

Uso de Vulkan para el trazado de rayos

Joaquín Andrés Fontana Albornoz

Pasantía de Iniciación a la Investigación de PEDECIBA Informática. Módulo Taller de la carrera de Ingeniería en Computación.

Proponente: Grupo de Computación Gráfica, Instituto de Computación, Fing, Udelar

Responsables: Eduardo Fernández, José Pedro Aguerre

5 de abril de 2024

Índice

1. Introducción	3
1.1. ¿Qué es Vulkan?	3
1.2. Antes de comenzar	3
2. Repositorios a utilizar	4
3. Pipeline de <i>ray tracing</i>	5
3.1. Shaders de <i>ray tracing</i>	6
3.2. Variables y funciones integradas (built-in functions) por GLSL_EXT_ray_tracing	8
4. Repositorio	10
4.1. Estructura	10
4.2. ImGui	11
4.3. Ejemplos	12
4.3.1. Vk1_launchID Visualización LaunchID	13
4.3.2. Vk2_primitiveID Visualización PrimitiveID	13
4.3.3. Vk3_material Evaluar materiales	14
4.3.4. Vk4_rebotes Rebote de un rayo	14
4.3.5. Vk5_random Números aleatorios	15
4.4. Vk_path_tracer Path tracer	16
5. Estudio de perforamnce	17
6. Lecturas adicionales	18
Referencias	19

1. Introducción

A lo largo de este informe explicaremos la implementación de un algoritmo de *ray tracing* en GPU utilizando Vulkan, desglosando su pipeline y analizando la estructura del repositorio que contiene ejemplos prácticos¹. Este repositorio no solo sirve como punto de partida para comprender la implementación del *ray tracing* en Vulkan, sino que también ofrece la posibilidad de personalizar y construir aplicaciones propias. Además, examinaremos cada ejemplo proporcionado y demostraremos cómo modificar el código para adaptarlo a necesidades específicas.

1.1. ¿Qué es Vulkan?

Vulkan es una API de gráficos desarrollada por el consorcio Khronos Group. Esta representa un salto adelante en rendimiento y eficiencia, ofreciendo un control directo sobre el hardware gráfico. Surgió en 2015 como sucesor de OpenGL, con el objetivo de proporcionar un marco moderno y poderoso para aplicaciones gráficas y de computación en paralelo. Vulkan es de código abierto y multiplataforma, lo que significa que está disponible para su uso en una variedad de sistemas operativos y arquitecturas de hardware. Su enfoque en el rendimiento, eficiencia y capacidad de optimización lo convierte en una herramienta clave para desarrolladores en campos que van desde videojuegos hasta simulación científica. [1] [2]

1.2. Antes de comenzar

Previo a comenzar con los detalles, es crucial asegurarnos que nuestra GPU cuente con soporte para *ray tracing* en Vulkan, o sea debe contar con las extensiones: VK_KHR_ray_tracing_pipeline, VK_KHR_acceleration_structure, VK_KHR_deferred_host_operations. Verificar la presencia de estas extensiones se puede realizar muy fácilmente al instalar el SDK de Vulkan o consultando la lista en la web proporcionada [3]. Es importante tener en cuenta que esta base de datos es mantenida por un usuario de la comunidad y podría estar desactualizada, por lo que se recomienda la primera opción.

¹Enlace al repositorio: <https://github.com/jquinfontana/VulkanRayTracing>

2. Repositorios a utilizar

En los ejemplos presentados utilizaremos los repositorios: **Nvpro Core**, para facilitar la inclusión de todos los recursos necesarios, y **Nvvk helper** para facilitar la manipulación de Vulkan.

- **Nvpro Core** de Nvidia es un repositorio que recopila código fuente y bibliotecas tanto de Nvidia como de terceros, útiles para el desarrollo de aplicaciones gráficas. Algunas de las bibliotecas que utilizaremos son: **Nvvk Helper**, glm, imgui, OBJLoader, entre otras [4].
- **Nvvk helper** es una colección de funciones auxiliares para trabajar con la API de Vulkan. Esta biblioteca proporciona una serie de utilidades y clases para simplificar el trabajo con Vulkan, facilitando tareas comunes y reduciendo la cantidad de código repetitivo [5].
Es importante destacar que, si bien la biblioteca **Nvvk Helper** fue desarrollada por Nvidia, también es **compatible con dispositivos de AMD**, debido a que esta consiste en envolver código Vulkan, que es portable.
- **ImGui** es una biblioteca de interfaz de usuario ligera y eficiente. Su enfoque minimalista facilita su integración en proyectos existentes, ofreciendo una amplia gama de widgets personalizables para crear interfaces de usuario de manera sencilla.
- **GLM** es una biblioteca de matemáticas utilizada comúnmente para gráficos en 3D. Diseñada para ser compatible con las especificaciones de **GLSL**, **GLM** proporciona una amplia gama de funciones y clases para realizar operaciones matemáticas comunes en gráficos, como transformaciones, rotaciones, escalas y proyecciones, facilitando el manejo de la geometría en aplicaciones 3D.

3. Pipeline de *ray tracing*

El pipeline de *ray tracing* consiste en múltiples etapas programables (shaders), que interactúan entre sí, de la forma que muestra el diagrama (ver Figura 1), destinadas a el uso de los desarrolladores para controlar el renderizado. En el contexto del *ray tracing*, un shader puede ser responsable de una variedad de tareas, incluyendo la generación de rayos, la determinación de cómo los rayos interactúan con los objetos en la escena, y la combinación de estos resultados para producir la imagen final.

