

FAST GPU GENERATION OF DISTANCE FIELDS FROM A VOXEL GRID

by

NICOLAS FEDOR

URN: 6683787

A dissertation submitted in partial fulfilment of the
requirements for the award of

BACHELOR OF SCIENCE IN COMPUTER SCIENCE

May 2025

Department of Computer Science
University of Surrey
Guildford GU2 7XH

Supervised by: Joey Sik Chun Lam

I declare that this dissertation is my own work and that the work of others is acknowledged and indicated by explicit references.

Nicolas Fedor
May 2025

© Copyright Nicolas Fedor, May 2025

Abstract

Write a summary of the work presented in your dissertation. Introduce the topic and highlight your main contributions and results. The abstract should be comprehensible on its own, and should not contain any references. As far as possible, limit the use of jargon and abbreviations, to make the abstract readable by non-specialists in your area. Do not exceed 300 words.

Acknowledgements

Write any personal words of thanks here. Typically, this space is used to thank your supervisor for their guidance, as well as anyone else who has supported the completion of this dissertation, for example by discussing results and their interpretation or reviewing write ups. It is also usual to acknowledge any financial support received in relation to this work.

Contents

1	Introduction	11
1.1	Problem Statement	11
1.2	Aims and Objectives	12
1.2.1	Performance Metrics	12
1.3	Scope and Limitations	13
2	Literature Review	14
2.1	Voxel-Based Representations and Optimization	15
2.1.1	Sparse Voxel Octrees	15
2.1.2	Compression and Optimization Techniques	16
2.1.3	Challenges in Dynamic Scenes	16
2.2	Distance Fields as an Alternative Representation	17
2.2.1	Real-Time Generation and Updates	17
3	Methodology	19
3.1	World Representation	19
3.2	Distance Field Computation	20
3.3	Rendering and Ray Marching	20
3.4	Demonstration Application: Falling Sand Simulation	21
3.5	Performance Evaluation	22

3.5.1	Frame Performance Metrics	22
3.5.2	Distance Field Computation Metrics	22
3.6	Limitations and Considerations	23

List of Figures

2.1	Illustration of how a collection of triangles get rasterized onto a screen as demonstrated by Lafruit et. al 2016	14
2.2	Visualisation of how ray-casting is used to rasterize a 3D world. Source: Wikipedia	15
2.3	Illustration of the structure of a Sparse Voxel Octree, nodes with more details have additional subdivisions. (Truong-Hong & Laefer 2014)	16
2.4	Comparison of a octree versus the compressed format of a SVDAG, illustrated as a quad tree for brevity. (Dolonijs 2018)	17
2.5	Illustration of a discrete signed distance field grid. Negatives values indicate a cell inside the shape, positive values indicate a cell outside of the shape.	18
2.6	Example of a ray marching from a camera using sphere ray marching and a signed distance function. The ray will query the function for the distance to the nearest object, and advance by that amount.	18

List of Tables

Glossary

P Placeholder

Abbreviations

SVO	Sparse Voxel Octree
JFA	Jump Flooding Algorithm
CPU	Central Processing Unit
GPU	Graphics Processing Unit
DAG	Directed Acyclic Graphs
SVDAG	Sparse Voxel Directed Acyclic Graph
FPS	Frames Per Second

Chapter 1

Introduction

In recent years, real-time computer graphics applications have increasingly adopted distance fields as a fundamental representation for rendering and physics simulations (Jones & Satherley 2001). Distance fields, which encode the minimum distance from any point to the nearest surface, provide an elegant solution for various graphics operations including collision detection (Fuhrmann, Sobotka & Groß 2003), soft shadows (Tan, Chua, Koh & Bhojan 2022), and ambient occlusion (Wright 2015). While techniques exist for generating distance fields in real-time from a triangle mesh (Kramer 2015), techniques covering distance field generation from discrete voxel data are uncommon.

1.1 Problem Statement

Current approaches to distance field generation from voxel grids present various tradeoffs that limit their effectiveness in dynamic scenes. The Jump Flooding Algorithm (JFA) (Rong & Tan 2006, Rong & Tan 2007, Wang, Ino & Ke 2023), while efficient for parallel computation, introduces accuracy issues particularly at larger distances from surfaces and near feature edges. Scan, or prefix sum-based, approaches (Erleben & Dohlmann 2008) provide accurate results but suffer from inherent sequential dependencies that limit GPU parallelization. Wavefront propagation methods can efficiently update local regions but may struggle with concurrent updates in complex scenes (Teodoro, Pan, Kurc, Kong, Cooper & Saltz 2013).

