Accurate Low-Contrast 3D Cone-Beam Reconstruction With Algebraic Methods

Klaus Mueller[1,2], Roni Yagel[1], and John J. Wheller[2]

[1] Department of Computer and Information Science, The Ohio State University, Columbus, OH
[2] Department of Pediatrics, The Ohio State University, Columbus, OH

Abstract

This paper examines the use of the Algebraic Reconstruction Method (ART) and related techniques to reconstruct 3D objects from a relatively sparse set of cone-beam projection data. Although ART has been widely used for cone-beam reconstruction of high-contrast objects, e.g. in computed angiography, we are interested in the more challenging low-contrast case which represents a little investigated scenario for ART. Preliminary experiments indicate that for cone angles greater than 20˚, traditional ART produces reconstructions with strong aliasing artifacts, obliterating much object detail. By analyzing the reconstruction process using signal processing principles it is revealed that the source of these artifacts is the non-uniform reconstruction grid sampling and correction by the cone-beam rays. To eliminate these errors, we devise a new way of computing the weights of the reconstruction matrix. This new method is more adequate for cone-beam and replaces the usual constant-size interpolation filter by one whose size and amplitude is dependent on the source-voxel distance. Doing so enables us to generate alias-free reconstruction at little extra cost. An alternative analysis reveals that Simultaneous ART (SART) has also great potential to produce reconstruction without cone-beam related artifacts, however at greater computational cost. Putting everything together, we thoroughly investigate the influence of various ART parameters, such as volume initialization, relaxation coefficient λ, correction scheme, and number of iterations, on reconstruction quality. Using a 3D version of the Shepp-Logan phantom, we find that ART typically requires only three iterations to render a reconstruction result close to the optimum (given proper parameter settings). Thus we conclude that ART is potentially not any costlier than Filtered Backprojection (FBP) techniques, particularly if one considers the fact that ART only requires a fraction of FBP's projection set.

1. Introduction

The field of 2D and 3D reconstruction methods can be roughly divided into two main categories. On one side there is the domain of direct methods that capitalize on the Fourier Slice Theorem [4], while on the other side lies the domain of iterative methods that seek to solve the reconstruction problem by solving a system of simultaneous linear equations. The most prominent member of the former group is the Filtered Backprojection (FBP) algorithm. Here, the reconstruction is achieved by filtering the projection images with a ramp filter in frequency space, and then backprojecting the filtered projections onto a reconstruction grid [5]. The first and still most prevalent representative of the iterative methods is the Algebraic Reconstruction Technique (ART), attributed to Gordon et. al. [8]. Here, a reconstruction grid is iteratively updated by a projection-backprojection procedure until a convergence criterion is satisfied.

In clinical, slice-based CT, FBP is nowadays exclusively used, since it is non-iterative and therefore potentially faster than the iterative ART, at least in the case when the number of projections is identical. Note, however, that clinical CT scanners usually acquire more than 500 line projections per slice, which approximates the continuous form of the inverse Radon integral rather well. The true power of ART is revealed in cases where one does not have a large set of projections available, when the projections are not distributed uniformly in angle, when the projections are sparse or missing at certain orientations, or when one wants to model some of the photon scattering artifacts in the reconstruction procedure [1][15][20]. Noise is also better handled by ART-type methods [15].

While 2D fan-beam tomographic reconstruction has been routinely used for over two decades, 3D cone-beam reconstruction is still in the research stage. As yet, there are no clinical cone-beam scanners, although a variety of cone-beam algorithms based on FBP have been proposed. The most notable ones are Feldkamp, Davis, and Kress [6], Gangreat [7], and Smith [25], all published in the mid and late 1980s. While the research on general 2D ART [8][12][11][9] and 2D fan-beam ART [10][20] is numerous, the published literature on 3D reconstructors using ART-type algorithms is rather sparse. One exception is the work by Matej et. al. [19], whose studies in the field of fully-3D reconstruction of noisy PET data indicate that ART produces quantitatively better reconstruction results than the more popular FBP and MLE (Maximum Likelihood Estimation) methods. In this case, however, not a cone-beam reconstructor was used, but the projection rays were re-binned which simplified ART to the parallel-beam case. Another group of researchers has successfully applied ART for SPECT data [24] not long ago. However, the most popular use of 3D cone-beam ART is in 3D computed angiography. Here one acquires images of blood-perfused structures injected with radio-opaque dye while rotating the cone-beam X-ray source-detector pair around the patient. For toxicity reasons, the dye injection time is limited and thus the number of projections that can be acquired is restricted as well. Ning and Rooker [23] and Saint-Felix et. al. [26] all use ART-type methods to 3D reconstruct vascular trees in the head and abdominal regions. Note, however, that the objects reconstructed in 3D computed angiography are of rather high contrast, which poses the reconstruction problem as almost a binary one.

In this work, we analyze ART for the low-contrast cone-beam setting, a scenario that has not been studied much in the past. While in this paper we aim to provide algorithms that make ART an attractive alternative to cone-beam FBP methods with regards to accuracy, another paper will focus on fast projection algorithms that make ART competitive to FBP in speed [21]. Addressing the former goal, we analyze the cone-beam reconstruction procedure of ART from a sampling theory point of view, which lets us identify and eliminate the additional set of problems that ART-type algorithms have in the cone-beam setting. Most of the problems stem from the circumstance that in cone-beam the diverging rays sample and update the reconstruction grid in a non-uniform fashion, which leads to a considerable amount of aliasing-related noise artifacts for cone angles greater than 20°. We will hence devise several strategies that eliminate these artifacts. In our analysis, we will investigate two ART-type algorithms that are relatively different in their view of the reconstruction process but yet of similar efficiency: ART and Simultaneous ART (SART) [2].

In summary, the better noise handling of ART, its capability to work well even with a small number of projections and the opportunity to apply spatial and numerical constraints on the reconstructed object make ART a serious contender for the current FBP-type algorithms. The circumstance that ART requires only half of the projection data of FBP [9] weakens the argument that ART must always be slower than FBP. We do not attempt in this paper to compare FBP with ART-type algorithms in a qualitative sense. However, we will show that ART can generate a reconstruction of sufficiently good quality in only three iterations, while requiring only a small dataset, which readily proofs the fitness of ART for routine 3D reconstruction. Considering that FBP brings with it the overhead for prior Fourier domain filtering or convolution of the projection data, it is quite possible that three ART iterations with half the projection data in conjunction with some caching of re-usable computations [21] can be performed at a similar cost than FBP.

The outline of this paper is as follows. Section 2 gives a short recap on the workings of ART-type algorithms. Section 3 then moves ART into the cone beam setting, analyzes its shortcomings and presents solutions to overcome these deficiencies. Next, Section 4 discusses the use of SART for cone-beam reconstruction. Finally, Section 5 puts everything together and presents a variety of results obtained with our ART testbed software. In this section, the effects of a wide selection of ART parameters on both reconstruction quality and speed are investigated. The studied factors include: the value and functional variation of the ART relaxation coefficient λ, the relevance of volume initialization, and the effect of the ART correction algorithm (ART vs. SART). Lastly, Section 6 concludes our studies.

2. Preliminaries

In this section, we will give a brief review of ART and present relevant previous work. While ART originally was proposed for the 2D case, the mathematical notation translates trivially to the 3D case.

As was mentioned before, ART poses the reconstruction problem as a system of linear equations:

$$\begin{aligned} w_{11}v_1 + w_{12}v_2 + w_{13}v_3 + \ldots + w_{1N}v_N &= p_1 \\ w_{21}v_1 + w_{22}v_2 + w_{23}v_3 + \ldots + w_{2N}v_N &= p_2 \\ &\ldots \\ w_{M1}v_1 + w_{M2}v_2 + v_{M3} + \ldots + w_{MN}v_N &= p_M \end{aligned} \tag{1}$$

Here, the v_j are the values of the reconstruction grid elements (called *voxels* from now on), the p_i are the values of the pixels in the acquired projection images, and the weight factors w_{ij} represent the amount of influence a voxel j has on a ray passing from the source through image pixel i.

Usually one reconstructs on a cubic voxel grid with a sidelength of n voxels, thus the number of grid voxels $N=n^3$. Also, for a 3D single-orbit reconstruction we generally assume a spherical reconstruction region. In this case we have $(1/6)\pi n^3$ unknown voxel values and $(1/4)\pi n^2$ relevant pixels per image. For the equation system (1) to be determined, the number of projection images S has to be:

$$S = \frac{(1/6)\pi n^3}{(1/4)\pi n^2} = 0.67n, \tag{2}$$

This means that for n=128, 86 projections images are required. As was shown by [9] for the 2D case, the number of required projections in ART is half the number of the projections required for FBP. This happens because the Fourier Slice Theorem arranges the projections onto a polar grid in frequency space, and in order to provide adequate sampling in the periphery, one must oversample in the interior frequency regions. This is part of the reason for the strength of ART in the limited projection data case.

However, it is not always the case that $S=0.67 \cdot n$. Sometimes (1) is overdetermined, or, more often, it is underdetermined. In either case, the large magnitude of (1) does not allow its solution by matrix inversion or least-squares methods. In addition, usually noise and sampling errors in the ART implementation do not provide for a consistent equation system anyhow. Thus, an iterative scheme proposed by Kaczmarz [14] is used. Starting from an initial guess for the volume vector, $\boldsymbol{V}=\boldsymbol{V}^{(0)}$, we select at each iteration step k, $k>0$ one of the equations in (1), say the one for p_i. A value $p_i^{(k)}$ is measured which is the value of pixel i computed using the voxel values as provided by the present state of the vector $\boldsymbol{V}=\boldsymbol{V}^{(k)}$. A factor related to the difference of $p_i^{(k)}$ and p_i is then distributed back onto $\boldsymbol{V}^{(k)}$ which generates $\boldsymbol{V}^{(k+1)}$ such that if a $p_i^{(k+1)}$ were computed from $\boldsymbol{V}^{(k+1)}$, it would be closer to p_i than $p_i^{(k)}$. Thus, we can divide each grid update into three phases: a projection step, a correction factor computation, and a backprojection step.

The correction process for one element of $\boldsymbol{V}$, i.e. v_j, can be expressed by:

$$v_j^{(k+1)} = v_j^{(k)} + \lambda \frac{p_i - \sum_{n=1}^{N} w_{in} v_n^k}{\sum_{n=1}^{N} w_{in}^2} w_{ij} \tag{3}$$

where λ is the *relaxation factor* typically chosen within the interval (0.0,1.0], but usually much less than 1.0 to dampen correction overshoot. This procedure is repeated in an iterative fashion for all equations in (1).

