Image quality of dual-energy cone-beam CT with total nuclear variation regularization

Lukas Schröder, Uros Stankovic, Simon Rit, Jan-Jakob Sonke

Image quality of dual-energy cone-beam CT with total nuclear variation regularization

Lukas Schröder[1], Uros Stankovic[1], Simon Rit[2], Jan-Jakob Sonke[1]*

[1]Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[2]Univ Lyon INSA-Lyon Universite Claude Bernard Lyon 1, UJM-SaintEtienne CNRS, Inserm CREATIS UMR 5220, U1206,F-69373, Lyon, France

*Corresponding author: j.sonke@nki.nl

Abstract

Despite the improvements in image quality of cone beam computed tomography (CBCT) scans, application remains limited to patient positioning. In this study, we propose to improve image quality by dual energy (DE) imaging and iterative reconstruction using least squares fitting with total variation (TV) regularization.

The generalization of TV called total nuclear variation (TNV) was used to generate DE images. We acquired single energy (SE) and DE scans of an image quality phantom (IQP) and of an anthropomorphic human male phantom (HMP). The DE scans were dual arc acquisitions of 70kV and 130kV with a variable dose partitioning between low energy (LE) and high energy (HE) arcs. To investigate potential benefits from a larger spectral separation between LE and HE, DE scans with an additional 2 mm copper beam filtration in the HE arc were acquired for the IQP. The DE TNV scans were compared to SE scans reconstructed with FDK and iterative TV with varying parameters. The contrast-to-noise ratio (CNR), spatial frequency, and structural similarity (SSIM) were used as image quality metrics.

Results showed largely improved image quality for DE TNV over FDK for both phantoms. DE TNV with the highest dose allocation in the LE arm yielded the highest CNR. Compared to SE TV, these DE TNV results had a slightly lower CNR with similar spatial resolution for the IQP. A decrease in the dose allocated to the LE arm improved the spatial resolution with a trade-off against CNR. For the HMP, DE TNV displayed a lower CNR and/or lower spatial resolution depending on the reconstruction parameters. Regarding the SSIM, DE TNV was superior to FDK and SE TV for both phantoms. The additional beam filtration for the IQP led to improved image quality in all metrics, surpassing the SE TV results in CNR and spatial resolution.

Keywords: cone beam CT, image reconstruction, dual-energy, total nuclear variation, image quality

1. Introduction

Recent years have seen significant improvements in the image quality of the cone-beam computed tomography (CBCT) scans [1-3]. Despite that, CBCT scans are mostly used for patient positioning [4-6] and, to some extent, for dose accumulation using deformed planning CT [7]. While some advanced applications, such as dose calculations on CBCT, could be feasible now [8], the image quality of CBCT still needs to be improved to unlock the full potential of adaptive radiotherapy (ART). Some of these

applications are for example image segmentation and automatic contour propagation. In these cases, a good contrast-to-noise ratio (CNR) and the ability to discriminate between low contrast tissues is required.

Dual-energy (DE) imaging aims to exploit the energy dependence of X-ray attenuation. Different scanning energies lead to different attenuations and therefore to different images. The difference in attenuation between various materials defines their contrast. Since the change of attenuation from one energy to another varies for different materials, and a lower energy leads to higher attenuation in the energy range typically used for (CB)CT imaging (60kV-140kV), the contrast increases with decreasing energy. The disadvantage is an increased noise level for the same imaging dose due to fewer photons reaching the detector. In single energy (SE) imaging, the beam energy is selected to maximize the CNR for a constant imaging dose. Low energy (kV setting) imaging has a high soft tissue contrast but also high noise. Its CNR is inferior to some higher energy representing the optimum for the same imaging dose. Therefore, the idea of using dual energy (DE) imaging to combine the advantages of different energies has been around for a long time [9, 10] and it might lead to better image quality and dose calculation accuracy [11]. An important concept for DE imaging is the spectral separation. An x-ray tube does not emit a mono-energetic beam but a spectrum. Different kV settings at the x-ray tube generate different spectra. In between these spectra, a number of particles have the same energy i.e. there is a spectral overlap. These particles have the same attenuation, hence they provide the same information. To increase the benefit of dual-energy imaging, the overlap should be minimized [12].

Previous research on DE imaging has focused on the linear mixing of the low and high energy acquisitions using conventional filtered back-projection (FBP) [13], but no clear benefit has been shown yet. A common alternative to the FBP based algorithms are the iterative reconstruction algorithms. They have shown to be very well suited for the reconstruction of low-dose and undersampled images [14,15]. Iterative reconstruction requires a regularization term to converge to a plausible solution. A common regularization term for iterative reconstruction is total variation (TV) [16]. One advantage of TV is that it preserves edges while decreasing noise. A second one is that TV can produce accurate reconstructions from under-sampled data [17].

In order to extend the concept of TV based iterative reconstruction to DE CBCT, the TV regularization term needs to be generalized to spectral data. Holt [18] as well as Rigie and Riviere [19] have proposed total nuclear variation (TNV) and have shown its favourable properties for spectral CT reconstruction. The TNV correlates the different energy channels and therefore considers the similarities between the channels, e.g. edge locations. It has a shared-direction bias, which means that TNV favours the gradients of the channels to point in the same direction.

In this study, we investigated the combination of DE with TNV for CBCT, making it an extension of the work of Rigie and Riviere [19] to CBCT. We evaluated the image quality in terms of CNR, spatial resolution, and structural similarity index measure (SSIM), and compared it to single-energy (SE) FDK and iterative reconstruction with total variation (TV).

2. Materials and Methods

2.1 Iterative reconstruction

Iterative reconstruction often minimizes a cost function consisting of a data fidelity and one or more regularization terms. Minimizing solely the data fidelity term

$$\arg\min_f \|\mathrm{R}f - p\|_2^2 \tag{1}$$

where R is the forward projector, f the image, and p the measured projections, leads to non-plausible images at convergence and regularization is therefore required. As regularization term, total variation (TV) was chosen. TV has been broadly studied and used for (CB)CT reconstruction [15, 20-22], where the gradient sparsity of CT images is exploited. We define the isotropic TV of an image f as the mixed l_1/squared l_2-norm, which means that the l_1-norm over space (spatial indices: x, y, z) of the squared l_2-norm over the gradient of the image is taken:

$$\mathrm{TV}(f) = \|\nabla f(x,y,z)\|_{1,2} = \sum_{x,y,z} \|(\nabla f)(x,y,z)\|_2 \,. \tag{2}$$

Since we want to study DE scans, we have to reconstruct 4D (spectral) data with the energy at the fourth dimension. A straightforward way to generalize TV to spectral data is the use of the Jacobian J and a suitable norm. Holt investigated several possibilities and concluded that the Schatten 1-norm, called the nuclear norm (NN), yields the desired properties and the best image quality [18]. The Schatten p-norm of a matrix **A** is the l_p-norm of the vector of the singular values (SV) σ of **A**. For the NN, this means that it is the sum of its SV

$$\|\mathbf{A}\|_* = \|\vec{\boldsymbol{\sigma}}_{\boldsymbol{A}}(x,y,z)\|_1 \,, \tag{3}$$

and the resulting total nuclear variation (TNV) is

$$\mathrm{TNV}(\boldsymbol{f}) = \|\mathrm{J}\,\boldsymbol{f}(x,y,z)\|_{1,*} = \sum_{x,y,z} \|(\mathrm{J}\,\boldsymbol{f})(x,y,z)\|_* \,. \tag{4}$$

Compared to a simple channel-by-channel TV, this approach couples the different energy channels and for one channel, TNV reduces to the isotropic TV. Because the NN is the best convex approximation of rank(J $\boldsymbol{f}(x,y,z)$), the TNV favours rank-sparsity in J $\boldsymbol{f}(x,y,z)$. Hence, the TNV is not only channel- and direction-coupled (ratio between directional derivatives in different spatial directions is constant over space) but also has a shared direction (gradients for all channels all point in the same directions) due to the rank minimization [18]. In other words, the (gradient-)coupling is achieved by promoting common edges and alignment of their gradients [19]. Therefore it is more advantageous than e.g. the Frobenius norm (Schatten 2-norm). A more detailed description of the NN and other TV generalizations and their properties can be found in Holt [18].

