

Métodos Fotométricos para Visão Computacional 3D

Este exemplar corresponde à redação da Tese apresentada para a Banca Examinadora antes da defesa da Tese.

Campinas, 26 de março de 2009.

Jorge Stolfi
IC-Unicamp (Orientador)

Tese apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Métodos Fotométricos para Visão Computacional 3D

Rafael Felipe Veiga Saracchini¹

Março de 2009

Banca Examinadora:

- Jorge Stolfi
IC-Unicamp (Orientador)
- Hélio Pedrini
IC-Unicamp
- José R. A. Torreão
IC-UFF
- Wu Shin-Ting
FEEC-Unicamp
- Neucimar J. Leite
IC-Unicamp
- Helena C. G. Leitão
IC-UFF

¹Suporte financeiro de: Bolsa FAPESP (processo 07/59509-9) 2007–2009 e Universidade Estadual de Campinas 2007–2009.

Resumo

O principal objetivo deste projeto de doutorado é o desenvolvimento e implementação de técnicas inovadoras para *visão estereoscópica* ou *estéreo* — a reconstrução de um modelo tridimensional de uma cena, a partir de duas ou mais imagens bidimensionais da mesma. Especificamente, planejamos estudar e aprimorar métodos que utilizam imagens tomadas sob várias condições de iluminação (*estéreo fotométrico*) e a partir de vários pontos de vista (*estéreo geométrico*). Planejamos também investigar aplicações práticas dessas técnicas, em particular na montagem automática de fragmentos arqueológicos.

Sumário

Resumo	iii
1 Introdução	1
1.1 Técnicas de estéreo computacional	1
1.1.1 Estéreo Geométrico	1
1.1.2 Estéreo Fotométrico	2
1.1.3 Limitações e comparação	2
1.2 Objetivos	3
1.2.1 Aceleração do estéreo fotométrico	3
1.2.2 Estéreo fotométrico robusto	3
1.2.3 Combinação de estéreo geométrico e fotométrico	3
2 Estéreo Fotométrico	5
2.1 Estéreo Fotométrico baseado em Gabaritos	5
2.2 Equações Fundamentais	8
2.3 Tabela de assinaturas	10
3 Aceleração por Hashing 2D	11
3.1 Trabalhos anteriores	11
3.2 Formato da tabela de assinaturas	12
3.3 Construção da estrutura	12
3.3.1 Determinação do plano principal	12
3.3.2 Definição das células	13
3.3.3 Divisão da tabela em baldes	13
3.4 Busca na grade	13
3.4.1 Descrição geral do algoritmo de busca	14
3.4.2 Ordem de busca na grade e término antecipado	15
3.5 Análise de complexidade	16
3.5.1 Soluções alternativas	17

3.6	Testes	17
3.6.1	Testes com imagens sintéticas	18
3.6.2	Testes com fotos de uma cena real	20
4	Sombras e Highlights	31
4.1	Anomalias de iluminação e acabamento	31
4.1.1	Iluminação externa não uniforme	31
4.1.2	Sombras projetadas	32
4.1.3	Anomalias de acabamento	32
4.2	Pontos canônicos	32
4.2.1	A hipótese de dados suficientes	33
4.2.2	Fator de anomalia	33
4.3	Estéreo fotométrico na presença de anomalias	34
4.3.1	Enumeração exaustiva	34
4.3.2	Descarte prévio de anomalias	34
4.4	Nossa abordagem: uma nova métrica	35
4.4.1	A métrica Δ	36
4.4.2	A métrica Δ_2	37
4.4.3	A métrica Δ_9	38
4.5	Testes	40
4.5.1	Testes com cena sintética	40
4.5.2	Teste com cenas reais	40
5	Aplicação - Captura de geometria facial	46
5.1	Soluções anteriores	46
5.2	Validação do estéreo fotométrico para face humana	47
5.3	Detalhamento de nossa abordagem	48
5.3.1	Escolha dos padrões de iluminação	49
5.3.2	Captura das Imagens	49
5.3.3	Alinhamento das Imagens	50
5.3.4	Gabaritos de iluminação	50
5.3.5	Cálculo de normais	50
5.3.6	Recuperação da superfície	51
5.4	Testes	51
6	Conclusões	54
6.1	Contribuições	54
6.1.1	Aceleração de estéreo fotométrico	54
6.1.2	Interpolação e suavização de gabaritos	54

6.1.3	Estéreo fotométrico robusto	54
6.2	Aplicações	55
6.3	Trabalhos Futuros	55
6.3.1	Aceleração de estéreo fotométrico robusta	55
6.3.2	Associação com estéreo geométrico	55
	Bibliografia	56

Capítulo 1

Introdução

A captura de modelos tridimensionais de objetos reais é um problema computacional com inúmeras aplicações. Podemos citar a documentação e restauração de monumentos, a engenharia reversa de peças ou prédios, a operação autônoma de robôs, a fabricação de próteses, o monitoramento de animais ou ambientes, etc..

Na última década, foram desenvolvidos vários tipos de *scanners tridimensionais*, aparelhos que capturam a geometria 3D de um objeto diretamente — por exemplo, medindo o tempo de trânsito de pulsos *laser* dirigidos ao objeto. Entretanto, tais equipamentos ainda são caros e volumosos, e seu campo de aplicação ainda é relativamente restrito. Uma alternativa a *scanners* 3D são técnicas baseadas em *luz estruturada* em que uma textura conhecida (linha, grade, rede de pontos) é projetada obliquamente sobre o objeto que é fotografado com uma câmera comum. O deslocamento aparente da textura na imagem permite determinar a profundidade. Estas técnicas, embora mais baratas que *scanners* 3D, exigem um projetor de boa resolução e são limitadas a objetos de baixo relevo.

Assim, ainda há interesse em métodos de *visão estereoscópica* (ou *estéreo*) *computacional*, que permitem obter a terceira dimensão a partir de fotos digitais comuns, dispensando equipamentos especializados.

1.1 Técnicas de estéreo computacional

As técnicas que podem ser usadas para esse fim podem ser divididas em duas classes principais: *estéreo geométrico* e *estéreo fotométrico*.

1.1.1 Estéreo Geométrico

Em estéreo geométrico, utilizam-se duas ou mais fotos do objeto, geralmente tiradas em condições semelhantes de iluminação, mas de posições diferentes. O deslocamento da

câmera entre uma foto e outra dá origem a deslocamentos aparentes (*paralaxes*) das várias partes da cena na imagem, que variam conforme a profundidade dessas partes. Assim, é possível recuperar a terceira dimensão da cena identificando os pontos das imagens dadas que correspondem ao mesmo ponto na cena, e analisando os deslocamentos relativos desses pontos.

1.1.2 Estéreo Fotométrico

Em estéreo fotométrico, utilizam-se três ou mais fotos do objeto, todas tiradas do mesmo ângulo e distância, mas com iluminações diferentes. Em princípio, os valores de de um mesmo pixel em todas essas fotos permitem determinar a inclinação do ponto da superfície do objeto visível nesse pixel. O *mapa de inclinações* que resulta desta análise pode ser então ser integrado para se obter um *mapa de profundidade*, que, para cada pixel, fornece a terceira dimensão do ponto correspondente da cena.

1.1.3 Limitações e comparação

No estéreo geométrico, a identificação de pontos correspondentes exige a comparação de áreas de contorno com vários pixels de largura. Portanto, a resolução lateral do estéreo geométrico é geralmente bem menor que a das imagens de entrada. Além disso, a precisão do estéreo geométrico é inversamente proporcional ao ângulo entre as direções de visão, o que tipicamente significa erros de vários milímetros no mapa de profundidade obtido.

Em comparação, métodos de estéreo fotométricos permitem obter a inclinação de uma superfície em cada pixel individualmente. A resolução lateral é portanto limitada pela resolução das imagens digitais da cena. Por exemplo: uma câmera digital com sensor de 10 megapixels permite resolver detalhes de 0,5mm de diâmetro em um objeto de 50cm de diâmetro.

Métodos de estéreo fotométrico funcionam muito bem para objetos com superfícies lisas, uniformes e sem textura, onde o estéreo geométrico não pode ser usado. Neste, mesmo em objetos que possuem textura, determinar a correspondência entre pontos de duas imagens não é um problema trivial.

O estéreo fotométrico fornece apenas a inclinação da superfície e necessita da integração do mapa de inclinações para produzir o mapa de profundidade da cena, o que pode amplificar erros. Além disso a perspectiva fixa da câmera implica que não há informação suficiente para determinar a profundidade de degraus abruptos (descontinuidades) do mapa de profundidades.

Métodos de estéreo geométrico não precisam supor quase nada sobre a natureza da superfície. Por exemplo, podem ser utilizados em superfícies muito rugosas ou esponjosas como tecidos e cascalho, que respondem a variações de luz de maneira muito com-

plexa. Outra diferença do estéreo geométrico é sua menor sensibilidade a anomalias de iluminação, como sombras, penumbras, inter-reflexões, *highlights* e iluminação não uniforme.

1.2 Objetivos

O tema central desta tese de doutorado é o desenvolvimento de métodos inovadores para estéreo fotométrico e sua aplicação a problemas práticos selecionados.

1.2.1 Aceleração do estéreo fotométrico

Métodos estéreo fotométricos nem sempre utilizam apenas 3 diferentes iluminações. Com o intuito de obter de eliminar defeitos causados por ruído, melhor precisão ou garantir uma quantidade mínima de informação confiável por pixel, métodos estéreo fotométricos normalmente utilizam uma quantidade maior de imagens sob diferentes condições de iluminação. Em métodos convencionais o tempo de cálculo das normais cresce de maneira linear de acordo com a quantidade de imagens fornecidas. Neste trabalho apresentamos um método de aceleração proporcional ao grau de precisão desejado, não crescendo de maneira linear em relação à quantidade de imagens, baseado no comportamento que o conjunto de amostras de um objeto de referência apresenta em um espaço \mathbb{R}^m .

1.2.2 Estéreo fotométrico robusto

Métodos de estéreo fotométrico são extremamente sensíveis a anomalias de iluminação. Para muitas aplicações importantes, como por exemplo: reconstrução de objetos arqueológicos ou faces humanas, é necessário melhorar a análise de modo a tornar estes métodos mais robustos. Neste trabalho apresentamos um método eficiente para utilizar apenas imagens válidas para um determinado ponto descartando informação proveniente de pixels afetados por anomalias.

1.2.3 Combinação de estéreo geométrico e fotométrico

As duas abordagens, fotométrica e geométrica, foram muito estudadas nas últimas décadas. Há entretanto relativamente poucos trabalhos que tentam combinam as duas abordagens. Possivelmente, isso tem a ver com a natureza das informações obtidas. O estéreo geométrico produz diretamente a profundidade de cada ponto, enquanto que o estéreo fotométrico essencialmente produz as derivadas dessa informação. Considerando as características complementares dos dois métodos, discutidos na seção 1.1.3, acreditamos

que a combinação aproveitando os pontos fortes destas técnicas, proporcionará resultados mais confiáveis e precisos do que cada um separadamente.

Capítulo 2

Estéreo Fotométrico

Neste capítulo vamos descrever em detalhe os princípios do estéreo fotométrico de iluminação variável (EFIV). Neste método, o objetivo é determinar a geometria tridimensional de uma cena a partir de uma lista de $m \geq 3$ imagens digitais monocromáticas S_1, S_2, \dots, S_m obtidas em diferentes condições de iluminação com um ponto de vista fixo em uma determinada posição. R. J. Woodham em 1980 [22] demonstrou que analisando-se as m intensidades de pixel $S_i[p]$ em um ponto p da imagem, é possível recuperar-se o vetor unitário $\vec{s}[p]$ perpendicular à superfície visível no ponto p . Este resultado inicial foi bastante discutido nos anos seguintes [5, 15, 19, 25, 4, 24].

Para efetuar a análise supracitada é em princípio necessário conhecer a *função bidirecional de distribuição da radiância*²(FBDR) da superfície, e o *campo de iluminação* Φ_i em cada imagem S_i . A FBDR de uma superfície é uma função $\sigma(\vec{n}, \vec{u}, \vec{v})$ que fornece o brilho aparente (*radiância*) da superfície quando orientada perpendicularmente à direção \vec{n} , observada a partir da direção \vec{u} e iluminada por uma luz uni-direcional de intensidade unitária vindo da direção \vec{v} .

2.1 Estéreo Fotométrico baseado em Gabaritos

Na variante do estéreo fotométrico conhecida por *estéreo fotométrico baseado em gabaritos*(EFBG) as informações necessárias sobre a FBDR e o fluxo de luz são fornecidas indiretamente por m imagens G_i, \dots, G_m de um objeto de forma e cor conhecida, o *gabarito de iluminação*. É conveniente usar gabaritos esféricos com albedo uniforme, preferivelmente branco ou cinza claro. Cada imagem G_i deve ser obtida sob as mesmas condições de iluminação da cena S_i correspondente. Dependendo da aplicação é conveniente incluir o gabarito de iluminação como parte da cena; neste caso, G_i passa a ser um subconjunto

²Bidirectional radiance distribution function - BRDF

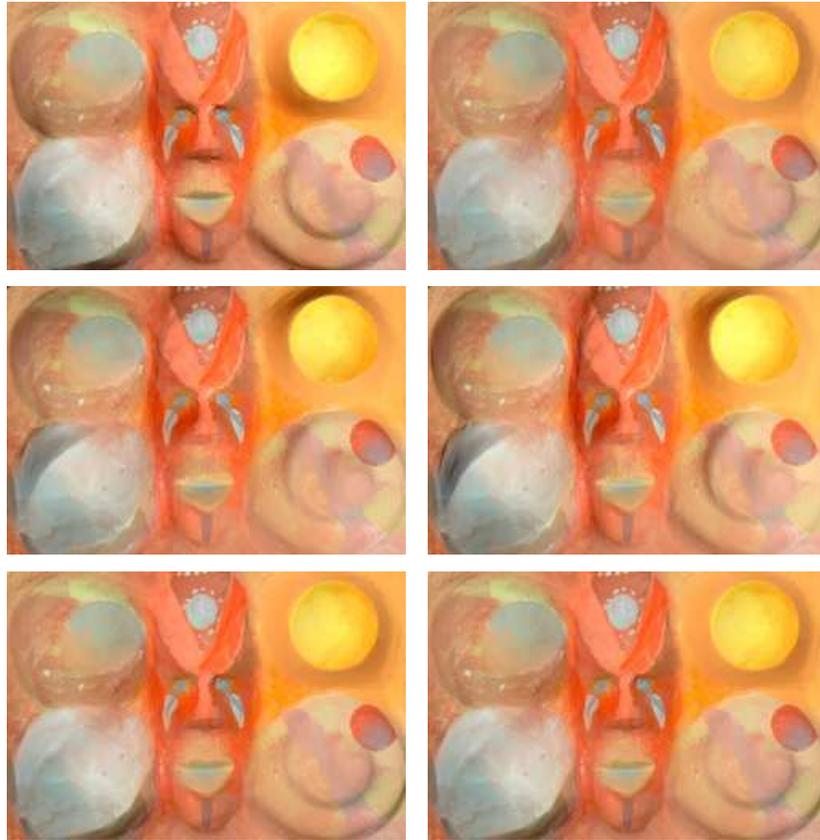


Figura 2.1: Imagens de objetos sob diferentes condições de iluminação.

de S_i . Veja figura 2.2.

Neste capítulo vamos supor que todas as imagens S_1, \dots, S_m estão geometricamente corrigidas, cortadas e alinhadas, o que implica que cada ponto p em seu domínio comum \mathcal{S} corresponde ao mesmo ponto na parte visível superfície da cena. Vamos supor que estas condições valem também para as imagens do gabarito G_1, \dots, G_m , cujo domínio será denominado \mathcal{G} .

Finalmente, vamos supor que também que a codificação dos pixels é linear, ou seja, que o valor de cada pixel é diretamente proporcional à intensidade luminosa que incide sobre o elemento sensor correspondente.

Em sua forma básica, o estéreo fotométrico baseado em gabaritos é viável apenas se todos os pontos da superfície visível da cena e dos gabaritos possuírem o mesmo acabamento, exceto por variações na cor intrínseca. Isto é, as FBDRs $\sigma[p]$ de cada ponto p da cena e $\gamma[q]$ de cada ponto q do gabarito, devem ser múltiplas de uma FBDR fixa $\bar{\sigma}$:

$$\begin{aligned} \sigma[p](\vec{n}, \vec{u}, \vec{v}) &= s[p] \bar{\sigma}(\vec{n}, \vec{u}, \vec{v}) \\ \gamma[q](\vec{n}, \vec{u}, \vec{v}) &= g[q] \bar{\sigma}(\vec{n}, \vec{u}, \vec{v}) \end{aligned} \tag{2.1}$$

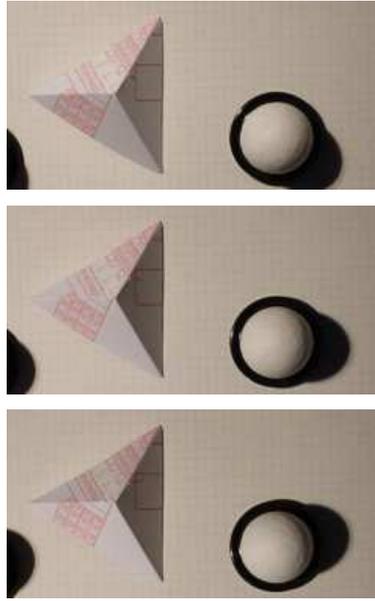


Figura 2.2: Gabaritos de iluminação como parte da cena

Os fatores $\bar{s}[p]$ e $\bar{g}[p]$ nestas fórmulas são os *albedos* da cena e do gabarito, respectivamente, nos pontos p e q .

Uma condição necessária para estéreo fotométrico baseado em gabaritos é que a que a FBDR $\bar{\sigma}$ seja dominada por termos difusivos, ou seja, a superfície deve distribuir a luz incidente de qualquer direção \vec{u} sobre um leque amplo de direções \vec{v} . O exemplo canônico é a da FBDR lambertiana:

$$\bar{\sigma}(\vec{n}, \vec{u}, \vec{v}) = \max \{0, -\vec{u} \cdot \vec{n}\} \quad (2.2)$$

Note que, para superfícies lambertianas, a radiância não depende da direção de observação \vec{v} . O método funciona para qualquer FBDR, contanto que esta seja suave, sem concentrações excessivas de luz em certas direções. Para a análise é necessário conhecer a direção $\vec{g}[q]$ e o albedo $\bar{g}[q]$ para todo $q \in \mathcal{G}$.