The sum term in the nominator of (3) requires us to compute the integration of a ray across the volume. The integration process can be performed by using raycasting, i.e., sampling the volume at equidistant locations with an interpolation kernel h and accumulating the interpolated values. Since accurate integration requires many sampling points, this is very time consuming. A more efficient way was proposed by many authors [10][16][17][30], and consists of reordering the ray integral so that each voxel's contribution to the integral can be viewed isolated from the other voxels. The interpolation kernel is placed at the voxel locations, and the volume grid can be viewed as a field of overlapping interpolation kernels which, as an ensemble, make up the continuous object representation. A voxel v_j's contribution is then given by $v_j \cdot \int h(s)\, ds$, where s follows the integration of the interpolation kernel in the direction of the ray. Here, $\int h(s)\, ds$ represents a voxel weight factor in (3). If the viewing direction is constant for all voxels or if the interpolation kernel is radially symmetric, we can pre-integrate $\int h(s)\, ds$, often analytically, into a lookup-table (also called the *kernel footprint*). We can then map all voxel footprints to the screen where they accumulate into a projection image, as is done in the splatting approach of Westover [30]. Alternatively, we can map the footprint tables onto polygons and place these into volume space at each voxel position. Here, it is very important for accurate ray integration to align the footprint polygons perpendicularly to the ray direction. The integration process then consists of traversing the field of footprint polygons for each ray, indexing the lookup-table for each voxel along the ray, multiplying the looked-up weights by the voxel value, and accumulating it into the raysum in the nominator of (3). (See a future paper for more detail [21]). Backprojection is performed in a similar way, only that here the voxels receive (corrective) energy, scaled by their weight factors, instead of emitting it.

The choice of h varies in the existing ART implementations. We will be using a kernel based on the Bessel-Kaiser window, as proposed by Matej and Lewitt [16][17]. Multidimensional Bessel-Kaiser functions have many desirable properties, such as fast decay for frequencies past the Nyquist rate and radial symmetry. They can also be tuned so that the kernel's frequency spectrum is at a minimum at multiples of the sampling frequency, where the signal's aliases are largest.

Instead of updating the volume on a ray-basis, ART-type methods exist that correct the volume on an image-basis. One representative of these block-iterative methods is Simultaneous ART (SART), developed by Andersen and Kak [2], which was shown to significantly reduce the noise artifacts that were observed with ray-iterative ART. The projection step of SART performs a summed volume rendering [13] of the reconstruction grid, then subtracts the rendered image from the acquired projection image, normalizes the result, and backprojects the image in an inverse volume rendering process. More formally, the SART correction equation is as follows:

$$v_j^{(k+1)} = v_j^{(k)} + \lambda \frac{\sum_{p_i \in P_\varphi} \left(\frac{p_i - \sum_{n=1}^{N} w_{in} v_n^{(k)}}{\sum_{n=1}^{N} w_{in}} \right) w_{ij}}{\sum_{p_i \in P_\varphi} w_{ij}} \tag{4}$$

In this equation, the correction term for voxel j depends on a weighted average of all rays of a projection P_φ that traverse the voxel j. (Here, φ denotes the orientation angle at which the projection was taken.) Thus, in SART, all rays of a projection P_φ are simultaneously processed, hence the name *Simultaneous* ART.

Before choosing one method over the other, however, one should also contrast the potential computational efficiency of ART vs. SART. As other research [21] has indicated, ART is considerably easier to accelerate than SART, and runtime ratios T_{SART}/T_{ART} of approximately 1.5 have been found.

Apart from the correction algorithm and from the projection access order as well (see e.g. [22]), other important parameters that influence the reconstruction quality include: the relaxation coefficient λ and the initial state $\boldsymbol{V}^{(0)}$ of the volume. A smaller λ generally provides for a less noisy reconstruction, but increases the number of iterations required for convergence. There is also the issue if λ should be set to a constant value, or if it should vary over some function of time, as suggested by Andersen [1]. With respect to $\boldsymbol{V}^{(0)}$, it was proven by Tanabe [29] that, with ART, the reconstruction will always converge to the solution closest to the initialized volume state. From this we realize that the initialization may play a large role in the quality of the reconstruction. Since, in the general case, it is hard to predict beforehand what the reconstructed volume will look like, proper volume initialization is difficult.

We shall not leave this section without mentioning that the ART methods outlined in this paper for cubic grids also fully extend to the dodecahedral grids that were proposed in [18]. These grids were shown to reduce the number of voxels to be processed by about 30%.

3. A Modification for ART to Enable Accurate Cone-Beam Reconstruction

In this section, we investigate the accuracy of ART in the context of low-contrast 3D cone-beam reconstruction. We will find that ART in its present form is unsuitable in the cone-beam setting as it produces reconstructions with significant reconstruction artifacts. Henceforth, we will prescribe a number of modifications of ART's projection and backprojection mechanisms with which accurate reconstructions can be obtained, and which do not compromise the efficiency of ART. For quality assessment we will use a 3D extension of the Shepp-Logan brain phantom [27], similar to the one due to Axelsson [3]. The definition of our phantom is given in Table 1, while two orthogonal slices across the phantom are shown in Fig. 1. From this phantom, we analytically compute 80 projection images of 128×128 pixels each, forcing (1) to be slightly underdetermined. The projections are obtained at equidistant angles φ within a range of $[0, 180^{\circ}+\gamma]$, where $\gamma/2$ is the cone half-angle.

	Center (mm)			Half-axis (mm)			Angle (degrees)		conc.
	x	y	z	x	y	z	θ	ϕ	
1	0.0	0.0	0.0	69.0	90.0	92.0	0.0	0.0	2.0
2	0.0	0.0	-1.84	66.24	88.0	87.4	0.0	0.0	-0.98
3	-22.0	-25.0	0.0	41.0	21.0	16.0	72.0	0.0	-0.02
4	22.0	-25.0	0.0	31.0	22.0	11.0	-72.0	0.0	-0.02
5	0.0	-25.0	35.0	21.0	35.0	25.0	0.0	0.0	0.01
6	0.0	-25.0	10.0	4.6	4.6	4.6	0.0	0.0	0.01
7	-8.0	-25.0	-60.5	4.6	2.0	2.3	0.0	0.0	0.01
8	6.0	-25.0	-60.5	4.6	2.0	2.3	90.0	0.0	0.01
9	6.0	6.25	-10.5	5.6	10.0	4.0	90.0	0.0	0.02
10	0.0	62.5	10.0	5.6	10.0	5.6	0.0	0.0	-0.02
11	0.0	-25.0	-10.0	4.6	4.6	4.6	0.0	0.0	0.01
12	0.0	-25.0	-60.5	2.3	2.3	2.3	0.0	0.0	0.01

Table 1. The definition of a 3D extension of the Shepp-Logan phantom [27], similar to the one used by [3]. The angles θ and ϕ are the polar and azimuthal angles of the ellipsoid z-axis. The scanner rotates about the y-axis.

Although detailed quantitative results are postponed to Section 5, we would like to illustrate the material in this section by the use of some examples. These examples will assume certain settings of parameters such as λ, and $\boldsymbol{V^{(0)}}$, which will later be shown to be a good compromise between

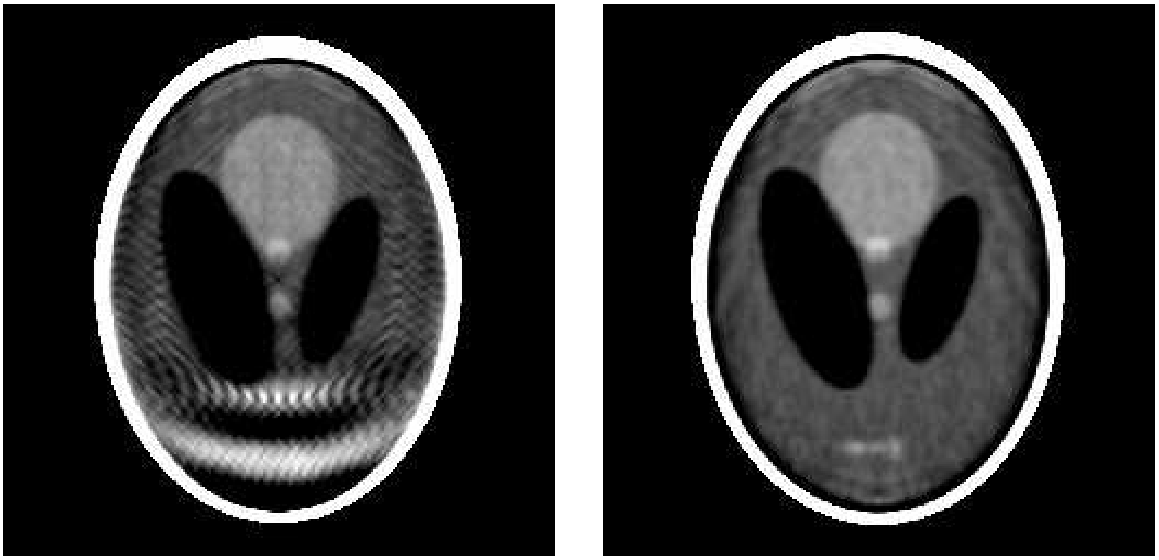
Fig. 1. Slices across the 3D Shepp-Logan brain phantom: (a) y=-25mm, (b) z=8mm. The pixel intensities in these and all other slice images presented in this paper were thresholded to the interval [1.0,1.04] to improve the displayed contrast of the brain features.

accuracy and speed of convergence.

3.1 Reconstruction artifacts in traditional cone-beam ART

Let us now use the ART algorithm of (3) to reconstruct a 128^3 volume from 80 projections with γ=60˚. λ is set to 0.08 and $\boldsymbol{V^{(0)}}$**=0**. The reconstruction result of slice y=-25mm after 3 iterations is shown in Fig. 2a. Here, we observe significant reconstruction artifacts which obliterate the small tumors in the lower image regions almost completely. For a comparison, Fig. 2b shows a 3D reconstruction from parallel beam data with the same algorithm and parameter settings. No significant artifacts are present in this case.

Thus the artifacts must result from the cone-beam configuration in which rays emanate from the source and traverse the volume in a diverging fashion before finally hitting the projection plane. It may be suspected that it is this diverging nature of the rays that causes the reconstruction artifacts in Fig. 2a. And indeed, more evidence is provided by Fig. 3, where we show the reconstruction volume of a solid sphere (diameter=0.75n) after the first correction image (at φ=0˚) was applied, for both 60˚ cone-beam data (Fig. 3a) and parallel-beam data (Fig. 3b). In these figures, we choose the z-axis to coincide with the beam direction and z_c to be the location of the volume center slice perpendicular to the beam direction (see also Fig. 4). In the cone-beam correction of Fig. 3a, the non-uniform density distribution in the volume center slice along the cone direction (Fig. 3a, side view) can be easily noticed. We see that much more energy is deposited in the volume slices close to the source where the ray density is high (Fig. 3a, near slice, $z=z_n=z_c-0.25n$), while only little energy is deposited in the volume slices further away from the source where the ray density is low (Fig. 3a, far slice, $z=z_f=z_c+0.25n$). In particular, the far slice displays a grid-like pattern which indicates an undersampling of the volume by the rays in this slice. This inadequate ray sampling rate potentially gives rise to aliasing, which is very likely to have caused the reconstruction artifacts of Fig. 2a. The effects of aliasing are amplified since in ART the volume is projected and updated con-



(a) (b)

Fig. 2. (a) The slice of Fig. 1a reconstructed with traditional ART from cone-beam projection data (λ=0.08, γ=60°, $\boldsymbol{V^{(0)}}$**=0,** 3 iterations). Notice the significant stripe artifacts that completely obliterate the small tumors. (b) A reconstruction of the same slice from parallel beam data using the same algorithm and parameter settings. This reconstruction does not have the strong artifacts of (a).

tinuously, with every projection introducing additional aliasing into the reconstruction.

