

PLACAS GRÁFICAS – RAY TRACING

João Sá

ISPGAYA-Portugal

José Lopes

ISPGAYA-Portugal

Leonel Ribeiro

ISPGAYA-Portugal

Manuel Marques

ISPGAYA-Portugal

Resumo:

O Ray Tracing é uma técnica de renderização que simula a iluminação em 3D, reproduzindo o comportamento natural da luz, e é considerado o método mais preciso para criar efeitos realistas, como sombras e reflexos. Utilizado em placas gráficas avançadas, como no *Cyberpunk 2077*, proporciona gráficos imersivos, mas com elevado impacto no desempenho e na memória gráfica. A insuficiência de memória pode causar perdas de fluidez. A Inteligência Artificial (IA) complementa esta tecnologia, reduzindo a carga computacional com redes neuronais treinadas em milhares de cenários, criando atalhos que otimizam os cálculos intensivos. Em conjunto, Ray Tracing e IA permitem alcançar gráficos de alta qualidade e realismo, mantendo um equilíbrio entre desempenho e eficiência, transformando as experiências visuais.

Palavras-chave: placas gráficas, ray tracing, inteligência artificial

GRAPHICS CARDS – RAY TRACING

Abstract:

Ray Tracing is a rendering technique that simulates 3D lighting, replicating natural light behavior, and is considered the most precise method for realistic effects like shadows and reflections. Used in advanced graphics cards, as seen in *Cyberpunk 2077*, it delivers immersive graphics but significantly impacts performance and memory usage. Insufficient memory can lead to reduced fluidity. Artificial Intelligence (AI) enhances this technology by alleviating computational loads through neural networks trained on thousands of scenarios, creating shortcuts to optimize intensive calculations. Together, Ray Tracing and AI achieve high-quality, realistic graphics while balancing performance and efficiency. This synergy transforms visual experiences, making cutting-edge rendering accessible without compromising system functionality.

Keywords: graphics card, ray tracing, artificial intelligence

Introdução

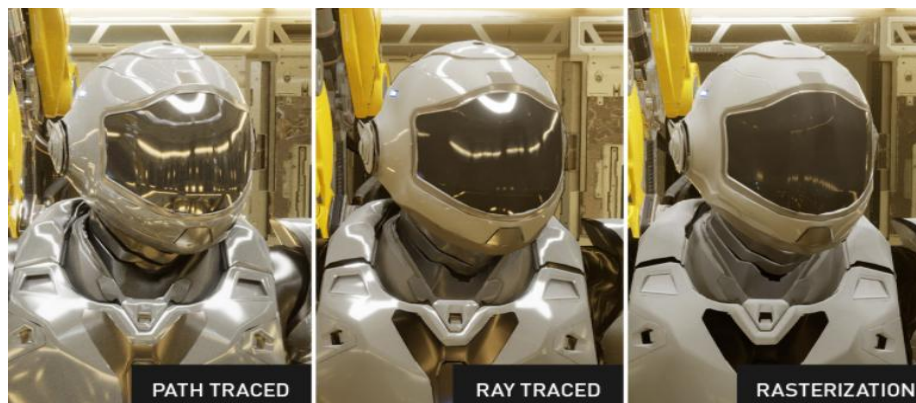
Unidades de processamento gráfico (GPU) sempre foram a origem de grandes avanços tecnológicos para o espaço 3D e as possibilidades quase sempre foram limitadas pelo hardware e não pelo software. Gerando um nível de procura e desenvolvimento bastante alto entre as duas principais empresas de fabrico de GPUs (AMD e NVIDIA). Um dos avanços que revolucionou este mundo tecnológico foi a tecnologia Ray Tracing em tempo real.

