

2 Trabalhos Relacionados

A visão humana e sua capacidade fantástica de extrair informação do espaço tridimensional à sua volta é uma área de pesquisa que envolve campos como Biologia, Física, Psicologia, Robótica e Computação.

Dentre as possíveis informações que a visão humana se utiliza para estimar a forma de um objeto tridimensional a partir de imagens bidimensionais temos: textura, movimento, sombreamento, estereoscopia, foco, sobreposição etc. No campo da Visão Computacional foram criados métodos computacionais baseados nestas informações e ao conjunto deles damos o nome de *Shape-from-X*.

Neste capítulo são apresentadas breves notas sobre algumas técnicas de *Shape from X*, que podem ser encontradas de maneira mais abrangente na tese de Jong [1]. Ao final é dada ênfase na área de *Shape from Texture*, tema desta pesquisa, mostrando assim um histórico das principais pesquisas desenvolvidas sobre o tema.

2.1 Visão Estéreo (*Shape from Stereopsis*)

A estereoscopia é baseada na diferença relativa entre posições de objetos em duas ou mais imagens, a qual chamamos disparidade. Diz respeito à percepção de formas 3D e distâncias que resultam de diferentes imagens de uma mesma cena produzidas por pontos de vista distintos. A fusão das imagens capturadas por nossos olhos permite uma forte percepção da profundidade da cena.

A visão estereo envolve dois processos: a fusão das características observadas pelos olhos e a reconstrução de seu modelo tridimensional. O primeiro consiste em encontrar características ou pontos de interesse que sejam facilmente localizáveis em ambas as imagens. Em uma imagem temos milhares de *pixels*; porém, nem todos podem ser usados como pontos de interesse, uma vez que seu par correspondente na segunda imagem (par homólogo) não seria adequadamente

localizado. Este é chamado de “problema da fusão binocular”, ilustrado na Figura 4. Exemplos de bons pontos de interesse incluem bordas e cantos.

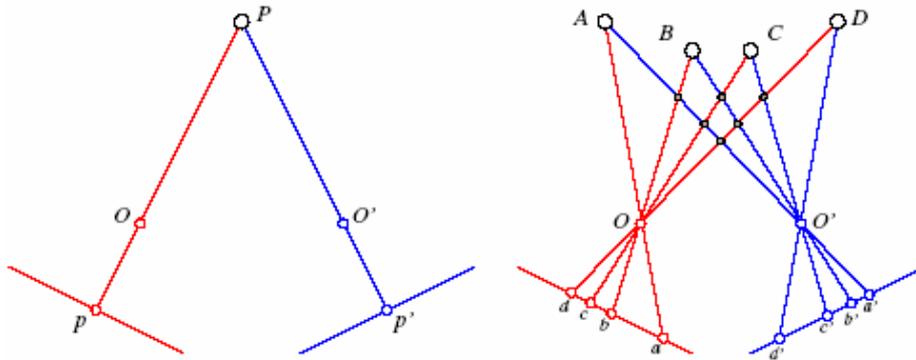


Figura 4 - O problema da fusão binocular: no diagrama da esquerda temos apenas um ponto e seu homólogo facilmente encontrado. No da direita temos quatro pontos e seus homólogos podem ser qualquer um dos quatro.

A segunda fase de um sistema estéreo envolve a reconstrução do modelo tridimensional dados os pontos de interesse e seus homólogos na outra imagem. Na Figura 5 temos exemplos de pares estereoscópios de objetos. Note que até aqui foi considerado um sistema estereoscópico baseado na visão humana, ou seja, com apenas duas câmeras, porém podem ser empregadas mais de duas, como um sistema trinocular [2].



Figura 5 - Exemplos de pares estereoscópios.

2.2

Forma a partir da intensidade (*Shape from Intensity*)

A visão humana consegue perceber diferenças de profundidade a partir da intensidade, em função da iluminação que um objeto recebe. Exemplos de imagens bidimensionais podem ser vistos na Figura 6, onde temos uma clara percepção de forma apesar da única informação presente ser a de iluminação na superfície. Este recurso de sombreamento já é empregado há muito tempo por pintores e artistas da área de computação gráfica, conseguindo passar com maestria a sensação de profundidade.

