

Exame de qualificação ao mestrado

1 INTRODUÇÃO

Há diferentes maneiras de reconstruir a forma tridimensional de um objeto a partir de imagens bidimensionais. A mais desenvolvida, a estéreo geométrica, usa duas ou mais imagens alinhadas de diferentes pontos de vista. Outra técnica é a estéreo fotométrica, que utiliza três ou mais imagens com o mesmo ponto de vista, mas sob diferentes condições de iluminação. Outra ainda é a estéreo multifocal, que usa fotos alinhadas com o mesmo ponto de vista e iluminação, mas com diferentes distâncias focais.

O objetivo da proposta é combinar as técnicas de estéreo multifocal e estéreo fotométrico de modo que cada uma compense as limitações da outra.

1.1 ESTÉREO MULTIFOCAL

Um microscópio óptico geralmente pode focar apenas nos pontos do objeto-alvo que se encontram em uma faixa estreita de distâncias da lente objetiva, isto é, entre dois planos próximos perpendiculares ao eixo óptico. A distância entre esses planos, chamada profundidade de campo, é inversamente proporcional à ampliação do microscópio e ao diâmetro efetivo da lente (CICILIATO; CHIQUITO, 1995). Se o objeto de interesse não puder se encaixar entre esses dois planos, é necessário obter várias imagens, variando a distância entre o objeto e o microscópio, para que cada ponto visível do objeto seja focalizado em pelo

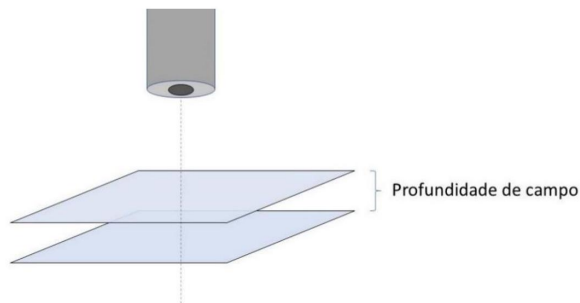


Figura 1: Representação da variação de coordenadas para aplicação EM.

menos uma das imagens. Essas imagens podem ser compostas para formar uma única imagem mostrando o objeto inteiro em foco.

Utilizando essa técnica também é possível obter informações sobre a forma do objeto ao longo da direção do eixo óptico (eixo Z). Essas informações têm a forma de um mapa aproximado da altura, que, para cada pixel da imagem combinada, especifica a coordenada Z do ponto visível até a superfície do objeto. Esse método para recuperar a forma 3D do objeto é conhecido como shape from focus (FERNANDES; TORREÃO, 2010; NAYAR, 1989), (ZAMOFING; HUGLI, 2004), Depth from focus (GROSSMANN, 1987), ou estéreo multifocal.

A tecnologia estéreo multifocal (EM) tem sido amplamente utilizada no campo da medição microscópica e macroscópica tridimensional,

Grossmann em 1987 (GROSSMANN, 1987), descreveu um método para estimar a profundidade de diferentes partes de um objeto a partir do grau de desfoque de uma única imagem. Shree Nayar (NAYAR, 1989) descreveu em 1989, como combinar pilhas de imagens multifocais.

1.1.1 TEORIA E NOTAÇÃO

O método de EF obtém o plano focal ideal de cada ponto na superfície alvo, utilizando a profundidade de campo limitada e o método de varredura vertical do sistema óptico e, em seguida, obtém a estrutura 3D de destino através da técnica de empilhamento de foco para realizar a medição 3D.