Ray Tracing é uma tecnologia de renderização gráfica que usa traçados de raios para criar efeitos de luz mais realistas. A técnica está presente em GPUs modernas e melhora o aspeto visual de jogos e animações. (Tecnoblog - Ray Tracing, 2020)

O método de Ray Tracing já existia antes deste avanço, mas não era possível ter uma animação com alta taxa de fotogramas com este cálculo avançado de linhas de luz, seria preciso esperar e deixar o GPU calcular o cenário, isto só para obter uma imagem. O que a NVIDIA apresentou foi uma forma de obter esta tecnologia sem as desvantagens incutidas nela.

Figura 7

Exemplo de Ray Tracing vs Rasterização



Fonte: (What Is Path Tracing? | NVIDIA Blog, n.d.)

As raízes do traçado de raios remontam ao passado, tirando inspiração nos métodos engenhosos de Albrecht Dürer. David Luebke, Vice-Presidente de Pesquisa Gráfica da NVIDIA, gosta de começar a história no século XVI com Albrecht Dürer, uma das figuras mais importantes do renascimento do norte da Europa, que usou cordas e pesos para replicar uma imagem 3D em uma superfície 2D. (Caulfield, 2022) O momento crucial na evolução das técnicas de renderização chega com as contribuições inovadoras de Jim Kajiya em 1986. Um dos primeiros trabalhos no caminho da formalização do processo de renderização foi (Kajiya, 1986), que define a Equação de Renderização (Tori, 2018). A introdução da equação de renderização por Kajiya marcou uma fusão entre a computação gráfica e a física, elucidando o comportamento da luz nas cenas. O engenhoso algoritmo de seguimento de caminhos, inspirado em estudos de transferência de calor radiante, enfrentou o desafio de resolver a complexa equação de renderização empregando técnicas estatísticas como a integração de Monte Carlo. Graças aos avanços significativos nas técnicas de renderização e ao aumento do poder computacional, o algoritmo de traçado de caminhos, também chamado de traçado de raios com Monte Carlo, tornou-se o método de renderização predominante usado em campos que priorizam o realismo, como na produção de filmes e na visualização foto realista de exames médicos. (Faria Dutra Fragoso, 2022)

Queremos realmente avaliar o quão longe o Ray Tracing nos pode levar, ver como está no presente, o que o possibilita e também o que poderá melhorar e que áreas de trabalho podem beneficiar desta tecnologia.

A base do Ray Tracing em tempo real é construir hardware focado no tipo de cálculos necessários para ter uma imagem fiel á realidade. Deste modo o GPU pode fazer outras tarefas sem

atrasos. Enquanto temos os núcleos principais do GPU a fazer a rasterização para transformar uma imagem vetorial numa imagem rasterizada, também vamos ter núcleos RT (dedicados ao Ray Tracing) para fazer a simulação de cada raio de luz sem rasterizar e depois o GPU junta as duas camadas numa só para a imagem final.

Alguns exemplos de como o Ray Tracing em tempo real está a ser usado atualmente:

- Jogos: títulos como o “Forza Horizon 5” e “Cyberpunk 2077” beneficiam desta tecnologia para fornecer uma simulação fiel á realidade com iluminação global, reflexos, sombras e iluminação dinâmica de objetos.
 - Cyberpunk 2077 is already one of the most technologically advanced games available, using several ray tracing techniques to render its neon-illuminated environments and vast Night City visuals at incredible levels of detail. (NVIDIA - Cyberpunk, 2023)
- Filmes: alguns filmes e séries recentes como o “The Mandalorian” e o “Spider-Man: No Way Home” usaram o Ray Tracing para gerar visuais gerados pelo computador autênticas e realistas de forma a gerar uma experiência harmoniosa para o expectador.
- Arquitetura: para ajudar a criar modelos 3D realistas com a iluminação nos vários tipos de materiais usados nos edifícios e estruturas.

A integração de abordagens baseadas em Inteligência Artificial (IA) na reconstrução de raios surgiu como um fator de mudança, ao melhorar significativamente a precisão e a eficácia dos processos de cálculo das linhas de luz.

