Assessment of metal artifact reduction methods in pelvic CT

Mehrsima Abdoli
Department of Radiation Oncology, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands

Abolfazl Mehranian
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva CH-1211, Switzerland

Angeliki Ailianou and Minerva Becker
Division of Radiology, Geneva University Hospital, Geneva CH-1211, Switzerland

Habib Zaidi[a)]
Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva CH-1211, Switzerland; Geneva Neuroscience Center, Geneva University, Geneva CH-1205, Switzerland; and Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, Hanzeplein 1, Groningen 9700 RB, The Netherlands

(Received 1 July 2015; revised 21 December 2015; accepted for publication 14 February 2016; published 3 March 2016)

Purpose: Metal artifact reduction (MAR) produces images with improved quality potentially leading to confident and reliable clinical diagnosis and therapy planning. In this work, the authors evaluate the performance of five MAR techniques for the assessment of computed tomography images of patients with hip prostheses.
Methods: Five MAR algorithms were evaluated using simulation and clinical studies. The algorithms included one-dimensional linear interpolation (LI) of the corrupted projection bins in the sinogram, two-dimensional interpolation (2D), a normalized metal artifact reduction (NMAR) technique, a metal deletion technique, and a maximum a posteriori completion (MAPC) approach. The algorithms were applied to ten simulated datasets as well as 30 clinical studies of patients with metallic hip implants. Qualitative evaluations were performed by two blinded experienced radiologists who ranked overall artifact severity and pelvic organ recognition for each algorithm by assigning scores from zero to five (zero indicating totally obscured organs with no structures identifiable and five indicating recognition with high confidence).
Results: Simulation studies revealed that 2D, NMAR, and MAPC techniques performed almost equally well in all regions. LI falls behind the other approaches in terms of reducing dark streaking artifacts as well as preserving unaffected regions ($p < 0.05$). Visual assessment of clinical datasets revealed the superiority of NMAR and MAPC in the evaluated pelvic organs and in terms of overall image quality.
Conclusions: Overall, all methods, except LI, performed equally well in artifact-free regions. Considering both clinical and simulation studies, 2D, NMAR, and MAPC seem to outperform the other techniques. [http://dx.doi.org/10.1118/1.4942810]

Key words: pelvic CT, metal artifacts, metal artifact reduction, comparison, simulation

1. INTRODUCTION

Over the past few decades, x-ray computed tomography (CT) has emerged as one of the leading cross-sectional imaging modalities offering a broad range of clinical applications in diagnostic radiology, radiation oncology, and multimodal molecular imaging.[1,2] Despite the acknowledged value of this imaging modality, CT image quality and quantitative accuracy can be impaired by a number of physical degrading factors,[3] mainly artifacts arising from the presence of metallic objects in the field-of-view, such as dental fillings, hip or knee prostheses, cardiac pacemakers, and spinal cages.[4] The appearance of streaking artifacts often obscures the anatomical structures surrounding the implants, leading to reduced diagnostic CT confidence and potentially equivocal findings.[5–7] Moreover, severe dark and bright streaking artifacts can result in imperfect dose calculation due to impaired organ boundary delineation in radiation therapy treatment planning[8] and over/underestimation of activity concentration in CT-based attenuation correction of positron emission tomography data.[9,10] As a result, many attempts have been directed toward developing metal artifact reduction (MAR) techniques capable of reducing the artifacts and thus enhancing CT image quality.[11]

Since polychromatic x-ray beams used in x-ray CT pass through the patient, soft and low energy x-rays are preferentially absorbed to a great extent compared to high-energy

photons. A direct consequence of this selective absorption is an increase in patient's absorbed dose, compared to an ideal monoenergetic beam, and a nonlinear increase in the beam's average energy, which is often referred to as beam hardening effect.[12] The metallic objects with a high atomic number strongly attenuate the crossing x-ray photons resulting in severe beam hardening and thus photon starvation at the detectors. This nonlinear propagation of x-ray beams over metallic objects renders the corresponding transmission projection data inconsistent and corrupted. The filtered backprojection (FBP) algorithm, which still is the most widely used algorithm in CT image reconstruction, however, assumes a linear propagation model for the detected photons and as such, is susceptible to data inconsistencies leading to dark and bright streaking artifacts in the reconstructed images. The beam hardening, scatter and streaking metal artifacts can be intrinsically suppressed to an acceptable diagnostic level using polychromatic statistical iterative image reconstruction algorithms.[13] However, such algorithms are memory demanding and computationally intensive. As a result, MAR algorithms have been mostly developed for FBP and algebraic/monochromatic iterative image reconstruction algorithms with the aim of correcting for sinogram corruption and data inconsistency. In this regard, metal artifact reduction techniques can generally be divided into five categories: (i) interpolation-based sinogram correction (i.e., direct interpolation in the sinogram domain), (ii) non-interpolation-based sinogram correction, (iii) hybrid sinogram correction, (iv) iterative image reconstruction, and (v) image-based approaches.[11] Most these MAR techniques possess some advantages and drawbacks in terms of computational simplicity, accuracy, and efficiency in reducing the artifacts produced by different metallic objects and introducing new artifacts. The different characteristics and potential performance of these algorithms, therefore, calls for a comprehensive comparative evaluation in order to understand the benefits and shortcomings of each algorithm.

In this work, five MAR techniques, belonging to the first three categories of the above-mentioned list, were compared and evaluated using simulation and clinical studies of patients with primary hip replacement. To the best of our knowledge, this is the first work that compares sinogram interpolation, noninterpolation, and hybrid approaches in quantitative simulation studies as well as a scored quality assessment by expert readers. Hip replacement is the second most common joint replacement surgery with an increasing number of procedures in the aging population.[14] This comparative study aimed to compare the performance of various MAR algorithms for the assessment of adjacent and distant pelvic organs and global image quality.

