

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS - ICEX
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

**MODELAGEM AUTOMÁTICA DE CENAS COM
ILUMINAÇÃO NÃO LOCAL A PARTIR DE IMAGENS**

JOSÉ PINHEIRO DE QUEIROZ NETO

Belo Horizonte

03 de Junho de 2005

JOSÉ PINHEIRO DE QUEIROZ NETO

**MODELAGEM AUTOMÁTICA DE CENAS COM
ILUMINAÇÃO NÃO LOCAL A PARTIR DE IMAGENS**

Tese apresentada ao Curso de Pós-Graduação
em Ciência da Computação da Universidade
Federal de Minas Gerais como requisito parcial
para a obtenção do grau de Doutor em Ciência
da Computação.

Belo Horizonte

03 de Junho de 2005

UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Modelagem Automática de Cenas com Iluminação Não Local a
Partir de Imagens

JOSÉ PINHEIRO DE QUEIROZ NETO

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Dr. RODRIGO LIMA CARCERONI - Orientador
Departamento de Ciência da Computação – UFMG

Prof. Dr. MÁRIO FERNANDO MONTENEGRO CAMPOS - Co-Orientador
Departamento de Ciência da Computação – UFMG

Prof. Dr. LUIZ CARLOS PACHECO RODRIGUES VELHO
Instituto de Matemática Pura e Aplicada - IMPA

Prof. Dr. ALEXEI MANSO CORRÊA MACHADO
Pontifícia Universidade Católica de Minas Gerais – PUC/MG

Prof. Dr. RENATO CARDOSO MESQUITA
Departamento de Engenharia Elétrica – UFMG

Prof. Dr. GERALDO ROBSON MATEUS
Departamento de Ciência da Computação – UFMG

Belo Horizonte, 03 de Junho de 2005 DCC/UFMG.

Resumo

Esta tese trata do problema de adquirir os modelos de uma cena a partir de imagens, onde a correspondência entre as intensidades dos pixels das imagens e as cores reais da cena não podem ser explicadas a partir de um modelo de iluminação local da interação da luz com a cena. Mais especificamente, o problema é abordado em dois casos particulares tratados em cenários distintos.

No primeiro caso, têm-se cenas submersas em água turva onde a luz interage globalmente com o meio, ao qual é desenvolvida uma metodologia para extrair a geometria da cena, utilizando-se um modelo de propagação da luz na água integrado a um algoritmo de visão estéreo, recuperando o mapa de disparidades a partir de um par estéreo da cena adquirido em pontos de vista distintos. Resultados experimentais comparam favoravelmente o método proposto nesta tese a um método de estéreo denso bem estabelecido. O método permite, ainda, restaurar as imagens da cena subaquática.

No segundo caso, têm-se cenas onde a luz interage globalmente com objetos com superfícies homogêneas, reflexivas e translúcidas, apresentando sombras e inter-reflexão entre objetos, ao qual é desenvolvido um método para adquirir os parâmetros de um modelo de reflectância de cada objeto da cena, resolvendo um sistema de equações fatorado obtido a partir de uma única renderização e tendo como restrições uma imagem da cena, cuja solução é mais estável e eficiente em comparação a métodos similares.

Abstract

This thesis deals with the problem of acquiring the models of a scene from its images, where the correspondence between the image pixels intensities and the real colors of the scene cannot be determined using a local illumination model of the interaction of the light with the scene. More specifically, the problem is treated in two particular cases in distinct scenes.

In the first case, we have the underwater scenes in muddy water where the light interacts globally with the environment, to which we developed a methodology to extract the scene's geometry, by integrating a light propagation model into a stereo vision algorithm and recovering the disparity maps from images pairs acquired from distinct viewpoints. Experimental results are compared to known methods. The method allows to restore the images of the underwater scene.

In the second case, we have a scene where the light interacts globally with objects that has homogeneous, reflexives and translucent surfaces, presenting shades and Inter-reflection between objects, to which we developed a method of recovering the reflectance parameters to each object of the scene, by solving a factorized equation system obtained using just one rendering and having as restrictions an image of the scene, whose solution is more steady and efficient in comparison to similar methods.

Agradecimentos

À Deus, que me tem acompanhado ao longo da minha caminhada.

Aos meus queridos familiares, minhas filhas Larissa e Juliana, minha esposa Michelle, minha mãe D. Joana e a todos meus irmãos, em especial à Josse e Jerson que muito me ajudaram cuidando das minhas coisas em Manaus.

Ao meu orientador Prof^o Rodrigo Lima Carceroni, pela amizade, confiança e dedicação na orientação deste trabalho, e ao meu co-orientador Prof^o Mário Campos que contribuiu para meu crescimento ao longo do doutorado.

Ao CEFET/AM, cujo auxílio financeiro permitiu a dedicação necessária à realização deste trabalho.

Aos amigos Lara Cristina e Wagner Barros, pelo empenho e dedicação nos trabalhos que desenvolvemos em conjunto, muitos dos quais contribuíram diretamente para a conclusão desta tese.

A todos os amigos com quem convivi em Belo Horizonte, em especial ao Nakamura, Fabíola, Maurício, Ingrid, Pedro, Aracelly, Ruitier, Ruth, Vilar, Gisele, Michele, Aldri, Pio, Menoti, David, Elbena, Silvana, Kissia, Dani, Gurvan, Cris e outros que não recordo agora, pelo companheirismo, amizade e bons momentos vividos ao longo de todo o período em que estivemos juntos.

Aos colegas do laboratório de Visão e Robótica que sempre me apoiaram, meus amigos Cardeal, Guilherme, Shiroma, Andréia, Rogério e Chaimo.

Ao Centro de Pesquisas Hidráulicas e Recursos Hídricos, na pessoa do Prof^o Carlos Martinez, e aos demais colegas Hersília, Edna, Viviane e Paulo Henrique. A todos agradeço pelo apoio e companheirismo que sempre me foram prestados.

A todos MUITO OBRIGADO!

Sumário

Lista de Figuras	x
Lista de Tabelas	xi
Lista de Símbolos	xiii
Lista de Algoritmos	xiii
1 Introdução	1
1.1 Motivação	1
1.1.1 Obtendo a Geometria da Cena	4
1.1.2 Obtendo a Reflectância da Cena	7
1.2 Descrição do Problema	9
1.2.1 Cena Subaquática	9
1.2.2 Cena com Inter-reflexões e Objetos Translúcidos	11
1.3 Abordagem de Solução dos Problemas	12
1.3.1 Estéreo Subaquático	12
1.3.2 Reflectância Inversa Eficiente	13
1.4 Contribuições	14
1.5 Aplicações	14
1.5.1 Projeto Sistema de Contagem Automática de Peixes	15
1.5.2 Projeto Estrada Real Digital	16
1.6 Organização da Tese	18
2 Trabalhos Relacionados	19
2.1 Conceitos de Radiometria e Reflectância	19
2.1.1 Radiometria	19
2.1.2 Modelos de Iluminação Local e Global	23
2.2 Renderização e Modelagem Baseada em Imagens	25
2.2.1 Métodos com Iluminação Local	26
2.2.1.1 Utilizando Múltiplas Imagens	27
2.2.1.2 Utilizando Uma Imagem	29
2.2.2 Métodos com iluminação global	30
2.2.2.1 Radiosidade Inversa	30

2.2.2.2	Métodos Baseados em <i>Ray Tracing</i>	32
2.2.2.3	Métodos de Formulação Geral	32
2.2.3	Utilizando Dispositivos Especiais	34
2.3	Conceitos de Visão Estéreo	35
2.3.1	O Problema da Reconstrução 3D	36
2.3.2	O Problema da Correspondência	37
2.3.3	Método de Minimização via Corte de Grafos	40
2.4	A Visão Computacional em Imagens Subaquáticas	41
2.4.1	Extraindo Dados sem a Melhoria das Imagens	44
2.4.2	Melhoria das Imagens sem Considerar um Modelo de Propagação da Luz na Água	45
2.4.3	Melhoria das Imagens Utilizando um Modelo de Propagação da Luz na Água	46
2.4.3.1	Imagens da Mesma Cena em Condições Ambientais Diferentes	47
2.4.3.2	Imagens da Mesma Cena com o Uso de um Filtro Polarizador	48
2.4.3.3	Utilizando um Par Estéreo da Cena	49
3	Obtendo Informações 3D em Cenas Subaquáticas Utilizando Estéreo	50
3.1	Modelo de Propagação da Luz na Água	51
3.2	Estimativa dos Parâmetros do Modelo de Propagação da Luz na Água	54
3.2.1	Cálculo do Parâmetro de Dispersão	54
3.2.2	Cálculo do Parâmetro de Atenuação	55
3.2.3	Refinamento dos Parâmetros do Modelo	56
3.3	Estéreo com Estimativa Experimental dos Parâmetros do Modelo de Propagação da Luz	57
3.4	Estéreo com Estimativa Automática dos Parâmetros do Modelo de Propagação da Luz	59
3.5	Análise de Complexidade dos Métodos	61
3.5.1	Complexidade de Custo	61
3.5.2	Complexidade de Espaço	62
3.6	Limitações dos Métodos	62
4	Obtendo os Parâmetros de Reflectância de uma Cena com Iluminação Global	63
4.1	Modelo de Formação da Imagem	64
4.2	Fatoração Fotométrica	68
4.3	Recuperando as Propriedades Fotométricas	70
4.3.1	Otimização Polinomial	71
4.3.2	Otimização Não Polinomial	73
4.4	Análise de Complexidade do Método	75
4.4.1	Complexidade de Tempo	75
4.4.2	Complexidade de Espaço	75
4.5	Limitações do Método	76

