



Universidade do Minho

Escola de Engenharia

Departamento de Informática

Tiago Manuel da Silva Santos

Mobile Ray Tracing

October 2018



Universidade do Minho

Escola de Engenharia

Departamento de Informática

Tiago Manuel da Silva Santos

Mobile Ray Tracing

Master dissertation

Master Degree in Computer Science

Dissertation supervised by

Professor Doutor Luís Paulo Peixoto dos Santos

October 2018

ACKNOWLEDGEMENTS

Agradeço a todos que me apoiaram, direta ou indiretamente, neste longo percurso académico. Em especial, ao meu orientador, professor Doutor Luís Paulo Peixoto dos Santos que propôs este tema, deu-me sempre boas ideias e por todo o seu apoio. Também, ao professores dos perfis de Computação Gráfica e de Computação Paralela e Distribuída pela excelente formação que deram.

A todos os meus amigos que sempre me apoiaram neste longo percurso. Em especial, destaco cinco amigos: João Costa, André Pereira, Miguel Esteves, Eduardo Mendes e Miguel Rego que desde que os conheci me ajudaram muito a ultrapassar diversos obstáculos tanto a nível pessoal como na minha formação académica.

Ao grupo Embedded Systems Research Group (ESRG) por me terem acolhido durante uma bolsa de investigação, me tratarem como parte dessa grande família e pelo apoio que sempre me deram.

E por último mas não menos importante, à minha família, em especial aos meus irmãos Pedro Santos e Paulo Santos pelos muitos sacrifícios que fizeram para me apoiar e permitir chegar onde cheguei.

ABSTRACT

The technological advances and the massification of information technologies have allowed a huge and positive proliferation of the number of libraries and APIs. This large offer has made life easier for programmers in general, because they easily find a library, free or commercial, that helps them solve the daily challenges they have at hand.

One area of information technology where libraries are critical is in Computer Graphics, due to the wide range of rendering techniques it offers. One of these techniques is ray tracing. Ray tracing allows to simulate natural electromagnetic phenomena such as the path of light and mechanical phenomena such as the propagation of sound. Similarly, it also allows to simulate technologies developed by men, like Wi-Fi networks. These simulations can have a spectacular realism and accuracy, at the expense of a very high computational cost.

The constant evolution of technology allowed to leverage and massify new areas, such as mobile devices. Devices today are increasingly faster and replace and / or complement tasks that were previously performed only on computers or on dedicated hardware. However, the number of image rendering libraries available for mobile devices is still very scarce, and no ray tracing based image rendering library has been able to assert itself on these devices. This dissertation aims to explore the possibilities and limitations of using mobile devices to execute rendering algorithms that use ray tracing, such as progressive path tracing. Its main goal is to provide a rendering library for mobile devices based on ray tracing.

RESUMO

Os avanços tecnológicos e a massificação das tecnologias de informação permitiu uma enorme e positiva proliferação do número de bibliotecas e APIs. Esta maior oferta permitiu facilitar a vida dos programadores em geral, porque facilmente encontram uma biblioteca, gratuita ou comercial, que os ajudam a resolver os desafios diários que têm em mãos.

Uma área das tecnologias de informação onde as bibliotecas são fundamentais é na Computação Gráfica, devido à panóplia de métodos de renderização que oferece. Um destes métodos é o ray tracing. O ray tracing permite simular fenômenos eletromagnéticos naturais como os percursos da luz e fenômenos mecânicos como a propagação do som. Da mesma forma também permite simular tecnologias desenvolvidas pelo homem, como por exemplo redes Wi-Fi. Estas simulações podem ter um realismo e precisão impressionantes, porém têm um custo computacional muito elevado.

A constante evolução da tecnologia permitiu alavancar e massificar novas áreas, como os dispositivos móveis. Os dispositivos são hoje cada vez mais rápidos e cada vez mais substituem e/ou complementam tarefas que anteriormente eram apenas realizadas em computadores ou em hardware dedicado. Porém, o número de bibliotecas para renderização de imagens disponíveis para dispositivos móveis é ainda muito reduzido e nenhuma biblioteca de renderização de imagens baseada em ray tracing conseguiu afirmar-se nestes dispositivos. Esta dissertação tem como objetivo explorar possibilidades e limitações da utilização de dispositivos móveis para a execução de algoritmos de renderização que utilizem ray tracing, como por exemplo, o path tracing progressivo. O objetivo principal é disponibilizar uma biblioteca de renderização para dispositivos móveis baseada em ray tracing.

CONTENTS

1	INTRODUCTION	1
1.1	Context	1
1.2	Motivation	2
1.3	Goals	3
1.4	Applications of Ray Tracing	3
1.5	Document Structure	4
2	STATE OF THE ART	5
2.1	Ray Tracing	5
2.2	Typical CPU features	6
2.3	Key features of Ray Tracing for this work	8
2.3.1	Type of software license	9
2.3.2	Platform	9
2.3.3	Interactivity	9
2.3.4	Progressive	10
2.3.5	Types of Rendering Components	10
2.4	Related work	10
2.4.1	Conclusions	11
3	SOFTWARE ARCHITECTURE	12
3.1	Approach	12
3.2	Methodology	14
3.3	Library	14
3.3.1	Third parties dependencies	15
3.3.2	Renderer	17
3.3.3	Scene	18
3.3.4	Shapes	18
3.3.5	Acceleration Structures	22
3.3.6	Primitive	26
3.4	Rendering Components	27
3.4.1	Shaders	27
3.4.2	Samplers	39
3.4.3	Lights	42
3.4.4	Cameras	45
3.4.5	Object Loaders	50
4	ANDROID CHALLENGES	53

4.1	Android specifics	53
4.2	User Interface	54
4.3	Programming decisions	55
4.3.1	Android benefits	55
4.3.2	Android challenges	56
5	DEMONSTRATION: GLOBAL ILLUMINATION	57
5.1	Results obtained	57
5.1.1	Whitted Shader	60
5.1.2	Path Tracing Shader	67
5.2	Comparison with Android CPU Raytracer (Dahlquist)	73
6	CONCLUSION & FUTURE WORK	74
6.1	Conclusions	74
6.2	Future Work	74
7	BIBLIOGRAPHY	77
A	USER DOCUMENTATION	82
A.1	API	82
A.2	Loading a scene	92
	Appendices	82

LIST OF FIGURES

Figure 1	Illustration of the most common mobile devices - tablet and smartphone (du Net).	2
Figure 2	Illustration of a variety of mobile devices compatible with Android (COOLFINESSE (2017)).	3
Figure 3	Illustration of a typical ray tracer algorithm.	5
Figure 4	Illustration of a typical cache memory inside a microprocessor (Karlo and Aps.).	6
Figure 5	Illustration of the Classic RISC multilevel CPU pipeline (Dictionaries and Encyclopedias).	7
Figure 6	Illustration of a simple superscalar pipeline (Wikipedia (b)).	7
Figure 7	Illustration of a typical execution of SIMD extension (Bishop).	8
Figure 8	Illustration of the developed User Interface.	12
Figure 9	Illustration of the three layers in the application.	13
Figure 10	The Conference scene rendered with OpenGL ES 2.0 before ray tracing the scene.	16
Figure 11	Illustration of tiling the image plane.	17
Figure 12	Illustration of shadow rays casting (Foundry).	18
Figure 13	Illustration of a ray intersecting a plane (Vink).	20
Figure 14	Illustration of a ray intersecting a sphere in two points (Chirag).	21
Figure 15	Illustration of a ray intersecting a triangle (Scratchapixel).	22
Figure 16	Illustration of traversing a regular grid acceleration structure (Torsten Möller).	24
Figure 17	Illustration of traversing a Bounding Volume Hierarchy acceleration structure (Kilgard).	25
Figure 18	Illustration of how Surface Area Heuristic works in Bounding Volume Hierarchy acceleration structure (Lahodiuks).	26
Figure 19	Illustration of the rendering equation describing the total amount of light emitted from a point x along a particular viewing direction and given a function for incoming light and a BRDF (Wikipedia (e)).	27
Figure 20	DepthMap shader.	29
Figure 21	DiffuseMaterial shader.	30
Figure 22	NoShadows shader.	31

Figure 23	Reflections projected on the floor and on the spheres (Vona and Pan).	32
Figure 24	Refraction of light at the interface of two media of different refractive indices (Physics).	32
Figure 25	Whitted shader.	34
Figure 26	The spherical coordinates (Wikipedia (f)).	36
Figure 27	The possible directions in secondary rays (Thompson).	37
Figure 28	PathTracer shader.	38
Figure 29	Halton sequence in a 2D image plane (Wikipedia (d)).	40
Figure 30	Pseudorandom sequence in a 2D image plane (Wikipedia (d)).	42
Figure 31	Hard shadows (Barsegyan).	43
Figure 32	Soft shadows (Barsegyan).	44
Figure 33	Calculation of point P by using barycentric coordinates starting at point A and adding a vector AB and a vector AC (Jacobson).	45
Figure 34	Illustration of how sampling the plane image with jittering works (Waters).	46
Figure 35	Aliasing effect (Science and Engineering).	47
Figure 36	Stratified sampling.	47
Figure 37	Jittered sampling.	47
Figure 38	Perspective camera with hFov and vFov (Schim).	48
Figure 39	Orthographic camera with sizeH and sizeV (Schim).	50
Figure 40	Diagram explaining how the JVM can call native code (Krajci and Cummings).	54
Figure 41	Execution flow of UI thread and Render Task thread.	55
Figure 42	Illustration of conference scene, rendered with shader Whitted in MobileRT.	58
Figure 43	Illustration of Porsche 911 GT2 scene, rendered with shader Whitted in MobileRT.	58
Figure 44	Illustration of Cornell Box scene, rendered with shader Whitted in MobileRT.	59
Figure 45	Time in the Conference scene	60
Figure 46	Time with Samsung smartphone	61
Figure 47	Time to construct accelerators in Samsung Galaxy smart phone	61
Figure 48	Memory usage in the Samsung Galaxy smart phone	62
Figure 49	Time to construct accelerators in the Raspberry Pi 2 Model B	63
Figure 50	Memory usage in the Raspberry Pi 2 Model B	64
Figure 51	Time to construct accelerators in the MINIX NEO X8-H PLUS	65
Figure 52	Memory usage in the MINIX NEO X8-H PLUS	66

Figure 53	Time in the Conference scene	67
Figure 54	Speedup in the Conference scene	68
Figure 55	Time to construct accelerators in the Samsung Galaxy smart phone	68
Figure 56	Memory usage in the Samsung Galaxy smart phone	69
Figure 57	Time to construct accelerators in the Raspberry Pi 2 Model B	70
Figure 58	Memory usage in the Raspberry Pi 2 Model B	71
Figure 59	Time to construct accelerators in the MINIX NEO X8-H PLUS	72
Figure 60	Memory usage in the MINIX NEO X8-H PLUS	73

LIST OF TABLES

Table 2	Comparison of different applications/frameworks that use ray tracing.	11
Table 3	Devices specifications.	57

ACRONYMS

ALU	Arithmetic logic unit.
API	Application Programming Interface.
App	Application.
BSD	Berkeley Software Distribution.
CG	Computer Graphics.
CISC	Complex instruction set computer.
CPU	Central processing unit.
cstdlib	C Standard Library.
DI	Departamento de Informática.
DLP	Data Level Parallelism.
FOSS	Free and open-source software.
GI	Global Illumination.
GLM	OpenGL Mathematics.
GPL	GNU General Public License.
GPU	Graphics processing unit.
GTest	Google Test.
GUI	Graphical User Interface.
ILP	Instruction-level parallelism.
JNI	Java Native Interface.
JVM	Java virtual machine.
Lib	Library.
Libstdc++	C++ Standard Library.
MEI	Mestrado em Engenharia Informática.

MiEI	Mestrado Integrado em Engenharia Informática.
NDK	Native Development Kit.
OpenGL	Open Graphics Library.
OS	Operating system.
RAM	Random-access memory.
RISC	Reduced instruction set computer.
RT	Ray Tracing.
SDK	Software development kit.
SIMD	Single instruction, multiple data.
SMT	Simultaneous multithreading.
STL	Standard Template Library.
UM	Universidade do Minho.

1

INTRODUCTION

1.1 CONTEXT

Programming is like building something with primitive blocks and it can be a very difficult task if we have to program every aspect of the application without some “blocks” already built for us to use. That’s why in the programming world there is a whole panoply of free and commercial libraries with APIs that provide a huge quantity of functions for the programmer to use.

In computer graphics, there are many 3D graphics libraries that provide functionalities to render images based in different techniques. Two of the most used render techniques are rasterization and ray tracing. Ray tracing can be used to calculate and simulate the path of particles and waves. It can simulate the propagation of sound, the path of light and even allows to simulate technologies developed by men like the Wi-fi networks. It can render an image by tracing the path of light through pixels in an image plane and simulating the effects of its interactions with virtual objects. It can render much more realistic shadows, reflections and refractions and in an easier way than rasterization. This technique is capable of producing a very high degree of visual realism but at a great computational cost.

Since nowadays we have more and more mobile systems whose computational power increases every year, the need for more libraries to help the programmers develop applications for these systems is increasing too. However, there are not many options to choose from for developing a renderer based on ray tracing. That’s why this dissertation focuses on assessing and providing a ray tracer library for mobile systems.

It is important to mention that this dissertation is not focused on rendering techniques other than ray tracing. It is not focused in assess different integrators (numerical solutions to the rendering equation) and / or assess different approaches in ray tracing like Packet Traversal. It is also not focused on assess different quasi random numbers generators neither assess the performance of different computer systems.

These rendering algorithms based in ray tracing are very useful because they allow rendering photo-realistic images with a degree of realism much better than the graphics cards allow by default through rasterization. As previously stated, the computing power in

mobile devices processors has been increasing and this may allow a mobile device with a mid-range processor to run these algorithms in a useful time.

Figure 1.: Illustration of the most common mobile devices - tablet and smartphone (du Net).



1.2 MOTIVATION

Among other factors, the productivity of a programmer depends on what libraries he can have access to. But, there are almost no rendering libraries based in ray tracing available today for mobile systems like Android, iOS, Windows 10 Mobile, BlackBerry 10, Tizen, Sailfish OS, Symbian and Ubuntu Touch. And it is likely that these systems already have enough processing power to render images with fairly complex scenes in an acceptable time by using algorithms based in ray tracing.

It is also important to note that there is not much documentation about the advantages and limitations of executing different rendering algorithms in these devices.

Last but not least, it is important to mention that of all the operating systems available for mobile devices, the one chosen for this ray tracing library was Android. Because nowadays it is the operating system with the most market share. More than 85% of the mobile devices use Android, and only 14% of these devices use iOS which is the second most used operating system in these devices. And, besides that, it allows to be executed on a variety of different devices: smart phones, tablet computers, smart TVs, smart watches and even laptop / desktop computers.

Figure 2.: Illustration of a variety of mobile devices compatible with Android ([COOLFINESSE \(2017\)](#)).



1.3 GOALS

The main goal of this dissertation is to assess and demonstrate the advantages of running rendering algorithms in mobile devices.

It is also intended to promote and facilitate the development of applications for mobile systems that use ray tracing techniques, with special emphasis on rendering applications. To do this, it will be developed a library that supports the fundamental operations of a ray tracing engine. This library will allow the agile development of diverse applications, by using components that invoke the functionality of the library itself.

Additionally, it is intended to supply rendering components at a higher abstraction level, like the camera, scene and the integrator that facilitates further the development of applications.

Finally, it is important to do a demonstration of the rendering application with some interface layer to let the user assess the performance of several functionalities provided by the library.

1.4 APPLICATIONS OF RAY TRACING

The subject of this dissertation is about ray tracing because it is a technology widely used in several areas of Computer Graphics.

Ray tracing is a rendering technique that is widely used in the areas of theater and television lighting. It allows set and lighting designers, actors, and directors to develop and visualize complex lighting setups months before a production ever opens.

Ray tracing is capable of considering all of the light in a given environment which is called Global Illumination. Global illumination is a physically correct model, which accurately simulates light's behavior in a real physical environment. Global illumination is a valuable engineering tool in that it allows us to quantitatively analyze the distribution and directionality of light and research radiant heat transfer.

Finally, ray tracing can also be used in the Animation market. It can be used to add "fancy" effects such as reflection and shadowing that are often difficult and time consuming for traditional artists to produce. Graphical technology is also capable of rendering photo-realistic images that would be nearly impossible to produce without computerized ray tracing.

1.5 DOCUMENT STRUCTURE

This dissertation is organized in six chapters: Introduction, State of the Art, Software Architecture, Android challenges, Demonstration: Global Illumination and Conclusion & Future work.

The first chapter describes the context and motivation behind this work, as well as its goals. Its main purpose is to identify the problem at hand and set up goals that should be accomplished.

The second chapter introduces the main concepts of ray tracing and compares different implementations of ray tracers already available in the world wide web. The reason behind this comparison is to show how many ray tracers are already available to mobile devices and highlight the differences between the features they provide. It also provides some information about the processors of today that helps realize their ability to execute computationally demanding algorithms like ray tracing.

The third chapter explains the proposed approach and explains each module developed in the library, as well as each of the rendering components.

The fourth chapter starts with an explanation of some Android specifics, such as the user interface by characterizing its work flow and mentioning some of the benefits and challenges overcame during the development of the application. It also describes some programming decisions made during the development of the ray tracing library.

The fifth chapter summarizes the key results obtained by executing different algorithms with different number of threads and different acceleration structures. And concludes with a small comparison of the features provided in this library and an Android ray tracing engine developed by third parties called [Android CPU Raytracer \(Dahlquist\)](#).

Finally, the last chapter ends this dissertation with the conclusions that can be withdrawn from this work and proposes some future work.

Last but not least, the Appendix contains this library's API and an example code of how to load a scene from a wavefront obj file.

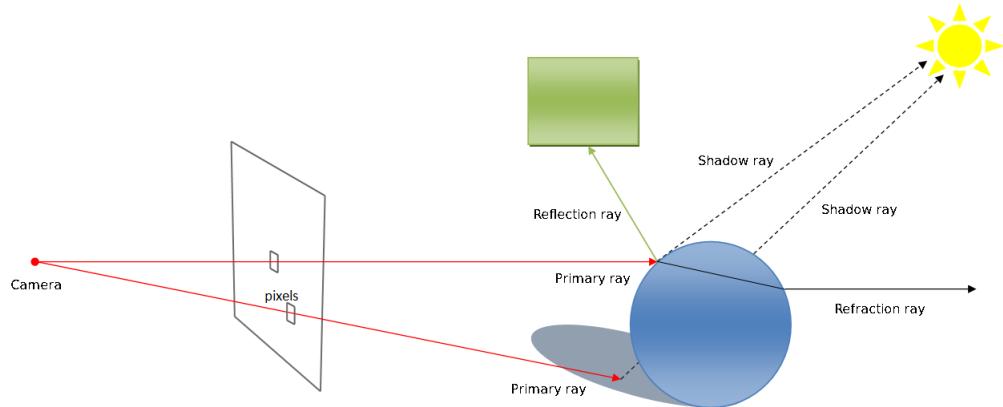
2

STATE OF THE ART

2.1 RAY TRACING

Ray tracing one frame can be, simultaneously, a computationally demanding task and an embarrassingly parallel task. As tracing rays is a recursive process which aims to calculate the luminance of light of each individual pixel separately. At least one ray is shot per pixel and in a naive approach each ray would be intersected with all the objects in the scene in order to determine which one is the closest primitive intersecting that given ray. To evaluate the light intensity that an object scatters towards the eye, the intensity of the light reaching that object has to be evaluated as well. Ray tracing achieves this by shooting additional secondary rays, because when a ray hits a reflecting or transparent surface, one or more secondary rays are cast from that point, simulating the reflection and refraction effects.

Figure 3.: Illustration of a typical ray tracer algorithm.



A typical image of 1024×1024 pixels tends to cast at least a million primary rays and a multiple of that as shadow, reflection, refraction and secondary rays. This is why ray tracing can't provide interactive frame rates with ease.

2.2 TYPICAL CPU FEATURES

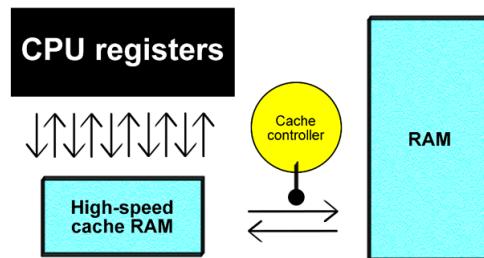
Fortunately, the present state of available technology provides affordable machines with multiple CPU cores that can work in parallel and with great performance. This is achieved thanks to Micro-architectural techniques developed inside the processor that are used to exploit ILP, like: the CPU cache, the Instruction pipelining, the Cache prefetching, the Superscalar execution, the Out-of-order execution, the Register renaming, the Speculative execution, the Branch prediction and the SIMD instruction set already available in the current processors.

The cache is a very small and fast multilevel memory inside the processor that temporarily stores the data read from the main memory.

This memory allows reading its content at a very low latency in the order of magnitude of around 1 nanosecond in the first level, 10 nanoseconds at second level and 50 nanoseconds at third level compared to the typical 100 nanoseconds of a DDR main memory. Like it was said, the CPU cache can be around 100x faster than the RAM.

The downside is that it is very small because its cost are much higher than a typical RAM memory. Nowadays, the level 1 has 64kB, the level 2 has 256kB and the level 3 has 8 MB which is very little compared to the typical 16GB provided by the RAM.

Figure 4.: Illustration of a typical cache memory inside a microprocessor ([Karbo and Aps.](#)).



The Instruction pipelining is a feature inside a processor which allows to reduce the processor's own latency by allowing a form of parallelism called instruction-level parallelism within a single processor. Basically, one instruction is divided in some stages and it allows the possibility to execute different stages of different instructions simultaneously. In the early days, the Classic RISC pipeline were typically divided in five stages:

- Fetching the instruction
- Decoding the instruction
- Executing the instruction where the arguments can be fetched from registers (1 cycle latency) or from memory (2 cycle latency)

- Memory access where it ensured that writing in memory was always performed in the same stage and allowed to use that value in another instruction before it was written in memory
- Writing back the result value in the register file

Nowadays, the current microprocessors have a pipeline with 8-14 stages and can be more complex than the Classic RISC, but the operating principle is the same. This allows faster CPU throughput, which means the number of instructions that can be executed in a unit of time is greater, than it would otherwise be possible at a given clock rate.

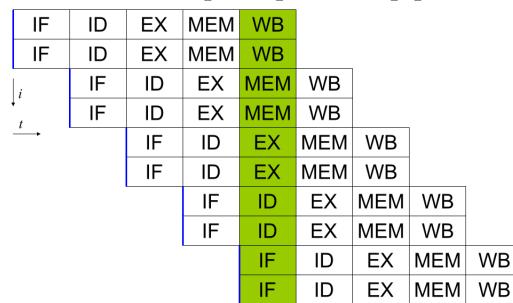
Figure 5.: Illustration of the Classic RISC multilevel CPU pipeline ([Dictionaries and Encyclopedias](#)).

Instr. No.	Pipeline Stage					
1	IF	ID	EX	MEM	WB	
2		IF	ID	EX	MEM	WB
3			IF	ID	EX	MEM
4				IF	ID	EX
5					IF	ID
Clock Cycle	1	2	3	4	5	6
						7

The Cache prefetching is, as the name implies, a feature in the processor that makes the processor fetch the data and the instructions from the RAM to the CPU cache before it really needs to read or execute. This allows to reduce the time which the processor spends waiting for the main memory data.

Superscalar execution is when the CPU uses multiple functional units simultaneously for the same type of task. For example, the CPU can fetch, decode and even execute two instructions at the same time.

Figure 6.: Illustration of a simple superscalar pipeline ([Wikipedia \(b\)](#)).



Out-of-order execution is the name of the technique that allows the CPU to reorder the instructions that are gonna be executed. This allows the CPU to execute the instructions out of order when they don't have a dependency and so it makes use of instruction cycles that would otherwise be wasted. Usually this technique uses the register renaming which

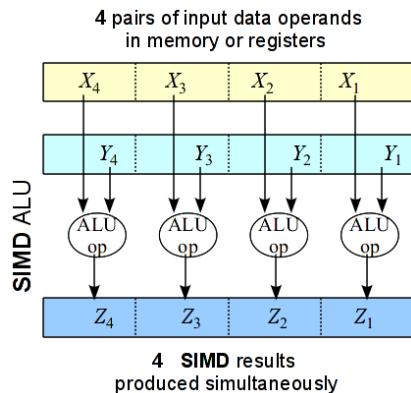
is another technique that is used to avoid unnecessary serialization of program operations imposed by the reuse of registers by those operations.

Speculative execution is the technique where the CPU performs some task that may not be needed. This prevents the CPU from having a delay that would have to be incurred by doing the work after it is known that it is needed.

Branch prediction is the technique where the CPU starts to execute the instructions of a path of a branch before it even knows which branch it needs to take.

Finally, the SIMD instruction set is a set of special instructions that allows the processor to read or write bigger sizes of data. Typically, a 64 bit processor can only work with 64 bits of information at a time during the execution of an instruction. This feature can allow to read or write up to 256 or even 512 bits by using special registers and additional ALUs provided in the processor.

Figure 7.: Illustration of a typical execution of SIMD extension (Bishop).



Nowadays, even mobile devices like smart phones and tablets have multiple CPU cores with features similar to those described above. This opens the possibility to execute more computationally demanding algorithms, like ray tracing, in these devices. And it is important to mention that all the optimization techniques described above are or can be transparent to the software programmer, and which only the processor and the compiler have to know they exist.

2.3 KEY FEATURES OF RAY TRACING FOR THIS WORK

Ray tracing is an algorithm that can have a panoply of features or optimizations in order to reduce the time required to trace all the rays and / or to show the results as fast as possible. It also, like any application, can have different types of software licenses and be executed in different platforms.

For this work, the most important features in a ray tracer are: the type of software license, the platform where can be executed, interactivity, if it is progressive and the type of rendering components.

2.3.1 *Type of software license*

A software application can have different type of software license, but in order to simplify this dissertation, the software license were grouped into three types: Free, Commercial and Open Source.