A entrada para a técnica EF é uma sequência de imagens J_1, J_2, \dots, J_n tiradas com objeto e microscópio na mesma posição relativa, exceto na distância vertical entre microscópio e objeto no eixo z , e as mesmas condições de iluminação. Supõe-se que as imagens estejam perfeitamente alinhadas e dimensionadas para que o pixel $J_i[x, y]$, para toda imagem j , mostre o mesmo ponto na superfície do objeto. Os pixels de cada imagem j que estão efetivamente focados são aqueles em que a coordenada z da superfície visível fica em um intervalo $\bar{z}_i \pm \varepsilon/2$ (LI; YANG, 2008) para o plano focal médio específico \bar{z}_i em cada imagem, onde ε é a profundidade de campo do equipamento. Veja a figura 2. Tipicamente, as alturas são igualmente espaçadas por um passo $\delta \leq \varepsilon$ e espaçam completamente a faixa relevante de coordenadas \bar{z}_i . Definiremos arbitrariamente a primeira posição do plano médio \bar{z}_1 como a origem do eixo z , de modo que $\bar{z}_i = (i - 1)\delta$ para todo i igual ou maior que um.

O primeiro passo no MFI é identificar os pixels de cada imagem em que o objeto está em foco. Essa classificação baseia-se no fato de que as partes que não estão em foco não podem ter brilho e variações de cores abruptas, como pontas afiadas e pequenos pontos. Portanto, se os gradientes da imagem em torno de um pixel $[x, y]$ são maiores em uma imagem J' do que em outra imagem J'' , então podemos concluir que essa parte de J' está melhor focada. Veja a figura 3.

Usando esse critério, uma pontuação numérica do foco $\alpha_i[x, y]$ é fornecido para cada pixel $J_i[x, y]$ de cada imagem na pilha. Uma forma simples de seleção pode ser através da magnitude absoluta ou relativa do gradiente em uma pequena vizinhança do pixel $[x, y]$, magnitude relativa à intensidade mé-

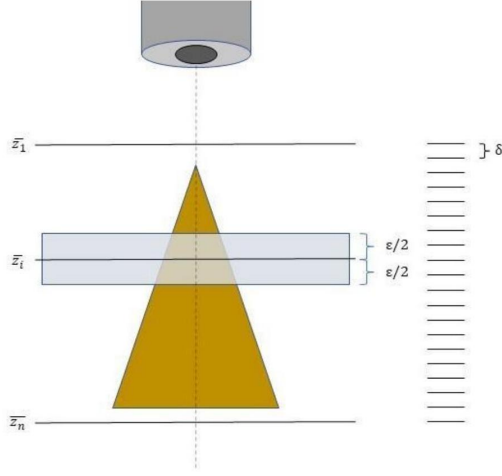


Figura 2: Representação da variação de coordenadas para aplicação EM.

da imagem. Um estudo comparativo dos operadores de pontuação de foco foi recentemente fornecido por Pertuz et al. (PERTUZ; PUIG; GARCÍA, 2012). Muitas outras fórmulas foram propostas, usando derivadas ou componentes de alta frequência de operadores do tipo wavelet (BERCOVICI; HADLEY; VILLANUEVA-AMADOZ, 2009).

Na abordagem mais simples, para cada posição de pixel $[x, y]$, o valor $\tilde{J}[x, y]$ da imagem combinada é copiado do pixel correspondente $J_i[x, y]$ com a maior pontuação $\alpha_i[x, y]$. A altura aproximada $\tilde{Z}[x, y]$ desse pixel é então definida como \tilde{Z}_i , com incerteza implícita de $\pm\delta/2$.

Os melhores resultados podem ser obtidos combinando os pixels $J_i[x, y]$ em algum tipo de média, em vez de simplesmente selecionando aquele com a maior pontuação.

$$\tilde{J}[x, y] = \sum_{i=1}^n \lambda_i[x, y] J_i[x, y] \quad (1)$$

Onde os coeficientes de mistura $\lambda_i[x, y]$, para cada pixel $[x, y]$ são uma partição da unidade derivada dos indicadores $\alpha_i[x, y]$. Por exemplo, pode-se simplesmente pegar para todos os pixels $[x, y]$:

$$\lambda_i = \frac{\alpha_i[x, y]}{\sum_{k=1}^n \alpha_k[x, y]} \quad (2)$$

A altura $\tilde{Z}[x, y]$ e sua incerteza podem ser estimadas pela média ponderada e pelo desvio da altura do plano de foco \tilde{Z}_i , com pesos $\lambda_i[x, y]$.