2. MATERIALS AND METHODS

2.A. Metal artifact reduction approaches

In this study, we selected representative MAR techniques from the three most commonly used MAR categories, namely, three interpolation-based sinogram correction techniques, one non-interpolation-based method, and one hybrid sinogram correction approach. These techniques are described in Subsections 2.A.1–2.A.3. In all MAR techniques used in this work, a virtual sinogram was utilized rather than the original raw CT data to overcome the challenges associated with the usage of proprietary raw data, usually encrypted by the manufacturers. The virtual sinogram is generated by forward projection of the reconstructed images,[15,16] as a post-processing procedure, which provides an easily accessible version of the original sinogram, in a smaller size. This facilitates fast data correction and eliminates the need to transfer and decrypt large raw data of the original sinogram. The term "*virtual*" is adopted since the original sinogram is not used in these implementations. Although the virtual sinogram is generated from the reconstructed images, these MAR implementations do not belong to the image-based category since the data correction is performed in the projection domain, rather than the image domain. It must be noted that using the original sinogram is to be preferred in case the above mentioned challenges can be easily handled.

2.A.1. Interpolation-based sinogram correction

The majority of the proposed MAR approaches employ this category of techniques due to simplicity and easy and fast implementation. These techniques consist of two steps: (i) metal trace identification, in which the corrupted sinogram bins, namely, missing projection bins, are identified and (ii) missing projection bin interpolation.[17–19] The missing projection bins can be either identified directly in the sinogram domain using dedicated segmentation techniques[19,20] or through forward-projection of segmented metallic objects on the image space.[6,21] Kalender *et al.*[6] proposed a simple linear interpolation (LI) based MAR algorithm in which missing projection bins are identified by forward-projection of manually segmented metallic objects and interpolated along projection profiles using a one-dimensional (1D) linear interpolation algorithm. The 1D linear interpolation takes the data on each column of the sinogram and uses those that are not affected by the metallic object and concatenates linear interpolants between each pair of these data points. The missing data will be found on the concatenated lines, which generates a continuous curve. In the current study, we slightly modified this approach by using a simple thresholding technique to delineate the metallic objects. Since the Hounsfield units (HUs) corresponding to high atomic number objects, such as metals, are considerably higher than that of human body tissues, they can be easily distinguished. However, in regions where the bright streaking artifacts are quite intense, their CT values might be very high, in which case the differentiation, by any means, becomes a challenge.

One-dimensional interpolation is, however, known to generate new artifacts in the reconstructed CT images, mainly due to the discontinuity of the interpolated bins along the second dimension (i.e., the dimension on which the interpolation is not performed) of the sinogram matrix.[15] In an attempt to improve the smoothness and continuity of the sinogram, two-dimensional (2D) interpolation techniques

were proposed.[22,23] The challenge associated with application of 2D interpolation schemes is that the corrupted projection bins are eliminated from the sinogram grid. When some data points are eliminated from a regular grid, it becomes irregular in nature, and as such, the sinogram grid (which is originally a square-based grid) is not regular anymore. Such irregular grid is not compatible with any 2D interpolation method. To tackle this issue, the irregular sinogram grid was rearranged into a triangle-based grid, known as Delaunay triangulated grid.[24] This 2D interpolation approach, compatible with such triangle-based grid, is referred to as Clough–Tocher interpolation.[25] Another reason for using a triangulated grid is that closed-form 2D or 3D interpolation techniques require the transformation of the data into a triangulated mesh (or another irregular gird), where missing data can be interpolated in all possible directions. When the interpolation is performed, the data are transformed back to a regular grid and the corrected image is reconstructed.

Another attempt to improve the performance of the simple linear interpolation approach was investigated by Meyer *et al.*[26] In their approach, known as normalized metal artifact reduction (NMAR), a normalized sinogram is generated by forward projection of a tissue-classified prior image. A multithreshold segmentation is applied to the original CT image in order to obtain the prior image. A linear interpolation is then performed on the normalized sinogram and the corrected image is generated by reconstruction of the denormalized sinogram. The prior images were derived by segmentation of the corrected CT images using the linear interpolation method into two tissue classes, bones and soft tissues using thresholding-based segmentation. As reported in Ref. 26, this would result in a more accurate prior image compared to the segmentation of uncorrected CT images. For the identification of the missing projections, the metallic implant was segmented by the thresholding of the uncorrected CT image at threshold of 2500 HU, and the resulting metal-only image was forward projected. To obtain the prior image, the LI corrected CT images were then segmented for soft tissues (body contour) at threshold on −501 HU was employed. The bones were identified from the LI CT image by a threshold of 200 HU. The metallic implants were excluded from the identified bones using a binary masking. The prior image was then obtained by assigning the CT values of 1070 HU to the soft tissue label, followed by superimposition of the identified bones (having the same CT values as the uncorrected images). Finally, the prior image was smoothed by 8 mm FWHM Gaussian filter.

2.A.2. Noninterpolation-based sinogram correction

Alongside the presented MAR approaches, which make use of interpolation methods to correct for the artifacts in the sinogram domain, a number of sinogram-based strategies aiming at correcting the affected projection bins using methods other than interpolation were investigated. In this context, Mehranian *et al.* proposed a MAR technique based on maximum *a posteriori* completion (MAPC) of the corrupted projection bins.[27] In this technique, a tissue-classified prior image, which is incorporated into a novel prior potential function, serves as a prior knowledge about the missing projection bins. Subtraction of the unknown target and the prior sinograms, which is a measure of sparsity of the residual sinogram, provides a prior knowledge. The MAPC problem was solved as a constrained optimization problem using an accelerated projected gradient algorithm. In this study, for the MAPC, we used the same prior image as used for NMAR. The MAPC method aims to iteratively estimate the missing projections from neighboring projections, and adjacent slices (in 3D case) and also the prior projections obtained from the forward projection of the prior image. In this algorithm, the impact of the prior projections is controlled by a weighting factor α, which in this study was set to one, since in the case of hip prosthesis a large amount of projection data are corrupted and therefore the prior images play an important role in the recovery of the corrupted data. In this algorithm there are two more parameters, including convergence parameter $\eta = 1 \times 10^{-4}$ and the diffusity contrast parameter $\delta = 5 \times 10^{-5}$. These parameters were selected according to previous results in Ref. 27.

2.A.3. Hybrid sinogram correction

Many correction strategies combine various techniques to achieve an improved performance. In this context, a combination of interpolation- and non-interpolation-based sinogram correction was proposed.[28] This method builds on the well-established metal deletion technique (MDT) to substitute the corrupted projection bins in an iterative filtered backprojection process. An initial image is generated using the linear interpolation approach. Thereafter, four iterations of FBP are conducted in each of which, the corrupted bins are replaced by their value in the previous iteration. In this study, the CT images corrected by LI MAR methods were filtered using a 3D median filtering, as an edge-preserving filter. The filtered images were forward projected and the corrupted projections of the original sinograms (identified following the forward projection of the metal only images) were then replaced by the projection of the resulting sinograms. This procedure was repeated four times as described in Ref. 28.