Para simplificar a exposição vamos supor neste capítulo que as imagens da cena foram obtidas sob projeção quase paralela e iluminada por uma fonte luminosa distante, ou seja, que a direção \vec{v} do ponto de vista e as condições de iluminação da cena serão os mesmos para cada ponto de \mathcal{S} e \mathcal{G} . Quando esta condição não é válida, as normais calculadas precisam sofrer uma correção possivelmente geométrica (que não envolve considerações fotométricas).

A técnica EFBG pode ser estendida também para imagens com vários canais de cor (por exemplo, RGB). Neste caso o resultado é um único mapa de normais $\vec{s}[p]$ porém um mapa diferente de albedo $\bar{s}_\lambda[p]$ para cada canal λ . A combinação destes mapas de

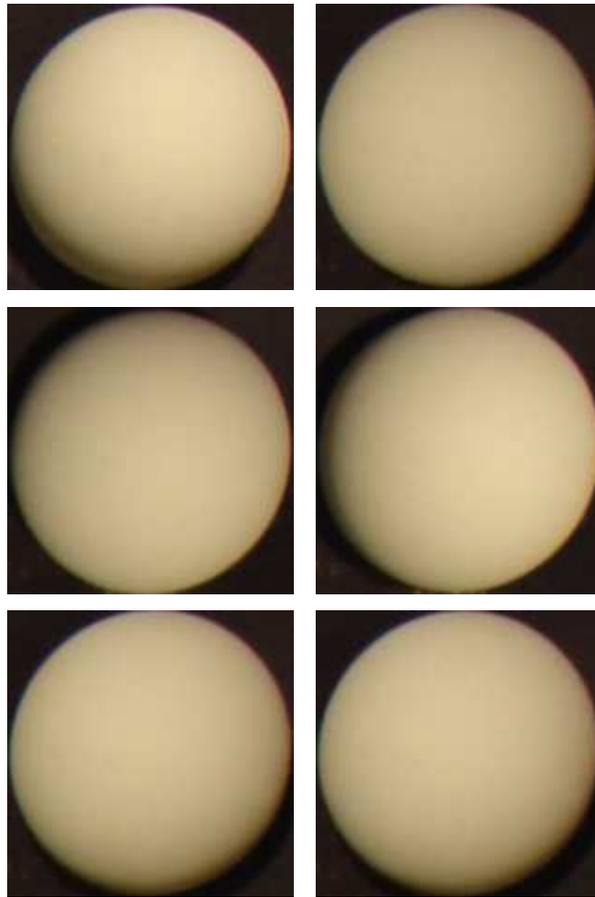


Figura 2.3: Imagens de gabaritos sob as mesmas condições de iluminação dos objetos da figura 2.1.

albedo fornece um mapa de cor intrínseco que fornece a cor própria de cada pixel da cena, independente da iluminação.

2.2 Equações Fundamentais

A idéia principal do estéreo fotométrico baseado em gabaritos é que a intensidade em um ponto de uma imagem da cena \mathcal{S}_i (ou de um gabarito \mathcal{G}_i) pode ser decomposta no produto de dois fatores: o albedo intrínseco \bar{s} ou \bar{g} , que depende apenas do material ou acabamento da superfície no ponto, e o *fator de iluminação* $L_i(\vec{n})$, que depende apenas do índice i e da normal \vec{n} da superfície. Mais especificamente:

$$\begin{aligned} S_i[p] &= \bar{s}[p] L_i(\vec{s}[p]) \\ G_i[q] &= \bar{g}[q] L_i(\vec{g}[q]) \end{aligned} \tag{2.3}$$

Nestas equações L_i é uma *função de tonalização*, determinada pelo campo de iluminação Φ_i e pela FBDR $\bar{\sigma}$, que mapeia cada vetor unitário \vec{n} para a radiância de uma superfície branca perpendicular a \vec{n} . A função L_i é dada pela fórmula:

$$L_i(\vec{n}) = \int_{\mathbb{S}^2} \Phi_i(\vec{u}) \bar{\sigma}(\vec{n}, \vec{u}, \vec{v}) d\vec{u} \quad (2.4)$$

O fator $\Phi_i(\vec{u})$ nesta fórmula é a intensidade da luz que flui na direção \vec{u} . A hipótese da iluminação uniforme permite supor que o fluxo luminoso $\Phi_i(\vec{u})$ não depende explicitamente da posição do ponto da superfície, mas apenas da direção \vec{u} .

Observe que o valor de $\Phi_i(\vec{u})$ é irrelevante para direções \vec{u} que apontam na direção oposta à normal \vec{n} da superfície, ou em pontos que não estão na superfície visível da cena. Isto permite supor que existe apenas uma função $\Phi_i(\vec{u})$ para toda a cena independentemente da orientação da superfície. Em particular, este modelo de iluminação permite sombras próprias e é adequado para cenas que consistem de apenas um objeto simples e convexo. Por outro lado este modelo não leva em conta sombras projetadas, iluminação não uniforme, efeitos de iluminação secundária por luz espelhada por outras partes da cena (radiosidade).

Baseado nas equações (2.3) podemos determinar a normal $\vec{s}[p]$ no ponto p das imagens da cena encontrando um ponto na imagem do gabarito que reage da mesma maneira que p a mudanças de direção na iluminação, exceto por diferença de albedo. Mais precisamente, devemos encontrar $q \in \mathcal{G}$ tal que os os vetores

$$\begin{aligned} \mathbf{S}[p] &= (S_1[p], S_2[p], \dots, S_m[p]) \\ \mathbf{G}[q] &= (G_1[p], G_2[p], \dots, G_m[p]) \end{aligned} \quad (2.5)$$

sejam múltiplos um do outro. Os vetores $\mathbf{S}[p]$ e $\mathbf{G}[q]$ são denominados *vetores de observação* dos pontos p e q . Uma vez localizado o ponto q adequado, podemos recuperar o vetor normal $\vec{s}[p]$ e o albedo $\hat{s}[p]$ do ponto p da cena pela fórmula abaixo:

$$\vec{s}[p] = \vec{g}[q] \quad \hat{s}[p] = |\mathbf{S}[p]| \frac{\hat{g}[q]}{|\mathbf{G}[q]|} \quad (2.6)$$

Nesta equação, $|\cdot|$ é qualquer norma de \mathbb{R}^m , por exemplo, a norma euclidiana $|\mathbf{X}| = \sqrt{\sum_{i=1}^m X_i^2}$.

Este método pode falhar se existirem dois pontos q', q'' no gabarito que possuem normais diferentes (ou seja, $\vec{g}[q'] \neq \vec{g}[q'']$) mas vetores de observação colineares (isto é, $\mathbf{G}[q'] = \alpha \mathbf{G}[q'']$ para algum escalar α). Para evitar este problema, o número de imagens m deve ser no mínimo 3, e os campos de iluminação devem ser suficientemente variados para eliminar qualquer tipo de ambigüidade. No restante deste texto, vamos supor que estas condições estão satisfeitas.

2.3 Tabela de assinaturas

No estéreo fotométrico baseado em gabaritos, o problema fundamental é a determinação do ponto $q \in \mathcal{G}$ tal que o vetor de observação $\mathbf{S}[p]$ é múltiplo de $\mathbf{G}[q]$. Se nenhum dos dois vetores for nulo esta condição é equivalente a dizer que p e q tem as mesmas *assinaturas de observação* $\mathbf{s}[p]$ e $\mathbf{g}[q]$, definidas por

$$\mathbf{s}[p] = \frac{\mathbf{S}[p]}{|\mathbf{S}[p]|} \quad \mathbf{g}[q] = \frac{\mathbf{G}[q]}{|\mathbf{G}[q]|} \quad (2.7)$$

Observe que a posição q não é relevante em si, sendo apenas utilizada para associar a assinatura de observação $\mathbf{g}[q]$ a normal $\vec{g}[q]$ e ao fator de albedo $\dot{g}[q] = \dot{g}/|\mathbf{G}[q]|$. Portanto, podemos substituir as imagens do gabarito por um conjunto não ordenado \mathbf{T} de triplas da forma $t = (t.\mathbf{g}, t.\vec{g}, t.\dot{g})$:

$$\mathbf{T} = \{ (\mathbf{g}[q], \vec{g}[q], \dot{g}[q]) : q \in \mathcal{G} \} \quad (2.8)$$

O conjunto \mathbf{T} é a uma *tabela de assinaturas* derivada das imagens $\mathbf{G}_1 \dots \mathbf{G}_m$. A determinação de $\vec{s}[p]$ torna-se então um problema de busca do valor mais próximo nesta tabela, onde procuramos pelo elemento t da tabela \mathbf{T} que minimiza a distância $\text{dist}(t.\mathbf{g}, \mathbf{s}[p])$. A métrica $\text{dist}(\cdot, \cdot)$ é qualquer métrica definida sobre a esfera \mathbb{S}^{m-1} . Por exemplo, podemos usar a própria a métrica Euclidiana de \mathbb{R}^m , ou seja: $\text{dist}(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$.

Capítulo 3

Aceleração por Hashing 2D

A etapa mais custosa no estéreo fotométrico é a determinação da entrada mais próxima na tabela de assinaturas, ou seja a que minimiza a distância $\text{dist}(t, \mathbf{g}, \mathbf{s}[p])$. A solução “força bruta” deste problema consiste em varrer a tabela \mathbf{T} , calcular a distância $\text{dist}(t, \mathbf{g}, \mathbf{s}[p])$ para cada entrada t , lembrando a entrada com melhor aproximação. Entretanto, prover uma boa cobertura de todas as possíveis direções é geralmente necessário usar uma tabela \mathbf{T} com dezenas de milhares de entradas. Uma vez que a busca deve ser efetuada para cada pixel do domínio \mathcal{G} da cena, esta etapa pode levar horas, mesmo para um conjunto pequeno de imagens. Neste capítulo descreveremos um método para acelerar a busca da entrada t em tempo $O(1)$ utilizando uma grade de *hashing* bidimensional.

3.1 Trabalhos anteriores

Várias técnicas foram propostas para acelerar etapa de busca na tabela. Woodham [23, 24] utilizou uma grade m -dimensional regular em forma de um hipercubo $[0, 1]^m$ com 2^b células em cada eixo, para algum inteiro $b \geq 1$. Na etapa de pré-processamento cada vetor de observação $\mathbf{G}[q]$ do gabarito é quantizado utilizando-se b bits por coordenada. O resultado é a tupla de m índices de uma célula da grade; nessa célula é armazenado o vetor normal $\vec{g}[q]$ (Woodham supôs albedos uniformes e iguais $\vec{s} = \vec{g}$, sendo então desnecessário normalizar os vetores de observação). Na etapa de busca, cada vetor de observação $\mathbf{S}[p]$ é quantizado da mesma maneira, sendo a normal $\vec{s}[p]$ lida da célula correspondente. Uma desvantagem óbvia deste método é o tamanho da estrutura: 2^{mb} células, o que significa em torno de 250.000 para $m = 3$ e $b = 6$.

Trabalhos mais recentes propuseram o uso de algoritmos genéricos para busca do ponto mais próximo em espaços m -dimensionais. Por exemplo. Hertzmann e Seitz [14] utilizam

um algoritmo de vizinho mais próximo aproximado¹ de Arya et al. [1], enquanto Zhong e Little [25] utilizam hashing localmente sensível (*locally sensitive hashing*) de Indyk e Motwani [16]. Contudo, todos estes métodos supõe que o conjunto das assinaturas do gabarito, $\mathbb{G} = \{\mathbf{g}[q] : q \in \mathcal{G}\}$, é uma nuvem genérica de pontos espalhados no espaço m -dimensional. Nestas condições, os algoritmos gerais de busca de vizinho mais próximo são bastante custosos em termos de tempo e/ou espaço[16].

3.2 Formato da tabela de assinaturas

Nosso método para aceleração da busca se baseia em uma observação não trivial: o fato de que conjunto de assinaturas \mathbb{G} do gabarito é essencialmente um subconjunto *bidimensional* de \mathbb{R}^m . Portanto, podemos reduzir a etapa de busca na tabela por uma busca de ponto mais próximo em um plano, o qual pode ser feito eficientemente por uma estrutura de dados conhecida como *grade bidimensional de baldes* (*2D bucket grid*) [10].

Para justificar esta observação, notamos que, devido às fórmulas (2.3) e (2.7), as assinaturas de observação $\mathbf{s}[p]$ e $\mathbf{g}[q]$ podem ser escritas como $\mathbf{l}(\vec{s}[p])$ e $\mathbf{l}(\vec{g}[q])$ respectivamente, onde \mathbf{l} é uma função de *luminância da assinatura*

$$\mathbf{l}(\vec{n}) = \frac{\mathbf{L}(\vec{n})}{|\mathbf{L}(\vec{n})|} \quad (3.1)$$

e $\mathbf{L}(\vec{n}) = (L_1(\vec{n}), \dots, L_m(\vec{n}))$. Note que a função \mathbf{l} , que mapeia normais da superfície para assinaturas de iluminação, é definida apenas no hemisfério \mathbb{H}^2 de \mathbb{S}^2 que consiste das direções que fazem ângulo menor que 90 graus com a direção do ponto de vista \vec{v} . Por outro lado, um bom gabarito deve fornecer uma amostragem densa e uniforme do hemisfério \mathbb{H}^2 (razão pela qual são utilizadas esferas como gabaritos). Segue-se então que o conjunto de assinaturas \mathbb{G} do gabarito deve ser uma cobertura densa e uniforme do conjunto $\mathbb{K} = \mathbf{l}(\mathbb{H}^2)$, que é o conjunto-imagem da função \mathbf{l} . Para superfícies predominantemente difusoras, a função \mathbf{l} é bastante suave. Concluí-se que \mathbb{G} deve ser também um conjunto bidimensional, ainda que curvo, do espaço \mathbb{R}^m .

3.3 Construção da estrutura

3.3.1 Determinação do plano principal

Em nosso método, a tabela de assinaturas é preprocessada através do procedimentos descritos abaixo. Inicialmente determinamos um plano P de \mathbb{R}^m . Especificamente, de-

¹approximate nearest neighbour - ANN

terminamos o centróide \mathbf{b} do conjunto \mathbb{G} de assinaturas (visto como um conjunto de pontos de \mathbb{R}^m), e dois vetores unitários ortogonais $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ que definem as direções de extensão máxima de \mathbb{G} . Estes vetores são obtidos construindo-se a matriz M (com dimensões $m \times m$) de momentos de coordenadas, dos deslocamentos $\mathbf{g} - \mathbf{b}$, para todo $\mathbf{g} \in \mathbb{G}$ $M_{ij} = \sum_{\mathbf{g} \in \mathbb{G}} (g_i - b_i)(g_j - b_j)$. Os vetores \mathbf{u}, \mathbf{v} são então os os autovetores unitários associados aos dois maiores autovalores de M .

O ponto \mathbf{b} e os vetores \mathbf{u}, \mathbf{v} definem o plano P , denominado *plano de projeção de assinaturas*, o qual é aproximadamente coplanar com o conjunto \mathbb{G} . A projeção ortogonal em P de uma dada assinatura de observação \mathbf{g} será denotada por $\mathbf{g} \downarrow P$.

3.3.2 Definição das células

Em seguida, definimos uma grade regular de $N \times N$ células quadradas no plano de projeção P . Esta grade é centrada no ponto \mathbf{b} , possui seus lados paralelos aos vetores \mathbf{u} e \mathbf{v} , e é um pouco maior que o necessário para conter a projeção $\mathbf{g} \downarrow P$ de qualquer assinatura de observação \mathbf{g} de \mathbb{G} . Mais precisamente, a grade é um quadrado de tamanho $2R$, onde:

$$R = \varepsilon + \max \{ |(\mathbf{g} - \mathbf{b}) \cdot \mathbf{u}|, |(\mathbf{g} - \mathbf{b}) \cdot \mathbf{v}| : \mathbf{g} \in \mathbb{G} \} \quad (3.2)$$

e ε é uma pequena margem de segurança. Cada célula da grade é então um quadrado com tamanho $\tau = 2R/N$.

3.3.3 Divisão da tabela em baldes

Tendo definido a grade, construímos então, para cada célula $C[i, j]$ da grade, uma lista encadeada $B[i, j]$ com toda entrada t em T tal que a projeção $t.\mathbf{g} \downarrow P$ de suas assinaturas caem no interior da célula. Esta lista é o *balde* da célula $[i, j]$. Calculamos também $\mu[i, j]$ definido o *centróide* de todas as assinaturas $t.\mathbf{g}$ na lista de $B[i, j]$, e $\rho[i, j]$ definido como a maior distância de $\mu[i, j]$ de qualquer assinatura $t.\mathbf{g}$ nesta lista. Veja a figura 3.1.

O formato bidimensional de \mathbb{G} significa que as entradas da tabela em $B[i, j]$ estão muito próximas umas das outras, mesmo se a distancia média do plano P é grande comparada com o tamanho da célula. Esta propriedade é confirmada pelos nossos testes (seção 3.6) e continua válida mesmo quando m é bem maior que 3.

3.4 Busca na grade

Uma vez construída a grade de baldes B , cada assinatura $\mathbf{s}[p]$ da cena é processada através do procedimento 1 abaixo. Suas etapas, incluindo a lista Δ e a função Λ são descritas nas seções 3.4.1 e 3.4.2.

