The entire available training set as well as every feature attribute is considered at each node (i.e. no bagging and no randomised attribute selection). A fixed value for $|\mathcal{T}_j|$ is used for all nodes. Trees are grown to a maximum depth D_T and leaf nodes are generated if the information gain falls below a minimum threshold IG_{min}. The value for D_T is chosen empirically using a validation set, to avoid fully grown trees.

5.1.4 Classification

The classification task considered in this work is binary in nature, whereby images are classified according to the presence or absence of a particular target item. The performance of two classifiers are evaluated and compared using the aforementioned combinations of sampling strategies and feature-encoding techniques: 1) Support Vector Machines (SVM) [Vap00] and 2) random forests [Bre01]. The fundamentals of both approaches have been discussed in Section 3.3.2. An SVM classifier using a Radial Basis Function (RBF) kernel is implemented here.

A random-forest classifier composed of trees constructed in a greedy, top-down manner (beginning at the root) is used here. In order to maximise training and classification speed, a common approach in recent literature has been to construct random forests using very basic pixel-level node tests (e.g. pixel differences [SJC08, WC06]). Similarly, in order to optimise the runtime of the random-forest classifier presented here, a simple linear classifier on the encoded feature vector $\mathbf{v}$ is used as the node-split function [BZM07]:

$$f(\mathbf{v}, \theta_j) = \begin{cases} 0 & \mathbf{n}^T\mathbf{v} + b \leq 0 \\ 1 & \text{otherwise} \end{cases} \tag{5.5}$$

where $\theta_j = \{\mathbf{n}, b\}$; $\mathbf{n}$ is a vector of the same dimensions as $\mathbf{v}$ and b is a constant. Randomness is incorporated via randomised node optimisation. That is to say, the vector $\mathbf{n}$ is randomly populated with values in the range $[-1, 1]$ and the constant b is randomly selected. Despite its simplicity, the linear classifier-based node split function has demonstrated good performance in previous studies [BZM07]. Trees are again grown to a maximum depth D_T and leaf nodes are generated if the information gain falls below a minimum threshold IG_{min}. The optimality criterion used for node splitting is the Information Gain (IG) as per Equation 5.3.

Given a test data point $\mathbf{v}$ (representing a test image), the output posterior probability of the random-forest classifier (composed of T trees) is computed as the average of the individual leaf-node predictions [Bre01]:

$$p(c|\mathbf{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\mathbf{v}) \tag{5.6}$$

5.1.5 Subvolume Data

The proposed techniques are evaluated on the classification of two target objects (handguns and bottles) in complex 3D baggage-CT imagery. The two object

Target object	Num. Images
Pistol	184
Revolver	100
Bottle	113
Clutter	179

Table 5.1: Whole volumes from which subvolumes were cropped. Handgun and bottle volumes were manually cropped with 30mm margin around the target object. Clutter subvolumes were automatically cropped. The handgun and bottle whole volume sets are not mutually exclusive.

classes are considered independently of one another. The objective of this chapter is to evaluate classification performance against the baseline visual cortex model [FBM12]. The identical subvolume dataset (composed of handgun subvolumes, bottle subvolumes and clutter subvolumes) used in this baseline study [FBM12] is thus used here.

Flitton *et al.* constructed the handgun and bottle subvolumes by manually cropping whole-volume scans in the dataset described in Section 1.2 such that each subvolume contained a single handgun (pistol or revolver) or a single empty or liquid-containing bottle (with variable liquids). A margin of 30mm was extended around each of the cropped objects. The handguns and bottles were originally scanned in random poses to obtain rotational invariance. The final handgun dataset contained 284 handgun subvolumes, while the bottle dataset contained 534 bottle subvolumes. Clutter subvolumes were obtained by automatically subdividing bottle and handgun-free whole volumes into subvolumes of sizes spanning the range of sizes in the handgun and bottle subvolume datasets. As the handguns and bottles were scanned in numerous orientations, the subvolume sizes varied considerably. Although the whole volumes from which the clutter subvolumes were cropped contained a variety of items representative of that typically found in packed luggage (e.g. books, clothing, shoes etc.), the cropping was unsupervised and the resulting subvolumes were thus not guaranteed to contain any whole objects. In total 971 clutter subvolumes were generated in this way. Additionally, 199 handgun subvolumes were used as clutter subvolumes in the bottle classification experiments. The anisotropic subvolumes (resolution $1.56 \times 1.61 \times 5$mm) were resampled to create isotropic voxel resolutions of $2.5 \times 2.5 \times 2.5$mm using cubic spline interpolation.

The final handgun dataset used here and in [FBM12] is thus composed of 971 clutter subvolumes, while the bottle dataset is composed of 1170 clutter subvolumes. The reasoning behind the choice of sizes for the handgun and bottle datasets is not known. The datasets are summarised in Tables 5.1 and 5.2 and several example subvolumes are shown in Figure 5.2.

Dataset	Target	Clutter
Handgun	284	971
Bottle	534	1170

Table 5.2: Subvolume test datasets. Handgun dataset contains 284 handgun (target) subvolumes and 971 clutter subvolumes. Bottle dataset contains 534 bottle (target) subvolumes and 1170 clutter subvolumes (of which 199 are handgun subvolumes).

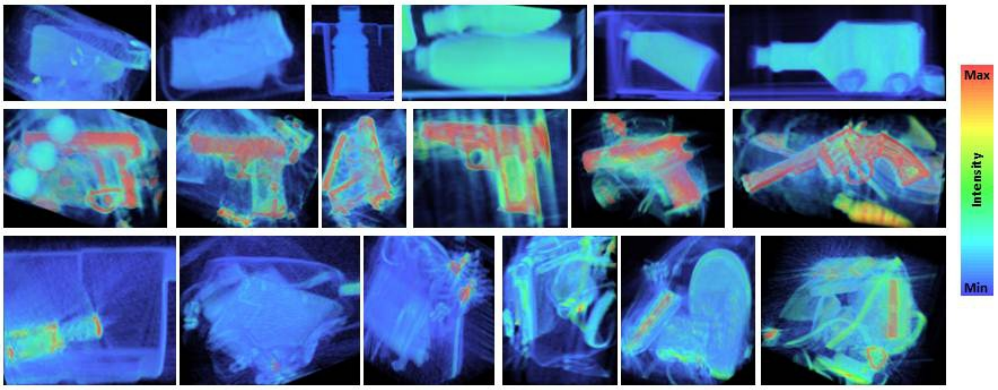


Figure 5.2: Example subvolume data: bottles (top); handguns (middle); clutter (bottom). Handgun subvolumes are created to contain a single handgun; bottle subvolumes a single bottle and clutter subvolumes are not guaranteed to contain any whole objects.

5.1.6 Summary of Methodology

In summary, the following six classification techniques are evaluated and compared (Table 5.3):

1. **Codebook1:** A codebook model built using sparsely sampled 3D-SIFT interest points [FBM10], k-means clustering feature encoding [FMMB13] and an SVM classifier [BL03].

2. **Codebook2:** A codebook model built using sparsely sampled 3D-SIFT interest points, ERC forest encoding [MTJ07] and an SVM classifier.

3. **Codebook3:** A codebook model built using sparsely sampled 3D-SIFT interest points, ERC forest encoding and a random-forest classifier [BZM07].

4. **Codebook4:** A codebook model built using densely sampled feature points, ERC forest encoding and an SVM classifier.

5. **Codebook5:** A codebook model built using densely sampled feature points, ERC forest encoding and a random-forest classifier.

6. **Cortex:** The 3D visual cortex model [FBM12].

Alias	Sampling	Model	Classifier
Codebook1	Sparse SIFT	k-means codebook	SVM
Codebook2	Sparse SIFT	ERC forest codebook	SVM
Codebook3	Sparse SIFT	ERC forest codebook	Random forest
Codebook4	Dense	ERC forest codebook	SVM
Codebook5	Dense	ERC forest codebook	Random forest
Cortex	Dense Gabor	3D visual cortex	SVM

Table 5.3: Summary of classification techniques compared.

5.2 Results

Testing was conducted via a 10-fold cross-validation procedure, using the identical data-splits used in the baseline study [FBM12]. This allowed for a direct performance comparison between methods. Performance was quantified via traditional measures (true-positive rate, false-positive rate and precision).

The cost C and the kernel width γ of the RBF kernel used in the SVM classifier were optimised using a standard grid-search cross-validation procedure [BHW10]. The k-means clustering-based codebooks were generated using $k = 1024$ clusters for the handgun target class and $k = 512$ clusters for the bottle target class. A kernel-based cluster assignment methodology was adopted for both classes (with kernel widths of $\sigma = 0.04$ and $\sigma = 0.08$ for handguns and bottles respectively). These parameters were based on the extensive experimental comparison performed on the same dataset in [FMMB13]. The optimal ERC forest parameters were determined empirically using a small validation set. The number of tests performed for each node split was set to $|\mathcal{T}_j| = 30$ - this value was fixed for all nodes. Trees were grown to a maximum depth of $D_T = 10$, with a lower bound of $IG_{min} = 10^{-4}$ on the information gain (see Section 3.3.2 for explanations of these parameters). It was found that using these settings resulted in tree growth terminating prior to maximum depth and thus no tree pruning was performed. The settings resulted in trees with approximately 1000 leaf nodes each. For a forest containing $T = 25$ trees, codebooks therefore typically contained approximately 25000 codewords. The optimal classification forest parameters were determined in a similar manner. Forests were composed of 30 trees, grown to maximum depths of $D_T = 20$ and used a lower bound of $IG_{min} = 10^{-4}$ on the information gain.

Experiments were performed on an Intel Core i5 machine running a 2.30GHz processor with 6GB of RAM. The random forest clustering and classification methods were implemented in C++ using the Sherwood decision forest library [CS13].

Method	Time (s)
Codebook1	-
Codebook2	94.36
Codebook3	92.83
Codebook4	186.89
Codebook5	161.47
Cortex	$> 3.6 \times 10^3$

Table 5.4: Classification processing times - averaged over handgun and bottle experiments.

The processing times, measured over the entire 10-fold cross-validation procedure and averaged over the two experiments (bottles and handguns) are recorded in Table 5.4. Use of the average is justified by the fact that all subvolumes considered in this chapter are of similar sizes and hence result in similar-sized codebooks (codebook sizes being the main factor impacting processing times). As the Codebook1 and Cortex approaches were not directly implemented in this study, their corresponding processing times are not known. It is known, however, that the processing time for the construction of the 3D visual cortex model using the current dataset, is in the order of hours [Fli12]. As expected, the sparse-feature sampling strategy (for both the SVM and random-forest classifiers) led to considerably lower processing times ($\sim 90s$) relative to the dense sampling strategy ($\sim 175s$). For each sampling strategy, the random-forest classifier resulted in a marginal improvement in processing time ($\sim 2s$ for sparse sampling; $\sim 25s$ for dense sampling).

Tables 5.5 - 5.8 summarise the results of the experiments averaged over the 10 folds. Table 5.5 shows the confusion matrices obtained in the handgun classification experiments for each of the six methods tested. While there was no major variance in performance across all six methods, the most significant improvements over the baseline Cortex approach [FBM12] were produced by the Codebook2, Codebook4 and Codebook5 approaches, each of which employ ERC forests. The use of the random-forest classifier resulted in a decline in performance, particularly in terms of the number of false-negative classifications (see Codebook2 (SVM) vs. Codebook3 (random forest) and Codebook4 (SVM) vs. Codebook5 (random forest)). Codebook4 (ERC forest, dense-feature sampling and SVM classifier) produced the optimal performance, with only 4 erroneous classifications (1 false negative and 3 false positives).

The results of the bottle classification experiments are shown in the confusion matrices in Table 5.6. The codebooks constructed using sparse-feature sampling resulted in the highest numbers of misclassifications (Codebook1 (92 misclassifications); Codebook2 (58 misclassifications) and Codebook3 (73 misclassifications)), although the use of ERC forests resulted in an improvement over k-means clus-

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	954	17
Handgun (actual)	8	276

(a) Codebook1

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	964	7
Handgun (actual)	4	280

(b) Codebook2

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	965	6
Handgun (actual)	13	271

(c) Codebook3

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	968	3
Handgun (actual)	1	283

(d) Codebook4

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	965	6
Handgun (actual)	6	278

(e) Codebook5

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	960	11
Handgun (actual)	9	275

(f) Cortex

Table 5.5: Handgun classification confusion matrices (284 handgun and 971 clutter images)

tering (Codebook1 vs. Codebook2). The most significant improvements over the baseline Cortex approach (30 misclassifications) were produced by the densely sampled codebooks (Codebook4 (13 misclassifications) and Codebook5 (22 misclassifications)). Optimal performance was again achieved using the ERC forest with a dense sampling strategy and an SVM classifier (Codebook4), with only 6 false-negative and 7 false-positive classifications. Similarly to the handgun experiments, the SVM classifier outperformed the random-forest classifier (see Codebook2 vs. Codebook3 and Codebook4 vs. Codebook5).

The true-positive, false-positive and precision rates for the handgun and bottle experiments are shown in Tables 5.7 and 5.8 respectively. The performance gains of the ERC codebook using dense sampling and an SVM classifier (Codebook4 - in bold) over the state-of-the-art Cortex approach are clear with improvements of $> 3\%$ and $> 2\%$ in the true-positive rates for handgun and bottle recognition respectively and reductions of 70% and 40% in the corresponding false-positive rates.

5.3 Discussion

The ERC forest codebook using dense-feature sampling and an SVM classifier (Codebook4) has offered improvements in both processing times as well as classification performance over the current state-of-the-art [FBM12]. Several trends are revealed upon closer examination of the aforementioned classification results. Note that each of the subvolumes shown in Figure 5.2 were correctly classified by all six methods.

Figure 5.3 illustrates several handgun misclassifications produced by each of the six methods. In terms of false-positive (FP) classifications, the only obvious trend is the presence of high-density objects (coloured red/orange), particularly in the false-positive instances for the ERC forest codebooks (Codebook2, Codebook3, Codebook4 and Codebook5). The k-means clustering codebook (Codebook1) and the Cortex false positives bear minimal similarities to the handgun training data. The only possible trend in the false-negative handgun classifications is the lack of prominence of the handles of the handguns relative to the barrels for examples from the Codebook2, Codebook3 and Codebook4 approaches (note that both the Codebook2 and Codebook3 approaches had additional false negatives which did not exhibit these characteristics). These examples have been magnified for clarity in Figure 5.4. The handgun handles are noticeably more prominent in the correctly classified handguns (Figure 5.2) as well as the remaining false-negative handguns in Figure 5.3 (especially the Codebook1 and Cortex examples).

Considering the classification of bottles (Figure 5.5) the performance gains of all four ERC forest-based codebooks over the traditional k-means codebook (Code-

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	1135	35
Bottle (actual)	57	477

(a) Codebook1

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	1148	22
Bottle (actual)	36	498

(b) Codebook2

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	1127	43
Bottle (actual)	30	504

(c) Codebook3

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	1163	7
Bottle (actual)	6	528

(d) Codebook4

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	1162	8
Bottle (actual)	14	520

(e) Codebook5

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	1158	12
Bottle (actual)	18	516

(f) Cortex

Table 5.6: Bottle classification confusion matrices (534 bottle and 1170 clutter images)

Method	TPR (%)	FPR (%)	Precision
Codebook1	97.34 ± 3.41	1.81 ± 1.70	0.942 ± 0.053
Codebook2	98.60 ± 1.52	0.70 ± 0.31	0.976 ± 0.028
Codebook3	95.61 ± 3.30	0.61 ± 0.72	0.978 ± 0.023
Codebook4	**99.71 ± 0.51**	**0.28 ± 0.21**	**0.990 ± 0.013**
Codebook5	97.74 ± 2.13	0.57 ± 0.53	0.979 ± 0.018
Cortex	96.81 ± 2.64	1.10 ± 0.93	0.962 ± 0.029

Table 5.7: Overall handgun classification performance for six tested methods. Optimal performance indicated in bold.

Method	TPR (%)	FPR (%)	Precision
Codebook1	89.33 ± 5.52	3.01 ± 1.44	0.932 ± 0.029
Codebook2	93.31 ± 3.10	1.88 ± 1.22	0.958 ± 0.042
Codebook3	94.23 ± 3.31	3.70 ± 2.00	0.921 ± 0.037
Codebook4	**98.88 ± 0.68**	**0.60 ± 0.25**	**0.987 ± 0.021**
Codebook5	97.44 ± 0.66	0.69 ± 0.43	0.985 ± 0.009
Cortex	96.62 ± 3.23	1.01 ± 1.63	0.977 ± 0.034

Table 5.8: Overall bottle classification performance for six tested methods. Optimal performance indicated in bold.

book1) are more substantial than observed in the handgun experiments. Closer examination of the bottle misclassifications has not indicated any obvious sources of error or notable trends within the false-negative classifications (i.e. missed bottles) produced by all six approaches. The two most obvious consistencies in the false-positive bottle classifications (again for all six methods) are: 1) the presence of items with circular cross sections similar to that of a full bottle and 2) the presence of image regions that are similar in density to the liquids used in the training set. It is worth noting that these observations are in accordance with those made in the previous works of Flitton *et al.* [FBM12, FMMB13].

While codebook approaches, by nature, do not capture spatial/geometric relations between codewords, it appears as if the increase in the amount of information captured by a dense sampling strategy compensates for this limitation. This is illustrated by the gain in performance of Codebook4 (dense sampling with SVM) over Codebook2 (sparse sampling with SVM) and Codebook5 (dense sampling with random-forest classifier) over Codebook3 (sparse sampling with random-forest classifier). It is likely that the k-means codebook-classification results (Codebook1) would improve using dense sampling, but at a significant increase in computational cost. Furthermore, it is suspected that these gains would not match those offered by the ERC forest codebooks using dense sampling, judging from the superior performance of the Codebook2 (ERC forest) over

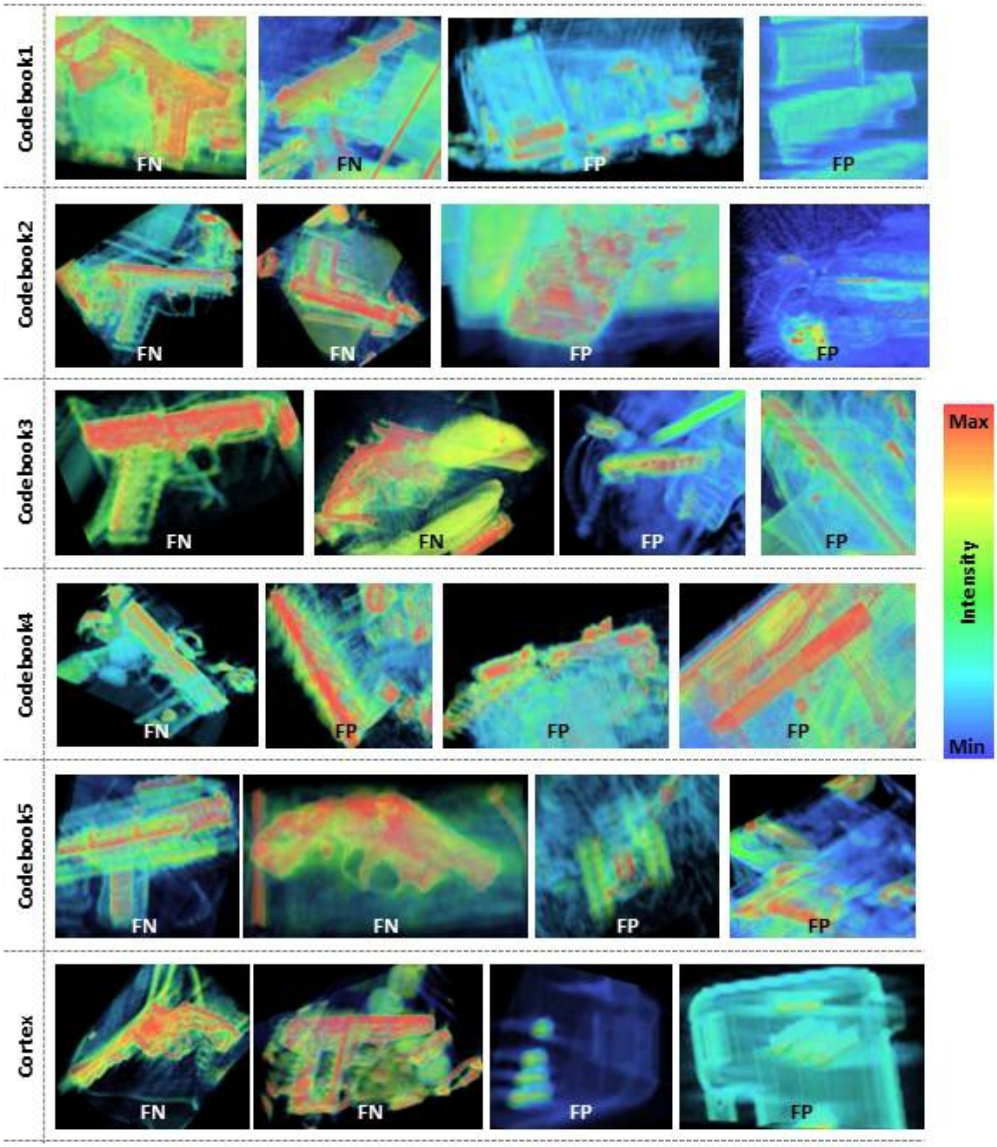


Figure 5.3: Handgun misclassifications: FN indicates false-negative classifications (handgun-containing subvolumes incorrectly classified as not containing handguns) and FP indicates false-positive classifications (handgun-free subvolumes incorrectly classified as containing handguns).

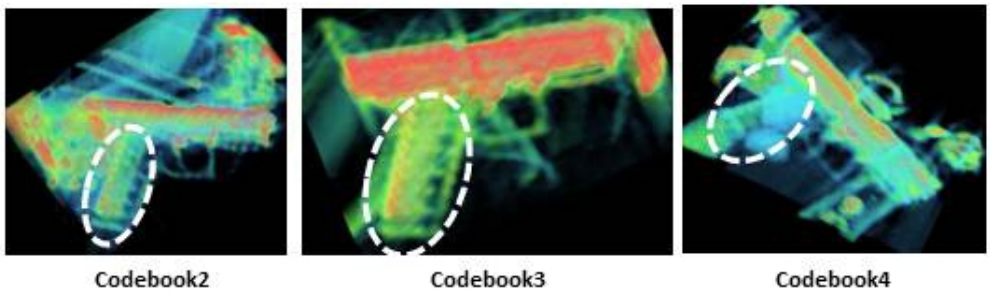


Figure 5.4: Magnified displays of missed handguns (false negatives) in Figure 5.3 resulting from low-density handles.

Codebook1 (k-means clustering) when using identical sparse features but different feature-encoding techniques.

Interestingly, despite the marginal increase in processing time, the SVM classi-

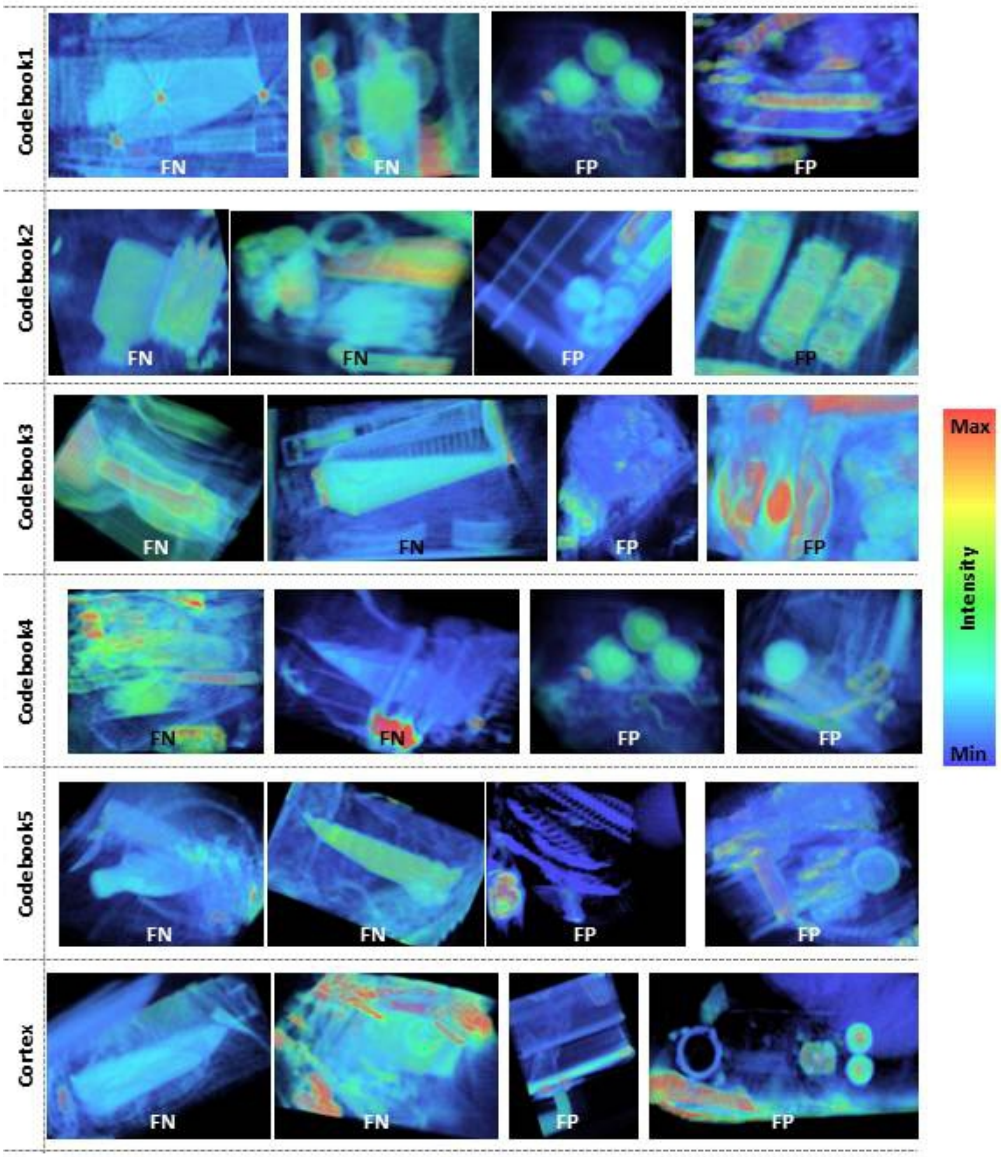


Figure 5.5: Bottle misclassifications: FN indicates false-negative classifications (bottle-containing subvolumes incorrectly classified as not containing bottles) and FP indicates false-positive classifications (bottle-free subvolumes incorrectly classified as containing bottles).

fier consistently outperforms the random-forest classifier in terms of classification accuracy (Codebook2 vs. Codebook3 and Codebook4 vs. Codebook5). This is in contrast to what has been observed in the prior image classification literature [CKY08] (Section 3.3.2), where a random-forest-based classifier is shown to outperform a variety of popular binary classifiers (including the SVM). It has however, been noted (Section 3.3) that there does exist prior work (especially within the bioinformatics literature) that demonstrates that SVMs consistently outperform random forests in some classification problems [SWA08, OPSS11]. Furthermore, Criminisi [Cri11] emphasises that the potential benefits of random-forest-based image classification over popular techniques (e.g. SVM and boosting) are most prominent in multiclass and high-dimensional classification problems (as opposed

to the two-class classification problems considered here). Although it is thus likely that the random-forest classification performance would improve relative to the SVM performance when considering the classification of multiple threats, it is reasonable to conclude that SVMs are the preferred mode of classification within the current context.

It also is worth emphasising that no pre-processing in the form of noise and/or metal artefact reduction (Chapter 4) has been considered in this chapter - demonstrating the robustness of the dense sampling-based codebook approaches to background noise and artefacts. It is expected, however, that such techniques will be of greater importance in the automation of the segmentation process and is addressed accordingly in Chapter 7 of this work. Finally, the results and observations presented in this chapter have demonstrated an improvement over the state-of-the-art in the automated classification of threats in low-resolution, complex volumetric baggage-CT imagery [FBM12], both in terms of classification accuracy as well as runtime.

5.4 Conclusion

This chapter has investigated the feasibility of a codebook approach for the automated classification of threats in manually segmented low-resolution, subvolumes of complex 3D volumetric baggage-CT imagery by comparing the performance of five codebook models to the current state-of-the-art 3D visual cortex approach [FBM12]. The codebook models are constructed using various combinations of sampling strategies (salient and dense [NJT06b]), feature-encoding techniques (k-means clustering and Extremely Randomised Clustering (ERC) forests [MTJ07]) and classifiers (Support Vector Machines (SVM) [Vap00] and random forests [BZM07]). An improvement over the current state-of-the-art, both in terms of accuracy as well as runtime, is achieved using a codebook constructed using an ERC forest, a dense-feature sampling strategy and an SVM classifier. The research presented in this chapter has extended the current state-of-the-art in the following ways:

1. State-of-the-art true-positive rates have been improved by $> 3\%$ and $> 2\%$ for handgun and bottle classifications respectively, with corresponding reductions of 70% and 40% in the false-positive rates.

2. Runtime has been decreased by several orders of magnitude.

3. A high-speed feature-encoding technique (the ERC forest [MTJ07]) has been implemented in the previously unconsidered domain of complex volumetric baggage-CT imagery.

4. Classification performance in the presence of noise and artefacts has been improved using a dense feature-point sampling strategy (which was previously not feasible using k-means clustering).

These improvements make the proposed approach an attractive option for the classification of threats in 3D baggage-CT imagery. The final step towards a fully-automated classification framework is the automation of the manual subvolume generation procedure currently employed by the methods proposed in this chapter. The automated segmentation of cluttered volumetric baggage-CT imagery is addressed in Chapter 7 of this work.

Chapter 6

Dual-Energy Techniques

Dual-Enery Computed Tomography (DECT) techniques have formed the basis for the majority of the state-of-the-art Explosives Detection Systems (EDS) used within the aviation-security domain [Sin03]. DECT techniques may generally be grouped into one of three categories (Section 3.4): post-reconstruction techniques; pre-reconstruction techniques and iterative-reconstruction techniques.

Post-reconstruction techniques operate in the image domain (directly on the low and high-energy scans). While such techniques are the most straightforward and computationally efficient, they are known to be sensitive to image noise and artefacts and the literature indicates that they offer comparatively limited discriminative power (compared to more advanced techniques). Nonetheless, techniques such as image fusion [EHIS+08] and the Dual-Energy Index (DEI) [Joh11] have been successfully employed for a variety of clinical material-differentiation tasks [CZLY13, GJCM09].

Pre-reconstruction DECT techniques are the most widely implemented techniques, particularly in the security-screening domain, where DECT decomposition and subsequent effective atomic number computations have been successfully used for materials-based explosives detection [Rod79, Sin03, SMK11, YNC06, NBC03]. Such techniques perform decompositions in the projection domain to estimate effective atomic number and electron density equivalent reconstructions, which provide indications of the chemical makeup of the objects present in a scan. This information has been shown to be invaluable in the discrimination of common explosive materials. The two primary drawbacks of pre-reconstruction DECT are 1) their susceptibility to reconstruction artefacts and 2) the computational overhead associated with two additional FBP reconstructions required per image.

DECT based on iterative-reconstruction techniques provides superior performance, particularly considering the reduction of artefacts. Improved performance does however, come at an increase in computational demand. Nonetheless, such techniques are gaining popularity with the ever-increasing computational power of modern hardware.

Property	Al	PMMA	Units
Mass density	2.7	1.21	g/cm^3
Z_{eff}	13	6.47	-
Compton coeff.	0.390	0.1952	cm^{-1}
Photoelectric coeff.	69734.0	3309.0	keV^3/cm
Min. thickness	0.16	1.27	cm
Max. thickness	2.56	21.59	cm
Thickness step	0.16	1.27	cm

Table 6.1: Calibration materials for DECT decomposition

DECT applications in aviation security-screening have, for the most part, relied on pre-reconstruction decomposition techniques [SMK11, YNC06]. Such techniques typically rely on calibration data and knowledge of the geometric and spectral configurations of the CT scanner (Section 3.4). The less complex and faster post-reconstruction Dual-Energy Index (DEI) has not been considered previously in this domain - predominantly due to the increased levels of noise and metal-streaking artefacts characteristic of baggage-CT imagery. The effectiveness of metal artefact reduction when used in conjunction with DECT is addressed in this chapter.

While the chemical characteristics (e.g. effective atomic numbers and densities) of the materials in a scan are likely to be of value in both the segmentation and classification of baggage-CT images, the configuration and calibration information necessary to perform accurate dual-energy decompositions has not been readily accessible in this work. This chapter presents an experimental evaluation of the potential discriminative capabilities of the previously-unconsidered DEI within the baggage-CT domain. Performance is compared to the traditional effective atomic number Z_{eff} obtained via pre-reconstruction dual-energy decompositions, performed using simulated calibration and configuration data. The DEI demonstrates meaningful discrimination of five material classes (with differing chemical properties) and is shown to outperform the effective atomic number both in terms of (estimated) accuracy as well as discriminative power. Crucially, high-quality discrimination relies on sufficient metal artefact reduction.

6.1 DECT Experiments

The iso-transmission method of Chuang and Huang [CH87] (Section 3.4.2) was used to perform the dual-energy decompositions. Similarly to [YNC06], simulated spectra were used to generate the required calibration data, using aluminium and plastic (Table 6.1) as the chosen calibration materials. The effective atomic number images were computed according to Equation 3.20 with $n = 3$.

The CT-80DR scanner is an Explosives Detection System (EDS) designed for

Parameter	**Value**
Focus-to-isocentre distance	550mm
Focus-to-detector distance	1008.4mm
FoV diameter	800mm
Rotation speed	90rpm
Detector elements	768
Views per rotation	960 (360° rotation)
Fan angle	95.875°
Spatial resolution	$1.56 \times 1.61 \times 5.00$mm

Table 6.2: Reveal CT-80DR scanner geometry and reconstruction parameters

the inspection of checked baggage. The geometric and reconstruction parameters of the scanner are summarised in Table 6.2. As discussed (Section 3.4.2), the reliable generation of calibration data using simulated energy spectra still requires knowledge of various scanner parameters. Unfortunately, without direct access to the scanner, these parameters are not readily available or are stored in a proprietary format. Those parameters which were not explicitly known for the CT-80DR scanner were estimated, based on related literature [YNC06, SMK11].

The CT-80DR scanner employs a fan-beam geometry with a focus-to-isocentre distance of 550mm, a focus-to-detector distance of 1008.4mm and an optimal spatial resolution of $1.56 \times 1.61 \times 5.00$mm. It is assumed that the dual-energy X-ray spectra are generated by applying a high-voltage power supply to the X-ray tube. The power supply is characterised by a sinusoidally-modulated waveform:

$$V(t) = V_{dc} + V_{ac} \sin(2\pi f t) \tag{6.1}$$

where the nominal values of V_{dc} and V_{ac} were estimated at $V_{dc} = 110$kVp and $V_{ac} = 40$kVp. A waveform frequency of $f = 54$ Hz was assumed. These values were estimated based on related literature [YNC06], as well as the limitations of the spectrum-simulation software [NH85]. The dual-energy X-ray spectra were simulated using the *Xcomp5r* software package [NH85], which generates X-ray spectra based on an input DC voltage, a source-to-detector distance and absorber filtration types and thicknesses. The following beamline filtration was assumed: 2.5mm aluminium and 0.2mm beryllium. Since the *Xcomp5r* software generates spectra based on DC voltages, a single cycle of the voltage supply in Equation 6.1 was discretised into 100 equally-spaced samples. The low and high-energy spectra were then computed by averaging the 50 X-ray spectra corresponding to the negative and positive half-cycles respectively [YNC06]. The resulting spectra (normalised to unit area) are illustrated in Figure 6.1.

In addition to the Z_{eff} reconstructions, Dual-Energy Index (DEI) images were

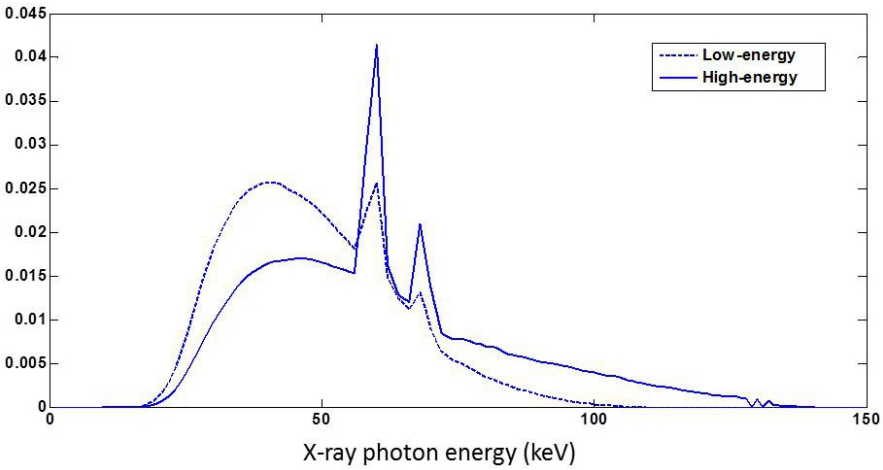


Figure 6.1: Normalised dual-energy spectra used in energy decompositions.

also generated. In order to evaluate the accuracy of both the Z_{eff} and DEI images, knowledge of the contents and corresponding chemical characteristics of the scanned objects are required. Such ground-truth information was not available for the dataset under investigation here. Performance evaluation was thus performed predominantly in a qualitative manner. Some degree of quantitative analysis was performed by manually annotating objects composed predominantly of certain material types. In particular, five object classes were considered (Figure 6.2): outsoles of shoes; clothing and books; bottles; handguns and grenades and bullets. Considering these object classes, several assumptions were made. The outsoles of shoes are assumed to be composed predominantly of vulcanised rubbers which typically have effective atomic numbers in the range of $Z_{\text{eff}} = 5$ to $Z_{\text{eff}} = 7$ [SSAW83]; clothing and books are assumed to be composed predominantly of cellulose (molecular formula: $(C_6H_{10}O_5)_n$) which has an approximate effective atomic number of $Z_{\text{eff}} = 7$; bottles are assumed to be constructed predominantly of plastic and containing liquids which are assumed to have a similar chemical composition to water ($Z_{\text{eff}} = 7.42$); the handguns considered are composed predominantly of stainless steel (main component is iron - $Z_{Fe} = 26$) and to a lesser degree plastic; similarly the grenades are assumed to be fragmentation grenades composed of varying combinations of common raw materials which typically include polycarbonates, steel and ammonium nitrate explosives etc. - handguns and grenades are thus expected to have effective atomic numbers greater than that of water but less than pure metals; bullets are assumed to be composed predominantly of lead alloyed with tin and antimony or zinc (i.e. $Z_{\text{eff}} \geq Z_{Al}$). The object classes and their assumed characteristics are summarised in Table 6.3. It is worth noting that using Equation 3.20 to compute the effective atomic numbers imposes lower and upper limits of $\min\{Z_1, Z_2\}$ and $\max\{Z_1, Z_2\}$ respectively on the computed value, where Z_1 and Z_2 are the effective atomic numbers of the calibration materials.

The aforementioned annotations were used to perform quantitative analyses

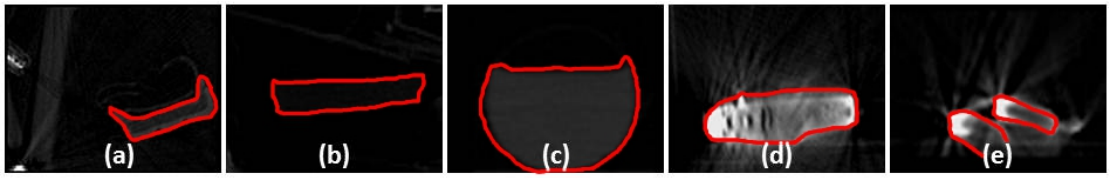


Figure 6.2: Examples of image classes (outlined regions) used in histogram analyses: a) sole of shoe; b) paperback book; c) bottle; d) handgun; e) bullets.

Class	Annotations	Main Component	Approx. Z_{eff}
Shoe soles	69	Vulcanised rubber	$6.4 - 7$
Clothing & books	102	Cellulose	7
Bottles	118	Liquid	7.42
Handguns & grenades	115	Iron, plastic	$7 - 13$
Bullets	32	Lead	13

Table 6.3: Object classes used in DEI and Z_{eff} histogram analyses. The estimated atomic numbers are based on the upper limit of 13 (imposed by the calibration procedure).

by computing the mean Z_{eff} and DEI values for each individual annotation and recording the results in class-specific histograms. These histograms should theoretically provide an indication of both the accuracy of the measure as well as the potential of the measure to provide a robust separation of the typical object classes found in baggage scans.

Finally, all experiments were conducted with and without the application of Metal Artefact Reduction (MAR) techniques. For these purposes, the MAR technique presented in Section 4.2 was employed.

6.1.1 DECT Data

The dual-energy techniques are evaluated on two cluttered volumetric baggage scans obtained on the Reveal CT-80DR dual-energy baggage-CT scanner (Section 1.2). The data is characterised by anisotropic voxel resolutions of $1.56 \times 1.61 \times 5$mm and dimensions of $512 \times 512 \times 99$ (Figure 6.3 (a)) and $512 \times 512 \times 87$ (Figure 6.3 (c)). The scans were chosen to contain a broad range of objects (handguns, shoes, spectacles, golf balls, magazines, alkaline batteries, cups etc.), expected to have distinguishable chemical characteristics. Performance is evaluated on individual 2D 512×512 axial slices as well as the whole volumes.

6.2 Results

Figures 6.4 and 6.5 show the results of the DECT decompositions and DEI computations applied to the two test images without and with metal artefact reduction

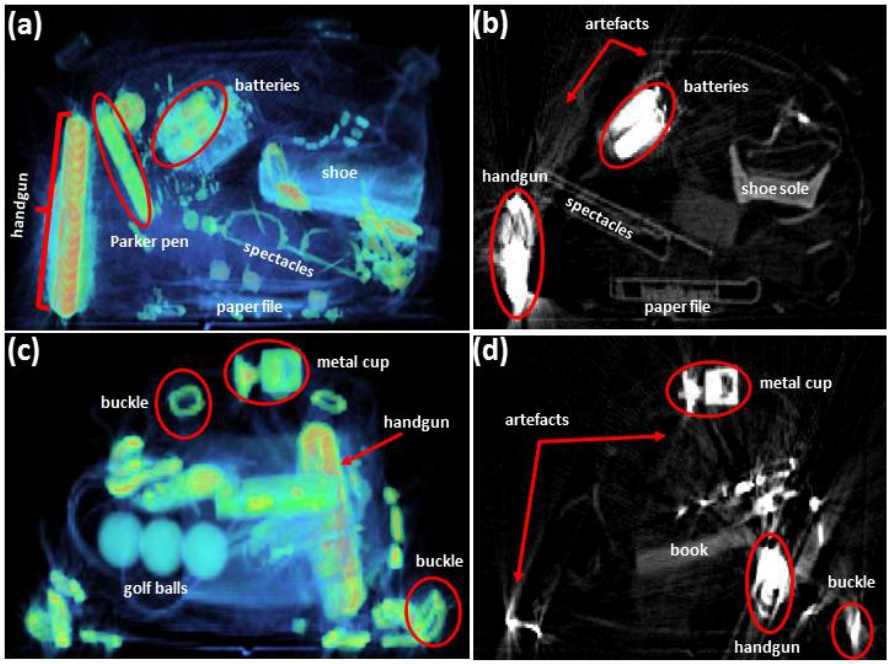


Figure 6.3: DECT test data. (a) and (c): visualisations of test volumes. (b) and (d): 2D axial slices obtained from volumetric data (a) and (b) respectively. Several objects of interest and artefacts have been labelled.

respectively. In both examples in Figure 6.4, the aluminium and plastic equivalent images appear to have emphasised the appropriate object types - those with higher expected atomic numbers for the aluminium images (e.g. metallic handguns and belt buckles) and those with lower expected atomic numbers for the plastic images (e.g. shoes). The impact of the metal-streaking artefacts is however, quite obvious - resulting in the greater than expected aluminium contributions from objects such as the soles of shoes (predominantly rubber). Consequently, the resulting Z_{eff} images are heavily corrupted by streaking artefacts and almost certainly overestimate the atomic numbers of the majority of the objects in the scans. Similarly, the DEI images also contain significant amounts of streaking. The positive impact of metal artefact reduction is clear in Figure 6.5. Observations worth noting include reduced streaking in all the reconstructions; lesser aluminium contributions from metal-free obects (e.g. shoe soles) and a more obvious separation of the constituent objects into plastic and aluminium components. While the resulting Z_{eff} images are considerably clearer (i.e. fewer artefacts), closer inspection indicates only a subtle variation in the intensities (i.e. atomic numbers) of the objects. The paperback in the second test image, for example, appears to have a very similar atomic number (i.e. similar intensity) to the handgun - in reality, one would expect the atomic numbers of the paperback to be noticeably lower than that of the handgun. The DEI images appear to provide superior visualisation of the objects in both test images (compared to the input high-energy scans), with minimal streaking artefacts. This is encouraging for the separating power of the DEI for tasks such as segmentation.

