Real-time Rendering of 3D "Fractal-like" Geometry

Deliverable 1: Final Year Dissertation

22/11/2021

by

Solomon Baarda

Meng Software Engineering

Heriot-Watt University

Supervisor: Dr Benjamin Kenwright

Second Reader: Ali Muzaffar

Abstract

A fractal is a recursively created never-ending pattern that is usually self-similar. Separate from Euclidean geometry, fractal geometry describes the more non-uniform shapes found in nature, like clouds, mountains, and coastlines. Fractal patterns exist everywhere in the universe, whether we can see them or not. From DNA molecules to the structure of galaxies, and everything in between. Fractals appear everywhere in nature and many technological breakthroughs have been made through studying their patterns.

With the increasing popularity in fractal geometry, and increasing computing power, fractal rendering software has become far more common in the last decade. However, only a small number of these programs are capable of rendering 3D fractals in real time, and those that are capable, are mostly written using graphics shaders which contain lots of code duplication between scenes. This makes it hard for a beginner to get into rendering 3D fractals as they must be competent in the chosen shader language and understand must the complex theory of rendering fractals. This project aims to produce a real-time 3D fractal geometry renderer, for which it is easy for a user to create new scenes and add geometry to it.

Declaration

I, Solomon Baarda confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed: 

Date: 16/11/2021

Table of Contents

Table of Figures

Table of Tables

Common Definitions

Table 1.1.i Common definitions

Word	Definition
Complex number	
Euclidian geometry	
Fractal	
Frame	
Geometry	
Method/constant overloading	
Quaternion	
Ray	
Render	
Polyhedrons	
Polygon	
Regular polygon	
Convex polyhedron	

Common Abbreviations

Table 1.1.ii Common abbreviations

Word	Abbreviation
CPU	Central Processing Unit
FPS	Frames per second
GPU	Graphics Processing Unit
PC	Personal computer
SDF	Signed distance function
DE	Distance estimation function

1 INTRODUCTION

All project materials can be found online in the GitHub repository
https://github.com/SolomonBaarda/fractal-geometry-renderer

1.1 PROJECT DESCRIPTION

A fractal is a recursively created never-ending pattern that is usually self-similar [1]. Separate from Euclidean geometry, fractal geometry describes the more non-uniform shapes found in nature, like clouds, mountains, and coastlines. Beniot Mandelbrot, inventor of the concept of fractal geometry, famously wrote "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" [2]. Fractal patterns exist everywhere in our lives [3], whether we can see them or not. From DNA molecules to the structure of galaxies, and everything in between.

Fractals have many applications in the real world [1]–[4]. In medicine, fractals have been used to help distinguish between cancerous cells which grow abnormally, and healthy human blood vessels which typically grow in fractal patterns. In fluid mechanics, fractals have been used to help model both complex turbulence flows and the structure of porous materials. In computer science, fractal compression is an efficient method for compressing images and other files and uses the fractal characteristic that parts of a file will resemble other parts of the same file. Fractal patterns are also used in the design of some cell phone and Wi-Fi antennas, as the fractal design allows them to be made more powerful and compact than other designs. Even losses and gains in the stock market have been described in terms of fractal mathematics. All these applications of fractals demonstrate that this geometry is a fundamental part of our universe, both in physical materials and in theoretical concepts.

With the increasing popularity in fractal geometry, and increasing computing power, fractal rendering software has become far more common in the last decade [5]. However, only a small number of these programs are capable of rendering 3D fractals in real time, and those that are capable, are mostly written using graphics shaders which contain lots of code duplication between scenes. This makes it

hard for a beginner to get into rendering 3D fractals as they must be competent in the chosen shader language and understand must the complex theory of rendering fractals. This project aims to produce a real-time 3D fractal geometry renderer, for which it is easy for a user to create new scenes and add geometry to it.

The term fractal has been used throughout this report to describe a pattern or geometry that displays the recursive self-similarity characteristic of fractals. The term fractal-like has been used to describe something that appears to display the fractal characteristics but may not truly contain infinite detail.

1.2 Aims & Objectives

This project aims to develop an application which can render 3D fractal geometry in real-time. The render will include common optical effects like ambient occlusion, soft shadows, and lighting. In addition, it should be possible to create scenes and view them using the application, and this process should be as straight forward as possible.

Listed below are the key objectives that this project sets out to achieve. These objectives will help guide the project in the correct in the correct direction to achieve its aims.

Objective 1: Research topic

Background research into the project area to gain a better understanding of the chosen topic and the scope of the project.

Objective 2: Investigate existing solutions

Several existing solutions already exist and an analysis of their strengths and flaws could benefit the design of this project's application.

Objective 3: Develop basic functionality

Implement the base functionality of the fractal renderer. This forms the safe core to the project.

Objective 4: Add additional functionality

Implement additional functionality and quality of life improvements for the application. This leaves room for increasing and decreasing the scope of the project.

Objective 5: Evaluation

Benchmark the performance of the application across various systems and evaluate how successfully the project aims were achieved.

Objective 6: Create user documentation

Create documentation to assist users or future developers of the application.

It is necessary to complete many of the objectives in the order they are specified, as they build upon previous objectives. These objectives have been created bearing the SMART properties [6] in mind. SMART stands for: Specific, Measurable, Achievable, Realistic and Time constrained.

These objectives form the main tasks to be completed during the duration of this project and the requirements specification in section 3.2 and the Gantt chart in section 6.5 have been structured around these.

1.3 SCOPE

The scope of the project has been carefully considered, and several stretch goals have been included in the requirements specification if good progress is made. Objective 4 leaves large amounts of flexibility and can be extended or cut back depending on time constraints.

Objectives 1 and 2 have been completed, and so has part of objective 3. Some initial experimentation rendering still images of the Mandel bulb fractal and other geometry has been very successful. Some of these renders can be viewed in the appendix, section 9. In addition, some experimentation with OpenCL, a library that allows GPU parallel code to be written and executed, has been completed. These experiments were done to gain familiarity with this style of programming in the hope to reduce the learning curve of this new software.

1.4 Document Structure

Continuing from section 1, section 2 discusses relevant background literature for the project. Section 3 contains the requirements specification for the application and Section 4 discusses the technical design of the application. Section 5 discusses the strategy for testing and evaluating the application. Section 6 contains a plan for the project and Section 7 concludes the document.

2 Literature Review

This literature review contains the relevant information for understanding the complexity of this project and details of how the problem will be tackled. While this review contains explanations of all key concepts related to, basic knowledge of the following topics is assumed: recursion, vector maths, and complex numbers.

This review is split into several sections, first the theory of fractals and their 3D counterparts will be discussed. Then the key concepts of ray marching, the chosen method of rendering fractals, will be discussed. And finally, a brief introduction GPU parallel programming will be given.

2.1 3D Fractals

As discussed in the introduction, a fractal is a recursively created never-ending pattern that is usually self-similar [1]. This concept defines a fractal geometry, which describes up the more non-uniform shapes found in nature, like clouds, mountains, and coastlines. There exist several ways of creating a fractal, from manually defining a simple repeating pattern to studying the convergence of complex equations. This section will discuss several common 3D fractals (and their 2D counterparts) and the methods used for creating them.

2.1.1 Sierpiński Tetrahedron

The Sierpiński tetrahedron, also known as the Sierpiński pyramid, is a 3D representation of the famous 2D Sierpiński tringle fractal, named after the Polish mathematician Wacław Sierpiński [7]. The Sierpiński triangle is one of the most simple and elegant fractals and has been a popular decorative pattern for centuries. This pattern is created by recursively each splitting each solid equilateral triangle

into four smaller equilateral triangles and removing the middle one. Theoretically, these steps are repeated forever, but in practice some maximum depth must be specified as computers only have finite memory.

