

AN EXPERIMENTAL SURVEY OF METAL ARTEFACT REDUCTION IN COMPUTED TOMOGRAPHY

Andre Mouton*, Najla Megherbi*, Katrien Van Slambrouck⁺, Johan Nuyts⁺, Toby P. Breckon*

*School of Engineering, Cranfield University, Bedfordshire, UK.

⁺ Dept. Nuclear Medicine and Medical Imaging Research Centre, KU Leuven, Belgium

ABSTRACT

We present a survey of techniques for the reduction of streaking artefacts caused by metallic objects in X-ray Computed Tomography (CT) images. A comprehensive review of the existing state-of-the-art Metal Artefact Reduction (MAR) techniques, drawn predominantly from the medical CT literature, is supported by an experimental comparison of twelve MAR techniques. The experimentation is grounded in an evaluation based on a standard scientific comparison protocol for MAR methods, using a software generated medical phantom image as well as a clinical CT scan. The experimentation is extended by considering novel applications of CT imagery consisting of metal objects in non-tissue surroundings acquired from the aviation security screening domain. We address the shortage of thorough performance analyses in the existing MAR literature by conducting a qualitative as well as quantitative comparative evaluation of the selected techniques. We find that the difficulty in generating accurate priors to be the predominant factor limiting the effectiveness of the state-of-the-art medical MAR techniques when applied to non-medical CT imagery. This study thus extends previous works by: comparing several state-of-the-art MAR techniques; considering both medical and non-medical applications and performing a thorough performance analysis, considering both image quality as well as computational demands.

Index Terms— Medical CT, security CT, Metal Artefact Reduction (MAR)

1. INTRODUCTION

Metal artefacts can corrupt CT images so severely that they become extremely challenging 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 [1], and only yields satisfactory reconstructions in ideal conditions. In the presence of high-density objects, such as metals, however, 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 and star-shaped artefacts (see Fig. 1). 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 [2]. Efficient strategies to minimise the impact of metal artefacts on the readability of CT images are thus 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 [3, 4] - 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.* [5] 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.* [6] 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. Beyond the work of Mouton *et al.* [7], where a novel technique for reducing the metal artefacts encountered in security CT imagery is presented, the development of new MAR techniques or the evaluation of existing, medical MAR techniques in novel applications of CT imagery (outside of medicine) are extremely limited [7, 8, 9]. 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 [10]. 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 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 [7].

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. This study thus extends the prior work of [5, 6] in the following ways: by providing a comprehensive literature review of Metal Artefact Reduction (MAR) in CT; by supporting this review with an experimental comparative study; by considering performance under both the standard medical protocol as well as a novel protocol outside of the medical domain, encountered in the aviation security setting and by performing a thorough qualitative as well as quantitative performance analysis, considering both image quality as well as computational cost. The study demonstrates the following key findings: the performance of 12 state-of-the-art MAR techniques (11 medically-based) are shown to decline when applied to non-medical imagery; the challenge of generating accurate prior information is shown to be the major factor limiting the application of medical MAR techniques to non-medical CT images; several state-of-the-art methods are shown to be sensitive to parameter tuning; the feasibility of the reference-free quantitative performance measure of Kratz *et al.* [11] is validated and the tradeoff between image quality and computational cost is shown to be an important factor (especially in non-medical applications) when determining overall performance.

2. ARTEFACT REDUCTION APPROACHES

In X-ray CT the most widely implemented reconstruction technique is the analytical Filtered Back Projection (FBP) algorithm where transformations between the sinogram (projection) and image domains are performed using the forward and inverse Radon transforms respectively [1]. 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 any other imperfections [12]. When the projections are incomplete for a certain range of arguments, however, the FBP reconstructions may become corrupted by artefacts.

In reality, projections are only approximations of the ideal case due to: the finite number of projection samples; Poisson noise in the projection data resulting from the quantum nature of X-ray photons; beam hardening due to the polychromatic nature of the spectra emitted by X-ray tubes and the impact of scattered radiation on the measured readings. 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 may become corrupted by severe artefacts [13].

Metal objects in particular cause significant artefacts in CT images [13]. In an extensive simulation study, De Man *et al.* [14] cite beam hardening (the preferential attenuation of low-energy photons in a polychromatic X-ray beam [15]), scattered radiation, photon (projection) noise and the exponential edge-gradient effect (transaxial 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 [13, 15], we focus on the aforementioned factors in the experimental stage of this study.

The majority of the published Metal Artefact Reduction (MAR) techniques fall into one of four categories: sinogram (or projection) completion methods (Section 2.1.2); iterative methods (Section 2.3); hybrid methods (Section 2.4) and miscellaneous methods (Section 2.4). The first part of this report is a comprehensive review of the current state-of-the-art MAR techniques in each of the aforementioned sub-categories as well as an overview of prior work in comparative evaluation methodologies (Section 3.1).

2.1. Sinogram Completion Methods

Considering a parallel-beam coordinate system (see Fig. 2), a sinogram $\rho(r, \theta)$ is obtained by log-converting the measured intensity profile:

$$\rho(\theta, r) = -\ln \frac{I(\theta, r)}{I_0} \quad (1)$$

where $I(\theta, r)$ is the measured intensity value at an angle θ (between the X-ray beam and the y -axis) and at a distance r (from the origin) and I_0 is the intensity of the unattenuated X-ray beam. The computational problem in CT reconstruction is then to compute (or reconstruct) the original attenuation distribution $\mu(x, y)$. The most common approach to approximating $\mu(x, y)$ from a measured sinogram $\rho(r, \theta)$ is via the transform-based analytical approach known as Filtered Back-Projection (FBP). As the name suggests, FBP combines back-projection with ramp-filtering. This essentially involves applying the inverse Radon Transform [1] to the sinogram data.

The vast majority of sinogram completion-based approaches to MAR rely on reconstructions using the FBP approach. Despite its popularity however, the FBP algorithm suffers from several shortcomings - especially when applied to incomplete projection data.

Streaking artefacts usually arise due to errors of isolated projection readings [16]. The errors are enhanced during reconstruction and then mapped into intense dark and bright lines across the image during back-projection. 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 then combined, ensuring that no straight lines appear in the final image. If there are inconsistencies in the projection data, however, the positive and negative contributions are not combined correctly by the reconstruction process, resulting in streaks in the reconstructed image [17]. When streaking artefacts appear in large quantities they can result in an image which is significantly degraded (see Fig. 1).

Sinogram completion methods typically regard these inconsistencies in the projection data as missing 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 ultimately 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 [3, 18, 19, 20, 21, 22, 23, 24, 25]. The final, artefact reduced image is created by back-projecting the interpolated sinogram and re-inserting the metal-only image into this corrected background 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 [18, 22, 25, 26, 27]. Although such direct interpolation-based MAR approaches were popular in early studies [3, 28], many recent studies have highlighted their limitations [16, 18, 20, 23, 26, 29, 30].

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 [3, 7, 18, 22, 25, 26]. 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 thus been widely implemented, even in some of the most complex MAR techniques [18, 25]. Nonetheless, several studies have claimed that minor segmentation errors may have significant detrimental effects on the overall performance of the MAR technique [19, 20, 24, 31]. These effects generally manifest as a loss of information from both the structures surrounding the metal objects as well as the metal objects themselves, leading to the generation of secondary streaking artefacts in the reconstructed image [14, 30]. More sophisticated segmentation processes such as the mean-shift technique [19, 32] and Mutual Information Maximised Segmentation (MIMS) [24] claim to better preserve edge and contour 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- and/or post-filtering etc.), making it unclear if the improved segmentation alone contributes to the improved final results.

2.1.2. Sinogram Completion

The simplest sinogram-completion techniques employ basic interpolation-based approaches to replace the corrupted data. Variations include, but are not limited to: linear interpolation [33]; spline interpolation [26] and cubic spline interpolation [29].

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 [30] 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 [14, 30]. 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 [29]. 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.

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 [34] and multiple thresholding [20]) to generate priors and then focus on variations in the sinogram completion phase to improve MAR results. Meyer *et al.* [20, 35] and Muller and Buzug [30] 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 [20]. Finally, Meyer *et al.* [36] propose frequency-splitting techniques to reduce the characteristic, undesired blurring 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 [15]. While the use of priors is intended to better preserve edge and contrast information and thereby minimise secondary artefacts [15, 20, 34], 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.* [15] 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, 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 thus expected to be less effective in such domains.

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 [13]. This exacerbates the effects of beam hardening and scatter, usually leading to more prominent streaks in the regions connecting the metal objects [14, 16, 23]. The presence of multiple metal objects also complicates the process of multiclass segmentation, particularly in the regions of and near to the metal objects [15]. Poor segmentations invariably result in poor priors, characterised by a loss of edge information of the anatomical structures surrounding the metal objects [15]. 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) [19]. Several studies have proposed interesting approaches to deal with the challenge of multiple metal objects in particular.