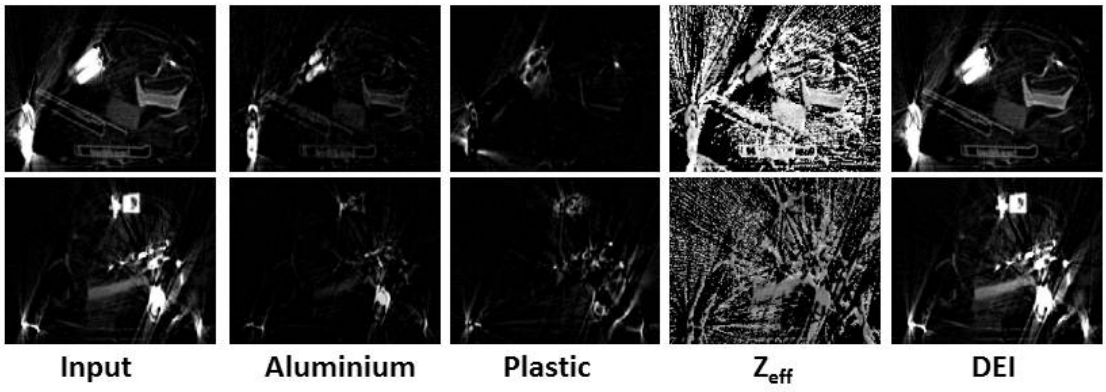


Figure 6.4: Results of 2D DECT decomposition and DEI computation without metal artefact reduction.

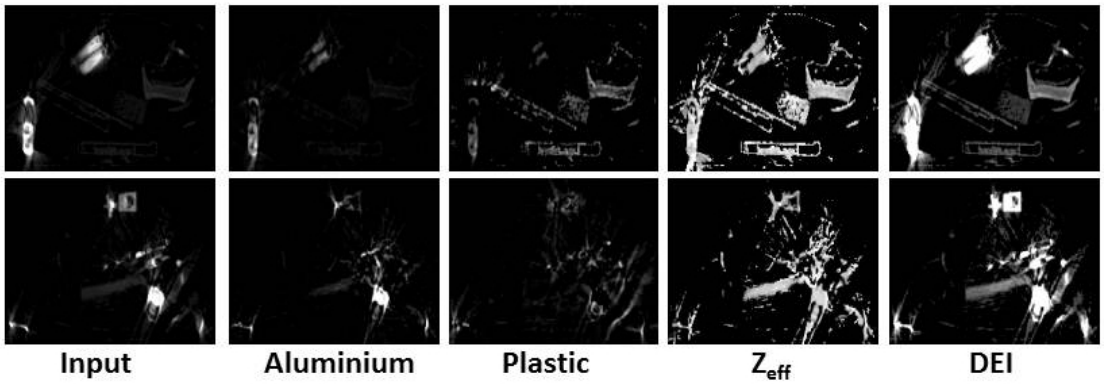


Figure 6.5: Results of 2D DECT decomposition and DEI computation with metal artefact reduction.

Additionally, Figures 6.6 and 6.7 show the results of the Z_{eff} and DEI computations applied to the original volumetric images (with and without MAR). Prior to MAR (Figure 6.6), metal-streaking artefacts appear to have resulted in greater than expected aluminium contributions from predominantly plastic-like objects (e.g. the golf balls which are composed predominantly of rubber and thermoplastics). Furthermore, all meaningful information appears to have been corrupted by noise and artefacts in the Z_{eff} volumes, where there is little (if any) distinction between the effective atomic numbers of various objects. The DEI volumes, although containing some streaking, do separate the objects fairly accurately - low atomic number objects (e.g. book) are, for example, characterised by negative DEIs (see scale), while the high atomic number objects (e.g. handgun) are characterised by high DEIs (~ 0.3). The application of MAR (Figure 6.7) improves the quality of all the volumes - the visualisation of the plastic-like golf balls, for example, is noticeably improved in the plastic-equivalent volume. It is worth noting, however, that the plastic volumes remain difficult to interpret, with considerable amounts of noise. Consequently, the Z_{eff} volumes exhibit very little variation in the computed Z_{eff} values, with the majority of the Z_{eff} values towards the upper end of the scale. In contrast, the post-MAR DEI volumes exhibit good separation of

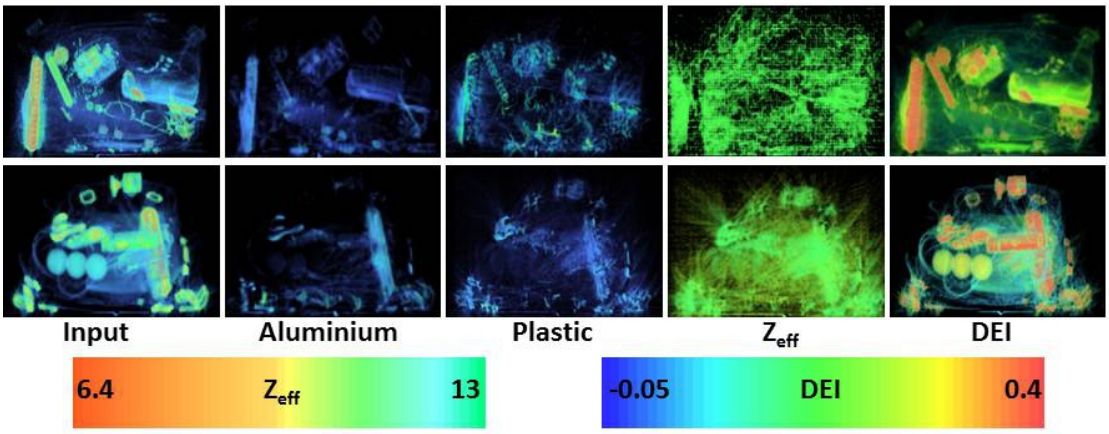


Figure 6.6: Results of 3D DECT decomposition and DEI computation without metal artefact reduction (volumetric visualisations).

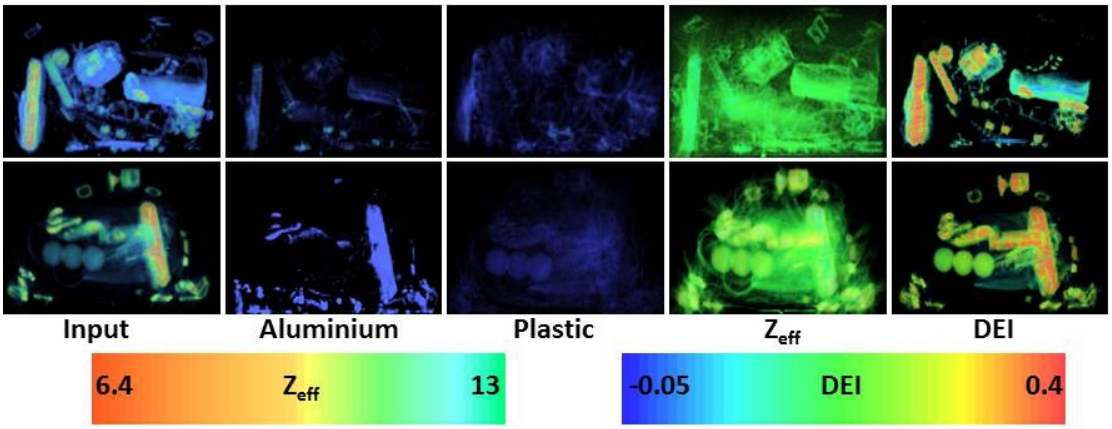


Figure 6.7: Results of 3D DECT decomposition and DEI computation with metal artefact reduction (volumetric visualisations).

the object types and contain very little noise and/or artefacts, suggesting strong discriminative capabilities.

Figures 6.8 - 6.10 display the results of the quantitative histogram analyses of the various reconstructions with and without metal artefact reduction. Figure 6.8 displays the histograms for the selected objects based on the initial high-energy CT values only. The streaking artefacts make the separation of objects, based on intensity alone, challenging (Figure 6.8 (a)). This is especially true for the low density objects (bottles, clothing and shoes), which exhibit very similar intensities. As expected, the predominantly metallic objects (bullets, handguns and grenades) are comparatively easy to distinguish. The application of MAR has a significant impact on this separability. Two positive factors are worth noting: firstly, the MAR procedure has resulted in a decrease in the mean intensities of the low density objects (represented by a shift to the left on the x-axis in Figure 6.8 (b)) - indicating the removal of high intensity streaking and secondly, the overlap between the shoes, clothing and bottle histograms has been reduced. Despite

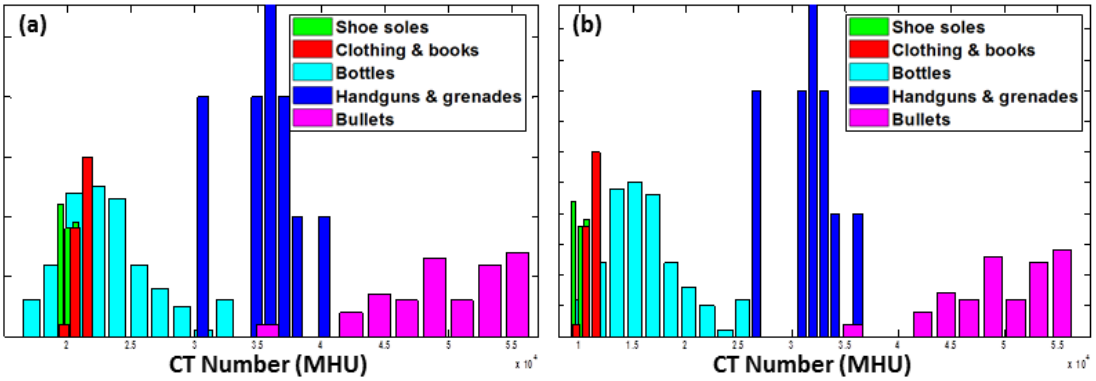


Figure 6.8: Histogram analysis using high-energy intensities only: a) without MAR; b) with MAR.

this reduction, the overlap has not been removed entirely, indicating that the separation of the low-density objects will remain a non-trivial task.

The results of the DEI computations are shown in Figure 6.9. The accuracy of the DEI computations may best be determined by evaluating the DEI of each object type relative to that of water (DEI of water ~ 0) - materials with atomic numbers less than that of water should return negative values, while materials with atomic numbers greater than water should result in positive DEIs. Therefore, under the assumption that the chemical characteristics of the liquids in the test images approximate those of water, it is expected that the liquid DEIs be approximately zero, the shoes and clothing DEIs less than zero and the handguns, grenades and bullets DEIs greater than zero. Prior to the application of MAR, the DEIs of liquids are centred around approximately 0.05 (Figure 6.9 (a) and (c)). While the DEIs of both the low-atomic-number objects (shoes, clothing) and the high-atomic-number objects (handguns, grenades and bullets) are correctly positioned relative to that of water, the fact that all the values are greater than zero is an indication that the streaking artefacts have resulted in a slight overestimation of the DEIs. As the various object groupings do contain common materials, some overlap in their DEIs is expected. Nonetheless, the degree of overlap in Figure 6.9 (a) and (c) indicate that separation of the selected object classes based on the DEI will be challenging. The application of MAR again has a positive impact on the results (Figure 6.9 (b) and (d)). The mean of the liquid DEI histogram is nearer to 0 and both the low and high atomic number object classes are better separated, indicating that the DEI may be a valuable measure in object classification and/or segmentation tasks. An important advantage of the DEI approach (compared to the DECT decomposition) is the low computational demand - since both the low and high-energy images are available by default, the DEI computation requires only simple arithmetic and is thus performed in real-time.

Finally, Figure 6.10 illustrates the results of the DECT decomposition and subsequent effective atomic number reconstructions. As indicated by the poor

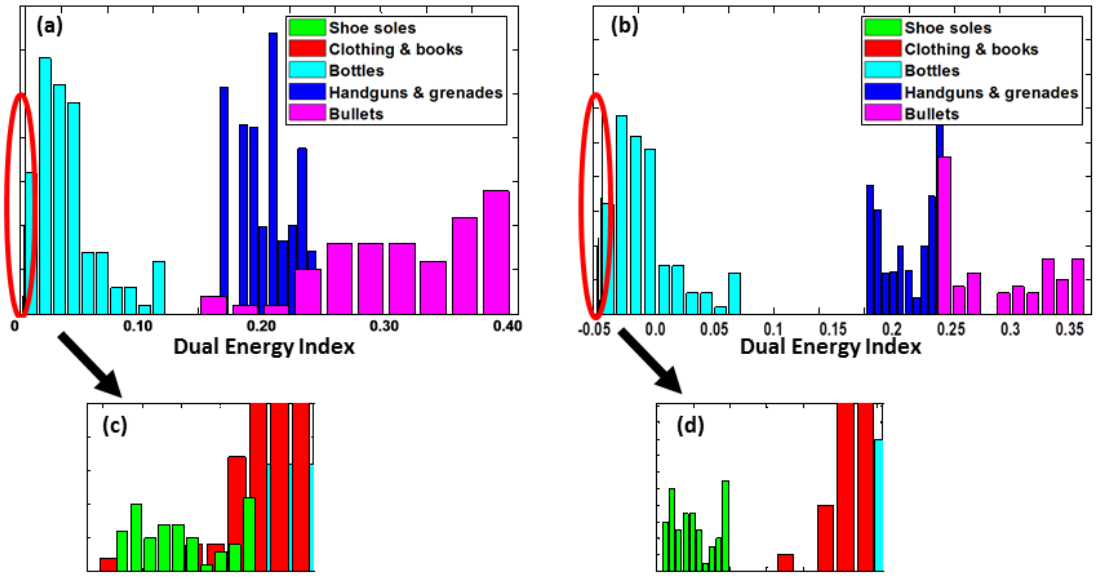


Figure 6.9: Histogram analysis using dual-energy index: a) without MAR; b) with MAR. Magnified regions shown in (c) and (d).

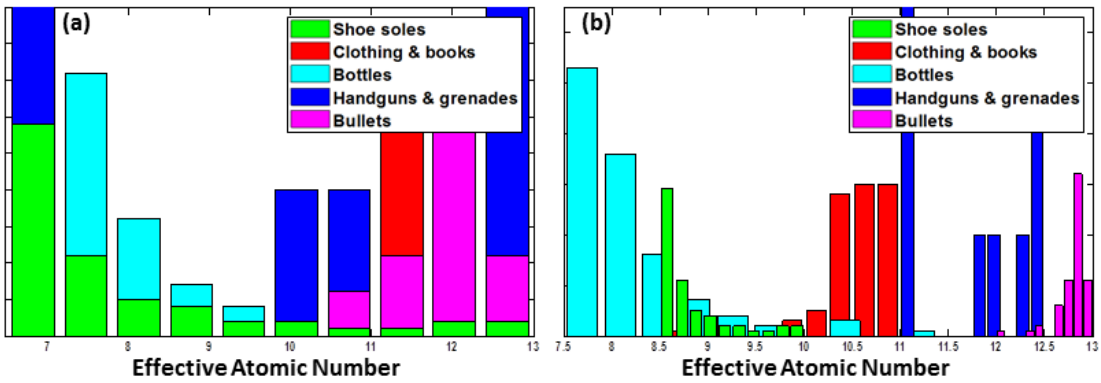


Figure 6.10: Histogram analysis using effective atomic number: a) without MAR; b) with MAR.

quality Z_{eff} images in Figure 6.4, the effective atomic numbers computed from images without applying MAR provide little discriminative power. In fact, there does not appear to be any meaningful order to the computed numbers, with almost all object classes containing atomic numbers spanning the entire range of possible values. While the application of MAR does lead to some improvement, the results are generally poor. Most apparent are the overestimation of the atomic numbers of the shoe soles and clothing items. These object classes are expected to be dominated by plastic in the decomposition process and should thus generate effective atomic numbers towards the lower end of the scale $Z_{\text{eff}} \in [6.47, 13]$. The actual values computed (especially for clothing and books - mean of approximately 10.7) are almost certainly too high. The Z_{eff} measure, as computed in this study, thus appears to contain little discriminative power. It may be argued that the bullets and handguns classes (which are correctly classified by high Z_{eff}) are exceptions to this conclusion. It is worth noting however, that these object classes are charac-

terised by considerably higher CT numbers relative to the other object classes and are thus fairly easily distinguishable based on intensities alone (Figure 6.8). As discussed in Sections 3.4.2 and 3.4.4, DECT decomposition requires an intensive calibration procedure, a table look-up process as well as two FBP reconstructions (which is the most computationally-intensive component of traditional CT imaging). Consequently, the computation of the Z_{eff} images requires considerably more computational power and longer processing times.

6.3 Conclusion

This chapter has provided an experimental investigation into the efficacy of pre-reconstruction and post-reconstruction DECT techniques when applied to the low-resolution, cluttered baggage-CT data obtained from the Reveal CT-80DR scanner. In particular, the potential discriminative capabilities of the previously unconsidered post-reconstruction Dual-Energy Index (DEI) and the traditional pre-reconstruction effective atomic number Z_{eff} have been evaluated on five manually-annotated object classes (with differing chemical characteristics). Furthermore, the impact of metal artefact reduction on each technique has been evaluated. Due to the shortage of empirical data, the DECT decompositions have been performed using simulated energy spectra and approximated scanner configurations.

While absolute quantitative analysis has not been possible, due to the lack of ground-truth data, a rough analysis has been performed using histogram analysis. It is important to emphasise that the precise contents of each of the test images are not known - while several of the objects present in the scans are easily identifiable (e.g. handguns and shoes), others range from challenging to virtually impossible. It has thus not been plausible to make a definitive conclusion on the accuracy of the reconstructions using the available information. At best, the results give an indication of the likely benefit of each measure in subsequent processing.

Considering the broader objective of this work (automated segmentation and classification), the most important observations of the aforementioned experimentation have been:

- The dual-energy index outperforms the effective atomic number both in terms of (estimated) accuracy as well as discriminative power.
- Performance is improved with metal artefact reduction.
- Errors resulting from the approximations of the scanner configuration and the use of simulated energy spectra appear to be the predominant cause of the comparatively poor performance of the Z_{eff} measure.

It is likely that this work will benefit from performing calibration using at least some empirical data (e.g. actual scans of materials with known properties and/or

the measured energy spectra of the scanner). Alternatively, as is indicated by [SM12], an iterative reconstruction DECT approach is likely to produce superior results. These alternatives are left as areas for future work.

A novel 3D volumetric-CT segmentation technique, which incorporates the DEI-based methods considered here, is presented and evaluated in Chapter 7 of this work.

Chapter 7

Segmentation

The broader objective of this work is the development of a fully-automated framework for the classification of objects in volumetric baggage-CT imagery. The current state-of-the-art in this regard relies on the manual generation of subvolumes containing at most a single target object [FBM12, FBM13]. This chapter presents the development and implementation of a segmentation framework to automate the generation of these subvolumes.

The dependence of the majority of the state-of-the-art medical-segmentation techniques on *a priori* information (Section 3.5.1) detracts from their suitability to the aviation security-screening domain, where the segmentation of multiple, unknown objects is required. This has been substantiated by the comparative work of Megherbi *el al.* [MBFM13], where popular medical-segmentation techniques are shown to perform poorly on baggage-CT data. The blind segmentation of unknown objects in cluttered volumetric imagery is considerably more challenging and comparatively few solutions exist. The majority of the prior work in this domain [Gra06, WGW12] has originated from the US Department of Homeland Security's ALERT baggage-segmentation initiative [CMP13] which considered the segmentation of 3D volumetric baggage-CT scans. The ALERT initiative led to the development of five dedicated baggage-CT segmentation techniques. Although each technique demonstrated high-quality segmentations, the study considered single-energy, medical-grade CT imagery, with comparatively low levels of noise and metal-streaking artefacts [CMP13]. Such data is not representative of that encountered in the current security-screening domain. The segmentation of low, anisotropic resolution volumetric baggage-CT imagery in the presence of multiple metal objects (such as that considered in this thesis) has not been considered previously.

Drawing from these prior works and from the dual-energy techniques presented in Chapter 6, a novel dual-energy-based segmentation technique is presented in this chapter. Within the proposed framework, four novel contributions are made:

1. A materials-based coarse segmentation technique using the Dual-Energy In-

dex (DEI) [Joh11] and connected component analysis.

2. A random-forest-based model for measuring the quality of individual object segments, which is used to guide the segmentation process.

3. A random-forest-based model for measuring the quality of entire segmentations.

4. An efficient segmentation-refinement procedure for splitting fused objects.

The segmentation framework is shown to produce fast, high-quality segmentations of low-resolution volumetric baggage-CT images. Based on previous performance and available information, three techniques (two region-growing [WGW12, WH03] and one graph-based [Gra06]), derived from the ALERT initiative [CMP13], are selected as benchmarks for comparative evaluations. The proposed approach is shown to outperform both region-growing-based methods in terms of segmentation quality and speed, but produces lower-quality segmentations than the graph-based approach [Gra06]. These do however, come at a notable reduction in processing time.

Portions of this chapter have been submitted for publication and are currently under review [MB14b].

7.1 A Definition of Image Segmentation

Although the precise objective of any given segmentation algorithm is dependent on a number of problem-specific variables, its core is typically based on some variant of the following formal definition of image segmentation.

Consider a digital image I defined in an N-dimensional discrete coordinate system $\mathcal{Z}^n$ such that $I \subset \mathcal{Z}^n$. The objective of image segmentation is then to partition the image into M disjoint Regions Of Interest (ROIs) such that the segmented image S takes the form [WH03]:

$$S = \bigcup_{i=1}^{M} R_i, \quad \text{where } R_i \cap R_j = \emptyset \quad \text{for } i \neq j \tag{7.1}$$

where:

- R_M is reserved for the background image region and is set to zero in S.
- Each ROI $R_i, i \in [1, M)$, consists of a only one connected component.
- Individual regions are distinguished by region labels.

As image-segmentation algorithms have been developed for a wide range of applications, the definition of what constitutes an object (i.e. what it is that needs to be segmented) is problem-specific. This task is inherently trivial in applications which require the segmentation of pre-specified objects (e.g. the segmentation of specific organs from medical images). In this work, however, a segmentation algorithm that does not require knowledge of the quantity and/or physical properties of the objects to be segmented, is required. Considering the particular domain of baggage security-screening, there are essentially two possible *object-philosophies* [CMP13]:

1. **Segment-all:** requiring the segmentation of all the objects in a scan (a universal definition of an object is then required).

2. **Segment-threats:** requiring the segmentation of threat-like objects only. This allows for a simpler and more precise definition of an object, based upon some prior knowledge of common threat items.

The task of defining an object is complicated further when considering objects composed of multiple parts and deciding whether or not the object should be segmented as one or into its constituents (e.g. bottle containing liquid). Several techniques (e.g. splitting and merging [WH03], hierarchical object representations [WGW12] etc.) have been presented to address such ambiguities and are discussed in further detail in Section 3.5.

The current state-of-the-art in baggage-CT-image classification [FBM12] relies on the manual generation of subvolumes containing at most a single target object. The segmentation techniques developed in this chapter seek to automate this procedure by segmenting every object in a given scan and generating a corresponding subvolume for each.

7.2 Proposed Segmentation Algorithm

The review of the literature in Section 3.5 has indicated that many popular segmentation techniques (particularly in the medical-imaging domain) operate on an initial *coarse* segmentation of the foreground (where the foreground refers to those image regions/voxels that require segmenting) [HHR01, UR05, SPvG05]. A similar framework has been successfully applied to the task of baggage-CT segmentation [GSK+12]. Particularly, connected-component analysis is performed on an initial estimation of the foreground. Each of the connected components in the foreground is then iteratively separated into individual objects using a global splitting algorithm and a segmentation-quality measure. The combination of all of the individual objects constitutes the final segmentation (Equation 7.1). A similar segmentation framework, consisting of three components, is proposed here:

1. Coarse segmentation/foreground determination.

2. Segmentation refinement.

3. Segmentation quality measure.

7.2.1 Coarse Segmentation

Simple intensity thresholding is typically sufficient to provide an initial coarse segmentation for the majority of medical applications [HHR01, UR05, SPvG05]. A greater degree of clutter, noise and artefacts, however, detracts from the suitability of this approach in the baggage-CT domain. Grady *et al.* [GSK+12] apply a Mumford-Shah functional [GA09] to an artefact-reduced input image [Tuy93] to generate a binary foreground-mask (where object voxels are labelled as 1). It has been demonstrated that the availability of dual-energy data provides a means for a coarse separation of the objects in a scan based on their material characteristics (see Chapter 6). In particular, in the absence of dual-energy calibration data, the use of the Dual-Energy Index (DEI) [Joh11] has been shown to exhibit the most potential in terms of discriminative power. Based on these observations, a DEI-based coarse segmentation is computed in the following way.

The Metal Artefact Reduction (MAR) technique described in Chapter 4 and [MMB+13] is applied to the low and high-energy CT images. The artefact-reduced images are then used to compute the DEI image using the techniques described in Chapters 6 and 3 (Equation 3.5). The observations made in Chapter 6 indicate that the objects in a scan are typically well separated by their DEI. The DEI image is thus subjected to a multiple thresholding procedure:

$$I_{\tau_i} = \begin{cases} 1 & \text{if } \tau_{i-1} \leq I_{\text{dei}} \leq \tau_i \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \ldots, N_\tau \tag{7.2}$$

where the number of thresholds N_τ is a user specified parameter. Connected component analysis is then performed on each of the N_τ thresholded images. All connected components smaller than a predefined minimum object size are discarded. The remaining connected components are assigned individual labels and represent the image foreground to be passed to the segmentation-refinement algorithm. While it is likely that an improvement in the quality of the coarse segmentations will result from a case-by-case fine-tuning of the DEI thresholds, to ensure automation and maintain low processing times, a predefined, constant set of thresholds are used. A constant uniformly-space threshold range is chosen to optimise processing times and to ensure that consistent material types are segmented across all images.

The final segmented image is obtained by refining (i.e. further segmenting) each of the coarsely segmented components. In order to achieve this refinement, a measure of segmentation quality is required.

7.2.2 Segmentation Quality Measures

The segmentation-quality measure is intended to provide a quantification of the likelihood that a given segmentation represents a single object (i.e. does not require further segmentation). Various segmentation evaluation metrics were discussed in Section 3.5.3. For the purposes of online segmentation evaluation (i.e. evaluation in the absence of ground-truth data), the feature-based generative model of 'high-quality' segmentations presented by Kohlberger *et al.* [KSA+12] has been met with success in related studies [GSK+12].

More specifically, Grady *et al.* [GSK+12] present the Automated QUality Assessment (AQUA) measure (see Section 3.5.3) to quantify segmentation quality and control the splitting of connected components within their graph-partitioning algorithm. Using a variety of segmentation metrics as features [GSK+12, KSA+12], the AQUA module is trained (using a Gaussian Mixture Model (GMM)) on a training set of *good* object features to recognise high-quality segmentations. The model is shown to provide meaningful separation of good object segments (i.e. high likelihood of consisting of only a single object) and bad object segments [CMP13].

As an equivalent measure is desired here, the suitability of the AQUA measure, using the Reveal CT-80DR dataset, and the 42-dimensional feature vector described in [KSA+12, GSK+12], is examined. The feature vector quantifies the shape and appearance of object segments by computing 42 shape and appearance attributes falling into one of five categories: 1) unweighted geometric features (quantifying the size and regularity of the segmentation); 2) weighted geometric features (locally emphasising the geometric features when intensity values are similar to each other); 3) intensity features (measuring absolute intensity and intensity distributions within segmentations); 4) gradient features and 5) ratio features (computed as ratios of previously computed features). The proposed feature attributes have been inspired by metrics used in prior segmentation-based literature [GSK+12, KSA+12]. For example, the geometric features are employed in several early segmentation studies [MS89]; the concept of weighted geometric features was first proposed by Caselles *et al.* [CKS97] in their work on geodesic active contours and several of the ratio features are variations of the *cut-over-volume* ratio, which has been used extensively as an objective function in graph-cut-based segmentation algorithms [SM00, GS06a]. For a detailed description of the features used, the reader is referred to Appendix B.

Similarly to [GSK+12], the aforementioned features are extracted from a set of manually segmented good object segments (i.e. containing only a single object

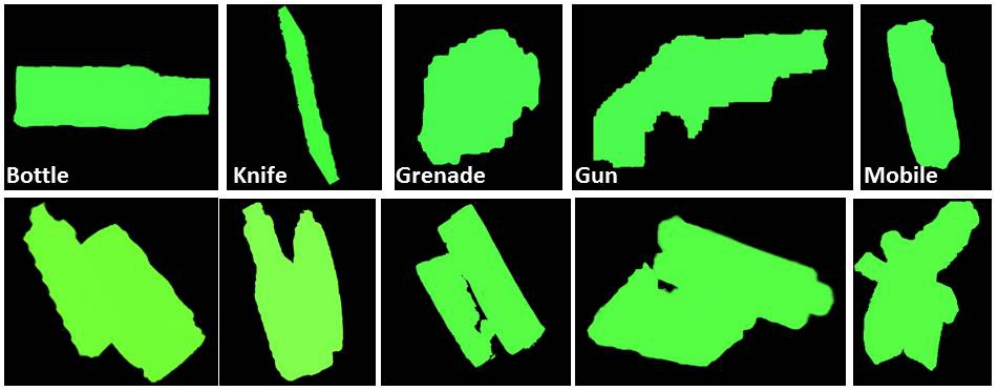


Figure 7.1: Object segment mask examples. Top row: manually generated single-object segments containing a single target object only. Bottom row: manually generated multi-object segments containing two or more connected objects (these would require further partitioning).

each). A large range of objects are included in this training set (see Figure 7.1) . Each feature is normalised by subtracting the mean over the entire feature space and dividing by the standard deviation. Principal Component Analysis (PCA) [JMF11] is then applied to reduce the dimensionality of the feature space. The GMM is fitted over the PCA coefficients of all the segments in the training set using the Expectation-Maximisation (EM) algorithm [Moo96]. When fitting statistical models, the likelihood (or accuracy) improves with increasing model complexity (i.e. by adding parameters). Increased complexity, however, carries the risk of overfitting. Statistical model selection criteria (used to select the most suitable model parameters) such as the Akaike Information Criterion (AIC) and the Bayesian (or Schwarz) Information Criterion (BIC)[MP04] compensate for the risk of overfitting by introducing a penalty term, which penalises the complexity of the model. The cost functions of the AIC and BIC are both composed of two terms: 1) a log-likelihood or accuracy term (favouring complexity) and 2) a penalty term (penalising complexity):

$$\mathrm{AIC} = -2\ln\hat{L}(\hat{\theta}) + 2m \tag{7.3}$$

$$\mathrm{BIC} = -2\ln\hat{L}(\hat{\theta}) + m\ln(n) \tag{7.4}$$

where $\hat{L}(\hat{\theta}) = p(x|\hat{\theta}, M)$ is the maximised log-likelihood of the model M for the parameter set $\hat{\theta}$ and the observed data x; m is the number of model parameters to be estimated and n is the number of observations in x. In determining the order of a mixture model, it is known that the AIC is order-inconsistent and thus tends to overestimate the correct number of components (i.e. overfits) [KM88, CS96]. The BIC penalises complexity more heavily than the AIC and tends to better avoid overfitting [KM88]. For these reasons, the BIC is used here to determine

the optimal GMM parameters. In determining the optimal GMM, the number of Gaussians and the covariance matrix type (diagonal or full-rank) are varied. The model minimising the BIC cost function (Equation 7.4) is selected. Finally, the AQUA measure of a given object segment (i.e. partition) $S \subseteq V$ (where V is the volume) is given by [GSK+12]:

$$\text{AQUA}(S) = \sum_{i=1}^{10} w_i \mathcal{N}(f(S); \mu_i, \Sigma_i) \tag{7.5}$$

As an alternative to the GMM-based generative model, the efficacy of a discriminative model to distinguish between good and bad object segments is evaluated. To this end, a random-forest classification model [BZM07] (Section 3.3.2) is used to provide a probabilistic classification of good and bad object segments. For the purposes of building a discriminative model to distinguish between good and bad object segments, a random forest composed of binary classification trees is employed. The forest is constructed in the same manner as the classification forest presented in Chapter 5. Training data points are represented by fixed-length feature vectors extracted from a set of manually segmented good and bad object segments. Trees are constructed in a top-down recursive manner using a simple thresholding function as the node split function for all internal nodes. The optimality criterion used for node splitting is the classical Information Gain (IG) [Cri11]. Randomness is injected into the trees via random node optimisation, whereby a random subset of the available node test parameter values is considered at each node. Trees are grown to a maximum depth D_T and leaf nodes are generated if the information gain falls below a minimum threshold IG_{min}. The value for D_T is chosen empirically using a small validation set (thereby avoiding fully grown trees). The quality of any given segmentation - denoted the Random Forest Score (RFS) - may then be computed by averaging the corresponding posterior probabilities of each of the leaf nodes reached in the forest:

$$\text{RFS} = p(c|\mathbf{v}) = \frac{1}{T} \sum_{t=1}^{T} p_t(c|\mathbf{v}) \tag{7.6}$$

where T is the number of trees in the forest; $p(c|\mathbf{v})$ is the estimated conditional probability that a given test data point $\mathbf{v}$ belongs to the class c and c is a discrete class label (i.e. $(0, 1) \rightarrow$ (bad,good)). The distribution is conditional on the specific leaf node reached by the data point [Cri11].

In the random forest model, data points (i.e. object segments) are represented by descriptor vectors. The 42-dimensional feature vector described in [KSA+12] has been shown to provide a good representation of segmentation quality - both

in the context of baggage segmentation [GSK+12] as well as in the evaluation of segmentation error in the absence of ground-truth [KSA+12]. Based on its previous success and its relative simplicity, the efficacy of this feature vector when used for the representation of object segments in the random forest model is investigated here. Furthermore, it is expected that the characteristics of the surface of a single-object segment will differ from that of a segment representing multiple objects. Based on this assumption, it may be possible to distinguish between single-object (good) and multi-object (bad) segments using a description of the object shape. 3D shape-based descriptors have been successfully applied to a variety of similar object-recognition, retrieval and classification tasks [YHQ04, ZdFFeR07, MFB10]. Based on these prior works, the following three 3D shape-based descriptors are evaluated in addition to the 42-dimensional feature vector used in the AQUA measure [GSK+12]: 1) 3D Zernike descriptors [NK04]; 2) the Histogram-of-Shape Index (HSI) [DJ95] and 3) a hybrid 3D shape descriptor [MFB10].

Several recent studies have demonstrated the efficacy of 3D Zernike descriptors for characterising 3D shape, particularly in proteomics (the study of proteins and their structures) [NK04, VCK+09, SMPW, KSCER11]. Novotni and Klein [NK04] developed the 3D Zernike descriptors by expanding upon the mathematical concepts of 3D Zernike moments as laid out by Canterakis [Can99]. The 3D Zernike polynomials are a set of basis functions that are orthogonal on the unit sphere. The 3D Zernike descriptor is an extension of the spherical harmonics-based descriptors of Kazhdan *et al.* [KFR03] and have been shown to be compact, robust to noise and invariant to rotation [NK04]. Canterakis [Can99] first introduced the concept of 3D Zernike moments as a means for describing objects in a 3D Cartesian coordinate system (as opposed to a spherical coordinate system, as used by spherical harmonics [KFR03]). By combining a set of radial basis functions with spherical harmonics, the n^{th}-order 3D Zernike functions may be computed as [Can99]:

$$Z_{nl}^{m}(r,\theta,\phi) = R_{nl}(r)Y_{l}^{m}(\theta,\phi) \tag{7.7}$$

where Y_l^m are complex valued spherical harmonics defined on the spherical coordinate system given by (θ,ϕ); n,l,m are integers such that $|m| \leq n$ and $n-|m|$ is even and $R_{nl}(r)$ are orthogonal radial basis polynomials [VCK+09]. Equation 7.7 may be rewritten in Cartesian coordinates:

$$Z_{nl}^{m}(\mathbf{x}) = \sum_{v=0}^{k} q_{kl}^{v}|\mathbf{x}|^{2v}e_{l}^{m}(\mathbf{x}) \tag{7.8}$$

where $2k = n-l$ and $e_l^m = r^l Y_l^m(\theta,\phi)$ are the harmonic polynomials as

defined by [KFR03]. The coefficients q_{kl}^{v} are computed such that the functions are orthonormal in the unit sphere [NK04]. The Zernike moments are then the projection of a given 3D shape function onto this orthonormal basis:

$$\Omega_{nl}^{m} = \frac{3}{4\pi} \int_{||x||\leq 1} f(\mathbf{x}) \overline{Z_{nl}^{m}(\mathbf{x})} d\mathbf{x} \tag{7.9}$$

where the voxelised 3D shape function $f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^3$ is a binarised representation of the object surface, defined on a regular cubic grid. While these moments are *not* invariant under rotations, Novotni and Klein [NK04] achieved rotational invariance by collecting the moments into $(2l+1)$-dimensional vectors: $\Omega_{nl} = (\Omega_{nl}^{l}, \Omega_{nl}^{l-1}, \ldots, \Omega_{nl}^{-l})^t$. The rotationally invariant, 3D Zernike descriptors F_{nl} are subsequently defined as the norms of the vectors Ω_{nl} [NK04]:

$$F_{nl} = ||\Omega_{nl}||_2 = \sqrt{\sum_{m=-l}^{l} (\Omega_{nl}^{m})^2} \tag{7.10}$$

For a detailed description of the mathematical bases of 3D Zernike moments and descriptors, the reader is referred to the original works of [Can99] and [NK04].

Dorai and Jain [DJ95] present the Histogram-of-Shape Index (HSI) for the representation and recognition of arbitrarily curved rigid 3D objects. The Shape Index (SI) is a scalar-valued quantitative measure of the shape of a surface at a point p [DJ95]:

$$\mathrm{SI}(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{7.11}$$

where κ_1 and κ_2 $(\kappa_1 \geq \kappa_2)$ are the principal curvatures of the surface at the point p. Given a smooth, plane unit-speed (parametrised) curve $\gamma(t)$, where t is the arc-length, the curvature is defined as:

$$K(t) = ||\ddot{\gamma}|| := \frac{1}{r} \tag{7.12}$$

where r is the radius of the osculating (kissing) circle. If one defines a normal vector field $\mathbf{n}$ along the curve γ, the curvature is denoted as positive when the curve turns in the same direction as the normal vector and negative otherwise. If M is defined as differentiable surface in $\mathbb{R}^3$, then at each point p on M, the surface has two principal curvatures κ_1 and κ_2. These are computed in the following way [DZ11]:

1. Let $\mathbf{n}$ be the unit normal vector to M at p.
2. Determine the normal plane $P \in \mathbb{R}^3$ containing $\mathbf{n}$.
3. Define the unit-speed curve: $\gamma_P(t) = P \cap M$ (i.e. the plane curve resulting from the intersection of P and M).
4. Compute the curvature K_P of $\gamma_P(t)$ according to Equation 7.12.
5. The principal curvatures are then defined as $(\kappa_1, \kappa_2) = (\min_P K_P, \max_P K_P)$

According to this definition, every distinct shape may be mapped onto a unique value in the interval SI $\in [0, 1]$. The exception to this is the planar shape, for which SI is undefined since $\kappa_1 = \kappa_2 = 0$ for all points on a planar surface [DJ95].

Megherbi *et al.* [MFB10] propose combining the HSI and Zernike descriptors (by direct concatenation) yielding a hybrid 3D shape descriptor. The proposed descriptor demonstrates potential in the classification of threats in CT-baggage imagery and is included in the evaluation here.

Prior to the extraction of the aforementioned shape-based features, the object segments are pose-normalised to ensure invariance to changes in scale and translation. This is achieved by translating and rescaling (voxel resampling by nearest-neighbour interpolation) each object segment based on its approximate minimum bounding box within the original CT image [MFB10].

7.2.3 Segmentation Refinement

Based on the results of preliminary experimentation regarding the aforementioned segmentation-quality measures, a simple yet efficient technique for refining the initial, coarse segmentation using a random-forest-based approach is proposed. The Random Forest Score (RFS) is computed for each of the N_c components (or objects) in a given coarse segmentation. Those components yielding an RFS below a given threshold τ_{RFS} are considered to be composed of multiple objects and are passed to a partitioning algorithm. Coarse components with RFS$> \tau_{RFS}$ are left unchanged and assigned a unique label in the final image.

Poor-quality objects (RFS$< \tau_{RFS}$) are partitioned at the estimated intersection (or touching) points of the multiple objects comprising the given segment. These points are found by detecting the perimeter voxels of the original object that are likely to be common to two objects. Non-zero (i.e. object) voxels are assumed to lie on the perimeter of the object if they are connected (see voxel connectivity relationships in Figure 7.2) to at least one zero (i.e. background) voxel. The assumption is made that those voxels corresponding to the intersections of multiple objects will be surrounded by higher numbers of object voxels compared to regular (non-intersection) perimeter points [CMP13]. The total number of object voxels

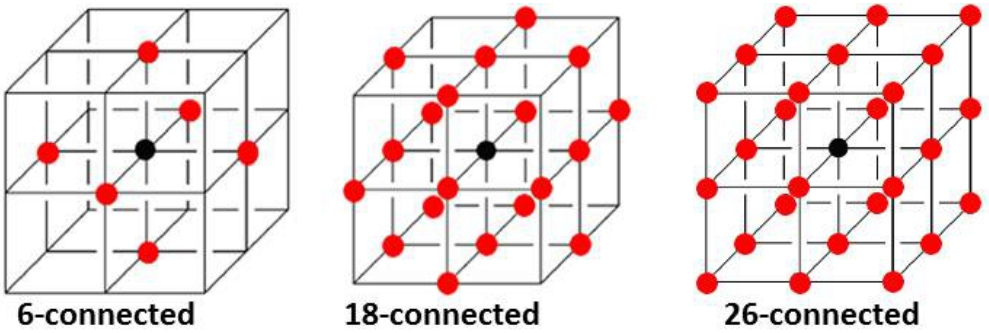


Figure 7.2: 3D voxel connectivity relationships used to map volumetric images onto lattice graphs.