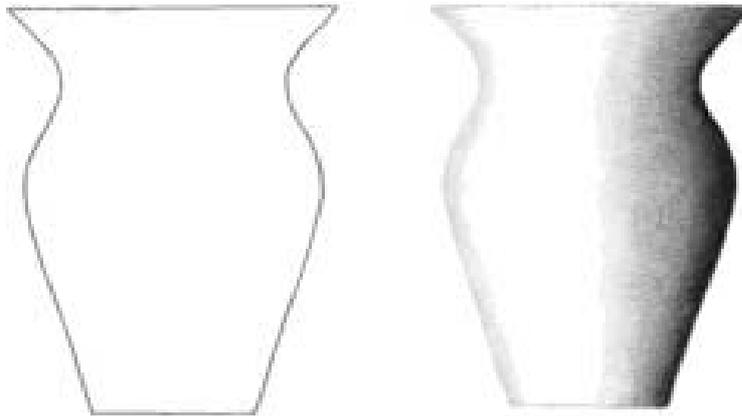
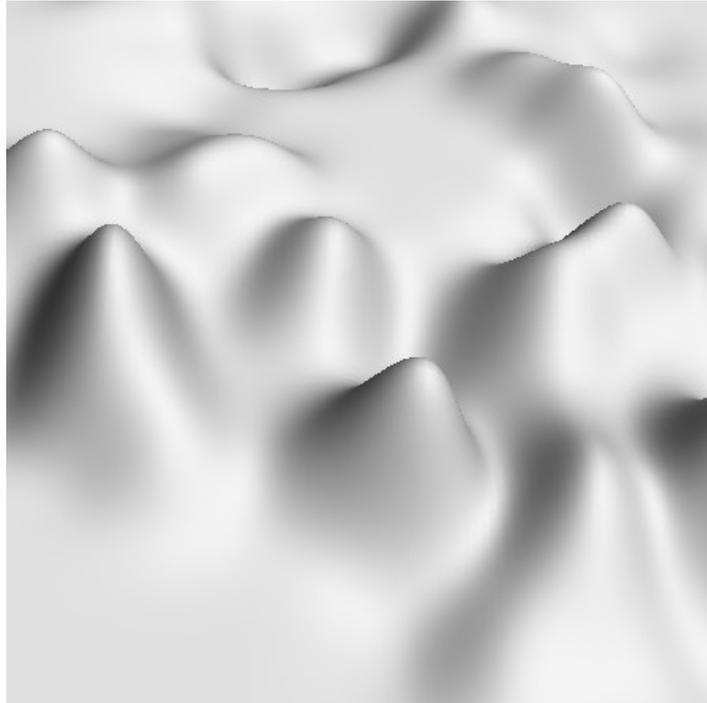


Figura 6 - Exemplos de imagens onde o sombreamento evidencia a noção de profundidade.

Partindo desta característica da visão humana surgiram na Visão Computacional técnicas que exploram este efeito. Esta abordagem é dividida em dois grupos: Reconstrução a partir do sombreamento, *Shape from Shading (SFS)* e Reconstrução a partir de Estereoscopia Fotométrica, *Shape from Photometric Stereo (SFPS)*.

O primeiro, *SFS*, teve seu trabalho inicial com Horn [3] em 1970. Consegue medir profundidade através do sombreamento partindo de uma única imagem. Normalmente admite uma superfície Lambertiana (reflete a luz incidente com a mesma luminância em todas as direções) e que existe uma única fonte de luz

incidente, cuja direção é conhecida. O problema desta abordagem é que, na prática, poucas são as superfícies Lambertianas e as imagens teriam de ser adquiridas em ambiente controlado, pois senão mais de uma fonte de iluminação teria de ser considerada, com parâmetros distintos, além da reflexão das superfícies. Com isso, *SFS* não é um método muito acurado para toda e qualquer situação real.

O segundo, *SPFS*, utiliza mais de uma imagem, ainda em ambiente controlado, mas sendo que agora cada imagem é adquirida usando uma fonte de iluminação com direção diferente, mantendo a posição da câmera e do objeto constantes. Algumas técnicas são descritas em [4, 5].

2.3

Forma a partir do Movimento (*Shape from Motion/Vídeo*)

A reconstrução a partir do movimento [6] é similar a estereoscopia, pois parte de mais de uma imagem de um objeto tomada de diferentes pontos de vista. É possível estimar a forma de um objeto móvel partindo de uma seqüência de imagens de uma câmera fixa, assim como o contrário também é verdadeiro, partindo de um objeto fixo e uma câmera móvel.

2.4

Forma a partir do Foco (*Shape from Focus/Defocus*)

A análise do Foco apresenta uma poderosa ferramenta para recuperar a forma tridimensional de superfícies. Esta técnica possui emprego em superfícies ásperas microscópicas, onde outras técnicas de Visão Computacional não possuem bons resultados. Um exemplo pode ser visto no trabalho de Nayar e Nakagawa [7].

