

Comparative Methods for Metal Artifact Reduction in x-ray CT

Mehrsima Abdoli, Member, IEEE, Abolfazl Mehranian, Angeliki Ailianou, Minerva Becker, and Habib Zaidi, Senior Member, IEEE

Abstract– To assess the performance of five metal artefact reduction (MAR) techniques for the assessment of computed tomography (CT) images of patients with hip prostheses. Five MAR algorithms were evaluated using simulation and clinical studies. The algorithms included one-dimensional linear interpolation (LI) of the corrupted projections in the sinogram, two-dimensional interpolation (2D), a normalized metal artefact reduction (NMAR) technique, a metal deletion technique (MDT), and a 3D prior image constrained projection completion approach (MAPC). The algorithms were applied to 10 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 artefact severity, as well as pelvic organ recognition for each algorithm, respectively. The simulated studies revealed that 2D, NMAR and MAPC techniques performed almost equally well in regions with dark streaking artefacts. However, in regions with bright streaking artefacts, LI outperformed the other techniques ($p < 0.05$). Visual assessment of clinical datasets confirmed the superiority of NMAR and MAPC in the evaluated pelvic organs and in terms of overall image quality. Overall, all methods performed equally well in artefact-free regions. However, NMAR and MAPC outperformed the other techniques in regions affected by artefacts.

I. INTRODUCTION

DUE to the high atomic number of metallic objects, x-ray photons are severely attenuated and generate gaps in the projection data. Reconstruction of such incomplete data results in appearance of bright and dark streaking artifacts on the reconstructed CT image. Metal induced streaking artifacts deteriorate the diagnostic quality and quantitative value of CT

This work was supported by the Swiss National Science Foundation under grant SNSF 31003A-149957.

M. Abdoli is with Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands (e-mail: mehrsima.abdoli@gmail.com).

A. Mehranian is with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland (e-mail: Abolfazl.Mehranian@etu.unige.ch)

A. Ailianou is with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland (e-mail: angeliki.ailianou@hcuge.ch)

M. Becker is with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland (e-mail: minerva.becker@hcuge.ch)

H. Zaidi is with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland and Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands (e-mail: habib.zaidi@hcuge.ch)

images. They obscure the anatomical structures surrounding the metallic objects and preclude confident diagnosis of the disease. Therefore, delineation of the organs for radiotherapy treatment planning using such artifactual CT image might become challenging. Moreover, in CT-based attenuation correction of PET images, streaking artifacts might introduce over- and/or under-estimation of activity concentration in regions corresponding to bright and dark streaking artifacts. Such over- and/or under-estimations might lead to false diagnosis of the disease. Therefore, reduction of metal-induced streaking artifacts has a significant role in improvement of diagnosis and treatment, and as such various approaches have been proposed to tackle the problem [1].

In this study five metal artifact reduction (MAR) techniques, belonging to different MAR categories, are comprehensively compared and evaluated. This comparison is intended to provide a deeper insight over the pros and cons of each category of MAR approaches.

II. MATERIALS AND METHODS

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 the following sub-section.

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 projections, are identified and ii) missing projection interpolation [2-4]. The missing projections can be either identified directly in the sinogram domain using dedicated segmentation techniques [4, 5] or through forward-projection of segmented metallic objects on the image space [6, 7]. Kalender et al [6] proposed a simple linear interpolation (LI) based MAR algorithm in which missing projections are identified by forward-projection of manually segmented metallic objects and interpolated along projection profiles using a one-dimensional (1D) linear interpolation algorithm. In the current study, we slightly modified this approach by using a simple thresholding technique to delineate the metallic

objects. Since the Hounsfield units (HU) 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 artefacts are quite intense, their HUs might be as high as metallic objects, in which case the differentiation, by any means, becomes a challenge.

1D interpolation is, however, known to generate new artefacts in the reconstructed CT images, mainly due to the discontinuity of the interpolated bins along the second dimension of the sinogram matrix [8]. In an attempt to improve the smoothness and continuity of the sinogram, a two-dimensional (2D) interpolation technique was proposed by Abdoli et al [9]. The challenge associated with application of 2D interpolation schemes is that the corrupted projection bins are eliminated from the sinogram grid, 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 [10]. This 2D interpolation approach, compatible with such triangle-based grid, is referred to as Clough-Tocher interpolation [11].