Takahashi *et al.* [21] 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.* [18] 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) [37]. Yu *et al.* [19] 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 [22] 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.* [23, 38] 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.* [39] treat the problem of CT MAR as a scattered data interpolation problem and perform interpolation based on the 2D Nonequispaced Fast Fourier Transform (NFFT) [40]. The technique is, however, met with limited visual improvements. Mouton *et al.* [7] refine the output of a conventional sinogram completion-based approach by im-

posing an upper limit on the intensity of the corrected images and by performing post-filtering using the non-local means filter. The study is targeted particularly at aviation security CT imagery and is the only example of a non-medical application of MAR considered in this survey.

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 [24, 25, 41]. Image inpainting is the process of replacing or modifying (i.e. reconstructing) 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 [42]. Li *et al.* [24] combine Non-Local (NL) inpainting and linear interpolation to replace corrupted sinogram data while Duan *et al.* [25] employ Total Variation (TV) inpainting within the conventional sinogram completion framework. Zhang *et al.* [41] 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.* [25]. The main drawback of using inpainting as opposed to interpolation is the significant increase in computational cost.

2.1.6. The Virtual Sinogram

The vast majority of sinogram completion methods operate directly on the raw sinogram (projection) data. This approach may pose considerable 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.* [18, 26] 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.

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.* [19] propose using an adaptive scale and filter scheme (originally developed by Chen *et al.* [43]) 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.* [29] 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.

2.2. Pre-and Post-Processing Operations

Several studies have implemented interesting pre-and/or 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 [24, 30]. A variety of common 2D image denoising filters have been used for these purposes such as: adaptive filtering [24, 34], Total Variation (TV) filtering [22] and Non-Local Means (NLM) filtering [7].

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) [44, 45].

The problem of iterative reconstruction is usually solved by optimising some 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: 1) an initial reconstruction is estimated (often just an image of zeros); 2) the virtual sinogram of the estimate is computed; 3) the error between the virtual sinogram and 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) [46, 47] is a special case of the Projection Onto Convex Sets (POCS) [48] iterative operator. POCS iteratively solves for f in the reconstruction problem:

$$p = A \cdot f \quad (2)$$

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) [49] and the Simultaneous Iterative Reconstruction Technique (SIRT) [50] 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 [51]. 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 [52]. 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 [16]. 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 [53].

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 [45]. 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 [44]. Ollinger [54] has, however, demonstrated that the EM algorithm converges extremely slowly for transmission reconstructions. Consequently, several ML [55] and Penalised Likelihood (PL) [56] 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 [57] present an alternate simultaneous update algorithm to the transmission EM algorithm of Lange and Carson [44]. 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 [58] and Maximum Likelihood for Transmission (MLTR) [59].

Hudson and Larkin [60], 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 [57, 61]. 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 major 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 [60]. Convergence can, however, be enforced by reducing the step size at each iteration (relaxed update schemes [62]) or by gradually reducing the number of subsets during reconstruction [63]. 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 [64].

A major advantage of ML reconstruction algorithms is that *a priori* information about the image to be reconstructed can be 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 [16], 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 [27]). 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.* [65] 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 [59]. While this Iterative Maximum-Likelihood

PolyChromatic Algorithm for CT (IMPACT) is not a dedicated MAR technique, preliminary experimentation does strongly 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 [66]. Elbakri and Fessler [67, 68] and Menveille *et al.* [69] describe ML methods which, similarly to the IMPACT algorithm, incorporate a polychromatic acquisition model for multiple materials. Van Slambrouck and Nuyts [66] demonstrate that the computational cost of iterative reconstruction using complex reconstruction models (e.g. [59, 65]) 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 [59]; IMPACT - a fully polychromatic model [65] 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 [3, 4], NMAR [20] and FS-MAR [36]). 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 and/or distortion 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 [27, 60, 62, 64, 70, 71, 72, 73]. Despite the development of optimised approaches such as Ordered Subset Expectation Maximisation (OSEM) [60], the Row-Action Maximum Likelihood Algorithm (RAMLA) [62], Model-Based Iterative Reconstruction (MBIR) approaches [72], Iterative Coordinate Descent (ICD) optimisation [74, 75], Block-Iterative (BI) modifications [76] and numerous hybrid methods [27, 64, 70], high computational cost remains the major 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 [16].

The work of Wang *et al.* [53] may perhaps be considered the benchmark for iterative reconstruction techniques aimed specifically at metal artefact reduction in CT. While several earlier studies [46, 77] 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.* [53] present modifications to the Expectation Maximisation (EM) approach [78] and the Simultaneous Iterative Reconstruction Technique (SIRT) [50] (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 [53] adopts simultaneous iterations: while the approach is similar in many ways to the Multiplicative Algebraic Reconstruction Technique (MART)

[46], the MART is derived from the Kaczmarz-method [79] for solving a system of linear equations and is thus 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.* [53] 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 [49]. 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 [53]. The second factor contributing to the superiority (in terms of MAR) of the methods proposed in [53] 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 thus did not compensate for the nonuniform 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 [80] and the Euclidean distance between the measured and the reprojected data [53]. The differences in the computational complexity of the two proposed approaches are shown to be negligible.

The majority of statistically-based iterative reconstruction methods are formulated as unconstrained optimisation models that minimise or maximise some data-dependent cost function [81]. Zhang *et al.* [81] 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 [48] (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 [3], ART [53] and EM [53] 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.

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 [82] propose merging the outputs of a conventional linear interpolation-based approach with that of a Multi-dimensional Adaptive Filter (MAF) [83] 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 [84, 85] present a modified ML-EM approach that uses a weighted ML-EM 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. The authors demonstrate in an overall improvement in image quality. High computational expense is again highlighted as a major concern.

Lemmens *et al.* [31] 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 [12], Bruyant *et al.* [86] 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 low pass filter in the reconstruction process.

The concept of the virtual sinogram [18, 26] has been mentioned above as a potential approach for cases where raw projection data is unavailable. An alternative approach to dealing with such scenarios involves performing all processing in the image domain and ignoring projection data entirely. Naranjo *et al.* [87, 88] propose two approaches for dealing with situations where raw projection data is not available. The first approach [88] involves morphological filtering in the polar domain (to exploit the inherent symmetry in FBP reconstructed images). A second, more recent approach [87] uses a comparison between the current, artefacted 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.

2.5. Summary

Previous work, where the relative performance of several state-of-the-art MAR techniques are compared, is limited [5, 6]. There are presently no comprehensive reviews or quantitative comparative studies of the state-of-the-art MAR techniques. Furthermore, with the exception of the work of Mouton *et al.* [7] (which considers security CT imagery), all of the MAR studies discussed in this review have been intended for use in the medical 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.

The remainder of this study thus aims to provide a more comprehensive performance comparison of several recent state-of-the-art MAR techniques using a standard medical comparison protocol and also investigating how this performance translates to a non-medical protocol encountered in typical engineering and security-type applications.

3. COMPARATIVE METHODOLOGIES

3.1. Prior Methodologies

Performance evaluation is an important yet challenging task. With the exception of the work of Mouton *et al.* [7] (which considers security CT imagery), all of the MAR studies discussed in the preceding review 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 for measuring performance. This makes it rather 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/or organs). In the medical domain, the use of physical phantoms [19, 20, 23, 24, 89] as well as software generated phantoms [16, 19, 27, 31, 65] 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 [90].

Meyer *et al.* [20] 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 [21]. 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 thus 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 [19], 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 [10].

Abdoli *et al.* [18] perform quantitative performance analysis using Bland-Altman plots [91]. 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.* [92] 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.* [11] 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. Due to its nature, this is not the case for the virtual sinogram however [26]. 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 [92], the approach is deemed superior in that it is fully automated.

Mouton *et al.* [7] introduce a novel quantitative analysis technique based on the ratio of noisy 3D SIFT feature points [93] identified. The technique is focussed particularly on quantifying the likely impact of MAR on 3D automated object recognition. To date, this appears to be the only approach that considers quantitative performance analysis of MAR in true 3-dimensional space as opposed to on a per-slice basis.

3.2. Proposed Methodology

3.2.1. Data

Prior work (Section 3.1) 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 [16, 19, 27, 31, 65, 66].

We employ a 2D phantom 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 tissue. The remainder of the phantom is water. The artefact free phantom is shown in Fig. 3 (a). The numerical simulation of the phantom was performed using a simulator based on that presented by De Man *et al.* [14], extended with a distance driven projector [94]. This 2D simulator has been used extensively in previous MAR studies [14, 27, 31, 65, 66]. 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.* [14] 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 1 mm) and a 5-times subsampling of the projection views (to model the continuous rotation of the gantry). The effects of afterglow and detector-crosstalk were not considered. Scatter is simulated according to the following formula:

$$s_i = F_0 \cdot C_c \cdot y_i \cdot m_i \quad (3)$$

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:

$$m_i = \ln(b_i/y_i) \quad (4)$$

where b_i is 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 672 detectors and 1160 views per rotation (360°). We used a focus-to-isocentre distance of 570 mm, a focus-to-detector distance of 1040 mm, 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 Xcomp5 software [95] at a nominal tube voltage of 140 kV (see Fig. 5). After simulation, the fan-beam sinograms were rebinned to parallel beam data [96]. Filtered Back Projection (FBP) was used to create the reconstructed image with 512×512 pixels. The final, artefact corrupted simulation is illustrated in Fig. 3 (c). A reference image (Fig. 3 (a)) was generated using a monochromatic simulation at 70 keV (which approximates the mean energy of the polychromatic spectrum in Fig. 5) - this image is used as a reference in the qualitative (visual comparison) analyses. To allow for quantitative analysis, a second reference image was generated by an identical polychromatic simulation but without metal inserts (Fig. 3 (b)).

