Evaluating the Potential and Problems of Three-Dimensional Computed Tomography Measurements of Arterial Stenosis

David S. Ebert, David G. Heath, Brian S. Kuszyk, Linglan Edwards, Christopher D. Shaw, James Kukla, Ted Bedwell, and Elliot K. Fishman

Volume visualization is gaining widespread acceptance in medical applications. As its use increases, the issue of accuracy becomes critical. There have been very few studies examining the accuracy of volume rendering techniques. We studied the accuracy of hardware-assisted volume rendering for measurement of arterial stenosis in computed tomography (CT) data. The results of our study reveal that accurate measurements can be made from volume rendered CT data. However, error is present (absolute average error from 5.1% to 13.6%) and there is some variability, even for experts (standard deviation ranged from 4.8% to 15%). The evidence suggests that the choice of volume rendering (transfer function) parameters greatly affects the accuracy of the results. Accurate transfer function parameter selection is a difficult problem. Parameters that produce realistic images often provide inaccurate measurements. As the use of volume visualization grows and more inexperienced users begin using these tools for medical diagnosis and staging, new guidelines, aids, and techniques must be developed to ensure reliable, accurate visualization results.

Copyright © 1998 by W.B. Saunders Company

KEY WORDS: Volume visualization, accuracy, CT angiography

WITH THE WIDESPREAD availability of hardware-based volume rendering,¹ volume visualization is increasingly being used in many applications. However, there has been little work addressing the accuracy of volume visualization techniques and most of the work has concentrated on the accuracy of reconstruction filters and volume sampling.^{2,3} Accuracy in visualization is important in many applications. The visualization accuracy in diagnostic medical applications is critical, since patient treatment may be based on measurements from visualizations. We have conducted an experiment to determine if hardware texture-based volume visualization is accurate for a specific medical application: arterial stenosis (narrowing) from computed tomography (CT) data. Five radiologists at Johns Hopkins University Medical Center were asked to measure the amount of stenosis in nine phantom datasets.

Visualization of CT Data

With the introduction of spiral CT in the early 1990s, high quality three-dimensional (3D) CT datasets became available. The main advantage to spiral CT is its speed, allowing for single breath-hold scanning and for acquisition of a volume dataset during optimal enhancement from iodinated contrast injection.

Because the data acquisition times are short and true volume data sets can be generated, several new applications for CT are possible. The peripheral injection of an iodinated contrast agent intravascularly coupled with the timing of data acquisition has created a whole new application, commonly referred to as computed tomographic angiography (CTA). CTA has a number of potential advantages over conventional angiography: it is faster, less expensive, easier on the patient, and has fewer potential complications. Many institutions have, therefore, explored the option of replacing conventional examinations with spiral CTA. Such varied applications as evaluation of renal artery stenosis, renal transplant evaluation, liver transplant evaluation, Circle of Willis aneurysms, and carotid artery stenosis have all been studied with varying degrees of success. However, for spiral CTA to become the primary imaging study for vascular applications, it must be shown to be as accurate as the conventional study, which is currently the gold standard. Significant work has been done to optimize data acquisition protocols but relatively little work has been done to evaluate the accuracy of 3D rendering techniques. Several papers have compared shaded

From the Computer Science and Electrical Engineering Department, University of Maryland Baltimore County, Baltimore, MD; Department of Radiology and Radiological Science, The Johns Hopkins University, Baltimore, MD; and the Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada.

Address reprint requests to Elliot K. Fishman, MD, Professor of Radiology and Oncology, Department of Radiology and Radiological Science, The Johns Hopkins University, 601 N. Caroline St., Baltimore, MD 21287.

*Copyright © 1998 by W.B. Saunders Company
0897-1889/98/1103-0006\$8.00/0*

surface display (SSD) techniques to maximum intensity projection (MIP) techniques and have shown the latter to be more accurate in several applications. Yet there has been no publications comparing SSD to volume rendering techniques (VRT) in these specific applications. Furthermore, little work has been done with proven data sets to determine the true accuracy of any of the techniques. The potential problems with variability of results between different users or reliability of results with 3D or spiral data sets in these applications has been all but ignored.

We have previously looked at the accuracy of VRT in relation to bone phantoms and have shown the superiority of this technique. Drebin et al⁴ showed that for simulated femoral neck fractures with varying amounts of distraction, fracture gaps were accurately visible to 0.0 mm with VRT. Shaded surface or binary classification techniques caused spurious fusions of gaps <2.0mm or, at settings preserving these gaps, produced false holes in adjacent bone. Similarly, Kuszyk et al⁵ defined some of the advantages of VRT over shaded surface rendering in a variety of orthopedic applications.

