Design and Optimization of 3D VSHARP® Scatter Correction for Industrial CBCT using the Linear Boltzmann Transport Equation

Kevin Holt[a], Devang Savaliya[b], Amy Shiroma[b], Martin Hu[a], David Nisius[a], Steve Hoelzer[a], Mingshan Sun[b], Don Vernekohl[b], Josh Star-Lack[b]

[a]Varex Imaging, 3835 Carnation St, Franklin Park, IL 60131; [b]Varex Imaging, 683 River Oaks Parkway, San Jose, CA 95131

ABSTRACT

We have developed VSHARP®, a suite of scatter correction solutions that have been incorporated into the commercially available cone-beam software development toolkit, CST (Varex Imaging, Salt Lake City, UT) enabling scatter correction to be applied as part of an entire CBCT reconstruction pipeline. The suite includes 2D VSHARP®, a deconvolution correction using asymmetric Gaussian kernels, 2D VSHARP-ML, a U-NET machine-learning correction, and 3D VSHARP®, a correction using a rapid finite-element Linear Boltzmann Transport Equation (LBTE) solver to estimate scatter in a manner similar to traditional stochastic Monte Carlo (MC) simulations. Of the three corrections, 3D VSHARP is the most accurate and flexible since it can be readily applied to arbitrary scanner geometries, protocols, and scan parts while the 2D VSHARP models may need to be regenerated for each configuration. On the other hand, 3D VSHARP is inherently slower since a minimum of two reconstruction passes are needed and the LBTE solver, while much faster than traditional MC, is still computationally intensive. The goal of this work was to minimize LBTE run times for (typically large) industrial datasets by optimizing parameter settings, particularly the choice of the sampling grid dimensions. This was achieved by applying a multi-objective genetic algorithm to find the Pareto front characterizing the tradeoff between speed and accuracy and identifying key operating points on the curve. Testing with 720 frames of 3720x3720 projection data to make a reconstruction volume of size 500x500x600, we found that excellent image quality can be obtained by using a coarse scatter grid size of 27x27x32 volume and 44x44 detector and a primary grid size of 246x246 x295 volume and 295x295 detector, both over 42 frames for a grand total of 21 seconds LBTE computation time. We show the Pareto characterization, as well as demonstrations of 3D VSHARP image quality with significantly reduced scatter-induced artifacts such as streaking and shading.

Keywords: CBCT, Scatter Correction, Linear Boltzmann Transport Equation (LBTE), Genetic Algorithms, Pareto Optimization

1.INTRODUCTION

Recent algorithmic and computational advances have made MC-like scatter correction approaches much more practical. The Acuros platform[1,2] rapidly solves the linear Boltzmann Transport Equation (LBTE) using finite element methods to determine the scatter distribution directly rather than stochastically as in conventional MC. Recently, we unveiled 3D VSHARP[3,4] that uses Acuros to achieve accuracies comparable to MC methods in a tiny fraction of the time. Acuros' accuracy and run-time are both highly dependent on the choice of sampling grid used for the finite element solution. If the grid is too coarse then results are inaccurate, and if the grid is too fine then time is wasted. Wang et al[2] have addressed this issue for medical use by optimizing the Pareto front over the set of sampling parameters. In this work we perform a similar Pareto optimization for an industrial case of an aluminum motorcycle cylinder head.

2.METHODS

2.1 CST Framework

The correction utilizes CST, Varex's CT reconstruction SDK, which allows for flexible connection of modular plugins to perform a reconstruction. CST includes over 30 bundled plugins, including those necessary to implement 2D VSHARP[5], 2D VSHARP-ML[6], and 3D VSHARP. CST also includes a Physics Library that contains user-selectable cross-sections as well as x-ray spectra and detector response files required by 3D VSHARP.

2.2 Pipeline with 3D VSHARP and 2-pass FDK reconstruction

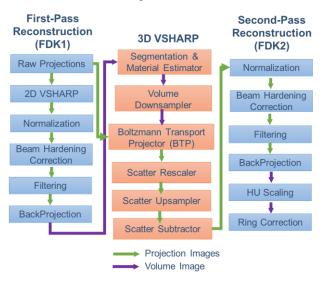An example pipeline with 3D VSHARP is shown in Figure 1.

Figure 1. Two-pass pipeline for FDK reconstruction with 3D VSHARP. The data flow green lines indicate projection data and purple lines indicate volume data.

Processing for 3D VSHARP is in 6 basic stages:

1.   Acquisition and Pre-processing: Read data from disk or from the detector, and perform pre-processing operations such as offset correction, bad-pixel correction, or lag correction.