Another attempt to improve the performance of the simple linear interpolation approach was investigated by Meyer et al [12]. In their approach, known as normalized metal artefact reduction (NMAR), a normalized sinogram is generated by forward projection of a tissue-classified prior image. A multi-threshold 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 de-normalized sinogram.

Non-interpolation based sinogram correction

Alongside the presented MAR approaches, which make use of interpolation methods to correct for the artefacts 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 of the corrupted projections (MAPC) [13]. 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 projections. 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.

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 by Boas et al [14]. This method, referred to as metal deletion technique (MDT),

substitutes 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.

B. Simulated and clinical studies

The five selected MAR techniques were evaluated using simulated and clinical datasets. Since in the clinical setting the ground truth is not known, an accurate quantitative evaluation of the techniques is not feasible in such setting. Therefore, we have followed the simulation procedure applied in the literature [13] in order to obtain realistic artifact-simulated CT images. We have simulated 10 metal-induced artifacts generated by hip implants on real clinical artifact-free hip CT images. First the image is segmented into three classes: air, soft tissue and bone and is superimposed by metallic prostheses. By modeling the polychromatic propagation of X-ray beams in the patient body, the beam hardening and the resulting streaking artifacts are simulated. Theoretical linear attenuation coefficient of each tissue class is assigned to the corresponding segmented compartment. 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 projection data is then log-processed and reconstructed utilizing FBP technique. We have simulated five double hip implant and five single implant CT images.

The representative MAR approaches were further evaluated on 30 clinical CT studies of patients with double or single hip implants. These patients were scanned on a 64-slice Biograph mCT scanner (Siemens Medical Solutions, Erlangen, Germany).

C. Evaluation strategy

In the presence of a ground truth, the processed images can be accurately evaluated using quantitative evaluation metrics. Therefore, for the simulated datasets, where the ground truth is available, three different metrics were measured using region of interest (ROI) analysis. Six ROIs were defined on each image, two on the regions corresponding to bright streaking artifacts, two in regions corresponding to dark streaking artifacts, and two on artifact-free regions. The same ROIs were used on the original artifact-free images (ground truth) as well as the artifact corrected images. The three evaluation metrics include mean relative error (MRE), normalized root mean square difference (NRMSD) and mean absolute deviation (MAD) between the corrected image ($I^{\text{corrected}}$) and the ground truth (I^{true}):

$$\%MRE = \frac{I^{\text{corrected}} - I^{\text{true}}}{I^{\text{true}}} \times 100 \quad (1)$$

$$\%NRMSD = 100 \times \sqrt{\frac{\sum_{i \in \text{ROI}} (I_i^{\text{corrected}} - I_i^{\text{true}})^2}{\sum_{i \in \text{ROI}} (I_i^{\text{true}})^2}} \quad (2)$$

$$MAD = \frac{1}{N} \sum_{i \in \text{ROI}} |I_i^{\text{corrected}} - I_i^{\text{true}}| \quad (3)$$

For the clinical datasets, a blind qualitative evaluation of the 5 MAR CT images and the original image by two experienced radiologists was performed who ranked the overall severity of artifacts for each algorithm, with a rank of 0 indicating the most severe artifacts and a rank of 5 indicating the least severe artifacts.

III. RESULTS

A. Quantitative analysis

Ten simulated hip datasets were used for performance assessment of the five MAR techniques. Figure 1 illustrates an example of a simulated artefacted CT image together with the corresponding ground truth and corrected images. It can be observed that all MAR approaches reduce the apparent streaking artefacts to some 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 artefact reduction process, and as such, the corrected images using all MAR approaches look slightly blurred.

