Multi-view Stereo Ranging via Distributed Ray Tracing

Thayne R. Coffman[1,2]
[1]21st Century Technologies
Austin, TX, USA
tcoffman@21technologies.com

Alan C. Bovik[2]
[2]Dept. of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX, USA
bovik@ece.utexas.edu

Abstract— **We explore the use of Distributed Ray Tracing (DRT), an anti-aliasing technique from computer graphics, in multi-view computational stereo. As an example, we study ABM, a multi-view stereo algorithm based on a set of Hough transform accumulation operations. Augmenting ABM with DRT improves both internal signal quality and reconstruction accuracy. Results are given for both fundamental and complex "super-resolution reconstruction" tasks, where the voxel side length is less than the image ground sample distance. DRT improves ABM accuracy by 18% and can be generalized to improve other stereo algorithms.**

Keywords-computational stereo; multiview stereo; super-resolution; Hough transform; distributed ray tracing

I. INTRODUCTION

Passive 3D reconstruction from multiple images or video remains an attractive but challenging technical task after nearly 40 years of research. It is a key enabler for ubiquitous 3D modeling because it often has lower size, power, and weight requirements than active approaches.

Aliasing limits stereo reconstruction accuracy by corrupting internal signals and data. While important to traditional stereo reconstruction, its effect is even more pronounced in super-resolution (SR) reconstruction, where the generated 3D models have voxel sizes smaller than the ground sample distance (GSD) of the input imagery.

This paper describes how distributed ray tracing (DRT) is used to improve the Accumulation Based Modeling (ABM) multi-view stereo algorithm. The ABM example provides insight into how the use of DRT can be generalized to other computational stereo or computational geometry algorithms.

II. BACKGROUND

Computational stereo reconstructs 3D models of a scene by analyzing the scene's appearance in 2D images captured from multiple viewpoints. Background on multi-view computational stereo can be found in [1].

Two-frame methods focus on analyzing input images in pairs. The core problem is to determine correspondences between the pixels in the pair. This computation typically causes the majority of complexity and runtime, although various methods exist for improving its efficiency [2].

Multi-view methods process more than two images simultaneously. These methods often avoid correspondence matching by interpreting data in an object-centric 3D model. Some methods, like ABM, extend the Hough transform (HT) to fuse data in a voxel representation [3][4]. A survey of multi-view techniques is given in [5] and the HT in [6].

It is well known that the HT domain has strong aliasing artifacts. The ideal HT domain signal is non-bandlimited, and the necessary use of discrete accumulators (e.g., voxel models) results in aliasing [7].

Aliasing has been studied extensively in standard HT applications. The vast majority of anti-aliasing strategies use traditional signal processing techniques [8][9]. These are problematic for stereo because HT-domain trajectories vary with the camera flight path and voxel range. It is impossible to simultaneously avoid under-sampling or over-sampling HT representations of non-homogeneous objects [8].

HT-domain interference between nearby features has also been noted in standard applications and in HT-based stereo [3][4]. Reducing aliasing in the HT representation can improve the results of separate anti-interference techniques.

Aliasing limits the accuracy of non-HT-based algorithms as well. Among others, Kutulakos and Seitz note that the accuracy of the Space Carving algorithm can be increased "only up to the point where image discretization effects (i.e., finite pixel size) become a significant source of error" [10].

Ray tracing generates synthetic images by simulating the visual contributions of the scene along a ray from the camera origin, through each pixel, and into the scene (including reflections and refractions)[11]. Aliasing occurs because the algorithms represent non-bandlimited phenomena on digitized media using discrete algorithms.

Distributed ray tracing (DRT) [12] is an anti-aliasing technique from computer graphics that has yet to be exploited in other computational geometry algorithms. DRT casts multiple rays through each pixel, each with minor variations in direction. Pixel appearance is the averaged contributions from each ray. This improves the modeling of pixels, lenses, lights, and objects to provide superior renderings of gloss, translucency, shadows, and other "soft" visual effects. Most broadly, it is a technique for mitigating the effects and appearance of aliasing in computer graphics.