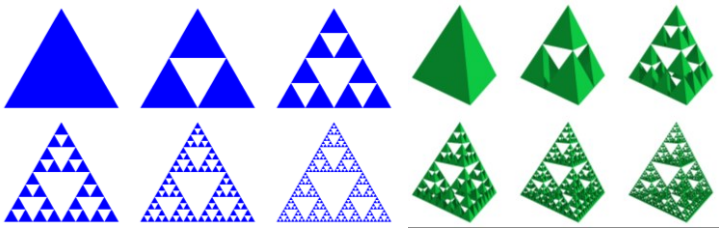


Figure 2.1.i Sierpiński triangle (left) [8] *and tetrahedron (right)* [8] *both of recursive depth 5*

As the recursive depth of the fractal increases, so does the number of objects (either triangles or tetrahedrons) in the scene. This is the limiting factor when rendering the Sierpiński tetrahedron as computer memory is finite and can only store limited number of objects. The total number of objects in the Sierpiński triangle increases by a factor of 3 each iteration and the tetrahedron by a factor of 4.

2.1.2 Menger Sponge

The Menger sponge, also known as the Menger cube or Sierpiński cube, is another 3D representation of one of Sierpiński's 2D fractals. This one is known as the Sierpiński carpet, which follows very similar recursive rules to the Sierpiński triangle but uses squares instead of triangles.

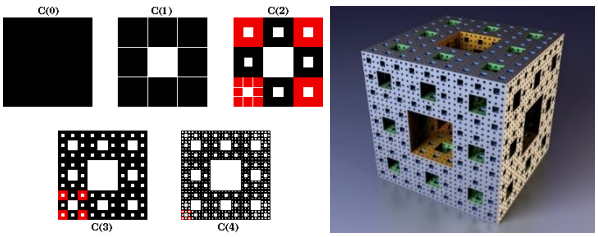


Figure 2.1.ii Sierpiński carpet (left) [9] *and Manger sponge (right)* [10] *both of recursive depth 4*

The number of objects required to create these fractals at various recursive depths increases similarly to the Sierpiński triangle and tetrahedron, but at a different rate. The Sierpiński carpet increases by a factor of 8 each iteration and the Menger sponge by a factor of 20. A Menger cube at recursive depth n is made up of 20^n smaller cubes, each with a side length of $(\frac{1}{3})^n$ [11].

In addition to the Sierpiński tetrahedron and Menger sponge, Sierpiński variations of other convex polyhedrons exist. A convex polyhedron is the 3D equivalent of a 2D regular polygon. While there are an infinite number of regular polygons, there are only fixe possible convex polyhedrons [12]. These are the tetrahedron, cube, octahedron, dodecahedron, and icosahedron. These polyhedrons are known as the platonic solids, and their fractal counterparts are called the platonic solid fractals.

2.1.3 Mandel Bulb

While the platonic solid fractals are created by making many copies of a primitive shape, another method for creating fractals is to plot the convergence of values for certain mathematical equations. This can instead generate much more "natural" looking fractal patterns. One of the first fractals of this type to be discovered was the Mandelbrot set, named after Beniot Mandelbrot, the inventor of the concept of fractal geometry. The Mandelbrot fractal is defined as the set of complex numbers c for which the equation $f_c(z) = z^2 + c$ does not diverge to infinity, when iterated from $z = 0$ [13].

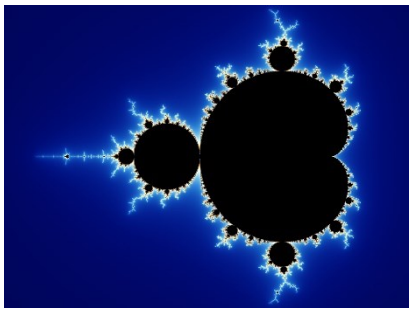


Figure 2.1.iii Mandelbrot set [13]

The Mandel bulb is a commonly used 3D representation of the 2D Mandelbrot fractal, created by Daniel White and Paul Nylander. For many years, it was thought that a true 3D representation of the Mandelbrot fractal did not exist, since there is no 3D representation of the 2D space of complex numbers, on which the Mandelbrot fractal is built upon [14].

White and Nylander considered some of the geometrical properties of the complex numbers. The multiplication of two complex numbers is a kind of rotation, and the addition is a kind of transformation. White and Nylander experimented with ways of preserving these characteristics when converting from 2D to 3D, and their solution was to change the squaring part of the formula to instead

use a higher power, a practice sometimes used with the 2D Mandelbrot fractal to produce snowflake type results [15]. This change leads us to the equation $f(z) = z^n + c$. The Mandel bulb is defined as the set of points c in $\mathbb{R}^3$ such that sequence $f^n(0)$ does not diverge.

White and Nylander's formula for the nth power of a point [15], [16] is given as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}^n = r^n \begin{bmatrix} \sin(n\theta)\cos(n\varphi) \\ \sin(n\theta)\sin(n\varphi) \\ \cos(n\theta) \end{bmatrix}$$

$$where\ r = \sqrt{x^2 + y^2 + z^2}, \theta = atan2(\sqrt{x^2 + y^2}, z), \varphi = atan2(y, x)$$

And the addition of two points is given as:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{bmatrix}$$

Using both formulas, the equation $f(z) = z^n + c$ can now easily be solved when $z = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

2.1.4 Julia Bulb

TODO left out for now as might not have space

The Julia set is a set of fractals named after the French mathematician Gaston Julia [17]. Fractals in the Julia set come from the convergence of the system given by the quadratic function $f(z) = z^2 + c$, where different values of c produce different fractals.

Similar to the Mandelbrot fractal, 2D Julia sets use the complex number plane as the domain of the quadratic function f. Unlike the Mandel bulb, the Julia set can be extended into 4D using quaternions as the domain of f [18], and then 3D slices can be taken of the quaternions to view the fractal in 3D space.

Some recent work, such as Quilez [19], uses pixel shaders to render a 3D Julia set.

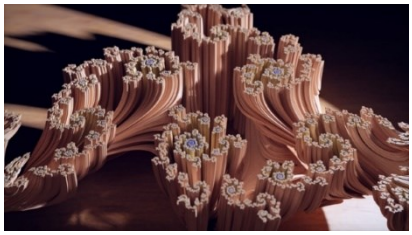


Figure 2.1.iv Ray marched Julia set, cut in half to expose the fractal interior [19]

The recent work da Silva et al in 2021 [14] took this a step further, using Nvidia DirectX Raytracing (DXR) shaders to render the visualisations of the 3D Julia set and Mandel bulb fractal.

TODO talk about pros/cons of existing work

Colour – orbit trap, as surface point transforms, look at how far away it gets from origin as it iterates through the transformation, min, max, sum, x,y,z etc

2.1.5 3D Fractals Summary

TODO

2.2 Ray Tracing

In computer graphics, ray tracing is a method of rendering an image of a 3D scene, often with photorealistic detail. This is done by tracing the paths of light and simulating their effects on geometry by taking into consideration reflections, light refraction, and reflections of reflections [20].

When rendering an image using ray tracing, for each pixel in the camera, a ray (simply a line in 3D space) is extended or traced forwards from the camera position until it intersects with the surface of an object. From there, the ray can be absorbed or reflected by the surface and more rays can be sent out recursively, which can be used to take into consideration light absorption, reflection, refraction, and fluorescence.

Ray tracing is ideal for photorealistic graphics, as it takes into consideration many of the properties of light, but because of this, it is computationally expensive. Often, ray tracers do not render images in real-time, and they can take hours to render a couple seconds of video. To make a ray tracer capable of rendering in real-time, many approximations must be made, or hybrid approaches used.

One of the limitations of ray tracing is that an accurate ray-surface intersection function must exist for every object in the scene. This is well suited for Euclidian geometry, such as primitive shapes or meshes, as points of intersection can be calculated relatively easily on these shapes. However, these functions do not exist for fractal geometry [21]. Instead, a slightly different approach must be used.