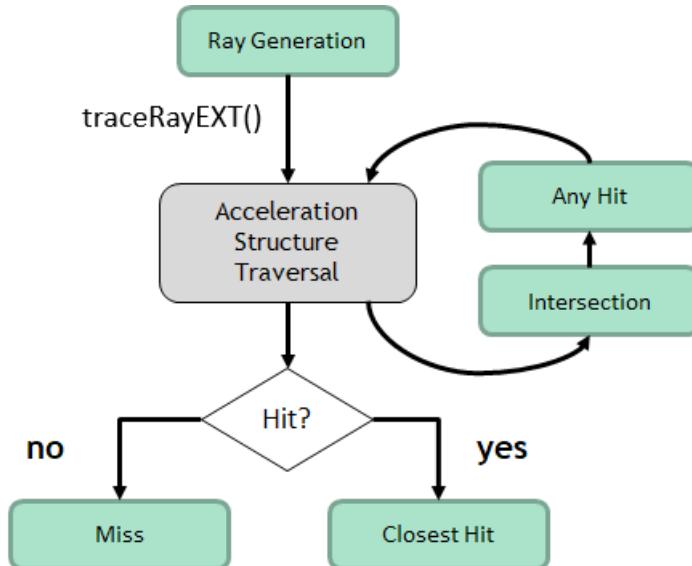


Figura 1: Orden de ejecución de shaders

Antes de comenzar con el trazado de rayos, deberemos construir una estructura de aceleración para nuestra escena. Esta es construida por el propio Vulkan y consta de una estructura de aceleración de nivel superior (TLAS) y múltiples estructuras de aceleración de nivel inferior (BLAS). Cada BLAS puede ser una malla de triángulos o una colección definida por el usuario de cajas delimitadoras (AABB).

Un BLAS puede ser instanciado en el TLAS y recibe un *gl_InstanceID* único. Además, cada triángulo en una malla de triángulos y cada AABB en la colección de cajas de intersección recibe un *gl_PrimitiveID*.

Cada BLAS tiene su propia matriz de transformación a espacio global, que se asigna inicialmente cuando el BLAS se añade al TLAS. Esta transformación es accesible en el shader a través de las variables integradas *gl_ObjectToWorldEXT* y *gl_WorldToObjectEXT*. Cuando se produce un impacto en el recorrido del rayo y se llama al *intersection shader*, *any-hit* o *closest-hit*, estas variables mencionadas se establecen en consecuencia.

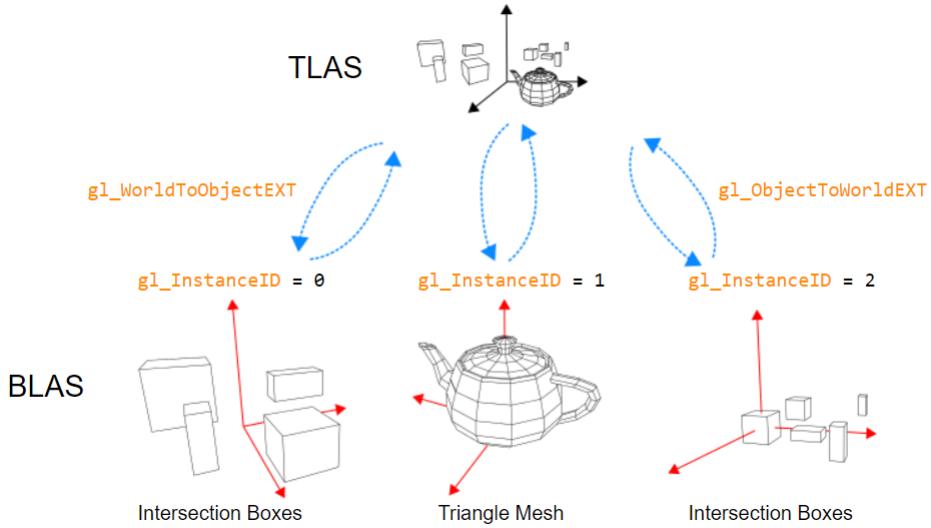


Figura 2: Diagrama de estructura de aceleración

3.1. Shaders de *ray tracing*

En esta sección daremos una descripción general de los diferentes tipos de shaders que podrías encontrar en un pipeline de *ray tracing*, además explicaremos la variable **payload**, ya que es un concepto central para entender como estos trabajan en conjunto.

- La variable **payload** es una variable global que puede ser accedida desde cualquier shader y contiene información sobre las propiedades de un rayo. Por ejemplo, suele ser un *struct* con las propiedades: color, origen, dirección, etc. Es el usuario quien define el tipo de esta variable y puede modificarse en las etapas de shaders que se llaman posteriormente en el recorrido del rayo. Conceptualmente es una estructura de datos que representa un rayo y permite comunicar información entre distintas etapas del pipeline.
- El ***ray-generation shader*** es el punto de partida para el pipeline de trazado de rayos. Especifica y lanza rayos a través de la “*Acceleration Structure*” llamando a la función `traceRayEXT(...)`. La función `traceRayEXT()` lanza un solo rayo en la escena para probar intersecciones y, al hacerlo, puede activar otros shaders, que se presentarán en breve. El parámetro más importante de la función `traceRayEXT()` es la variable **payload**. Una vez que el recorrido del rayo se completa, la función `traceRayEXT()` retorna el control al *ray-generation shader*. En general, luego que esto ocurre, se evalúa y procesan los datos del **payload** en el *ray-generation shader* para producir una imagen de salida, ya que en este se encuentra la información sobre la interacción del rayo con la escena proporcionada por los otros shaders.

- El ***closest-hit shader*** se ejecuta cuando se determina que un rayo ha golpeado el objeto más cercano desde el origen del rayo. Por lo que típicamente se utilizan para calcular o recopilar las propiedades del objeto (cálculos de iluminación, evaluación de materiales, etc) en el punto de intersección. Para lograr esto, el shader puede acceder a varias variables integradas, como *gl_PrimitiveID* o *gl_InstanceID*, que se establecen luego de cada impacto y son un identificador tanto de la geometría, como del triángulo impactado por el rayo. De lo contrario, si no se produjo ningún impacto, se llama al *miss shader*.
- El ***miss shader*** usualmente muestrea un mapa de entorno, o simplemente devuelve un color. Se le puede dar otro uso para implementar un rayo de sombra, lanzando un rayo e indicando a través de un **payload** que un objeto no está ocluido si se ejecuta este shader.