2.B. Simulation and clinical studies

The above referenced MAR techniques (LI, 2D, NMAR, MAPC, and MDT) were evaluated using simulated and clinical studies. Since the ground truth is not available in clinical setting, we followed a simulation procedure described by Mehranian *et al.* for the simulation of metal artifacts on clinical artifact-free CT images.[27] In this procedure, the image is segmented into four classes: air, adipose, soft tissue, and bone and is superimposed by metallic prostheses [Fig. 1(a)]. A polychromatic x-ray spectrum was generated using the SpekCal software for a tube voltage of 120 kVp, 2.5 mm aluminum filtration, 10° anode angle and a tube output of 123.8 μGy/mAs at 1 m. The spectrum was then uniformly sampled into 35 mono-energetic beams, with intensities and average energies calculated over each energy interval to preserve the tube

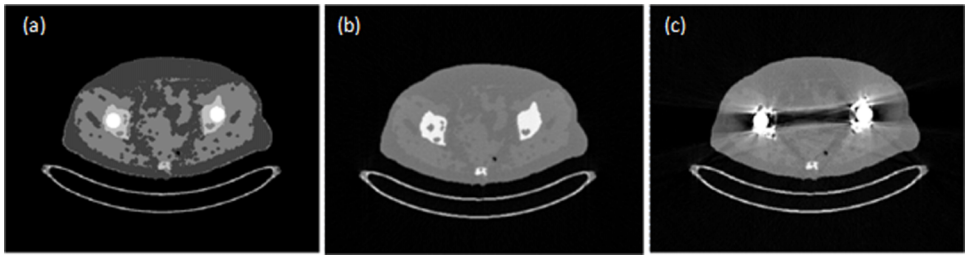

FIG. 1. Artifact simulation process: (a) The CT image is segmented into four classes: air, adipose, soft tissue, and bone. This image presents a labeled map rather than actual intensities; (b) artifact-free CT image (ground truth). (c) Simulated image containing metallic artifacts. The metal is inserted in the segmented image when the artifacted image is to be reconstructed (a) and is excluded otherwise. The window width/window level in (b) and (c) is 2100/1200 HU.

output. We assumed that tissues within each tissue class (air, adipose, soft tissue, bone, and iron) are homogenous, as shown in Fig. 1(a). Since the NIST XCOM photon cross section library provides mass attenuation coefficients of tissues at different energies, we obtained linear attenuation coefficients of each tissue class by multiplying mass attenuation coefficients by corresponding tissue densities. The densities of air, adipose, soft tissue, bone, iron tissues were set to 1.205×10^{-03}, 0.95, 1.06, 1.92, 7.874 g/cm^3. For each tissue class, 35 different attenuation values were obtained based on the NIST XCOM library. Simulated hip implants were adopted from the segmentation of hip implants in a patient. The material of the implant was assumed to be iron with a density of 7.874 g/cm^3. Thereafter, the attenuation map is forward projected and Poisson noise is added to the resulting sinogram to obtain projection data acquired using polychromatic x-ray beams. The amount of added noise mimics the typical noise level observed in clinical studies. The projection data are then log-processed and reconstructed utilizing a FBP algorithm using a ramp filter with a cut-off frequency of 1 cycles/pixel. In this first-order simulation, the CT scanner was assumed to have a linear, energy-invariant and count-independent detector response. Moreover, in this simulation, only the artifacts caused by beam hardening and photon starvation were simulated. Other causes of streaking artifacts, such as nonlinear partial volume effects, exponential edge-gradient, and scatters[29] were not considered. Figures 1(b) and 1(c) show a sample of simulated CT image produced according to the above described procedure. In this study, we simulated metal artifacts in ten patients with hip replacements (five bilateral and five unilateral hip implants). For data simulation, we used the following specifications: fan-beam geometry of a simulated single-slice CT scanner with 888 detector channels, 984 angular samples over a 360 orbit, detector pitch of 1 mm, 949 mm source to detector distance, 541 mm source to isocenter distance, 408 mm isocenter to detector distance. The geometric system matrix describing this scanner was generated by the Image Reconstruction Toolbox, running in MATLAB 2010a (The MathWorks, Inc., Natick, MA).

The various MAR approaches were further evaluated using 30 clinical studies of patients with single or double hip implants (13 males and 17 females with age range, 42–90 yr, and average age 65 yr). These patients were scanned on a 64-slice Biograph mCT scanner (Siemens Healthcare, Erlangen, Germany), using the following parameters: 100 kVp, 30 mAs (using CARE Dose4D automatic tube current modulation), 16×1.2 collimation and a pitch of 1.5.

2.C. Evaluation strategy and statistical analysis

Using the ground truth artifact-free CT image, the impact of MAR can be accurately evaluated using quantitative metrics. Therefore, for the simulated datasets, three metrics were calculated using region of interest (ROI)-based analysis. Six ROIs were defined on the original image, two on regions corresponding to bright streaking artifacts, two on regions corresponding to dark streaking artifacts, and two on artifact-free regions, and the same ROIs were applied to the corrected images. The same ROIs were used on both ground truth images and corrected images. The ROIs are defined manually on the original scan, where the artifacts are clearly visible. The ROIs corresponding to the dark and bright artifacts are defined on areas close to the metallic object, where the artifacts are the strongest. The ROIs corresponding to artifact-free regions are defined in areas close to the skin, where almost no artifact exists. In Fig. 2(b), three representative ROIs are illustrated on different regions, as explained above. The evaluation metrics include the mean relative error (MRE), normalized root mean square difference (NRMSD), and mean absolute deviation (MAD) between the corrected ($I^{\text{corrected}}$) and ground truth image (I^{true}),

$$\text{MRE} = \frac{I^{\text{corrected}} - I^{\text{true}}}{I^{\text{true}}}, \tag{1}$$

$$\text{NRMSD} = \sqrt{\frac{\sum_{i\in\text{ROI}} \left(I_i^{\text{corrected}} - I_i^{\text{true}}\right)^2}{\sum_{i\in\text{ROI}} \left(I_i^{\text{true}}\right)^2}}, \tag{2}$$

$$\text{MAD} = \frac{1}{N}\sum_{i\in\text{ROI}} \left|I_i^{\text{corrected}} - I_i^{\text{true}}\right|. \tag{3}$$

$I^{\text{corrected}}$ and I^{true} represent the mean intensity (original HU + 1024 to prevent zero denominator) within each ROI and i is the index of pixels belonging to each ROI. N is the number of voxels in a given ROI.