5	Implementação e Resultados	77
5.1	Estéreo Subaquático	77
5.1.1	Ambiente Experimental	77
5.1.2	Calibração das Câmeras	78
5.1.3	Retificação das Imagens	79
5.1.4	Adaptação do Algoritmo de Estéreo Convencional no Estéreo Subaquático	81
5.1.5	Resultados do Estéreo Subaquático com Estimativa Experimental dos Parâmetros da Água	82
5.1.5.1	Cena com Água Limpa (Turbidez Baixa)	82
5.1.5.2	Cena com Água + Leite (Turbidez Média)	84
5.1.5.3	Cena com Água + Argila (Turbidez Alta)	86
5.1.6	Resultados do Estéreo Subaquático com Estimativa Automática dos Parâmetros da Água	87
5.1.7	Restauração de Imagens Subaquáticas	88
5.2	Reflectância Inversa Eficiente	92
5.2.1	Aplicando o Método a uma Cena Sintética	92
5.2.2	Aplicando o Método a uma Cena Real	96
5.2.2.1	Ambiente Experimental	96
5.2.2.2	Calibração das Câmeras e das Fontes de Luz	98
5.2.2.3	Modelagem 3D da Cena	99
5.2.2.4	Resultados para Cena Real	99
6	Conclusão e Trabalhos Futuros	104
A	Métodos para Solução do Problema da Correspondência	108
A.1	Correlação	109
A.2	Programação Dinâmica	110
A.3	Fluxo Máximo e Cortes Mínimos em Grafos	114
A.3.1	Redes de Fluxo e Fluxos	114
A.3.2	O Método de Ford-Fulkerson	115
A.3.3	Fluxo Máximo, Cortes Mínimos e Problema da Correspondência	118
B	Estéreo com Minimização de Energia via Corte de Grafos	121
B.1	Representação do Problema	121
B.2	O algoritmo de Estéreo Utilizando Movimentação de Expansão da Disparidade	122
B.3	O Funcional de Energia	122
B.4	A Minimização via Corte de Grafos	124
	Referências Bibliográficas	137

Lista de Figuras

1.1	Método básico de renderização de uma imagem.	2
1.2	Renderização com iluminação local e iluminação global.	4
1.3	Exemplo da disparidade binocular.	5
1.4	Exemplo de cenas onde a propagação da luz no meio influencia nas cores da imagem.	6
1.5	Par estéreo de exemplo em ambiente subaquático.	6
1.6	Imagem de uma cena fotometricamente complexa.	8
1.7	Efeitos da redução de contraste de uma cena em água turva	10
1.8	A UHE de Igarapava e o mecanismo de transposição de peixes	16
1.9	Imagens sintéticas da cidade de Ouro Preto.	17
1.10	Matriz N. Sa. da Conceição.	17
2.1	Aspectos geométricos da radiometria.	20
2.2	Geometria do estéreo com imagens paralelas ao <i>baseline</i> (T) e imagens retificadas utilizando a geometria epipolar.	36
2.3	Resultados comparativos de estéreo com a imagem da Universidade de Tsukuba.	39
3.1	Exemplo dos efeitos da água nas imagens.	51
3.2	Modelo simplificado de propagação da luz na água.	52
3.3	Evolução das distâncias escaladas αd e βd	55
3.4	Alvos e resultados da estimativa dos parâmetros da água.	58
4.1	Visão geral do método de reflectância inversa proposto.	64
4.2	Árvore de Computação da Intensidade do Pixel	65
4.3	Geometria do modelo de Phong.	67
5.1	Ambiente experimental do estéreo subaquático.	78
5.2	Posições dos alvos e resultado das calibrações geométricas.	80
5.3	Alvo de calibração radiométrica.	81
5.4	Imagens adquiridas (a) e (b) do experimento com turbidez baixa.	83
5.5	Imagens retificadas (a) e (b) do experimento com turbidez baixa	83
5.6	Mapas de disparidade da cena do experimento na água com turbidez baixa.	83
5.7	Imagens adquiridas (a) e (b) do experimento com turbidez média.	85
5.8	Imagens retificadas (a) e (b) do experimento com turbidez média.	85

5.9	Mapas de disparidade da cena do experimento na água com turbidez média.	85
5.10	Imagens da cena utilizada no experimento com turbidez alta.	86
5.11	Mapas de disparidade da cena do experimento na água com turbidez alta. .	86
5.12	Mapas de disparidade da cena do experimento na água com turbidez média.	88
5.13	Par estéreo temporalmente alinhado.	90
5.14	Mapas de disparidade da cena do experimento de restauração.	90
5.15	Resultado da restauração de imagens	90
5.16	Exemplos de restauração de imagens subaquáticas.	91
5.17	Resultados comparativos - cena sintética	93
5.18	Erro fotométrico para diferentes entradas dos parâmetros de reflectância. . .	94
5.19	<i>Speedup</i> entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.	95
5.20	Tempo de CPU para os métodos RBO, NF-TBO e F-TBO da cena sintética	96
5.21	Alvo de calibração geométrica.	98
5.22	Resultados comparativos - cena real	100
5.23	Erro fotométrico RMS em níveis de cinza para três diferentes entradas dos parâmetros de reflectância / transmitância.	101
5.24	<i>Speedup</i> entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.	102
5.25	Tempo de CPU para os métodos RBO, NF-TBO e F-TBO da cena real . . .	102
5.26	Em (a), (b) apresentam novas vistas da cena sintética obtida pelo método F-TBO. Em (c) a mesma cena sintética sob novas condições de iluminação.	103
A.1	Estereograma formado por pontos aleatórios.	109
A.2	Correlação por meio de uma janela retangular. Um pixel p_1 na imagem I_1 , corresponderá a um pixel p_2 , na região de busca da imagem I_2 , que estiver contido na janela de maior similaridade com relação à janela de p_1	110
A.3	Correspondência via Programação Dinâmica. Representação do caminho de custo ótimo na computação da correspondência entre linhas epipolares correspondentes nas imagens de entrada.	112
A.4	Arcos possíveis que levam à correspondência (i, j) na correspondência via Programação Dinâmica.	113
A.5	Correspondência estéreo como um problema de fluxo máximo.	119

Lista de Tabelas

2.1	Comparação entre os diversos métodos de IBMR.	26
2.2	Métodos de estéreo baseados em busca local da correspondência.	38
2.3	Métodos de estéreo baseados em busca global da correspondência.	38
3.1	Distancias escaladas (cm) e cálculo de β médio (cm^{-1}) em cada canal RGB.	56
3.2	Distancias escaladas (cm), cálculo de α médio (cm^{-1}) em cada canal RGB.	56
3.3	otimização dos parâmetros α , β em cm^{-1} e $L(\infty)$ em intensidade de pixels.	57
5.1	Parâmetros das câmeras utilizadas na aquisição do par estéreo.	79
5.2	Parâmetros α , β em cm^{-1} obtidos experimentalmente - turbidez média.	84
5.3	Parâmetros α , β em cm^{-1} obtidos automaticamente.	87
5.4	Parâmetros α , β em cm^{-1} estimados experimentalmente - cena dinâmica.	89
5.5	<i>Speedup</i> entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.	95
5.6	Valores dos parâmetros de reflectância do modelo obtidos com os métodos F-TBO e RBO/NF-TBO para uma cena sintética	97
5.7	<i>Speedup</i> entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.	101

Lista de Algoritmos

1	Algoritmo da estimativa dos parâmetros da água.	54
2	Algoritmo do estéreo subaquático.	59
3	Algoritmo de estéreo subaquático com estimativa automática dos parâmetros do modelo de propagação da luz na água.	61
4	Computando os parâmetros lineares ótimos $\tilde{x}$, $\tilde{r}$, $\tilde{t}$ e o erro fotométrico residual ϵ , para um dado vetor $\tilde{\theta}$ de parâmetros não lineares.	71
5	Algoritmo de normalização do sistema de equações.	72
6	Valores iniciais de $\tilde{x}$, $\tilde{r}$, $\tilde{t}$ a partir de y	73
7	Reflectância inversa eficiente.	74
8	Algoritmo para resolver o problema da correspondência a partir de técnicas de Programação Dinâmica.	112
9	Método de Ford-Fulkerson para o problema do fluxo máximo.	115
10	Algoritmo básico de Ford-Fulkerson.	117
11	Algoritmo de estéreo convencional (EMGC).	123

