

# 1

## Introdução

A formação de reservatórios naturais de petróleo se dá pela acumulação de hidrocarbonetos em rochas sedimentares. A indústria de petróleo extrai esses recursos perfurando poços no domínio do reservatório, que são conectados a uma linha de produção na superfície para processamento e armazenamento. Dado o uso de pressões naturais e induzidas, força-se os fluidos como o óleo e o gás a fluir na direção dos poços e em seguida para a superfície.

A indústria de petróleo vem utilizando a simulação numérica computacional desses reservatórios na tentativa de prever o fluxo dos fluidos ao longo dos seus domínios. As estimativas de produção ao longo de dezenas de anos são utilizadas num processo de busca pelas melhores disposições e planos de produção para os poços, com o objetivo de maximizar o valor econômico do reservatório ao longo do seu tempo de operação.

O modelo de simulação empregado atualmente pela indústria é definido por uma malha discreta de células hexaédricas cobrindo o domínio do reservatório, que é organizada numa grade tridimensional contendo  $n_i \times n_j \times n_k$  células. A geometria associada às células é no entanto bastante irregular, devido à presença de falhas geológicas, que causam descontinuidades na elevação, e devido a algumas células serem consideradas inativas, resultando em grupos irregulares e possivelmente desconectados de células. O reservatório é caracterizado pela associação de propriedades físicas às células de sua malha, como pressão e saturação de óleo, gás e água. Essas propriedades variam conforme o tempo, sendo gravadas para cada passo da simulação numérica.

A simulação numérica desse modelo produz portanto conjuntos grandes de dados, que devem ser analisados em ferramentas de visualização 2D e 3D. Estas devem prover diversas formas de analisar o comportamento do reservatório ao longo do tempo, o que é feito pela inspeção cuidadosa dos campos escalares e vetoriais associados às propriedades das células. As ferramentas de visualização também devem prover uma renderização eficiente do modelo, dada pela geração contínua de quadros (imagens finais do modelo) com taxas interativas.

Essas ferramentas de visualização podem se beneficiar do uso de uma representação compacta e hierárquica da malha de reservatório. O uso de estruturas hierárquicas permite a redução da complexidade de vários algoritmos que operam sobre a malha do reservatório. Esse tipo de estrutura é explorado por diversas técnicas de aceleração conhecidas na literatura de renderização em

tempo real, como por exemplo o descarte de primitivas por volume de visão ou oclusão. Essas técnicas conseguem identificar rapidamente conjuntos grandes de primitivas que não contribuem para a imagem final do modelo, dada a posição e direção de observação e os parâmetros de câmera utilizados.

No Brasil, dois fatores têm causado recentemente uma demanda por um enorme aumento no tamanho do modelo de simulação. A recente descoberta de reservatórios gigantes localizados na camada de pré-sal do país tem forçado a criação de modelos que cobrem uma larga extensão territorial. Além disso, há uma demanda crescente pelo uso de uma maior precisão numérica na simulação, motivada pela geração de resultados com um grau maior de certeza em relação ao problema real.

Os avanços recentes na tecnologia de simulação paralela de reservatórios naturais de petróleo têm começado a atender a essa demanda ao permitir a simulação numérica de domínios cada vez mais discretizados (45). No entanto, essas simulações, que atualmente têm utilizado modelos que estão chegando na ordem de dezenas de milhões de células, passam a produzir um volume de dados sem precedentes, inviabilizando o uso de técnicas convencionais de visualização para a análise desses resultados. Esse contexto introduz uma grande demanda pelo desenvolvimento de soluções escaláveis de visualização.

A necessidade de se visualizar dados tão complexos introduz diversos problemas computacionais que precisam ser tratados para renderizar o modelo com taxas interativas. O primeiro grande obstáculo é a impossibilidade de se armazenar todas as estruturas de dados necessárias para representar o modelo na memória principal de um só computador. Além do problema de armazenamento e representação, a quantidade de dados a serem visualizados excede largamente a capacidade de processamento dos sistemas gráficos convencionais, como os processadores gráficos de prateleira (*GPUs*).

Apesar das técnicas convencionais de aceleração permitirem uma redução significativa do número de primitivas gráficas enviadas para o *pipeline* gráfico, o seu uso isolado não é nem minimamente suficiente para lidar com modelos de geometria massivamente complexa. Há duas áreas de pesquisa em computação gráfica com foco na obtenção de taxas interativas na renderização de modelos muito complexos: renderização distribuída e renderização com multi-resolução.

O uso de sistemas de renderização distribuída visa tratar o problema aumentando a capacidade total de processamento disponível. Esses sistemas fazem uso do poder combinado de um conjunto de processadores comuns (*CPUs*) e gráficos (*GPUs*) interligados por um canal de comunicação de alta velocidade. Tipicamente, o esforço computacional para a geração de um quadro é particionado entre esses diversos processadores. Ao subdividir e paralelizar

a solução do problema, permite-se que cada processador gráfico lide com um (sub)modelo de médio porte. Isso introduz uma capacidade de processamento e armazenamento que cresce linearmente com a quantidade de computadores e processadores gráficos utilizados. No entanto, como toda solução paralela, é necessário lidar com problemas de sincronização de dados, balanceamento de carga e uso eficiente do canal de comunicação entre os processadores, fatores que afetam diretamente a escalabilidade da solução (37, 33, 60, 53, 2).