A IA, em particular os modelos de aprendizagem automática, tem demonstrado um potencial notável no aperfeiçoamento das técnicas de reconstrução de linhas de luz. Estes modelos tiram partido de vastos conjuntos de dados e de processos de aprendizagem iterativos para decifrar padrões complexos nos dados apresentados, permitindo a reconstrução de imagens visualmente coerentes e sem ruído a partir de amostras de raios esparsos.

Estas técnicas de redução de ruído baseadas em IA utilizam redes neuronais treinadas em conjuntos de dados extensos para remover eficazmente os artefactos de ruído e as imperfeições inerentes às imagens com raios de luz. Uma imagem gerada usando poucos raios apresenta bastante ruído, como uma foto tirada em um ambiente com pouca luz. Mas aplicando o “AI denoiser” (um filtro de remoção de ruído baseado em Inteligência Artificial) a imagem sai nítida e

limpa como se tivesse sido processada traçando milhares de raios por pixel. (*Precisamos Falar Sobre Ray Tracing* | NVIDIA, 2021).

Evolução das placas gráficas

No início das placas gráficas, na década de 1970, surgiram dispositivos rudimentares projetados principalmente para gráficos 2D. Nesse contexto, destacam-se placas como a IBM 3270, pioneira na manipulação de elementos gráficos simples. A verdadeira revolução, no entanto, veio na década de 1990, com a introdução das primeiras placas aceleradoras 3D. A 3dfx Voodoo, lançada em 1996, tornou-se um ícone ao proporcionar experiências visuais tridimensionais inovadoras, marcando o início da era das GPUs dedicadas.

O início do século XXI testemunhou uma intensa competição entre NVIDIA e ATI (posteriormente adquirida pela AMD). A série GeForce, da NVIDIA, e a família Radeon, da ATI, dominaram o mercado, introduzindo GPUs mais poderosas e eficientes. No entanto, a verdadeira revolução no presente é liderada pela NVIDIA com a introdução da arquitetura Turing e suas placas RTX. A NVIDIA GeForce RTX 20 e 30 séries, com sua tecnologia de Ray Tracing, representam um marco na busca por gráficos mais realistas em tempo real.

Olhando para o futuro das placas gráficas, a tecnologia de Ray Tracing continuará a desempenhar um papel central. A NVIDIA, com suas GPUs Ampere, e a AMD, com a arquitetura RDNA, estão na vanguarda da corrida pela excelência gráfica. Superar os desafios computacionais do Ray Tracing em tempo real permanece uma meta crucial. No horizonte, a promessa de placas gráficas mais eficientes e poderosas abre caminho para experiências visuais ainda mais imersivas, não apenas em jogos, mas também em simulações e design. O futuro da computação gráfica promete ultrapassar as fronteiras atuais com a utilização da inteligência artificial, explorando novas possibilidades e levando a tecnologia de placas gráficas a patamares inéditos.

Integração da Inteligência Artificial na Evolução do Ray Tracing

Além das inovações tecnológicas discutidas anteriormente, a Inteligência Artificial (IA) emerge como uma força impulsionadora significativa na evolução contínua da tecnologia de Ray Tracing. A aplicação estratégica da IA oferece oportunidades promissoras para otimizar a eficiência e a qualidade do processamento gráfico em tempo real.

A IA pode ser usada para otimizar algoritmos de Ray Tracing, tornando-os mais eficientes e rápidos. Algoritmos de aprendizagem de máquina podem analisar padrões complexos nos

cálculos de iluminação e refração, identificando atalhos e aprimorando a execução de tarefas computacionais intensivas.

A redução de serrilhados (anti-aliasing) é crucial para criar imagens visualmente atraentes. Técnicas avançadas baseadas em IA, como o uso de redes neurais convulsionais, podem aprimorar significativamente a qualidade da renderização, proporcionando resultados mais suaves e naturais.