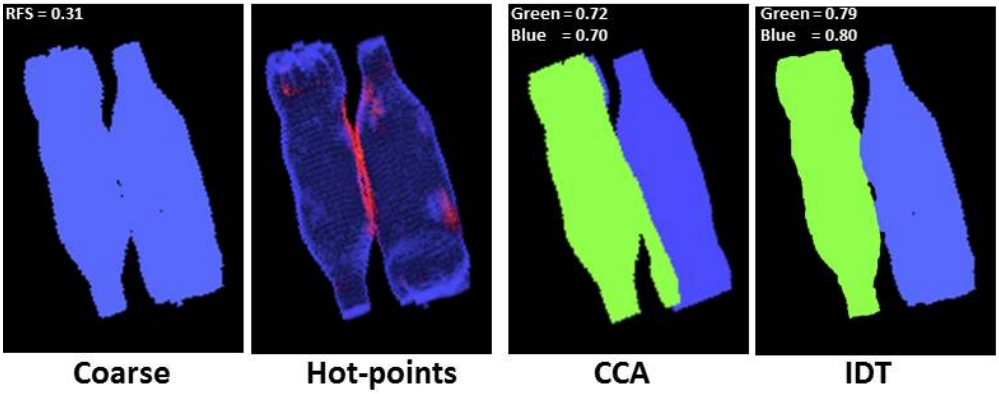


Figure 7.3: Segmentation refinement examples using CCA (Section 7.2.3) and IDT [GSK+12] algorithms. Hot-points (red) and RFS indicated.

in a predefined local cubic neighbourhood ($11 \times 11 \times 11$) of each perimeter voxel is thus determined. If this number is greater than a predefined threshold, τ_{HP}, the perimeter voxel is considered to be an intersection point [CMP13] (red points in Figure 7.3). For a given object, this analysis may result in multiple clusters of such points (denoted *hot-points*), in which case it is assumed that the object requires splitting at multiple regions. Each cluster of hot-points is considered individually. It has previously been suggested that splitting of touching objects may be performed by fitting a plane (e.g. by RANSAC [FB81]) to such hot-points [CMP13]. Such planes, however, are likely to intersect the object at multiple regions (not just at the locations of the hot-points) leading to over-segmentations. Restricting the planes to local regions is challenging, especially when determining which voxels lie above or below the plane. The plane-based approach becomes particularly problematic when an object requires splitting at multiple locations.

A simpler approach is proposed here. A Connected Component Analysis (CCA) is performed on a mask obtained by removal (i.e. setting to zero) of the hot-points. The two connected components returning the highest RFS (computed with the hot-points reinserted) are retained. If the CCA results in only a single connected component, morphological dilation of the zeroed-out region is performed until the CCA returns at least two components. If the RFS of one of

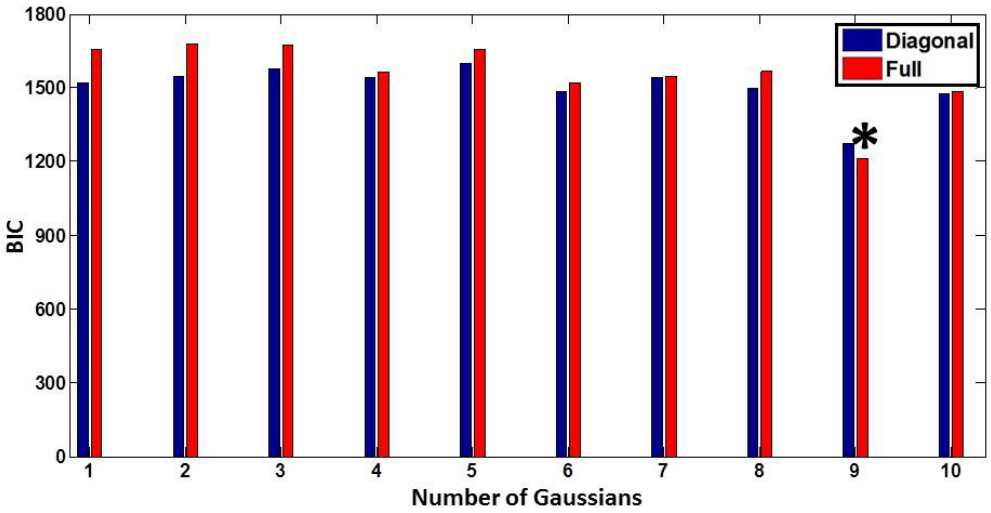


Figure 7.4: Bayesian Information Criteria (BIC) for different Gaussian mixture model parameters. Optimal model using 9 Gaussians and full-rank covariance matrix (indicated by asterisk).

the regions falls below the RFS of the original region, or if the region is smaller than the minimum permissible size of an object, then the region is assumed to be noise or artefact-induced and is discarded. If both components result in a decrease relative to the original RFS, then the original object is retained (Figure 7.3). Although objects split in this way are not guaranteed to produce segments with RFS$> \tau_{RFS}$, only splits resulting in improved scores are permitted. The procedure thus performs both splitting as well as denoising of the coarse segmentations. For objects containing multiple hot-point clusters, the quality of the final split objects are affected by the order in which the clusters are dealt with. As the described splitting procedure is fast and the number of hot-point clusters per object is generally low (≤ 3), the optimal order (i.e. that which results in the individual objects with the highest RFS) may be determined by testing all possible orders.

7.3 Comparative Methodologies

The performance of the segmentation algorithm proposed in Section 7.2 is compared to three segmentation techniques chosen based on their success in related works: 1) the isoperimetric distance tree algorithm [Gra06]; 2) a symmetric region-growing algorithm [WH03] and 3) a 3D flood-fill region-growing algorithm [WGW12].

The graph-partitioning Isoperimetric Distance Tree (IDT) algorithm [Gra06] is evaluated as an alternative to the proposed segmentation-refinement procedure (Section 7.2.3). The IDT algorithm has previously demonstrated success as part of an automated segmentation algorithm for medical-grade CT imagery [Gra06, GSK+12]. As opposed to applying the identical techniques used in [GSK+12] (Section 3.5.2), the IDT algorithm is used as a direct alternative to the splitting algorithm proposed in Section 7.2.3 - that is to say, the algorithm is applied to the

DEI coarse segmentations (Section 7.2.1) and is driven by the random forest quality measure (Section 7.2.2). The reader is referred to Appendix A.1 for a detailed explanation of the IDT algorithm. The IDT algorithm is applied recursively to each individual connected component in the DEI mask until one of three criteria is met [GSK+12]: 1) the input mask has a sufficiently high quality (as determined by the chosen quality measure - Section 7.2.2); 2) the partitioned objects have a quality measure below a given threshold τ_γ and an isoperimetric ratio greater than that of the input mask (i.e. further partitioning produces low-quality segmentations) or 3) the input mask is smaller than the minimum permissible object size. The second criterion is motivated by the fact that the partitioning of a connected component, composed of multiple (fused) objects, may result in two components which are themselves composed of multiple fused objects (which will by definition have low quality scores) [GSK+12]. In order to prevent the early termination of the recursive IDT algorithm in such scenarios, the recursion is only terminated if the quality measure is less than the threshold τ_γ *and* the isoperimetric ratio of the split component is greater (i.e. worse) than that of the input component. Grady *et al.* [GSK+12] propose that, owing to the efficiency of the IDT algorithm, several candidate segmentations for each connected component in a mask may be generated by executing multiple runs of the IDT algorithm with different (randomly selected) reference (ground) points (Appendix A.1) and selecting the highest-quality candidate as the final result.

Song *et al.* [CMP13] applied a multi-stage segment-and-carve algorithm for the segmentation of medical-grade baggage-CT images. As previously discussed (Section 3.5.2), the predominant limitation of the technique is its complexity (owing to the number of separate stages in the algorithm). Consequently, the technique is characterised by a large parameter set which is shown to require careful, case-by-case tuning to produce optimal results. Nonetheless, the core of the technique - the (seedless) Symmetric Region Growing (SymRG) algorithm [WH03] - is a parameter-free efficient region-growing technique which has demonstrated success in a variety of 3D segmentation tasks [WH03, WKRH00]. Owing to its low runtime and its fully-automated nature, the SymRG algorithm is an attractive option for the segmentation of complex 3D imagery and is thus evaluated here. The reader is referred to Appendix A.2.1 for a detailed description of the SymRG algorithm.

The 3D flood-fill region-growing method traverses a volume in a flood-fill manner using a 3D spherical kernel of varying sizes. The seed-points and kernel dimensions are determined automatically based on image content (e.g. local gradients), while the kernel movement criteria for specific kernels are inferred from a set of training examples. The algorithm is composed of five stages: 1) Definition of 3D kernel; 2) Determination of movement criteria; 3) Seed initialisation; 4) Flood-fill and 5) Merging. For a detailed description of each of the stages, the reader is

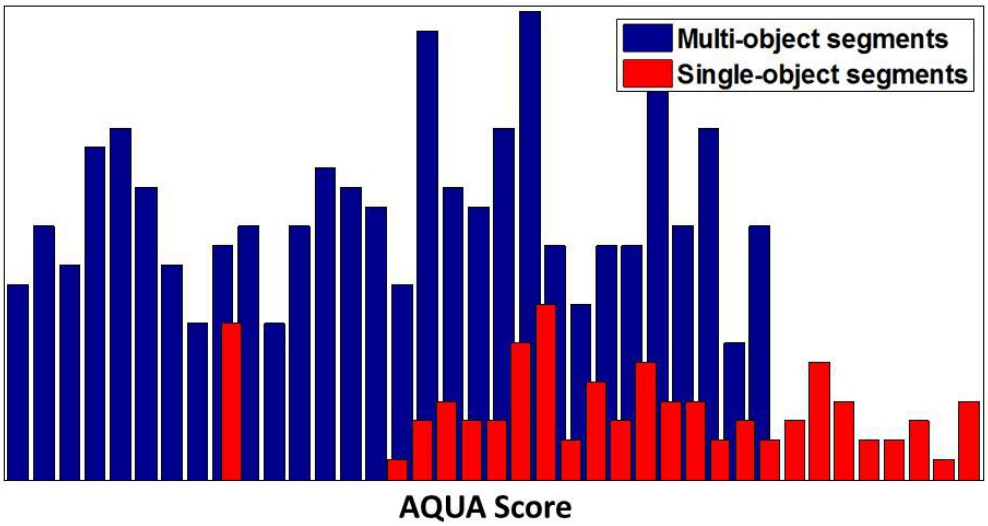


Figure 7.5: Separation of single and multi-object segments using AQUA score [GSK+12]

.

referred to Appendix A.2.2.

As discussed in Section 3.5.3, the most widely adopted approach for the evaluation of segmentation techniques is the comparison of algorithm-generated segmentations to manually delineated ground-truth data using some similarity measure [Zha96, NBVV00]. The manual delineation of volumetric data, however, is a laborious task - particularly when the dataset is large and each data sample contains a large number of objects (as is the case in this study). Consequently, due to the lack of ground-truth data, the aforementioned segmentation techniques are evaluated and compared predominantly in a qualitative manner. As suggested by Grady *et al.* [GSK+12], an overall measure of segmentation quality for a given image may be obtained by averaging the quality scores for each segmented object in that image. It was noted earlier (Section 3.5.3) that this approach may fail in cases where too few objects are segmented from an image. To address this shortcoming, quantitative analysis is performed on a set of test images containing known numbers of objects (no knowledge of the actual object boundaries are required) allowing for the overall segmentation score for a given image to be computed as the average RFS (of each segmented object) multiplied by the error in the number of segmented objects:

$$\mathrm{RFS}_S = \underbrace{\left(\frac{1}{N_S}\sum_{i=1}^{N_S}\mathrm{RFS}_i\right)}_{\text{Average RFS}} \times \underbrace{\left|1 - \frac{|N_T - N_S|}{N_T}\right|}_{\text{Penalty Term (PT)}} \tag{7.13}$$

where N_T is the true number of objects in the image; N_S is the number of segmented objects and RFS_i is the quality score for the object i. Note that $PT \in [0, 1)\ \forall N_S \neq N_T$ and $PT = 1$ otherwise. All segmentations containing the incorrect

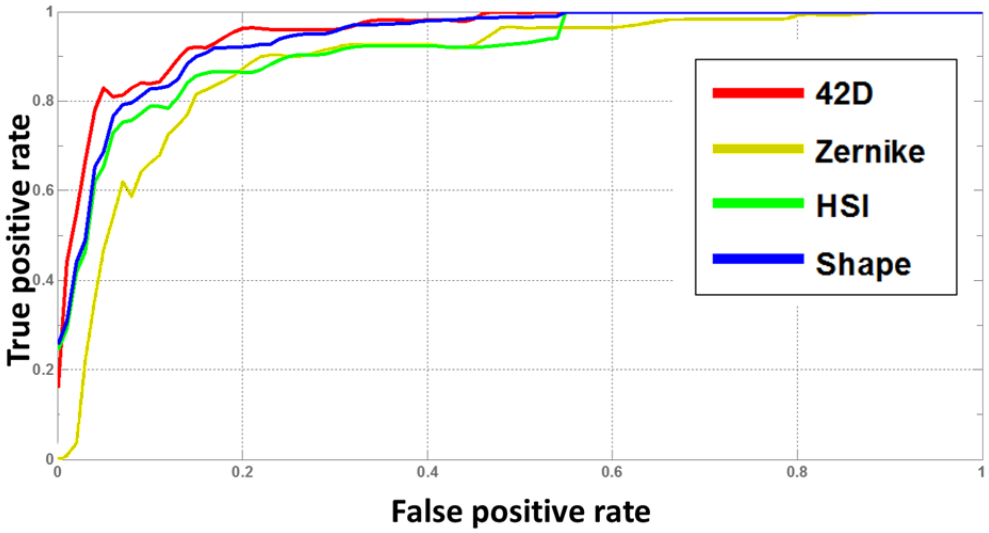


Figure 7.6: ROC curve for leave-one-out cross validation testing of random forest segmentation-quality measures. Conducted as a binary classification experiment - segments are classified as either single-object segments (positive class) or multi-object segments (negative class). The curves were generated by varying the discrimination threshold on the classification posterior.

number of components will thus be penalised. It should be noted that the measure in Equation 7.13 is likely to be biased in favour of the segmentation-refinement methods (as both are inherently controlled by it). Finally, all evaluations are performed with and without metal artefact reduction.

7.3.1 Segmentation Data

Various datasets were used in the experiments presented in this chapter. All images are represented in Modified Hounsfield Units (MHU) and have voxel resolutions of $1.56 \times 1.61 \times 5$mm.

Quality measure evaluations: The GMM used in the AQUA model is built using a training set composed of 80 manually cropped single-object segments. In order to evaluate each of the quality measures (AQUA and random-forest based models) a separate test set containing 194 manually-cropped single-object segments and 415 manually-cropped and algorithm-generated multi-object segments has been created (e.g. Figure 7.1). Algorithm-generated multi-object segments are obtained using the DEI procedure described in Section 7.2.1. The test samples vary in size (depending on the object(s) in the scan).

The performance of the AQUA measure is evaluated by examining the histogram of AQUA scores for each of the samples in the test set. In order for the AQUA measure to be successfully incorporated into the proposed segmentation framework, it is required to provide a good separation between the good and bad object segments in this histogram.

The performance of each of the random-forest-based scores is evaluated by a

Descriptor	AUC	Optimal Operating Point (TPR, FPR)
42D [KSA+12]	0.971	(0.960, 0.098)
Zernike [NK04]	0.863	(0.862, 0.240)
HSI [DJ95]	0.901	(0.871, 0.160)
Shape [MFB10]	0.942	(0.800, 0.036)

Table 7.1: LOO cross validation results for random forest segmentation-quality measures: Area Under (ROC) Curves (AUC) and optimal operating points. Optimal operating points were determined using equal costs for false-positive and false-negative classifications.

Leave-One-Out (LOO) cross-validation procedure. Note that the process is considered a binary classification task where single-object segments represent the positive class. Receiver Operating Characteristic (ROC) curves (computed by varying the discrimination threshold on the forest posterior - Equation 7.6), the Area Under the Curve (AUC) and the optimal (false-positive;true-positive) operating points are computed to illustrate the performance of each of the descriptor types. Finally, the histograms of RFS are generated for each method to illustrate the separation of single and multi-object segments.

Segmentation evaluations: Qualitative analysis of the four segmentation algorithms is performed using four, cluttered whole volume baggage-CT scans obtained on the Reveal CT-80DR scanner (Figure 7.11 (a) - (d)). Quantitative analysis of the four segmentation algorithms is performed using a set of 30 cropped baggage-CT scans obtained on the Reveal CT-80DR scanner. Each of the volumes in the set are cropped such that they contain a known number of objects. All volumes are composed of 512×512 axial slices and the number of slices in each volume ranges from 92 to 112. The random forests used to guide the segmentation refinements in these experiments are built using a set of 80 manually-cropped single-object segments (the same set used to build the GMM) and 80 manually-cropped multi-object segments (a subset of the 415 multi-object test set mentioned above).

7.4 Results

With reference to the notion of object-philosophy introduced in Section 7.1, a 'segment-all' approach (as opposed to 'segment-threats') was adopted in this study. That is to say, all objects meeting specified criteria in a given volume were segmented. In particular, all voxels with intensities lower than a predefined threshold of 1000 MHU were considered to belong to the background and thus set to zero. Additionally, the minimum permissible object volume was set to 50 cm^3 (computed based on the voxel resolution of the data). All objects in the final segmentation

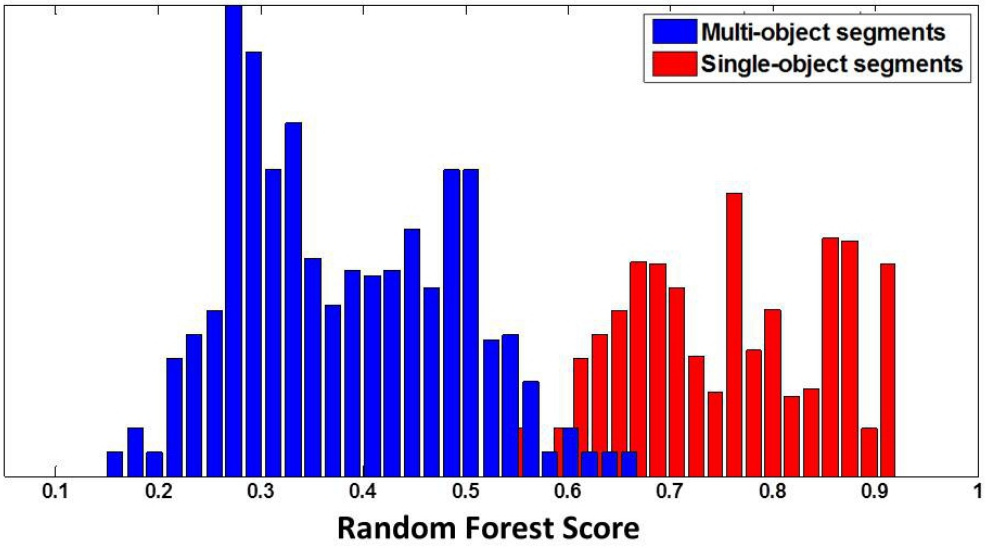


Figure 7.7: Separation of single and multi-object segments using random forest score and 42D descriptor [KSA+12]

.

smaller than 50 cm^3 were thus discarded (set to zero).

Hereafter, the four segmentation algorithms compared (Section 7.3) are denoted as follows: *CCA* - the connected component analysis-based segmentation refinement (Section 7.2.3); *IDT* - the isoperimetric distance tree refinement [GS06a, Gra06] (Appendix A.1); *SymRG* - the symmetric region-growing algorithm [WH03] (Appendix A.2.1) and *FloodFill* - the 3D flood-fill region-growing algorithm [WGW12] (Appendix A.2.2).

7.4.1 Quality Measure Results

In order to evaluate the performance of the AQUA measure when applied to the CT-80DR data, a GMM using 9 Gaussians and a full-rank covariance matrix (see BIC results in Figure 7.4) was fitted over the PCA-reduced feature vectors of the GMM training segments (Section 7.3.1). The number of PCA coefficients retained was selected such that approximately 99% of the feature variance in the training set was retained. The resulting AQUA scores for the 609 object test set (Section 7.3.1) are shown in Figure 7.5. Considerable overlap between the two object classes (single and multi-object) is evident, such that no clear separation boundary can be established. It is unlikely that this formulation of the AQUA measure will lead to satisfactory segmentations using the current dataset.

In the evaluation of the random-forest-based quality measures, the forest parameters were fixed for all feature types. The number of tests performed for each node split was set to $0.7Dim_F$ (where Dim_F is the dimensionality of the feature vector under consideration) - this value was fixed for all nodes in a given forest; trees were grown to a maximum depth of $D = 10$, with a lower bound of $IG_{min} = 10^{-4}$ on the information gain and forests contained 30 trees (see Section

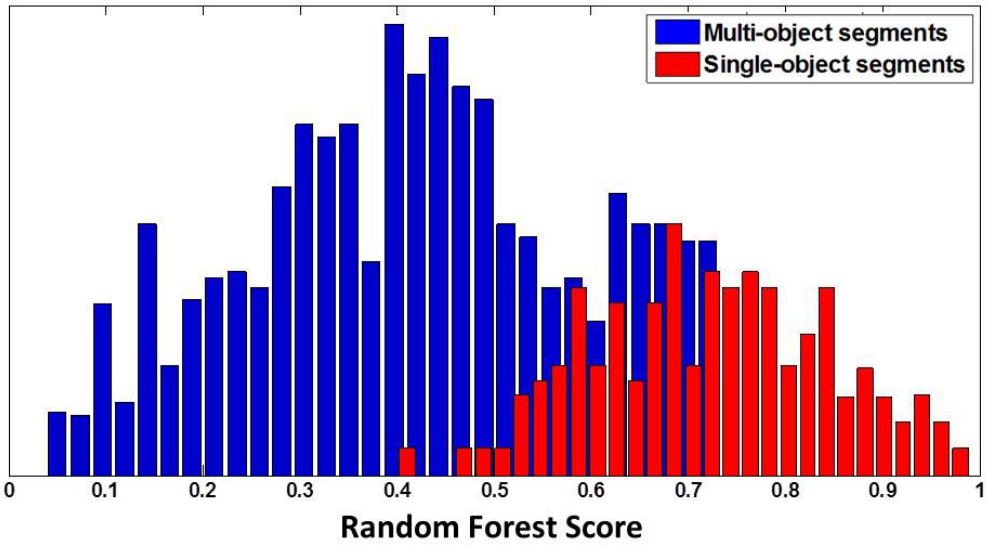


Figure 7.8: Separation of single and multi-object segments using random forest score and 3D Zernike descriptor [NK04]

.

3.3.2 for explanations of these parameters). It was found that using these settings resulted in tree growth terminating prior to maximum depth and thus no tree pruning was performed.

The ROC curves and corresponding AUC and optimal operating points for the LOO evaluation of the random forest quality measures are shown in Figure 7.6 and Table 7.1 respectively. Additionally, the histogram analysis results for each descriptor type are shown in Figures 7.7 - 7.10. Four descriptor types were considered: the 42-dimensional segmentation-based descriptor of Kohlberger *et al.* [KSA+12] (denoted *42D*); the 3D Zernike descriptor [NK04] (denoted *Zernike*); the histogram of shape-index [DJ95] (denoted *HSI*) and the hybrid 3D shape descriptor of Megherbi *et al.* [MFB10] (denoted *Shape*). Based on the results and recommendations of Megherbi *et al.* [MFB10], the 3D Zernike descriptors were generated using a maximal order of 20, yielding a 121-dimensional descriptor. The HSI was computed using a bin-width of 0.005, resulting in a 200-dimensional HSI descriptor. These settings resulted in a 321-dimensional combined 3D shape descriptor.

The 42D descriptor yielded the best LOO cross-validation results (Figure 7.6 and Table 7.1) with an AUC = 0.971 and an optimal operating point on the ROC curve of (0.098, 0.960) - significantly outperforming all 3 shape-based descriptors. All four random-forest-based measures yielded superior separations of the single and multi-object segments (Figures 7.7 - 7.10) compared to the GMM-based AQUA results (Figure 7.5). The 42D descriptor, in particular, resulted in good separation of the classes, despite the relatively high false-positive rate at its optimal operating point (Table 7.1). It is also worth noting that the computation of the 42D descriptor [KSA+12] is considerably less computationally demanding than the

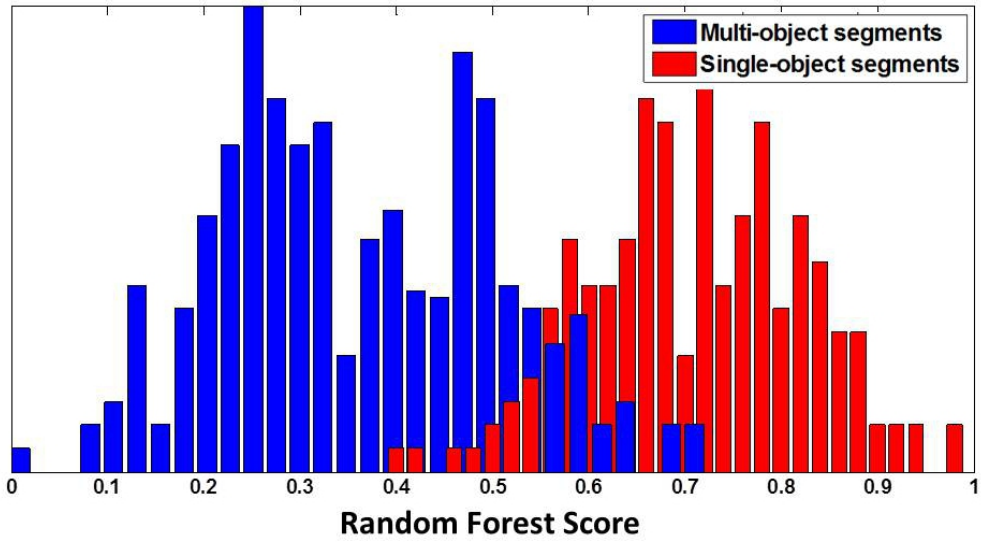


Figure 7.9: Separation of single and multi-object segments using random forest score and HSI descriptor [DJ95]

.

Zernike [NK04], HSI [DJ95] and 3D shape [MFB10] descriptors. Based on the results of this preliminary experimentation, the random forest measure using the 42D descriptor was used in all subsequent evaluations. The optimal operating point for the 42D descriptor occurred at a threshold of 0.73 (i.e. $(\mathbf{v} \in \mathbb{R}^{42})$ = single-object if $p(c|\mathbf{v}) > 0.73$). This threshold was used for τ_{RFS} in the segmentation-refinement procedure (Section 7.2.3).

7.4.2 Segmentation Results

The coarse segmentations were created using $N_\tau = 10$ equally-spaced thresholds. Image refinement using the method proposed in Section 7.2.3 was performed using an RFS threshold of $\tau_{RFS} = 0.73$ and a hot-points threshold of $\tau_{HP} = 300$ (chosen empirically). IDT [Gra06] was implemented using a lattice-connectivity of 6 (Figure 7.2) and a hot-points threshold of $\tau_{HP} = 300$. The optimal value for the quality threshold τ_γ, used as a termination criterion in the recursive IDT algorithm (Section 7.3) was determined empirically by visually comparing candidate segmentations. The flood-fill region-growing algorithm [WGW12] (Appendix A.2.2) was performed using four spherical kernels with radii $r = \{1, 2, 3, 4\}$. The movement polynomial was fitted over 70 training points obtained from 25 separate scans. SymRG was implemented using the following symmetric function:

$$g(p,q) = \begin{cases} \text{TRUE} & \text{if} |f(p) - f(q)| \leq \tau_c \\ \text{FALSE} & \text{otherwise} \end{cases} \tag{7.14}$$

where $f(p)$ and $f(q)$ are the intensities of voxels p and q respectively and τ_c is

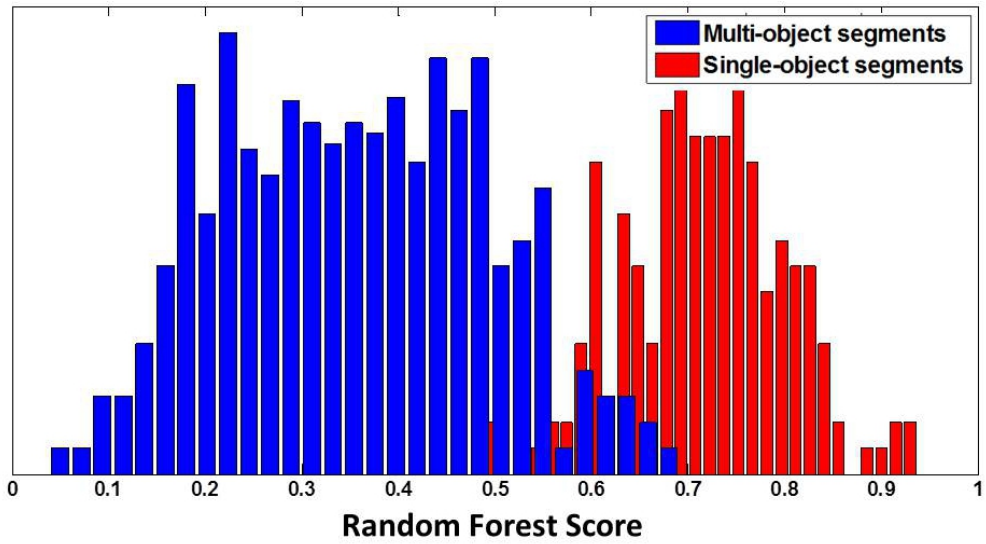


Figure 7.10: Separation of single and multi-object segments using random forest score and 3D shape descriptor [MFB10]

.

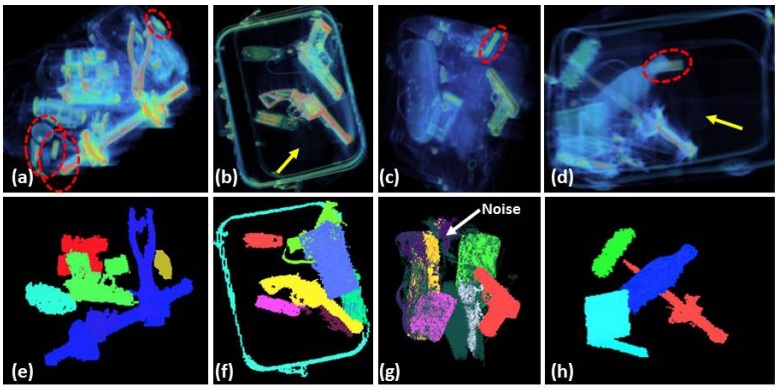


Figure 7.11: Volumetric visualisations of segmentation test images and corresponding coarse segmentations: (a) - (d) Input baggage-CT scans used in qualitative evaluation of segmentation algorithms. (e) - (h) Coarse image segmentation / foreground determination using DEI thresholding and CCA (with MAR). Objects missed by coarse segmentation have been indicated (circles and arrows).

a user-defined constant threshold.

Figures 7.11 (e) - (h) show the coarse segmentations produced by the DEI thresholding process, which were used as input to the CCA and IDT segmentation-refinement procedures. Metal artefact reduction was applied to the input images prior to generating the coarse segmentations. As expected, several objects are under-segmented (e.g. pliers and dumbbells in (a) and (e)). While the majority of the objects appear to have been well segmented in all four examples, two object types were commonly eliminated/missed by the coarse segmentations: 1) small cylindrical objects (encircled in red in Figure 7.11 (a),(c),(d)) and 2) thin, low-density magazines (indicated with arrows in Figure 7.11 (b) and (d)). The paperback book in Figure 7.11 (c) was well segmented in (g) - indicating that it is not the material characteristics alone of the magazines that resulted in their elimina-

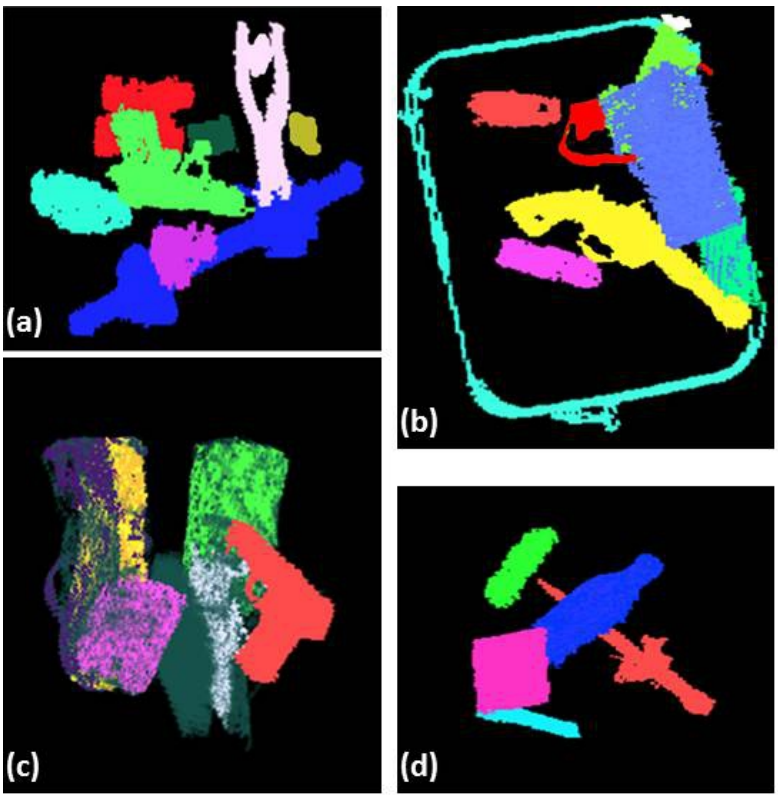


Figure 7.12: CCA segmentation (Section 7.2.3) results for test images in Figure 7.11 with MAR.

tion. A more likely cause is the positioning and geometry of the magazines: in both scenarios, the magazines are lying flat against the bottom of the case/bag making them difficult to distinguish (even for the human observer) from the actual bag (on account of similar densities, their lack of bulk and noise). It is worth noting that low-density objects and thin, sheet-like objects have been known to pose difficulties for all previous baggage-segmentation algorithms [CMP13, WGW12, GSK+12].

The final segmentation results produced by each of the algorithms (with metal artefact reduction) are shown in Figures 7.12 - 7.16. CCA (Figure 7.12) and IDT (Figure 7.13) produced similar results as both algorithms rely on the same coarse segmentations and RFS to determine which components require refinement. The results thus differed only in those components which required refinement. In general, IDT produced superior refinements. This is especially evident in the test images (a) and (b). Considering, for example, the test image in Figure 7.12 (a) (CCA) and Figure 7.13 (a) (IDT), where IDT produced superior partitions in five hot-points regions. Figure 7.14 illustrates these regions in the original coarse segmentations and shows the computed hot-points and the post-refinement RFS. The object boundaries produced by IDT refinement are better defined in all five cases, resulting in higher RFS (for the individual objects). Nonetheless, CCA correctly split the coarse segmentations at all hot-points regions (with the exceptions of regions 3 and 4) and produced corresponding improvements in the RFS. CCA refinements at regions 3 and 4 (Figure 7.14) were most likely rejected based

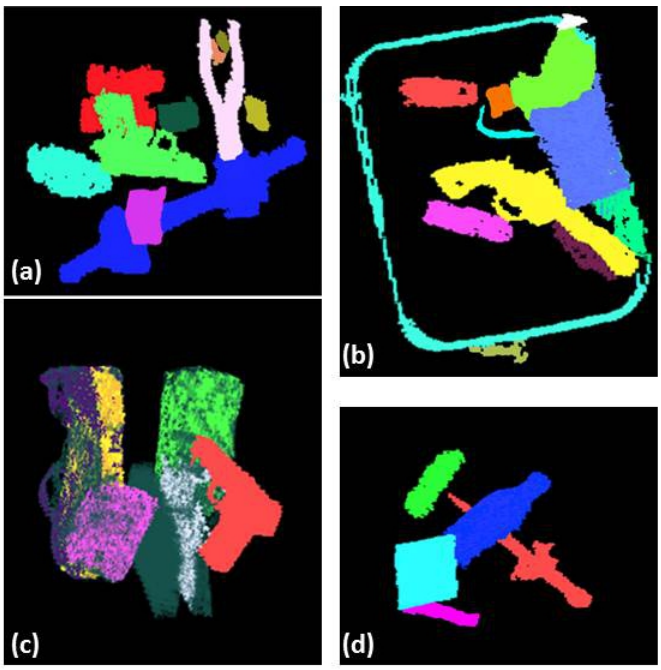


Figure 7.13: IDT segmentation [Gra06] results for test images in Figure 7.11 with MAR.

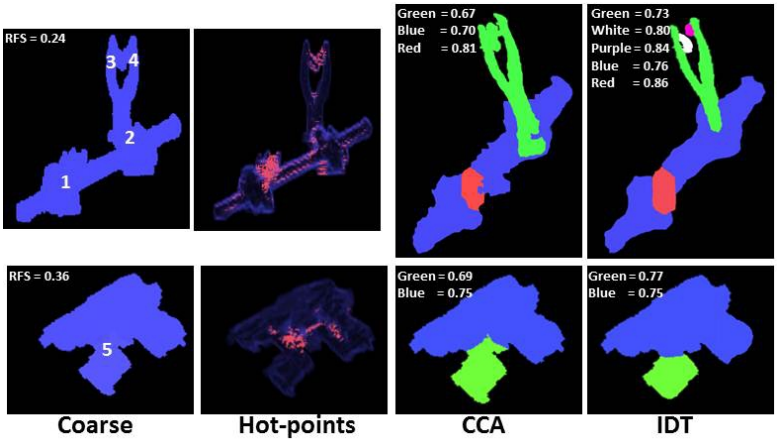


Figure 7.14: Segmentation refinement using CCA (Section 7.2.3) and IDT [GSK+12] techniques. Hot-points (red) and RFS indicated.

on the resulting components not meeting the minimum permissible object size. As discussed in Section 7.2.3, the refinement procedure, in addition to splitting merged objects, possesses denoising characteristics. This is illustrated in test image (c): the coarse segmentation (Figure 7.11 (g)) exhibits what appears to be noise/artefacts to the right of the sole of the shoe. This noise has been removed in the corresponding regions in both CCA (Figure 7.12 (c)) and IDT (Figure 7.13 (c)) refinements.

The segmentations produced by SymRG (Figure 7.15) were noticeably poorer compared to CCA and IDT. In particular, the results are characterised by under-segmentations (indicated by solid circles) and missed segmentations - where object regions have been incorrectly set as background (indicated by dotted circles). The results suggest the necessity for post-segmentation splitting and merging opera-

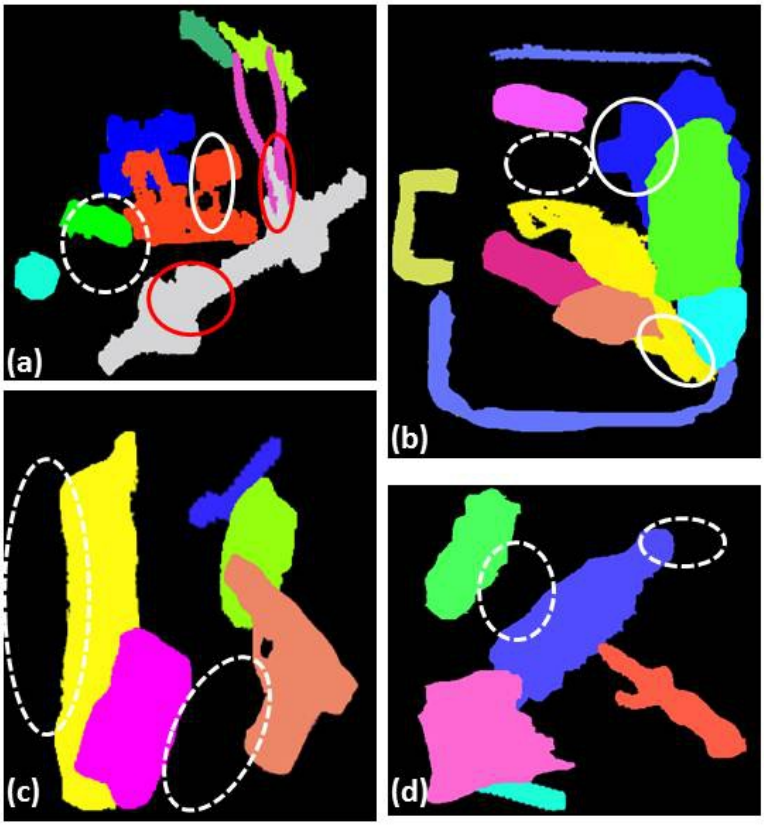


Figure 7.15: SymRG segmentation [WH03] results for test images in Figure 7.11 with MAR. Examples of under-segmentations (solid circles) and incorrect background assignments (dotted circles) indicated.

tions and explain the complexity in this regard of the segment-and-carve baggage-segmentation algorithm of Song *et al.* [CMP13], which employs a total of 5 splitting and/or merging operations. As discussed in Section 3.5.2, refining the segmentations in this way significantly expands the input parameter space and hence the degree of user interaction. Accurate segmentations consequently depend on careful parameter tuning and suffer from increased computational demand. It is worth noting, however, that SymRG did capture several objects (or parts thereof) which were missed by the DEI coarse segmentations (Figures 7.11) - notably, the cylindrical structures in Figures 7.15 (a) (turquoise) and (c) (blue). Furthermore, the segmentations of the regions corresponding to the hot-points labelled 3 and 4 in Figure 7.14, were more accurately segmented by SymRG compared to CCA (Figure 7.12 (a)).

Similarly to SymRG, FloodFill [WGW12] (Figure 7.16) produced segmentations characterised by ill-defined object boundaries, under-segmentations (indicated by solid circles) and missed segmentations (indicated by dotted circles). The most evident shortcomings of the FloodFill approach, however, are the poorly-defined object boundaries. This is particularly apparent for the handguns in Figures 7.16 (a) and (c) (compared to the equivalent CCA (Figure 7.12) and IDT (Figure 7.13) segmentations). Similarly to the SymRG segmentation, the Flood-

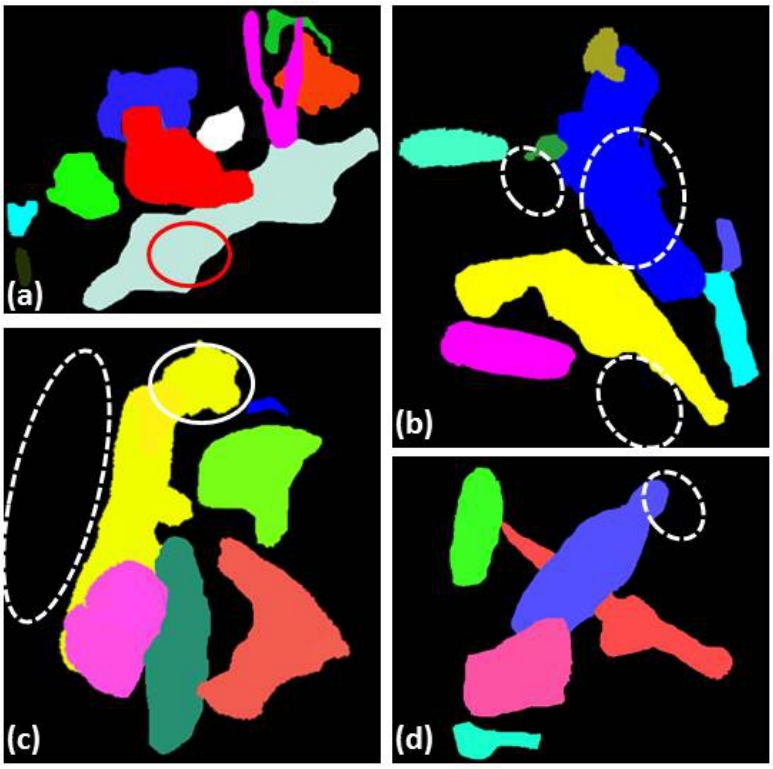


Figure 7.16: FloodFill segmentation [WGW12] results for test images in Figure 7.11 with MAR. Examples of under-segmentations (solid circles) and incorrect background assignments (dotted circles) indicated.