The difference between mean intensities within the defined regions of interest on the corrected and ground truth images was also statistically analyzed using a two-tailed paired t-test.

For the clinical datasets, where the ground truth is not available, a blind qualitative evaluation of the original anonymized

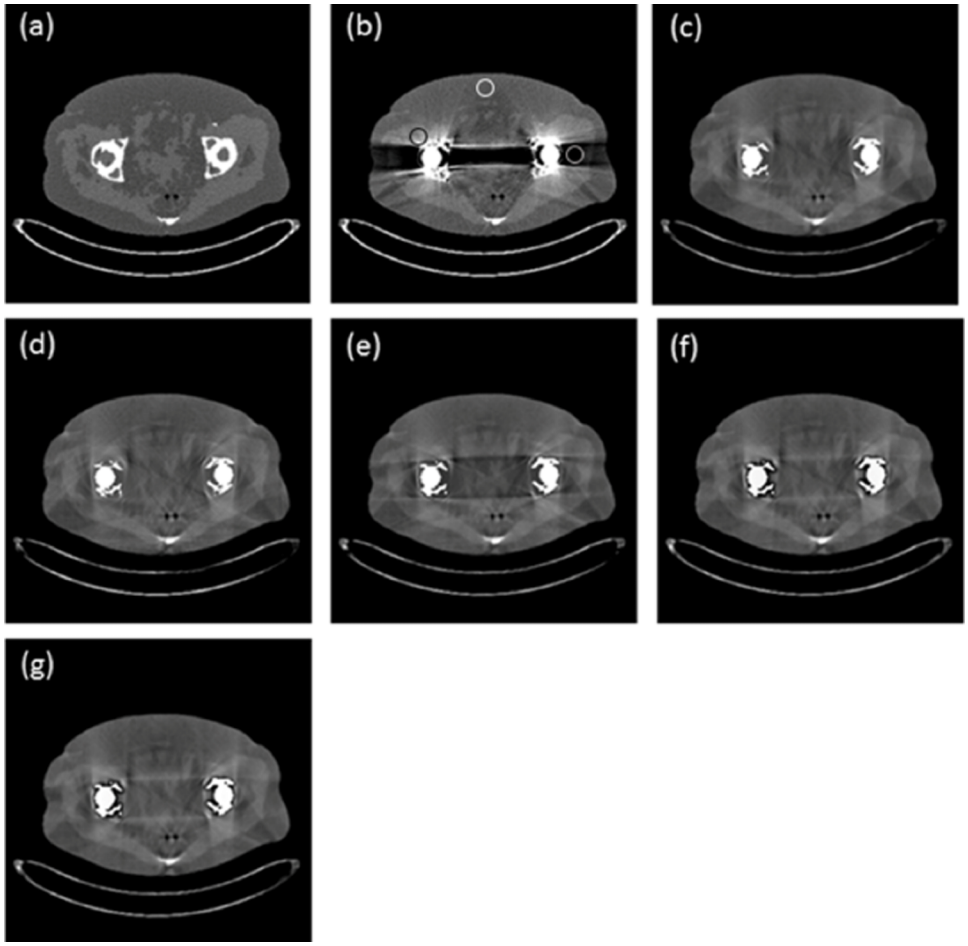

FIG. 2. (a) Representative artifact-free CT image without metal implants, (b) artifacted CT image with simulated metal implants, and same image shown in (b) corrected using the different MAR techniques: (c) LI, (d) 2D, (e) MDT, (f) NMAR, and (g) MAPC. The ROIs used for the quantitative analysis are shown in (b), where the black ROI corresponds to the bright region, the grey ROI to the dark region and the white ROI to the unaffected region. Window width/window level = 2100/1200 HU.

CT images and those produced using the five MAR techniques was performed by two experienced radiologists (10 and 25 yr of experience) separately who ranked the distinction of five organs in the pelvic area: bladder, rectum, prostate (if applicable), seminal vesicles (if applicable), and vagina (if applicable). The ranking was performed by assigning scores from zero to five, with rank zero indicating totally obscured organs, with no structures identifiable and five indicating recognition with high confidence.[28] The overall quality of the images was also similarly scored. Typical windowing adopted in clinical pelvic organ evaluation was used to evaluate the datasets. The window level/width was chosen such that they allow for the best visual inspection and represent the closest setting to what the clinicians use in routine. The same fixed settings were used across all comparisons.

Interobserver agreement regarding ratings for organ evaluation and for overall image quality was assessed using the kappa statistics, kappa values ranging from zero to one (slight agreement: 0.01–0.2; fair agreement: 0.21–0.4; moderate agreement: 0.41–0.6; good agreement: 0.61–0.8; excellent agreement: 0.81–1). All statistical analyses (t-test and kappa) were performed using SPSS for windows (SPSS, Inc., version 14.0).

3. RESULTS

3.A. Quantitative analysis

Figure 2 illustrates an example of a simulated artifacted CT image together with the corresponding ground truth and corrected images. It can be observed that all MAR approaches reduce the apparent streaking artifacts to a large extent. However, since the simulated datasets contain less diverse assortment of soft tissue compartments compared to real clinical studies, there is less information to be retrieved through the artifact reduction process.

Figure 3 presents the quantitative analysis of the simulated images using the three defined metrics. Each metric is presented in a separate graph where every MAR technique is plotted as a bar with a different gray level. Looking at the three plots, it can be observed that three methods (2D, NMAR, and MAPC) seem to share the highest performance in correcting for dark streaking artifacts, while LI falls behind the other methods in all regions.