2.3 Ray Marching

Ray marching is a variation of ray tracing, which only differs in the method of detecting intersections between the ray and objects. Instead of a using a ray-surface intersection function which returns the position of intersection, ray marching uses a distance estimation (DE) function, which simply returns the distance from any given position in the scene, to the closest object. Instead of shooting the ray in one go, ray marching uses an iterative approach, where the current position is moved/marched along the ray in small increments until it lands on the surface of an object. For each point on the ray that is sampled, the DE function is called and marched forward by that distance, and the process is repeated until the ray lands on the surface of an object. If the distance function returns 0 at any point (or is close enough to an arbitrary epsilon value), then then the ray has collided with the surface of the geometry.

The diagram below shows a ray being marched from position p0 in the direction to the right. The distance estimation for each point is marked using the circle centred on that point.

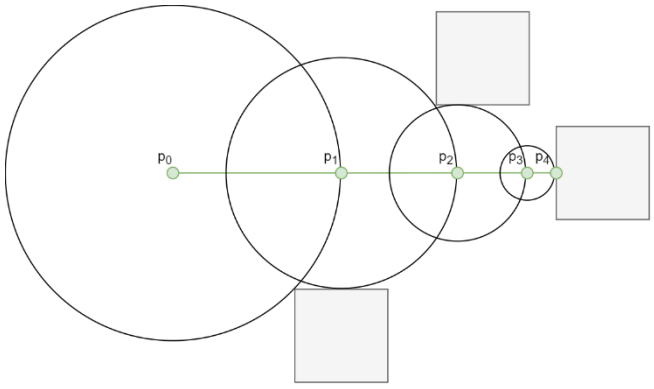


Figure 2.3.i Ray marching diagram

Technically, the DE does not have to return the exact distance to an object, as for some objects this may not be computable, but it must never be larger than the actual value. However, if the value is too small, then the ray marching algorithm becomes inefficient, so a fine balance must be found between accuracy and efficiency.

2.3.1 Benefits of Ray Marching

Ray marching may sound more computationally complex than ray tracing since it must complete multiple iterations of an algorithm do what ray tracing does in a single ray-surface intersection function, however, it does provide several benefits. As mentioned before, ray marching does not require a surface intersection function like ray tracing does, so it can be used to render geometry for which these functions do not exist. This property will be used to render 3D fractal geometry in this project. While many effects such as reflections, hard shadows and depth of field can be implemented almost identically to how they are in ray tracing, there are several optical effects that the ray marching algorithm can compute very cheaply.

Ambient occlusion is a technique used to determine how exposed each point in a scene is to ambient lighting [22]. This means that the more complex the surface of the geometry is (with creases, holes etc), the less ways ambient light can get into it those places and so the darker they should be. With ray marching, the surface complexity of geometry is usually proportional to the number of steps taken by the algorithm [23]. This property can be used to implement ambient occlusion and comes with no extra computational cost at all.

Soft shadows can also be implemented very cheaply, by keeping track of the minimum angle from the distance estimator to the point of intersection, when marching from the point of intersection towards the light source [24]. This second round of marching must be done anyway if any type of lighting is to be taken into consideration, so minimum check required for soft shadows is practically free.

A glow can also be applied to geometry very cheaply, by keeping track of the minimum distance to the geometry for each ray. Then, if the ray never actually intersected the geometry, a glow can be applied using the minimum distance the ray was from the object, a strength value and colour specified [23].

2.3.2 Signed Distance Functions

A signed distance function (SDF) for a geometry, is a function which given any position in 3D space, will return the distance to the surface of that geometry. If the distance contains a positive sign if the position is outside of the object, and a negative sign if the position is inside of the object. If a distance function returns 0 for any position, then the position must be exactly on the surface of an object. Every single geometry in a scene must have its own SDF. The scenes distance estimation (DE) function will loop through all of the SDF values for the geometry in the scene and will return the minimum.

The sign returned by the SDF is useful as it allows the ray marcher to determine if a camera ray is inside of a geometry or not, and from there it can use that information to render the objects differently. We may want to render geometry either solid or hollow, or potentially add transparency.

Signed distance functions are already known for most primitive 3D shapes [25], such as spheres, boxes, and planes. Some of these functions are trivial, such as the SDF for a sphere with radius R, positioned on the origin.

$$sphereSDF(p) = |p| - R$$

$$where\ p = \begin{bmatrix} x \\ y \\ z \end{bmatrix},\ |p|\ is\ the\ magnitude\ of\ the\ vector\ p,\ R\ is\ the\ circle\ radius\ in\ world\ units$$

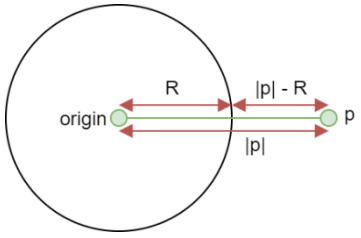


Figure 2.3.ii Sphere SDF diagram

2.3.3 Alterations & Combinations

Signed distance functions can be translated, rotated, and scaled. In addition, they can also be combined using the union, subtraction, and intersection operations. The union, intersection and subtraction of two SDFs can be taken as:

$$union(a, b) = \min(a, b),$$

$$subtraction(a, b) = \max(-a, b),$$

$$intersection(a, b) = \max(a, b),$$

$where\ a, b\ \in \mathbb{R}\ are\ the\ values\ returned\ from\ object\ a\ and\ b's\ SDF$

There also exist variations of formulas above which can be used to apply a smoothing value to the operation. Formulas have been included in the appendix section 9.1. The images below were rendered using an early version of the application.

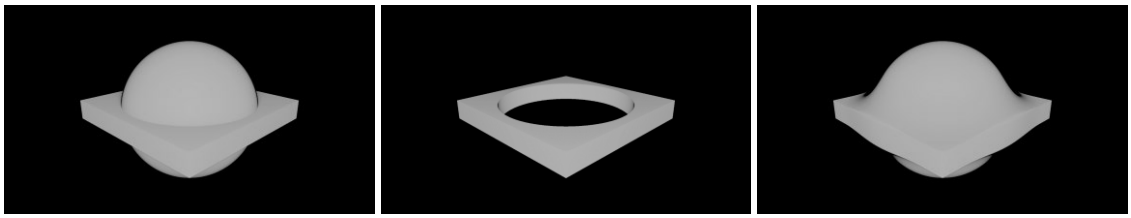


Figure 2.3.iii Ray marched sphere and box scene experiment union (left), subtraction (middle) and smooth union(right)

There are several additional alterations that can be applied to primitives once we have their signed distance function. A primitive can be elongated along any axis, its edges can be rounded, it can be extruded, and it can be "onioned" – a process of adding concentric layers to a shape. All these operations are relatively cheap. Signed distance functions can also be repeated, twisted, bent, and surfaces displaced using an equation e.g., a noise function or sin wave, though these alterations are more expensive. All these techniques mentioned will be essential when creating more complex geometry.

2.3.4 Surface Normal

The surface normal of a position on the surface of a geometry is a normalised vector that is perpendicular to the that surface [26]. This information is is essential for most lighting calculations.

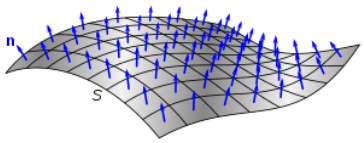


Figure 2.3.iv Surface normal vectors for a curved surface [26]

The surface normal of any point on the geometry of a scene can be determined by probing the SDF function on each axis, using an arbitrary epsilon value.

$$normal = normalise\left(\begin{bmatrix} DE(p+\begin{bmatrix}e\\0\\0\end{bmatrix}) - DE(p-\begin{bmatrix}e\\0\\0\end{bmatrix}) \\ DE(p+\begin{bmatrix}0\\e\\0\end{bmatrix}) - DE(p-\begin{bmatrix}0\\e\\0\end{bmatrix}) \\ DE(p+\begin{bmatrix}0\\0\\e\end{bmatrix}) - DE(p-\begin{bmatrix}0\\0\\e\end{bmatrix}) \end{bmatrix}\right)$$

$$where\ p\ is\ a\ vector\ in\ the\ form\ \begin{bmatrix}x\\y\\z\end{bmatrix}, e\ is\ an\ arbitrary\ epsilon\ value,$$

$$DE\ is\ the\ scene\ distance\ estimation\ function$$

The value of a normal can be converted to a colour by mapping the x, y, and z values to r, g, and b.