III. METHOD

DRT is used to improve how Accumulation Based Modeling (ABM) models and fuses sampled input images into sampled signals and voxel models by reducing aliasing. ABM is one of many computational geometry algorithms that can benefit from DRT.

A. Accumulation Based Modeling (ABM)

ABM computes scene structure in a 4-stage sparse-to-dense approach. It performs two fusion-interpretation cycles, first for sparse (wireframe) structure and then for dense. Evidence fusion uses principles of HT accumulation. Evidence interpretation uses specialized techniques.

Stage 1 begins by extracting features (typically edge pixels) from each image. Each feature adds evidence that 3D space is occupied somewhere along the ray from the camera origin through that pixel. Rays are cast through each feature pixel into a voxel model (Fig. 1). ABM walks the rays in small steps and accumulates the steps ending in each voxel to approximate the total length of ray segments that intersect it. This estimates the evidence in the input that is consistent with each voxel being occupied.

The voxel model is the quantized HT parameter space. Pixel rays are trajectories through HT space. Voxels are accumulators for a particular parameterization. As with other HT strategies, many rays will converge on occupied voxels in the wireframe reconstruction.

Stage 2 walks the same trajectories to interpret the fused evidence. Evidence interferes constructively in occupied voxels to form clear peaks. Fig. 2 shows an observed evidence signal relative to the ideal signal (derived below). Using a set of 1D envelope and peak estimation techniques, ABM identifies the most significant peak in each evidence signal. Additional tests are performed to reject inconclusive or badly behaved peaks. ABM increments a count in a second voxel model for each passing peak.

At the end of Stage 2, peak counts are thresholded to generate a wireframe 3D model. Thresholding peak counts requires the position of the occupied voxel to be consistent when evidence is interpreted from different viewpoints. This second accumulation yields more accurate reconstructions [3]. It follows similar HT extensions for line detection [13].

Stages 3 and 4 perform similar processing to convert the wireframe model to a dense model. The wireframe model is used to render a sparse range image from each viewpoint. Sparse ranges are interpolated to form dense range estimates, which are accumulated as in Stage 1. Stage 4 uses a variety of 3D morphological filters and thresholds to compute the dense reconstruction from the interpolated estimates.

B. Modifying ABM with Distributed Ray Tracing

Using DRT, ABM represents each feature pixel with a frustum instead of a ray. A frustum's cross section grows with increasing range from the camera. We model each cross section to have equal evidence, so *evidence density* decreases as range increases. Evidence is fused by integrating evidence density over the intersection of frustum and voxel so the voxel values approximate *evidence mass*.

An analytical computation would be very expensive, and would require solid geometry intersections, density calculation at many points, and 3D integration of a non-constant function over irregular volumes.

All of these quantities can be approximated using DRT. Instead of casting one ray per feature pixel, DRT casts many rays with small perturbations (Fig. 3). Increasing the number of rays improves the accuracy of the approximations.