A resolução é um fator essencial na qualidade visual. Técnicas de up-scaling baseadas em IA, como o DLSS (Deep Learning Super Sampling), podem aumentar a resolução de maneira eficiente, mantendo a fidelidade visual. Isso é especialmente valioso em situações onde o Ray Tracing pode impactar o desempenho.

A IA pode ser usada para aprimorar a simulação de materiais, capacitando as placas gráficas a reproduzirem texturas e reflexões de maneira mais precisa. Redes neurais podem aprender a representar características complexas de materiais, contribuindo para ambientes virtuais mais realistas. ((Watson, 2020)

A IA pode monitorar dinamicamente as necessidades gráficas em tempo real e ajustar automaticamente as configurações de Ray Tracing para equilibrar o desempenho e a qualidade visual. Isso permite uma experiência de usuário mais fluida, adaptando-se às variações na complexidade da cena.

Ao integrar inteligência artificial à tecnologia de Ray Tracing, abre-se uma janela de oportunidades para superar desafios computacionais e proporcionar experiências visuais ainda mais envolventes e realistas. À medida que a IA continua a evoluir, sua contribuição para aprimorar a computação gráfica, especialmente no contexto do Ray Tracing, promete definir os padrões para as futuras gerações de placas gráficas.

Comparação Ray Tracing vs Rasterização

Para podermos realmente perceber as diferenças principais entre o Ray Tracing e a Rasterização decidimos fazer um estudo com recurso a um jogo previamente referido, Cyberpunk 2077.

Este jogo tem servido como plataforma para demonstrar as tecnologias mais avançadas no mundo do Ray Tracing e então seria a melhor escolha para avaliar as alterações visuais e o seu reflexo no desempenho do GPU.

Dentro do Cyberpunk 2077 existe uma ferramenta que simula um cenário no jogo durante

uns segundos, neste cenário a camara virtual percorre desde o interior do edifício até ao exterior, servindo como referência para testar não só o desempenho dos fotogramas com as várias definições gráficas do menu, como também, podemos observar os gráficos e as respetivas mudanças.

Figura 8: Menu de Configurações Cyberpunk 2077



Nota: imagem capturada no menu de configurações gráficas dentro do jogo Cyberpunk 2077.

Dentro das definições podemos escolher as várias categorias do Ray Tracing, ao ativar, temos também a escolha de usar a tecnologia do Ray Reconstruction apoiada pela rede neuronal para melhorar a fidelidade visual, mais a frente iremos comparar as diferenças que esta tecnologia proporciona ao utilizador.

i. Path Tracing

Apesar do Ray Tracing se basear na simulação dos raios de luz, dentro do jogo, esta tecnologia não é a única a ser usada, mesmo com a definição ativada, ainda são usados alguns métodos tradicionais para renderizar cada cenário dentro do jogo. O Path Tracing é Ray Tracing, técnicas anteriores abordaram separadamente sombras por Ray Tracing, reflexões e iluminação global para um pequeno número de fontes de luz. O Ray Tracing completo modela as propriedades da luz de um número virtualmente ilimitado de fontes emissivas, fornecendo sombras fisicamente corretas, reflexos e iluminação global em todos os objetos. (*Uma Prévia Da Nova Tecnologia Ray Tracing: Overdrive Mode Chegará Ao Cyberpunk 2077 Dia 11 de Abril | Notícias GeForce | NVIDIA, 2023*).

ii. Diferenças visuais em reflexões

Foi escolhida uma zona em que fosse possível verificar o efeito do Ray Tracing nas reflexões. Esta zona noturna com luzes néon de várias cores demonstra bem o efeito.

Podemos ver no letreiro cor-de-rosa que é muito mais nítido na Fig.3 do lado esquerdo, também verificamos que as reflexões das pessoas a caminhar estão menos perceptíveis na Fig.3, do lado direito, pois não são baseadas nos traçados dos raios de luz mas em métodos mais tradicionais e menos dinâmicos.