The license Free means that the user has the right to execute freely the application but cannot copy, modify or distribute the implemented code. This includes the Freeware, Shareware and Freemium types of software licenses.

The license Commercial means that the user cannot even execute the application without buying it first and cannot copy, modify or distribute the implemented code. This includes the Proprietary and Trade secret types of software licenses.

The license Open Source means that the user has the right to execute freely the application and can even copy, modify and distribute the implemented code. This includes the Public-domain software, and the Permissive and Copyleft types of software licenses like the BSD and GPL.

The software license is very important in this work because it lets the user know if he can develop something over the provided software or if he can just use the application.

2.3.2 *Platform*

An application can only be executed in the platforms that the developers compiled the code for. So, a ray tracer that can be executed in a desktop may also or may not be executed on another platform, like in a mobile device. Usually, the software is specific for a single Operating System. This information is very important for this work because it inform us whether a ray tracer can be executed in a mobile device with the typical Operating System like Android, iOS or Windows 10 Mobile.

2.3.3 *Interactivity*

In ray tracing, interactivity means rendering an image with a very low response time, like a few milliseconds per frame. An interactive ray tracer can render a scene with multiple frames per second.

2.3.4 *Progressive*

A typical ray tracer shows the rendered image only after the whole process is complete. A progressive ray tracer is a ray tracer that updates the color of pixels in the screen as soon as the rays are traced, instead of waiting for the rendering process to complete. This means that an image is rendered quickly with some aliasing or noise and it is progressively improved over time.

2.3.5 *Types of Rendering Components*

A ray tracer can be developed with the different rendering components programmed separately. Some examples of rendering components are: Integrators, Cameras, Scenes, Samplers, Shapes, Lights and Acceleration Structures. In some ray tracers, these rendering components can even be programmable by the user. This is very important because it allows the user to develop his own renderer based in ray tracing without having to develop every feature in the ray tracer.

2.4 RELATED WORK

The possibility of rendering an image with ray tracing was demonstrated in 1980 by Whitted ([dos Santos](#)) and since then the number of libraries that provide basic ray tracing functionalities increased greatly. There is a wide range of different ray tracers available today for the programmer to use, yet the majority can only be used with the traditional personal computer hardware (desktop or laptop).

The table [2](#) shows some applications or frameworks that use ray tracing available today and compares them according to their type of license, platform compatibility, interactivity, if they are progressive and whether they allow development of your own rendering components like the integrator and sampler. Note that some of these ray tracers provide only the engine with the basic ray tracing functions, such as creating rays and intersecting them with geometric primitives, so the rendering components are only programmable if the application that uses the engine supports it. During the research of the available ray tracers, others than the ones presented in the table were found, but they were excluded because the documentation was very poor without explaining the basic functionalities provided or because there was no documentation at all.

Table 2.: Comparison of different applications/frameworks that use ray tracing.

Product	License	Platform	Mobile	Interactivity	Progressive	Programmable Components
Optix (Nvidia (a))	F	Nvidia GPU	N	Y	Y	Y
Optix Prime (Nvidia (b))	F	CPU & Nvidia GPU	N	Y	Y	Y
RenderMan RIS (Pixar)	C	CPU	N	Y	Y	Y
OctaneRender (Inc)	C	GPU	N	Y	Y	N
Embree (Intel)	O	Intel CPU	N	Y	Y	Y
Radeon Rays (AMD)	O	CPU & GPU	N	Y	Y	Y
PBRT (Matt Pharr)	O	CPU	N	N	N	N
Visionaray (Zellmann)	O	CPU & Nvidia GPU	N	Y	Y	N
YafaRay (Gustavo Pichorim Boiko)	O	CPU	N	N	N	N
tray_rust (Usher (c))	O	CPU	N	N	N	N
micro-packet (Usher (a))	O	Intel CPU	N	N	N	N
tray (Usher (b))	O	CPU	N	N	N	N
The G3D Innovation Engine (Morgan)	O	CPU	N	N	N	N
HRay (Kenneth)	O	CPU	N	N	N	N
Mitsuba (Jakob)	O	CPU	N	N	N	N
Indigo RT (Limited)	O	CPU & GPU	N	Y	N	N
jsRayTracer (Chedea)	O	CPU	N	Y	Y	N
Android CPU Raytracer (Dahlquist)	O	CPU	Y (Android)	Y	N	N

2.4.1 Conclusions

As the table 2 shows, there is a lack of generic ray tracing libraries for the mobile devices. Although there are some closed-source ray tracing demo applications, only one ray tracer already available has some sort of documentation and is compatible with mobile devices, in this case with Android. This ray tracer is open source, uses only the CPU of the device and has a good performance. It only renders spheres and even allows interactions with them during the render process. But, it doesn't support progressive rendering and also does not allow the programmers to use their own rendering components.

This dissertation aims to fix this lack of libraries, by providing one that contains ray tracing basic functionalities and the ability to let the programmer be able to develop his own rendering components like the sampler, integrator, camera, types of lights and even develop his own object loader. It also studies the drawbacks that these mobile devices may have comparing with the average multi-core personal computer hardware. And finally, a small comparison was made with the Android CPU Raytracer ([Dahlquist](#)) in order to illustrate the advantages and disadvantages of both. This comparison is important because it enriches all the work done, as it demonstrates if the performance of the provided features are relatively efficient.

3

SOFTWARE ARCHITECTURE

3.1 APPROACH

The development of this demonstration application involves the distinction of three layers of abstraction: user interface, rendering components and the library itself.

The top layer is the User Interface which is obviously application, and eventually device, dependent. The user interface of this dissertation's demo application is very simple and just allows the user to see the rendered image and choose some rendering components to use, like the integrator, the sampler, the number of threads and samples and choose the scene to render. Although this layer is useful, this dissertation focuses only on the middle and bottom layers.

Figure 8.: Illustration of the developed User Interface.

Spheres	NoShadows	3
Cornell2	Whitted	4
Spheres2	PathTracer	1
RegGrid	100	100 800x976
BVH	1	1 992x1200
None	4	2 48x48

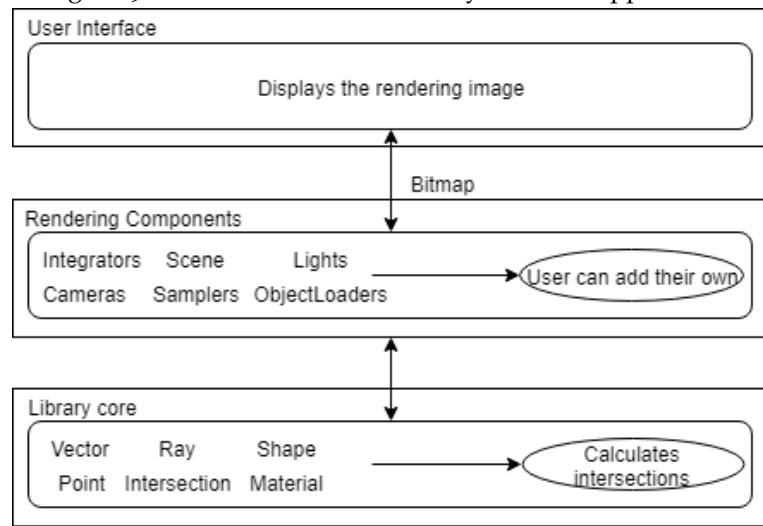
The middle layer provides the rendering components which are abstract concepts about rendering which use functionalities that the library itself offers to the programmers. Some of these rendering components are the camera, the light, the sampler, the integrator and the object loader. These rendering components are useful for the programmer since it allows them to use features without having the need to know how these were developed. And, of course, this facilitates and accelerates the development of new rendering applications for the programmers.

Lastly, the bottom layer is the library itself, which contains the business logic of basic features in a renderer, that the rendering components use. These features are the basic functionalities of a ray tracing engine. Those functionalities can be: create vectors and

points; create different shapes like triangles, spheres and planes; create different primitives with different materials and with different shapes; cast rays; and even intersect rays with the primitives.

It is also important to mention that the demo application was built in order to show the developed features, the performance achieved in the mobile devices and also to help promote the library. This demo application is a good way to test it in several Android devices, like: a smart phone, a tablet computer, a smart TV, or even a smart watch.

Figure 9.: Illustration of the three layers in the application.



Besides the abstraction layers, there are some important strategic decisions made in order to guide the progress of the development of this library.

The first decision was: the primary rays always have origin in the camera. This decision was made in order to not mix the code of the integrator with the ray tracer renderer engine.

Other decision made was to make the rendering process progressive, which means that the rendered image is incrementally refined with more and more traced rays. As the integrator will be converging to better values. This is important in order to give the user a fast rendered image, with some noise or aliasing, and converge it progressively to a better solution with higher details and practically without any visual noise or aliasing.

Another thought aspect that was studied is the permission of dynamic scenes and / or dynamic cameras, which means to let the programmer modify the camera or the scene while the ray tracer is rendering it. This makes possible to build challenging applications and also provide more interesting scenes and more eye candy applications for the final user.

Last, but not least, is that the code was developed in a modular way, in this case was programmed in an object oriented manner. This allows the programmers to code their own rendering components, like the integrator, camera, sampler and light, without having to develop the basic features in the renderer engine.

3.2 METHODOLOGY

In order to take advantage of most of the mobile CPU resources and give a good performance for the applications, the library and the rendering components were developed using the native programming language C++. This was achieved by using the Native Development Kit (NDK) provided by the Integrated Development Environment (IDE) Android Studio. This way, the mobile device can provide better performance in computationally intensive applications at the expense of the programmer having to manage manually the memory. And it facilitates the porting of existing C/C++ code to Android and promotes developing multiplatforms applications. But, it won't ensure device portability despite processor architecture and the number of libraries compatible with the NDK is very scarce. It is also important to mention that the User Interface was developed in Java using the traditional Software Development Kit (SDK) provided by the Android Studio because there is no framework in the NDK that helps the programmer design his own user interface. Despite that, its performance is not very important because it doesn't interfere significantly with the others layers of the application.

3.3 LIBRARY

As stated above, this library was implemented in an object-oriented fashion. The most important classes that provide functionalities already developed for the user to use are:

- Renderer: class that starts the rendering process and stores the calculated pixels colors in a C style array.
- Scene: class that stores the geometry information in vectors and provides methods to trace rays in the scene without any acceleration structures.
- Shapes: set of classes that allows to create triangles, spheres and planes.
- Material: class that stores all four types of material color:
 - Emission light color
 - Diffuse reflection color
 - Specular reflection color
 - Specular refraction color
- Primitive: class that stores the shape and material of each primitive in the scene.
- Ray: class that represents a ray casted into the scene.

3.3.1 Third parties dependencies

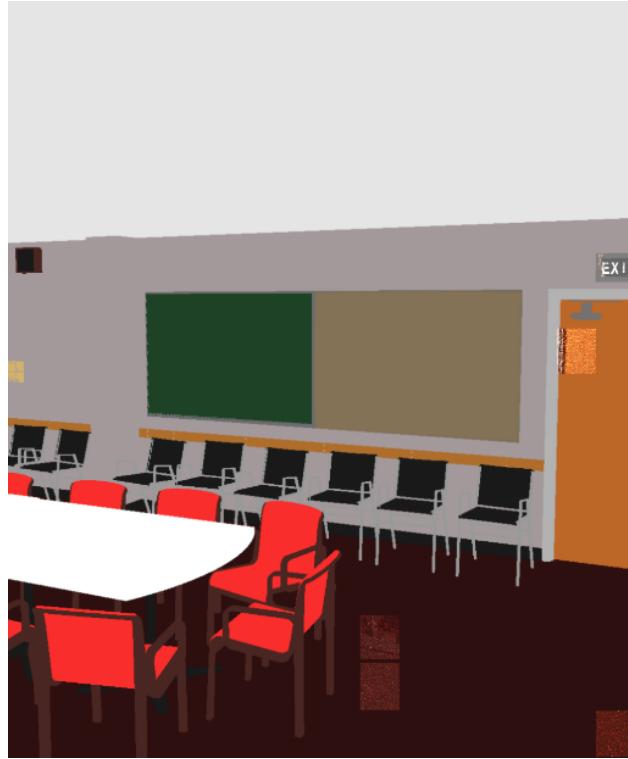
Before describing each functionality provided in each class developed, it is important to mention that this library uses three other libraries developed in C++ by third parties. Those libraries are:

- OpenGL Mathematics ([Creation](#)): used to create 3d points and vectors, perform geometry calculations and store pixels and primitives' colors. Before using this library, it was implemented our own mathematics library and its performance was very bad compared with this library.
- tinyobjloader ([syoyo](#)): used to load scenes from wavefront obj files into the main memory.
- Google Test ([Google \(b\)](#)): used to create some unit tests and to mock some classes. This is important in order to have a safe and maintainable code.
- C++ Standard Library ([Wikipedia \(a\)](#)): used in order to save the data in popular containers like the array and the vector. It also provides some useful utilities like the unique pointers in the Dynamic memory management utilities, the String and the Atomic operations used in the Samplers.

All these libraries are Android compatible and are reliable in terms of performance and maintenance. Besides being dependent on these four libraries, this library also depends on OpenGL ES 2.0 ([Inc.](#)) from Android SDK and depends on CMake ([kitware](#)) in order to compile the application.

The OpenGL ES 2.0 is necessary in order to let the application rasterize one frame of the scene so the user can view it without having to wait a long time to render it with the ray tracer. For example, in the Conference scene, the OpenGL in the Samsung Galaxy Fresh Duos can rasterize one frame in less than 17ms compared with 500ms that the NoShadows shader takes. This is done by copying the scene data from native code to the OpenGL of JVM with the help of JNI.

Figure 10.: The Conference scene rendered with OpenGL ES 2.0 before ray tracing the scene.



The CMake was used so that the application code could be compiled on any operating system. As it is a cross-platform family of tools designed to build, test and package software developed for a variety of operating systems like Windows, Linux, Mac OS X, and even FreeBSD. The main goal of CMake is to control the software compilation process using simple platform and compiler independent configuration files, and generate native makefiles and workspaces that can be used in the compiler environment of the programmer's choice. This facilitates the integration of third party libraries that use CMake to compile into our own project. But, unfortunately the NDK only provides CMake up to the version 3.6.0, so projects that use latter versions of CMake will not work.

The CMake generate makefiles that call the available C++ compilers in the Android Environment. The NDK in Android Studio uses, by default, the LLVM Project, which is a collection of modular and reusable compiler and toolchain technologies. This toolchain provides Clang, which is C/C++/Objective-C compiler that aims to deliver amazingly fast compiles and extremely useful error and warning messages for the programmers. Unfortunately, this compiler only provides features up to C++ 17, so developers can't use the most recent features of C++ 20.

Last, but not least, it is important to mention that all the code of this application, including the ray tracer library, the rendering components and the Graphical User Interface are available in the [github](#), which is a web-based hosting service for version control using

Git. It is 100% open source and it is accessed in the web page: <https://github.com/PTPuscas/MobileRayTracer>.

3.3.2 Renderer

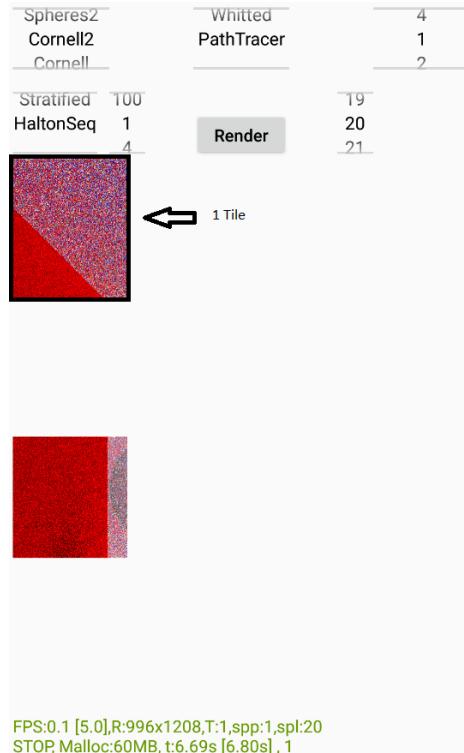
The Renderer is the closest class to the application that starts the rendering process. This class provides two main methods:

```
1 void renderFrame(std::uint32_t *bitmap, std::int32_t numThreads, std::uint32_t stride) noexcept;
2 void stopRender() noexcept;
```

Listing 3.1: Main methods in Renderer

The *renderFrame* method starts the rendering process and writes the calculated light luminance of each pixel in the parameter *bitmap*. This method also allows to choose the number of threads that will render the image into the bitmap, and needs the *stride* of that bitmap array. Besides rendering the scene in a progressive way, the image plane is divided into 256 tiles of pixels and it is traced one primary ray per pixel in each tile. This allows the user to wait less time in order to view the scene being rendered.

Figure 11.: Illustration of tiling the image plane.



The `stopRender` method only serves to stop the rendering process without cleaning the pixels' colors already calculated.

3.3.3 Scene

The Scene is the class that handles the process of intersecting a ray with the primitives and source lights in the scene. Besides providing a vector to add the light sources and a vector to add primitives to the scene, it also provides two main methods:

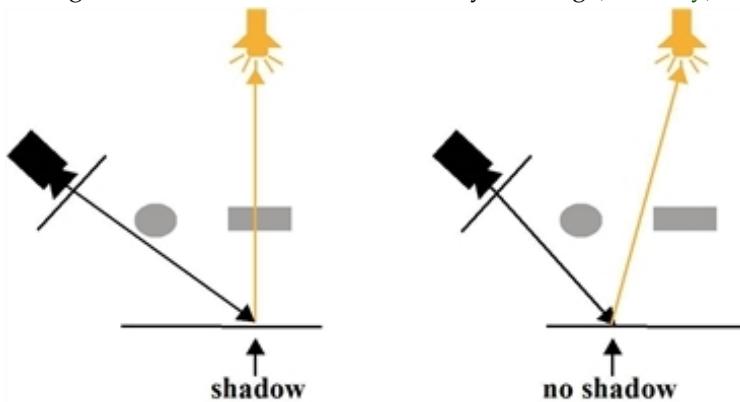
- 1 Intersection trace(Intersection intersection, `const Ray &ray`) noexcept;
- 2 Intersection shadowTrace(Intersection intersection, `const Ray &ray`) noexcept;

Listing 3.2: Main methods in Scene

The *trace*, as the name implies, is a method that tries to intersect a ray with all the primitives and light sources in the scene. And It returns the closest intersection to the origin of that ray. This method is used to determine the intersection of the casted ray with the primitives in the scene.

The *shadowTrace*, is similar to the *trace* method but with the difference that returns the first intersection found. The purpose of this method is to simulate the shadows in the scene, like it is illustrated in figure 12. The shadow ray is represented by the yellow line pointed to the light source.

Figure 12.: Illustration of shadow rays casting (Foundry).



3.3.4 Shapes

In order to make possible to generate scenes with all kind of objects, it was developed three types of shapes: plane, sphere and triangle. One way to allow this would be to use the process of inheritance available in C++. Which means, for example, create a class

called Shape and create sub classes like Triangle, Plane or Sphere that are derived from that class. And each of those classes would implement different intersect methods. In this way, it would be possible for the programmer to develop his own custom shapes as new Rendering Components. But of course, the benefit of this flexibility also brings the downside of performance loss as the application will have at runtime to figure out which of the derived classes is required to call. Because, calling a virtual function requires a v-table lookup, and can make it 2-3 times slower than a normal function. The other way to do it is by developing each shape class without inheritance and being each class independent of each other. This can have greater performance but the downside is that the programmer cannot add new shapes to the library. This ray tracer was implemented this way because it is not important to let the user add his own shapes as rendering components as with only the triangles it is possible to render many types of custom shapes. Also, the performance is important in the intersect method of the shapes because it is a method that can be called millions of times while rendering a scene.

So, all shapes provide three important methods:

- 1 Intersection intersect(`const` Intersection &intersection, `const` Ray &ray) `const` noexcept;
- 2 `bool` intersect(`const` AABB &box) `const` noexcept;
- 3 AABB getAABB() `const` noexcept;

Listing 3.3: Main methods in Shape

The first *intersect* method determines if a ray intersects the shape and returns the intersection. It is important to mention that, obviously, each shape was developed with different algorithms.

The second *intersect* method checks if an AABB contains the shape, and returns only true or false. This is useful to determine in which AABB the shape belongs to in the regular grid accelerator.

Finally, the *getAABB* method, as the name implies returns the smallest AABB that surrounds the shape. This is used in the building of the acceleration structures because it allows to check if a ray intersects multiple shapes inside an AABB as it is simpler to intersect an AABB than a shape.

Plane

The plane is an essential primitive shape because it allows the user to build indoor scenes.

The construction of a plane requires just an arbitrary point in the plane and the normal of the plane. So, the intersect method implemented has the following algorithm taken from [belthaczar](#):

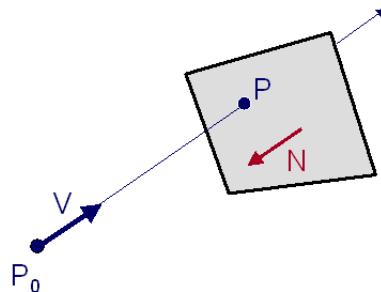
Let

$$\begin{aligned}
 intersectionPoint &= rayOrg + rayDir * dist \\
 n \cdot (q - x) &= 0 \\
 n \cdot (q - intersectionPoint) &= 0 \\
 n \cdot (q - rayOrg - rayDir * dist) &= 0 \\
 n \cdot (q - rayOrg) &= (n \cdot rayDir) * dist \\
 dist &= [n \cdot (q - rayOrg)] / (n \cdot rayDir)
 \end{aligned} \tag{1}$$

where

n	: normal of the plane
q	: known point on the plane
x	: any point on the plane
$rayOrg$: origin of the ray
$rayDir$: direction of the ray
$dist$: distance of the ray
$intersectionPoint$: intersection point of the ray in the plane

Figure 13.: Illustration of a ray intersecting a plane (Vink).



Sphere

The sphere is also an important primitive shape because it allows the user to build some common objects with a shape of a ball.

The construction of a sphere requires just the point in the center and the radius of the sphere. It also should be noted that a ray can intersect a sphere at two points and therefore it is necessary to determine the closest intersection point to the origin of the ray. The algorithm used in this library was the algebraic version. So, the intersect method implemented has the following algorithm taken from [Feminella](#):

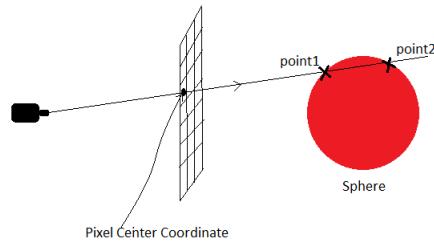
Let

$$\begin{aligned}
 intersectionPoint &= rayOrg + rayDir * dist \\
 dist &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
 \end{aligned} \tag{2}$$

where

a	: $\ rayDir\ ^2$
b	: $2 * rayOrgToCenter \cdot rayDir$
c	: $\ rayOrgToCenter\ ^2 - sphereRadius^2$
$rayOrgToCenter$: vector from the origin point of the ray to the center of the sphere
$sphereRadius$: radius of the sphere
$sphereCenter$: center point of the sphere
$rayOrg$: origin point of the ray
$rayDir$: direction of the ray
$dist$: distance of the ray
$intersectionPoint$: intersection point of the ray in the sphere

Figure 14.: Illustration of a ray intersecting a sphere in two points (Chirag).



Triangle

And finally, obviously the triangle has also been implemented because, as it is the simplest primitive with an area, it allows to build many different object shapes. The implemented Ray-Triangle Intersection algorithm was the Möller-Trumbore ([Moller and Trumbore \(1997\)](#)) algorithm, which was named after its inventors Tomas Möller and Ben Trumbore.

The construction of a triangle requires three points $[A, B, C]$, two vectors $[AB, AC]$ and the normal of the triangle. This is needed in order to execute the Möller-Trumbore algorithm for the intersection of a ray with a triangle. The Möller-Trumbore ray-triangle intersection algorithm is a fast method for calculating the intersection of a ray and a triangle in three dimensions without the needing to precompute the plane equation containing the triangle. It only needs to store 1 triangle vertex and 2 triangle vectors, which means that

each triangle needs only to store 9 floating point values, or 36 bytes. So, the intersect method implemented has the following algorithm:

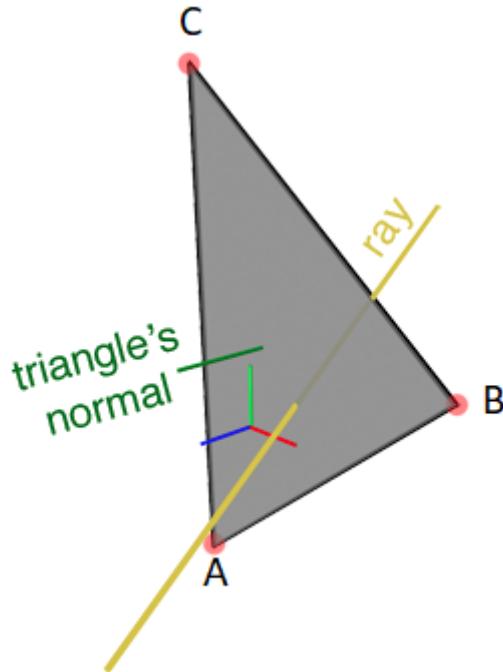
Let

$$\begin{aligned} intersectionPoint &= rayOrg + rayDir * dist \\ dist &= (AC \cdot ((rayOrg - A) * AB)) / (AB \cdot (rayDir * AC)) \end{aligned} \quad (3)$$

where

A	: point A of the triangle
AB	: vector of the triangle from point A to B
AC	: vector of the triangle from point A to C
$rayOrg$: origin point of the ray
$rayDir$: direction of the ray
$dist$: distance of the ray
$intersectionPoint$: intersection point of the ray in the triangle

Figure 15.: Illustration of a ray intersecting a triangle (Scratchapixel).



3.3.5 Acceleration Structures

As previously stated, the rendering algorithms based on ray tracing can be very computationally demanding, like Path Tracing and Bidirectional Path Tracing. These

algorithms are very computationally demanding because if the scene is very complex, made with millions of primitives, it is necessary to try to intersect every ray casted into the scene with those millions of primitives. This task is, obviously, very time consuming and can waste a lot of battery power on the mobile device. Luckily, there are already known techniques, called acceleration structures, that helps to accelerate this process by reducing the number of intersections.