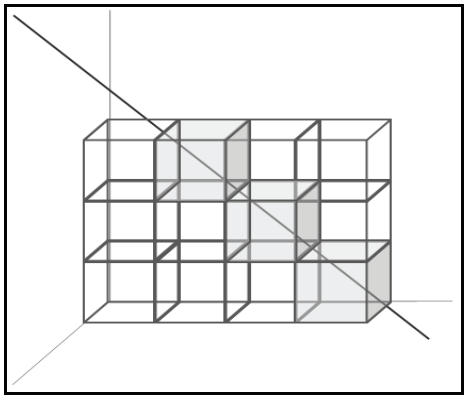

Fig. 1: ABM sparse ray casting without DRT

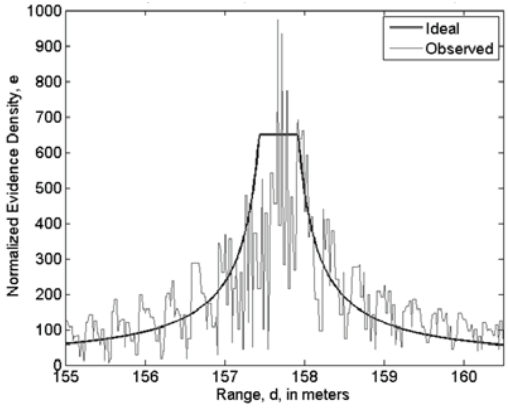

Fig. 2: Accumulated evidence without DRT

ABM is modified in three ways. First, DRT is used to accumulate the evidence provided by each pixel. Second, ABM uses the same ray distribution when extracting the accumulated evidence signal for peak detection. The evidence signal is the average of all voxel values at each range. Third, when a peak is detected, the same distribution is used to convert its range into voxel locations to accomplish peak accumulation.

Fig. 2 depicts a badly aliased evidence density signal generated without DRT. Fig. 4 shows an equivalent signal computed using DRT, and aliasing is clearly reduced.

The technique can be extended to incorporate camera parameter uncertainty. With perfect calibration, the frusta are known exactly. Parameter uncertainty alters the origin and orientation of each frustum. DRT can easily model distributions that exploit all available knowledge.

The technique is conceptually simple, fast, and requires only integer accumulations and integer voxel models. It is parallelizable, flexible, and a tractable way to approximate complex modeling and fusion computations.

C. Ideal Evidence Signals

In order to quantify the benefits of DRT, we define "ideal" evidence signals and compare observed evidence signals to them using signal-to-noise ratio (SNR).

Assume the scene has a single point object at point p, which is distance d_p from the camera. The point is visible in

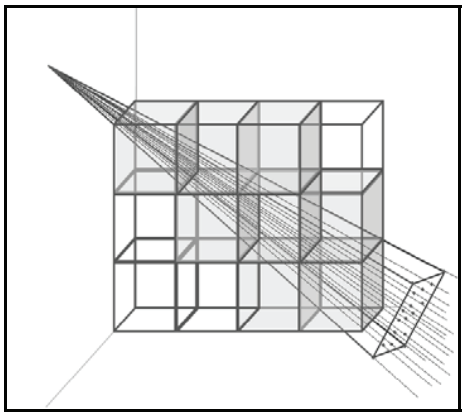

Fig. 3: ABM sparse ray casting with DRT

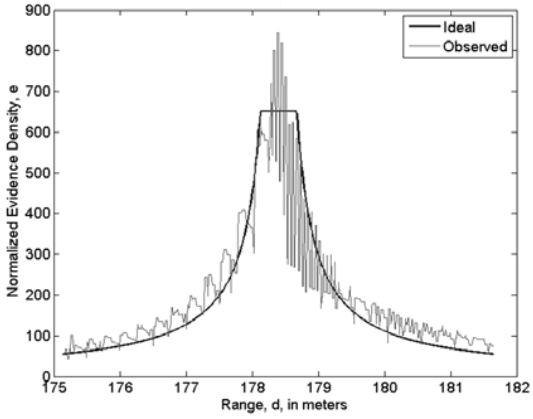

Fig. 4: Accumulated evidence with DRT

imagery captured along a linear flight path of length L_0 that is symmetric about p. We seek the evidence density $e^*(d)$ along a ray $\vec{r}$ from the camera to p, after all evidence has been accumulated. Evidence density on the flight path is set to 1.0m^{-1}. As d approaches d_p, $e^*(d)$ grows to infinity. While accurate, this is undesirable, so density is limited to e_{max}.

The ideal density $e(d) = e(x)$ for each point x on $\vec{r}$ at distance d from the camera origin is thus defined as

$$e(d) = \min\left\{\frac{L_0}{L(d)}, e_{\max}\right\} = \min\left\{\frac{d_p}{d_p - d}\ ,\ e_{\max}\right\}.$$

Ideal density signals are shown in Fig. 2 and Fig. 4. In theory, the choice of e_{max} is arbitrary. In practice, the voxel size, L_0, and other factors affect the maximum achievable density. In the results, e_{max} was set empirically.