Los shaders que se presentan a continuación, son opcionales:

- El ***intersection shader*** se utiliza para intersecar rayos con geometría definida por el usuario. Las intersecciones de rayo/tríangulo tienen soporte incorporado, por lo tanto, no requieren un *intersection shader*, pero si se detecta una intersección del rayo con una caja AABB (axis-aligned bounding box) definida por el usuario o un triángulo de una malla de triángulos, se llama al *intersection shader*. Si el shader determina que ha ocurrido una intersección rayo/primitiva dentro de la caja delimitadora, se notifica con la función *reportIntersectionEXT(...)*. Además, el *intersection shader* puede llenar una variable *hitAttributeEXT* (que puede ser el usuario quien define su tipo). En el caso de triángulos, ya hay un *intersection shader* por defecto. Este proporciona coordenadas baricéntricas de la ubicación del impacto dentro del triángulo con la variable “*hitAttributeEXT vec2 baryCoord*”. Para primitivas geométricas que no son triángulos (como cubos, cilindros, esferas, superficies paramétricas, etc.), el usuario debe programar un *intersection shader* acorde a la primitiva geométrica.
- El ***any-hit shader*** se ejecuta en cada intersección del rayo con una primitiva (no solo en la más cercana). Se puede utilizar para descartar intersecciones, por ejemplo, para realizar alfa testing, mediante una búsqueda de textura e ignorando la intersección si el valor obtenido no cumple con un criterio específico. El *any-hit shader* por defecto devuelve información sobre las intersecciones para que se pueda determinar la intersección más cercana.

En resumen: Los algoritmos de *ray tracing* se implementan mediante la programación de los diferentes shaders, los cuales se ejecutan en el orden especificado en el diagrama de arriba (ver Figura 1), y donde cada uno, en general, se encargara de una tarea específica. Gracias a esta metodología de trabajo, se aprovecha la capacidad de paralelización entre etapas del pipeline, así como la creación de la estructura de aceleración y el trazado eficiente de

rayos sobre ella, ambas provistas por la GPU. Esto resulta en un notable incremento en la performance de nuestra aplicación.

3.2. Variables y funciones integradas (built-in functions) por GLSL_EXT - ray_tracing

Mediante una directiva al comienzo de nuestros shader podremos habilitar la extensión *GLSL_EXT_ray_tracing*. Gracias a esto se definirán nuevas variables, constantes y funciones ya integradas en la GPU para el uso del pipeline de *ray tracing*. En esta sección daremos una lista de todas estas variables (ver Figura 3) y explicaremos las más importantes.

Ray generation	Closest-hit	Miss	Intersection	Any-hit
<code>uvec3 gl_LaunchIDEXT</code>	✓	✓	✓	✓
<code>uvec3 gl_LaunchSizeEXT</code>	✓	✓	✓	✓
<code>int gl_PrimitiveID</code>	✓		✓	✓
<code>int gl_InstanceID</code>	✓		✓	✓
<code>int gl_InstanceCustomIndexEXT</code>	✓		✓	✓
<code>int gl_GeometryIndexEXT</code>	✓		✓	✓
<code>vec3 gl_WorldRayOriginEXT</code>	✓	✓	✓	✓
<code>vec3 gl_WorldRayDirectionEXT</code>	✓	✓	✓	✓
<code>vec3 gl_ObjectRayOriginEXT</code>	✓		✓	✓
<code>vec3 gl_ObjectRayDirectionEXT</code>	✓		✓	✓
<code>float gl_RayTminEXT</code>	✓	✓	✓	✓
<code>float gl_RayTmaxEXT</code>	✓	✓	✓	✓
<code>uint gl_IncomingRayFlagsEXT</code>	✓	✓	✓	✓
<code>float gl_HitTEXT</code>	✓			✓
<code>uint gl_HitKindEXT</code>	✓			✓
<code>mat4x3 gl_ObjectToWorldEXT</code>	✓		✓	✓
<code>mat4x3 gl_WorldToObjectEXT</code>	✓		✓	✓

Figura 3: Lista de variables integradas

La extensión *GLSL_EXT_ray_tracing* define varias funciones integradas, siendo la más utilizada *traceRayEXT()*, que inicia una operación de trazado de rayo. Esta puede ser llamada desde un *ray-generation*, *closest-hit* o *miss shader*. Las aplicaciones típicas son disparar los rayos primarios desde la posición de la cámara en el *ray-generation shader*, y disparar un rayo “de sombra” en dirección a la fuente de luz para determinar si la superficie está ocluida por otros objetos en el *closest-hit shader*.

Debemos tener especial cuidado con llamar a la función *traceRayEXT()* en shaders que no sean el de generación de rayos, puesto que una llamada en otro shader se considera **recursión**. Esto no es deseable por dos motivos: Dependiendo la GPU que utilicemos puede variar

el número de llamadas recursivas soportado, por ejemplo, una Nvidia Quadro RTX 6000 soporta 31 llamadas recursivas, mientras que una AMD Radeon RX 7600 soporta una sola, por lo que si realizamos llamadas recursivas, nuestra aplicación, podría no funcionar sobre cierto hardware. El otro motivo es que los algoritmos recursivos, si bien son fáciles de programar, suelen ser menos eficientes que su contraparte iterativa, por lo que la programación iterativa impactará positivamente en el rendimiento de la aplicación.