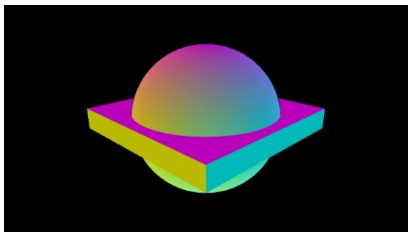


Figure 2.3.v Surface normal of ray marched sphere and box scene experiment

2.3.5 Ray Marching Summary

TODO

2.4 Review of Existing Applications

A review of some existing 3D fractal renderers was completed, and the key features of each application were recorded. A summary containing points to take away from the review is included at the end.

2.4.1 Fragmentarium [27], [28]

Key notes:

- Features an editor within the application, which allows scenes to be created and previewed in real time
- Contains many example scenes
- Provides very detailed parameters to edit the scene
- Rendering quality of the output image is acceptable (see appendix 9.2)

2.4.2 Mandelbulb3D [29], [30]

Key notes:

- User interface is clunky, and features are hard to find
- Only contains a couple example scenes
- The performance of the real-time preview feature is sub optimal, making it hard to navigate the scene
- The output image is very detailed, is coloured beautifully and contains many advanced optical effects (see appendix 9.3)

2.4.3 Existing Applications Summary

After reviewing existing 3D fractal rendering applications, there are several key points that should be taken away.

- The application must contain lots of example scenes, and they must be easy to find
- The real time preview of the scene must have good performance and be easy to control
- It would be nice to be able to output an image of the scene

- It must be relatively easy to create a new scene and preview it

Additionally, Fragmentarium and Mandelbulb3D are both built using OpenGL with GLSL shaders, as this makes creating scenes simple as it uses an already existing language and format. However, using OpenGL is unnecessary as not many of the features it provides are being used. The application could instead be implemented using a different approach, which should provide better performance.

2.5 GPU Computing

GPU computing is when a GPU (Graphics Processing Unit) is used in combination with a CPU (Central Processing Unit) to execute some code [31]. The correct use of GPU computing improves the overall performance of the program by offloading some computation from the CPU to GPU. CPUs and GPUs have been designed to achieve different goals. A CPU is designed for executing a sequence of operations, called a thread. A CPU can execute a few tens of threads in parallel and is designed to execute them as fast as possible. On the other hand, a GPU is designed to execute a few thousand of these threads in parallel but does it slower than the CPU would. However, a GPU achieves a higher overall throughput than a CPU does. This difference in architecture between CPUs and GPUs means that CPUs are better suited for computations requiring data caching and flow control, while GPUs are better suited for highly parallel computations.

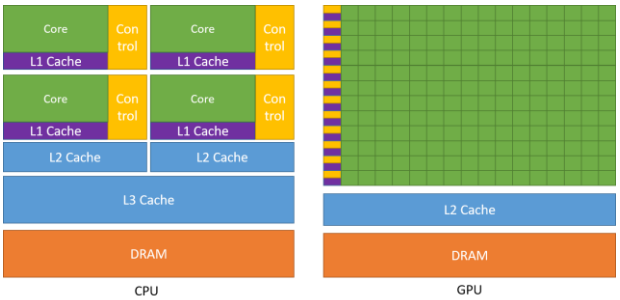


Figure 2.5.i CPU and GPU architecture [32]

Implementing a graphics renderer using GPU computing is the perfect choice, as this is a massively parallel task requiring a computation to be executed for every single pixel on the screen. Making use of this architecture is the only feasible way to implement a real-time renderer. When implementing a real-time fractal renderer, the extra features that libraries like OpenGL [33] provide, such as compute

shaders, aren't needed. Instead, it makes more sense to make use of general-purpose GPU computing interfaces, as the overhead of this software will be significantly less, giving us better performance.

2.5.1 CUDA vs OpenCL

Currently, there are two suitable general-purpose GPU computing interfaces, CUDA (Compute Unified Device Architecture) and OpenCL (Open Computing Language). The table below breaks down the main advantages and disadvantages [32], [34], [35] of the two interfaces.

Table 2.5.i CUDA and OpenCL comparison

Comparison	CUDA	OpenCL
Portability	Only runs on NVIDIA hardware	Runs on pretty much all hardware - NVIDIA, AMD, Intel, Apple, Radeon, etc
Open Source	Propriatery framework of NVIDIA	Open source standard
Technicalities		Compiles kernel code at runtime, which could take a lot of time, though this also means that the compiled code is more optimised for the device currently running it
Interface Languages	C, C++, Fortran, Java Python Wrappers, DirectCompute, Directives (e.g. OpenACC)	C, C++, Python wrapper
Kernel Language	KUDA C++	OpenCL C (C99) and OpenCL C++ (C++17)
Libraries	Extensive and powerful libraries	Good libraries, but not as extensive as CUDA
Performance	No clear advantage	No clear advantage
OS Support	Runs on Windows, Linux and MacOS	Runs on Windows, Linux and MacOS

The main difference between CUDA and OpenCL, is CUDA can only run on NVIDIA branded GPUs. While this technically does allow CUDA to make full use of the hardware, it makes the application deployment far less portable as the code will not run on other types of GPUs. OpenCL is, therefore, the superior choice when aiming to write portable code to be executed on the GPU.

2.6 SUMMARY

3 Requirements Analysis

3.1 Use Cases

TODO

Relate to real world - model objects in nature? Cell structure, coastlines, etc

3.2 Requirements Specification

The tables below display the key functionality to be implemented in order to achieve the projects aims and objectives, specified in section 1.2. Over the course of the project, this table will be updated to contain the specific test IDs for each associated requirement, and its test/fail status.

Requirements were prioritised using the following strategy:

- MUST – a requirement that is of the highest priority to the project
- SHOULD – a requirement that is not essential, but it would be good if the project had it
- COULD – a requirement that is optional
- WON'T – a requirement that would be implemented if the project could run for more time

Table 3.2.i Functional requirement specification

ID	Name	Description	Aim/ Objective	Priority	Testing strategy
	Game Loop	The application must make use of a game loop so that it updates in real time	Objective 4	MUST	Unit Test
F-1	Real-time	The performance of the application must be as good as possible	Aim 1	MUST	Benchmark
F-2	Scene requirements	A scene must contain: • Geometry • Lights • Camera	-	MUST	Unit test
F-3	Example scenes	The application must contain multiple example scenes, some of which could include: • Julia set fractal • Mandel bulb fractal • Sierpinski tetrahedron fractal • Menger sponge fractal	Objective 8	MUST	Unit test
	Performance benchmark	The application must contain a performance benchmark scene, as specified in section 5.2	Objective 7	MUST	
F-4	Mandatory optical effects	The application must support the following optical effects: • Ambient occlusion • Hard and soft shadows • Glow	Objective 6	MUST	Unit test

F-5	Optional optical effects	The application could support the following optical effects: • Reflections • Depth of field • Transparency	Objective 6	COULD	Unit test
F-6	Controllable camera	The user must be able to control the scene camera using a keyboard and mouse to move it around the scene	Objective 5	MUST	
F-7	Fixed camera paths	The application camera could support fixed camera paths	Objective 5	COULD	
	Image output	It could be possible for the application to output a high-resolution image of the current view		COULD	
	Video output	It could be possible for the application to output a video sequence for the current scene		COULD	