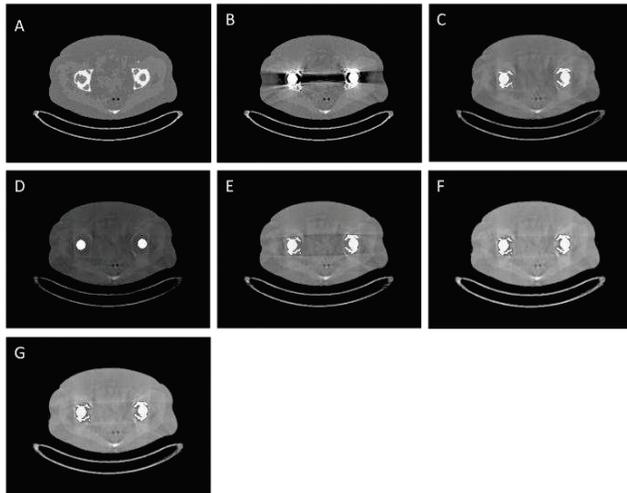


FIG. 1. (A) REPRESENTATIVE ARTEFACT-FREE CT IMAGE, (B) ARTEFACTED CT IMAGE, AND SAME IMAGE SHOWN IN (B) CORRECTED USING THE DIFFERENT MAR TECHNIQUES: (C) LI, (D) 2D, (E) MDT, (F) NMAR AND (G) MAPC.

Figure 2 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) share the highest performance in correcting for dark streaking artefacts, while LI outperforms the other methods in terms of reducing bright streaking artefacts. In regions with no artefacts, the errors and deviations associated with all five MAR methods remain similarly low. It must be noted that error bars are not shown in figures 3(b) and 3(c) since they both present a single measurement for the whole dataset, and thus there is no variability in the presented results. Table 1 summarizes the results of the statistical analysis of these datasets in different regions using on a 2-tailed paired t-test. As the results presented in Table 1 suggest, the differences in regions corresponding to dark streaking

artefacts are all significant (P -value < 0.05). However, there is no proof of statistical significance between regions corresponding to bright streaking artefacts and artefact-free areas for all five correction techniques (P -value > 0.05).

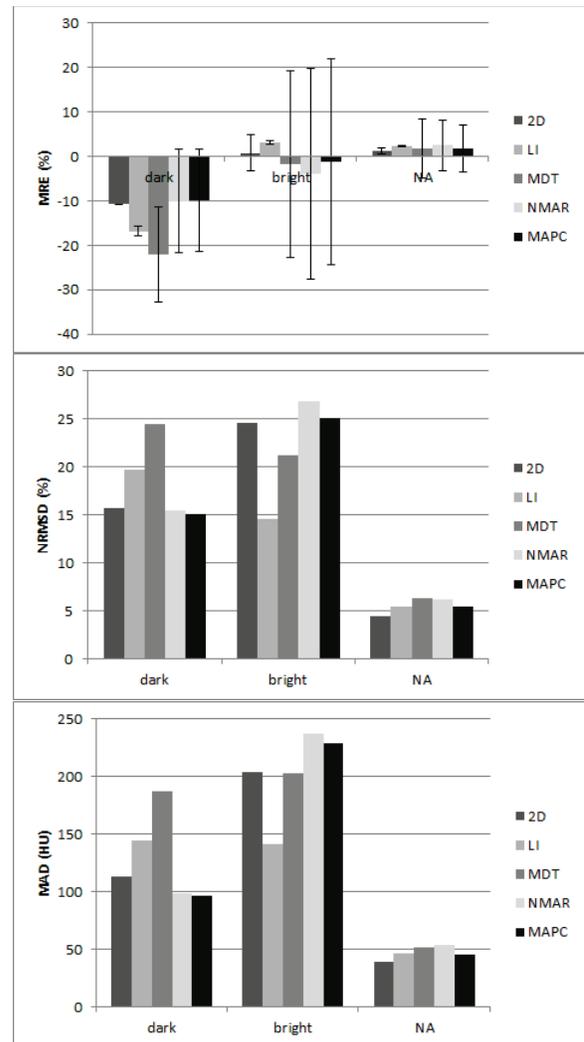


FIG. 2. QUANTITATIVE ANALYSIS OF THE SIMULATED CT IMAGES USING THREE METRICS. TOP: MEAN RELATIVE ERROR (MRE), MIDDLE: NORMALIZED ROOT MEAN SQUARE DIFFERENCE (NRMSD), AND BOTTOM: MEAN ABSOLUTE DEVIATION (MAD).

TABLE I. P-VALUES OBTAINED BY STATISTICAL ANALYSIS OF THE SIMULATED STUDIES USING PAIRED 2-TAILED T-TEST.