Procedure 1 (Busca na Tabela) *Dada uma assinatura de observação \mathbf{s} , encontrar a entrada $t_{best} \in T$ com assinatura $t_{best}.\mathbf{g}$ mais similar a \mathbf{s} .*

1. $i \leftarrow \lfloor ((\mathbf{g} - \mathbf{b}) \cdot \mathbf{u} + R) / \tau \rfloor$;
2. $j \leftarrow \lfloor ((\mathbf{g} - \mathbf{b}) \cdot \mathbf{v} + R) / \tau \rfloor$;
3. $dbest \leftarrow +\infty$; $t_{best} \leftarrow \mathbf{null}$;
4. Para cada par (δ_i, δ_j) na lista Δ , na ordem, faça
 5. Se $dbest^2 \leq \tau^2 \Lambda^2(\delta_i, \delta_j)$, retorne t_{best} .
 6. $(i', j') \leftarrow (i, j) + (\delta_i, \delta_j)$;
 7. Se $0 \leq i' < N$ e $0 \leq j' < N$; então
 8. Se $dbest > \text{dist}(\mathbf{s}, \mu[i', j']) - \rho[i', j']$, então
 9. Para cada t em $B[i', j']$, faça
 10. Defina $d \leftarrow \text{dist}(\mathbf{s}, t.\mathbf{g})$;
 11. Se $d < dbest$, defina $dbest \leftarrow d$ e $t_{best} \leftarrow t$.
12. Retorne t_{best} .

3.4.1 Descrição geral do algoritmo de busca

O Procedimento 1 começa calculando os índices (i, j) da célula que contém a projeção $\mathbf{s} \downarrow P$ de uma dada assinatura \mathbf{s} (passos 1–2). Nos passos 3–11 o procedimento busca a entrada mais próxima de \mathbf{s} começando pelo balde $B[i, j]$. Se necessário, a busca prossegue analisando buckets $B[i', j']$ vizinhos em uma ordem apropriada. Observe que alguns buckets podem estar vazios, e que o melhor casamento para \mathbf{s} pode não estar no balde $B[i, j]$, mesmo que este não esteja vazio.

No passo 8, os atributos $\mu[i, j]$ e $\rho[i, j]$ permitem ignorar buckets que não tem possibilidade de oferecer um melhor par do que já encontrado para a assinatura \mathbf{s} . Mais precisamente, um bucket $B[i', j']$ deve ser examinado apenas se a assinatura \mathbf{s} está mais próxima da esfera delimitadora do bucket do que o melhor par já encontrado até o momento na tabela. Ou seja, apenas se

$$\text{dist}(\mathbf{s}, \mu[i', j']) - \rho[i', j'] < \text{dist}(\mathbf{s}, t_{best}.\mathbf{g}) = dbest \quad (3.3)$$

sendo que $\text{dist}(\mathbf{s}, t_{best}.\mathbf{g})$ é o valor da variável $dbest$. Denominamos a desigualdade (3.3) como *condição de percurso de bucket*.

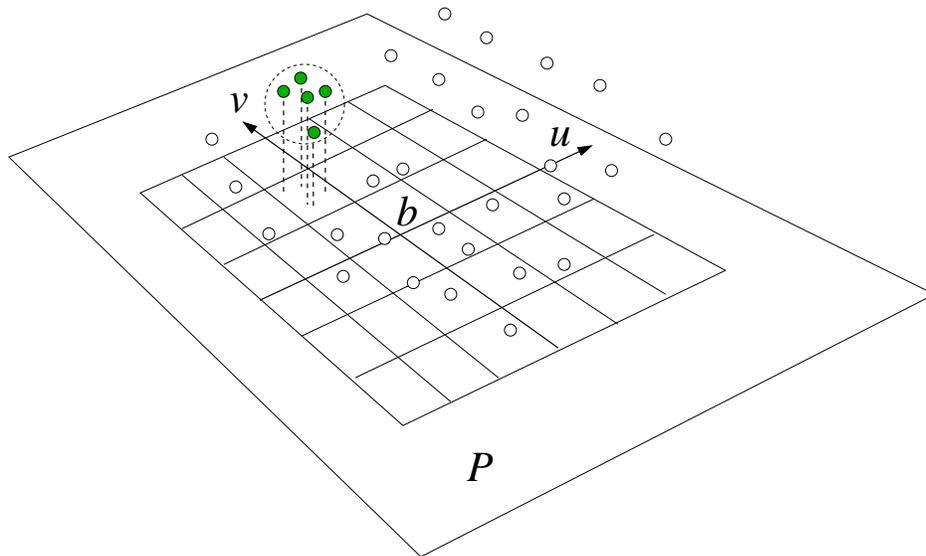


Figura 3.1: Ilustração esquemática da grade de baldes, para $m = 3$, mostrando algumas assinaturas de observação em \mathbb{G} (pequenos círculos vazios), um balde $B[i, j]$ (pequenos círculos preenchidos) e a esfera delimitadora do balde (círculo pontilhado maior) definida pelo centróide $\mu[i, j]$ e raio $\rho[i, j]$.

3.4.2 Ordem de busca na grade e término antecipado

A condição de percurso de bucket (3.3) nos permite ignorar um bucket $B[i', j']$ sem examinar suas entradas. Entretanto, se meramente aplicarmos este critério para todos os buckets individualmente, o tempo de processamento de uma assinatura ainda seria proporcional ao número de células na grade (N^2). Uma vez que o número de células deve ser proporcional ao tamanho da tabela, o custo do Procedimento 1 seria proporcional ao custo da busca de força bruta.

Para evitar este custo, o Procedimento 1 enumera os buckets $B[i', j']$ em uma ordem específica, começando do bucket que contém a projeção $\mathbf{s} \downarrow P$ e movendo gradualmente para longe dele (passo 4). Um segundo critério (passo 5) permite então abandonar a busca tão logo fique evidente que nenhum dos buckets ainda por examinar podem conter uma assinatura melhor que a encontrada até o momento. Tipicamente isto acontece após examinar uma pequena fração dos buckets.

Mais precisamente, considere duas assinaturas \mathbf{s} e \mathbf{s}'' projetadas ortogonalmente em P nas células $C[i', j']$ e $C[i'', j'']$ respectivamente. É fácil ver que

$$\text{dist}(\mathbf{s}', \mathbf{s}'') \geq \text{dist}(C[i', j'], C[i'', j'']) \quad (3.4)$$

Denominamos esta desigualdade de *condição para término de busca*. Nesta fórmula,

$\text{dist}(C[i', j'], C[i'', j''])$ é a menor distância entre duas células, vistas como subconjuntos do plano P . Esta distância é

$$\text{dist}(C[i', j'], C[i'', j'']) = \tau\Lambda(i' - i'', j' - j'') \quad (3.5)$$

onde

$$\Lambda(\delta_i, \delta_j) = \sqrt{(\max\{0, |\delta_i| - 1\})^2 + (\max\{0, |\delta_j| - 1\})^2} \quad (3.6)$$

Note que $\Lambda(\delta_i, \delta_j)$ é um pouco menor que a norma euclidiana $|(\delta_i, \delta_j)| = \sqrt{\delta_i^2 + \delta_j^2}$. Pode-se concluir que é inútil examinar um balde $B[i', j']$ se o teste (3.4) exclui a possibilidade de haver uma assinatura mais próxima dentro dele, isto é, se

$$\tau\Lambda(i' - i, j' - j) \geq \text{dist}(\mathbf{s}, t\text{best.g}) \quad (3.7)$$

onde $t\text{best}$ é o melhor candidato encontrado até o momento.

Observe que a condição de término de busca (3.7) é mais fraca que a condição de percurso do bucket (3.3). Contudo a condição (3.7) depende apenas do melhor par $t\text{best}$ encontrado e da diferença $(i' - i, j' - j)$ entre índices de células. Portanto, se analisarmos os buckets (i', j') em uma determinada ordem tal que $\Lambda(i' - i, j' - j)$ seja crescente, podemos interromper a busca tão logo a condição (3.7) seja satisfeita (passo 5).

Assim, uma parte do pré-processamento da tabela consiste no cálculo de uma lista Δ de todos os vetores (δ_i, δ_j) em $J \times J$, onde $J = \{-(N - 1)..+(N - 1)\}$, ordenados por valor crescente de $\Lambda(\delta_i, \delta_j)$ (e desempatados por valor crescente de $|(\delta_i, \delta_j)|$). Na prática, é mais conveniente calcular e computar $\Lambda^2(\delta_i, \delta_j)$ em vez de $\Lambda(\delta_i, \delta_j)$. Veja a figura 3.2.

Nos passos 6 do Procedimento 1 usamos a lista Δ , para enumerar os baldes $B[i', j']$ em ordem de distância $\tau\Lambda$ crescente. O teste no passo 7 restringe a busca apenas aos casos em que i' e j' caem dentro da grade.

3.5 Análise de complexidade

O tempo de execução médio do Procedimento 1 é aproximadamente $Bb + Dd + O(1)$, onde b é a média de buckets examinados na etapa 5, d é a média de entradas da tabela testadas (etapa 10), e B, D são os custos associados a cada uma destas operações.

Em um caso extremo onde $N = 1$, teremos $b = 1$ e $d = |\mathbf{T}|$, o que é equivalente a uma busca de força bruta em \mathbf{T} . À medida que N aumenta, d diminui tendendo para um limite que depende do espalhamento das entradas em \mathbb{R}^m na direção perpendicular a \mathbb{R} .

Ao mesmo tempo, b aumenta imediatamente para aproximadamente 10, pois $\Lambda(\delta_i, \delta_j)$ é zero para os primeiro nove pares (δ_i, δ_j) na lista Δ ; proporcionalmente a N^2 , uma vez que

45	37	29	25	30	38	46
39	21	13	09	14	22	40
31	15	05	01	06	16	32
26	10	02	00	03	11	27
33	17	07	04	08	18	34
41	23	19	12	20	24	42
47	43	35	28	36	44	48

Figura 3.2: A ordem de análise dos buckets fornecida pela lista Δ para o subconjunto das 7×7 células adjacentes à célula inicial (localizada no centro). O valor da função de distância da célula $\Lambda^2(\delta_i, \delta_j)$ é indicado pela cor das células: 0(vermelho), 1.0(laranja-escuro), 1.42 (laranja-claro), 2.0 (amarelo), 2.23 (verde) e 2.82 (azul).

o procedimento tem que examinar cada vez mais baldes vazios até encontrar o próximo balde não vazio.

3.5.1 Soluções alternativas

Outras formas de projeção podem ser usadas, como por exemplo projeção em uma superfície esférica contida em \mathbb{R}^m , em vez de um plano. Tal projeção alternativa pode resultar em um espalhamento mais eficiente das entradas na grade. Porém, neste esquema as distâncias entre as células não serão mais uniformes, invalidando a técnica da lista Δ . Seria necessário então usar uma fila de prioridades para enumerar as células na ordem correta.

Esta análise indica que existe um valor ótimo de N que minimiza o tempo de processamento. Este valor depende da razão B/D . Em nossos testes observamos que o tempo total de nossa implementação foi minimizado quando N era em torno de $2\sqrt{|T|}$, o que significa uma média de 0.25 entradas da tabela por balde.

3.6 Testes

Para avaliar o nosso método, efetuamos testes em dois conjuntos de dados: imagens sintéticas e fotos digitais de uma cena real. Embora testes com imagens sintéticas não sejam aceitáveis para uma validação completa de métodos estéreo fotométricos de ilu-

minação variável, eles são úteis para análise da eficiência do procedimento de busca. Isto porque elas fornecem um conjunto \mathbb{G} com forma e tamanho realistas, que são os principais fatores que afetam a eficiência do algoritmo, porém livres de certos defeitos que podem afetar a performance do algoritmo mas que não são inerentes ao mesmo, como resposta não-linear do sensor da câmera, erros de alinhamento de imagem, luz não uniforme e ruído de compressão. Por outro lado, o impacto destes defeitos no comportamento e eficiência do algoritmo também devem ser investigados, o que exige testes com fotos de cenas reais.

Em cada teste, real ou sintético, as imagens dos gabaritos foram amostradas com interpolação de modo obter uma tabela de assinaturas T com aproximadamente 10000 entradas. Utilizamos grades de $N \times N$ baldes com 5 diferentes valores de N : 422, 299, 211, 149 e 105. O número médio $\kappa = |T|/N^2$ de assinaturas por balde foi, respectivamente 0.0625, 0.125, 0.25, 0.50, e 1.00. Cada conjunto de imagens também foi processado com $N = 1$, o que equivale fazer a busca por força bruta. Em todos os testes efetuados foi verificado que as assinaturas de observação retornadas pelo procedimento de busca na tabela foram sempre idênticos aos obtidos por força bruta.

3.6.1 Testes com imagens sintéticas

As imagens sintéticas foram geradas com o software POV-Ray [6]. Usamos uma cena virtual simples, consistindo de blocos geométricos simples de vários formatos e cores posicionados sobre uma superfície plana com textura quadriculada. Veja a figura 3.3. Todos os objetos possuíam acabamento lambertiano, e a cena era iluminada por uma fonte de luz pontual. A distância desta fonte luminosa era cerca de 10000 vezes a largura da parte visível da cena, isto é, distante o suficiente para garantir uma iluminação uniforme para todos os elementos da cena. A câmera virtual foi configurada para capturar uma projeção paralela (ortográfica) da cena. O recurso `shadowless` do POV-Ray foi utilizado para eliminar sombras projetadas mantendo porém sombras próprias.

Foram geradas 48 imagens da cena em formato PNG [18] de 16 bits por canal com dimensões de 630×450 pixels, 24 delas com a fonte luminosa formando um ângulo de $\theta = 25^\circ$ com a vertical e 24 formando um ângulo de $\theta = 45^\circ$, com ângulos de azimute igualmente espaçados. As imagens foram posteriormente convertidas para o formato PPM [13] de 16 bits, reduzidas em 50% nos dois eixos e finalmente convertidas para escala de cinza pela fórmula $Y = 0.2989 R + 0.5866 G + 0.1145 B$. A tabela T foi gerada internamente a partir das direções conhecidas das fontes luminosas, da geometria do gabarito e das fórmulas de *shading* lambertiano utilizadas pelo POV-Ray.

As figuras 3.4 mostram a saída do nosso programa para $m = 24$ imagens de entrada e ângulo $\theta = 25^\circ$ para iluminação. Nas figuras 3.4(a-c), os componentes X, Y e Z da normal calculada são exibidos como níveis de cinza. A figura 3.4(d) mostra o mapa de

albedo \bar{s} calculado.

A figura 3.5(a) mostra o mapa de erro de normais, isto é, a diferença absoluta entre a normal calculada pelo programa e a normal real (conhecida a partir do modelo do POV-Ray). Estes erros ocorrem devido a vários fatores: mistura de cores de pixels que contém arestas e siluetas dos objetos; amostragem inadequada do gabarito em regiões onde a superfície é quase vertical; e ruído de quantização nas imagens de entrada. Nota-se que os erros de quantização são significativos apenas nas regiões onde o albedo é muito baixo.

A figura 3.5(b) mostra a visualização 3D do mapa de alturas $Z[p]$ obtido pela integração do mapa de normais calculado (a). Este mapa de alturas foi transformado em uma malha triangular com 4 triângulos para cada pixel da imagem de entrada. A malha foi renderizada com o POV-Ray usando iluminação arbitrária. Como era esperado, a maior parte dos erros de integração ocorre ao longo de descontinuidades no mapa de alturas, onde o mapa de normais não oferece informação suficiente a respeito da altura de regiões adjacentes.

A figura 3.6 ilustra o formato do conjunto de assinaturas \mathbb{G} para cada caso. A base quadrada ao fundo representa a região do plano de projeção coberta pela grade. A altura de cada palito corresponde à distância entre o centróide $\mu[i, j]$ das entradas de cada bucket ao centro $c[i, j]$ da célula correspondente no plano da grade. Observe que os deslocamentos $\mu[i, j] - c[i, j]$ não necessariamente precisam ser paralelos em \mathbb{R}^m . O raio da esfera no topo de cada palito é proporcional ao raio $\rho[i, j]$ do bucket. Usamos a mesma escala para todas estas dimensões. Esta visualização confirma a tese de que o conjunto \mathbb{G} é bastante curvo, porém pouco “espesso“. Portanto a grade bidimensional proporciona um hashing eficiente na busca da assinatura mais próxima.

A tabela 3.1 mostra diversas medidas de custo para cada operação de busca em tabela para diversas configurações de teste. Todos os valores são médias calculadas sobre todos os pixels de \mathcal{S} . As colunas são: o tempo t de cada busca em microssegundos, o número b de buckets $B[i', j']$ que foram examinados e o número d de entradas da tabela que foram realmente testadas (isto é, o número de vezes que $\text{dist}(\mathbf{s}, \mathbf{g})$ foi calculada) Os testes foram feitos em um PC padrão com relógio de 2.2Ghz; o tempo absoluto t obviamente depende da plataforma e implementação.

A figura 3.7 mostra o número de entradas de cada bucket $B[i, j]$ observado em dois dos testes (para $m = 10, \theta = 25^\circ$, e $m = 24, \theta = 45^\circ$, respectivamente). Em ambos os casos o tamanho da grade foi $N \times N = 211 \times 211$, correspondendo a uma média de $\kappa = 25\%$ entradas por bucket. A altura de cada palito é proporcional ao número de entradas no balde; cada faixa de cor nos palitos representa uma entrada na lista. Observe que as assinaturas de observação estão espalhadas sobre uma porção substancial da grade, de maneira razoavelmente uniforme.

θ	m	N	κ	t	d	b
45	3	1	—	2160.8	11172.0	1.00
45	3	105	1.01	5.8	22.1	9.97
45	3	149	0.50	4.9	17.0	9.98
45	3	211	0.25	4.2	13.3	10.73
45	3	299	0.12	4.7	11.0	12.70
45	3	422	0.06	4.8	10.4	15.31
45	10	1	—	3560.8	11172.0	1.00
45	10	105	1.01	7.6	13.9	12.48
45	10	149	0.50	6.9	9.9	14.76
45	10	211	0.25	7.1	7.1	19.78
45	10	299	0.12	7.9	4.8	29.79
45	10	422	0.06	10.6	4.4	48.11
45	24	1	—	6068.0	11172.0	1.00
45	24	105	1.01	13.3	14.1	12.50
45	24	149	0.50	11.7	9.3	14.79
45	24	211	0.25	10.7	5.4	19.73
45	24	299	0.12	12.0	4.9	29.86
45	24	422	0.06	14.7	4.5	48.12
25	3	1	—	2182.9	11172.0	1.00
25	3	105	1.01	10.4	50.3	10.00
25	3	149	0.50	7.5	33.9	10.01
25	3	211	0.25	4.9	17.7	10.04
25	3	299	0.12	4.1	12.3	10.19
25	3	422	0.06	3.8	9.0	10.84
25	10	1	—	3547.9	11172.0	1.00
25	10	105	1.01	15.2	47.0	10.08
25	10	149	0.50	10.4	27.9	10.17
25	10	211	0.25	8.5	19.2	10.36
25	10	299	0.12	7.1	13.3	10.80
25	10	422	0.06	6.2	8.9	11.96
25	24	1	—	6059.5	11172.0	1.00
25	24	105	1.01	27.5	51.1	10.09
25	24	149	0.50	17.1	24.6	10.19
25	24	211	0.25	14.1	17.5	10.40
25	24	299	0.12	12.4	13.3	10.87
25	24	422	0.06	10.5	7.8	12.15

Tabela 3.1: Média de custos e operações do procedimento de busca na tabela para imagens sintéticas com $\theta = 25^\circ$ e $\theta = 45^\circ$ e diversos valores de m e N . As entradas para $N = 1$ correspondem á uma busca de força bruta na tabela.