Common optimization strategies, such as spatial partitioning into smaller chunks for localized updates (Naylor 1992), introduce their own challenges including boundary artifacts and increased

memory management overhead. While these techniques work well in isolation for specific use cases, there remains a fundamental gap in solutions that can handle arbitrary dynamic scene modifications while maintaining both accuracy and performance. This research investigates whether a novel hybrid approach—combining elements of existing techniques or developing new algorithmic patterns—could better address these challenges.

1.2 Aims and Objectives

This research aims to develop and evaluate novel GPU-based techniques for rapid distance field generation from voxel grid representations. The primary objectives are:

- To analyze and classify existing approaches to distance field generation, with particular focus on GPU-accelerated methods.
- To develop new algorithms that optimize the conversion process from voxel grids to distance fields.
- To implement and validate these algorithms on modern GPU architectures.
- To establish a comprehensive comparison framework for evaluating different distance field generation techniques.

1.2.1 Performance Metrics

The evaluation of the proposed methods will be conducted against sparse voxel octree implementations, which currently represent the state-of-the-art in many graphics applications. Key performance metrics include:

1. Computation time for initial distance field generation.
2. Memory consumption during generation and storage.
3. Update latency for localized geometric changes.
4. Scalability with increasing voxel grid resolution.
5. Accuracy of distance field values compared to analytical solutions.
6. GPU resource utilization, including memory bandwidth and compute occupancy.

1.3 Scope and Limitations

While this research addresses the core challenge of distance field generation, several related aspects fall outside its scope:

- The optimization of ray marching techniques for distance field rendering.
- The development of new compression methods for distance field storage.
- The optimization of the underlying renderer, which will include features like:
 - Memory management between the CPU and GPU.
 - Synchronization with the windowing framework.
 - Optimizing presentation of ray marching output image.

The focus remains specifically on the GPU-based generation process and its performance characteristics in dynamic scenarios. The research assumes access to modern GPU hardware and primarily targets real-time graphics applications where frequent distance field updates are required.

A sample “Falling Sand” graphics application will be implemented that will serve as a demo and benchmark for real-time performance of the distance field regeneration. The simulation itself will not be optimized and include a bare minimum of sand and water voxels that can fall and spread out in the world.

Chapter 2

Literature Review

Traditional real-time rendering has predominantly relied on triangle meshes as the fundamental primitive, with modern graphics hardware specifically optimized for rasterizing triangles efficiently (Akenine-Moller, Haines & Hoffman 2019).