In regions with no artifacts, the errors and deviations associated with all MAR methods remain similarly low, except LI which presents a slightly higher error. Table I summarizes the results of the statistical analysis of these datasets in different

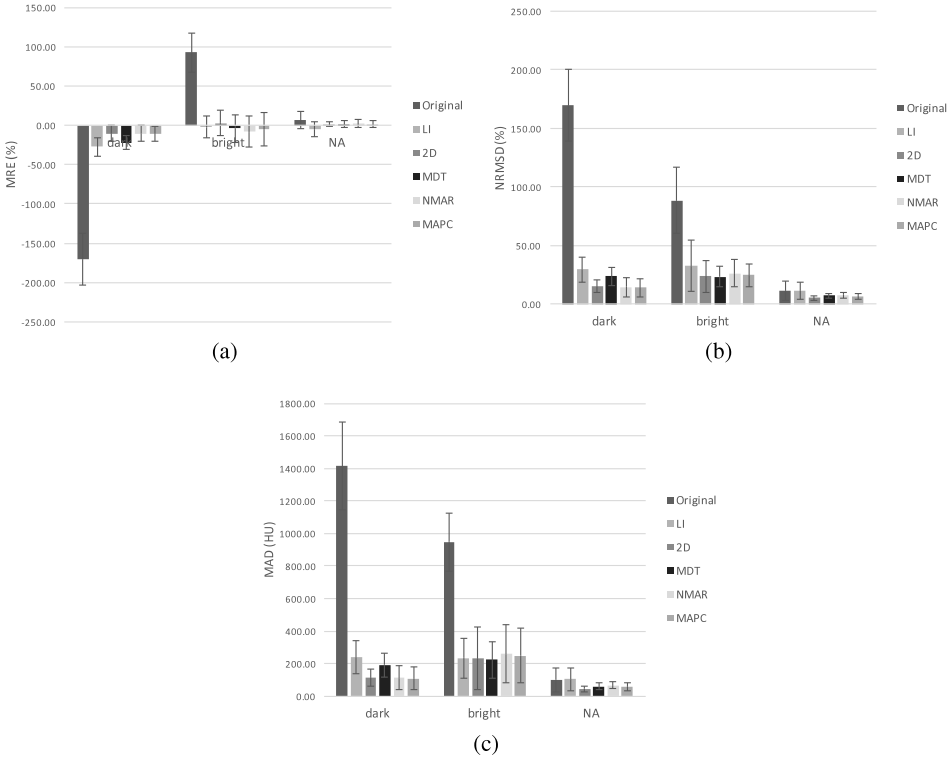

FIG. 3. Quantitative analysis of the simulated CT images using three metrics: (a) mean relative error (MRE), (b) normalized root mean square difference (NRMSD), and (c) mean absolute deviation (MAD). Error bars represent the standard deviation over the group of (simulated) patients.

regions using a two-tailed paired t-test. As the results presented in Table I suggest, the differences between regions associated with dark streaking artifacts and the corresponding regions on the ground truth image are all significant (P-value < 0.05). However, there is no proof of statistical significance between regions associated with bright streaking artifacts and artifact-free areas and the corresponding regions on the artifact-free images for all five correction techniques (P-value > 0.05). It must be noted that failure to prove statistically significant differences is not sufficient to confirm that the results are statistically identical.

3.B. Qualitative data analysis

A representative clinical study is shown in Fig. 4, where the original image and the corresponding corrected images using the five MAR methods are presented. It can be noticed that almost all methods leave a trace of the streaking artifacts after correction. However, they all enhance the quality of the images to a great extent. Since the ground truth for the clinical datasets is not known, the images were independently analyzed by two expert radiologists and scored from 0 to 5 for five organs in the pelvic area based on how the organ can be recognized. The different organs used for scoring are marked with an arrow [see Fig. 4(f)]. The overall quality of the images was also ranked to reflect the severity of the remaining artifacts. Figure 5 illustrates the average scores and standard deviations assigned to each organ, as well as the ranking reflecting overall image quality.

A representative clinical study is shown in Fig. 6 where the streaking artifacts in the original image [Fig. 6(a)] completely obscure the prostate, seminal vesicles, and bladder. The average scores assigned by the radiologists to these organs are 0, 0, and 0.5, respectively. As can be seen in Figs. 6(b)–6(f), all five MAR techniques have considerably improved the diagnostic value of CT images and the pelvic organs are clearly recognizable after image processing, with scores varying between 4 and 5. The subtraction of corrected and original images is illustrated in Figs. 6(g)–6(k) to highlight the effect of each MAR method. Interobserver agreement for scoring of each organ, as well as for overall image quality were excellent with kappa values as follows: 0.87 for the prostate, 0.89 for the seminal vesicles, 0.86 for the vagina, 0.80 for the bladder, and 0.83 for overall image quality. Although there is a very good agreement between the observations of the two raters, the calculation of kappa was not possible for the rectum since the results did not meet the kappa statistics criteria given that the ranges of scores for both observers were different (one radiologist ranked it in the range of two–five, while the second ranked it between four and five).

TABLE I. P-values obtained by statistical analysis of the different MAR techniques in dark, bright, and unaffected regions for the simulated studies using paired two-tailed t-test.

	MAR method				
Region	LI	2D	MDT	NMAR	MAPC
Dark	<0.005	<0.001	<0.001	<0.002	<0.002
Bright	0.79	0.57	0.38	0.22	0.40
Unaffected	0.06	0.24	0.23	0.08	0.14

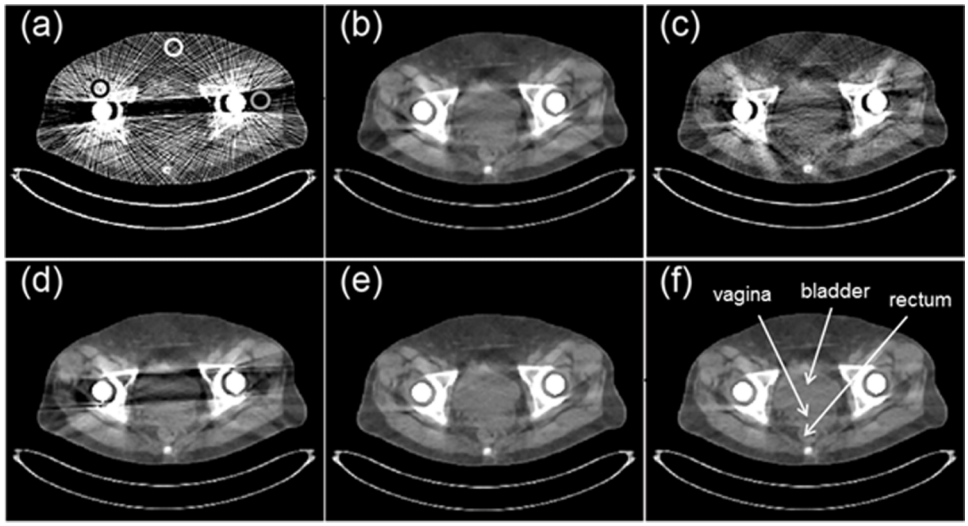

FIG. 4. Representative clinical case showing (a) the original artifacted CT image and the corrected images using the various MAR approaches: (b) LI, (c) 2D, (d) MDT, (e) NMAR, and (f) MAPC. The organs used for scoring are also depicted in (f). Window width/window level = 500/20 HU.