3.6.2 Testes com fotos de uma cena real

Para testes com dados reais, foi utilizada uma cena consistindo de uma escultura de gesso pintada com tinta guache. Veja a figura 3.8. Para gabaritos usamos quatro bolinhas de *mouse* de borracha cinza, sendo que duas delas foram pintadas com tinta *spray* branca fosca. Os gabaritos foram colocados dentro de cones rasos feitos de papel preto, para reduzir a interferência de luz refletida de outros objetos da cena. A escultura e os gabaritos foram posicionados sobre uma folha de papel branco, contendo uma grade impressa a *laser* com células de 1×1 cm.

Foram obtidas 24 imagens digitais desta cena com uma câmera de 3 Mega-pixels (Sony CyberShot DSC-V1). A câmera foi montada em um tripé de fotógrafo e configurada para captura de imagem por timer para reduzir vibrações. Veja a figura 2.1. A iluminação foi fornecida por uma lâmpada halogênica cilíndrica de 500W com um refletor de $8\text{cm} \times 16\text{cm}$ posicionado a aproximadamente 1 metro da cena. A luminária foi posicionada manualmente em cada foto de maneira a variar o ângulo θ com a normal variasse 10° e 25° e o azimutes variando entre 0° e 270° .

Essas fotos foram alinhadas por translação com precisão de ± 0.35 pixels, cortadas de 2048×1536 para 672×1000 pixels, reduzida para 168×250 para remover a maior

parte do ruído da compressão JPEG e convertida para escala de cinza, como nas imagens sintéticas.

Durante o processamento das imagens, constatamos que os gabaritos eram perceptivelmente irregulares e marcados por pequenos defeitos, como mostrado na figura 3.9. Para resolver este problema combinamos as imagens $G_i^{(1)}, G_i^{(2)}, G_i^{(3)}, G_i^{(4)}$ dos quatro gabaritos de cada foto i de modo a obter uma imagem G_i de um "gabarito médio". Mais especificamente, as imagens dos quatro gabaritos foram redimensionadas, alinhadas e tiveram seu brilho e contraste ajustados de modo a obterem mesma média e variância na região central. Feito isto, para cada pixel q do gabarito e iluminação i , ordenamos as intensidades $G_i^{(1)}[q], G_i^{(2)}[q], \dots, G_i^{(4)}[q]$, e definimos o valor $G_i[q]$ como a média ponderada dos valores ordenados com pesos 1, 5, 5, 1. O resultado é mostrado na figura 3.10.

A figura 3.11 mostra o mapa de normais da cena e a figura 3.12 o mapa de albedo calculados pelo nosso método. Para obter o mapa de albedo, executamos o programa separadamente para cada canal de cor (R,G e B). A figura 3.13 é uma reconstrução tridimensional do mapa de alturas obtido pela integração do mapa de normais.

A tabela 3.2 mostra as medidas de performance de várias configurações de teste com as versões em escala de cinza das imagens de entrada. As colunas são as mesmas da tabela 3.1.

θ	m	N	κ	t	d	b
10-25	3	1	—	2183.6	11172.0	1.00
10-25	3	105	1.01	7.6	30.5	10.15
10-25	3	149	0.50	5.7	18.8	10.35
10-25	3	211	0.25	4.3	11.9	10.70
10-25	3	299	0.12	3.8	7.4	11.74
10-25	3	422	0.06	3.8	4.6	14.17
10-25	10	1	—	3563.1	11172.0	1.00
10-25	10	105	1.01	15.2	46.6	13.24
10-25	10	149	0.50	12.9	33.0	18.07
10-25	10	211	0.25	12.1	27.1	27.71
10-25	10	299	0.12	13.3	22.9	43.31
10-25	10	422	0.06	16.7	19.6	70.79
10-25	24	1	—	6111.4	11172.0	1.00
10-25	24	105	1.01	29.0	50.3	20.64
10-25	24	149	0.50	27.6	43.6	32.76
10-25	24	211	0.25	27.6	38.1	50.12
10-25	24	299	0.12	30.5	33.8	84.66
10-25	24	422	0.06	28.8	29.4	146.68

Tabela 3.2: Custo médio e contagem das operações efetuadas no procedimento de busca para vários valores de θ , m e N . As entradas com $N = 1$ correspondem a uma busca seqüencial, sem a aceleração oferecida pela grade de baldes.

A figura 3.14 mostra o comprimento da lista de baldes em $B[i, j]$ para os testes com $m = 10$ e $m = 24$, respectivamente, com as mesmas convenções da figura 3.7. Em ambos os casos o tamanho da grade era $N \times N = 211 \times 211$, correspondendo a uma ocupação média de $\kappa = 25\%$ entradas por bucket. A figura 3.15 mostra o formato do conjunto \mathbb{G} , com as mesmas convenções da figura 3.6.

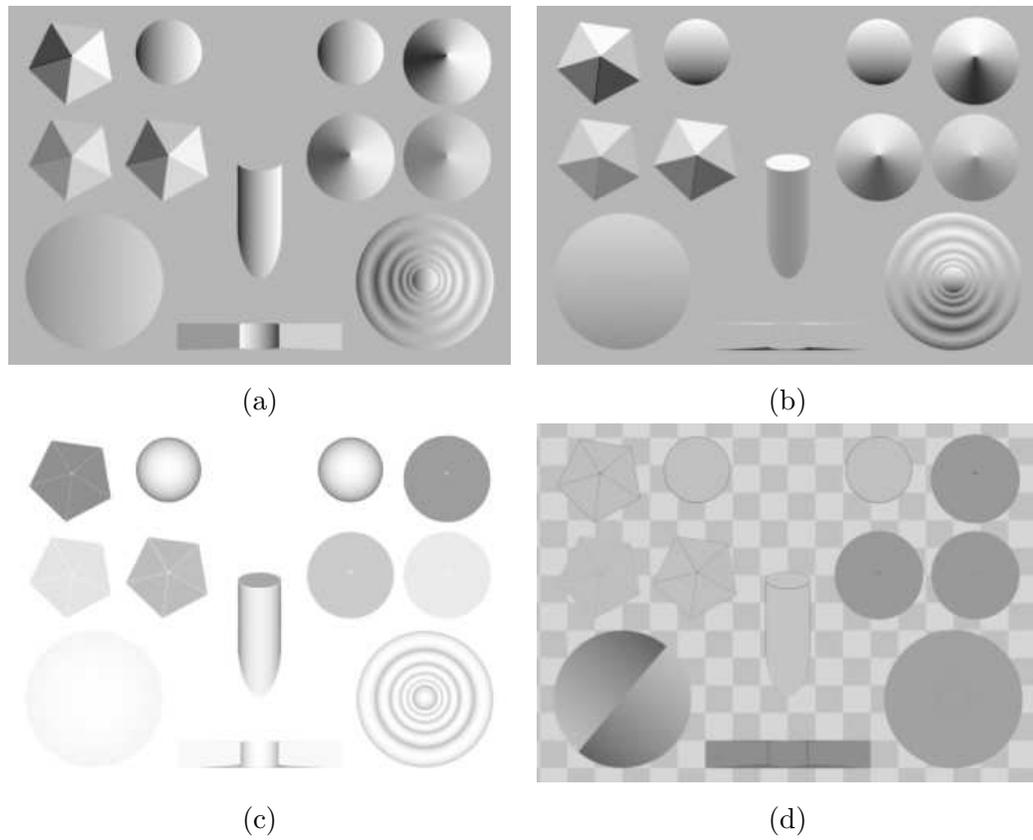


Figura 3.4: Resultados para o teste com imagens: (a–c) os componentes X,Y e Z do mapa de normais; (d) o mapa de albedo

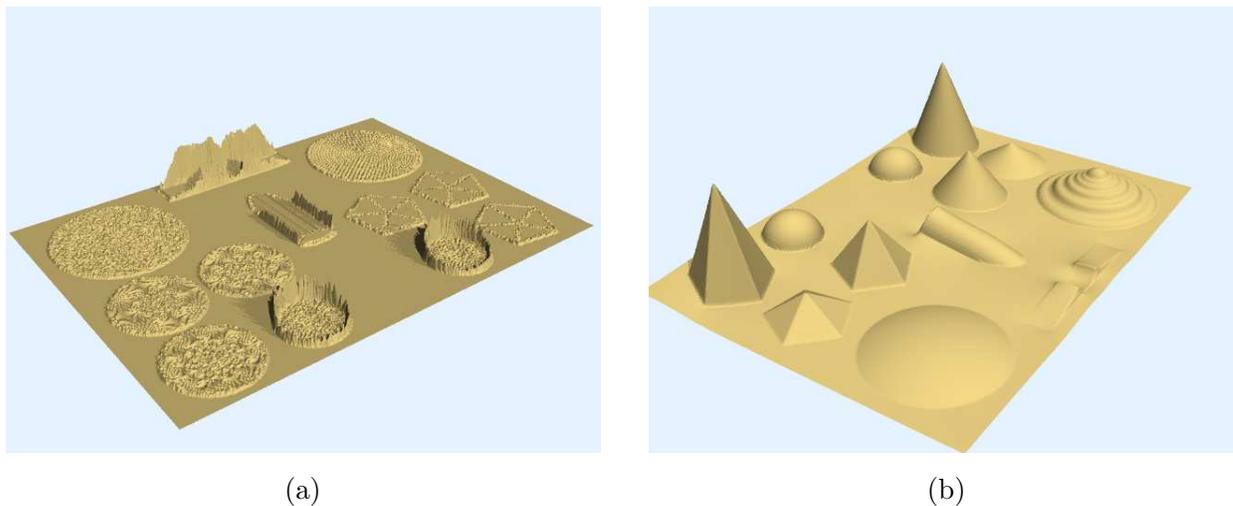


Figura 3.5: (a) Visualização dos erros das normais calculadas. O maior erro é 0.21 radianos ($\approx 12.5^\circ$) e a média do quadrado do erro é 0.012 ($\approx 0.66^\circ$). O erro inferior a 0.0001 ($\approx 0.005^\circ$) em toda superfície de base quadriculada. (b) Visualização 3D do mapa de alturas obtido pela integração do mapa de normais.

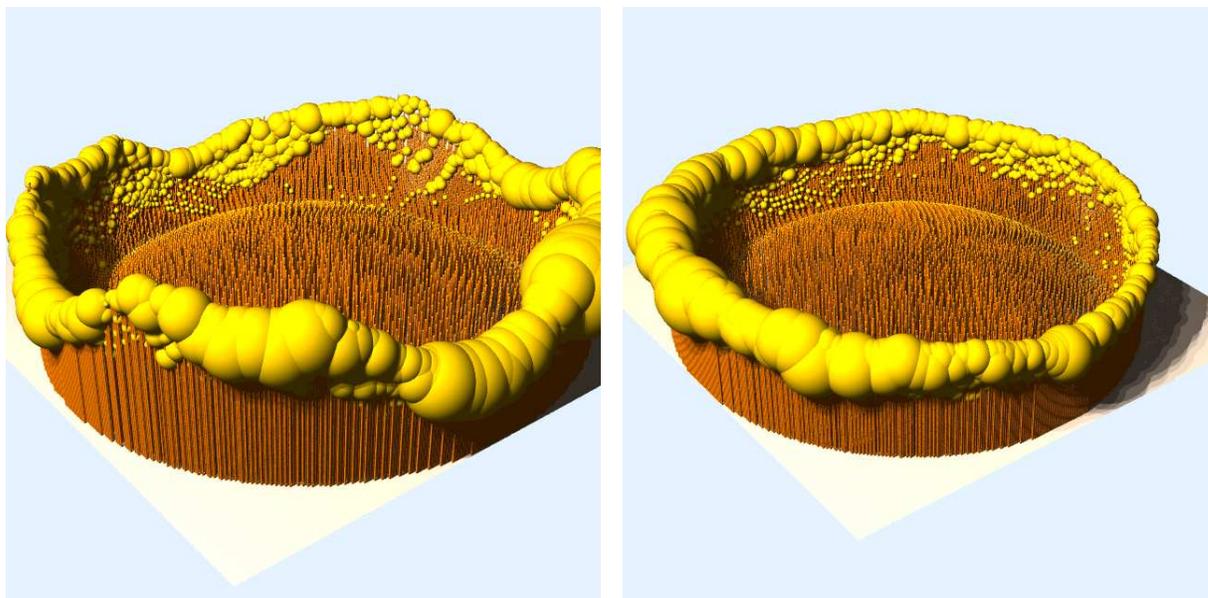


Figura 3.6: Formato do conjunto de assinaturas \mathbb{G} para conjunto de imagens sintéticas. O comprimento de cada palito é a distância de $\mu[i, j]$ do plano da grade; o raio da esfera é o valor de $\rho[i, j]$.

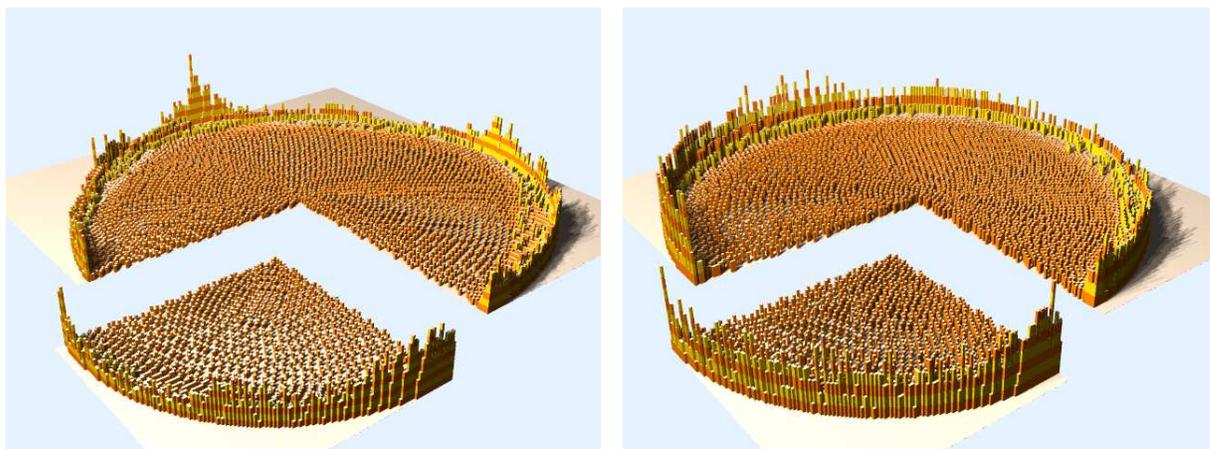


Figura 3.7: Número de entradas de cada balde da grade para teste com imagens sintéticas, com $m = 10$ imagens de entrada (esquerda) e $m = 24$ (direita). A mais longa lista possui 9 entradas.

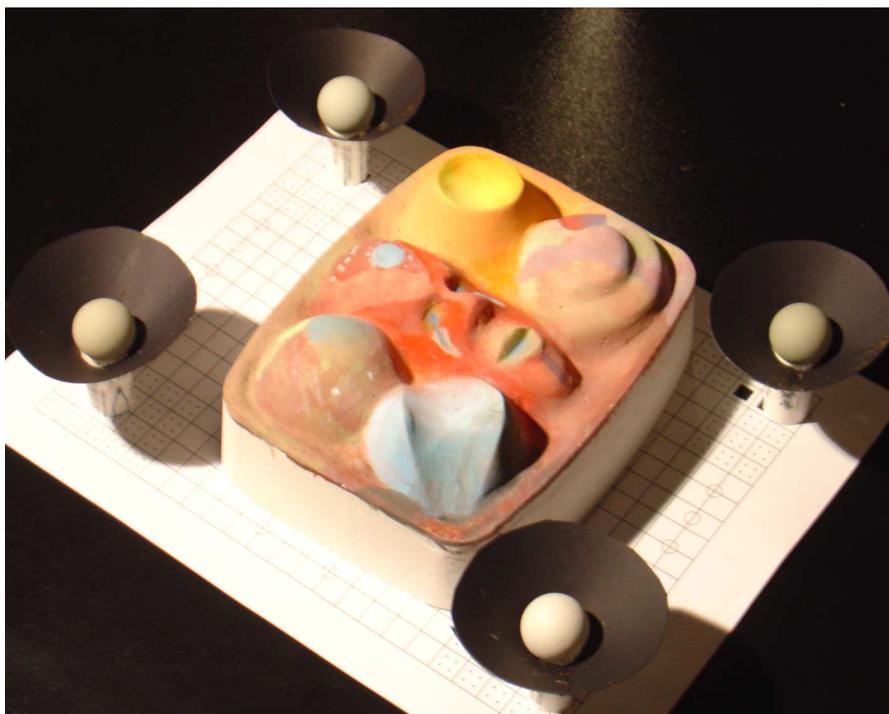


Figura 3.8: Foto em perspectiva do objecto utilizado nos testes com cena real. Observe os quatro gabaritos de iluminação posicionados nos cones de proteção.