There are two types of approaches that acceleration structures can have:

- Subdivision of Space: the 3D space of the scene is divided in smaller portions of space which the volume elements can be uniform or irregular:
 - Regular Grids
 - Octrees
 - Kd-trees
- Subdivision of Objects: the 3D space of the scene is divided by aggregating the 3D primitives in groups which are next to each others:
 - Bounding Volume Hierarchy (BVH)

The idea in Subdivision of Space is allowing the intersections to start with the nearest primitives of the ray and is only intersected with the further ones, if these are not in the same direction of the ray. These structures also allow the rapid and simultaneous rejection of groups of primitives.

In contrast, structures based on Subdivision of objects only allow the rapid and simultaneous rejection of groups of primitives.

The developed library provides only two types of structures, one of each approach: the regular grid and the bounding volume hierarchy. Both structures are built with Axis Aligned Bounding Volumes (AABB), which are volume elements (voxels) that can cover, one or more primitives from the 3D scene. An AABB is a voxel, like the name implies, that is aligned to the axis of the scene. This type of voxels are great to use in acceleration structures because it is very fast to test whether or not a ray intersects that voxel and it only needs to store, in memory, two points of that box. Besides AABB, it is possible to build the structures with other types of voxels, like:

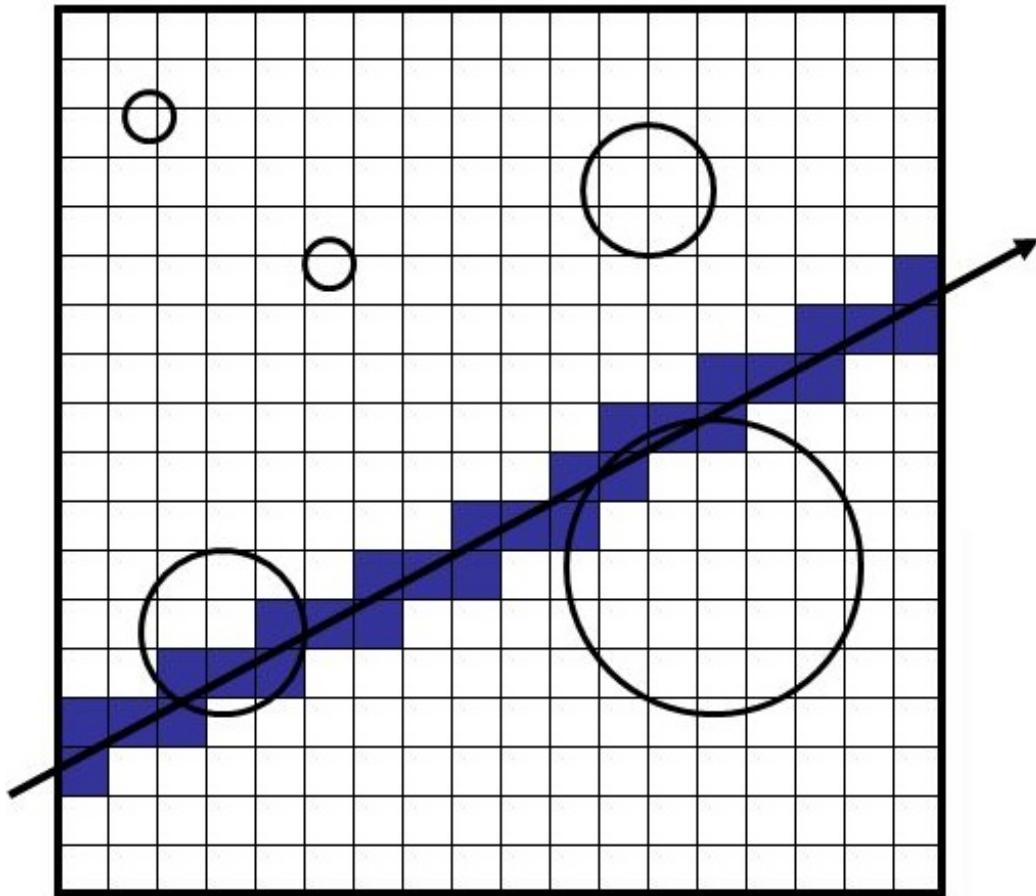
- Sphere: intersection with a sphere is simpler than AABB but covers more 3D space than the scene primitive needs.
- Oriented Bounding Volume: its like an AABB but rotated according to the orientation of the primitive.
- k-Discrete Oriented Prototype (k-DOP): a generalization of AABBs defined by k hyperplanes with normals in discrete directions.

- Convex Hull: is the smallest and more complex convex volume containing the object.

Regular Grid

A regular grid is an acceleration structure where the scene space is subdivided into equal voxels. Each voxel is a volume element where it could have inside one or more primitives of the scene, if a ray intersects that voxel, then it tries to intersect the primitives inside it. The order of testing each voxel is from the nearest of the ray into the furthest in the direction of that ray. This type of structure is very fast to build but its traversal is poorly efficient due to poor distribution of primitives by voxels. The figure 16 shows the process of intersecting a ray with the primitives in a scene.

Figure 16.: Illustration of traversing a regular grid acceleration structure (Torsten Möller).



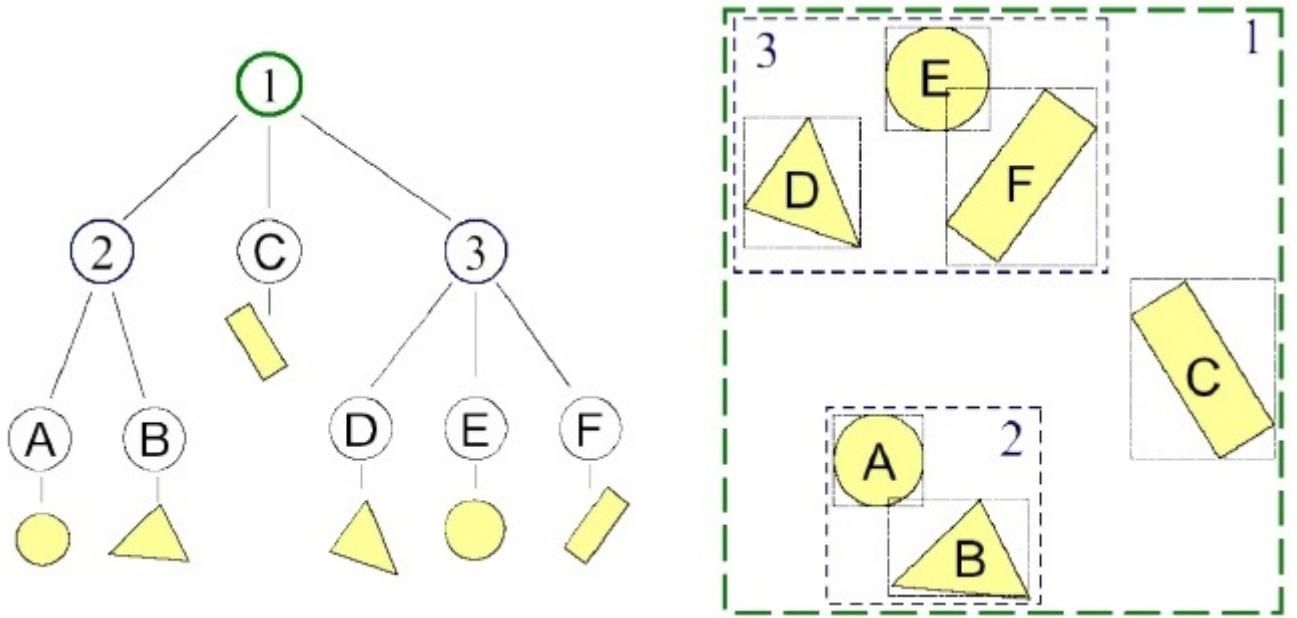
The developed ray tracer library provides a regular grid as an acceleration structure and its operation is as shown in the figure.

Bounding Volume Hierarchy

Bounding Volume Hierarchy is another acceleration structure provided in the developed library. This structure is different from the regular grid because the space is not divided into voxels of the same size. The space is divided in a way where the primitives of the scene are grouped according to the nearest primitives of each. This allows to build a structure of voxels where a voxel has, for sure, one or more primitives inside.

The figure 17 shows an example of a BVH structure of a scene. As it can be seen, each primitive is covered by an AABB, and then two or more AABBs are covered by one larger AABB. In this library, this is built in a top down manner, where the scene is divided by a fake plane that divides the scene into two AABBs. The calculation of those planes is performed with the Surface Area Heuristic (SAH) algorithm which is a very popular heuristic commonly used in ray tracers.

Figure 17.: Illustration of traversing a Bounding Volume Hierarchy acceleration structure (Kilgard).



In this library, the BVH class is a template, because this way, it is possible to create multiple BVH classes where each one store a vector of primitives with different shapes.

Surface Area Heuristic

As it was said above, the library uses the Surface Area Heuristic in order to split the scene. The algorithm used in the library is very simple, it just needs to calculate the minimum area surface of the sum of left and right subtrees, as following:

$$SAH_{optimal} = \min(S_L * N_L + S_R * N_R) \quad (4)$$

N_L = number of polygons in left subtree

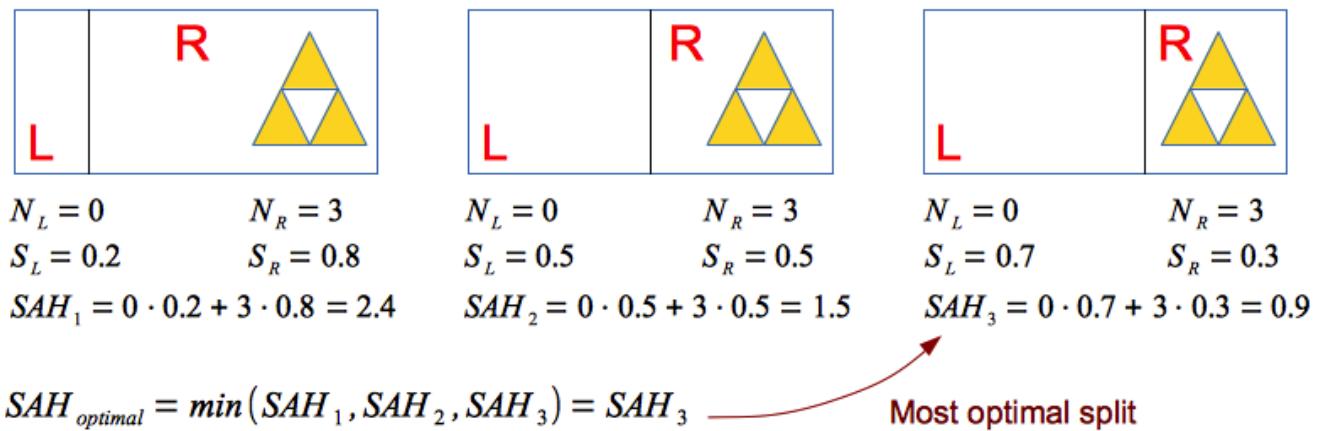
N_R = number of polygons in right subtree

S_L = surface area of left subtree

S_R = surface area of right subtree

The figure 18 shows a very simple example of how the SAH algorithm works.

Figure 18.: Illustration of how Surface Area Heuristic works in Bounding Volume Hierarchy acceleration structure (Lahodiuk).



3.3.6 Primitive

The Primitive is a template class that needs a Shape and wraps it with the Material together. It is useful to be a template class because this way it allows to different shapes have different types of material.

- 1 Intersection intersect(Intersection intersection, `const Ray &ray`) noexcept;
- 2 AABB getAABB() `const` noexcept;

Listing 3.4: Main methods in Primitive

The *intersect*, as the name implies, is a method that tries to intersect a ray with the primitive. This method calls the *intersect* of the Shape and it returns the closest intersection to the origin of that ray. It is used to determine the intersection of the casted ray with the primitives in the scene.

The *getAABB* is the method that returns the AABB voxel that surrounds the primitive. It also calls the *getAABB* from the Shape and it is used to construct the Acceleration Structures.

3.4 RENDERING COMPONENTS

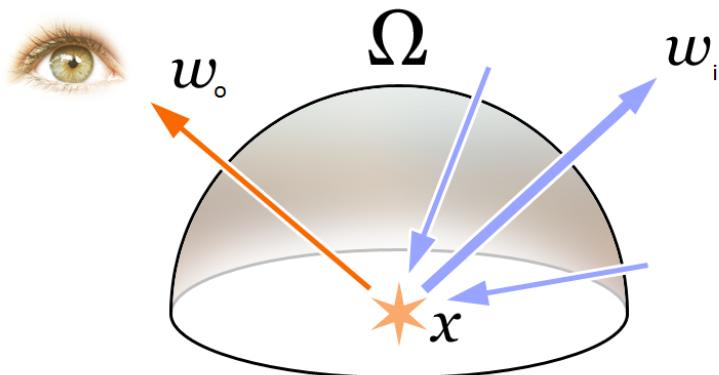
In order to show the functionalities provided by the library, it was developed a few Rendering Components. It is provided the perspective and orthographic cameras, the area and point lights, six Samplers, five Shaders and one Object loader.

3.4.1 Shaders

The implemented ray tracer was programmed in an object oriented fashion, so each Rendering Component was developed separately, in a different class. This allows the user to develop his own Rendering Components without having to develop the ray tracer engine.

A shader is the most important Rendering Component because it is the Rendering Component that describes how the Rendering Equation is approximated. The Rendering Equation describes how the total radiance reflected by any point p of a surface in a direction ωr is calculated. The Bidirectional Reflectance Distribution Function (BRDF) is a function that tries to approximate the Rendering Equation. In summary, a shader, in this context, is the algorithm of a BRDF.

Figure 19.: Illustration of the rendering equation describing the total amount of light emitted from a point x along a particular viewing direction and given a function for incoming light and a BRDF ([Wikipedia](#) (e)).



There were developed five shaders: NoShadows, Whitted, PathTracer, DepthMap and DiffuseMaterial. All shaders provide only one main method that allows the user to calculate the RGB color of a pixel with the information of a ray casted into the scene and its nearest calculated intersection.

```

1 virtual bool shade(glm::vec3 *rgb, const Intersection &intersection, const Ray &ray) noexcept = 0;
2 glm::vec3 getCosineSampleHemisphere(const glm::vec3 &normal) const noexcept;
3 bool rayTrace(glm::vec3 *rgb, const Ray &ray) noexcept;
4 bool shadowTrace(Intersection intersection, const Ray &ray) noexcept;
```

Listing 3.5: Main methods of Shader

The method *shade* calculates the color of the pixel that will be stored in the parameter *rgb* and the *intersection* contains the calculated information about the intersection of a ray with the nearest primitive of the scene like the intersection point, its normal and the material of the intersected material. Finally the parameter *ray* is where the information about the ray like the origin of the ray and its direction is accessed.

The method *getCosineSampleHemisphere* calculates a random direction in World coordinates around an hemisphere with a given *normal*.

This method is used to cast new secondary rays in the scene in order to simulate the indirect light. This is typically used in Path Tracers.

DepthMap

DepthMap is a very simple shader that calculates an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint ([Wikipedia \(c\)](#)). It is useful to various things like simulating the effect of uniformly dense semi-transparent media within a scene - such as fog, smoke or large volumes of water; or even to create Shadow Maps.

The algorithm is as much as simple as:

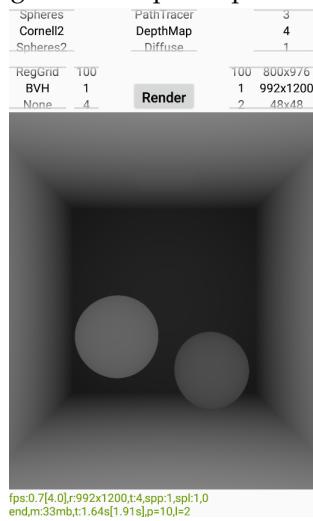
```

1 float maxDist = glm::length(maxPoint - rayOrigin) * 1.1
2 float depth = std::max((maxDist - intersectionLength) / maxDist, 0.0)
3 rgb = {depth, depth, depth}
```

Listing 3.6: Algorithm of DepthMap Shader

Where the *maxPoint* is the furthest 3D point that the user send as parameter in the DepthMap constructor.

Figure 20.: DepthMap shader.



DiffuseMaterial

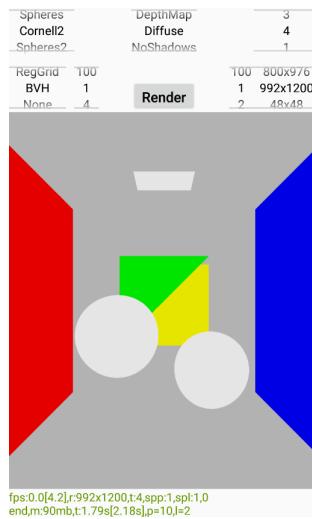
The *DiffuseMaterial* is another very simple shader that only computes directly the diffuse material color of the intersected primitive. But, in case the primitive material does not have diffuse color, the specular color is computed or the transmission color or even the light emission. This is can be useful to debug the application because it renders the scene, including all types of materials, very fast. And the algorithm is very simple as illustrated below:

```

1 if (kD > o) {
2   outputRGB = kD;
3 } else if (kS > o) {
4   outputRGB = kS;
5 } else if (kT > o) {
6   outputRGB = kT;
7 } else if (Le > o) {
8   outputRGB = Le;
9 }
```

Listing 3.7: Algorithm of *DiffuseMaterial* Shader

Figure 21.: DiffuseMaterial shader.



NoShadows

NoShadows is a simple shader because, as the name implies, it does not synthesize the shadows and it only simulates the direct lighting. This BRDF only simulates direct lighting in primitives with diffuse surfaces and it does not take into account the light coming from other points. As already said, the indirect lighting is not fully simulated but rather approximated with a fixed ambient light of about 10% of the diffuse color of the objects. This can be useful in order to let the user render a scene with good details and without having to wait as long as the Whitted algorithm.

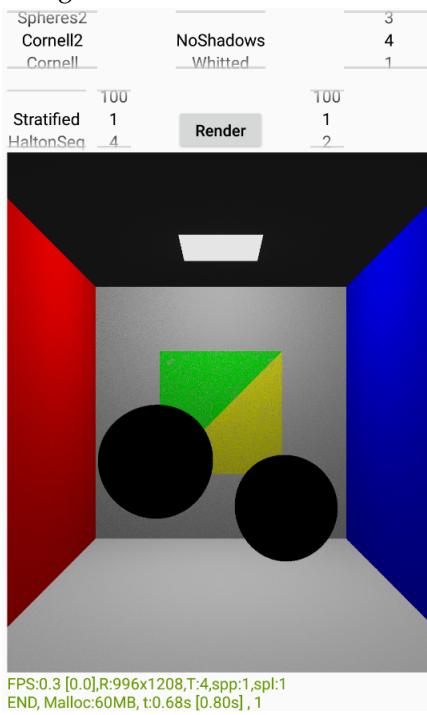
The algorithm developed is as following:

```

1 if (intersected material is diffuse) {
2   for each light source
3     for each sample
4       vectorToLight = lightPosition - intersectionPoint
5       cos_N_L = vectorToLight . intersectionNormal
6       if (cos_N_L > 0) rgb += kD * radLight * cos_N_L
7       rgb /= #samples
8       rgb /= #lights
9       rgb += kD * 0.1 //Ambient light
10  }
```

Listing 3.8: Algorithm of NoShadows Shader

Figure 22.: NoShadows shader.



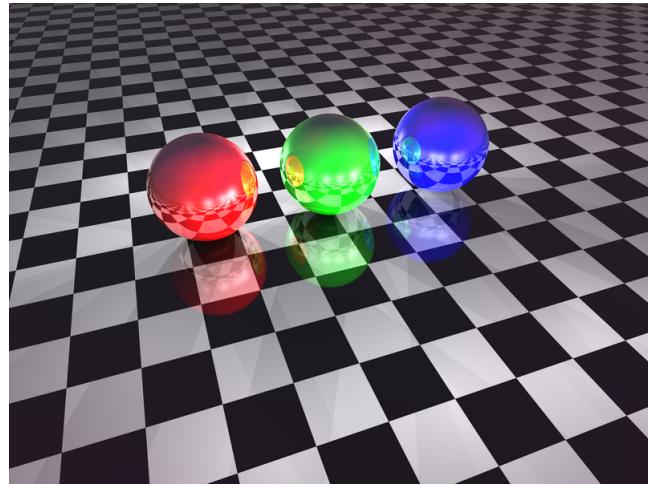
Whitted

As the name of this shader implies, Whitted is the algorithm presented by John Turner Whitted in the 1980s. Like the previous shader, it doesn't simulate indirect lighting. This BRDF recursively perform the ray casting to simulate refraction and reflection on diffuse and specular surfaces. In order to simulate a reflective and refractive surfaces, the algorithm was divided into each case. This makes it possible to simulate refractive surfaces, reflective surfaces and even refractive and reflective surfaces.

The reflected light in a diffuse surface is calculated by adding the casting rays in direction to the lights. These rays are called shadow rays and their radiance are multiplied by the dot product between vector to the light and the shading normal. At the end, the summation is multiplied by the diffuse color of the intersected primitive and divided by the number of samples taken.

The reflected light in a specular surface is calculated in a different way. The specular ray is casted and ray traced, then the obtained radiance is multiplied by the specular color of the intersected primitive. The reflection direction is the subtraction between the double of dot product between the inverse of the ray direction and the shading normal multiplying by the shading normal, and the inverse of the ray direction.

Figure 23.: Reflections projected on the floor and on the spheres (Vona and Pan).



Finally, the refracted light in a transmission surface is calculated by using the refractive index of the intersected primitive and the ray direction and shading normal.

The specular refraction is calculated as follows:

```

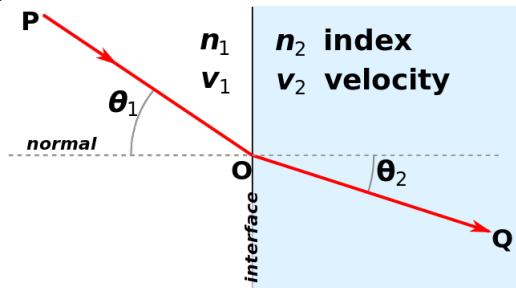
1 float refractiveIndiceInv = 1.0 / refractiveIndice
2 glm::vec3 refractDir = glm::refract(rayDirection, shadingNormal, refractiveIndiceInv)
3 Ray transmissionRay (refractDir, intersectionPoint, rayDepth + 1, intersectionPrimitive)
4 glm::vec3 LiT_RGB
5 rayTrace(&LiT_RGB, transmissionRay)
6 glm::vec3 specularTransmissionColor = kT * LiT_RGB

```

Listing 3.9: Algorithm of Specular Transmission

And, as usual, the transmission ray is traced and its radiance is multiplied by the transmission component of the intersected primitive.

Figure 24.: Refraction of light at the interface of two media of different refractive indices (Physics).



The algorithm developed is as following:

```

1 if (intersected material is diffuse) {
2   for each light source

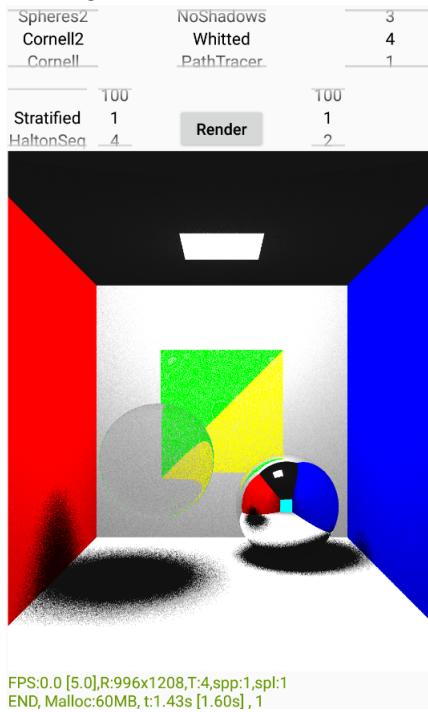
```

```

3   for each sample
4     vectorToLight = lightPosition - intersectionPoint
5     cos_N_L = vectorToLight . shadingNormal
6     if (cos_N_L > 0) {
7       Ray shadowRay(intersectionPoint, vectorToLight, distanceToLight, rayDepth + 1)
8       if (!shadowTrace(shadowRay)) rgb += radLight * cos_N_L
9     }
10    rgb *= kD
11    rgb /= #samples
12  }
13
14 if (intersected material is specular reflective) {
15   glm::vec3 reflectionDir = glm::reflect(rayDirection, shadingNormal)
16   Ray specularRay (reflectionDir, intersectionPoint, rayDepth + 1, intersectionPrimitive)
17   glm::vec3 LiS_RGB
18   rayTrace(&LiS_RGB, specularRay)
19   rgb += kS * LiS_RGB
20 }
21
22 if (intersected material is specular refractive) {
23   float refractiveIndiceInv = 1.0 / refractiveIndice
24   glm::vec3 refractDir = glm::refract(rayDirection, shadingNormal, refractiveIndiceInv)
25   Ray transmissionRay (refractDir, intersectionPoint, rayDepth + 1, intersectionPrimitive)
26   glm::vec3 LiT_RGB
27   rayTrace(&LiT_RGB, transmissionRay)
28   rgb += kT * LiT_RGB
29 }
```

Listing 3.10: Algorithm of Whitted Shader

Figure 25.: Whitted shader.



PathTracer

Finally, the last developed shader is the canonical Path Tracer which is a Monte Carlo method of rendering images of three-dimensional scenes such that the global illumination is faithful to reality. This renderer algorithm can synthesize very realistic scenes. Unlike the previous shaders, this one fully simulates both direct and indirect lighting. But, like the previous shader, this algorithm simulates the light reflected on diffuse and specular surfaces, as well as the light refracted on transparent primitives.