Table 3.2.ii Non-functional requirement specification

ID	Name	Description	Aim/ Objective	Priority	Testing strategy
NF-1	Executable	The application must run from a compiled executable		MUST	
NF-2	Display resolutions	The application must support the following common display resolutions: 1366x768, 1920x1080, 2560x1440 and 3840x2160		MUST	Unit test
	GPU Parallel Computing	The application must run in parallel on the GPU	Objective 3	MUST	Unit test

4 SOFTWARE DESIGN

4.1 TECHNOLOGIES

The application will be developed using the following technologies:

Table 4.1.i Application technologies

Technology	Description	Justification
OpenCL	Programming language which allows code to be run in parallel on the GPU	• GPU parallel computing gives a massive performance boost when executing the same piece of code simultaneously for many different input values • GPU parallelism is far better suited for this task than CPU parallelism as the same piece of code must be executed for every pixel on the screen

		• OpenCL was chosen as it has good documentation and examples, contains C and C++ programming interfaces, and allows deployments to different platforms
C++	Common system programming language	• C++ is a low-level language with good performance • C++ was chosen over C to allow an object-oriented style of programming
SDL2	Cross platform C++ library for manipulating windows and reading user input	• Cross platform libraries provide an abstraction layer over platform specific libraries, which allows the program implementation to remain decoupled from the deployment platform • SDL2 was chosen as it provides both window display interaction and user input event polling, and has good documentation and examples

Development of the application and documentation will be assisted the following technologies:

Table 4.1.ii Development technologies

Technology	Description	Justification
Visual Studio 2019	Code editor	• Contains powerful development tools such as refactoring options, code snippets, documentation preview etc
GitHub	Version control software	• Version control is essential for any large coding project • GitHub is being used to store all project materials, including documents, papers, and the coding project
Microsoft Word	Word processing software	• Powerful and easy to use word processing software • Documents will be edited locally and backed up to GitHub
Mendeley	Reference manager	• This reference manager has a web interface which is very convenient when researching • It also has a Microsoft Word add in which manages all referenced sources automatically
Microsoft Teams	Video communication software	• For communication with the project supervisor and second readers

4.2 Class Structure

The application will be structured using several key classes:

Table 4.2.i Class responsibilities

Class name	Responsibilities
Application	Contains the run method, the main application loop which drives the application This class contains instances of Display, Renderer and Controller
Display	Setting pixels in the display window and controlling any GUI elements
Renderer	Calculating the colour for each pixel of the display window
Controller	Reading keyboard and mouse input from the user

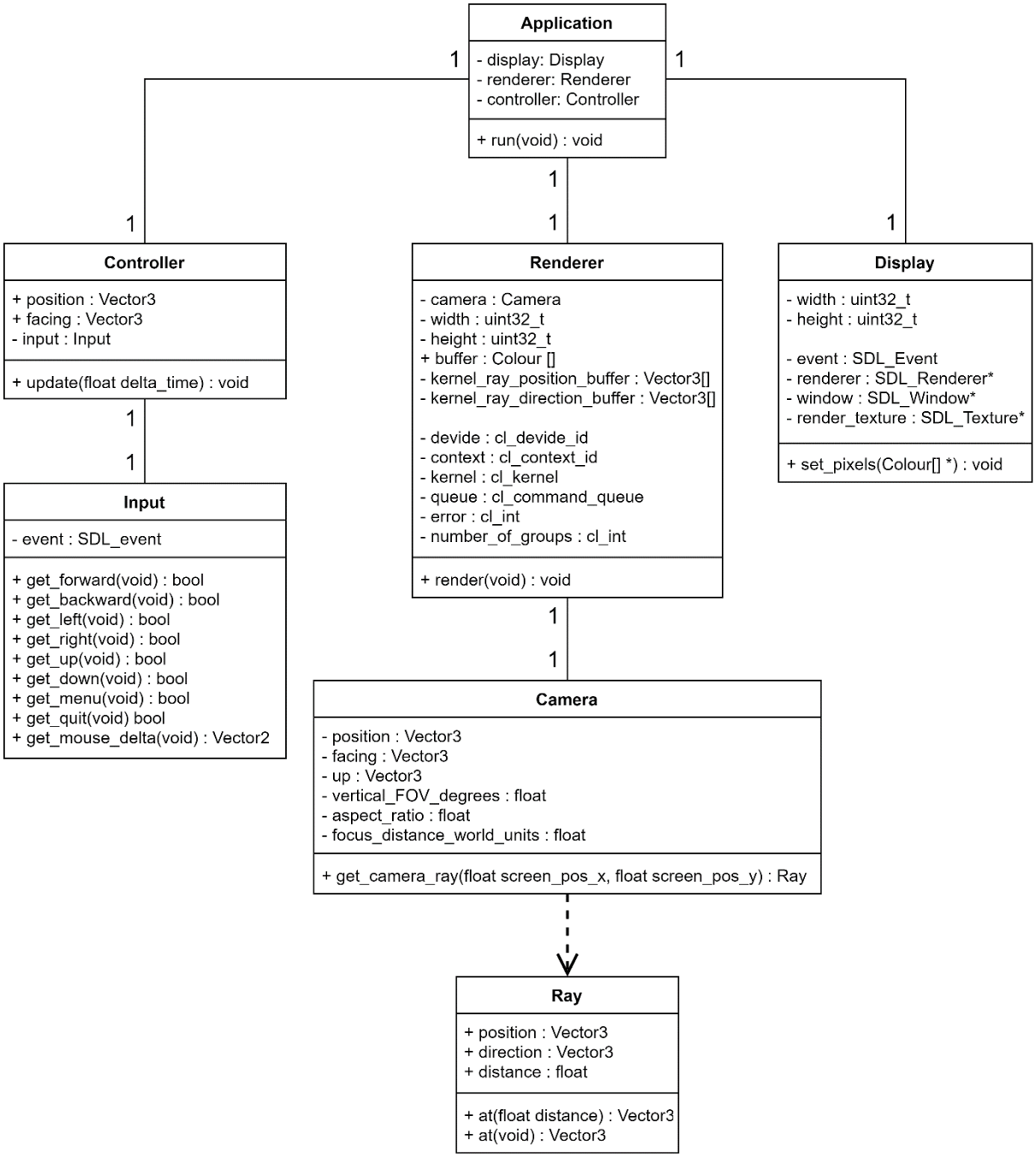


Figure 4.2.i Application class diagram

The Display and Controller classes are basic and only provide an interface to some SDL2_Event, SDL2_Renderer, SDL2_Window and SDL2_Texture instances. The Renderer class, however, is much more complex and requires discussion.

The Renderer class provides an interface to get the current pixels to be displayed on the screen, which are calculated using an OpenCL kernel. The OpenCL kernel is a piece of code written in one of the

OpenCL kernel languages, and is the code that is executed in parallel on the GPU. Most of the ray marching code should be written in this kernel language to give the best performance to the application. Each scene will be defined within its own kernel file and will be loaded into the application at runtime. However, this makes it hard to reuse code between kernel files as the implementations of several methods, and the values of several constants will differ between scenes. The tables below show the main methods and constants used in the kernel, and their reusability status between scenes.