Lista de Símbolos

$\mathcal{F}$	função que descreve a luz refletida em um ponto da cena	7
$\mathcal{K}$	função que descreve a reflectância do material em um ponto da cena	7
$\mathcal{H}$	função que descreve a luz incidente em um ponto da cena	7
$\mathcal{G}$	função que descreve a fração de luz refletida a partir de outro ponto da cena	7
Φ	fluxo radiante: energia radiante por unidade de tempo (W)	20
L	radiância: fluxo de energia radiante refletido em uma superfície (W/(m ² sr))	20
ρ	albedo ou coeficiente de reflectância lambertiana (sr ⁻¹)	25
B	radiosidade: energia radiante refletida em um ponto que sofre influencia da luz direta e indireta (W/m ²)	25
Be	emitância: energia radiante emitida por um ponto (uma fonte de luz) (W/m ²)	25
$\mathcal{D}$	disparidade (pixels)	40
$\mathcal{E}$	funcional de energia	40
λ	comprimento de onda da luz que se propaga em um meio (m)	42
β	coeficiente de dispersão da luz em um meio volumétrico (m ⁻¹)	42
α	coeficiente de atenuação da luz em um meio volumétrico (m ⁻¹)	52
E	irradiância: fluxo de energia radiante incidente em uma superfície (W/m ²)	66
I	intensidade da fonte de luz (W)	66
η	índice de refração: é a razão entre a velocidade da luz no vácuo pela sua velocidade em outro meio	66
$\mathbf{x}$	vetor de parametros fotométricos lineares	66
θ	vetor de parametros fotométricos não lineares	66
k_d	parâmetro de reflectância difusa do modelo de Phong	67
k_s	parâmetro de reflectância especular do modelo de Phong	67
n_s	coeficiente de shininess: parâmetro não linear do modelo de Phong	67

Capítulo 1

Introdução

1.1 Motivação

A maneira como o homem interage com o meio em que vive é de fundamental importância para sua sobrevivência, crescimento e manutenção da espécie. Entre os sentidos que permitem perceber o meio em que vivemos encontra-se a visão. Tão importante quanto complexo, este sentido tem sido estudado por diversas áreas da ciência, entre essas a computação.

Compreender o processo de aquisição, formação e interpretação das imagens é uma tarefa não trivial, que envolve áreas como a biologia, psicologia, química e física, entre outras. Embora não seja possível reproduzir fielmente o sistema de visão humano, um conjunto de técnicas têm sido desenvolvidas para interpretar imagens adquiridas por sensores (e.g. câmeras) de forma a permitir algumas funções de reconhecimento e interpretação do ambiente, utilizando recursos computacionais. A área que desenvolve pesquisas neste sentido é chamada de *visão computacional*.

Basicamente, a visão computacional utiliza um ou mais sensores para adquirir imagens do ambiente, similarmente aos olhos do ser humano que captam as imagens do mundo, e um conjunto de técnicas que extraem e manipulam dados que estão contidos nas imagens, ditos características, tais como: forma, cor e textura, entre outras. Estas características, separadas ou integradas, são interpretadas e se transformam em informações que descrevem em parte o ambiente. A este ambiente de onde se adquirem as imagens é dado o nome de *cena*, que pode conter um ou mais objetos.

É possível afirmar, então, que a visão computacional permite a um sistema computacional “reproduzir” o sentido humano da visão, ainda que de forma limitada, e possui diversas aplicações no campo da ciência, tais como auxiliar na navegação autônoma de robôs móveis, inspeção qualitativa e quantitativa de processos industriais, interpretação de imagens

médicas, sistemas de segurança como reconhecimento de faces e impressões digitais.

Uma vez que se dispõe de informações suficientes sobre uma dada cena, é possível reproduzi-la através de métodos que utilizam modelos matemáticos para gerar as formas e cores da cena em uma imagem no computador. Estes métodos se encontram numa área conhecida por Computação Gráfica.

Extrair informações que permitam modelar uma cena real é um objetivo fundamental em aplicações da visão computacional e computação gráfica moderna, tais como produção de ambientes virtuais, entretenimento, treinamento e simulações, efeitos especiais, entre outras. Mas como obter estas informações que permitam reproduzir estas cenas?

Para responder melhor a essa questão, é importante entender quais são estas informações e como utilizá-las no processo de formação de uma imagem. O termo *renderização* – neologismo criado a partir da palavra inglesa *render* – é dado ao processo de criar uma imagem conforme um ou mais modelos matemáticos que representam uma cena. Basicamente, a renderização é obtida a partir de três informações primárias da cena (Figura 1.1):

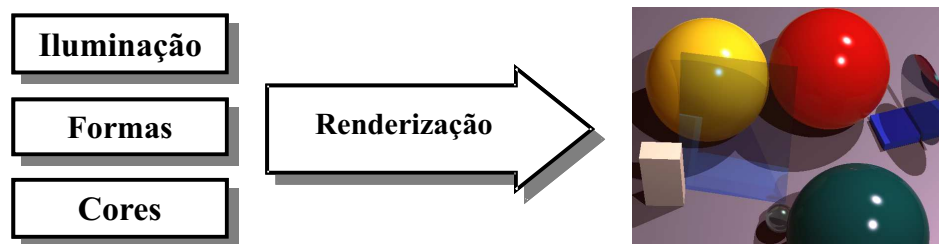


Figura 1.1: Método básico de renderização de uma imagem. Utiliza como entrada os dados sobre iluminação, forma e cor da cena.

Iluminação: pode ser ambiente, onde a luz se mantém praticamente constante em toda a cena, como no caso de um dia nublado onde os raios de sol são dispersos pelas nuvens e a iluminação é aproximadamente a mesma em uma área não sombreada; ou devido a uma ou mais fontes de luz, onde se faz necessário conhecer a posição e intensidade de cada uma delas.

Forma: dada pela geometria e posicionamento de todos os objetos contidos na cena, seja por meio de primitivas como pontos, linhas e curvas, ou de formas tridimensionais mais complexas.

Cor: em cada objeto na cena, tem relação com a luz incidente sobre a superfície do mesmo e o quanto do espectro desta luz é refletida em direção ao observador. Esta relação é uma propriedade de cada material do qual é composta a superfície do objeto, e é

chamada de *reflectância*. Em geral, modelos matemáticos são criados para explicar esse fenômeno e representá-lo de uma forma mais adequada ao uso em computação, e são conhecidos como *funções bidirecionais de distribuição da reflectância - BRDF* (*Bidirectional Reflectance Distribution Function*).

É importante considerar ainda a maneira como a iluminação interage com os objetos, que pode ser primária ou secundária. Na iluminação primária a luz total proveniente de uma ou mais fontes de luz incide diretamente em um ponto numa superfície da cena; e na iluminação secundária a luz total incidente na superfície é resultado da interação da luz primária em outros objetos da cena, também conhecida como iluminação indireta. Essa interação pode ser a luz refletida pela superfície de um objeto da cena em direção a outro objeto e vice-versa, causando um efeito conhecido como inter-reflexão, ou a luz transmitida através de um objeto translúcido em direção a outro objeto da cena.

Devido às diferentes maneiras de interação da luz com os objetos, a renderização pode ser feita de duas formas distintas. A primeira utiliza um modelo de iluminação local, onde a cor de um ponto na cena depende apenas das propriedades materiais do objeto e da iluminação primária, seja ambiente ou de uma fonte de luz específica. A segunda forma utiliza um modelo de iluminação global, onde a cor de um ponto na cena depende das propriedades materiais do objeto, da iluminação primária e da iluminação secundária incidentes, considerando os efeitos de inter-reflexividade e transmissividade resultante da interação da luz com outros objetos da cena. A Figura 1.2 apresenta um exemplo de ambos os casos.

Obter informações que permitam renderizar uma cena a partir de imagens não é tarefa trivial, entretanto vários métodos baseados em imagem foram propostos nos últimos anos para recuperar a forma tridimensional (3D) dos objetos, seus parâmetros de reflectância (BRDF) e informações sobre as fontes de iluminação [Corrêa et al., 2002][Oliveira, 2002].

Esses métodos pertencem a uma área chamada Modelagem e Renderização Baseada em Imagens (*Imaged-Based Modelling and Rendering - IBMR*) que, basicamente, consiste em obter informações do modelo de uma cena utilizando uma ou mais imagens. De acordo com [Marschner e Greenberg, 1998], os principais problemas de modelagem e renderização baseada em imagens são: a geometria inversa, onde se busca obter o modelo 3D de cada objeto da cena; a reflectância inversa, que consiste em obter o modelo de reflectância das superfícies dos objetos da cena; e a iluminação inversa, cujo problema é extrair informações sobre as fontes de luz que iluminam a cena. Entretanto, não estaremos dando enfoque a esse último problema neste trabalho.

