A picture containing tool

Description automatically generatedMusschoot Adriaan

Graduation Work Title

Supervisor: Samyn Koen

Coach: Geeroms Kasper

Graduation Work 2023-2024

Digital Arts and Entertainment

Howest.be

A close up of a card

Description automatically generated

Contents

[Contents 1](#_Toc186709649)

[1 Abstract & Key words 2](#_Toc186709650)

[2 Preface 3](#_Toc186709651)

[3 List of Figures 4](#_Toc186709652)

[4 Introduction 5](#_Toc186709653)

[5 Literature Study / Theoretical Framework 6](#_Toc186709654)

[5.1 Ray Tracing 6](#_Toc186709655)

[5.1.1 Fundamentals 6](#_Toc186709656)

[5.1.2 Acceleration structures 7](#_Toc186709657)

[5.2 sphere tracing (ray marching) 8](#_Toc186709658)

[5.2.1 Fundamentals 8](#_Toc186709659)

[5.2.2 Usage 11](#_Toc186709660)

[5.2.3 FRAMEWORK’s considered 11](#_Toc186709661)

[6 Research 13](#_Toc186709662)

[7 case study 14](#_Toc186709663)

[8 Discussion 15](#_Toc186709664)

[9 Conclusion 16](#_Toc186709665)

[10 Future work 17](#_Toc186709666)

[11 Critical Reflection 18](#_Toc186709667)

[12 References 19](#_Toc186709668)

[13 Acknowledgements 21](#_Toc186709669)

[14 Appendices 22](#_Toc186709670)

# Abstract & Key words

**An abstract explains the outline of the paper concisely (the methods, results, etc.). Maximum length of 250 words, preferably both in English and Dutch.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis. Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# Preface

***A preface is a statement of the author's reasons for undertaking the work and may include personal comments that are not directly relevant to other sections of the thesis or dissertation.* No word count limit.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis. Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# List of Figures

[Figure 1: Example of a triangle mesh 7](#_Toc186546142)

[Figure 2: The process of ray casting and shading pixels 7](#_Toc186546143)

[Figure 3: AABB for a single triangle 8](#_Toc186546144)

[Figure 4: Complex mesh with BVH applied 9](#_Toc186546145)

[Figure 5: Tree structure for BVH 9](#_Toc186546146)

[Figure 6: Example of SDF output 10](#_Toc186546147)

[Figure 7: Ray marching basic and a missed surface 11](#_Toc186546148)

[Figure 8: Example of both scenarios when raymarching (Hart, 1996) 11](#_Toc186546149)

# Introduction

**In the introduction, you write the background of your topic and discuss the observation that spurred you on to do this research project. Explain the purpose of the paper and present your research question(s) and the hypothesis at the end of this section. This section is typically a couple of pages long.**

# Literature Study / Theoretical Framework

## Ray Tracing

### Fundamentals

#### Triangle meshes

In computer graphics, we need to visualize 3D objects. One of the approaches is a triangle mesh. These consist of a bunch of triangles (Figure 1), each triangle consist of 3 vertices and 3 edges connecting these vertices. (“Triangle Mesh,” 2024)



Figure 1: Example of a triangle mesh

#### Ray tracing algorithm

The ray tracing algorithm uses these meshes to fill the pixels on the screen with meaningful colour. For every pixel on the screen the algorithm casts a ray from the origin through each pixel in an image plane into the scene (the origin is often referred to as the camera or eye point). If the ray intersects with one or multiple meshes, it calculates the shading at that point using the material properties of the closest mesh. The result is used to fill in the pixel the ray was cast for. (Figure 2) This process can be repeated with different origins and directions to incorporate light, reflections and shadows. (*Ray Tracing | NVIDIA Developer*, n.d.)



Figure 2: The process of ray casting and shading pixels

### Acceleration structures

#### WHY?

If you have very complex geometries and many objects in your scene, then the algorithm executes a lot of unnecessary calculations.

Scenario 1: You have a lot of objects in your scene. The algorithm needs to calculate the intersection for every object just to conclude the ray didn’t hit anything.

Scenario 2: You have a super detailed object with 100 000 primitives. If the algorithm must go over every triangle in that object to find the one it is intersecting with. Then again you have a lot of unnecessary calculations.

To avoid this there are the following acceleration structures.

#### aabb

AABB or Axis Aligned Bounding Box is a common acceleration structure used when ray tracing primitives.