Figura 9: *Reflexões com Ray Tracing vs Sem Ray Tracing*



Nota: Imagem retirada dentro do jogo e no mesmo local e hora do mapa.

iii. Diferenças visuais em sombras

Esta escadaria metálica perfurada demonstra o potencial do Ray Tracing em simular os pequenos detalhes nas sombras. Na Fig.4, as perfurações estão bem visíveis na sombra debaixo do primeiro degrau.

Figura 10: *Sombras com Ray Tracing*



Nota: Imagem retirada dentro do jogo, na mesma hora para obter as sombras a partir do mesmo ângulo da fonte de luz.

iv. Diferenças visuais com o Ray Reconstruction

Com Ray Reconstruction, é possível observar que as reflexões são mais nítidas sem perdas de desempenho. Portanto, este é um dos benefícios da AI com o Ray Tracing. A simulação dos raios causa bastante ruído na imagem por causa da dispersão da luz e por sua vez a imagem final fica com artefactos provenientes do ruído.

“O Ray Reconstruction reconhece padrões de iluminação a partir de dados de treino, como a iluminação global ou a oclusão ambiente, e recria-os no jogo à medida que joga. Os resultados são superiores à utilização de denoisers ajustados manualmente.” [Traduzido do Inglês] (Burns, 2023)

Figura 11: *Diferenças Visuais com o Ray Reconstruction*



Nota: Imagem retirada enquanto estava a ser usada a ferramenta de teste de desempenho fornecida pelo próprio jogo.

v. Desempenho

Para medir o desempenho, foi usada uma ferramenta de medição de fotogramas por segundo (fps) denominada por CapFrameX. (*CapFrameX*, n.d.)

Para manter os dados consistentes e comparáveis, os testes foram executados sempre no mesmo computador, apenas com o jogo (*Cyberpunk 2077*) a executar em primeiro plano. Como anteriormente mencionado, foi usada a ferramenta de simulação dentro das definições do jogo para manter os testes dentro do mesmo grau de exigência computacional.

Especificações do computador:

- CPU: AMD Ryzen 5 5600 (6 núcleos)
- RAM: 16GB DDR4
- GPU: NVIDIA RTX 3060 TI 8GB

Foram feitos 3 tipos de testes com as seguintes definições:

- Rasterização
- Path Tracing
- Path Tracing + Ray Reconstruction

Em todos estes testes foi utilizada a tecnologia DLSS da NVIDIA para melhorar o desempenho do GPU através de Machine Learning. Esta tecnologia também é importante para tornar o Ray Tracing numa realidade permitindo equilibrar as reduções no desempenho com a otimização do up-scaling.

O desempenho do jogo está diretamente relacionado com os FPS do jogo e, portanto, quando maior o número de FPS melhor o desempenho. O melhor desempenho num jogo caracteriza-se numa melhor perceção de fluidez do movimento pois o tempo de espera entre cada fotograma gerado pelo GPU é reduzido.

