Utilising ray marching and signed distance functions to render a scene of primitives

By Anthony Sturdy (18015368)

BSc (Hons) Computer Games Programming

Supervisors: Dr. David White, Craig Weightman

May 2022

Contents

[Glossary 3](#_Toc94970614)

[Abstract 3](#_Toc94970615)

[Introduction 4](#_Toc94970616)

[Aim 4](#_Toc94970617)

[Objectives 4](#_Toc94970618)

[Literature Review 5](#_Toc94970619)

[3D Wireframe Rendering 5](#_Toc94970620)

[Raycast Rendering 6](#_Toc94970621)

[Rasterised Rendering 7](#_Toc94970622)

[Shaders 7](#_Toc94970623)

[Ray Traced Rendering 8](#_Toc94970624)

[Ray Marching 9](#_Toc94970625)

[Algorithm 9](#_Toc94970626)

[Lighting and Shading 10](#_Toc94970627)

[Rendering Volumes 10](#_Toc94970628)

[Point-cloud Rendering 10](#_Toc94970629)

[Constructive Solid Geometry 11](#_Toc94970630)

[Fractals 11](#_Toc94970631)

[Research Methodology 13](#_Toc94970632)

[Project Management 13](#_Toc94970633)

[Waterfall 13](#_Toc94970634)

[Kanban 13](#_Toc94970635)

[Scrum 13](#_Toc94970636)

[Conclusion 14](#_Toc94970637)

[Quantitative vs. Qualitative 14](#_Toc94970638)

[Platform 14](#_Toc94970639)

[Unity 14](#_Toc94970640)

[DirectX11 14](#_Toc94970641)

[Vulkan 15](#_Toc94970642)

[Conclusion 15](#_Toc94970643)

[Project Plan 15](#_Toc94970644)

[Gantt Chart 17](#_Toc94970645)

[Analysis 18](#_Toc94970646)

[Design 18](#_Toc94970647)

[Required Software 18](#_Toc94970648)

[Required Hardware 19](#_Toc94970649)

[Implementation 20](#_Toc94970650)

[Development 20](#_Toc94970651)

[Application Framework 20](#_Toc94970652)

[Basic Ray Marcher 20](#_Toc94970653)

[Testing 22](#_Toc94970654)

[Results 23](#_Toc94970655)

[Critical Evaluation 24](#_Toc94970656)

[Bibliography 25](#_Toc94970657)

[Appendices REMOVE THIS / CHANGE TO PICTURE BIBLIOG. Appendices should be Figures instead, most likely need reordering (i.e. Appendix 5 is not the 5th image displayed, so should be not just be figure 5) 30](#_Toc94970658)

[Appendix 1: Rasterisation 30](#_Toc94970659)

[Appendix 2: Recursive Sphere Reflections 30](#_Toc94970660)

[Appendix 3: Point-Cloud Rendering 30](#_Toc94970661)

[Appendix 4: Constructive Solid Geometry 30](#_Toc94970662)

[Appendix 5: Elite 1984 30](#_Toc94970663)

[Appendix 6: Ray casting diagram and image 30](#_Toc94970664)

[Appendix 7: Comanche: Maximum Overkill 30](#_Toc94970665)

# Glossary

* **3D** – Shorthand for Three Dimensional
* **2D** – Shorthand for Two Dimensional
* **SDF** – Signed Distance Function
* **CSG** – Constructive Solid Geometry
* **FOV** – Field of View
* **API** – Application Programming Interface

# Abstract

This research explores using ray marching as a method of rendering. Ray marching steps along a ray toward a specified direction, until an intersection, is found. The distance of each step is determined by the shortest distance to any object in the scene. The signed distance function which each object uses defines the shape of the object.

For each pixel on the screen, a direction can be determined by its position on the screen and the cameras FOV. Each object in the scene is sampled for its distance from the starting point and the shortest distance is used to determine how far the ray can be stepped without intersecting with any object. This step is repeated until the distance becomes less than a specified threshold, which is then considered to be an intersection.

This method of using a signed distance function to determine the shape of objects in the scene allows for many effects which are difficult to achieve using traditional rasterised rendering. For example, object smoothing/morphing, real-time CSG boolean operations, rendering fractals, and more. Ray marching is commonly used for rendering volumetric objects, such as clouds.

# Introduction

Recent advances in graphics processing hardware have brought ray tracing into the forefront of real-time rendering (Akenine-Möller et al., 2018). Ray tracing can be used to accurately simulate how light travels which can result in a higher fidelity render than traditional methods of rendering.

Whereas ray tracing tests for an intersection between a ray segment and objects in the scene to find the nearest intersection (Whitted & Foley, 1980), ray marching differs by stepping along a ray incrementally until an intersection is found.

A common ray marching technique is known as sphere tracing. Sphere tracing utilises signed distance functions (SDFs) to calculate the distance of each step along the ray. At each ray step, all objects are sampled for their distance to the current point, the ray can then be stepped forward by the shortest distance which ensures the ray will not travel inside, nor skip over, any object in the scene (Hart, 1996). Due to the type of SDF defining the shape of that object, it is possible to render fractals or modify the output distance to manipulate the shape or appearance of the object.

## Aim

This research aims to explore the areas needed to produce a game engine that utilises ray marching as a method of rendering a 3D scene and to take advantage of the characteristics of ray marching to achieve various visual effects.

## Objectives

To achieve this aim, the research will enable the development of:

* A performant, real-time ray marching renderer.
* An engine framework surrounding the renderer, providing a platform to produce games/real-time applications.
* GUI controls to interact with the engine and scene at runtime.
* A 3D modelling tool that allows for primitives to be combined and visual effects to be applied to output a detailed model, which can then be imported to the engine.

# Literature Review

## 3D Wireframe Rendering

Original real-time 3D video games, such as Elite (Frontier Developments, 1984), for example, used a wireframe rendering technique to render polygons. This resulted in the outline of polygons (triangles) to be rendered without being filled, contrary to modern rasterised rendering which does fill the rendered polygons (see *Rasterised Rendering* below).