2.2 Reconstruction algorithms

All reconstructions were performed using the Reconstruction toolkit (RTK v.1.4) [23]. For the SE and DE, the 4Drooster (4D RecOnstructiOn using Spatial and TEmporal Regularization) algorithm was used,

which is based on the work of Mory et al. [24] and includes spatial (3D) and temporal TV. Compared to the original algorithm described in Mory et al. [24], RTK allows disabling the temporal TV regularization and enabling TNV. In the latter case, spatial TV was disabled as well. The 4Drooster algorithm sequentially solves different problems. In every main loop iteration k, the data fidelity term (Eq. 1) is (channel-wise) minimized first using the conjugate gradient method (CG) with n=10 iterations. This results in $f_0^{(k)}$ for SE. For DE, all spectral components are reconstructed separately generating $\mathbf{f_0^{(k)}}=(f_{0,LE}^{(k)}, f_{0,HE}^{(k)})$. Here and onwards, bold letters indicate 4D spectral data. Afterwards, a non-negativity constraint is applied which sets all negative voxels to 0. For SE scans, the new image was then computed in m iterations of TV denoising, i.e.

$$f^{(k)} = \arg\min_f \|f - f_0\|_2^2 + \theta\, \mathrm{TV}(f)\,. \tag{5}$$

For DE scans, the new image was regularized with m iterations of TNV denoising, i.e.:

$$\boldsymbol{f}^{(k)} = \arg\min_{\boldsymbol{f}} \|\boldsymbol{f} - \boldsymbol{f_0}\|_2^2 + \theta\, \mathrm{TNV}(\boldsymbol{f})\,. \tag{6}$$

resulting in a 4D image with a low energy (LE) and a high energy (HE) component $\mathbf{f^{(k)}} = (f_{LE}^{(k)}, f_{HE}^{(k)})$. To solve the minimization, both TV and TNV denoising employed the basis pursuit dequantization algorithm [25]. The parameter θ regulates the denoising of the image, the larger θ, the smoother the image. Initially, three values were experimentally chosen so that with the minimum value ($\theta = 10^{-4}$) an effect of the total (nuclear) variation term is apparent and so that higher values than the maximum ($\theta = 10^{-3}$) do not show large improvements in image quality. For a subset of the acquisitions, reconstructions with more θ values were necessary to get a more accurate image quality comparison. The maximum number of main loop iterations N was chosen so that a near convergence was reached. To limit the parameter space, the number of CG and TV/TNV iterations was kept constant at n=m=10. Furthermore, the two 4Drooster reconstructions, SE with TV and DE with TNV, are denoted as SE TV and DE TNV. The selected parameters of the RTK algorithms are shown in Table 1. In RTK, the number of iterations has to be pre-defined before reconstruction. An exemplary command line for every utilized RTK algorithm is provided in the supplementary material (Appendix A).

Table 1 Utilized RTK algorithms and their parameters

RTK algorithm	Scans	Parameters		
		Name	**Symbol**	**Values**
FDK	SE	Cut-off frequency (Hann window) of the 1D ramp filter relative to the Nyquist frequency (unitless)	h	0.1, 0.2 …1.0, No window
4DRooster	SE/DE	Number of main loop iterations	N	1,2,…10
		Number of CG iterations	n	10
		Number of TV/TNV iterations	m	10
		TV/TNV weighting	θ	…,10^{-4}, …., 5x10^{-4}, 10^{-3}

2.3 CBCT acquisition

Since it was not possible to perform experimental DE scans of patients, a CBCT Electron Density & Image Quality Phantom (IQP) (Model 062MQA) and a Virtually Human Male Pelvis Phantom (HMP) (Model 801-P) (CIRS Inc., Norfolk, Virginia, USA) were scanned on a linac integrated CBCT scanner (Synergy, XVI 5.0, Elekta Ltd, Crawley, UK), augmented with in-house developed software, utilizing a bow-tie filter and a fibre-interspaced anti-scatter grid (ASG) (N44r15, Philips Medical Systems, Best, The Netherlands) [1]. The ASG had a line frequency of 44 cm^{-1} and a grid ratio of 15:1. A head and neck (H&N) and a pelvis phantom configuration of the IQP were scanned with a field-of-view (FOV) of 25.6 cm (small FOV) and 40 cm (medium FOV) in the radial direction, respectively. The pelvis phantom was oval with a width (left-right) of 33cm and a height (anterior-posterior) of 27cm. The H&N phantom consisted of a circular head part with a diameter of 18cm and a shoulder part with the same extent as the pelvis phantom. For the HMP, the medium FOV was utilized. According to clinical protocols for H&N and pelvis tumours, the scans were acquired with a 360° gantry rotation and gantry speeds of 1 rpm and 0.5 rpm for the small and medium FOV, respectively.

The DE scans were acquired in a dual arc (DA) with changed parameters (kV, mA, ms) between the arcs. The settings were performed in such a way that the combined dose of LE and HE acquisitions best matched but preferably not exceed the dose resulting from commonly used clinical protocols using 120kV. The LE and HE were 70kV and 130kV, which are the lowest and highest enabled kV generator settings. The larger the difference between the two energy spectra, the more complementary information (attenuation) the different energy acquisitions provide, leading to a higher benefit of using DE. To determine if a larger spectral separation between the LE and HE improves image quality, HE acquisitions of the pelvis IQP phantom were also performed with an additional 2 mm copper filtration (+Cu).

2.4 CT dose index

We used the CT dose index (CTDI) [26] to determine the imaging dose (CTDI phantom by PTW, Freiburg, Germany). The weighted CTDI ($CTDI_w$) is calculated as:

$$CTDI_{\mathrm{W}} = \frac{1}{3} CTDI_{\mathrm{centre}} + \frac{2}{3} CTDI_{\mathrm{periphery}}\,, \tag{7}$$

where the $CTDI_{centre}$ is the CTDI value in the centre of the phantom and $CTDI_{periphery}$ the average of the CTDI values at the four peripheral locations of the CTDI phantom. In our study, the $CTDI_w$ was determined for the two phantom sizes (H&N, a cylinder with 16 cm diameter and pelvis, cylinder with 32 cm diameter) and three energies, LE of 70kV, HE of 130kV, and clinical (medium, ME) 120kV. For the pelvis 130kV, the $CTDI_w$ was also determined with the added Cu filtration. After testing the repeatability of the dose measurements and verification of linearity (R^2>0.999) of the $CTDI_w$ - mAs dependence (Appendix B), a linear function for each setting was established with only one measurement (Fig. B.1). These linear functions were applied to determine the mAs for the low- and high-energy acquisitions (Appendix B).

2.5 Measurements

For the SE acquisitions, 70kV (SE_{70}), 120kV (SE_{120}), and 130kV (SE_{130}) were used. To find the optimal image quality for DE, the dose was partitioned differently between LE and HE. The reference dose was the clinically used SE_{120}. Due to technical limitations, i.e. allowed mA-ms settings and heating of the x-ray source, the possible proportions of LE and HE varied between phantom sizes and the dose deviations were $D(SE_{120kV}) \pm 13\%$. In an additional experiment, the option of halving the number of projections but doubling the mAs for every projection was investigated for possible combinations. Hereby, every second projection was discarded to keep a regular angular sampling over the 360° rotation and the same $CTDI_w$. Ignoring image lag characteristics, this approach can be seen as a substitute for a fast kV-switching (FS) method where the gantry rotation speed was kept the same. Discarding half of the projections might lead to undersampling. Therefore, we also investigated how well TNV is able to handle undersampling.

2.6 Reconstruction

Before reconstruction, the projections were pre-processed. This pre-processing included several steps:

1. Multi-angle gain correction to prevent grid line artefacts.
2. Down-sampling from 512x512 to 256x256 as routinely done in clinical practice.
3. As the H&N phantom configuration had a shoulder mimicking part in the caudal region extending outside the FOV which is incompatible with the iterative reconstruction data fidelity term, the projections for this configuration were cropped accordingly to mask phantom parts extending outside of the small FOV.

The dimensions of the reconstructed 3D (4D for DE) images were 400x400x256(x2) for the pelvis IQP and HMP (medium FOV) and 256x256x133(x2) for the H&N IQP. The dimension in the cranial caudal direction was reduced from 256 to 133 for the H&N IQP due to the cropping. The voxel size was $1x1x1mm^3$.

2.7 Image quality evaluation

Due to the nature of iterative algorithm, performing a trade-off between the edge sharpness and noise, the image quality was investigated through the relation between the contrast-to-noise ratio and spatial resolution. The CNR was calculated according to the following formula:

$$CNR = \frac{|\overline{CT}_i - \overline{CT}_s|}{\sqrt{\frac{1}{2}(\sigma_i^2 + \sigma_s^2)}}, \tag{8}$$

where $\overline{CT}_i$ and $\overline{CT}_s$ represent the mean CT numbers of cylindrical volumes of interest (VOI) with a diameter and height of 1.8cm. For the IQP, the VOI were placed in the linearity inserts (i), ranging from air to Teflon, and in their water equivalent surroundings (s). The σ are the standard deviations of these VOIs. For the HMP, cylindrical VOIs with a diameter and height of 1.2cm were placed in two neighbouring, distinguishable areas, with a contrast of 104 HU on a conventional CT. These two VOIs represented insert and surrounding (Supplementary material Fig. C.1 A).