In contrast to the cone-beam case, the parallel-beam correction, shown in Fig. 3b, provides a homogeneous density distribution. No excess density is deposited in the near slice at z=z_n and no aliasing-prone, grid-like pattern is generated in the far slice at z=z_f. Thus reconstruction artifacts are unlikely to occur and, indeed, have not been observed in Fig. 2b. In the following section, we will now investigate our observations more formally.

3.2 A new scheme for projection and backprojection to prevent reconstruction artifacts

Both the projection and the backprojection algorithms must be adapted to avoid the aliasing problems outlined in the previous section. These enhancements make it necessary to modify ART's basic correction algorithm. We now describe these new concepts in more detail.

3.2.1 Adapting the projection algorithm for cone-beam ART

In the usual implementation of ART, a pixel value $p_i^{(k)}$ is computed by the ray integral:

$$p_i^{(k)} = \sum_{i=1}^{N} v_j w_{ij} = \sum_{i=0}^{N} v_j \bar{h}(r_i) \tag{5}$$

where r_i is the ray going from the source to image pixel i, and $\bar{h}$ is the interpolation kernel function h pre-integrated in the direction of ray r_i. The ART weight factor w_{ij} that determines the amount of influence of voxel v_j on the pixel sum $p_i^{(k)}$ is thus given by $\bar{h}(r_i)$.

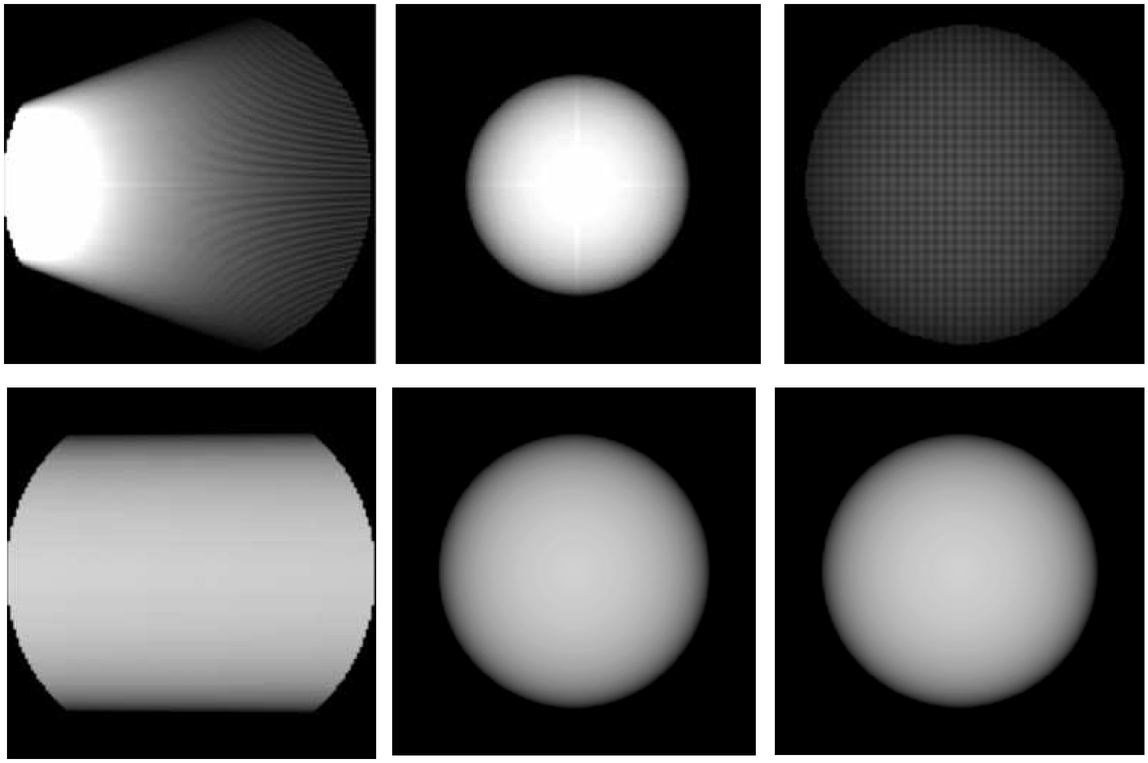


Fig. 3. Reconstruction of a solid sphere (diameter=0.75n) after one correction at φ=0˚ was applied : (a) cone angle γ=60˚, (b) parallel beam, γ=0˚. The side view shows a cut across the center of the volume along z, e.g. the direction of the cone beam. With z=z_c being the location of the volume center slice perpendicular to the cone direction (see also Fig. 4), the near slice is the volume slice at z=z_n=z_c-0.25n and the far slice is the volume slice at z=z_f=z_c+0.25n. Notice the uneven density distribution for the cone-beam reconstruction, while for parallel-beam the density is uniformly distributed..

Although in ART a volume is updated for each ray separately, it is convenient for our discussion to treat all rays that belong to one projection image as an ensemble and act as if grid correction is performed only after all image rays have completed their forward projection. Doing so allows us to use principles from sampling theory to explain and subsequently eliminate the reconstruction artifacts observed before. We admit that this approach is slightly incorrect since in ART the projection sum of a ray belonging to a particular projection image always contains the grid corrections performed by the previous ray(s) of the same projection image. Due to this circumstance the projections and corrections obtained with ray-based ART and image-based ART are not strictly the same. However, this simplification has only a minor effect on the correctness of our analysis.

Consider now Fig. 4 where the 2D case is illustrated. Here, the dashed lines denote the linear rays along which the volume is integrated. The rays that emanate from the source traverse the volume in form of a curved rayfront with radius $s(z)$. The rate ω_r=1/T_r at which the ensemble of rays samples the grid within a rayfront is a linear function of $s(z)$. The larger $s(z)$, the smaller the ray ensemble's grid sampling rate. At z=z_c the ray front samples the grid at exactly the grid sampling rate ω_g, e.g., ω_r=ω_g=1/T_g. However, for z<z_c, the rayfront sampling rate is higher than the grid sampling rate, while for z>z_c, the rayfront sampling rate is lower than the grid sampling rate. For our discussion, we approximate the curved rayfronts by planar rayfronts or slices. Thus, in Fig. 4, z_n is the location of the near slice of Fig. 3 and z_f is the location of the far slice.

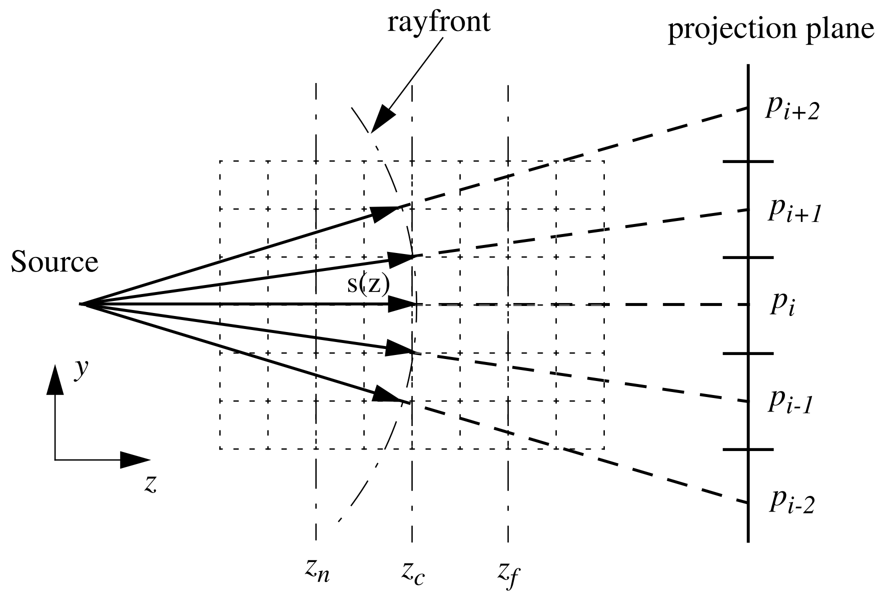


Fig. 4. Perspective projection for the 2D case. Rays emanate from the source into the volume along a rayfront of radius $s(z)$. For our discussion, we approximate the curved rayfronts by planar rayfronts or slices. Then, z_c is the location of the volume center slice, z_n is the location of the near slice of Fig. 3 and z_f is the location of the far slice of Fig. 3.

We mentioned earlier that by placing an interpolation kernel h at each grid voxel j and scaling it by the grid voxel's value v_j, we obtain a field of overlapping interpolation kernels that reconstructs the discrete grid function f_s into a continuous function f. Let us now decompose the volume into an infinite set of parallel slices along z. The contribution of a voxel j to the function $g(z)$ represented by a slice is then given by the 2D intersection of its interpolation kernel and the slice, denoted $h(z)$, while the sum of all scaled kernel intersections $h(z)$ produces the continuous slice function $g(z)$. A ray integral for pixel value $p_i^{(k)}$ is then computed by sampling all slices in the ray direction along z, which changes equation (5) into:

$$p_i^{(k)} = \int_z \sum_{j=1}^{N} v_j w_{ij}(z) = \int_z \sum_{j=1}^{N} v_j h(z, r_i) \tag{6}$$

The rayfront as a whole produces a sampled slice image at each depth z (see Fig. 4). Hence, a complete projection image can be formed by adding these sampled slice images $g_s(z)$. This leads to an alternate expression for $p_i^{(k)}$:

$$p_i^{(k)} = \int_z g_s(z, iT_r) \tag{7}$$

Note that the ray sampling rate $1/T_r$ within each sampled slice image is not constant but a linear function of z.

The process in which an ensemble of rays in a rayfront at depth z generates a sampled slice image $g_s(z)$ can be decomposed into two steps:

1. Reconstruction of the discrete grid signal $f_s(z)$ into a continuous signal $f(z)$ by convolving $f_s(z)$ with the interpolation filter h.
2. Sampling $f(z)$ by a comb function with period T_r=$T_g z/z_c$.