Elite 1984 (Appendix 5)

Elite’s non-horizontal lines were rendered using Bresenham’s line algorithm (Bresenham, 1965). Bresenham’s algorithm works by calculating the delta (difference) between positions *x1* and *x2* (*deltaX*), and the delta between positions *y1* and *y2* (*deltaY*), then for each position between *x1* and *x2* (*currentX*), the y position of the pixel (*currentY*) is incremented by *deltaY ÷ deltaX*. A pixel is plotted at the position *(currentX, currentY)* on each loop iteration. This technique would typically use real/floating point numbers, but to avoid this, the algorithm used in Elite would multiply all real numbers by 256 so that 256 would be equivalent to 1.0 (Moxon, 2020).

After polygons were rendered to the screen, Elite would ‘undraw’ the polygons to clear the screen for the next frame. This is due to a full screen clear being too slow for the hardware of 1984, and double buffering was not feasible due to memory constraints. The implemented solution was for the line rendering algorithm to invert each pixel drawn, which allowed for the same routine to be used for drawing and un-drawing, and for the algorithm to be re-run after the frame had been rendered to minimise the number of pixels being cleared (Braben, 2011).

Convex shapes are shapes which will only return a maximum of two intersections when a straight line is passed through any part of the shape (Underground Mathematics, 2022), a ball is a convex shape, for example. However, a bowl is not. The hull of each ship in Elite are convex, as convex shapes cannot occlude themselves partially. This means that lines would not need to be half-rendered if part of the hull is partly occluding another (Braben, 2011), thus increasing the render speed. A line draw could be completely skipped if drawing the back-face of a polygon, which allows for the ship to appear to occlude parts of itself (as the polygons facing away from the camera would not be seen) and time when rendering.

Elite also optimised the matrix multiplications required to project vertices into world and screen space by avoiding using multiplication operations, as CPUs in 1984 lacked the hardware for multiplication and division (Retro64, 2017), which meant multiplication and division operations were computationally slow. The method used for multiplication, without using a multiplication operation, was to use logarithms - a method of approximately multiplying and dividing two numbers with only a simple, and computationally fast, addition or subtraction and look up of a pre-defined log and antilog table (Sears, 2020). Ships in Elite were also symmetrical, sometimes in more than one axis, which allowed for values of matrices to be negated, resulting in the same vertex position as if the matrix and had been calculated using the real negative position. This resulted in almost-computationally-free matrix multiplication for many vertices of the ships (Braben, 2011).

## Raycast Rendering

Wolfenstein 3D (id Software, 1992) was one of the original first-person shooter games but is not a truly 3D game. Ray casting is a technique which renders a 3D projection of a 2D grid (Vandevenne, 2020). The ray casting algorithm casts a ray for each column of the screen from the players perspective and is stepped across a 2D grid until an intersection is found. The distance in which the ray has travelled determines how many pixels are filled for that column of the screen (Sanglard, 2017).

|  |  |
| --- | --- |
| Diagram of rays being cast from camera (appendix 6) | Resulting rendered image (appendix 6) |

Unlike its predecessors such as Hovertank 3D (id Software, 1991), Wolfenstein 3D’s environments featured textured walls which increased the technical complexity of the engine. Due to the game-world being a stored as a 2D grid, the mantissa of the ray intersection point can be used for a horizontal texture coordinate, and the vertical texture coordinate can be calculated by dividing 1 by the current column height, for each vertical pixel drawn, the current vertical texture coordinate can then be incremented by that amount (Vandevenne, 2020).

An optimisation id Software used for Wolfenstein 3D to increase texturing performance, was to buffer pixel columns if they were similar – columns are considered similar if they share the same horizontal texture coordinate - and batch render them later at the same height regardless of their actual distance from the player, which resulted in slight distortion, but less write operations (Sanglard, 2017).

|  |
| --- |
| comanche-maximum-overkill screenshot for dos  Comanche: Maximum Overdrive (appendix 7) |

Other engines utilised similar ray casting techniques for different styles of rendering. The Voxel Space engine (Novalogic, 1992) rendered a 3D projection of a 2D heightmap, which allowed for rendering of a high detail textured 3D terrain, featuring valleys, shadows, and man-made structures. The algorithm used, often referred to as floorcasting or heightmap ray casting (Deschaseaux, 2018), begins by sampling the heightmap and colour map for each horizontal pixel, at the furthest point in the view frustrum and drawing a vertical column of pixels of the sampled colour of a height determined by the sampled heightmap value (Macke, 2020). This algorithm repeats sampling increasingly closer to the camera, overriding previous pixels to guarantee occlusion, this type of algorithm is referred to as a painters algorithm (Hughes et al., 2013). An optimisation for this technique is to store a y-buffer for each screen column and sample the height and colour maps increasingly further from the camera, referencing the y-buffer to determine whether the current column is entirely or partly occluded, and drawing only the non-occluded portion of each column.

## Rasterised Rendering

Over time, graphics pipelines and hardware have improved to meet the market demand for a higher visual fidelity in real-time applications. Dedicated graphics processing hardware, also known as graphics processing units (GPUs), gained popularity in the 1990’s and have since evolved to be programmable (Verschelde, 2021) and include various hardware acceleration chips (Walton, 2020) alongside the ever-improving hardware for efficiently rendering triangles – for rasterisation. The rasterisation rendering technique is the most common rendering technique for 3D real-time applications (Scratchapixel, 2015).

The rasterisation process fundamentally begins by projecting vertices into screen space via the vertex shader. The vertex shader multiplies each vertex position by the World and View transformation matrices, which brings each vertex into world space then into space relative to the camera respectively. Finally, the camera space vertex position is multiplied by a Projection matrix which brings the coordinate into projection space (Akenine-Möller et al., 2008). The Projection matrix is where a perspective or orthographic camera effect can be applied, depending on how the Projection matrix is constructed.

The graphics pipeline goes through other programmable and fixed-function stages before eventually reaching the pixel shader. The pixel shader is another programmable stage, similarly to the vertex shader stage, which is run in parallel once for each pixel which is occupied by a triangle. The pixel shader, also referred to as the fragment shader, processes the data provided by the vertex shader and calculates a colour for that pixel (Caulfield, 2018).