Os modelos de geometria muito complexa possuem um grande número de primitivas gráficas, que quando projetadas para a tela ocupam poucos *pixels* ou até menos que um *pixel*. Esse fato é explorado pelas técnicas de renderização com multi-resolução e níveis de detalhe, cujo objetivo principal é prover a imagem mais fiel ao modelo original que for possível dados os limites do *hardware* disponível. Esses limites são dados pelo volume de dados armazenados em memória principal e processados pelo *pipeline* gráfico a cada quadro. Essas técnicas armazenam em memória secundária (disco) uma representação hierárquica contendo diversas simplificações do modelo original. Dado que é possível extrair uma boa aproximação do modelo original da estrutura armazenada em disco, utiliza-se os recursos computacionais disponíveis para representar a malha original do modelo da melhor forma possível.

Ambas as linhas de tratamento do problema possuem vantagens e desvantagens quando comparadas. Os sistemas de renderização distribuída permitem que os algoritmos de visualização e técnicas de aceleração sejam aplicados diretamente em um modelo de médio porte, o que simplifica de forma significativa a parte de tratamento do modelo. Outra grande vantagem de reduzir a complexidade local de cada modelo é a facilidade de se gerar uma imagem com 100% de fidelidade ao modelo original, o que é um dos grandes desafios de técnicas de multi-resolução.

No entanto, a solução distribuída pode possuir um custo alto, devido à sua infraestrutura complexa de *hardware*: múltiplos PCs, múltiplas GPUs e uma conexão possivelmente cara, apesar de, no contexto da indústria de petróleo, a existência de grandes centros de processamentos de dados ser uma realidade. A solução como um todo é mais difícil de manter não só pela complexidade de *hardware* como pela complexidade de *software*, pois ela é normalmente composta de um número de componentes clientes e servidores.

Sistemas de visualização distribuída também herdam alguns problemas que ocorrem em praticamente toda solução paralela. O primeiro problema é o de balanceamento de carga: é desejável que o modelo seja particionado igualmente entre os nós do agrupamento. Ainda assim, sabemos que diversos outros fatores influenciam o desempenho de sistemas gráficos, como o estágio de raste-

rização, a eficiência de técnicas de aceleração, entre outros, gerando inevitavelmente um desbalanceamento entre os processadores gráficos. O uso do canal de comunicação entre os processadores também deve ser feito de forma eficiente. O tratamento de entradas do usuário e a consequente troca de parâmetros de visualização devem ser sincronizados entre todos os componentes distribuídos relevantes. Por último, o problema mais difícil: a escalabilidade da solução. A estratégia de renderização paralela por partição do modelo requer a transmissão e composição de um número de imagens parciais que aumenta conforme são adicionados processadores gráficos à solução do problema. Isso pressiona as partes do sistema responsáveis pela transmissão e composição de resultados parciais, limitando a escalabilidade do sistema.

Por sua vez, os sistemas de visualização com multi-resolução permitem utilizar uma única estação de trabalho para visualizar um modelo gigante, que é uma arquitetura de *hardware* muito mais simples que a de soluções por visualização distribuída. A garantia de um tamanho máximo para a representação do modelo também possui diversas aplicações, como a possibilidade de se limitar o uso de memória e de se forçar uma taxa de renderização mínima, praticamente garantindo uma interatividade contínua no uso do sistema. Como essa solução trabalha com a troca entre esforço computacional e qualidade de representação, ela permite uma grande escalabilidade, pois o detalhe de representação muitas vezes só é perceptível para partes do modelo próximas ao observador, e idealmente será oferecido tanto detalhe quanto houver recurso computacional para visualizá-lo.

Apesar de possuir vantagens muito atraentes quando comparadas a soluções distribuídas, os sistemas de multi-resolução também possuem sérias desvantagens. A topologia original das células que compõem o modelo é modificada, o que precisa ser tratado em algoritmos que utilizam essa informação. Um cuidado especial deve ser tomado para que não haja perda na qualidade da imagem, visto que haverá simplificação do suporte geométrico do modelo. Como é necessário lidar com uma grande quantidade de dados para a criação de uma estrutura de multi-resolução, essa solução sofre de longos tempos de pré-processamento e gera representações grandes em armazenamento secundário. Dada essa conjuntura, é necessário otimizar a criação da estrutura e a sua posterior transferência do disco para a memória principal.

Além disso, um problema específico da área de visualização científica se apresenta em tais sistemas: o modelo, que possui um suporte geométrico gigantesco, composto por dezenas a centenas de milhões de células, também possui diversas propriedades físicas associadas a cada uma das células ou vértices do modelo. Cada uma dessas propriedades pode ainda possuir valores

para cada passo da simulação numérica. Torna-se então fundamental tratar da independência entre a estrutura de multi-resolução e as propriedades associadas ao modelo, visando a possibilidade de se reutilizar essa estrutura ao longo de diversas simulações diferentes do mesmo reservatório.