It is a simple box which represents the closed region that contains a triangle mesh. The edges of the box are always aligned with the world axes, meaning that they are parallel to the x, y, and z axes in three-dimensional space. To define the AABB for a triangle mesh you must find the minimum and maximum points along each axis that defines the mesh. Then you can construct the AABB, using the minimum and maximum as two opposite corners of the box. (“Bounding Volume,” 2024; *CSSE451 Advanced Computer Graphics*, n.d.; *What Is AABB in Computing?*, 2024)

The ray tracing algorithm can then calculate ray intersections using the AABB of a mesh, if and only if that is the case, does it calculate further intersections with any of the triangles from the mesh. This method of early out speeds up ray intersection calculations in the case the ray misses an object.

A computer screen shot of a colorful pyramid

Description automatically generated

Figure 3: AABB for a single triangle

#### bvh

BVH or Bounding Volume Hierarchy is another common approach to speed up the ray intersection process.

When you have complex mesh such as in Figure 4 calculating intersections with all triangles is very costly. To avoid this, you can subdivide the initial AABB into smaller boxes and intersect with those first. This process can be repeated until one smaller box contains the desired number of triangles. (Sebastian Lague, 2024)

A white dragon with a green laser pointing at it

Description automatically generated

Figure 4: Complex mesh with BVH applied

You can think of this structure as many smaller AABBs parented to each other. First you check if you intersect with the root node (which is also the AABB). If that is the case, you evaluate its child nodes. If the child node is a leaf node the algorithm can start calculating intersections with the primitives. This approach allows you to determine the level of detail for the boxes and the minimum number of triangles per box.



Figure 5: Tree structure for BVH

## sphere tracing (ray marching)

### Fundamentals

#### Signed distance fields (sdf)

Another way to visualize 3D objects is by using signed distance fields (SDF).

There are 2 ways of storing SDF data. One being baked volumetric data and the other being a mathematical function. The latter is explained here after and is the one referred to when writing about SDFs.

SDFs take in a point and calculate the smallest distance from the surface of the SDF to that point considering the SDFs origin is at (0,0,0) in cartesian coordinate system. If a point is outside the surface of the SDF, the function will return a positive distance value. If it is inside the surface, it will return a negative distance value. When a point is perfectly aligned with the surface the returned value will be 0. Figure 6 is an example of what values such a function returns.



Figure 6: Example of SDF output

Let’s look at the simplest shape to calculate: a sphere.

1. float GetDistance(vector3 point)

2. {

3. return length(point) – radius;

4. }

It takes in a point (the current position along the ray, this is explained later. It calculates the length to the point and subtracts the radius. If this result is negative, the point is inside the sphere, if it is positive the point is outside the sphere. This is assuming the sphere is at position (0, 0, 0).

If you want to have translations, rotations or scaling you must manipulate the point before calculating the final value. (Quilez, n.d.)

The biggest benefit of SDFs is that they are usually fully implemented in code in the fragment shader. Meaning this requires way less memory than triangle meshes. As there is no need to store all the vertices that make up those meshes. Only one line of code in the case of a sphere.

#### Ray marching algorithm

##### The marching loop

We need a slightly different algorithm to ray tracing, but with some similar principles. We still use a camera to cast rays through an image plane. However, we cannot calculate any intersection points due to the nature of the SDFs. The most default approach is to step forward along a ray at a fixed increment. If at any point the returned value of the SDF is negative, we can consider that we have hit an object. In Figure 7 one can see that if the increment size is too big the algorithm can easily miss a surface. However, if the step size is too small then it can cause reduced performance. (Donnelly, 2005)

A diagram of a graph

Description automatically generated

Figure 7: Ray marching basic and a missed surface

There is however a variation called sphere tracing that is more efficient and precise, presented in (Hart, 1996). The process goes as follows:

1. We can calculate the distance to the SDFs in the scene by giving passing in the origin of the ray.
2. Then we can displace the origin of the ray along its direction, using the distance value returned by the SDF.
3. Repeat steps 1 and 2 until one of 2 possible outcomes:
   1. The distance we can travel is smaller than an arbitrary value (0.001), meaning we hit an object
   2. The value of the total distance travelled by the ray has exceeded the scene boundaries.

The pseudo code

1. distance travelled = 0

2. origin = camera point

3. while (true)

4. point = origin + distance travelled \* direction

5. distance able to travel = distance to scene based on point //loops over all SDFs

6. distance travelled = total distance + distance able to travel

7. if distance able to travel < 0.001

8. break //hit object

9. if total distance > scene boundaries

10. break //no objects hit

In Figure 8 both rays are being marched towards the triangle from left to right. The top ray represents outcome ‘a’ where the SDF result approaches 0. As a result, the pixel for which this ray was cast should be colored with the triangle’s color properties. The bottom ray represents outcome ‘b’ where the distance the ray travelled exceeded the scene bounds (in this case the image bounds). (Devred, 2022)

A black and white drawing of a triangle

Description automatically generated

Figure 8: Example of both scenarios when raymarching (Hart, 1996)

While this approach is more memory friendly than using triangle meshes, it does mean the entire scene needs to be reconstructed every frame. And as SDFs become more complex they require more computation time .