In order to evaluate the MAR methods on clinical data, we have included a CT scan of a patient with a double hip prosthesis. 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 nominal tube voltage of 120 kVp, a Computed Tomography Dose Index (CTDI) of 11.9, using a collimation of 16×1.5 mm. A reconstruction slice thickness of 3 mm was used. Prior to reconstruction, the spiral data was rebinned to parallel beam data [96]. The FBP-reconstructed slice used in this study is shown in Fig. 3 (d).

Furthermore, two ‘real-world’, non-medical scans obtained using the CT-80 model dual-energy baggage scanner (Reveal Imaging Inc.), 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). A fan-beam geometry was employed with a focus-to-isocentre distance of 550 mm, a focus-to-detector distance of 1008.4 mm at a nominal tube voltage of 160 kVp. The optimal spatial resolution of the scanner is $1.56 \times 1.61 \times 5.00$ mm. For this study, we have used individual 512×512 axial slices from the volumes generated by the CT-80 scanner (Fig. 4 (c) and (d)). Prior to reconstruction, the fan-beam sinograms were rebinned to parallel beam data [96]. The first of the two real-world security scans contains two metallic objects (firearms) in a container with no background clutter (Fig. 4 (a)). The second real-world scan is of a cluttered passenger bag containing multiple

metallic objects (firearm, belt buckles, metallic zipper etc.) and a variety of commonly encountered objects of varying density (e.g. clothing, bottles etc.) (Fig. 4 (b)). In the context of this study, we consider all non-metallic objects in the scan as clutter.

Security screening CT scans are generally of a poorer quality than the state-of-the-art medical CT data. While sub-millimetre isotropic resolutions in all three dimensions has become the norm in medical CT scanners [97, 98], the real-world data obtained from the CT-80 baggage scanner used in this study, yields an optimal spatial resolution of $1.56 \times 1.61 \times 5.00$ mm. Furthermore, the demand for a higher scan speed in the security screening setting leads to compromises in image quality in both resolution and noise [99]. The resolution of security CT data is thus anisotropic and significantly worse than the state-of-the-art medical data. Anisotropic voxel resolution and poor resolution in the axial plane in particular, compound the effects of image noise and artefacts [2].

The content of security CT scans also differ substantially to that encountered in the medical setting [7]. Medical CT scans generally present with relatively low degrees of clutter and the metallic objects encountered are surrounded by human tissue with known and hence predictable characteristics. The content of a typical security CT scan, however, is highly unpredictable and complex, exhibiting high degrees of clutter. The metallic objects encountered in this setting are not surrounded by human tissue, but rather by any range of materials, which are not known prior to the scan. This makes the fine-tuning of algorithms more challenging. It is widely accepted that the performance of both humans and computer algorithms are severely affected by complexity and clutter [99]. Furthermore, the metal objects encountered in the security setting are both larger and more variable in appearance than those found in the medical setting (e.g. firearm vs. dental filling). As a result, the metal artefacts generated are often more severe.

The purpose of including these real-world security images in the experimental study is thus to evaluate the suitability of the current state-of-the-art medical MAR techniques when applied to images obtained from alternate CT domains.

3.2.2. Performance Analysis

Quantitative performance analysis for the phantom study is performed by computing the Normalised Root Mean Squared Error (NRMSE) [100, 101] of the difference between the FBP-reconstructed image without metal inserts (Fig. 3 (b)) and the MAR-corrected images:

$$\text{NRMSE} = \sqrt{\frac{\sum_{j=1}^{\tilde{N}} (f_j - f_j^{\text{ref}})^2}{\sum_{j=1}^{\tilde{N}} (f_j^{\text{ref}} - \mu)^2}} \quad (5)$$

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 [101].

In the real-world experiments (patient and security data), quantitative performance analysis is performed using the reference-free ground truth metric of Kratz *et al.* [11]. The technique has been described in further detail in Section 3.1. In order to determine the veracity of the reference-free metric, the reference-free and normalised root mean square errors for the phantom data are computed and compared.

While absolute computational times are presented, we emphasise that in the implementation of each of the compared techniques little attention has been paid to optimisation. We therefore acknowledge 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).

Finally, standard qualitative performance analyses (visual comparisons) are performed for all the data. In order to remain consistent with the majority of MAR-based literature, performance is considered on a 2D (per-slice) basis only.

4. APPROACHES EVALUATED

Based on the review in Section 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.* [53] 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 [3] has also been included in the comparison. In summary, 11 sinogram completion-based approaches [3, 7, 18, 19, 20, 22, 23, 24, 34, 35, 36] and 1 iterative reconstruction approach [53] are compared. Hereafter, these 12 techniques are referred to using the following descriptors: Kalender [3]; Wang [53]; Zhao [23]; Bal [34]; Yu [19]; Jeong [22]; Li [24]; Abdoli [18]; Meyer1 [20]; Meyer2 [35]; Meyer3 [36]; Mouton [7]. Unless specified otherwise, algorithm parameters have been selected according to the optimal settings presented in the original works.

5. RESULTS

All the software for this study was developed in MATLAB (The MathWorks Inc., Natick, Massachusetts, USA) and based on the information available in the original publications. The experiments were performed on a regular PC (Intel Core i5, 2.30GHz CPU).

5.1. Medical CT Results

Tables 1 and 2 and Figs. 6, 7 and 9 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.* [53] yielded the lowest error (NRMSE = 0.167) 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 [3], Meyer1 (NMAR) [20], Meyer2 (ANMAR) [35], and to a lesser degree, Mouton [7] approaches) yielded images which, at first glance, appear smooth and apparently free of artefacts. Upon closer inspection however, 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.

While the Mouton image displays a significant reduction in streaking, the method performs poorly when dealing with the dark bands which are common in the regions connecting multiple metal objects. The Meyer3 (FSMAR) [36] approach resulted in a better 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 [23] and Abdoli [18] 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 Mouton (0.221) 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 Kalender approach.

To emphasise the difference in image quality, the metal regions of the best (iterative as well as sinogram completion) and worst methods (in terms of NRMSE) have been magnified and are shown in Fig. 8. The Wang and Zhao images show significant reductions in streaking, while 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 normalised errors (reference-free [11] and NRMSE) for the phantom data are compared in Table 2 and Fig. 9. To allow for a direct comparison, the errors are represented as a percentage of the unprocessed (FBP) error. While smaller variations in the reference-free metric are observed between the different methods, the two metrics resulted in a very similar performance ranking.

Table 3 and Fig. 10 and 11 show the results of the patient experiments. Quantitative error analysis was performed using the reference free metric of Kratz *et al.* [11]. For the most part, the results are fairly similar to the phantom data experiments, and are again largely in agreement with the observations made in the literature. The Wang (17.21%), Zhao (19.63%) and Abdoli (20.18%) approaches yielded the lowest errors, while the Jeong approach (84.22%) was again the worst performing method, performing considerably worse than the linear interpolation approach of Kalender (36.88%) and only marginally better than standard FBP reconstruction (100%) (Table 3). The remaining approaches all resulted in significant reductions in the normalised error (compared to the FBP reconstruction). With reference to the qualitative results in Fig. 11, the interpolation-based approaches (Kalender, Meyer1, Meyer2, Bal and Mouton) 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 Mouton approach again failed to correct for the dark bands in the regions connecting the metal objects. The improvement in image quality yielded by the Jeong approach is limited and despite noticeable secondary streaking (see Fig. 12), the approach still yielded a reasonable reduction in the error (Table 3) 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 (see Tables 1 and 3). 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 approach and within a manageable range. Of the sinogram completion-based approaches, the Bal [34] and Li 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 pre-filtering stage of the algorithm. The plots in Figs. 6 and 10 indicate that performance (in terms of error) is not correlated with computational expense. Excluding the iterative reconstruction approach, 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 1 and 10. Judging performance based on these products alone, emphasises the major drawback of the high computational cost associated with iterative reconstruction (Wang approach).

Fig. 12 shows magnified regions in the patient images produced by the best performing methods (in terms of error). The reduction in streaking evident in the Wang and Zhao images is perhaps less than that observed in the phantom experiments. In fact, the Wang image appears to contain a greater degree of streaking than the Zhao image, despite a lower error. 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.

