1
Introdução

Esta tese apresenta propostas para processamento paralelo adequadas à implementação em placas gráficas para um conjunto selecionado de problemas de visão computacional. A estrutura desta tese segue a formulação em camadas de abstração de Marr sobre como os problemas de processamento de informação devem ser explorados (Mar80).

David Marr foi um neurocientista e psicólogo que reuniu resultados interdisciplinares das áreas de psicologia, inteligência artificial e neurofisiologia para estabelecer uma estrutura que descreve sistemas complexos de processamento de informação.

Seguindo a proposição de Marr (Mar80), a compreensão de um sistema complexo de processamento de informações deve ser analisada em diferentes níveis de abstração. Propôs quatro níveis de abstração, correspondendo respectivamente (do nível mais alto de abstração para o mais baixo): à formulação teórica do problema, ao estabelecimento dos algoritmos para resolver tal problema, à análise de mecanismos para operação do algoritmo, e aos circuitos e componentes (nível do hardware) pelos quais tais mecanismos são compostos.

Ao repensar problemas de visão computacional com o intuito de propor formulações paralelas coerentes à computação utilizando hardware gráfico fazem-se necessárias análises nos quatro níveis de abstração de Marr.

Na medida em que descemos nos níveis de abstração de Marr, as formulações tradicionais existentes na literatura para algoritmos de processamento de imagens e visão computacional passam a não ser adequadas à arquitetura das Unidades de Processamento Gráfico disponíveis. Por esse motivo, se faz necessário descer ao nível mais baixo de abstração de Marr e compreender o funcionamento de componentes da arquitetura do hardware gráfico.

No outro extremo de abstração, dando suporte à formulação dos problemas abordados nesta tese, o nível mais alto apresenta a base teórica de como o mundo físico se comporta, de forma a descrever processos capazes de construir a representação do problema para extração das características desejadas.

Esta tese apresenta o hardware gráfico utilizado e transita desde a descrição em alto nível da natureza dos problemas abordados, até a descrição

PUC-Rio - Certificação Digital Nº 0510966/CA

efetiva de formulações propícias a tal hardware.

As placas gráficas modernas possuem arquiteturas com alto poder de processamento paralelo, projetadas para processar grandes volumes de dados. O modelo de arquitetura vigente é conhecido como arquitetura de aritmética de fluxo paralelo com respeito aos dados ("streaming data-parallel arithmetic architecture"). A computação nesse tipo de arquitetura é feita executando um conjunto similar de operações sobre um fluxo de dados. São especialmente eficientes para computações envolvendo operações similares predominantemente aritméticas e de forma independente entre os dados.

As tarefas de visão computacional são frequentemente compostas por operações intensivas e repetitivas sobre grandes volumes de dados (obtidos de uma ou mais imagens), o que as torna candidatas beneficiárias das características da GPU. Mapear algoritmos de visão computacional para abordagens paralelas é de especial interesse, uma vez que para diversas aplicações, ainda não existem formulações sequenciais suficientemente rápidas de serem computadas.

Enquanto a computação gráfica recebe uma descrição numérica da cena e produz (renderiza) uma imagem, a visão computacional analisa uma ou mais imagens para extrair representações numéricas da cena. Por esse motivo, utilizar o hardware gráfico e através de seu "pipeline" gráfico executar tarefas de visão computacional pode ser visto como um uso inverso da proposição inicial para o qual esse hardware foi desenvolvido (GPU Gems 2). Entretanto, com o aumento da programabilidade das placas gráficas e o surgimento de linguagens de propósito genérico, o desafio que antes exigia a transcrição dos algoritmos para as etapas do pipeline passa a se concentrar na proposição de algoritmos operando em grandes fluxos de dados de forma paralela e independente entre seus elementos.

Dada a interdisciplinariedade presente no desenvolvimento desta tese, o leitor encontra ao longo dos capítulos os fundamentos necessários a compreensão das formulações propostas sem que seja indispensável conhecimento a priori de sistemas gráficos 3D ou de sistemas paralelos.

1.1 Visão Computacional

A pesquisa em visão computacional procura auxiliar ou automatizar o processo de reconhecimento visual. Para desempenhar tal tarefa, são utilizados métodos científicos tais como abordagens estatísticas e de aprendizado, modelos geométricos e físicos. A teoria de visão computacional fundamenta-se no conhecimento do processo físico de formação da imagem, desenvolvendo in-

PUC-Rio - Certificação Digital Nº 0510966/CA

ferências a partir da análise de valores de pixels ou conjunto de pixels de uma ou mais imagens.

Ao explicar a visão humana de um ponto de vista da teoria computacional, ou seja, formular quais problemas a mente humana resolve para que possamos enxergar, Marr (Mar80) considera que a percepção visual é um problema de processamento de informações e assume que a visão é a construção de descrições simbólicas a partir de imagens do mundo.

Em sua proposta, Marr identifica três camadas de representação para o conteúdo visual, a saber: a camada do esboço primitivo ("the primal sketch"), a camada do esboço $2\frac{1}{2}D$ ("the $2\frac{1}{2}D$ sketch"), e a camada de representação 3D do modelo ("the 3D model").

Ainda segundo Marr, a formulação de uma teoria em visão computacional que permita a definição de processos capazes de construir tais representações é, na verdade, o descobrimento de regras ou restrições ("constraints") que descrevam a forma que o mundo funciona ao mesmo tempo que permitam recuperar a característica desejada da informação visual.