Considerando a Figura 1.1, o caso ideal seria aquele em que as imagens são fornecidas como entrada e as informações de iluminação (fontes de luz), forma (geometria) e co-

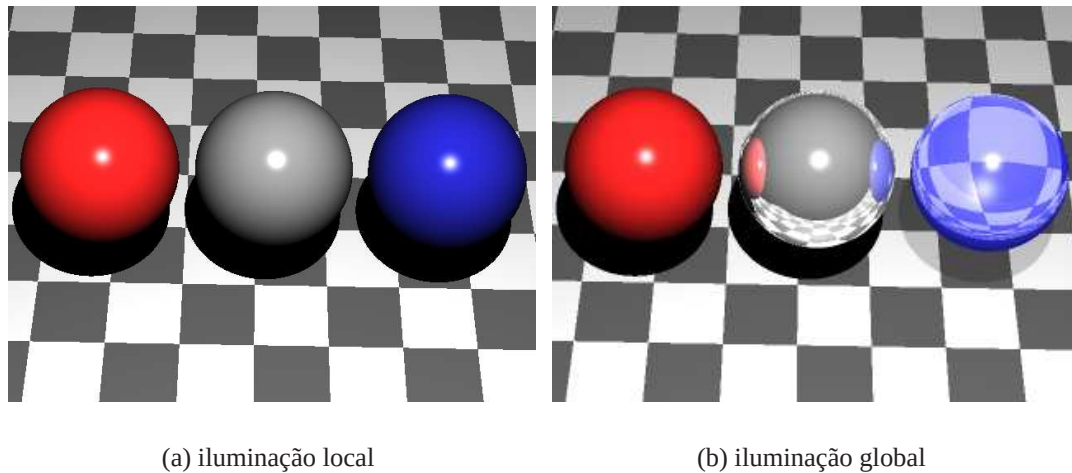


Figura 1.2: Imagens sintetizadas de uma mesma cena com objetos de características difusas, reflexivas e transmissivas, respectivamente. Em (a) a imagem renderizada utilizando um modelo de iluminação local. Observe que os objetos não apresentam inter-reflexão e transmissão, embora dois deles tenham essas características. Em (b) a mesma cena renderizada com um modelo de iluminação global. Neste caso é possível perceber as características de inter-reflexão e transmissão entre os objetos da cena.

res (reflectância) são obtidas como resultado. Entretanto, dentre as principais dificuldades deste método está o fato de que a reflectância e a forma, juntamente com a iluminação, são informações que estão integradas aos dados da imagem. Assim, uma vez conhecidos os dados referentes à forma e à iluminação, é possível obter informações da reflectância e, conseqüentemente, uma vez conhecido algo a respeito da reflectância e iluminação, é possível obter informações a respeito da forma.

1.1.1 Obtendo a Geometria da Cena

Obter o modelo 3D implica em extrair uma representação tridimensional da cena, a partir de uma representação bidimensional que é dada por uma imagem. Como fazer para recuperar a informação que foi perdida no processo de geração da imagem? Uma maneira é utilizar as próprias características da imagem, como linhas paralelas contidas na cena que possam dar a noção de profundidade, e que permitam inferir a transformação projetiva utilizada na geração da imagem [Criminisi et al., 1999]. Entretanto, essa técnica é bastante limitada e aplicada apenas a cenas com características muito peculiares.

Contudo, é possível “imitar” a visão humana e obter informações da geometria 3D de uma cena mais geral. Considere a Figura 1.3, onde cada imagem simula uma visão ins-

uma cena subaquática (Figura 1.4); na segunda a cena possui inter-reflexões ou transmissões entre os objetos da qual é composta (Figura 1.2(b)). Ambos os casos são problemas ainda não resolvidos sob o ponto de vista de visão estéreo, e neste trabalho estamos propondo responder a pergunta acima para o primeiro deles.

Figura 1.4: Exemplo de cenas onde a propagação da luz no meio influencia nas cores da imagem. Em (a) uma cena em uma região com névoa. Em (b) uma cena no fundo do oceano.

(a) Câmera da esquerda

(b) Câmera da direita

Figura 1.5: Par estéreo de exemplo em ambiente subaquático, onde as câmeras estão ortogonais entre si. A busca por pontos correspondentes é dificultada devido à influência do meio subaquático.

Considere o caso em que a propagação da luz no meio impede que a cena possa ser explicada por um modelo de iluminação local. Assim, o problema de se obter a correspondência entre os pontos da cena em um par estéreo torna-se uma tarefa difícil com as técnicas atuais. Tome-se como exemplo o par estéreo da Figura 1.5. Neste caso a cor e intensidade do pixel percebido pela câmera sofrem influência da reflexão da luz no ponto da cena, e da dispersão

e absorção da luz pelo meio no qual a cena está inserida. Como fazer então a busca por pixels correspondentes que não possuem mais a mesma intensidade? Esta é uma das perguntas que esta tese se propõe a responder.

1.1.2 Obtendo a Reflectância da Cena

A retina do olho humano possui células nervosas especiais que captam a intensidade e as cores de uma cena. Os *cones* são as células responsáveis pela informação da cor, e os *bastonetes* pela informação da luminância [Gonzalez e Woods, 2000]. Neste caso, diferentemente da visão estéreo, é difícil “imitar” a natureza humana. Esta é uma questão importante da *reflectância inversa*, que consiste em extrair a reflectância da cena, ou seja, de que forma a luz interage com cada material dos objetos da cena proporcionando a informação de intensidade e cor. Um equipamento utilizado para medir a reflectância de um material é o *reflectômetro*, entretanto isto seria equivalente a utilizar uma fita métrica para obter a geometria da cena, um método demorado e não automatizado. Como resolver então essa questão?

Considere a Equação Geral da Renderização [Marschner e Greenberg, 1998] dada por:

$$\mathcal{F} = \mathcal{K}(\mathcal{H} + \sum \mathcal{G}_i \mathcal{F}_i) \quad (1.1)$$

onde:

- $\mathcal{F}$ descreve a luz refletida em um ponto da cena que se propaga em várias direções, inclusive na direção do sensor, e que é utilizada no processo de formação da imagem;
- $\mathcal{K}$ representa a reflectância do material nesse mesmo ponto da cena;
- $\mathcal{H}$ descreve a luz incidente em cada ponto devido à iluminação direta, e depende da geometria da cena.
- $\sum \mathcal{G}_i \mathcal{F}_i$ é a iluminação indireta no ponto devido à interação da luz com outros objetos da cena, onde $\mathcal{G}$ descreve a fração de luz refletida/transmitida por um outro ponto i da cena;

Dessa forma, renderizar significa resolver a Equação (1.1) para se obter $\mathcal{F}$, dado que são conhecidos os valores das outras variáveis. É possível, então, extrair os valores de $\mathcal{F}$ de um conjunto de imagens obtidas da cena e, conhecidos $\mathcal{G}$ e $\mathcal{H}$, encontrar a reflectância $\mathcal{K}$. Esta é a base da maioria dos métodos de reflectância inversa.

No caso de uma cena explicada por um modelo de iluminação local, como no exemplo da Figura 1.2(a), o termo relativo à contribuição da luz indireta é nulo e este caso configura uma simplificação do caso mais geral, limitando-se a cenas que apresentem essas características.

Contudo, se a cena só pode ser explicada por um modelo de iluminação global, como no exemplo da Figura 1.2(b), encontrar a reflectância inversa torna-se uma tarefa difícil e computacionalmente cara. Os maiores problemas associados a essa tarefa são:

- Definir um modelo global que descreva adequadamente a Equação (1.1), considerando os diversos tipos de superfícies, como espelhos (que possuem alta reflectividade) e objetos translúcidos (que possuem alta transmissividade);
- Adquirir diversas imagens a partir de vários pontos de vista da cena, todas controladas de forma a obter as informações necessárias de $\mathcal{F}$.
- Solucionar o sistema formado pelas equações obtidas na renderização inversa, através de um método que apresente um custo computacional razoável;

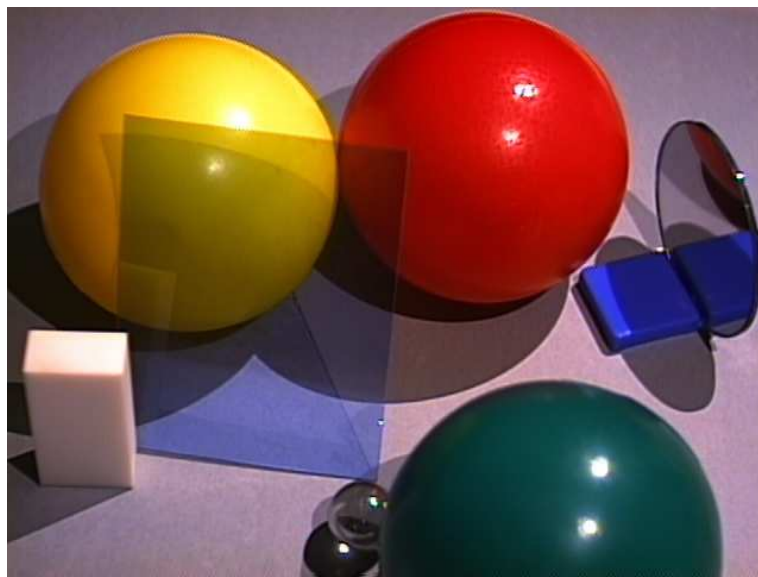


Figura 1.6: Imagem de uma cena fotometricamente complexa, com superfícies reflexivas (espelho) e transmissivas (plástico sobre a esfera amarela).