The light reflected on diffuse surfaces is divided in two parts: direct lighting and indirect lighting. The direct lighting is calculated in a similar way to the Whitted shader but with the difference that the samples are not taken from all sources of light but rather only one. The light source in each sample is chosen randomly. Then the obtained light radiance is multiplied by the Probability Density Function (PDF) in order to approximate the expected value. In this case the PDF is as simple as $1/\#lights$.

Finally, the indirect lighting reflected in diffuse surfaces is calculated in a much more complex way. First it is generated, from the intersected point, a random direction on an unit hemisphere with a PDF proportional to cosine-weighted solid angle ($PDF : p(\Theta) = \cos\theta/\pi$, $x = \cos(2\pi r_1)\sqrt{1-r_2}$, $y = \sin(2\pi r_1)\sqrt{1-r_2}$, $z = \sqrt{r_2}$). This kind of PDF is one of the best ways to render images with path tracing that the rendering equation converges to the expected value as fast as possible.

Then to generate a new ray direction in a world coordinates from a hemisphere it is needed to distinguish the six different possible materials:

- Diffuse reflection
- Glossy reflection
- Specular reflection
- Diffuse transmission
- Glossy transmission
- Specular transmission

The figure 27 illustrates those six possible scenarios.

To calculate the secondary ray direction in a diffuse reflective material, it was implemented a generator of random directions on unit hemisphere proportional to cosine-weighted solid angle around the normal of that hemisphere. Sampling in a cosine weighted hemisphere is better than uniform sampling over the hemisphere about the normal because the rays casted at the bottom of the hemisphere will not contribute as much as the other directions because of the multiplication by the cosine of theta.

To do this, it was implemented a method called `getCosineSampleHemisphere`, which generates a random direction around an hemisphere oriented by an arbitrary normal.

```
1 glm::vec3 getCosineSampleHemisphere(glm::vec3 normal);
```

Listing 3.11: Method in Shader which generates a random direction in an hemisphere.

The algorithm to generate a random direction in an hemisphere was taken from Wolfe, which is as follows:

```
1 float phi = glm::two_pi<float>() * uniformRandom1;// azimuthal angle
2 float r2 = uniformRandom2;// random distance from center
3 float cosTheta = std::sqrt(r2);// cos(theta) = cos(elevation angle)
4
5 glm::vec3 u = std::abs(normal.x) > 0.1 ? glm::vec3 {0.0, 1.0, 0.0} : glm::vec3 {1.0, 0.0, 0.0};
6 u = glm::normalize(glm::cross(u, normal));// second axis
7 glm::vec3 v = glm::cross(normal, u); // final axis
8
9 glm::vec3 newDirection = glm::normalize(u * (std::cos(phi) * cosTheta) + v * (std::sin(phi) * cosTheta)
10 + normal * std::sqrt(1.0 - r2));
```

Listing 3.12: Method in Shader which generates a random direction in an hemisphere.

The figure 26 shows the spherical coordinate system with the respective polar angle θ (theta) and the azimuthal angle ϕ (phi).

And the indirect lighting on diffuse reflective surfaces is then the multiplication of the ray traced light radiance with the material diffuse color of the primitive and π , and divided by the probability of continuing the Russian roulette. The set of equations 5 shows the simplification of the reflected diffuse color equation.

The resolution of diffuse reflected color of an object.

$$\cos(\theta) = \cos(\overrightarrow{\text{NewDirection}}, \vec{N})$$

$$\begin{aligned} PDF &= \cos(\theta)/\pi \\ PDF &= \cos(\overrightarrow{\text{NewDirection}}, \vec{N})/\pi \end{aligned} \quad (5)$$

$$\begin{aligned} LiD+ &= kD * RGB_{SecondaryRay} * \cos(\overrightarrow{\text{NewDirection}}, \vec{N}) / (PDF * ContinueProbability) \\ LiD+ &= kD * RGB_{SecondaryRay} * \pi / ContinueProbability \end{aligned}$$

Figure 26.: The spherical coordinates ([Wikipedia \(f\)](#)).

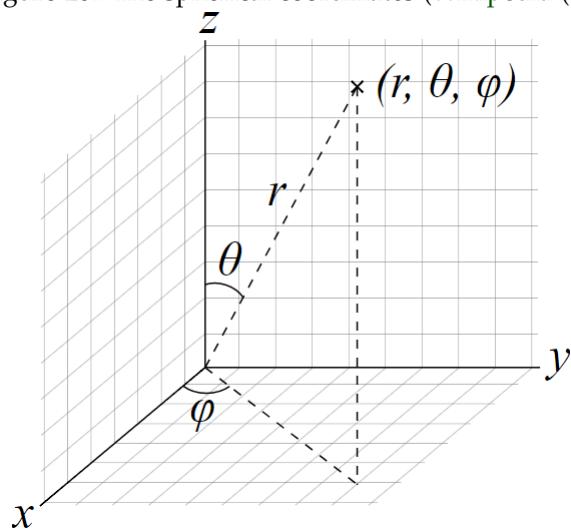
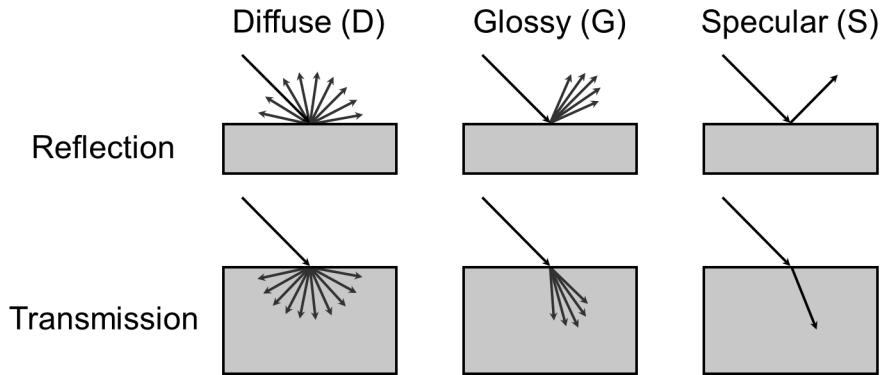


Figure 27.: The possible directions in secondary rays (Thompson).



The reflected and refracted light in a specular surface are simulated in a similar way to the Whitted algorithm presented previously.

So, the algorithm developed in Path Tracer Shader is as following:

```

1 if (intersected material is diffuse) {
2     for each light
3         for each sample
4             light = samplerLight.getSample()
5             vectorToLight = lightPosition - intersectionPoint
6             cos_N_L = vectorToLight . shadingNormal
7             if (cos_N_L > 0) {
8                 Ray shadowRay(intersectionPoint, vectorToLight, distanceToLight, rayDepth + 1)
9                 if (!shadowTrace(shadowRay)) rgb += radLight * cos_N_L
10            }
11            rgb *= kD
12            rgb *= #lights
13            rgb /= #samples
14
15        if (rayDepth <= RAY_DEPTH_MIN || uniform_dist(gen) > finish_probability) {
16            glm::vec3 newDirection = getCosineSampleHemisphere(shadingNormal)
17            Ray normalizedSecondaryRay (newDirection, intersectionPoint, rayDepth + 1,
18            intersectionPrimitive)
19
20            glm::vec3 LiD_RGB
21            intersectedLight = rayTrace(&LiD_RGB, normalizedSecondaryRay)
22
23        if (rayDepth > RAY_DEPTH_MIN) LiD /= continue_probability
24        if (Ld > 0 && intersectedLight) LiD = 0
25
26        rgb += LiD

```

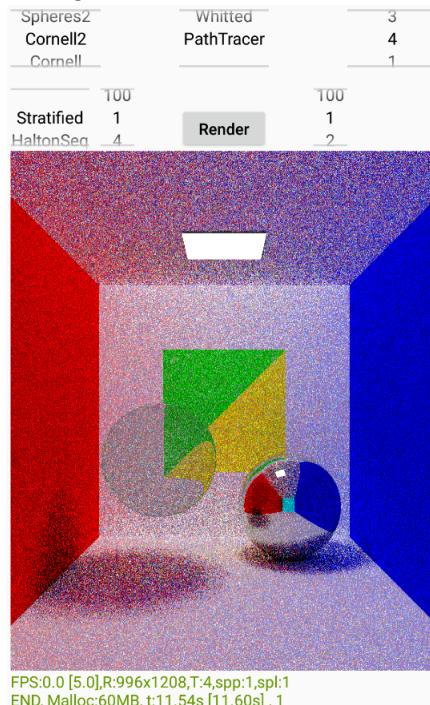
```

27 }
28
29 }
30
31 if (intersected material is specular reflective) {
32     glm::vec3 reflectionDir = glm::reflect(rayDirection, shadingNormal)
33     Ray specularRay (reflectionDir, intersectionPoint, rayDepth + 1, intersectionPrimitive)
34     glm::vec3 LiS_RGB
35     rayTrace(&LiS_RGB, specularRay)
36     LiS += kS * LiS_RGB
37 }
38
39 if (intersected material is specular refractive) {
40     float refractiveIndiceInv = 1.0 / refractiveIndice
41     glm::vec3 refractDir = glm::refract(rayDirection, shadingNormal, refractiveIndiceInv)
42     Ray transmissionRay (refractDir, intersectionPoint, rayDepth + 1, intersectionPrimitive)
43     glm::vec3 LiT_RGB
44     rayTrace(&LiT_RGB, transmissionRay)
45     LiT += kT * LiT_RGB
46 }

```

Listing 3.13: Algorithm of Path Tracer Shader

Figure 28.: PathTracer shader.



3.4.2 Samplers

There were implemented four samplers: Constant, Stratified, HaltonSequence and MersenneTwister.

All samplers just provide one method for the user to use:

```
1 float getSample(unsigned sampleNumber);
```

Listing 3.14: Main methods of Sampler

The *getSample* method receives as parameter the number of the current sample and returns the the actual sample. An atomic variable *sample* is used to count the current index of the samples.

Constant

This sampler is the simplest because it always returns the same number passed to the constructor.

This is used when the user only needs one sample per pixel and want all the samples to be in the middle of each pixel.

So, the algorithm of the *getSample* is just:

```
1 return value
```

Listing 3.15: Algorithm of Constant Sampler

Stratified

This sampler makes each sample at equal distance ($1/\text{domainSize}$) and in ascending order. For example, for a domain size of 4, the samples taken are going to be: 0, 0.25, 0.5 and 0.75.

This is useful for taking samples when supersampling without any noise, because each sample is not randomly chosen.

The algorithm of the *getSample* is then:

```
1 current = sample++ //atomic operation
2 if (current >= (domainSize * (sampleNumber + 1))) {
3     sample--
4     return 1
5 }
6 unsigned task = current - (sampleNumber * domainSize)
7 resSample = task / domainSize
8 return resSample
```

Listing 3.16: Algorithm of Stratified Sampler

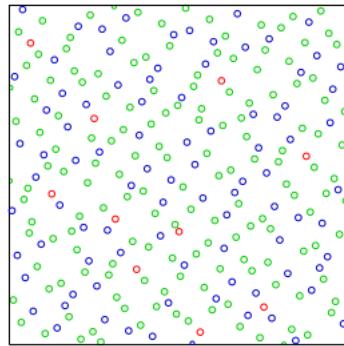
This algorithm ensures that if the current sample is greater than the domain size, then it will always return 1 which means the end of the sampler.

HaltonSequence

As the name implies, this sampler generates the Halton sequence.

Halton sequence is a quasi random number sequence which is a deterministic sequence with low discrepancy. These sequences are usually good for Rendering Algorithms like Path Tracing because it can make the rendering equation converge faster, that is, with fewer samples.

Figure 29.: Halton sequence in a 2D image plane ([Wikipedia \(d\)](#)).



The algorithm of the *getSample* is as following:

```

1 current = sample++ //atomic operation
2 if (current >= (domainSize * (sampleNumber + 1))) {
3     sample--
4     return 1
5 }
6 unsigned task = current - (sampleNumber * domainSize)
7 float resSample = MobileRT::haltonSequence(task, 2)
8 return resSample

```

Listing 3.17: Algorithm of HaltonSequence Sampler

As it can be seen in listing 3.17, most of the Halton Sequence Sampler is similar to the Stratified Sampler. The only difference is the call to the *haltonSequence* method where it computes the Halton sequence for an arbitrary index and base number.

The Halton sequence algorithm is very simple as shown in listing 3.18 and it was taken from [Wikipedia \(d\)](#).

```

1 float haltonSequence(unsigned index, unsigned base) {
2     float f = 1.0
3     float result = 0.0
4     while (index > 0) {
5         f /= base
6         result += f * (index % base)
7         index = std::floor(index / base)
8     }
9     return result;
10 }
```

Listing 3.18: Developed method of HaltonSequence

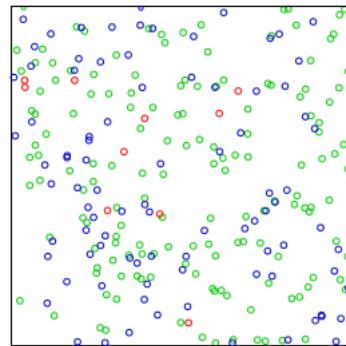
MersenneTwister

This sampler is just a wrapper to call the constructor of `std::uniform_real_distribution()` and the constructor of `std::random_device()` of the standard C++ library.

This allow to construct a random number generator which calculates a random number between 0 and 1 with an uniform distribution. This generator is a pseudo random number generator (PRNG), which is an algorithm for generating a sequence of numbers whose properties approximate the properties of sequences of random numbers. Although the PRNG-generated sequence is not truly random, because it is completely determined by an initial value, called the PRNG's seed, it commonly used a `std::random_device` for the generator seed because it produces non-deterministic random numbers, when supported. These sequences are usually used in simulations that use Monte Carlo methods like the Monte Carlo Ray Tracing.

It is a good PRNG because it produces uniformly distributed numbers, it doesn't repeat the same sequence and it does not exhibit correlation between successive numbers.

Figure 30.: Pseudorandom sequence in a 2D image plane ([Wikipedia \(d\)](#)).



So the algorithm of the *getSample* is just:

```

1 static std::uniform_real_distribution<float> uniform_dist {0.0, 1.0}
2 static std::mt19937 gen(std::random_device {}())
3 float res = uniform_dist(gen)
4 return res

```

Listing 3.19: Algorithm of MersenneTwister Sampler

As can be seen in listing 3.19, the algorithm is very simple because it just constructs two objects in the first time the method is called:

1. `std::uniform_real_distribution` (`paramA`, `paramB`) - which produces random floating-point values, uniformly distributed on the interval $[paramA, paramB]$, that the user specifies.
2. `std::mt19937` (`std::random_device ()`) - pseudo-random generator of 32-bit numbers with a state size of 19937 bits.

Then every time the user calls *getSample*, the MersenneTwister Sampler just calls the `operator()` of `std::uniform_real_distribution` generator, which generates the next random number in the distribution.

3.4.3 Lights

There were implemented two types of light sources: point light and area light. These light sources provide the user two main methods:

```

1 glm::vec3 getPosition()
2 Intersection intersect(Intersection intersection, Ray ray)

```

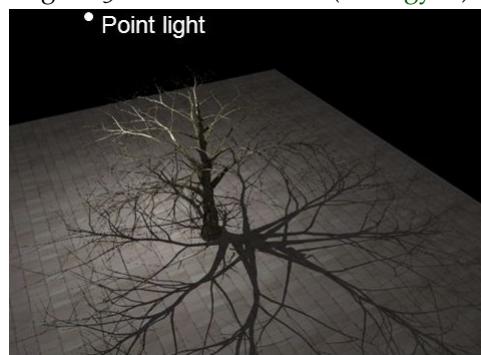
Listing 3.20: Main methods in Light

The *getPosition* method just returns the position of the light source and the *intersect* method determines whether a given *ray* as a parameter intersects this light source and, if it intersects, writes the result to the *intersection* parameter.

Point light

The Point light is the simplest form of a light because, as the name implies, it is just a point of light which emits light in all directions at once. This type of light is useful to render fast shadows because it just simulates hard shadows, in that each shadow is only constituted with an umbra, which is the fully dark shadow.

Figure 31.: Hard shadows (Barsegyan).



Therefore, the *getPosition* method is very simple because it only returns the position of the light source determined in its constructor:

```
1 return lightPosition
```

Listing 3.21: Algorithm of getPosition in Point light

And the *intersect* method is also quite simple because the ray parameter never intersects the point light. This is because the probability of a ray intersecting a single 3D point in the world is practically null.

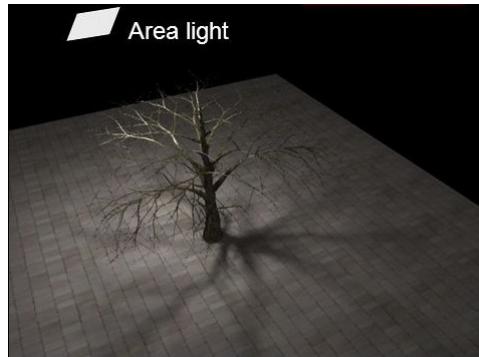
```
1 return intersection
```

Listing 3.22: Algorithm of intersect in Point light

Area light

The Area light is another type of light source where the light comes from an area. It can have many different shapes, like square, triangle, circle, etc., and this allows the simulation of soft shadows like the figure 32.

Figure 32.: Soft shadows (Barsegyan).



The Area light implemented for this application has a shape of a triangle. This shape is intended as the triangle is the simplest shape which allows to construct more complex shapes. And, with barycentric coordinates, it is very easy to generate a random point in the triangle.

A triangle with three points: A, B and C. We get vectors AB and AC, being:

1 $\text{AB} = [\text{Bx} - \text{Ax}, \text{By} - \text{Ay}, \text{Bz} - \text{Az}]$ and $\text{AC} = [\text{Cx} - \text{Ax}, \text{Cy} - \text{Ay}, \text{Cz} - \text{Az}]$.

Listing 3.23: Vectors AB and AC in a triangle

These vectors tell us how to get from point A to the other two points in the triangle, by telling us what direction to go and how far. So, with barycentric coordinates $R=1/3$, $S=1/3$ and $T=1/3$, we get the point in the center of the triangle. To generate a random point in the triangle, we have to generate two random numbers between 0 and 1 (R and S). Then we have to make sure we stay inside the triangle by checking if they are larger than one:

```
1 if (R + S >= 1) {
2   R = 1 - R
3   S = 1 - S
4 }
```

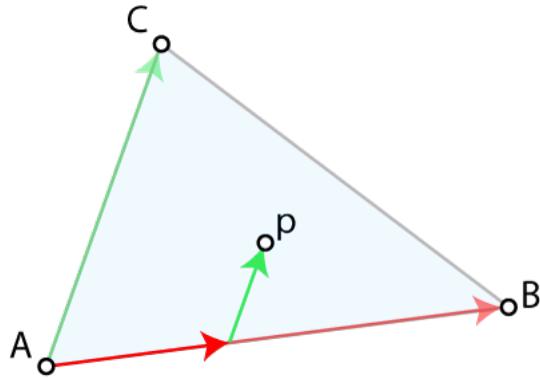
Listing 3.24: Algorithm of Area light

Finally we can obtain a random point in the triangle by starting at point A, then getting a random percentage along vector AB and a random percentage along vector AC:

1 $\text{RandomPointPosition} = \text{A} + \text{R} * \text{AB} + \text{S} * \text{AC}$

Listing 3.25: Algorithm of Area light

Figure 33.: Calculation of point P by using barycentric coordinates starting at point A and adding a vector AB and a vector AC (Jacobson).



So, the *getPosition* algorithm is as following:

```

1 float R = samplerPointLight->getSample()
2 float S = samplerPointLight->getSample()
3 if (R + S >= 1.0) {
4     R = 1.0 - R;
5     S = 1.0 - S;
6 }
7 glm::vec3 position = triangle.pointA + R * triangle.AB + S * triangle.AC
8 return position

```

Listing 3.26: Algorithm of getPosition in Area light

And the *intersect* method just calls the intersection of a ray with a triangle method presented earlier, and if it intersects updates the material.

```

1 float lastDist = intersection.length
2 intersection = triangle.intersect(intersection, ray)
3 intersection.material_ = intersection.length < lastDist? &radiance : intersection.material
4 return intersection

```

Listing 3.27: Algorithm of intersect in Area light

3.4.4 Cameras

Only two types of cameras were implemented: perspective and orthographic cameras.

The camera only provides one method for the user to cast a ray from the camera position in direction to the image plane:

```
1 Ray generateRay(float u, float v, float deviationU, float deviationV);
```

Listing 3.28: Main methods in Camera

The method *generateRay* needs four parameters to create a ray. The *u* and *v* are used to choose the targeted pixel in the image plane. Being *u* the inverse of the index of the pixel in its line, that is, x/width , and *v* the inverse of the index of the pixel in its column, that is, y/height . In order to allow the reduction of aliasing in the generated images of the scene, the camera also accepts two extra parameters *deviationU* and *deviationV* which are variances inside a pixel. The *deviationU* is a horizontal variance of the pixel, that is, $[-0.5 * \text{pixelWidth}, 0.5 * \text{pixelWidth}]$ and *deviationV* the variance in the vertical of the pixel $[-0.5 * \text{pixelHeight}, 0.5 * \text{pixelHeight}]$. This technique is called jittering, as shown in figure 34, and allows to reduce the aliasing effect by introducing noise into the output image. Although the final image gets some noise, this effect turns out to be visually more appealing than aliasing because it is an effect without patterns that are easily detectable by the human eye. It is important to reduce the aliasing effect because, it is an effect that the human eyes can detect very fast because the generated image will have quite regular patterns, as shown in figure 35.

Figure 34.: Illustration of how sampling the plane image with jittering works (Waters).

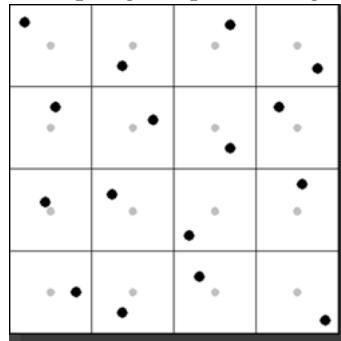
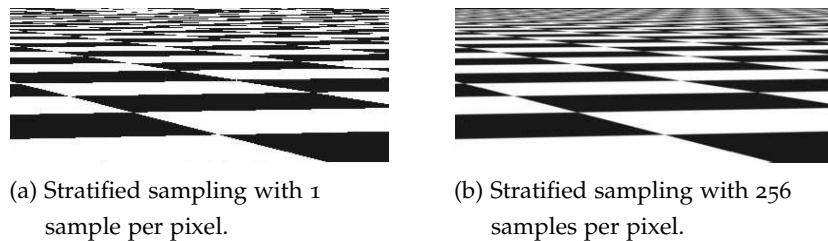


Figure 35.: Aliasing effect (Science and Engineering).



Sampling with jittering is, as previously stated, a good technique to avoid aliasing in the image. The figure 36 shows an example of stratified sampling with 1 and 256 samples per pixel. As can be seen in the figure on the left, there is a noticeably aliasing in the image, and we can't even perceive the correct positions of the black squares on the background. Of course, with 256 samples per pixel, the aliasing effect is greatly reduced, but also, the execution time is linearly increased, which is a downside to the user experience.

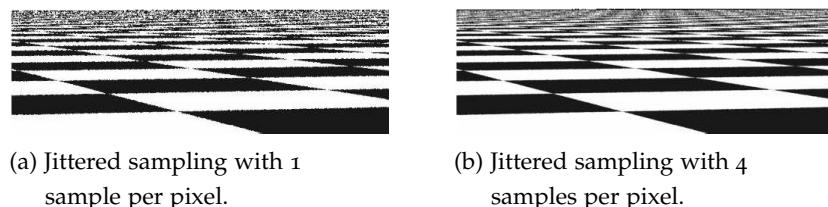


(a) Stratified sampling with 1 sample per pixel.

(b) Stratified sampling with 256 samples per pixel.

Figure 36.: Stratified sampling.

Sampling with jittering is, obviously not perfect, as shown in the figure 37. But, it produces more visually appealing images, although it introduces some noise. Even, with just 1 sample per pixel, the generated image is less "jaggy" in the background. And, with 4 samples per pixel, the noise is greatly reduced and its quality is visually comparable to the stratified sampling with 256 samples per pixel. This means that by using jittering it is possible to obtain images pleasing to the human eye with fewer samples per pixel, and consequently in less execution time.



(a) Jittered sampling with 1 sample per pixel.

(b) Jittered sampling with 4 samples per pixel.

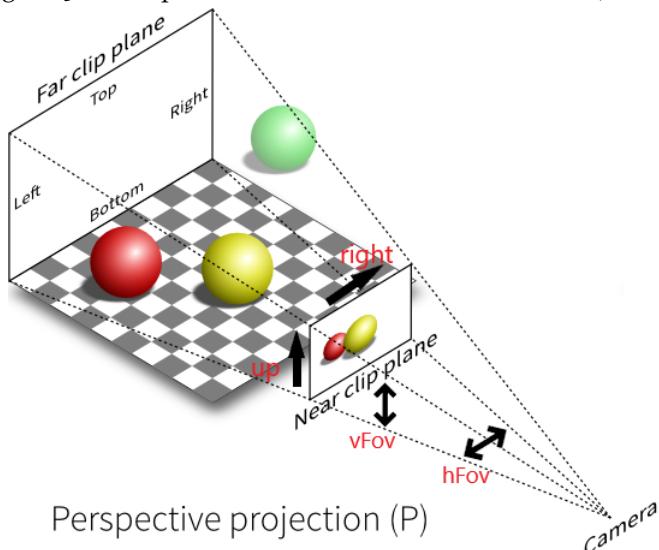
Figure 37.: Jittered sampling.

Perspective Camera

This type of camera is the most used in renderers because it uses perspective projection. With it, it can simulate images being seen by the human eyes, which means, it simulates the depth of the objects, and produces 2D images with a 3D projection.

To obtain an image plane with perspective, it is necessary to have a Field of View. In order to accept any resolution of the image plane, we have to divide the field of view in two parts: horizontal and vertical. This way, we can obtain the aspect ratio of the image plane we want.

Figure 38.: Perspective camera with hFov and vFov (Schim).



The algorithm to generate a ray from the perspective camera is very simple. By knowing the horizontal Field of View and vertical Field of View in radians, and with u and v as parameters, it is possible to calculate the direction of that ray.

