Graphics Processor Unit – Unidade de Processamento Gráfico.

Curso: Análise e desenvolvimento de sistemas - 1° Semestre

Nome: Gabriel Henrique Rosa.

Caracterização da GPU:

A GPU surgiu para "aliviar" o processador principal do computador (CPU) da pesada tarefa de gerar imagens. Por isso, é capaz de lidar com um grande volume de cálculos matemáticos e geométricos, condição trivial para o processamento de imagens 3D (utilizadas em jogos, exames médicos computadorizados, entre outros).

Para que as imagens possam ser geradas, a GPU trabalha executando uma sequência de etapas, que envolvem elaboração de elementos geométricos, aplicação de cores, inserção de efeitos e assim por diante. Essa sequência, de maneira bastante resumida, consiste no recebimento pela GPU de um conjunto de vértices (o ponto de encontro de dois lados de um ângulo); no processamento dessas informações para que elas obtenham contexto geométrico; na aplicação de efeitos, cores e afins; e na transformação disso tudo em elementos formados por pixels (um pixel é um ponto que representa a menor parte de uma imagem), processo conhecido com *rasterização*. O passo seguinte é o envio dessas informações à memória de vídeo (*frame buffer*) para que então o conteúdo final possa ser exibido na tela.

As GPUs podem contar com vários recursos para a execução dessas etapas, entre eles:

- Pixel Shader: *shader* é um conjunto de instruções utilizado para o processamento de efeitos de renderização de imagens. Pixel Shader, portanto, é um programa que trabalha com a geração de efeitos com base em pixels. Esse recurso é amplamente utilizado em imagens 3D (de jogos, por exemplo) para gerar efeitos de iluminação, reflexo, sombreamento.
- Vertex Shader: semelhante ao Pixel Shader, só que trabalha com vértices em vez de pixels. Assim sendo, Vertex Shader consiste em um programa que trabalha com estruturas formadas por vértices, lidando, portanto, como figuras geométricas. Esse recurso é utilizado para a modelagem dos objetos a serem exibidos;
- Render Output Unit (ROP): basicamente, manipula os dados armazenados na memória de vídeo para que eles se "transformem" no conjunto de pixels que formará as imagens a serem exibidas na tela. Cabe a essas unidades a aplicação de filtros, efeitos de profundidade, entre outros;
- Texture Mapping Unit (TMU): trata-se de um tipo de componente capaz de rotacionar e redimensionar bitmaps (basicamente, imagens formadas por conjuntos de pixels) para aplicação de uma textura sob uma superfície.

Função da GPU:

As GPUs são as peças responsáveis pelo processamento de imagem nos nossos computadores. Elas possuem componentes de hardware especializados em trabalhar com imagens 2D ou 3D, em operações como cálculo de geometria, mapeamento de texturas, transformação de pixels, entre outros.

As imagens 3D que vemos nos jogos e aplicativos são tradicionalmente criadas a partir do que chamamos de gráficos vetoriais. A GPU plota uma série de coordenadas (chamadas de vértices), utilizando a conexão entre esses pontos para formar os polígonos de uma cena. No processo, ela aplica texturas, iluminação e uma série de detalhes complementares a essas formas.

Depois de calcular toda a cena no espaço virtual, a GPU realiza o processo de rasterização, convertendo a cena em dados de pixels. Essas informações são então enviadas para os nossos monitores, onde a imagem final é exibida. Esse processo se repete inúmeras vezes por segundo.

Diferenças entre GPU e CPU:

Apesar de o processador (CPU) e a GPU serem construídos a partir de bilhões de transistores, a forma como eles são estruturados internamente e suas funções se diferem consideravelmente.

Os processadores (CPUs) precisam lidar com uma série de tarefas distintas e geralmente não repetitivas, que vão desde planilhas de Excel às mais variadas rotinas do sistema operacional. Isso exige versatilidade e uma compatibilidade mais ampla. A execução desses processos também acontece de forma seriada, uma por vez.

Já as placas de vídeo são unidades extremamente especializadas. Sua função é primordialmente o processamento gráfico (2D ou 3D). Para isso, ela usa centenas de unidades de processamento menores, todas trabalhando em funções similares, como a renderização de polígonos, texturas e a conversão das informações em pixels.

As GPUs atacam um grande volume de tarefas simultaneamente, executando aquilo que chamamos de processamento paralelo.

CPU (Processador):

- Unidade central de processamento
- Diversos Núcleos
- Foco em baixa latência de operação
- Bom para processamento em série
- Pode realizar diversas operações de uma vez

GPU (Placa gráfica):

- Unidade de processamento gráfico
- Muitos núcleos
- Foco em altíssima taxa de transferência
- Bom para processamento paralelo
- Pode processar milhares de operações de uma só vez

Fornecedores:

Há vários fabricantes de GPU no mercado, mas as empresas do ramo mais conhecidas são NVIDIA, AMD e INTEL, sendo que as duas primeiras são as mais populares no que se refere a chips gráficos mais sofisticados.