A pesquisa em visão computacional utiliza-se dessas camadas de representação do conteúdo visual para se estruturar entre os seguintes módulos:

- Conversão Fotônica: aborda o processo de aquisição da imagem pela câmera, pela conversão fotométrica da luz para a formação da imagem digital.
- Processamento de Baixo Nível: considera o conteúdo visual no nível da variação de intensidade e natureza bidimensional da imagem digital Este módulo utiliza diversos tipos de manipulação de imagens tais como filtragem, detecção de arestas, agrupamento de pixels, etc. para gerar um esboço primitivo.
- Detecção de Formas: organiza uma estruturação do conteúdo visual para extração das formas, extraindo informações tais como profundidade, orientação e descontinuidade das superfícies visíveis em relação ao ponto de vista do observador. Este módulo utiliza informações de textura, movimento, sombreamento e visão estéreo para gerar um esboço $2\frac{1}{2}D$;
- "Parsing": responsável pela extração da representação 3D do modelo, seja na forma de uma descrição de cena ou de modelos tridimensionais hierarquizados em coordenadas do objeto. Tal representação permite uma descrição e organização centrada no objeto da estrutura tridimensional.
- Processamento de Alto Nível (ou Reconhecimento de Objetos): engloba a formação, recuperação, classificação e reconhecimento de objetos ou ima-

PUC-Rio - Certificação Digital Nº 0510966/CA

gens, pela comparação com protótipos definidos como imagens, modelos ou descrições estatísticas previamente classificadas.

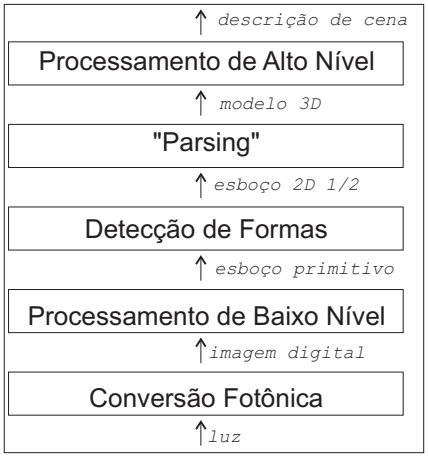

Figura 1.1: Paradigmas da Visão Computacional: do sinal aos símbolos

1.2 Contribuições e contexto da pesquisa

Os temas selecionados para esta tese estão contextualizados nas camadas denominadas Processamento de Baixo Nível e Detecção de Formas (Figura 1.1). No decorrer dos capítulos, apresentamos algoritmos em formulações paralelas adequadas ao processamento em hardware gráfico. A seguinte organização é adotada: o Capítulo 2 apresenta os fundamentos sobre o hardware gráfico e sua programação; o Capítulo 3 apresenta o primeiro grupo de algoritmos sobre a manipulação do conteúdo visual em baixo nível; e o Capítulo 4 apresenta um segundo grupo de algoritmos sobre problemas relacionados à detecção de formas. Finalmente, conclusões e trabalhos futuros são apresentados no Capítulo 5.

As contribuições da pesquisa desenvolvida foram publicadas em artigos incluídos nos anexos desta tese. Por este motivo, este texto foi elaborado de maneira a complementar o material das publicações, fornecendo uma introdução ao processamento de imagens e visão computacional em GPU ao leitor não familiarizado com o processamento paralelo em hardware gráfico e uma revisão dos algoritmos propostos baseados em processamento paralelo.

O Capítulo 3 expõe abordagens paralelas para problemas de processamento e análise de imagens. As principais contribuições apresentadas em tal capítulo incluem: a descrição de um operador paralelo para localização no espaço da imagem de objetos identificáveis por sua cor, denominado MOCT; as variações, análise de seus padrões de acesso à memória e a formalização do padrão de programação, denominado operador de redução múltipla, o qual

PUC-Rio - Certificação Digital Nº 0510966/CA

viabiliza o processamento paralelo de um fluxo por um conjunto ordenado de operadores independentes; e, finalmente, a descrição de um operador para distribuir em computações independentes o processamento do conteúdo de regiões de interesse, denominado operador de redução regional múltipla. As publicações relacionadas aos temas deste capítulo são apresentadas nos anexos A e B.

Em seguida, o Capítulo 4 aborda problemas de detecção de formas, mais especificamente, de segmentação de imagens, também em abordagens para processamento paralelo. As primeiras contribuições apresentadas no Capítulo 4 descrevem: a construção em GPU de quadtrees regionais com representação hierárquica pelo agrupamento de pixels semelhantes; a utilização de tal construção para responder consultas a árvore; e um novo modelo de representação de vizinhança entre folhas, denominado QuadN4, o qual viabiliza o levantamento do conjunto de folhas vizinhas de uma folha qualquer de forma ótima (uma leitura por vizinha).

Ainda no Capítulo 4 é apresentada uma formulação geral sobre algoritmos baseados em minimização de energia via corte de grafo. Como contribuição, são feitas propostas para acelerar a computação de tal otimização, através de pré-processamentos em GPU. Uma das formas de aceleração em pré-processamento proposta apresenta a computação em GPU das parcelas que compõem os termos da função de energia adotada em processamentos independentes. Em outra aceleração proposta, o tamanho do grafo criado para avaliar uma imagem é reduzido por meio da formulação de um problema de minimização no qual as variáveis correspondem a folhas de uma quadtree regional, representando a mesma imagem, como variáveis da otimização. As publicações relacionadas aos temas deste capítulo foram incluídas nos anexos C e D).

PUC-Rio - Certificação Digital Nº 0510966/CA