Refinement of Clinical X-ray Computed Tomography (CT) Scans Containing Metal Implants

Graham Treece

CUED/F-INFENG/TR 701

April 2016

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

Corresponding e-mail: gmt11@eng.cam.ac.uk

Abstract

X-ray Computed Tomography (CT) data contains artefacts from many sources and these are sufficiently prominent to affect diagnostic utility when metal is present in the scans. Several techniques have been proposed to reduce these artefacts, the most successful involving the removal and filling in of any sinogram data which has been affected by metal. Most such techniques are prone to introducing new artefacts into the CT data or may take a long time to correct the data. In this paper, a new technique is proposed which is fast, yet can effectively remove most artefacts without introducing significant new ones. This is thoroughly tested against both published and commercial alternatives, quantitatively on phantom data, and qualitatively on a selection of clinical scans, mostly of the hip. The phantom data is from two recently published studies, enabling direct comparison with the technique presented here. The results show an improvement in reducing artefacts on phantom data, in most cases significant at $p < 0.001$, with similar qualitative improvements on clinical data, particularly for bony features close to metal implants.

1 Introduction

X-ray Computed Tomography (CT) imaging has been ubiquitous within medicine for many years, with numerous advances in image quality and reduction of the radiation dose. However, the reconstructed cross-sectional linear attenuation data can still contain multiple artefacts which, though well understood, have the potential to impact diagnostic utility (Suetens, 2002). These artefacts arise from discrepancies between the actual behaviour of the imaging system and the assumptions made in reconstructing material properties from the measured data. They are always present to some extent, but particularly prominent in medical scans containing metal implants, which have much higher attenuation than the surrounding tissue or bone. Since there are many scenarios where a CT scan is required and the likelihood of metal implants (for instance due to a previous fracture) is fairly high, the reduction of these 'metal artefacts' is an important research area.

There are four common sources of CT artefacts due to the presence of metal: beam hardening, beam width, increased scatter or low signal counts, and photon starvation. These are described in more detail by Boas and Fleischmann (2012); Suetens (2002); Verburg and Seco (2012). In summary, beam hardening and photon starvation both result in extended areas leading radially away from the metal with an apparent attenuation which is either too high or too low. The former is due to the mistaken assumption, for instance inherent in Filtered Back-Projection (FBP), that attenuation at each point is independent: this is only true for a strictly mono-energetic X-ray source. The latter is due to the inherent inability of the reconstruction to represent the apparently infinite attenuation implied by receiving no photons at all at the detector. Increased scatter and low signal counts (due to the greater interaction of metal with the incident X-rays) introduce higher levels of detector noise, which appear in the reconstructed data as alternating light/dark radial lines emanating from the metal throughout the rest of the data. Likewise, beam width effects generate radial lines from transitions between metal and the surrounding medium.

Of these sources, beam hardening has probably received the most attention, and most medical CT scans include a physical filter to narrow the range of energies from the X-ray source in order to lessen the consequent effects. Beam hardening artefacts can be further diminished by measuring attenuation independently at different energies (Jeon et al., 2015; Schmidt, 2009; Wu et al., 2014). In medical CT imaging, attenuation is dominated by the photo-electric effect and Compton scattering, and hence it is potentially possible to characterise its energy dependence by measurement at only two different energies (or of two different materials) (Van de Casteele et al., 2002). Mono-energetic images can then be derived at various energies, which ideally would not be affected by beam hardening (Guggenberger et al., 2012). A further advantage if using a mixture of mega-electron volt (MV) and kilo-electron volt (kV) imaging is that there is less attenuation in the MV range and hence photon starvation is unlikely to be an issue, though contrast is also reduced (Wu et al., 2014).

For conventional poly-energetic imaging, beam hardening artefacts are present even for soft tissue, necessitating 'cupping correction', or linearisation of the attenuation variation with material thickness (Kachelrieß et al., 2006). Where measurement at multiple energies is either not desirable or not available, poly-energetic reconstructions can be further improved by allowing for independent linearisation of two or more materials (for instance tissue and bone). Such methods determine the presence and location of each material from an initial uncorrected reconstruction, and use this knowledge to linearise the attenuation and hence refine the reconstruction (Krumm et al., 2008; Menvielle et al., 2005; Vedula and Munshi, 2008), possibly requiring many iterations of this procedure (Van Gompel et al., 2011). Alternatively, the presence and location of bone or metal can be used to predict the effect of beam hardening on the reconstruction, so long as the material and scanner characteristics are well defined (Park et al., 2016). Such predictions consist of both the direct effect of the metal on the transmitted beam energy and the secondary effect on any soft tissue cupping correction already applied to the data (Hsieh et al., 2000). Once modelled, the effects can be removed directly from the reconstructed data. More complex iterative algebraic reconstructions (Wang et al., 1996) or fully poly-energetic spectral reconstruction models (Van Slambrouck and Nuyts, 2012) have also been proposed.

1.1 Sinogram interpolation techniques

One might expect that improving the reconstruction model would be the best approach for artefact reduction, since they arise from incorrect model assumptions. Nevertheless, beam hardening correction methods are generally slow and in any case do not remove all of the artefacts associated with metal in CT data. This is evident from the thorough comparison in Van Slambrouck and Nuyts (2012) which contrasts their poly-energetic reconstruction with other alternatives, and is equally true of dual-energy methods (Jeon et al., 2015). Much of the research into metal artefact reduction (MAR) has instead focused on correcting sinogram data, i.e. the summed attenuation along rays at different angles and offsets, which is the radon transform of the reconstructed data and is closely related to the measured transmitted X-ray energy at the detectors. Rays which pass through metal are considered corrupted, and replaced using surrounding data which has not been affected by metal. Simple FBP reconstruction is then used on the corrected sinogram. Working in the sinogram domain makes sense, since here the errors are highly localised, whereas they have global scope in the reconstructed data.

The basic technique, proposed by Kalender et al. (1987), uses an initial FBP reconstruction to determine where metal is present, and any sinogram values which are seen to pass through metal are replaced by linear interpolation from neighbouring un-corrupted values. This will be referred to as LI (Linear Interpolation); an early review of similar approaches is in Wang et al. (1996). LI is usually performed independently for each CT slice, but can also be implemented in the X-ray domain (Zhang et al., 2007), or by detecting metal directly in the sinogram rather than in an initial reconstruction (Veldkamp et al., 2010), with both variations leading to similar results.

LI is fast and very successful at removing metal artefacts, however it can remove real data, and also introduce new artefacts due to the linearly interpolated values not matching the real attenuation. Subsequent work has aimed to reduce these additional artefacts by improving the interpolation step, whilst retaining the otherwise good performance. For instance, Normalised Metal Artefact Reduction (NMAR) (Meyer et al., 2010) coarsely thresholds the reconstructed image, and removes any obvious metal, to generate a prior which is then forward-projected and used to improve the linear interpolation step in LI. Whilst this does considerably reduce the new artefacts from LI, it also generates more in some cases (Lell et al., 2012) since the original artefacts lead to inevitable mistakes in the prior. NMAR also shows loss of detail near to metal, and this is addressed in FSMAR (Frequency-Split MAR) which re-introduces information above a certain frequency threshold to NMAR from the original reconstruction (Meyer et al., 2012). It is clear from clinical tests on hip prostheses in (Morsbach et al., 2013) that this also re-introduces some of the original metal artefacts, which

limits the overall performance.

The IMAR (Intelligent MAR) algorithm implemented by Siemens[1] is developed from FSMAR and hence has similar characteristics (Axente et al., 2015). Philips[2] O-MAR (Orthopaedic-MAR) seems essentially similar to NMAR, in that it also makes use of a coarsely-thresholded prior to improve the linear interpolation (Philips CT Clinical Science, 2012). An assessment of O-MAR focusing on the effect of artefacts on radiotherapy planning (Li et al., 2012) shows results which have slightly different characteristics, but comparable performance, with good corrections in most regions but additional artefacts still clearly present. This is tested against both dual energy methods and another commercial MAR algorithm from GE in Huang et al. (2015). All these methods are reasonably straightforward extensions of LI and hence processing times, though not always reported, should all be fairly fast.

Performance can be improved by using a prior which is less affected by the original metal artefacts, and also by improving this prior from the corrected images, repeating the correction, and then iterating the whole process to convergence. In the Metal Deletion Technique (MDT) (Boas and Fleischmann, 2011, 2012) an initial LI reconstruction is processed with a very wide (radius of 100-200 pixels) edge-preserving filter which is sensitive to differences in data value as well as location[3]. This image is then forward-projected and used as a better prior to improve on the interpolation for the next iteration of sinogram correction. Total-Variation (TV) minimisation can alternatively be used in place of the edge-preserving filter (Zhang et al., 2013). The very large extent of such filters (and the coarse thresholding used in the previous methods) is necessary to reduce artefacts in the prior which are otherwise re-introduced into the data. Edge-preservation when smoothing over such a large extent is difficult, but crucial, since the iterative forward and backward projections (between the reconstructed and the sinogram data) introduce significant new radial artefacts if edges have been blurred. Hence, whilst the results of the MDT technique are very good, with a measurable improvement in clinical utility (Boas and Fleischmann, 2011), both MDT and TV minimisation take considerably longer to apply to the data.

In addition to the difficulty of avoiding new artefacts, and inability to successfully correct small details close to metal, these techniques also commonly throw away any sinogram data affected by metal (even those with priors initialised using LI as a first step when the metal artefact is significant). An exception to this is in Verburg and Seco (2012), where a combination of beam hardening prediction and sinogram interpolation was proposed. The apparent attenuation of the metal determines which process to follow: either correcting using a prediction of the beam hardening effect if the metal is less attenuating, or sinogram interpolation where it is expected that there was complete photon starvation and hence no useful residual information. However, whilst the combination is novel, neither of these two processes offer an improvement on the aforementioned techniques.