It starts to calculate the distance to go through the right vector and the up vector of the camera. That can be done with the arctangent of each Field of View of the camera (horizontal and vertical) and multiplying it with $u - 0.5$ in the right vector and with $0.5 - v$ in the up vector. This makes sure that we go through every pixel, starting with the pixel from the top left corner to the bottom right corner of the image plane. Then the destination point of the ray is just the sum:

$destinationPoint = cameraPosition + cameraDirection + rightVector + upVector$, and so its direction is just:

$rayDirection = destinationPoint - cameraPosition$. The origin of the ray is the position of the camera, because in a perspective camera, all rays come from the same point: the point where the camera is located.

```

1 float rightFactor = arctan(hFov * (u - 0.5)) + deviationU
2 glm::vec3 rightVector = right * rightFactor
3 float upFactor = arctan(vFov * (0.5 - v)) + deviationV
4 glm::vec3 upVector = up * upFactor
5 glm::vec3 destinationPoint = cameraPosition + cameraDirection + rightVector + upVector
6 glm::vec3 rayDirection = destinationPoint - cameraPosition
7 Ray ray = Ray(glm::normalize(rayDirection), cameraPosition)
8 return ray

```

Listing 3.29: Algorithm of generateRay in Perspective Camera

Orthographic Camera

In this projection mode, an object's size in the rendered image stays constant regardless of its distance from the camera. This can be useful for rendering 2D scenes and UI elements, amongst other things.

The orthographic camera removes the sense of perspective by drawing the image plane without simulating the depth of the objects, making all the objects looking flat. This is achieved by inverting the logic of the perspective camera. Instead of calculating the direction and always having the same origin, in the orthographic camera, the direction is always the same for all rays, and the origin of the ray varies.

It starts to calculate the distance to go through the right vector and up vector of the camera, like the perspective camera. But, instead of using the Field of View, we now use the *sizeH* and *sizeV*, which are the horizontal and vertical sizes of the image plane. Then, to make sure that we go through all pixels from top left pixel to the bottom right, we need to use the *u* and *v* values from the parameters. Then, the origin of the ray is: *rayOrigin = cameraPosition + rightVector + upVector*.

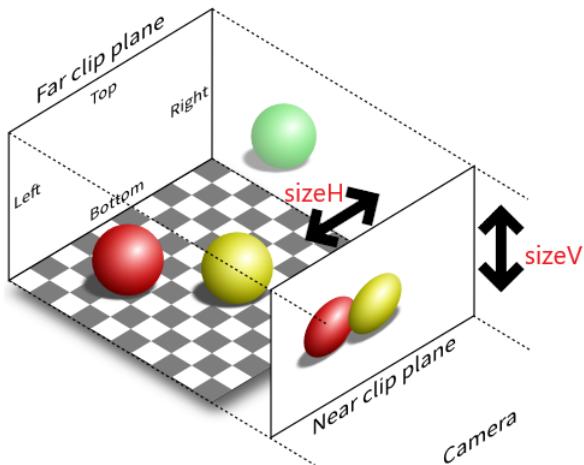
```

1 float rightFactor = (u - 0.5) * sizeH
2 glm::vec3 rightVector = right * rightFactor + right * deviationU
3 float upFactor = (0.5 - v) * sizeV
4 glm::vec3 upVector = up * upFactor + up * deviationV
5 glm::vec3 rayOrigin = cameraPosition + rightVector + upVector
6 Ray ray = Ray(cameraDirection, rayOrigin)
7 return ray

```

Listing 3.30: Algorithm of generateRay in Orthographic Camera

Figure 39.: Orthographic camera with sizeH and sizeV (Schim).



Orthographic projection (O)

3.4.5 Object Loaders

Last, but not least, the library allows the programmer to develop his / her own Object Loaders. An Object Loader is, as the name implies, a class that allows the ray tracer to import meshes from different Model file formats. There are many different types of file formats for storing meshes:

- 3ds
- FBX
- Wavefront .obj file
- COLLADA
- SketchUp
- AutoCAD DXF

All these file formats allows the user to store the positions of many triangles in a file, which together form the geometry of the scene.

This library only provides one Object Loader that allows loading the scene geometry from Wavefront obj files, which was achieved using a third party library called "tinyobjloader". This type of Model file format is a good choice because, besides being simple, it is open and has been adopted by many 3D graphics application vendors, like, 3ds Max and Blender.

OBJ Loader

The class "OBJLoader" is a Rendering Component that allows the user to load 3D scenes from Wavefront obj files.

The OBJ file format is a simple file data format that allows to store only the position of each vertex, the UV position of each texture coordinate vertex, each vertex normal, and the faces that make each polygon defined as a list of vertices, and texture vertices. Vertices are stored in a counter-clockwise order by default, making explicit declaration of face normals unnecessary.

```

1 # This is a comment
2
3 # List of geometric vertices
4 # x y z w
5 v 1.0 2.0 3.0 1.0
6 v 5.0 6.0 7.0 1.0
7 ...
8
9 # List of texture coordinates
10 # x y w
11 vt 0.5 1.0 0.0
12 vt 0.1 0.3 0.0
13 ...
14
15 # List of vertices normals
16 # x y z
17 vn 0.3 0.0 0.7
18 vn 0.2 1.0 0.3
19 ...
20
21 # List of polygonal face elements
22 # v1/vt1/vn1 v2/vt2/vn2 v3/vt3/vn3
23 f o3/o1/o1 o4/o2/o3 o5/o3/o4
24 f o6/o6/o3 o7/o1/o2 o8/o2/o6
25 ...

```

Listing 3.31: .OBJ file format

The Wavefront OBJ file format is described as shown in [3.31](#), and is accompanied by another file, Material Template Library (MTL), that describes the visual aspects of the polygons. It is this file that describes surface shading (material) properties of objects within one or more .OBJ files. It allows to define multiple materials, with, ambient color, diffuse

color, specular reflective color and specular refractive color. And even allows to load texture maps stored in TGA files. The listing 3.32 shows an example of a description of a material.

```
1 # This is a comment
2
3 # Define a material named 'Colored'
4 newmtl Colored
5
6 # Ambient color
7 Ka 1.0 0.0 0.0 # red
8
9 # Diffuse color
10 Kd 0.0 1.0 0.0 # green
11
12 # Specular color
13 Ks 0.0 0.0 1.0 # blue
14
15 # Dissolved (transparency)
16 d 0.4
```

Listing 3.32: .MTL file format

4

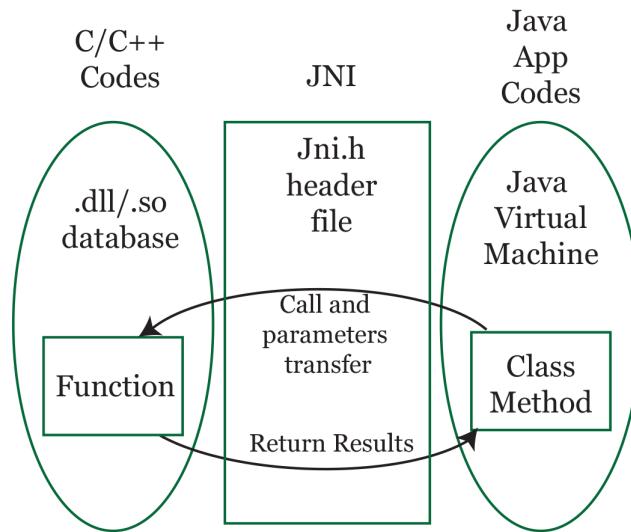
ANDROID CHALLENGES

4.1 ANDROID SPECIFICS

Before developing the ray tracer to Android, it is necessary to understand how an Android application works.

A typical Android application is programmed in Java programming language. And it usually needs to communicate with an User Interface which also is programmed in Java. The code is compiled with Android Software Development Kit (SDK) along with any data and resource files into an Android package (APK). But, in order to avoid using the Java Virtual Machine (JVM) in our code and program it in native code, it is necessary to use Android Native Development Kit (NDK). Unfortunately, the NDK tool-chain doesn't provide any User Interface libraries like Qt, GTKmm, or wxWidgets, that facilitates the build of a User Interface in native code. So, in order to obtain the best performance possible in a mobile device, the ray tracer library and components were programmed in C++ and the User Interface was programmed in Java. The User Interface calls the ray tracer available methods using the Java Native Interface (JNI) which is a programming standard that allows Java code running on a JVM to call native applications (programs specific to a hardware platform and operating system) and libraries written in other languages such as C, C ++ and assembly. The use of JNI to call native applications is not guaranteed to have better performance than having all the application written in Java, because each JNI call have a performance cost associated with the fact that the JVM needs to use call external native code. But, in this case as the ray tracer can be a very computing demanding application, the fact it is written in native code (C++) is better than having all the code written in Java and executed by the JVM.

Figure 40.: Diagram explaining how the JVM can call native code (Krajci and Cummings).



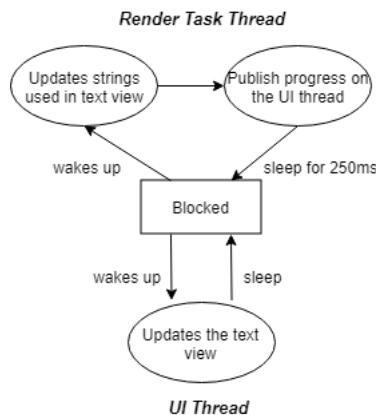
4.2 USER INTERFACE

The Android User Interface is programmed in Java and only one thread can refresh the User Interface (UI thread).

There are only two goals in the User Interface: should be as simple as possible and should be able to allow progressive ray tracing, which means, refreshing the image while it is being rendered.

In order to allow progressive ray tracing, it was used a pool of one thread called Render Task thread that every 250 ms wakes up the thread and updates the strings used in the text view. Before it finishes, it publishes the progress on the UI thread. Then the UI thread wakes up and updates the text view.

Figure 41.: Execution flow of UI thread and Render Task thread.



4.3 PROGRAMMING DECISIONS

The main goal of programming this library in an object oriented manner is to have a library with relatively good performance and with a maintainable code.

Besides that, it was used the smart pointers available in the C++ standard library. The smart pointer used was the unique pointer, as this was enough for this project and also because it is the smart pointer with best performance. This smart pointer automatically free the resources it contains. This greatly reduces the chance of the programmer to forget to free a resource or trying to free it multiple times. Of course this comes with a little performance cost of about 1% to 5% relatively to the manual resource management by the programmer. But for this project, some extra code correctness is more important than more performance.

4.3.1 *Android benefits*

As mentioned above, this dissertation focuses only on ray tracing in Android. This is intentional because, as also mentioned above, it is the Operating System for mobile devices with the most market share.

Developing apps in Android it's cheaper than in other platforms. The company provides its SDK for free, so all the costs are destined to the app testing and deploying. That means you don't have to make a big investment in that part. Investing less on your app development means that you'll have a higher ROI and your project will be more profitable.

Software developers, commonly learn to develop in Java, and it is easier for them to adapt to that programming language for mobile app development. Android is built mostly in Java, so its adaptation becomes faster and easier.

4.3.2 *Android challenges*

Developing applications for mobile devices have different challenges compared with the traditional personal computer hardware.

The Android User Interface has some particularities like only one thread can modify the UI, and is usually called the UI thread.

The size of RAM available for executing applications is typically smaller. In the 2010s the most expensive mobile devices had only about 512MB of RAM, and nowadays most of them have more than 4GB. But, still, comparing them with personal computers which typically have 16GB it is a big difference. This can make a big impact in the maximum number of vertices that a scene can have in the ray tracer.

And the CPU microarchitecture is generally simpler and with smaller computational power. Also the Operating System (OS) is shaped for the mobile world, making a lot of restrictions in the performance of the applications in order to save battery. The amount of main memory available for the applications can also be affected by the OS.

Other challenges are related with the communication mechanism between the SDK and the NDK, because two different languages environments need to "communicate" in runtime (GUI in Java and the library in C++). This involves learning how to use the Java Native Interface (JNI), so that the native code can send and receive information from Java Virtual Machine (JVM).

Because ray tracing may involve rendering complex scenes, placing all this information in the main memory may not be possible, and so this memory management can be a worthy challenge.

5

DEMONSTRATION: GLOBAL ILLUMINATION

5.1 RESULTS OBTAINED

The developed application was tested in three different devices:

- Samsung Galaxy Fresh Duos GT-S7392
- Raspberry Pi 2 Model B
- MINIX NEO X8-H PLUS (k200)

Table 3.: Devices specifications.

Device	CPU	Cache(L1/L2/L3)	GPU	RAM
Samsung Galaxy Fresh	1xARM Cortex A9 @1GHz	64KB/Unknown	1xBroadcom VideoCore IV	512MB
Raspberry Pi 2 Model B	4xARM Cortex A7 @900MHz	64KB/1MB	1xBroadcom VideoCore IV	1GB
MINIX NEO X8 PLUS	4xARM Cortex A9r4 @2.0GHz	64KB/Unknown	8xMali-450	2GB

As you can see in the table 3, the application has been tested on a few devices. Unfortunately, it was only possible to test it on one mobile device, the Samsung Galaxy Fresh Duos GT-S7392. This device is a low-end smartphone with a low-end single core CPU.

It was also possible to test it on two different computers that are portable and even smaller than a common laptop. The Raspberry Pi 2 Model B and the MINIX NEO X8-H PLUS which are devices with the Android operating system installed.

In order to assess the performance of the ray tracer, it was tested with three different scenes and with Whitted and Path Tracing shaders. One large scene with some hundred of thousands triangles, one moderate with few thousand of triangles and one small scene with just some dozens triangles.

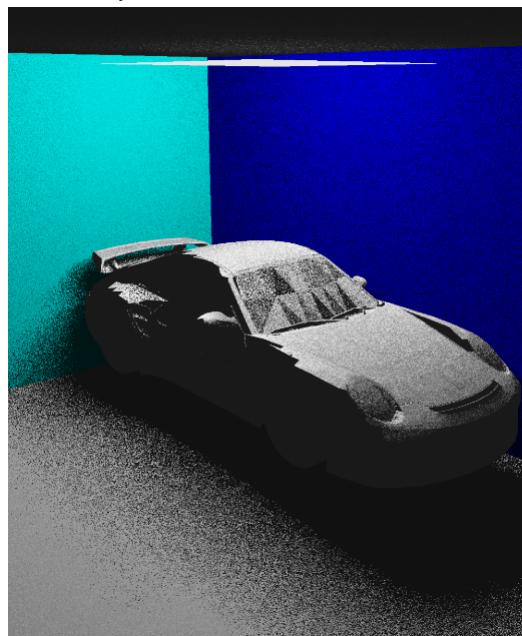
The large scene used for testing was the conference scene, as illustrated in figure 42. This scene consists of 331179 triangles and has two area lights in form of triangles.

Figure 42.: Illustration of conference scene, rendered with shader Whitted in MobileRT.



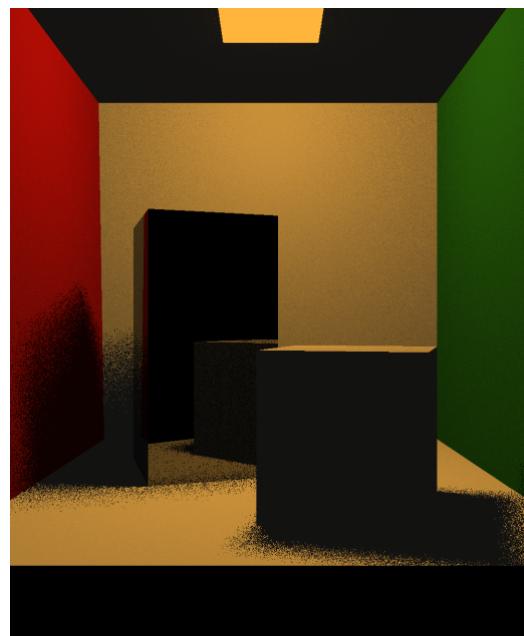
The medium scene used for testing was the Porsche 911 GT2 scene, as illustrated in figure 43. This scene consists of 22017 triangles and has two area lights in form of triangles.

Figure 43.: Illustration of Porsche 911 GT2 scene, rendered with shader Whitted in MobileRT.



The small scene used for testing was the Cornell Box scene, as illustrated in figure 44. This scene consists of 34 triangles and has two area lights in form of triangles.

Figure 44.: Illustration of Cornell Box scene, rendered with shader Whitted in MobileRT.



All of the used scenes models were downloaded from [http://casual-effects.com/
data/index.html](http://casual-effects.com/data/index.html) (McGuire (2017)).

5.1.1 Whitted Shader

Samsung Galaxy Fresh Duos GT-S7392

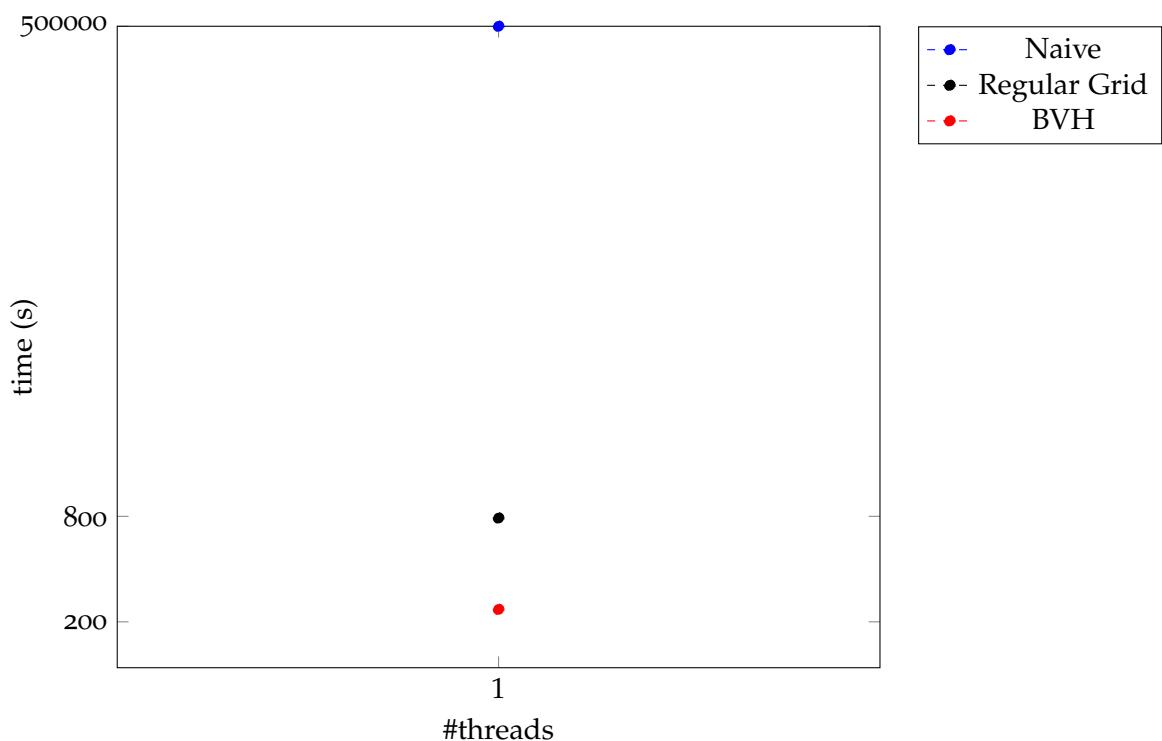


Figure 45.: Time in the Conference scene

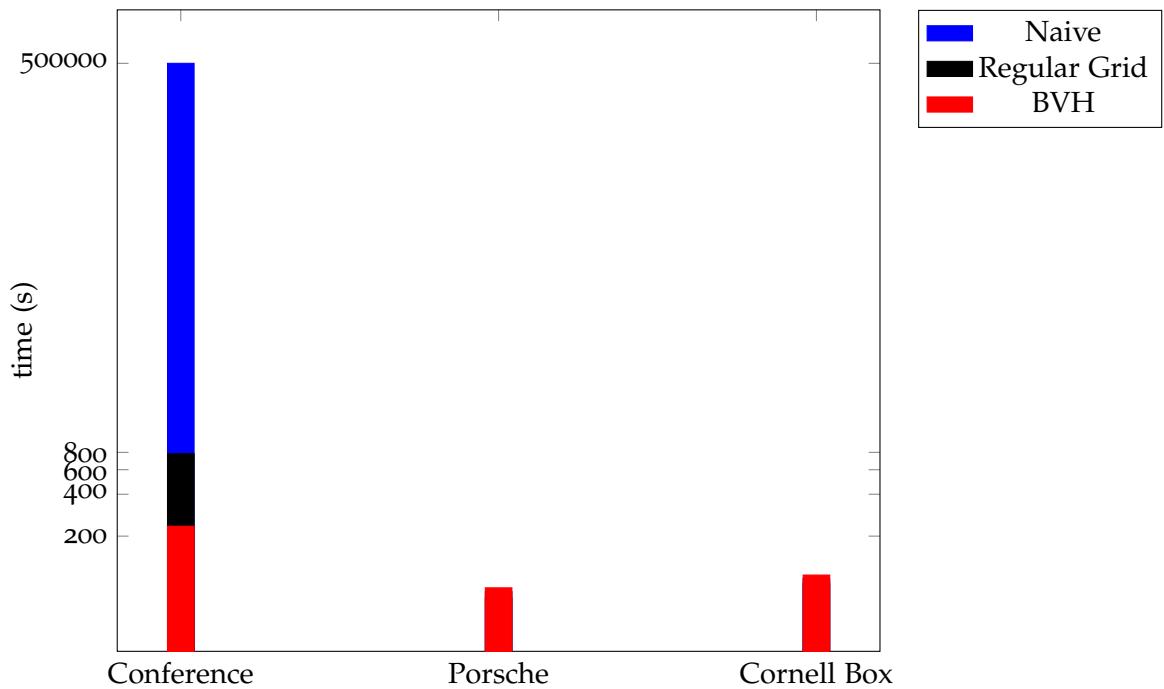


Figure 46.: Time with Samsung smartphone

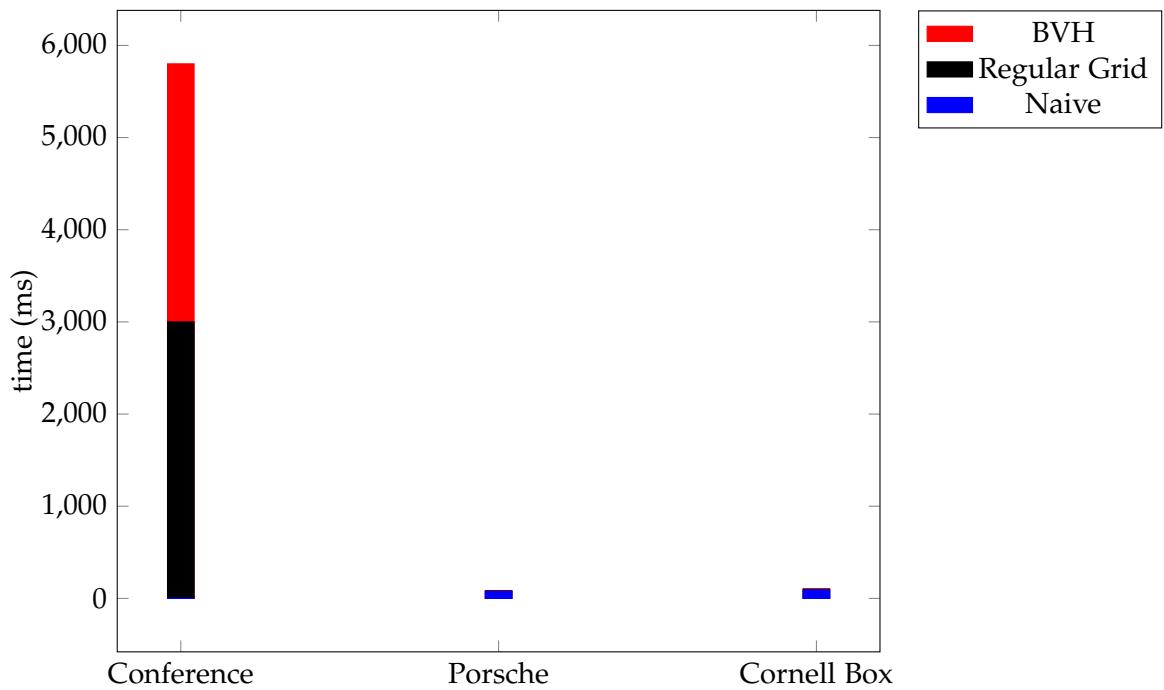


Figure 47.: Time to construct accelerators in Samsung Galaxy smart phone

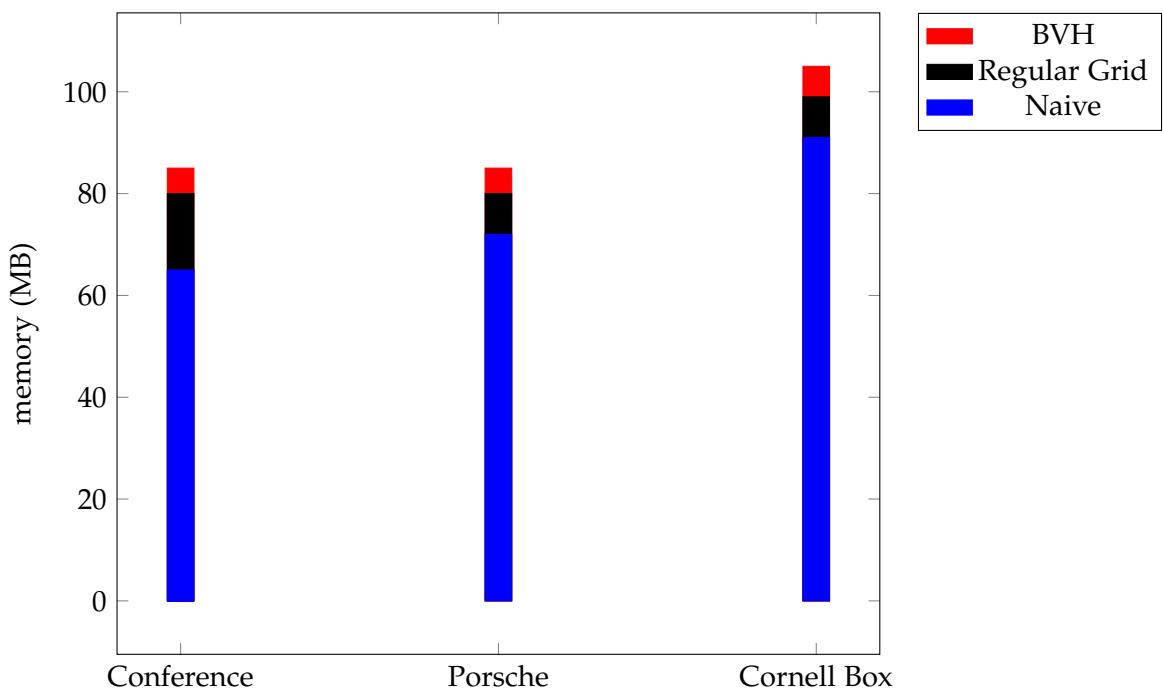
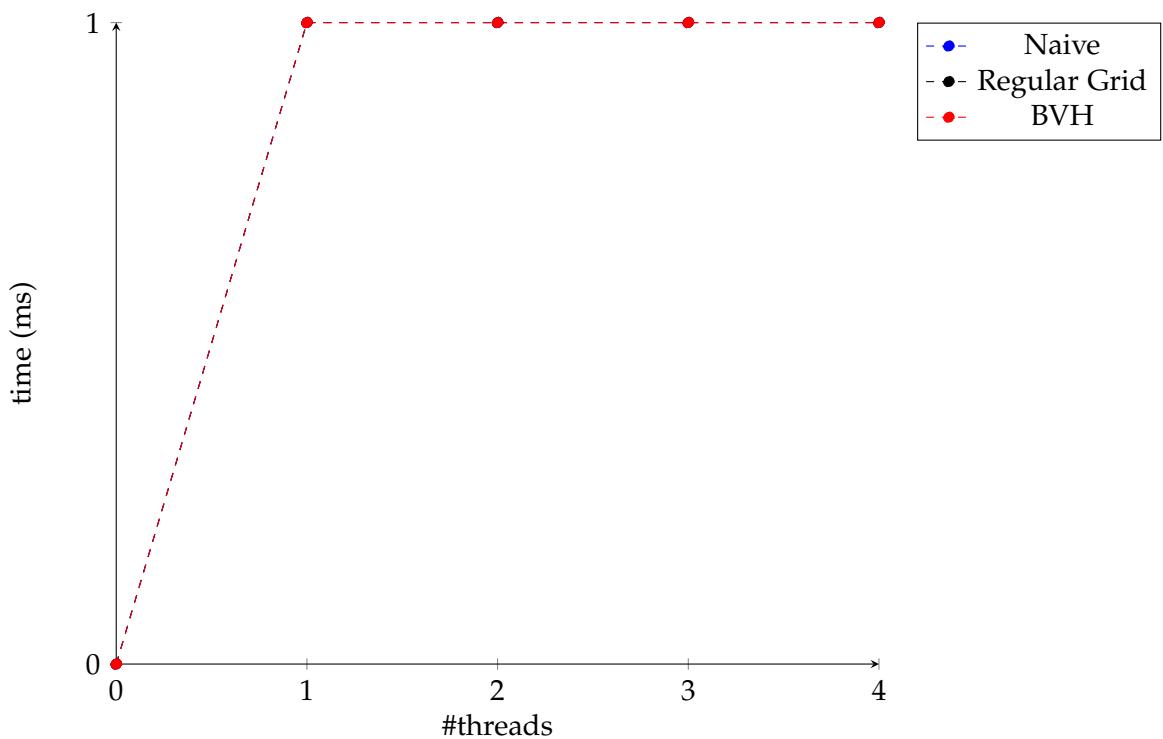