As a measure for the spatial resolution, the spatial frequency was used. For the IQP, a region of interest (ROI) of a central slice of the phantom containing the Teflon insert was chosen. A 2D cumulative normal distribution (CND) representing a 2D edge response function (ESF) was fitted to this ROI:

$$ESF = A - B * CND\left(\sqrt{(x-x_0)^2 + (y-y_0)^2}, \mu, \sigma\right) + LinRegX * x + LinRegY * y \tag{9}$$

Here, x and y are the coordinates of the voxels in the ROI, and μ and σ are the mean and the standard deviation of the CND. Additionally, the fitting function was expanded with linear regression terms in x and y direction to take non-uniformities ("cupping artefacts") into account. An example is shown in the Supplementary material Fig. C.2 A. The derivative of the ESF yields the line spread function (LSF), and with the Fourier transform, the LSF can be used to calculate the modulation transfer function (MTF). Therefore, as derived by Krah et al. [27], the standard deviation of the ESF (σ) can be used to calculate the spatial frequency $f_{10\%}$ at which the MTF falls below 10% as:

$$f_{10\%} = \frac{\sqrt{2\,ln10}}{2\pi\sigma}, \quad \text{with} \quad \frac{|MTF(f_{10\%})|}{|MTF(0)|} = \frac{1}{10}, \tag{10}$$

On a CT scan with typical clinical acquisition and reconstruction parameters, the $f_{10\%}$ was 7.8cm^{-1} for the IQP pelvis configuration. The slice thickness of the CT was 1.5mm but since the inserts are directed in the slice direction, the influence of the difference in slice thickness can be neglected. For the HMP, an area of 1cm x 2.5cm across the border of two different tissues was evaluated (Supplementary material Fig. C.1 B). After averaging in the 1cm direction, a 1D ESF similar to Eq. 9 was fitted (Supplementary material Fig. C.2 B). On a CT scan of the HMP, $f_{10\%}$ was 6.8cm^{-1} but the slice thickness was 2mm. Compared to the 1mm of the CBCT, the spatial resolution was, therefore, underestimated because the border between the two tissues is not exactly oriented in the slice direction. The image analysis was performed with Python (version 3.8.2) and the SciPy package was utilized for the fit computation [28].

The size and distance of the VOIs were chosen such that the CNR measurements were only impacted by poor spatial resolution. This occurred for images with an $f_{10\%}$ below ~1.1-1.5cm^{-1}. Then, the blurring of the images was so pronounced that the insert blurred into the VOI of the background. Therefore, only results with an $f_{10\%}$ above 1.1cm^{-1} were evaluated. For the DE scans, both the LE and HE component were evaluated.

For the pelvis configuration of the IQP and the HMP, the SSIM was evaluated as a third image quality metric. With an equal weighting of 1 for luminance, contrast, and structure, the SSIM of image A and B can be expressed as:

$$SSIM(A,B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)} \tag{11}$$

Here, μ and σ are the mean and the standard deviation of the respective image and σ_{AB} their covariance. For the variables c_1 and c_2, the default values of $(0.01*L)^2$ and $(0.03*L)^2$ were used, respectively, with L as the dynamic range of the images. As ground truth images, diagnostic CT scans (120kV) of the phantoms were utilized.

3. Results

For the IQP, the polystyrene insert had the lowest CNR of all inserts and therefore, these results are shown. The behaviour of the other inserts was similar with only varying CNR. For the IQP as well as for the HMP, the spatial resolution improved with increasing number of iterations, although the improvements for the last iterations were very small because convergence was almost reached. The behaviour of the CNR depended on the parameter θ. Generally, the CNR dependence on the number of iterations reduced with increasing θ. More specifically, for lower θ (θ=10^{-4}), the CNR usually decreased with increasing N. This effect was reduced with increasing θ and resulted in a CNR increase for higher θ (θ=10^{-3}). For clarity, only the results for N=10 are shown.

3.1. CBCT Electron Density & Image Quality Phantom

3.1.1 Single energy TV and dual energy TNV

The direct comparisons between the best SE and selected DE results for H&N and pelvis phantom configuration are shown in Fig. 1. In both cases, the TV and TNV clearly outperformed the clinically used FDK algorithm. They generated images with CNRs close to the best FDK results but with a much higher spatial resolution. For the H&N configuration, DE TNV yielded a bit lower CNRs than SE TV with similar spatial resolutions. For the pelvis configuration without additional copper filtration, the results are alike with only showing a slightly lower spatial resolution for DA with θ=$5x10^{-4}$. Reconstruction examples for the pelvis configuration for FDK, SE TV and DE TNV are shown in Fig. 2 and for the H&N configuration in the Supplementary material (Fig. D.1). Since they are all 70kV images, their contrasts are the same and therefore, the CNR differences are caused by the different noise characteristics. This is clearly visible in the elevated noise of the FDK reconstruction (Fig. 2A and Fig. D.1A).Comparisons between SSIM and spatial resolution for LE and HE component of the pelvis configuration are shown in Fig. 3. Here as well, the best SSIM results are shown for the different reconstruction methods. The behaviour of the SSIM was very similar to the CNR with the difference that the best DE results ($FS_{20\%/80\%}$) surpassed the SE TV results and for HE, additionally, all FDK results.

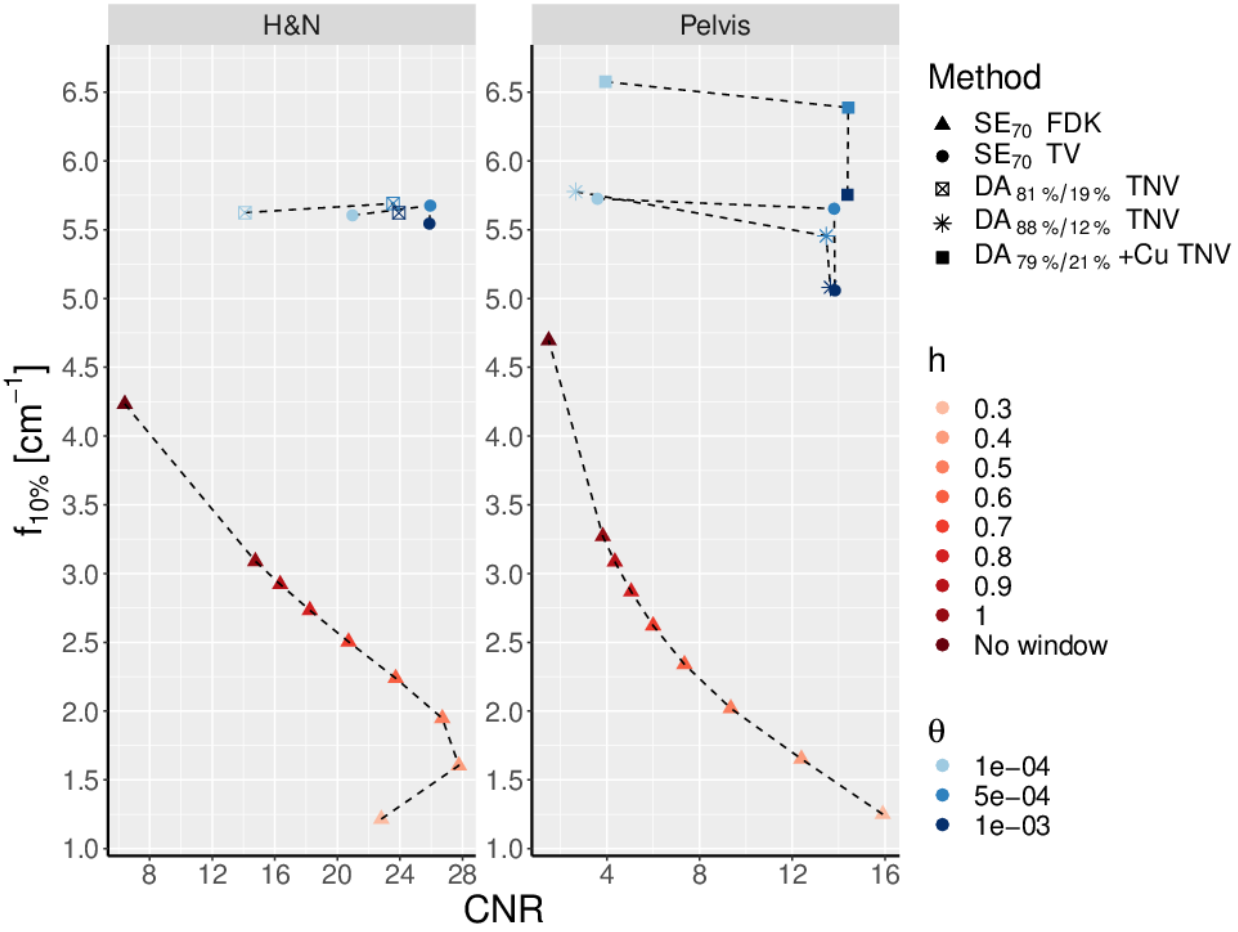

Figure 1 CNR and spatial frequency for the best scans of FDK, SE TV, and DE (DA) TNV (LE component, N=10) for the H&N (left) and pelvis (right) phantom configuration. For the pelvis (right) phantom configuration, an example with additional beam filtration is displayed as well.