In many cases the original sinogram data is not available, but this can be forward-projected from reconstructed data (e.g. in DICOM format) if necessary. The only disadvantages are the consequent inability to change the filter used in the initial FBP reconstruction, and also the possibility that the reconstructed data has been clipped, typically to the range -1024 to $3072\,\mathrm{HU}$, and hence the measured sinogram data may not be precisely recovered.

1.2 Assessment of MAR algorithms

Assessment of MAR algorithms is complicated since artefacts are highly dependent on the particular distribution of materials in clinical scans, but it is not realistic to generate comparative *in vivo* clinical data with and

[1] Siemens Medical Solutions USA, Martinez, CA, USA

[2] Philips Healthcare, Cleveland, OH, USA

[3] This is similar to a bilateral filter, see (Buades et al., 2005) for a more thorough review of this and other de-noising filters.

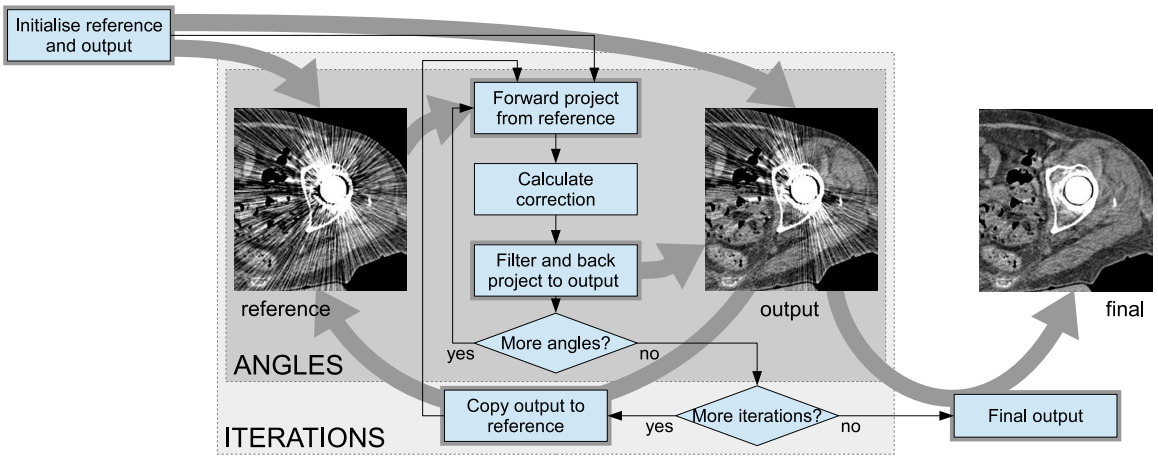


Figure 1: Overview of the metal refinement algorithm. The reference and output images are both initialised from the original reconstructed data. A sinogram is forward-projected from this, and corrections are calculated for each reconstruction angle and back-projected in turn over the output image. The whole process is iterated typically three or four times.

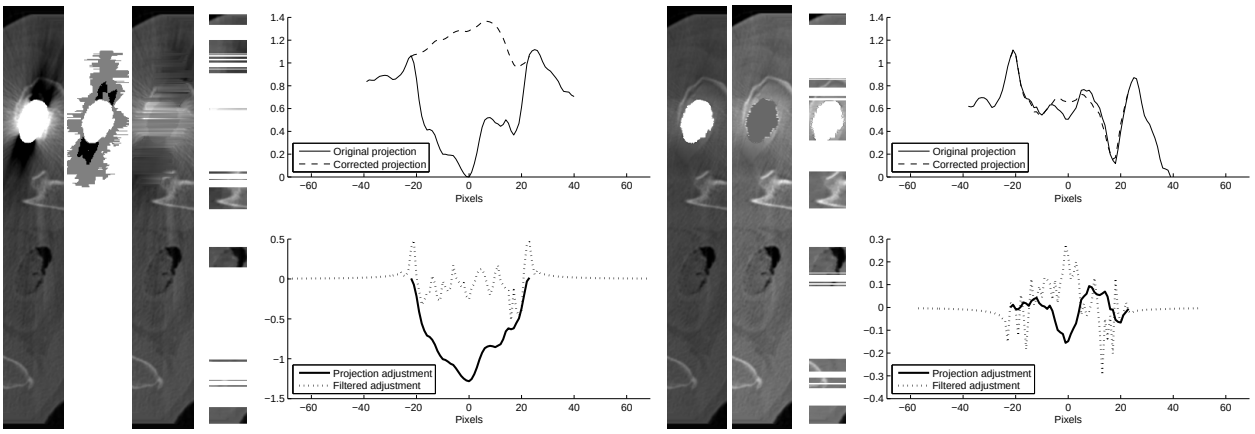
without metal present. A clinical evaluation can still be achieved by either looking for improvements in clinical reading of CT data with and without metal artefact correction, or presenting sample cases for the reader to judge for themselves (Li et al., 2012; Boas and Fleischmann, 2011; Lell et al., 2012). Whilst such tests are able to demonstrate the usefulness of MAR, they are not sufficiently sensitive to quantify the differences between the various techniques. Simulated CT data can be generated with and without metal present, however the homogeneity and artefacts in such data are not always a realistic model of clinical CT scans, and tend to favour algorithms which also presume homogeneity.

Though still not ideal, the next best approach is to use a real CT phantom which has removable metal sections. For instance Huang et al. (2015) tested O-MAR and GE dual-energy and MAR algorithms by repeat scans of several (hip, spine and dental) anthropomorphic phantoms. Positioning of the phantoms between scans was sufficiently repeatable for quantitative per-pixel assessment of MAR performance. CT phantoms were also used to assess the Siemens IMAR algorithm (Axente et al., 2015), though this was not compared to other techniques. Poly-energetic reconstruction has also been assessed against LI, NMAR and FSMAR using real phantoms (Van Slambrouck and Nuyts, 2012).

It is the dual aim of this paper to present a new metal artefact reduction technique which addresses some of the current limitations, and to quantitatively compare this with a representative range of techniques (LI, NMAR and MDT) using real scans of phantoms. The commercial IMAR and O-MAR algorithms are also evaluated by use of the same data from the relevant publications (Axente et al., 2015; Huang et al., 2015), enabling direct comparison with previously reported results. Qualitative clinical scans are also included in order to demonstrate different features of some of these algorithms in real clinical scenarios.

2 Method

The proposed method follows an overall procedure shown in Fig. 1, presuming that only the reconstructed data is available. This has much in common with other iterative sinogram interpolation methods, with forward-projection to create a sinogram, correction of the sinogram given information in the reconstructed image, and back-projection of this correction over the output. The presence of metal is determined with a simple



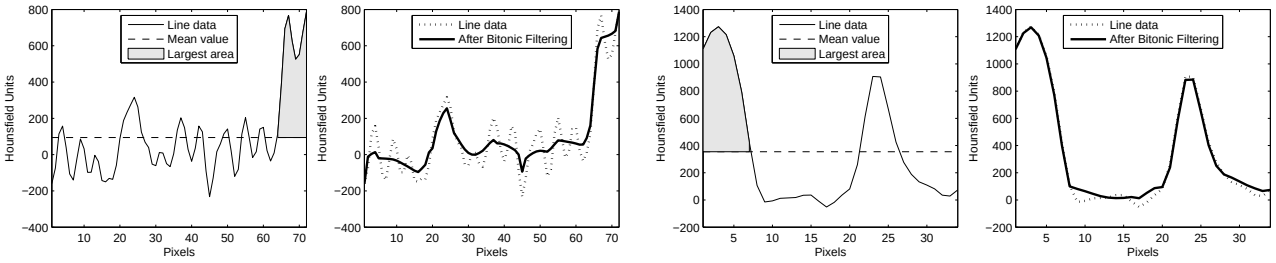
(a) (b) (c) (d) (e) Correction, first iteration (f) (g) (h) (i) Correction, second iteration

Figure 2: Example of metal correction for one reconstruction angle over the first two iterations. (a) As the first stage in forward projection, the image is re-sampled local to the metal (white region) on a grid aligned with the direction of projection (the vertical axis). (b) Metal (very light grey) is found by thresholding this image, and likely regions of incorrect background level (grey) are found by searching outwards from the metal. Low-signal areas which have been clipped in the initial reconstruction (black) are found by thresholding within this region. (c) A new background level is estimated for the grey or black regions by linear interpolation along the horizontal axis. (d) Only horizontal lines which contain strong transitions contribute to the correction of the sinogram. These are smoothed using a Bitonic filter in the horizontal direction to remove noise but preserve features (see Fig. 3). (e) This partial smoothed data (dashed line, upper graph) is summed vertically and fitted to a forward projection from the original data (with the metal removed, solid line), by adding a suitable linear trend. The difference between these two projections is filtered and back-projected over the output image (lower graph). (f) to (i) The following iterations are similar, except that no data is re-estimated as in (c), and slightly more data is allowed to contribute to the correct projection each time. Convergence is fast since corrections are of a significantly lower magnitude for each new iteration.

threshold of the initial reconstructed data (2800 HU was used for all the results in this paper) since it is very rare for artefacts to generate data values high enough to be confused with metal. Only the sinogram data very close to metal is changed, and hence these are the only regions which need to be forward-projected. Forward-projection is more time consuming than back-projection, since it involves sampling the data at sufficient spatial and angular resolution to preserve all the original frequency content, so this is a useful cost saving[4].

Since the output image will in most cases be very similar to the input, it is also more efficient to back-project only the *difference* between the corrected and uncorrected sinogram data, which is then added to the original reconstruction to form the output image. This difference is constrained to the area fairly close to metal, though the effect of the filter in FBP means the extent of back-projection will be larger than that of the corrections to the sinogram. Back-projection is only required where the filtered difference is non-zero, and taken together with the partial forward-projection, a single iteration of the forward and back-projection process in Fig. 1 can be completed in less than a second for typical clinical data and on a typical Intel Core i7 processor, though this will grow if there is considerable metal content. Once corrected, the output reconstruction is then used as a new input and the process iterated until convergence, which is typically only three or four iterations, in common with that cited for MDT (Boas and Fleischmann, 2011).