Therefore, we decided to determine if VRT would be accurate in vascular imaging where display of information requires a quality image, both qualitatively (ie, one that clearly delineates potentially complex three dimensional anatomic relationships, such as those among the vessels in CTA) and quantitatively (accurate size measurement).

Arterial Stenosis

The use of spiral CTA for the evaluation of vascular stenosis has been explored for several specific applications. Dillon,⁶ in a review of 50 carotid arteries in 27 patients, found that CT angiography could provide most of the information needed before endarterectomy. The 3D technique used in that study was shaded surface rendering. Link,⁷ in a review of 28 patients found that spiral CTA was superior to color coded duplex sonography for carotid stenosis and in 89% of cases agreed with the results of catheter angiography. That study used maximum intensity projection (MIP) for the 3D technique. However, other articles are quick to point out the limitations of these 3D rendering techniques despite the impressive results in these two recent studies. Taskahashi et al⁸ recently re-

ported the artifacts and pitfalls of shaded surface display in spiral CTA of the carotid bifurcation.

Similarly, the results for evaluation of suspected renal artery stenosis with spiral CTA have been promising. Recent articles by Rubin et al⁹ have found spiral CTA to compare favorably with classic angiography. Rubin and colleagues have shown that MIP is favored over shaded surface display for accuracy in evaluating the renal arteries. Neither author had volume rendering available for use in their analyses.

The accuracy of the 3D rendering technique is critical if spiral CTA is to replace classic angiography. For example, in carotid stenosis, the North American Symptomatic Carotid Endarterectomy Trial (NASCET) classifies therapy based on degree of stenosis. In patients with less than 70% stenosis, medical management with careful follow-up is preferred. If the stenosis is 70% or greater, but less than 100%, then surgery is recommended. With 100% stenosis, nothing is done. Patients placed in the wrong category may either get needless surgery or, more importantly, be denied potential curative surgery. Similarly, in renal artery stenosis, depending on the degree of stenosis, patient management will be altered.

Experimental Design

We conducted an experiment to determine the accuracy of arterial stenosis measurements made from VRT images. Three contrast-filled phantoms, simulating varying amounts of stenosis, were scanned in three orientations as described below. VRT software was then used by five radiologists to produce arterial stenosis measurements of these datasets. The results of their measurements were analyzed to assess the accuracy of VRT for clinical assessment of CTA.

MATERIALS

Tubes with carefully controlled inner diameters were created to accurately simulate CT readings for arterial stenosis. The material used in the tube construction had a density of approximately 100 HU. The tubes were filled with a contrast material having a density of approximately 375 HU. We constructed in vitro 9.5 mm diameter vascular models with 33%, 66%, and 83% stenoses which we imaged with helical CT. Scans were performed with 2 mm collimation with a pitch of 1, a reconstruction interval of 1 mm, and a scan time of 1 second.

Each phantom was scanned with the tube in three orientations: in the scanning plane, perpendicular to the scanning plane, and at a 45 degree angle to the scanning plane. This provided us

with nine data sets. Each dataset was replicated to make a total of 18 datasets.

The VRT rendering system used in the experiment is a version of the Silicon Graphics *volren* program. The program was modified to support direct rendering of 12 bit data and to include a movable and rotatable two dimensional calibrated grid overlay which was used by the test subjects to measure the tube diameters.¹⁰ The renderer uses 3D texture mapping hardware to extract 2D slices at arbitrary orientations from volume datasets using trilinear interpolation. A user-adjustable transfer function maps the CT numbers to color and opacity, and the resulting slices are composited in the frame buffer. On high-end graphics hardware, such as the Silicon Graphics InfiniteReality, images can be interactively generated at 10 to 20 frames per second, allowing the user to fully explore the volume rendering parameter space and to optimize the parameters for best possible display.

For the experiment, surface shading was not used, and all trials were done with a monochromatic *window width and level* transfer function. This transfer function is similar to that available on most medical scanners and workstations. However, because volume rendering allows for partial obscuration of objects in the back of the image by objects in the front, two additional parameters are added. The user had control over 4 parameters of the transfer function: w (window width), l (window level), b (brightness), and a (opacity).