For SE, the LE yielded the best results for FDK and TV. The overall image quality of the FDK results was lower than the iterative reconstructions with the parameter h controlling the trade-off between spatial resolution and CNR. SE_{70} TV (θ=5x10^{-4}) resulted in a higher spatial resolution while reaching a high CNR for both phantom configuration. A visualization of all SE results can be found in the Supplementary material (Fig. E.1).

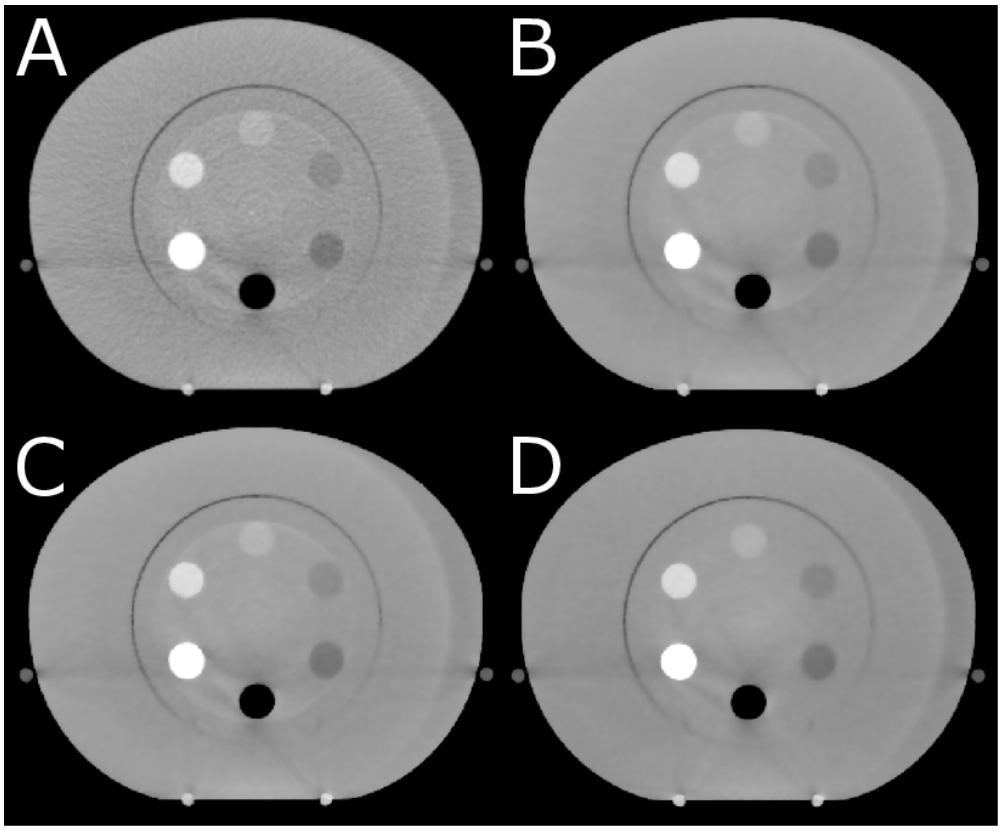

Figure 2 Reconstruction examples of the pelvis configuration of the image quality phantom for (A) FDK (SE_{70}, h=1.0), (B) SE_{70} TV (N=10, θ = 5x10^{-4}), (C) DE TNV ($DA_{88\%/12\%}$, LE component, N=10, θ = 10^{-3}), and (D) DE +Cu TNV ($DA_{79\%/21\%}$, LE component, N=10, θ = 5x10^{-4}). The window (393 HU) and level (1343 HU) were kept constant for all images.

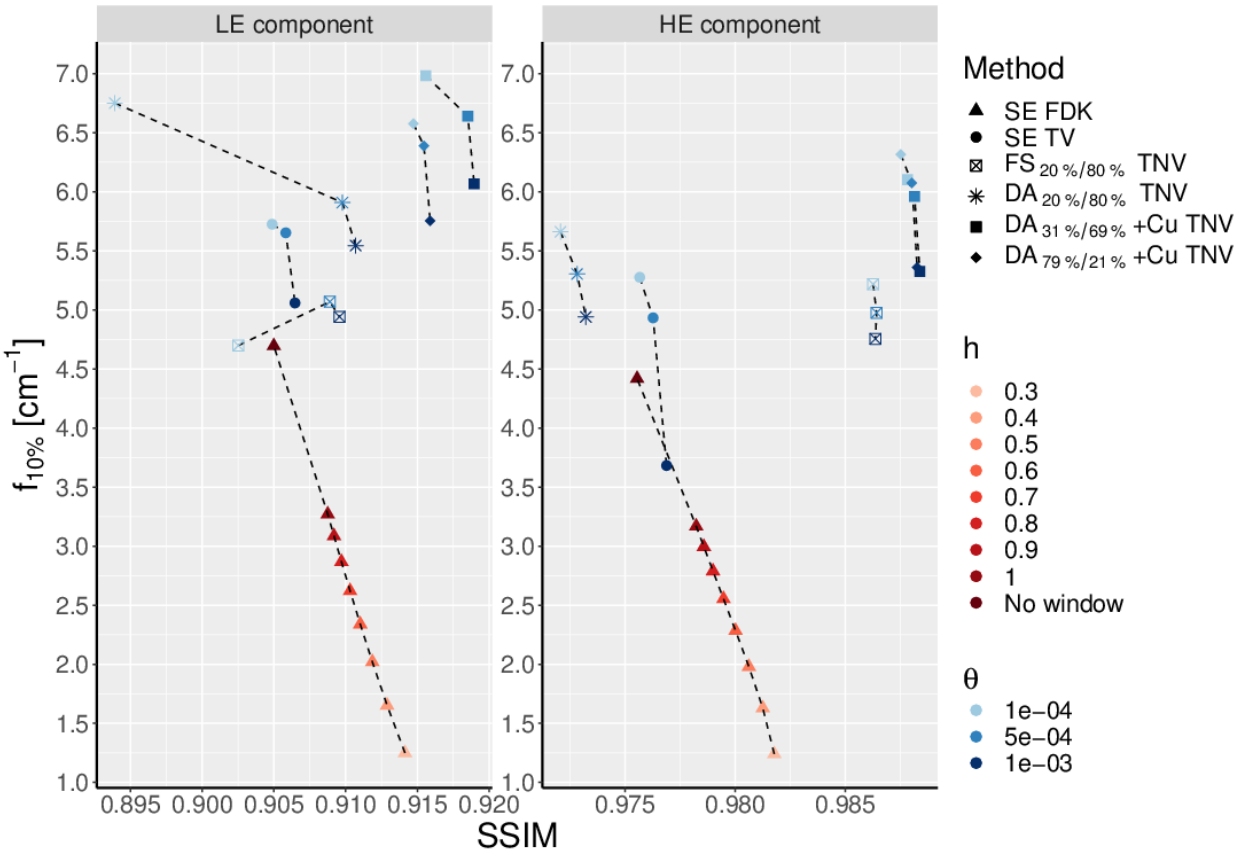

Figure 3 SSIM and spatial frequency for the best pelvis IQP scans of FDK, SE TV, and DE TNV (N=10). LE (left) and HE (right) result in considerable SSIM differences, which is visible in the different axes. For DE TNV, the best results with and without additional beam filtration are displayed.

The DE TNV results in Fig. 1 were obtained with the highest dose in the LE arc as it provided superior CNR results to other dose allocations. More specifically, for the LE component, the CNR increased with increasing dose allocated in the LE arc. Therefore, the best CNR was obtained with the highest dose in the LE arc. This corresponds to a dose partitioning of 81%/19% for the H&N and 88%/12% for the pelvis phantom configuration. Similar behaviour was observed in the HE component even though to a lesser extent. Generally, the larger the dose allocation in the HE arc, the better the CNR. However, the LE component demonstrated CNRs about twice as high as the HE component. Mostly, the CNR increased with increasing θ. This increase was considerable from θ=1x10^{-4} to θ=5x10^{-4} but much smaller from θ=5x10^{-4} to θ=1x10^{-3}. In some cases, the CNR decreased slightly. For the pelvis configuration, the spatial resolution improved with decreasing θ but for the H&N configuration the spatial resolution decreased from θ=5x10^{-4} to θ=1x10^{-4}. In both configurations, the LE component demonstrated slightly higher spatial resolution. The results for all dose allocations can be found in the Supplementary material (Fig. E.2-E.4).