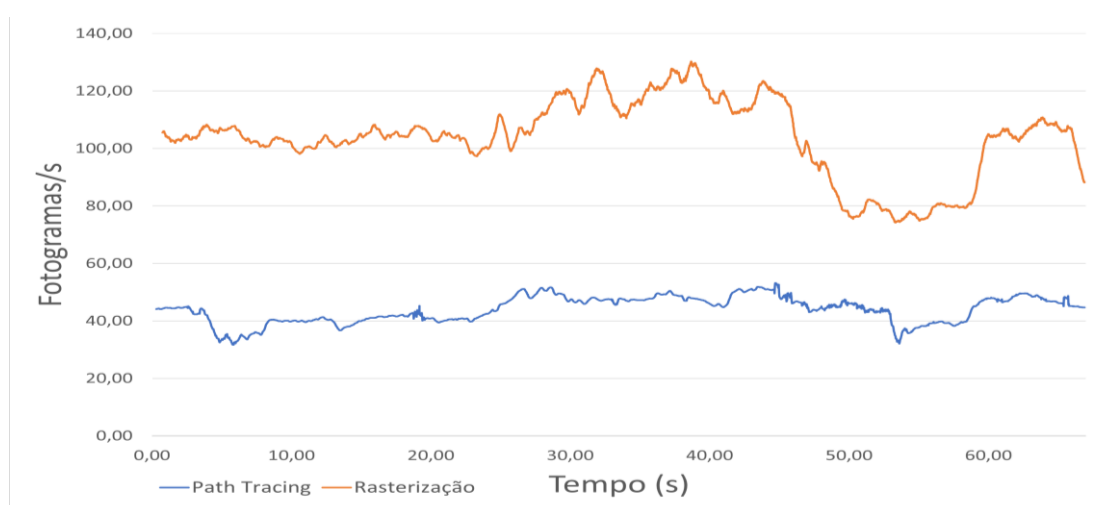
Outro parâmetro que pode reduzir o desempenho é a utilização de memória gráfica (VRAM), “o comportamento dos jogos pode variar quando a VRAM é realmente utilizada ao máximo. Em alguns, a memória é eliminada, causando uma falha de desempenho perceptível enquanto a cena atual é recarregada na memória. Noutros, apenas serão carregados e descarregados dados selecionados, sem impacto visível. E, em alguns casos, os novos recursos podem ser carregados mais lentamente, uma vez que estão a ser trazidos da RAM do sistema”. (A

Deeper Look At VRAM On GeForce RTX 40 Series Graphics Cards, 2023)

Para comparar os resultados dos testes realizados, foi elaborado um gráfico de linhas com os dados obtidos em cada teste. Estes dados são a média dos fotogramas por segundo no eixo vertical e o instante de tempo em segundos no eixo horizontal.

A diferença mais notável tanto em termos gráficos como em desempenho seria entre a rasterização e o Path Tracing.

Figura 12: *Fotogramas por segundo médio Path Tracing vs Rasterização*

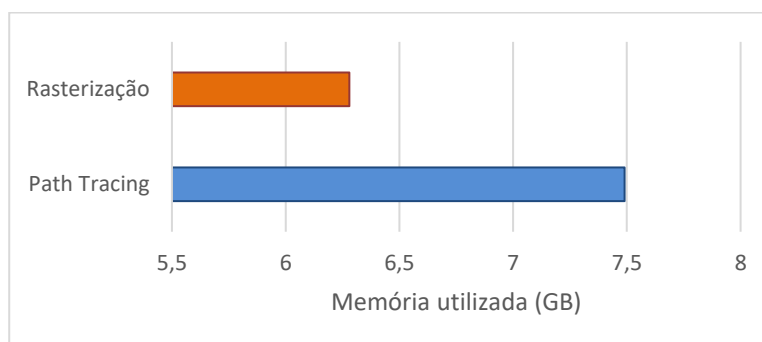


Fonte: Dados obtidos a partir do software CapFrameX

De acordo com a Fig.6 podemos inferir que existe uma diferença nominal entre os 2 tipos de processamento gráfico, em que a rasterização supera com praticamente o dobro de fotogramas por segundo em comparação com o Path Tracing. Enquanto a rasterização teve uma média global de 103,06 fotogramas por segundo (FPS), o Path Tracing teve apenas uma média de 44,24 FPS. Como a rasterização tem uma complexidade visual inferior, é de esperar que esta tenha melhor desempenho visto que usa menos recursos do GPU.

Em relação á utilização da memória gráfica, existe uma correlação entre a maior quantidade utilizada e o Path Tracing.