Figure 48.: Memory usage in the Samsung Galaxy smart phone

Raspberry Pi 2 Model B



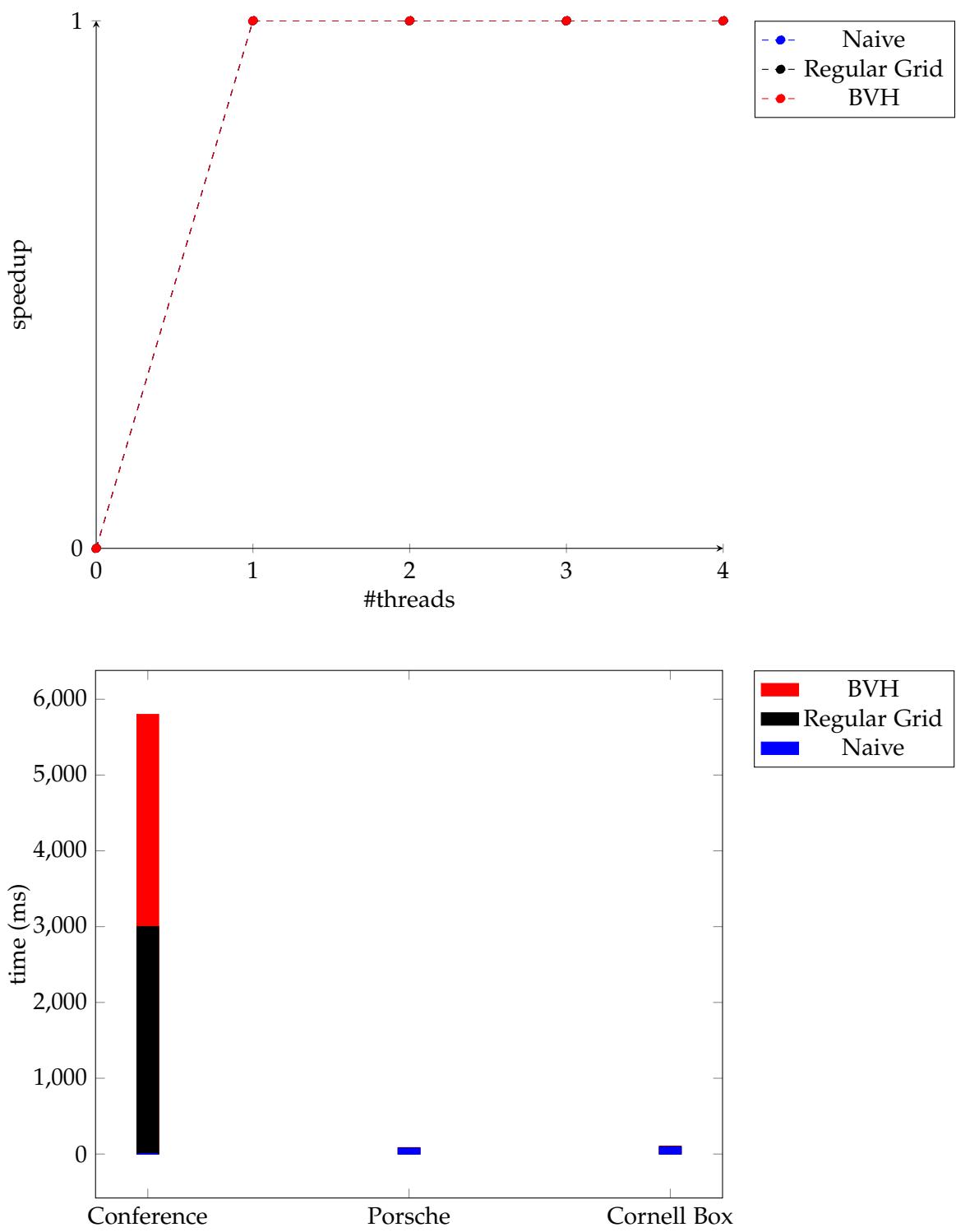


Figure 49.: Time to construct accelerators in the Raspberry Pi 2 Model B

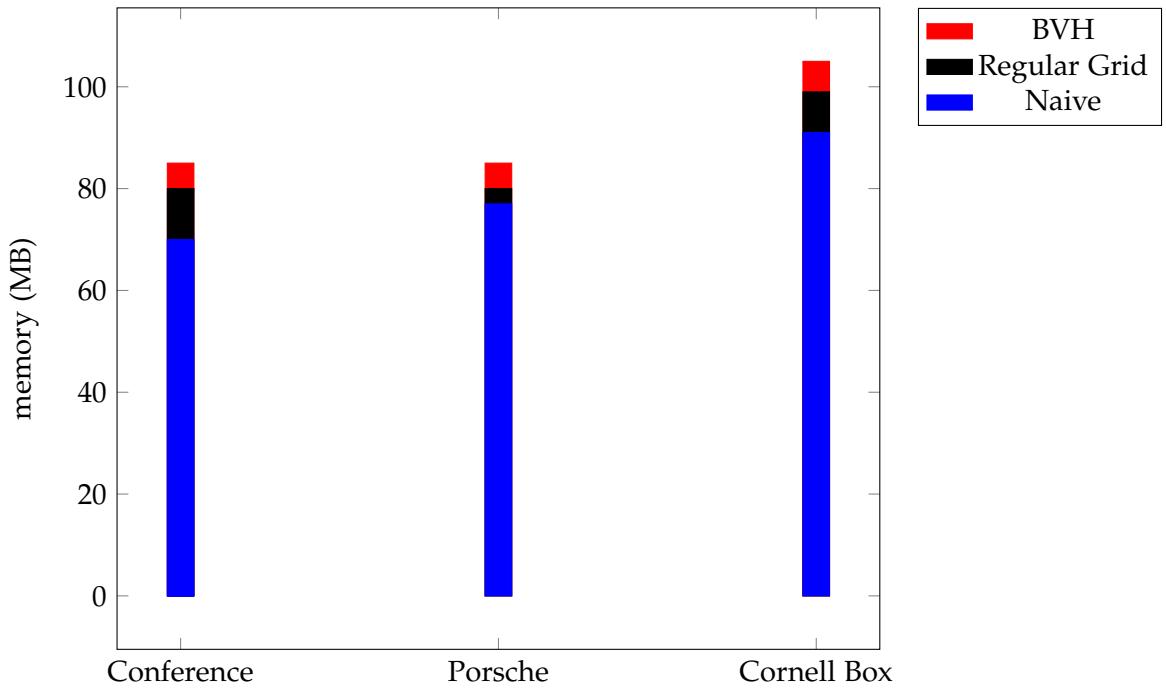
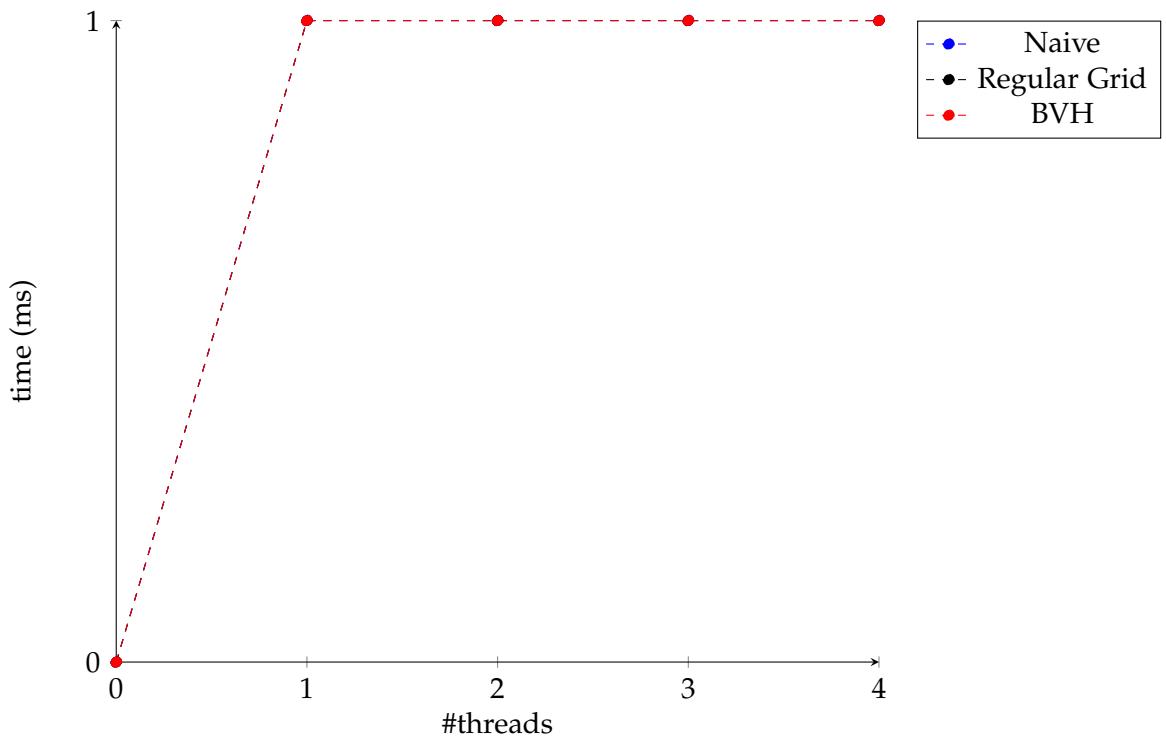