Table 4.2.ii Kernel method reusability matrix

Method name	Purpose	Reusable across scenes?
render()	Calculates the colour for all pixels in the display and puts the values into a buffer	YES
calculatePixelColour(Ray)	Calculates the colour for the camera pixel with the specified ray direction	YES
DE(Vector3)	Calculates the distance to the nearest geometry surface in the current scene	NO
calculateNormal(Vector3)	Calculates the surface normal vector of the geometry for the position specified	YES

Table 4.2.iii Kernel constant reusability matrix

Constant name	Purpose	Reusable across scenes?
MAXIMUM_MARCH_STEPS	Maximum number of iterations the ray marching algorithm can make	NO
MAXIMUM_MARCH_DISTANCE	Maximum distance the ray can be marched in the scene	NO
SURFACE_INTERSECTION_EPSILON	A very small value used to determine when the DE has converged to zero	YES
SURFACE_NORMAL_EPSILON	Arbitrary distance to probe the DE function when calculating the surface normal	YES

A solution to reducing code duplication between the kernel files is to use the new OpenCL C++ kernel language, which supports most C++17 features. Method and constant overloading will be used within the kernel file for each scene to override the implementation of the distance estimation (DE) function and MAXIMUM_MARCH_STEPS and MAXIMUM_MARCH_DISTANCE constants defined in a main kernel

file. This main kernel file will contain the implementation of all other methods, such as the render and calculatePixelColour methods, and will contain an empty DE method for the other kernels to overload.

The OpenCL C++ kernel language is a new addition to OpenCL, released in March 2021, and as such there are few examples of C++ kernels, though the official documentation [36] is good.

4.3 PSEUDO CODE

MAYBE?

5 EVALUATION STRATEGY

A goal-based evaluation strategy will be used to determine if the project aims and objectives have been achieved. Unit tests and a benchmark scene will be used to determine how many of the requirements specified in section 3.2 have been fully implemented. For an objective to be considered achieved, all requirements related to that objective must have been implemented. For an aim to be considered achieved, all objectives related to that aim must have been achieved.

5.1 UNIT TESTING

Requirements (and hence objectives and aims) will be tested using several strategies. Each requirement is assigned a testing strategy, which defines how the implemented functionality for that requirement will be evaluated for correctness. A performance benchmark of the final application will be completed. See the full breakdown of this scene in section 5.2. In addition, unit tests will used to test for functionality and code correctness and will be implemented in parallel to the application.

5.2 PERFORMANCE BENCHMARK SCENE

A performance benchmark of the application must be implemented in order to evaluate if the real-time aspect of aim 1 has been achieved. This section will specify the results that the benchmark scene must output, and calculations that will be completed with the data.

Table 5.2.i Results recorded from benchmark

Description	Units
Window display resolution	-

Duration of the benchmark scene	Seconds
Total number of frames rendered	-
Minimum frame time	Milliseconds
Maximum frame time	Milliseconds

Table 5.2.ii Results calculated from benchmark

Description	Units	Calculation
Average frame time	Milliseconds	$\frac{duration\ of\ the\ benchmark\ scene}{total\ number\ of\ frames\ rendered} \times 1000$
Average frames per second	Frames per second	$\frac{1000}{average\ frame\ time}$
Minimum frames per second	Frames per second	$\frac{1000}{minimum\ frame\ time}$
Maximum frames per second	Frames per second	$\frac{1000}{maximum\ frame\ time}$

In addition, the PC specs of the computer running the benchmark must be recorded. This information can be exported to a file by running the "System information" program on windows and selecting the File>Export option. The exported file contains a list of the computer specifications and relevant driver information. The important information that will be extracted from the file is included in the table below.

Table 5.2.iii Relevant PC specification values

Description	Units	Line number in exported file
Operating system name	-	6
Processor name	-	15
Processor base clock speed	Megahertz	15
Processor number of cores	-	15
Processor number of logical cores	-	15
Total physical memory	Gigabytes	34
Total virtual memory	Gigabytes	36
Graphics card name	-	Varies, located under [Display] heading
Graphics card dedicated memory	Gigabytes	N/A
Graphics card shared memory	Gigabytes	N/A

Unfortunately, the dedicated and shared memory of the graphics card is not included in the file and must be recorded manually from viewing the task manager.

The benchmark scene has yet to be fully defined, but it must be non-trivial to render. This means it should contain multiple geometries (both fractal and primitive) and multiple lights while also making

use of advanced rendering features like ambient occlusion, soft shadows, and reflections. It is important that the scene is consistent as possible between separate runs, therefore, the camera should be either stationary or move through the scene on a fixed path to view the geometries. The benchmark scene should run for a fixed duration so that it takes the same amount of time to run on all machines.

The benchmark should be run multiple times and averages taken of the results. This is to reduce impact of any background tasks that may be running on the machine.

5.3 Unique Characteristics

- Users will have different experiences

Details of the evaluation and analysis to be conducted.

stats, analysis, what and I going to do, how to compare results

emphasise visualising

pros and cons

5.4 Current Achievements

Table 5.4.i Aim 1 progress

Objective	Testing Strategy	Testing IDs	Status
1			Achieved
3			
4			
5			

6			
7			

Table 5.4.ii Aim 2 progress

Objective	Testing Strategy	Testing IDs	Status
2			Achieved
8			

6 Project Plan

6.1 Project Management

The project will be developed using an Agile [37] approach, which uses small sprints of work to complete specific and defined tasks. This approach allows teams to respond to change quickly as requirements and plans are updated regularly. While this an individual project and the Agile approach is normally used in teams, the continuous improvement gained through this approach will be incredibly valuable for the project.

6.2 Design Methodology

The application will be developed following the main object-oriented principles [38]. These include encapsulation, abstraction, inheritance, and polymorphism. An object-oriented style of programming has been chosen to promote code reusability within the application in the hopes that this will allow the kernel code to be as simple and easy to use as possible.

6.3 Professional, Legal, Ethical & Social Issues

TODO

open source standards

BCS

License

6.4 Risk Analysis

Table 6.4.i Risk analysis matrix

ID	Description	Probability	Severity	Mitigation plan	Rating
R-1	Loss of work	LOW	HIGH	All work will be backed up regularly using online version control	**MEDIUM**
R-2	Change in requirements	LOW	MEDIUM	A thorough requirements specification has been prepared to reduce the probability of this happening In addition, an Agile development approach will be used, which by design minimises the effects of changed requirements and plans	**MEDIUM**
R-3	Change of deadlines	LOW	HIGH	Communication from the course leaders will be regularly monitored This risk is very unlikely to happen as assurances have been made that the course deadlines will not change	**LOW**
R-4	Delays due to learning new software	HIGH	MEDIUM	Time was assigned during the planning stage to experiment with new software like OpenCL and GLSL shaders Additionally, some free time has been allocated at the end of the project timetable to allow for delays	**MEDIUM**
R-5	Delays due to illness	LOW	MEDIUM	Free time has been allocated at the end of the timetable to allow for delays	**LOW**
R-6	Delays due to bugs	MEDIUM	MEDIUM	Free time has been allocated at the end of the timetable to allow for delays	**MEDIUM**