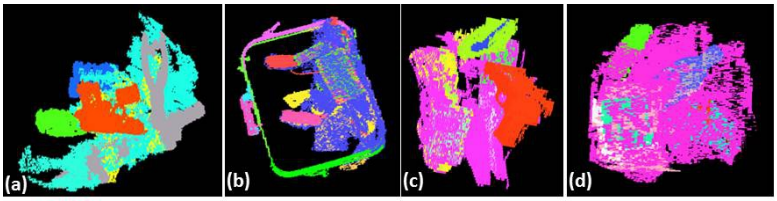


Figure 7.17: Coarse DEI image segmentations of test images in Figure 7.11 without MAR.

Fill segmentation of test image (a) captured objects which were missed by the DEI coarse segmentations (and hence the CCA and IDT results). In general, the segmentations produced by SymRG and FloodFill were of an inferior quality to the corresponding CCA and IDT segmentations.

The equivalent results without the application of metal artefact reduction are shown in Figures 7.17 - 7.21. Similarly to the observations made in Chapter 6, the discriminative power of the coarse DEI segmentations (Figure 7.17) deteriorated significantly when metal artefact reduction was not applied. As expected, CCA (Figure 7.18) and IDT (Figure 7.19) segmentations suffered as a result. CCA produced segmentations characterised by a considerably higher number of under-segmentations (multiple objects labelled as a single object) and background noise (compared to the corresponding results with MAR - Figure 7.12). The results suggest that the high-frequency streaking artefacts result in the merging of nearby objects, making object spitting by simple connected component analysis less ef-

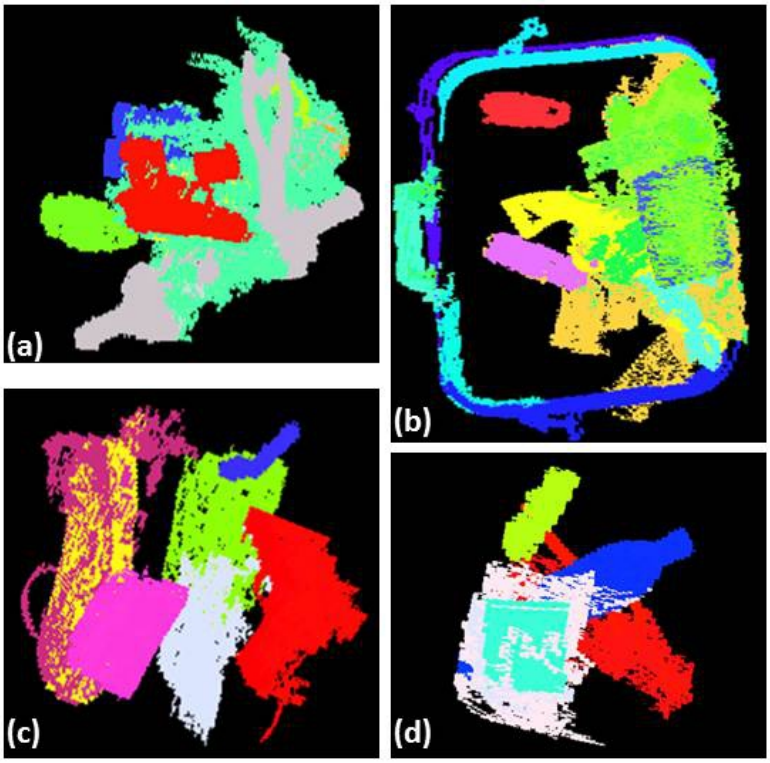


Figure 7.18: CCA segmentation (Section 7.2.3) results for test images in Figure 7.11 without MAR.

Method	$\overline{\mathbf{RFS}}_S$	$\overline{\lvert N_T - N_S \rvert}$
CCA (Section 7.2.3)	0.89 ± 0.05	0.53 ± 0.57
IDT [GSK+12]	0.94 ± 0.02	0.10 ± 0.31
SymRG [WH03]	0.51 ± 0.08	1.73 ± 1.14
FloodFill [WGW12]	0.57 ± 0.11	1.37 ± 1.25

Table 7.2: Quantitative results of four segmentation algorithms with MAR: total Random Forest Score (Equation 7.13) and error in number of objects segmented. Results averaged over 30 volumes containing known numbers of objects.

fective. Although IDT (Figure 7.19) was able to successfully split several fused objects which CCA could not (e.g. pliers and dumbbell in test image (a)), the segmentations are similarly corrupted by background noise and exhibit an increase in the number of under-segmentations.

SymRG (Figure 7.20) and FloodFill (Figure 7.21) showed a similar decline in performance in the absence of MAR. In addition to several cases of under-segmentations and missed-segmentations, similar to those produced by CCA and IDT, the SymRG and FloodFill segmentations are further characterised by several examples of over-segmentations (e.g. dumbbell in Figures 7.20 and 7.21 (a) and handgun in Figures 7.20 and 7.21 (c)).

The processing times of each of the segmentation techniques when applied to the test images with and without MAR are shown in Figure 7.22. With the exception of the SymRG technique [WH03], all techniques exhibited an increase in

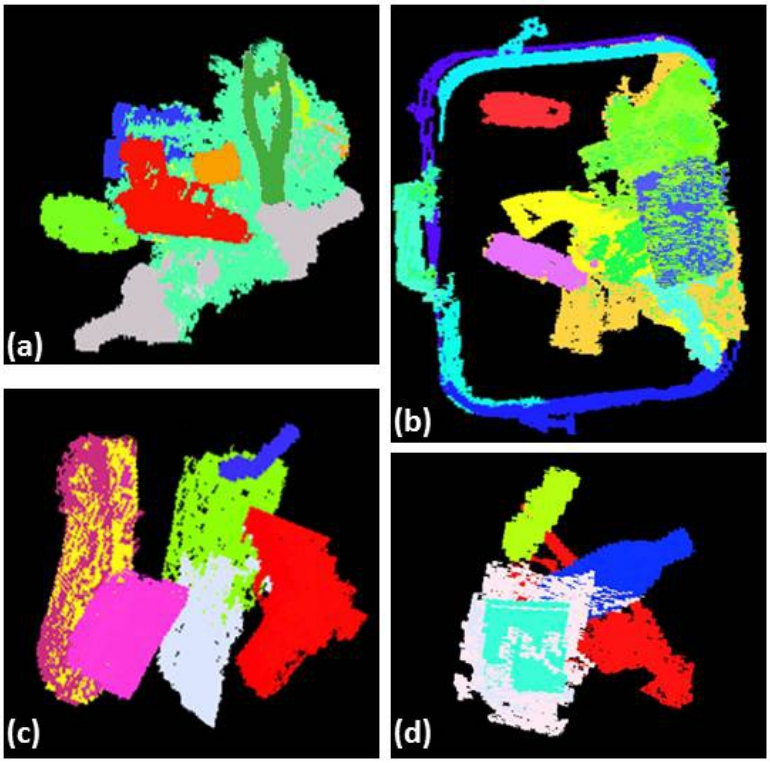


Figure 7.19: IDT segmentation [Gra06] results for test images in Figure 7.11 without MAR.

Method	$\overline{\mathbf{RFS}}_S$	$\overline{\lvert N_T - N_S \rvert}$
CCA (Section 7.2.3)	0.58 ± 0.09	1.80 ± 1.13
IDT [GSK+12]	0.69 ± 0.09	1.13 ± 0.78
SymRG [WH03]	0.39 ± 0.17	2.93 ± 1.31
FloodFill [WGW12]	0.41 ± 0.21	2.80 ± 2.33

Table 7.3: Quantitative results of four segmentation algorithms without MAR: total Random Forest Score (Equation 7.13) and error in number of objects segmented. Results averaged over 30 volumes containing known numbers of objects.

processing time when metal artefact reduction was not performed. The computational demand of SymRG is by nature dependent only on the dimensions of the input image [WH03]. The resulting processing times were thus consistent ($\sim 165s$) for the artefact-reduced and original volumes. CCA (Section 7.2.3) was the most efficient of the four techniques when operating on the artefact-reduced images, with processing times ranging from 94s to 155s. These times, however, increased by approximately 90% when MAR was not performed, making it less efficient than SymRG. The processing times of FloodFill [WGW12] varied significantly from image-to-image (ranging from 249s to 548s for the artefact-reduced images) and appear to be largely dependent on the complexity of the image. IDT [GSK+12] was, as expected, the most computationally intensive and yielded consistently high processing times (ranging from 352s to 1238s for the artefact-reduced volumes). Both IDT and FloodFill exhibited significant increases in processing times when

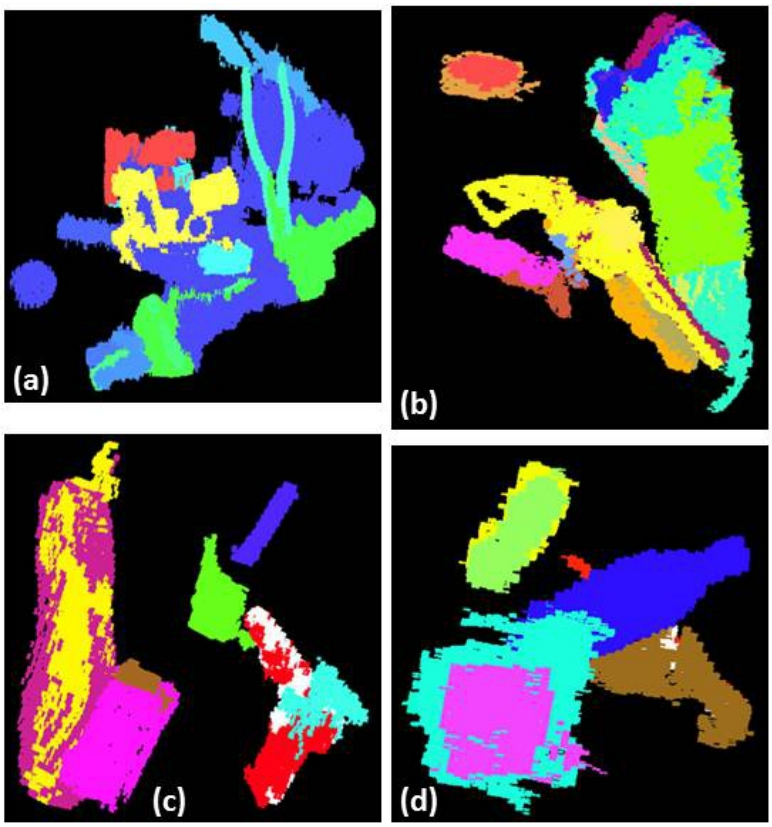


Figure 7.20: SymRG segmentation [WH03] results for test images in Figure 7.11 without MAR.

applied to the original volumes, with times in excess of 20 minutes. Such high processing times detract from the practical usability of these approaches (in their current states), particularly in the security-screening domain, where the demands on low processing times are paramount.

The quantitative results, with and without metal artefact reduction, are illustrated in Figures 7.23 - 7.26 and summarised in Tables 7.2 - 7.3. Figure 7.23 shows the total segmentation RFS for the artefact-reduced test images computed according to Equation 7.13. CCA (Section 7.2.3) and IDT [GSK+12] yielded significantly higher segmentation scores compared to SymRG [WH03] and FloodFill [WGW12] for all 30 test images. In particular, IDT produced on average the highest quality segmentations ($\overline{\text{RFS}}_S = 0.94$), which may be attributed to both the high quality of the individual components in each segmentation (as observed in the qualitative results) as well as the high accuracy in the number of objects segmented in each test image (an average error of only 0.1 - Table 7.2). Figure 7.24 shows that IDT segmented the correct number of objects in 27/30 images and the remaining 3 images (test images 11,23,24) each contained a discrepancy of only a single object. Although CCA produced the incorrect number of objects in 15/30 test images (Figure 7.24), the discrepancies were low (≤ 2). Furthermore, the mean segmentation quality remained high ($\overline{\text{RFS}}_S = 0.89$ - Table 7.2), indicating that the segmentation quality of the individual objects in each image were

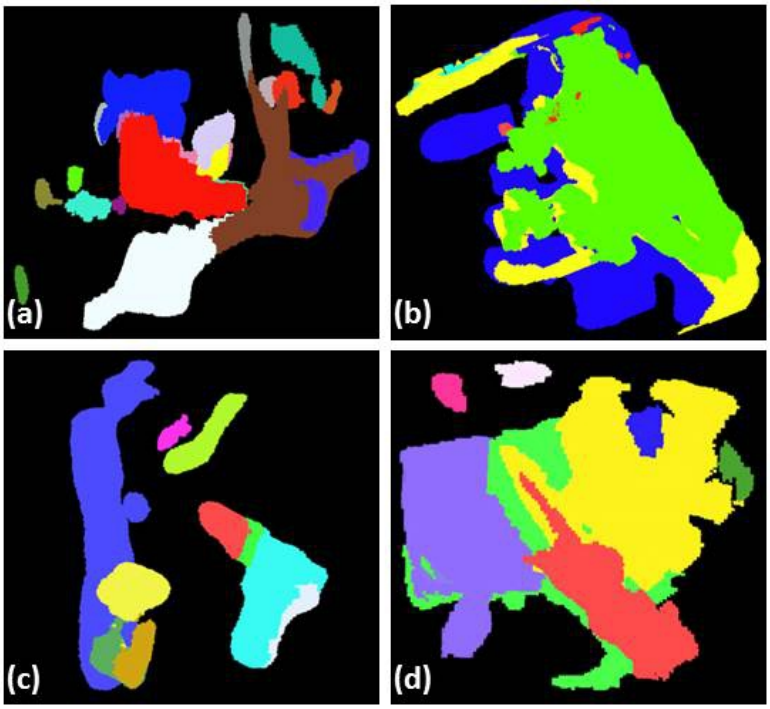


Figure 7.21: FloodFill segmentation [WGW12] results for test images in Figure 7.11 without MAR.

high. SymRG and FloodFill performed significantly poorer with mean scores of ($\overline{\text{RFS}}_S = 0.51$) and ($\overline{\text{RFS}}_S = 0.57$) respectively (Table 7.2) and the incorrect number of segmented objects in 26/30 images and 23/30 images respectively (Figure 7.24).

Figure 7.25 shows a decline in the segmentation quality for each method for all 30 images when metal artefact reduction was not applied. Figure 7.26 additionally shows that the number of over and/or under-segmented images also increased for all four methods (CCA = 27/30; IDT = 24/30; SymRG = 30/30; FloodFill = 25/30). The decline in performance was fairly consistent for all four techniques, with IDT again producing on average the highest quality segmentations ($\overline{\text{RFS}}_S = 0.69$), followed by CCA ($\overline{\text{RFS}}_S = 0.58$), FloodFill ($\overline{\text{RFS}}_S = 0.41$) and SymRG ($\overline{\text{RFS}}_S = 0.39$) (Table 7.3). The significant decline in performance, coupled with the increase in processing time (Figure 7.22), demonstrates the detrimental effects that image noise and artefacts have on the segmentation process and emphasises the importance of an effective metal-artefact-reduction process.

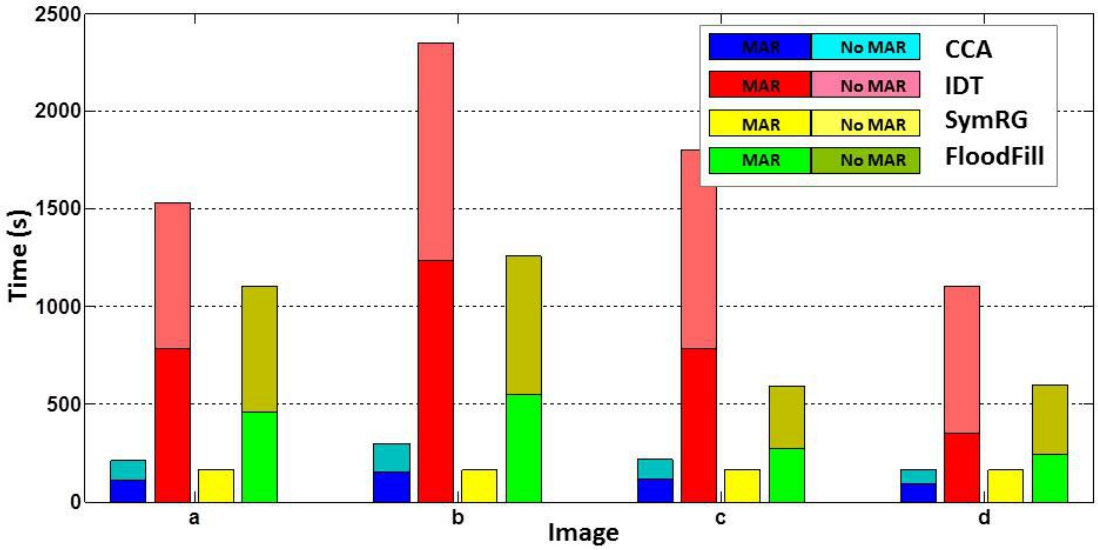


Figure 7.22: Segmentation processing times for test images in Figure 7.11 with and without MAR.

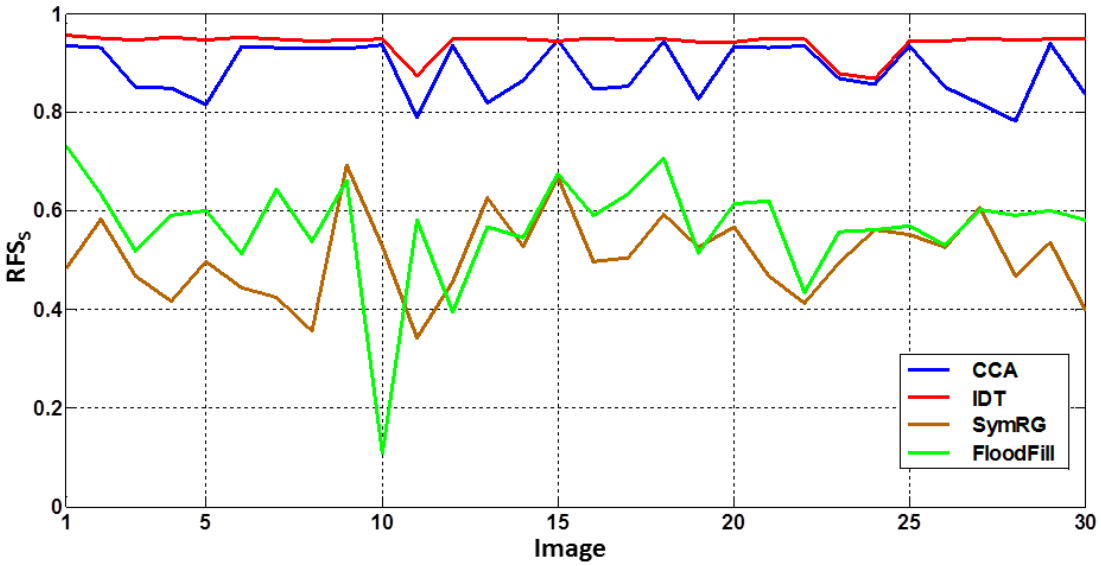


Figure 7.23: Overall image-segmentation quality scores (Equation 7.13) for 30 artefact-reduced test images containing known numbers of objects.

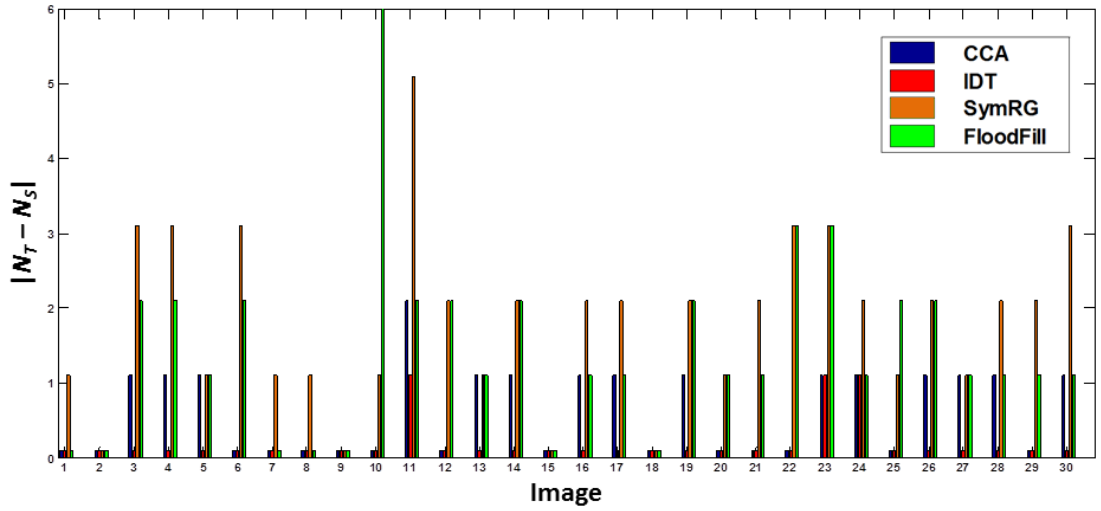


Figure 7.24: Errors in numbers of objects segmented for 30 artefact-reduced test images containing known numbers of objects.

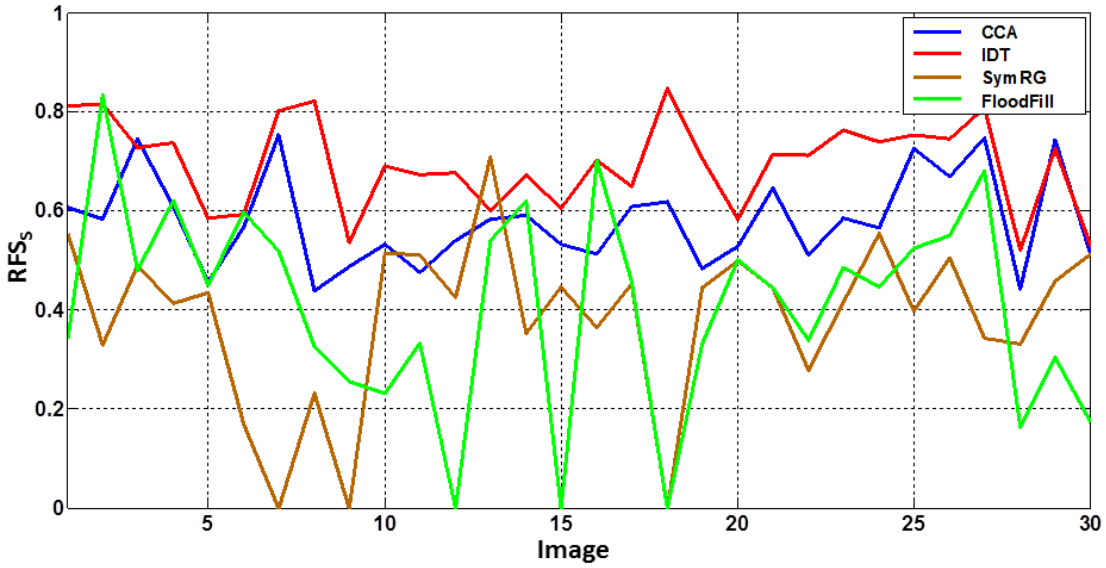


Figure 7.25: Overall image-segmentation quality scores (Equation 7.13) for 30 test images without MAR.

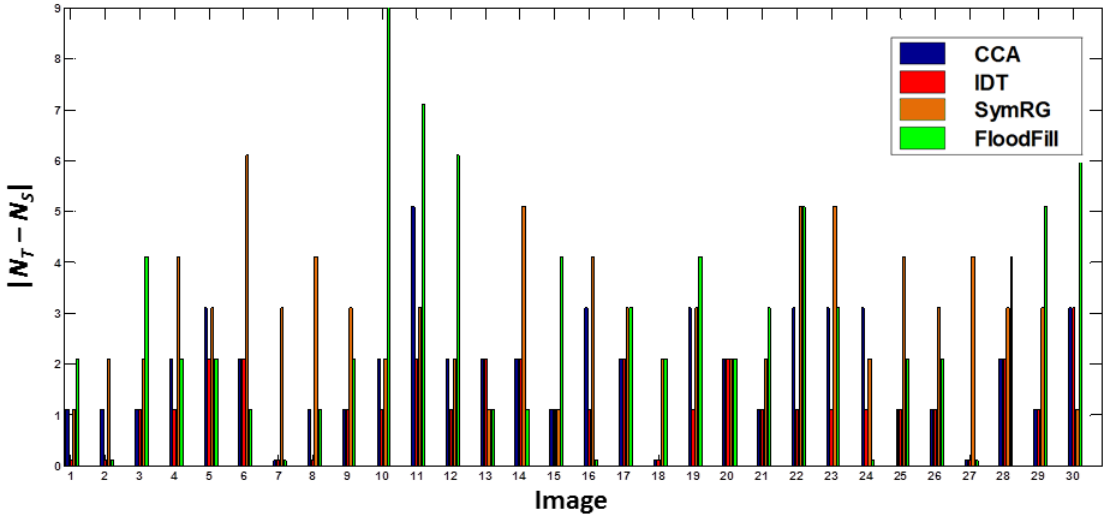


Figure 7.26: Errors in numbers of objects segmented for 30 test images without MAR.

7.5 Conclusion

This chapter has addressed the segmentation of unknown objects from low-resolution, cluttered volumetric baggage-CT data. Based on the dual-energy techniques presented in Chapter 6, a novel materials-based segmentation technique has been presented.

The proposed segmentation algorithm is composed of three stages: 1) coarse segmentation; 2) segmentation quality evaluation and 3) segmentation refinement. Coarse segmentations are generated using a characterisation of the chemical composition of an image (using the Dual-Energy Index (DEI) [Joh11]), simple thresholding operations and connected component analysis. The quality of the individual components of the coarse segmentations are evaluated using the Random Forest Score (RFS) - which is trained to recognise high-quality (single-object) object segments. Preliminary experimentation has demonstrated the superiority, in the current context, of the RFS over the related generative Automated QUality Assessment (AQUA) measure [GSK+12]. Segmented objects are represented using the descriptor described in [KSA+12] (Appendix B), which is shown to outperform more complex 3D shape-based descriptors (the 3D Zernike descriptor [NK04], the Histogram-of-Shape Index [DJ95] and a hybrid 3D shape descriptor [MFB10]). Based on the RFS of a given coarse segmentation, low-quality individual object segments are subjected to an object-partitioning operation which splits fused objects at automatically-detected regions using a simple connected component analysis. A second segmentation-quality measure is presented for quantifying the quality of a full segmentation (as opposed to individual object segments). The measure only requires prior knowledge of the number of objects in a given image (as opposed to a fully-annotated reference image) to provide a measure of segmentation quality.

Within the proposed framework, four novel contributions have been made: 1) a materials-based coarse segmentation technique; 2) a random-forest-based model for measuring the quality of individual object segments; 3) a random-forest-based model for measuring the quality of entire segmentations and 4) an efficient segmentation-refinement procedure for splitting fused objects.

An experimental comparison between the proposed segmentation algorithm (denoted CCA) and three state-of-the-art volumetric segmentation techniques (the Isoperimetric Distance Tree (IDT) [GSK+12]; Symmetric Region Growing (SymRG) [WH03] and 3D flood-fill region growing (FloodFill) [WGW12]) has been performed using low-resolution, complex volumetric baggage-CT data (Section 1.2). Qualitative performance analysis, using four realistic, cluttered baggage scans, has demonstrated that IDT and CCA generate higher (visual) quality segmentations relative to SymRG and FloodFill. Although IDT is shown to outperform CCA in partitioning fused objects in the DEI-generated coarse segmentations,

it is characterised by high processing times and is significantly outperformed by CCA in this regard. Low-density, sheet-like objects (e.g. magazines) are shown to pose difficulties for all four methods (an observation which has been made in the majority of related studies [CMP13, WGW12, GSK+12]).

A quantitative analysis, using the proposed Random Forest Score (RFS) for image segmentations and a set of volumes containing known numbers of objects, substantiates the observations made in the qualitative analysis. Particularly, IDT and CCA consistently outperform SymRG and FloodFill in terms of segmentation quality and in terms of segmentation accuracy (with reference to the number of objects segmented from each image). Finally, the importance of MAR is demonstrated by the significant decline in performance for all four segmentation techniques, across all evaluation metrics (qualitative and quantitative) when MAR is not applied.

The observations made in this chapter indicate that the proposed CCA segmentation algorithm (Section 7.2.3) is well-suited to the task of volumetric image segmentation - particularly in the baggage security-screening domain, where the demands for low processing times are paramount. Chapter 8 investigates the incorporation of the proposed segmentation algorithm into a fully-automated 3D object-classification framework.

Chapter 8

3D Object Classification

The current state-of-the-art in 3D object classification in low-quality, complex 3D volumetric imagery [FBM12] relies on the manual segmentation of the input data, incurs large computational overhead (in building the model) and suffers a decline in performance in the presence of image noise and artefacts. Although the need for manual segmentations have been eliminated in the fully-automated approach of [FMMB13], the technique is computationally expensive and leads to false-positive classification rates in excess of 15%.

Each of the aforementioned limitations have been addressed individually in the preceding chapters of this work. In this chapter, the presented techniques are combined to create an efficient, yet fully-automated framework for the classification of objects in complex, volumetric imagery (Figure 8.1). The resulting framework is shown to improve on the current state-of-the-art [FBM12] by reducing the detrimental effects of image noise and artefacts (methods from Chapter 4); automating the segmentation process (Chapters 6, 7); decreasing computational cost and increasing classification accuracy (Chapter 5).

Portions of this chapter have been submitted for publication and are currently under review [MB14c].

8.1 Methods

The object-classification framework proposed here is composed of three stages (Figure 8.1): 1) noise and/or artefact reduction; 2) segmentation and 3) classification.

8.1.1 Noise and Artefact Reduction

Flitton *et al.* [FBM10, FBM13, FBM12] cite image noise and artefacts as the two major factors limiting the performance of object classification in complex volumetric imagery. A comprehensive review and evaluation of noise and artefact reduction in low-quality 3D baggage-security-CT imagery has been presented in

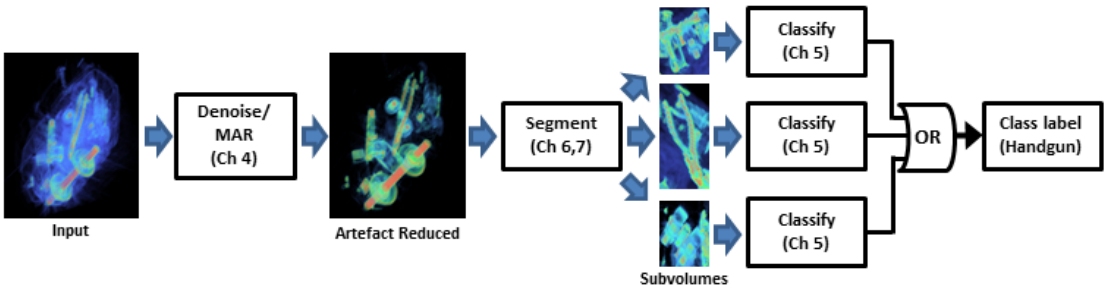


Figure 8.1: Fully-automated object classification pipeline. Chapter contributions indicated.

Chapters 3 and 4 (see also [MMFB13, MMFB12, MMvS+13, MMB+13]). Most notably, it is shown that the superiority of the state-of-the-art techniques, from the medical literature, over simpler techniques is reduced as a result of the complexity and poor quality of the imagery considered in this work. It is thus not obvious if the performance gains (if any) of such techniques merit their additional computational overhead. In the development of the object-classification framework presented here, the impact of four noise/artefact reduction techniques (of varying complexities) are considered and compared: 1) simple intensity thresholding; 2) Non-Local Means (NLM) filtering [BCM05a, BCM05b]; 3) MAR by linear interpolation [KHE87] (denoted LIMar) and 4) distance-driven MAR [MMB+13] (denoted DDMar). NLM filtering and MAR are applied on a per-slice basis. The linear-complexity NLM implementation of Mahmoudi and Sapiro [MS05a] (Chapter 4), is used to optimise computational efficiency and reduce processing times. Denoising and MAR are considered pre-processing operations to be applied prior to the chosen segmentation algorithm (Figure 8.1).

8.1.2 Segmentation

The current state-of-the-art in object classification in complex, volumetric-CT imagery employs subvolumes generated by the manual isolation (i.e. segmentation) of target objects [FBM13, FBM12]. The segmentation techniques presented in Chapter 7 are considered here as a means of automating the generation of these subvolumes.

In Chapter 5 of this work (as well as in the related literature [FBM13, FBM12]), classification of objects in cluttered volumetric imagery was performed on manually generated subvolumes, containing at most a single object of interest. In view of the high classification rates achieved using such volumes ($> 98\%$ - Chapter 5), the output label maps generated using the segmentation method(s) presented in Chapter 7, are used to create a set of single-object subvolumes for each given input volume. These subvolumes are then passed to the chosen classifier. The subvolumes are generated as to represent those used in the experimentation presented

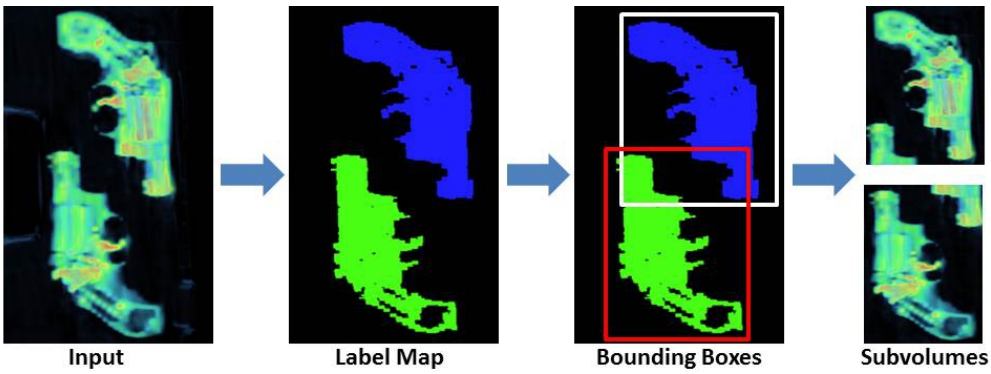


Figure 8.2: Generation of single-object subvolumes for object classification.

in Chapter 5 and [FBM12].

In particular, for each labelled object in a given segmentation, a subvolume is generated by computing the minimum bounding box encompassing that object and then increasing the dimensions of the box by approximately 30mm (in all 3 dimensions) [FBM12]. The high correct classification rates presented in Section 5.2 of this work, were achieved without considering noise and/or artefact reduction. This suggests that the classification performance of the proposed approach is robust to image noise and artefacts, provided a given subvolume is dominated by a single object. The final subvolume used for classification is thus obtained by extracting the entire region corresponding to the expanded bounding box from the denoised volume (as opposed to setting the non-object (background) voxels to zero in the subvolume). This strategy is adopted to ensure that contextual information is not lost in the subsequent feature extraction and description process. A segmentation composed of N labelled objects will thus result in N subvolumes (each theoretically containing a single, distinct object). Figure 8.2 illustrates the subvolume generation process for an input volume composed of two objects.

The experimental comparison of the four segmentation algorithms presented in Chapter 7 (CCA - Section 7.2, IDT [GSK+12], SymRG [WH03], FloodFill [WGW12]) indicated a clear superiority of the CCA and IDT segmentation techniques over the SymRG and FloodFill techniques. Owing to its low processing time, the CCA segmentation technique proposed in Chapter 7 is used here to generate the subvolumes for object classification.

8.1.3 Classification

The codebook classification framework proposed in Chapter 5 is used to independently classify each of the subvolumes comprising a given input volume. In particular, object classification is accomplished via a Support Vector Machine (SVM) classifier [Vap00] using a Radial Basis Function (RBF) kernel and operating on quantised feature vectors built using Extremely Randomised Clustering (ERC) forests [MTJ07, MNJ08] and densely sampled, multi-scale Density His-

togram (DH) descriptors [FBM13]. This classification framework produced the optimal results in the experimental comparison presented in Section 5.2 (where it was denoted *Codebook4* - Table 5.3).

The class label of a given input volume, composed of N segmented objects (i.e. N subvolumes) is computed as the logical 'OR' of the class labels of each of its N constituent subvolumes. While the classification of each of the N subvolumes is easily parallelised the processing time of a serial classification of a given N-object input volume may be improved by only classifying the i^{th} subvolume if the $(i-1)^{th}$ subvolume (where $i = 2, \ldots, N$) has returned a negative class label (i.e. classification is terminated as soon as a positive label is produced).

8.1.4 Test Data

Similarly to the frameworks presented in Chapter 5 and the baseline study of Flitton *et al.* [FBM12], the classification of two independent object types (handguns and bottles) is considered. Five separate datasets are used in this chapter. All intensities are represented in MHU.

1. The dataset used to build the random forest model which guides object partitioning. This set is composed of 80 manually-cropped single-object segments and 80 manually-cropped multi-object segments of varying sizes.

2. The dataset used to train the SVM classifier for the handgun experiments (Figure 8.3). This set is composed of 101 manually-cropped handgun (positive) subvolumes and 134 manually-cropped clutter (negative) subvolumes.

3. The dataset used to train the SVM classifier for the bottle experiments (Figure 8.3). This set is composed of 88 manually-cropped bottle (positive) subvolumes and 90 manually-cropped clutter (negative) subvolumes.

4. The test set used in the handgun experiments. This set is composed of 208 handgun-containing (positive) whole volumes and 150 handgun-free clutter (negative) whole volumes (Figure 8.4).

5. The test set used in the bottle experiments. This set is composed of 146 bottle-containing (positive) whole volumes and 190 bottle-free clutter (negative) whole volumes (Figure 8.4).

Dataset (1) above is identical to that used to build the random forest models in the experimentation presented in Chapter 7. In contrast to the experimentation presented in Chapter 5 and [FBM12], the clutter subvolumes in sets (2) and (3) have all been manually cropped to prevent the inclusion of meaningless subvolumes (e.g. subvolumes containing no whole objects). The whole volumes comprising

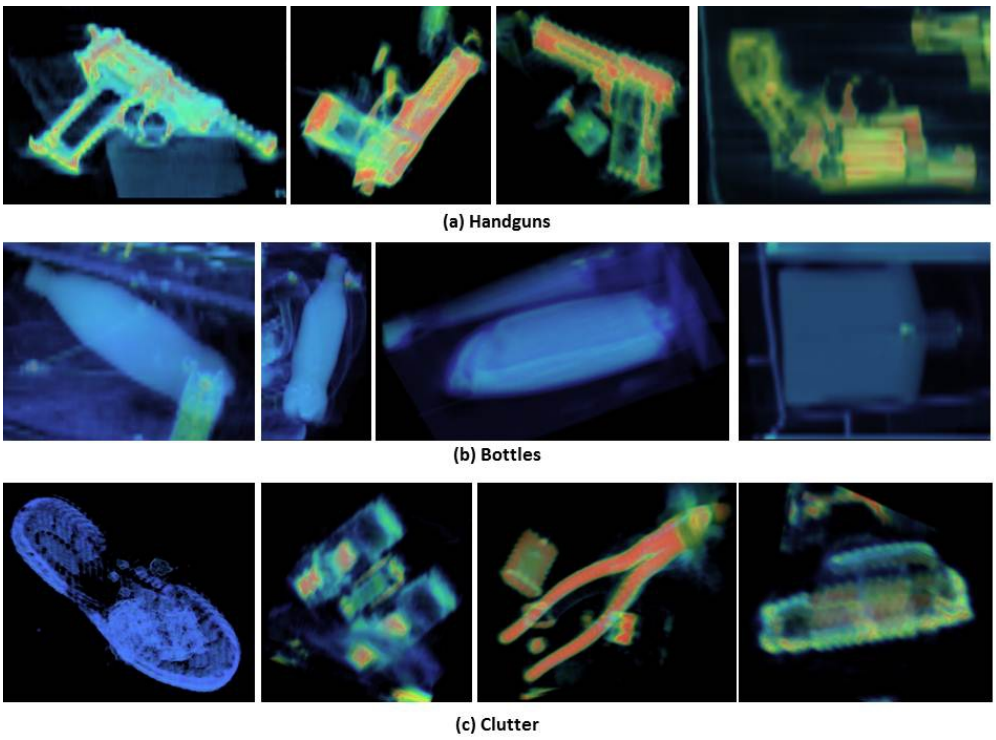


Figure 8.3: Examples of manually segmented training data. Handgun subvolumes are created to contain a single handgun; bottle subvolumes a single bottle and clutter subvolumes a single whole object which is neither a handgun nor a bottle.

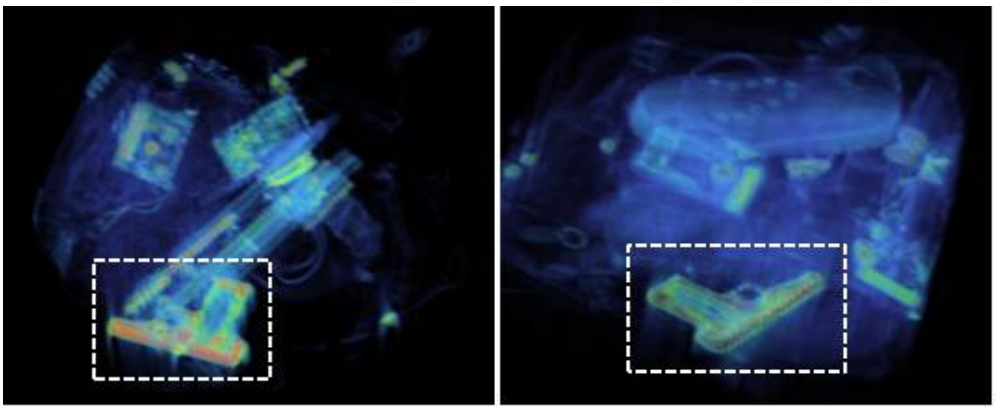
the test sets in (4) and (5) are used to evaluate the end-to-end performance of the proposed classification algorithm.

All non-target objects are considered as clutter and are chosen to provide an environment that is comparable to that encountered within the transport infrastructure. Typical clutter items include both low density items (e.g. clothing, books etc.) and high density items (e.g. belt buckles, batteries, pliers, dumbbells etc.). In the handgun classification experiments, bottles are considered as clutter items and *vice versa*.

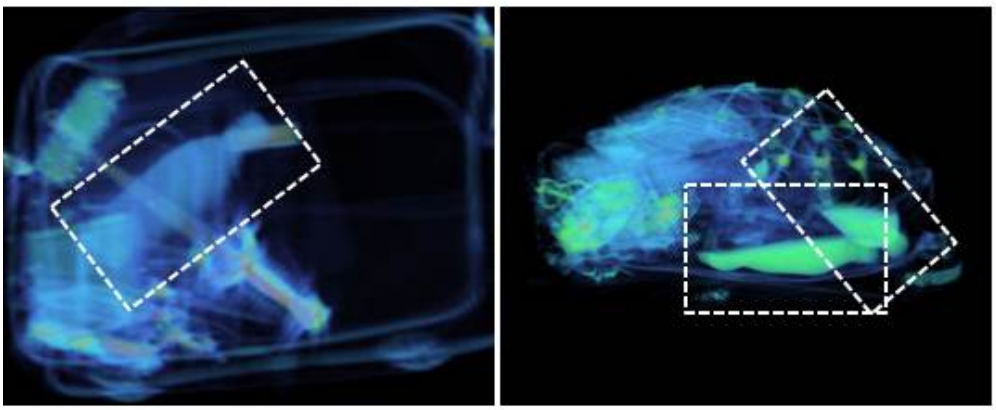
8.2 Results

Performance was quantified via traditional measures (true-positive rate, false-positive rate, precision, accuracy and processing time). Processing times were measured for all experiments performed on an Intel Core i5 machine running a 2.30GHz processor with 6GB of RAM.