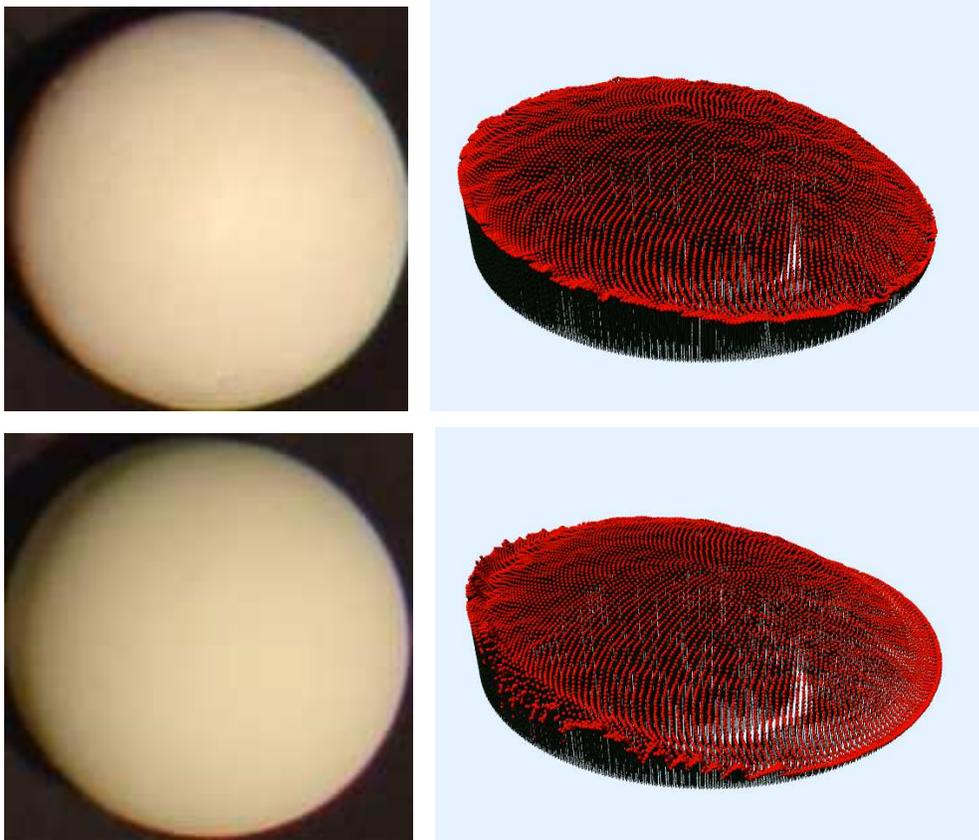


Figura 3.9: Imagens de dois gabaritos da cena real (esquerda) sob as mesmas condições de iluminação da figura 2.1 e a visualização 3D das intensidades luminosas dos pontos amostrados na tabela de busca (direita).

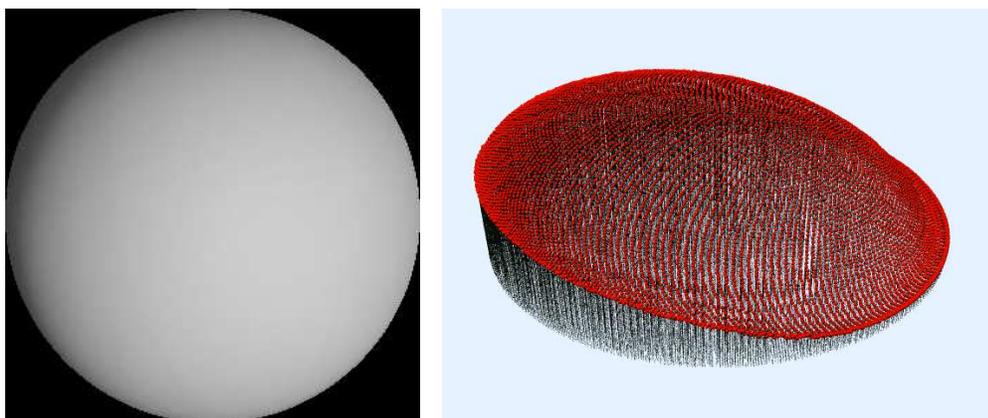


Figura 3.10: Gabarito composto pela média dos quatro gabaritos da cena real (esquerda) e visualização 3D dos seus pontos amostrados (direita).

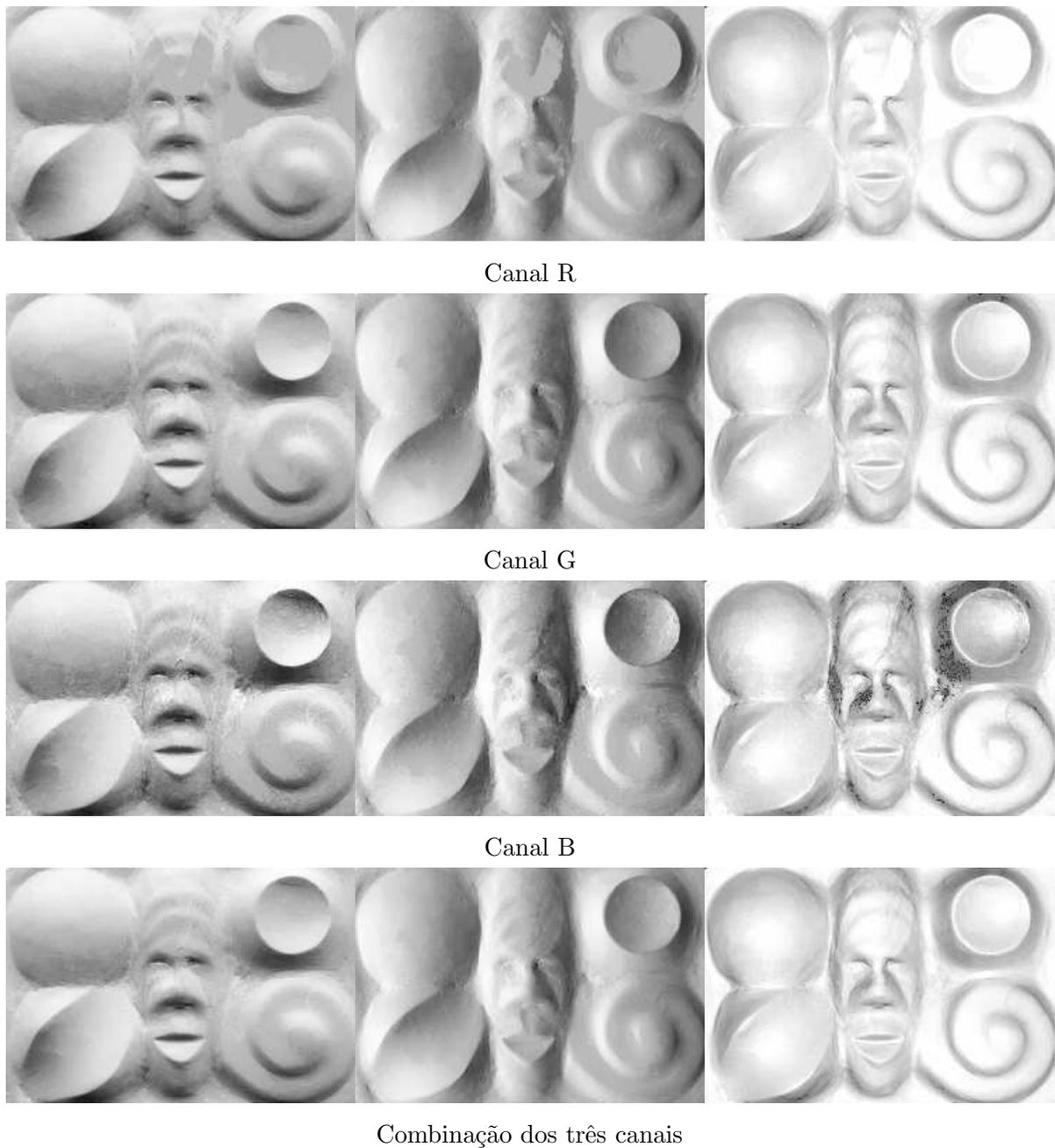


Figura 3.11: Resultados do teste com cena real: componentes X, Y, e Z do mapa de normais para cada um dos canais e para a combinação dos três canais



Figura 3.12: O mapa de cor intrínseca para o teste com cena real; composto dos mapas de albedo dos 3 canais.

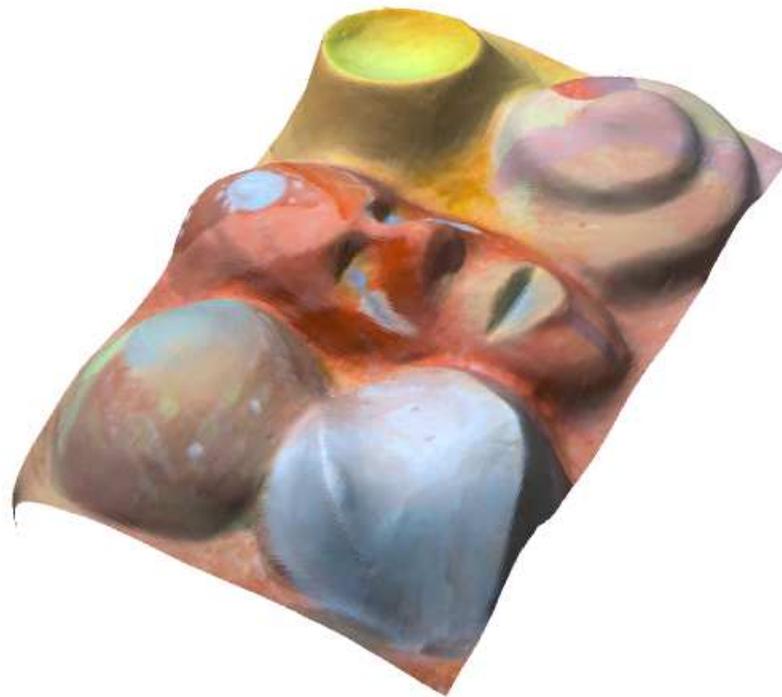


Figura 3.13: Visualização 3D do mapa de alturas obtido pela integração do mapa de normais (figura 3.11) colorido pelo mapa de cor intrínseca 3.12.

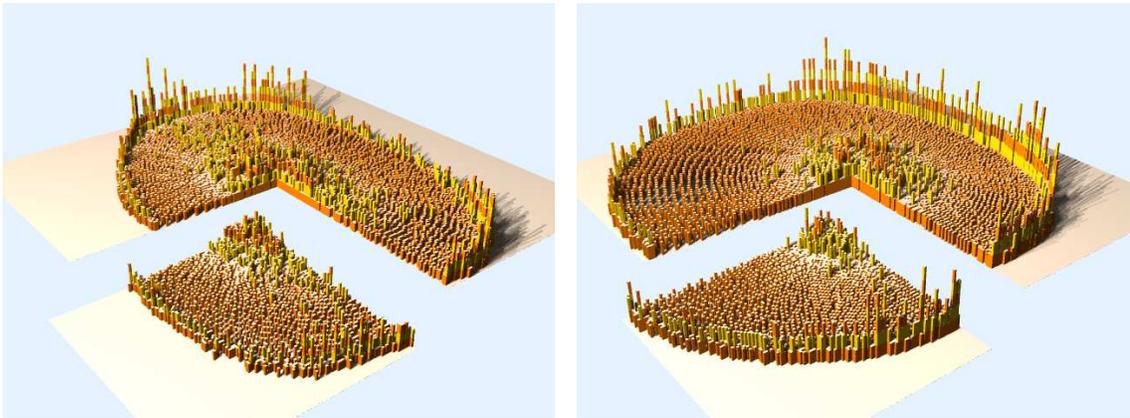


Figura 3.14: Número de entradas em cada balde para um teste com cena real e parâmetros $m = 10$ (esquerda), $m = 24$ (direita), $N = 211$, $\theta = 10^\circ\text{--}25^\circ$ com as mesmas convenções da figura 3.7.

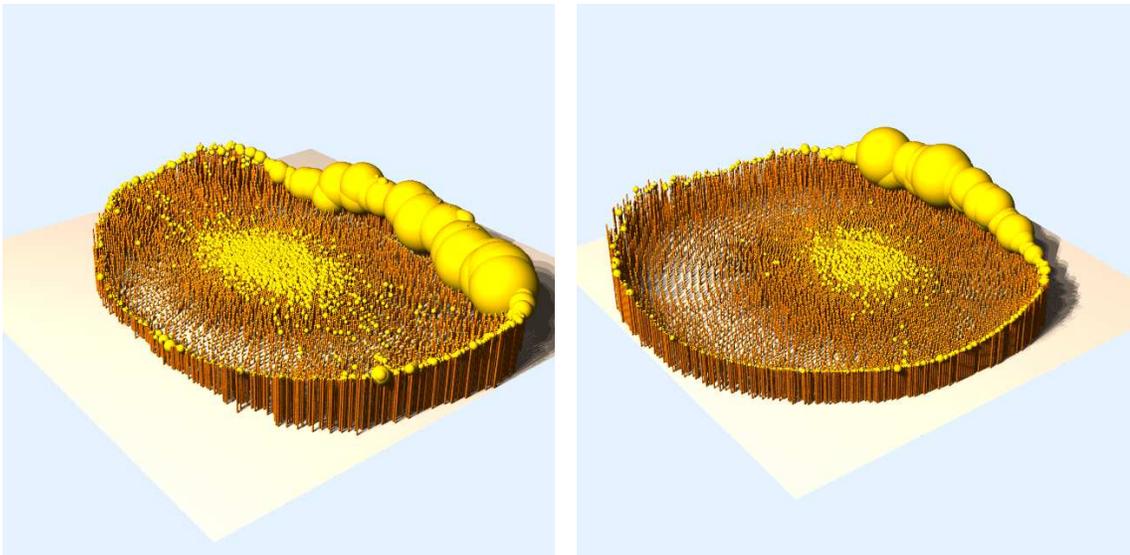


Figura 3.15: Formato do conjunto \mathbb{G} para teste com cena real, com as mesmas convenções da figura 3.6

Capítulo 4

Sombras e Highlights

É muito difícil obter iluminação uniforme, mesmo em cenas simples e com posicionamento cuidadoso dos objetos e fontes de iluminação. É difícil também garantir que o gabarito tenha FBDR exatamente proporcional à da cena. *Anomalias de iluminação* ou *anomalias de acabamento* provocam alterações no valor de intensidade aparente $S_i[p]$, deixando-a maior ou menor que o previsto pela equação (2.3). Quando isto ocorre, os vetores de observação $\mathbf{S}[p]$ e $\mathbf{G}[q]$ de dois pontos com a mesma normais não serão colineares e o método descrito na seção 2.1 não funciona. Neste capítulo vamos analisar a natureza e as consequências dessas anomalias e descrever nossa solução para esse problema.

4.1 Anomalias de iluminação e acabamento

4.1.1 Iluminação externa não uniforme

Nas situações que consideramos neste trabalho, o fluxo de luz $\Phi[p](\vec{n})$ em um ponto p da superfície da cena e com uma direção \vec{u} poder ser decomposto em um *fluxo primário* $\Phi^*[p](\vec{u})$, devido fontes externas à cena, e um fluxo *fluxo secundário* devido a luz espalhada pela própria cena.

Uma condição básica para estéreio fotométrico é que o fluxo $\Phi^*[p](\vec{u})$ seja independente de p , ou que sua dependência seja conhecida. Isso exige que as fontes de luz estejam situadas a uma distância D grande comparada com o tamanho W da parte visível da cena. Por exemplo, se a imagem digital tem b bits por amostra, é necessário que D seja maior que $2^{b+2}W$, para que a diferença da intensidade de dois pixels situados em extremos opostos da imagem com a mesma normal e albedo, seja desprezível.

Outro tipo de não-uniformidade é a *iluminação secundária* que incide em uma parte da cena após ter sido dispersa por outra parte da mesma. Para reduzir os efeitos de iluminação secundária, a superfície deve ser tal que a luz dispersa deixe a cena de forma

a não atingir alguma outra parte da superfície. Esta exigência praticamente restringe a análise a cenas consistindo de um único objeto convexo ou com concavidades pequenas e/ou rasas. Vale notar, entretanto, que iluminação indireta é proporcional ao albedo da cena, de modo que o impacto dessa iluminação na imagem, em termos relativos, é também proporcional ao albedo. Ou seja, para cenas com albedo muito baixos, o efeito da luz indireta é desprezível

4.1.2 Sombras projetadas

Um tipo importante de não-uniformidade são as *sombras projetadas*, que ocorrem quando uma parte da cena bloqueia, total ou parcialmente, a luz proveniente de uma direção \vec{u} que de outra forma incidiria em outra parte da cena.

Para evitar sombras projetadas, o campo de iluminação primário $\Phi_i^*(\vec{u})$ — a componente do fluxo de luz $\Phi_i(\vec{u})$ devida a fontes luminosas externas à cena — deve ser zero se a direção \vec{u} pode produzir sombras projetadas, isto é, se existir uma linha com direção \vec{u} que atinge um ponto visível da cena depois de atingir algum outro objeto. Esta restrição limita severamente a geometria da cena e o posicionamento das fontes de luz.

A rigor, a condição de iluminação uniforme também exclui *sombras próprias*, regiões onde a normal da superfície do objeto faz um ângulo obtuso com a direção do fluxo luminoso externo, de maneira que a luz é bloqueada pelo próprio objeto. Porém acontece que, o valor de $\Phi_i[p](\vec{u})$ é irrelevante para direções \vec{u} que apontam para fora da superfície, pois o fluxo luminoso nestas direções é excluído de qualquer forma na integral (2.4) pelo fator $\max\{0, \vec{u} \cdot \vec{n}\}$ incluído na BRDF. Portanto, em qualquer ponto p situado em uma sombra própria, pode-se supor que $\Phi_i[p]$ é o campo luminoso Φ_i que seria observado em p se o objeto em si não existisse. Ou seja, as sombras próprias (mas não as projetadas) são compatíveis com a hipótese de iluminação uniforme.

4.1.3 Anomalias de acabamento

Além de anomalias de iluminação, é difícil evitar anomalias de acabamento, pontos da superfície da cena cuja FBDR difere do gabarito de iluminação. Dentre estes fenômenos, os mais comuns são *highlights*, manchas claras causadas por reflexos de fontes de luz em superfícies polidas ou cobertas com cera, verniz ou esmalte.

4.2 Pontos canônicos

De qualquer maneira, é necessário que as imagens do gabarito não apresentem anomalias de iluminação. Isto é, o fluxo $\Phi_i[q]$ em cada ponto q do gabarito deve ser o mesmo fluxo

fixo Φ_i^* . Por este motivo, os gabaritos devem ser convexos.