5.2. Security Screening Results

The quantitative analysis results for the security CT scans are shown in Tables 4 and 5 and Fig. 13 and 16. 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: 8.91% for clutter-free, 18.91% for cluttered), Zhao (8.71% and 23.11%), Mouton (9.01% and 25.17%) and Abdoli (15.12% and 29.22%) approaches showed the most significant improvements. Notably, the Zhao approach outperformed (albeit only marginally) the Wang approach for the clutter free scenario. As was the case in the phantom experiments, the Jeong (42.01% and 89.01%) and Li (35.33% and 65.28%) approaches were the two worst performing methods - yielding significantly higher errors than the Kalender approach (20.13% and 44.73%). 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.* [20, 35, 36] (Meyer1 (24.32% and 44.43%), Meyer2 (22.40% and 43.98%) and Meyer3 (34.81% and 42.81%)) and Bal and Spies [34] (19.67% and 33.60%) which use multi-class segmentation to generate priors. The small performance gains are emphasised further by the normalised products in the third column of Tables 4 and 5, 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 linear interpolation approach were not as significant as indicated in the medical literature.

The images in Fig. 14 to Fig. 18 show the qualitative results for the real-world scans. In all cases the reduction in streaking is considerably less than in the phantom image. While the Wang, Zhao and Mouton approaches yielded significant reductions in streaking for both the clutter-free (Fig. 14) and cluttered (Fig. 17) scenarios, all of the sinogram completion-based approaches led to a loss of and/or distortion in image details to some degree. This is especially evident in the cluttered scenario where, despite the notable reductions in streaking, the Zhao, Bal, Abdoli, Meyer1, Meyer2 and Mouton

images are characterised by a noticeable blurring of the low density regions of the image (Fig. 17). In both the clutter-free and cluttered scenarios, the Jeong approach produced images characterised by rather intense secondary streaking with little, if any, improvement in image quality. As expected, performance on a whole was poorer in the cluttered scenario (Fig. 17), where even the best performing methods (Wang and Zhao) contained noticeable streaking and/or loss of image detail in the non-metal regions. It is difficult to determine whether or not the Mouton approach again performed poorly in image regions corrupted by dark bands since the security scans have a higher dynamic range than medical scans, making the background regions and regions in between metal objects naturally dark (even in the cluttered images).

Fig. 15 and Fig. 18 show magnifications of the best performing techniques in the clutter-free and cluttered scenarios respectively. For the clutter-free case (Fig. 15), although the best performing method in terms of error (Zhao method) produced a considerably cleaner image, a noticeable loss of edge and contrast information is evident (in the form of blurring). The Jeong image is again heavily corrupted by secondary streaking. For the cluttered case (Fig. 18), the Wang and Zhao images again exhibit a significant reduction in streaking relative to the FBP image, but image detail (for example the outline of the bag and low density objects in the direct vicinity of the metal objects) are compromised. This is especially noticeable in 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 and Zhao approaches are negligible, the Wang images appear to be of a slightly better 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 and 5, thereby further verifying the veracity of the reference-free error metric.

Judging from the shapes of the curves in Fig. 13 and 16 as well as the readings in Tables 4 and 5, the processing times in both real-world experiments followed a very similar pattern to that seen in the phantom experiment (Table 1 and Fig. 6).

6. DISCUSSION

It is worth elaborating on several of the aforementioned results. The results of the phantom experiment have highlighted a shortcoming of the Mouton approach [7]. The approach is shown to perform poorly in image regions corrupted by dark bands. A phenomenon which typically occurs in the (straight line) regions connecting the multiple metal objects. This observation reveals a fundamental flaw in the Mouton approach - namely that imposing an intensity limit on the corrected image may undo the corrections of dark bands. Since the approach was originally developed for use on security CT images [7], it is likely that this limitation was overlooked due to the typically high dynamic range of security images, which results in naturally dark background regions, thereby making dark bands less prominent. A possible solution to this could be to reduce or ignore the intensity limiting procedure in the regions connecting the metal objects and is left as an area for future work.

While the interpolation-based approaches (especially the Kalender [3], Meyer1 [20], Meyer2 [35], Bal [34] and Mouton[7]) successfully remove streaking, a common shortcoming is the loss of and/or distortion in edge and contrast information. The Meyer3 approach (FSMAR) [36] 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 [22] and Li [24] 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 [53]), 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. [65, 66, 85]) will lead to further reductions in the errors. The Zhao [23], Abdoli [18] and Mouton [7] 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 approach yields the lowest errors and performs comparably to the iterative approach of Wang [53]. The processing times of the majority of the sinogram completion-based approaches remain manageable and considerably lower than the iterative approach. We do however, emphasise 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.* [3] 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 applications (e.g. baggage checkpoints at airports) for example, processing times are of greater relative importance when 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 ([20, 23, 34, 35, 36]). 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 in those domains. 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.* [15] 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 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 [11] and the NRMSE, as well as the correlation between image quality and error in the real-world studies, gives credence to the veracity 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 - this is a factor which has been neglected to a large degree in the literature and has possibly contributed to exaggerated performance claims.

7. CONCLUSIONS AND FUTURE WORK

This study has provided an overview of techniques for the reduction of streaking artefacts caused by metallic objects in X-ray Computed Tomography (CT) images. A comprehensive literature review of the existing state-of-the-art Metal Artefact Reduction (MAR) techniques, drawn almost exclusively from the medical CT literature, has been presented. The literature review has been supported by 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 MAR literature is limited. In particular, we have considered CT images obtained from the aviation security screening domain which consist of metal objects with no surrounding tissue in both isolated and cluttered environments. We perform a comprehensive quantitative analysis by considering both image quality as well as computational cost and demonstrate the importance (especially in non-medical applications) of considering the tradeoff between the two measures when determining overall performance. A comparative study of this nature, that considers both medical and non-medical applications, has not been conducted previously.

MAR approaches generally fall into one of four categories: sinogram completion methods; iterative reconstruction methods; hybrid methods and miscellaneous methods. The most predominant drawback of standard interpolation-based sinogram completion approaches is the introduction of secondary artefacts in the corrected images and a loss of information, especially in regions in the direct vicinity of the metal objects. Therefore, the majority of recent MAR-based literature has focussed on reducing the occurrence of secondary artefacts in the corrected images. Although iterative reconstruction techniques are known to produce superior results, high computational costs are their major drawback. The majority of recent iterative reconstruction approaches have thus focussed on computational optimisation as opposed to improving image quality.

Based on the review of the literature, the second part of this

study has been an experimental performance comparison of several ‘state-of-the-art’ MAR techniques when applied to one, state-of-the-art numerical medical phantom and two aviation security CT images of varying complexity. While performance trends 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 aviation security domain); 2) sinogram completion-based approaches are generally sensitive to input parameters and require manual tuning. Finally, our experimental analysis has supported the claims of Kratz *et al.* [11] regarding the reliability and effectiveness of their reference-free quantitative performance measure.

The development of techniques to automatically determine optimal parameters may be a valuable contribution of future work. The importance of determining an appropriate tradeoff between image quality and computational cost is highlighted and left as an area for future work. Judging from 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, 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.

Acknowledgements

This project was funded under the Innovative Research Call in Explosives and Weapons Detection (2010) initiative, a cross-government programme sponsored by a number of departments and agencies under the UK Government’s CONTEST strategy, in partnership with US Department of Homeland Security. The authors are grateful for additional support from Reveal Imaging Technologies Inc. (USA).

Table 1. Quantitative analysis results of phantom experiment.

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
Mouton	0.221	9.62	0.020

Table 2. Comparison of NRMSE and reference-free error [11] for phantom data. Errors represented as percentage of FBP error.

Method	NRMSE (% original error)	Ref. Free Error (% 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
Mouton	25.34	21.67

Table 3. Quantitative analysis results for patient data. Errors computed using the reference-free metric of [11].

Method	Normalised Error	Time (seconds)	Normalised Product
FBP	100.00	0.21	0.001
Kalender	36.88	1.32	0.003
Wang	17.21	149.80	0.172
Zhao	19.63	9.22	0.012
Bal	26.40	46.81	0.083
Yu	24.19	7.77	0.013
Jeong	84.22	8.00	0.045
Li	39.28	73.31	0.192
Abdoli	20.18	9.83	0.013
Meyer1	25.01	10.10	0.017
Meyer2	24.90	10.39	0.017
Meyer3	33.71	19.07	0.043
Mouton	23.79	10.25	0.016

Table 4. Quantitative analysis results for real-world data: multiple metallic objects, no clutter. Errors computed using the reference-free metric of [11].

Method	Normalised Error	Time (seconds)	Normalised Product
FBP	100.00	0.13	0.001
Kalender	20.13	0.73	0.001
Wang	8.91	143.94	0.089
Zhao	8.71	5.10	0.003
Bal	19.67	25.81	0.035
Yu	19.83	8.55	0.012
Jeong	42.01	6.56	0.019
Li	35.33	42.14	0.103
Abdoli	15.12	5.33	0.006
Meyer1	24.32	5.24	0.009
Meyer2	22.40	5.23	0.008
Meyer3	34.81	11.18	0.027
Mouton	9.01	6.25	0.004

Table 5. Quantitative analysis results for real-world data: multiple metallic objects, cluttered background. Errors computed using the reference-free metric of [11].