Como exemplo, considere a Figura 1.6. Como saber neste caso que o espelho é uma superfície que apresenta uma alta reflexão, e não um quadro cuja cor e textura é como vista na imagem? E como saber que o objeto sobre a esfera amarela é translúcido? Uma vez que se consiga responder perguntas como estas, resta saber como obter a reflectância de

uma maneira eficiente, sem que para isso sejam necessárias várias imagens da mesma cena? Existem diversos métodos que possibilitam responder, em maior ou menor grau, algumas destas questões, alguns dos quais são apresentados no Capítulo 2. O fato é que este é um problema muito difícil para o qual não existe ainda uma solução geral, principalmente porque seria necessário resolver um sistema com milhares/milhões de equações não lineares com centenas/milhares de variáveis. Além disso, a aproximação da reflectância real através de uma BRDF e imperfeições do processo de aquisição tornam o sistema mal condicionado. Em função disso, os métodos de renderização atuais possuem limitações quanto à complexidade da cena.

1.2 Descrição do Problema

Esta tese se propõe avançar o estado da arte da solução do problema geral: *como extrair informações de cor e forma de uma cena a partir de imagens, onde um modelo de iluminação local não seja suficiente para explicar as cores da cena.*

Como exposto na Seção 1.1, o problema geral é complexo e de difícil solução quando se pretende obter tanto a geometria quanto a reflectância para uma dada cena globalmente iluminada, pois a modelagem deste problema é uma tarefa não trivial e de solução desconhecida, principalmente devido ao número de ambigüidades existentes. Ainda que se consiga obter uma modelagem para o problema geral, possivelmente seria necessário resolver um sistema com milhares de equações e incógnitas, tornando o método computacionalmente inviável.

Portanto, nesta tese o problema será abordado em dois casos particulares que são complementares em relação ao problema geral, onde as questões relacionadas ao mesmo são tratadas a partir de dois cenários específicos. No primeiro caso tem-se uma cena submersa em água turva da qual pretende-se obter informações sobre a sua geometria. No segundo caso tem-se uma cena com inter-reflexão entre objetos, sendo um altamente reflexivo (um espelho) e outro translúcido (plástico), onde se busca extrair os parâmetros da reflectância de cada objeto.

1.2.1 Cena Subaquática

Neste primeiro cenário, além de outros fatores de menor relevância/impacto, os processos de *absorção* e *dispersão* da luz na água contribuem fortemente para a degradação das imagens em meio subaquático e dificultam o uso eficiente de algoritmos tradicionais da visão

computacional [Born, 1999][Jenkins e White, 1997]. A absorção é o processo físico em que a luz interage com a matéria e é convertida em outras formas de energia, fazendo com que os objetos tenham sua intensidade aparente reduzida quanto maior for a distância entre estes e o sensor utilizado no processo de captura da imagem. A dispersão é o processo físico onde os fótons “colidem” com as partículas do meio e mudam sua direção de propagação, em geral por um ângulo pequeno (*forward scattering*) ou por até 180 graus (*backscattering*). Em consequência deste último fenômeno, os objetos escuros tendem a parecer menos escuros quanto mais se afastam do sensor, devido ao fato de parte da luz refletida na água se propagar na direção do sensor. Combinados, estes dois efeitos causam redução significativa no contraste da imagem, mas podem ser explicados por um modelo aproximado de propagação da luz na água. Como exemplo, a Figura 1.7 apresenta as imagens dos mesmos objetos no ar e dentro de um recipiente com água turva (partículas em suspensão).

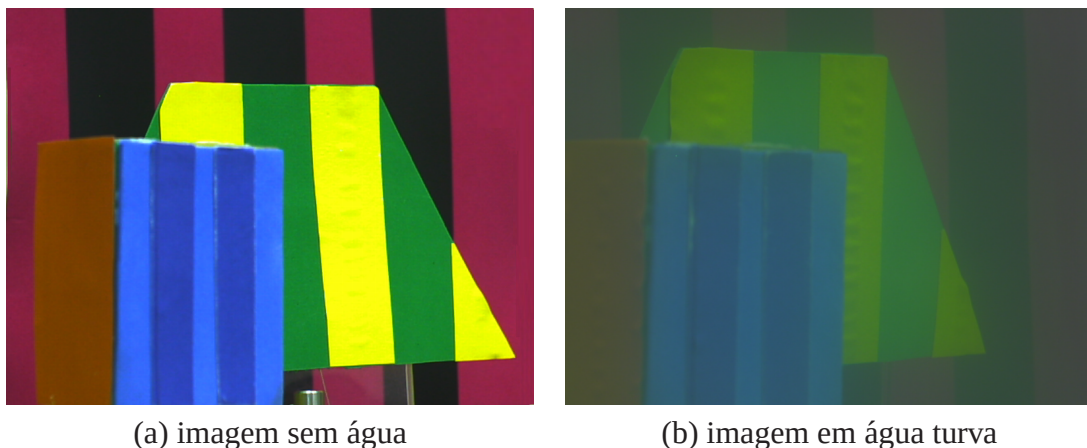


Figura 1.7: Exemplo dos efeitos da redução de contraste na imagem de uma cena em água turva.

Sendo assim, o problema consiste em recuperar informações da estrutura 3D de uma cena subaquática utilizando visão estéreo, considerando que a mesma está inserida em um ambiente que interfere na propagação da luz e impede que a cor de um ponto da cena seja explicado por um modelo de iluminação local. Mais especificamente:

- Dada uma ou mais imagens subaquáticas de entrada, definir um modelo físico que descreva a atenuação e dispersão da luz na água, necessário para restaurar a qualidade da imagem degradada devido às propriedades do meio, considerando a imagem adquirida de uma cena com uma dispersão uniforme do meio e uma região mínima do horizonte (fundo) da cena sem oclusão.
- Dado um par estéreo de imagens, obter um modelo 3D simplificado da cena a partir de

mapas de disparidade, considerando os efeitos da atenuação e dispersão da luz refletida na cena, devido às características próprias de um meio subaquático.

1.2.2 Cena com Inter-reflexões e Objetos Translúcidos

No segundo cenário, tem-se uma cena onde os objetos possuem inter-reflexão, e pelo menos um deles possui uma superfície translúcida, como no exemplo dado na Figura 1.6. Para tratar estes diversos tipos de superfícies, o método de iluminação global precisa ser tal que permita identificar o tipo de material, seja um espelho ou um objeto translúcido.

Como a cena é explicada por um modelo de iluminação global, isto implica que a luz refletida por uma superfície é dada pela soma da luz incidente e das contribuições que recebe da luz refletida e transmitida de outras superfícies. Nesse caso, a componente de luz indireta da Equação (1.1) aumenta com a quantidade de objetos e de inter-reflexões da cena, gerando um sistema complexo que torna lento o processo de obtenção da reflectância.

Portanto, para resolver o problema de se obterem os parâmetros de reflectância do modelo adequados a cada superfície de cada objeto da cena, é importante que o método proposto não somente obtenha resultados foto-realísticos, mas seja computacionalmente eficiente.

Além disso, os métodos de reflectância inversa diferem quanto à quantidade de imagens da cena, tendo como entrada múltiplas imagens [Fournier et al., 1993][Drettakis et al., 1997][Marschner e Greenberg, 1998][Debevec, 1998][Yu et al., 1999][Loscos et al., 1999] ou uma imagem simples [Ikeuchi e Sato, 1991][Boivin e Gagelowicz, 2001]. No primeiro caso, é necessário adquirir diversas imagens de uma cena, cada uma com a câmera em uma posição específica e devidamente calibrada, em muitos casos utilizando um aparato mecânico complicado para mover a câmera em torno da cena. Isto torna o processo experimental lento e caro, e portanto este cenário considera uma única imagem como entrada do método, evitando os problemas operacionais de adquirir várias imagens, e permitindo que o método seja facilmente aplicável a outras cenas.

Em termo mais específicos, o problema consiste em:

- Dada uma imagem de uma cena real globalmente iluminada, seu modelo 3D, e as informações inerentes às fontes de luz, obter os parâmetros do modelo de reflectância (BRDF) das superfícies de cada objeto da cena, de maneira computacionalmente eficiente, considerando objetos reflexivos e ou translúcidos cujas superfícies sejam homogêneas.