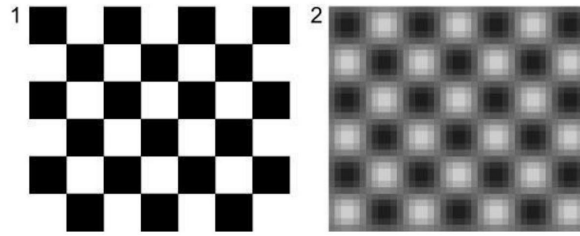


Figura 3: 1-Imagem com bordas bem definidas, indicando que o objeto alvo está em foco. 2- Bordas não definidas, indicando que o objeto alvo está fora da distância entre os planos focais.

1.2 ESTÉREO FOTOMÉTRICO

Outra maneira de adquirir a forma tridimensional de um objeto é capturar várias imagens com a câmera e o objeto na mesma posição relativa, variando a direção da luz que incide sobre o objeto (COLEMAN; JAIN, 1982; WOODHAM, 1980). A análise de como a cor aparente do objeto muda em cada pixel produz a inclinação de sua superfície visível nesse pixel, bem como sua cor "intrínseca". Essas informações de inclinação podem ser integradas para recuperar a coordenada Z em cada pixel (SARACCHINI et al., 2012). Essa técnica, conhecida como estéreo fotométrico (EF).

Os métodos de estéreo fotométrico visam recuperar a forma dos objetos a partir de coleções de imagens imóveis tiradas com iluminação variável. O EF é desafiador devido a várias razões. Em particular, a refletância do objeto pode variar de lambertiana a especular, com graus variados de rugosidade. Além disso, as luzes ao redor podem ser complexas e incluir várias fontes de luz junto com reflexos de objetos ao redor. Em muitos casos práticos, nem a configuração da luz nem as propriedades de refletância são conhecidas.

Os princípios físicos do EF são conhecidos desde o desenvolvimento da fotometria em 1760 por Lambert (NIEDERÖSTA; NIEDERÖSTA; ŠČUČKAB, 2003) e seu uso pelos astrônomos no início do século 20 (RINDFLEISCH, 1966).

A primeira descrição do método estéreo fotométrico, usando várias imagens sob diferentes campos de luz, foi fornecida por Woodham em 1980 (WOODHAM, 1980). Essa ideia foi pesquisada extensivamente desde então, com melhorias significativas, como restrições reduzidas na natureza das fontes de luz, propriedades ópticas da superfície e redução na necessidade de calibração.

Trabalhos clássicos em EF geralmente assumem que as condições de iluminação são conhecidas, por exemplo, através da especificação de uma função de refletância (WOODHAM, 1980). O campo de pesquisa continua propondo diferentes abordagens através de trabalhos que se concentram em lidar com objetos não-lambertianos como (COLEMAN; JAIN, 1982) e (BARKSKY; PETROU, 2003), utilizando fontes de luz distribuídas (NAYAR; IKEUCHI; KANADE, 1990) e em situações em que a fonte de luz está próxima ao objeto e próxima à câmera (KIM; BURGER, 1991).