Diagram demonstrating rasterisation of a triangle (Appendix 1)

## Shaders

Graphics Processing Units are used for rendering, over CPUs, due to the focus on a parallel architecture which allows for much faster processing on specialised tasks such as rendering. A CPU has fewer, but more powerful cores, whereas a GPU has many smaller cores which allows for many simultaneous programs, known as Shaders, to be executed at once (Vivo & Lowe, 2015). The ability to execute many shaders at once is beneficial when the application requires a Shader to be executed once per pixel and/or per vertex.

Modern GPUs can be programmed by executing Shaders, which are created using C-Style languages such as HLSL or GLSL which are then compiled to an intermediate language to be executed (Akenine-Möller et al., 2008). GPUs were originally not programmable and were entirely fixed function (Tamasi, 2008), but over time have become more flexible and programmable, leading to the GPUs of today which have evolved to become more than hardware for rendering games. Although real-time rendering is their primary use, modern GPUs can be utilised as general-purpose parallel processors (Intel, 2021) and can be applied to many different applications outside of gaming/real-time rendering.

Since the dawn of programmable GPUs, other shader types have been introduced which provides developers with more control over the GPU and how it is used. After the introduction of programmable vertex and pixel/fragment shaders, a geometry shader stage was introduced which is responsible for creating new primitives from the output of the vertex shader. Compute shaders were also introduced but are not part of the graphics pipeline, they allow for general computation to be performed on the GPU (Rodríguez, 2013) and the highly parallelised architecture to be taken advantage of outside of graphics rendering.

## Ray Traced Rendering

Ray tracing is a rendering technique which attempts to emulate the basic principle of vision. Our eyes see the environment around us because of light rays reflecting from surfaces and hitting our eyes. Ray tracing flips the direction of these rays, projecting them from the camera (eye) and shading the pixel based on the lighting and material at the intersection point. This ray reversal is an optimisation so light rays which would not hit the camera (eye) are not considered (Deng et al., 2017).

Finding the intersection between a ray and an object is the fundamental operation of ray tracing (Glassner et al., 1989), the intersection point paired with the surface normal allows for various visual effects to be simulated by bouncing the ray from surface to surface, similarly to how light bounces from surface to surface in real life. Ray tracing algorithms often use a recursive trace function which runs for a specified number of bounces (Buck, 1999). This allows the colour of the pixel to accumulate every time the ray bounces from one object to another. The accumulation of the colour through the recursive function allows for effects such as the reflections of a surface also having their own reflections (Szirmay-Kalos et al., 1995).

A picture containing plate, several, surrounded

Description automatically generated

Recursive ray traced reflections (Appendix 2)

Due to the general computational expense (Rinker, 1991) required to utilise ray tracing in real-time applications, ray tracing was not a viable technique for mainstream video games to integrate for an improvement in fidelity until NVIDIA released their 20 series GPUs, which used the Turing architecture containing new ray tracing (RT) cores for hardware-accelerated ray tracing (Kilgarif et al., 2018). Many games utilise these cores to implement a hybrid-rendering approach, which renders the scene with traditional rasterised techniques and utilises the RT cores for visual effects which provide the most benefit, such as reflections or diffuse global illumination (Sjoholm, 2018).

## Ray Marching

### Algorithm

The fundamental concept of ray marching is to step along a ray in a specified direction until an intersection has been found. The naïve approach to implement this would be to step along the ray linearly using a fixed step size (Biagioli, 2016), this algorithm is simple but introduces potential problems. The accuracy of this algorithm is tied to the step size – a shorter step is more accurate but computationally more expensive – which means a perfectly accurate step size is unachievable.

One of the potential problems with using a linear step size is that an intersection can be found inside the geometry, as seen in the following visual representation:

Another problem is that geometry can be entirely stepped over, meaning it would not be rendered. This would occur when trying to render, for example, a thin wall or an object with a high amount of fine detail (such as a fractal).

An improved algorithm is to ray march across distance fields, where the step size at each increment is determined by the shortest distance from the current position to any object in the scene. As the ray approaches a surface, the distance travelled becomes increasingly shorter (Hart et al., 1989), when the distance becomes lower than a specified threshold, the ray is considered to have intersected with the geometry. This method provides increased accuracy as it is impossible for the ray to pass inside, or through, any geometry. This method is referred to as sphere tracing.

The following diagram is a 2D visual representation of sphere tracing, each circle’s radius represents the shortest distance to a surface at each step, which is how far the ray is stepped at each iteration along the direction of the ray.

### Lighting and Shading

* Shadows
  + Soft Shadows
* Ambient Occulsion
  + Fast AO
  + Real AO
* Surface Shading
  + BRDF

### Rendering Volumes

Ray marching is commonly used for the real-time rendering of volumetric objects, such as clouds. Volumes can be simulated using traditional rasterised rendering techniques by layering multiple flat surfaces, but this lacks visual fidelity especially when up-close or inside the volume.

Using ray marching for volume rendering allows for a much more realistic representation of the volume, as the ray can sample density multiple times from within the volume and calculate the colour for the current pixel (Häggström, 2018). The technique of rendering volumetric objects with ray marching has been used alongside traditional rasterised rendering for games (Schneider & Vos, 2015).

### Point-cloud Rendering

Point-cloud rendering is where a collection of points in space are represented visually, instead of an interconnected topology (Shahrabi, 2019). There are various reasons why a point-cloud rendering technique would be used, in the case of this research it would be used for performance – per-pixel rendering requires each pixel to be ray marched, point-cloud rendering does not require a ray for each pixel – and aesthetic purposes. Media Molecule utilises SDFs for rendering in their game “Dreams”. The game allows the user to sculpt objects and apply effects to them. A point-cloud rendering technique is used to achieve a unique painterly effect by rendering the points as paint splats (Evans, 2015). The source of the following images can be found in appendix 3.

|  |  |
| --- | --- |
| A picture containing indoor, decorated  Description automatically generated  Scene rendered with a high-density point-cloud | A map of the world  Description automatically generated with medium confidence  Scene rendered with a low-density point-cloud |

### Constructive Solid Geometry

Constructive solid geometry is a method of creating complex objects from simple primitives using Boolean operations. A Boolean operation takes two objects – or in the case of ray marching SDFs, the distance to two objects – and returns a new result, a new distance value (Wong, 2016).

