

1

Introdução

A reconstrução de cenas a partir de imagens é um dos problemas mais importantes em visão computacional e tem recebido atualmente um grande interesse por parte dos pesquisadores das áreas de computação gráfica e modelagem geométrica.

A origem da pesquisa em reconstrução de cenas pode ser relacionada, de certa maneira, aos primeiros estudos sobre o processo de percepção de forma nos sistemas visuais de animais superiores e de seres humanos. Nesta fase inicial, pesquisadores das áreas de inteligência artificial, psicologia e neurologia buscavam encontrar modelos que pudessem explicar como o cérebro humano é capaz de reconstituir e interpretar as informações sobre a forma de uma cena codificadas através de um par de imagens [6, 8]. A maioria dos primeiros modelos propostos procurava descrever uma abstração da arquitetura existente nos sistemas visuais biológicos através de um conjunto de interações entre sensores e um sistema de processamento visual primitivo. Nesta fase, era bastante comum a modelagem de tais sistemas através de redes neuronais ou através de sistemas cooperativos, os quais podem ser compreendidos como casos particulares de redes de autômatos celulares [7].

À medida que conhecimentos mais profundos sobre tais processos foram sendo adquiridos, houve uma mudança gradual do enfoque para a investigação de técnicas que pudessem ser empregadas no auxílio à execução de tarefas que requerem algum tipo de conhecimento sobre as formas existentes no espaço de trabalho.

Os primeiros métodos para determinação da forma de uma cena a partir de imagens foram naturalmente influenciados pelos primeiros trabalhos sobre os sistemas visuais biológicos, sendo a maior parte deles baseados na visão estéreo [9, 10, 14, 15, 17]. Esta abordagem foi e ainda é, em certo nível, adotada principalmente por pesquisadores das áreas de visão computacional e robótica.

O processo de reconstrução de forma através de visão estereoscópica

requer a solução de um problema de correspondências entre feições existentes em duas ou mais imagens obtidas de uma cena a partir de pontos de vista distintos. Uma vez determinadas estas correspondências, e conhecidos os parâmetros intrínsecos e extrínsecos das câmeras que forneceram as imagens, é possível calcular, através de um processo de triangulação, as coordenadas tridimensionais dos pontos da cena associados às feições consideradas. Em uma etapa posterior, efetua-se a interpolação dos pontos 3D obtidos com base em alguma restrição sobre a forma da superfície (normalmente uma restrição de suavidade), de maneira que a geometria e a topologia que descrevem a cena possam ser reconstituídas. Finalmente, com o objetivo de tornar as reconstruções mais realistas, é efetuado um processo de colorização através da aplicação de uma combinação das texturas provenientes das diferentes imagens. Para maiores detalhes ver [35, 36, 62].

Dentre as principais limitações associadas aos métodos baseados em visão estéreo, podemos citar: reconstrução restrita a superfícies $2,5D$, aplicabilidade a cenas cujos efeitos de oclusão e paralaxe são limitados e a necessidade de se introduzir na formulação do problema conhecimentos sobre a forma da cena através da adição de restrições. Algumas destas restrições são inerentes à geometria do problema como, por exemplo, a restrição epipolar, enquanto que outras são adicionadas com o objetivo de garantir a unicidade da solução ou possibilitar o desenvolvimento de métodos simples e eficientes.

Paralelamente, pesquisadores das áreas de computação gráfica e modelagem geométrica começaram a se interessar por técnicas de modelagem a partir de imagens. Inicialmente, as técnicas principais envolviam a determinação do espaço de ocupação da cena através da interseção de volumes cônicos, os quais eram determinados por raios passando pelos conjuntos de pixels delimitados pelas silhuetas dos objetos nas imagens e seus respectivos centros de projeção.

Podemos citar, como exemplos clássicos, os métodos propostos por Martin e Aggarwal [11], Chien [12, 18] e Potmesil [23]. Estes primeiros trabalhos limitavam bastante as vistas de entrada, já que requeriam procedimentos de interseção relativamente complexos.