Dizemos que o fluxo luminoso $\Phi_i[p]$ em um ponto p da cena é *canônico* se for igual ao fluxo do gabarito Φ_i^* no ponto q de G_i com mesma normal de p , exceto possivelmente pelos efeitos de sombras próprias e erros de medida e quantização. Também dizemos que um pixel p de uma imagem S_i é *canônico* se o fluxo $\Phi_i[p]$ for canônico, e a FBDR $\sigma[p](\vec{n}, \vec{u}, \vec{v})$ da superfície for proporcional à FBDR $\gamma[p](\vec{n}, \vec{u}, \vec{v})$ do gabarito, exceto para direções \vec{u} onde $\Phi[p](\vec{u})$ é desprezível. Por exemplo, se a cena é lustrosa, p pode ser canônico desde que o reflexo (*highlights*) não caia sobre p .

4.2.1 A hipótese de dados suficientes

O estéreo fotométrico baseado em gabaritos ainda é possível na presença de anomalias de iluminação ou de acabamento, se, para cada ponto p , existe um subconjunto $K[p]$ dos índices $\{1, \dots, m\}$ com pelo menos 3 elementos tal que p é canônico em S_i para todo $i \in K[p]$. Observe que este subconjunto pode ser diferente para cada ponto p de S . Esta condição será denominada como *hipótese de dados suficientes*.

4.2.2 Fator de anomalia

Para modelar as anomalias em geral, introduziremos um *fator de anomalia* $\kappa_i[p]$, definido pela fórmula

$$\kappa_i[p] = \frac{S_i[p]/\bar{s}[p]}{G_i[q]/\bar{g}[q]} \quad (4.1)$$

Onde q é o ponto do gabarito com mesma normal $\vec{g}[q] = \vec{s}[p]$. Podemos então substituir as equações (2.3) por

$$\begin{aligned} S_i[p] &= \bar{s}[p] L_i^*(\vec{s}[p]) \kappa_i[p] \\ G_i[q] &= \bar{g}[q] L_i^*(\vec{g}[q]) \end{aligned} \quad (4.2)$$

onde L_i^* é a função de *shading* do gabarito, que é supostamente independente da posição do ponto q .

Note que o fator de anomalia $\kappa_i[p]$ é 1 se o ponto p é canônico em S_i . Em imagens S_i onde p está localizado em uma região de sombra projetada, o fator $\kappa_i[p]$ será menor que 1. Reciprocamente, em imagens S_i onde o fluxo luminoso em p foi aumentado devido a *highlights* ou iluminação secundária, o valor de $\kappa_i[p]$ será maior que 1. A hipótese de dados suficientes pode então ser reformulada como: para cada ponto p em \mathcal{S} , existem no mínimo três índices i tais que $\kappa_i[p] = 1$.

4.3 Estéreo fotométrico na presença de anomalias

4.3.1 Enumeração exaustiva

O teste de colinearidade da equação (2.3) continua válido se de alguma forma for possível identificar o subconjunto $K[p]$ de imagens “boas” para cada ponto p . Em teoria podemos aplicar métodos básicos do capítulo 2 para todos os subconjuntos K com 3 ou mais imagens, e selecionar a resposta que fornece o melhor casamento entre os vetores de observação do gabarito e da cena. Porém, o número de combinações $\binom{m}{3}$ é aproximadamente $m^3/6$, o que torna este método extremamente lento quando se tem-se mais de uma dezena de iluminações distintas.

4.3.2 Descarte prévio de anomalias

Na literatura encontramos várias propostas de heurísticas para detectar se um pixel p é afetado por sombras e/ou highlights em uma imagem S_i . Entretanto estas heurísticas são pouco confiáveis ou exigem restrições que tornam a aplicação do método pouco prática.

Por exemplo, Barski e Petrou [4] sugerem excluir imagens S_i onde $S_i[p]$ está abaixo de um determinado limiar (quer absoluto, quer relativo a outros valores de $S_j[p]$). Esta abordagem foi adotada nos trabalhos de Funk [11] e Clark [8]. Entretanto, este critério pode levar à exclusão de imagens com sombras próprias. Lembramos que sombras próprias não invalidam o algoritmo de comparação de assinaturas. Na verdade, elas fornecem informação útil para identificação da normal; e a inclusão de imagens com sombra própria no conjunto K pode ser necessária para satisfazer a hipótese de dados suficientes.

Além disso, fontes de luz reais normalmente produzem penumbras em vez de sombras com limites bem definidos. Portanto um pixel afetado por sombra projetada pode ser substancialmente mais claro que um pixel canônico de albedo baixo. Ou seja, “estar na sombra projetada” não é um predicado binário (verdadeiro ou falso), como suposto por Chandraker, Agarwal e Kriegman [7]; e as regiões de sombras projetadas não podem ser adequadamente modeladas por subconjuntos de pixels. A presença de iluminação secundária aumenta as dificuldades desta abordagem.

Estritamente falando, é *impossível* determinar com total confiança se um fluxo Φ_i é canônico em um ponto p analisando-se apenas as imagens da cena S_1, \dots, S_m . Por outro lado qualquer análise que utilize também as imagens de gabarito G_1, \dots, G_m para esta decisão deve necessariamente identificar o ponto q do gabarito com a mesma normal de p . Ou seja, a identificação do conjunto κ não pode ser uma etapa de pre-processamento.

4.4 Nossa abordagem: uma nova métrica

Dadas as limitações práticas e teóricas do uso de heurísticas para escolha do conjunto κ , optamos por uma abordagem “holística” diferente. Na busca do ponto q desejado do gabarito, utilizamos todos os valores das imagens $S_1[p], \dots, S_m[p]$, sem tentar excluir as imagens onde p não é canônico; porém ao compararmos os vetores de observação $\mathbf{S}[p]$ e $\mathbf{G}[q]$, usamos uma métrica $\Delta(\mathbf{S}, \mathbf{G})$ que não é afetada por anomalias ocasionais na iluminação.

Para melhor entender este conceito, considere as figuras 4.1, 4.2 e 4.3. Estas figuras referem-se a uma instância do problema de estéreo fotométrico com $m = 9$ diferentes iluminações. Cada gráfico (à esquerda) refere-se a um par de pixels $p \in \mathcal{S}$ e $q \in \mathcal{G}$, identificados por cruzes vermelhas nas imagens S e G ao lado. Cada marca ‘•’ no gráfico representa um par de intensidades $a_i = (\mathbf{S}_i, \mathbf{G}_i) = (S_i[p], G_i[q])$ para $i \in \{1, \dots, m\}$.

A figura 4.1 mostra uma situação ideal, onde p e q possuem a mesma normal e o pixel p é canônico em todas as imagens. Observe que, os m pontos a_i formam uma linha reta com a origem, exceto por pequenos desvios devidos à quantização e ruído. A inclinação dessa reta é o albedo relativo $\bar{s}[p]/\bar{g}[q]$ do ponto p .

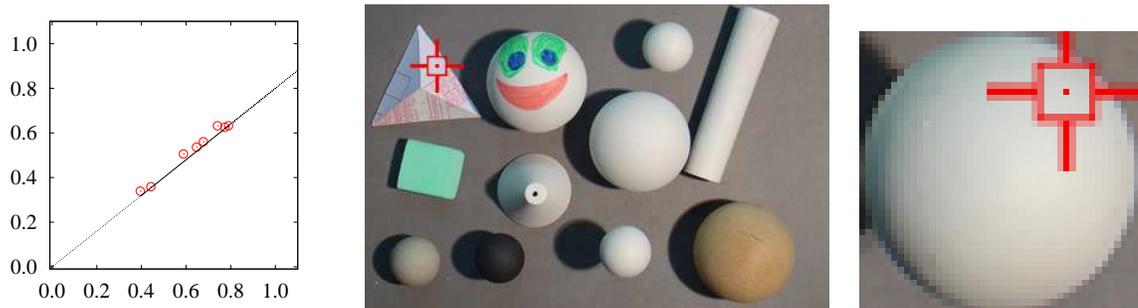


Figura 4.1: Pares $(\mathbf{S}_i[p], \mathbf{G}_i[q])$ para uma situação ideal, com p canônico e $\bar{s}[p] = \bar{g}[q]$

A figura 4.2 mostra a situação onde p e q possuem a mesma normal, mas p é canônico em apenas quatro das nove imagens. Os quatro pares a_i correspondentes a estas imagens continuam formando uma linha reta com a origem, enquanto que os demais caem abaixo da linha (devido a sombras projetadas) ou acima dela (devido a *highlights* ou iluminação secundária).

Finalmente, a figura 4.3 é típica de um caso onde p e q possuem normais diferentes. Note que, neste caso, não há um conjunto significativo de pares a_i que sejam colineares com a origem.

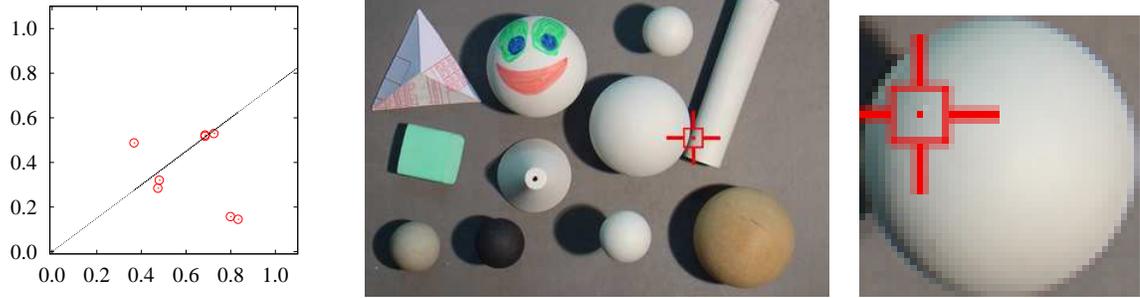


Figura 4.2: Pares $(\mathcal{S}_i, \mathcal{G}_i)$ em imagens com presença de sombras.

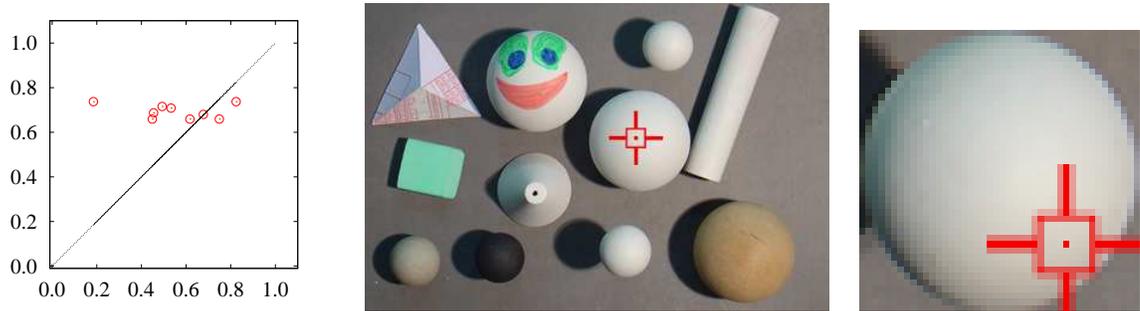


Figura 4.3: Pares $(\mathcal{S}_i, \mathcal{G}_i)$ para pontos p e q com normal distinta

4.4.1 A métrica Δ

Nossa solução para o problema das anomalias é definir uma função $\Delta(\mathcal{S}, \mathcal{G})$ é que tenha valores baixos nas situações das figuras 4.1 e 4.2, e valores altos na situação da figura 4.3. Mais especificamente, a função deve encontrar um conjunto $K[p, q] \subseteq \{1, .. n\}$ de índices tais que os pares a_i correspondentes estejam suficientemente próximos a uma linha reta partindo da origem. Se tal conjunto existir, a função Δ deve retornar um valor pequeno (muito menor que 1), cuja magnitude depende de quantos pares existem neste conjunto e o quão próximos eles estão da linha. Caso contrário a função deve retornar um valor elevado arbitrário.

O significado preciso de “suficientemente grande” e “suficientemente próximo” no parágrafo acima obviamente depende do número de imagens m , das intensidades de \mathcal{G}_i , e da precisão dos dados das imagens. Contudo, observamos que variando o ponto q do gabarito podemos fazer com que as duas abscissas $G_i[q]$ e $G_j[q]$ de dois pares distintos a_i, a_j variem independentemente uma da outra. Portanto espera-se que exista uma variedade unidimensional de pontos $q \in \mathcal{G}$ para os quais os pares a_i e a_j são alinhados com a origem. Pode-se concluir então que o alinhamento de apenas dois pares a_i, a_j com a origem não é

significativo.

Por outro lado, a ocorrência de três pares a_i, a_j, a_k alinhados com a origem é uma situação que pode ocorrer apenas por acidente, em pontos com sombras ou *highlights*, ou porque duas normais são similares. Como a primeira hipótese é relativamente improvável, esta coincidência pode ser considerada com uma evidência que p e q possuem a mesma normal. A evidência fica tanto mais forte quanto maior for o número de pares a_i alinhados.

Nesta abordagem a normal $\vec{s}[p]$ para um ponto p da cena é determinada calculando-se a distância $\Delta(S[p], G[q])$ para todos os pontos q do gabarito, e escolhendo-se o ponto q de menor Δ . Em princípio o custo é proporcional a $T\delta(m)$ onde T é a tabela de assinaturas do gabarito e $\delta(m)$ é o custo de calcular $\Delta(S, G)$ para um par de vetores.

Observe que esta abordagem difere das abordagens baseadas em descarte prévio, discutidas na seção 4.3. Em vez de tentar escolher um subconjunto K *a priori* examinando apenas o vetor $S[p]$ da cena, nossa métrica de distância escolhe um subconjunto K que depende tanto de $S[p]$ quanto do vetor $G[q]$ que está sendo considerado.

4.4.2 A métrica Δ_2

Para esclarecer este conceito, vamos descrever a seguir uma das métricas que desenvolvemos ao decorrer do projeto e que denominamos de métrica Δ_2 .

Se o albedo relativo $\beta = \vec{s}[p]/\vec{g}[q]$ fosse conhecido, poderíamos calcular a semelhança entre dois vetores $\mathbf{S} = \mathbf{S}[p]$ e $\mathbf{G} = \mathbf{G}[q]$ traçando o gráfico da fórmula $\mathbf{S} = \beta\mathbf{G}$ e contando quantos pares α_i caem próximos a mesma. Esta contagem pode ser feita pela fórmula

$$W(\beta) = \sum_{j=1}^m \mathbf{G}_j \exp \left[-\frac{1}{2}(\mathbf{S}_j - \beta\mathbf{G}_j)^2 \right] \quad (4.3)$$

Onde σ é o desvio padrão do ruído presente nas intensidades S_i, G_i . Chamamos a função $W(\beta)$ de *pontuação do albedo* β . Em particular, se k pares $(\mathbf{S}_i, \mathbf{G}_i)$ com $\mathbf{G}_i \simeq 1$ caem precisamente sobre a reta, e os demais $m - k$ pares ficam bem distantes da mesma, a pontuação $W(\beta)$ será aproximadamente k . O fator \mathbf{G}_j na fórmula 4.3 compensa o fato de que pares com \mathbf{G}_j baixo tendem a cair mais próximos da reta por mera coincidência.

Uma vez que o valor do albedo relativo β do ponto p não é conhecido de antemão, o que podemos fazer é determinar o valor β_{\max} que tem maior pontuação $W(\beta_{\max})$, e supor que W_{\max} é uma indicação da semelhança entre os vetores $\mathbf{S}[p]$ e $\mathbf{G}[q]$, e portanto da probabilidade de que q tenha a mesma normal que o ponto p .

Uma dificuldade é o fato que $W(\beta)$ depende do parâmetro β de maneira não-linear, e pode ter vários máximos locais. A heurística que usamos na função Δ_2 é calcular $W(\beta)$ apenas para os albedos relativos $\beta_i = \mathbf{S}_i[p]/\mathbf{G}_i[q]$ (ou seja, apenas para retas que passam por alguns dos pontos a_i) e tomar o máximo dentre os valores $W(\beta_i)$. Embora o valor de

β com pontuação máxima não seja necessariamente um desses β_i , espera-se que seja algo bem próximo de algum deles.

Uma vez que as intensidades $\mathbf{S}_i, \mathbf{G}_i$ geralmente são afetadas por erros de medida e quantização, a razão $\mathbf{S}_i[p]/\mathbf{G}_i[q]$ pode ter erros muito grandes quando $\mathbf{G}_i[q]$ é baixo. Portanto utilizamos um estimador mais seguro $\beta_i = \sqrt{\mathbf{S}_i^2 + \sigma^2}/2/\sqrt{\mathbf{G}_i^2 + \sigma^2}$, que tende a $\sqrt{1/2}$ quando ambas intensidades tendem a zero. Além disso, supomos que o albedo relativo β não pode ser maior que 1; portanto excluimos os pares $\mathbf{S}_i, \mathbf{G}_i$ sempre que $\mathbf{S}_i \geq \mathbf{G}_i + 3\sigma$.

Tendo encontrado a pontuação máxima W_{\max} , calculamos a distância Δ_2 de tal maneira que Δ_2 é grande quando W_{\max} é menor ou igual a 2, e tende rapidamente à zero quando W_{\max} é maior ou igual a 3. Utilizamos para tanto a fórmula

$$\Delta_2(\mathbf{S}, \mathbf{G}) = \exp \left[-\frac{1}{2} \left(\frac{W_{\max}}{W_{\text{ref}}} \right)^2 \right] \quad (4.4)$$

onde W_{ref} é um parâmetro ajustado experimentalmente. No nossos testes, obtivemos melhores resultados com W_{ref} .

O custo médio de calcular $\Delta_2(\mathbf{S}, \mathbf{G})$ segundo as fórmulas 4.3 e 4.4 é proporcional a m^2 . O custo pode ser facilmente reduzido para $m \log m$ ordenando-se os pares a_i pelo albedo estimado β_i e considerando no somatório (4.3) apenas os termos mais significativos, onde $\beta_j \simeq \beta_i$. Com esta otimização nosso método é viável mesmo quando m chega a uma centena ou mais.