Os métodos mais confiáveis para determinar as normais usam um objeto de forma conhecida como gabarito, fotografado sob as mesmas condições de luz que o objeto de destino, e possui características de superfície semelhantes, exceto possivelmente por cores intrínsecas. Essa técnica já foi usada por Woodham em 1980 (WOODHAM, 1980) e foi aprimorada em velocidade, precisão e generalidade por Hertzmann (GOLDMAN et al., 2010), Saracchini (SARACCHINI; STOLFI; DA GAMA LEITÃO, 2011) e Quéau (QUÉAU; LAUZE; DUROU, 2015)

1.2.1 TEORIA E NOTAÇÃO

No EF, as imagens $J^{(1)}, J^{(2)}, \dots, J^{(m)}$ são obtidas com o objeto e o microscópio na mesma posição relativa, mas usando fontes de luz diferentes $L^{(1)}, L^{(2)}, \dots, L^{(m)}$ em diferentes direções. Como na MFI, supomos que as imagens estão perfeitamente alinhadas e dimensionadas, de modo que o pixel $J^{(j)}[x, y]$, para todos j , mostre o mesmo ponto $p[x, y]$ na superfície do objeto.

Para simplificar a descrição, vamos supor que as fontes de luz são monocromáticas com a mesma cor espectral, e o valor $J^{(j)}[x, y]$ de cada pixel é um número real proporcional à altura do brilho aparente da região de superfície em torno de $[x, y]$; isto é, à potência luminosa por unidade de área espalhada por esse pedaço de superfície em direção à lente objetiva do microscópio. Ou seja, assumimos uma codificação de brilho linear, em vez da codificação "gama" usada na maioria dos arquivos de imagem.

Seja $\vec{u}[x, y]$ o local normal, isto é, o vetor de direção unitária médio perpendicular à superfície em $p[x, y]$ e apontando para fora do objeto. De acordo com a óptica física, o brilho aparente dos pixels da imagem pode ser expresso como:

$$J^{(j)}[x, y] = A[x, y]L^{(j)}[x, y]\vec{u}[x, y] \quad (3)$$

Aqui, $A[x, y]$ é um coeficiente de dispersão total da superfície nesse ponto, com valor entre 0 e 1, representa a porcentagem de luz incidente que é espalhada de volta ao ambiente; e $L^{(j)}[x, y]$ é a função de sombreamento para essa parte da superfície. O albedo deve ser uma propriedade intrínseca da superfície naquele ponto, que descreve sua "cor" intrínseca, distinguindo uma superfície branca ($A = 1$) de uma cinza ($A < 1$) ou preta ($A = 0$).

A função de iluminação deve ser independente do albedo, mas pode depender do acabamento da superfície (fosco, brilhante etc.). Seja $\Phi^{(j)}[x, y](\vec{v})$ responsável em denotar o fluxo de luz que chega da direção \vec{v} ; isto é, o brilho aparente do ambiente como seria visto por um minúsculo observador em pé naquele elemento da superfície e olhando na direção \vec{v} . Seja também $\vec{w}[x, y]$ a direção de $p[x, y]$ para a câmera na posição nominal. O valor de $L^{(j)}[x, y](\vec{u})$, para uma superfície com orientação \vec{u} , é determinado por essas quantidades e pelas propriedades ópticas intrínsecas da superfície naquele ponto, de acordo com a equação

$$L^{(j)}[x, y](\vec{u}) = \int_{s^2} \Phi^{(j)}[x, y](\vec{v})\beta[x, y](\vec{u}, \vec{w}[x, y], \vec{v})d\vec{v} \quad (4)$$

Aqui $\beta[x, y]$ é a função normalizada de distribuição de radiação bidirecional ou bidirectional radiance distribution function (BRDF), que informa qual fração da luz que chega da direção \vec{v} que não é absorvida pela superfície será espalhada na direção $\vec{w}[x, y]$.

O objetivo principal do estéreo fotométrico é resolver a equação (3) para cada pixel $[x, y]$, a fim de recuperar o valor $A[x, y]$ e a superfície normal $\vec{u}[x, y]$. Deve-se considerar que a equação (3) é inútil se o ponto $p[x, y]$ estiver sombreado em relação à luz j , de modo que não receba luz (ou apenas luz ambiente, que geralmente é isotrópica e indeterminada). Mesmo que as funções $L^{(j)}[x, y]$ sejam totalmente conhecidas por todos j , ainda existem três incógnitas (A e os dois componentes horizontais de \vec{u}) e, portanto, pelo menos três das equações devem ser independentes e não triviais. Concluímos que o problema pode ser resolvido apenas se $p[x, y]$ estiver suficientemente iluminado por pelo menos três fontes de luz. Mesmo para objetos convexos mais simples, seis fontes de luz são o mínimo necessário para atender a esse requisito.