Figura 13: Média de utilização de memória gráfica



Fonte: Dados obtidos a partir do software CapFrameX

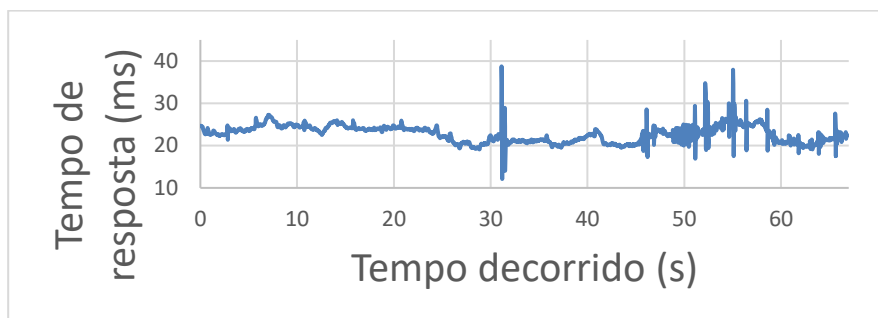
Neste cenário pode-se apontar que ao ativar o Path Tracing a quantidade média de memória utilizada é 7,49 GB, enquanto com a rasterização é 6,28 GB. Isto representa um aumento de 19,26% ao escolher ativar o Path Tracing.

Sendo neste caso que a quantidade máxima de memória disponível é 8GB, com o Path Tracing, a utilização encontra-se perto do limite e possivelmente reduzindo mais o desempenho. Isto pode-se verificar pelo tempo de resposta entre cada fotograma.

O tempo de resposta entre cada fotograma (frametime) mostra o tempo total em que o GPU ficou a renderizar cada fotograma individual e por isso quanto menor for o número maior a fluidez da imagem. No entanto, a consistência deste tempo também é importante pois pode dar a perceção que o jogo “encrava” ou tem solavancos. Quando um fotograma demora mais tempo a aparecer no ecrã, o fotograma anterior fica mais tempo visível ao observador, dando a sensação de que houve uma pausa no movimento. Tal como a água num rio, com o movimento consistente aparenta ser mais fluida do que num rio com várias correntes diferentes.

Na Fig.8 podemos visualizar o frametime durante todo o teste com o Path Tracing ativo. Durante o teste foi observado no jogo pequenos solavancos afetando a fluidez do movimento.

Figura 14: Tempo de Resposta Entre Cada Fotograma com Path Tracing

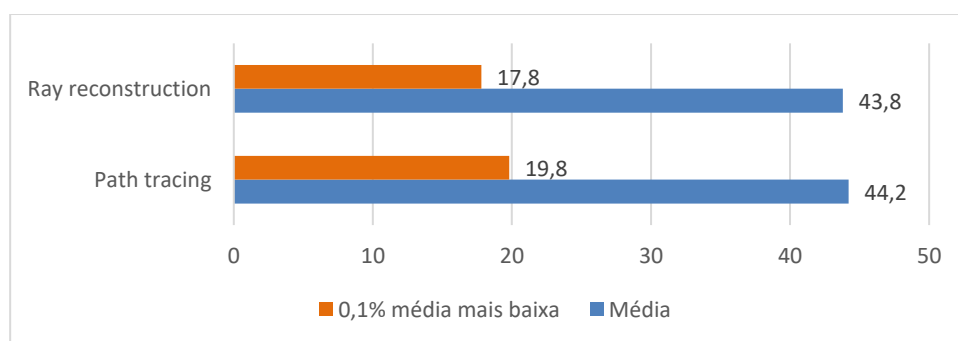


Fonte: Dados obtidos a partir do software CapFrameX

Avaliando os valores na Fig.8, existem certos momentos no tempo, durante o teste, em que a variação de frametime aumenta consideravelmente entrando em linha com a observação visual do jogo durante os testes. Uma das causas para isto acontecer pode ser de facto a utilização da memória gráfica se aproximar do limite como foi demonstrado na Fig.7.

Durante os testes não foi perceptível a diferença de desempenho entre usar o Ray Reconstruction ou não. Na visualização dos testes juntamente com a Fig.9 podemos verificar os benefícios da Inteligência Artificial, fornecendo uma maior nitidez nas reflexões visualmente sem sacrificar o desempenho.