The main bottleneck of performance is acquiring the distance to the scene for every step of the ray. In its most primitive form of execution the pseudo code for it looks something like this.

1. minimum distance = infinity

2.

3. for object in scene

4. distance = distance to object based on point

5. if distance < minimum distance

6. minimum distance = distance

7.

8. return minimum distance

This is a heavy computation happening every frame for every ray at each step it takes. If you have a screen of 100 by 100 pixels with an average ray step of 10 to traverse the scene. You would evaluate this function 100 000 times each frame. This linearly increases with the number of objects in the scene. To speed this up can we apply the same acceleration structures commonly used for ray tracing?

### Usage

The technique has been around for a while (Wyvill & Wyvill, 1989) in different shapes. It was mostly used as baked volumetric data. Recently, GPUs were evolving their computational/ALUs abilities faster than their memory band width. This meant purely mathematical SDFs became competitive against 3D-texture based SDFs. (Quilez, n.d.)

When compared to rasterization, which consists of the vertex shader, rasterization and the fragment shader, the ray marching algorithm can be fully implemented in the fragment shader, allowing for integration with existing pipelines. Its use cases are cheap ambient occlusion, screen space reflections, rendering volumetrics, rendering clouds. (Papaioannou et al., 2010; “Ray Marching,” 2024; Schneider, 2024; Tomczak, 2012)

In Unreal Engine 5 ray marching is used for ambient occlusion and distance field shadows. (Unreal Engine 5, n.d.; Ureal Engine 5, n.d.)

There is a game called “Claybook” (<https://claybookgame.com/>) which handles the entire rendering pipeline using ray marching and fully exploiting the benefits and quirks of raymarching.

The website “shadertoy” (<https://www.shadertoy.com/>) developed by Inigo Quilez and Pol Jeremios is a tool used to teach and create demo scenes. It purely uses the ray marching algorithm. Writing code happens in GLSL.

There is also the “demoscene” community, they specialize in creating non-interactive audio-visual presentations executed in real time on computers. (“Demoscene,” 2019) Which has embraced “shadertoy” and SDFs in general for their scenes.

### APPROACHES considered

Creating a ray marcher from scratch is a hefty task and will take a considerable amount of time. The main goal is to compare the performance with and without optimization techniques. The following approaches were considered.

#### Custom framework / own implementation

The framework can completely be run on the CPU, meaning that debugging would be easy.

Running completely on the CPU is of course also a drawback. There would be a considerable drop in performance, as we can’t use the computational power of the GPU.

Writing regular C++ code is a huge benefit. There is more functionality than when writing shader code. In C++ you can also use recursion, classes, pointers, containers, etc. In C++ there is no built-in support for vector math and matrix operations.

#### Graphics api

Graphics APIs allow the user to communicate with the GPU. This will get us all the frames we could ever wish for. However, there are some drawbacks. Debugging support for graphics applications is very limited, not built-in and often not as straightforward and requiring external tools such as “RenderDoc”.

Feature support for languages such as GLSL or HLSL is also limited as it has a different execution environment. They do have built-in support for vector math and matrix operations.

As mentioned above, the ray marching algorithm only must be implemented in the fragment shader of a rasterization pipeline. This means all that is needed is a singular compute shader. So once that has been set up using the API, all that is left to write is in the shader. This is of course a nightmare for organizational purposes, if everything is in one file.

Vulkan and OpenGL were considered. OpenGL for its ease of use and its wide usage across the C++ community. Vulkan is where I have the most experience, and I already have an existing framework there.

#### shadertoy

The “shadertoy” website would require the least amount of work to execute this research in. It runs entirely on the ray marching algorithm already. It uses GLSL, which is simple to understand, and it has its own little compiler to tell you what is wrong.

However, there aren’t any methods of capturing frames and looking at the output, like with “RenderDoc”.

There is no possibility for me to track anything, the user only has access to the shader.

#### choice and motivation

Custom implementation is the approach of choice. It is the approach that will allow the easiest gathering of data while keeping the amount of work relatively small.

Before I explored this topic, I had created my own implementation of the ray-tracing algorithm. As discussed earlier, the ray marching algorithm has some of the same traits as the ray tracing algorithm. So, this could potentially be a good building block to set up the experiment, I am also familiar with this framework. Unlike setting up something entirely from scratch using the graphics API.