There are three primitive Boolean operations: Intersection, union, and difference/subtract. Intersection returns only the space in which both objects occupy. Union returns both objects combined. Difference returns a subtraction of the second object from the first object. The following image only uses three primitive types of objects, but results in a relatively complex object:



CSG demonstration (Appendix 4)

### Fractals

Fractals are commonly found in nature, such as in snowflakes, plants, coastlines, and galaxies. Benoit Mandelbrot, the person who coined the name ‘Fractal’ and pioneered early research into fractals, explored the occurrence of fractals in nature in his book ‘The Fractal Geometry of Nature’ where he states *“Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line. More generally, I claim that many patterns of Nature are so irregular and fragmented, that, compared with [standard geometry,] Nature exhibits not simply a higher degree but an altogether different level of complexity.”* (Mandelbrot, 1982). An example of a computer-generated fractal which mimics nature is the Barnsley Fern, which was first described by Michael Barnsley in his book ‘Fractals Everywhere’ (Barnsley, 1993), where the drawn fractal closely resembles a black spleenwort fern.

As described by Della-Bosca et al. (2014), virtual game worlds are dominated by the Cartesian concept of space. Built rectilinearly, polygon-by-polygon, into a collective, familiar experience. The natural world contrasts this as it is filled with rough edges, asymmetries, and complex interactions. Fractals can provide rich and interesting visuals, not only for the real natural world, but also in virtual worlds, particularly to provide the ability for shapes to appear more natural.

Ray marching enables the rendering of 3D deterministic fractals, which can provide unique, complex, and visually appealing geometry in games as explored by the open-source physics video game ‘Marble Marcher’ (HackerPoet, 2019). The goal of the video game is to traverse various fractal terrains to reach the flag in the shortest time possible. Many of the terrains featured in Marble Marcher appear organic, whether in a familiar, earth-like manner, or an alien appearance. Others appear industrial, with straight edges and grid patterns.

# Research Methodology

## Project Management

For this research, three project management frameworks have been considered. The reasonings for choosing between Waterfall, Kanban, and Scrum, are that each framework offers advantages and disadvantages, and they are each unique. Kanban and Scrum are both Agile frameworks, yet they are both very different in how they operate, which is why both are considered.

Add more methodologies.

### Waterfall

The Waterfall methodology is a linear approach to project management. Each stage of development must be complete before the next one begins as each stage depends on the previous one. Once a stage of development is complete, it should not be returned to.

The advantages of using the Waterfall methodology are that the structure is clear and easy to understand. The following phase of development cannot be started until the current one is finished. Another advantage is that the end goal is defined early in the project’s life cycle, which helps to keep the team aligned to the goal and prevents the project from growing out of proportion.

The disadvantages of using the Waterfall methodology are due to the linearity. It is difficult to make changes as it would deviate from the planned goal. Another disadvantage is that testing does not start until after completion. (Lucid Content Team, 2018)

### Kanban

Kanban is an agile methodology that focuses on flexibility and adaptability. Agile frameworks allow for multiple stages of development to be in progress simultaneously, as the project is broken down into small segments which allow for iteration. Kanban uses cards to represent tasks, which can be moved from stage-to-stage of development, the ‘work-in-progress’ stage of development must have limits to how many tasks can be assigned simultaneously.

The advantages of using Kanban come from its flexibility. Due to each task having its own development cycle, there is constant progress being made even when other features may still be in an early phase.

Flexibility is also what gives Kanban disadvantages too, as an improperly maintained board can result in issues with development and the board can become overcomplicated. Kanban also does not have specified time-frames alike Scrum, which can more easily lead to delays or roadblocks. (PLANETTOGETHER, 2019)

### Scrum

Scrum is also an agile methodology, which allows for tasks to be completed incrementally and simultaneously. Scrum differs by using sprints of a specified time (usually two weeks) to complete tasks. Tasks are then reviewed as a team following the sprint, where they can be evaluated, and team members can sync to avoid any interference with development progress. A scrum master is assigned to ensure the team is following the correct processes and to remove obstacles.

Scrum offers more transparency and visibility across the entire workflow than other frameworks, which is an advantage as changes are easy, and the team is always in sync. This leads to improved productivity for a team.

The disadvantages of using Scrum for this research stems from Scrum having many different roles. Scrum can only be taken full advantage of when working in a team, and due to this research being a solo project, Scrum would be ineffective (PLANETTOGETHER, 2019).

### Conclusion

Kanban will be the methodology used for research and development throughout this project.

Waterfall will be too restrictive for this project due to its linearity. There would be a risk of planning and researching more than is viable to implement within the time constraint of the project, which is not a problem that is present for other, agile, methodologies.

The Scrum methodology is well-suited for use in a team, as it allows for iteration and simultaneous stages of development whilst keeping all members in sync. Due to the nature of this research project being a solo effort, Scrum would not be suitable.

Kanban is a flexible methodology that doesn’t require a team to be fully taken advantage of. Tasks can be researched and developed simultaneously, which mitigates the risk of planning more than can be implemented within the time constraint.

## Quantitative vs. Qualitative

## Platform

Three platforms have been considered for the development of this project. Each platform is considerably unique and offers advantages and disadvantages. DirectX11 and Vulkan are both graphics APIs that would allow for the creation of a rendering engine or other applications which need to utilise graphics hardware. Unity differs, as it is already a complete real-time development platform (Unity Technologies, 2021).

### Unity

Unity is a real-time development platform that includes tools to assist the creation of games or other real-time rendered media. Ray marching in Unity would be possible, as it allows for user-made shaders and control over the rendering pipeline via the scriptable render pipeline.

The advantage of using Unity for the development of this project would be the collection tools already built into the engine, which could be utilised with the ray marching renderer to provide the user with a familiar environment.

These tools and extra and unused features could lead to overhead on performance, which would be one of the disadvantages of using Unity for this project. This would not likely be a problem if using DirectX or Vulkan as the engine would be built from the ground-up to only include what is necessary.

Compare with Unreal Engine, differences, advantages/disadvantages.

### DirectX11

DirectX11 is an API that can be used by an application to utilise graphics hardware, it supports multiple shader types, such as pixel and compute shaders (Microsoft, 2020), which could be utilised for ray marching. An alternative to DirectX11 would be OpenGL, which is comparable in terms of functionality and design.