4.4.3 A métrica Δ_9

Nos testes relatados mais adiante, utilizamos uma versão mais elaborada da métrica, que denominamos Δ_9 . A função Δ_9 utiliza inferência Bayesiana para estimar a probabilidade dos pontos p e q terem a mesma normal. Ela supõe que para pontos com a mesma normal e na ausência de anomalias, o valor de \mathbf{S}_i é $\beta\mathbf{G}_i + N$ onde N é um ruído com distribuição gaussiana. Por outro lado, supõe que, na presença de sombras ou *highlights*, o valor de \mathbf{S}_i pode estar acima ou abaixo de $\beta\mathbf{G}_i$ com probabilidade uniforme. A função Δ_9 usa uma aproximação constante-por-partes para calcular a função de distribuição de probabilidade para cada β testado.

Mais precisamente, a métrica Δ_9 supõe que a intensidade \mathbf{S}_i é uma variável aleatória que pode resultar de três processos. Com certa probabilidade ω_1 , o valor de \mathbf{S}_i está numa sombra projetadas e portanto pode ter um valor arbitrário entre 0 e $\beta\mathbf{G}_i$, onde β é o albedo relativo. Com probabilidade ω_2 o valor de \mathbf{S}_i está sob efeito de *highlights* ou iluminação secundária, possuindo valor entre $\beta\mathbf{G}_i$ e 1. Com probabilidade $1 - \omega_1 - \omega_2$, p é canônico em \mathbf{S}_i , e portanto seu valor é $\beta\mathbf{G}_i$ mais um ruído gaussiano N_i com desvio

padrão σ .

Pela fórmula de Bayes, a probabilidade do albedo relativo ter β dados os valores de $\mathbf{S}_i, \mathbf{G}_i$, e supondo que as normais coincidem é

$$Pr(\beta|\mathbf{S}_i, \mathbf{G}_i) = \frac{Pr(\mathbf{S}_i|\beta\mathbf{G}_i)Pr(\beta)}{\int \alpha Pr(\mathbf{S}_i|\alpha, \mathbf{G}_i)Pr(\alpha)d\alpha} \quad (4.5)$$

Como de praxe, o denominador não precisa ser calculado pois é apenas um fator de normalização. Neste cálculo supomos $Pr(\beta)$ uniforme em $[0, 1]$ e utilizamos uma aproximação constante por partes para $Pr(\mathbf{S}_i|\beta, \mathbf{G}_i)$. Supomos também que as probabilidades são independentes para cada iluminação e portanto

$$Pr(\beta|S, G) = \prod_i^m Pr(\beta|\mathbf{S}_i, \mathbf{G}_i) \quad (4.6)$$

Cada fator desta produtória é uma função constante-por-partes definida por três fatores e dois degraus em albedos β'_i, β''_i como na figura 4.4

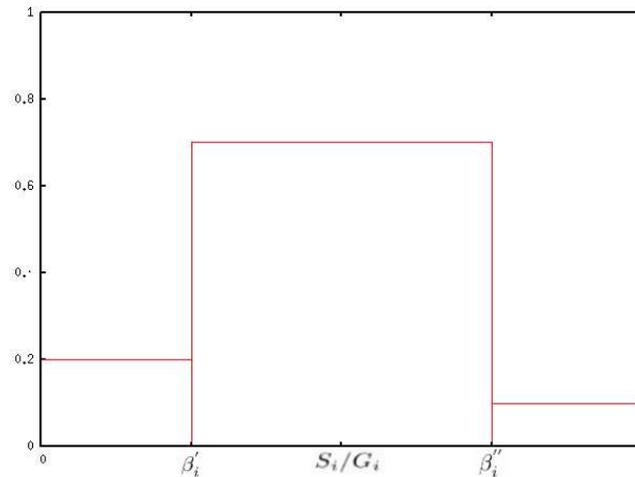


Figura 4.4: Gráfico típico da aproximação para $Pr(\beta|\mathbf{S}_i, \mathbf{G}_i)$ em função de β . O pico está aproximadamente centrado em $\beta_i = \mathbf{S}_i/\mathbf{G}_i$

Para calcular esta produtória de maneira eficiente, ordenamos os albedos relativos β'_i, β''_i para as m iluminações e em seguida percorremos esta lista ordenada atualizando o valor da produtória a cada degrau.

Nesta métrica a pontuação do albedo W_{\max} é igual a $1/\log(Pr(\beta_{\max}|\mathbf{S}_i, \mathbf{G}_i))$ onde β_{\max} é o β com maior probabilidade para o par $\mathbf{S}_i, \mathbf{G}_i$. O valor da distância é então calculado pela função 4.4.

O custo médio para calcular $\Delta_9(S, G)$ é proporcional a $2m \log 2m$, dominado pela

ordenação dos $2m$ albedos relativos. O cálculo de produtória tem tempo proporcional a m . Esta característica permite o uso de Δ_9 mesmo em situações com dezenas de iluminações distintas.

4.5 Testes

Para verificar a eficácia de nosso método de tratamento de anomalias, realizamos testes em dois tipos de dados: (1) imagens sintéticas, (2) fotos digitais de uma cena real. Utilizou-se a função Δ_9 para todos os testes.

4.5.1 Testes com cena sintética

Como discutido na seção 3.6.1, testes com imagens sintéticas permitem verificar a eficácia do método de busca na tabela sem a interferência de defeitos indesejados encontrados em imagens reais.

As imagens deste lote foram produzidas pelo software POV-Ray [6] a partir de uma cena virtual semelhante à usada na seção 3.6.1, consistindo de blocos de diversos formatos e cores posicionados sobre uma base plana quadriculada. Utilizamos onze imagens para este teste: duas delas renderizadas com iluminação quase vertical e livre de sombras projetadas, as outras apresentando sombras projetadas em regiões extensas da cena, tanto na base quadriculada quanto nos blocos. Veja a figura 4.5.

Todos os objetos possuem acabamento lambertiano e foram iluminados por uma fonte de luz pontual, distante aproximadamente 100 vezes a largura da parte visível da cena. A câmera foi ajustada para oferecer uma projeção quase paralela da cena.

Para construção da tabela de assinaturas do gabarito foram utilizados 3771 pontos q de amostragem em um gabarito virtual gerado internamente a partir modelo de *shading* do POV-Ray. Os mapas resultantes de normal e de albedo são apresentados nas figuras 4.6 e 4.7.

4.5.2 Teste com cenas reais

Para o teste com imagens de uma cena real, utilizamos a mesma escultura de gesso e gabaritos utilizados na seção 3.6.2, em disposição similar. Obtivemos 17 imagens digitais desta cena com a mesma câmera de 3 megapixels (Sony CyberShot DSC-V1) e a mesma fonte de luz, segundo o mesmo protocolo. Veja a figura 4.9. A luminária foi posicionada de modo a satisfazer a hipótese de dados suficientes para toda a cena com um subconjunto de imagens κ diferente para cada pixel. O ângulo θ da luz com a normal foi variada entre 25° a 45° , para azimutes irregularmente espaçados entre 0° e 270° .

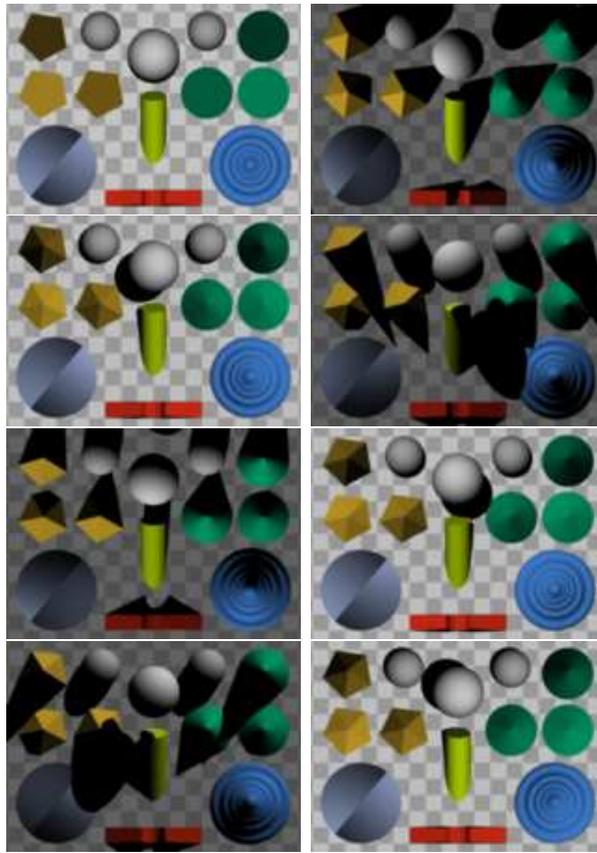


Figura 4.5: Imagens da cena sintética.

As fotos foram alinhadas manualmente por translação com precisão de ± 0.35 pixels, cortadas de 2048×1536 para 672×1000 pixels, reduzidas para 225×163 pixels e convertidas para tons de cinza. Os mapas de normais e de albedo resultantes são apresentados nas figuras 4.10 e 4.11.

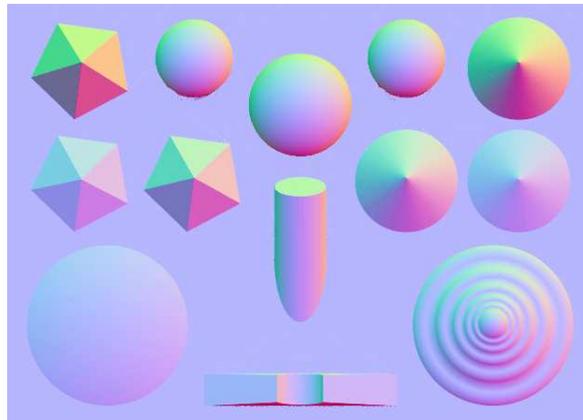


Figura 4.6: Mapa de normais calculado à partir do conjunto de imagens sintéticas.

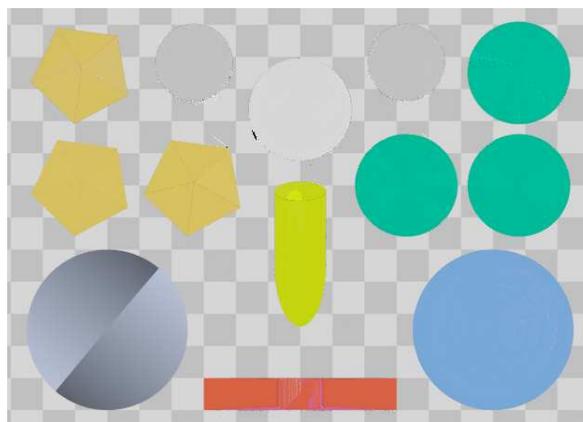


Figura 4.7: Mapa de albedo calculado a partir do conjunto de imagens sintéticas.

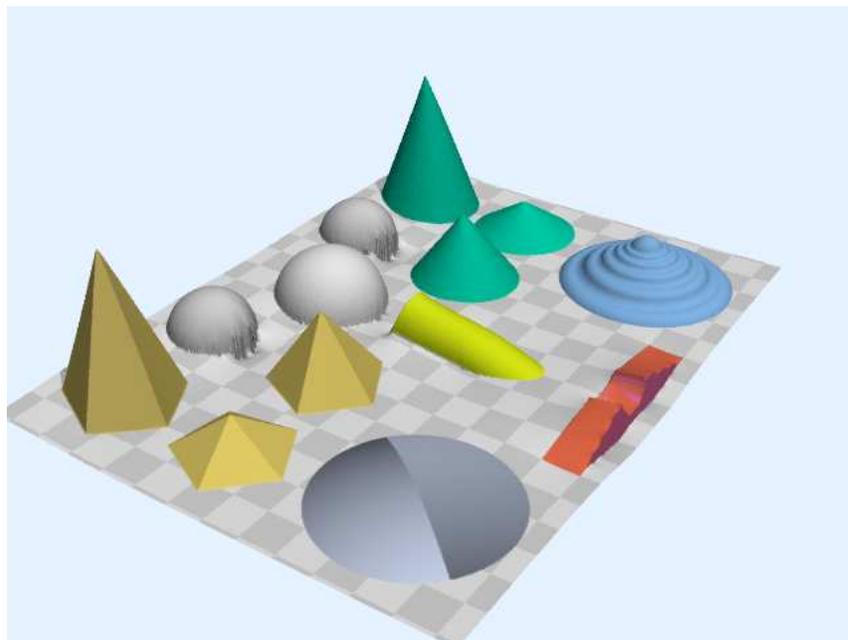


Figura 4.8: Visualização em perspectiva do mapa de alturas calculado a partir do conjunto de imagens sintéticas.



Figura 4.9: Algumas das imagens utilizadas para teste de cena real



Figura 4.10: Mapa de normais calculado à partir de cenas reais.



Figura 4.11: Mapa de albedo calculado à partir de cenas reais.

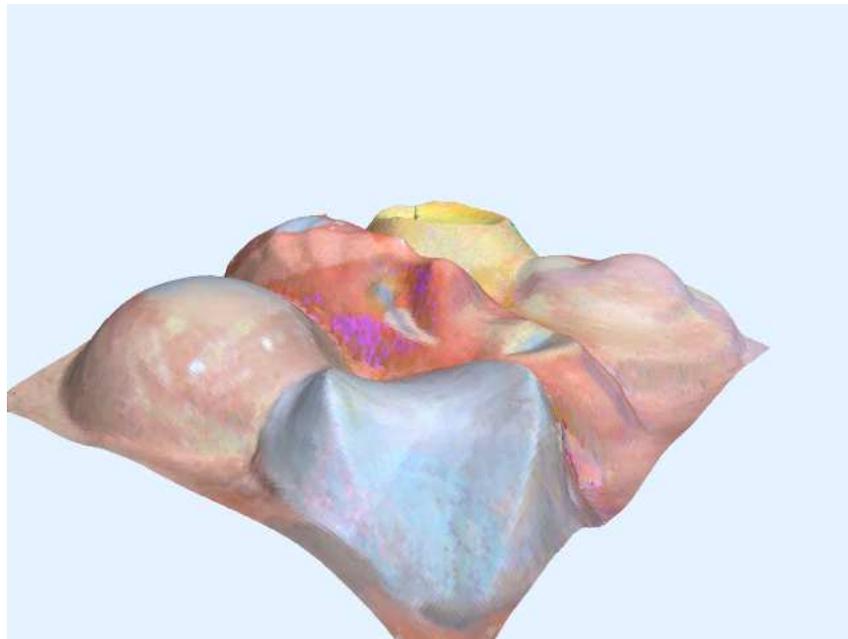


Figura 4.12: Superfície calculada à partir de cenas reais.

Capítulo 5

Aplicação - Captura de geometria facial

Neste capítulo será descrita a utilização dos métodos desenvolvidos neste projeto para captura de geometria 3D da uma face humana, usando uma *webcam* e um monitor como fonte de iluminação variável. Obviamente, a técnica também pode ser usada para qualquer objeto de dimensões comparáveis. Esta combinação fornece um método fácil e de baixo custo para captura 3D de objetos sem equipamento especial ou fontes de iluminação separadas, que pode ser uma ferramenta útil em aplicações de segurança, identificação ou autenticação.

5.1 Soluções anteriores

A captura da geometria 3D de faces humanas popularizou-se após o surgimento de *scanners* 3D a *laser* [21], que continuam sendo o principal meio de obtenção deste tipo de dado. Uma limitação destes equipamentos, além de seu custo elevado, é a baixa resolução espacial (da ordem de alguns mm), insuficiente para uma modelagem realista ou detecção de expressões faciais. Para contornar esta limitação são necessárias informações adicionais, por exemplo, um modelo facial paramétrico da face em questão, como proposto por Bickel [2].

Em vista destas limitações vários pesquisadores investigaram métodos diferenciados para captura de faces. Os métodos considerados incluem estéreo geométrico usando marcadores [17], luz estruturada [20] ou *shape-from-silhouettes* [3]. Como estas técnicas não são suficientes para determinar o mapa denso de normais ou alturas, eles geralmente usam modelos estatísticos da geometria facial para preencher lacunas na informação obtida.

O uso de estéreo fotométrico com iluminações múltiplas para este problema foi inicialmente proposto por Georgiades *et al* [12]. Sua solução supõe fontes de luz pontuais

distantes. Para fornecer as diferentes iluminações o autor utilizou um equipamento especial consistindo de 64 fontes de luz pontuais controladas por computador, afixadas a uma esfera envolvendo o usuário. A captura de um conjunto de 64 imagens leva em torno de 2 segundos para reconstrução são necessárias 7 ou mais imagens. Embora seu método produza resultados de boa qualidade, o custo e tamanho do equipamento limitam bastante sua aplicação.

A utilização de monitores como fonte de iluminação alternativa para estéreo fotométrico foi descrita em 2006 por Funk [11]. Ele utilizou um monitor LCD ou CRT e uma câmera digital controlada por computador colocadas em uma cabine para eliminar luz ambiente. Para reconstrução, Funk utilizou a variante clássica de estéreo fotométrico, supondo fontes de luz pontuais a distância infinita. Por essa razão, cada iluminação era produzida por um pequeno quadrado branco com fundo negro exibida pelo monitor. Devido à baixa intensidade da luz, a captura era bastante demorada — de 5 a 30 segundos por imagem. Esta característica praticamente inviabiliza este método para captura de geometria facial.

A limitação de fontes de iluminação pontual foi parcialmente levantada por Clark [8] em 2008, Seu algoritmo supõe fontes luminosas retangulares (em vez de pontuais) . Em seu método, cada luz é na verdade proveniente de uma imagem arbitrária exibida no monitor. Em uma etapa de pré-processamento, cada imagem é analisada para determinar as dimensões e direções de uma fonte retangular aproximadamente equivalente. Esta abordagem permite o uso de padrões luminosos com grande luminosidade total e captura em tempo relativamente pequeno. Entretanto o método de cálculo requer o conhecimento da posição do objeto analisado em relação ao monitor, somente dá bons resultados para objetos pequenos, e não funciona na presença de sombras, nem mesmo próprias. Portanto o mesmo não é adequado para captura de geometria facial.

Determinamos que as técnicas de estéreo fotométrico baseado em gabaritos dos capítulos 3 e 4 fornecem uma solução para este problema. Nossa abordagem tem várias vantagens sobre os métodos citados. Em particular nossa solução não impõe restrições sobre a posição e forma das fontes de luz pois utilizamos gabaritos. Nossos algoritmos também oferecem rapidez e robustez no cálculo de normais na presença de sombras e outras anomalias.