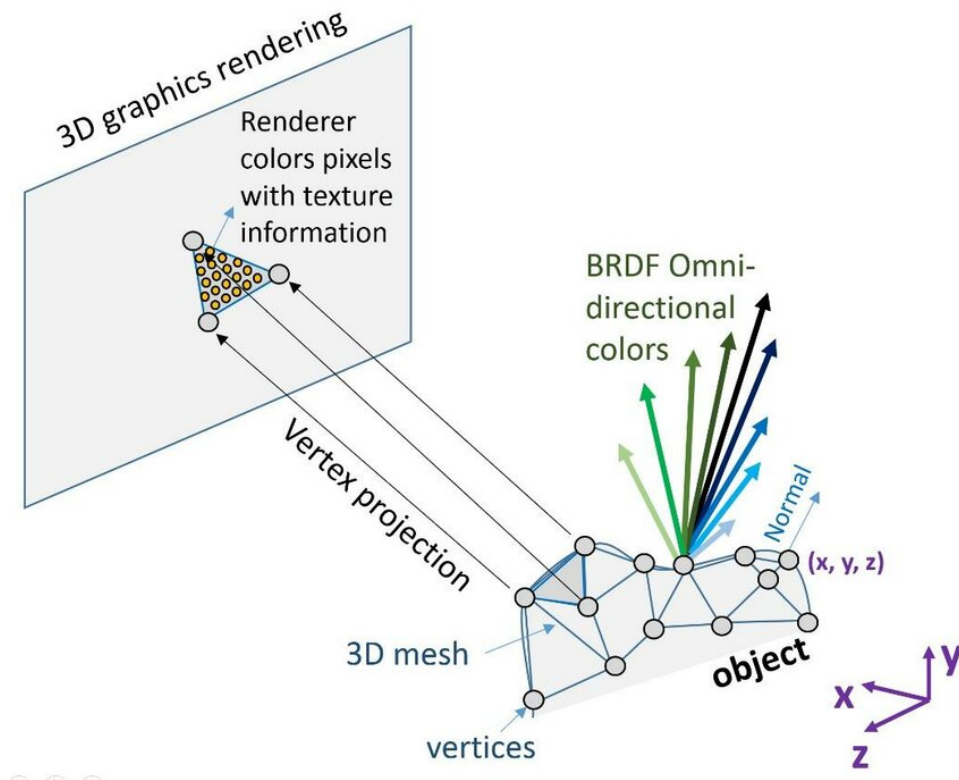


Figure 2.1: Illustration of how a collection of triangles get rasterized onto a screen as demonstrated by Lafruit et. al 2016

While this approach remains widespread, recent advances in hardware-accelerated ray tracing, particularly with NVIDIA’s RTX series (2018) and AMD’s RDNA2 architecture (2020), have enabled real-time ray tracing in commercial applications. Games like “Cyberpunk 2077” and “Metro Exodus Enhanced Edition” demonstrate that hybrid approaches combining rasterization and ray tracing can achieve photorealistic effects such as global illumination and accurate reflections at interactive framerates (Keller, Viitanen, Barré-Brisebois, Schied & McGuire 2019). However, both rasterization and ray tracing face similar challenges when representing highly detailed or volumetric content, leading to the exploration of alternative representations.

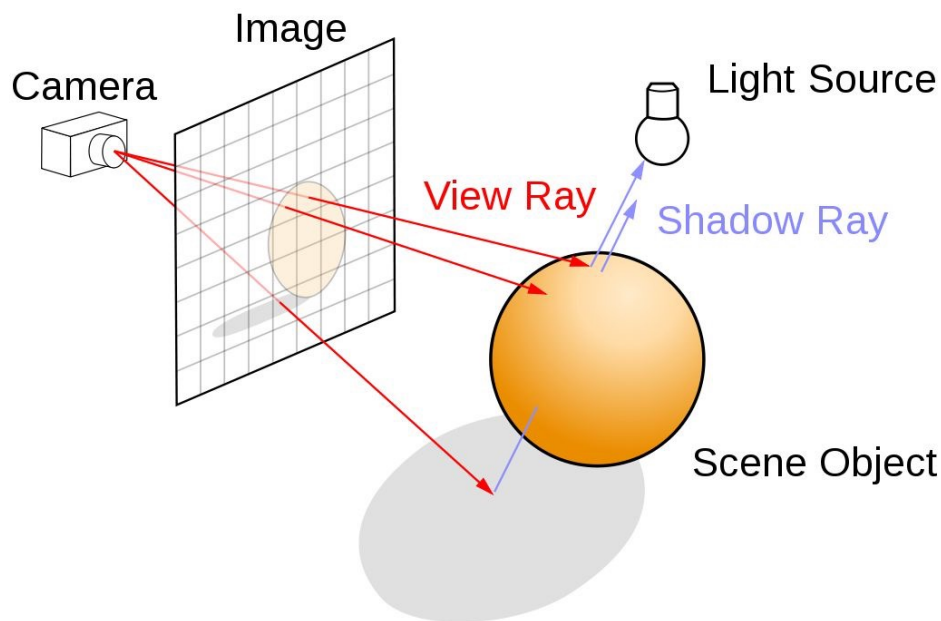


Figure 2.2: Visualisation of how ray-casting is used to rasterize a 3D world. Source: Wikipedia

2.1 Voxel-Based Representations and Optimization

2.1.1 Sparse Voxel Octrees

Sparse Voxel Octrees (SVOs) have emerged as a powerful solution for managing large-scale voxel worlds (Crassin, Neyret, Lefebvre & Eisemann 2009). Crassin et al. (2009) introduced GigaVoxels, a groundbreaking approach that demonstrated efficient rendering of highly detailed voxel scenes through hierarchical structure and streaming. Their work showed that SVOs could effectively compress empty space while maintaining quick traversal times for ray casting. Building

on this foundation, Laine and Karras (2010) developed an efficient sparse voxel octree (Laine & Karras 2010) implementation that improved upon previous approaches by introducing a novel node structure and traversal algorithm. Their method significantly reduced memory requirements while maintaining high rendering performance, making it particularly suitable for static scenes with complex geometry.