4. DISCUSSION

The reduction of metallic streaking artifacts on CT images proved to play a significant role in improving diagnostic image quality and quantitative capability of this imaging modality. The confident usage of x-ray CT in a broad range of clinical applications requires proper correction for image degrading factors to reduce the artifacts they produce. Several correction approaches have been proposed during the past few decades,[11] which motivated the comparison of their performance in an attempt to highlight their advantages and drawbacks. In this paper, we compared five different MAR techniques belonging to different categories of MAR approaches using simulated and clinical studies of patients bearing hip metallic implants. All selected MAR methods operate in the sinogram domain, which are generally superior to image-based approaches.[11]

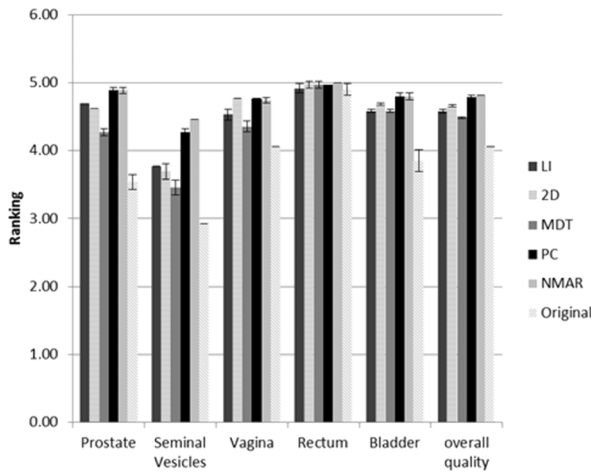

FIG. 5. Average scoring of the clinical studies in five selected organs of the pelvic area and overall image quality (observations of the two radiologists are averaged). Error bars represent the standard deviation.

The results obtained using simulated studies demonstrated that the performance of the selected MAR algorithms alters in different regions of the images. In regions corresponding to dark streaking artifacts, two interpolation-based approaches (i.e., 2D and NMAR) and a noninterpolation-based approach (i.e., MAPC) perform almost equally well. A lower MAD is the evidence of a lower deviation in the intensity of the pixels (in HU) in the investigated regions. Since the ROIs were carefully defined in areas where only one tissue type is present, a low MAD was expected. MRE and NRMSD are two measures of relative error (in percentage) and as such, lower values are desired. As a consequence, all three quantitative validation metrics confirm that 2D, NMAR, and MAPC outperform the other methods in terms of correcting for dark streaking artifacts. LI appears to have the lowest performance in these regions.

In regions corresponding to bright streaking artifacts, all methods have comparable performance. However, LI has slightly inferior performance in terms of NRMSD. NMAR appears to fall slightly behind the other three methods in correcting the bright streaking artifacts. The performance of all selected methods, except LI, is approximately similar in artifact-free regions. MRE and NRMSD remain within about ±5% and deviations of intensities are less than 50 HU. The average HU deviation in artifact-free regions goes up to 100 HU when using LI and NRMSD. MRE measures are also higher relative to other four methods. The results of the statistical analysis revealed that the mean differences between the corrected images and the corresponding ground truth are not statistically significant in regions corresponding to bright streaking artifacts and artifact-free regions. On the other hand, all techniques seem to have equally poor performances in dark streaking areas ($P > 0.05$). It would have been beneficial if a statistically powered analysis could be performed to back up the results of this comparative study. However, the limited number of samples did not enable us to conduct this type of rigorous statistical analysis.