Figura 15: Comparação de desempenho Ray Reconstruction vs Path Tracing



Fonte: Dados obtidos a partir do software CapFrameX

Conclusão

Refletindo sobre os resultados obtidos no nosso estudo, o Ray Tracing está bastante próximo de ser um requisito no futuro. As perdas de desempenho podem ser notáveis, mas as diferenças visuais causam um grande impacto na percepção do realismo pelo utilizador final.

Adicionando o poder da Inteligência Artificial, com modelos treinados para otimizar o ruído da imagem construída com a simulação de raios de luz, podemos dar um grande salto não só em jogos como também noutras áreas que possa ser útil, como por exemplo na Arquitetura e Modelos 3D. Também podemos ter a possibilidade de aplicar esta tecnologia não só à simulação de luz como também ao som, sendo uma onda, visto que a luz é em simultâneo uma partícula e uma onda.

A única limitação está ao nível do hardware do GPU, sendo esta tecnologia intensivamente focada em simulação, com o passar do tempo será mais acessível se seguirmos a observação da Lei de Moore, em que, a densidade de transístores duplica a cada 2 anos e por sua vez a capacidade computacional dos componentes vai aumentando.

Bibliografia

- A Deeper Look At VRAM On GeForce RTX 40 Series Graphics Cards.* (2023, May 18). NVIDIA. <https://www.nvidia.com/en-gb/geforce/news/rtx-40-series-vram-video-memory-explained/>
- Burns, A. (2023, August 23). *NVIDIA DLSS 3.5: Enhancing Ray Tracing With AI; Coming This Fall To Alan Wake 2, Cyberpunk 2077: Phantom Liberty, Portal with RTX & More.* NVIDIA. <https://www.nvidia.com/en-us/geforce/news/nvidia-dlss-3-5-ray-reconstruction/>
- CapFrameX. (n.d.). CapFrameX. Retrieved January 18, 2024, from <https://www.capframex.com/>
- Caulfield, B. (2022, May 10). *O que é Path Tracing? | Blog da NVIDIA.* Blog oficial NVIDIA Brasil. <https://blog.nvidia.com.br/2022/05/10/o-que-e-path-tracing/>
- Cyberpunk 2077: Technology Preview Of New Ray Tracing Overdrive Mode Out Now.* (2023, April 11). NVIDIA. <https://www.nvidia.com/en-us/geforce/news/cyberpunk-2077-ray-tracing-overdrive-update-launches-april-11/>
- Faria Dutra Fragoso, B. (2022). *IMPLEMENTAÇÃO DE UM RENDERIZADOR COM TRAÇADO DE CAMINHOS* [MESTRE EM INFORMÁTICA, PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO]. <https://doi.org/10.17771/PUCRio.acad.61873>
- O que é ray tracing e como essa tecnologia na GPU deixa os jogos mais realistas.* (2020). Tecnoblog. <https://tecnoblog.net/responde/o-que-e-ray-tracing/>
- Precisamos falar sobre Ray Tracing | NVIDIA.* (2021). <https://www.nvidia.com/pt-br/drivers/prbr-05282018/>
- Tori, R. (with Hounsell, M. da S.). (2018). *Introdução a realidade virtual e aumentada.* Sbc.
- Uma prévia da nova tecnologia Ray Tracing: Overdrive Mode chegará ao Cyberpunk 2077 dia 11 de Abril | Notícias GeForce | NVIDIA.* (2023, April 11). <https://www.nvidia.com/pt-br/geforce/news/cyberpunk-2077-ray-tracing-overdrive-update-launches-april-11/>
- Watson, A. (2020). *Deep Learning Techniques for Super-Resolution in Video Games* (arXiv:2012.09810). arXiv. <https://doi.org/10.48550/arXiv.2012.09810>