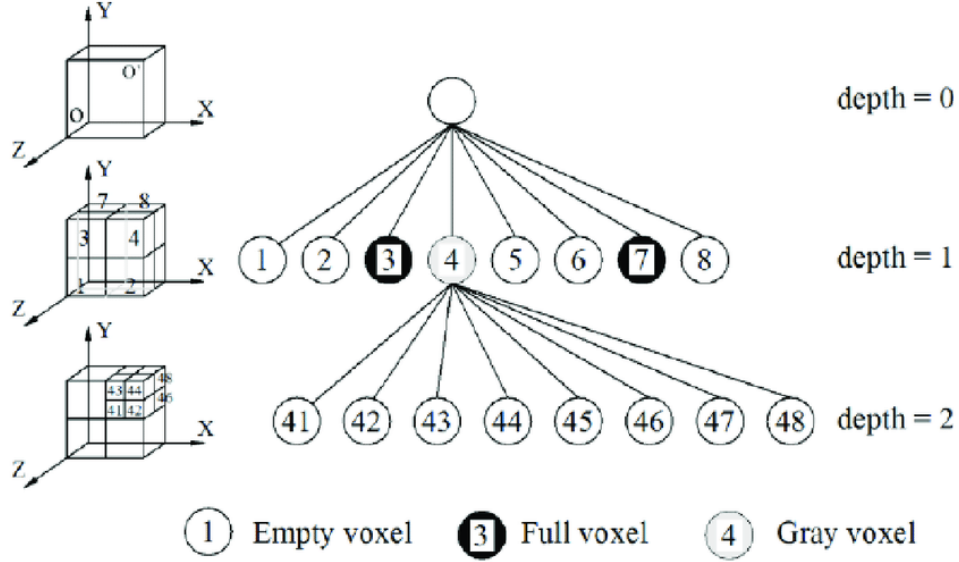


Figure 2.3: Illustration of the structure of a Sparse Voxel Octree, nodes with more details have additional subdivisions. (Truong-Hong & Laefer 2014)

2.1.2 Compression and Optimization Techniques

Several researchers have explored various compression techniques to further optimize voxel storage. Kämpe et al. (2013) introduced directed acyclic graphs (DAGs) for voxel scenes (Kämpe, Sintorn & Assarsson 2013), achieving compression ratios of up to 50:1 compared to standard SVOs while maintaining real-time rendering capabilities. This approach proved particularly effective for architectural and synthetic scenes with repeated structures.

2.1.3 Challenges in Dynamic Scenes

The primary challenge in dynamic voxel environments lies in maintaining data structures that can efficiently support modifications. Updating traditional SVOs in real-time presents significant

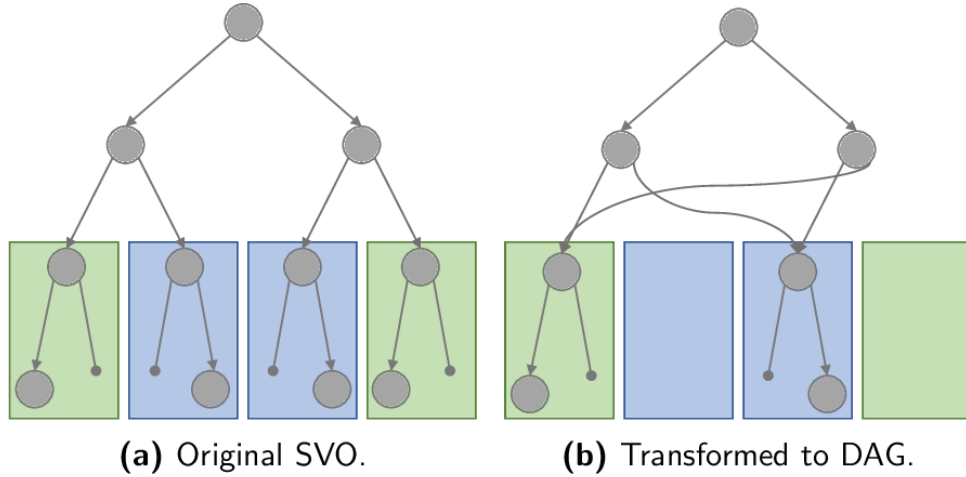


Figure 2.4: Comparison of an octree versus the compressed format of an SVDAG, illustrated as a quad tree for brevity. (Dolonius 2018)