Esta extensión también define a las siguientes variables:

- *gl_LaunchIDEXT* y *gl_LaunchSizeEXT* se usan para identificar un hilo lanzado por la GPU para un *ray-generation shader*. *gl_LaunchIDEXT* es un vector donde sus primeras dos entradas indican las coordenadas del píxel para el que se está ejecutando el shader. Mientras que *gl_LaunchSizeEXT* contiene en sus primeras dos coordenadas el ancho y alto de la imagen de salida.
- *gl_PrimitiveIDEXT* y *gl_InstanceIDEXT* se usan para identificar una primitiva (en general un triángulo) y una instancia (en general una malla de triángulos) respectivamente, al procesar una intersección con un *closest-hit shader*, *any-hit* o *intersection*.
- *gl_HitTEXT* contiene la distancia desde el origen del rayo al punto de intersección (accesible en shaders *closest-hit* y *any-hit*).

A demás la extensión define nuevos calificadores de almacenamiento para la información que necesitan compartirse entre los shaders. Los calificadores *rayPayloadEXT* y *rayPayloadInEXT* se utilizan para definir variables globales, que se acceden desde distintos shaders para la comunicación de estos. Donde *rayPayloadEXT* indica que la variable se crea en ese shader y *rayPayloadInEXT* indica que la variable viene como una entrada en ese shader. Se permite utilizar en cualquier etapa que pueda ser invocada durante la ejecución de *TraceRayEXT()*. El tipo de variable utilizado en *rayPayloadEXT* y *rayPayloadInEXT* debe coincidir entre el llamador y el receptor.

El calificador *hitAttributeEXT* instancia una variable con almacenamiento para datos de intersección rayo/primitiva. La declaración explícita de un *hitAttributeEXT* solo es necesaria cuando el pipeline incluye *intersection shaders* personalizados. Las variables *hitAttributeEXT* es de solo lectura para shaders *any-hit* y *closest-hit* y de lectura-escritura en *intersection shaders*.

4. Repositorio

En esta sección explicaremos el contenido del repositorio que se encuentra en [Github](#), este cuenta con varias aplicaciones construidas en *C++* utilizando Vulkan, cuyo objetivo es facilitar recursos útiles para el desarrollo de algoritmos de trazado de rayos.

Si bien se dio un pantallazo general a las librerías utilizadas, todos los ejemplos tienen una estructura ya definida para la aplicación en *C++*. Por esta razón los ejemplos se enfocarán en la modificación y programación de los shaders. Por supuesto, la implementación de algunas funcionalidades implicarán la modificación de la aplicación, pero esto se mantendrá acotado.

4.1. Estructura

Para cada proyecto de ejemplo encontraremos una serie de archivos con un objetivo concreto, los cuales describiremos a continuación:

- **Aplicación:** Los archivos *main.cpp*, *hello_vulkan.h* y *hello_vulkan.cpp*, son los correspondientes al código de cada aplicación. Dentro de *main.cpp* se inicializan los pipelines, se modifica la interfaz gráfica de usuario, se construye la escena a partir de archivos .obj y sus matrices de transformación (Figura 4), etc. En los archivos restantes simplemente se definen y declaran las funciones relacionadas al *ray tracing* en Vulkan.

```
// Creation of the example

//helloVk.loadModel(nvh::findFile("media/scenes/CornellBox-Sphere.obj", defaultSearchPaths, true));

{ //cornell dragon
    helloVk.loadModel(nvh::findFile("media/scenes/CornellBox-Empty-CO.obj", defaultSearchPaths, true));
    helloVk.loadModel(nvh::findFile("media/scenes/dragon.obj", defaultSearchPaths, true),
                      glm::translate(
                          glm::rotate(
                              glm::scale(glm::mat4(1.0f), vec3(1.5,1.5,1.5)), (float)1.5, vec3(0, 1, 0)),vec3(0, 0.5, 0)));
}
```

Figura 4: Ejemplo de creación de escena, cargando archivos .obj, con sus matrices de transformación

En la carpeta *shaders* encontraremos lo siguiente:

- **Host_device.h:** Este archivo está incluido tanto por los shaders como por la aplicación, por lo que en él se establecen definiciones que deben ser conocidas por ambas partes: la aplicación que corre en la PC (host) y los shaders ejecutados por la GPU (device). Se utiliza por ejemplo, para la transferencia de estructuras de datos, en la que tanto la aplicación y los shaders deben ser conscientes del formato de estas.
- **Post.frag shader:** Este shader se ejecuta con la imagen obtenida del pipeline de *ray tracing* como input. Por defecto, no hace nada, asigna a un píxel el color que ya tiene, pero su propósito es ser utilizado para la implementación de un tone mapping, por ejemplo, una corrección gamma.
- **shaders de ray tracing (raytrace.rgen, raytrace.rchit y raytrace.rmiss) y rasterizado (vert_shader.vert y frag_shader.frag):** Obviamente en cada ejemplo se encontrarán los shaders correspondientes al pipeline de *ray tracing*, además siempre tendremos disponibles shaders de rasterizado con sombreado de *Phong* para renderizar nuestra escena con este algoritmo si lo deseamos.
- **Raycommon.glsl:** En este archivo se define la estructura del **payload** del rayo, este debe ser incluido en todos los shaders.

4.2. ImGui

También tenemos la posibilidad de utilizar la librería **ImGui** (ya integrada en la aplicación), esta nos permitirá cambiar parámetros de la escena en tiempo real, desde una interfaz gráfica (ver Figura 5) y es muy fácil de usar. Se recomienda ver los ejemplos en su documentación [6] e indagar en el código de los ejemplos.