The advantages of using DirectX11 are that there would not be unused feature overhead that could be present when using Unity. DirectX is relatively (to Vulkan) high level, which means development progress would be faster than using an API such as Vulkan.

The disadvantages of using DirectX11 are that there would be more work than if using Unity, as some of the existing tools would have to be made from scratch. The performance also may not be as fast as if the project was developed using a lower-level API such as Vulkan, performance advantages of Vulkan will be further covered in the following Vulkan consideration.

Compare with OpenGL, differences, advantages/disadvantages.

### Vulkan

Vulkan is also an API to utilise graphics hardware, it differs from DirectX11 as it is much lower-level and verbose (NVIDIA, 2021), giving the developer more control over the entire rendering process. An alternative to Vulkan would be DirectX12, which is similarly low-level and too offers fine control to the developer.

The advantage of using Vulkan, over another API such as DirectX11, is that the finer control which the API provides can lead to improved performance, as driver overhead is mitigated due to driver software having to make less assumptions on what the application is doing. Another performance advantage Vulkan provides is the batch processing feature, which can help to reduce CPU load (Furmaniak, 2018).

This control is also one of Vulkan’s disadvantages. Having more control over the API could cause the program, if improperly utilised, to operate slower than a higher-level API. The verbosity of the code also makes using Vulkan much more complex than a higher-level API, which is likely to slow the development progress.

Compare with DirectX12, differences, advantages/disadvantages.

### Conclusion

The platform on which this project will be built is DirectX11, as it strikes a balance between being lightweight to avoid unnecessary overhead, but high-level enough that the project will be able to progress at a reasonable rate.

Using DirectX11 would also enable research into implementing other standard features of a game engine, for example, Gizmos, Scene Management, Camera Controls, etc. which Unity would have otherwise handled.

## Project Plan

The project will be planned and managed using Trello. A Kanban board can be created with Trello to break the project into individual tasks and subtasks which gives an overview of the status of the project. The task limit, per stage, will be two tasks. This will prevent new features being worked on before other features are finished but allows a degree of flexibility and simultaneous development.

The Trello project has four lists to categorise the stage of each task: To Do; Pending; Blocked; Complete. There is a fifth list to contain extra project resources.

The ‘To Do’ stage will be for features which have not yet started and is where all tasks begin. Once a task has begun development, it will be moved into the ‘Pending’ stage. The ‘Pending’ stage is where the task limit is applied. The blocked stage will contain tasks which are not yet finished and cannot be finished yet. For example, if Task A could not progress any further without the completion of Task C, Task A would then become blocked until Task C is complete. Finally, tasks are moved into the ‘Complete’ stage once each subtask is finished and there is nothing more which is required.



### Gantt Chart

## Analysis

A description of the choice of analysis or problem-solving method. It is important to describe the process by which the method is chosen to show that it is appropriate for the problem situation.

A narrative description of the application of the analysis method, indicating the problems which arose during this process and how they were identified and overcome. Most projects will include models, charts, or diagrams at this stage. These may be included in the chapter or in an appendix.

In short, investigate all possible options available to you to successfully achieve the scope of the project in the time given.

## Design

Here you will justify your chosen tools and processes from your previous research into the background of the project and the options available to you in the analysis.

Identify and justify:

* The choice of an appropriate method.
* Your experience of its application.

### Required Software

#### Visual Studio 2022

Visual Studio 2022 will be the IDE used for the development of this project. Visual Studio 2022 is the most recent version of the software which provides features such an upgrade to 64-bit architecture which provides increased IDE performance, better cross-platform development tools and the latest C++ build tools which support C++20 (Microsoft, 2021c).

Another popular C++ IDE is CLion (JetBrains, 2022), which is developed by JetBrains. CLion was also considered due to the known quality of other products developed by JetBrains and direct focus on being a C/C++ IDE. The final decision to use Visual Studio is because it supports multiple programming language and build tools, is widely used within the industry, and, alike the other considerations, offers a free non-commercial license which can be used for this research project.

#### C++

The language of choice for the implementation of this research project is C++. The C++ language is cross-platform and allows low-level interaction with the hardware. An alternative to C++ would be the C language, but C++ is preferred as it allows for object-oriented architecture (DataFlair, 2020) which would be advantageous for structuring the codebase.

C++ is natively supported with major graphics API’s (KhronosGroup, 2022; Microsoft, 2021a; Apple, 2022)which gives an advantage over many other languages which offer similar low-level access and performance, but not a direct implementation of graphics API’s and other useful libraries.

#### ImGui

#### DirectXTK

#### Git / GitHub

#### Trello

### Required Hardware

# Implementation

## Development

### Application Framework

To begin developing the ray marcher engine, a DirectX 11 framework was needed as a basis. The minimum requirements of this framework were: mesh rendering, shader compilation, camera matrix calculations, and a GUI. These topics had previously been researched for the module ‘Advanced Graphics and Realtime Rendering’, so the base code for this engine was referenced from the existing codebase and stripped down to only the minimum required features. The project uses DirectXTK (Microsoft, 2021b) for the initial codebase, which provides a general structure for the engine and fundamental features such as window handling.

ImGui (Cornut, 2021) has been used for the GUI library due to its extensive number of features, wide use across the games industry, and ease of use. As ImGui is open source, many add-ons are available which integrate seamlessly into the library and provide extra functionality. For example, ImGuizmo is an add-on which will be used later in the project, for the implementation of game object and camera control gizmos.

Before beginning on the ray marching implementation, a quad had to be rendered to the entire screen. A pixel shader, contain the ray marching code, can then execute over each pixel. This can be achieved by creating a quad with vertex positions ranging from minus one to one. In the Vertex shader, matrix multiplications would usually be performed to move the vertex into world, view, and projection space, if none of these calculations are performed the quad covers the screen. This quad is then rendered to an ImGui window which allows for it to be docked and moved to different regions of the screen.

Background pattern

Description automatically generated

Quad rendered to the viewport window. Texture coordinates output as a colour via the pixel shader.