computational overhead, as changes often require rebuilding portions of the tree structure. Pan (2021) explored a novel technique for merging SVOs and dynamically creating nodes where updates are needed (Pan 2021), while this showcases SVOs have the potential to support large dynamic scenes, the results show that real-time updates, such as in a video game application, are hard to achieve.

2.2 Distance Fields as an Alternative Representation

Distance fields have gained attention as an alternative to direct voxel storage, offering several advantages for both rendering and collision detection. Several recent works have demonstrated the advantages of using distance fields for real-time rendering of implicit surfaces (Hadj-Kyriacou & Arandjelović 2021) and function grids (Söderlund, Evans & Akenine-Möller 2022).

2.2.1 Real-Time Generation and Updates

The challenge of generating and updating distance fields in real-time remains an active area of research. Techniques such as Jump Flooding (Rong and Tan, 2006) provide fast approximate solutions but suffer from accuracy issues (Rong & Tan 2006, Rong & Tan 2007); improvements to Jump Flooding are being researched that allow for it to be used in dynamic contexts where recalculation of distance fields is needed (Stevenson & Navarro 2022).

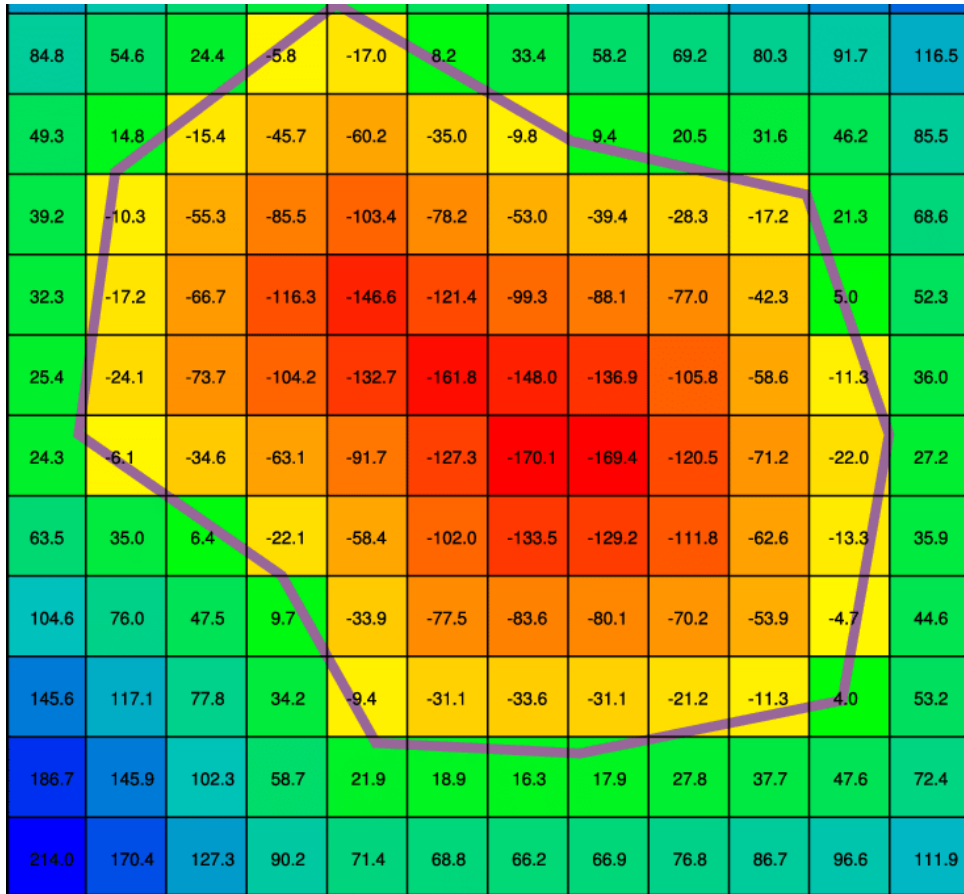


Figure 2.5: Illustration of a discrete signed distance field grid. Negatives values indicate a cell inside the shape, positive values indicate a cell outside of the shape.