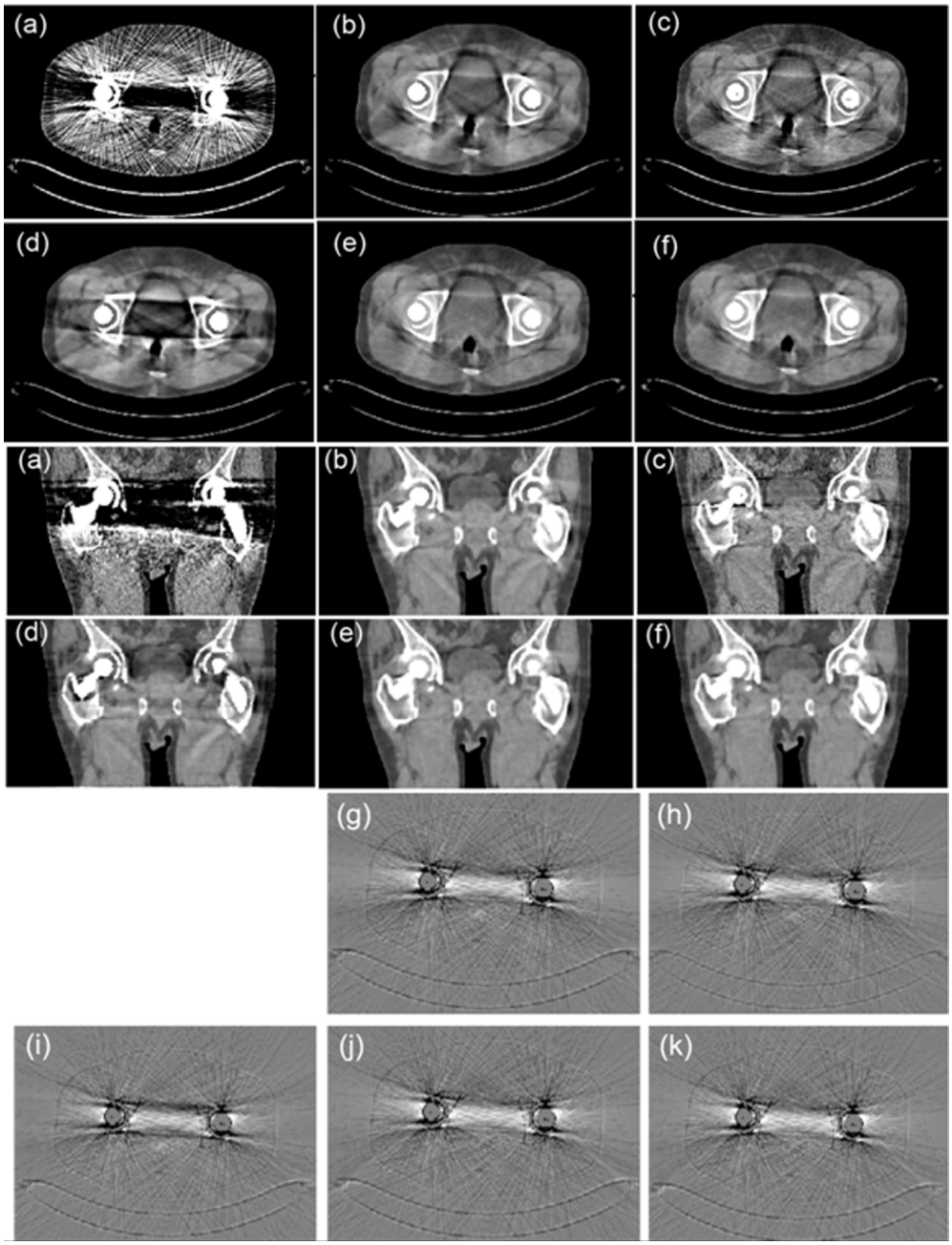

FIG. 6. Representative clinical study showing in two planes (top: axial, middle: coronal, and bottom: subtraction images in axial view) the considerable enhancement of the diagnostic value of CT examinations contributed by metal artifact reduction. (a) The original artifacted CT image and the corrected images generated using the various MAR approaches: (b) LI, (c) 2D, (d) MDT, (e) NMAR, and (f) MAPC. Window width/window level = 500/20 HU. (g)–(k) Illustrate the subtraction of (b)–(f) from the original image (a) in axial view. The window width/level for the subtracted images is 2300/120 HU.

The results of the clinical studies enabled us to observe that, except the rectum, which is almost never influenced by streaking artifacts and as such is ranked close to five in all image sets, the other organs have been scored approximately between three and four before correction (Fig. 5). This means that all organs in the pelvic area are more or less recognizable, with different levels of certainty.

Although the differences between the ranking of the different MAR techniques were not considerable, NMAR and MAPC appear to outperform the other techniques, particularly for the prostate, where the ranking is very close to the maximum rank. These observations are in agreement with the simulated studies in the regions corresponding to dark streaking artifacts. In this analysis, the LI algorithm does not outperform the other techniques, neither in the five different organs nor the overall image quality. The 2D approach achieves an intermediate ranking among other approaches. LI and MDT techniques have shown the lowest performance using both simulation and clinical studies.

The contrasting performance of MAR techniques belonging to the same category (e.g., NMAR and LI) indicates that the approaches cannot be ranked based on the category they

belong to. Since the more complicated MAR techniques, such as NMAR and MAPC, are able to outperform simple interpolation schemes, such as LI, it seems advisable to invest in more complex approaches. On the other hand, there is still room for improvement of the more sophisticated approaches, such as NMAR and MAPC, which might enable them to achieve improved performance in the regions corresponding to bright streaking artifacts. On the other hand, the potential of rather simple approaches, such as 2D, must not be overlooked. Hybrid approaches enabling to combine various techniques taking the advantages of each one into account seem to be the way to go to achieve an ultimate solution addressing the shortcomings of each technique. Recent studies reporting on the evaluation of this type of techniques seem to confirm their strength and robustness in clinical setting.[30] In addition, utilizing enhancement approaches, such as the frequency split method proposed by Meyer *et al.* to preserve structures adjacent to the metallic object,[31] can contribute to the improvement of the performance of each MAR technique.

The scope of this study was limited to metal hip implants. Further investigation of other common sources of streaking artifacts, such as dental fillings, is required to generalize the conclusions of these findings regarding the performance of various MAR approaches. It should however be noted that metallic hip implants have been chosen in this work since they produce the most severe streaking artifacts and, as such, this work is deemed to be a proper representative sample of various sources of metallic artifact.

Another known limitation of this work is that artificial sinograms are used rather than the original sinograms produced by the CT scanner. The virtual sinogram is a reprojection of previously reconstructed CT images using a simple projection model ignoring some physical factors, such as beam-hardening and scatter. Although previous studies demonstrated that the use of virtual sinograms does not bring in significant bias in the outcome of MAR techniques,[11,15] it is worthwhile to compare the performance of the various MAR algorithms on the original sinograms using a larger number of clinical studies to increase the statistical significance of the study.

5. CONCLUSION

In this work, we compared the performance of five MAR techniques for the reduction of metal artifacts in CT images of patients presenting with hip prostheses using simulation and clinical studies. The simulation studies demonstrated that the 2D, NMAR, and MAPC methods achieved the best performance in reducing dark streak artifacts, which are dominant in patients with bilateral hip implants. In regions contaminated with bright streaking artifacts, all five methods exhibited comparable performance. The results of the clinical studies showed that NMAR and MAPC methods outperform the other techniques, particularly for the prostate, which is usually obscured in dark streaking artifacts. It was concluded that the reduction of artifacts induced by hip prosthesis, enabling improved diagnostic confidence in CT images, can be reliably achieved by NMAR and MAPC techniques.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under Grant No. SNSF 31003A-149957 and 320030_135728/1 and the Indo-Swiss Joint Research Programme No. ISJRP-138866.

[a]Author to whom correspondence should be addressed. Electronic mail: habib.zaidi@hcuge.ch; Telephone: +41 22 372 7258; Fax: +41 22 372 7169.