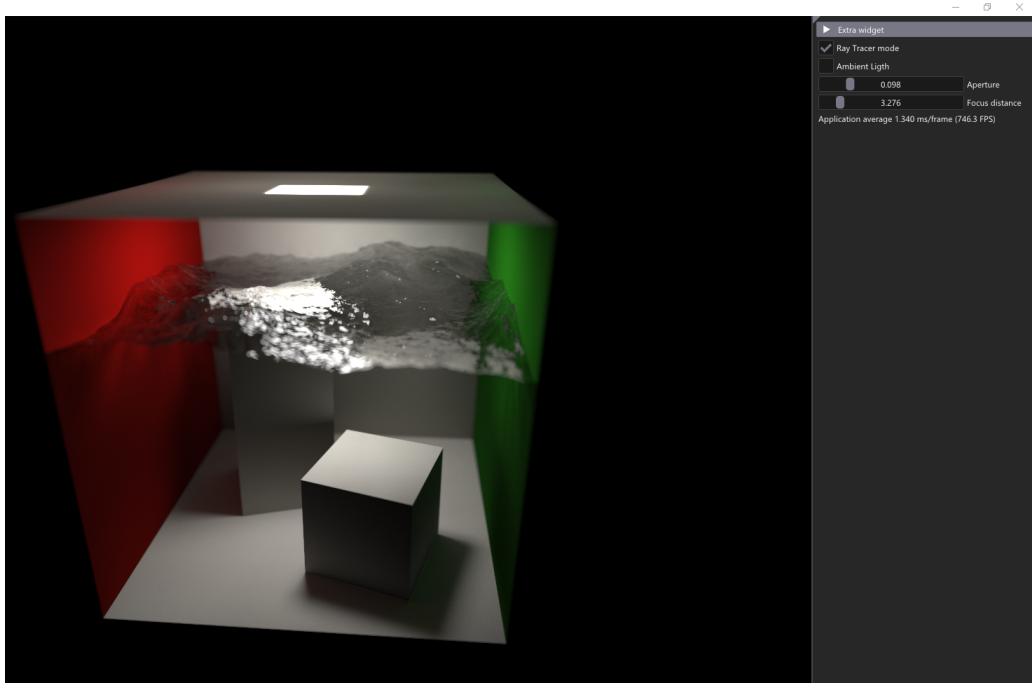


Figura 5: Ejemplo de interfaz gráfica que permite cambiar la distancia focal, la apertura de la cámara y el modelo de iluminación

Los ejemplos ya vienen con una GUI básica, que permite cambiar parámetros de la cámara, algoritmo de renderizado, entre otras cosas. Para activarla se debe descomentar el código correspondiente del archivo *main.cpp*, similar al de la imagen (Figura 6)

```
// Show UI window.
if(helloVk.showGui()) {
    ImGuiH::Panel::Begin();
    renderUI(helloVk);

    ImGui::Checkbox("Ray Tracer mode", &useRaytracer); // Switch between raster and ray tracing
    ImGui::Checkbox("Ambient Light", &helloVk.m_pcRay.ambientLigh); //enable ambient light
    ImGui::SliderFloat("Aperture", &helloVk.m_pcRay.camAperture, 0.001f, 0.5f); //camera parameters
    ImGui::SliderFloat("Focus distance", &helloVk.m_pcRay.focusDist, 1.f, 20.f);

    ImGui::Text("Application average %.3f ms/frame (%.1f FPS)", 1000.0f / ImGui::GetIO().Framerate, ImGui::GetIO().Framerate);
    ImGuiH::Panel::End();
}
```

Figura 6: Código que genera el panel de la Figura 5

4.3. Ejemplos

Estos ejemplos fueron desarrollados para la realización del presente informe. Se recomienda revisar y repasar las secciones anteriores mientras analizan los códigos de los siguientes ejemplos.

4.3.1. [Vk1_launchID](#) Visualización LaunchID

Este ejemplo es el más simple de todos. Solo utiliza el *ray-generation shader*. Este se ejecuta en paralelo para cada píxel de la imagen final, con distintos valores para la variable *gl_LaunchIDEXT*, por lo que este ejemplo nos ayuda a visualizar esto, modulando los canales rojo y verde del píxel, con el valor de la variable *gl_LaunchIDEXT*.



Figura 7: Captura de pantalla del ejemplo vk1_launchID

4.3.2. [Vk2_primitiveID](#) Visualización PrimitiveID

A continuación, se introducen distintos elementos, el *closest-hit shader*, el *miss shader* y el uso del **payload** del rayo para transferir el color desde estos shaders al *ray-generation shader*. El *miss shader* se encarga exclusivamente de asignar el color negro al fondo, mientras que el *closest-hit* asigna un color correspondiente al valor de primitiveID.

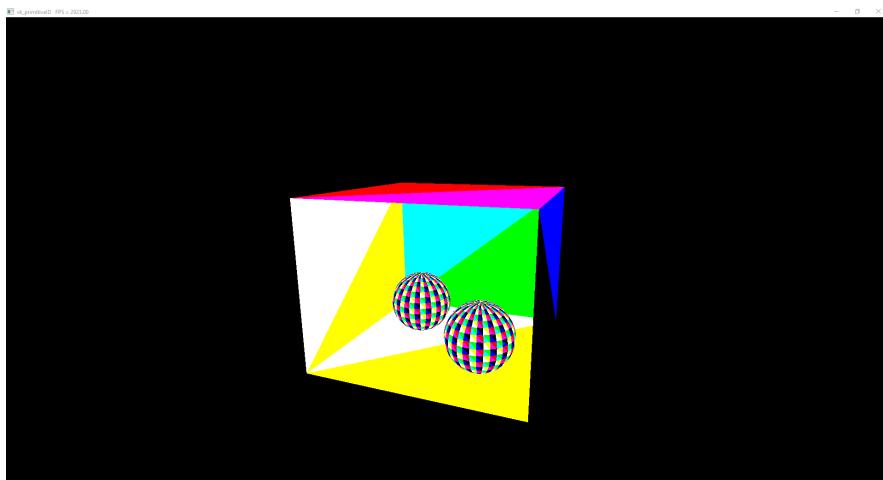


Figura 8: Captura de pantalla del ejemplo vk2_primitiveID

4.3.3. [Vk3_material](#) Evaluar materiales

Por otro lado mostraremos como procesar una intersección, accediendo a los buffers de materiales, vértices, normales, y otros datos transferidos al *closest-hit* shader, como texturas, posición de la luz, etc. Estos datos están disponibles para su uso en el cuerpo del shader. En este caso se utilizan para realizar un sombreado difuso.