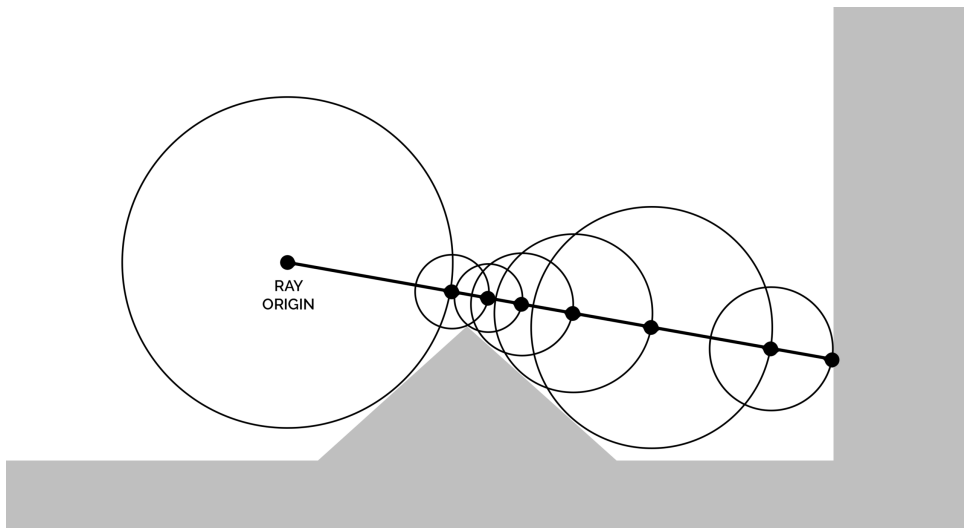


Figure 2.6: Example of a ray marching from a camera using sphere ray marching and a signed distance function. The ray will query the function for the distance to the nearest object, and advance by that amount.

Chapter 3

Methodology

The research methodology focuses on developing a dynamic distance field generation system using Vulkan, a low-level graphics API that provides precise control over GPU resources and computation. The system is designed to address the challenges of efficient distance field generation and rendering in dynamic voxel-based environments.

3.1 World Representation

The world is stored in a GPU buffer with host and device accessibility. This design choice prioritizes flexibility in world modification while minimizing performance overhead. Unlike traditional rendering approaches, the world buffer is not directly rendered, which mitigates potential performance penalties associated with host-visible memory. The choice of host-visible memory means that the host is able to update the world buffer as needed and the updates will be visible to the device as well reducing complexity in staging buffers.

The world buffer will be stored in an uncompressed and dense format; this means each voxel in the world will be present in the world buffer with all of its associated data. In this case a voxel will be a 32-bit unsigned integer, a value of 0 indicates an “air” voxel that should not be visible when rendered, while all other values indicate some form of solid voxel.

3.2 Distance Field Computation

The computation of a distance field, given a voxel grid, is the primary focus of this paper. To accomplish this a compute shader is implemented that will output a buffer containing the discrete distance field grid for a corresponding input voxel grid. This implementation is what will change throughout this paper as new algorithms and optimizations are added. The distance field will contain the manhattan distance to the nearest solid voxel, this is important for accurate ray marching of the distance field 3.3.

The computation of a distance field should not occur every frame as that would negatively impact the frame rate of an application. Instead the CPU will update the world state to “dirty” to indicate that an update to the world buffer has occurred that needs an update to the corresponding distance field so that it can be accurately rendered. This results in a computational workflow:

1. World buffer updated.
2. World is marked as “dirty”.
3. Distance field computation shader is dispatched.
4. Renderer waits for the distance field to be updated before rendering begins.

To facilitate voxels having colors, colour information is encoded into the distance field. The distance field buffer will be a 1-dimensional unsigned integer array. The first 8 bits define the distance to the nearest solid voxel using the manhattan distance, the remaining 8 bits define the colour of the voxel in a compressed RGB332 format; voxel colours are hard-coded into the distance field.