6.5 PROJECT TIMELINE

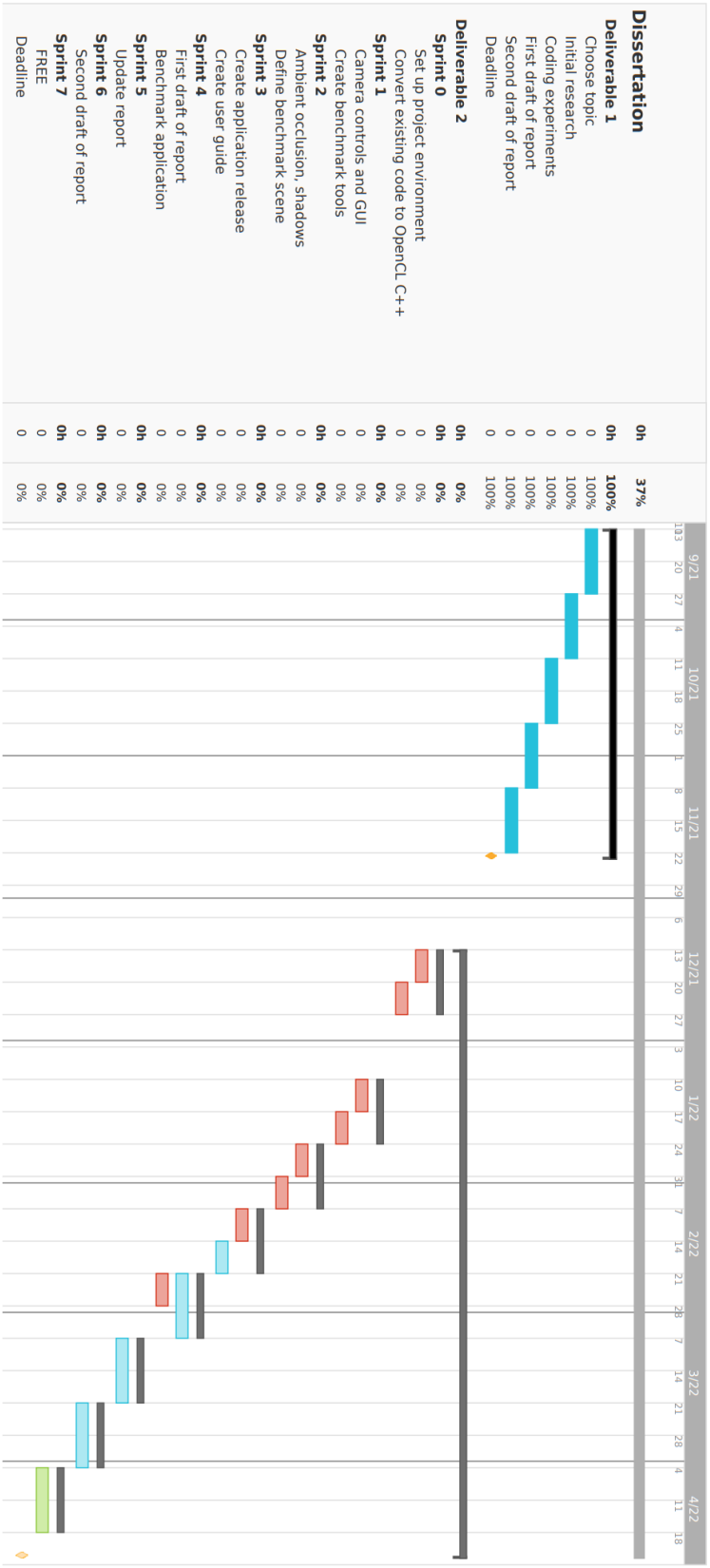


Figure 6.5.i Project timeline Gantt chart

The project timeline for the second deliverable has been split into 8 sprints, each 2 weeks long. Project objectives one and two have already been completed, leaving objectives three to eight.

Sprint 0 will be completed during the Christmas holiday, to set up the project environment and refactor the existing experimentation code into something usable. Once the semester starts, sprint

7 CONCLUSION

TODO

8 REFERENCES

[1] "Top 5 applications of fractals | Mathematics | University of Waterloo." https://uwaterloo.ca/math/news/top-5-applications-fractals (accessed Nov. 10, 2021).

[2] "How Mandelbrot's fractals changed the world - BBC News." https://www.bbc.co.uk/news/magazine-11564766 (accessed Nov. 10, 2021).

[3] "Fractals in nature and applications." https://kluge.in-chemnitz.de/documents/fractal/node2.html (accessed Nov. 10, 2021).

[4] "Fractal Foundation Online Course - Chapter 12 - FRACTAL APPLICATION." http://fractalfoundation.org/OFC/OFC-12-2.html (accessed Nov. 10, 2021).

[5] "Fractal-generating software - Wikipedia." https://en.wikipedia.org/wiki/Fractal-generating_software (accessed Nov. 14, 2021).

[6] "Writing Dissertations: Aims and objectives." https://learn.solent.ac.uk/mod/book/view.php?id=116233&chapterid=15294 (accessed Nov. 11, 2021).

[7] "Wacław Sierpiński - Wikipedia." https://en.wikipedia.org/wiki/Wac%C5%82aw_Sierpi%C5%84ski (accessed Nov. 11, 2021).

[8] H. Segerman, "Fractals and how to make a Sierpinski Tetrahedron", Accessed: Nov. 11, 2021. [Online]. Available: http://www.segerman.org

[9] "Sierpinski Carpet." https://larryriddle.agnesscott.org/ifs/carpet/carpet.htm (accessed Nov. 11, 2021).

[10] "Menger Sponge | Visual Insight." https://blogs.ams.org/visualinsight/2014/03/01/menger-sponge/ (accessed Nov. 11, 2021).

[11] "Menger sponge - Wikipedia." https://en.wikipedia.org/wiki/Menger_sponge (accessed Nov. 11, 2021).

[12] "n-flake - Wikipedia." https://en.wikipedia.org/wiki/N-flake (accessed Nov. 16, 2021).

[13] "Mandelbrot set - Wikipedia." https://en.wikipedia.org/wiki/Mandelbrot_set (accessed Nov. 13, 2021).

[14] V. da Silva, T. Novello, H. Lopes, and L. Velho, "Real-time rendering of complex fractals," Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.01747

[15] "Mandelbulb: The Unravelling of the Real 3D Mandelbrot Fractal." https://www.skytopia.com/project/fractal/mandelbulb.html (accessed Nov. 12, 2021).

[16] R. Englund, S. Seipel, and A. Hast, "Rendering Methods for 3D Fractals," 2010.

[17] "Gaston Julia - Wikipedia." https://en.wikipedia.org/wiki/Gaston_Julia (accessed Nov. 12, 2021).

[18] A. Norton, "Julia sets in the quaternions," *Computers & Graphics*, vol. 13, no. 2, pp. 267–278, Jan. 1989, doi: 10.1016/0097-8493(89)90071-X.

[19] Inigo Quilez, “3D Julia sets.” https://www.iquilezles.org/www/articles/juliasets3d/juliasets3d.htm (accessed Nov. 04, 2021).

[20] J. Peddie, “Ray Tracing: A Tool for All,” 2019.

[21] J. C. Hart, D. J. Sandin, and L. H. Kauffman, “Ray Tracing Deterministic 3-D Fractals,” 1989.

[22] “Ambient occlusion - Wikipedia.” https://en.wikipedia.org/wiki/Ambient_occlusion (accessed Nov. 03, 2021).

[23] Mikael Hvidtfeldt Christensen, “Distance Estimated 3D Fractals,” 2011. http://blog.hvidtfeldts.net/index.php/2011/08/distance-estimated-3d-fractals-ii-lighting-and-coloring/ (accessed Nov. 04, 2021).

[24] Inigo Quilez, “Soft Shadows in Raymarched SDFs,” 2010. https://iquilezles.org/www/articles/rmshadows/rmshadows.htm (accessed Nov. 04, 2021).

[25] Inigo Quilez, “distance functions,” 2013. https://www.iquilezles.org/www/articles/distfunctions/distfunctions.htm (accessed Oct. 28, 2021).

[26] “Normal (geometry) - Wikipedia.” https://en.wikipedia.org/wiki/Normal_(geometry) (accessed Nov. 13, 2021).

[27] “Syntopia/Fragmentarium: Fragmentarium is a cross-platform application for exploring pixel based graphics on the GPU.” https://github.com/Syntopia/Fragmentarium (accessed Nov. 17, 2021).

[28] “3Dickulus/FragM: Derived from https://github.com/Syntopia/Fragmentarium/.” https://github.com/3Dickulus/FragM (accessed Nov. 17, 2021).

[29] “Mandelbulb 3D (MB3D) fractal generator / rendering software | Mandelbulb.com.” https://www.mandelbulb.com/2014/mandelbulb-3d-mb3d-fractal-rendering-software/ (accessed Nov. 17, 2021).

[30] “thargor6/mb3d: Mandelbulb3D.” https://github.com/thargor6/mb3d#Coloring (accessed Nov. 17, 2021).

[31] “What Is GPU Computing?” https://www.boston.co.uk/info/nvidia-kepler/what-is-gpu-computing.aspx (accessed Nov. 16, 2021).