Method	Normalised Error	Time (seconds)	Normalised Product
FBP	100.00	0.23	0.001
Kalender	44.73	1.46	0.004
Wang	22.41	167.80	0.224
Zhao	23.11	10.12	0.014
Bal	33.60	51.68	0.104
Yu	32.75	16.74	0.033
Jeong	89.01	13.01	0.069
Li	65.28	76.12	0.296
Abdoli	29.22	10.33	0.018
Meyer1	44.43	11.04	0.029
Meyer2	43.98	11.23	0.029
Meyer3	42.81	21.38	0.055
Mouton	25.17	13.25	0.002

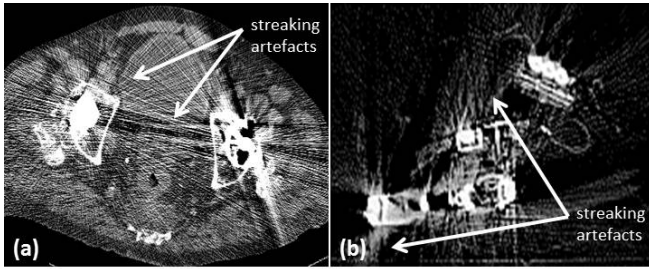


Fig. 1. Example of streaking artefacts caused by metallic objects. (a) Medical scan containing metallic hip replacements (b) Security screening scan containing metallic firearm

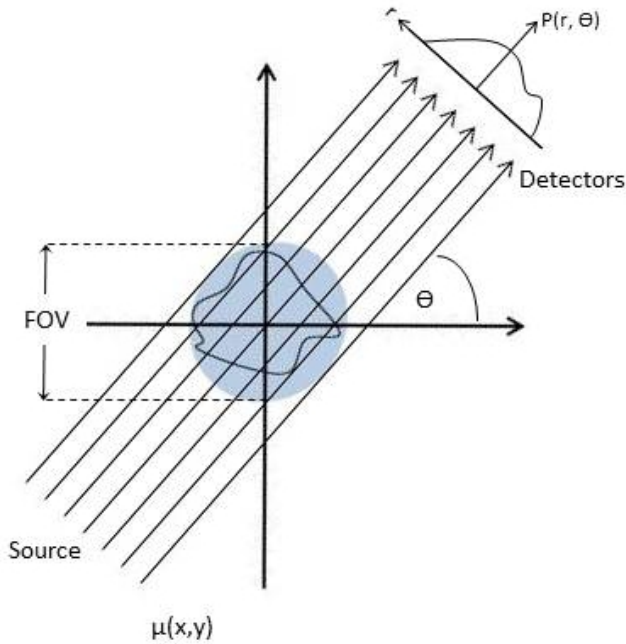


Fig. 2. A parallel beam geometry. The X-ray beams make an angle of θ with the y -axis and are at a distance r from the origin

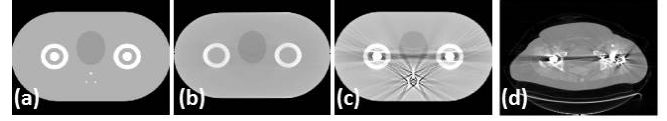


Fig. 3. Simulated and clinical medical data. a) Monochromatic FBP reconstruction of software simulated phantom image (no artefacts). b) Polychromatic FBP reconstruction of metal-free phantom (used as a reference image in quantitative analysis). c) Polychromatic FBP reconstruction of phantom with metal inserts (test image). d) FBP reconstruction of patient scan with double hip prosthesis.

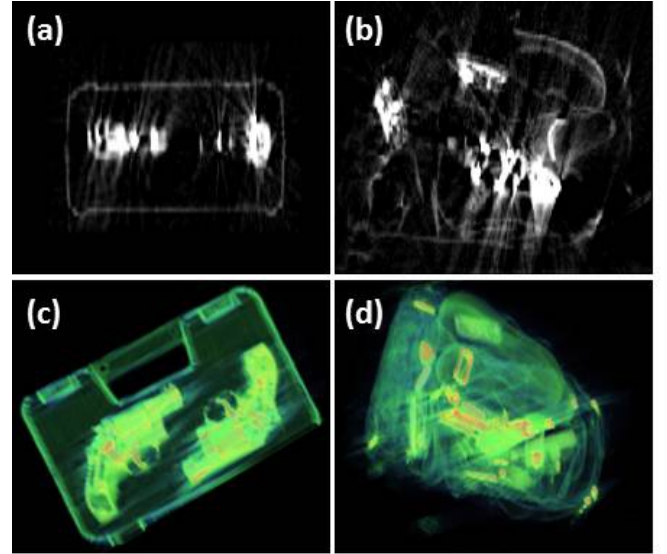


Fig. 4. Real-world security screening CT data. a) Axial slice of two firearms in clutter-free environment. b) Axial slice of multiple metal objects in cluttered environment. c) Source volume of slice in (a). d) Source volume of slice in (b)

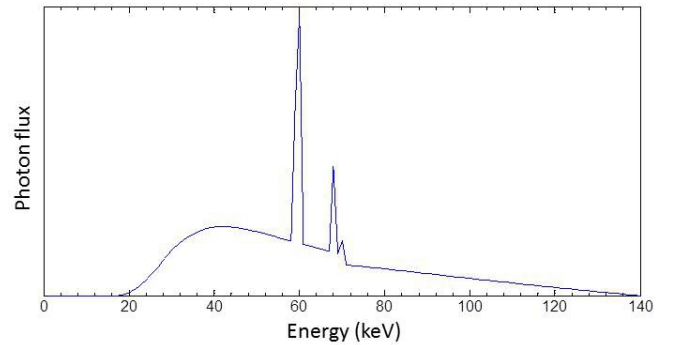


Fig. 5. Simulated spectrum at a nominal tube voltage of 140 kV used in polychromatic simulations. Generated using the Xcomp5 software [95].

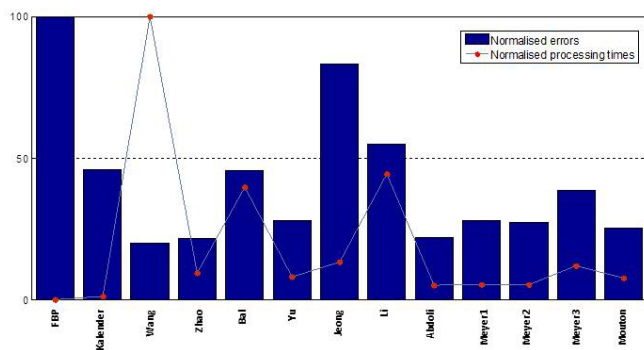


Fig. 6. Graphical comparison of quantitative analysis results of phantom experiments.

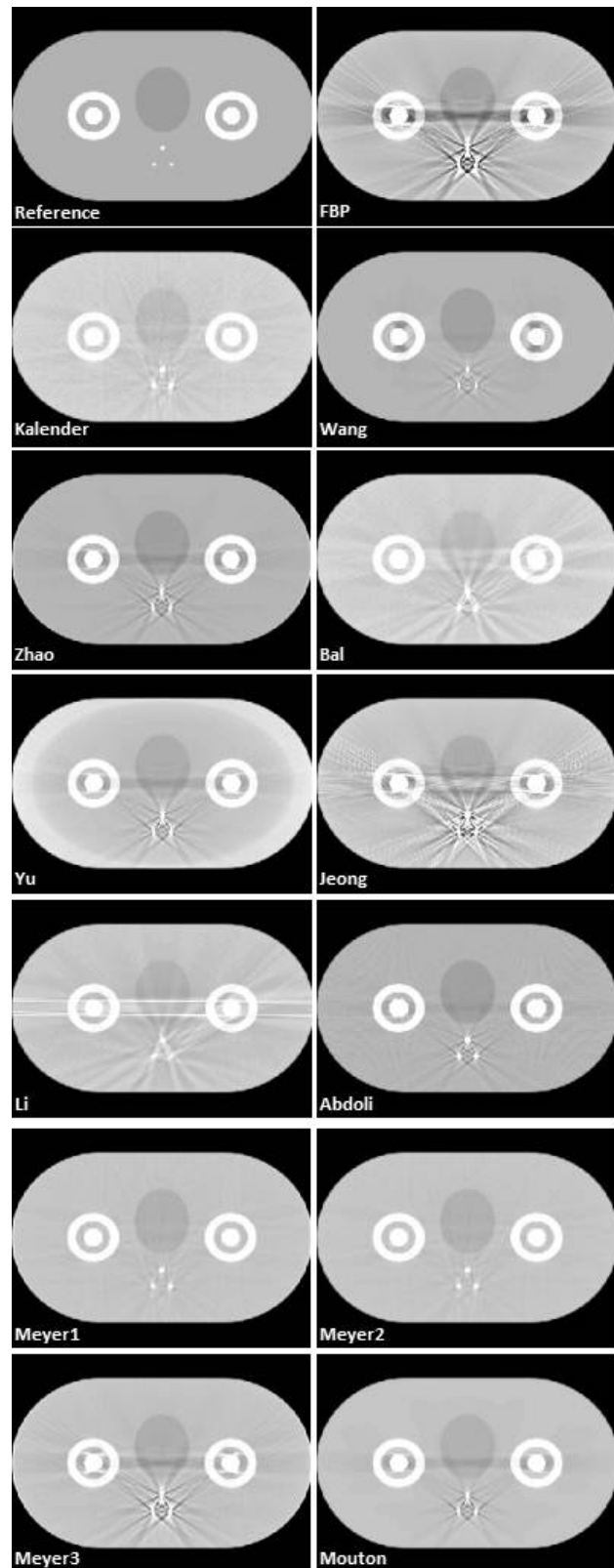


Fig. 7. Qualitative results of phantom experiments. (Window = 800 HU, Centre = 0 HU)