3.1.2 Additional beam filtration

For the pelvis IQP configuration, the effect of additional 2mm Cu filtration was evaluated. Such filtration increased both CNR and resolution as illustrated in Fig. 1. With a dose allocation of 79%/21%, the added filtration resulted in a spatial resolution above the one of SE TV and also in a slightly larger CNR.

Consequently, it also surpassed the best results of the pelvis phantom configuration without Cu filtration. Adding the Cu filtration also increased the SSIM for both energies (Fig. 3), bringing the DE SSIM beyond all FDK SSIMs for LE as well. A reconstruction example for DE TNV with added Cu filtration is shown in Fig. 2D. The results for all dose allocations can be found in the Supplementary material (Fig. E.5 and Fig. E.6).

3.2.2. Fast-switching

For the H&N phantom configuration, the FS-like acquisition scheme decreased image quality primarily by a decrease in CNR (Fig. 4 a). The spatial resolution was similar for FS and DA acquisition but the spatial resolution decrease for θ=10^{-4} was amplified. For the pelvis phantom configuration, the FS-like acquisition scheme slightly increased the CNR (Fig. 4 b). For a 39%/61% dose allocation, the spatial resolution improved as well. For a lower dose partition in the LE arm (20%/80%), the spatial resolution decreased for the FS. The results for both image components from both phantom configurations are visualized in the Supplementary material (Fig. E.2 and Fig. E.3). Regarding the SSIM, FS led to a small decrease in the LE component in most cases but to a much larger increase for the HE component. The SSIM results are displayed in Fig. E.4.

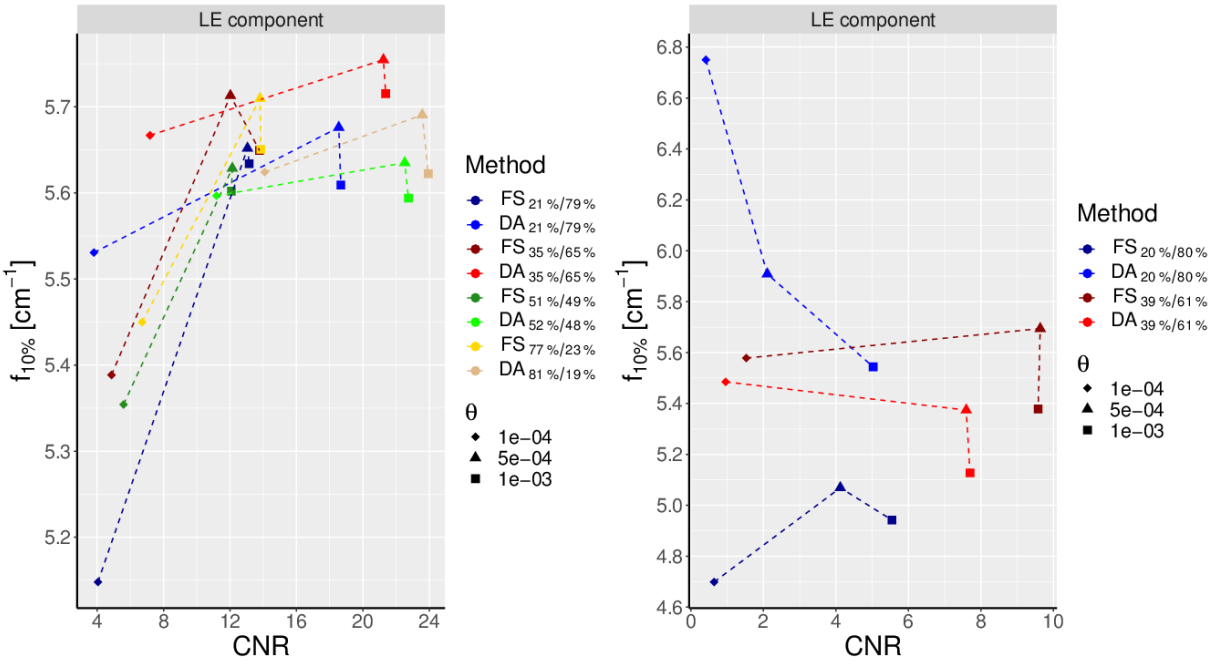

Figure 4 CNR and spatial frequency for DE TNV (LE component) with dual-arc (DA) and fast-switching (FS) acquisition scheme for the H&N (a) and pelvis (b) phantom configuration. Fig. E.2 sand Fig. E.3 show both the LE and HE component for H&N and pelvis phantom configuration, respectively.

3.2 Virtually Human Male Pelvis Phantom

Most findings from the IQP were also observed in the HMP, showing that those findings also hold in a human anatomy with more variation and more dense material. These include the superiority of TV and TNV over FDK, the θ dependent behaviour of the DE image quality, and that a higher dose allocation in

the respective energy arm resulted in higher CNR for DE TNV. This effect was also here larger in the LE component. Furthermore, DE TNV and SE TV showed similar image quality. The spatial resolution of the HMP was in the same range as the IQP. In Fig. 5, the best FDK, SE TV, and DE TNV results for CNR and spatial resolution are compared. Here, θ was varied from 0 to 10^{-3} to allow a more extensive comparison between SE TV and DE TNV. Reconstruction examples for FDK, SE_{120} TV, and DE TNV are shown in Fig. 6. The best SSIM results for all reconstruction methods are displayed in Fig. 7. As for the IQP, FS usually generated superior SSIMs than DA and the DE SSIM is larger than for FDK and SE TV, but contrary to the IQP, here also for the LE. The SSIM results for all dose allocations can be found in the Supplementary material (Fig. E.7). Further differences to the IQP are, that, the best SE CNR results were seen for images using 120kV (cf. Fig. E.8). Additionally, the LE and HE components of the HMP showed similar CNR results, which were comparable to the values of the IQP ones (pelvis) (Fig. 8). In the LE component, the CNRs for the various dose allocations were more similar and not as widespread as for the IQP. For a dose allocation of 39%/61%, FS slightly increased CNR as for the IQP but for a dose allocation of 19%/81%, the increase in the LE component was large.

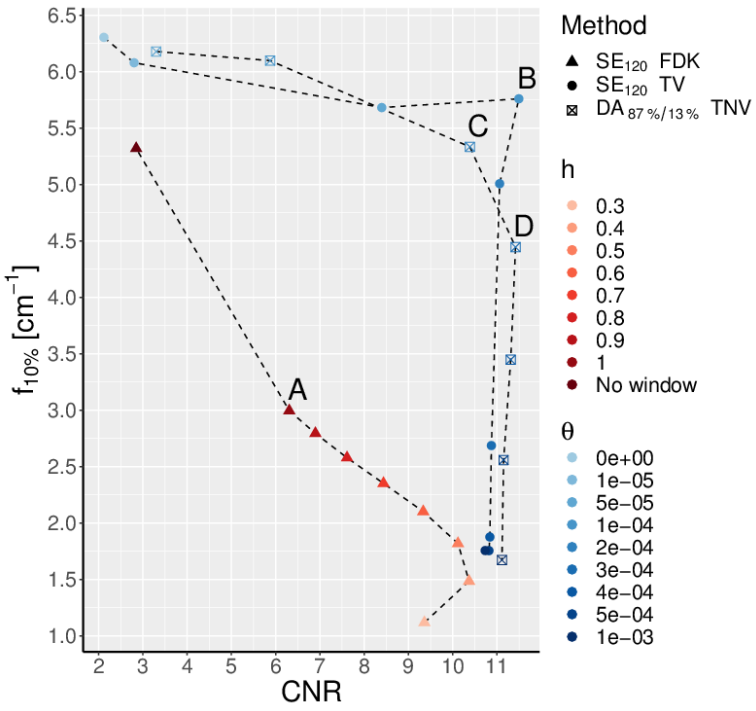

Figure 5 CNR and spatial frequency for the best scans of FDK, SE TV, and DE (DA) TNV for the virtual human male pelvis phantom. The letters A-D indicate the image reconstructions shown in Fig. 6.