[4]For a 512×512 image, single-threaded on an Intel Core i7, full forward-projection (over all reconstruction angles) would take typically 5 secs, compared to less than 0.5 secs for back-projection.



(a) Strong transitions with metal artefacts (b) Strong transitions with few artefacts

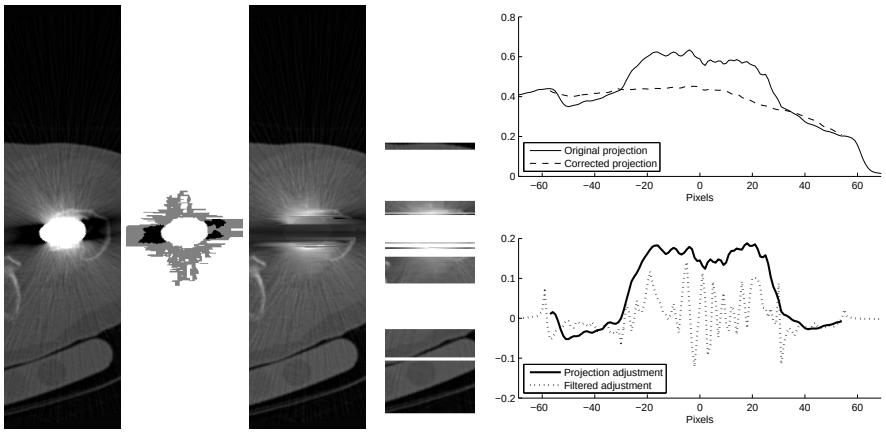
Figure 3: Detection and smoothing of strong transitions. (a) Shows the process for one horizontal line, such as those included in Fig. 2(d), containing a transition and also metal artefacts, (b) shows the same for a line with few artefacts. Strong transitions are detected by looking for the largest area above or below the mean value (left-hand graphs). If this is below a threshold, the line is not included in the forward projection, and hence is implicitly replaced with a linear trend. Otherwise, artefacts are reduced using the Bitonic filter (right-hand graphs: the same size filter is used in both cases) which is good at preserving real transitions.

Where this process differs from previous techniques is in the correction applied to the sinogram in the centre box of Fig. 1 and outlined in Fig. 2. Rather than basing this correction on a prior from a complete reconstruction, it is instead calculated individually for each reconstruction angle from the sampled data used to forward-project at that angle, i.e. Fig. 2(a). In order to minimise new artefacts, any strong horizontal transitions in the data at the same vertical location where metal is present must contribute to the correction of the sinogram data. However, any existing artefacts at these locations must also be removed. Previous methods (NMAR, MDT, etc.) achieve this by drastic filtering of the prior but with highly edge-preserving techniques, or ignore the step entirely (LI) and hence *do* introduce new artefacts.

Use of the forward-projected data for each reconstruction angle means that only strong *horizontal* transitions need to be included in the prior for that angle, and filtering is also only necessary in this direction, as in Fig. 2(d) and Fig. 3. Vertical transitions, and any vertical filtering which preserves the mean value, have no effect since the data is summed in this direction to create the new sinogram values at this angle. This leads to significant performance improvements compared to the wide-ranging two-dimensional filters required for MDT and TV minimisation.

A key difficulty of this approach is the need to distinguish between strong transitions due to genuine changes in material attenuation, and those which are actually due to the very artefacts which we are trying to correct. Disambiguation is not possible with a forward prediction of beam hardening from the presumed location and type of the metal, since as already discussed this is only one source of the artefacts and hence does not match them sufficiently well. Instead, data close to metal is inspected to determine if it is likely to have been affected by either beam hardening or photon starvation, both of which would change the local mean attenuation, as in Fig 2(b). Fortunately, most real attenuation data is piecewise-homogeneous (i.e. independent materials which each have relatively constant linear attenuation values) whereas the aforementioned artefacts generate effects which gradually lessen with distance from metal. Hence, starting at the metal location, artefacts are presumed to exist if the attenuation values are gradually approaching a notional background level (normally set to 0 HU) and marked until the values are constant or start to move away from this level: see the grey region in Fig 2(b).

Strong transitions are only found outside of this marked region, and will usually be at other vertical locations. In the rare instance where there might be a strong transition alongside such a region, so that we need to include data in the forward projection from this marked area, it is replaced by a linear interpolation from the nearest unmarked values, as in Fig. 2(c). In most cases, this marking only serves to indicate which



(a) (b) (c) (d) (e) Correction, iteration 1

Figure 4: Metal correction with clipping in the initial reconstruction. The sub-figures have the same meaning as in Fig. 2, but here there are areas of signal loss (black in (b)) either side of the metal. Projections through metal and through regions of signal loss both require correction: however there is no need to smooth projections which do not pass through metal, since the corruption is caused only by clipping of the reconstructed data value, not due to metal artefacts. (d) shows the subsequent smoothing of only the centre columns. Such correction is only necessary in the (usual) case that the original sinogram data is not available.

regions of the original data are corrupted and should be ignored completely. In other locations, a horizontal transition is considered to be strong enough if there are regions of consistent difference from the mean value for that line, i.e. the summed difference from the mean between zero crossings is sufficiently large (see Fig. 3). The threshold t (in HU $\times$ image pixels) for this difference goes down with each iteration:

$$t = \min\left\{200\frac{\sqrt{p_m}}{i+2}, 2000\right\} \tag{1}$$

where p_m is the number of image pixels which are marked as metal, and i is the iteration, starting from zero.

Having decided which horizontal lines contain transitions and must be included in the forward projection, any noise due to scatter, low signal count or beam width effects needs to be removed. Such noise can have a very large magnitude but is usually zero-mean, consisting of alternating light and dark radial streaks. The lines are hence smoothed with a newly developed Bitonic filter (Treece, 2015), which is able to remove such noise whilst preserving other transitions, even if the noise is of higher magnitude than the transition. Figure 3 shows examples of the effect of this filter on transitions with and without significant metal artefacts. The extent of smoothing is controlled by the width of this filter, the default value of 13 image pixels being appropriate for the vast majority of situations, though this can be reduced if there is very little metal in the scans. Much of the performance of the metal reduction technique is the result of the effectiveness of this new filter.

The smoothed lines are summed in the vertical direction and a linear trend is added, in order to ensure that the replacement data matches the projected data just outside the metal region. Data either side of the metal region, at least two pixels away from the metal, is used to determine the required linear trend. In contrast to the reconstructed data, the forward projected sinogram data away from metal is completely free from metal artefacts and hence only a single sample is required at each side. Figure 2(e) shows this corrected data, and also the difference from the original data (ignoring the contribution of any pixels actually marked as metal). It is this difference which is back-projected over the output image for each reconstruction angle.

One further step is required to account for possible clipping of the initial reconstructed values to some minimum, for instance -1024 HU. Any forward projections through such data will not then represent the

actual measured value in the original sinogram. Such clipped regions, an example shown in black in Fig. 4(b), therefore do not give reliable projected values and must also be replaced; however projections through this data will not necessarily contain metal artefacts, in which case they will not need filtering. Figure 4(e) shows a typical correction, demonstrating that the replacement values now cover both the metal and clipped regions and the linear trend is calculated from forward projected values beyond both these regions.

3 Results

3.1 Phantom scans

In addition to the new method described in Section 2, referred to as RMAR or 'Refined' MAR, LI and NMAR were also implemented and tested. MDT was included via software[5] made available by the authors of this technique (Boas and Fleischmann, 2011), using default settings. The commercial IMAR and O-MAR algorithms do not have freely available implementations, however phantom data was acquired to which these techniques had already been applied. The 'rods' phantom (for which IMAR reconstructions were also available) was scanned with a Siemens SOMATOM Definition AS+ in Biograph mCT configuration, with scans provided by Axente et al. (2015), and the 'hip', 'dental' and 'spine' anthropomorphic phantoms (for which O-MAR reconstructions were also available) were scanned with a Philips Brilliance scanner at 120 kVp, with scans provided by Huang et al. (2015). Details of their phantoms are as follows:

spine This was designed by the MDADL[6] and contains lung, heart and an additional target structure. The usual spine insert was replaced with a high impact polystyrene insert including two titanium rods (9.5 mm diameter), mimicking spinal fixation rods. These were replaced with high impact polystyrene rods for the scan without metal.

hip This was also designed by the MDADL and contained prostate, bladder, and rectum structures in a central water-filled imaging insert, and femoral head structures in a similar water-filled outer insert. The phantom was modified to hold a cobalt-chromium hip prosthesis (6.84 g/cm^3).