2.5

Forma a partir de sensores de distância (*Shape from Range Sensor*)

As outras técnicas em *Shape from X* são ditas passivas, porque utilizam sensores como câmeras para captura de imagem e então aplicam algoritmos de processamento de imagem. Porém, existe uma área baseada em sensores ativos [8, 9], que emprega sensores lasers ou ultra-sônicos. Por exemplo, usando esses sensores como radares *time-of-flight* (tempo de voo), pode-se medir com precisão a distância entre o sensor e o ponto no objeto, mensurando a diferença entre um pulso eletromagnético recebido e um transmitido.

2.6

Forma a partir da textura (*Shape from Texture*)

A idéia de que a visão humana usava como uma de suas medidas de profundidade a deformação de textura foi apontada inicialmente por Gibson [10], em 1950. Foi ele que cunhou o termo gradiente de textura, que descreve o fenômeno pelo qual pequenos recortes (*patches*) de uma superfície que possui um padrão repetido, ou seja, uma textura, projetam no plano da imagem os mesmos recortes, porém com aparência diferente devido à orientação, distância e forma da superfície. Gibson sugere que se pudermos medir este gradiente, essa diferença entre os recortes, podemos encontrar a orientação local da superfície na cena. Foi a partir desse trabalho que começaram as pesquisas sobre *Shape from Texture* (*SFT*) em Visão Computacional.

Os métodos de *SFT* se dividem inicialmente em dois grupos de abordagens: medir a orientação de uma superfície plana inclinada; e, medir a orientação local de pequenos recortes de uma superfície curva, podendo assim integrar essa informação e obter a forma da superfície. Os métodos são apresentados nas subseções a seguir, divididos pelas duas abordagens.

2.6.1 Estimativa da orientação de planos inclinados

O primeiro método computacional na área foi o de Bajscy e Lieberman [11] em 1976, apesar deste não calcular a orientação de um plano. Na verdade, eles desenvolveram uma forma de medir distâncias relativas entre objetos em cenas reais, estes sobre um plano texturizado, a partir da diferença de tamanho entre elementos de textura (*texels*) ao longo do plano. Para medir essa diferença é empregada análise do espectro de frequência, usando a transformada de Fourier em janelas [12]. O método é limitado a planos inclinados naturais como pisos, solos e mar.

Para medir o gradiente de textura nos *texels* utilizam-se informações como mudança na densidade, área, escala, compressão (*foreshortening*), e ainda uma abordagem estatística usando desvio da isotropia - uma característica física, que significa que um objeto possui a mesmas propriedades em todas as direções -, quando uma textura isotrópica é projetada perde esta propriedade. Na literatura o primeiro trabalho que apontou e classificou essas informações distintas foi o de Stevens [13], em 1981.

Muitos trabalhos, no início, basearam-se em extração de *texels*, usando detecção de bordas, por exemplo. Exemplos incluem os trabalhos de Witkin [14] em 1981, que utiliza um algoritmo de detecção de bordas para encontrar os *texels* e busca medir dois efeitos: que eles aparecem menores de acordo com a distância; e que são comprimidos na direção da inclinação da superfície. Usando um modelo de projeção ortográfica, que equivale a uma transformação afim na imagem, são aplicadas uma série de possíveis transformações ao plano. Com os possíveis planos reconstruídos, a solução se resume em determinar qual destes é o mais similar à imagem original. O método é estendido para superfícies curvas, porém não apresenta bons resultados.

No trabalho de Aloimonos [15] a orientação do plano é calculada assumindo que a densidade dos *texels* é uniforme. Aloimonos emprega um modelo de projeção distinto, o paraperspectivo, que é mais simples que o perspectivo porém mais acurado e complexo do que o ortográfico. Apresenta bons resultados, mas apenas voltado para planos inclinados.

Blostein e Ahuja [16], em 1989, propõem um método voltado apenas para superfícies inclinadas, contendo uma única textura natural. O método extrai os *texels* usando um algoritmo de detecção de região multiescala. Uma relação entre o tamanho da textura e um plano texturizado hipotético é derivada do modelo de projeção, o problema recai em uma procura pela orientação que melhor se ajusta aos *texels* encontrados.

Dentre outros exemplos temos o de Kender [17], Ohta et al. [18], Ikeuchi [19], Kanatani e Chou [20]. A grande desvantagem destes métodos é a abordagem por extração de *texels*, uma vez que esse é um passo difícil, que requer normalmente intervenção do usuário, além do que, uma textura pode conter sub-texturas, dependendo da escala em que se trabalha. Esses métodos também acabam por não utilizar toda a informação contida na imagem, somente as existentes nos *texels*.