The derivation can be applied to non-linear paths by defining a 2D manifold containing the flight path and occupied voxel. Points, distances, voxels, and densities are mapped from the linear derivation onto that manifold.

IV. Results

A. Fundamental Scenes

Internal signal quality is measured on a scene containing a point target as described above. Evidence from 200 images (720x480 pixels) is fused from viewpoints along a 75m flight path at ranges of 150-220m. The sensor's nominal ground sample distance is 0.08m. The modeling region is 7.0m x 4.7m x 4.5m, divided into cubic voxels 0.04m per side. This gives a super-resolution reconstruction with a GSD-to-voxel-side ratio of 2.0. Aliasing effects are significant.

Results are shown in Fig. 5 for ABM without DRT, with DRT on evidence accumulation (only), and with DRT on all three accumulations described in Section III.B. Without DRT, ABM achieves an average SNR of 7.67dB on its sparse evidence signals. DRT improves average SNR to as much as 13.00dB when it is used on all three sparse accumulation processes. Most benefits are achieved using 100-250 DRT rays (note the logarithmic X-axis scale).

Fig. 5 uses ±0.5 pixel perturbations from a uniform distribution. Other distributions and parameterizations were tested. Most distributions improve performance relative to operating without DRT. Best performance occurs with modest perturbations and requires parameter tuning. DRT performance degrades (eventually worse than the baseline) as perturbation magnitude becomes large.

B. Complex Scenes

Reconstruction accuracy is measured against sparse ground truth on a dataset provided by the Air Force Research Laboratory (Fig. 6). The camera path gives true ranges of 150-220m. Video was captured at 60 frames per second (interlaced) and 720x480 pixels resolution. Ground truth is known for 301 locations, including building corners, markers, and vehicles. Extrinsic camera parameters are known for each frame and intrinsic parameters are estimated. Analysis was based on a representative 200-frame sequence.

Inaccuracies in camera parameters led to the use of an atypical error metric. Mean absolute error (MAE) measured the distance to the closest 3D position estimate in a 5-pixel radius (E_5) near a ground truth point's true projection onto each frame. E_5 was selected by inspection based on re-projection error after parameter estimation.

Additional detail on dataset and metrics is given in [2].

Input video was downsampled to 180x120 resolution to yield a nominal GSD of 0.32m. The modeling region was 70m x 47m x 45m region, divided into voxels 0.20m on a side. This results in a super-resolution reconstruction with GSD-to-voxel-side ratio of 1.6.

Baseline ABM performance without DRT was E_5=0.71m. Thresholds and other algorithm parameters were set empirically. DRT was applied with 100 rays per pixel perturbed in a uniform distribution with ±0.5 pixels deviation in horizontal and vertical. Thresholds and other parameters were adjusted for the additional rays. Uniform distribution DRT achieved E_5=0.62m (a 15% improvement in accuracy). DRT was also applied with 100 rays per pixel with a blurred uniform distribution, which achieved E_5=0.58m (an 18% improvement in accuracy).

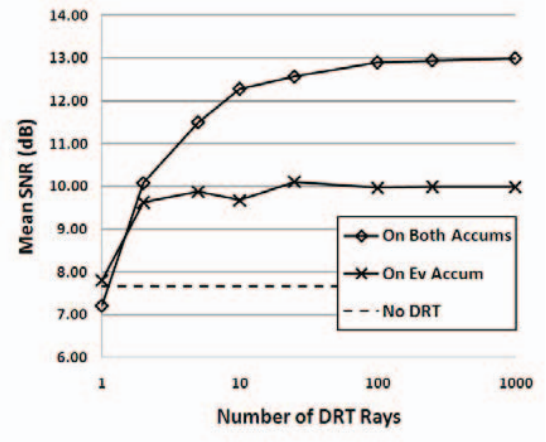

Fig. 5: DRT performance on fundamental scenes

Fig. 6: Example test data frame with ground truth locations marked

V. Conclusions

DRT was shown to improve the performance of ABM, which is a HT-based multi view stereo approach. DRT reduces aliasing to improve internal ABM signal quality by up to 5dB. The effects of aliasing become more significant as voxel sizes are reduced to achieve super-resolution 3D reconstruction. DRT provided an 18% improvement in ABM accuracy on a complex super-resolution reconstruction dataset. It was most effective and efficient when using 100-250 rays per pixel.