### Basic Ray Marcher

To ray march the scene, a sphere tracing algorithm is used for its improved accuracy over a linear step algorithm. For each pixel rendered, a ray direction needs to be calculated for the initial ray. This direction will determine visual aspects such as a perspective effect, the camera’s field of view, and the direction in which the camera is facing. The ray direction vector can be calculated using the following equations:

Ray direction calculation

is the transposed view matrix, which is calculated using the DirectX XMMatrixLookAtLH() function. and are the current floating point screen coordinates ranging from -1 to 1. is the camera’s field of view value in radians (Scratchapixel, 2014). is the normalised ray.

Once the ray direction had been calculated, the ray marching algorithm could begin. An SDF is needed to test the algorithm, so a sphere was used as it is one of the simplest distance functions and would be immediately clear if there were any problems. The distance to a sphere (Quilez, n.d.) can be calculated using the following equation:

Signed distance function of a sphere

The ray marching loop can then begin. The ray is stepped along by the distance to the sphere at each point. On each iteration of the loop, before the ray has been stepped forward, a check takes place to see if the distance to the sphere is less than a specified threshold value, if it is, the sphere is considered to have been intersected and the loop can exit. The following pseudocode summarised the algorithm:

|  |
| --- |
| depth **=** **0**  **for** **(**i **=** **0,** i **<** MAX\_STEPS**,** i**++):**  distance **=** SphereDist**(**origin **+** direction **\*** depth**)**  **if** **(**distance **<** THRESHOLD**):**  **return** true  depth **+=** distance  **return** false |

Sphere tracing algorithm

Once this algorithm had been implemented in the pixel shader, a colour could be selected based on the Boolean return value and resulted in the following image:

Shape

Description automatically generated

Sphere with a radius of 1

The image is outputting the ray direction as a colour if no intersection is found, and a bright yellow colour if an intersection is found.

Only having a Boolean return value is very limiting, and there is much more intersection data to be utilised, so the ray marching function has been modified to return a struct containing other useful information such as position, depth, and number of steps.

The following image is the output when the number of steps is used for the colour. It results in a Fresnel effect around the sphere as the algorithm must take more, shorter, steps to reach the sphere on the edges.



Sphere with a radius of 2

To implement effects such as lighting/shading, the surface normal of the intersection would need to be calculated. The gradient of the surface can be approximated by sampling points around the given intersection point, and then return the surface normal. This calculation can be represented by the following equation (Wong, 2016), where is a defined epsilon value and is the distance function, and pseudocode snippet:

Approximate normal at given point

|  |
| --- |
| float3 CalculateNormal**(**float3 p**):**  float2 e **=** **{** **0.001**f**,** **0.0**f **}**  float3 n **=** **{** SphereDist**(**p **+** e**.**xyy**)** **-** SphereDist**(**p **-** e**.**xyy**),**  SphereDist**(**p **+** e**.**yxy**)** **-** SphereDist**(**p **-** e**.**yxy**),**  SphereDist**(**p **+** e**.**yyx**)** **-** SphereDist**(**p **-** e**.**yyx**)** **}**  **return** normalise**(**n**)** |

Pseudocode function to approximate surface normal at given point

## Testing

This chapter should address the evaluation of the solution against its objectives and success criteria.

Consider:

* A description of the testing strategy and the choice of testing method.
* The planning and application of the tests. How have you concluded this is the correct type of test to run?

## Results

The conclusions that may be drawn from the results of the tests and any modifications to the design and implementation that could be recommended.

What are you results?

What analysis can you identify from them?

Did you find something you did not expect?

Or was it exactly what you expected?

# Critical Evaluation

This chapter is of **crucial importance** to the whole work. It deals with the success of the project in academic terms, rather than the success criteria for the solution. Even the best analysis, design and implementation will be let down by an inadequate critical evaluation. The examiners will look at this chapter most carefully when determining the success (or otherwise) of the project. Although the exact nature of the evaluation will vary between projects, it is possible to identify certain issues which should be addressed:

* Your evaluation of the degree of success in carrying out the project
* What you have learned by doing the project
* What you would do differently if the project were to be repeated
* Any extra features you would recommend if the project could be extended
* The value to you of the learning process and the extent to which the project has added to your professional and academic expertise
* What future projects open the next chapter, should this projects research be continued further into Master Degree for example?

# Bibliography

Akenine-Möller, T., Haines, E. & Hoffman, N. (2008). *Real-Time Rendering: Third Edition*. Third. Taylor & Francis.

Akenine-Möller, T., Haines, E., Hoffman, N., Pesce, A., Hillaire, S. & Iwanicki, M. (2018). *Real-Time Rendering: Fourth Edition*. Fourth. [Online]. A K Peters/CRC Press. Available from: https://www.realtimerendering.com/Real-Time\_Rendering\_4th-Real-Time\_Ray\_Tracing.pdf. [Accessed: 25 October 2021].

Apple (2022). *Getting started with Metal-cpp*. [Online]. Available from: https://developer.apple.com/metal/cpp/. [Accessed: 20 January 2022].

Barnsley, M. (1993). *Fractals Everywhere*. 2nd Ed. AP Professional.

Biagioli, A. (2016). *Raymarching Distance Fields: Concepts and Implementation in Unity*. [Online]. 1 October 2016. Available from: https://adrianb.io/2016/10/01/raymarching.html. [Accessed: 30 October 2021].

Braben, D. (2011). *Classic Game Postmortem - ELITE*. [Online Video]. 2011. Available from: https://www.gdcvault.com/play/1014628/Classic-Game-Postmortem. [Accessed: 30 January 2022].

Bresenham, J.E. (1965). Algorithm for computer control of a digital plotter. *IBM Systems Journal*. 4 (1). p.pp. 25–30.

Buck, J. (1999). *The Recursive Ray Tracing Algorithm*. [Online]. Available from: http://www.geocities.ws/jamisbuck/raytracing.html#RayTracingOverview. [Accessed: 20 January 2022].

Caulfield, B. (2018). *What’s the Difference Between Ray Tracing and Rasterization?*

Cornut, O. (2021). Dear ImGui. *GitHub repository*. [Online]. Available from: https://github.com/ocornut/imgui.

DataFlair (2020). *Advantages and Disadvantages of C++*.

Della-Bosca, D., Patterson, D. & Costain, S. (2014). *Fractal Complexity in Built and Game Environments*. In: [Online]. pp. 167–172. Available from: https://hal.inria.fr/hal-01408518. [Accessed: 12 December 2021].