Figure 50.: Memory usage in the Raspberry Pi 2 Model B

MINIX NEO X8-H PLUS



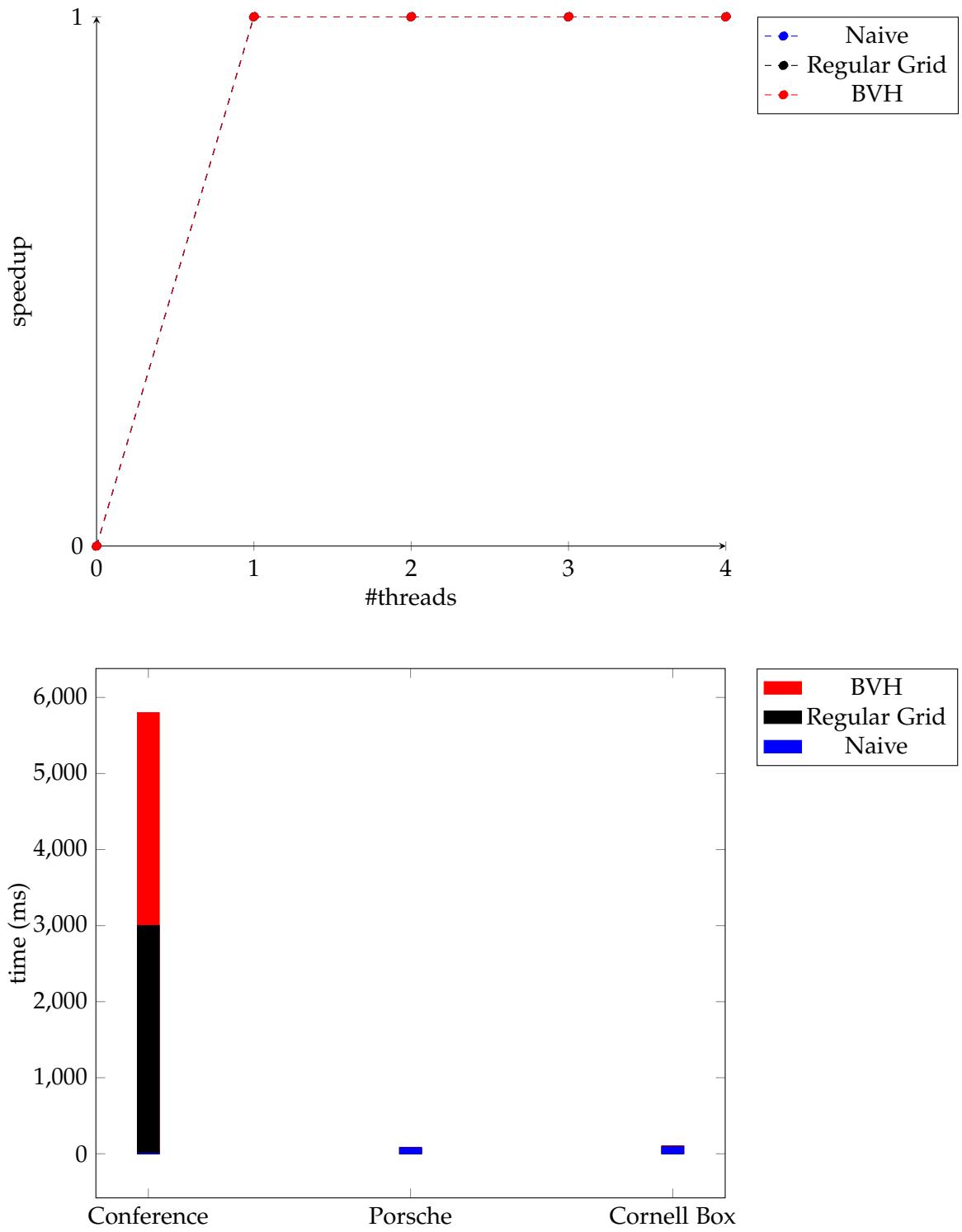


Figure 51.: Time to construct accelerators in the MINIX NEO X8-H PLUS

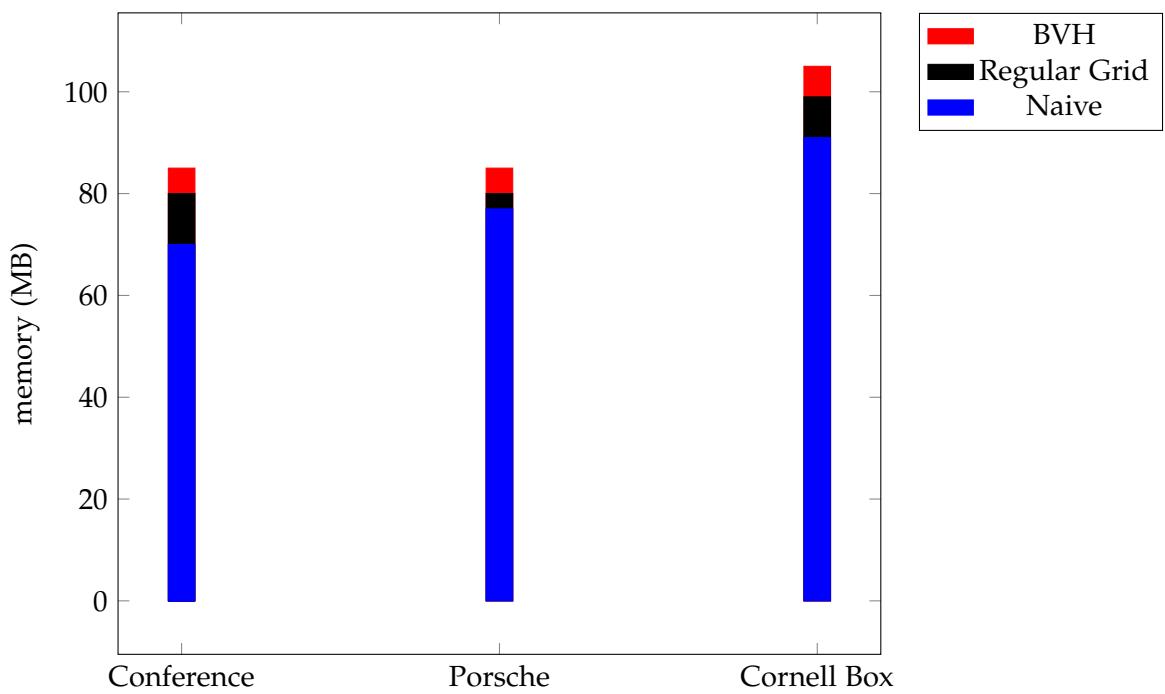


Figure 52.: Memory usage in the MINIX NEO X8-H PLUS

5.1.2 Path Tracing Shader

Samsung Galaxy Fresh Duos GT-S7392

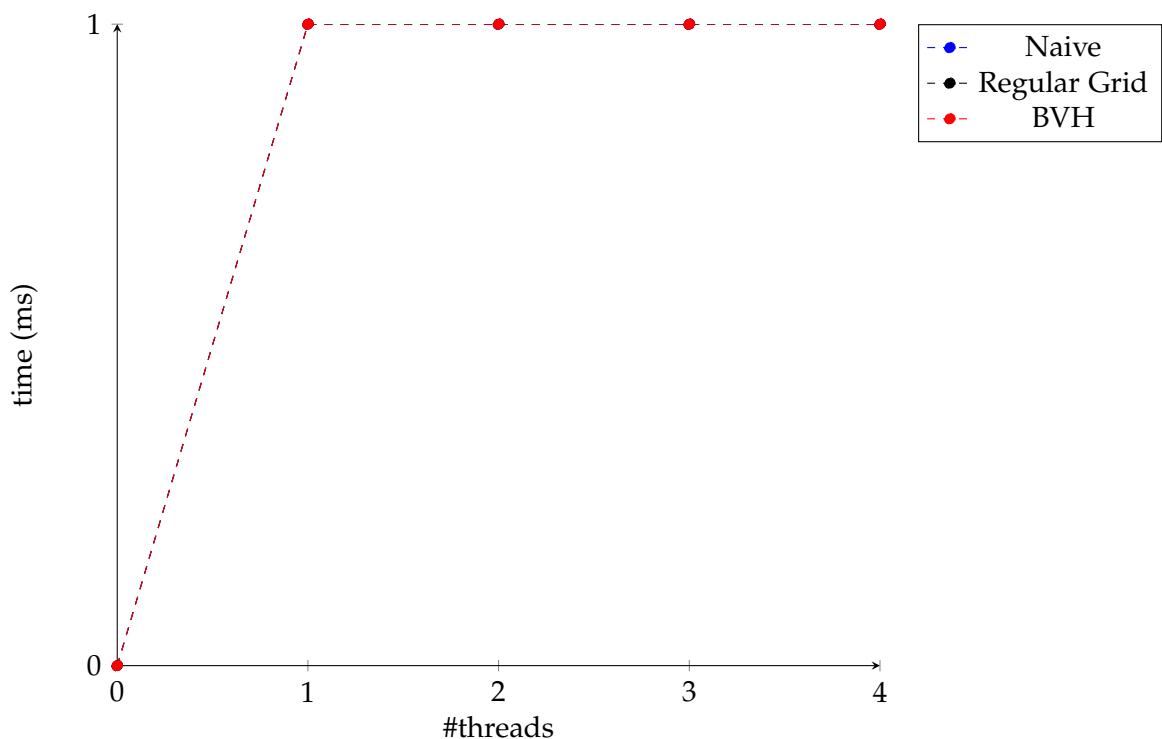


Figure 53.: Time in the Conference scene

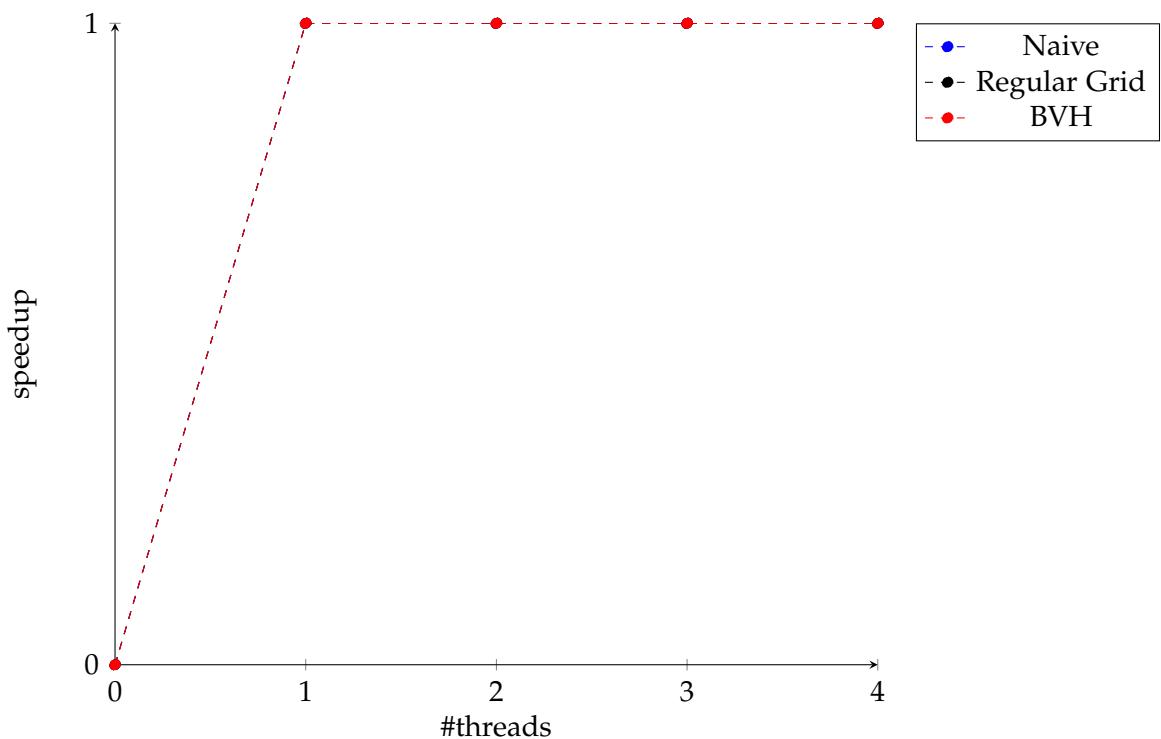


Figure 54.: Speedup in the Conference scene

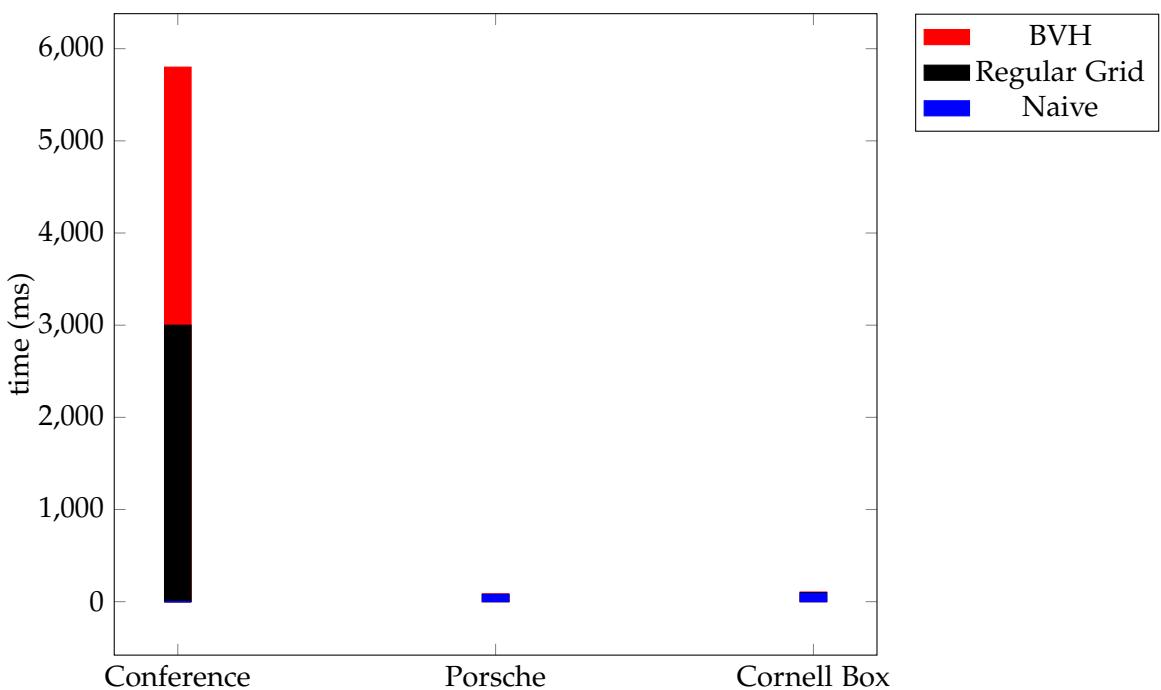


Figure 55.: Time to construct accelerators in the Samsung Galaxy smart phone

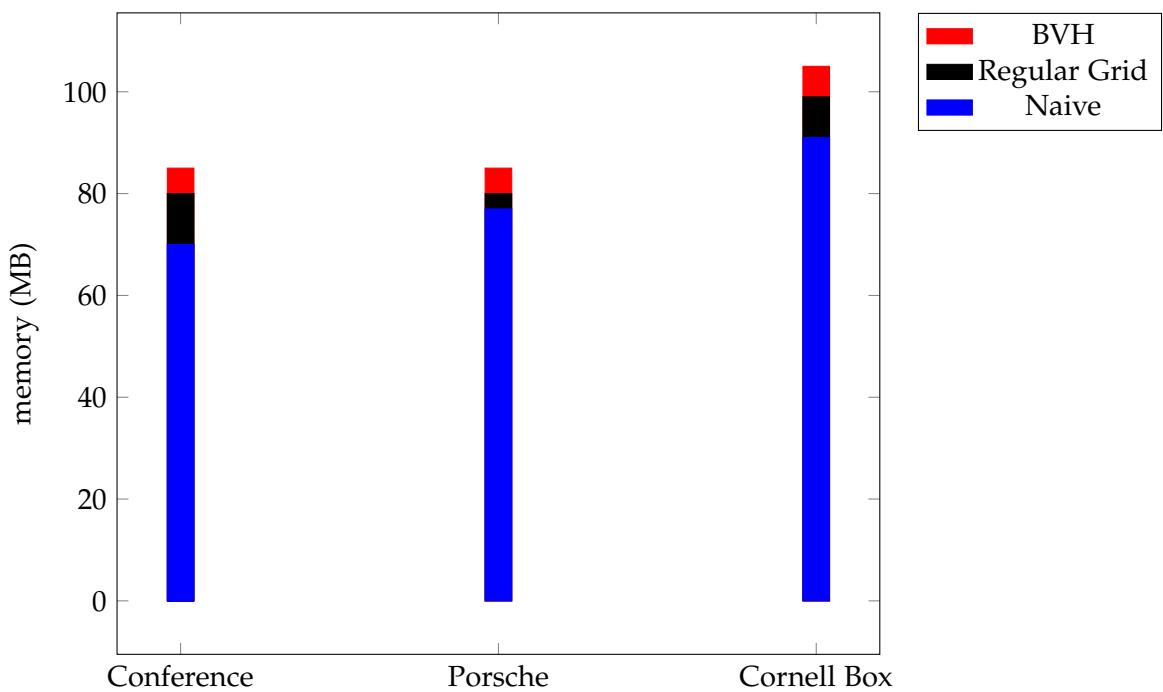
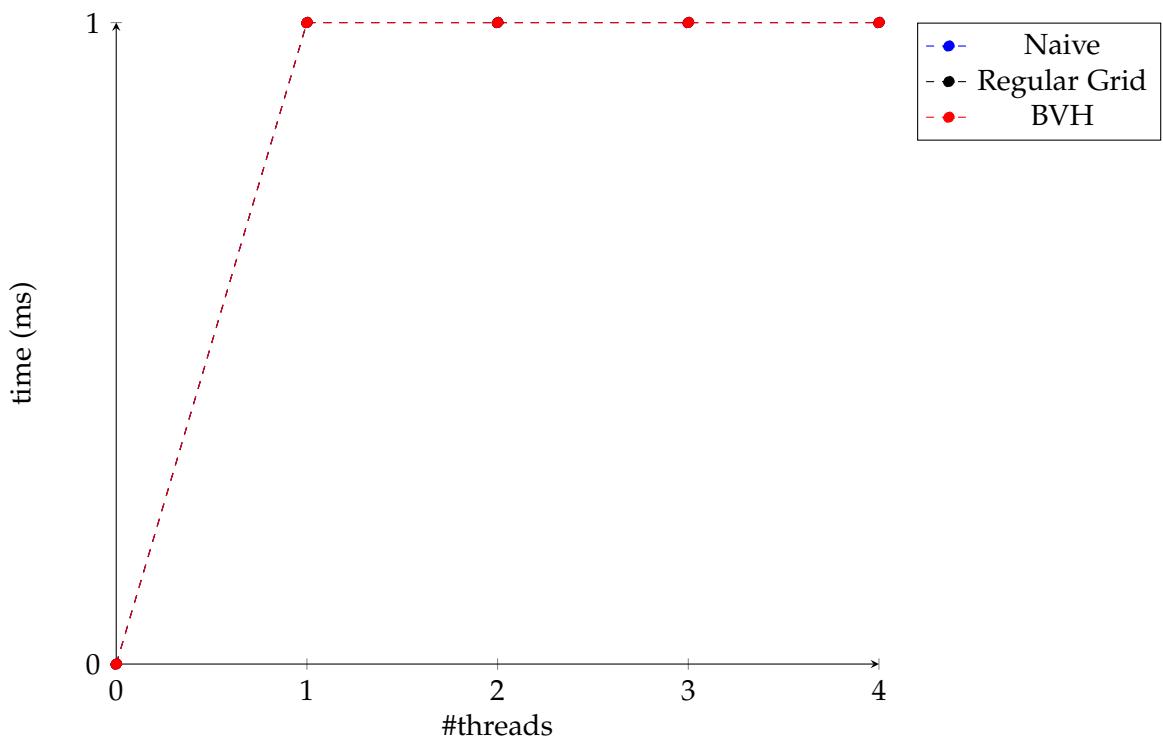


Figure 56.: Memory usage in the Samsung Galaxy smart phone

Raspberry Pi 2 Model B



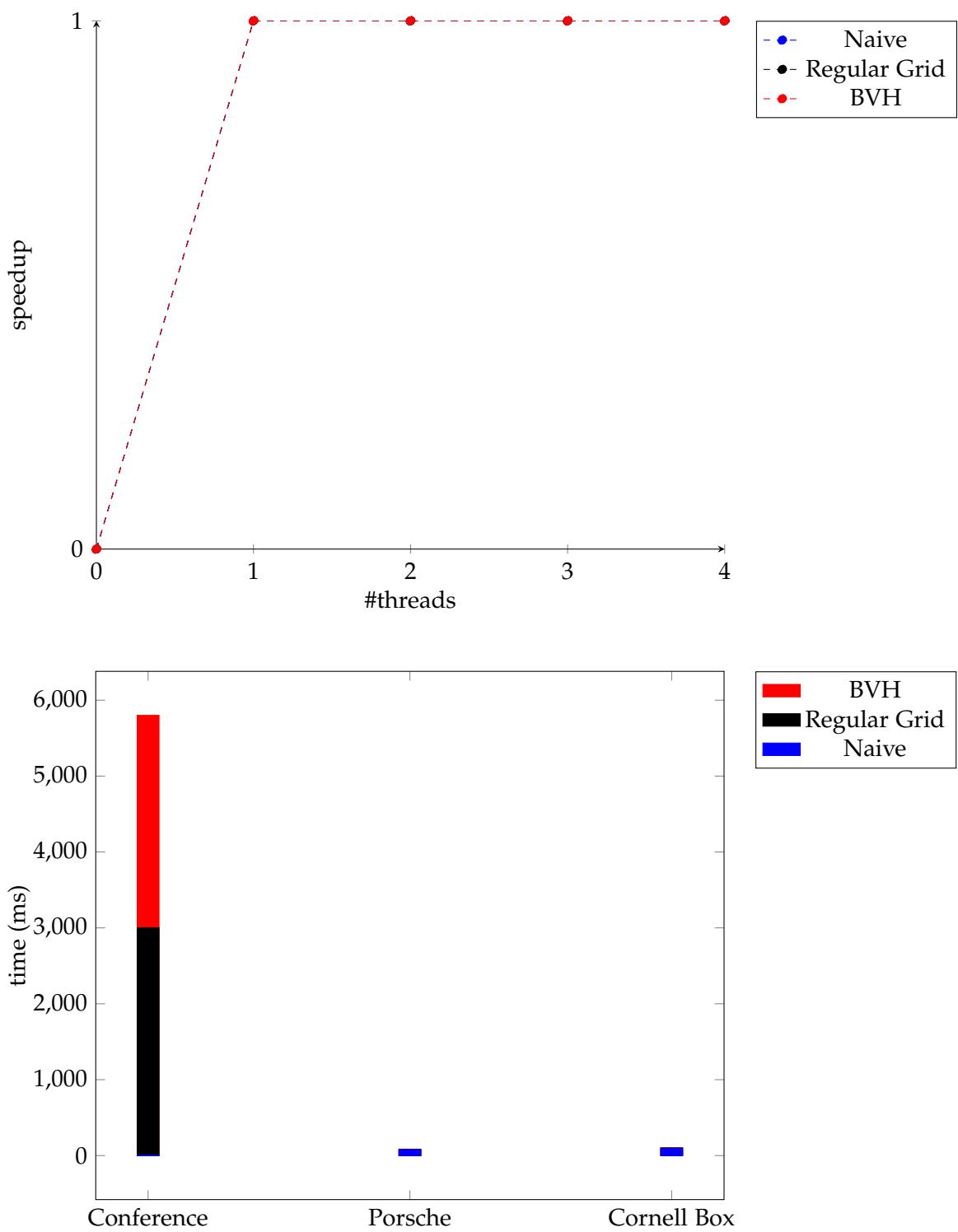


Figure 57.: Time to construct accelerators in the Raspberry Pi 2 Model B

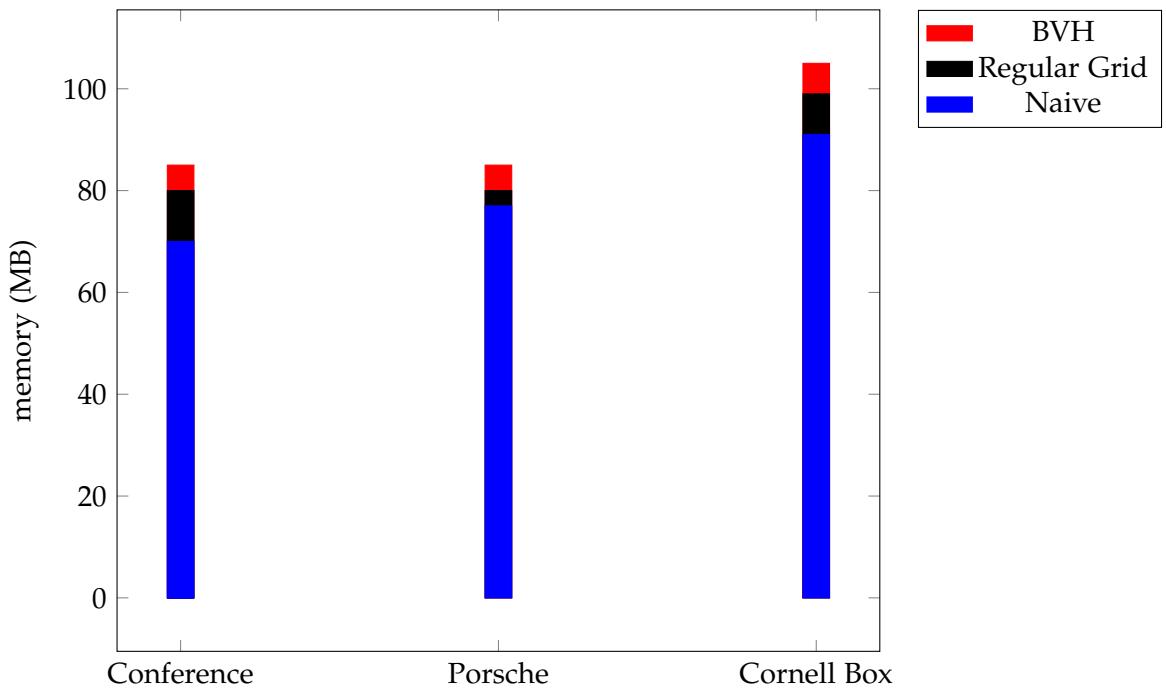
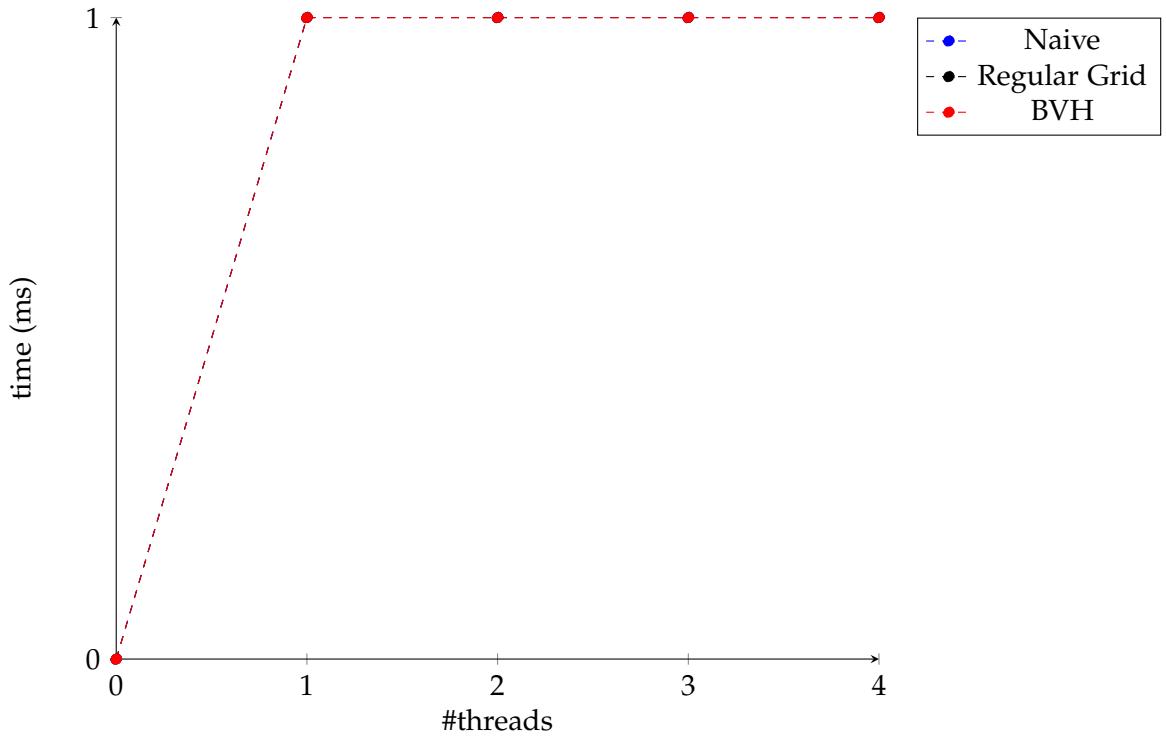


Figure 58.: Memory usage in the Raspberry Pi 2 Model B

MINIX NEO X8-H PLUS



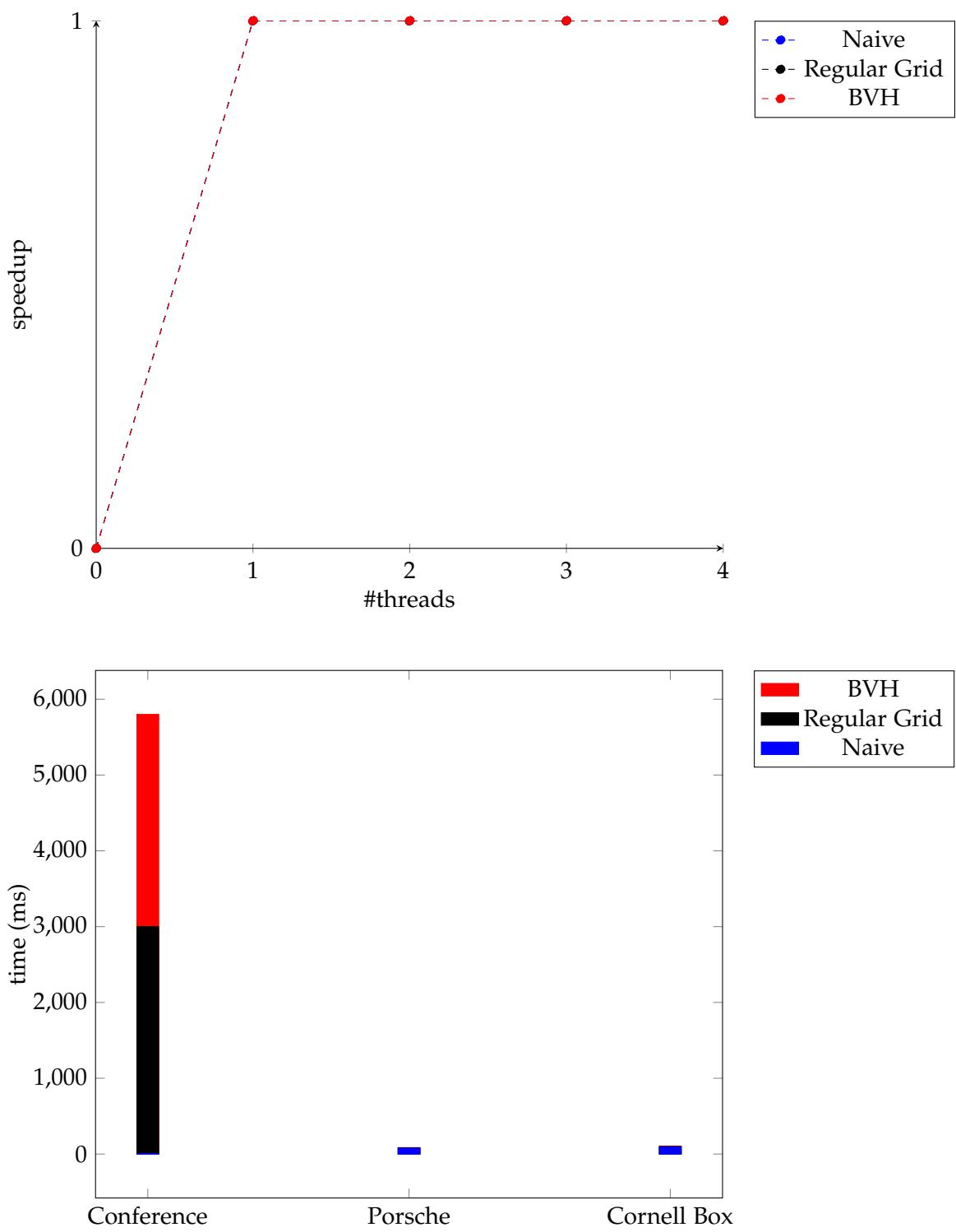


Figure 59.: Time to construct accelerators in the MINIX NEO X8-H PLUS

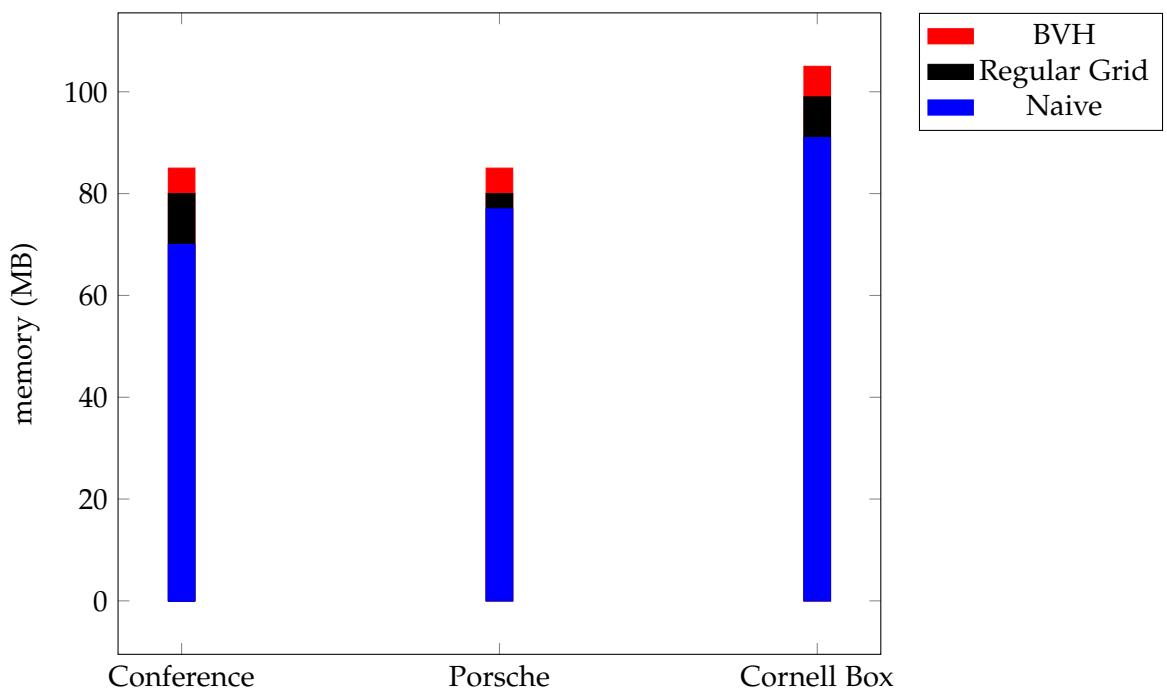


Figure 6o.: Memory usage in the MINIX NEO X8-H PLUS

5.2 COMPARISON WITH ANDROID CPU RAYTRACER (DAHLQUIST)

Comparison ...

6

CONCLUSION & FUTURE WORK

6.1 CONCLUSIONS

The constant evolution of technology allowed to leverage and massify the mobile devices. With, more and increasingly powerful, mobile devices, it is possible to perform more and more complex and useful tasks in them. This is a huge market where the programmers can develop their applications to, and where the development time can be a huge factor in their career success.

But as noted in the section [2.4](#), there is clearly a lack of well documented rendering libraries for mobile systems. There is then the need to change this reality, since more and more the Internet of Things is more present each passing year, because there are increasingly powerful mobile systems.

During the development of this ray tracer library, it was identified some challenges caused by the fact that it was developed for an Android mobile device, like, the smaller amount of RAM available for the software applications. And, also, the simpler CPUs microarchitectures and smaller computational power available from these devices can make computational demanding applications, like ray tracers, difficult to perform the required calculations in a useful time for the user. This fact is corroborated by the obtained results as demonstrated in section [5.1](#).

6.2 FUTURE WORK

With respect to future work, it should be the development of the library for Android and the analysis of possibilities and challenges that may arise. Upon completion of the development process, besides writing the dissertation and the article, it is expected to provide a demo application in order to illustrate the functionalities and performance that the library will offer.

The library was implemented as planned, having been developed:

- two acceleration structures

- three primitive shapes
- intersection
- material
- primitive
- ray
- renderer
- scene
- possibility to develop different types of cameras
- possibility to develop different types of lights
- possibility to develop different types of object loaders
- possibility to develop different types of samplers
- possibility to develop different types of shaders

All these developed components allow the execution of the basic features of a ray tracer, such as the intersection of rays with the different shapes of primitives. Besides that, it allows the user the possibility to develop several types of rendering components, like cameras, lights, object loaders, samplers and shaders. With this, the user can create simple rendering applications that just use the rendering components offered together with the library, or even very complex applications, that use rendering components developed by the user.

Besides the library, it was also developed several rendering components:

- two cameras
- two lights
- one object loader
- six samplers
- five shaders

This allows the user to develop rendering applications in an easy, safe and fast way, as was intended in this dissertation. Unfortunately, this does not allow rapid development of all kinds of rendering applications. One of the things that can be improved in this library is allowing to put more than one camera in the scene, so that you can cast rays to the scene from more than one different point. And thus open the possibility of allowing the user

to develop more different shaders, such as Bidirectional Path Tracing, or even allowing to render the 3D scene into a 3D image.

Other thing that could be implemented is a texture loader for the ray tracer. This allows the possibility to render more beautiful scenes.

And last but not least, there is always the possibility to improve the ray tracer performance by redeveloping the application using a Data-oriented design instead of the traditional Object-oriented design used. This approach was motivated by cache coherency, used in video game development, usually in the programming languages C or C++.