The optimal algorithm parameters were determined independently for each object class and then kept constant for every instance in that experiment (i.e. one set of parameters used for the entire handgun experiment and another set for the entire bottle experiment). The pre-processing parameters were determined using a small set of validation volumes and several different sets of input parameters. Those parameters that subsequently produced the most visually satisfying segmen-



(a) Handgun volumes



(b) Bottle volumes

Figure 8.4: Example test volumes - target object indicated.

tations were chosen. The block-wise linear-time NLM implementation [MS05a] was used to reduce the characteristically-high processing time of NLM filtering [BCM05a, BCM05b]. Window sizes of 11×11 and 7×7 were used for gradient and intensity similarity computations respectively. Due to the efficiency of the random forest-based classification process and the relatively small dataset, the optimal forest and SVM parameters were determined empirically via 5-fold cross-validation performed over the entire dataset. The number of tests performed for each node split in the random forest was set to $|\mathcal{T}_j| = 10$ - this value was fixed for all nodes. Trees were grown to a maximum depth of $D_T = 5$, with a lower bound of $IG_{min} = 10^{-4}$ on the information gain (Section 3.3.2). It was found that using these settings resulted in tree growth terminating prior to maximum depth and thus no tree pruning was performed. The settings resulted in trees with approximately 300 leaf nodes each. For a forest containing $T = 5$ trees, codebooks therefore typically contained approximately 1500 codewords.

Tables 8.1 - 8.4 summarise the results of the classification experiments. Table 8.1 shows the confusion matrices obtained in the handgun classification experiments for each of the four pre-processing methods tested (thresholding, NLM filtering [BCM05b], LIMar [KHE87] and DDMar [MMB+13] (Section 4.2)). Pre-processing by DDMar and NLM filtering yielded the optimal results, with only 7

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	141	9
Handgun (actual)	19	189

(a) Intensity thresholding.

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	148	2
Handgun (actual)	7	201

(b) NLM filtering [BCM05a, BCM05b].

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	146	4
Handgun (actual)	9	199

(c) Linear interpolation-based MAR [KHE87].

	Clutter (predicted)	**Handgun** (predicted)
Clutter (actual)	148	2
Handgun (actual)	5	203

(d) Distance-driven MAR (Section 4.2).

Table 8.1: Handgun classification confusion matrices (208 handgun and 150 clutter images)

and 9 incorrect classifications respectively. While LIMar (13 errors) outperformed simple intensity thresholding (28 errors), the latter still performed surprisingly well given its simplicity.

The results of the bottle classification experiments are shown in the confusion matrices in Table 8.2. NLM filtering correctly classified all positive (bottle) instances and resulted in only 4 false positives. DDMar again yielded high correct classification rates (2 false negatives and 3 false positives). The superiority of NLM filtering and DDMar over LIMar (19 errors) was more pronounced compared to the handgun experiments. Intensity thresholding performed significantly poorer than all three methods, with a particularly high false-negative rate (115/146 bottles missed).

The aforementioned results are summarised in Table 8.3 which illustrates the

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	183	7
Bottle (actual)	115	31

(a) Intensity thresholding.

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	186	4
Bottle (actual)	0	146

(b) NLM filtering [BCM05a, BCM05b].

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	187	3
Bottle (actual)	16	130

(c) Linear interpolation-based MAR [KHE87].

	Clutter (predicted)	**Bottle** (predicted)
Clutter (actual)	187	3
Bottle (actual)	2	144

(d) Distance-driven MAR (Section 4.2)

Table 8.2: Bottle classification confusion matrices (146 bottle and 190 clutter images)

True-Positive Rates (TPR), False-Positive Rates (FPR), precision and accuracy for all of the experiments.

The mean, per-volume processing times for each of the four pre-processing methods (averaged over both sets of experiments) are shown in Table 8.4. Similarly to Chapter 5, the use of the average is justified by the fact that similar volumes were used in both experiments (with only the target objects differing) and thus processing times across the two experiments were relatively consistent. As expected, simple intensity thresholding was performed with virtually no computational overhead (0.23s per volume). As suggested by the experimentation in Chapter 4, the processing times of the LIMar and DDMar (49.05s and 401.42s per volume respectively) were higher than that of the NLM filter (34.01s per volume) due to the computational expense associated with the Filtered Back-Projection (FBP) reconstructions. DDMar was, on average, significantly more computa-

Method	Class	TPR (%)	FPR (%)	Precision	Accuracy
Intensity threshold	Handgun	90.87	6.0	0.955	0.922
	Bottle	21.23	3.68	0.816	0.637
NLM filter [BCM05b]	Handgun	96.63	1.33	0.990	0.975
	Bottle	100.0	2.11	0.973	0.988
LIMar [KHE87]	Handgun	95.70	2.67	0.980	0.964
	Bottle	89.04	1.58	0.977	0.944
DDMar (Section 4.2)	**Handgun**	**97.60**	**1.33**	**0.990**	0.980
	Bottle	**98.63**	**1.58**	**0.980**	0.985

Table 8.3: Overall classification performance for tested methods.

Method	Avg. Processing Times (s/volume)		
	Denoising	Segmentation	Total
Intensity threshold	0.23	245.87	246.10
NLM filter [BCM05b]	34.01	127.31	161.32
LIMar [KHE87]	49.05	145.67	194.72
DDMar (Section 4.2)	401.42	129.44	530.86

Table 8.4: Mean per-volume processing times for stages of automated classification (actual classification times negligible, thus not shown). Times have been averaged over both experiments (handguns and bottles).

tionally demanding than the other three methods. For the reasons discussed in Chapter 7, segmentation times were lower for volumes with higher signal-to-noise ratios (i.e. better denoising/artefact reduction). The mean segmentation times for the NLM pre-processed volumes (127.31s) and the LIMar and DDMar volumes (145.67s and 129.44s respectively) were thus significantly lower than those of the thresholded volumes (245.87s). Although the processing times associated with the final stage of the proposed framework (random forest clustering and SVM classification) are theoretically dependent on the number of objects segmented from the original volume (i.e. number of subvolumes), the average times for all four methods were negligible ($< 1.0s$) relative to the pre-processing and segmentation stages and are thus not shown in Table 8.4. The overall mean, per-volume processing time was thus lowest for the NLM-filtered volumes (161.32s).

8.3 Discussion

The relatively high correct handgun classification rates obtained using simple intensity thresholding may be attributed to the predominantly metallic nature of the handguns in the dataset. Their correspondingly high atomic numbers lead to significantly higher intensity values compared to the majority of other commonly encountered, low-density items (e.g. clothing and books) as well as high-density streaking artefacts. High-density, metallic items such as handguns, are thus fairly

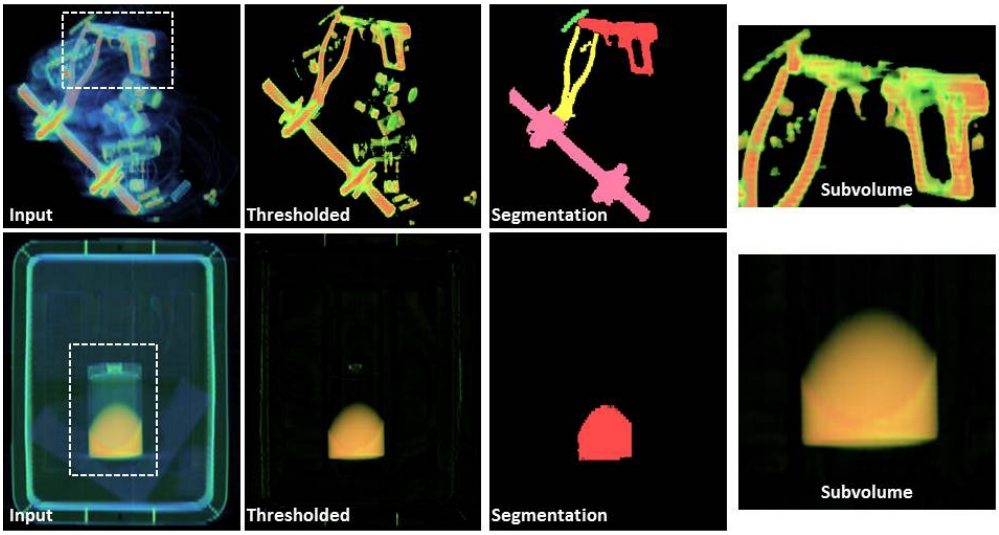


Figure 8.5: Threshold pre-processing examples. Top row: The handgun is correctly classified as the high-density information is not eliminated by the threshold. Bottom row: The bottle is misclassified as the threshold eliminates the empty (top) half of bottle.

easily isolated, even in cluttered and noisy volumes, using only a single threshold (a property which is exploited in most MAR techniques - see Chapter 4). The obvious shortcoming of segmentations performed in this manner, however, is the elimination of the majority of the contents of the scans. The top row of Figure 8.5 shows an example of a threshold which has been chosen such that the handgun is successfully segmented from a cluttered bag. Note, however, that only the high-density objects (the handgun, pliers and dumbbell) in the original scan are accurately depicted in the segmentation map, while the majority of the remaining items are eliminated. This limitation of pre-processing by thresholding is further emphasised by the massive decline in performance when applied to the bottle classification task (where the correct classification rate is significantly lower than random guessing - Table 8.3). The bottom row in Figure 8.5 illustrates an example of a bottle-containing volume which was incorrectly classified using intensity thresholding. Note that only the liquid-containing region of the bottle is retained in the segmentation while the remaining part of the bottle is eliminated by the threshold. The corresponding subvolume bears little resemblance to a bottle and is thus misclassified. Thresholding is thus only suitable when considering objects with very high densities - it is important to emphasise that these densities need to be higher than the high density noise/artefacts in the image. In the vast majority of scenarios such an approach will not suffice (as illustrated by the poor performance on the bottle dataset). At best, thresholding may be used as an initial screening for high-density threats.

Perhaps the most important observation that can be made from the results in Section 8.2, is the high-quality performance of the NLM filtering-based approach relative to the two MAR-based approaches (both in terms of classification perfor-

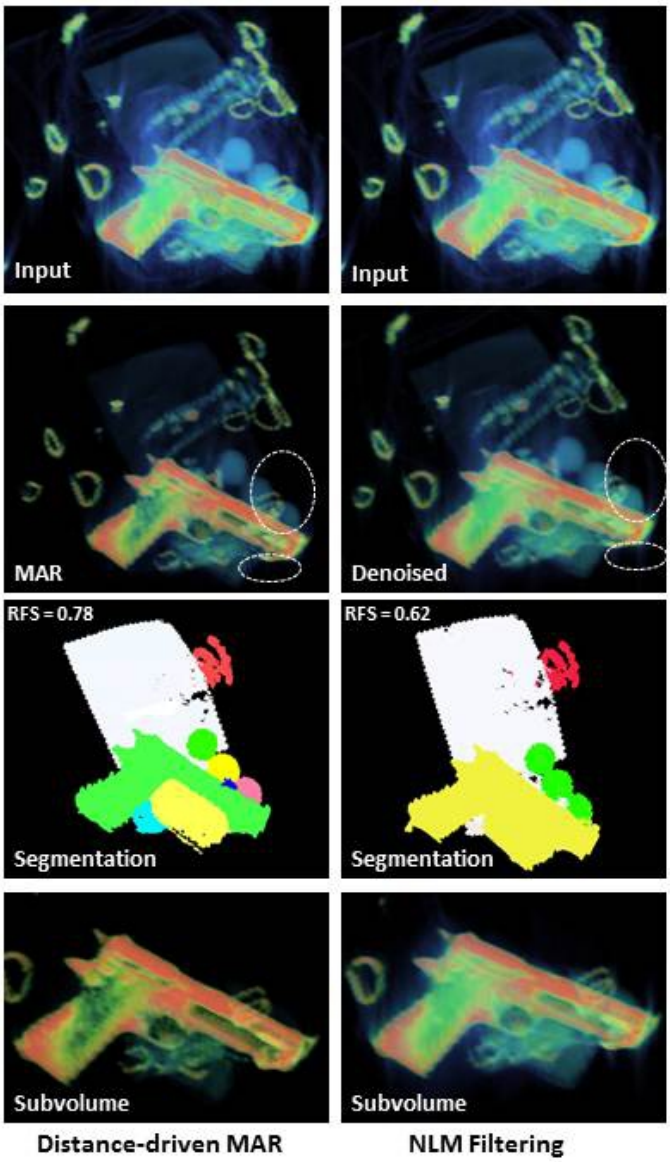


Figure 8.6: Comparison of DDMar and NLM pre-processed handgun segmentations which led to correct classifications. Superior artefact removal by MAR shown in second row. Although NLM leads to an under-segmentation of the handgun (third row, right column), the error is small and a positive classification is still returned.

mance as well as processing time). The experimentation presented in Chapter 4 suggested that dedicated MAR techniques (such as those considered in this chapter) outperform simple denoising filters (e.g. NLM filter) in the reduction of noise and metal-streaking artefacts in low-quality, complex volumetric imagery. The results in Section 8.2, however, bring into question the relevance of this superior artefact reduction. Figure 8.6, for example, shows a handgun-containing volume that was correctly classified by both the DDMar and NLM-filtering approaches. The MAR volume, however, contains notably less streaking than the NLM volume (clearly visible in the final subvolumes) and leads to a handgun segmentation that is superior to the under-segmented NLM handgun - both qualitatively as well as quantitatively (as determined by a higher Random Forest Score (RFS) - see

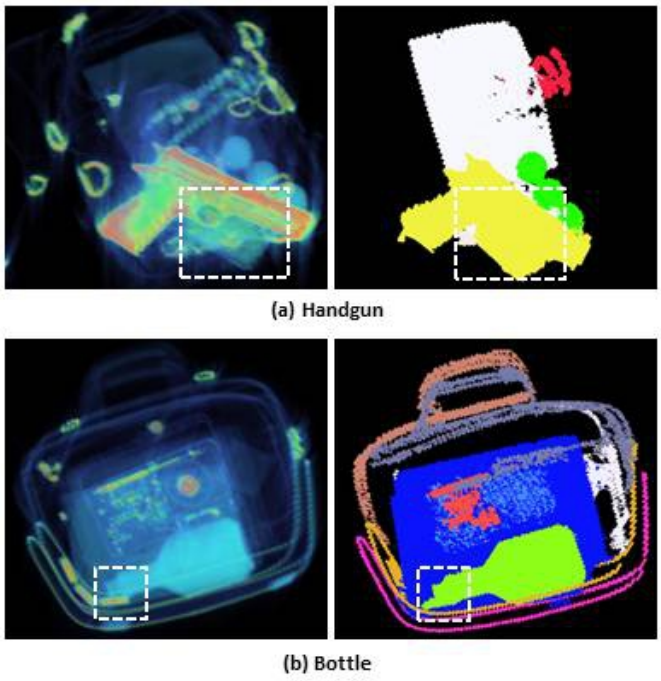


Figure 8.7: Under-segmentations (outlined regions) resulting from NLM filtering. Since errors are small, classification is not affected.

Chapter 7). The key factor appears to be in the generation of the subsequent subvolumes. Since the original (denoised) volume information is retained in these subvolumes (as opposed to retaining only the foreground/object information) the errors related to the incorrect labelling in the under-segmented object are largely eliminated - note that both the MAR and NLM subvolumes contain the object which was incorrectly labelled in the NLM segmentation. In Chapter 5 it was shown that classification performance using randomised clustering forests is relatively robust to background noise and clutter, provided the clutter objects are small in relation to the target object (i.e. the target object forms the main part of the subvolume). It is apparent (as illustrated by the additional examples in Figure 8.7), that for both the handgun and bottle datasets, the NLM segmentations produced sufficiently small under-segmentations to allow for correct classifications. This is an indication that the degree of artefact reduction offered by more complex MAR techniques (which comes at a significant increase in computational cost - Table 8.4) as well as very precise segmentations (such as those provided by the computationally demanding IDT algorithm [Gra06]) are not necessary for the successful classification of objects in complex volumetric imagery. This observation is particularly encouraging in the context of security screening, where the demands for high throughput are paramount [Sin03]. It is worth emphasising, however, that some degree of denoising is still important, as illustrated by the comparatively poor performance of simple intensity thresholding.

Considering the comparable classification performance of NLM and DDMar, the fact that DDMar employs the NLM filter [MMB+13] and the comparatively

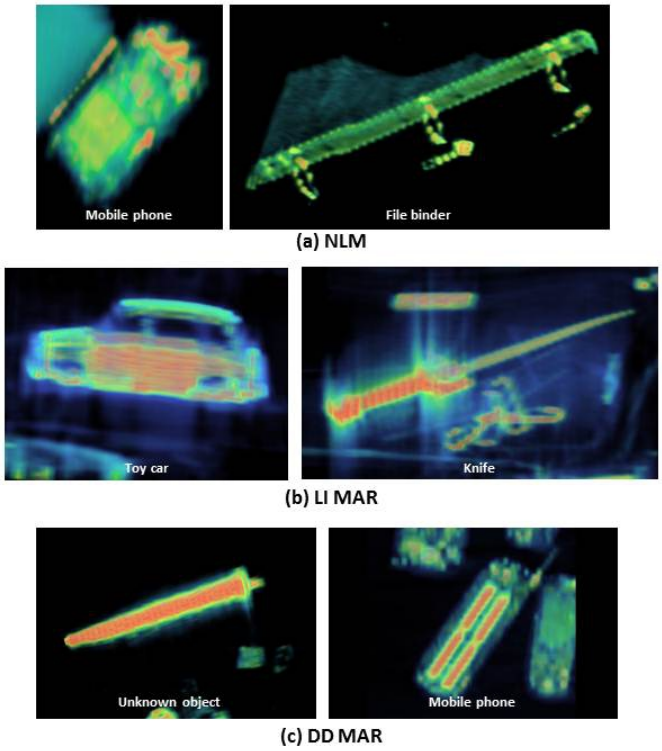


Figure 8.8: Handgun false-positive examples caused by high-density objects.

poor performance of LIMar, suggests that the most beneficial component (in the context of object classification) of the MAR approach is in fact the pre-filtering and post-filtering using the NLM filter.

A more detailed analysis of the classification errors produced by the NLM, LIMar and DDMar pre-processing techniques is important. Figures 8.8 - 8.10 show examples of misclassified handgun volumes for each of the three methods.

Figure 8.8 illustrates examples of two subvolumes that resulted in false-positive handgun classifications. The most obvious trend (evident for all three pre-processing techniques) is the presence of high-density objects. Beyond this, the subvolumes bear little obvious resemblance to the handguns in the training set (e.g. Figure 8.3), making it difficult to determine, with any confidence, the root of the missed classifications.

Further investigation of the results have indicated apparent trends in the characteristics of the missed NLM and DDMar handguns (false negatives). In particular, the missed handguns contain relatively large low-density regions in their grips/handles, resulting in over-segmentations of the handguns into separate handle and barrel components. Examples of such handguns (together with their corresponding segmentation maps and relevant subvolumes) are shown in Figure 8.9 (a) and (b) for NLM and DDMar respectively. Interestingly, the handgun barrel subvolumes that have led to false negatives are notably similar in appearance to the false-positive generating subvolumes for all three methods (Figure 8.8) and certainly bear resemblance to the whole gun subvolumes. This suggests that the

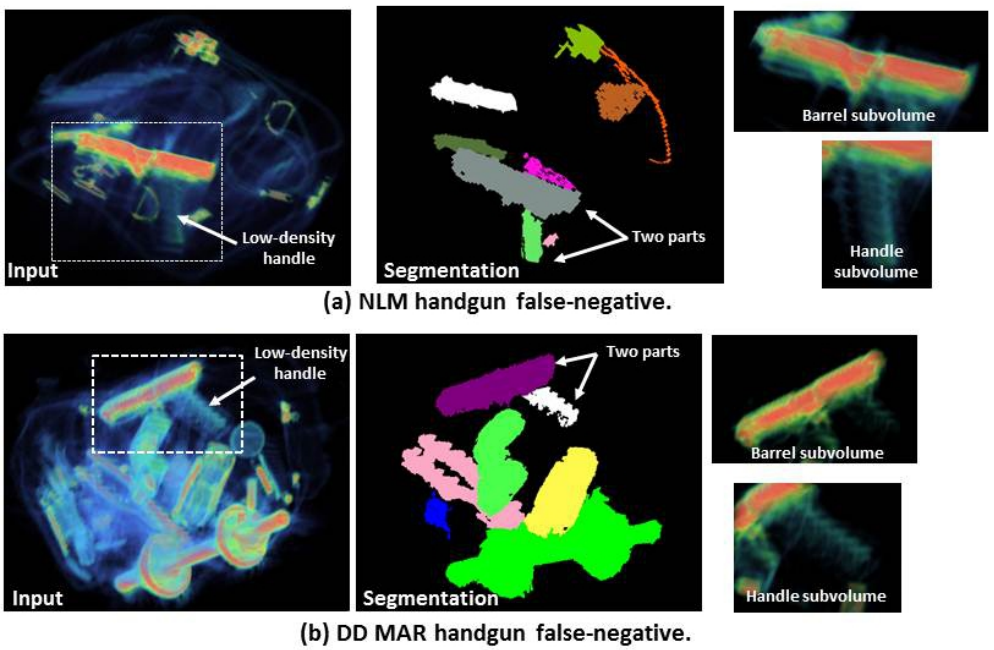


Figure 8.9: NLM and DDMar handgun false-negative examples. Handguns (outlined) are over-segmented into barrels and handles due to uncharacteristically low-density handles. Resulting subvolumes classified as clutter.

false-positive and false-negative instances lie very close to the decision boundary established by the classifier. It is likely that a larger and more diverse training set and/or a finer tuning of the input parameters may alleviate these errors. Furthermore, it is worth noting that the fact that handguns with these properties (low-density handles) were included in the classification training data is immaterial, as the errors essentially occur in the segmentation phase of the pipeline. This again highlights the challenge of defining a suitable 'object philosophy' in the development of a segmentation algorithm [CMP13]. Hierarchical approaches to segmentation, whereby the relation between the individual parts of composite objects are stored in tree-like structures, allow for multi-part objects to be represented both by their constituent parts as well as single composite objects [WGW12]. The incorporation of such techniques into the proposed framework is likely to be beneficial in the aforementioned scenarios and is left as an area for future work.

It has been found that the main cause of false-negative handgun classifications for the LIMar method is an introduction of new streaking artefacts that arise from the FBP reconstructions of linearly interpolated sinograms (see Chapters 3 and 4 for details on the mechanism of this phenomenon). The effective increase in streaking in turn leads to under-segmentations (i.e. multiple objects segmented as one). Two such examples are shown in Figure 8.10. In both cases, the subvolumes containing the handguns also contain the majority of the large items present in the original scans. While the NLM under-segmentations are typically small (Figure 8.7), the LIMar under-segmentations contain objects similar in size (or larger)

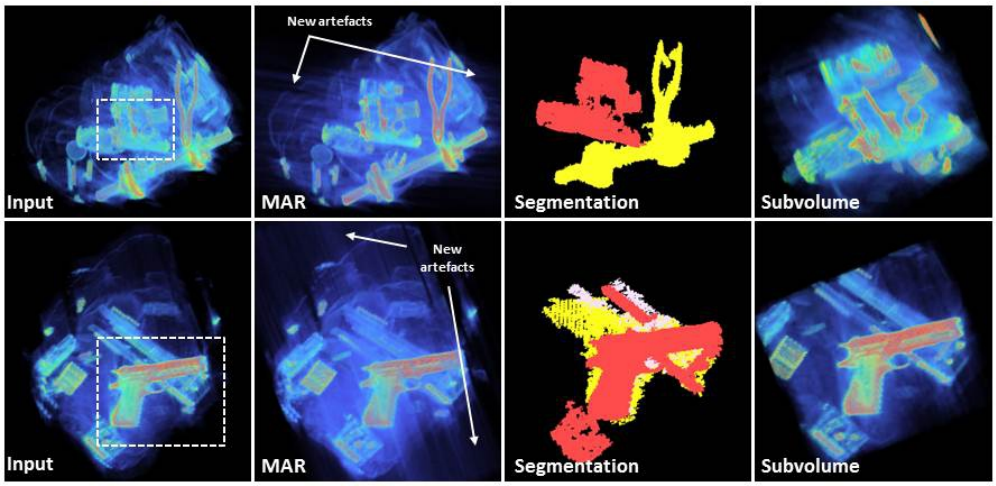


Figure 8.10: LIMar handgun false-negative examples. New streaking introduced in MAR procedure leads to over-segmentations of the handguns and hence incorrect classifications. Handguns outlined in inputs.

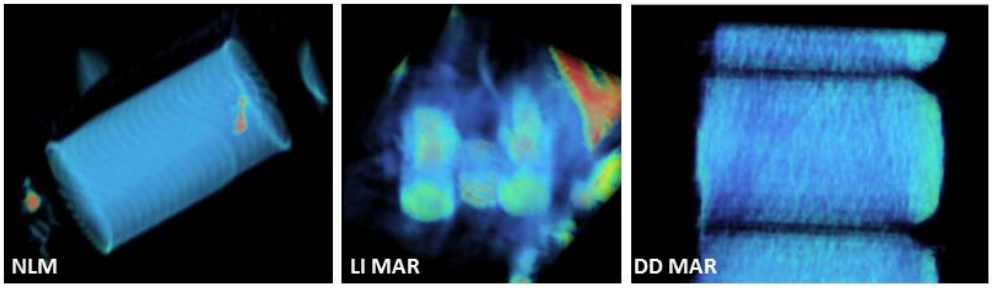


Figure 8.11: Bottle false-positive examples. Caused by objects with circular cross-sections and similar densities to liquids.

than the target objects, leading to erroneous classifications.

Figures 8.11 - 8.13 show examples of misclassified bottle volumes for the NLM filtering and two MAR-based pre-processing methods. The subvolumes that led to false-positive classifications for all three pre-processing techniques are dominated by objects with circular cross-sections and densities in the range of common liquids (Figure 8.11). This is a similar observation to that made in Chapter 5. It is worth noting that not all such objects resulted in false-positive classifications and the reasons for these particular misclassifications are not clear.

NLM filtering returns a perfect classification of the positive (bottle-containing) test volumes, while DDMar results in only two false negatives. A closer examination of these two cases has indicated that both false positives are caused by half-filled bottles surrounded by high-density objects and hence corrupted (i.e. intersected) by streaking artefacts. One of the two volumes is shown in Figure 8.12 in addition to the corresponding denoised volume, the segmentation map and the subvolume for NLM filtering. Due to the presence of four high-density objects in the input volume (indicated in Figure 8.12), the degree of streaking is severe, making it challenging to distinguish the upper border of the bottle. The bottom half of the bottle is more pronounced due to the presence of a higher-density liq-

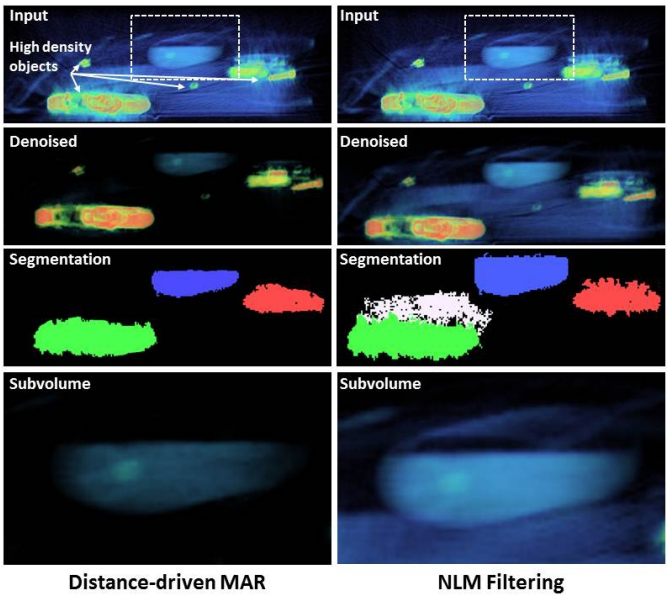


Figure 8.12: Comparison of DDMar (false negative) and NLM (true positive) bottle segmentation/classification. DDMar removes greater degree of artefacts, eliminating the empty region in the bottle and causing a misclassification.

uid (relative to the plastic of the bottle). While DDMar successfully removes the artefacts, it also appears to have over-compensated and removed some degree of important image information. Most importantly, the upper border of the bottle is almost entirely eliminated. Consequently, only the liquid-containing region of the bottle is segmented and represented in the corresponding subvolume. In contrast, while NLM filtering removes considerably less streaking (as expected), the entire bottle is retained. The resulting segmentation map, although including background noise in the bottle region, captures all of the relevant information and hence the subvolume is correctly classified. This again illustrates the point that despite an under-segmentation of the bottle and surrounding noise, since the error is relatively small, classification is not affected. It is worth noting that, in the case of the MAR volume, the information related to the upper half of the bottle is already eliminated prior to the segmentation (since it is the pre-processed volume that is used in the generation of the subvolume). A possible solution may thus be to use the original, unprocessed volume as the input to the subvolume generation.

The false negatives resulting from the LIMar pre-processed volumes again appear to stem from large under-segmentations, caused by the introduction of new streaking artefacts in the MAR procedure. Figure 8.13 illustrates two such examples. In both instances the post-MAR volumes contain new streaking artefacts, leading to under-segmentations. The limitations of linear interpolation-based MAR in the presence of multiple metal objects, which are already well documented [MMvS+13], are further substantiated here.

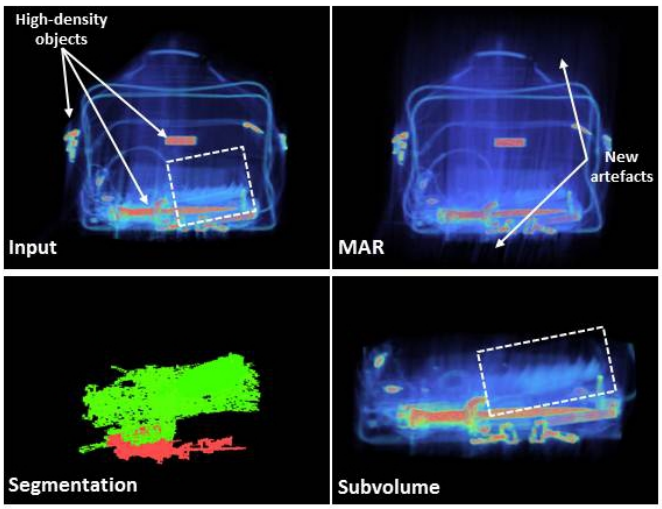


Figure 8.13: Example of LIMar false-negative classification of a bottle (outlined).

8.4 Conclusion

The limitations in the current state-of-the-art in 3D object classification within complex volumetric imagery [FBM12] have been addressed by the development of an efficient, fully-automated three-stage classification framework. The proposed classification pipeline is constructed by combining the pre-processing, CCA segmentation and random forest-based SVM classification techniques presented in the preceding chapters of this work. Correct classification rates in excess of 97% with false-positive rates of less than 2% are obtained for the classification of two object classes (handguns and bottles) in low-quality, complex volumetric imagery.

An experimental comparison to investigate the impact on classification performance of four denoising and/or Metal Artefact Reduction (MAR) techniques (intensity thresholding, NLM filtering [BCM05a, BCM05b], LIMar [KHE87], DDMar [MMB+13]), has demonstrated the superiority of the NLM filtering and DDMar over LIMar and intensity thresholding techniques. Furthermore, NLM filtering is shown to outperform both MAR-based approaches in terms of runtime. Considering the comparable classification performance obtained using the NLM and DDMar pre-processed volumes, it is reasonable to conclude that the benefits of superior artefact and noise reduction of the MAR process do not sufficiently justify the large associated increase in processing time. These experimental results may be considered a system-level quantitative evaluation of the denoising and artefact reduction techniques.

Although it may be argued that the target objects considered in this study are comparatively easy to classify (particularly the high-density handguns), the currently available dataset has limited the study to these two classes. A seemingly ubiquitous demand in the computer-vision community is for larger, more standardised datasets. Naturally, due to the sensitivity of the security data considered in this study, the data gathering process is not straightforward. Nonetheless, the

expansion of the current dataset (which is comparable to those used in the earlier works of Flitton *et al.* [FBM12, FBM13]) and the inclusion of a broader range of target classes is left as an area for future work.

The proposed classification framework has improved on the current state-of-the-art [FBM12, FMMB13] in the following ways: 1) the effects of noise and artefacts have been addressed by pre-filtering and metal artefact reduction; 2) the previously manual generation of subvolumes has been automated by the development of a dual-energy-based segmentation technique and 3) classification processing time has been reduced by several orders of magnitude using randomised-clustering forests. Furthermore, classification performance remains comparable to that achieved by the semi-automated visual-cortex approach of [FBM12] and improves upon the previous state-of-the-art in fully-automated object classification [FMMB13].

Chapter 9

Conclusions

The automated analysis of low-resolution, complex 3D baggage-CT imagery has been addressed through the development of techniques for image denoising, artefact reduction, segmentation and classification. The outcome of this thesis has been a novel, fully-automated framework for the classification of objects within this domain.

This chapter presents an overview of the research conducted in this thesis (Section 9.1), acknowledges concurrent and more recent studies in the wider research community (Section 9.2), reviews the most important original contributions (Section 9.3) and suggests avenues for future developments (Section 9.4).

9.1 Summary of Research

Three-dimensional X-ray Computed Tomography (CT), initially developed for use in the medical-imaging domain, has fairly recently been incorporated into the aviation security infrastructure in the form of sophisticated Explosives Detection Systems (EDS). The objective of this thesis has been the development of a fully-automated, yet computationally efficient 3D object-classification framework for low-resolution, complex baggage-CT imagery. The development of this framework is shown to rely on three components: 1) image quality improvement; 2) 3D object segmentation and 3) 3D object classification. These three topics form the foundation of the research conducted in this thesis.

Although prior work considering the application of computer-vision techniques to 3D volumetric baggage-CT imagery is relatively limited, a vast resource of related studies have been conducted in the medical domain. A review of these related studies was presented in (Chapter 3) and led to several important observations which formed the bases of the novel techniques presented in the latter chapters of this work. In particular, prior denoising and artefact-reduction techniques were shown to be almost entirely limited to the medical domain. Furthermore, the overwhelming trend observed in the medical literature was the exploitation

of *a priori* knowledge related to the properties of the anatomical regions being scanned (e.g. size, shape, density, spatial context etc.), allowing for the adjustment algorithm parameters for specific anatomical regions or abnormalities. This detracts from the suitability of these techniques in the security-screening domain, where the imagery is of a poorer quality, is more cluttered and where no *a priori* information related to the objects being scanned exists. The unsupervised segmentation of cluttered volumetric baggage-CT imagery has been considered previously, but high-resolution medical-grade imagery with relatively low levels of noise and metal-streaking artefacts were employed [CMP13]. The segmentation of low, anisotropic resolution volumetric imagery in the presence of multiple metal objects had not been considered previously. Finally, the state-of-the-art in 3D object classification in baggage-CT imagery [FBM12] was shown to be computationally demanding, sensitive to image noise and artefacts and dependent on manual segmentations.

The reduction of noise and metal-streaking artefacts in the previously unconsidered context of low-resolution, cluttered baggage-CT imagery was then addressed (Chapter 4). Two experimental comparisons were conducted, which considered both denoising as well as metal artefact-reduction. In contrast to the majority of existing comparative studies, a broad range of techniques was considered and qualitative as well as quantitative performance analyses were performed. The two most important observations were: 1) the notable declines in the performance of the state-of-the-art medical MAR techniques when applied in imaging domains where the generation of accurate priors is difficult and 2) a characteristic blurring effect common to all sinogram-completion-based MAR approaches. Comparative studies of this nature had not previously been conducted (in either the security-screening or medical-imaging domains) and represent notable contributions to knowledge. Additionally, a novel method for quantifying the performance of denoising or MAR algorithms when applied to 3D imagery was presented. The performance measure quantifies denoising performance based on the ratio of object to noise 3D-SIFT interest points and provides an indication of the likely impact of denoising or MAR on subsequent object recognition performance. Finally, a novel sinogram-completion MAR technique was presented, designed particularly for baggage-CT imagery containing multiple metal objects. The proposed method, which employs a distance-driven weighting scheme to adjust the corrected voxel intensities according to their locations relative to the metallic objects within the scan, was shown to perform comparably to state-of-the-art medical MAR techniques, particularly within the baggage-CT domain.

The feasibility of a codebook approach for 3D object classification in low-resolution, cluttered baggage-CT imagery was then investigated (Chapter 5). To this end, five codebook models were constructed using various combinations of

sampling strategies, feature-encoding techniques and classifiers. A codebook model constructed using an Extremely Randomised Clustering (ERC) forest [MTJ07] for feature encoding, a dense-feature sampling strategy and an SVM classifier resulted in improvements over the state-of-the-art 3D visual cortex model [FBM12] both in terms of classification accuracy as well as processing time. High-speed feature-encoding techniques had not been considered previously in this domain, thereby preventing the incorporation of dense-feature sampling strategies.

An experimental investigation into the efficacy of Dual-Energy Computed Tomography (DECT) techniques, when applied to low-resolution, cluttered baggage-CT data, was then presented (Chapter 6). The discriminative capabilities of the Dual-Energy Index (DEI) [Joh11] (previously unconsidered in the security-screening domain) and the effective atomic number Z_{eff} [WdB69] were qualitatively and quantitatively evaluated using five manually annotated object classes. The DEI was shown to outperform the effective atomic number both in terms of accuracy as well as discriminative power. The comparatively poor performance of the Z_{eff} measure was attributed to errors resulting from the approximations of the scanner configuration and the use of simulated energy spectra. Crucially, high-quality material discrimination was shown to be dependent on sufficient metal artefact reduction. The observations made in this chapter laid the foundation for the development of a novel materials-based segmentation technique.

A novel materials-based technique was proposed for the segmentation of unknown objects from low-resolution, cluttered volumetric baggage-CT data (Chapter 7). To this end, four novel contributions were made: 1) a materials-based coarse segmentation technique; 2) a random-forest-based model for measuring the quality of individual object segments; 3) a random-forest-based model for measuring the quality of entire segmentations and 4) a segmentation-refinement procedure for splitting fused objects. Coarse segmentations are generated using a characterisation of the chemical composition of an image (based on the DEI [Joh11]), simple thresholding operations and connected component analysis. The quality of the individual components of the coarse segmentations are then evaluated using the Random Forest Score (RFS) - which is trained to recognise single-object segments. Based on the RFS of a given coarse segmentation, low-quality individual object segments are subjected to an object-partitioning operation which splits fused objects at automatically-detected regions using a simple connected component analysis. A second segmentation-quality measure was presented for quantifying the quality of a full segmentation. The measure only requires prior knowledge of the number of objects in a given image. In a comparative evaluation, the proposed segmentation method was shown to perform only marginally worse than the state-of-the-art [GSK+12] in terms of segmentation quality (both qualitatively as well as quantitatively) but at a reduction in computational cost.

The importance of MAR was demonstrated by the decline in performance for all segmentation techniques when MAR is not applied.

Finally a novel, fully-automated and efficient framework for 3D object classification within cluttered baggage-CT imagery was presented by combining the pre-processing (denoising and MAR), segmentation and classification techniques developed in the preceding chapters of this work (Chapter 8). Dedicated MAR resulted in only marginal improvements in classification accuracy over simple Non-Local Means (NLM) filtering [BCM05b] but at a significant increase in processing time. It was concluded that the benefits of superior reduction of artefacts and noise of dedicated MAR do not sufficiently justify the large associated increase in processing time. The fully-automated classification framework presented represents an extension to the current state-of-the-art [FBM12, FMMB13]. In particular, classification performance in the presence of noise and artefacts was improved via pre-filtering; the previously manual generation of subvolumes was automated by the development of a dual-energy-based segmentation technique and classification processing times were reduced using randomised-clustering forests.

9.2 Concurrent Work

Since the commencement of this research, concurrent work in the wider research community has been conducted in related fields. Most relevant to this thesis are the recent developments in MAR and tomographic-reconstruction for baggage-CT data.

Metal artefact reduction in CT is an open problem and additions to the state-of-the-art are frequent. Since the completion and publication of the novel research presented in Chapters 3.2 and 4, a particularly relevant advancement in metal artefact reduction designed specifically for baggage-CT imagery has been presented by Karimi *et al.* [KCM13]. The proposed approach is the first of its kind to employ a prior image in this domain. Prior-images are constructed as solutions to constrained numerical optimisation problems. Particularly, a regularised Weighted Least Squares (WLS) error is minimised, where the regularisation is performed via the total variation norm. Artefact reduction is predominantly achieved via the weighting scheme (which is chosen to de-emphasises metal) and the chosen constraint (which exploits the fact that low-frequency metal artefacts are caused by beam hardening and scattered radiation). To reduce computational overhead, the size of the convex problem is decreased by solving for a smaller image. The sinograms are filtered and downsampled in views and samples. The employment of this so-called 'miniature image' is shown to lead to a reduction in reconstruction time by a factor of 16^3. Once the miniature image has been constructed, a second miniature image is constructed in a similar manner but ignoring the afore-

mentioned weights and constraints. This second image represents the original, artefact-corrupted image in the same coordinate space as the first miniature image. An artefact-only miniature image is then computed as the difference between the two miniature images and upsampled to the original dimensions, yielding the so-called prior image. The sinogram of this prior image is computed and used to guide the replacement of the metal trace in the original corrupted sinogram (using a standard interpolation-based approach). The technique is shown to outperform the linear-interpolation-based approach of Kalender *et al.* [KHE87] as well as a more recent iterative projection replacement method [VS12] - particularly in terms of preservation of image details. As with all interpolation-based approaches, some degree of blurring is observed. It is also worth noting that experiments were performed using medical-grade imagery (obtained from the ALERT initiative [CMP13]) which is not representative of that encountered in the aviation security domain. Nonetheless, this study represents the current state-of-the-art in sinogram-completion-based MAR in the baggage-CT domain and its incorporation into the automated classification framework presented in this thesis is likely to be beneficial and is left as an area for future work.

Tomographic reconstruction: The ALERT initiative [CMP13], which led to the development of several dedicated baggage-CT segmentation algorithms (Section 3.5) has recently released the results of a second phase of the initiative addressing the role of CT reconstruction in explosives detection [CKM]. Due to the late release of this report (relative to the time frame of this thesis), the techniques presented have not been included in this research. The most pertinent observations of the initiative are discussed below.

In total nine independent medical research groups were tasked with developing advanced reconstruction algorithms to improve image quality and explosives detection in baggage-CT imagery. Of these nine groups, eight used raw projection data to directly develop reconstruction techniques, while the ninth group was tasked with developing simulation tools to mitigate the computational expense of complex reconstruction techniques. The initiative has led to the development of several novel or modified reconstruction algorithms based on iterative-reconstruction techniques, sinogram pre-processing; dual-energy techniques and modifications to the FBP process. These methods (which employ both single-energy and dual-energy techniques) are shown to offer varying degrees of improvements in image quality (in terms of MAR and contrast enhancement). The degree of quality improvement is shown to correlate with the complexity of the reconstruction technique (i.e. better image quality comes at an increased computational expense). The reconstruction techniques are further shown to benefit subsequent explosive detection rates (according to a feature-based performance metric). This is largely attributed to the associated reduction in streaking artefacts and the improvements

in image contrast. Similarly to the earlier ALERT initiative [CMP13] and the recent MAR study of Karimi *et al.* [KCM13], these studies employed medical-grade CT imagery.