1.3 Abordagem de Solução dos Problemas

Considerando os dois cenários definidos na Seção 1.2, esta seção apresenta de maneira geral a solução adotada para resolver o problema de extrair a geometria 3D de uma cena subaquática, o qual será denominado *Estéreo Subaquático*, e outra solução para equacionar o problema de obter a refletância em uma cena com objetos reflexivos e translúcidos de forma eficiente, que será denominado *Refletância Inversa Eficiente*.

1.3.1 Estéreo Subaquático

Algoritmos de estéreo geralmente assumem que as câmeras utilizadas na aquisição das imagens são geométrica e radiometricamente calibradas *a priori*, que as correspondências entre as linhas da imagem são conhecidas e que os pixels das imagens que se encontram na mesma posição 3D na cena possuem a mesma cor. Esta última consideração, em particular, impede que a maioria dos algoritmos de estéreo existentes sejam aplicados a cenas subaquáticas pois, como será apresentado neste trabalho, a cor de cada pixel da imagem destas cenas depende não apenas da iluminação e das propriedades radiométricas das superfícies da cena, mas também das propriedades radiométricas e da distância em espessura de água que a luz precisa atravessar para ir da cena até o sensor. Então, se as imagens em um par estéreo subaquático são adquiridas em diferentes distâncias da cena, pixels correspondentes que deveriam ter a mesma cor terão cores diferentes.

Contudo, se as propriedades de absorção e dispersão do meio forem uniformes e conhecidas *a priori* e as distâncias entre cada sensor e o ponto da cena forem também conhecidas, é possível aplicar uma equação inversa do modelo de propagação da luz na água, de forma a criar uma imagem pré-processada que será uma boa aproximação da imagem real adquirida em um meio sem atenuação e dispersão.

A idéia proposta é, então, (1) obter o modelo e os parâmetros de propagação da luz na água, e (2) utilizar essas informações em um algoritmo de estéreo modificado de modo a corrigir as cores de cada pixel quando estes forem comparados em busca da correspondência. É importante observar que essas correções devem ser efetuadas durante a execução do algoritmo de estéreo, quando a profundidade de cada ponto na cena pode ser calculada e utilizada, juntamente com os parâmetros do modelo de propagação da luz, para corrigir as cores dos pixels [Queiroz-Neto et al., 2004].

1.3.2 Reflectância Inversa Eficiente

Extraír a reflectância de uma cena utilizando uma única imagem e um modelo de iluminação global implica em obter os parâmetros da BRDF utilizando algum método de otimização não linear, como em [Boivin e Gagalowicz, 2001][Boivin e Gagalowicz, 2002]. Isto ocorre devido ao fato de que uma única imagem não provê restrições suficientes para resolver linearmente um sistema formado pela aplicação inversa da Equação (1.1). Nesses casos, um algoritmo de renderização gera uma imagem sintética a cada passo da otimização dos parâmetros da BRDF, que é comparada com a imagem da cena visando reduzir a diferença entre ambas. Uma característica que pode ser explorada é o fato de que a geometria da cena, utilizada a cada renderização, não se modifica. Isto significa que a luz em um ponto segue uma árvore de caminhos fixos que “mapeia” a geometria da iluminação primária e secundária. Portanto, as informações da geometria destes caminhos podem ser guardadas em uma estrutura de dados e utilizadas a cada passo do algoritmo, sem que haja a necessidade de uma nova renderização.

Outra questão interessante está no fato de que as inter-reflexões podem ser explicadas adequadamente se o sistema de equações é formado considerando a BRDF de todos os objetos da cena ao mesmo tempo. Isto ocorre porque, a cada iteração da otimização, a influência da alteração na cor de um objeto em outros objetos é imediatamente considerada no sistema, o que facilita a tarefa de identificar objetos altamente reflexivos, como um espelho, e objetos translúcidos.

Finalmente, tem-se que o sistema formado a partir da renderização possui termos referentes à iluminação local e à iluminação global, tal qual na Equação (1.1). A não linearidade do sistema se deve, principalmente, às componentes de iluminação global. Com os dados do sistema contidos em uma estrutura de dados adequada, é possível manipular este sistema de modo a “simular” um sistema onde todos os parâmetros são tidos como locais e resolver o sistema linearmente para obter uma solução inicial, e em seguida refinar essa solução inicial com um método de otimização não linear.

A idéia proposta é, então, utilizar uma única renderização para obter e armazenar os fatores geométricos constantes da cena, tais como intersecções, raios de reflexão e refração, entre outros, e a partir destes fatores estabelecer um sistema de equações onde as variáveis são os parâmetros da BRDF em todos os pontos interceptados, e a restrição é a cor do pixel da cena real. A partir daí, temos então um sistema de equações não linear que pode ser resolvido através de um método híbrido que envolve parte da solução como um sistema linear, e outra parte como um sistema não linear, tornando a solução do sistema computacionalmente mais eficiente.

1.4 Contribuições

As contribuições que este trabalho oferece à comunidade científica são listados em seguida, e a grande maioria destas contribuições foram apresentadas no *XVII Brazilian Symposium on Computer Graphics and Image Processing* ([Queiroz-Neto et al., 2004]) e na *X IEEE International Conference on Computer Vision* ([Queiroz-Neto et al., 2005]).

1. Método experimental para calibração dos parâmetros de atenuação e dispersão de um modelo de propagação da luz na água, utilizando um conjunto de imagens conhecidas (alvos), sem que haja necessidade de utilização de algum equipamento de medição específico (*e.g.*, turbidímetro);
2. Algoritmo de estéreo para ambientes subaquáticos utilizando a calibração *a priori* dos parâmetros do modelo de propagação da luz na água, onde o meio influencia negativamente os algoritmos de estéreo tradicionais, aos quais os resultados são comparados favoravelmente;
3. Método de restauração de imagens subaquáticas aplicável a cenas dinâmicas, considerando um modelo de propagação da luz na água;
4. Método de reflectância inversa de uma cena com inter-reflexões, contendo simultaneamente objetos com superfícies reflexivas e translúcidas, cenário para o qual não se tem conhecimento de um método semelhante;
5. Método de solução da reflectância inversa utilizando uma única imagem, que é eficiente se comparado a métodos semelhantes que utilizam renderização iterativa.

1.5 Aplicações

Além de avançar em direção à solução do problema geral de modelagem de cenas com iluminação global, este trabalho é motivado também por duas aplicações práticas mais específicas, que são projetos multidisciplinares desenvolvidos com a participação do Laboratório de Visão Computacional e Robótica (VeRLab) do DCC/UFMG.

É importante ressaltar que os resultados desta tese foram obtidos em ambientes controlados visando desenvolver os métodos para aplicações em geral, sendo necessárias futuras adaptações às necessidades específicas de cada um dos projetos citados para permitir suas utilizações.

1.5.1 Projeto Sistema de Contagem Automática de Peixes

As hidrelétricas são uma necessidade no que diz respeito à produção de energia elétrica e ao desenvolvimento econômico e social do nosso país, entretanto a construção destas implica, em geral, na alteração de cursos de rios e, conseqüentemente, na forma de viver dos peixes da região em questão. Como forma de reduzir os impactos causados, faz-se necessária a criação de meios artificiais que permitam aos peixes subir os cursos dos rios para desova e alimentação. Algumas medidas legais foram adotadas em São Paulo e Minas Gerais¹, tornando obrigatória a construção de uma estrutura para a transposição dos peixes, chamada de *mecanismo de transposição de peixes*. A instrumentação dessas estruturas possibilita a obtenção de informações concretas sobre a movimentação das espécies migradoras bem como o desenvolvimento de tecnologia própria para a aplicação nas barragens.

Alguns resultados desta tese foram desenvolvidos no projeto “Utilização Racional de Mecanismos de Transposição de Peixes”, *Transpeixe*, patrocinado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq-521259/2001-0, e pela Companhia Energética do Estado de Minas Gerais, CEMIG. O projeto é aplicado à Usina de Igarapava (Figura 1.8(a)), Rio Grande, SP/MG, onde existe um mecanismo de transposição do tipo escada de ranhura vertical com 320 m de comprimento e 16,5 m de desnível, composta de 80 tanques de 3 m de largura, 3 m de comprimento e 4,5 m de altura, e cada ranhura vertical mede 40 cm de largura (Figura 1.8(b)). Na saída, a escada possui um fosso com um visor (janela de visualização) que permite a contagem dos peixes que alcançam o reservatório, atualmente feita de modo manual por funcionários da Hidrelétrica (Figuras 1.8(c) e 1.8(d)).