The brightness $L(v)$ and opacity $\alpha(v)$ of a voxel with value v

after application of the transfer function are given by:

$$L(v) = b \begin{cases} 0 & \text{if } v < l - w/2 \\ 1 & \text{if } v > l + w/2 \\ \frac{v - l + w/2}{w} & \text{otherwise} \end{cases} \quad (1)$$

$$\alpha(v) = a \begin{cases} 0 & \text{if } v < l - w/2 \\ 1 & \text{if } v > l + w/2 \\ \frac{v - l + w/2}{w} & \text{otherwise} \end{cases} \quad (2)$$

Clinical Evaluation Experiment

The experiment to assess clinical judgment was a complete, within-subjects design. The subjects were five radiologists at The Johns Hopkins University Medical Center who were experienced users of the VRT rendering system. Each subject measured the diameters of the normal and stenosed sections of our phantom arterial data sets for 18 cases. (To prevent the users from guessing actual dimensions, the absolute scale of the voxels in millimeters was assigned a random value, but the actual volume data and aspect ratios were not changed. Thus, the rendering process was the same for all datasets, but the

Fig 1. Two example volume rendered images from a 66% stenosed data set showing the range of images that can be produced from the same data.

Table 1. Measurement Results and Error Analysis for Clinical Study

Case #	Percent Stenosis						Error				
	Actual	Rad 1	Rad 2	Rad 3	Rad 4	Rad 5	Rad 1	Rad 2	Rad 3	Rad 4	Rad 5
3	33	21	29	28	37	37	-12	-4	-5	4	4
6	33	30	31	36	40	37	0	1	11	7	3
9	33	33	34	44	40	36	2	-6	7	7	-5
12	33	35	27	40	40	28	3	-3	0	7	17
17	33	36	30	33	40	50	9	2	13	7	4
18	33	42	35	46	40	37	-3	-2	3	7	4
1	66	58	60	69	75	77	-8	-6	3	9	11
4	66	61	60	75	83	80	2	-11	9	4	0
5	66	62	60	69	75	54	-2	-5	9	4	0
7	66	64	61	75	70	66	0	-10	9	0	0
16	66	66	56	75	66	66	-4	-6	3	9	-12
2	66	68	55	75	70	66	-5	-6	9	17	14
15	83	66	25	92	85	86	-8	-30	-3	-8	-6
8	83	75	53	80	75	77	-6	-27	5	-3	-3
10	83	77	56	88	80	80	7	-30	7	-8	-1
11	83	90	53	90	75	82	7	-20	1	1	-2
13	83	90	63	84	84	81	7	-17	7	1	-3
14	83	90	66	90	84	80	-17	-58	9	2	3
Signed Avg Error		-1.56	-13.2	5.39	3.72	1.56					
Absolute Avg Error		5.67	13.6	6.28	5.83	5.11					
Standard Deviation		7.08	15.1	4.89	6.12	7.10					

measurement tools of the renderer reported different values. The presentation sequence was randomly generated, with the restriction that no dataset was followed or preceded by another scaled replicant of itself.)

Each subject was asked to record the diameters of the phantom tubes before and after narrowing, as is the case with actual stenosis data. The radiologists were also asked to record the rendering setting used for each case. Each subject was presented these cases in the same randomly-generated order.

All measurements were made with a two-dimensional grid overlay. The users were instructed to rotate the grid overlay so that one axis was parallel to the distance to be measured. Figure 1 shows a sample volume rendering of a phantom dataset with 66% stenosis.

RESULTS

The measurements of the five radiologists for each of the 18 cases were collected and analyzed. These results are summarized in Table 1. This table also provides an error analysis of the data, which was derived by subtracting the known stenosis measurement from the measurement supplied by the subject. The mean absolute error in percent of stenosis for all subjects was 7.29 percentage points. Radiologist 2 consistently underestimated stenosis

for the 83% stenosed phantom tube. If the subject is excluded, mean absolute error in percent of stenosis for the remaining four radiologists was 5.7 percentage points. The standard deviation across all subjects was 10.9 percentage points.

The error data were analyzed by means of a 3-factor analysis of variance. The independent variables were actual degree of stenosis, scanning angle, and radiologist. The dependent variable is error in percentage points, or deviation from the actual stenosis percentage amount. The reason for including the radiologist as an independent variable is that only 10% of the variance can be explained if the radiologists are excluded from the analysis.

The $3 \times 3 \times 5$ analysis of variance yielded two significant main effects and one significant interaction. The first main effect was for radiologist ($F(4, 44) = 32.48, P < .001$). Follow-up analysis with Bonferroni pair-wise tests showed that Radiologist 2 significantly underestimated stenosis, while the others were grouped nearer 0.