Deng, Y., Ni, Y., Li, Z., Mu, S. & Zhang, W. (2017). Toward Real-Time Ray Tracing. *ACM Computing Surveys*. 50 (4). p.pp. 1–41.

Deschaseaux, P. (2018). *Heightmap rendering using a floorcasting algorithm*. [Online]. Available from: https://observablehq.com/@ehouais/heightmap-rendering-using-a-floorcasting-algorithm. [Accessed: 3 February 2022].

Evans, A. (2015). *Learning from failure: A Survey of Promising, Unconventional and Mostly Abandoned Renderers for ‘Dreams PS4’, a Geometrically Dense, Painterly UGC Game*. [Online]. Available from: http://advances.realtimerendering.com/s2015/AlexEvans\_SIGGRAPH-2015-sml.pdf. [Accessed: 2 November 2021].

Frontier Developments (1984). *Elite*.

Furmaniak, L. (2018). *Stuck on OpenGL ES? Time to move on! Why Vulkan is the future of graphics*. [Online]. Available from: https://www.imaginationtech.com/blog/stuck-on-opengl-es-time-to-move-on-why-vulkan-is-the-future-of-graphics/. [Accessed: 24 November 2021].

Glassner, A., Haines, E., Hanrahan, P., Arvo, J., Cook, R., Heckbert, P. & Kirk, D. (1989). *An Introduction to Ray Tracing*. [Online]. Available from: https://www.realtimerendering.com/raytracing/An-Introduction-to-Ray-Tracing-The-Morgan-Kaufmann-Series-in-Computer-Graphics-.pdf. [Accessed: 19 January 2022].

HackerPoet (2019). MarbleMarcher. *GitHub repository*. [Online]. Available from: https://github.com/HackerPoet/MarbleMarcher. [Accessed: 12 December 2021].

Häggström, F. (2018). *Real-time rendering of volumetric clouds*. [Online]. Umeå. Available from: http://www.diva-portal.org/smash/get/diva2:1223894/FULLTEXT01.pdf. [Accessed: 2 November 2021].

Hart, J.C. (1996). Sphere tracing: a geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer*. [Online]. 12 (10). Available from: https://www.researchgate.net/publication/2792108\_Sphere\_Tracing\_A\_Geometric\_Method\_for\_the\_Antialiased\_Ray\_Tracing\_of\_Implicit\_Surfaces. [Accessed: 26 October 2021].

Hart, J.C., Sandin, D.J. & Kauffman, L.H. (1989). Ray tracing deterministic 3-D fractals. *ACM SIGGRAPH Computer Graphics*. [Online]. 23 (3). Available from: https://www.cs.drexel.edu/~david/Classes/Papers/rtqjs.pdf. [Accessed: 30 October 2021].

Hughes, J.F., van Dam, A., Mcguire, M., Sklar, D.F., Foley, J.D., Feiner, S.K. & Akeley, K. (2013). *Computer Graphics: Principles and Practice*. Third Edition. Addison-Wesley Professional.

id Software (1991). *Hovertank 3D*.

id Software (1992). *Wolfenstein 3D*.

Intel (2021). *CPU vs GPU: What’s the Difference*. [Online]. Available from: https://www.intel.co.uk/content/www/uk/en/products/docs/processors/cpu-vs-gpu.html. [Accessed: 19 January 2022].

JetBrains (2022). *CLion: A cross-platform IDE for C and C++*. [Online]. Available from: Jetbrains.com/clion/. [Accessed: 20 January 2022].

KhronosGroup (2022). Vulkan-Hpp. *GitHub repository*. [Online]. Available from: https://github.com/KhronosGroup/Vulkan-Hpp.

Kilgarif, E., Moreton, H., Stam, N. & Bell, B. (2018). *NVIDIA Turing Architecture In-Depth*. [Online]. Available from: https://developer.nvidia.com/blog/nvidia-turing-architecture-in-depth/#:~:text=Turing%20GPUs%20introduce%20new%20RT,tracing%20approaches%20of%20the%20past. [Accessed: 19 January 2022].

Lucid Content Team (2018). *The Pros and Cons of Waterfall Methodology*. [Online]. 10 August 2018. Available from: https://www.lucidchart.com/blog/pros-and-cons-of-waterfall-methodology. [Accessed: 7 November 2021].

Macke, S. (2020). *Voxel Space - Terrain rendering algorithm in less than 20 lines of code*. [Online]. Available from: https://github.com/s-macke/VoxelSpace. [Accessed: 3 February 2022].

Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman and Co.

Microsoft (2021a). *Direct3D 12 programming environment setup*. [Online]. Available from: https://docs.microsoft.com/en-us/windows/win32/direct3d12/directx-12-programming-environment-set-up. [Accessed: 20 January 2022].

Microsoft (2021b). DirectXTK. *GitHub repository*. [Online]. Available from: https://github.com/microsoft/DirectXTK.

Microsoft (2020). *Graphics Pipeline*. [Online]. 4 November 2020. Available from: https://docs.microsoft.com/en-us/windows/win32/direct3d11/overviews-direct3d-11-graphics-pipeline. [Accessed: 8 November 2021].

Microsoft (2021c). *What’s new in Visual Studio 2022*. [Online]. Available from: https://docs.microsoft.com/en-us/visualstudio/ide/whats-new-visual-studio-2022?view=vs-2022. [Accessed: 20 January 2022].

Moxon, M. (2020). *Bresenham’s line algorithm*. [Online]. Available from: https://www.bbcelite.com/deep\_dives/bresenhams\_line\_algorithm.html. [Accessed: 30 January 2022].

Novalogic (1992). *Voxel Space*.

NVIDIA (2021). vk\_mini\_path\_tracer. *GitHub repository*. [Online]. Available from: https://github.com/nvpro-samples/vk\_mini\_path\_tracer. [Accessed: 8 November 2021].

PLANETTOGETHER (2019). *Agile Methodologies: Kanban Vs Scrum - Advantages and Disadvantages*. [Online]. 31 December 2019. Available from: https://www.planettogether.com/blog/agile-methodologies-kanban-vs-scrum-advantages-and-disadvantages. [Accessed: 7 November 2021].