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

B. Qualitative data analysis

A representative clinical study is shown in figure 3, 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 artefacts

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 overall quality of the images was also ranked to reflect the severity of the remaining artefacts. Figure 4 illustrates the average scores and standard deviations assigned to each organ, as well as the ranking reflecting overall image quality.

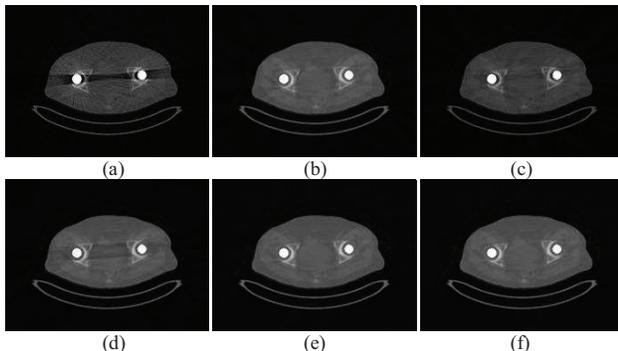


FIG. 3. REPRESENTATIVE CLINICAL CASE SHOWING (A) THE ORIGINAL ARTEFACTED CT IMAGE AND THE CORRECTED IMAGES USING THE VARIOUS MAR APPROACHES: (B) LI, (C) 2D, (D) MDT, (E) NMAR AND (F) MAPC.

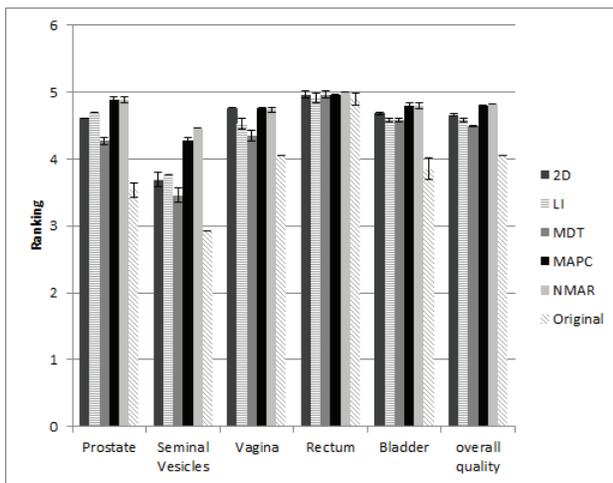


FIG. 4. AVERAGE SCORING OF THE CLINICAL STUDIES IN FIVE SELECTED ORGANS OF THE PELVIC AREA AND OVERALL IMAGE QUALITY (OBSERVATIONS OF THE TWO RADIOLOGISTS ARE AVERAGED).

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. 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 that 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 2 to 5, while the second ranked it between 4 and 5).

IV. DISCUSSION

The reduction of metallic streaking artefacts 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 artefacts they produce. Several correction approaches have been proposed during the past few decades [1], 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 proved to be superior to image-based approaches.

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 artefacts, two interpolation-based approaches (i.e. 2D and NMAR) and a non-interpolation 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 artefacts. MDT appears to have the lowest performance in these regions.

However, in regions corresponding to bright streaking artefacts, LI is outperforming the rest of selected techniques. It is less pronounced in terms of mean relative error, but the standard deviation (error bars) clearly illustrates the superior performance of this simple interpolation-based algorithm (figure 2.a). The 2D interpolation-based approach takes the second place in terms of standard deviation of the mean relative error. In terms of NRMSD and MAD, however, its performance is not considerably different from that achieved by the three other algorithms. NMAR appears to fall behind other methods in correcting the bright streaking artefacts.

The performance of all five selected methods is approximately similar in artefact-free regions. MRE and NRMSD remain within about $\pm 5\%$ and deviations of intensities are less than 50 HU, which is quite acceptable. 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 artefacts and artefact-free regions. On the other hand, all techniques seem to have a poor performance in dark streaking areas. Nevertheless, this analysis is based on only 10 simulated datasets and the

significance might be influenced by the low number of samples.