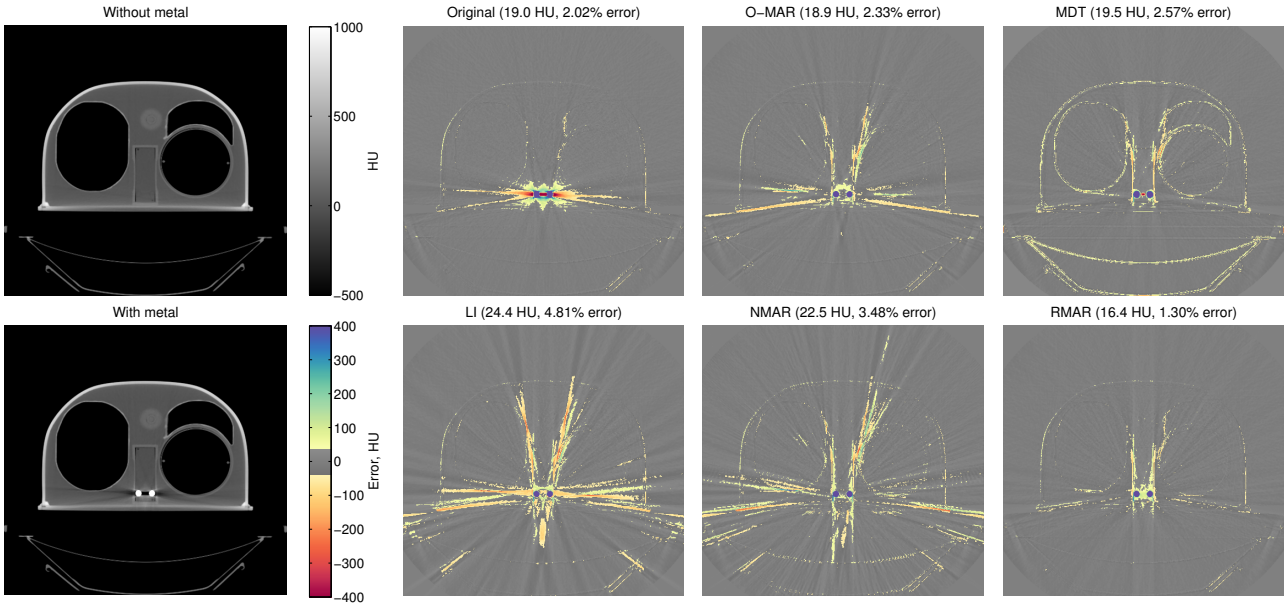
dental This was a CIRS[7] Model 606 head phantom with articulating lower jaw, tongue, teeth, and air cavities, and two removable teeth which could be interchanged with teeth containing dental restoration materials. In addition, a metal crown was taped on top of one of the original non-metal teeth.

rods This was a CIRS Model 062MA standard electron density phantom, into which were inserted three different configurations of stainless steel rods of diameters 10, 15, 20 and 25 mm.

All phantoms were scanned with and without the metal inserts with minimal movement between scans. Any residual movement was compensated by rigid registration between the data to sub-voxel accuracy. Unfortunately the non-metal scan for the rods phantom was not available, so this was substituted for a scan with only a single small 10 mm metal rod, which was easily corrected by IMAR. There were some residual artefacts, as can be seen in the top image of Fig. 8(a), but these were of much lesser extent than the metal artefacts under investigation, and hence this could still act as a useful reference.

Analysis and presentation of this data follows a similar style to Huang et al. (2015). After alignment, data from the non-metal scans was subtracted from the metal scans and from any subsequent data sets with metal artefact correction applied, to give a set of difference data. A small 3×3 2D median filter was applied to each

[5] http://www.revisionrads.com/
[6] MD Anderson Dosimetry Laboratory, University of Texas, Houston, TX, USA
[7] Computerized Imaging Reference Systems, Inc, Norfolk, VA, USA



(a) CT Images (b) Difference between metal and non-metal data

Figure 5: Performance for one frame of the spine phantom data. (a) CT data with and without the metal insert. (b) 'Original' (top-left) shows the difference between the two CT images in (a). The other images show the difference between the metal scan with correction applied, and the scan without metal. Differences less than 40 HU are shown in grey, larger differences are coloured as shown in the colour key on the left. The average error in HU and percentage of pixels regarded as containing metal artefacts are recorded in the title for each image.

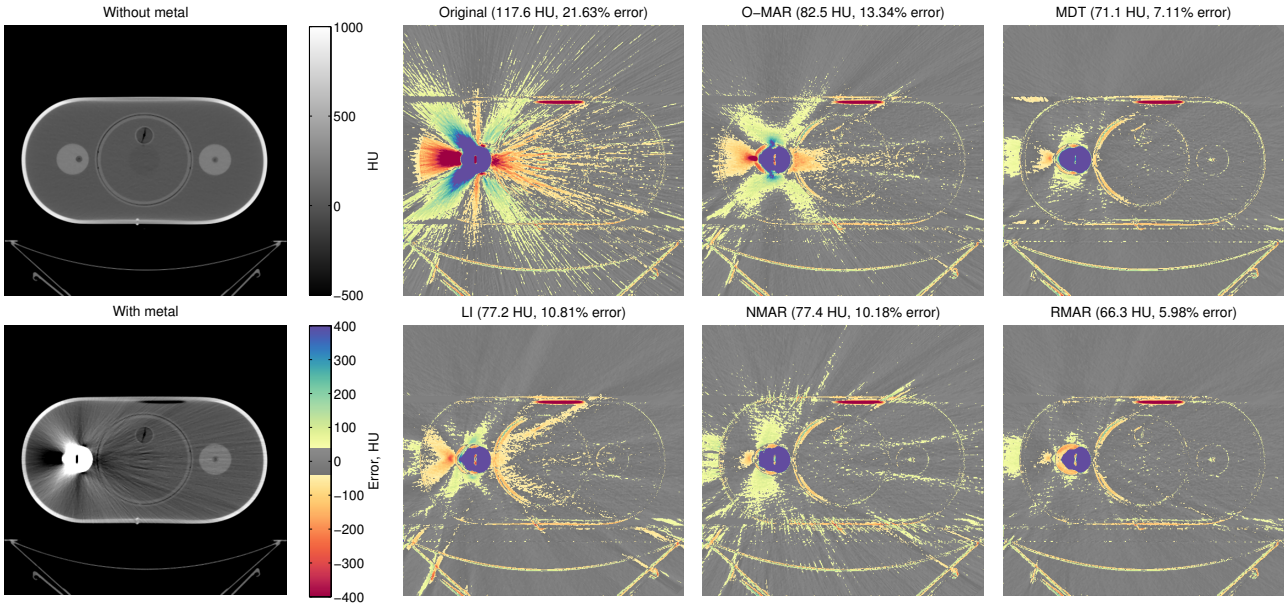
slice to remove random CT noise which was judged to be the effect of repeat scanning rather than the presence of metal, and any residual absolute differences above 40 HU were regarded to be due to metal artefacts. Data differences where the original values were either above 2700 HU or both metal and non-metal scans were below −900 HU were also ignored. In the former case this excluded the actual location of the metal insert, and in the latter ignored small changes to the data in regions where there was no material present, so long as there was still no apparent material present after correction.

Summary data for these four phantoms is in Table 1, aggregated over every CT slice containing metal. Both the average HU difference and percentage of pixels affected by metal artefacts are tabulated, recorded as absolute values and also in decibels (dB) relative to the scan with metal but without any attempt to correct the metal artefacts.

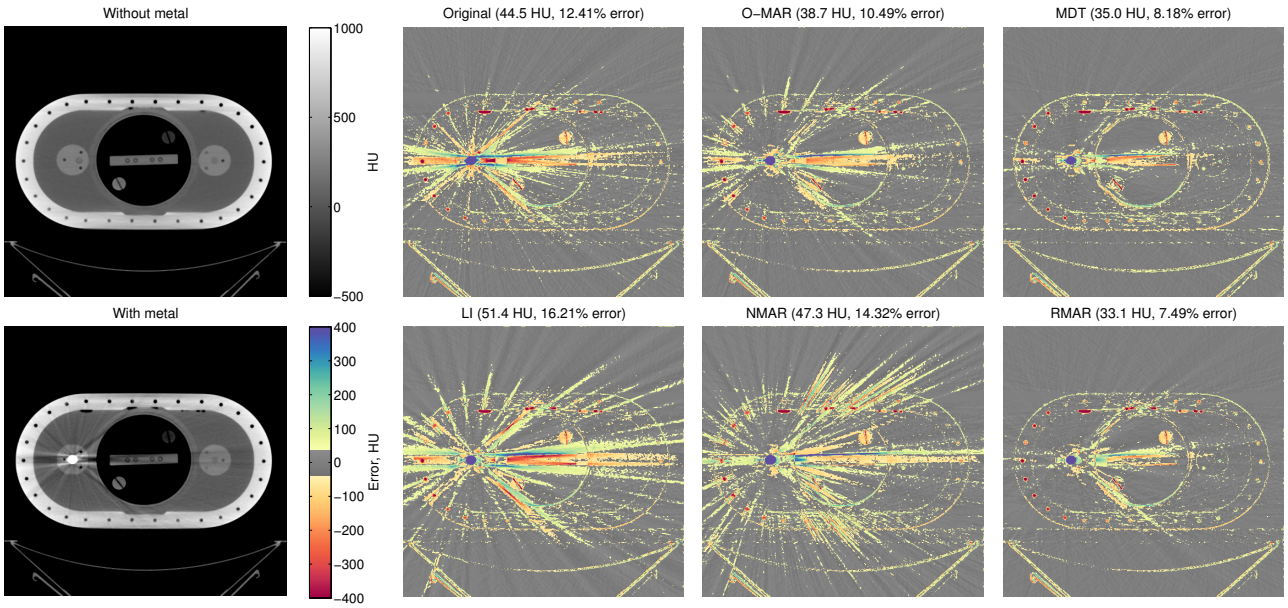
Representative frames from each of the phantom data sets are shown in Figs. 5, 6, 7 and 8. There are some residual discrepancies between the metal and non-metal images which are not metal artefacts, for instance the horizontal line coloured red at the top of the phantom in Fig. 6, where air is only present in the metal scans, and slight mis-registrations at the edges of all phantoms. However, these will have equally affected the results for all types of metal artefact correction. A slightly lower level of smoothing (Bitonic filter width of 9 rather than 13) was used for RMAR in the spine data since there was substantially less metal. The dental data was substantially different from the other phantoms, since the teeth are particularly dense and immediately surrounded by air, hence in this case the metal background correction was not applied (i.e. the process started directly at the second iteration, see Fig. 2): this choice is discussed in the next section. In all other scans the default parameters were used as described in Section 2.