Aos poucos, os novos métodos propostos começaram a aproveitar todo o potencial fornecido pelo desenvolvimento tecnológico relacionado à aquisição de imagens, se libertando definitivamente do paradigma fornecido pela visão estéreo binocular. Surgem então diversos métodos baseados em visão estéreo capazes de determinar mapas de profundidade a partir de múltiplos pontos de vista [47, 48, 49, 76].

Um outro grande avanço surgiu com o aprimoramento da tecnologia de aquisição de dados 3D, seja através de técnicas de estéreo ativo [33], seja através do surgimento de sistemas de varredura 3D de alta precisão [19, 24, 37, 39, 40, 41, 42, 43]. A partir deste momento, os métodos de reconstrução baseados em imagens de profundidade (*range images*) se tornaram bastante utilizados, principalmente quando são necessárias reconstruções detalhadas e com grande precisão.

Com o surgimento de câmeras digitais de baixo custo e de boa qualidade, renasce o interesse pela reconstrução de cenas a partir de fotos provenientes de câmeras calibradas. Devido às limitações existentes nas técnicas baseadas em estéreo, foram propostas novas alternativas que trabalham no espaço da cena e não requerem a solução do problema de correspondências como um pré-requisito.

Seitz e Dyer [44], provavelmente se inspirando nas técnicas de reconstrução volumétrica iniciais, propuseram uma nova abordagem baseada em uma representação do espaço da cena através de voxels na qual não é necessário o conhecimento explícito das silhuetas dos objetos, pois são utilizados apenas critérios de foto-consistência para determinar a pertinência de um voxel à cena. Como os demais métodos, a técnica proposta por Seitz restringe-se apenas à reconstrução de superfícies opacas. Além disso, introduz uma nova restrição: a de que as cenas a serem reconstruídas possam ser modeladas por superfícies lambertianas. A grande inovação introduzida por estes métodos é sua capacidade de integrar, em um único arcabouço, as diversas etapas necessárias à reconstrução que normalmente são efetuadas separadamente nos demais métodos de reconstrução. Tanto a reconstrução da geometria e da topologia de uma cena, quanto a sua colorização, são efetuadas simultaneamente no decorrer do processo (Figura 1.1).

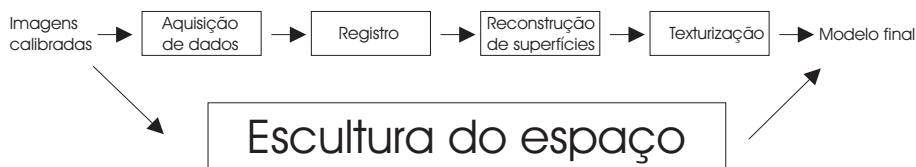


Figura 1.1: Etapas de um processo de reconstrução

A primeira implementação do método de coloração de voxels se restringia a configurações de câmera que se posicionavam fora do fecho convexo da cena. Esta restrição foi introduzida com o intuito de simplificar o problema de determinação da visibilidade, um aspecto fundamental para a realização da técnica. Através desta restrição é possível determinar uma

ordenação global dos voxels em relação ao conjunto de câmeras permitindo desta forma que o método possa ser realizado em um único passo. Em seguida, Kutulakos e Seitz [63] desenvolveram uma teoria sobre reconstrução de formas com base na satisfação de restrições de foto-consistência e generalizaram o método proposto para configurações de câmeras arbitrárias. Esta generalização passou a ser conhecida como Escultura do Espaço (*Space Carving*).