In conjunction with the ALERT initiative [CKM], Eger *et al.* [EIKP11, EDI+11] have demonstrated improved explosives detection using machine learning techniques based on Multi-Energy Computed Tomography (MECT). MECT is claimed to provide superior characterisation of the chemical composition of the materials in a scan relative to conventional DECT. Low-dimensional features are extracted from the high-dimensional 'Linear Attenuation Coefficient (LAC) vs. energy' curves of materials and are shown to outperform the traditional photoelectric and Compton coefficients in terms of discriminative capabilities. The studies suggest improved detection performance relative to conventional dual-energy X-ray systems [EIKP11, EDI+11].

While the detection of explosives was not considered in this thesis, the results of the aforementioned CT reconstruction studies [CKM, EIKP11, EDI+11] indicate potential benefits of incorporating such techniques into the CCA segmentation framework presented in Section 7.2 of this work and this is again left as an area for future work.

9.3 Review of Contributions

The most important original contributions of this thesis may be summarised as follows:

- A novel interest-point-based quantitative performance measure is presented, extending traditional denoising performance evaluation approaches by evaluating the potential benefits of denoising on the application of more complex operations (e.g. 3D object classification) within the current imaging context (published as [MMFB12]).

- A novel Metal Artefact Reduction (MAR) technique, designed specifically for cluttered baggage-CT imagery containing multiple metal objects, is presented and shown to perform comparably to state-of-the-art medical techniques when applied to cluttered baggage-CT imagery (published as [MMB+13]).

- A comprehensive comparative performance evaluation (which has not previously been considered in medical or non-medical CT domains) is conducted for seven image denoising techniques and twelve artefact-reduction techniques (published as [MMFB13, MMvS+13]).

- A novel dual-energy-based segmentation technique is presented and shown to provide fast, high-quality segmentations of complex baggage-CT imagery.

Within the proposed framework, four novel contributions are made: 1) a materials-based coarse segmentation technique; 2) a random-forest-based model for measuring the quality of individual object segments; 3) a random-forest-based model for measuring the quality of entire segmentations and 4) an efficient segmentation-refinement procedure for splitting fused objects. In a comparative performance evaluation, the proposed technique is shown to perform comparably to the state-of-the-art [WGW12, Gra06, CMP13] (submitted as [MB14b]).

- A codebook image classification model constructed using random-forest-based feature encoding, a dense-feature sampling strategy and a Support Vector Machine (SVM) classifier is presented and shown to outperform the current state-of-the-art [FBM12] both in terms of accuracy as well as processing time (submitted as [MBF14]).

- A novel, fully-automated and efficient framework for 3D object classification within cluttered baggage-CT imagery is presented and shown to improve on the current state-of-the-art [FBM12] by reducing the detrimental effects of image noise and artefacts; by automating the segmentation process; by improving accuracy and by reducing computational cost (submitted as [MB14c]).

9.4 Future Work

The research presented in this thesis has led to the identification of several areas where future developments are likely to be beneficial.

9.4.1 Image Quality Improvement

The overwhelming consensus in the literature is that iterative-reconstruction techniques provide superior image quality to conventional FBP (particularly in terms of artefact reduction). Furthermore, despite the broad range of existing metal-artefact-reduction techniques, these perform comparatively poorly in the security-screening domain (Section 4.4.2), an observation which has been substantiated by the surprisingly small degree of improvement in classification performance produced by the dedicated baggage-CT MAR approach (Section 4.2) over NLM filtering [BCM05b] (Section 8.2). These observations are an indication that future baggage-CT systems will benefit from improved CT reconstruction (as opposed to the development of further sinogram-completion-based MAR techniques). In terms of baggage-CT image quality, the following areas are outlined as directions for future work:

- The development of iterative-reconstruction techniques suitable for the security-screening domain, with a particular focus on minimising processing times.

- The observations presented in Section 4.5 of this work have indicated that the current state-of-the-art medical MAR techniques benefit from the use of accurate priors but typically require careful parameter tuning. An exhaustive optimisation of the MAR parameter space and/or the development of techniques to automatically determine optimal MAR algorithmic parameters are thus recommended as areas for future work. Furthermore, the more recent work of Karimi *et al.* [KCM13] (Section 9.2) has demonstrated that the development of useful prior data is possible in the baggage-CT domain without significant compromises in computational cost. These techniques were however, evaluated predominantly on medical-grade CT imagery, which is not representative of the imagery encountered in the aviation security-screening domain. An evaluation of their performance in the current imaging domain is thus a necessary task.

- Further concurrent research discussed in Section 9.2 has revealed several important advancements in CT reconstruction for baggage-CT imagery. Although these advancements have led to improved image quality (particularly in terms of artefact reduction and contrast enhancement), the degree of this improvement has been shown to correlate with processing time (i.e. better image quality at higher processing times). This observation has been substantiated by the broader resource of iterative reconstruction-based literature. The demand for high throughput in security-screening settings has been highlighted throughout this thesis. The importance of determining an appropriate tradeoff between image quality and computational cost is thus a vital factor in future developments in tomographic reconstruction.

9.4.2 Segmentation

The observations made in Chapter 7 have indicated that the proposed dual-energy-based segmentation algorithm is potentially well-suited to the task of 3D object segmentation within baggage imagery, particularly owing to its comparatively low computational overhead. Nonetheless, several limitations in the current formulation suggest the following areas would benefit from further development:

- The manner in which composite objects (composed of multiple parts) are segmented is a source of ambiguity (e.g. over-segmentation of handguns in barrel and handles). Hierarchical approaches to segmentation, whereby the relation between the individual parts of composite objects are stored in tree-like structures, allow for multi-part objects to be represented both

by their constituent parts and as single composite objects [WGW12]. The incorporation of such techniques into the proposed framework is likely to be beneficial in the aforementioned scenarios.

- The proposed segmentation algorithm is based on the relatively crude material representation offered by the Dual-Energy Index (DEI). Due to the lack of accurate calibration data and the use of sub-optimal simulated data, an accurate estimation of effective atomic number has not been possible in this work. Calibration using at least some empirical data (e.g. actual scans of materials with known properties and/or the measured energy spectra of the scanner) is thus an important area for future work.

- The recent work of Eger *et al.* [EIKP11, EDI+11] has demonstrated the superior performance of feature-based multi-energy CT (compared to conventional DECT) in the materials-based discrimination of objects in baggage-CT imagery. Since the segmentation algorithm proposed in Section 7.2 is based on feature-based dual-energy techniques, it is likely that the incorporation of multi-energy CT techniques into this framework will lead to higher quality segmentations.

- Harvey *et al.* (University of East Anglia) [CMP13] present a baggage-CT segmentation algorithm based on multiscale sieves [BCPL96] (Section 3.5.2). It is noted that the sieves segmentation algorithm segments all objects at all scales, resulting in at least one channel always containing a segmentation of an object. It is proposed [CMP13] that instead of merging all channels into a single segmentation, a more intuitive approach would be to pass the channel images directly into some artificial intelligence system (e.g. a classifier, object detector or salient region detector). This recommendation is in line with the segmentation objectives considered in this thesis (i.e. to provide subvolumes for classification). An evaluation of the feasibility of the sieves segmentation algorithm is thus recommended as an area for future work.

9.4.3 Classification

Although improvements over the state-of-the-art in classification performance [FBM12] have been demonstrated in Sections 5.2 and 8.2, due to limitations in the current dataset, the classification tasks have been restricted to two object classes. In order for a baggage-CT classification tool to be of value in industry a broad range of threats must be detectable. The most important direction for future classification-related work is thus an extension to multiclass problems. While it has been suggested that the true benefits of random-forest-based classification [Cri11] have not been exploited due to this restriction to two object classes, it is likely that such techniques will be of increased value in multiclass problems.

9.4.4 Data

Owing to the sensitivity of security data and the related challenges in data gathering, a relatively limited dataset has been used throughout this study. Perhaps the most important direction for future work is the expansion of the current dataset to include a greater number of total images; to contain a broader range of target and clutter items; to perform a more stringent documentation of the data-gathering process (e.g. exact contents of bags; geometric properties; material properties etc.) and to capture relevant calibration and spectral measurements. An expanded dataset of this nature is likely to benefit all facets of future work.

Bibliography

[AAA+10] M. Abdoli, M. R. Ay, A. Ahmadian, R. A. J. O. Dierckx, and H. Zaidi. Reduction of dental filling metallic artifacts in CT-based attenuation correction of PET data using weighted virtual sinograms optimized by a genetic algorithm. *Medical Physics*, 37:6166–6177, 2010.

[AAAZ10] M. Abdoli, M. R. Ay, A. Ahmadian, and H. Zaidi. A virtual sinogram method to reduce dental metallic implant artefacts in computed tomography-based attenuation correction for PET. *Nuclear Medicine Communications*, 31(1):22–31, 2010.

[AAIW05] G. A, S. G. Armato III, and C. Wu. Vessel tree reconstruction in thoracic CT scans with application to nodule detection. *IEEE Transactions on Medical Imaging*, 24(4):486–499, 2005.

[AB83] D. Altman and J. Bland. Measurement in medicine: the analysis of method comparison studies. *The Statistician*, 32:307–317, 1983.

[ABGG+09] A. Akselrod-Ballin, M. Galun, J. M. Gomori, M. Filippi, P. Valsasina, R. Basri, and A. Brandt. Automatic segmentation and classification of multiple sclerosis in multichannel MRI. *IEEE Transactions on Biomedical Engineering*, 56(10):2461–2469, 2009.

[ADES07] A. R. Abdel-Dayem and M. R. El-Sakka. Fuzzy c-means clustering for segmenting carotid artery ultrasound images. In *Image Analysis and Recognition*, pages 935–948. Springer, 2007.

[AGDL09] A. Adams, N. Gelfand, J. Dolson, and M. Levoy. Gaussian KD-trees for fast high-dimensional filtering. *ACM Transactions on Graphics*, 28(3):21, 2009.

[AHH+09] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[AHR+11] J. Altenbernd, T. A. Heusner, A. Ringelstein, S. C. Ladd, M. Forsting, and G. Antoch. Dual-energy-CT of hypervascular liver lesions in patients with hcc: investigation of image quality and sensitivity. *European Radiology*, 21(4):738–743, 2011.

[AIGM98] S. G. Armato III, M. L. Giger, and H. MacMahon. Automated lung segmentation in digitized posteroanterior chest radiographs. *Academic Radiology*, 5(4):245–255, 1998.

[AK84] A. Andersen and A. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging*, 6(1):81–94, 1984.

[AM76] R. E. Alvarez and A. Macovski. Energy-selective reconstructions in X-ray computerised tomography. *Physics in Medicine and Biology*, 21(5):733–744, 1976.

[ASB+12] S. K. Adhikari, J. Sing, D. Basu, M. Nasipuri, and P. Saha. Segmentation of MRI brain images by incorporating intensity inhomogeneity and spatial information using probabilistic fuzzy c-means clustering algorithm. In *Proceedings of the IEEE International Conference on Communications, Devices and Intelligent Systems*, pages 129–132, 2012.

[ASM+11] O. Acosta, A. Simon, F. Monge, F. Commandeur, C. Bassirou, G. Cazoulat, R. de Crevoisier, and P. Haigron. Evaluation of multi-atlas-based segmentation of CT scans in prostate cancer radiotherapy. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1966–1969, 2011.

[AST04] R. E. Alvarez, J. A. Seibert, and S. K. Thompson. Comparison of dual energy detector system. *Medical Physics*, 31:556–565, 2004.

[AT13] A. Andreopoulos and J. K. Tsotsos. 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, 2013.

[AVV08] P. Anbeek, K. L. Vincken, and M. A. Viergever. Automated ms-lesion segmentation by k-nearest neighbor classification. *MIDAS Journal*, 2008.

[AZGA06] B. R. Abidi, Y. Zheng, A. V. Gribok, and M. A. Abidi. Improving weapon detection in single energy X-ray images through pseudocoloring. *IEEE Transactions on Systems, Man, and Cybernetics*, 36(6):784–796, 2006.

[Bab98] R. Babuska. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Norwell, MA, USA, 1st edition, 1998.

[Bal81] D. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[BB96] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

[BBHM81] W. R. Brody, G. Butt, A. Hall, and A. Macovski. A method for selective tissue and bone visualization using dual energy scanned projection radiography. *Medical Physics*, 8:353–357, 1981.

[BBLS05] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging*, 24(9):1151–1169, 2005.

[BCM05a] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005.

[BCM05b] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modeling and Simulation*, 4(2):490–530, 2005.

[BCM10] A. Buades, B. Coll, and J.-M. Morel. Image denoising methods. a new nonlocal principle. *SIAM Review*, 52(1):113–147, 2010.

[BCPL96] J. A. Bangham, P. Chardaire, C. J. Pye, and P. D. Ling. Multiscale nonlinear decomposition: The sieve decomposition theorem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):529–539, 1996.

[BCZX08] W. Bi, Z. Chen, L. Zhang, and Y. Xing. A volumetric object detection framework with dual-energy CT. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, pages 1289–1291, 2008.

[BCZX09] W. Bi, Z. Chen, L. Zhang, and Y. Xing. Fast detection of 3D planes by a single slice detector helical CT. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, pages 954–955, 2009.

[BD09] C. Belcher and Y. Du. Region-based SIFT approach to iris recognition. *Optics and Lasers in Engineering*, 47(1):139–147, 2009.

[BDM10] R. S. Brock, A. Docef, and M. J. Murphy. Reconstruction of a cone-beam CT image via forward iterative projection matching. *Medical Physics*, 37(12):6212, 2010.

[BdP96] J. Browne and A. B. de Pierro. A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. *IEEE Transactions on Medical Imaging*, 15(5):687–699, 1996.

[Ber97] D. P. Bertsekas. A new class of incremental gradient methods for least squares problems. *SIAM Journal on Optimization*, 7(4):913–926, 1997.

[Ber99] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[BETG08] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[Bez81] J. C. Bezdek. Models for pattern recognition. In *Pattern Recognition with Fuzzy Objective Function Algorithms*, pages 1–13. Springer, 1981.

[BF11] F. E. Boas and D. Fleischmann. Evaluation of two iterative techniques for reducing metal artifacts in computed tomography. *Radiology*, 259(3):894–902, 2011.

[BGMG+00] M. S. Brown, J. G. Goldin, M. F. McNitt-Gray, L. E. Greaser, A. Sapra, K.-T. Li, J. W. Sayre, K. Martin, and D. R. Aberle. Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function. *Medical Physics*, 27:592, 2000.

[BH87] M. J. Berger and J. Hubbell. *XCOM, Photon cross sections on a personal computer*. Center for Radiation Research, National Bureau of Standards, 1987.

[BHHC98] J. A. Bangham, J. R. Hidalgo, R. Harvey, and G. Cawley. The segmentation of images via scale-space trees. In *Proceedings of British Machine Vision Conference*, volume 1, pages 33–43, 1998.

[BHW10] A. Ben-Hur and J. Weston. A user's guide to support vector machines. *Methods in Molecular Biology*, 609:223–239, 2010.

[Big93] N. Biggs. *Algebraic graph theory*. Cambridge University Press, 1993.

[BJ01] Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. In *Proceedings IEEE International Conference on Computer Vision*, volume 1, pages 105–112, 2001.

[BK01] F. J. Beekman and C. Kamphuis. Ordered subset reconstruction for X-ray CT. *Physics in Medicine and Biology*, 46(7):1835–1844, 2001.

[BK04] J. F. Barrett and N. Keat. Artifacts in CT: Recognition and avoidance. *Radiographics*, 24(6):1679–1691, 2004.

[BL03] H. Byun and S.-W. Lee. A survey on pattern recognition applications of support vector machines. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(03):459–486, 2003.

[BL07] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[BM76] J. A. Bondy and U. S. R. Murty. *Graph theory with applications*, volume 290. Macmillan London, 1976.

[BMFU+07] S. Bouix, M. Martin-Fernandez, L. Ungar, M. Nakamura, M.-S. Koo, R. W. McCarley, and M. E. Shenton. On evaluating brain tissue classifiers without a ground truth. *Neuroimage*, 36(4):1207–1224, 2007.

[BN08] J. D. Bustard and M. S. Nixon. Robust 2D ear registration and recognition based on SIFT point matching. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2008.

[BNKF97] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone. *Genetic Programming: An Introduction: On the Automatic Evolution of Computer Programs and Its Applications (The Morgan Kaufmann Series in Artificial Intelligence)*. Morgan Kaufmann Publishers, 1997.

[BP02] V. Butler and R. W. Poole. Rethinking checked-baggage screening. Technical report, Reason Public Policy Institute Los Angeles, CA, 2002.

[Bra90] R. N. Bracewell. Numerical transforms. *Science*, 248(4956):697–704, 1990.

[Bre01] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[BS96] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Transactions on Image Processing*, 5(3):480–492, 1996.

[BS06] M. Bal and L. Spies. Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering. *Medical Physics*, 33:2852, 2006.

[BSM00] P. P. Bruyant, J. Sau, and J. J. Mallet. Streak artifact reduction in filtered backprojection using a level line based interpolation method. *Journal of Nuclear Medicine*, 41(11):1913–1919, 2000.

[BT09] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[BV08] R. C. Bilcu and M. Vehvilainen. Combined non-local averaging and intersection of confidence intervals for image de-noising. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1736–1739, 2008.

[Byr97] C. L. Byrne. Convergent block-iterative algorithms for image reconstruction from inconsistent data. *IEEE Transactions on Image Processing*, 6(9):1296–1304, 1997.

[BZ10] S. Belhassen and H. Zaidi. A novel fuzzy c-means algorithm for unsupervised heterogeneous tumor quantification in pet. *Medical Physics*, 37:1309, 2010.

[BZM07] A. Bosch, A. Zisserman, and X. Muoz. Image classification using random forests and ferns. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[CA79] G. B. Coleman and H. C. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.

[Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):679–698, 1986.

[Can99] N. Canterakis. 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition. In *In 11th Scandinavian Conference on Image Analysis*. Citeseer, 1999.

[CCFT06] L. Cheng, Y. Chen, T. Fang, and J. Tyan. Fast iterative adaptive reconstruction in low-dose CT imaging. In *Proceedings of the IEEE International Conference on Image Processing*, pages 889–892, 2006.

[CDAO95] R. R. Coifman, D. L. Donoho, A. Antoniadis, and G. Oppenheim. Translation-invariant de-noising. *Wavelets and Statistics*, pages 125–150, 1995.

[CDMC90] T. S. Curry, J. E. Dowdey, R. C. Murry, and E. E. Christensen. *Christensen's physics of diagnostic radiology*. Lippincott Williams and Wilkins, 4 edition, 1990.

[CEJB+07] D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. Random forests for classification in ecology. *Ecology*, 88(11):2783–2792, 2007.

[CET01] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[CGB06] W. Chen, M. L. Giger, and U. Bick. A fuzzy c-means (fcm)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced mr images. *Academic Radiology*, 13(1):63–72, 2006.

[CH87] K. Chuang and H. Huang. A fast dual-energy computational method using isotransmission lines and table lookup. *Medical Physics*, 14:186–192, 1987.

[Cha04] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in Analysis*, 625:195–199, 1970.

[Che91] L. Chen. The lambda-connected segmentation and the optimal algorithm for split-and-merge segmentation. *Chinese Journal of Computers*, 14:321–331, 1991.

[CHK97] D. E. Cullen, J. H. Hubbell, and L. Kissel. Tables and graphs of photon-interaction cross sections from 10 eV to 100 GeV derived from the LLNL evaluated photon data library. *Report UCRL-50400 (Lawrence Livermore National Laboratory, Livermore, CA). Vol 6, Parts: A and B*, 1997.

[CHTH94] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–365, 1994.

[CKM] C. Crawford, C. Karl, and H. Martz. Research and development of reconstruction advances in CT-based object detection systems - final report. Northeastern University, Boston. This report can be found at: https://myfiles.neu.edu/groups/ALERT/strategic_studies/TaskOrder03/TO3_FinalReport.pdf.

[CKS97] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[CKY08] R. Caruana, N. Karampatziakis, and A. Yessenalina. An empirical evaluation of supervised learning in high dimensions. In *Proceedings of the 25th ACM International Conference on Machine Learning*, pages 96–103, 2008.

[CLSR02] L. M. Chen, Y. Liang, G. A. Sandison, and J. Rydberg. Novel method for reducing high-attenuation object artifacts in CT reconstructions. In *Proceedings of SPIE*, volume 4684, page 841, 2002.

[CLVZ11] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. In *Proceedings British Machine Vision Conference*, 2011.

[CM02] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[CMP13] C. Crawford, H. Martz, and H. Piena. Segmentation of objects from volumetric CT data, 2013. Northeastern University, Boston. This report can be found at: https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf.

[CMS12] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 3642–3649, 2012.

[CNB+10] C. A. Coursey, R. C. Nelson, D. T. Boll, E. K. Paulson, L. M. Ho, A. M. Neville, D. Marin, R. T. Gupta, and S. T. Schindera. Dual-energy multidetector CT: how does it work, what can it tell us, and when can we use it in abdominopelvic imaging? 1. *Radiographics*, 30(4):1037–1055, 2010.

[Coo98] K. Cooke. X-ray interactions with matter. *Practical Radiotherapy: Physics and Equipment*, page 39, 1998.

[CPB+04] M. B. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J.-G. Villemure, and J.-P. Thiran. Atlas-based segmentation of pathological mr brain images using a model of lesion growth. *IEEE Transactions on Medical Imaging*, 23(10):1301–1314, 2004.

[Cri11] A. Criminisi. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*, 7(2-3):81–227, 2011.

[CRLR12] A. Y.-S. Chia, D. Rajan, M. K. Leung, and S. Rahardja. Object recognition by discriminative combinations of line segments, ellipses, and appearance features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1758–1772, 2012.

[CS96] G. Celeux and G. Soromenho. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2):195–212, 1996.

[CS13] A. Criminisi and J. Shotton. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer Publishing Company, Incorporated, 2013.

[CSB09a] A. Criminisi, J. Shotton, and S. Bucciarelli. Decision forests with long-range spatial context for organ localization in CT volumes. In *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, 2009.

[CSB09b] A. Criminisi, J. Shotton, and S. Bucciarelli. Decision forests with long-range spatial context for organ localization in CT volumes. In *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, 2009.

[Csi75] I. Csiszar. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.

[CSRK11] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu. Regression forests for efficient anatomy detection and localization in CT studies. *Medical Computer Vision.Recognition Techniques and Applications in Medical Imaging*, pages 106–117, 2011.

[CTCG95] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[CV01] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[CZLY13] W. Cai, D. Zhang, J.-G. Lee, and H. Yoshida. Low-dose dual-energy electronic cleansing for fecal-tagging CT colonography. In *SPIE Medical Imaging*, pages 86700W–86700W. International Society for Optics and Photonics, 2013.

[DB03] M. Dettling and P. Bühlmann. Boosting for tumor classification with gene expression data. *Bioinformatics*, 19(9):1061–1069, 2003.

[DBST02] A. Demiriz, K. P. Bennett, and J. Shawe-Taylor. Linear programming boosting via column generation. *Machine Learning*, 46(1-3):225–254, 2002.

[Dea93] S. R. Deans. *The Radon Transform and Some of Its Applications*. Krieger Publishing Company, 1993.

[DFC+11] J. A. Dowling, J. Fripp, S. Chandra, J. P. W. Pluim, J. Lambert, J. Parker, J. Denham, P. B. Greer, and O. Salvado. Fast automatic multi-atlas segmentation of the prostate from 3D mr images. In *Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions*, pages 10–21. Springer, 2011.

[Dic45] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[Die00] T. G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2):139–157, 2000.

[DJ94] D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425, 1994.

[DJ95] C. Dorai and A. K. Jain. COSMOS-a representation scheme for free-form surfaces. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1024–1029, 1995.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society.Series B (Methodological)*, pages 1–38, 1977.

[DOM02] S. Dreiseitl and L. Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35(5):352–359, 2002.

[Don95] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Technology*, 41(3):613–627, 1995. ID: 1.

[Don07] B. Y. Dong. Image reconstruction using EM method in X-ray CT. In *Proceedings of the IEEE International Conference on Wavelet Analysis and Pattern Recognition.*, volume 1, pages 130–134, 2007.

[DUDA06] R. Díaz-Uriarte and S. A. De Andres. Gene selection and classification of microarray data using random forest. *BioMed Central Ltd. Jounral of Bioinformatics*, 7(1):3, 2006.

[DXDL10] W. Deng, W. Xiao, H. Deng, and J. Liu. Mri brain tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve. In *Proceedings of the IEEE International Conference on Biomedical Engineering and Informatics*, volume 1, pages 393–396, 2010.

[DYC03] J. Dehmeshki, X. Ye, and J. Costello. Shape based region growing using derivatives of 3D medical images: application to semiautomated detection of pulmonary nodules. In *Proceedings IEEE International Conference on Image Processing*, volume 1, pages I–1085, 2003.

[Dys05] N. A. Dyson. *X-rays in Atomic and Nuclear Physics*. Cambridge University Press, 2005.

[DZ11] M. C. Delfour and J.-P. Zolésio. *Shapes and geometries: metrics, analysis, differential calculus, and optimization*, volume 22. SIAM, 2011.

[DZS+07] X. Deng, L. Zhu, Y. Sun, C. Xu, L. Song, J. Chen, R. D. Merges, M.-P. Jolly, M. Suehling, and X. Xu. On simulating subjective evaluation using combined objective metrics for validation of 3D tumor segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 977–984. Springer, 2007.

[DZX+08] X. Duan, L. Zhang, Y. Xiao, J. Cheng, Z. Chen, and Y. Xing. Metal artifact reduction in CT images by sinogram TV inpainting. In *Nuclear Science Symposium Conference Record*, pages 4175–4177, 2008.

[EBPP12] I. Ersoy, F. Bunyak, J. Peng, and K. Palaniappan. HEp-2 cell classification in IIF images using Shareboost. In *Proceedings of the IEEE International Conference on Pattern Recognition*, pages 3362–3365, 2012.

[Edd96] S. R. Eddy. Hidden Markov Models. *Current opinion in structural biology*, 6(3):361–365, 1996.

[EDI+11] L. Eger, S. Do, P. Ishwar, W. C. Karl, and H. Pien. A learning-based approach to explosives detection using multi-energy X-ray computed tomography. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2004–2007, 2011.

[EF99] H. Erdogan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Physics in Medicine and Biology*, 44(11):2835, 1999.

[EF02] I. A. Elbakri and J. A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Transactions on Medical Imaging*, 21(2):89–99, 2002.

[EF03] I. A. Elbakri and J. A. Fessler. Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation. *Physics in Medicine and Biology*, 48(15):2453–2477, 2003.

[EHIS+08] C. Eusemann, D. R. Holmes III, B. Schmidt, T. G. Flohr, R. Robb, C. McCollough, D. M. Hough, J. E. Huprich, M. Wittmer, H. Siddiki, et al. Dual energy CT: How to best blend both energies in one fused image? In *Medical Imaging*, pages 691803–691803. International Society for Optics and Photonics, 2008.

[EIKP11] L. Eger, P. Ishwar, W. Karl, and H. Pien. Classification-aware dimensionality reduction methods for explosives detection using multi-energy X-ray computed tomography. In *SPIE Electronic Imaging*, pages 78730Q–78730Q. International Society for Optics and Photonics, 2011.

[EK93] R. F. Eilbert and K. D. Krug. Aspects of image recognition in vivid technologies' dual-energy X-ray system for explosives detection. In *Applications in Optical Science and Engineering*, pages 127–143. International Society for Optics and Photonics, 1993.

[EK01] T. Elomaa and M. Kaariainen. An analysis of reduced error pruning. *Journal of Artificial Intelligence Research*, 15:163–187, 2001.

[EKB10] S. Ens, B. Kratz, and T. M. Buzug. Automatische beurteilung von artefakten in tomographischen bilddaten. In *Biomed Tech*, volume 55, page BMT.2010.550, Rostock, Germany, 2010.

[EVGW+10] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[FBM10] G. Flitton, T. Breckon, and N. Megherbi. Object recognition using 3D SIFT in complex CT volumes. In *Proceedings British Machine Vision Conference*, pages 11.1–11.12, 2010.

[FBM12] G. Flitton, T. Breckon, and N. Megherbi. A 3D Extension to Cortex Like Mechanisms for 3D Object Class Recognition. In *IEEE Proceedings International Conference on Computer Vision and Pattern Recognition*, pages 3634–3641, June 2012.

[FBM13] G. Flitton, T. P. Breckon, and N. Megherbi. A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery. *Pattern Recognition*, 2013.

[Fen78] A. Fenster. Split xenon detector for tomochemistry in computed tomography. *Journal of Computer Assisted Tomography*, 2(3):243–252, 1978.

[FFCL97] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Transactions on Medical Imaging*, 16(2):166–175, 1997.

[Fli12] G. Flitton. Extending computer vision techniques to recognition problems in 3D volumetric baggage imagery [PhD Thesis], Cranfield University UK, 2012.

[FM04] G. M. Foody and A. Mathur. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6):1335–1343, 2004.

[FMMB13] G. Flitton, A. Mouton, N. Megherbi, and T. P. Breckon. A codebook approach to object detection in 3D computed tomography baggage imagery (submitted to pattern recognition - under review), 2013.

[FNVV98] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Interventation*, pages 130–137. Springer, 1998.

[FO09] T. Furuya and R. Ohbuchi. Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 26, 2009.

[FR98] C. Fraley and A. E. Raftery. How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal*, 41(8):578–588, 1998.

[Fre01] Y. Freund. An adaptive version of the boost by majority algorithm. *Machine Learning*, 43(3):293–318, 2001.

[FS95] Y. Freund and R. E. Schapire. A desicion-theoretic generalization of online learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[FS99] Y. Freund and R. E. Schapire. Large margin classification using the perceptron algorithm. *Machine Learning*, 37(3):277–296, 1999.

[GA09] L. Grady and C. V. Alvino. The piecewise smooth Mumford-Shah functional on an arbitrary graph. *IEEE Transactions on Image Processing*, 18(11):2547–2561, 2009.

[GBH70] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical biology*, 29(3):471, 1970.

[GC11] J. C. Goswami and A. K. Chan. *Fundamentals of wavelets: theory, algorithms, and applications*, volume 233. Wiley. com, 2011.

[GCM+11] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache. Spatial decision forests for ms lesion segmentation in multichannel magnetic resonance images. *NeuroImage*, 57(2):378–390, 2011.

[geH] GE Healthcare discovery CT750 HD http://www.gehealthcare.com/euen/ct/products/ [Jan. 01, 2012].

[GEW06] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006.

[GHM+10] R. T. Gupta, L. M. Ho, D. Marin, D. T. Boll, H. X. Barnhart, and R. C. Nelson. Dual-energy CT for characterization of adrenal nodules: initial experience. *American Journal of Roentgenology*, 194(6):1479–1483, 2010.

[Gib85] A. Gibbons. *Algorithmic Graph Theory*. Cambridge University Press, 1985.

[GJC01] G. Gerig, M. Jomier, and M. Chakos. Valmet: A new validation tool for assessing and improving 3D object segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 516–523. Springer, 2001.

[GJCM09] A. Graser, T. R. Johnson, H. Chandarana, and M. Macari. Dual energy CT: preliminary observations and potential clinical applications in the abdomen. *European Radiology*, 19(1):13–23, 2009.

[GJH+09] A. Graser, T. R. Johnson, E. M. Hecht, C. R. Becker, C. Leidecker, M. Staehler, C. G. Stief, H. Hildebrandt, M. C. Godoy, M. E. Finn, et al. Dual-energy CT in patients suspected of having renal masses: Can virtual nonenhanced images replace true nonenhanced images? 1. *Radiology*, 252(2):433–440, 2009.

[GKKJ92] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, 1992.

[GL13] J. Gall and V. Lempitsky. Class-specific Hough forests for object detection. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 143–157. Springer, 2013.

[Glo82] G. Glover. Compton scatter effects in CT reconstructions. *Medical Physics*, 9:860, 1982.

[GMB+11] C. Golden, S. R. Mazin, F. E. Boas, G. Tye, P. Ghanouni, G. Gold, M. Sofilos, and N. J. Pelc. A comparison of four algorithms for metal artifact reduction in CT imaging. In *Proceedings SPIE*, volume 7961, page 79612Y, 2011.

[GMC+10] E. Geremia, B. Menze, O. Clatz, E. Konukoglu, A. Criminisi, and N. Ayache. Spatial decision forests for ms lesion segmentation in multi-channel mr images. *Medical Image Computing and Computer-Assisted Intervention*, pages 111–118, 2010.

[GMKP11] Y. Gao, J. F. Mas, N. Kerle, and J. A. N. Pacheco. Optimal region growing segmentation and its effect on classification accuracy. *International Journal of Remote Sensing*, 32(13):3747–3763, 2011.

[GOM+12] S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, and F. Meriaudeau. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. *Computer Methods and Programs in Biomedicine*, 2012.

[Gon72] R. Gonnor. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36:105–117, 1972.

[GP81] G. H. Glover and N. J. Pelc. An algorithm for the reduction of metal clip artifacts in CT reconstructions. *Medical Physics*, 8:799, 1981.

[GPKC12] B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi. Joint classification-regression forests for spatially structured multi-object segmentation. In *European Conference on Computer Vision*, pages 870–881. Springer, 2012.

[Gra06] L. Grady. Fast, quality, segmentation of large volumes - isoperimetric distance trees. In *Proceedings European Conference on Computer Vision*, volume 3, pages 449–462, 2006.

[Gre96] K. D. Gremban. Combinatorial preconditioners for sparse, symmetric, diagonally dominant linear systems [PhD Thesis] Carnegie Mellon University, 1996.

[GS06a] L. Grady and E. L. Schwartz. Isoperimetric graph partitioning for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):469–475, 2006.

[GS06b] L. Grady and E. L. Schwartz. Isoperimetric partitioning: A new algorithm for graph partitioning. *SIAM Journal on Scientific Computing*, 27(6):1844–1866, 2006.

[GSK+12] L. Grady, V. Singh, T. Kohlberger, C. Alvino, and C. Bahlmann. *Automatic segmentation of unknown objects, with application to baggage security*, pages 430–444. European Conference on Computer Vision. Springer, 2012.

[GSZ03] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Texture preserving variational denoising using an adaptive fidelity term. In *Proceedings Variational and Level Set Methods*, volume 5, pages 137–144, 2003.

[Gun10] B. K. Gunturk. Bilateral filter: Theory and applications. *Computational Photography: Methods and Applications*, 2:339, 2010.

[GVR+10] O. Gambino, S. Vitabile, G. L. Re, G. L. Tona, S. Librizzi, R. Pirrone, E. Ardizzone, and M. Midiri. Automatic volumetric liver segmentation using texture based region growing. In *Proceedings of the IEEE International Conference on Complex, Intelligent and Software Intensive Systems*, pages 146–152, 2010.

[GWI+97] B. M. Gordon, H. Weedon, I. Izrailit, T. R. Fox, and J. F. Moore. Dual energy power supply. *US Patent 5,661,774*, filed: August 26 1997.

[Han81] M. Hanson. *Noise and contrast discrimination in computed tomography*, volume 5. C.V. Mosby, 1981.

[HC94] R. V. Hogg and A. Craig. *Introduction to mathematical statistics*. Prentice Hall, 5th edition, 1994.

[Her09] G. T. Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer, 2009.

[HES+07] J. Huang, S. Ertekin, Y. Song, H. Zha, and C. L. Giles. Efficient multiclass boosting classification with active learning. In *SIAM International Conference on Data Mining*, pages 297–308, 2007.

[HHR01] 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*, 20(6):490–498, 2001.

[HJWL13] J. Huang, F. Jian, H. Wu, and H. Li. An improved level set method for vertebra CT image segmentation. *Biomedical engineering online*, 12(1):48, 2013.

[HL80] G. T. Herman and H. P. Lung. Reconstruction from divergent beams: a comparison of algorithms with and without rebinning. *Computers in biology and medicine*, 10(3):131–139, 1980.

[HL94] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4):601–609, 1994.

[HLC12] S. Hu, Z. Liao, and W. Chen. Sinogram restoration for low-dosed x-ray computed tomography using fractional-order Perona-Malik diffusion. *Mathematical Problems in Engineering*, 2012.

[HLN76] G. Herman, A. Lakshminarayanan, and A. Naparstek. Convolution reconstruction techniques for divergent beams. *Computers in biology and medicine*, 6(4):259–271, 1976.

[HM09] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical image analysis*, 13(4):543–563, 2009.

[Hou72] G. N. Hounsfield. Apparatus for examining a body by radiation such as X or gamma radiation. *Patent Specification 1283915, London, England*, 1972.

[Hou73] G. N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *British Journal of Radiology*, 46(552):1016–1022, 1973.

[HR03] W. R. Hendee and E. R. Ritenour. Production of X-rays. *Medical Imaging Physics, Fourth Edition*, pages 69–90, 2003.

[Hsi03] J. Hsieh. *Computed tomography: principles, design, artifacts, and recent advances*. SPIE and John Wiley and Sons, 2003.

[HSKR96] W. E. Higgins, W. J. T. Spyra, R. Karwoski, and E. Ritman. System for analyzing high-resolution three-dimensional coronary angiograms. *IEEE Transactions on Medical Imaging*, 15(3):377–385, 1996.

[HTF01] T. Hastie, R. Tibshirani, and J. J. H. Friedman. *The elements of statistical learning*, volume 1. Springer New York, 2001.

[HvGS+09] T. Heimann, B. van Ginneken, M. A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009.

[Ise09] A. Iserles. *A first course in the numerical analysis of differential equations*, volume 44. Cambridge University Press, 2009.

[ISNC05] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide Second Edition*. Kitware Inc., 2005.

[Jin11] Y. Jin. Implementation and optimization of dual energy computed tomography [PhD Thesis], University of Erlangen-Nuremberg, 2011.

[JKK05] N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate discrete distributions*, volume 444. John Wiley & Sons, 2005.

[JKS95] R. Jain, R. Kasturi, and B. G. Schunck. *Machine Vision*, volume 5. McGraw-Hill New York, 1995.

[JKS+07] T. R. C. Johnson, B. Krauss, M. Sedlmair, M. Grasruck, H. Bruder, D. Morhard, C. Fink, S. Weckbach, M. Lenhard, and B. Schmidt. Material differentiation by dual energy CT: initial experience. *European Radiology*, 17(6):1510–1517, 2007.

[JMF11] R. A. Johnson, I. Miller, and J. E. Freund. *Probability and statistics for engineers*. Prentice-Hall, 2011.

[Joh11] T. R. Johnson. *Medical Radiology/Diagnostic Imaging: Dual Energy CT In Clinical Practice*. Springer, 2011.

[Jor02] A. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. *Advances in neural information processing systems*, 14:841–848, 2002.

[JR09] K. Y. Jeong and J. B. Ra. Reduction of artifacts due to multiple metallic objects in computed tomography. In *Proceedings SPIE*, volume 7258, page 72583E, 2009.

[JSX11] Z.-X. Ji, Q.-S. Sun, and D.-S. Xia. A modified possibilistic fuzzy¡ i¿ c¡/i¿-means clustering algorithm for bias field estimation and segmentation of brain mr image. *Computerized Medical Imaging and Graphics*, 35(5):383–397, 2011.

[JT05] F. Jurie and B. Triggs. Creating efficient codebooks for visual recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 1, pages 604–610, 2005.

[JWP+12] C.-C. Jia, S.-J. Wang, X.-J. Peng, W. Pang, C.-Y. Zhang, C.-G. Zhou, and Z.-Z. Yu. Incremental multi-linear discriminant analysis using canonical correlations for action recognition. *Neurocomputing*, 83:56–63, 2012.

[Kac37] S. Kaczmarz. Angenherte auflsung von systemen linearer gleichungen. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres*, 35:355–357, 1937.

[Kai80] T. Kailath. *Linear systems*, volume 1. Prentice-Hall Englewood Cliffs, NJ, 1980.

[Kal95] W. A. Kalender. Thin-section three-dimensional spiral CT: is isotropic imaging possible? *Radiology*, 197(3):578, 1995.

[KB01] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[KBK03] M. Kachelriess, T. Berkus, and W. Kalender. Quality of statistical reconstruction in medical CT. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, volume 4, pages 2748–2752, 2003.

[KCM13] S. Karimi, P. Cosman, and H. Martz. Metal artifact reduction for CT-based luggage screening, October 2013. Lawrence Livermore National Laboratory Technical Report, LLNL-TR-645205.

[KCWM12] S. Karimi, P. Cosman, C. Wald, and H. Martz. Segmentation of artifacts and anatomy in CT metal artifact reduction. *Medical Physics*, 39(10):5857–5868, 2012.

[KE63] D. E. Kuhl and R. Q. Edwards. Image separation radioisotope scanning. *Radiology*, 80(4):653–662, 1963.

[KEMB11] B. Kratz, S. Ens, J. Mller, and T. M. Buzug. Reference-free ground truth metric for metal artifact evaluation in CT images. *Medical Physics*, 38:4321–4328, 2011.

[KFR03] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proceedings Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 156–164, 2003.

[KHE87] W. A. Kalender, R. Hebel, and J. Ebersberger. Reduction of CT artifacts caused by metallic implants. *Radiology*, 164(2):576, 1987.

[KKK+08] P. Korfiatis, C. Kalogeropoulou, A. Karahaliou, A. Kazantzi, S. Skiadopoulos, and L. Costaridou. Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. *Medical Physics*, 35:5290, 2008.

[KKM+08] B. Kratz, T. Knopp, J. Muller, M. Oehler, and T. M. Buzug. Non-equispaced Fourier Transform vs. polynomial-based metal artifact reduction in computed tomography. *Bildverarbeitung fur die Medizin*, pages 21–25, 2008.

[KKRH+00] A. F. Kopp, K. Klingenbeck-Regn, M. Heuschmid, A. Kuttner, B. Ohnesorge, T. Flohr, S. Schaller, and C. D. Claussen. Multislice computed tomography: basic principles and clinical applications. *Electromedica-Erlangen*, 68(2):94–105, 2000.

[KKSF90a] E. Klotz, W. A. Kalender, R. Sokiransky, and D. Felsenberg. Algorithms for the reduction of CT artifacts caused by metallic implants. In *Proceedings of SPIE*, volume 1234, page 642, 1990.

[KKSF90b] E. Klotz, W. A. Kalender, R. Sokiransky, and D. Felsenberg. Algorithms for the reduction of CT artifacts caused by metallic implants. In *Proceedings SPIE*, volume 1234, page 642, 1990.

[KM88] A. B. Koehler and E. S. Murphree. A comparison of the akaike and schwarz criteria for selecting model order. *Applied Statistics*, pages 187–195, 1988.

[KMA97] K. Krissian, G. Malandain, and N. Ayache. *Directional anisotropic diffusion applied to segmentation of vessels in 3D images*, volume 1252. Springer, 1997.

[KP03] C. Kotropoulos and I. Pitas. Segmentation of ultrasonic images using support vector machines. *Pattern Recognition Letters*, 24(4):715–727, 2003.

[KPVK86] W. Kalender, W. Perman, J. Vetter, and E. Klotz. Evaluation of a prototype dual-energy computed tomographic apparatus. i. phantom studies. *Medical Physics*, 13:334–339, 1986.

[Kru56] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48–50, 1956.

[KS88] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. IEEE Press , New York, 1988.

[KS04] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–506, 2004.

[KSA+12] T. Kohlberger, V. Singh, C. Alvino, C. Bahlmann, and L. Grady. Evaluating segmentation error without ground truth. In *Medical Image Computing and Computer-Assisted Intervention*, pages 528–536. Springer, 2012.