[1]W. A. Kalender, "X-ray computed tomography," Phys. Med. Biol. **51**, R29–R43 (2006).
[2]A. Bockisch, L. S. Freudenberg, D. Schmidt, and T. Kuwert, "Hybrid imaging by SPECT/CT and PET/CT: Proven outcomes in cancer imaging," Semin. Nucl. Med. **39**, 276–289 (2009).
[3]J. F. Barrett and N. Keat, "Artifacts in CT: Recognition and avoidance," Radiographics **24**, 1679–1691 (2004).
[4]E. Klotz, W. A. Kalender, R. Sokiransky, and D. Felsenberg, "Algorithms for the reduction of CT artifacts caused by metallic implants," Proc. SPIE **1234**, 642–650 (1990).
[5]D. D. Robertson, J. Yuan, G. Wang, and M. W. Vannier, "Total hip prosthesis metal-artifact suppression using iterative deblurring reconstruction," J. Comput. Assisted Tomogr. **21**, 293–298 (1997).
[6]W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants," Radiology **164**, 576–577 (1987).
[7]A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi, "X-ray CT metal artifact reduction using wavelet domain L_0 sparse regularization," IEEE Trans. Med. Imaging **32**, 1707–1722 (2013).
[8]Y. Kim, W. A. Tome, M. Bal, T. R. McNutt, and L. Spies, "The impact of dental metal artifacts on head and neck IMRT dose distributions," Radiother. Oncol. **79**, 198–202 (2006).
[9]P. E. Kinahan, D. W. Townsend, T. Beyer, and D. Sashin, "Attenuation correction for a combined 3D PET/CT scanner," Med. Phys. **25**, 2046–2053 (1998).
[10]H. Zaidi and B. H. Hasegawa, "Determination of the attenuation map in emission tomography," J. Nucl. Med. **44**, 291–315 (2003).
[11]M. Abdoli, R. A. J. O. Dierckx, and H. Zaidi, "Metal artifact reduction strategies for improved attenuation correction in hybrid PET/CT imaging," Med. Phys. **39**, 3343–3360 (2012).
[12]W. R. Hendee and E. R. Ritenour, *Medical Imaging Physics*, 4th ed. (Wiley-Liss, Inc., New York, NY, 2002).
[13]J. Nuyts *et al.*, "Modelling the physics in the iterative reconstruction for transmission computed tomography," Phys. Med. Biol. **58**, R63–R96 (2013).
[14]H. J. Griffiths, D. R. Priest, D. M. Kushner, and D. Kushner, "Total hip replacement and other orthopedic hip procedures," Radiol. Clin. North Am. **33**, 267–287 (1995).
[15]M. Abdoli, M. Ay, A. Ahmadian, R. Dierckx, and H. Zaidi, "Reduction of dental filling metallic artefacts in CT-based attenuation correction of PET data using weighted virtual sinograms optimized by a genetic algorithm," Med. Phys. **37**, 6166–6177 (2010).
[16]M. Kidoh, T. Nakaura, S. Nakamura, S. Tokuyasu, H. Osakabe, K. Harada, and Y. Yamashita, "Reduction of dental metallic artefacts in CT: Value of a newly developed algorithm for metal artefact reduction (O-MAR)," Clin. Radiol. **69**, e11–e16 (2014).
[17]G. H. Glover and N. J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT reconstructions," Med. Phys. **8**, 799–807 (1981).
[18]T. Hinderling, P. Ruegsegger, M. Anliker, and C. Dietschi, "Computed tomography reconstruction from hollow projections: An application to *in vivo* evaluation of artificial hip joints," J. Comput. Assisted Tomogr. **3**, 52–57 (1979).
[19]C. Xu, F. Verhaegen, D. Laurendeau, S. A. Enger, and L. Beaulieu, "An algorithm for efficient metal artifact reductions in permanent seed," Med. Phys. **38**, 47–56 (2011).
[20]W. J. H. Veldkamp, R. M. S. Joemai, A. J. van der Molen, and J. Geleijns, "Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT," Med. Phys. **37**, 620–628 (2010).
[21]M. Bazalova, L. Beaulieu, S. Palefsky, and F. Verhaegen, "Correction of CT artifacts and its influence on Monte Carlo dose calculations," Med. Phys. **34**, 2119–2132 (2007).
[22]M. Abdoli, J. R. de Jong, J. Pruim, R. A. Dierckx, and H. Zaidi, "Reduction of artefacts caused by hip implants in CT-based attenuation-

corrected PET images using 2-D interpolation of a virtual sinogram on an irregular grid," Eur. J. Nucl. Med. Mol. Imaging **38**, 2257–2268 (2011).
[23]D. Prell, Y. Kyriakou, M. Beister, and W. A. Kalender, "A novel forward projection-based metal artifact reduction method for flat-detector computed tomography," Phys. Med. Biol. **54**, 6575–6591 (2009).
[24]P. Delaunay, "Sur la sphère vide," Bull. Acad. Sci. USSR **6**, 793–800 (1934).
[25]R. Clough and J. Tocher, "Finite element stiffness matrices for analysis of plates in blending," in *Proceedings of Conference on Matrix Methods in Structural Analysis* (Air Force Flight Dynamics Laboratory Technical Report: AFFDL-TR-66-80, Wright-Patterson AFB, OH, 1965), Vol. 515-545, see http://contrails.iit.edu/DigitalCollection/1966/AFFDLTR66-080.html.
[26]E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Normalized metal artifact reduction (NMAR) in computed tomography," Med. Phys. **37**, 5482–5493 (2010).
[27]A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi, "3D prior image constrained projection completion for x-ray CT metal artifact reduction," IEEE Trans. Nucl. Sci. **60**, 3318–3332 (2013).
[28]F. E. Boas and D. Fleischmann, "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography," Radiology **259**, 894–902 (2011).
[29]B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Metal streak artifacts in x-ray computed tomography: A simulation study," IEEE Trans. Nucl. Sci. **46**, 691–696 (1999).
[30]M. Axente, A. Paidi, R. Von Eyben, C. Zeng, A. Bani-Hashemi, A. Krauss, and D. Hristov, "Clinical evaluation of the iterative metal artifact reduction algorithm for CT simulation in radiotherapy," Med. Phys. **42**, 1170–1183 (2015).
[31]E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Frequency split metal artifact reduction (FSMAR) in computed tomography," Med. Phys. **39**, 1904–1916 (2012).