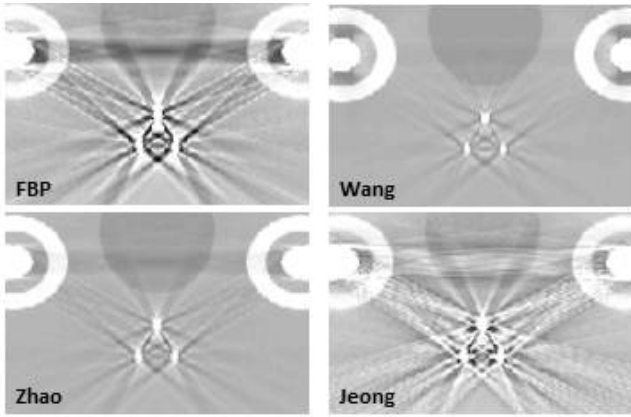


Fig. 8. Magnification of FBP, Wang, Zhao and Jeong phantom images. (Window = 800 HU, Centre = 0 HU)

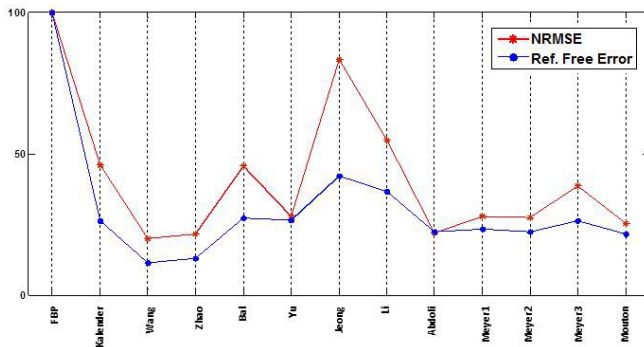


Fig. 9. Graphical comparison of NRMSE and reference-free error [11] for phantom experiments. Errors represented as % of FBP error.

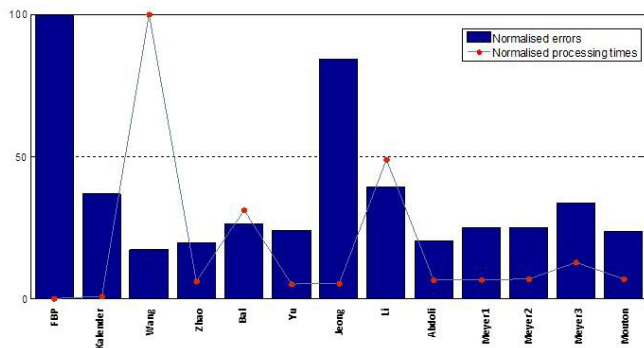


Fig. 10. Graphical comparison of quantitative analysis results of patient experiments.

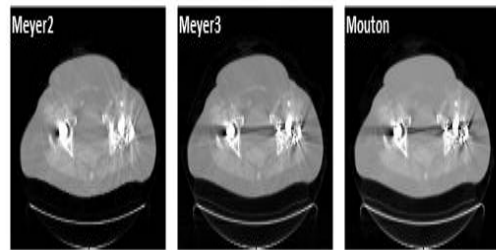
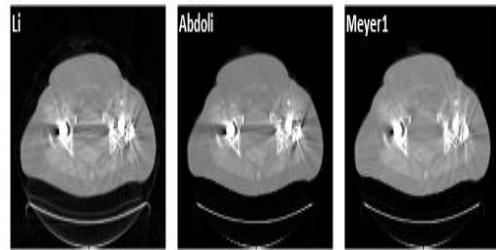
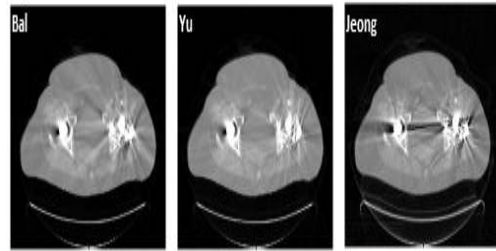
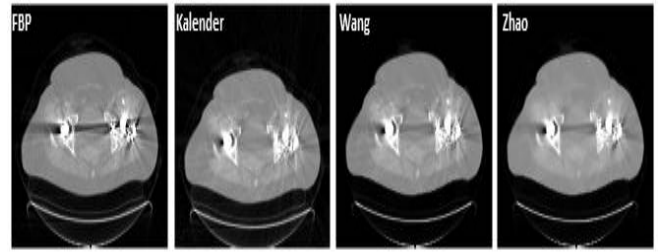


Fig. 11. Qualitative results of patient experiments. (Window = 500 HU, Centre = 0 HU)

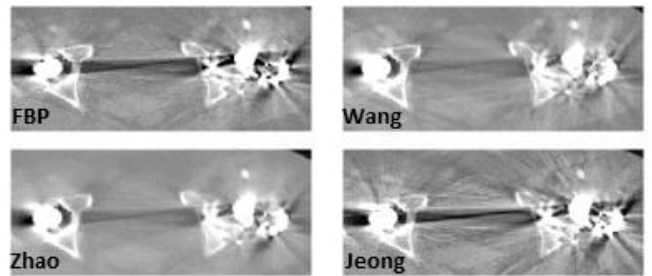


Fig. 12. Magnification of FBP, Wang, Zhao and Jeong patient images. (Window = 500 HU, Centre = 0 HU)

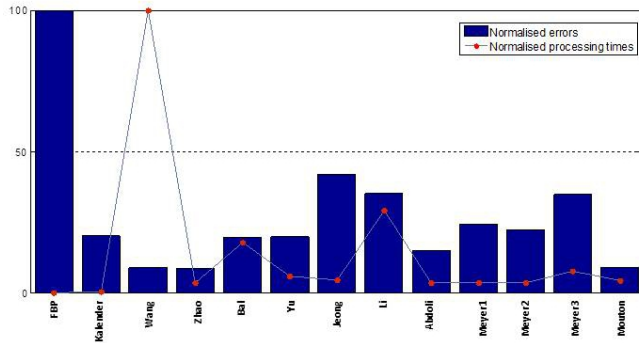


Fig. 13. Graphical comparison of quantitative analysis results for real-world data: clutter-free scenario.

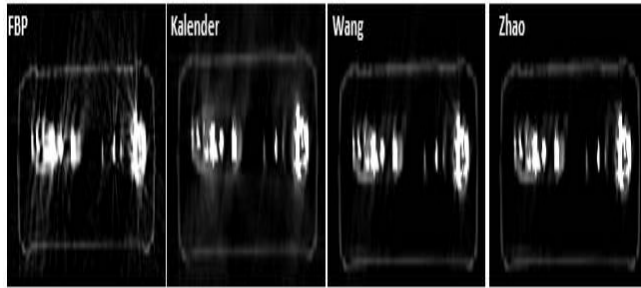


Fig. 14. Qualitative results for real-world data: clutter-free scenario.

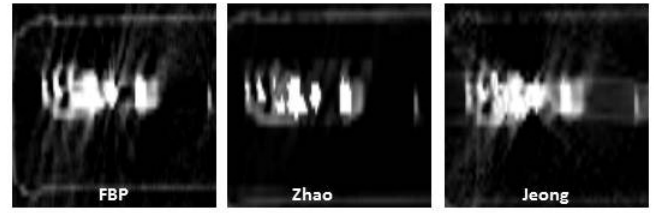


Fig. 15. Magnification of FBP, Zhao and Jeong images: clutter-free scenario

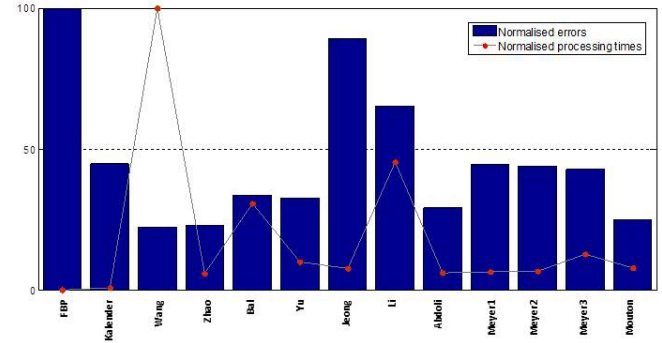


Fig. 16. Graphical comparison of quantitative analysis results for real-world data: cluttered scenario.