This can be written as:

$$\begin{aligned} g_s(z, iT_r) &= comb\left(\frac{y}{T_r}\right) \cdot \left(f_s(z, kT_g) * h\left(\frac{y}{T_g}\right)\right) \\ &= \left(comb\left(\frac{y}{T_g z/z_c}\right) \cdot \left(f_s(z, kT_g) * h\left(\frac{y}{T_g}\right)\right)\right) \end{aligned} \tag{8}$$

In the frequency domain, this is expressed as follows:

$$\begin{aligned} G_s(z, v) &= \frac{1}{\omega_g z_c/z} comb\left(\frac{v}{\omega_g z_c/z}\right) * \left(\sum_{-\infty}^{\infty} F(v - k \cdot \omega_g) \cdot H\left(\frac{v}{\omega_g}\right)\right) \\ &= \sum_{-\infty}^{\infty} F(v - k \cdot \omega_r) = \sum_{-\infty}^{\infty} F(v - k \cdot \omega_g z_c/z) \end{aligned} \tag{9}$$

using the relationship

$$X \cdot comb(vX) = \sum_{i = \infty}^{\infty} \delta\left(v - \frac{k}{X}\right) \tag{10}$$

If the backprojection is performed correctly, we must assume that the grid contains frequencies of up to but not greater than $\omega_g/2$. Then, since h has bandwidth ω_g and subsequently $f=h*f_s$ has also bandwidth ω_g, there is a chance that the signal aliases of G_s overlap in frequency domain whenever $\omega_r<\omega_g$, i.e., $z>z_c$. Thus each slice with $z>z_c$ potentially contributes an aliased signal to the composite projection image.

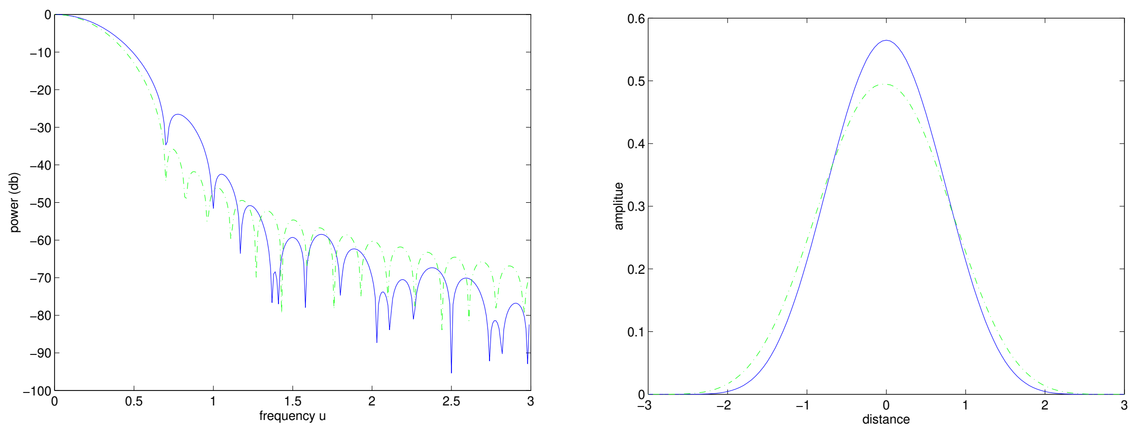
We can fix this problem by adding a lowpass filter L_P between H and the sampling process for all slices with $z>z_c$:

$$(z, v) = \frac{1}{\omega_g z_c/z} comb\left(\frac{v}{\omega_g z_c/z}\right) * \left(\sum_{-\infty}^{\infty} F(v - k \cdot \omega_g) \cdot H\left(\frac{v}{\omega_g}\right) \cdot L_P\left(\frac{v}{\omega_g z_c/z}\right)\right) \qquad z > z_c \tag{11}$$

An efficient way of implementing this concept is to pre-convolve H with a lowpass filter L_p, say a boxfilter B of width z/z_c, and use this new filter HB in place of H for interpolation. An alternative method is to simply decrease the bandwidth of H to $\omega_g z_c/z$. which gives rise to a filter H':

$$G_s(z, v) = \frac{1}{\omega_g z_c / z} comb\left(\frac{v}{\omega_g z_c / z}\right) * \left(\sum_{-\infty}^{\infty} F(v - k \cdot \omega_g) \cdot H\left(\frac{v}{\omega_g z_c / z}\right)\right) \qquad z > z_c \tag{12}$$

This technique was also used in [28] to achieve accurate perspective volume rendering with the splatting technique. The frequency response of *H'* is shown in Fig. 5a for the slice at $z=z_f$ where $z_f/z_c=1.42T_g$. The frequency response of *HB* is also shown. We notice that the frequency responses of both filters are similar and we also observe that both effectively limit the bandwidth to $\omega_r=0.7\omega_g$. Note, however, that although reducing the bandwidth of *h* for $z>z_c$ properly separates the signal aliases, the upper frequency bands of the grid signal in these regions have been filtered out by the lowpass operation and will not contribute to the projection images.



(a) (b)

Fig. 5. (a) Frequency response of the interpolation filter *H'* and the combined filter *HB*, at $z/z_c=1.42T_g$, (b) Impulse response of the same filters *h'* and *hb* at $z/z_c=1.42T_g$ (dash-dot *h'*, solid:*hb*).

According to the Fourier scaling property, *h'* is obtained from *h* by stretching and attenuating it in the spatial domain. The sampled slice signal is then:

$$g_s(z, iT_r) = comb\left(\frac{y}{T_r}\right) \cdot \left(f_s(z, kT_g) * \frac{1}{z/z_c} h\left(\frac{y}{T_g z/z_c}\right)\right) \qquad z > z_c \tag{13}$$

This filter *h'* is shown in Fig. 5b and is also contrasted with *hb*. The spatial extent of *hb* is $|hb|=|h|+|b|$, while the spatial extent of *h'* is $|h'|=|b||h|=z/z_c|h|$ (for $T_g=1.0$). Thus for $|b|<|h|/(|h|-1)$, $|h'|<|hb|$. Therefore, if $|h|=4.0$, then as long as $z/z_c<1.33$, *h'* is more efficient to use for interpolation. Since the majority of all interpolations occur in this range, the use of *h'* is to be preferred.

Finally, although one must use a stretched and attenuated version of *h* to lowpass *f* before sampling it into g_s when $z>z_c$, one cannot use a narrow and magnified version of *h* when $z<z_c$. Doing so would increase *h*'s bandwidth above ω_g and would introduce higher order aliases of f_s into the reconstructed *f*. Hence, we must use *h* in its original width when $z<z_c$.

Not only the ART projection step is affected by the non-uniform grid sampling rate of the cone-beam rays, the backprojection step must also be adapted. This is discussed next.

3.2.2 Adapting the backprojection algorithm for cone-beam ART

In backprojection, just like in forward projection, the volume is traversed by an ensemble of divergent rays. However, in backprojection the rays do not traverse the volume to gather densities, instead they deposit (corrective) density into the volume. In other words, the volume now samples and interpolates the ensemble of (correction) rays instead of the rays sampling and interpolating the volume. For the following discussion recall that, for convenience, we assume that all N_r rays of a projection P_φ first complete their projection and then all simultaneously correct the volume voxels. Let us now again decompose the volume into an infinite set of slices along the z axis, oriented perpendicular to the beam direction, and consider the corrections for the voxels within each slice separately. Each ray r_i carries with it a correction factor $corr_i$, computed by the nominator of equation (3). Like in the projection phase, we use $h(z)$ as the interpolation filter within a slice. Then the total correction dv_j to update v_j is given by the sum of all ray corrections $corr_i$, $1 \le i \le N_r$, for voxel j within a slice, integrated over all slices. This gives rise to the following equation which is similar to equation (6):

$$dv_j = \int_z \sum_{i=1}^{N_r} corr_i \cdot w_{ij}(z) = \int_z \sum_{i=1}^{N_r} corr_i \cdot h(z, r_i) \tag{14}$$

This equation is in line with the traditional viewpoint of ART. Let us now consider an alternative representation that will help us to explain the aliasing artifacts of cone-beam. We observe from Fig. 4 that the intersection of the ensemble of rays with a slice, say the one at $z=z_f$, gives rise to a discrete image $c_s(z)$ with the pixel values being the correction factors $corr_i$. Note that the pixel rate in the $c_s(z)$ is a linear function of z. In order to compute the correction factors dv_j, the voxel grid then samples and interpolates each of the slice images $c_s(z)$ into a correction image $d_s(z)$ with constant pixel period T_g, and each voxel j integrates all its correction factors in the $d_s(z)$ along z. Again, we use $h(z)$ as the interpolation filter within a slice, only now it is indexed by the discrete voxel locations and not by the traversing rays. The correction dv_j for voxel j can then be expressed as follows:

$$dv_j = \int_z c_s(z, iT_r)\, h(z, v_j) = \int_z d_s(z, jT_g) \tag{15}$$

Let us concentrate on one interpolated correction image $d_s(z)$. Since the interpolated signal is now the ensemble of rays and the sampling process is now the volume grid, the roles of ω_r and ω_g in equations (8) and (9) are reversed. The interpolation of the rayfront slice image $c_s(z)$ into a correction image $d_s(z)$ by the voxel grid can be decomposed into two steps:

1. Reconstruction of the discrete correction rayfront signal $c_s(z)$ into a continuous signal $c(z)$ by convolving $c_s(z)$ with the interpolation filter h. Notice that, for now, in order to capture the whole frequency content of c_s, we set the bandwidth of h to the bandwidth of the rayfront grid, $1/T_r = z_c/zT_g$. This is a new approach to ART, as usually the bandwidth of h is always set to ω_g, along with an amplitude scale factor of 1.0.
2. Sampling $c(z)$ into $d_s(z)$ by a comb function with period T_g.

This is formally written as:

$$\begin{aligned} d_s(z, jT_g) &= comb\left(\frac{y}{T_g}\right) \cdot \left(c_s(z, kT_r) * h\left(\frac{y}{T_r}\right)\right) \\ &= comb\left(\frac{y}{T_g}\right) \cdot \left(c_s(z, kT_g z/z_c) * h\left(\frac{y}{T_g z/z_c}\right)\right) \end{aligned} \tag{16}$$

In the frequency domain, this is expressed as follows:

$$\begin{aligned} D_s(z, v) &= \frac{1}{\omega_g} comb\left(\frac{v}{\omega_g}\right) * \left(\sum_{-\infty}^{\infty} C(v - k \cdot \omega_r) \cdot H\left(\frac{v}{\omega_r}\right)\right) \\ &= \sum_{-\infty}^{\infty} C(v - k \cdot \omega_g) = \sum_{-\infty}^{\infty} C(v - k \cdot \omega_r z / z_c) \end{aligned} \tag{17}$$

again using the relationship of (10). Since C's bandwidth is always ω_r, constrained by H, we get overlapping aliases in D_s when $z<z_c$. However, when $z \geq z_c$, no overlap occurs. This means that only for $z<z_c$ we need to limit H to the bandwidth of the grid, ω_g, resulting in a filter H'. For all other z, we use the bandwidth of the rayfront ω_r. More formally:

$$\begin{aligned} D_s(z, v) &= \frac{1}{\omega_g} comb\left(\frac{v}{\omega_g}\right) * \left(\sum_{-\infty}^{\infty} C(v - k \cdot \omega_r) \cdot H\left(\frac{v}{\omega_g}\right)\right) & z < z_c \\ &= \frac{1}{\omega_g} comb\left(\frac{v}{\omega_g}\right) * \left(\sum_{-\infty}^{\infty} C(v - k \cdot \omega_r) \cdot H\left(\frac{v}{\omega_r z / z_c}\right)\right) & z < z_c \end{aligned} \tag{18}$$

Again, according to the Fourier scaling property, h' is obtained from h by stretching and attenuating it in the spatial domain. The sampled slice signal is then:

$$\begin{aligned} d_s(z, jT_g) &= comb\left(\frac{y}{T_g}\right) \cdot \left(c_s(z, kT_r) * \frac{1}{z_c / z} h\left(\frac{y}{T_r z_c / z}\right)\right) & z < z_c \\ &= comb\left(\frac{y}{T_g}\right) \cdot \left(c_s(z, kT_r) * \frac{1}{z_c / z} h\left(\frac{y}{T_g}\right)\right) & z < z_c \end{aligned} \tag{19}$$

Thus for $z<z_c$ the bandwidth of H' is ω_g and the scaling factor in the spatial domain is z/z_c. Reducing the amplitude of h when $z<z_c$ prevents the over-exposure of corrective density in the volume slices near the source (as displayed in Fig. 3a). On the other hand, for $z>z_c$, the bandwidth of H' is $\omega_r=\omega_g z_c/z$ and the scaling factor in the spatial domain is 1.0. Increasing the width of h when $z>z_c$ prevents the under-exposure of some voxels in the volume slices far from the source and facilitates a homogenous distribution of the corrective energy. If we used a more narrow filter for $z>z_c$ then we would introduce the higher order aliases of $c_s(z)$ into the reconstructed $c(z)$, which is manifested by the aliased grid-like pattern in the far-slice of Fig. 3a.