As mentioned before there is more added functionality when writing C++ instead of shader languages. Debugging anything will also be easier on the CPU, definitely when it comes to acceleration structures as they don’t provide any visual output.

The first drawback of this approach is that it uses a custom math library which is slower and incomplete. It still needs to be swapped for OpenGL Mathematics (GLM).

Secondly, nothing could ever run as fast on the CPU as on the GPU. The application will most definitely have to be multi-threaded to execute the research at more than 1 FPS.

# Research

**In the research section, you detail the elements of your experiment(s), the tests, objects you will test upon and subjects you will test with, the data gathering, data cleaning or feature extraction, measurements, … and you present the results obtained in an objective manner for each of the tests you conducted.**

## aabbs for sdfs

### aabbs for sdfs are early outs

In the theoretical framework creating the AABB for triangle meshes was discussed. However, when working with those, every point is known in world space, allowing for easy construction of the axis aligned bounding box.

The only data we have from SDFs is that we can get a distance from the surface to a point. Given this property we can approach the maximum boundaries that contain this SDF. Two approaches were implemented and compared for constructing the shape of the early out SDF.

What does the axis aligned bounding box look like in ray marching? Unlike ray tracing we cannot intersect with a box first using the ray marching algorithm. We must march towards the early out shape along the ray. This means we need an additional SDF as early out. Two SDF shapes were considered a box, and a sphere.

To use this additional SDF, first the distance to the additional SDF is calculated. If the distance to the additional SDF is smaller than a certain threshold (0.001), the distance to the object enclosed by the additional SDF must be computed and returned. If the distance to it is bigger than the threshold that distance is returned and acts as a form of early out, and the need to use the enclosed object is redundant.

1. float sdf::Object::GetDistance(glm::vec3 const& point)

2. {

3. float const earlyOutDistance{ EarlyOutTest(point) };

4. if (earlyOutDistance >= 0.001f)

5. {

6. return earlyOutDistance;

7. }

8. return GetDistanceUnoptimized(point);

9. }

Only if the result of the early out test is greater than the threshold value, should it return the early out value. This speeds up the ray marching process greatly. However, it does mean that once the ray is marching inside the early out-SDF extra and unnecessary calculations will happen. This is to be expected and no different than what happens in ray tracing. The benefits of an early out however hugely outweigh those extra calculations when actually approaching and hitting an object.

### Box early out

#### Constructing the boundaries

The following approach was used to approximate the boundaries of a box.

For each direction on every axis in the cartesian coordinate space (x, -x, y, -y, z, -z) a “wall” of points is constructed in a rectangular grid perpendicular to its direction. The initial distance at which the walls are constructed should be the maximum scene boundary. Then each wall marches forward opposite its initial direction (so towards the origin) using the distance value, retrieved by subtracting the minimum distance from the wall to the SDF from the length of the vector representing the point for which the minimum distance was valid. Each wall repeats this process until one of the points on the wall has a distance value smaller than a certain threshold (0.001). This can be considered as reaching the surface.

Once all 6 values were retrieved, we compare the absolute values of the x, y and z components of each point to determine the box extent of the SDF box. Which will be used to enclose the SDF used to compute the box.

#### formula

Below you can find the formula used for the distance of a box.

1. float sdf::Box::EarlyOutTest(glm::vec3 const& point)

2. {

3. vec3 const q{ abs(point) – boxExtent };

4. return length(max(q, 0.0f)) + min(max(q.x, max(q.y, q.z)), 0.0f);

5. }

Calculating the distance to a box is quite cheap but not the utmost cheapest in ray marching.

### Sphere early out

#### constructing the boundary

At the core of calculating the boundary you sample points on a sphere and request the distance for each point to the SDF.

For every intersection of a spheres 360 longitude lines and its 360 latitude lines, a point was created and retrieved its distance value to the SDF. After having executed this for the initial sphere radius, which should encompass the maximum scene boundary, the smallest possible distance from any point to the SDF is used to set up the next sphere. Through this method we can approach the furthest possible distance from the origin (0,0,0) in the cartesian coordinate system to the surface area.

Once one of the points has a distance value smaller than a certain threshold (0.001), we consider the point to have reached the surface. We can then calculate the length of the vector representing this point and use it to construct a sphere which represents our “axis aligned bounding box”.

//insert the code for approximating the sphere

#### Formula

Below you can find the formula used for the distance of a sphere.