The use of DRT generalizes to other stereo and computational geometry algorithms. Not every algorithm will benefit, but many can use DRT to reduce aliasing. We have already begun modifying Space Carving [10] with DRT by enhancing its pixel consistency test.

Future work will include application of DRT to other computational geometry algorithms, including comparisons in standard HT tasks against other anti-aliasing techniques. Parameter uncertainty knowledge should be incorporated into the DRT distribution. Evaluations should be performed on industry standard datasets, including [14]. Finally, detailed analysis of accumulator interference patterns in two-point and one-line scenes would be enlightening.

We find DRT to be an effective technique for mitigating aliasing in ABM, with potential applications in many other computational stereo algorithms. Mitigating aliasing is critical to achieving accurate super-resolution reconstruction.

Acknowledgment

This work was supported by the U.S. Army Aviation and Missile Command and the Defense Advanced Research Projects Agency under contract W31P4Q-09-C-0464. The views, opinions, and/or findings contained in this article/presentation are those of the author/presenter and should not be interpreted as representing the official views or policies, either express or implied, of the Defense Advanced Research Projects Agency or the Department of Defense. Approved for Public Release, Distribution Unlimited.

References

[1] R. I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge: Cambridge Univ. Press, 2000.

[2] T. Coffman and A. C. Bovik, “Efficient stereoscopic ranging via stochastic sampling of match quality,” IEEE Trans. Img. Proc., vol. 19, pp.451-460, 2010.

[3] T. Hamano, T. Yasuno, and K. Ishii, “Direct Estimation of Structure from Non-linear Motion by Voting Algorithm without Tracking,” Proc. Int. Conf. on Pattern Recognition, pp. 505-508, 1992.

[4] S. Kawato, “Hough transform to Extract 3D Information from images of Different View Points,” Proc. Int. Conf. on Comp. Analysis of Images and Patterns, pp. 528-532, 1993.

[5] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” Proc. IEEE CVPR, 2006.

[6] J. Illingworth and J. Kittler, “A Survey of the Hough Transform,” Computer Vision, Graphics, and Image Processing: Image Understanding, vol. 44, pp. 87-116, 1988.

[7] N. Kiryati and A. M. Bruckstein, “Antialiasing the Hough transform,” Computer Vision, Graphics, and Image Processing, vol. 53, pp. 213-222, 1991.

[8] W. C. Y. Lam, L. T. S. Lam, K. S. Y. Yuen, and D. N. K. Leun, “An analysis on quantizing the Hough space,” Pattern Recognition Letters, vol. 15, pp. 1127-1135, 1994.

[9] W. Niblack and D. Petkovic, “On improving the accuracy of the Hough transform,” Proc. IEEE CVPR, pp. 574-579, 1988.

[10] K. N. Kutulakos and S. M. Seitz, “A Theory of Shape by Space Carving,” Int. J. of Comp. Vis., vol. 38, pp. 199-218, 2000.

[11] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, Computer Graphics: Principles and Practice, 2nd ed., Boston: Addison-Wesley, 1996.

[12] R. L. Cook, T. Porter, and L. Carpenter, “Distributed ray tracing,” Computer Graphics, Vol. 18, pp. 165-174, 1984.

[13] J. Princen, H. K. Yuen, J. Illingworth, and J. Kittler, “A comparison of Hough transform methods,” Proc. Int. Conf on Img. Proc. and its Applications, pp. 73-77, 1989.

[14] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” Proc. IEEE CVPR, pp. 519-526, 2006.