8. REFERENCES

- [1] Stanley R. Deans, *The Radon Transform and Some of Its Applications*, Krieger Publishing Company, 1993.
- [2] A. F. Kopp, K. Klingenberg-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*, vol. 68, no. 2, pp. 94–105, 2000.
- [3] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants," *Radiology*, vol. 164, no. 2, pp. 576, 1987.
- [4] G. H. Glover and N. J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT reconstructions," *Medical physics*, vol. 8, pp. 799, 1981.
- [5] 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*, vol. 32, no. 4, pp. 621, 2008.
- [6] 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 *Proc. SPIE*, 2011, vol. 7961, p. 79612Y.
- [7] A. Mouton, N. Megherbi, G.T. Flitton, and T.P. Breckon, "A novel intensity limiting approach to metal artefact reduction in 3D CT baggage imagery," in *IEEE Inter. Conf. on Image Processing*, 2012, pp. 2057–2060.
- [8] 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 *IEEE Nuclear Science Symposium Conference Record*, 2009, pp. 904–908.
- [9] L. Grady, V. Singh, T. Kohlberger, C. Alvino, and C. Bahlmann, "Automatic segmentation of unknown objects, with application to baggage security," *European Conference on Computer Vision*, in press.
- [10] G. van Kaick and S. Delorme, "Computed tomography in various fields outside medicine," *European Radiology Supplements*, vol. 15, pp. 74–81, 2005.
- [11] B. Kratz, S. Ens, J. Miller, and T. M. Buzug, "Reference-free ground truth metric for metal artifact evaluation in CT images," *Medical Physics*, vol. 38, pp. 4321–4328, 2011.
- [12] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*, SPIE and John Wiley and Sons, 2003.
- [13] J. F. Barrett and N. Keat, "Artifacts in CT: Recognition and avoidance," *Radio-graphics*, vol. 24, no. 6, pp. 1679–1691, 2004.
- [14] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Metal streak artifacts in x-ray computed tomography: a simulation study," in *IEEE Trans. on Nuclear Science*, 1999, vol. 46, pp. 691–696.
- [15] Seemeen Karimi, Pamela Cosman, Christoph Wald, and Harry Martz, "Segmentation of artifacts and anatomy in CT metal artifact reduction," *Medical Physics*, vol. 39, no. 10, pp. 5857–5868, 2012.

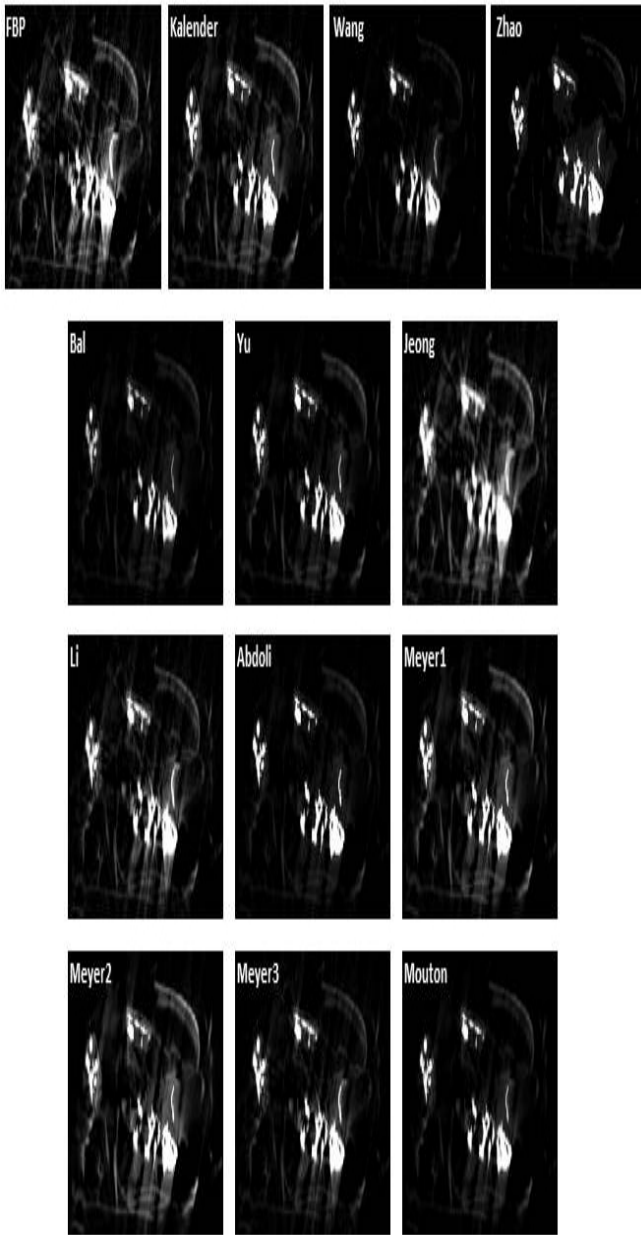


Fig. 17. Qualitative results for real-world data: cluttered scenario.

- [16] B. De Man, *Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography [PhD Thesis]*, Ph.D. thesis, Katholieke Universiteit Leuven, 2001.
- [17] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, New York: IEEE Press, 1988.
- [18] 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*, vol. 37, pp. 6166, 2010.
- [19] 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*, vol. 14, no. 4, pp. 495–504, 2007.
- [20] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical physics*, vol. 37, pp. 5482, 2010.
- [21] 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*, vol. 24, no. 2, pp. 133–138, Feb 2006.
- [22] K. Y. Jeong and J. B. Ra, "Reduction of artifacts due to multiple metallic objects in computed tomography," in *Proc. SPIE*, 2009, vol. 7258, p. 72583E.

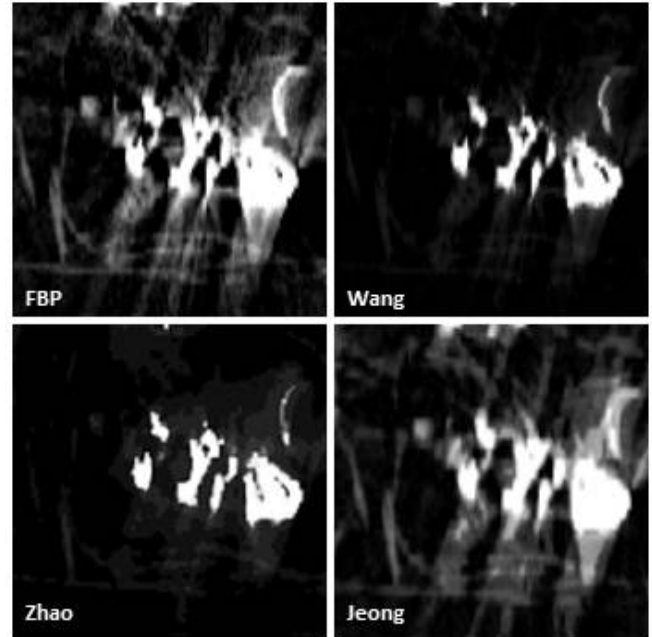


Fig. 18. Magnification of FBP, Wang, Zhao and Jeong images: cluttered scenario

- [23] 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*, vol. 10, no. 2, pp. 67–76, 2002.
- [24] 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 *Biomedical Engineering and Informatics, 2010 3rd Inter. Conf. on*, 2010, vol. 1, pp. 380–383.
- [25] 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*, 2008, pp. 4175–4177.
- [26] 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*, vol. 31, no. 1, pp. 22, 2010.
- [27] B. De 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 Trans. on Nuclear Science*, vol. 47, no. 3, pp. 977–981, 2000.
- [28] 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*, 1990, vol. 1234, p. 642.
- [29] 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*, vol. 2, no. 4, pp. 207–214, 2003.
- [30] J. Muller and T.M. Buzug, "Spurious structures created by interpolation-based CT metal artifact reduction," in *Proc. of SPIE*, 2009, vol. 7258, p. 72581Y.
- [31] C. Lemmens, D. Faul, and J. Nuyts, "Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion," *IEEE Trans. on Medical Imaging*, vol. 28, no. 2, pp. 250–260, 2009.
- [32] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [33] E. Klotz, W. A. Kalender, R. Sokiransky, and D. Felsenberg, "Algorithms for the reduction of CT artifacts caused by metallic implants," in *Proc. SPIE*, 1990, vol. 1234, p. 642.
- [34] M. Bal and L. Spies, "Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering," *Medical Physics*, vol. 33, pp. 2852, 2006.
- [35] E. Meyer, R. Raupach, B. Schmidt, A. H. Mahnken, and M. Kachelriess, "Adaptive normalized metal artifact reduction (ANMAR) in computed tomography," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2011 IEEE, 2011, pp. 2560–2565.
- [36] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Frequency split metal artifact reduction (FSMAR) in computed tomography," *Medical physics*, vol. 39, no. 4, pp. 1904–1916, Apr 2012.