1. float sdf::Sphere::EarlyOutTest(glm::vec3 const& point)

2. {

3. return length(point) - radius;

4. }

# case study

**Alternatively, as opposed to research, you might have opted for a case study. Whichever you choose, you detail the elements of your experiment(s), the tests, objects you will test upon and subjects you will test with, the data gathering, data cleaning or feature extraction, measurements, … and you present the results obtained in an objective manner for each of the tests you conducted.**

# Discussion

**In this section, you offer an interpretation of the results you obtained and try to relate them to the theoretical framework you presented. This is typically not a very long section, but obviously one of the most important ones.**

# Conclusion

**In this section, you ascertain the demonstrable outcomes of your study and outline the merits of the project for the academic field and the discourse community. This is typically not a very long section, but obviously also one of the more important ones.**

# Future work

**This section is sometimes standalone, sometimes incorporated in the conclusion. It looks at the shortcomings of the study, alternative strategies, and what could be the next course of action in the research field. This is typically not a very long section.**

# Critical Reflection

**This section is typically associated with a bachelor paper, not other forms of serious writing. It allows the student to reflect on the learning outcomes, both academically and in terms of personal growth.**

# References

Bounding volume. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Bounding\_volume&oldid=1226824951

*CSSE451 Advanced Computer Graphics*. (n.d.). Retrieved December 29, 2024, from https://www.rose-hulman.edu/class/cs/csse451/AABB/

Demoscene. (2019). In *Wikipedia*. https://nl.wikipedia.org/w/index.php?title=Demoscene&oldid=53157976

Devred, M. (2022). *Rendering 3D cross-sections of 4D Fractals*. https://allpurposem.at/paper/2022/fractals.pdf

Donnelly, W. (2005, April). *Per-Pixel Displacement Mapping with Distance Functions*. NVIDIA Developer. https://developer.nvidia.com/gpugems/gpugems2/part-i-geometric-complexity/chapter-8-pixel-displacement-mapping-distance-functions

Hart, J. C. (1996). *Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces*. http://link.springer.com/10.1007/s003710050084

Papaioannou, G., Menexi, M., & Papadopoulos, C. (2010). *Real-Time Volume-Based Ambient Occlusion*. https://www.researchgate.net/publication/45113929\_Real-Time\_Volume-Based\_Ambient\_Occlusion

Quilez, I. (n.d.). *Inigo Quilez*. Retrieved December 29, 2024, from https://iquilezles.org

Ray marching. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Ray\_marching&oldid=1265162208

*Ray Tracing | NVIDIA Developer*. (n.d.). Retrieved December 29, 2024, from https://developer.nvidia.com/discover/ray-tracing

Schneider, A. (2024, July 25). *Synthesizing Realistic Clouds for Video Games*. https://www.guerrilla-games.com/read/synthesizing-realistic-clouds-for-video-games

Sebastian Lague (Director). (2024, June 12). *Coding Adventure: Optimizing a Ray Tracer (by building a BVH)* [Video recording]. https://www.youtube.com/watch?v=C1H4zIiCOaI

Tomczak, L. J. (2012). *GPU Ray Marching of Distance Fields* [Technical University of Denmark]. https://www2.imm.dtu.dk/pubdb/edoc/imm6392.pdf

Triangle mesh. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Triangle\_mesh&oldid=1234853041

Unreal Engine 5. (n.d.). *Mesh Distance Fields in Unreal Engine | Unreal Engine 5.5 Documentation*. Epic Games Developer. Retrieved October 19, 2024, from https://dev.epicgames.com/documentation/en-us/unreal-engine/mesh-distance-fields-in-unreal-engine

Ureal Engine 5. (n.d.). *Distance Field Ambient Occlusion in Unreal Engine | Unreal Engine 5.5 Documentation*. Epic Games Developer. Retrieved October 19, 2024, from https://dev.epicgames.com/documentation/en-us/unreal-engine/distance-field-ambient-occlusion-in-unreal-engine

*What is AABB in Computing? (Axis Aligned Bounding Box)*. (2024, April 16). https://60sec.site/buyer-guides/what-is-aabb-in-computing-axis-aligned-bounding-box

Wyvill, B., & Wyvill, G. (1989). *The Visual Computer*. SpringerLink. https://link.springer.com/journal/371

# Acknowledgements

**In this section, you can thank people who contributed to your work in a meaningful way.**

# Appendices

**In many cases, there are items that were developed for a research paper that can’t go into the actual paper in full. Things suc as code, art pieces, output of statistical analysis, questionnaires, … In this section, you can present these elements; use the first page to list and number the items, then paste them sequentially. If some items are too large, you can store them online, and link to them. Common practice is to keep those links active at least one year after the publication of the thesis.**