A pesquisa consiste em definir um sistema de contagem e identificação automática de peixes, que se encontra dividido em três etapas principais: aquisição de imagens; rastreamento e contagem; e identificação. A fase de aquisição de imagens busca reduzir o impacto de fatores associados ao meio e que influenciam na qualidade das mesmas. Dentre estes fatores estão as condições de iluminação, a atenuação e a dispersão da luz. Neste ponto, os conceitos desenvolvidos nesta tese serão de grande aplicação para a melhoria das imagens a serem utilizadas nas demais etapas do sistema, pois a obtenção de um modelo de propagação da luz na água em conjunto com a geometria da cena são suficientes para restaurar as cores originais da imagem da cena. A fase de rastreamento e contagem consiste em extrair os objetos em movimento na cena, no caso os peixes, onde a cada objeto rastreado é selecionado um conjunto de imagens que será passado à etapa seguinte. A fase de identificação é a etapa responsável pela classificação das espécies, dado um conjunto de imagens [Nery, 2004].

Apesar de não fazer parte do escopo inicial do projeto, o estêreo subaquático pode ainda

¹Leis 9.798 de 7/10/1997 e 12.488 de 9/4/1997, respectivamente.

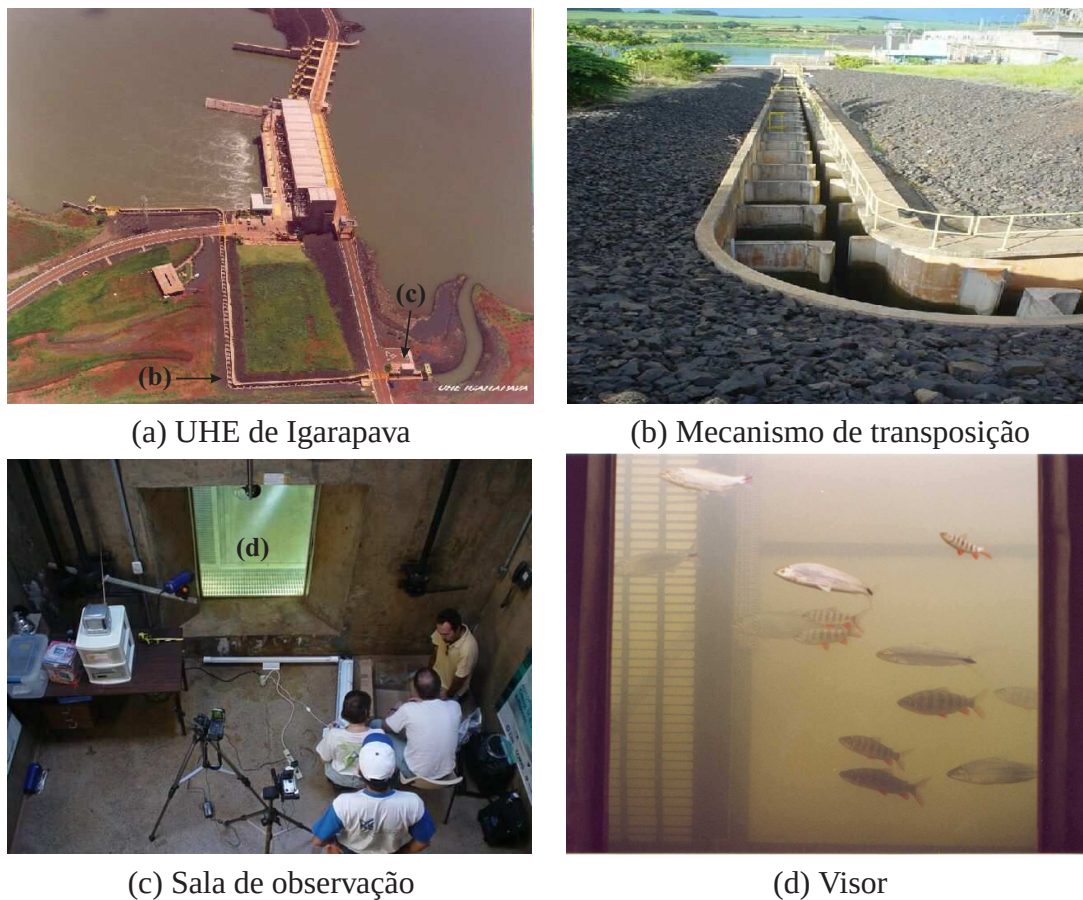


Figura 1.8: (a) A UHE de Igarapava, (b) um detalhe do mecanismo de transposição de peixes, (c) a sala de observação abaixo do nível da água, e em (d) o detalhe da janela de observação.

ser utilizado no processo de identificação e rastreamento dos peixes, utilizando a forma do peixe como uma característica discriminante entre indivíduos de espécies diferentes.

1.5.2 Projeto Estrada Real Digital

Na época áurea do ciclo do ouro no Brasil-colônia, as únicas vias autorizadas de acesso à região das reservas auríferas e diamantíferas da capitania das Minas Gerais, os caminhos reais adquiriram, já a partir da sua abertura, natureza oficial. A circulação de pessoas, mercadorias, ouro e diamante era obrigatoriamente feita por eles. Durante todo o século XVIII, e também em parte do XIX, quando a era mineradora já se fora e os caminhos se tornaram livres e empobrecidos, estes caminhos foram os troncos viários principais do centro-sul do território colonial do Brasil. As vias hoje reunidas sob o nome de Estrada Real foram, assim, fundamentais na história do povoamento e da colonização de vastas regiões do território brasileiro, tornando-se verdadeiros eixos históricos-culturais de construção de parte da nossa



Figura 1.9: Imagens sintéticas da cidade Ouro Preto, que é um dos cenários do jogo.

história [Santos, 2001].

Equipes do Departamento de Ciência da Computação (DCC) e do Centro Cultural da UFMG trabalham no desenvolvimento de um jogo educativo. A aventura se passa na Estrada Real com suas rotas do ouro e diamante, e visa criar um jogo gratuito onde o jogador possa aprender sobre a história, culinária e atrativos turísticos das regiões por onde passa a Estrada Real. As imagens estão sendo capturadas por alunos da graduação, pós-graduação e pesquisadores da UFMG, e a histórica Ouro Preto já virou cenário 3D do jogo (Figura 1.9). O projeto é fruto de uma parceria com as Secretarias de Ciência, Tecnologia e Ensino Superior e de Turismo do estado de Minas Gerais.



Figura 1.10: Fotografia interna da Matriz Nossa Senhora da Conceição, que apresenta diversas esculturas/pinturas a ouro gerando superfícies reflexivas.

O jogo conterà de diversos cenários urbanos e naturais que serão gerados a partir de fotos de cenários reais e digitalizados utilizando um processo de fotogrametria. Neste sentido, o método de renderização inversa para obter os parâmetros da reflectância pode ser utilizado para aumentar o foto-realismo de algumas cenas internas de igrejas (Figura 1.10), como por exemplo cenas com esculturas e pinturas feitas em ouro, que possuem pouca ou nenhuma textura e são bastante reflexivas, vindo a auxiliar os desenvolvedores dos cenários.

1.6 Organização da Tese

A tese está organizada da seguinte forma: o Capítulo 2 apresenta o estado da arte dos trabalhos que se relacionam com o problema a ser tratado, divididos em *Renderização e modelagem baseada em imagens (IBMR)* e *Visão computacional de imagens subaquáticas*; o Capítulo 3 descreve os conceitos principais utilizados e o método proposto para obtenção do modelo 3D simplificado de uma cena subaquática; no Capítulo 4 temos o mesmo referente ao problema da reflectância inversa de uma cena; No Capítulo 5 são dados detalhes de implementação e resultados comparativos dos métodos propostos; e por fim, no Capítulo 6 têm-se as conclusões e os trabalhos futuros.

Capítulo 2

Trabalhos Relacionados

Este capítulo busca apresentar uma visão sobre os conceitos e os principais trabalhos relacionados à modelagem de cenas baseada em imagens. Para uma melhor compreensão o capítulo está dividido em quatro seções principais. A primeira expõe conceitos gerais importantes para o entendimento das demais seções, a segunda se relaciona com o problema da obtenção dos parâmetros de reflectância, um dos cenários específicos a ser tratado. A terceira seção é voltada para trabalhos de obtenção de informações 3D utilizando visão estéreo, e a última seção apresenta trabalhos desenvolvidos em ambientes sub-aquáticos. Essas duas últimas seções tem relação com o segundo cenário em estudo neste trabalho.

2.1 Conceitos de Radiometria e Reflectância

Nesta Seção são apresentados alguns conceitos básicos, importantes ao entendimento das demais Seções. Longe de esgotar o conteúdo pertinente a cada tema, a idéia central é proporcionar um conhecimento direcionado que, embora limitado, será de grande valia ao leitor.

2.1.1 Radiometria

Diversas técnicas em visão computacional e computação gráfica estão relacionadas à maneira como a luz se propaga em dado meio, que é discutida pela *Teoria do Transporte Radiativo* [Chandrasekhar, 1960]. Um caso especial dessa teoria é a forma como a luz incide e reflete em um dado material com características que influenciam a propagação da luz.

Isto nos leva à *Radiometria*, que é a medida de como os materiais refletem a luz em uma parcela do espectro eletromagnético. Na prática, o termo é geralmente limitado à medida do espectro eletromagnético visível, infravermelho e ultravioleta.