Este trabalho apresenta em primeiro lugar uma breve introdução sobre o modelo de simulação de reservatórios de petróleo, incluindo o processo de simulação do modelo e seus dados associados. Após apresentadas as técnicas de visualização comumente empregadas para analisar os resultados de uma simulação, é feita uma proposta para a representação e visualização de modelos de médio porte. Essa proposta apresenta as técnicas convencionais de aceleração adaptadas para o contexto de reservatórios, o que permite obter uma visualização eficiente para modelos com poucos milhões de células. Em seguida, esse trabalho apresenta soluções para a visualização interativa de modelos massivos de reservatórios naturais de petróleo seguindo cada uma das duas linhas de pesquisa aqui mencionadas.

## 1.1 Contribuições

Na linha de renderização distribuída, este trabalho propõe um sistema distribuído para a renderização de modelos massivos de reservatório de petróleo. Seguindo algumas tendências na área de computação de alto desempenho, o sistema foi projetado para permitir a visualização de tais modelos em clientes leves em redes corporativas, disponibilizando o poder agregado de processamento de um agrupamento de PCs de prateleira (*off-the-shelf*), onde cada PC é equipado com múltiplas GPUs. O sistema proposto utiliza uma estratégia de ordenação no fim para oferecer uma melhor escalabilidade. Dado um uso eficiente de cada GPU e um estágio de composição parcial em cada nó do agrupamento, nossa proposta trata bem os problemas de escalabilidade típicos de sistemas que utilizam agrupamentos de médio a grande porte. O sistema proposto (1) traz as seguintes contribuições:

- permite a conexão de clientes leves e utiliza técnicas que aumentam o grau de paralelismo entre os diversos componentes do sistema;
- tira proveito de nós com múltiplas GPUs, dada a nossa proposta de efetuar um estágio de composição parcial de imagens em cada nó;
- escala bem com o tamanho do modelo. Os experimentos feitos em um agrupamento de médio porte demonstraram que o sistema consegue lidar com modelos com dezenas de milhões de células;

Os resultados experimentais obtidos comprovam a eficácia das técnicas apresentadas e a escalabilidade da solução proposta.

Já na linha de técnicas de multi-resolução, este trabalho apresenta uma proposta para a geração e representação hierárquica de múltiplos níveis de detalhe da malha do modelo de reservatório. A nossa proposta traz as seguintes contribuições:

- é apresentado um novo algoritmo de simplificação feito especialmente para malhas de reservatórios de petróleo;
- é proposta uma estimativa menos conservadora para a percepção do erro geométrico de cada simplificação quando comparada a trabalhos anteriores;
- é proposto um algoritmo para a renderização com multi-resolução que garante uma taxa mínima de renderização, efetivamente oferecendo o máximo de detalhe geométrico dados os limites do *hardware* disponível.

Após a avaliação experimental das decisões tomadas no desenho do algoritmo de simplificação, é comprovada experimentalmente a melhoria de desempenho obtida pela estimativa menos conservadora do erro de simplificação. Por fim, é discutido o desempenho do sistema, que nos experimentos computacionais efetuados garantiu uma taxa de renderização mínima de 30 quadros por segundo na visualização de modelos com até 240 milhões de células ativas. Esse resultado permitirá que se lide com modelos com grande extensão territorial e grandes discretizações de domínio, garantindo a análise dos resultados em ferramentas 3D pelos próximos anos.

Este trabalho propõe ainda um conjunto de técnicas de mapeamento do *grid* original e das propriedades associadas às células da malha original do reservatório sobre o suporte geométrico criado pelo nosso algoritmo de simplificação. Todas as técnicas têm como base o mapeamento de texturas sobre a malha do reservatório, que podem ser construídas rapidamente em tempo de visualização. Com isso, a estrutura de multi-resolução se torna independente das propriedades associadas à grade topológica do modelo, permitindo o reuso dessa estrutura em diversas simulações de um mesmo reservatório, o que é uma característica considerada fundamental em termos da praticidade para o uso da ferramenta proposta.

## 1.2

### Organização do Documento

O restante deste trabalho está estruturado da seguinte forma. No Capítulo 2 é feita uma recapitulação dos trabalhos relacionados às áreas de aceleração da visualização em tempo real, de visualização distribuída e de visualização com multi-resolução. Em seguida, no Capítulo 3, é apresentado o modelo de simulação de reservatórios naturais de petróleo, as técnicas de visualização usualmente empregadas para analisar o resultado da simulação e uma proposta para a visualização de modelos de médio porte. O Capítulo 4 traz a nossa proposta para a renderização distribuída de modelos massivos de reservatório. A nossa proposta para a visualização com multi-resolução segue no Capítulo 5. O Capítulo 6 apresenta a nossa proposta para o mapeamento do *grid* original e das propriedades associadas às células da malha original do reservatório sobre as malhas simplificadas geradas pelo nosso sistema de multi-resolução. Por fim, o Capítulo 7 traz as conclusões da pesquisa realizada e os trabalhos futuros.