Table 1: Aggregate results for all data sets compared to ground truth (no metal present) data, from per-frame measurements of the average HU difference, and percentage of pixels affected by metal artefacts. Absolute values are recorded as well as the relative difference from the uncorrected data, in dB ($20 \log_{10} \frac{\text{corrected}}{\text{none}}$), with negative values hence showing improvements. †Significantly worse than RMAR at $p < 0.05$. ‡Significantly worse than RMAR at $p < 0.001$. $^{\diamond}$Significantly better than RMAR at $p < 0.001$. In each case the number of frames of data n is shown in addition to the number of frames for which each method performed the best.

Data	**Reduction technique**	**Absolute error**				**Relative to original, dB**	
		Average (HU)		**Pixels** > 40 **(%)**		**Average (HU)**	**Pixels** > 40 **(%)**
		mean $\pm$ **std**	**n best**	**mean** $\pm$ **std**	**n best**	**mean** $\pm$ **std**	**mean** $\pm$ **std**
Spine	None	$21.6 \pm 5.3^{\ddagger}$	0	$3.49 \pm 2.94^{\ddagger}$	2	-	-
Phantom	LI	$27.0 \pm 4.7^{\ddagger}$	0	$6.58 \pm 3.19^{\ddagger}$	0	$2.03 \pm 0.63^{\ddagger}$	$6.56 \pm 2.12^{\ddagger}$
($n = 84$)	O-MAR	$22.2 \pm 5.6^{\ddagger}$	0	$4.05 \pm 3.19^{\ddagger}$	0	$0.20 \pm 0.44^{\ddagger}$	$1.38 \pm 1.14^{\ddagger}$
	NMAR	$26.6 \pm 7.1^{\ddagger}$	0	$5.64 \pm 3.76^{\ddagger}$	0	$1.78 \pm 0.50^{\ddagger}$	$4.71 \pm 1.33^{\ddagger}$
	MDT	$21.9 \pm 5.0^{\ddagger}$	0	$4.20 \pm 2.97^{\ddagger}$	0	$0.15 \pm 0.24^{\ddagger}$	$2.06 \pm 0.88^{\ddagger}$
	RMAR	19.2 ± 5.3	84	2.82 ± 2.91	82	-1.05 ± 0.29	-2.57 ± 1.15
Hip	None	$102.3 \pm 82.5^{\ddagger}$	0	$16.17 \pm 7.31^{\ddagger}$	6	-	-
Phantom	LI	$82.5 \pm 81.5^{\ddagger}$	6	$10.92 \pm 7.90^{\ddagger}$	0	$-2.24 \pm 2.13^{\ddagger}$	$-4.05 \pm 3.84^{\ddagger}$
($n = 63$)	O-MAR	$84.7 \pm 83.5^{\ddagger}$	1	$11.36 \pm 7.02^{\ddagger}$	0	$-2.06 \pm 1.22^{\ddagger}$	$-3.37 \pm 2.04^{\ddagger}$
	NMAR	$83.2 \pm 81.5^{\ddagger}$	0	$10.86 \pm 7.81^{\ddagger}$	0	$-2.14 \pm 2.02^{\ddagger}$	$-4.00 \pm 3.73^{\ddagger}$
	MDT	$79.6 \pm 88.8^{\ddagger}$	5	$8.75 \pm 7.17^{\ddagger}$	8	$-2.91 \pm 1.92^{\ddagger}$	$-5.94 \pm 4.12^{\ddagger}$
	RMAR	76.9 ± 83.9	51	8.25 ± 7.20	49	-3.18 ± 1.82	-6.63 ± 4.13
Dental	None	63.0 ± 33.8	0	5.23 ± 2.20	0	-	-
Phantom	LI	$85.0 \pm 17.7^{\dagger}$	0	$8.43 \pm 1.28^{\dagger}$	0	$3.34 \pm 2.92^{\dagger}$	$4.62 \pm 2.42^{\dagger}$
($n = 4$)	O-MAR	55.8 ± 14.0	1	4.81 ± 0.95	0	-0.37 ± 2.35	-0.31 ± 1.99
	NMAR	$74.2 \pm 17.2^{\dagger}$	0	$6.96 \pm 1.70^{\dagger}$	0	2.13 ± 2.71	$2.84 \pm 1.59^{\dagger}$
	MDT	57.2 ± 15.5	0	4.74 ± 0.53	1	-0.19 ± 2.19	-0.34 ± 2.60
	RMAR	55.7 ± 14.0	3	4.73 ± 1.94	3	-0.74 ± 0.94	-0.85 ± 0.39
Rod	None	$114.2 \pm 19.5^{\ddagger}$	0	$21.95 \pm 3.07^{\ddagger}$	0	-	-
Phantom	LI	$77.2 \pm 4.27^{\ddagger}$	0	$17.46 \pm 1.57^{\ddagger}$	0	$-3.30 \pm 1.01^{\ddagger}$	$-1.94 \pm 0.60^{\ddagger}$
($n = 48$)	IMAR	$63.5 \pm 2.2^{\dagger}$	2	$11.18 \pm 1.09^{\dagger}$	2	$-4.98 \pm 1.32^{\dagger}$	$-5.82 \pm 0.69^{\ddagger}$
	NMAR	$66.9 \pm 3.4^{\ddagger}$	0	$12.14 \pm 1.62^{\ddagger}$	0	$-4.54 \pm 1.10^{\ddagger}$	$-5.14 \pm 0.73^{\ddagger}$
	MDT	$58.4 \pm 3.1^{\diamond}$	46	$9.12 \pm 1.56^{\diamond}$	46	$-5.71 \pm 1.26^{\diamond}$	$-7.68 \pm 0.72^{\diamond}$
	RMAR	62.4 ± 4.0	0	10.84 ± 1.44	0	-5.15 ± 1.10	-6.12 ± 0.67

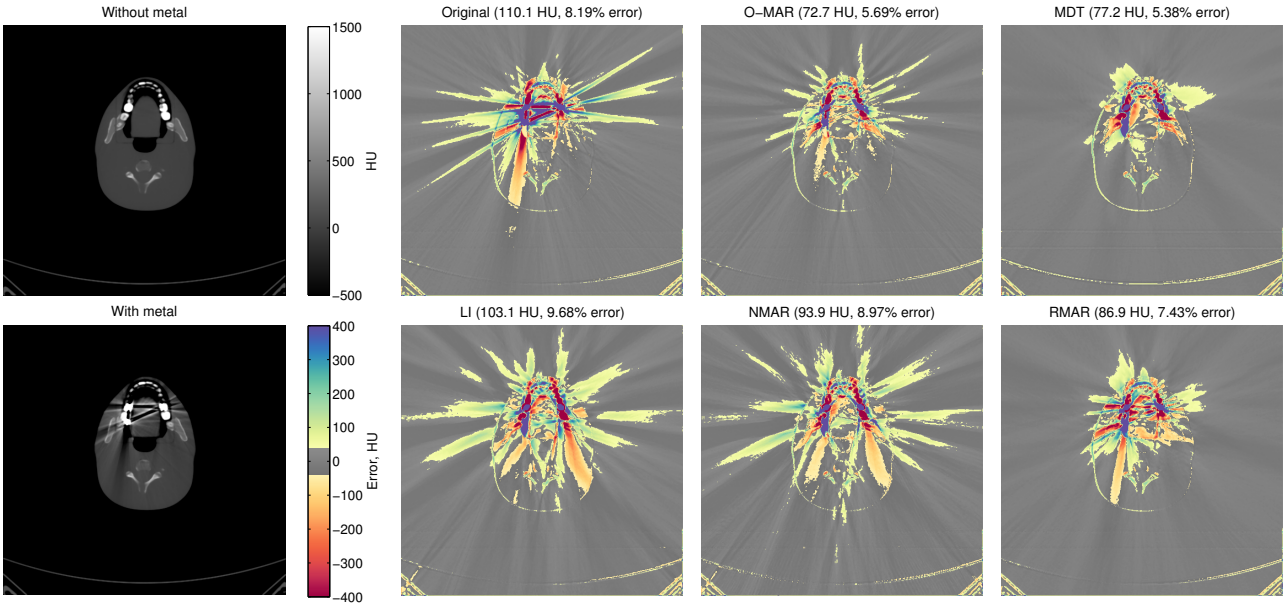


(a) CT Images (b) Difference between metal and non-metal data

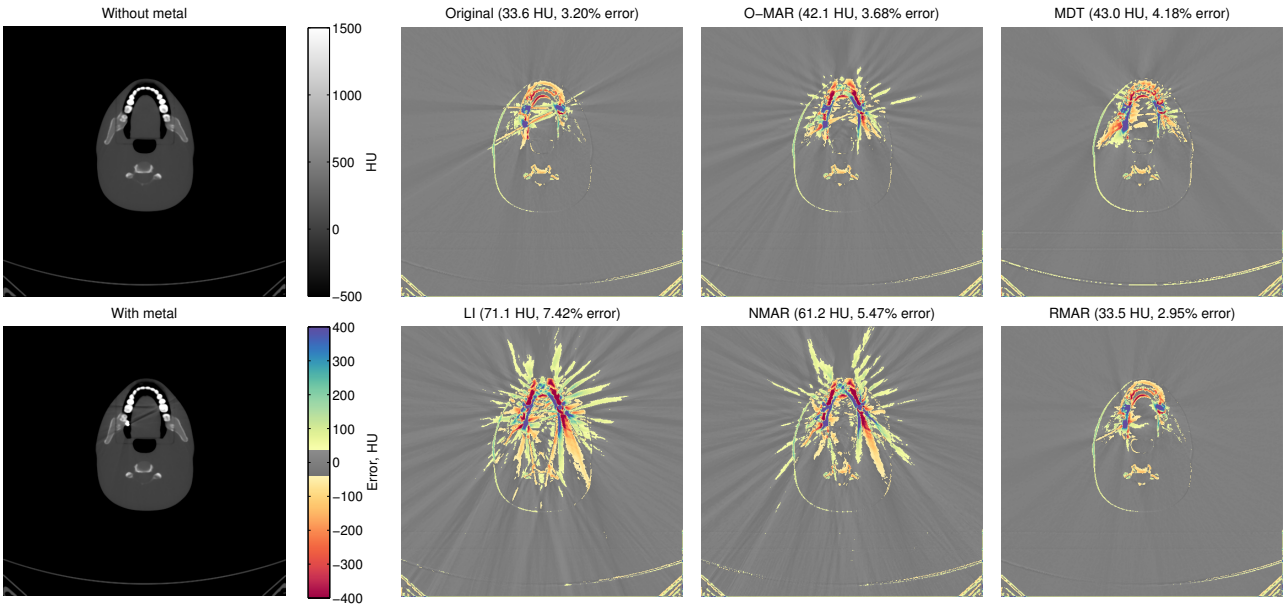


(c) CT Images (d) Difference between metal and non-metal data

Figure 6: Performance for two frames of the hip phantom data. The presentation is as in Fig. 5.