Quilez, I. (n.d.). *Article on Distance Functions*. [Online]. Available from: https://www.iquilezles.org/www/articles/distfunctions/distfunctions.htm. [Accessed: 24 November 2021].

Retro64 (2017). *An introduction to 6502 math: addition, subtraction and more*. [Online]. Available from: https://retro64.altervista.org/blog/an-introduction-to-6502-math-addiction-subtraction-and-more/. [Accessed: 1 February 2022].

Rinker, R. (1991). *Reducing Computational Expense of Ray-Tracing Using Surface Oriented Pre-Computation*.

Rodríguez, J. (2013). *GLSL Essentials*.

Sanglard, F. (2017). *Game Engine Black Book: Wolfenstein 3D*.

Schneider, A. & Vos, N. (2015). The Real-time Volumetric Cloudscapes of Horizon: Zero Dawn. *Siggraph 2015*. [Online]. Available from: https://www.guerrilla-games.com/read/the-real-time-volumetric-cloudscapes-of-horizon-zero-dawn. [Accessed: 15 January 2022].

Scratchapixel (2015). *Rasterization: a Practical Implementation*. [Online]. Available from: https://www.scratchapixel.com/lessons/3d-basic-rendering/rasterization-practical-implementation. [Accessed: 18 January 2022].

Scratchapixel (2014). *Ray-Tracing: Generating Camera Rays*. [Online]. Available from: https://www.scratchapixel.com/lessons/3d-basic-rendering/ray-tracing-generating-camera-rays/generating-camera-rays. [Accessed: 23 November 2021].

Sears, D.W. (2020). *Introduction to Multiplication and Division by Logarithmic Algebra*. [Online]. Available from: https://biosci.mcdb.ucsb.edu/biochemistry/tw-lig/logarithms/logarithmic-algebra.htm. [Accessed: 2 February 2022].

Shahrabi, S. (2019). *Point cloud rendering*. [Online]. Available from: https://medium.com/realities-io/point-cloud-rendering-7bd83c6220c8. [Accessed: 27 November 2021].

Sjoholm, J. (2018). *Effectively Integrating RTX Ray Tracing into a Real-Time Rendering Engine*. [Online]. Available from: https://developer.nvidia.com/blog/effectively-integrating-rtx-ray-tracing-real-time-rendering-engine/#:~:text=Effectively%20Integrating%20RTX%20Ray%20Tracing%20into%20a%20Real%2DTime%20Rendering%20Engine,-By%20Juha%20Sjoholm&text=RTX%20is%20NVIDIA’s%20new%20platform,number%20of%20games%20and%20engines. [Accessed: 20 January 2022].

Szirmay-Kalos, L., Márton, G., Dobos, B., Horváth, T., Risztics, P. & Kovács, E. (1995). *Theory of Three Dimensional Computer Graphics*.

Tamasi, T. (2008). The Evolution of Computer Graphics. *NVISION 08*. [Online]. Available from: https://www.nvidia.com/content/nvision2008/tech\_presentations/Technology\_Keynotes/NVISION08-Tech\_Keynote-GPU.pdf. [Accessed: 19 January 2022].

Underground Mathematics (2022). *Convex shape*. [Online]. Available from: https://undergroundmathematics.org/glossary/convex-shape. [Accessed: 1 February 2022].

Unity Technologies (2021). *Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine*. [Online]. 2021. Available from: https://unity.com/. [Accessed: 8 November 2021].

Vandevenne, L. (2020). *Raycasting*. [Online]. Available from: https://lodev.org/cgtutor/raycasting.html. [Accessed: 2 February 2022].

Verschelde, J. (2021). Evolution of Graphics Pipelines. *Introduction to Supercomputing*. [Online]. Available from: http://homepages.math.uic.edu/~jan/mcs572/gpu\_evolution.pdf. [Accessed: 18 January 2022].

Vivo, P.G. & Lowe, J. (2015). *The Book of Shaders*. [Online]. Available from: https://thebookofshaders.com/. [Accessed: 19 January 2022].

Walton, J. (2020). *Nvidia RTX 30-Series Ampere Architecture Deep Dive: Everything We Know*.

Whitted, T. & Foley, J.D. (1980). An Improved Illumination Model for Shaded Display. *Graphics and Image Processing*. [Online]. 23. p.pp. 343–349. Available from: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.156.1534. [Accessed: 26 October 2021].

Wong, J. (2016). *Ray Marching and Signed Distance Functions*. [Online]. Available from: http://jamie-wong.com/2016/07/15/ray-marching-signed-distance-functions. [Accessed: 26 November 2021].

# Appendices REMOVE THIS / CHANGE TO PICTURE BIBLIOG. Appendices should be Figures instead, most likely need reordering (i.e. Appendix 5 is not the 5th image displayed, so should be not just be figure 5)

## Appendix 1: Rasterisation

Image by raywenderlich.com (n.d.):

<https://www.raywenderlich.com/books/metal-by-tutorials/v2.0/chapters/3-the-rendering-pipeline>

## Appendix 2: Recursive Sphere Reflections

Image by David Kuri (2018):

<http://blog.three-eyed-games.com/2018/05/03/gpu-ray-tracing-in-unity-part-1/>

## Appendix 3: Point-Cloud Rendering

Images by Media Molecule (2015):

<http://advances.realtimerendering.com/s2015/AlexEvans_SIGGRAPH-2015-sml.pdf>

## Appendix 4: Constructive Solid Geometry

Image by Alec Jacobson (2016):

<http://www.alecjacobson.com/weblog/?tag=constructive-solid-geometry>

## Appendix 5: Elite 1984

Image by Robert August de Meijer (2016):

<https://whengameswereking.wordpress.com/2016/12/08/elite-1984/>

## Appendix 6: Ray casting diagram and image

Image by Fabien Sanglard (2017):

(Sanglard, 2017)

## Appendix 7: Comanche: Maximum Overkill

Image by abandonwaredos.com (2022):

<https://www.abandonwaredos.com/abandonware-screenshot.php?gid=1409&idi=YWJhbl9pbWdfc2NyZWVucy9jb21hbmNoZW1heGltdW0tMy5qcGc=&tit=comanche-maximum-overkill>