Determinar as funções $L^{(j)}[x, y]$ é um problema difícil. A equação (4) requer conhecimento do BRDF $\beta[x, y]$ e do campo de luz recebido $\Phi^{(j)}[x, y]$. Uma maneira de atacar esse problema é restringir a situação, exigindo que $\Phi^{(j)}$ seja essencialmente unidirecional e uniforme e / ou exigindo que a superfície seja lambertiana.

Alternativamente, as funções $L^{(j)}[x, y]$ podem ser estimadas através da imagem de um ou mais exemplos de objetos nas mesmas condições de iluminação. Um desses medidores é suficiente e considerar que o campo de luz Φ que entra é uniforme; isto é, essencialmente o mesmo para todos os pontos totalmente iluminados $p[x, y]$. Essa suposição pode ser feita se o objeto e a fonte de luz forem pequenos comparados à distância entre os dois.

Quando o mapa normal $\vec{u}[x, y]$ for determinado, o mapa \hat{Z} será obtido por um dos métodos de integração à superfície (SARACCHINI et al., 2012) e (FRANKOT; CHELLAPPA, 1988; SARACCHINI; STOLFI, [s.d.]).

2 COMBINANDO ESTÉREO MULTIFOCO E ESTÉREO FOTOMÉTRICO

Os dois métodos possuem limitações que muitas vezes impedem a determinação da profundidade em partes significativas da cena. Nestas partes, a xxx da imagem não depende da profundidade. Felizmente os métodos EM e EF são muitas vezes complementares.

O método estéreo multifocal não funciona na parte da cena em que a superfície é lisa (sem textura). Essas partes receberão uma pontuação de foco baixa mesmo se a imagem de entrada estiver perfeitamente focada naquela região. A imagem composta continuará válida, porque qualquer seleção, ou critério, de combinação fornecerá o mesmo valor para o pixel de saída. No entanto, a altura estimada será totalmente incerta. Por outro lado, EF funciona bem onde a superfície é lisa e tem a distribuição de radiação bidirecional (DRB) uniforme.

Nesses locais, as normais calculadas e as alturas integradas relativas possuem maior probabilidade de estarem corretas. Por outro lado o EF falha onde a cena tem descontinuidades de profundidade ou sombras projetadas.

EM e EF podem interagir sinergicamente para a produção de um método estéreo melhor do que apenas o uso de um ou outro em partes diferentes da imagem.

3 A DESCOBRIR

Descontinuidades e regiões significativas com normal mal definida (como fundo distante) podem ser geralmente detectadas a partir de operadores morfológicos simples, como limiar, dilatação e erosão. A máscara de validação necessária pode ser derivada do mapa de altura aproximada, sem intervenção manual do usuário. Mesmo se as descontinuidades dividirem o domínio em várias regiões desconectadas, o mapa fornece as alturas relativas aproximadas dessas áreas.

Em regiões com textura, o EM é usado para estimar a profundidade, pois fornece resultados mais confiáveis e robustos. Já em regiões lisas, o EF é utilizado, pois é mais adequado para lidar com essas superfícies. A combinação dos métodos garante uma estimativa de profundidade mais confiável em diferentes tipos de superfícies, reduzindo a propagação de erros no mapa de altura e gerando mapas de altura mais completos e precisos.

Dadas as possibilidades da criação de uma solução aprimorada por meio da integração dos dois métodos apresentados, este projeto propõe combinar os métodos estéreo multifocal e estéreo fotométrico para a criação de modelos tridimensionais a partir de imagens capturadas de objetos-alvo.