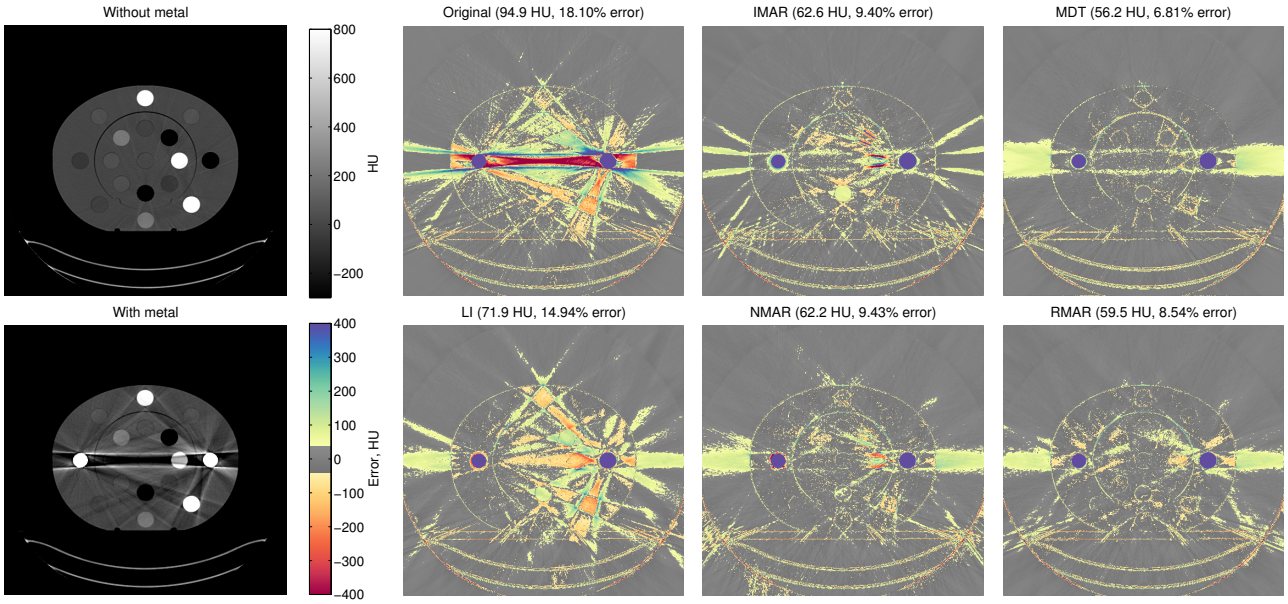


(a) CT Images (b) Difference between metal and non-metal data



(c) CT Images (d) Difference between metal and non-metal data

Figure 7: Performance for two frames of the dental phantom data. The presentation is as in Fig. 5.



(a) CT Images (b) Difference between metal and non-metal data

Figure 8: Performance for one frame of the rod phantom. The presentation is as in Fig. 5, except that in this case (b) contains IMAR rather than O-MAR results (top-centre).

3.2 Clinical data

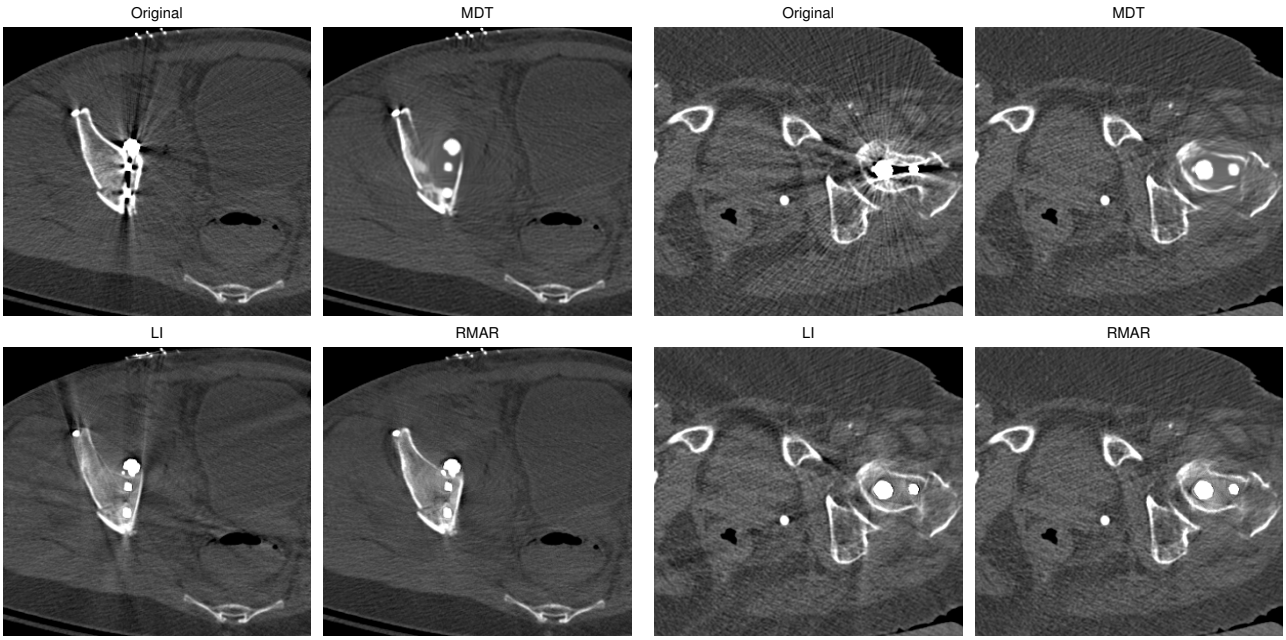
Post-operative hip CT scans were collected from a variety of patients, each of whom underwent a hip repair or replacement procedure. The CT scans were part of a retrospective study which received approval by the local institutional Research Ethics Committee (LREC 99/076 and LREC 04/0108). All patients provided written, informed, consent. Figures 9(b), 10, 11 and 12 show examples of the performance of LI, MDT and RMAR on a selection of CT frames which are representative of the test data set of 16 patients, each scan containing typically 100-200 frames affected by metal artefacts. Fig. 9(a) shows similar results from an online CT data set[8] available for research and teaching use.

LI demonstrates what can be achieved very simply, whereas MDT and RMAR are overall the best performing techniques on the phantom data (and for which implementations were available such that the clinical data could be processed). The MDT and RMAR techniques were used with default parameter settings except for Fig. 9(a), where RMAR used slightly lower smoothing, exactly as for the spine phantom data.

3.3 Processing time

The LI, NMAR, MDT and RMAR algorithms were all run without hardware acceleration, single-threaded on similar Intel Core i7 processors. For a typical single-frame 512×512 image with some metal present, LI took less than one second, NMAR between 1 and 2 seconds, RMAR between 3 and 8 seconds, and MDT between 15 and 25 minutes. This means a typical clinical data set with 150 frames of data including metal might be processed in about 15 minutes with the new RMAR technique, whilst it would take more than two days using MDT. All algorithms are easily scalable and could be equally sped up by use of multiple cores. Times for IMAR and O-MAR are unknown, since these were neither implemented nor run locally, but there is little reason to suspect they would be substantially different to those for NMAR or RMAR.

[8] 'PELVIX' data from http://www.osirix-viewer.com/datasets/



(a) Fixation after pelvic fracture (b) Two hip screws

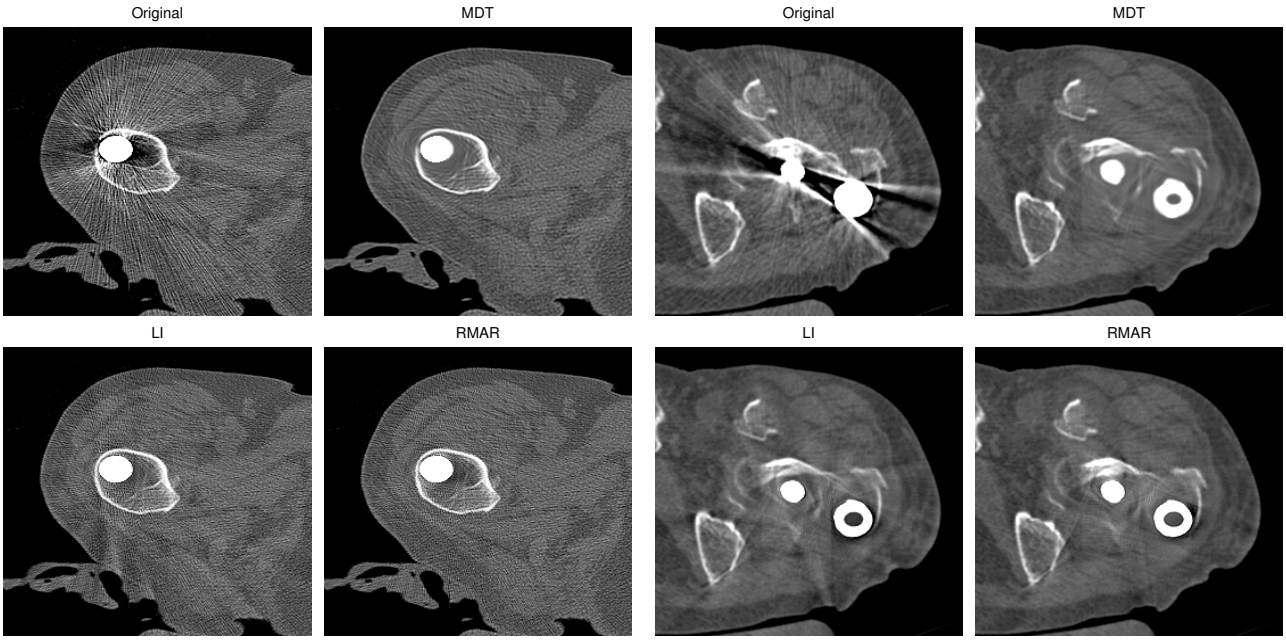
Figure 9: Performance on clinical data. Both data sets show prominent artefacts using LI which are of similar magnitude to the original metal artefacts. Both MDT and RMAR reduce these artefacts, but RMAR preserves more detail in the bone close to metal, whereas MDT replaces much of the bone with HU values more reminiscent of soft tissue.