7

BIBLIOGRAPHY

- AMD. Radeon rays technology for developers. URL <http://developer.amd.com/tools-and-sdks/graphics-development/radeonpro/radeonrays-technology-developers/>. Accessed: January 2017.
- Armen Barsegyan. Reality check. what is nvidia rtx technology? what is directx dxr? here's what they can and cannot do. URL <http://cgicoffee.com/blog/2018/03/what-is-nvidia-rtx-directx-dxr>. Accessed: July 2018.
- belthaczar. Ray/plane intersection point. URL <http://www.idevgames.com/forums/thread-5744.html>. Accessed: August 2018.
- Russ Bishop. Swift 2: Simd. URL <http://www.russbishop.net/?page=5&x=272>. Accessed: May 2018.
- Christopher Chedeau. jsraytracer. URL <http://blog.vjeux.com/2012/javascript/javascript-ray-tracer.html>. Accessed: January 2017.
- Chirag. Ray sphere intersection. URL <http://ray-tracing-concept.blogspot.pt/2015/01/ray-sphere-intersection.html>. Accessed: May 2018.
- COOLFINESSE. 5 reasons we love android devices, 2017. URL <https://iheanyiigboko.wordpress.com/2017/09/18/5-reasons-we-love-android-devices/>. Accessed: July 2018.
- G-Truc Creation. Opengl mathematics. URL <https://github.com/g-truc/glm>. Accessed: May 2018.
- Nic Dahlquist. Android cpu raytracer. URL <https://github.com/ndahlquist/raytracer>. Accessed: January 2017.
- Academic Dictionaries and Encyclopedias. Instruction pipeline. URL <http://enacademic.com/dic.nsf/enwiki/141209>. Accessed: May 2018.

- Luís Paulo Peixoto dos Santos. Ray tracing clássico. URL gec.di.uminho.pt/psantos/VI2/Aacetatos/02-RayTracing.ppsx. Accessed: May 2018.
- Le Journal du Net. Gérer et sécuriser ses flottes de smartphones et tablettes. URL <https://www.journaldunet.com/solutions/mobilite/mobile-device-management-comparatif-des-offres>. Accessed: May 2018.
- John Feminella. How to set up quadratic equation for a ray/sphere intersection? URL <https://stackoverflow.com/questions/1986378/how-to-set-up-quadratic-equation-for-a-ray-sphere-intersection>. Accessed: August 2018.
- António José Borba Ramires Fernandes. Ray-triangle intersection. URL <http://www.lighthouse3d.com/tutorials/math/intersection/>. Accessed: July 2018.
- The Foundry. Casting shadows. URL https://learn.foundry.com/nuke/8.0/content/user_guide/3d_compositing/casting_shadows.html. Accessed: May 2018.
- Google. Android developers, a. URL <https://developer.android.com/index.html>. Accessed: May 2017.
- Google. Google test, b. URL <https://github.com/google/googletest>. Accessed: May 2018.
- Rodrigo Placencia & David Bluecame & Olaf Arnold & Michele Castigliego Gustavo Pichorim Boiko. Yafaray. URL <http://www.yafaray.org/>. Accessed: January 2017.
- The Khronos™ Group Inc. Opengl es overview. URL <https://www.khronos.org/opengles/>. Accessed: May 2018.
- © OTOY Inc. Real-time 3d rendering. URL <https://home.otoy.com/render/octane-render/>. Accessed: January 2017.
- Intel. High performance ray tracing kernels. URL <https://embree.github.io/index.html>. Accessed: January 2017.
- Alec Jacobson. Barycentric coordinates and point-triangle queries. URL <http://www.alecjacobson.com/weblog/?p=1596>. Accessed: July 2018.
- Wenzel Jakob. Mitsuba renderer. URL <http://www.mitsuba-renderer.org>. Accessed: January 2017.
- Michael Karbo and ELI Aps. Chapter 28. the cache. URL <http://www.karbosguide.com/books/pcarchitecture/chapter28.htm>. Accessed: May 2018.

- Kenneth. Hray - a haskell ray tracer. URL <http://kejo.be/ELIS/Haskell/HRay/>. Accessed: January 2017.
- Mark Kilgard. Cs 354 acceleration structures. URL https://www.slideshare.net/Mark_Kilgard/26accelstruct. Accessed: May 2018.
- kitware. Cmake. URL <https://cmake.org/>. Accessed: May 2018.
- Iggy Krajci and Darren Cummings. Android on x86: Java native interface and the android native development kit. URL <http://www.drdobbs.com/architecture-and-design/android-on-x86-java-native-interface-and/240166271>. Accessed: July 2018.
- Yurii Lahodiuk. Kd tree for triangle meshes is too slow. URL <https://stackoverflow.com/questions/20019110/kd-tree-for-triangle-meshes-is-too-slow>. Accessed: May 2018.
- Glare Technologies Limited. Indigo rt. URL http://www.indigorenderer.com/indigo_rt. Accessed: January 2017.
- Greg Humphreys Matt Pharr, Wenzel Jakob. Physically based rendering. URL <http://pbrt.org/>. Accessed: January 2017.
- Morgan McGuire. Computer graphics archive, July 2017. URL <https://casual-effects.com/data>. Accessed: July 2018.
- Tomas Moller and Ben Trumbore. Fast minimum storage ray/triangle intersection, 1997. URL <https://cadxfem.org/inf/Fast%20MinimumStorage%20RayTriangle%20Intersection.pdf>. Accessed: July 2018.
- Michael Mara Morgan. The G3D innovation engine. URL <http://g3d.cs.williams.edu/>. Accessed: January 2017.
- Nvidia. Nvidia® optix™ ray tracing engine, a. URL <https://developer.nvidia.com/optix>. Accessed: January 2017.
- Nvidia. Baking with optix, b. URL <https://developer.nvidia.com/optix-prime-baking-sample>. Accessed: January 2017.
- Mini Physics. Refraction of light. URL <https://www.miniphysics.com/refraction-of-light.html>. Accessed: July 2018.
- Pixar. Pixar ris. URL <https://renderman.pixar.com/resources/current/RenderMan/risOverview.html>. Accessed: January 2017.

- Raph Schim. From perspective picture to orthographic picture. URL <https://stackoverflow.com/questions/36573283/from-perspective-picture-to-orthographic-picture>. Accessed: May 2018.
- Computer Science and Engineering. Anti aliasing computer graphics. URL <https://www.slideshare.net/DelwarHossain8/anti-aliasing-computer-graphics>. Accessed: May 2018.
- Scratchapixel. Ray tracing: Rendering a triangle. URL <https://www.scratchapixel.com/lessons/3d-basic-rendering/ray-tracing-rendering-a-triangle>. Accessed: May 2018.
- syoyo. tinyobjloader. URL <https://github.com/syoyo/tinyobjloader>. Accessed: May 2018.
- Daniel A. Thompson. Ray tracing glossy reflection: sampling ray direction. URL <https://stackoverflow.com/questions/32077952/ray-tracing-glossy-reflection-sampling-ray-direction>. Accessed: July 2018.
- Raghu Machiraju Torsten Möller. Ray intersection acceleration. URL <http://slideplayer.com/slide/7981872/>. Accessed: May 2018.
- Will Usher. upacket - a micro packet ray tracer, a. URL <https://github.com/Twinklebear/micro-packet>. Accessed: January 2017.
- Will Usher. tray - a toy ray tracer, b. URL <https://github.com/Twinklebear/tray>. Accessed: January 2017.
- Will Usher. tray_rust - a toy ray tracer in rust, c. URL https://github.com/Twinklebear/tray_rust. Accessed: January 2017.
- Menno Vink. Ray to plane intersection. URL <http://www.echo-gaming.eu/ray-to-plane-intersection/>. Accessed: May 2018.
- Marsette Vona and Tong Pan. Computer graphics (cs 4300) 2010s: Lecture 21. URL <https://course.ccs.neu.edu/cs4300old/s10/L21/L21.html>. Accessed: July 2018.
- Zack Waters. Realistic raytracing. URL https://web.cs.wpi.edu/~emmanuel/courses/cs563/write_ups/zackw/realistic_raytracing.html. Accessed: July 2018.
- Wikipedia. C++ standard library header files, a. URL <https://en.cppreference.com/w/cpp/header>. Accessed: October 2018.
- Wikipedia. Superscalar processor, b. URL https://en.wikipedia.org/wiki/Superscalar_processor. Accessed: October 2018.

- Wikipedia. Depth map, c. URL https://en.wikipedia.org/wiki/Depth_map. Accessed: July 2018.
- Wikipedia. Halton sequence, d. URL https://en.wikipedia.org/wiki/Halton_sequence. Accessed: July 2018.
- Wikipedia. Rendering equation, e. URL https://en.wikipedia.org/wiki/Rendering_equation. Accessed: July 2018.
- Wikipedia. Spherical coordinate system, f. URL https://en.wikipedia.org/wiki/Spherical_coordinate_system. Accessed: July 2018.
- Alan Wolfe. Randomcode. URL <https://github.com/Atrix256/RandomCode>. Accessed: July 2018.
- Stefan Zellmann. Visionaray. URL <https://github.com/szellmann/visionaray>. Accessed: January 2017.

A

USER DOCUMENTATION

A.1 API

```
1 public:  
2 explicit Renderer(::std::unique_ptr<Shader> shader,  
3 ::std::unique_ptr<Camera> camera,  
4 ::std::unique_ptr<Sampler> samplerPixel,  
5 ::std::uint32_t width, ::std::uint32_t height,  
6 ::std::uint32_t samplesPixel) noexcept;  
7  
8 Renderer(const Renderer &renderer) noexcept = delete;  
9  
10 Renderer(Renderer &&renderer) noexcept = delete;  
11  
12 ~Renderer() noexcept = default;  
13  
14 Renderer &operator=(const Renderer &renderer) noexcept = delete;  
15  
16 Renderer &operator=(Renderer &&renderer) noexcept = delete;  
17  
18 void renderFrame(::std::uint32_t *bitmap, ::std::int32_t numThreads, ::std::uint32_t stride) noexcept;  
19  
20 void stopRender() noexcept;  
21  
22 ::std::uint32_t getSample() const noexcept;
```

Listing A.1: Renderer API

```
1 public:  
2 explicit Scene() = default;  
3  
4 Scene(const Scene &scene) noexcept = delete;  
5
```

```

6 Scene(Scene &&scene) noexcept = default;
7
8 ~Scene() noexcept;
9
10 Scene &operator=(const Scene &scene) noexcept = delete;
11
12 Scene &operator=(Scene &&scene) noexcept = default;
13
14 Intersection trace(Intersection intersection, const Ray &ray) noexcept;
15
16 Intersection shadowTrace(Intersection intersection, const Ray &ray) noexcept;
17
18 Intersection traceLights(Intersection intersection, const Ray &ray) const noexcept;
19
20 void resetSampling() noexcept;

```

Listing A.2: Scene API

```

1 public:
2 explicit Ray(const ::glm::vec3 &dir, const ::glm::vec3 &origin,
3 ::std::int32_t depth, const void *primitive = nullptr) noexcept;
4
5 Ray(const Ray &ray) noexcept = default;
6
7 Ray(Ray &&ray) noexcept = default;
8
9 ~Ray() noexcept = default;
10
11 Ray &operator=(const Ray &ray) noexcept = delete;
12
13 Ray &operator=(Ray &&ray) noexcept = delete;

```

Listing A.3: Ray API

```

1 public:
2 Primitive(const T &shape, const Material &material) noexcept;
3
4 Primitive() noexcept = delete;
5
6 Primitive(const Primitive &primitive) noexcept = default;
7
8 Primitive(Primitive &&primitive) noexcept = default;
9

```

```

10 ~Primitive() noexcept = default;
11
12 Primitive &operator=(const Primitive &primitive) noexcept = default;
13
14 Primitive &operator=(Primitive &&primitive) noexcept = default;
15
16 AABB getAABB() const noexcept;
17
18 Intersection intersect(Intersection intersection, const Ray &ray) noexcept;

```

Listing A.4: Primitive API

```

1 public:
2 explicit Intersection() noexcept = default;
3
4 explicit Intersection(float dist, const void *primitive = nullptr) noexcept;
5
6 explicit Intersection(
7   const ::glm::vec3 &orig,
8   const ::glm::vec3 &dir,
9   float dist,
10  const ::glm::vec3 &normal,
11  const void *primitive = nullptr) noexcept;
12
13 explicit Intersection(
14   const ::glm::vec3 &intPoint,
15   float dist,
16   const ::glm::vec3 &sphereCenter) noexcept;
17
18 Intersection(const Intersection &intersection) noexcept = default;
19
20 Intersection(Intersection &&intersection) noexcept = default;
21
22 ~Intersection() noexcept = default;
23
24 Intersection &operator=(const Intersection &intersection) noexcept = delete;
25
26 Intersection &operator=(Intersection &&intersection) noexcept = default;

```

Listing A.5: Intersection API

```

1 public:
2 explicit Material(

```

```

3 const ::glm::vec3 &Kd = ::glm::vec3 {},
4 const ::glm::vec3 &Ks = ::glm::vec3 {},
5 const ::glm::vec3 &Kt = ::glm::vec3 {},
6 float refractiveIndice = 1.0f, const ::glm::vec3 &Le = ::glm::vec3 {}) noexcept;
7
8 Material(const Material &material) noexcept = default;
9
10 Material(Material &&material) noexcept = default;
11
12 ~Material() noexcept = default;
13
14 Material &operator=(const Material &material) noexcept = default;
15
16 Material &operator=(Material &&material) noexcept = default;

```

Listing A.6: Material API

```

1 public:
2 explicit Triangle(
3     const ::glm::vec3 &pointA, const ::glm::vec3 &pointB, const ::glm::vec3 &pointC,
4     const ::glm::vec3 &normal = ::glm::vec3 {}) noexcept;
5
6 Triangle(const Triangle &triangle) noexcept = default;
7
8 Triangle(Triangle &&triangle) noexcept = default;
9
10 ~Triangle() noexcept = default;
11
12 Triangle &operator=(const Triangle &triangle) noexcept = default;
13
14 Triangle &operator=(Triangle &&triangle) noexcept = default;
15
16 Intersection intersect(const Intersection &intersection, const Ray &ray) const noexcept;
17
18 void moveTo(float x, float y) noexcept;
19
20 float getZ() const noexcept;
21
22 AABB getAABB() const noexcept;
23
24 bool intersect(const AABB &box) const noexcept;

```

Listing A.7: Triangle API

```

1 public:
2 explicit Plane(const ::glm::vec3 &point, const ::glm::vec3 &normal) noexcept;
3
4 Plane(const Plane &plane) noexcept = default;
5
6 Plane(Plane &&plane) noexcept = default;
7
8 ~Plane() noexcept = default;
9
10 Plane &operator=(const Plane &plane) noexcept = default;
11
12 Plane &operator=(Plane &&plane) noexcept = default;
13
14 Intersection intersect(const Intersection &intersection, const Ray &ray) const noexcept;
15
16 void moveTo(float x, float y) noexcept;
17
18 float getZ() const noexcept;
19
20 AABB getAABB() const noexcept;
21
22 float distance(const ::glm::vec3 &point) const noexcept;
23
24 bool intersect(const AABB &box) const noexcept;

```

Listing A.8: Plane API

```

1 public:
2 explicit Sphere(const ::glm::vec3 &center, float radius) noexcept;
3
4 Sphere(const Sphere &sphere) noexcept = default;
5
6 Sphere(Sphere &&sphere) noexcept = default;
7
8 ~Sphere() noexcept = default;
9
10 Sphere &operator=(const Sphere &sphere) noexcept = default;
11
12 Sphere &operator=(Sphere &&sphere) noexcept = default;
13
14 Intersection intersect(const Intersection &intersection, const Ray &ray) const noexcept;
15
16 void moveTo(float x, float y) noexcept;

```

```

17
18 float getZ() const noexcept;
19
20 AABB getAABB() const noexcept;
21
22 bool intersect(const AABB &box) const noexcept;

```

Listing A.9: Sphere API

```

1 #define LOG(...) ::MobileRT::log(::MobileRT::getFileName(__FILE__), ":" __LINE__ ":" ,
2 __VA_ARGS__);
3
4 template<typename ...Args>
5 void log(Args &&... args) noexcept;
6
7 inline ::std::string getFileName(const char *filepath) noexcept;
8
9 template<typename T>
10 ::std::vector<T *> convertVector(::std::vector<T> &source) noexcept;
11
12 ::std::int32_t roundDownToMultipleOf(::std::int32_t value, ::std::int32_t multiple) noexcept;
13
14 float haltonSequence(::std::uint32_t index, ::std::uint32_t base) noexcept;
15
16 ::std::uint32_t incrementalAvg(
17     const ::glm::vec3 &sample, ::std::uint32_t avg, ::std::uint32_t numSample) noexcept;

```

Listing A.10: Utils API

```

1 public:
2 explicit AABB() noexcept = default;
3
4 explicit AABB(const ::glm::vec3 &pointMin, const ::glm::vec3 &pointMax) noexcept;
5
6 AABB(const AABB &AABB) noexcept = default;
7
8 AABB(AABB &&AABB) noexcept = default;
9
10 ~AABB() noexcept = default;
11
12 AABB &operator=(const AABB &AABB) noexcept = default;
13
14 AABB &operator=(AABB &&AABB) noexcept = default;

```

```

15
16 float getSurfaceArea() const noexcept;

```

Listing A.11: AABB API

```

1 public:
2 explicit BVH() noexcept = default;
3
4 explicit BVH<T>(
5     ::std::vector<Primitive<T>> && primitives) noexcept;
6
7 BVH(const BVH &bVH) noexcept = delete;
8
9 BVH(BVH &&bVH) noexcept = default;
10
11 ~BVH() noexcept = default;
12
13 BVH &operator=(const BVH &bVH) noexcept = delete;
14
15 BVH &operator=(BVH &&bVH) noexcept = default;
16
17 Intersection trace(Intersection intersection, const Ray &ray) noexcept;
18
19 Intersection shadowTrace(Intersection intersection, const Ray &ray) noexcept;

```

Listing A.12: BVH API

```

1 public:
2 explicit RegularGrid() noexcept = default;
3
4 explicit RegularGrid(AABB sceneBounds, Scene *scene, ::std::int32_t gridSize) noexcept;
5
6 RegularGrid(const RegularGrid &regularGrid) noexcept = delete;
7
8 RegularGrid(RegularGrid &&regularGrid) noexcept = default;
9
10 ~RegularGrid() noexcept = default;
11
12 RegularGrid &operator=(const RegularGrid &regularGrid) noexcept = delete;
13
14 RegularGrid &operator=(RegularGrid &&regularGrid) noexcept = default;
15
16 Intersection trace(Intersection intersection, const Ray &ray) noexcept;

```

```

17
18 Intersection shadowTrace(Intersection intersection, const Ray &ray) noexcept;
```

Listing A.13: RegularGrid API

```

1 protected:
2 virtual bool shade(::glm::vec3 *rgb, const Intersection &intersection,
3 const Ray &ray) noexcept = 0;
4
5 ::glm::vec3 getCosineSampleHemisphere(const ::glm::vec3 &normal) const noexcept;
6
7 public:
8 void initializeAccelerators(Camera *camera) noexcept;
9
10 public:
11 explicit Shader(Scene scene, ::std::uint32_t samplesLight, Accelerator accelerator) noexcept;
12
13 Shader(const Shader &shader) noexcept = delete;
14
15 Shader(Shader &&shader) noexcept = default;
16
17 virtual ~Shader() noexcept;
18
19 Shader &operator=(const Shader &shader) noexcept = delete;
20
21 Shader &operator=(Shader &&shader) noexcept = delete;
22
23 bool rayTrace(::glm::vec3 *rgb, const Ray &ray) noexcept;
24
25 bool shadowTrace(Intersection intersection, const Ray &ray) noexcept;
26
27 Intersection traceTouch(Intersection intersection, const Ray &ray) noexcept;
28
29 virtual void resetSampling() noexcept;
30
31 ::std::uint32_t getLightIndex();
```

Listing A.14: Shader API

```

1 public:
2 explicit Sampler() noexcept = default;
3
4 explicit Sampler(::std::uint32_t width, ::std::uint32_t height,
```

```

5 ::std::uint32_t samples) noexcept;
6
7 Sampler(const Sampler &sampler) noexcept = delete;
8
9 Sampler(Sampler &&sampler) noexcept = delete;
10
11 virtual ~Sampler() noexcept;
12
13 Sampler &operator=(const Sampler &sampler) noexcept = delete;
14
15 Sampler &operator=(Sampler &&sampler) noexcept = delete;
16
17 void resetSampling() noexcept;
18
19 void stopSampling() noexcept;
20
21 virtual float getSample(::std::uint32_t sample) noexcept = 0;
22
23 float getSample() noexcept;

```

Listing A.15: Sampler API

```

1 protected:
2 float degToRad(const float deg) const noexcept;
3 float radToDeg(const float rad) const noexcept;
4
5 public:
6 explicit Camera(const ::glm::vec3 &position,
7   const ::glm::vec3 &lookAt, const ::glm::vec3 &up) noexcept;
8
9 Camera(const Camera &camera) noexcept;
10
11 Camera(Camera &&camera) noexcept = delete;
12
13 virtual ~Camera() noexcept;
14
15 Camera &operator=(const Camera &camera) noexcept = delete;
16
17 Camera &operator=(Camera &&camera) noexcept = delete;
18
19 virtual Ray generateRay(float u, float v,
20   float deviationU,
21   float deviationV) const noexcept = 0;

```

```

22
23 float getBlock(:std::uint32_t sample) noexcept;
24
25 virtual AABB getAABB() const noexcept;
26
27 void resetSampling() noexcept;

```

Listing A.16: Camera API

```

1 public:
2 explicit Light(const Material &radiance) noexcept;
3
4 Light(const Light &light) noexcept = delete;
5
6 Light(Light &&light) noexcept = delete;
7
8 virtual ~Light() noexcept;
9
10 Light &operator=(const Light &light) noexcept = delete;
11
12 Light &operator=(Light &&light) noexcept = delete;
13
14 virtual ::glm::vec3 getPosition() noexcept = 0;
15
16 virtual void resetSampling() noexcept = 0;
17
18 virtual Intersection intersect(Intersection intersection, const Ray &ray) const noexcept = 0;

```

Listing A.17: Light API

```

1 public:
2 explicit ObjectLoader() noexcept = default;
3
4 ObjectLoader(const ObjectLoader &objectLoader) noexcept = delete;
5
6 ObjectLoader(ObjectLoader &&objectLoader) noexcept = delete;
7
8 virtual ~ObjectLoader() noexcept;
9
10 ObjectLoader &operator=(const ObjectLoader &objectLoader) noexcept = delete;
11
12 ObjectLoader &operator=(ObjectLoader &&objectLoader) noexcept = delete;
13

```

```

14 virtual void process() noexcept = 0;
15
16 bool isProcessed() const noexcept;
17
18 virtual bool fillScene(Scene *scene,
19 ::std::function<::std::unique_ptr<Sampler>()> lambda,
20 JNIEnv *env) noexcept = 0;

```

Listing A.18: ObjectLoader API

A.2 LOADING A SCENE

```

1 jboolean isCopy {JNI_FALSE};
2 const char *const objFileName {(env)->GetStringUTFChars(globalObjFile, &isCopy)};
3 const char *const matFileName {(env)->GetStringUTFChars(globalMatFile, &isCopy)};
4
5 const jboolean result{
6 env->CallStaticBooleanMethod(mainActivityClass, mainActivityMethodId, o);
7 if (result) {
8   return -1;
9 }
10
11 ::Components::OBJLoader objLoader{objFileName, matFileName};
12 {
13   const jboolean result{
14 env->CallStaticBooleanMethod(mainActivityClass, mainActivityMethodId, o);
15   if (result) {
16     return -1;
17   }
18 }
19 objLoader.process();
20 {
21   const jboolean result{
22 env->CallStaticBooleanMethod(mainActivityClass, mainActivityMethodId, o);
23   if (result) {
24     return -1;
25   }
26 }
27
28 if (!objLoader.isProcessed()) {
29   return -1;
30 }

```

```

31 const bool sceneBuilt{objLoader.fillScene(&scene_,
32 []() { return ::std::make_unique<Components::StaticHaltonSeq>(); }, env)};
33 if (!sceneBuilt) {
34     return -1;
35 }
36 {
37     const jboolean result{
38         env->CallStaticBooleanMethod(mainActivityClass, mainActivityMethodId, o)};
39     if (result) {
40         return -1;
41     }
42 }
43
44 const float ratio {
45     ::std::max(static_cast<float>(width) / height, static_cast<float>(height) / width);
46     const float hfovFactor{width > height ? ratio : 1.0f};
47     const float vfovFactor{width < height ? ratio : 1.0f};
48     const float fovX{45.0f * hfovFactor};
49     const float fovY{45.0f * vfovFactor};
50     maxDist = ::glm::vec3 {1, 1, 1};
51     const ::MobileRT::Material &lightMat{::glm::vec3 {0.0f, 0.0f, 0.0f},
52         ::glm::vec3 {0.0f, 0.0f, 0.0f},
53         ::glm::vec3 {0.0f, 0.0f, 0.0f},
54         1.0f,
55         ::glm::vec3 {0.9f, 0.9f, 0.9f}};
56
57 //cornellbox
58 if (::std::strstr(objFileName, "CornellBox") != nullptr) {
59     camera = ::std::make_unique<Components::Perspective>(
60         ::glm::vec3 {0.0f, 0.7f, 3.0f},
61         ::glm::vec3 {0.0f, 0.7f, -1.0f},
62         ::glm::vec3 {0.0f, 1.0f, 0.0f},
63         fovX, fovY);
64
65     ::std::unique_ptr<MobileRT::Sampler> samplerPoint1{
66         ::std::make_unique<Components::StaticHaltonSeq>()};
67     scene_.lights_.emplace_back(::std::make_unique<Components::AreaLight>(
68         lightMat,
69         ::std::move(samplerPoint1),
70         ::glm::vec3 {0.5f, 1.58f, 0.5f},
71         ::glm::vec3 {-0.5f, 1.58f, 0.5f},
72         ::glm::vec3 {-0.5f, 1.58f, -0.5f}));
73     ::std::unique_ptr<MobileRT::Sampler> samplerPoint2{

```

```
74 ::std::make_unique<Components::StaticHaltonSeq>());
75 scene_.lights_.emplace_back(::std::make_unique<Components::AreaLight>(
76     lightMat,
77     ::std::move(samplerPoint2),
78     ::glm::vec3 {0.5f, 1.58f, 0.5f},
79     ::glm::vec3 {-0.5f, 1.58f, -0.5f},
80     ::glm::vec3 {0.5f, 1.58f, -0.5f}));
81 }
```

Listing A.19: How to load a 3D scene