Note that if projection and backprojection are performed with the proper filtering as described here, then the reconstruction volume will never contain frequencies greater than ω_g. Thus our previous argument that the grid projection may assume that the volume frequencies never exceed ω_g is correct. Of course, there is a small amount of aliasing introduced by the imperfect interpolation filter, but these effects can be regarded negligeable as the Bessel-Kaiser filter has a rather fast decay in its stop-band (see Fig. 5a).

3.2.3 Putting everything together

Since we are using an interpolation kernel that is pre-integrated into a 2D footprint table we cannot represent the depth dependent kernel size accurately and still use the same footprint everywhere in the volume. An accurate implementation would have to first distort the 3D kernel with respect to the depth coordinate z and then integrate it along z. Since the amount of distortion changes for every z, we would have to do this process for every voxel anew. Since this is rather inefficient, we use an approximation that utilizes the generic footprint table of an undistorted kernel function and simply stretches it by z_v/z_c whenever appropriate. Here, z_v is the depth coordinate of the voxel center. Thus the volume regions covered by the voxel kernel for which $z<z_v$ are lowpassed too much, while for $z>z_v$ the amount of lowpassing is less than required. However, since the kernel extent is rather small and the volume bandlimit rarely reaches the Nyquist limit, this error is not significant.

Another issue is how one goes about estimating the local sampling rate ω_r of the grid of rays. Highest accuracy is obtained when the curved representation of the rayfronts is used (see Fig. 4). In this case one would compute the normal distance between the set of neighboring rays traversing the voxel for which the kernel needs to be stretched. An easier way that is nevertheless a good approximation is to just use the voxel's z-coordinate in conjunction with the method of similar triangles to estimate ω_r. This approximates the curved rayfronts to planar rayfronts, as we did in our theoretical treatment. Various fast algorithms for the estimation of ω_r are subject of another paper [21].

Since the interpolation kernel is stretched and scaled differently for projection and backprojection, the computed weights in these stages are different as well. Thus equation (3) changes to:

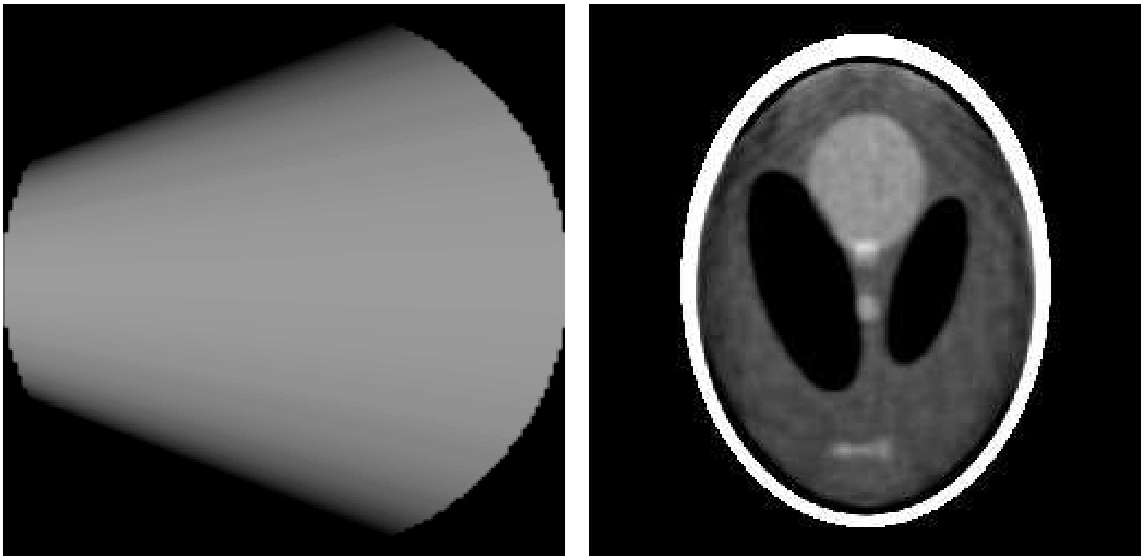
$$v_j^{(k+1)} = v_j^{(k)} + \lambda \frac{p_i - \sum_{n=1}^{N} w_{in}^F v_n^k}{\sum_{n=1}^{N} w_{in}^F w_{in}^B} w_{ij}^B \tag{20}$$

Here, w_{in}^F is the weight factor used for the forward projection and w_{in}^B is the weight factor used for backward projection.

Fig. 6 shows images of a reconstruction obtained with the same parameters settings than Fig. 2a, but using the new approach for cone-beam ART. Fig. 6a shows the cone-cut of Fig. 3a after one projection was processed. As was to be expected, the correction cone has now uniform intensity. Finally, the reconstructed slice of Fig. 1a is shown in Fig. 6b. Here we observe that this reconstruction no longer displays the strong aliasing artifacts that dominated Fig. 2a.

4. 3D Cone-Beam Reconstruction with SART

Let us now investigate SART and its behavior in the cone-beam setting. The bracketed part in the numerator of equation (4) is equivalent to the numerator of equation (3). Hence, the projection process of SART is identical to the one of ART and Section 3.2.1 applies again. However, SART's backprojection process differs from the one of ART. In contrast to ART, in SART, after the ray grid has been interpolated by the volume grid, a voxel correction is first normalized by the sum of influential interpolation filter weights before it is added to the voxel. This changes equation (14) to:



(a) (b)

Fig. 6. Reconstruction with ART utilizing the new variable-width interpolation kernels: (a) The cone-cut of Fig. 3a after one projection was processed: we observe a uniform distribution of the correction factors. (b) The slice of Fig. 1a (λ=0.08, γ=60°, $\boldsymbol{V^{(0)}=0}$, 3 iterations): the cone-beam reconstruction artifacts apparent in Fig. 2a were succesfully eliminated by using the new filter.

$$dv_j = \frac{\int_z \sum_{i=1}^{N_r} corr_i \cdot w_{ij}(z)}{\int_z \sum_{i=1}^{N_r} w_{ij}(z)} = \frac{\int_z \sum_{i=1}^{N_r} corr_i \cdot h(z, r_i)}{\int_z \sum_{i=1}^{N_r} h(z, r_i)} = \frac{\sum_{i=1}^{N_r} corr_i \cdot h(r_i)}{\sum_{i=1}^{N_r} h(r_i)} \tag{21}$$

We will see that this provides for a smoothly interpolated correction signal, even when the bandwidth of the interpolation filter is a constant ω_g, as is the case in the traditional ART and SART approaches. Let us now illustrate this behavior by ways of an example (see Fig. 7). Here, we utilize a uniform discrete correction signal (spikes, $c_s(z)$ in section 3.2.2), a 1D version of the one used in Fig. 3, and a linear interpolation filter h (dotted line) for ease of illustration. In this figure, we show the result of the signal reconstruction with h, i.e., the signal $c(z)$ *prior* to sampling into $d_s(z)$ by the voxel grid. Consider now Fig. 7a, where the situation for $z<z_c$ is depicted. In this case, the uniform correction signal has a rate $\omega_r>\omega_g$. In accordance to our previous results, we see that if $c_s(z)$ is reconstructed with traditional ART utilizing an h with constant bandwidth ω_g, we obtain a signal $c(z)_{\mathrm{ART}}$ with an amplitude much higher than the original correction signal. We know that we can avoid this problem by reducing the interpolation filter magnitude, which gives rise to the correctly scaled signal $c(z)_{ART}^{\text{var. kern.}}$. The correction signal reconstructed with SART, $c(z)_{\mathrm{SART}}$, on the other hand, does not require a reduction of the interpolation filter amplitude, since for each voxel the sum of interpolated ray corrections is normalized by the sum of the respective interpolation filter weights. Thus, with SART the reconstructed correction signal has always the correct amplitude. For $z>z_c$ (see Fig. 7b) when $\omega_r<\omega_g$, we observe an aliased signal of half the grid frequency for $c(z)_{\mathrm{ART}}$ instead of the expected constant signal. Sampling $c(z)_{\mathrm{ART}}$ into $d_s(z)$ would then result in a gridded pattern, similar to the one observed in the far slice of Fig. 3a. We know, however, from the previous discussion that we can eliminate this effect by widening h with respect to ω_r. This yields the correct signal $c(z)_{ART}^{\text{var. kern.}}$. For SART, on the other hand, a smooth and uniform $c(z)_{\text{SART}}$ of correct amplitude is obtained even with the original interpolation filter width - again due to the normalization step. Consider now Fig. 8a which shows the actual cone-cut of Fig. 3a for SART. We see that the distribution of correction energy is homogeneous, even though the bandwidth of the interpolation filter was set to ω_g everywhere in the volume. Finally, Fig. 8b shows a reconstruction obtained with SART under the same conditions than the previous ART reconstructions. We observe that no significant reconstruction artifacts are noticeable. Thus we can conclude that the SART approach is inherently more adequate for algebraic cone-beam reconstruction than the traditional, unmodified, ART approach.