The other main effect was for actual degree of

Table 2. 95% Confidence Intervals for the Mean for Each Radiologist at Each Level of Real Stenosis

% Stenosis	Rad 1	Rad 2	Rad 3	Rad 4	Rad 5
33	-7.6, 7.5	-5.0, 1.4	-2.1, 12.0	5.2, 7.4	-2.7, 11.8
66	-6.8, 1.0	-10.5, -4.8	3.4, 9.6	0.9, 12.9	-7.6, 11.8
83	-12.2, 9.1	-45.7, -14.7	-0.4, 9.5	-7.7, 2.7	-5.2, 1.4

Fig 2. Range of images chosen by radiologists (varying horizontally) for cases of 33%, 66% and 83% stenosis (varying vertically).

stenosis ($F(2, 44) = 23.19$, $P < .001$), with the 83% stenosis being underestimated by 6.3 percentage points, and the 33% and 66% stenoses being overestimated by 2.8 and 1.0 percentage points respectively. Scanning angle was not a significant factor.

The significant interaction was between actual degree of stenosis and radiologist ($F(8, 44) = 7.45$, $P < .001$). Radiologists 3, 4, and 5 tended to overestimate the 33% and 66% cases, and Radiologists 1, 4 and 5 slightly underestimated the 83% cases on average. Of more concern, the largest standard deviations occurred in the 83% stenoses.

In summary, a statistical analysis of our experimental data showed the following significant results:

- The mean absolute error in stenosis measurements was 7.29 percentage points with a standard deviation of 10.9 percentage points.
- One radiologist significantly underestimated stenosis, while the other four neither significantly overestimated nor underestimated stenosis.
- The 83% stenosis case was underestimated by 6.3 percentage points while the 33% and 66% stenoses were slightly overestimated (2.8 and 1.0 percentage points, respectively).
- There was large variation among radiologists, with three radiologists overestimating the 33% and 66% cases and three underestimating the 83% cases.

Table 3. Clinical Study Results for Images in Figure 2

Case Number	% Stenosis	Rad 1	Rad 2	Rad 3	Rad 4	Rad 5
18	33	42	35	46	40	37
1	66	58	60	69	75	77
8	83	75	53	80	75	77

- The largest standard deviation in stenosis measurement occurred in the 83% stenosis cases.

DISCUSSION

In CT angiography, accuracy is most important when it affects the staging of the stenosis. A measurement of 10% stenosis versus 20% stenosis is normally not clinically significant; however, a measurement of 65% versus 75% could affect the clinical stage and treatment (surgery versus medical management). A reasonable expectation of error would be ± 5 percentage points, but, as Table 2 shows, only 2 of the 15 entries indicate that this was achieved during the experiment. The 95% confidence interval for the mean is the range within which the real mean would occur 95% of the time. Radiologist 2 achieves the desired range in the 33% case, and Radiologist 5 almost achieves it for the 83% case. For the more relaxed range of ± 10 , 8 of the 15 cases would have achieved errors within this range.

If we use the NASCET as a model for our study, Radiologist 2 would have understaged the patients with high grade stenosis and moved them from the category of requiring surgery to watchful waiting. Based on current experience, this could be catastrophic. Radiologist 1 made a similar error on Case 15 that would have changed management. This study, although a simulated one with phan-

Table 4. Experimental Results Using the Actual Measurements

	Case 1	Case 2	Case 3
Actual	33	66	83
Test 1	33	78	100
Test 2	14	65	81

Fig 3. Images generated using known measurements for cases of 33%, 66% and 83% stenosis.

toms, clearly shows the importance of developing tools to control the variation between different users of volume rendering. Spiral CTA promises better patient care, less morbidity, lower cost, and is less invasive. However, if individual success with the technique can not be guaranteed, then the technique loses its value and the growth of this new technology will suffer. Successful imaging studies rely on reproducibility and consistency. These problems can be overcome if guidelines and aids for selection of accurate transfer function parameters are developed. (The transfer function converts the CT volumetric data to relative brightness and opacity in the image.)

Our data show that accurate measurements can be made from volume rendered CT angiography data. However, it also shows that there is high variability in accuracy even in highly trained radiologists who use volume visualization on a regular basis. Figure 2 shows the wide range of images used by each radiologist for measurements in three cases (33%, 66% and 83% stenosis). Table

3 shows the measurement results for each radiologist for the cases in Fig 2.

These results show that volume visualization is generally accurate for staging in a clinical setting. However, there is wide variation in images and measurements caused by the choice of different transfer function parameters.