2.   2D VSHARP: Perform 2D (kernel-based) scatter corrections for scatter from the detector housing and from the scanned object.

3.   FDK1: Complete a first pass reconstruction using the 2D VSHARP scatter-corrected projection data.

4.   3D VSHARP contains six main components:

a.   Segmentation and Material Estimation: The FDK1 volume is segmented into different regions corresponding to different materials, and a density is assigned to each material-voxel.

b.   Volume Downsampling: To reduce BTP computation time and GPU memory requirements, the volume is downsampled to make a lower resolution volume sampling grid.

c.   The Boltzmann Transport Projector (BTP) runs the LBTE solver for a subset of projection angles. In addition to using a lower resolution volume sampling grid, the BTP detector matrix size is typically smaller than the original detector matrix size.

d.   Scatter Rescaling: The simulated scatter output from BTP is rescaled to be at a comparable signal level to the measured data. To help determine the scaling factor, the simulated primary output signal is used as a reference since the measured data should be proportional to the sum of the primary and scatter.

e.   Scatter Upsampling: The scatter signal is upsampled in 3 dimensions (detector matrix U,V and projection angle) so that sampling matches the measured data.

f.   Scatter Subtraction: The subtraction also includes a scatter-fraction smoothing and clipping step to ameliorate noise amplification from the subtraction.

5.   FDK2: Perform a second pass reconstruction. Apply post-processing operations such as ring correction.

2.3 Data acquisition and reconstruction

A motorcycle cylinder head was scanned and reconstructed with the parameters shown in Table 1.

Table 1. Acquisition and reconstruction parameters.

Acquisition Parameters	
Number of Detector Pixels (Dexels)	3072 x 3072
Projection Detector Pixel (Dexel) Size	139 μm x 139 μm
Detector Size	427 mm x 427 mm
Number of Projection Frames	720
Source-Axis Distance	1084 mm
Source-Imager Distance	1302 mm
Tube Spectrum	450 kV + 2 mmCu
Detector Model	4343HE with DRZ Plus
Reconstruction Parameters	
Number of Voxels	500 (x) x 500 (y) x 600 (z)
Voxel Size	0.5 mm x 0.5mm x 0.5 mm

Reconstruction was performed on a PC with 2 Intel Xeon ES-2637v4 chips each containing 8 cores at 3.5 GHz, and an NVIDIA Titan RTX GPU with 4608 cores at 1.35 GHz.

2.4 Pareto optimization

To characterize the processing time-vs-error tradeoff, we used the NSGA2 algorithm[7], a genetic algorithm (GA) for discovering Pareto fronts in multi-objective problems. The two objectives were time and error. To measure error, a "golden" reconstruction was performed with all parameters set for maximum accuracy, then for each operating point the root mean square (RMS) error versus the golden reconstruction was measured.

The total search space included: 1. Primary Volume Matrix Size, 2 Scatter Volume Matrix Size, 3. Primary Detector Matrix Size, 4. Scatter Detector Matrix Size, 5. Number of Primary Projections, 6. Number of Scatter Projections.

The constraints on the search were:

1.   All voxels were isotropic and completely filled the prescribed reconstruction field-of-view (FOV) in the *x*-, *y*- and *z*-directions. As the z-axis FOV was somewhat larger than the transaxial FOV—because the first pass reconstruction was extrapolated to better capture second-order scattering events in the top and bottom portions of the object—the GA algorithm chose values for the number of voxels in the x-y direction, denoted as *Primary NumVoxelslXY* and *Scatter NumVoxelsXY*. and then automatically compute the corresponding number of primary or scatter voxels that spanned the entire z-axis FOV.

2.   All detector pixels (**dexels**) were isotropic and completely filled the (square) detector extent in the U and V directions. The number of scatter dexels, *Scatter NumDexelsUV*, was chosen by the GA but the number of primary

dexels depended on the number of primary voxels (*Primary NumVoxelsXY*) so that the Primary Dexel Size (mm) was the Magnification*PrimaryVoxelSize = SID/SAD*PrimaryVoxelSize = 1.2*PrimaryVoxelSize.

3.   The primary and scatter projection angles were equally spaced. The relevant native search parameter was termed the *Downsampling Factor* from which an *AngularIncrement* value, quantized to multiples of 0.5°, was computed. The number of frames then was then equal to Floor (720/*AngularIncrement*). The *Scatter FrameRate Downampling* factor was an integer multiple of the *Primary FrameRate Downsampling*.

Table 2 shows the native search space used by the GA.

Table 2. Native search space used by the GA.