5.2 Validação do estéreo fotométrico para face humana

Embora pele humana não seja exatamente lambertiana, verificamos que métodos de estéreo fotométrico proporcionam resultados bastante satisfatórios, desde que o número de imagens seja suficientemente grande. Para confirmar esta tese, aplicamos nosso método a um conjunto de 20 imagens (veja figura 5.1) do banco de faces B da Universidade de

Yale [12]. Este banco contém imagens de faces humanas de boa qualidade (iluminação com fontes de luz pontual, câmera de resposta linear, imagens perfeitamente alinhadas). Com a falta de um gabarito de iluminação, geramos a tabela de assinaturas do gabarito internamente usando a direção da fonte luminosa de cada imagem indicada no banco de faces.

A qualidade do resultado (veja figura 5.2) é muito boa, o que é surpreendente pois muitas das imagens apresentam *highlights*, às vezes bastante extensos.

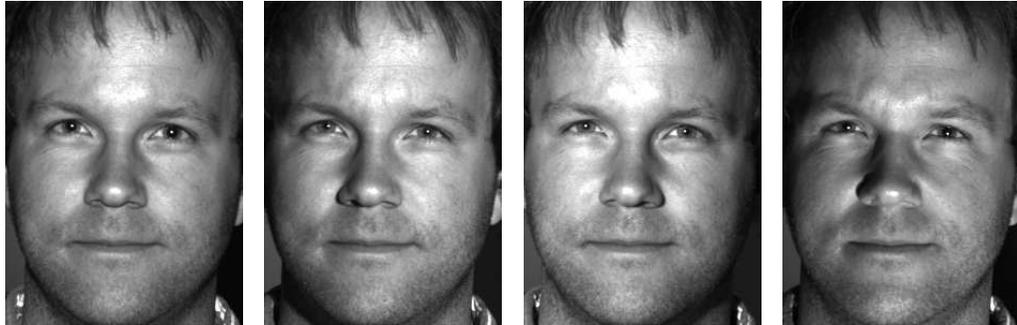


Figura 5.1: Quatro das imagens sob condições ideais da base de dados de Yale .

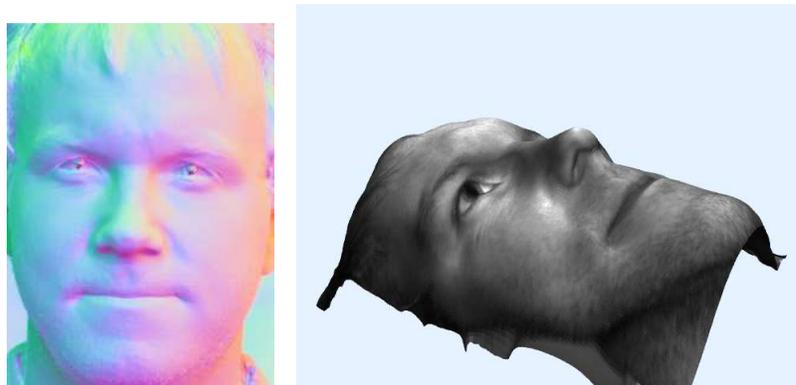


Figura 5.2: Mapa de inclinação e Mapa de alturas calculados a partir das imagens 5.1.

Como este exemplo mostra, a face humana geralmente é uma superfície suave, sem descontinuidades de profundidade (exceto nas narinas) nem superfícies polidas (exceto nos olhos). Vale notar também que, nas imagens deste banco quase todas as sombras são sombras próprias.

5.3 Detalhamento de nossa abordagem

Nossa solução para a captura de geometria facial consiste nas seguintes etapas:

1. Exibição de padrões de iluminação
2. Captura das imagens
3. Alinhamento de imagens
4. Extração dos gabaritos de iluminação
5. Determinação das normais por estéreo fotométrico
6. Integração do mapa de normais

5.3.1 Escolha dos padrões de iluminação

Na escolha dos padrões de iluminação há uma tensão entre fontes concentradas e difusas. Fontes de luz concentradas maximizam o contraste entre os vetores de observação de pontos com normais diferentes. Porém, tais fontes possuem pouco brilho, o que aumenta o nível de ruído da imagem e/ou exige tempos maiores de exposição.

Deve-se também cuidar para que a hipótese de dados suficientes (seção 4.2.1) seja válida sobre toda a face. Em vista destas considerações, usamos como padrões de iluminação quadrados brancos ocupando $\frac{1}{3}$ ou $\frac{1}{4}$ da área da tela em várias posições, contra um fundo cinza escuro. Veja a figura 5.3.

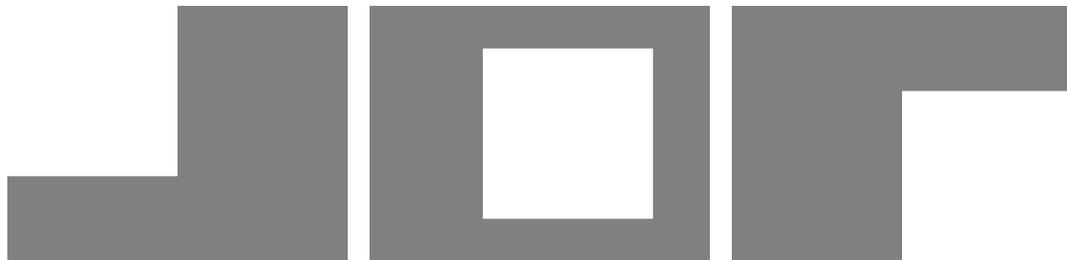


Figura 5.3: Três das imagens utilizadas como fonte de iluminação.

5.3.2 Captura das Imagens

Para aplicação de estéreo fotométrico é necessário que a cena não se altere durante a captura de todas as imagens. No caso de faces, a captura deve ser rápida o bastante para que o usuário permaneça imóvel durante toda sessão. Se a captura levar mais que 2-3 segundos, o usuário provavelmente vai mover a cabeça, ao respirar ou ao acompanhar inconscientemente o movimento da fonte de luz com os olhos. Além disso, a captura também deve ser acionada automaticamente, uma vez que o ato de pressionar uma tecla inevitavelmente leva o usuário a mudar de posição.

Note que para os métodos propostos o tempo necessário para captura de cada imagem pela câmera depende do hardware da câmera, do nível de iluminação e da taxa de transmissão de dados. Uma câmera digital (*webcam*) convencional de computador normalmente consegue uma taxa de captura de quadros em torno de 20 a 30 quadros por segundo. Porém, nas condições de iluminação fraca que consideramos, a taxa de captura pode ser bem menor, 4 quadros por segundo ou menos. Uma vez que 6 ou 10 imagens são mais que suficientes, a captura do conjunto completo pode ser feita em 2 a 3 segundos.

Algumas *webcams* possuem ajustes automáticos internos de brilho, contraste e tempo de exposição. O programa de captura deve levar estes detalhes em conta e procurar contornar esses ajustes. A maioria das *webcams* capturam as imagens pelo método da matriz de Bayer e convertem internamente a informação para o padrão YUV: 4:2:2. Neste padrão, a informação do brilho (Y) possui o dobro da resolução horizontal dos canais de cromaticidade (U e V). Para o cálculo das normais recomenda-se trabalhar diretamente com o canal Y destas imagens brutas, pois a conversão de YUV para RGB pode acarretar perda de informação.

5.3.3 Alinhamento das Imagens

Por mais rápida que seja feita a captura, é impossível evitar pequenos deslocamentos de alguns pixels entre as imagens, quer pelo movimento do usuário, quer por vibrações da câmera, entre as capturas da imagem. Portanto é geralmente necessário alinhar as imagens, de preferência automaticamente, antes da aplicação do estéreo fotométrico.

5.3.4 Gabaritos de iluminação

Nossos algoritmos para cálculo das normais com iluminação arbitrária exigem a inclusão de gabaritos de iluminação na cena. Em princípio basta um gabarito esférico. O gabarito pode ser segurado com a mão, desde que seja posicionado ao lado da face, de forma a evitar sombras projetadas da face nos gabaritos e vice-versa. Veja figura 5.4. Não é necessário que o gabarito esteja exatamente na mesma posição (quer absoluta, quer relativa à face) em todas as imagens.

A extração do gabarito na cena pode ser feita especificando-se manualmente seu raio e as coordenadas (x, y) do centro. Estes valores podem também ser determinados por diversos métodos de segmentação.

5.3.5 Cálculo de normais

Nossos experimentos indicam que até mesmo as técnicas de estéreo fotométrico simples do capítulo 2 produzem mapas de normais satisfatórios. Portanto o cálculo das normais



Figura 5.4: Posicionamento dos gabaritos de iluminação .

pode ser acelerado pelo uso de uma grade de baldes como descrito no capítulo 3.

5.3.6 Recuperação da superfície

O mapa de normais obtido pelos métodos descritos neste trabalho é convertido em um mapa de profundidade por um integrador multi-escala [9]. Observamos que o mapa de alturas resultante é geralmente correto, pois faces em visão frontal praticamente não tem oclusões nem descontinuidades. O maior problema é a presença de detalhes muito escuros, como a pupila dos olhos e pelos faciais (barba, bigode, cílios, etc...), onde o estéreo fotométrico não é aplicável. Felizmente há poucas regiões de descontinuidade e as regiões de albedo muito baixo recebem valores baixos no mapa de ponderação, minimizando o impacto dos erros das normais calculadas nessas regiões.

5.4 Testes

Nos testes realizados, usamos um notebook de baixo custo (Asus EEE PC 701), que possui uma camera digital embutida com resolução de 640x480 pixels com quadro codificado em padrão YUV 4:2:2 com gama suposto 2.2 e taxa de captura variando de 7 a 30 quadros por segundo dependendo do nível de iluminação da cena. Uma vez que a iluminação fornecida pelo monitor LCD do Asus de 7 polegadas não era suficiente para captura de imagens pela câmera, utilizamos um monitor CRT convencional de 17 polegadas (Philips 109B4) conectado à saída de vídeo do notebook.

Nos testes, o usuário estava posicionado a aproximadamente 50cm do monitor e da câmera do notebook. Como gabaritos, usamos uma esfera de madeira afixada em um cartão preto, e uma bola de ping-pong afixada em um pino de metal. Os gabaritos foram pintados com esmalte *spray* branco fosco, posicionados a aproximadamente 5cm do

usuário e à mesma distância do monitor.

Foram utilizadas 24 padrões de iluminações distintas, com o quadrado branco posicionado em 6 posições horizontais e 4 verticais distintas. A captura das imagens foi realizada em um ambiente escuro, predominando a luz do monitor. Nestas condições o tempo total de captura para as 24 imagens variou entre 4 a 5 segundos. Observamos que a câmera armazena intensidade $Y = 16$ (numa escala de 0 a 255) para pixels muito escuros, portanto aplicamos uma normalização nas imagens. Veja figura 5.5.



Figura 5.5: Quatro das imagens utilizadas para testes.

Para o cálculo das normais, construímos uma tabela de assinaturas com aproximadamente 10000 amostras do gabarito e utilizamos uma grade de baldes de tamanho $N \times N$ com $N = 211$. Os mapas de normais e albedo resultantes podem ser vistos na figura 5.6.



Figura 5.6: Mapa de normais e mapa de albedo calculados a partir das figuras 5.5

As lacunas no mapa de normais se devem a regiões de albedo baixo, como descrito na seção 5.2. Apesar da ocorrência das mesmas, o integrador conseguiu obter um mapa de alturas qualitativamente correto. Veja figura 5.7.

Um defeito notável do mapa de alturas é uma curvatura acentuada, aparentemente devida à variação espacial da iluminação, que por sua vez é devida à pequena distância



Figura 5.7: Mapa de alturas calculados a partir das figuras 5.5

entre a face e a tela. Outro fator que pode estar contribuindo para esta deformação é a resposta não-linear da câmera, que ainda não conseguimos determinar precisamente.

Capítulo 6

Conclusões

6.1 Contribuições

Os métodos desenvolvidos neste trabalho apresentam um aprimoramento considerável do método estéreo fotométrico baseado em gabaritos em relação á métodos encontrados na literatura.

6.1.1 Aceleração de estéreo fotométrico

A técnica de aceleração por *hashing* 2D proporciona uma redução considerável no cálculo de normais, proporcionando um método cujo tempo de execução cresce de maneira quase independente da quantidade de imagens usadas. A recuperação de normais com alto grau de precisão (em torno de 10000 entradas na tabela de assinaturas) pode ser realizada em poucos minutos.

6.1.2 Interpolação e suavização de gabaritos

A qualidade das informações extraídas da imagem do gabarito de iluminações é vital para o sucesso de um método EFBG. O desenvolvimento da interpolação de gabaritos permite evitar defeitos provocados por gabaritos que apresentem irregularidades em suas superfícies, garantindo maior qualidade na recuperação do mapa de normais.

6.1.3 Estéreo fotométrico robusto

A métrica de distância proporcionada pela função Δ permite efetuar o cálculo do mapa de normais de uma superfície mesmo na presença de sombras e *highlights*. Esta contribuição permite a utilização de métodos estéreo fotométricos em cenas onde não é possível evitar a presença destes efeitos, tornando maior o campo de aplicação prática destes meios.

6.2 Aplicações

A utilização dos métodos de aceleração e tratamento de imagens com sombreamento ou highlights discutidos nos capítulos anteriores possibilitam a aplicação prática em diversos problemas de reconstrução tridimensional sem requerir equipamento dispendioso.

Uma aplicação que pretendemos investigar é projeto Raab, cujo objetivo é desenvolver técnicas computacionais para reconstrução de objetos cerâmicos e líticos fragmentados.

6.3 Trabalhos Futuros

6.3.1 Aceleração de estéreo fotométrico robusta

O método de aceleração desenvolvido considera o pixel analisado livre de efeitos indesejados na iluminação. Uma abordagem baseada na combinação de RANSAC com a aceleração oferecida pelo hashing está sendo estudada, possibilitando associar robustez com pouco tempo de processamento, tornando a eficácia dos métodos discutidos ainda maior.

6.3.2 Associação com estéreo geométrico

A associação com métodos estereo geométricos ainda está sendo estudada como uma maneira de recuperar ou complementar informação de objetos que não podem ser convenientemente abordados por métodos estereo geométricos. Planeja-se também uma possível extensão dos métodos para cobrir aspectos como especularidade e reflexão, não cobertos pelo modelo Lambertiano assumido.

Referências Bibliográficas

- [1] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the Association of Computing Machinery*, 45(6):891–923, 1998.
- [2] M. Botsch M. Otaduy M. Gross B. Bickel, M. Lang. Multi-scale capture of facial geometry and motion. In *ACM Transactions on Graphics - Proceedings of the 2007 ACM SIGGRAPH conference*, number 23, 2007.
- [3] Hanspeter Pfister Baback Moghaddam, Jinho Lee and Raghu Machiraju. Model-based 3d face capture with shape-from-silhouettes. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, page 20, 2003.
- [4] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in presence of highlights and shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, October 2003.
- [5] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.
- [6] Chris Cason and POV-Team. Persistence of Vision Raytracer. Electronic document at <http://www.povray.org/>, edit date 2008/07/30, last accessed on 2008/10/25, July 2008.
- [7] Manmohan Chandraker, Sameer Agarwal, and David Kriegman. ShadowCuts: Photometric stereo with shadows. In *CVPR07*, pages 1–8, 2007.
- [8] James J. Clark. Photometric stereo using lcd displays. *Image and Vision Computing (Article in Press)*, November 2008.
- [9] Helena Cristina da Gama Leitão, Rafael Saracchini, and Jorge Stolfi. Depth from slope by multi-scale integration (paper being written).

- [10] Helena Cristina da Gama Leitão, Rafael Saracchini, and Jorge Stolfi. A bucket grid structure to speed up table lookup in gauge-based photometric stereo. In *Proceedings of the 20th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007)*, pages 221–227. IEEE Computer Society Press, October 2007.
- [11] Nathan Funk and Yeeh-Hong Yang. Using a raster display for photometric stereo. Technical Report TR06-13, Computing Science, University of Alberta, June 2006.
- [12] A.S. Georghiadis, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.
- [13] Bryan Henderson. Netpbm home page. Electronic document at <http://netpbm.sourceforge.net/>, edit date 2008/08/01, last accessed on 2008/10/22, August 2008.
- [14] Aaron Hertzmann and Steven M. Seitz. Shape and materials by example: A photometric stereo approach. In *Proceedings IEEE CVPR 2003*, volume 1, pages 533–540, June 2003.
- [15] Aaron Hertzmann and Steven M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, August 2005.
- [16] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbour: Towards removing the curse of dimensionality. In *Proceedings of the 13th Annual ACM Symposium on Theory of Computing (STOC'98)*, pages 604–613, 1998.
- [17] Brian Curless Li Zhang, Noah Snavely and Steven Seitz. Spacetime faces: High resolution capture for modeling and animation. pages 248–276, 2007.
- [18] Greg Roelofs. Portable Network Graphics: An open, extensible image format with lossless compression. Electronic document at <http://libpng.org/pub/png/>, edit date 2008/09/19, last accessed on 2008/10/22, September 2008.
- [19] Li Shen, Takashi Machida, and Haruo Takemura. Efficient photometric stereo for three-dimensional surfaces with unknown BRDF. In *Proceedings of the 5th International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, pages 326–333, 2005.
- [20] Terence Sim and Takeo Kanade. Combining models and exemplars for face recognition. In *Proceedings of CVPR Workshop on Models versus Exemplars in Computer Vision*, Lecture Notes in Computer Science, 2001.

- [21] Yasuhito Suenaga and Yasuhiko Watanabe. A method for the synchronized acquisition of cylindrical range and color data. In *MVA*, pages 137–142, 1990.
- [22] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [23] Robert J. Woodham. Determining surface curvature with photometric stereo. In *Proceedings of the 1989 IEEE International Conference on Robotics and Automation*, volume 1, pages 36–42, May 1989.
- [24] Robert J. Woodham. Gradient and curvature from the photometric stereo method, including local confidence estimation. *Journal of the Optical Society of America, Series A*, 11(11):3050–3068, 1994.
- [25] Lin Zhong and James J. Little. Photometric stereo via locality sensitive high-dimension hashing. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05)*, pages 104–111, 2005.