[32] “Programming Guide :: CUDA Toolkit Documentation.” https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html (accessed Nov. 17, 2021).

[33] “OpenGL Overview - The Khronos Group Inc.” https://www.khronos.org/opengl/ (accessed Nov. 17, 2021).

[34] “CUDA vs OpenCL: Which to Use for GPU Programming - Incredibuild.” https://www.incredibuild.com/blog/cuda-vs-opencl-which-to-use-for-gpu-programming (accessed Nov. 16, 2021).

[35] “OpenCL Overview - The Khronos Group Inc.” https://www.khronos.org/opencl/ (accessed Nov. 17, 2021).

[36] Khronos®, “The C++ for OpenCL 1.0 Programming Language Documentation,” 2021. https://www.khronos.org/opencl/assets/CXX_for_OpenCL.html#_the_c_for_opencl_programming_language (accessed Nov. 04, 2021).

[37] “What is Agile? | Atlassian.” https://www.atlassian.com/agile (accessed Nov. 13, 2021).

[38] “What are four basic principles of Object Oriented Programming? | by Munish Chandel | Medium.” https://medium.com/@cancerian0684/what-are-four-basic-principles-of-object-oriented-programming-645af8b43727 (accessed Nov. 13, 2021).

[39] Inigo Quilez, “Mandelbulb,” 2009. https://www.iquilezles.org/www/articles/mandelbulb/mandelbulb.htm (accessed Nov. 04, 2021).

9 APPENDICES

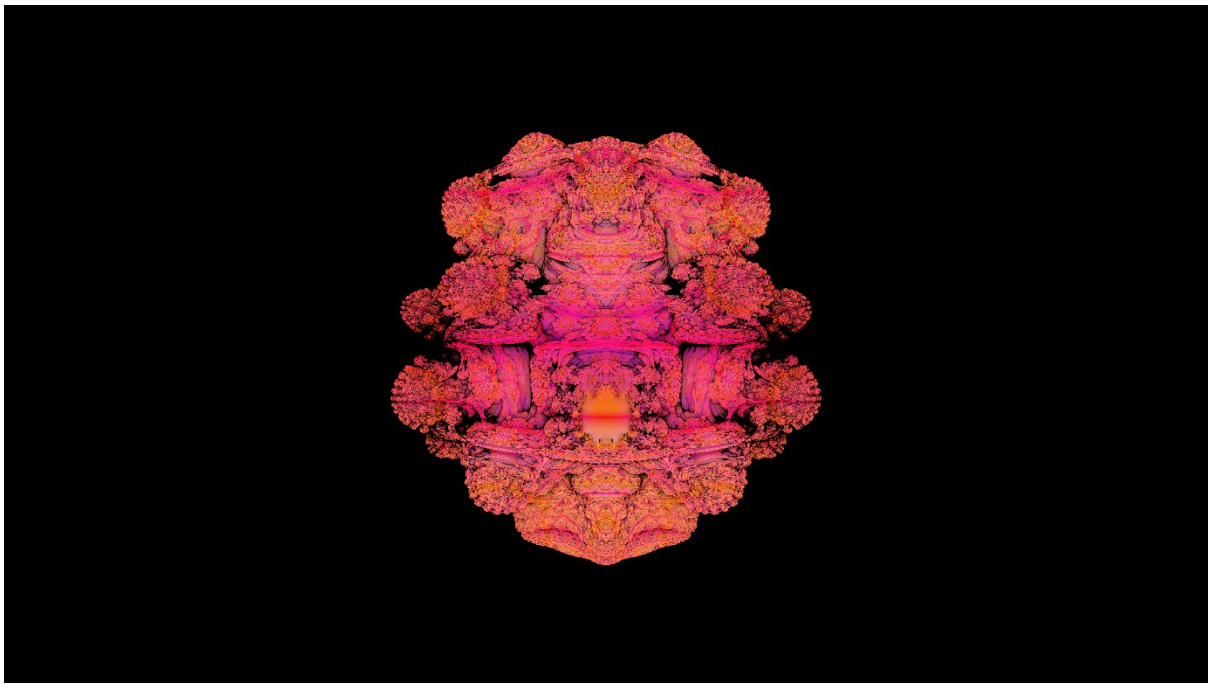


Figure 6.5.i Render of the Mandel bulb fractal, created using fractal equation from [39]

9.1 SMOOTH SDF COMBINATIONS

$$smoothUnion(a, b, s) = \min(a, b) - h^2 \times \frac{0.25}{k}$$

$where\ a, b\ \in \mathbb{R}, s\ \in \mathbb{R}\ is\ the\ smoothing\ value, h = \max(s - abs(a - b), 0)$

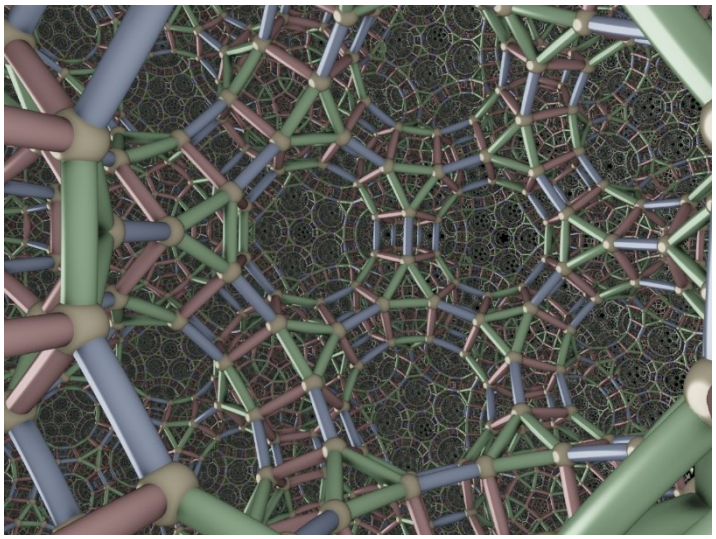
$$smoothSubtraction(a, b, s) = \max(-a, b) + h^2 \times \frac{0.25}{k}$$

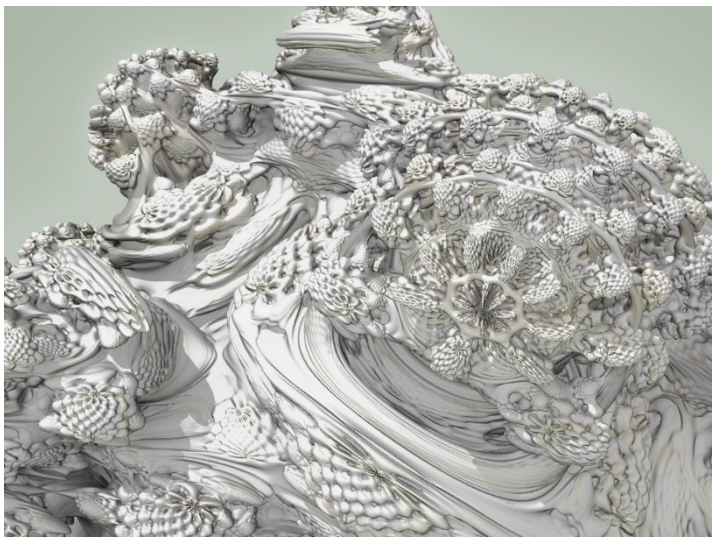
$where\ a, b\ \in \mathbb{R}, s\ \in \mathbb{R}\ is\ the\ smoothing\ value, h = \max(s - abs(-a - b), 0)$

$$smoothIntersection(a, b, s) = \max(a, b) + h^2 \times \frac{0.25}{k}$$

$where\ a, b\ \in \mathbb{R}, s\ \in \mathbb{R}\ is\ the\ smoothing\ value, h = \max(s - abs(a - b), 0)$

9.2 Renders from Fragmentarium





9.3 Renders from Mandelbulb3D