4 Discussion

It is clear from the phantom results in Table 1 that the new RMAR technique performs well on all this data, significantly better than all other techniques for the spine and hip phantoms, and in fact the only technique which actually reduces the overall artefacts for the spine phantom. It also performs slightly better (though not significantly so) for the dental phantom, and is only outperformed by MDT on the rod phantom. However it is also apparent that the MDT technique performs generally very well. In contrast, the results from LI and NMAR are universally less good (though NMAR is clearly an improvement on LI). O-MAR performs well on the spine and dental phantom, but considerably less so on the more significant artefacts present in the hip phantom, and IMAR generates good results on the rod phantom, though it should be noted that NMAR also performs well in this case, in contrast to its performance on the other phantoms, so it may be a less representative test.

The features which contribute to this aggregate performance are more apparent from the sample results in Figs. 5 to 8. For the spine phantom, the metal artefacts are relatively minor, and the remainder of the phantom is relatively complex, hence the performance rests just as much on not introducing new artefacts as on removing the existing ones. The results for O-MAR, NMAR, MDT and RMAR in Fig. 5 are quite similar very close to the metal inserts, but RMAR introduces far fewer additional artefacts in the far field. This is in contrast to the hip phantom, Fig. 6(b) where, except for NMAR, the various techniques do not introduce many new artefacts, but only MDT and RMAR succeed in removing most of the existing ones.

Fig. 6(d) shows both effects: here there are considerable artefacts against a very complex background and all techniques other than RMAR, and MDT to a lesser extent, either fail to correct or add new artefacts in the region of the small holes in the outer ring. These regions are not homogeneous, and the background is not



(a) Dynamic hip screw (b) Intramedullary hip screw

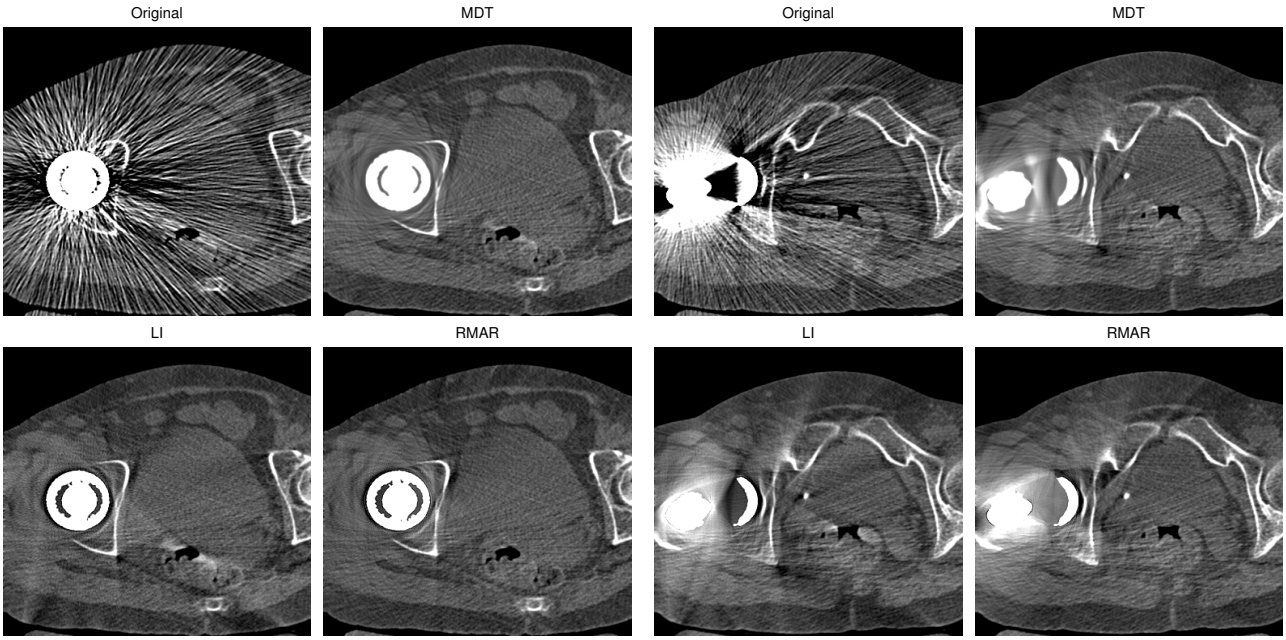
Figure 10: Performance on clinical data. (a) Higher resolution data is better preserved by RMAR which updates the original image, whereas MDT creates a new one with lower resolution. (b) Multiple large metal objects are a challenge and both RMAR and MDT create small artefacts between the objects, though less so than LI. The specific homogeneity imposed by MDT generates a better result overall in this case.

representative of soft tissue: in this case presumed homogeneity and the use of thresholds potentially enhances artefacts rather than suppressing them; an effect which has been noted before for the NMAR algorithm (Zhang et al., 2013). This hypothesis is supported by the rather better results of NMAR, O-MAR and MDT on the rod phantom in Fig. 8, where the background (other than the main rods) *is* homogeneous and at about 0 HU, and hence the assumptions are more applicable. RMAR performs well in both cases, since it is less sensitive to either of these assumptions, instead making use of the threshold-less Bitonic filter to determine whether data in the replaced forward projection should be included or not.

Correction of artefacts in the dental phantom is much more challenging (Huang et al., 2015), since it contains a mix of very low and very high densities close to metal artefacts with a very similar appearance. However, good correction can still be achieved by RMAR by turning off the initial background metal correction as previously described. Whilst this lowers the overall potential for correction, it also reduces the likelihood that any new artefacts will be generated. Visually pleasing results are still possible with the default settings, but these also increase the signal level in some of the regions of air close to the teeth.

Turning to the clinical data, three particular features of the new technique are apparent. Firstly, the RMAR results very close to metal, and particularly within surrounding bone, are generally better than with MDT. In particular, MDT (again because of the reliance on threshold-based filtering) tends to replace some nearby trabecular bone with values more consistent with soft tissue, as in Figs. 9(a) and (b), Fig. 10(a) and Fig. 12(b). Secondly, in cases where there are nearby areas of low signal, for instance rectal gas as in Fig. 11, the new technique is better at reducing bordering bright signal artefacts. Thirdly, both MDT and RMAR introduce far fewer artefacts than LI (and other techniques), and those they do introduce are of similar magnitude but in different locations, as can also be seen in Fig. 11.

RMAR is not as fast as LI and NMAR, but at only a few seconds per frame is still easily responsive



(a) Bipolar hemi-arthroplasty, frame 1 (b) Bipolar hemi-arthroplasty, frame 2

Figure 11: Performance on clinical data. (a) Even severe artefacts can successfully be removed. In this case RMAR is better at not introducing bright artefacts neighbouring the low-signal rectal gas (lower-middle). (b) RMAR has a better response to complete signal loss between the two metal parts, though also introduces a slight artefact at the top right which is not seen in the MDT data.

enough to enable interactive use on typical CT data. This is about two hundred times faster than MDT, which can only be performed offline, and requires a very considerable delay if an entire data set is processed.

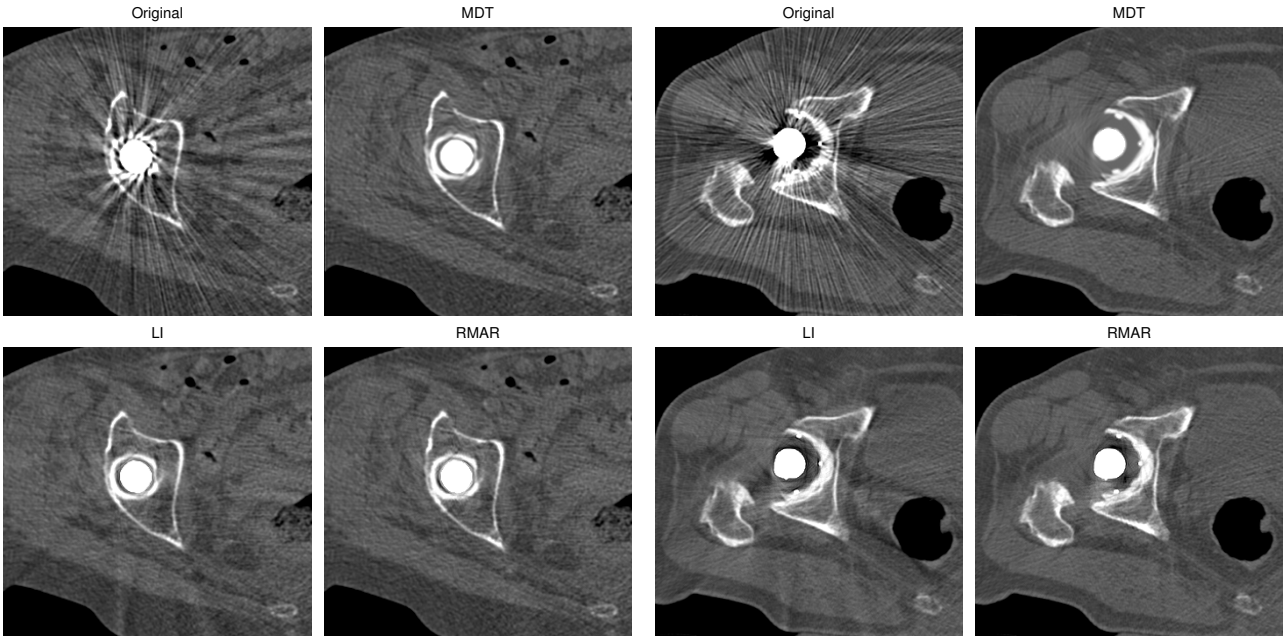
5 Conclusion

A new technique has been presented which corrects metal artefacts in CT scans by replacing corrupted values in a sinogram forward-projected from initial reconstructed data. This technique has been compared to a broad range of published and commercial alternatives, both quantitatively on various phantom data sets, and qualitatively on a range of clinical data sets of the hip. The performance is significantly better than alternatives on most phantom data sets, and in many clinical cases there is also a visual improvement over MDT.