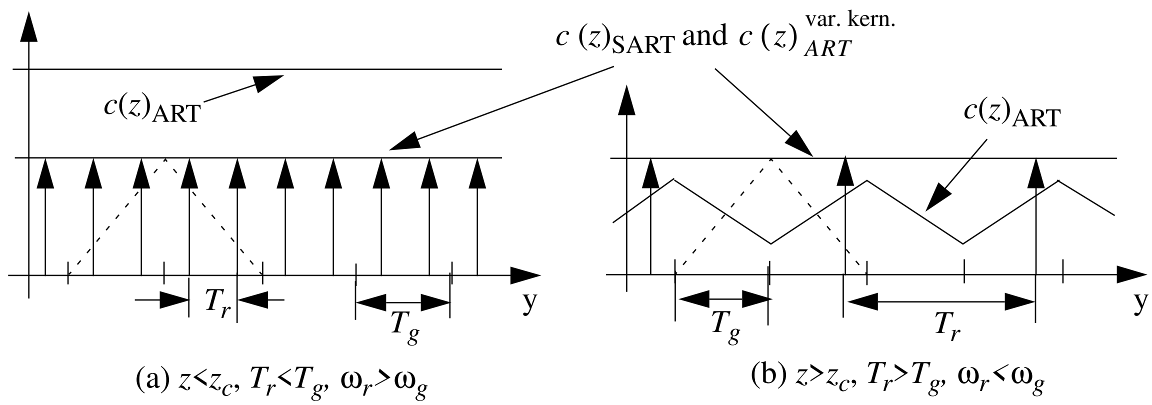


Fig. 7. Reconstructing a uniform discrete correction signal $c_s(z)$ (spikes) into the continuous signal $c(z)$ with ART and SART using a linear interpolation filter h (depicted here for a bandwidth of ω_g in a dotted line). Signal $c(z)_{\text{ART}}$ is the reconstruction of $c_s(z)$ with ART and the traditional interpolation filter of bandwidth ω_g. In case (a), when $z<z_c$, the correction signal is magnified, as in Fig. 3a, near slice. In case (b), when $z>z_c$, the correction signal is aliased, as in Fig. 3b, far slice. As previously discussed, we can avoid these effects using variable interpolation kernels, i.e., by scaling the amplitude of h in case (a) and by widening h in case (b), which yields the correct signal $c(z)_{ART}^{\text{var. kern.}}$ in both cases. The normalization procedure of SART prevents magnification and aliasing artifcats and thus in both cases (a) and (b) a correctly reconstructed signal $c(z)_{\text{SART}}$ identical to $c(z)_{ART}^{\text{var. kern.}}$ is produced.

5. Results

In order to assess the goodness of the reconstructions, we define two figures of merit: (i) the correlation coefficient CC between the original 3D Shepp-Logan brain phantom (SLP) and the reconstructed volume and (ii) the background coefficient of variation CV. These measures were also used by Ros et.al. [24].

We calculate CC within two selected regions of the SLP: (i) the entire inside of the head (without the skull) and (ii) an ellipsoidal volume containing the three small SLP tumors, as shown at the bottom of the brain section of Fig. 1a and described by ellipsoid 7, 8, and 12 in the SLP definition (see Table 1). While (i) measures the overall correlation of the reconstructed volume with the original, (ii) captures the sensitivity of the reconstruction algorithm to preserving small object detail. The CC is defined by:

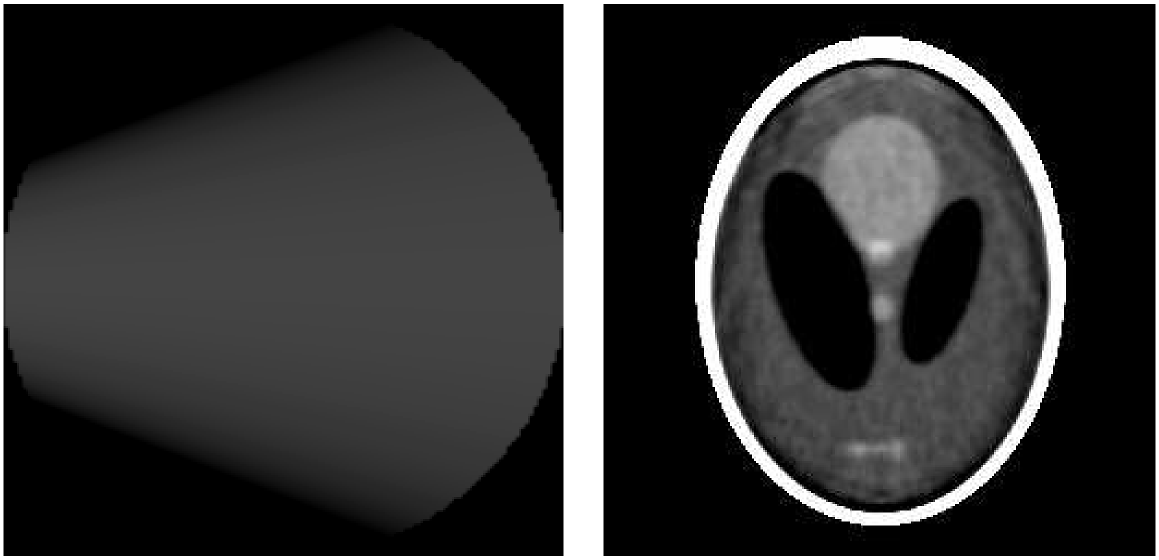


Fig. 8. Reconstruction with SART: (a) The cone-cut of Fig. 3a, after one projection was processed. (b) The slice of Fig. 1a (λ=0.3, γ=60°, $\boldsymbol{V^{(0)}=0}$, 3 iterations). The aliasing artifacts apparent in Fig. 2a do not exist.

$$CC = \frac{\sum_i (v_i - \mu_v)(o_i - \mu_o)}{\sqrt{\sum_i (v_i - \mu_v)^2 \sum_i (o_i - \mu_o)^2}} \tag{22}$$

where o_i and v_i are the values of the voxels in the original phantom volume and the reconstructed volume, respectively, and μ_o and μ_v are their means.

The background coefficient of variation CV is calculated within four ellipsoidal featureless regions located in different areas of the SLP. Since these regions are ideally uniform, CV represents a good measure for the noise present in the reconstructed SLP. The overall CV is the average of the four individual CVs and is defined as:

$$CV = \frac{1}{4}\sum_i^4 \frac{\sigma_i}{\mu_v} \tag{23}$$

where σ_i is the standard deviation of the voxel values within region i.

Using CV and CC to assess reconstruction quality, we have conducted a thorough study of the effects of various parameter settings:

- The initial state of the reconstruction volume: (i) zero, (ii) the average value of one of the projection images, properly scaled, and (iii) the average value of the SLP brain matter (e.g. 1.02). The last initialization method is hard to achieve in practice, since one usually does not know the average value of the reconstructed object in advance.
- The setting of the relaxation coefficient λ: experiments revealed that values in the range of [0.02,

0.5] yielded reconstructions that offered the best balance with respect to reconstruction quality and number of iterations. Therefore, reconstructions were obtained for λ=0.5, 0.3, 0.1, 0.08, 0.05, and 0.02.
- Time-weighted variation of λ during the first iteration: starting from 10% of the final value we have used: (i) immediate update to the final value, (ii) a linear increase and (iii) an increase due to a shifted cosine function (similar to [1]).
- The correction algorithm: (i) ART and (ii) SART.
- The size of the interpolation kernel: (i) constant and (ii) depth-dependent, as described in Section 3.2.
- The cone angle: 20˚, 40˚, and 60˚.
- The number of iterations: 1 through 10.

During our experiments we found that there is no significant difference in reconstruction quality whether λ is varied linearly or with a cosine function. Thus we only report results using a linear variation.

Fig. 9 shows a number of plots illustrating CC and CV for various reconstruction methods and parameter settings. Fig. 10 shows reconstructed SLP slice images after 3 iterations. The number of 3 iterations was chosen because the plots indicate that after this number of iterations both CC and CV have reached a level close to their final value. In both Fig. 9 and Fig. 10 the following terminology is used: Each parameter combination is described by a 3 digit code. The first digit codes the size of the reconstruction kernel (C for constant, V for variable, depth-dependent). The second digit codes the volume initialization method (- for zero, I for projection average, I* for SLP average). Lastly, the third digit codes the function at which λ is varied during the first iteration (- for no variation, G for linear increase). Reconstructions were obtained for each combination of parameter settings over the length of 10 iterations for all six settings of λ. In order to keep the amount of presented data manageable, we do not plot the effect of λ in Fig. 9. Instead, at every iteration and parameter combination we pick the lowest CV or highest CC, respectively, that was obtained in the set of six λ-dependent reconstructions. As the ideal λ is somewhat object-dependent anyway [12], this does not represent a serious trade-off. During the course of our experiments, we found that for ART and the SLP the λ that produces good reconstructions within four iterations is somewhere around 0.08, for SART this λ setting is around 0.3.

Fig. 9a-c shows the CC for the SLP tumors at a cone angle of 40˚ for the three main correction methods used: ART using the constant interpolation kernel, ART using the variable-size interpolation kernel, and SART using both kernels. In Fig. 9a we see that the reconstructions for G are always better than the respective ones without G. Since it is not any costlier to do G, the following plots will always use G. In Fig. 9b we see that reconstruction results with VIG are similar to the ones with VI*G. Since VI*G is not realistic anyway, we will use only VIG for the remaining discussion. In the same figure we also observe that V-G, VI- and VI*- yield only marginally worse results than VIG, thus we will eliminate these settings from further plots. Finally, in Fig. 9c, we see that SART C-G is either similar or always better than SART C--, and the same is true for SART V-G and SART V--. Thus we will be using only SART C-G and SART V-G. Also, preceding experiments revealed that SART is not dependent on the initial state of the volume. This is largely due to the circumstance that SART corrects the volume on an image basis, which provides a good initialization after the first image was spread onto the volume.

The following plots (Fig. 9d-l) illustrate the effects of the settings of the remaining parameters on

CC and CV. (Please use the legends inserted into Fig. 9f and i). In Fig. 9d, g, and j, we observe that for a 60° cone angle CC and CV improve significantly as volume initialization is made more accurate. This can also be seen in the reconstructions shown in the first column of Fig. 10. If the volume is initialized to zero, the small SLP tumors (and other portions of the image) are completely obliterated by aliasing artifacts (ART C--). The artifacts are reduced somewhat if the volume is initialized to the average value of one of the projection images (ART CIG). The artifacts are reduced further, but still well noticeable, if the volume is initialized to the average value of the SLP brain matter (ART CI*G). It is apparent that volume initialization alone cannot remove all artifacts. However, the reconstructions obtained with the variable sized interpolation kernel in conjunction with ART and the ones obtained with SART are all artifact-free. The plots support these observation with only SART and the ART V-methods having good CC and low CV. The plots also indicate that reconstruction quality increases when SART is used with a V-kernel instead of a C-kernel, however, the improvements are not large. At the same token, reconstruction quality also improves when the ART-V methods are used in conjunction with volume initialization and gradually increasing relaxation coefficient, but the rate of improvement is at a much smaller scale than in the ART-C case.

From the images in the second column in Fig. 10 we observe that for a cone angle of 40° considerable artifacts around the tumors still remain for ART C-- and ART CIG. Again notice the improvements with more accurate volume initialization. For ART CI*G, the artifacts are more attenuated but are still visible (however, note again that ART CI*G may not always be realizable). On the other hand, with ART VIG and SART C-G, the artifacts are completely eliminated. The plots of Fig. 9e, h, and k support these observations: the CCs are consistently higher, especially for small object detail like the SLP tumors, and the CVs are consistently lower with the ART V-methods and SART than for the ART C-methods (with ART CI*G being closest to ART V and SART).