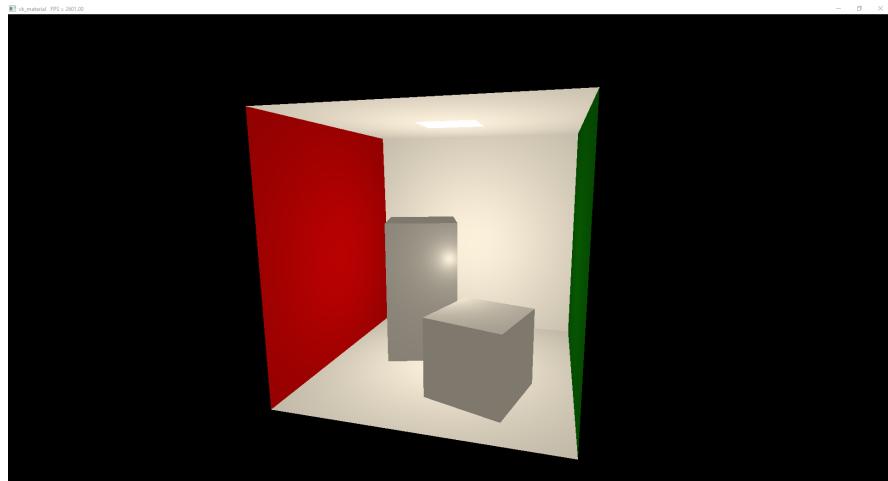


Figura 9: Captura de pantalla del ejemplo vk3_material

4.3.4. [Vk4_rebotes](#) Rebote de un rayo

Pasemos ahora a agregar reflejos en ciertos materiales. En este ejemplo se expandió el **payload** del rayo para que pueda guardar el lugar de impacto, y la dirección con que se va a reflejar el rayo, esta información es comunicada al *ray-generation shader* para trazar un segundo rayo a partir de estos parámetros.

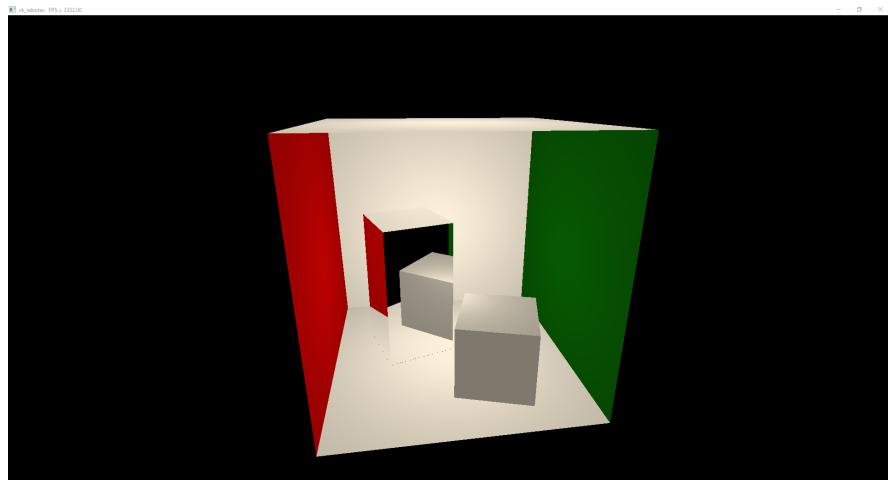


Figura 10: Captura de pantalla del ejemplo vk4_rebotes

4.3.5. [Vk5_random](#) Números aleatorios

Este ejemplo es el más complejo. En él se agrega un archivo *random.gsls* con funciones para la generación de un número semilla y de números aleatorios de punto flotante. Además se agrega el número de frame que se está procesando, en el archivo *host_device.h*, esto nos permitirá transferirlo desde la aplicación, y así utilizarlo en combinación con la variable *LaunchID* para generar una semilla distinta por cada píxel y cada frame (Figura 11).

```
uint seed = InitRandomSeed(gl_LaunchIDEXT.y * gl_LaunchSizeEXT.x + gl_LaunchIDEXT.x, pcRay.frame);
```

Figura 11: Generación de semilla

En el ejemplo, en lugar de trazar cada rayo primario en dirección al centro del píxel, se utiliza la semilla para trazar el rayo en una dirección aleatoria en un entorno del centro del píxel. Como ya mencionamos, esto cambia píxel a píxel y frame a frame, generando el efecto de ruido en la imagen (Figura 12).



Figura 12: Zoom en la escena de vk5_random

4.4. `Vk_path_tracer` Path tracer

Con el fin de demostrar el potencial de los recursos existentes en el repositorio, se elaboró un path tracer basado en el libro Ray Tracing in One Weekend [7], en él se implementan técnicas y efectos como: *defocus blur*, *gamma correction*, materiales *glossy*, *fresnel effect*, *gaussian filter antialiasing*, etc.