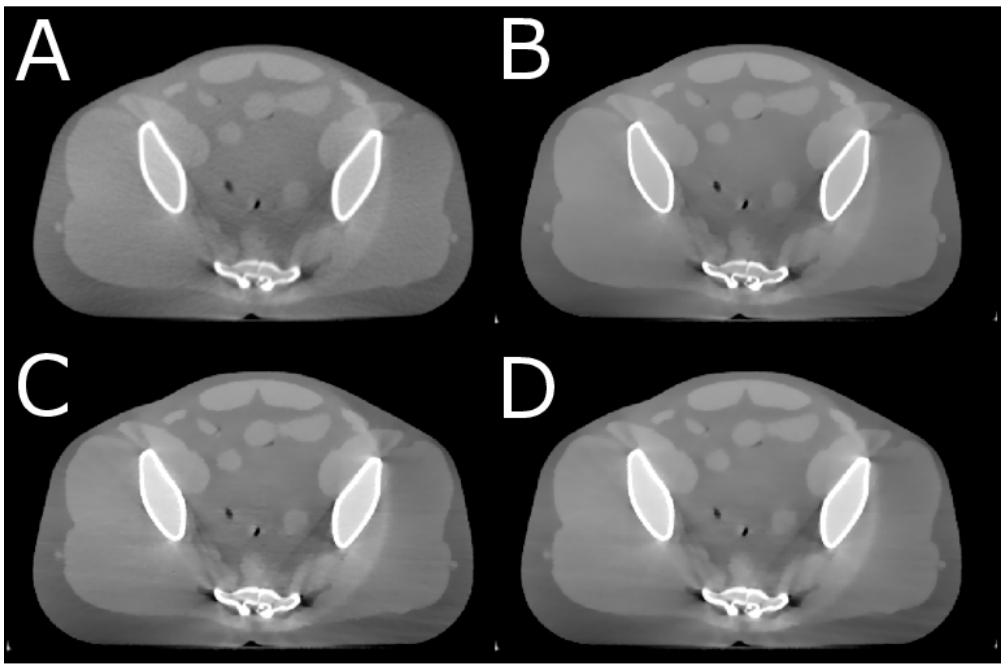

Figure 6 Reconstruction examples of the virtual human male pelvis phantom for (A) SE_{120} FDK (h=1.0), (B) SE_{120} TV (N=10, θ=1x10^{-4}), and DE TNV ($DA_{87\%/13\%,,}$ LE component, N=10) with (C) θ=2x10^{-4} and (D) θ=3x10^{-4}. The window and level of the DE images were kept constant but changed by means of histogram matching for the SE images due to differing attenuation values.

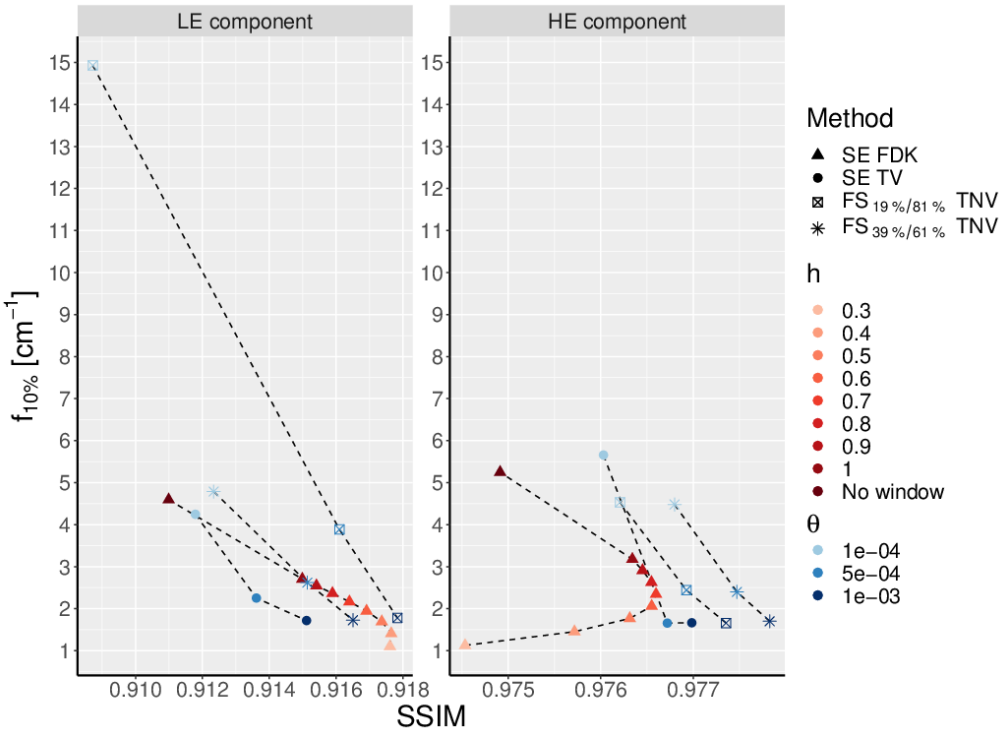

Figure 7 SSIM and spatial frequency for the best HMP scans of FDK, SE TV, and DE TNV (N=10) for LE (left) and HE (right).

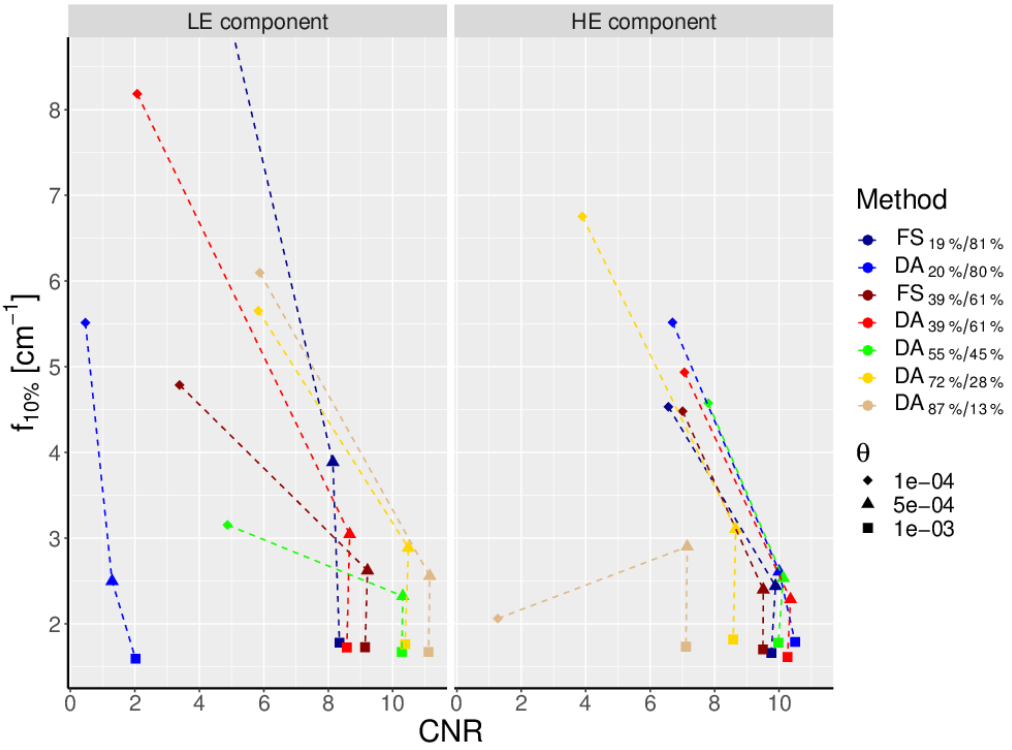

Figure 8 CNR and spatial frequency of the LE and HE component of the DE TNV reconstructions of the virtual human male pelvis phantom. Both dual-arc (DA) and fast-switching (FS) acquisition scheme are displayed.

4. Discussion

In this paper, we investigated the combination of DE CBCT with spectral TV iterative reconstruction (TNV). We evaluated the image quality and compared it to the image quality of SE scans reconstructed with different algorithms. While DE TNV yielded better SSIM results than FDK and SE TV reconstructions, the SE TV resulted in similar image quality in terms of CNR and spatial resolution. Increasing the spectral separation between LE and HE for the DE acquisition (by adding extra beam filtration) improved these image quality metrics beyond SE TV.

For the IQP DE scans, we observed an increase in CNR with increasing percentage of dose in the respective energy arm. This is due to the decrease of noise with higher imaging dose. Because of the higher density of the HMP, the LE and HE components showed more similar CNRs than the IQP. Higher density leads especially for lower kV to higher attenuation and therefore to more noise, decreasing the CNR.

The FS option means that a lower number of projection was used but with a higher mAs per projection. Despite similar dose, this led to an increase in CNR due to a decrease in noise for the large sized phantoms. The reason could be that part of the "noise" in images with low mAs was also due to streak artefacts and these streaks were reduced for the FS option. This lower noise could also have been the reason for the better SSIM results for FS compared to DA in most cases. For the H&N configuration where the gantry rotation was twice as fast as for the pelvis configuration, FS resulted in under-sampling which introduced aliasing, and therefore decreasing the CNR. Images with a larger θ might not have been so affected since larger θsmooth the images more and hence reducing the streaks. To mitigate this problem, the gantry rotation speed could be reduced from 1 rpm to 0.5 rpm. Since a 200° acquisition is used clinically for several body sites in our institute, a reduction in the angular range of the gantry rotation from 360° to 200° while maintaining the number of projections constant is also a feasible option if a reduction in the FOV is acceptable.