Há empresas como, Gigabyte, Asus, Zorac, XFX, entre outras que fabricam as placas, mas não produzem GPUs. Cabe a elas inserir GPUs em suas placas assim como outros recursos, tais como memória e conectores. Por outro lado, é importante frisar também que GPUs podem estar embutidas diretamente em placas-mãe de computadores, sendo popularmente chamadas de "placas de vídeo onboard" neste casos.

Processamento 3D:

Nos dias atuais, é praticamente impossível falar de placas de vídeo sem considerar gráficos 3D. Esse tipo de recurso é essencial para a indústria do entretenimento, onde filmes e jogos em 3D fazem grande sucesso. Quanto mais avançadas forem as imagens de uma aplicação, mais processamento gráfico será necessário.

Conceitos relacionados ao processamento 3D:

- Fillrate: consiste na medida da quantidade de pixels que o chip gráfico é capaz de renderizar por segundo, sendo também chamado de "*pixel fillrate*". Em outras palavras, trata-se da medição do processamento de pixels. Isso ocorre porque, quando uma imagem 3D é gerada, ela deve posteriormente ser transformada em 2D para que possa ser visualizada na tela, tarefa que se dá pela transformação das informações em pixels. Note que também existe o "*texel fillrate*", que mede, também por segundo, a capacidade da placa de vídeo de aplicar texturas;
- Frames per Second (FPS): como o nome indica, o FPS indica a quantidade de *frames* por segundo que é exibida na tela. Quando um filme é executado, por exemplo, ele é, na verdade, composto por uma sequência de imagens, como se fossem várias fotografias seguidas. Cada uma dessas "fotografias" é um frame. Via de regra, quanto maior a quantidade de frames por segundo, melhor é a percepção do usuário para os movimentos que vê. Se o FPS for muito baixo, o usuário terá a impressão de que as imagens estão dando pequenas "congeladas" na tela. O FPS é especialmente importante em jogos, onde lentidão na geração da imagem pode prejudicar o desempenho do jogador. Para games, o ideal é que o FPS seja de, pelo menos, 30 FPS, assim é possível obter o mínimo de qualidade visual. No entanto, muitas vezes a geração de imagens é um trabalho tão pesado que o número de FPS cai. Nestes casos, o usuário pode desabilitar certos efeitos. Com placas de vídeo potentes, é possível manter o FPS em uma taxa mais elevada, sem que efeitos tenham que ser desativados. Daí o fato de muita gente preferir placas que lidam, por exemplo, com 100 FPS, pois quando essa taxa cair, a quantidade disponível ainda será satisfatória;
- V-Sync: uma maneira de tornar a exibição de imagens mais confortável aos olhos humanos é ativando o V-Sync. Trata-se de um recurso que sincroniza a taxa de FPS com a frequência de atualização do monitor de vídeo (*refresh rate*). Essa medida informa quantas vezes por segundo a tela atualiza a exibição de imagens. Se for 60 vezes, por exemplo, seu refresh rate é de 60 Hz. A sincronização pode fazer com que as imagens sejam exibidas com maior "naturalidade", pois ajuda a evitar efeitos desconfortáveis, como o *tearing*, que geralmente ocorre quando o FPS é maior que o refresh rate do monitor, causando uma sensação de "rasgo" em imagens bastante movimentadas;

- Antialiasing: este é um recurso extremamente importante para melhorar a qualidade da imagem a ser exibida. Muitas vezes, por causa de limitações de resolução do monitor, objetos 3D aparecem na tela com as bordas "tremidas", como se alguém tivesse passado uma tesoura por ali de maneira bem grosseira. Os filtros antialiasing conseguem amenizar esse problema de maneira bastante satisfatória, mas podem exigir muitos recursos de processamento;
- Anisotropic Filtering: também conhecido pela sigla *AF*, esse recurso permite eliminar ou amenizar o efeito de "imagem borrada" existente em texturas, principalmente quando estas representam superfícies inclinadas. Essa sensação piora quando o usuário aproxima a visualização da textura (como se aplicasse um "zoom"). O AF consegue reduzir esse problema a tal ponto que as imagens não só ficam mais nítidas como também apresentam melhor sensação de profundidade. O Anisotropic Filtering veio em substituição ao *Bilinear Filtering* e ao *Trilinear Filtering*, que tinham a mesma finalidade, mas não apresentavam a mesma eficiência.

Vantagens e Desvantagens:

Vantagens:

Permitem que os usuários montem sua própria configuração, definindo a quantidade de memória e ajustando acordo com a sua necessidade. Como elas utilizam um processador independente, o do computador fica livre para cuidar de outras tarefas e o desempenho geral do equipamento fica bem melhor. Também traz suporte a alguns recursos que não são encontrados em modelos onboard, como gráficos em 3D, por exemplo.

Desvantagens:

É necessário fazer um certo investimento para fazer uso de uma GPU no computador, pois seu custo de aquisição é elevado. Além do custo, é preciso ter um conector especifico disponível na placa mãe para conectá-la.

Bibliografia:

https://www.meupositivo.com.br/doseujeito/tecnologia/o-que-e-placa-de-video-gpu/https://www.infowester.com/placavideo.php