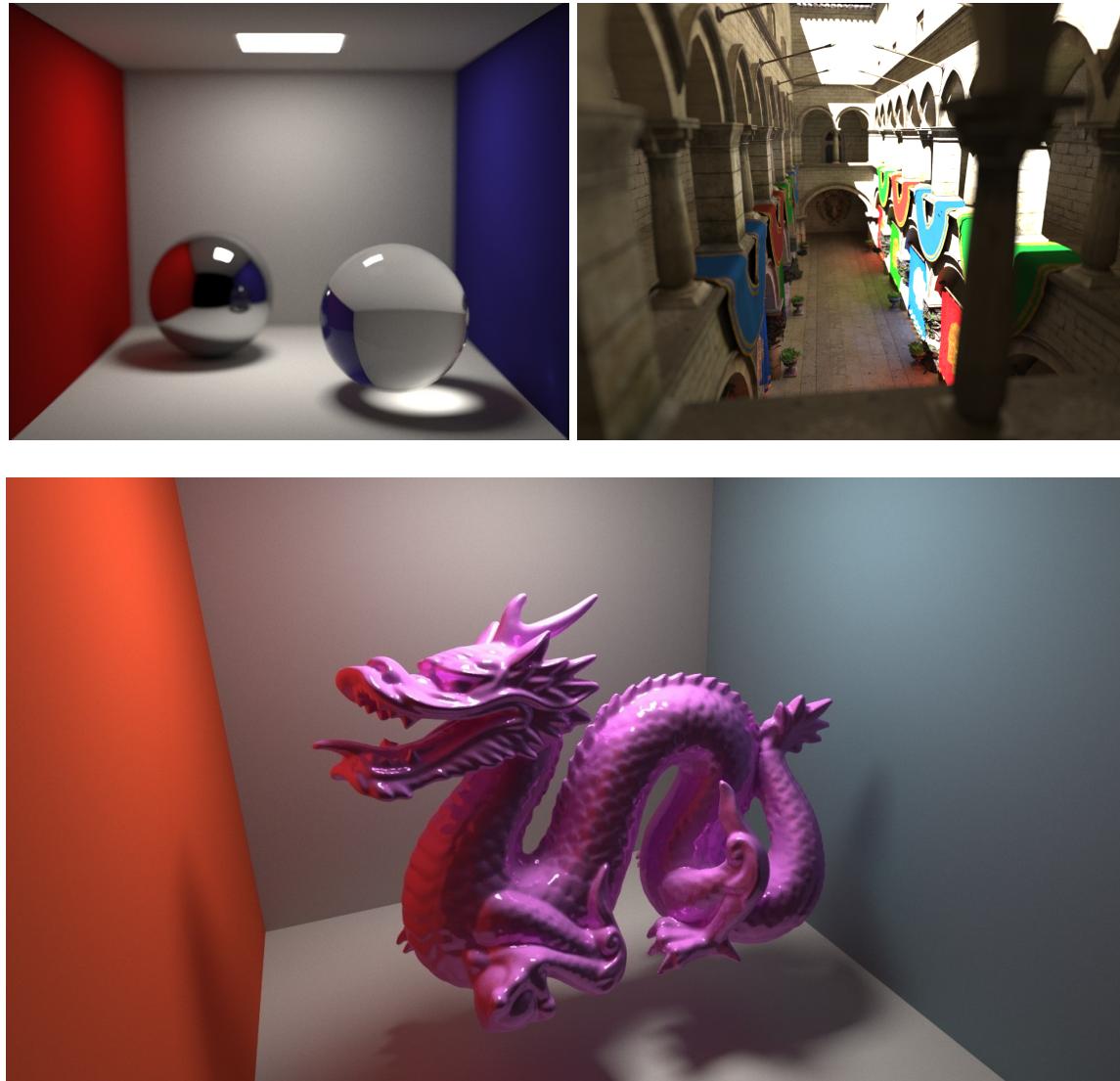


Figura 13: Resultados obtenidos por `vk_path_tracer`

El path tracing es un algoritmo utilizado en la renderización de imágenes por computadora para simular de manera realista el comportamiento de la luz en una escena tridimensional. Funciona trazando múltiples rayos de luz que se propagan desde la cámara a través de la escena, interactuando con los objetos y las superficies. Cada rayo de luz, al alcanzar una su-

perficie, puede ser absorbido, reflejado o refractado según las propiedades del material. Para simular la iluminación global y los efectos de luz indirecta, el algoritmo path tracing sigue recursivamente estos rayos, hasta alcanzar una condición de terminación, como un límite de profundidad o un objeto que emite luz. A través de este proceso iterativo y acumulando muestras para cada pixel, el path tracing calcula el color y la intensidad de cada pixel de la imagen, produciendo resultados visualmente realistas con efectos de sombras suaves, reflexiones y refracciones precisas.

5. Estudio de performance

Dado que Vulkan es multiplataforma, para conocer el desempeño de distintas GPU's aplicadas en raytracing, se realizó un estudio de performance de la aplicación vk_path_tracer para GPU's de Nvidia y AMD. Las dos tarjetas (Nvidia RTX 4060 y AMD Radeon RX 7600) son similares en cuanto a los *benchmarks* publicados en UserBenchmark [8].

El estudio de performance consistió en la comparación de fps al ejecutar el programa Vk_path_tracer con una escena cerrada que cuenta con 871 mil triángulos, 2,6 millones de vértices y materiales translúcidos (Figura 14 izquierda), donde a su vez se varía la cantidad máxima de rebotes de los caminos manteniendo fija la posición de la cámara en todas las pruebas (Cuadro 1). Por otro lado se realizaron pruebas fijando a 10 la cantidad de rebotes, pero cambiando la cantidad de tráigalos de la escena, para esto se utilizó una escena de 36 triángulos y 106 vértices (Figura 14 derecha) y la escena anterior (Cuadro 2).