For the IQP, the SE TV reconstructions produced slightly better image quality regarding CNR and spatial resolution than DE TNV. Adding an extra beam filtration for the HE increased CNR and spatial resolution showing that a larger spectral separation improves image quality. With the 2mm Cu filtration, we were able to improve the DE image quality beyond the image quality of SE. An enhancement in hardware (improved X-ray source cooling) and optimizing the filtration (e.g. material, reducing present filtration for the LE acquisition) might lead to further increase in image quality as well as make it possible to use additional beam filtration for denser anatomy. Without the possibility of additional beam filtration for the HMP due to the aforementioned limitations, the CNR and spatial resolution of the SE_{120} TV was slightly superior to DE TNV. When comparing the spatial resolution of DE CBCT to CT, DE achieved an approximate 10% lower spatial frequency for both phantoms. For the IQP, the CT scan resulted in a spatial frequency of $f_{10\%,CT}=7.5cm^{-1}$ and DE TNV in $f_{10\%}=6.6cm^{-1}$ for $DA_{79\%/21\%}$ (LE component) with additional Cu filtration. For the HMP, the frequencies were $f_{10\%,CT}=6.8cm^{-1}$ and $f_{10\%}=6.2cm^{-1}$.

In the SSIM calculation, the mean gray value of the image is needed (Eq. 11). The gray values (attenuation coefficients) are energy dependent and therefore, the mean gray values of LE and HE scans are different. Unlike CNR and spatial resolution, a ground truth image is required to calculate the SSIM. Since our ground truth image is a 120kV CT scan, LE images result in lower SSIMs than HE images. So

even if the SSIM comparison had to be performed separately for low and high energy, the various reconstruction methods of the same energy can be compared to each other. For both phantoms and energies, the DE TNV resulted in larger SSIMs than FDK and SE TV. For the HE component of the IQP, this was even accomplished without extra beam filtration. The added filtration improved the SSIM nonetheless. Since the ground truths were 120kV diagnostic CT scans, the HE yielded larger SSIMs. Performing a HU calibration would increase the SSIM considerably but due to present image artefacts like streaks and shading, a simple calibration based on the phantom inserts of the IQP would lead to erroneous HU in some regions of the images.

At first sight, it is surprising that for the IQP, the LE yielded larger CNRs than the clinically used ME. This is because the contrast of the polystyrene insert for 70kV was more than double the contrast for 120kV and 130kV. The noise for the higher energies was also lower but not to an extent to compensate for such a contrast difference. For the HMP, the decrease in contrast from 70kV to 120kV was lower but the decrease in noise was larger, leading to a better CNR for 120kV. Using T(N)V decreased the noise considerably while maintaining a high spatial resolution leading to a superior image quality.

Generally, the change in image quality for SE TV and DE TNV with varying θ behaved as expected. First, the CNR increased considerable with increasing θ due to denoising. With further increase, the CNR increase got smaller and even reversed in some cases. That is because the noise could not be further reduced and the images got so blurred that the regions of the VOIs started to overlap. The decrease in spatial resolution with increasing θ is also expected since the images become blurrier with increasing θ. Unusual is the lower spatial resolution for θ=1x10^{-4} for the H&N configuration. The reason could be related to the number of projections because the effect is increased for FS and also visible for the pelvis configuration with FS. In some cases with low mAs and θ, the spatial resolution is apparently very high (Fig. 4 b, Fig. 8 LE component). This might be due to the high noise levels and streak artefacts. These make it difficult to find a proper fit for the ESF and the spatial resolution can be over-estimated.

For diagnostic CT, Yu et al. used DECT to generate monochromatic images of phantoms and evaluated their image quality. They were only able to show a greater (iodine) CNR compared to SE_{120} but not compared to SE_{80} images [29]. Until now, DE CBCT has mostly been used for different applications like material classification [30] and proton stopping power determination [31]. Men et al. developed a method to reconstruct an electron density image using DE CBCT and two basis material attenuation measurements [13]. Applying this approach to a Catphan 503 image quality phantom, they showed increased uniformity and signal-to-noise ratio. In an (anthropomorphic) Alderson Rando phantom, the dose calculation accuracy was also improved [11]. Lee et al. studied the feasibility of a single-scan DE cone-beam CT [32]. They used a multi-slit filter to generate the high (filtered) and low energy (non-filtered) beams, decreasing their total imaging dose by 55%. The LE projections were reconstructed with TV minimization and the HE image was reconstructed with an iterative algorithm using the joint sparsity between the two images. For a Catphan 600, the CNRs for polystyrene were 2.31 and 3.63, for the LE and HE image, respectively. These values are well below the results presented in this study, despite having a dose similar to our pelvis sized phantoms. Cassetta et al. performed DE CBCT acquisition and used image decomposition to generate virtual mono-energetic (VM) images of a Catphan 604 [33]. VM images with the lowest kV (50kV) resulted in the highest CNR, which matches our findings, where the LE component showed higher CNRs. For acquisitions with a similar CTDI dose as our H&N IQP, Cassetta et al. showed an increase in CNR from FDK to iterative reconstruction (iCBCT by Varian) for all images. However, the

highest CNR for polystyrene was less than 8. Even if their higher reconstruction resolution with a voxel size of 0.5x0.5x2mm^3 is taken into account, that is much lower than the CNR we reached for the H&N IQP (CNR=23.9) and even below the CNR for the pelvis IQP (without additional Cu filtration, CNR=13.6). Reasons for these differences could be the acquisition, Cassetta et al. performed a real fast kV switching acquisition, or the difference in the applied reconstruction algorithms. Furthermore, material decomposition is an ill-conditioned process which amplifies noise. Shi et al. also acquired DE CBCT scans to create VM images [34] and found, besides other advantages, an increase in CNR if an additional beam filtration (1mm Cu) was used in the HE acquisition, which is in agreement with our findings.

Besides traditional methods to improve image quality such as the application of anti-scatter grids [1] and iterative reconstructions [14, 20] as applied in this work, more recently, learned improvements based on deep neural networks have been successfully applied [35, 36]. The limited number of data sets available in this work did not permit to train such deep neural nets. It is, however, plausible, that learning based approaches will benefit from improved input images in terms of model complexity and\or resulting image quality [37].

The main limitation in this study was the heat limitation of the kV tube. This made FS for higher dose allocations in the LE arm and adding additional beam filtration for the HMP infeasible. From the results shown, it could be inferred that both would improve the image quality.
Another limitation was the use of the CTDI. The CTDI was developed for CT and it is not entirely applicable to CBCT due to the extended FOV. The cone-beam expands farther in cranial-caudal direction than the ionization chamber in the CTDI phantom. Nonetheless, the CTDI concept can be used in a relative manner, i.e. comparing doses of different exposures. In this study, the maximal $CTDI_w$ deviations from the SE_{120} dose were -13% and 9%, leading to a CNR decrease and increase of up to -7% and 4%, respectively. However, the relevant findings tend to be supported by the differences in dose, e.g. the added copper filtration ($DA_{79\%/21\%}$+Cu) showed a CNR increase of about 4% compared to SE_{70} TV despite its lower $CTDI_w$ of about 4%.
For SE TV and DE TNV, the parameter space was limited for clarity and to limit reconstruction times. The number of CG and T(N)V iterations were kept constant. Moreover, θ was not optimized. As seen for the HMP, using a limited number of θs might lead to missing the best image quality. Finding the optimal values for all parameters might lead to improvements in CNR and spatial resolution.
Our image quality evaluation was limited to CNR, spatial resolution, and SSIM. Other image quality characteristics (e.g. uniformity and visual streak artefacts (Fig. 6 C, D)) were not investigated. However, these metrics cover important aspects of image quality and we showed which values are achievable for the evaluated metrics for each reconstruction method. This can be seen as a starting point to a specific task. For example, (auto-)segmentation may benefit from higher CNR and deformable image registration may benefit from higher spatial resolution. The potential impact of the image quality improvements of DE CBCT needs to be evaluated in future studies.
Furthermore, some non-ideal physics effects (e.g. scatter, image lag) were neglected. Although, ignoring these effects also meant ignoring their influence on the cost function value and hence, on the reconstructed image, their impact is reduced. The presence of an ASG reduced the scatter substantially [1]. Image-lag causes circular shading in all reconstructed images. This so-called radar artefact is only exhibited in the periphery. In the IQP, it is visible at the left edge of the axial view (Fig. 2) and for the HMP, it appears around the left pelvic bone in the axial view of the HMP (Fig. 6). Consequently, the impact of image-lag on the evaluated image quality metrics is minimal. Despite it being a known problem

with potential corrections for SE [38], image lag and also scatter are energy dependent and therefore, energy dependent correction strategies would have to be implemented to yield comparable results. Implementing energy dependent image lag and scatter correction strategies was out of the scope of this project and therefore not performed. Including scatter correction and mitigating image artefacts resulting from aliasing and image lag would nonetheless lead to improved image quality.
For DE, we only analysed the two components of the images separately. Initial linear combinations of LE and HE components did not show significant improvements in image quality. Since the applied reconstruction algorithm couples the two spectral channels, the noise of the LE and HE components are not independent. The noise correlation between the two components increases with the parameter θ. This correlation makes a combination of the components to generate a new image more complex and a more sophisticated combination might be necessary to yield a gain in image quality.