The plots of Fig. 9f, i, and l indicate that for a smaller cone angle of 20° the differences between the methods are not as pronounced as for the larger cone angles, as long as one initializes the volume at least with the average projection image value. The circumstance that ART VIG and SART maintain a marginally better CC for the SLP tumors in a quantitative sense could be relevant for automated volume analysis and feature detection. However, in a visual inspection the differences are hardly noticeable, as indicated in the reconstruction images in the third column of Fig. 10.

6. Conclusions

This paper examined the use of ART-type methods to reconstruct 3D objects from a relatively sparse set of cone-beam projection data. In particular, we were interested in the reconstruction of low-contrast objects (such as the 3D Shepp-Logan brain phantom), which puts a high demand on the accuracy of the reconstruction algorithm. Although ART has been widely used for cone-beam reconstruction of high-contrast objects, e.g. in computed angiography, the low-contrast cone-beam case represents a new scenario for ART. One purpose of this paper was to present ART as an attractive alternative to the various FBP-type approaches that are mainly used in today's cone-beam reconstruction research. ART appears promising to compete with these methods due to the multitude of advantages it offers, such as better noise handling, its capability to work well even with a small number of projections and the opportunity to apply spatial and numerical constraints on the reconstructed object. Even though ART is iterative in nature, which was often equated with slowness of computation, our research indicates that really only three iterations are necessary to render

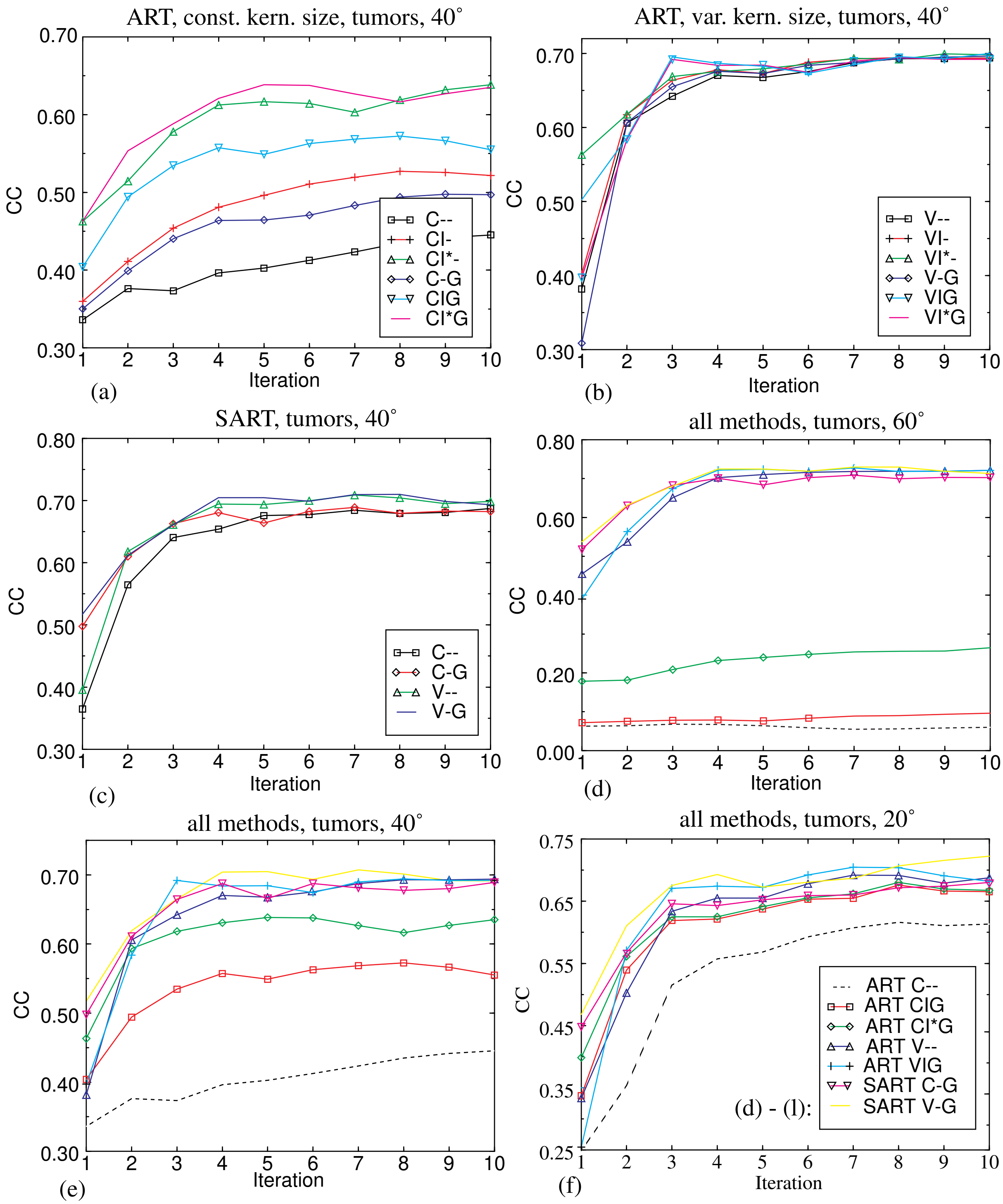


Fig. 9. Correlation Coefficient (CC) and Background Coefficient of Variation (CV) for ART and SART with constant and variable interpolation kernel size for 3 regions of the SLP (using 80 projections of 128^2 pixels each) Each reconstruction was performed for λ=0.5, 0.3, 0.1, 0.08, 0.05, and 0.02, the graphs plot the hightest CC/lowest CV of all λ settings at each iteration. (C: constant sized kernel, V: variable sized kernel, I: volume initialized to average of projection 0, I*: volume initialized to average of the SLP, G: linear increase of λ during the first iteration, -: volume initialization to zero or constant λ)

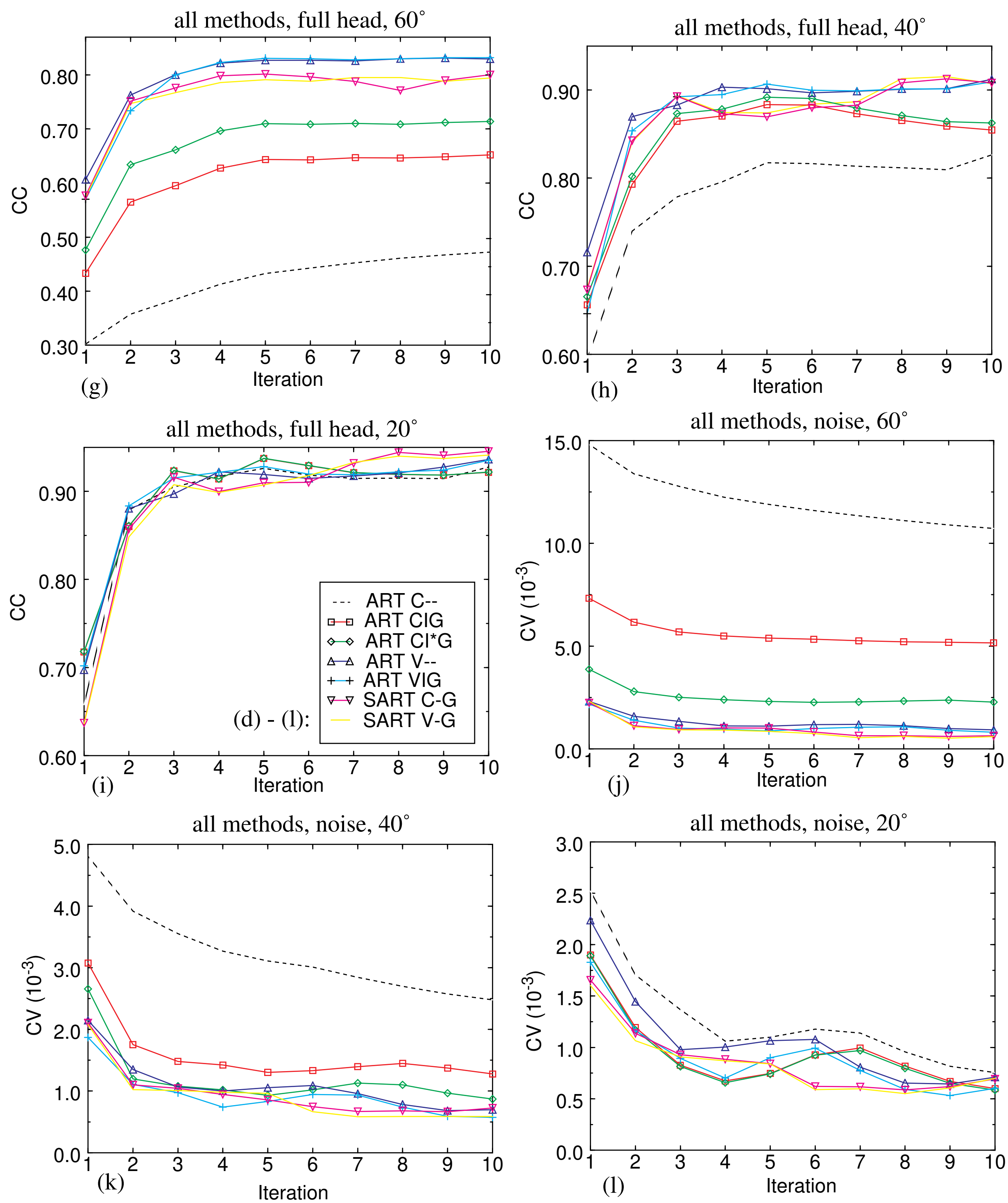


Fig. 15 (cont'd). CC and CV for ART and SART with constant and variable interpolation kernel size for three regions of the 3D Shepp-Logan phantom. (C: constant sized kernel, V: variable sized kernel, I: volume initialized to average of projection 0, I*: volume initialized to average of Shepp-Logan brain, G: linear increase of λ during the first iteration, -: volume initialization to 0 or constant λ)

Fig. 10. (next page): Slice of Fig. 1a reconstructed with various methods and cone angles