Both MDT and the new technique virtually eliminate metal artefacts whilst introducing very few new artefacts: the new technique provides better detail close to the metal, particularly in bony regions, whereas MDT generates slightly fewer artefacts in the far field. However the new technique is over two hundred times faster than MDT and can hence be used interactively on single frames, or on entire clinical scans in typically 10-20 minutes.

Software which implements this technique on DICOM data is freely available on the internet[9].

[9] wxDicom, http://mi.eng.cam.ac.uk/Main/GMT_wxDicom



(a) Bipolar hemi-arthroplasty (b) Cemented total hip arthroplasty

Figure 12: Performance on clinical data. (a) Shows considerable windmill artefact (due to the beam width in the out-of-plane direction) which can also be corrected by both MDT and RMAR with no discernible errors. (b) A structurally complex cemented implant with considerable signal enhancement between it and the large area of rectal gas. Both MDT and RMAR correct this, though the latter leaves slightly better bone definition over the pelvic region.

Acknowledgements

The author is very grateful to Stephen Kry from the MD Anderson Cancer Centre at the University of Texas and Jessie Huang-Vredevoogd from the University of Wisconsin Department of Human Oncology for the CT data in Huang et al. (2015). Equally to Dimitre Hristov and Pooja Pradhan from the Imaging Systems for Radiotherapeutics Group at Stanford School of Medicine for the CT data in Axente et al. (2015). The author would like to thank F. Edward Boas and Dominik Fleischmann from the Department of Radiology, Stanford University Medical Center, for providing the MDT software. The author would also like to thank Tristan Whitmarsh and Ken Poole from the University of Cambridge for access to the clinical data, and acknowledge all the other investigators and study participants involved in the clinical trials.

References

Axente, M., Paidi, A., Von Eyben, R., Zeng, C., Bani-Hashemi, A., Krauss, A., Hristov, D., 2015. Clinical evaluation of the iterative metal artifact reduction algorithm for CT simulation in radiotherapy. Medical Physics 42 (3), 1170–1183.

Boas, F. E., Fleischmann, D., 2011. Evaluation of two iterative techniques for reducing metal artifacts in computed tomography. Radiology 259 (3), 894–902.

Boas, F. E., Fleischmann, D., Apr. 2012. CT artifacts: causes and reduction techniques. Imaging in Medicine 4 (2), 229–240.

Buades, A., Coll, B., Morel, J.-M., 2005. A review of image denoising algorithms, with a new one. Multiscale Modeling & Simulation 4 (2), 490–530.

Guggenberger, R., Winklhofer, S., Osterhoff, G., Wanner, G., Fortunati, M., Andreisek, G., Alkadhi, H., Stolzmann, P., 2012. Metallic artefact reduction with monoenergetic dual-energy CT: systematic ex vivo evaluation of posterior spinal fusion implants from various vendors and different spine levels. European Radiology 22 (11), 2357–2364.

Hsieh, J., Molthen, R. C., Dawson, C. A., Johnson, R. H., 2000. An iterative approach to the beam hardening correction in cone beam CT. Medical Physics 27 (1), 23–29.

Huang, J. Y., Kerns, J. R., Nute, J. L., Liu, X., Balter, P. A., Stingo, F. C., Followill, D. S., Mirkovic, D., Howell, R. M., Kry, S. F., 2015. An evaluation of three commercially available metal artifact reduction methods for CT imaging. Physics in Medicine and Biology 60 (3), 1047.

Jeon, H., Park, D., Youn, H., Nam, J., Lee, J., Kim, W., Ki, Y., Kim, Y. H., Lee, J. H., Kim, D., et al., 2015. Generation of hybrid sinograms for the recovery of kV-CT images with metal artifacts for helical tomotherapy. Medical Physics 42 (8), 4654–4667.

Kachelrieß, M., Sourbelle, K., Kalender, W. A., 2006. Empirical cupping correction: a first-order raw data precorrection for cone-beam computed tomography. Medical Physics 33 (5), 1269–1274.

Kalender, W. A., Hebel, R., Ebersberger, J., 1987. Reduction of CT artifacts caused by metallic implants. Radiology 164 (2), 576–577.

Krumm, M., Kasperl, S., Franz, M., 2008. Reducing non-linear artifacts of multi-material objects in industrial 3D computed tomography. NDT & E International 41 (4), 242–251.

Lell, M. M., Meyer, E., Kuefner, M. A., May, M. S., Raupach, R., Uder, M., Kachelriess, M., 2012. Normalized metal artifact reduction in head and neck computed tomography. Investigative Radiology 47 (7), 415–421.

Li, H., Noel, C., Chen, H., Li, H. H., Low, D., Moore, K., Klahr, P., Michalski, J., Gay, H. A., Thorstad, W., et al., 2012. Clinical evaluation of a commercial orthopedic metal artifact reduction tool for CT simulations in radiation therapy. Medical Physics 39 (12), 7507–7517.

Menvielle, N., Goussard, Y., Orban, D., Soulez, G., 2005. Reduction of beam-hardening artifacts in X-ray CT. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the. IEEE, pp. 1865–1868.

Meyer, E., Raupach, R., Lell, M., Schmidt, B., Kachelrieß, M., 2010. Normalized metal artifact reduction (NMAR) in computed tomography. Medical Physics 37 (10), 5482–5493.

Meyer, E., Raupach, R., Lell, M., Schmidt, B., Kachelrieß, M., 2012. Frequency split metal artifact reduction (FSMAR) in computed tomography. Medical Physics 39 (4), 1904–1916.

Morsbach, F., Bickelhaupt, S., Wanner, G. A., Krauss, A., Schmidt, B., Alkadhi, H., 2013. Reduction of metal artifacts from hip prostheses on CT images of the pelvis: value of iterative reconstructions. Radiology 268 (1), 237–244.

Park, H. S., Hwang, D., Seo, J. K., Feb 2016. Metal artifact reduction for polychromatic X-ray CT based on a beam-hardening corrector. IEEE Transactions on Medical Imaging 35 (2), 480–487.

Philips CT Clinical Science, 2012. Metal artifact reduction for orthopedic implants (O-MAR). White paper, Philips Healthcare USA, Andover, MA.

Schmidt, T. G., 2009. Optimal "image-based" weighting for energy-resolved CT. Medical Physics 36 (7), 3018–3027.

Suetens, P., 2002. Fundamentals of Medical Imaging. Cambridge University Press.

Treece, G. M., Dec. 2015. The bitonic filter: linear filtering in an edge-preserving morphological framework. Tech. Rep. CUED/F-INFENG/TR 700, Cambridge University Engineering Dept.

Van de Casteele, E., Van Dyck, D., Sijbers, J., Raman, E., 2002. An energy-based beam hardening model in tomography. Physics in Medicine and Biology 47 (23), 4181.

Van Gompel, G., Van Slambrouck, K., Defrise, M., Batenburg, K. J., de Mey, J., Sijbers, J., Nuyts, J., 2011. Iterative correction of beam hardening artifacts in CT. Medical Physics 38 (S1), S36–S49.

Van Slambrouck, K., Nuyts, J., 2012. Metal artifact reduction in computed tomography using local models in an image block-iterative scheme. Medical Physics 39 (11), 7080–7093.

Vedula, V., Munshi, P., 2008. An improved algorithm for beam-hardening corrections in experimental X-ray tomography. NDT & E International 41 (1), 25–31.

Veldkamp, W. J., Joemai, R. M., van der Molen, A. J., Geleijns, J., 2010. Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT. Medical Physics 37 (2), 620–628.

Verburg, J. M., Seco, J., 2012. CT metal artifact reduction method correcting for beam hardening and missing projections. Physics in Medicine and Biology 57 (9), 2803.

Wang, G., Snyder, D. L., O'Sullivan, J. A., Vannier, M. W., 1996. Iterative deblurring for CT metal artifact reduction. IEEE Transactions on Medical Imaging 15 (5), 657–664.

Wu, M., Keil, A., Constantin, D., Star-Lack, J., Zhu, L., Fahrig, R., 2014. Metal artifact correction for X-ray computed tomography using kV and selective MV imaging. Medical Physics 41 (12), 121910.

Zhang, Y., Yan, H., Jia, X., Yang, J., Jiang, S. B., Mou, X., 2013. A hybrid metal artifact reduction algorithm for X-ray CT. Medical Physics 40 (4), 041910.

Zhang, Y., Zhang, L., Zhu, X. R., Lee, A. K., Chambers, M., Dong, L., 2007. Reducing metal artifacts in cone-beam CT images by preprocessing projection data. International Journal of Radiation Oncology * Biology * Physics 67 (3), 924–932.