Uma outra abordagem na literatura de *SFT* é baseada em um modelo probabilístico de textura. O problema de determinar a forma e orientação de uma superfície recai então em um problema de inferência estatística (estimativa dos parâmetros do modelo dadas amostras na imagem). Normalmente nesses casos, admite-se que a textura é isotrópica, isto é, possui a mesma aparência independente de rotação e quando é projetada perde esta propriedade. Então pode-se usar o desvio da isotropia como meio de inferir a orientação. Exemplos dessa abordagem são os trabalhos de Witkin [14], Davis et al. [21], Blake e Marinos [22], Garding [23] e Brown e Shvaytser [24], sendo que os últimos usam análise de Fourier e a função de autocorrelação para encontrar a orientação de planos.

O problema da hipótese de isotropia é a dificuldade de encontrar texturas com essas características no mundo real, pois normalmente as texturas são direcionais.

Krumm e Shafer [25] estimam a transformação afim presente na projeção no domínio da frequência, porém usando o conceito de *espectrograma* (melhor descrito no capítulo 4 desta dissertação). Esses autores procuraram por todas as combinações de orientação em um conjunto discreto para estimar a correta. Entretanto seu método é empregado somente para planos.

2.6.2 Estimativa da forma de superfícies curvas

Garding [26], em 1992, desenvolveu toda uma base matemática para *SFT* em superfícies curvas. O trabalho de Garding foi de extrema importância para a área, pois, além de desenvolver uma *metodologia* para análise dos gradientes baseado em geometria diferencial [27], também mostrou que estas simples informações de distorção não são suficientes para uma medida completa da curvatura da superfície (ex.: sinal da curvatura gaussiana).

Garding [28] e Super e Bovik [29] usam o domínio da frequência e momentos de segunda ordem para estimar a forma de superfícies curvas, sendo que o primeiro trabalho é focado na estrutura matemática da estimativa. Super e Bovik empregaram filtros de Gabor para a análise de textura e projeção ortográfica, que demonstraram resultados encorajadores. Também apontam uma maneira de encontrar um ponto visto frontalmente na imagem, possuindo assim um referencial para seu algoritmo, que estima a deformação comparando pontos na imagem.

Rosenholtz e Malik [30] desenvolveram um método de *SFT* para superfícies curvas usando a base matemática desenvolvida por Garding. O método utiliza transformada de Fourier e análise diferencial para estimar a transformação afim entre dois recortes pequenos da superfície curva. Com a análise diferencial e algumas aproximações usando álgebra linear e mínimos-quadrados, consegue chegar a uma solução para a orientação, sem a necessidade de determinar um ponto visto frontalmente, como no trabalho de Super e Bovik. Tanto o trabalho de Super e Bovik quanto parte do de Malik e Rosenholtz são empregados nesta dissertação e melhor descritos no Capítulo 4.

Dando continuidade ao trabalho de Malik e Rosenholtz, Ribeiro et al. [31] apresenta uma solução fechada para o método de Rosenholtz, sem empregar otimização numérica iterativa, através dos autovetores das matrizes de transformação afim.

Clerc e Mallat [32] apresentaram um método que usa *wavelets* como medida de textura espectral e modelam o gradiente de textura como uma deformação de um processo estocástico estacionário [33].

Loh e Kovési [34] apresentaram um método para superfícies curvas que utiliza o conceito de inércia local [35] sobre os momentos de segunda ordem da transformada de Fourier dos recortes de textura na imagem. Os resultados desse método são promissores porém ainda limitados, uma vez que apenas calcula um dos ângulos da orientação (*tilt*) e precisa que um ponto frontal na imagem seja indicado manualmente.

Forsyth [36] apresenta uma técnica bem mais elaborada, que extrai recortes pequenos na superfície curva e calcula a orientação relativa em cada ponto sem a necessidade de determinar um ponto frontal na imagem, visto que este não é tido como existente na imagem e sim, estimado a partir dos recortes extraídos. Forsyth vai além e sugere uma outra aplicação prática de *SFT*, em conjunto com a computação gráfica, na renderização de tecidos, que pode ser empregado em lojas virtuais de roupas.

Loh e Zisserman [37] apresentam um trabalho semelhante ao de Forsyth. Assim como ele, não encontram um ponto frontal na imagem e sim o estimam. Entretanto, empregam como cálculo da transformação afim entre o ponto frontal e os recortes na imagem um método próprio, apresentado em outro trabalho [38] (usado nesta dissertação e descrito em detalhes no Capítulo 4).

Até aqui foi apresentado um resumo sobre os artigos pesquisados neste trabalho, com ênfase em *SFT* voltado para superfícies curvas. O capítulo seguinte apresenta alguns conceitos amplamente usados em *SFT*. O capítulo 4 descreve em detalhes três artigos selecionados deste resumo, que foram analisados nos experimentos, apresentados no capítulo 5.