Search Parameter	Search Range (Integers Only)
Scatter NumVoxelsXY	25 to 90
Primary NumVoxelsXY	25 to 500
Scatter NumDexelsUV	10 to 42
Primary FrameRate	6 to 60
Scatter FrameRate	*Primary FrameRate Downsampling* ×

We ran 25 generations of NSGA2 with a population of 50, then another 25 generations with a population of 75.

3.RESULTS

3.1 Motorcycle cylinder head reconstructions

Figure 2 shows example reconstructions of a sagittal slice including A) an "Uncorrected" reconstruction, B) the first pass reconstruction (from the unoptimized 2D VSHARP), and two 3D VSHARP reconstructions: C) the "Golden" reconstruction performed at maximum LBTE resolution for a run time of about 6 hours, and D) the "Operating Point F" reconstruction using a coarser LBTE grid, requiring only 21 seconds of BTP time. Significant improvements in crispness and homogeneity are seen in the 3D VSHARP images with Operating Point F retaining similar image quality to the Golden image.
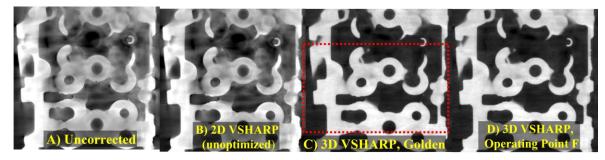

Figure 2. Central sagittal slice. The window for each is chosen so that the 10% and 90% gray values within the window correspond to the 5th and 95th percentiles of the voxel values. The red outline shows the zoomed-in region in Figure 4.

3.2 Pareto results

Figure 3 shows results from all generations of the NSGA2 run. Each operating point is shown as a blue dot, the Pareto front is shown as a red line, the convex hull of all the operating points is shown in green, and the set of points that were further studied is indicated by labeled circles.
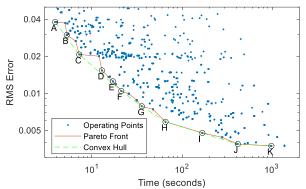
Figure 3. Log-Log plot of Pareto Optimization Results.

Table 2 shows search parameter and objective results for several operating points (A,D,F,I). As expected, RMS error decreases as 3D VSHARP runtime increases as do the sampling grid dimensions in general. Of note, is the exception that *Scatter NVoxelsXY* is relatively constant. This may partially reflect that its lower bound was 25. Also of note, *Scatter NDexelsUV* is more than 1.5x larger than *Scatter NVoxelsXY* which is greater than the 1.2x magnification one might have assumed. This may reflect the relatively high fidelity and angular resolution of each voxel's computed scatter distribution which is enabled by the use of Legendre polynomials to describe the profile and propagate scatter across the grid[1] which, in turn, might allow for coarser volume resolution than detector resolution. Finally of note is that *FrameRateDownsampling* for each operating point was the same for the primary and scatter computations even though higher voxel and dexel resolution was required for the primary estimate.

Table 3a. Results for selected operating points matching Figure 3.

Objectives	Operating Points			
	A	D	F	I
RMS Error	.038	.015	.010	.005
3DVSHARP time (s)	3.9	12.9	21.1	168

Table 3b. Results for selected operating points matching Figure 3.

Sampling Grid Search Parameter	GA results for selected Operating			
	A	D	F	I
Scatter NVoxelsXY	27	27	27	25
Primary NVoxelsXY	85	119	246	449
Scatter NDexelsUV	13	44	44	57
Number Primary Projections	14	37	42	120
Number Scatter Projections	14	37	42	120

3.3 Example images along the Pareto front

Figure 3 shows zoomed-in images, corresponding to the red outlined region in Figure 2c, for the Operating Points in Table 3. The red arrows point to inhomogeneities in the form of streaks or shading. The top image (Operating Point A) takes the least amount of BTP time, 3.9 seconds, but does show artifacts. Moving down the figure, artifacts decrease as execution time increases. While homogeneity steadily improves with increased BTP time, we note that for many applications, images D or F may be perfectly acceptable, requiring 13 and 21 seconds BTP time respectively.

3.4 Contrast analysis

The histograms of different reconstructed volumes are shown in Figure 4. In the Uncorrected Image, the air and aluminum peaks are poorly separated, with slightly better separation occurring in the 2D VSHARP image and good separation in 3D VSHARP images.