3.3 Rendering and Ray Marching

The rendering is handled by a ray marching compute shader; the distance field is the only input to this shader. The ray marcher utilizes a digital differential analyzer to traverse through a voxel grid quickly (Amanatides, Woo et al. 1987). This approach ensures a ray traversing the distance field grid will traverse each voxel along the ray, as other methods like sphere marching could result in artifacts due to a ray “missing” a voxel, when using the manhattan distance.

The ray marching compute shader will use a 3D perspective camera, that can move around the world. A ray can:

1. Miss the world entirely, this should result in a sky color being output at that pixel.
2. Hit the world, but not hit any solid voxels. A solid voxel is determined as a distance of 0. This will also result in a sky color being output at that pixel.
3. Hit the world, and hit a solid voxel. This will result in the color at that voxel being output to the pixel.

3.4 Demonstration Application: Falling Sand Simulation

A “Falling Sand” simulation (Castro & Chernick n.d.) will be implemented, this will be used in determining whether distance field updates are a viable approach for a dynamic voxel-based world. In our version of the falling sand simulation, an additional water particle will be implemented that can also move horizontally in addition to downwards like sand.

This simulation will be used to determine the performance of rendering the distance field per the defined frame performance metrics 3.5.1. In addition to this, the high amount of updates will be useful in collecting a large data set of distance field computation metrics 3.5.2.

The simulation will have a variable update frequency allowing control over the number of updates run per second. Additionally, the world will have a variable size to allow for testing the performance at different sizes. SVO implementations typically see upwards of 10 voxel levels resulting in a world size of at least 1024^3 , all the way to worlds of sizes 65536^3 . The simulation will be run at different world sizes but will be limited by memory consumption due to the uncompressed nature of the world representation and distance field.

The process for rendering will be:

1. Run an iteration of the falling sand simulation on the world buffer.
2. Recompute distance field if the world buffer has been modified.
3. Render the distance field to the screen.

3.5 Performance Evaluation

As part of the evaluation of the distance field generation in a dynamic environment, several performance metrics need to be gathered.

The performance test will be run multiple times with the same parameters to ensure accurate metrics are gathered; for each implementation of the distance field computation, the demonstration application will be run using differing world sizes, this may vary from implementation to implementation depending on their specific limitations. Results between implementations will also be evaluated.

3.5.1 Frame Performance Metrics

Frame performance metrics will help in determining whether the application can run in real-time in a “playable” speed. The minimum required frame rate for a game to be deemed playable is a heavily debated topic; however, player performance typically hits a plateau above 60 frame-per-second (FPS) (Claypool & Claypool 2007), with 30 FPS being a solid starting point. For this paper, a minimum target of 30 FPS will be set with the falling sand simulation; this should take into account the efficiency which the distance field computation is carried out and the effect it has on rendering.

3.5.2 Distance Field Computation Metrics

The primary metric used here will be the execution time. Assuming a 30 FPS target, the frame time is 33.33ms, this means for any given frame all of the applications operations must take less than 33.33ms to achieve the target FPS; the upper limit for the distance field execution time is 33.33ms which would assume all other factors like simulation update and rendering take no time in a frame.

Statistical analysis on the execution time will be carried out to provide insights into the computational consistency.

3.6 Limitations and Considerations

The performance of the distance field computation, and the rendering, is highly hardware dependant; factors such as:

1. GPU specifications
2. Memory configuration
3. Driver versions

all have an effect on performance. With this in mind the performance testing for this paper is carried out on a laptop:

CPU: AMD Ryzen 9 5900HS with Radeon Graphics @ 3.3GHz

GPU: NVIDIA GeForce RTX 3070 Laptop GPU

RAM: 16384MB @ 3200 MT/s

Bibliography

- Akenine-Moller, T., Haines, E. & Hoffman, N. (2019), *Real-time rendering*, AK Peters/crc Press.
- Amanatides, J., Woo, A. et al. (1987), A fast voxel traversal algorithm for ray tracing., *in* ‘Eurographics’, Vol. 87, Citeseer, pp. 3–10.
- Castro, J. & Chernick, S. (n.d.), ‘Real-time 3d sand simulation’.
- Claypool, K. T. & Claypool, M. (2007), ‘On frame rate and player performance in first person shooter games’, *Multimedia systems* **13**(1), 3–17.
- Crassin, C., Neyret, F., Lefebvre, S. & Eisemann, E. (2009), Gigavoxels: Ray-guided streaming for efficient and detailed voxel rendering, *in* ‘Proceedings of the 2009 symposium on Interactive 3D graphics and games’, pp. 15–22.
- Dolonius, D. (2018), *Sparse Voxel DAGs for Shadows and for Geometry with Colors*, Chalmers Tekniska Hogskola (Sweden).
- Erleben, K. & Dohlmann, H. (2008), ‘Signed distance fields using single-pass gpu scan conversion of tetrahedra’, *Gpu Gems* **3**, 741–763.
- Fuhrmann, A., Sobotka, G. & Groß, C. (2003), Distance fields for rapid collision detection in physically based modeling, *in* ‘Proceedings of GraphiCon’, Vol. 2003, Citeseer, pp. 58–65.
- Hadji-Kyriacou, A. & Arandjelović, O. (2021), ‘Raymarching distance fields with cuda’, *Electronics* **10**(22), 2730.
- Jones, M. W. & Satherley, R. (2001), Using distance fields for object representation and rendering, *in* ‘Proc. 19th Ann. Conf. of Eurographics (UK Chapter)’, London, pp. 37–44.
- Kämpe, V., Sintorn, E. & Assarsson, U. (2013), ‘High resolution sparse voxel dags’, *ACM Transactions on Graphics (TOG)* **32**(4), 1–13.

- Keller, A., Viitanen, T., Barré-Brisebois, C., Schied, C. & McGuire, M. (2019), Are we done with ray tracing?, *in* ‘SIGGRAPH Courses’, pp. 3–1.
- Kramer, L. (2015), Real-time sparse distance fields for games, *in* ‘Game Developers Conference’.
- Laine, S. & Karras, T. (2010), Efficient sparse voxel octrees, *in* ‘Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games’, pp. 55–63.
- Naylor, B. F. (1992), Interactive solid geometry via partitioning trees, *in* ‘Proc. Graphics Interface’, Vol. 92, pp. 11–18.
- Pan, Y. (2021), Dynamic update of sparse voxel octree based on morton code, Master’s thesis, Purdue University.
- Rong, G. & Tan, T.-S. (2006), Jump flooding in gpu with applications to voronoi diagram and distance transform, *in* ‘Proceedings of the 2006 symposium on Interactive 3D graphics and games’, pp. 109–116.
- Rong, G. & Tan, T.-S. (2007), Variants of jump flooding algorithm for computing discrete voronoi diagrams, *in* ‘4th international symposium on voronoi diagrams in science and engineering (ISVD 2007)’, IEEE, pp. 176–181.
- Söderlund, H. H., Evans, A. & Akenine-Möller, T. (2022), ‘Ray tracing of signed distance function grids’, *Journal of Computer Graphics Techniques Vol 11*(3).
- Stevenson, R. & Navarro, C. A. (2022), ‘Gpu voronoi diagrams for random moving seeds’, *arXiv preprint arXiv:2209.00117*.
- Tan, Y. W., Chua, N., Koh, C. & Bhojan, A. (2022), ‘Rtsdf: Real-time signed distance fields for soft shadow approximation in games’, *arXiv preprint arXiv:2210.06160*.
- Teodoro, G., Pan, T., Kurc, T. M., Kong, J., Cooper, L. A. & Saltz, J. H. (2013), ‘Efficient irregular wavefront propagation algorithms on hybrid cpu–gpu machines’, *Parallel computing* **39**(4-5), 189–211.
- Truong-Hong, L. & Laefer, D. F. (2014), ‘Octree-based, automatic building facade generation from lidar data’, *Computer-Aided Design* **53**, 46–61.

Wang, J., Ino, F. & Ke, J. (2023), Prf: A fast parallel relaxed flooding algorithm for voronoi diagram generation on gpu, *in* ‘2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS)’, IEEE, pp. 713–723.

Wright, D. (2015), Dynamic occlusion with signed distance fields, *in* ‘ACM SIGGRAPH’, Vol. 3.