[KSCER11] D. Kihara, L. Sael, R. Chikhi, and J. Esquivel-Rodriguez. Molecular surface representation using 3D Zernike descriptors for protein shape comparison and docking. *Current Protein and Peptide Science*, 12(6):520–530, 2011.

[KSS+11] A. Kreshuk, C. N. Straehle, C. Sommer, U. Koethe, M. Cantoni, G. Knott, and F. A. Hamprecht. Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images. *PloS one*, 6(10):e24899, 2011.

[Kun06] S. Kunis. *Nonequispaced FFT: generalisation and inversion*. Shaker, 2006.

[KWK01] M. Kachelrie, O. Watzke, and W. A. Kalender. Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT. *Medical Physics*, 28:475, 2001.

[KWT88] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[LAL05] Q. Li, S. Ahn, and R. Leahy. Fast hybrid algorithms for PET image reconstruction. In *Nuclear Science Symposium Conference Record, IEEE*, volume 4, pages 1851–1855, 2005.

[LAM+81] L. Lehmann, R. Alvarez, A. Macovski, W. Brody, N. Pelc, S. Riederer, and A. Hall. Generalized image combinations in dual kVp digital radiography. *Medical Physics*, 8:859–867, 1981.

[LBY+10] Y. Li, X. Bao, X. Yin, Y. Chen, L. Luo, and W. Chen. Metal artifact reduction in CT based on adaptive steering filter and nonlocal sinogram inpainting. In *Proceedings of the IEEE International Conference on Biomedical Engineering and Informatics*, volume 1, pages 380–383, 2010.

[LC84] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–316, 1984.

[LCNC98] B. Lichtenbelt, R. Crane, S. Naqvi, and H.-P. Company. *Introduction to volume rendering*. Prentice Hall PTR Upper Saddle River, NJ, 1998.

[LF95] K. Lange and J. A. Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. *IEEE Transactions on Image Processing*, 4(10):1430–1438, 1995.

[LFN09] C. Lemmens, D. Faul, and J. Nuyts. Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion. *IEEE Transactions on Medical Imaging*, 28(2):250–260, 2009.

[LG+12] O. Lézoray, L. Grady, et al. Graph theory concepts and definitions used in image processing and analysis. *Image Processing and Analysing With Graphs: Theory and Practice*, pages 1–24, 2012.

[LGF00] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, 2000.

[LHT10] Y. Lan, R. Harvey, and J. R. P. Torres. Finding stable salient contours. *Image and Vision Computing*, 28(8):1244–1254, 2010.

[Lin94] T. Lindeberg. *Scale-space theory in computer vision*. Kluwer Academic Publishers Netherlands, 1994.

[LLHL02] H. Lu, X. Li, T. Hsiao, and Z. Liang. Analytical noise treatment for low-dose CT projection data by penalized weighted least-square smoothing in the K-L domain. In *Proceedings of SPIE Medical Imaging*, pages 146–152, 2002.

[LLS08] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1-3):259–289, 2008.

[LMT+07] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. Lu. Person-specific sift features for face recognition. In *Proceedings of the IEEE International Conference on acoustics, speech and signal processing*, volume 2, pages 593–596, 2007.

[Low99] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157 vol. 2, 1999.

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[LSP03] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using affine-invariant regions. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–319 – II–324, 2003.

[LSP05] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278, 2005.

[Lu99] Q. Lu. The utility of X-ray dual-energy transmission and scatter technologies for illicit material detection [PhD Thesis] Virginia Polytechnic Institute and State University, 1999.

[LVNB09] V. Lempitsky, M. Verhoek, J. A. Noble, and A. Blake. Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In *Functional Imaging and Modeling of the Heart*, pages 447–456. Springer, 2009.

[LW10] G. Li and Y. Wan. Adaptive seeded region growing for image segmentation based on edge detection, texture extraction and cloud model. *Information Computing and Applications*, pages 285–292, 2010.

[LWK+10] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3):2352–2365, 2010.

[LY03] A. W.-C. Liew and H. Yan. An adaptive spatial fuzzy clustering algorithm for 3-d mr image segmentation. *IEEE Transactions on Medical Imaging*, 22(9):1063–1075, 2003.

[LYPM09] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough. Quantitative imaging of element composition and mass fraction using dual-energy CT: Three-material decomposition. *Medical Physics*, 36:1602, 2009.

[Mah02] M. Mahesh. Search for isotropic resolution in CT from conventional through multiple-row detector. *Radiographics*, 22(4):949–962, 2002.

[Man01] B. D. Man. Iterative reconstruction for reduction of metal artifacts in computed tomography [PhD Thesis] Katholieke Universiteit Leuven, 2001.

[Mar63] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[MB04] B. D. Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Physics in Medicine and Biology*, 49(11):2463, 2004.

[MB08] J. Muller and T. M. Buzug. Intersection line length normalization in CT projection data. *Bildverarbeitung fur die Medizin 2008*, pages 77–81, 2008.

[MB09] J. Muller and T. Buzug. Spurious structures created by interpolation-based CT metal artifact reduction. In *Proceedings of SPIE*, volume 7258, page 72581Y, 2009.

[MB14a] A. Mouton and T. P. Breckon. Automated Analysis of 3D Computed Tomography Imagery for Baggage Security Screening - A Review (Submitted to Machine Vision and Applications - under review), 2014.

[MB14b] A. Mouton and T. P. Breckon. Dual-Energy Based Segmentation of Low Resolution, Cluttered Volumetric Baggage Computed Tomography Imagery (Submitted to Pattern Recognition - under review), 2014.

[MB14c] A. Mouton and T. P. Breckon. On the Relevance of Denoising and Artefact Reduction in 3D Segmentation and Classification within Complex CT Imagery (Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence - under review), 2014.

[MBC+11] L. Monno, R. Bellotti, P. Calvini, R. Monge, G. B. Frisoni, and M. Pievani. Hippocampal segmentation by random forest classification. In *IEEE International Workshop on Medical Measurements and Applications*, pages 536–539, 2011.

[MBF14] A. Mouton, T. Breckon, and G. Flitton. 3D Object Classification in Complex Volumes using Randomised Clustering Forests (Submitted to IEEE International Conference on Computer Vision and Pattern Recognition, under review), 2014.

[MBFM13] N. Megherbi, T. Breckon, G. Flitton, and A. Mouton. Investigating existing medical CT segmentation techniques within automated baggage and package inspection. In *Proceedings SPIE Security and Defence: Optics and Photonics for Counterterrorism, Crime Fighting and Defence*, volume 8901, pages 1–8, 2013.

[MC11] R. Muralidharan and C. Chandrasekar. Object recognition using SVM-KNN based on geometric moment invariant. *International Journal of Computer Trends and Technology*, 1(1):215–220, 2011.

[MCUP04] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[MFB10] N. Megherbi, G. T. Flitton, and T. P. Breckon. A classifier based approach for the detection of potential threats in CT based baggage screening. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1833–1836, 2010.

[MGCA12] J. Margeta, E. Geremia, A. Criminisi, and N. Ayache. Layered spatio-temporal forests for left ventricle segmentation from 4d cardiac MRI data. In *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, pages 109–119. Springer, 2012.

[MGE11] T. Malisiewicz, A. Gupta, and A. A. Efros. Ensemble of exemplar-SVMs for object detection and beyond. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 89–96, 2011.

[MGOS05] N. Menvielle, Y. Goussard, D. Orban, and G. Soulez. Reduction of beam-hardening artifacts in X]-ray CT. In *Proceedings of the IEEE International Conference of the Engineering in Medicine and Biology Society.*, pages 1865–1868, 2005.

[MHBF12] N. Megherbi, J. Han, T. P. Breckon, and G. T. Flitton. A comparison of classification approaches for threat detection in CT based baggage screening. In *Proceedings of the IEEE International Conference on Image Processing*, pages 3109–3112, 2012.

[MLC86] O. Morris, M. d. J. Lee, and A. Constantinides. Graph theory for image analysis: an approach based on the shortest spanning tree. *Communications, Radar and Signal Processing, IEEE Proceedings on*, 133(2):146–152, 1986.

[MLS05] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *Tenth IEEE International Conference on Computer Vision, 2005.*, volume 2, pages 1792–1799, 2005.

[MMB+13] A. Mouton, N. Megherbi, T. Breckon, K. Van Slambrouck, and J. Nuyts. A distance driven method for metal artefact reduction in computed tomography. In *Proceedings IEEE International Conference on Image Processing*, pages 2334–2338, 2013.

[MMBB+05] J. Mourão-Miranda, A. L. Bokde, C. Born, H. Hampel, and M. Stetter. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *Neuroimage*, 28(4):980–995, 2005.

[MMFB12] A. Mouton, N. Megherbi, G. Flitton, and T. Breckon. A novel intensity limiting approach to metal artefact reduction in 3D CT baggage imagery. In *Proceedings of the IEEE International Conference on Image Processing*, pages 2057–2060, 2012.

[MMFB13] A. Mouton, N. Megherbi, G. Flitton, and T. Breckon. An evaluation of CT image denoising techniques applied to baggage imagery screening. In *Proceedings IEEE International Conference on Industrial Technology*, pages 1063–1068, 2013.

[MMK+09] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51(2):73–83, 2009.

[MMvS+13] A. Mouton, N. Megherbi, K. van Slambrouk, J. Nuyts, and T. Breckon. An experimental survey of metal artefact reduction in computed tomography. *Journal of X-Ray Science and Technology*, 21(2):193–226, 2013.

[MND+99] B. D. Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. Metal streak artifacts in X-ray computed tomography: a simulation study. In *IEEE Transactions on Nuclear Science*, volume 46, pages 691–696, 1999.

[MND+00] B. D. Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. Reduction of metal streak artifacts in X-ray computed tomography using a transmission maximum a posteriori algorithm. *IEEE Transactions on Nuclear Science*, 47(3):977–981, 2000.

[MND+01] B. D. Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Transactions on Medical Imaging*, 20(10):999–1008, 2001.

[MNJ08] F. Moosmann, E. Nowak, and F. Jurie. Randomized clustering forests for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1632–1646, 2008.

[Moo96] T. K. Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6):47–60, 1996.

[MP04] G. McLachlan and D. Peel. *Finite mixture models*. Wiley. com, 2004.

[MRL+10] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess. Normalized metal artifact reduction (NMAR) in computed tomography. *Medical Physics*, 37:5482, 2010.

[MRL+12] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess. Frequency split metal artifact reduction (FSMAR) in computed tomography. *Medical Physics*, 39(4):1904–1916, Apr 2012.

[MRS+11] E. Meyer, R. Raupach, B. Schmidt, A. H. Mahnken, and M. Kachelriess. Adaptive normalized metal artifact reduction (ANMAR) in computed tomography. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 2560–2565, 2011.

[MS89] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.

[MS04] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[MS05a] M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *Signal Processing Letters, IEEE*, 12(12):839–842, 2005.

[MS05b] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[MSK+10] M. Macari, B. Spieler, D. Kim, A. Graser, A. J. Megibow, J. Babb, and H. Chandarana. Dual-source dual-energy MDCT of pancreatic adenocarcinoma: initial observations with data generated at 80 kVp and at simulated weighted-average 120 kVp. *American Journal of Roentgenology*, 194(1):W27–W32, 2010.

[MSVGJ03] F. Mees, R. Swennen, M. Van Geet, and P. Jacobs. Applications of X-ray computed tomography in the geosciences. *Geological Society, London, Special Publications*, 215(1):1–6, 2003.

[MSW+11a] A. Montillo, J. Shotton, J. Winn, J. Iglesias, D. Metaxas, and A. Criminisi. Entangled decision forests and their application for semantic segmentation of CT images. In *Information Processing in Medical Imaging*, pages 184–196. Springer, 2011.

[MSW+11b] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, and A. Criminisi. Entangled decision forests and their application for semantic segmentation of CT images. In *Information Processing in Medical Imaging*, pages 184–196. Springer, 2011.

[MT96] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108, 1996.

[MTJ07] F. Moosmann, B. Triggs, and F. Jurie. Fast discriminative visual codebooks using randomized clustering forests. *Advances in Neural Information Processing Systems 19*, pages 985–992, 2007.

[MTJM10] Z. Ma, J. M. R. Tavares, R. N. Jorge, and T. Mascarenhas. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Computer Methods in Biomechanics and Biomedical Engineering*, 13(2):235–246, 2010.

[MTS+05] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[MV98] J. Maintz and M. A. Viergever. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998.

[NBC03] R. Naidu, I. Bechwati, and C. Crawford. Decomposition of multi-energy scan projections using multi-step fitting. *US Patent 7,197,172 B1*, Filed : July 1, 2003.

[NBVV00] W. J. Niessen, C. J. Bouma, K. L. Vincken, and M. A. Viergever. Error metrics for quantitative evaluation of medical image segmentation. In *Characterization in Computer Vision*, pages 275–284. Springer, 2000.

[NH85] R. Nowotny and A. Hfer. Program for calculating diagnostic X-ray spectra. *RoFo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*, 142(6):685, 1985.

[NHS01] T. D. Nguyen, T. B. Ho, and H. Shimodaira. A scalable algorithm for rule post-pruning of large decision trees. In *Advances in Knowledge Discovery and Data Mining*, pages 467–476. Springer, 2001.

[NJT06a] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Proceedings European Conference on Computer Vision*, pages 490–503, 2006.

[NJT06b] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *European Conference on Computer Vision*, pages 490–503, 2006.

[NJW+02] A. Y. Ng, M. I. Jordan, Y. Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[NK04] M. Novotni and R. Klein. Shape retrieval using 3D zernike descriptors. *Computer-Aided Design*, 36(11):1047–1062, 2004.

[NKY12] J. J. Näppi, S. H. Kim, and H. Yoshida. Automated detection of colorectal lesions with dual-energy CT colonography. In *SPIE Medical Imaging*, pages 83150Y–83150Y. International Society for Optics and Photonics, 2012.

[NLA+11] V. Naranjo, R. Llorens, M. Alcaniz, R. Verdu-Monedero, J. Larrey-Ruiz, and J. Morales-Sanchez. A new 3D paradigm for metal artifact reduction in dental CT. In *IEEE Conference on Image Processing*, pages 461–464, 2011.

[NLP+09] V. Naranjo, R. Llorens, P. Paniagua, M. Alcaniz, and S. Albalat. A new approach in metal artifact reduction for CT 3D reconstruction. *Bioinspired Applications in Artificial and Natural Computation*, pages 11–19, 2009.

[NMD+98] J. Nuyts, B. D. Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans. Iterative reconstruction for helical CT: a simulation study. *Physics in Medicine and Biology*, 43:729, 1998.

[NS06] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2161–2168, 2006.

[NWK+11] F. Nachtrab, S. Weis, P. Keßling, F. Sukowski, U. Haßler, T. Fuchs, N. Uhlmann, and R. Hanke. Quantitative material analysis by dual-energy computed tomography for industrial NDT applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 633:S159–S162, 2011.

[OB06] M. Oehler and T. Buzug. Modified MLEM algorithm for artifact suppression in CT. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, volume 6, pages 3511–3518, 2006.

[OB07] M. Oehler and T. M. Buzug. The λ-mlem algorithm: An iterative reconstruction technique for metal artifact reduction in CT images. *Advances in Medical Engineering*, pages 42–47, 2007.

[OEW08] J. Orchard, M. Ebrahimi, and A. Wong. Efficient nonlocal-means denoising using the SVD. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1732–1735, 2008.

[Old61] W. Oldendorf. Isolated flying spot detection of radiodensity discontinuities-displaying the internal structural pattern of a complex object. *IRE Transactions on Biomedical Electronics*, 8(1):68–72, 1961.

[Oll94] J. M. Ollinger. Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm. *IEEE Transactions on Medical Imaging*, 13(1):89–101, 1994.

[OPFA06] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer. Generic object recognition with boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):416–431, 2006.

[Opp77] B. E. Oppenheim. Reconstruction tomography from incomplete projections. *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine*, M.M. Ter-Pogossian et al., Eds. Baltimore, MD: University Park:155–183, 1977.

[OPSS11] J. O. Ogutu, H.-P. Piepho, and T. Schulz-Streeck. A comparison of random forests, boosting and support vector machines for genomic selection. In *Proceedings BioMed Central European Workshop on QTL Mapping and Marker Assisted Selection*, volume 5, page S11, 2011.

[OSE05] M. Omran, A. Salman, and A. Engelbrecht. Dynamic clustering using particle swarm optimization with application in unsupervised image classification. In *Fifth World Enformatika Conference Prague, Czech Republic*, pages 199–204. Citeseer, 2005.

[Par02] N. Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Computer Vision*, 50(3):345–362, 2002.

[PB99] M. Petrou and P. Bosdogianni. *Image Processing: The Fundamentals*. John Wiley and Sons, 1999.

[PBLL11] J. Pang, Z. Bai, J. Lai, and S. Li. Automatic segmentation of crop leaf spot disease images by integrating local threshold and seeded region growing. In *Proceedings of the IEEE International Conference on Image Analysis and Signal Processing*, pages 590–594, 2011.

[PFPB10] B. Park, A. Furlan, A. Patil, and K. T. Bae. Segmentation of blood clot from CT pulmonary angiographic images using a modified seeded region growing algorithm method. In *SPIE Medical Imaging*, pages 76234K–76234K–5. International Society for Optics and Photonics, 2010.

[PKTD07] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. A gentle introduction to bilateral filtering and its applications. In *ACM SIGGRAPH 2007 courses*, 05-09 August 2007.

[PL77] T. Peters and R. Lewitt. Computed tomography with fan beam geometry. *Journal of Computer Assisted Tomography*, 1(4):429–436, 1977.

[PM90] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[Pol94] P. A. Polski. International aviation security research and development. *Journal of Testing and Evaluation*, 22(3):267–274, 1994.

[Pre12] J. Prekeges. *Nuclear Medicine Instrumentation*. Jones and Bartlett Learning, 2012.

[PS01] A. L. Prodromidis and S. J. Stolfo. Cost complexity-based pruning of ensemble classifiers. *Knowledge and Information Systems*, 3(4):449–469, 2001.

[PT01] R. Pohle and K. D. Toennies. Segmentation of medical images using adaptive region growing. In *SPIE Medical Imaging*, volume 4322, pages 1337–1346, 2001.

[PT13] N. Payet and S. Todorovic. Hough forest random field for object recognition and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1066–1079, 2013.

[PV98] M. Pontil and A. Verri. Support vector machines for 3D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):637–646, 1998.

[PXP00] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation 1. *Annual review of biomedical engineering*, 2(1):315–337, 2000.

[Qui86] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[RC01] M. M. O. B. B. Richard and M. K. Y. S. Chang. Fast digital image inpainting. In *Proceedings of the International Conference on Visualization, Imaging and Image Processing*, 2001.

[RDF+08] J. Rinkel, W. P. Dillon, T. Funk, R. Gould, and S. Prevrhal. Computed tomographic metal artifact reduction for the detection and quantification of small features near large metallic implants: a comparison of published methods. *Journal of Computer Assisted Tomography*, 32(4):621, 2008.

[Ris01] I. Rish. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

[RK96] W. D. Richard and C. G. Keen. Automated texture-based segmentation of ultrasound images of the prostate. *Computerized Medical Imaging and Graphics*, 20(3):131–140, 1996.

[RLP+03] J. C. Roeske, C. Lund, C. A. Pelizzari, X. Pan, and A. J. Mundt. Reduction of computed tomography metal artifacts due to the Fletcher-Suit applicator in gynecology patients receiving intracavitary brachytherapy. *Brachytherapy*, 2(4):207–214, 2003.

[RLZ+08] B. Ruzsics, H. Lee, P. L. Zwerner, M. Gebregziabher, P. Costello, and U. J. Schoepf. Dual-energy CT of the heart for diagnosing coronary artery stenosis and myocardial ischemia-initial experience. *European Radiology*, 18(11):2414–2424, 2008.

[Rod79] F. L. Roder. Explosives detection by dual-energy computed tomography (CT). In *Techincal Symposium East*, pages 171–178. International Society for Optics and Photonics, 1979.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[RP66] A. Rosenfeld and J. L. Pfaltz. Sequential operations in digital picture processing. *Journal of the ACM*, 13(4):471–494, 1966.

[RP79] R. Ritchings and B. Pullan. A technique for simultaneous dual energy scanning. *Journal of Computer Assisted Tomography*, 3(6):842–846, 1979.

[S+04] M. Sezgin et al. Survey over image thresholding techniques and quantitative evaluation. *Journal of Electronic imaging*, 13(1):146–168, 2004.

[SB93] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 41(2):534–548, 1993.

[SB10] C. Solomon and T. Breckon. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley-Blackwell, 2010.

[SBB+93] J. Stoer, R. Bulirsch, R. Bartels, W. Gautschi, and C. Witzgall. *Introduction to numerical analysis*, volume 2. Springer New York, 1993.

[SBB12] S. Sun, C. Bauer, and R. Beichel. Automated 3-d segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach. *IEEE Transactions on Medical Imaging*, 31(2):449–460, 2012.

[SBHC+12] G. Schmid-Bindert, T. Henzler, T. Chu, M. Meyer, J. Nance Jr, U. Schoepf, D. Dinter, P. Apfaltrer, R. Krissak, C. Manegold, et al. Functional imaging of lung cancer using dual energy CT: how does iodine related attenuation correlate with standardized uptake value of 18FDG-PET-CT? *European Radiology*, 22(1):93–103, 2012.

[SBSA03] L. Szilagyi, Z. Benyo, S. M. Szilágyi, and H. Adam. Mr brain image segmentation using an enhanced fuzzy c-means algorithm. In *Engineering in Medicine and Biology Society. Proceedings of the 25th Annual International Conference of the IEEE*, volume 1, pages 724–726, 2003.

[SCS+13] G. Saiprasad, C.-I. Chang, N. Safdar, N. Saenz, and E. Siegel. Adrenal gland abnormality detection using random forest classification. *Journal of Digital Imaging*, pages 1–7, 2013.

[See01] E. Seeram. *Computed tomography: physical principles, clinical applications, and quality control*. WB Saunders Philadelphia, 2001.

[Set99] J. A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge university press, 1999.

[SEZ05] J. Sivic, M. Everingham, and A. Zisserman. Person spotting: video shot retrieval for face sets. In *Image and Video Retrieval*, pages 226–236. Springer, 2005.

[SFTA08] A. Suga, K. Fukuda, T. Takiguchi, and Y. Ariki. Object recognition and segmentation using SIFT and graph cuts. In *Proceedings of the IEEE International Conference on Pattern Recognition*, pages 1–4, 2008.

[SHS04] H. Shikata, E. A. Hoffman, and M. Sonka. Automated segmentation of pulmonary vascular tree from 3D CT images. In *SPIE Medical Imaging*, pages 107–116. International Society for Optics and Photonics, 2004.

[Sin03] S. Singh. Explosives detection systems (EDS) for aviation security. *Signal Processing*, 83(1):31–55, 2003.

[SJC08] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[SL91] S. R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3):660–674, 1991.

[SLT+03] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, 43(6):1947–1958, 2003.

[SM00] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[SM12] O. Semerci and E. Miller. A parametric level-set approach to simultaneous object identification and background reconstruction for dual-energy computed tomography. *IEEE Transactions on Image Processing*, 21(5):2719–2734, 2012.

[SMB00] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of computer vision*, 37(2):151–172, 2000.

[Smi03] S. W. Smith. *Digital signal processing: a practical guide for engineers and scientists*. Newnes, 2003.

[SMK11] J. Smith, H. Martz, and J. Kallman. Case for an improved effective atomic number for the electronic baggage screening program, December 2011. Lawrence Livermore National Laboratory Technical Report, LLNLTR-520312.

[SMPW] A. Sit, J. C. Mitchell, G. N. Phillips, and S. J. Wright. An extension of 3D Zernike moments for shape description and retrieval of maps defined in rectangular solids. *Molecular Based Mathematical Biology*, 1:75–89.

[SN12] K. V. Slambrouck and J. Nuyts. Metal artifact reduction in computed tomography using local models in an image block-iterative scheme. *Medical Physics*, 39(11):7080, 2012.

[Spe01] R. Speller. Radiation-based security. *Radiation Physics and Chemistry*, 61(3):293–300, 2001.

[Spi46] F. Spiers. Effective atomic number and energy absorption in tissues. *British Journal of Radiology*, 19(218):52–63, 1946.

[SPvG05] I. Sluimer, M. Prokop, and B. van Ginneken. Toward automated segmentation of the pathological lung in CT. *IEEE Transactions on Medical Imaging*, 24(8):1025–1038, 2005.

[SSAW83] E. Saion, Z. A. Sulaiman, A. Ahmad, and H. Wagiran. Determination of effective atomic number of rubber. *Pertanika*, 6(3):95–98, 1983.

[SSB+97] B. Scholkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing*, 45(11):2758–2765, 1997.

[SSO92] D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan. Deblurring subject to nonnegativity constraints. *IEEE Transactions on Signal Processing*, 40(5):1143–1150, 1992.

[SSW88] P. K. Sahoo, S. Soltani, and A. Wong. A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 41(2):233–260, 1988.

[SV82] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.

[SWA08] A. Statnikov, L. Wang, and C. F. Aliferis. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics*, 9(1):319, 2008.

[SWB+11] M. Sofka, J. Wetzl, N. Birkbeck, J. Zhang, T. Kohlberger, J. Kaftan, J. Declerck, and S. K. Zhou. Multi-stage learning for robust lung segmentation in challenging CT volumes. In *Medical Image Computing and Computer-Assisted Intervention*, pages 667–674. Springer, 2011.

[SWP05] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by the visual cortex. In *IEEE computer society conference on computer vision and pattern recognition*, volume 2, pages 994–1000, 2005.

[SZ03] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1470–1477, 2003.

[SZD06] X. Sun, H. Zhang, and H. Duan. 3d computerized segmentation of lung volume with computed tomography. *Academic Radiology*, 13(6):670–677, 2006.

[Sze06] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1):1–104, 2006.

[Sze10] R. Szeliski. *Computer vision: algorithms and applications*. Springer-Verlag New York Inc, 2010.

[Tab07] M. Tabakov. A fuzzy segmentation method for computed tomography images. *International Journal of Intelligent Information Database Systems*, 1(1):79–89, 2007.

[TB94] D. C. Taylor and W. A. Barrett. Image segmentation using globally optimal growth in three dimensions with an adaptive feature set. In *SPIE*, volume 2359, pages 98–107, 1994.

[TBH+08] S. F. Thieme, C. R. Becker, M. Hacker, K. Nikolaou, M. F. Reiser, and T. R. Johnson. Dual energy CT for the assessment of lung perfusion-correlation to scintigraphy. *European journal of Radiology*, 68(3):369–374, 2008.

[TKK+10] C. Thomas, A. Korn, B. Krauss, D. Ketelsen, I. Tsiflikas, A. Reimann, H. Brodoefel, C. Claussen, A. Kopp, U. Ernemann, et al. Automatic bone and plaque removal using dual energy CT for head and neck angiography: feasibility and initial performance evaluation. *European journal of Radiology*, 76(1):61–67, 2010.

[TM98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 839–846, 1998.

[TM08] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008.

[TMK+06] Y. Takahashi, S. Mori, T. Kozuka, K. Gomi, T. Nose, T. Tahara, M. Oguchi, and T. Yamashita. Preliminary study of correction of original metal artifacts due to 1-125 seeds in postimplant dosimetry for prostate permanent implant brachytherapy. *Radiation Medicine*, 24(2):133–138, Feb 2006.

[tos] Toshiba Medical Systems Corporation Aquilion 32 http://www.toshiba-medical.co.uk/ct-systems.asp [Jan. 01, 2012].

[TSBH07] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multislice helical CT. *Medical Physics*, 34:4526, 2007.

[TT94] D.-Y. Tsai and N. Tanahashi. Neural-network-based boundary detection of liver structure in CT images for 3-d visualization. In *IEEE World Congress on Computational Intelligence.*, volume 6, pages 3484–3489, 1994.

[TTMG12] J. C. Tilton, Y. Tarabalka, P. M. Montesano, and E. Gofman. Best merge region growing segmentation with integrated nonadjacent region object aggregation. In *IEEE Transactions on Geoscience and Remote Sensing*, number 11, pages 4454–4467, 2012.

[Tu05] Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 1589–1596, 2005.

[Tuy93] H. K. Tuy. A post-processing algorithm to reduce metallic clip artifacts in CT images. *European Radiology*, 3(2):129–134, 1993.

[TVG99] T. Tuytelaars and L. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Visual Information and Information Systems*, pages 493–500. Springer, 1999.

[UR05] S. Ukil and J. M. Reinhardt. Smoothing lung segmentation surfaces in three-dimensional X-ray CT images using anatomic guidance. *Academic Radiology*, 12(12):1502–1511, 2005.

[Urq82] R. Urquhart. Graph theoretical clustering based on limited neighbourhood sets. *Pattern Recognition*, 15(3):173–187, 1982.

[US96] J. K. Udupa and S. Samarasekera. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. *Graphical models and image processing*, 58(3):246–261, 1996.

[USS09] J. R. Uijlings, A. W. Smeulders, and R. J. Scha. Real-time bag of words, approximately. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 6, 2009.

[USV+09] L. G. Ugarriza, E. Saber, S. R. Vantaram, V. Amuso, M. Shaw, and R. Bhaskar. Automatic image segmentation by dynamic region growth and multiresolution merging. *IEEE Transactions on Image Processing*, 18(10):2275–2288, 2009.

[UWH+09] K. Uotani, Y. Watanabe, M. Higashi, T. Nakazawa, A. K. Kono, Y. Hori, T. Fukuda, S. Kanzaki, N. Yamada, T. Itoh, et al. Dual-energy CT head bone and hard plaque removal for quantification of calcified carotid stenosis: utility and comparison with digital subtraction angiography. *European Radiology*, 19(8):2060–2065, 2009.

[Vap00] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York Inc, 2000.

[VC02] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[VCK+09] V. Venkatraman, P. Chakravarthy, D. Kihara, et al. Application of 3D Zernike descriptors to shape-based ligand similarity searching. *J. Cheminformatics*, 1:19, 2009.

[VGFS+02] B. Van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, 2002.

[VGMU96] L. Van Gool, T. Moons, and D. Ungureanu. Affine/photometric invariants for planar intensity patterns. In *European Conference on Computer Vision*, pages 642–651. Springer, 1996.

[vGVSG10] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek. Visual word ambiguity. *IEEE transactions on pattern analysis and machine intelligence*, 32(7):1271–1283, 2010.

[VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, 2001.

[VJ04] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[vKD05] G. van Kaick and S. Delorme. Computed tomography in various fields outside medicine. *European Radiology Supplements*, 15:74–81, 2005.

[vRAvG07] E. van Rikxoort, Y. Arzhaeva, and B. van Ginneken. Automatic segmentation of the liver in computed tomography scans with voxel classification and atlas matching. In *MICCAI 2007 workshop proceedings: 3D segmentation in the clinic*, pages 101–108. Citeseer, 2007.

[vRdHV+09] E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*, 36:2934, 2009.

[vRvG13] E. M. van Rikxoort and B. van Ginneken. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. *Physics in medicine and biology*, 58(17):R187, 2013.

[VS91] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE transactions on pattern analysis and machine intelligence*, 13(6):583–598, 1991.

[VS12] J. M. Verburg and J. Seco. CT metal artifact reduction method correcting for beam hardening and missing projections. *Physics in Medicine and Biology*, 57(9):2803, 2012.

[W+01] D. B. West et al. *Introduction to graph theory*, volume 2. Prentice hall Englewood Cliffs, 2001.

[Wan77] L. Wang. Cross-section reconstruction with a fan-beam scanning geometry. *IEEE Transactions on Computers*, 100(3):264–268, 1977.

[Wan90] E. A. Wan. Neural network classification: A Bayesian interpretation. *IEEE Transactions on Neural Networks*, 1(4):303–305, 1990.

[WBB05] R. Wiemker, T. Bülow, and T. Blaffert. Unsupervised extraction of the pulmonary interlobar fissures from high resolution thoracic CT data. In *International Congress Series*, volume 1281, pages 1121–1126. Elsevier, 2005.

[WC06] J. Winn and A. Criminisi. Object class recognition at a glance. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2006.

[WdB69] J. Weber and D. V. den Berge. The effective atomic number and the calculation of the composition of phantom materials. *British Journal of Radiology*, 42(497):378–383, 1969.

[Weh96] L. Wehenkel. On uncertainty measures used for decision tree induction. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 1996.

[WFV00] G. Wang, T. Frei, and M. W. Vannier. Fast iterative algorithm for metal artifact reduction in X-ray CT. *Academic Radiology*, 7(8):607–614, 2000.

[WGW12] D. F. Wiley, D. Ghosh, and C. Woodhouse. Automatic segmentation of CT scans of checked baggage. In *Proceedings of the 2nd International Meeting on Image Formation in X-ray CT*, pages 310–313, 2012.

[WGY+06] J. Wang, Y. Guo, Y. Ying, Y. Liu, and Q. Peng. Fast non-local algorithm for image denoising. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1429–1432, 2006.

[WH03] S. Y. Wan and W. E. Higgins. Symmetric region growing. *IEEE Transactions on Image Processing*, 12(9):1007–1015, 2003.

[Wil09] D. F. Wiley. Analysis of anatomic regions delineated from image data. *US Patent 8,194,964 (US App. No. 12/430,545)*, 2009.

[WK04] O. Watzke and W. A. Kalender. A pragmatic approach to metal artifact reduction in CT: merging of metal artifact reduced images. *European Radiology*, 14(5):849–856, 2004.

[WK08] D. Withey and Z. Koles. A review of medical image segmentation: methods and available software. *International Journal of Bioelectromagnetism*, 10(3):125–148, 2008.

[WKRH00] S. Y. Wan, A. P. Kiraly, E. L. Ritman, and W. E. Higgins. Extraction of the hepatic vasculature in rats using 3-D micro-CT images. *IEEE Transactions on Medical Imaging*, 19(9):964–971, 2000.

[WLL09] J. Wang, F. Li, and Q. Li. Automated segmentation of lungs with severe interstitial lung disease in CT. *Medical Physics*, 36:4592, 2009.

[WP08] W. Wieclawek and E. Pietka. Fuzzy clustering in segmentation of abdominal structures based on CT studies. In *Information Technologies in Biomedicine*, pages 93–104. Springer, 2008.

[WRRB09] S. Witoszynskyj, A. Rauscher, J. R. Reichenbach, and M. Barth. Phase unwrapping of mr images using phi un-a fast and robust region growing algorithm. *Medical image analysis*, 13(2):257–268, 2009.

[WSOV96] G. Wang, D. L. Snyder, J. A. O'Sullivan, and M. W. Vannier. Iterative deblurring for CT metal artifact reduction. *IEEE Transactions on Medical Imaging*, 15(5):657–664, 1996.

[WZW04] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.

[XZX+09] H. Xue, L. Zhang, Y. Xiao, Z. Chen, and Y. Xing. Metal artifact reduction in dual energy CT by sinogram segmentation based on active contour model and TV inpainting. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, pages 904–908, 2009.

[YCSB09] Z. Yi, A. Criminisi, J. Shotton, and A. Blake. Discriminative, semantic segmentation of brain tissue in mr images. In *Medical Image Computing and Computer-Assisted Intervention*, pages 558–565. Springer, 2009.

[YHLL02] M.-S. Yang, Y.-J. Hu, K. C.-R. Lin, and C. C.-L. Lin. Segmentation techniques for tissue differentiation in MRI of ophthalmology using fuzzy clustering algorithms. *Magnetic Resonance Imaging*, 20(2):173–179, 2002.

[YHQ04] Y. Yubin, L. Hui, and Z. Qing. Content-based 3D model retrieval: a survey. *Chinese journal of computers*, 27(10):1297–1310, 2004.

[YMFS04] J. Yao, M. Miller, M. Franaszek, and R. M. Summers. Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models. *IEEE Transactions on Medical Imaging*, 23(11):1344–1352, 2004.

[YNC06] Z. Ying, R. Naidu, and C. R. Crawford. Dual energy computed tomography for explosive detection. *Journal of X-ray Science and Technology*, 14(4):235–256, 2006.

[YNI+12] M. Yaqub, R. Napolitano, C. Ioannou, A. Papageorghiou, and J. A. Noble. Automatic detection of local fetal brain structures in ultrasound images. In *IEEE International Symposium on Biomedical Imaging*, pages 1555–1558, 2012.

[YNSC07] Z. Ying, R. Naidu, S. Simanovsky, and C. R. Crawford. Method of and system for computing effective atomic number images in multi-energy computed tomography. *US Patent 7,190,757*, Filed: March 13, 2007.

[YTB+11] Z. Yu, J. B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. Fast model-based X-Ray CT reconstruction using spatially nonhomogeneous ICD optimization. *IEEE Transactions on Image Processing*, 20(1):161–175, 2011.

[YZB+07] H. Yu, K. Zeng, D. K. Bharkhada, G. Wang, M. T. Madsen, O. Saba, B. Policeni, M. A. Howard, and W. R. K. Smoker. A segmentation-based method for metal artifact reduction. *Academic Radiology*, 14(4):495–504, 2007.

[YZY08] Z. Yi, C. Zhiguo, and X. Yang. Multi-spectral remote image registration based on SIFT. *Electronics Letters*, 44(2):107–108, 2008.

[Zah71] C. T. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, 100(1):68–86, 1971.

[ZBWW02] S. Zhao, K. T. Bae, B. Whiting, and G. Wang. A wavelet method for metal artifact reduction with multiple metallic objects in the field of view. *Journal of X-Ray Science and Technology*, 10(2):67–76, 2002.

[ZCK+12] C. Zhou, H.-P. Chan, J. W. Kuriakose, A. Chughtai, J. Wei, L. M. Hadjiiski, Y. Guo, S. Patel, and E. A. Kazerooni. Pulmonary vessel segmentation utilizing curved planar reformation and optimal path finding (crop) in computed tomographic pulmonary angiography (ctpa) for cad applications. In *SPIE Medical Imaging*, pages 83150N–83150N. International Society for Optics and Photonics, 2012.

[ZCS+07] C. Zhou, H.-P. Chan, B. Sahiner, L. M. Hadjiiski, A. Chughtai, S. Patel, J. Wei, J. Ge, P. N. Cascade, and E. A. Kazerooni. Automatic multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications. *Medical Physics*, 34:4567, 2007.

[ZCZC06] G. Zhang, Z. Chen, L. Zhang, and J. Cheng. Exact reconstruction for dual energy computed tomography using an H-L curve method. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, volume 6, pages 3485–3488, 2006.

[ZdFFeR07] L. Zhang, M. J. da Fonseca, A. Ferreira, and C. R. A. e Recuperação. Survey on 3D shape descriptors. Technical report, Technical Report FCT POSC/EIA/59938/2004, 2007.

[ZG08] M. Zhang and B. K. Gunturk. A new image denoising framework based on bilateral filter. In *Proceedings SPIE*, volume 6822, pages 68221B–68221B–8, 2008.

[ZGC13] D. Zikic, B. Glocker, and A. Criminisi. Atlas encoding by randomized forests for efficient label propagation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 66–73. Springer, 2013.

[ZGK+12] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. Thomas, T. Das, R. Jena, and S. Price. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In *Medical Image Computing and Computer-Assisted Intervention*, pages 369–376. Springer, 2012.

[Zha96] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.

[ZPA10a] Y. Zhou, K. Panetta, and S. Agaian. 3D CT baggage image enhancement based on order statistic decomposition. In *Proceedings of the IEEE International Conference on Technologies for Homeland Security*, pages 287–291, 2010.

[ZPA10b] Y. Zhou, K. Panetta, and S. Agaian. CT baggage image enhancement using a combination of alpha-weighted mean separation and histogram equalization. In *Proceedings SPIE*, volume 7708, page 77080G, 2010.

[ZPH+11] Y. Zhang, Y. F. Pu, J. R. Hu, Y. Liu, and J. L. Zhou. A new CT metal artifacts reduction algorithm based on fractional-order sinogram inpainting. *Journal of X-ray Science and Technology*, 19(3):373–384, 2011.

[ZPV05] C. Zhang, J. C. Platt, and P. A. Viola. Multiple instance boosting for object detection. In *Advances in Neural Information Processing Systems*, pages 1417–1424, 2005.

[ZRW+00] S. Zhao, D. D. Robeltson, G. Wang, B. Whiting, and K. T. Bae. X-ray CT metal artifact reduction using wavelets: an application for imaging total hip prostheses. *IEEE Transactions on Medical Imaging*, 19(12):1238–1247, 2000.

[ZS12] Y. Zou and M. D. Silver. Method, apparatus, and computer-readable medium for pre-reconstruction decomposition and calibration in dual energy computed tomography. *US Patent 8,194,961*, Filed: June 5, 2012.

[ZWX11] X. Zhang, J. Wang, and L. Xing. Metal artifact reduction in X-ray computed tomography (CT) by constrained optimization. *Medical Physics*, 38(2):701, 2011.

[ZXLD11] Z. Zhengtao, Y. Xiongyi, H. Liuqian, and W. De. Fast capsule image segmentation based on linear region growing. In *Proceedings of the IEEE International Conference on Computer Science and Automation Engineering*, volume 2, pages 99–103, 2011.

[ZY96] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

Appendix A

Segmentation Algorithms

A description of three segmentation techniques (and related theory) that have been successfully incorporated into baggage-segmentation algorithms is presented.

A.1 The Isoperimetric Distance Tree Algorithm

The isoperimetric algorithm has demonstrated efficient and stable results in image-segmentation problems [GS06a]. Grady [Gra06] demonstrates that the algorithm may be applied in low-constant linear time by operating on a subgraph (termed the distance tree) of the lattice graph representing a connected component. The resulting Isoperimetric Distance Tree (IDT) algorithm for graph partitioning [Gra06] has been successfully incorporated into a framework for the segmentation of 3D CT-baggage imagery and large, medical volumes [GSK+12, Gra06].

A fundamental understanding of the concepts and terminology of graph-based image processing theory is necessary for the understanding of the IDT algorithm. A brief review of these concepts is provided here - for a more comprehensive review of the mathematics of graph theory, the reader is referred to the literature [BM76, Gib85, Big93, W+01].

A.1.1 Basic Graph Theory

Conceptually, a graph may be defined as a set of points (*vertices* or *nodes*) interconnected by a set of lines (*edges*). Formally, a graph G is defined by a non-empty set of vertices V and an edge set E (disjoint from V), such that $G = (V, E)$. The number of vertices, or *order*, of a graph may be denoted by $N_v = |V|$ and the number of edges, or *size*, as $N_e = |E| \leq N_v^2$, where $|.|$ denotes cardinality. If two vertices $v_i \in V$ and $v_j \in V$ are connected by an edge, that edge may be denoted by $e_{ij} \in E$. A *simple* graph (Figure A.1 (a)) is one containing no loops (edges connected at both ends to the same vertex) and only a single edge connecting any two vertices.

A *weighted* graph assigns a weighting to every edge in the graph, representing the strength of affinity between the connected vertices (Figure A.1 (a)) Given a graph $G = (V, E)$, the weight of an edge spanning two vertices v_i and v_j is denoted by $w(v_i, v_j)$ or simply w_{ij}. If $w_{ij} = 1$, $\forall e_{ij} \in E$ then the graph is said to be *unweighted*. An edge-weight of $w_{ij} = 0$ implies that $e_{ij} \notin E$.

If $w_{ij} \neq w_{ji}$ for a given edge e_{ij}, then that edge is said to be *directed*. A *directed* graph, is one in which at least one edge is directed. An *undirected* graph contains no directed edges - that is to say, $w_{ij} = w_{ji}$, $\forall e_{ij} \in E$. Unless specified otherwise, all graphs (and associated terminology) considered hereafter are assumed to be undirected, simple graphs (Figure A.1 (a)).