The investigated DE CBCT approach might be beneficial for online adaptive radiotherapy. The presented improvements in CNR, SSIM, and spatial resolution could facilitate further steps towards online ART, where several aspects of image quality are important.

5. Conclusions

In this study, we applied a generalization of total variation to spectral data, total nuclear variation, to dual-energy CBCT scans. For the standard hardware configuration, TNV CBCT outperformed traditional FDK but was similar to single energy total variation CBCT. However, adding an additional copper beam filtration to reach a larger spectral separation between low and high energy led to increased image quality beyond SE. Further mitigation of our study limitations and optimization of beam filtration, dose allocation, reconstruction algorithms and software corrections may lead to further image quality improvements.

Acknowledgments

We thank Georgios Sotiropoulos for the help with RTK, Ray Sheombarsing for mathematical advice, and Thijs Perik, Geert Streekstra, and Atilla Erogluer for the help with the CTDI measurements.

Conflict of interest

This study was partially funded by Elekta Oncology Ltd. Our department licenses software for cone-beam reconstruction and registration to Elekta Oncology Systems Ltd.

References

[1] Stankovic U., Ploeger L.S., van Herk M., et al. Optimal combination of anti-scatter grids and software correction for CBCT imaging. Med Phys. 2017 Sep;44(9):4437-4451.

[2] Hansen D.C., Landry G., Kamp F., et al. ScatterNet: A convolutional neural network for cone-beam CT intensity correction [published correction appears in Med Phys. 2019 May;46(5):2538]. Med Phys. 2018;45(11):4916-4926.

[3] Estler A., Bongers M., Thomas C., et al. Application of a Metal Artifact Reduction Algorithm for C-Arm Cone-Beam CT: Impact on Image Quality and Diagnostic Confidence for Bronchial Artery Embolization. Cardiovasc Intervent Radiol 2019;42:1449–1458.

[4] Jaffray D.A., Siewerdsen J.H., Wong J.W., et al. Flat-panel cone-beam computed tomography for image-guided radiation therapy. Int J Radiat Oncol Biol Phys 2002;53(5):1337–1349.

[5] McBain C.A., Henry A.M., Sykes J., et al. X-ray volumetric imaging in image-guided radiotherapy: the new standard in on-treatment imaging. Int J Radiat Oncol Biol Phys 2006;64(2):625–634.

[6] De Los Santos J., Popple R., Agazaryan N., et al. Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. Int J Radiat Oncol Biol Phys 2013;87(1):33–45.

[7] Cui Y., Piper J.. Deformable dose accumulation with image guided radiotherapy for final dose evaluation in pelvic cases. Journal of Nuclear Medicine & Radiation Therapy 2011 S3–e001.

[8] Schröder L., Stankovic U., Remeijer P., et al. Evaluating the impact of cone-beam computed tomography scatter mitigation strategies on radiotherapy dose calculation accuracy. Physics and Imaging in Radiation Oncology 2019;10:35–40.

[9] Genant H.K., Boyd D.. Quantitative bone mineral, analysis using dual energy computed tomography. Invest Radiol. 1977;12(6):545-551.

[10] Chiro G.D., Brooks R.A., Kessler R.M., et al. Tissue signatures with dual-energy computed tomography. Radiology. 1979;131(2):521-523.

[11] Men K., Dai J., Chen X., et al.. Dual-energy imaging method to improve the image quality and the accuracy of dose calculation for cone-beam computed tomography. Phys Med 2017;36:110–118.

[12] Primak A.N., Ramirez Giraldo J.C., Liu X., et al. Improved dual-energy material discrimination for dual-source CT by means of additional spectral filtration. Med Phys. 2009;36:1359–1369.

[13] Men K., Dai J.R., Li M.H., et al. A method to improve electron density measurement of cone-beam CT using dual energy technique. BioMed research international 2015(12):858907.

[14] Baumueller S., Winklehner A., Karlo C., et al. Low-dose CT of the lung: potential value of iterative reconstructions. Eur Radiol 2012;22(12):2597–2606.

[15] Chen G.H., Tang J., Leng S.. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. Med Phys 2008;35(2):660–663.

[16] Rudin L.I., Osher S., Fatemi E.. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 1992;60(1):259 – 268.

[17] LaRoque S. J., Sidky E.Y., Pan X.. Accurate image reconstruction from few-view and limited-angle data in diffraction tomography. J Xray Sci Technol 2008;14:119–139.

[18] Holt K.. Total Nuclear Variation and Jacobian Extensions of Total Variation for Vector Fields. IEEE Trans Image Process 2014;23(9):3975–3989.

[19] Rigie D.S., La Riviere P.J.. Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. Phys Med Biol 2015;60(5):1741–1762.

[20] Sidky E.Y., Pan X.. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys Med Biol 2008;53(17):4777–4807.

[21] Tian Z., Jia X., Yuan K., et al. Low-dose CT reconstruction via edge-preserving total variation regularization. Phys Med Biol 2011;56(18):5949–5967.

[22] Niu S., Gao Y., Bian Z., et al. Sparse-view x-ray CT reconstruction via total generalized variation regularization. Phys Med Biol 2014;59(12):2997–3017.

[23] Rit S., Oliva M.V., Brousmiche S., et al. The reconstruction toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the insight toolkit (ITK). Journal of Physics: Conference Series 2014;489:012079.

[24] Mory C., Auvray V., Zhang B., et al. Cardiac C-arm computed tomography using a 3D + time ROI reconstruction method with spatial and temporal regularization. Med Phys 2014;41(2):021903.

[25] Jacques L., Hammond D.K., Fadili J.. Dequantizing compressed sensing: when oversampling and non-Gaussian constraints combine. IEEE Trans 2011;57 559–71.

[26] Shope T.B., Gagne R.M., Johnson G.C.. A method for describing the doses delivered by transmission x-ray computed tomography. Med Phys 1981;8(4):488–495.

[27] Krah N., Khellaf F., Létang J.M., et al.. A comprehensive theoretical comparison of proton imaging set-ups in terms of spatial resolution. PhysMed Biol 2018;63(13):135013.

[28] Virtanen P., Gommers R., Oliphant T.E., et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 2020;17:261–272.

[29] Yu L., Christner J.A., Leng S., et al. Virtual monochromatic imaging in dual-source dual-energy CT: radiation dose and image quality. Med Phys 2011;38(12):6371–6379.

[30] Zbijewski W., Gang G., Xu, J., et al. Dual-energy cone-beam ct with a flat-panel detector: Effect of reconstruction algorithm on material classification. Medical physics 2014;41(2):021908.

[31] Wohlfahrt P., Möhler C., Richter C., et al. Evaluation of stopping-power prediction by dual-and single-energy computed tomography in an anthropomorphic ground-truth phantom. International Journal of Radiation Oncology* Biology* Physics 2018;100(1):244–253.

[32] Lee D., Lee J., Kim H., et al. A Feasibility Study of Low-Dose Single-Scan Dual-Energy Cone-Beam CT in Many-View Under-Sampling Framework. IEEE Trans Med Imaging. 2017;36(12):2578-2587.

[33] Cassetta R., Lehmann M., Haytmyradov M., et al. Fast-switching dual energy cone beam computed tomography using the on-board imager of a commercial linear accelerator. Phys Med Biol. 2020;65(1):015013.

[34] Shi L., Bennett NR., Shapiro E., et al. Comparative study of dual energy cone-beam CT using a dual-layer detector and kVp switching for material decomposition, Proc. SPIE 11312, Medical Imaging 2020: Physics of Medical Imaging, 1131220.

[35] Liang X., Chen L., Nguyen D., et al.. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. Phys Med Biol. 2019;64(12):125002

[36] Harms J., Lei Y., Wang T., et al.. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med Phys. 2019;46(9):3998-4009

[37] Landry G., Hansen D., Kamp F., et al.. Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. Phys Med Biol. 2019;24;64(3):035011

[38] Stankovic U., Ploeger L.S., Sonke J.J., et al. Clinical introduction of image lag correction for a cone beam CT system. Med Phys 2016;43(3):1057–1064.