To quantify the separation of air and aluminum, the contrast-to-noise ratio in each histogram was computed by segmenting the images into "Air" or "Object" voxels and using the equation $CNR = \frac{(\mu_{Obj}-\mu_{Air})\sqrt{2}}{\sqrt{\sigma_{Obj}^2 + \sigma_{Air}^2}}$ where μ_{air} and μ_{obj} are the respective typical linear attenuations of the Air and Object voxels and σ_{air} and σ_{obj} are the respective standard deviations. We used the histogram peaks for μ, and their midpoints as the segmentation thresholds, shown by o and x in Figure 4.
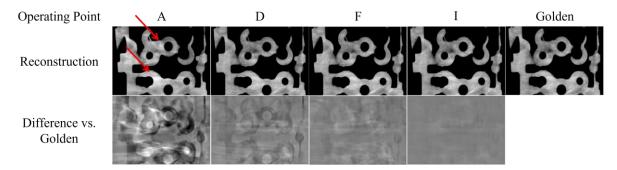

Figure 3a. Zoom-in on the central sagittal slice from various reconstructions. The window for each reconstruction is $[\mu_{Obj}-3\sigma_{Obj}, \mu_{Obj}+2\sigma_{Obj}]$. The window for each difference image is $[-0.02 \text{ mm}^{-1}, +0.02 \text{ mm}^{-1}]$.
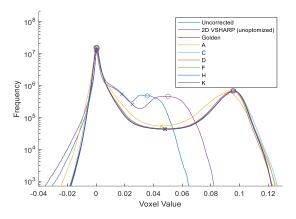

Figure 4. Histograms of reconstructed volumes.

The resulting CNRs for operating Points A, D, F, I and "Golden" were 8.1, 8.9, 9.0, 9.0, 9.0 respectively. The Uncorrected CNR of 4.7 was the lowest while the 2D VSHARP CNR was slightly improved at 5.3. Points F and I have CNRs comparable to the golden image, which again suggests that there is no significant benefit to spending more than 10-20 seconds on the LBTE solution.

4.DISCUSSION

MC or pseudo-MC scatter correction methods such as 3D VSHARP can produce highly accurate scatter estimates and, in fact, are used as a gold standard for training ML-scatter correction methods. A main advantage of MC methods is that they are versatile since all that is needed at runtime is the geometric specification of the CBCT system and a physics library which characterizes it. However, MC or even pseudo-MC methods are generally not as fast as Machine Learning or Kernel methods since a first pass reconstruction is required and the scatter transport calculation is computationally intensive.

For this data set, it was found that 10 to 20 second LTBE run times are sufficient if using an optimized sampling grid. We expect this result to be somewhat problem dependent, and may change with object size, complexity, or material, as well as with scanner geometry. However, it is interesting to note that our optimal time is roughly in line with the results of[2].

Of note is that the 2D VSHARP calibration was not optimized for this setup. Although proper tuning may improve 2D VSHARP image quality, we chose to leave it unoptimized to show that the segmentation algorithm is fairly forgiving.

For future work, there are still many interesting parameters left to optimize including looking into non-uniform angular sampling[2], optimizing interpolation and segmentation methods, and optimizing intrinsic LBTE parameters such as energy grouping.

5.CONCLUSION

3D VSHARP was shown to significantly reduce scatter artifacts and produce excellent results with 10-20 seconds of computation time. Although a first-pass reconstruction is still needed, for the second pass reconstruction the results show that the additional time added by 3D VSHARP is minimal especially given that CST permits the LBTE computation to be performed in parallel with other FDK operations such as filtering and backprojection.

REFERENCES

[1]   A. Maslowski et. al., "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter– Part I: Core algorithms and validation", Med. Phys. 45(5) (2018) 1899-1913.

[2]   A. Wang et. .al, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter –Part II: System modelling, scatter correction, and optimization", Med. Phys. 45(5) (2018) 1914-1925.

[3]   A. Shiroma et al. "Scatter Correction for Industrial Cone-Beam Computed Tomography (CBCT) Using 3D VSHARP, a fast GPU-Based Linear Boltzmann Transport Equation Solver", 9th Conference on Industrial Computed Tomography (iCT) 2019, 13-15 Feb, Padova Italy.

[4]   Star-Lack, et. al. "3D VSHARP®, a general-purpose CBCT scatter correction tool that uses the Linear Boltzmann Transport Equation." In Medical Imaging 2021: Physics of Medical Imaging, vol. 11595, International Society for Optics and Photonics, 2021.

[5]   M. Sun, J. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition", Phys. Med. Biol. 55 (2010) 6695-6720.

[6]   Maier, et. al. "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation." Medical physics 46, no. 1 (2019): 238-249.

[7]   K. Deb, et. al. "A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II", Lecture Notes in Computer Science, vol 1917, Springer, 2000