In the case of an undirected graph, a given edge e_{ij} is said to have two *endpoints* v_i and v_j. Vertices v_i and v_j are said to be *adjacent* if $\exists e_{ij} \in E$, which is denoted by $v_i \sim v_j$. An edge e_{ij} is *incident* to the vertices v_i and v_j. A vertex which is not incident to any edges is said to be *isolated*.

The *degree* of a vertex v_i is denoted by $d(v_i)$ or d_i and is equal to the sum of the weights of its incident edges: $d_i = \sum_{e_{ij} \in E} w_{ij}$. It follows that for an unweighted graph the degree d_i is simply equal to the number of edges that connect to the vertex v_i. An undirected graph is said to be *regular* if each of its vertices has the same degree. A regular graph with vertices of degree k is called a k-regular graph.

The graph $G' = (V', E')$ is called a *subgraph* of a graph $G = (V, E)$ if: 1) $V' \subseteq V$ and 2) every edge of G' is also an edge of G (i.e. $E' = \{e_{ij} \in E | v_i \in V' \text{ and } v_j \in V'\}$). The subgraph G' is a *spanning subgraph* of G if $V' = V$. G' is said to *span* G.

A graph is said to be *bipartite* if the vertex set V may be partitioned into two disjoint sets $V_1 \subset V$ and $V_2 \subset V$ where $V_1 \cap V_2 = \emptyset$ and $V_1 \cup V_2 = V$, such that no two vertices within the same set are adjacent - that is to say, every edge connects a vertex in V_1 to a vertex in V_2. Furthermore, if every vertex in V_1 is connected to every vertex in V_2 the graph is said to be a *complete bipartite* graph. A complete bipartite graph with $|V_1| = a$ and $|V_2| = b$ is denoted by K_{ab}.

A *fully-connected* graph contains an edge spanning every pair of vertices - that is to say: $N_e = N_v$.

The *adjacency matrix* is a symmetric $|V| \times |V|$ matrix given by (Figure A.2):

$$\mathbf{A}_{ij} = \begin{cases} w_{ij} & \text{if } e_{ij} \in E \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

The *degree matrix* is a $|V| \times |V|$ diagonal matrix containing the degree of each vertex (Figure A.2):

$$\mathbf{D}_{ij} = \begin{cases} d_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

The *Laplacian matrix* is a $|V| \times |V|$ matrix containing both degree and adjacency information (Figure A.2):

$$\mathbf{L}_{ij} = \begin{cases} d_i & \text{if } i = j \\ -w_{ij} & \text{if } e_{ij} \in E \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

L may be computed as the difference of the degree matrix and the adjacency matrix: $L = D - A$.

A *walk* in the graph $G = (V, E)$ is a finite sequence of alternating vertices and edges of G, beginning at one vertex and ending at another. Every vertex in the set is incident to both its preceding and its superseding edge in the set. The walk from the *initial* vertex v_{i0} to the *terminal* vertex v_{ik} is denoted by: $\pi(v_{i0}, v_{ik}) = (v_{i0}, e_{j1}, v_{i1}, e_{j2}, \ldots, e_{jk}, v_{ik})$. The length of the walk is given by $|\pi| = k$. If $v_{i0} = v_{ik}$, the walk is said to be *closed*, otherwise it is *open*. A walk is permitted to contain multiple instances of the same vertices and/or edges.

A walk where every edge in the sequence is traversed only once is called a *trail*. A closed trail is called a *circuit* and a circuit where all vertices are distinct (excluding the initial and terminal vertices) is a *cycle*. An open trail where all vertices are distinct is called a *path* (or simply an *open walk*) and a *subpath* is any sequential subset of a path. A graph containing no cycles is said to be *acyclic*.

A *tree* is a connected acyclic simple graph. A *forest* is an acyclic simple graph (i.e. the disjoint union of one or more disconnected trees). A *spanning tree* of a graph G is a tree that is also a spanning subgraph of G (Figure A.1 (b) - (d)). Only fully connected graphs have spanning trees - for a graph that is not connected, a *spanning forest* is the maximal acyclic subgraph of that graph (i.e. the graph consisting of a spanning tree in each connected component of the graph). A *Minimum* (or shortest) Spanning Tree (MST) of a weighted, undirected graph is a spanning tree for which the sum of its edge weights (the weight of the spanning tree) is less than or equal to the weight of every other spanning tree (Figure A.1 (d)).

Images may be mapped onto graphs in a number of ways. In this work, every voxel in a given volume represents a vertex of a graph (i.e. $|V|$ equals the number of voxels in the image) and edges join neighbouring voxels in n-connected neighbourhood or lattice (Figure 7.2). Edge weights then define some measure

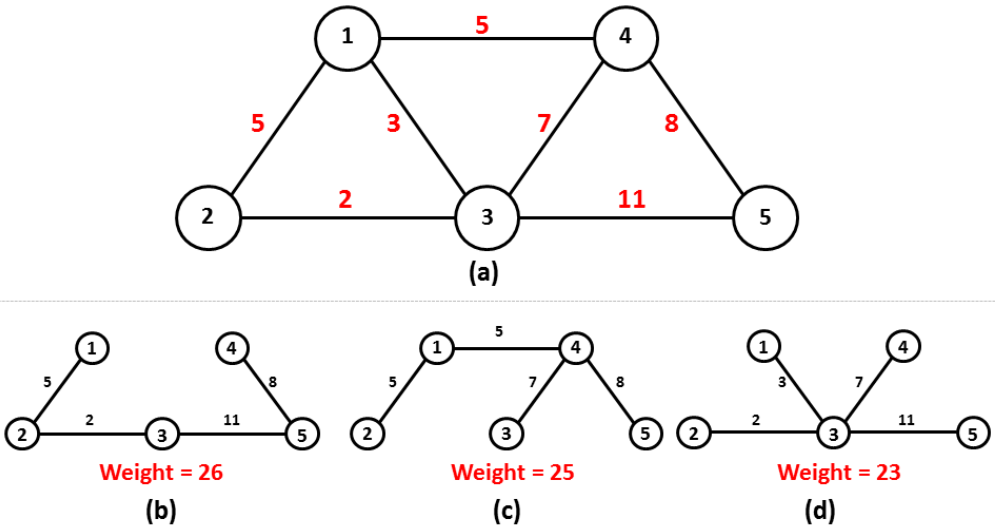


Figure A.1: Graph with corresponding spanning trees.: (a) simple, undirected, weighted graph with 5 vertices and 7 edges. Edge weights are indicated in red. (b) - (d) The three spanning trees of the graph in (a) with total weights indicated. The minimum spanning tree (d) has a weight of 23.

$$A = \begin{pmatrix} 0 & 5 & 3 & 5 & 0 \\ 5 & 0 & 2 & 0 & 0 \\ 4 & 2 & 0 & 7 & 11 \\ 5 & 0 & 7 & 0 & 8 \\ 0 & 0 & 11 & 8 & 0 \end{pmatrix}$$
Adjacency Matrix

$$D = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$
Degree Matrix

$$L = \begin{pmatrix} 3 & -5 & -3 & -5 & 0 \\ -5 & 2 & -2 & 0 & 0 \\ -4 & -2 & 4 & -7 & -11 \\ -5 & 0 & -7 & 3 & -8 \\ 0 & 0 & -11 & -8 & 2 \end{pmatrix}$$
Laplacian Matrix

Figure A.2: The adjacency, degree and Laplacian matrices of the graph in Figure A.1 (a).

of similarity between neighbouring voxels (i.e. connected vertices). Edge weights may be computed in a number of ways, depending on the similarity measure desired (e.g. simple voxel intensity difference).

The *partition* of a graph $G = (V, E)$ is defined as the assignment of every vertex in V into two disjoint subsets $V_1 \subseteq V$ and $V_2 \subseteq V$ where $V_1 \cap V_2 = \emptyset$ and $V_1 \cup V_2 = V$ (Figure A.3). A $p-$way partition of a graph $G = (V, E)$ is thus a mapping of the vertex set of G into p disjoint subsets: $P : V \rightarrow [1, \ldots, p]$. The most prominent graph partitioning methods in the image-segmentation literature have been those based on the normalised cuts approach [SM00] (spectral graph theory) and the max-flow/min-cut algorithm [BJ01]. Spectral graph-based approaches, however, rely on the computationally-intensive process of formulating and solving an eigenvector problem [SM00], while max-flow/min-cut algorithms typically require significant user interaction (e.g. setting of sinks/sources) [BJ01]. Consequently, both spectral and min-cut partitioning approaches are computationally intensive and have been predominantly limited to 2D segmentation problems [Gra06].

A significant reduction in processing time may be achieved by considering minimal or maximal trees as the underlying graphs in the partitioning process

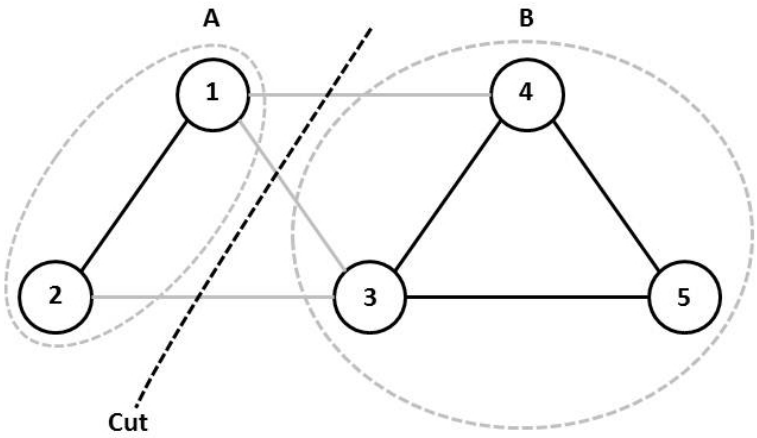


Figure A.3: A two-way partition (graph-cut) resulting in two disjoint partitions A and B. The edges removed by the cut are indicated in grey.

[Zah71, Urq82, MLC86]. A spanning tree of a given graph may be partitioned by removing (or *cutting*) edges in the tree, resulting in a spanning forest composed of a set of disjoint trees (each tree defining a particular partition) [MLC86, SM00]. The mapping of a given partition P back onto an image defines the resulting segmentation. A simple way of performing this (reverse) mapping is to generate a segmentation image wherein each pixel is assigned a constant value according to the partition to which it belongs [MLC86]. Despite the improved efficiency of tree-based partitioning methods, they are known to produce suboptimal segmentations if the underlying graphs are weakly connected [Zah71].

A.1.2 The IDT Algorithm

Grady and Schwartz [GS06a] present an efficient and stable image-segmentation algorithm motivated by the solution to the classical isoperimetric problem [Che70]: finding the shape having the smallest perimeter for a fixed area [Gra06]. While the solution in $\mathbb{R}^2$ is known to be the circle, on a discrete manifold (represented as a graph) the solution is NP-Hard [GS06b]. Grady and Schwartz [GS06b] demonstrate that an optimal approximation to the solution of the isoperimetric problem for graphs may be found by finding a partition that minimises the so-called *isoperimetric ratio* (the ratio of the perimeter of a vertex set to the number of vertices in the set). It is further demonstrated that high-quality segmentations (i.e. high-volume regions with small boundaries) may be obtained by finding the partitions for which this ratio is minimised [GS06a].

The original definition of the isoperimetric constant for an arbitrary, compact manifold is given by [Che70]:

$$h = \inf_{S} \frac{|\delta S|}{\mathrm{Vol}_S} \tag{A.4}$$

where S is a region in the manifold; $\text{Vol}_S \leq \frac{1}{2}\text{Vol}_{\text{Total}}$ is the volume of the region S; δS is the perimeter of S and h is the infimum of the ratio over all possible S. For a finite graph, the infimum in Equation A.4 becomes a minimum; the region S represents a set of vertices $S \subseteq V$ with a volume $\text{Vol}_S = |S| \leq \frac{1}{2}\text{Vol}_V$ and the boundary (perimeter) of the vertex set S is defined by:

$$|\delta S| = \sum_{e_{ij} \in \delta S} w(e_{ij}) \tag{A.5}$$

where $w(e_{ij})$ represents the edge weights [GS06b]. The isoperimetric number of a finite graph $G = (V, E)$ is thus given by [GS06b]:

$$h_G = \min_S \frac{|\delta S|}{|S|} \tag{A.6}$$

It may be further shown that the isoperimetric ratio of a given partition $S \subseteq V$ is given by [Gra06]:

$$h_G(x) = \min_x \frac{x^T L_x}{x^T r} \tag{A.7}$$

where $x^T r \leq |V|$; r is the vector of ones; L is the Laplacian matrix of G (Equation A.3) and x is an indicator vector defining the vertex membership in S:

$$x_i = \begin{cases} 0 & \text{if } v_i \in S \\ 1 & \text{if } v_i \in \overline{S} \end{cases} \tag{A.8}$$

Therefore, by asserting the cardinality constraint $x^T r = k$ for some constant $k \leq \frac{1}{2}\text{Vol}_V$, relaxing the binary definition of x to include non-negative real numbers and by introducing a Lagrange multiplier [Ber99] Λ, an optimal partition may be found by minimising the function [GS06a]:

$$Q(x) = x^T L x - \Lambda(x^T r - k) \tag{A.9}$$

$Q(x)$ will be minimal at its stationary points as L is by definition positive semi-definite [LG+12] and $x^T r >= 0$. The stationary points of $Q(x)$ are found by differentiation with respect to x:

$$\frac{dQ(x)}{dx} = 2Lx - \Lambda r \tag{A.10}$$

The solution to the minimisation of $Q(x)$ thus reduces to solving the linear system:

$$2Lx = \Lambda r \tag{A.11}$$

One is thus left with a singular system of equations (since L is singular) which may be converted to a non-singular system by assigning an arbitrary vertex v_g to S (i.e. setting $x_g = 0$) [Gra06]:

$$L_0 x_0 = r_0 \tag{A.12}$$

where L_0 is computed by removing the g^{th} row and column of L and (x_0, r_0) are computed by removing the g^{th} row of x and r respectively (a process termed *grounding* - by way of a circuit analogy [GS06b]). The scalar multiplier 2 and the Lagrange multiplier Λ may be ignored since only the relative values of the solution to Equation A.11 are of interest [GS06b]. The solution to Equation A.12 (which may be found using memory efficient methods such as conjugate gradients) is non-negative real-valued for x_0 which may by converted into an optimal partition by finding a threshold τ_x (from $n = |V|$ possible thresholds) that minimises the isoperimetric ratio (Equation A.7). The thresholding is performed by placing vertices with $x_i < \tau_x$ into S, and those with $x_i > \tau_x$ into $\overline{S}$. During thresholding, the denominator in Equation A.7 is set to $x^T r$ if $x^T r < \frac{n}{2}$ and to $(n - x^T r)$ otherwise. Grady and Schwartz prove that method ensures that the partition results in a connected object and that the ground vertex serves as a specification of the foreground, while the background is determined by the thresholding of x_0 [GS06b, Gra06].

While the aforementioned isoperimetric partitioning produces efficient solutions for 2D segmentation problems [GS06a], further optimisation is required when considering volumetric imagery [Gra06]. To this end, Grady [Gra06] presents the Isoperimetric Distance Tree (IDT) algorithm - whereby the standard lattice edge set is replaced with a so-called *distance tree*.

A popular and efficient technique for solving systems of linear equations is by the Cholesky decomposition, which is approximately twice as fast as the LU decomposition for symmetric, positive definite systems [Kai80].The solution to the

system of linear equations $A\mathbf{x} = \mathbf{b}$ may be found by first computing the Cholesky decomposition: $A = LL^*$; solving $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$ by forward substitution and then solving $L\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$ by backward substitution. If A is sparse (as is the case in Equation A.12), the objective of a good Cholesky decomposition is to minimise the fill-in of A. For a given symmetric ordering of linear equations, the fill-in of the system matrix A is the number of entries which change from an initial zero to a non-zero value after the decomposition. Determining an ordering that minimises the fill-in can significantly improve the efficiency and stability [Ise09]. The *perfect* ordering (or *zero-fill* ordering) is that for which every zero in A is retained [Ise09]. Such an ordering allows for a solution to the system of linear equations to be found in two passes, with memory storage equal to $n = |A|$. Gremban [Gre96] has shown that symmetric matrices that correspond to trees have orderings that permit Cholesky decompositions with zero-fill. This is the predominant motivation for employing the distance tree as the underlying structure in the isoperimetric partitioning algorithm [Gra06].

The IDT algorithm is thus composed of the following steps [Gra06]:

1. Compute image mask.
2. Compute a distance map on the mask.
3. Determine ground-vertex.
4. Compute the maximal spanning tree (distance tree) on the lattice.
5. Compute the zero-fill ordering of the tree.
6. Solve for x_0.
7. Select the threshold τ_{x_i}, $i = 1, \ldots, n$ minimising the isoperimetric ratio.

In this work, the image mask is computed using the methods described in Section 7.2.1. The distance map on the mask is computed using the fast L1 Chamfer-based approximation to the Euclidean distance [RP66]. For every voxel in the mask, the distance transform returns an integer value equal to its distance from the set of zeros. The zeros thus remain unchanged, the ones neighbouring the zeros remain unchanged; the ones neighbouring such ones become twos etc. The transform operates in $O(n)$.

The ground-vertex is selected as that which has the maximal weighted degree, where the weight of the edge connecting vertices v_i and v_j is given by [Gra06]:

$$w_{ij} = D(v_i) + D(v_j) \tag{A.13}$$

where $D(v_i)$ denotes the Chamfer distance map at v_i. The maximal spanning tree (spanning tree with maximal total weight) is computed using Kruskal's algorithm [Kru56]. Kruskal's algorithm is, in fact, a technique for finding the minimum

Input:	
$G = (V, E)$	input graph
W	edge weight matrix
Tree T	empty minimum spanning tree
for $v \in V$	
makeSet(v)	every edge is a separate component
end for	
E = sort(E, W)	sort edges by ascending weight
for $e_{ij} \in E$	
if findSet(v_i) != findSet(v_j)	edge does not create a cycle
T.Add(e_{ij})	add edge to MST
Union(v_i, v_j)	merge connected component
if T.NumEdges() == $\|V\| - 1$	tree complete
break	terminate
end for	
return T	

Table A.1: Kruskal's algorithm for finding the minimal spanning tree.

spanning tree (as opposed to the maximum). To ensure that the maximum spanning tree is found, the edge weights are thus negated. The maximal spanning tree computed in this way is denoted the *distance tree* [Gra06]. For a graph $G = (V, E)$, Kruskal's algorithm is performed as follows (pseudocode in Table A.1):

1. Begin with a graph consisting of only the vertices of G and no edges. This is essentially a graph of $n = |V|$ disjoint connected components, where each vertex is a connected component.
2. Sort all edges of G in ascending order of weight.
3. Select the smallest edge. If it forms a cycle with the current spanning tree then discard it, otherwise include it in the spanning tree.
4. Repeat step 3 until there are $|V| - 1$ edges in the spanning tree.

Grady [Gra06] presents a linear-time method for computing the zero-fill ordering. Initially, all vertices with unweighted degree of one (i.e. leaf vertices in the distance tree) are eliminated. All vertices which then have updated degrees of one are recursively eliminated until the selected ground node is reached (see Table A.2). Once the ordering has been computed the system in Equation A.12 may be solved in two passes (see Table A.3): 1) a forward pass to modify the right-hand side (i.e. elimination) and 2) a backward pass to compute the solution [Gra06]. The final partition is determined by determining the threshold on the x_0 that minimises the isoperimetric ratio.

Input:	
tree	input tree (each node contains index of one neighbour)
degree	input degree matrix
ground	selected ground vertex
ordering = zeros(n)	initialisation of ordering
$k = 0$	
degree[root]= 0	Fixed to avoid elimination of ground
ordering[$n-1$] = ground	
for $v \in V$	
while degree[v] = 1	
ordering[k] = v	
degree[v] = degree[v] - 1	
v = tree[v]	
degree[v] = degree[v] - 1	
$k = k + 1$	
end while	
$k = k + 1$	
end for	
return ordering	

Table A.2: Method for computing zero-fill ordering [Gra06].

A.2 Segmentation by Region Growing

Region-growing techniques are amongst the most popular approaches to image segmentation. Image segmentation by region-growing may be represented mathematically as follows [WH03]:

$$S(I, RG(\psi), \mathcal{S}) = \bigcup_{i=1}^{M} R_i, \quad \text{where } R_i \cap R_j = \emptyset \quad \text{for } i \neq j \tag{A.14}$$

where I is the input image; $RG(\psi)$ denotes the region-growing algorithm; ψ governs the growing and merging criteria, by specifying the properties that non-seed points must have to be included in the evolving segmented regions as well as the criteria for excluding certain image points from all regions of interest. $\mathcal{S}$ represents the criteria for defining seed points, where a seed point is defined as one that is known to belong to a particular region and specifies the location where the growth of the region should begin. Alternatively, $\mathcal{S}$ may be transformed into an explicit set of seed points $A = a_1, \ldots, a_{M-1} \subset I$, where each point a_i in A is the seed point for the corresponding region R_i. The region growth and merging criteria ψ and the seed criteria $\mathcal{S}$ each consist of a predicate formed as a combination of

Input:
ordering
diagonal (of L_0)
r_0
tree

1. Forward pass
$k = 0$
for each non-ground vertex
  r_0[tree[ordering[k]]] = r_0[tree[ordering[k]]] + r_0[ordering[k]] / f[ordering[k]]
  f[tree[ordering[k]]] = f[tree[ordering[k]]] - 1/f[ordering[k]]
  $k = k + 1$
end for

output[ordering[$n-1$]] = r_0[ordering[$n-1$]]/f[ordering[$n-1$]]

2. Backward pass
$k = n - 2$
for each non-ground vertex
  output[ordering[k]] = output[tree[ordering[k]]] + r_0[ordering[k]] / f[ordering[k]]
  $k = k - 1$
end for

return output

Table A.3: Method for solving $L_0 x_0 = r_0$ [Gra06].

Boolean operations of various image feature measures. Several example binary operations and corresponding region growth and region merging predicates are shown in Table A.5. A voxel would by added to the current region if the predicate P_G were true. Similarly, two regions would be merged if predicate P_M were true. A particular region-growing segmentation is thus described in its entirety by the pair $\langle RG(\psi), \mathcal{S} \rangle$. Segmenting an image using $\langle RG(\psi), \mathcal{S} \rangle$ involves the following steps: 1) definition of the seed points for all regions $\{R_i, \quad i = 1, \ldots, M-1\}$; 2) iterative application of region growing criteria ψ to each region; 3) termination of region growth when the application of ψ results in no further changes to the evolving image. The final, segmented image is then given by $S(I, RG(\psi), \mathcal{S})$. The process is illustrated in Figure A.4 (a).

The predominant focus of the majority of the region-growing literature has been on developing novel growing and merging criteria [PBLL11, DXDL10, TTMG12, PFPB10, ZY96], on incorporating image features into the algorithms [LW10, GVR+10, DXDL10] and on improving computational efficiency [GMKP11, ZXLD11, WRRB09]. The vast majority of region-growing techniques, however, are sensitive to the selection of the region starting points (or *seed* points) as well as the order in which the points in the image are examined. The growth of a region typically

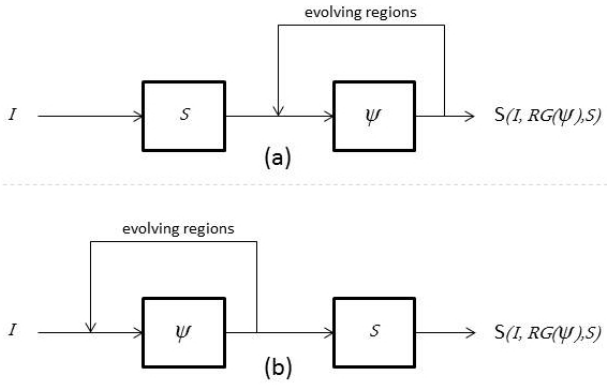


Figure A.4: Flows chart illustrating segmentation by region growing. (a) Conventional region-growing process and (b) region-growing by SymRG [WH03] for input image I; seed criteria $\mathcal{S}$; region growing/merging criteria ψ. The final segmented image is given by: $S(I, RG(\psi), \mathcal{S})$ [WH03].

requires that only small changes in its properties (e.g. mean intensity, gradient etc.) occur after adding new points to the region. Therefore, different starting points for the regions lead to different values for the evolving region information and ultimately to different segmentations. Typically, regions are required to be initiated in relatively homogeneous regions within objects of interest (as opposed to in background regions or near object edges) [WGW12]. Such starting points are generally challenging to locate automatically (especially in highly cluttered and complex images such as baggage-CT scans), making high-quality segmentations at least partially dependent on user input. Furthermore, since the nature of region-growing techniques requires the continual evaluation of image points, computational demands are typically high. This is especially prominent in 3D imagery [TB94, HSKR96].

A.2.1 Symmetric Region Growing

Wan *et al.* [WH03] present a computationally efficient Symmetric Region-Growing (SymRG) framework, whereby region-growing algorithms are made insensitive to their starting conditions. Particularly, the quality of the segmentations of region-growing algorithms that abide by the theoretical criteria of SymRG are insensitive to the both the set of initial seed points as well as the order in which the image points (pixels or voxels) are processed. It is worth noting, however, that it is not claimed that SymRG produces superior quality segmentations (compared to existing techniques), but rather that comparable segmentations may be obtained without the requirement of manual/good algorithm initialisation. Mathematically, the consequence of the SymRG framework may be represented by the following theorem:

$$S(I, RG(\psi), A) \equiv S(I, RG(\psi), B) \tag{A.15}$$

where A and B represent two different sets of seed points. In order for Equation A.15 to hold, it is necessary that the region growing criteria ψ be symmetric in every aspect [WH03]. That is to say, assuming ψ is defined as the function $g(p, q)$, where the parameters p and q represent some information about the evolving region (e.g. mean intensity, mean gradient etc.), then ψ is symmetric if and only if the function $g(p, q)$ is symmetric:

$$g(p, q) = g(q, p) \quad \forall p, q \in I \tag{A.16}$$

and the parameters p and q are independent of their previous states. Any region growing algorithm $RG(\psi)$ is then symmetric provided the above constraints on ψ are met. The symmetry of a region-growing algorithm is in no way dependent on the set of initial seed points A. In fact, the only impact that A has on the segmentation result, is the number of resulting segmented regions. The task now is to define a 3D SymRG framework. The algorithm is composed of two main stages: 1) 2D region growing on the individual slices of the CT volume and 2) region merging between consecutive slices to construct complete regions. To accomplish the task of constructing 3D regions, two issues regarding the merging of regions need to be addressed. Firstly, a method for merging the regions in consecutive slices is required and secondly, since 3D regions may span multiple slices, a method for continuing the growth of a region across multiple slices is also required.

Merging between adjacent slices is accomplished via the use of two global structures: a *region table* and an *equivalence table* [WH03]. The region table stores the information on individual grown regions. Each entry in the table stores the following information: [*region ID; region bounding box; number of pixels in region; number of 0-1 crossings; number of seeds for region*; *pointer to equivalence table*]. The equivalence table stores information on 3D regions after merging. The table is incrementally adapted after two *equivalent* (i.e. homogeneous) regions have merged. Each entry in the table represents a growing region and stores a linked-list of the region IDs of the equivalent 2D regions constituting the 3D region as well as accumulated region information gathered from the region table. Each entry in the region table is linked to its corresponding equivalence table via a pointer. On completion of the region-growing and merging process, the equivalence table is taken to be the final region table (containing all information describing segmented regions).

Region growth (i.e. merging) across multiple slices is achieved via a region labelling scheme. For example, once the region-growing algorithm has terminated on the first slice ($k = 0$), every entry in the region table (i.e. every 2D region in the first slice) is labelled as either: *interesting*, *pending* or *background*. *Interesting* regions contain seed points; *pending* regions contain no seed points but contain points satisfying some loosened criteria and *background* regions contain the remaining points. The region information for interesting and pending regions is stored in the region table (and accessed in the merging process for the next slice), while background regions play no further part in the region growing process and are thus neglected. Furthermore, each entry in an equivalence table is labelled as either *active* (involved in the merging process in the current slice) or *inactive* (not involved in merging). If a 3D region (an equivalence table entry) has been labelled as inactive after the merging process for the current slice (i.e. its growth is complete), a decision is made on whether or not the region will form part of the final segmentation. In particular, an inactive region is labelled as *desired* if it meets the minimum requirements to be accepted as a final region and as *deletable* if it does not. Once the final slice in the volume has been processed, every entry in the equivalence table will be labelled as either *desired* or *deletable*. All voxels in the deletable regions are considered background while the desired regions are kept as the final 3D regions (or objects) in the segmented volume. The complete 3D SymRG process is illustrated in Table A.4. Figure A.4 (b) illustrates the basic flow of the SymRG segmentation procedure (compared to conventional region-growing). Importantly, the SymRG seed criteria are not required until after the regions have been grown, when they are used to label the final regions in the segmentation. This is in contrast to conventional region-growing, that requires a set of seed points ($\mathcal{S}$) to initiate the growth. The final output of the SymRG algorithm is an image in which each voxel stores the ID of its member region and a region table containing all the relevant 3D region information.

The 3D SymRG algorithm makes use of the following functions:

- $Construct1DRegions(j, \psi)$: Constructs 1D line segments on the j^{th} row of a 2D image by applying growing criteria ψ. The output is an updated region table.
- $Merge(n, k, \psi)$: Merges overlapping and similar $(n-1)$-dimensional regions in the k^{th} and $(k-1)^{th}$ $(n-1)$-dimensional images using ψ. The output is an updated equivalence table.
- $LabelRegions(I, \mathcal{S})$: Assigns the final region labels in the image I to the regions that contain seeds satisfying $\mathcal{S}$. All other regions are labelled as background. The output is an updated equivalence table containing the final region labels.

A computationally efficient implementation of the SymRG algorithm can be

Function: $3DSymRG(I, RG, \psi, \mathcal{S})$
- Perform $2DSymRG(I_0, RG, \psi, \mathcal{S})$
For: slice I_k where $k = 1, \ldots, N_z - 1$
    - Perform $2DSymRG(I_k, RG, \psi, \mathcal{S})$
    - Perform $Merge(3, k, \psi)$
end For
- Perform $LabelRegions(I, \mathcal{S})$

Function: $2DSymRG(I_k, RG, \psi, \mathcal{S})$
- Perform $Construct1DRegions(0, \psi)$
For: row j where $j = 1 \cdots N_y - 1$
    - Perform $Construct1DRegions(j, \psi)$
    - Perform $Merge(2, j, \psi)$
end For
- Perform $LabelRegions(I_k, \mathcal{S})$

Table A.4: The complete 3D SymRG algorithm for a volumetric image composed of N_z axial slices.

obtained by creating a binary input image, where every voxel is labelled as either foreground or background. This can be achieved via simple thresholding or a more sophisticated approach, such as the combination of Metal Artefact Reduction (MAR) and the Mumford-Shah functional as used by [GSK+12]. Passing a binary input image to the SymRG algorithm and merely defining growing criteria that assigns all object points to a valid region (where the only constraint on regions is a minimum size), eliminates the need for seed criteria and reduces the segmentation to a simple connected component labelling of the image. Segmentation performed in this manner is completed in a single pass of the image. The symmetric region growing algorithm presents a framework for computationally efficient region-growing segmentations without a decline in segmentation quality. Considering its computational efficiency, the fact that it is fully automated (invariant to algorithm initialisation) and the fact that the technique has been successfully applied to 3D medical-CT segmentation [WH03, WKRH00], the SymRG framework is a potentially attractive option for baggage-CT segmentation.

A.2.2 3D Flood-Fill Region Growing

Wiley *et al.* [WGW12] present a 3D region-growing method based on the Stratovan Tumbler medical-segmentation technology [Wil09]. Segmentation is achieved irrespective of object shapes, topologies, and orientations. A 3D kernel traverses a volume in a flood-fill manner provided the traversed voxels satisfy some predefined

$P_0 = \left\{ I(x, y, k) \in [G_{seedMin}, G_{seedMax}] \right\}$	(x, y, k) is a valid seed
$\overline{P_0} = \{ I(x, y, k) \notin [G_{seedMin}, G_{seedMax}] \}$	(x, y, k) is not a valid seed
$P_1 = \{ I_k(x', y') \text{ undefined} \}$	(x', y', k) has not been examined yet
$P_2 = \{ I(x', y', k) \in [G_{seedMin}, G_{seedMax}] \}$	neighbour (x', y', k) is a valid seed
$P_3 = \{ I(x', y', k) \in [G_{min}, G_{max}] \}$	neighbour (x', y', k) is in allowed intensity range
$P_4 = \{ \lvert I(x, y, k) - I(x', y', k) \rvert \leq G_{tol} \}$	$I(x', y', k)$ is close enough to $I(x, y, k)$ to be added to region
$P_G = P_1$ AND $\{(P_0$ AND $(P_2$ OR $(P_3$ AND $P_4)))$ OR $(\overline{P_0}$ AND $(P_3$ AND $P_4))\}$	region growing predicate
$P_M = [P_0$ AND $(P_2$ OR $P_4)]$ OR $[\overline{P_0}$ AND $P_4]$	region merging predicate

Table A.5: Example Boolean rules and predicates controlling region growth and merging. The uppercase symbol G represents image intensity.

criteria. In particular the algorithm is composed of five stages: 1) Definition of 3D kernel; 2) Determination of movement criteria; 3) Seed initialisation; 4) Flood-fill and 5) Splitting and merging.

Definition of 3D kernel: A spherical kernel is proposed, provided the size of the kernel is smaller than the object being segmented but larger than any expected holes in its boundary. The size of the kernel for a particular seed voxel is determined automatically, based on the degree of clutter (quantified by the local image gradient) in the vicinity of that voxel. Smaller kernels are used in highly cluttered regions, while larger kernels are better suited to more homogeneous image regions.

Determination of movement criteria: The movement of a kernel from one voxel location to the next is controlled by the statistics of the voxels to be traversed. Particularly, the kernel is moved to a new location provided that some predefined criteria are met (e.g. the mean intensity of the voxels to be traversed are greater than some threshold). The specific movement criteria for a given kernel size are determined automatically using a training procedure. An initial bound on the mean intensity computed from the kernel when placed at each seed point is defined as: $[0.98\overline{I}, 1.02\overline{I}]$ - where $\overline{I}$ is the mean intensity computed from the kernel. This

While Queue not empty:
- Remove last queue entry
- Compute $min, max, mean, std$ of voxel intensities
- Compare to movement criteria

If criteria met:
- Move kernel to new position
- Mark traversed voxels as 'object'
- Add neighbours of new position to queue

Else
- Return to step 1.

end While

Table A.6: The flood-fill process.

threshold range is matured by manually improving errant segmentations, adding each improvement to a central training file. A polynomial is fitted to the training points and used to automatically determine the movement criteria at any voxel.

Seed initialisation: An ordering method for every voxel in a given volume is presented. The automatic determination of seed point priority is enabled by: running large kernel sizes first; running high intensity voxels first; starting in the centre of objects as opposed to on the edges (determined by local image gradient) and by increasing the priority of 'thin' objects. Once a seed point has been determined, the neighbouring voxel coordinates and the direction of movement are placed in a queue.

Flood-fill: The flood-process is driven by the aforementioned voxel queue. A single flood-fill cycle terminates when its corresponding queue is empty - this represents the segmentation of a single object. A new kernel is then initiated at the next unsegmented seed point in the volume (determined by the aforementioned ordering procedure) and the segmentation of the next object is commenced. The flood fill procedure is described Table A.6.

Splitting and merging: The splitting and merging phase of the algorithm is intended to deal with the ambiguities associated with the definition of an object and in particular with the issue of compound objects (discussed in Section 7.1. In fact, the authors claim that no dedicated 'splitting' step is required. Instead, hierarchical trees are used to define the relation of the multiple components constituting a compound object (i.e. each compound object is represented both as a set of individual parts and as a whole). In particular, objects are initially segmented into multiple parts. The degree of overlap between each pair of delineated parts is then examined. The parts are merged if they overlap by more than 10%. Once the merging is complete a compound object is represented both as a whole as well

as a set of parts (all the parts involved in the merge).

The algorithm is shown to produce high-quality segmentations for homogeneous objects and results in good separation of touching objects. Performance deteriorates considerably for low contrast objects (< 800 MHU) and thin objects (particularly for those touching other objects) and is shown to be sensitive to common CT image artefacts. The authors also indicate that performance relies on near isotropic voxel resolutions in all three dimensions.

Appendix B

Segmentation Descriptor

Kohlberger *et al.* [KSA+12] present a set of shape and appearance features falling into one of five categories: 1) unweighted geometric features (quantifying the size and regularity of the segmentation); 2) weighted geometric features (locally emphasising the geometric features when intensity values are similar to each other); 3) intensity features (measuring absolute intensity and intensity distributions within segmentations); 4) gradient features and 5) ratio features (computed as ratios of previously computed features). In the original studies [GSK+12, KSA+12] it is noted that several of the features have been inspired by metrics used in previous segmentation-based literature. For example, the geometric features are employed in several early segmentation studies [MS89]; the concept of weighted geometric features was first proposed by Caselles *et al.* [CKS97] in their work on geodesic active contours and several of the ratio features are variations of the *cut-over-volume* ratio, which has been used extensively as an objective function in graph-cut based segmentation algorithms [SM00, GS06a]. Feature weights are computed using the Cauchy distribution function [JKK05]:

$$w(i,j) = \frac{1}{1+\beta(\frac{I_i - I_j}{\rho})^2} \tag{B.1}$$

Where I_i and I_j are the image intensities of neighbouring voxels v_i and v_j; β is a constant that controls the sensitivity of the weight to the intensity difference and the normalisation factor $\rho = \max_{(x,y)\in S} || \bigtriangledown I(x,y)||_1$ is the maximum L1 norm of all the intensity gradients within the segmentation mask, $S \subseteq V$. Furthermore, $w_+(i,j)$ is defined as follows:

$$w_+(i,j) = \begin{cases} w(i,j) & \text{if } I_i > I_j \\ 1 & \text{otherwise} \end{cases} \tag{B.2}$$

Similarly, $w_-(i,j) = 1$ if $I_i > I_j$ and $w_-(i,j) = w(i,j)$ otherwise. Given these

definitions, the following features are computed for a given segmentation mask $S \subseteq V$.

Unweighted geometric features quantifying the size of the segmentation mask: segment volume, defined as the number of voxels in the segment mask; segment surface area, defined as the number of edges in the mask (those voxels in the mask who have a neighbouring voxel that is not in the mask) and total curvature of the segment mask, defined as the sum of the mean curvatures of the surface of S:

$$\text{Volume}(S) = |S| \tag{B.3}$$

$$\text{Surface Area}(S) = \sum_{i,j:v_i \in S, v_j \in \overline{S}} 1 \tag{B.4}$$

$$\text{Total Curvature}(S) = \sum_{i,j:v_i \in S, v_j \in \overline{S}} H(i,j) \tag{B.5}$$

Weighted geometric features again quantify the size of the segmentation mask, but are additionally emphasised or suppressed based on local intensity differences (according to Equation B.1): the weighted volume, defined as the sum over the weights of all the voxels in the segment S; the weighted cut, defined as the sum over all the edge weights along the boundary of S; the weighted curvature, defined as the sum of the mean curvatures weighted by the local edge weights; the low-high and high-low weighted cuts along the boundary of S:

$$\text{Weighted Volume}(S) = \sum_{i,j:v_i,v_j \in S} w(i,j) \tag{B.6}$$

$$\text{Weighted Cut}(S) = \sum_{i,j:v_i \in S, v_j \in \overline{S}} w(i,j) \tag{B.7}$$

$$\text{Total Weighted Curvature}(S) = \sum_{i,j:v_i \in S, v_j \in \overline{S}} w(i,j)H(i,j) \tag{B.8}$$

$$\text{Low-High Weighted Cut}(S) = \sum_{i,j:v_i \in S, v_j \in \overline{S}} w_+(i,j) \tag{B.9}$$

$$\text{High-Low Weighted Cut}(S) = \sum_{i,j:v_i \in S, v_j \in \overline{S}} w_(i,j) \tag{B.10}$$

Intensity features include mean intensity of the voxels in the segmentation mask; the median intensity of the voxels in the mask; the sum of the intensities over the mask; the minimum and maximum intensities in the mask; the Inter-Quartile Range (IQR) [JMF11] of the mask intensities, measuring the statistical

dispersion of the intensities in the segmentation mask - particularly, half the difference between the first quartile (25^{th} percentile) and the third quartile (75^{th} percentile values) and the standard deviation of the intensities in the mask:

$$\text{Mean Intensity: } \mu_I = \frac{1}{|S|} \sum_{v_i \in S} I_i \tag{B.11}$$

$$\text{Median Intensity}(S) = \text{median}(\{I_i : v_i \in S\}) \tag{B.12}$$

$$\text{Total Intensity}(S) = \sum_{v_i \in S} I_i \tag{B.13}$$

$$\text{Min Intensity}(S) = \min_{v_i \in S} I_i \tag{B.14}$$

$$\text{Max Intensity}(S) = \max_{v_i \in S} I_i \tag{B.15}$$

$$\text{Std Dev}(S) = \frac{1}{|S| - 1} \sum_{v_i \in S} (I_i - \mu_I)^2 \tag{B.16}$$

Gradient features include the sums of the L1 and L2 gradient norms; the means of the L1 and L2 gradient norms; the median of the L1 gradient norms; the minimum and maximum L1 gradient norms; the standard deviations of the L1 and L2 gradient norms and the IQR of the L1 gradient norms. In the following formulations, all gradients are computed via central differences:

$$\text{Total L1 Gradient Norm}(S) = \sum_{v_i \in S} || \bigtriangledown I_i||_1 \tag{B.17}$$

$$\text{Total L2 Gradient Norm}(S) = \sum_{v_i \in S} || \bigtriangledown I_i||_2 \tag{B.18}$$

$$\text{Mean L1 Gradient Norm: } \mu_{g1} = \frac{1}{|S|} \sum_{v_i \in S} || \bigtriangledown I_i||_1 \tag{B.19}$$

$$\text{Mean L2 Gradient Norm: } \mu_{g2} = \frac{1}{|S|} \sum_{v_i \in S} || \bigtriangledown I_i||_2 \tag{B.20}$$

$$\text{Median L1 Gradient Norm}(S) = \text{median}(\{|| \bigtriangledown I_i||_1 : v_i \in S\}) \tag{B.21}$$

$$\text{Min L1 Gradient Norm}(S) = \min_{v_i \in S} || \bigtriangledown I_i||_1 \tag{B.22}$$

$$\text{Max L1 Gradient Norm}(S) = \max_{v_i \in S} || \bigtriangledown I_i||_1 \tag{B.23}$$

$$\text{Std Dev L1 Norm}(S) = \frac{1}{|S| - 1} \sum_{v_i \in S} (|| \bigtriangledown I_i||_1 - \mu_{g1})^2 \tag{B.24}$$

$$\text{Std Dev L2 Norm}(S) = \frac{1}{|S| - 1} \sum_{v_i \in S} (|| \bigtriangledown I_i||_2 - \mu_{g2})^2 \tag{B.25}$$

Ratio features are defined as ratios of the previously computed features. Particularly, the following ratios are computed: all four weighted and unweighted combinations of cut divided by volume; all four combinations of low-high weighted cut or high-low weighted cut divided by unweighted or weighted volume; weighted cut divided by unweighted cut; all four combinations of low-high weighted cut or high-low weighted cut divided by unweighted or weighted cut; total L2 gradient norm divided by total L1 gradient norm; curvature divided by unweighted cut and weighted curvature divided by unweighted cut.