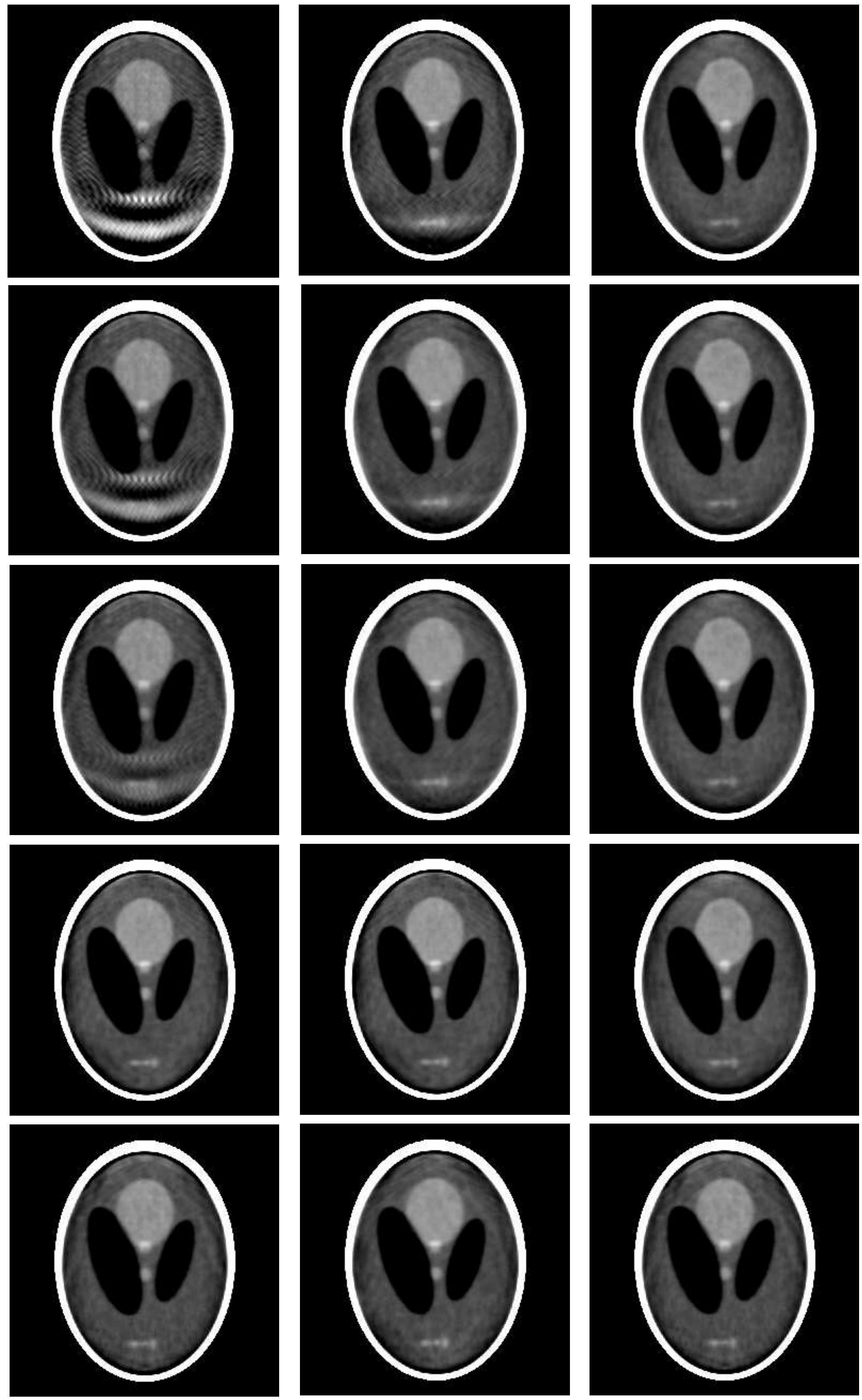

60˚ cone angle
40˚ cone angle
20˚ cone angle
ART C--
ART CIG
ART CI*G
ART VIG
SART C-G

Fig. 10

a reconstruction result close to the optimum (given proper parameter settings). Competitiveness with FBP is achieved due to the fact that ART requires substantially less projections than FBP and thus more than one iteration can be afforded.

With respect to accuracy, we found that ART's present mechanism of computing the weight coefficients was insufficient for cone angles greater than 20° in that it lead to strong aliasing artifacts in the reconstructed object. We showed examples where small object detail was completely obliterated by these aliasing artifacts. Even though these artifacts may have never been noticeable in high-contrast reconstructions, they become visible in the low-contrast case. To find the source of the artifacts, we analyzed the projection and correction process from a signal processing point of view and concluded that regions further away from the source were both sampled and updated below the Nyquist rate and needed to be filtered appropriately. In addition, regions close to the source received too much correction, leading to correction magnification artifacts. Subsequently, a new and efficient remedial solution that eliminates all artifacts was identified: the use of depth-dependent interpolation kernels. Using these kernels, the projection images no longer contain aliasing, and the correction energy is homogeneously distributed onto the reconstruction volume. This leads to reconstructions free of cone-beam related reconstruction artifacts.

Apart from the ART correction scheme, we also investigated the use of SART as an alternative correction algorithm. We found that SART, being a projection-based correction procedure, is also very suitable to prevent the aliasing artifacts of traditional ART. This can be attributed to the circumstance that SART, in contrast to ART, normalizes the voxel corrections by the sum of influential kernel weights before updating the voxel's value. This leads to a more homogeneous distribution of the correction image.

We then investigated the impact of volume initialization and the setting of the relaxation coefficient for the various methods developed. Our results indicate that for cone angles of 60°, as is often used for micro-tomography, strong aliasing artifacts prevail even with optimal volume initialization. These artifacts can only be eliminated when ART is used in conjunction with the depth-dependent kernel or by using SART. The same, though in a less dramatic way, is also true for cone angles of 40°, which are common-place in many clinical applications. For cone angles of 20° the aliasing effects are not visible with either method, but can still be measured numerically.

Our research suggests that SART conceptually represents a better correction scheme than the traditional ART approach, as it produces artifact-free reconstructions for all cone angles without the need for depth-dependent interpolation kernel sizes. However, SART's drawback is that it is not easily accelerated. As other research [21] has indicated, ART is considerably easier to speed up than SART, and runtime ratios $T_{\mathrm{SART}}/T_{\mathrm{ART}}$ of approximately 1.5 have been observed.

In the future, we plan to focus more on the computational efficiency of ART-type methods. In particular, we strive to make ART as efficient than FBP methods. Since grid projection and back-projection constitute the bulk of ART's calculations, efficient projections schemes are crucial to achieve this goal. We have already shown that no more than three iterations are necessary for a good quality reconstruction. Along with the fact that ART requires substantially fewer projections than FBP, an algorithm that reuses some of the early projection calculations for later projections and backprojections may have great promise to make three iterations of ART as efficient as one filtering step and one backprojection step with FBP.

References

[1] A. H. Andersen, "Algebraic Reconstruction in CT from Limited Views," *IEEE Trans. Med. Img.*, vol. 8, no.1, pp. 50-55, 1989.

[2] A.H. Andersen and A.C. Kak, "Simultaneous Algebraic Reconstruction Technique (SART): a superior implementation of the ART algorithm," *Ultrason. Img.*, vol. 6, pp. 81-94, 1984.

[3] C. Axelson, "Direct Fourier methods in 3D reconstruction from cone-beam data," *Thesis*, Linkoping University, 1994.

[4] R.N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed., McGraw-Hill, Inc., 1986.

[5] Z.H. Cho, J.P. Jones, and M. Singh, *Foundations of Medical Imaging,* Wiley, 1993.

[6] L.A. Feldkamp, L.C. Davis, and J.W. Kress, "Practical cone beam algorithm," *J. Opt. Soc. Am.*, pp. 612-619, 1984.

[7] P. Grangeat, "Mathematical framework of cone-beam 3D reconstraution via the first derivative of the Radon transform," *Proc. Mathematical Models in Tomography (Oberwohlfach, 1990), Springer Lecture Notes in Mathematics, 1497,* pp. 66-97.

[8] R. Gordon, R. Bender, and G.T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theoretical Biology*, vol. 29, pp. 471-481, 1970.

[9] H. Guan and R. Gordon, "Computed tomography using algebraic reconstruction techniques (ARTs) with different projection access schemes: a comparison study under practical situations," *Phys. Med. Biol*., no. 41, pp. 1727-1743, 1996.

[10] K.M. Hanson and G.W. Wecksung, "Local basis-function approach to computed tomography," *Applied Optics*, Vol. 24, No. 23, 1985.

[11] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. New York: Academic Press, 1980.

[12] G.T. Herman and L.B. Meyer, "Algebraic reconstruction can be made computationally efficient," *IEEE Trans. Med. Img.*, vol. 12, no. 3, pp. 600-609, 1993.

[13] A. Kaufman, Ed.*, Volume Visualization*, IEEE Press 1991.

[14] S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen," *Bull. Int. Acad. Pol. Sci. Lett*., A, vol. 35, pp. 335-357, 1937.

[15] A.C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.

[16] R.M. Lewitt, "Alternatives to voxels for image representation in iterative reconstruction algorithms," *Phys. Med. Biol*., vol. 37, no. 3, pp. 705-715, 1992.

[17] S. Matej and R.M. Lewitt, "Practical considerations for 3-D image reconstruction using spherically symmetric volume elements," *IEEE Trans. Med. Img.*, vol. 15, no. 1, pp. 68-78, 1996.

[18] S. Matej and R.M. Lewitt, "Efficient 3D grids for image reconstruction using spherically-symmetric volume elements," *IEEE Trans. on Nucl. Sci.*, vol. 42, no 4, pp 1361-1370, 1995.

[19] S. Matej, G.T. Herman, T.K. Narayan, S.S. Furuie, R.M. Lewitt, and P.E. Kinahan, "Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms," *Phys. Med. Bio*l, Vol. 39, pp. 355-367, 1994.

[20] E.J. Mazur and R. Gordon, "Interpolative algebraic reconstruction techniques without beam partioning for computed tomography," *Med. & Biol. Eng. & Comput.*, vol. 33, pp. 82-86, 1995.

[21] K. Mueller, R. Yagel and J.J. Wheller "Fast implementations of the Algebraic Reconstruction Technique (ART)," in review, 1997.

[22] K. Mueller, R. Yagel, and J.F. Cornhill, "The weighted distance scheme: a globally optimizing projection ordering method for the Algebraic Reconstruction Technique (ART)," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 223-230, April 1997.

[23] R. Ning and S.J. Rooker, "Image intensifier-based volume angiography imaging system: work in progress," *Proc. SPIE, vol. 1896, Medical Imaging 1993: Physics of Medical Imaging*, pp. 145-155, 1993.

[24] D. Ros, C. Falcon, I Juvells, and J. Pavia, "The influence of a relaxation parameter on SPECT iterative reconstruction algorithms," *Phys. Med. Biol.*, no. 41, pp. 925-937, 1996.

[25] B. Smith, "Image reconstruction from cone-beam projections: necessary and sufficient conditions and recon-

struction methods," *IEEE Trans. Med. Img.*, vol. 4, no. 1, pp. 14-25, 1985.

[26] D. Saint-Felix, Y. Trousset, C. Picard, C. Ponchut, R. Romeas, A. Rougee, "In vivo evaluation of a new system for 3D computerized angiography," *Phys. Med. Bio*l, Vol. 39, pp. 583-595, 1994.

[27] L.A. Shepp and B.F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. NS-21, pp. 21-43, 1974.

[28] J.E. Swan, K. Mueller, T. Moeller, N. Shareef, R. Crawfis, and R. Yagel, "An anti-aliasing technique for splatting," *Proceedings Visualization'97*, pp. 197-204, 1997.

[29] K. Tanabe, "Projection method for solving a singular system," *Numer. Math.*, vol. 17, pp. 203-214, 1971.

[30] L. Westover, "Footprint evaluation for volume rendering," *Computer Graphics (SIGGRAPH)* , vol. 24, no. 4, pp. 367-376, 1990.