Rebotes	Nvidia	AMD	Rendimiento de Nvidia sobre AMD
7	128 fps	45 fps	+184 %
10	79 fps	29 fps	+172 %
70	10.5 fps	4 fps	+163 %
100	7.7 fps	3 fps	+157 %
500	3 fps	1.5 fps	+100 %

Cuadro 1: Resultado de las pruebas con distinto numero de rebotes

Número de triángulos	Nvidia	AMD	Rendimiento de Nvidia sobre AMD
36	395 fps	243 fps	+63 %
871.000	114 fps	40 fps	+185 %

Cuadro 2: Resultado de las pruebas con distinto numero de triángulos en la escena

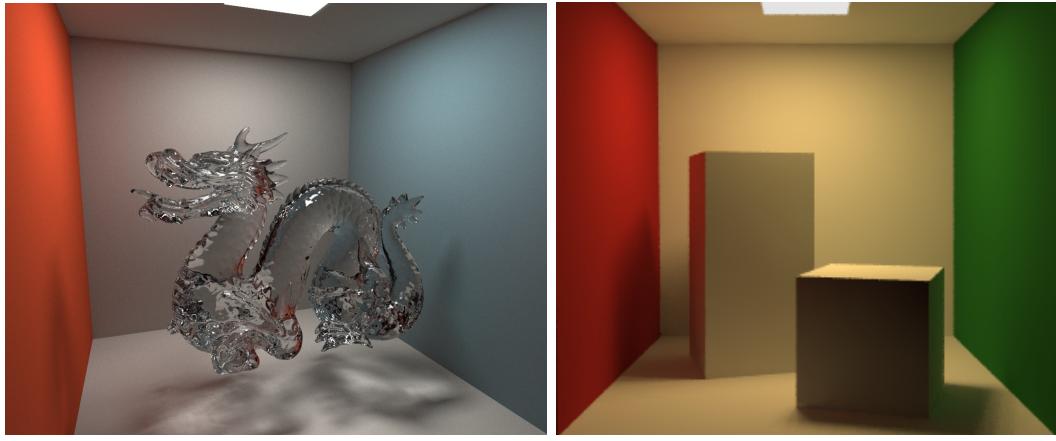


Figura 14: Escenas utilizadas en pruebas (escena con 817 mil triángulos a la izquierda, escena con 36 triángulos a la derecha)

Como podemos observar, si bien, en los *benchmarks* las dos tarjetas presentan resultados similares, se observa que: en particular para su aplicación en ray tracing, Nvidia da resultados significativamente superiores a AMD en cuanto a performance.

Un resultado adicional a resaltar es observado cuando la escena está compuesta por más de un archivo .obj: el comportamiento no se ajusta a lo que uno esperaría de una escena donde el número de triángulos debería ser la suma de los triángulos de cada objeto. Se sugiere que este fenómeno pueda atribuirse a una mala optimización en la construcción de la estructura de aceleración por parte del usuario (no por Vulkan). En otras palabras, es probable que en el código de carga del archivo se esté enviando el archivo completo a Vulkan para crear una estructura de aceleración, cuando en realidad sería más eficiente enviar mallas separadas de manera lógica.

6. Lecturas adicionales

Para terminar, dejamos tutoriales de Nvidia que pueden ser de utilidad para implementar funcionalidades complementarias o los shaders que no fueron utilizados en los ejemplos (*anyhit e intersection shader*) [9] [10] [11] [12]. Los ejemplos vistos anteriormente son una extensión del código base del que parten estos tutoriales, por lo que es muy fácil seguirlos e incorporarlos a nuestro proyecto si fuera necesario.

Referencias

- [1] Overvoorde, A. (2023). *Vulkan Tutorial*, <https://vulkan-tutorial.com>.
- [2] Khronos Group, *Vulkan*, consultado el 25 de abril del 2024, <https://www.vulkan.org>.
- [3] *Gpuinfo*, consultado el 25 de abril del 2024, https://vulkan.gpuinfo.org/listdevicescoverage.php?extension=VK_KHR_ray_tracing_pipeline.
- [4] Nvidia (2022). *nvpro_core*, GitHub, consultado el 25 de abril del 2024, https://github.com/nvpro-samples/nvpro_core/tree/master.
- [5] Nvidia (2022). *nvk_helper*, GitHub, consultado el 25 de abril del 2024, https://github.com/nvpro-samples/nvpro_core/blob/master/nvvk/README.md.
- [6] *Imgui* (2022). GitHub, consultado el 25 de abril del 2024, <https://github.com/ocornut/imgui?tab=readme-ov-file>.
- [7] Shirley, P., Black, D., & Hollasch, S. (2023). *Libro Ray Tracing in One Weekend* (Edición v4.0.0-alpha.1). <https://raytracing.github.io/>.
- [8] *UserBenchmark*, consultado el 25 de abril del 2024, <https://gpu.userbenchmark.com/>.
- [9] Nvidia (2022). *Intersection Shader - Tutorial*, GitHub, consultado el 25 de abril del 2024, https://github.com/nvpro-samples/vk_raytracing_tutorial_KHR/tree/master/ray_tracing_intersection.
- [10] Nvidia (2022). *Any Hit Shader - Tutorial*, GitHub, consultado el 25 de abril del 2024, https://github.com/nvpro-samples/vk_raytracing_tutorial_KHR/tree/master/ray_tracing_anyhit.
- [11] Nvidia (2022). *Callable Shader - Tutorial*, GitHub, consultado el 25 de abril del 2024, https://github.com/nvpro-samples/vk_raytracing_tutorial_KHR/tree/master/ray_tracing_callable.
- [12] Nvidia (2022). *Specialization Constants*, GitHub, consultado el 25 de abril del 2024, https://github.com/nvpro-samples/vk_raytracing_tutorial_KHR/tree/master/ray_tracing_specialization.
- [13] Nvidia (2022). *NVIDIA Vulkan Ray Tracing Tutorial*, GitHub, consultado el 25 de abril del 2024, https://nvpro-samples.github.io/vk_raytracing_tutorial_KHR/vkrt_tutorial.md.html.
- [14] Nvidia (2022). *NVIDIA vk_mini_path_tracer Tutorial*, GitHub, consultado el 25 de abril del 2024, https://nvpro-samples.github.io/vk_mini_path_tracer/index.html.
- [15] Marrs, A., Shirley, P. & Wald, I. (2021). *Ray Tracing Gems II*. Apress. <https://link.springer.com/book/10.1007/978-1-4842-7185-8>.