Ultimamente, algumas variações desta técnica têm sido propostas, dentre as quais podemos destacar as seguintes: a *Coloração de Voxels Generalizada (Generalized Voxel Coloring)* de Slabaugh [64], que propõe novas estruturas de dados para o tratamento da visibilidade dos voxels; a *Coloração de Voxels por Hipóteses Múltiplas* de Eisert [55], que propõe um novo algoritmo em dois passos, um envolvendo uma construção de uma lista de hipóteses e um outro no qual é efetuada uma remoção de hipóteses não consistentes com os dados de entrada; a *Visão Estereoscópica Aproximada através de Múltiplos Pontos de Vista (Approximated N-view Stereo)* de Kutulakos [65] que procura resolver o problema de como determinar uma família de formas que aproximam incrementalmente a forma de uma cena 3D arbitrária através de uma nova definição de foto-consistência baseada em uma classe de transformações, as *shuffle transforms*. Outras variantes importantes serão vistas na revisão bibliográfica apresentada no restante do documento.

As versões mais recentes foram propostas por Broadhurst e Cipolla [66, 71, 72] e paralelamente por Kutulakos e Bhotika [73], baseando-se em uma formalização estatística e probabilística para a modelagem do problema de reconstrução de cenas através de representações volumétricas.

Neste trabalho, apresentamos um novo método de escultura do espaço baseado no uso combinado de técnicas de mapeamento projetivo de textura e subdivisão adaptativa do espaço. Tal método é capaz de lidar apropriadamente com os problemas de amostragem existentes no processo de reconstrução, além de empregar esforços proporcionais às características locais da cena, que são descobertas à medida em que a reconstrução é realizada. O principal argumento desta tese é a de que métodos de escultura do espaço baseados no uso combinado de informações fotométricas e de segmentação, juntamente com estratégias de refinamento adaptativo, são capazes de produzir eficientemente reconstruções de boa qualidade, mesmo quando são utilizadas poucas imagens e sistema de aquisição mais modestos. A validade desta afirmação foi comprovada através os resultados obtidos através de sistema de reconstrução 3D baseado em *webcams*.

Apresentamos também as seguintes contribuições secundárias:

- Um estudo sistemático da aplicação de técnicas de visualização no desenvolvimento de métodos de escultura do espaço que trabalham diretamente no espaço da cena. Em particular, investigamos em detalhes o uso da técnica de mapeamento projetivo como mecanismo de registro de diversas informações no espaço da cena, como a visibilidade dos elementos do espaço, a segmentação dos objetos de interesse e a quantidade de ruído introduzido pelos sensores das câmeras usadas na aquisição das imagens.
- Um estudo sobre os testes estatísticos adequados para a determinação da foto-consistência dos elementos do espaço na presença de diferentes taxas de ruído em cada um dos sensores das câmeras.
- O desenvolvimento de um sistema de reconstrução 3D com base em *webcams* e diferentes estratégias para lidar com suas limitações.

O restante do trabalho está organizado da seguinte forma: no capítulo 2 apresentamos os principais conceitos e técnicas sobre reconstrução volumétrica de cenas, além de apresentar, de modo resumido, alguns dos trabalhos mais importantes no assunto; no capítulo 3, descrevemos como a técnica de mapeamento projetivo pode ser utilizada para registrar as informações necessárias à realização dos métodos de escultura do espaço no espaço em que se encontram os objetos de interesse, e discutimos suas diversas implicações; no capítulo 4, descrevemos o funcionamento dos métodos de reconstrução que trabalham diretamente no espaço 3D com base nas técnicas descritas no capítulo 3; no capítulo 5, apresentamos o método de reconstrução proposto que se baseia em uma estratégia de escultura do espaço por refinamento adaptativo; no capítulo 6, descrevemos o sistema de reconstrução 3D baseado em *webcams* que utiliza as técnicas de reconstrução baseadas em escultura do espaço, além de apresentarmos os principais problemas encontrados e as respectivas técnicas utilizadas para resolvê-los; no capítulo 7, descrevemos os resultados obtidos através de reconstruções a partir de imagens obtidas de cenas sintéticas e reais e finalmente, no capítulo 8, apresentamos as conclusões finais sobre o trabalho e alguns dos possíveis tópicos de pesquisa que podem ser investigados em trabalhos futuros.