- [37] Wolfgang Banzhaf, Peter Nordin, Robert E. Keller, and Frank 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.
- [38] S. Zhao, D. D. Robelton, G. Wang, B. Whiting, and K. T. Bae, "X-ray CT metal artifact reduction using wavelets: an application for imaging total hip prostheses," *IEEE Trans. on Medical Imaging*, vol. 19, no. 12, pp. 1238–1247, 2000.
- [39] 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 für die Medizin*, pp. 21–25, 2008.
- [40] S. Kunis, *Nonequispaced FFT: generalisation and inversion*, Shaker, 2006.
- [41] 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*, vol. 19, no. 3, pp. 373–384, 2011.
- [42] M. M. O. B. B. Richard and M. K. Y. S. Chang, "Fast digital image inpainting," in *Proc. of the Inter. Conf. on Visualization, Imaging and Image Processing*, 2001.
- [43] L. M. Chen, Y. Liang, G. A. Sandison, and J. Rydberg, "Novel method for reducing high-attenuation object artifacts in CT reconstructions," in *Proc. of SPIE*, 2002, vol. 4684, p. 841.
- [44] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *Journal of Computer Assisted Tomography*, vol. 8, no. 2, pp. 306–316, 1984.
- [45] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.
- [46] 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*, vol. 29, no. 3, pp. 471, 1970.
- [47] G. N. Hounsfield, "Apparatus for examining a body by radiation such as X or gamma radiation," *Patent Specification 1283915, London, England*, 1972.
- [48] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review*, vol. 38, no. 3, pp. 367–426, 1996.
- [49] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.
- [50] R. Gonnor, "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology*, vol. 36, pp. 105–117, 1972.
- [51] 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)*, pp. 1–38, 1977.
- [52] B. Y. Dong, "Image reconstruction using EM method in X-ray CT," in *Wavelet Analysis and Pattern Recognition. Inter. Conf. on*, 2007, vol. 1, pp. 130–134, IEEE.
- [53] G. Wang, D. L. Snyder, J. A. O'Sullivan, and M. W. Vannier, "Iterative deblurring for CT metal artifact reduction," *IEEE Trans. on Medical Imaging*, vol. 15, no. 5, pp. 657–664, 1996.
- [54] J. M. Ollinger, "Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm," *Medical Imaging, IEEE Trans. on*, vol. 13, no. 1, pp. 89–101, 1994.
- [55] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *Signal Processing, IEEE Trans. on*, vol. 41, no. 2, pp. 534–548, 1993.
- [56] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *Medical Imaging, IEEE Trans. on*, vol. 16, no. 2, pp. 166–175, 1997.
- [57] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine and Biology*, vol. 44, no. 11, pp. 2835, 1999.
- [58] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. on Image Processing*, vol. 4, no. 10, pp. 1430–1438, 1995.
- [59] 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*, vol. 43, pp. 729, 1998.
- [60] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 601–609, 1994.
- [61] F. J. Beekman and C. Kamphuis, "Ordered subset reconstruction for x-ray ct," *Physics in Medicine and Biology*, vol. 46, no. 7, pp. 1835–1844, 2001.
- [62] J. Browne and A. B. de Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Trans. on Medical Imaging*, vol. 15, no. 5, pp. 687–699, 1996.
- [63] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM Journal on Optimization*, vol. 7, no. 4, pp. 913–926, 1997.
- [64] Q. Li, S. Ahn, and R. Leahy, "Fast hybrid algorithms for PET image reconstruction," in *Nuclear Science Symposium Conference Record, IEEE*, 2005, vol. 4, pp. 1851–1855.
- [65] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Trans. on Medical Imaging*, vol. 20, no. 10, pp. 999–1008, 2001.
- [66] K. Van Slambrouck and J. Nuyts, "Metal artifact reduction in computed tomography using local models in an image block-iterative scheme," *Medical physics*, vol. 39, no. 11, pp. 7080, 2012.
- [67] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *IEEE Trans. on Medical Imaging*, vol. 21, no. 2, pp. 89–99, 2002.
- [68] 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*, vol. 48, no. 15, pp. 2453–2477, 2003.
- [69] N. Menvielle, Y. Goussard, D. Orban, and G. Soulez, "Reduction of beam-hardening artifacts in X-ray CT," in *IEEE Inter. Conf. of the Engineering in Medicine and Biology Society*, 2005, pp. 1865–1868.
- [70] F. E. Boas and D. Fleischmann, "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography," *Radiology*, vol. 259, no. 3, pp. 894–902, 2011.
- [71] G. Wang, T. Frei, and M. W. Vannier, "Fast iterative algorithm for metal artifact reduction in X-ray CT," *Academic Radiology*, vol. 7, no. 8, pp. 607–614, 2000.
- [72] 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 Trans. on Image Processing*, vol. 20, no. 1, pp. 161–175, 2011.
- [73] L. Cheng, Y. Chen, T. Fang, and J. Tian, "Fast iterative adaptive reconstruction in low-dose CT imaging," in *IEEE Inter. Conf. on Image Processing*, 2006, pp. 889–892, IEEE.
- [74] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *Image Processing, IEEE Trans. on*, vol. 5, no. 3, pp. 480–492, 1996.
- [75] 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*, vol. 34, pp. 4526, 2007.
- [76] C. L. Byrne, "Convergent block-iterative algorithms for image reconstruction from inconsistent data," *Image Processing, IEEE Trans. on*, vol. 6, no. 9, pp. 1296–1304, 1997.
- [77] B. E. Oppenheim, "Reconstruction tomography from incomplete projections," *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine*, vol. M.M. Ter-Pogossian et al., Eds. Baltimore, MD: University Park, pp. 155–183, 1977.
- [78] D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan, "Deblurring subject to non-negativity constraints," *IEEE Trans. on Signal Processing*, vol. 40, no. 5, pp. 1143–1150, 1992.
- [79] S. Kaczmarz, "Angenherte auflösung von systemen linearer gleichungen," *Bulletin International de l'Academie Polonaise des Sciences et des Lettres*, vol. 35, pp. 355–357, 1937.
- [80] I. Csiszar, "I-divergence geometry of probability distributions and minimization problems," *The Annals of Probability*, pp. 146–158, 1975.
- [81] X. Zhang, J. Wang, and L. Xing, "Metal artifact reduction in x-ray computed tomography (CT) by constrained optimization," *Medical physics*, vol. 38, no. 2, pp. 701, 2011.
- [82] O. Watzke and W. A. Kalender, "A pragmatic approach to metal artifact reduction in CT: merging of metal artifact reduced images," *European radiology*, vol. 14, no. 5, pp. 849–856, 2004.
- [83] 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*, vol. 28, pp. 475, 2001.
- [84] M. Oehler and T. M. Buzug, "Modified MLEM algorithm for artifact suppression in CT," in *Nuclear Science Symposium Conference Record*, 2006, vol. 6, pp. 3511–3518, IEEE.
- [85] 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*, pp. 42–47, 2007.
- [86] P. P. Bruyant, J. Sau, and J. J. Mallet, "Streak artifact reduction in filtered back-projection using a level line based interpolation method," *Journal of Nuclear Medicine*, vol. 41, no. 11, pp. 1913–1919, 2000.
- [87] 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 Conf. on Image Processing*, 2011, pp. 461–464.
- [88] 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*, pp. 11–19, 2009.
- [89] J. Muller and T. M. Buzug, "Intersection line length normalization in CT projection data," *Bildverarbeitung für die Medizin 2008*, pp. 77–81, 2008.

- [90] C.J. Solomon and T.P. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*, Wiley-Blackwell, 2010.
- [91] D.G. Altman and J.M. Bland, "Measurement in medicine: the analysis of method comparison studies," *The Statistician*, vol. 32, pp. 307–317, 1983.
- [92] S. Ens, B. Kratz, and T. M. Buzug, "Automatische beurteilung von artefakten in tomographischen bilddaten," in *Biomed Tech*, Rostock, Germany, 2010, vol. 55, p. BMT.2010.550.
- [93] G. Flitton, T.P. Breckon, and N. Megherbi, "Object recognition using 3D SIFT in complex CT volumes," in *Proc. British Machine Vision Conference*, 2010, pp. 11.1–11.12.
- [94] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology*, vol. 49, no. 11, pp. 2463, 2004.
- [95] R. Nowotny and A. Hfer, "Program for calculating diagnostic x-ray spectra," *RoFo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*, vol. 142, no. 6, pp. 685, 1985.
- [96] L. Wang, "Cross-section reconstruction with a fan-beam scanning geometry," *IEEE Trans. on Computers*, vol. 100, no. 3, pp. 264–268, 1977.
- [97] "Toshiba Medical Systems Corporation Aquilion 32 <http://www.toshiba-medical.co.uk/ct-systems.asp> [Jan. 01, 2012];"
- [98] "GE Healthcare discovery CT750 HD <http://www.gehealthcare.com/euen/ct/products/> [Jan. 01, 2012];"
- [99] S. Singh, "Explosives detection systems (EDS) for aviation security," *Signal Processing*, vol. 83, no. 1, pp. 31–55, Jan. 2003.
- [100] R. S. Brock, A. Docef, and M. J. Murphy, "Reconstruction of a cone-beam CT image via forward iterative projection matching," *Medical physics*, vol. 37, no. 12, pp. 6212, 2010.
- [101] G. T. Herman, *Fundamentals of computerized tomography: image reconstruction from projections*, Springer, 2009.