The results of the clinical studies enabled to observe that, except the rectum, which is almost never influenced by streaking artefacts and as such is ranked close to 5 in all image sets, the other organs have been scored approximately between 3 and 4 before correction (figure 4). 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 artefacts. In this analysis, the LI algorithm does not outperform the other techniques, neither in the five different organs nor the overall image quality. However, in the simulated studies we noticed that LI is superior to other methods in regions corresponding to bright streaking artefacts. One explanation for this inconsistency is that the strong bright streaking artefacts occur in close vicinity of the implant, and these areas were specifically analyzed for the simulated datasets. However, the five organs analyzed for the clinical studies are rather distant from the metallic object and are thus more influenced by the dark streaking artefacts. The MDT technique has shown the lowest performance using both simulation and clinical studies. LI and 2D approaches achieve an intermediate ranking among other approaches.

The inconsistent 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. Moreover, the fact that a simple interpolation scheme, such as LI, is capable of outperforming the other techniques only in specific areas in the vicinity of the implants indicates that such simplistic approaches are not the right choice for general purposes. 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 artefacts. 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.

The scope of this study was limited to metal hip implants. Further investigation of other common sources of streaking artefacts, 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 artefacts and, as such, this work is deemed to be a proper representative sample of various sources of metallic artefact.

Another known limitation of this work is that artificial sinograms are used rather than the original sinograms produced by the CT scanners. Although previous studies demonstrated that the use of virtual sinograms does not bring

in significant bias in the outcome of MAR techniques [1, 8, 9], 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 power of the study.

REFERENCES

- [1] M. Abdoli, R. A. J. Dierckx, H. Zaidi, "Metal artifact reduction strategies for improved attenuation correction in hybrid PET/CT imaging", *Med Phys.* vol. 39, no.6, pp. 3343-60, 2012.
- [2] G. H. Glover, N. J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT reconstructions", *Med Phys.* vol. 8, no. 6, pp. 799-807, 1981.
- [3] T. Hinderling, P. Ruegsegger, M. Anliker, C. Dietschi, "Computed tomography reconstruction from hollow projections: an application to in vivo evaluation of artificial hip joints", *J Comput Assist Tomogr.* vol. 3, no. 1, pp. 52-7, 1979.
- [4] C. Xu, F. Verhaegen, D. Laurendeau, S. A. Enger, L. Beaulieu, "An algorithm for efficient metal artifact reductions in permanent seed" *Med Phys.* vol. 38, no. 1, pp. 47-56, 2011.
- [5] W. J. H. Veldkamp, R. M. S. Joemai, A. J. van der Molen, J. Geleijns, "Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT" *Med Phys.* vol. 37, no. 2, pp. 620-8, 2010.
- [6] W. A. Kalender, R. Hebel, J. Ebersberger, "Reduction of CT artifacts caused by metallic implants", *Radiology*, vol. 164, no.2, pp. 576-577, 1987.
- [7] M. Bazalova, L. Beaulieu, S. Palefsky, F. Verhaegen, "Correction of CT artifacts and its influence on Monte Carlo dose calculations" *Med Phys.* vol. 34, no. 6, pp. 2119-32, 2007.
- [8] M. Abdol, M. Ay, A. Ahmadian, R. Dierckx, 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.* vol. 37, no. 12, pp. 6166-77, 2010.
- [9] M. Abdoli, J. R. de Jong, J. Pruijm, R. A. Dierckx, 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.* vol. 38, no. 12, pp. 2257-68, 2011.
- [10] P. Delaunay, "Sur la sphère vide", *Bulletin of Academy of Sciences of the USSR.* vol. 6, pp. 793-800, 1934.
- [11] R. Clough, J. Tocher, "Finite element stiffness matrices for analysis of plates in blending", *Proceedings of Conference on Matrix Methods in Structural Analysis.* Wright-Patterson A.F.B., p. 515-45, Ohio 1965.
- [12] E. Meyer, R. Raupach, M. Lell, B. Schmidt, M. Kachelriess, "Normalized metal artifact reduction (NMAR) in computed tomography", *Med Phys.* vol. 37, no. 10, pp. 5482-93, 2010.
- [13] A. Mehranian, M. R. Ay, A. Rahmim, H. Zaidi, "3D prior image constrained projection completion for x-ray CT metal artifact reduction", *IEEE Trans Nucl Sci.* vol. 60, no.5, pp. 3318-3332, 2013.
- [14] F. E. Boas, D. Fleischmann, "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography", *Radiology.* vol. 259, no. 3, pp. 894-902, 2011.