Further Experiments

To further validate our results, we conducted an experiment with the same CTA phantom datasets to ascertain if a user could determine a set of VRT transfer function settings that would provide accurate answers for all cases of stenosis if he/she knew the actual tube measurements. These settings could then be used for all CTA datasets, reducing the measurement errors. The user selected transfer function parameters to achieve accurate measurements for both percent stenosis and actual tube diameters for one case, then applied those parameters to the other two cases and made measurements. The results are summarized in Table 4 and

Fig 4. Range of images generated by the volume raycaster for cases of 33%, 66% and 83% stenosis.

Table 5. Experimental Results from a Volume Raycaster

	Case 1	Case 2	Case 3
Actual	33	66	83
Raycaster	36	72	63

Fig 3. In Test 1, the measurements for the 33% stenosis case were used to pick the best transfer function parameters. For Test 2, the measurements for the 83% stenosis case were used to pick the best transfer function parameters. These results indicate that the choice of a set of transfer function parameters that provide accurate measurements for one case may provide erroneous measurements for other cases. In Test 1, the other cases were off by an average of 15 percentage points. In Test 2, the 66% case was accurate; however, the 33% case was underestimated by 19 percentage points. These results were obtained under a relatively ideal setting, unlike actual cases where there is variation in signal strength for each case and the solution is unknown.

We have also experimented with making stenosis measurements from a different VRT software system, such as a volume raycaster. As can be seen in Fig 4 and Table 5, initial results with this technique indicate similar difficulties in choosing accurate transfer function parameters.

These results give further evidence that the choice of correct VRT transfer function parameters to provide accurate images is a difficult, open problem. This choice must take into consideration perceptual issues involved in the visual system's ability to detect edges and interpret shading. The

selection of transfer function parameters, in most cases, is an under-constrained problem where many choices may give correct results for one case, but not give correct results in other cases.

It is worth noting that, in any system which gives the user control of imaging parameters, potential for error is present. Even when viewing individual CT slices, improper choice of window width and level can result in misleading displays. Three dimensional rendering techniques (MIP, VRT, and SSD) introduce additional controls leading to other potential pitfalls. For example, a MIP image generated from one direction might completely hide a stenosis, while the stenosis might be quite apparent from another direction. The extraction of surfaces needed for SSD display usually requires threshold selection. As we discussed earlier, this can involve a tradeoff between loss of detail and erroneous closing of gaps.

In our experience, users of our VRT system often make the implicit assumption that if they select a set of rendering parameters to clearly and accurately convey three dimensional structural information, then these parameters will also be effective for making quantitative measurements. Our experiments show that this is not entirely untrue; that is, volume rendered CTA images can provide accurate measurements for arterial stenosis. However, with current techniques, the potential for error is present. The clinical success of these techniques depends on improved accuracy, reproducibility, and consistency through the development of guidelines and aids for selecting accurate transfer function parameters.

REFERENCES

1. Cabral B, Cam N, Foran J: Accelerated volume rendering and tomographic reconstruction using texture mapping hardware, in Kaufman A and Krueger W, (eds): 1994 Symposium on Volume Visualization. New York, NY, ACM Siggraph, 1994, pp. 91-98.
2. Machiraju R, Yagel R: Reconstruction error characterization and control: A sampling theory approach. *IEEE Trans Visual Comput Graph* 2:364-378, 1996
3. Marschner SR, Lobb RJ: An evaluation of reconstruction filters for volume rendering. *Proc IEEE Visual* 94:100-107, 1994
4. Drebin RA, Magid D, Robertson DD, et al: Fidelity of three dimensional CT imaging for detecting fracture gaps. *J Comput Assist Tomogr* 13:487-489, 1989
5. Kuszyk BS, Heath DG, Bliss DF, et al: Skeletal 3-D CT: Advantages of volume rendering over surface rendering. *Skeletal Radiol* 25:207-214, 1996
6. Dillon E, Leeuwen MV, Fernandez M, et al: CT angiography: Application to the evaluation of carotid artery stenosis. *Radiology* 189:211-219, 1993
7. Link J, Brossman J, Penselin V, et al: Common carotid artery bifurcation: Preliminary results of CT angiography and color-coded duplex sonography compared with digital subtraction angiography. *AJR Am J Roentgenol* 168:361-365, 1997
8. Takahashi M, Ashtari M, Patel M, et al: CT angiography of carotid bifurcation: Artifacts and pitfalls in shaded surface display. *AJR Am J Roentgenol* 168:813-817, 1997
9. Rubin G, Dake M, Napel S, et al: Spiral CT of renal artery stenosis: Comparison of three-dimensional rendering techniques. *Radiology* 190:181-189, 1994
10. Fishman EK, Kuszyk BS, Heath DG, et al: Surgical planning for liver resection. *IEEE Computer* 29:64-72, 1995