A luz é um tipo de energia, conhecida como *energia radiante*. Quando a luz é absorvida por um objeto, a mesma é convertida em outras formas de energia, como o calor, um caso bastante conhecido. A energia radiante por unidade de tempo é o que chamamos de *fluxo radiante*, cuja unidade é o Watt (W). Como exemplo, um feixe laser possui uma certo fluxo radiante de energia.

Em termos de fotometria, o que se mede de uma superfície é o fluxo radiante em uma certa área, que relaciona a quantidade de energia por unidade de tempo irradiando em uma superfície determinada. Essa energia é denominada *irradiância* (E) e medida em W/m^2 , dada por:

$$E(\theta_i, \phi_i) = \frac{d\Phi_i(\theta_i, \phi_i)}{dA}. \quad (2.1)$$

Sendo Φ_i o fluxo radiante incidente na superfície de área A na direção especificada pelas coordenadas de elevação-azimute (θ_i, ϕ_i) .

Para entender melhor os próximos conceitos, considere a Figura 2.1. Pode-se definir a irradiância como sendo um fluxo de luz que incide sobre um elemento de área, vinda de todas as direções de uma semi-esfera, cuja base coincide com uma superfície considerada.

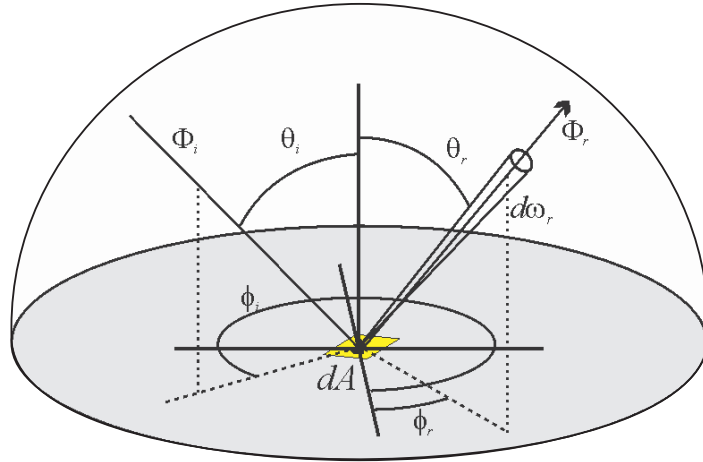


Figura 2.1: Aspectos geométricos da radiometria em uma esfera unitária. Um fluxo de luz incidente Φ_i atinge um elemento de superfície dA pela direção especificada pelas coordenadas de elevação-azimute (θ_i, ϕ_i) , gerando um fluxo de luz refletido Φ_r em um ângulo sólido $d\omega_r$ em uma direção de saída arbitrária especificada pelas coordenadas (θ_r, ϕ_r) . Figura obtida de [Barros, 2004].

A irradiância, após refletir na superfície, é conhecida como *emitância radiante* ou simplesmente *radiância*, que é definida como um fluxo de luz refletido que emana de uma de-

terminada superfície em direção a uma semi-esfera, a partir de um dado ângulo sólido, dada em $W/(m^2 sr)$ ¹ conforme:

$$L(\theta_r, \phi_r, \theta_i, \phi_i) = \frac{d\Phi_i(\theta_r, \phi_r, \theta_i, \phi_i)}{dA \cos\theta_r dw_r}. \quad (2.2)$$

O conceito de ângulo sólido pode ser dado da seguinte forma: quando se refere a um ângulo plano, é possível medir o mesmo em radianos e dizer que um ângulo tem um radiano quando delimita um arco cujo comprimento é igual ao raio da circunferência que contém esse arco; quando se refere a um ângulo sólido, tem-se uma abertura espacial em três dimensões, onde a forma de um cone (vide Figura 2.1) facilita a abstração, e pode-se definir que um ângulo sólido de um *steradian* é aquele que delimita uma área de esfera igual ao quadrado do raio dessa esfera [McCluney, 1994].

A luz incidente em um num ponto de uma dada superfície pode ser refletida, internamente dispersa e re-emitada, absorvida ou transmitida [Carceroni, 2001], o que vem a ser uma característica própria de cada material. Para simplificar o entendimento, considere o caso onde a luz é parcial ou totalmente refletida, e essa característica do material é chamada de *reflectância* e dada pela razão entre a radiância e a irradiância [Oren e Nayar, 1995], ou seja, a fração da luz incidente que é efetivamente refletida pelo ponto. Um modelo de reflectância é utilizado para obter esta razão através de uma equação analítica que depende do tipo de material da superfície. Mais especificamente, a reflectância é uma função com cinco graus de liberdade (depende do comprimento de onda eletromagnético λ) que quantifica a razão da radiância ao longo de cada direção (θ_r, ϕ_r) à irradiância ao longo de cada direção (θ_i, ϕ_i) (Figura 2.1), conforme a Equação (2.3).

$$f(\theta_r, \phi_r, \theta_i, \phi_i, \lambda) = \frac{L(\theta_r, \phi_r, \theta_i, \phi_i)}{E(\theta_i, \phi_i)}. \quad (2.3)$$

Um algoritmo de renderização utiliza o modelo de reflectância para calcular a intensidade da luz refletida em cada ponto da superfície, para todas as projeções destes pontos em um plano de imagem, obtendo assim uma imagem foto-realística da cena, considerando todas as diferentes superfícies de cada objeto [Hearn e Baker, 1994].

A função de reflectância é complexa e difícil de determinar com precisão, mas pode ser representada por um função conhecida como BRDF (*Bidirectional Reflectance Distribution Function*), que é uma descrição simplificada de como uma superfície reflete a luz. Uma vez que a BRDF caracteriza a reflectância de uma superfície em termos dos ângulos de incidência, da luz incidente, e dos ângulos de reflexão, e por ser uma propriedade do material,

¹ *sr* é a unidade *steradian*, que mede o ângulo sólido em uma esfera

não depende de fatores externos como a iluminação.

Quando a BRDF em um dado ponto não muda à medida que a superfície é rotacionada em torno do vetor normal neste ponto, esta superfície é chamada *isotrópica*. Por outro lado, temos superfícies denominadas *anisotrópicas*, que se caracterizam por uma BRDF que é função da orientação da superfície ao longo de sua normal. De acordo com a forma da BRDF, as superfícies podem ser divididas em dois tipos principais [Schlick, 1993]:

- Superfícies difusas: A luz é refletida em todas as direções. No caso ideal, quando a superfície é perfeitamente difusa ou superfície *Lambertiana*, a BRDF se torna uma função constante.
- Superfícies especulares: A luz é refletida somente em uma pequena região em torno da direção ideal de reflexão geométrica. No caso ideal, onde a superfície é perfeitamente especular, a BRDF torna-se uma função que é não nula em uma única direção.

Os modelos de BRDF que atualmente são utilizados em computação gráfica e visão computacional podem ser divididos em duas classes principais: *modelos sintéticos* [Blinn, 1977] [Phong, 1975] que possuem baixo custo computacional e são empíricos, e *modelos analíticos* [Cook e Torrance, 1981][Beckmann e Spizzichino, 1963][Torrance e Sparrow, 1967] com o custo computacional alto, porém mais precisos por se basearem em modelos físicos de interação da luz com a matéria.

Os modelos mais simples em geral são conhecidos pelos nomes de seus autores: o modelo de *Lambert* ou *Lambertiano* [Lambert, 1760] e o modelo de *Phong* [Phong, 1975]. O modelo de Lambert é aplicado à superfícies perfeitamente difusas; a BRDF é simplesmente uma constante para todas as direções, comprimento de onda e polarizações. O modelo de *Phong* expressa a BRDF como uma combinação linear de uma parte difusa e uma parte especular², e introduz um termo cossenoidal com expoente para modelar os efeitos de especularidade, em particular o *specular highlight*, que é a concentração da luz em determinado ponto.

Para representar superfícies mais complexas, outros modelos são utilizados. O modelo de *Cook-Torrance* [Cook e Torrance, 1981] é baseado na física da superfície e possui três componentes principais: o modelo de micro-facetadas da superfície, o termo coeficiente de *Fresnel* que descreve a reflectância, e o termo de rugosidade que parametriza a distribuição das micro-facetadas. O modelo de *Kajiya* [Kajiya, 1985], baseado na teoria de *Kirchoff*, e

²para algoritmos de renderização leva-se em conta uma terceira componente correspondente à luz proveniente do ambiente em que os objetos da cena estão inseridos

o modelo de *Poulin-Fournier* são bastante utilizados em superfícies anisotrópicas. O modelo de *Strauss*, o modelo Programável [Glassner, 1995] e o modelo de *Ward* [Ward, 1992] são modelos baseados em simulação física da estrutura das superfícies e geram uma boa aproximação para aplicações específicas.

2.1.2 Modelos de Iluminação Local e Global

Para que seja possível visualizar uma cena, é necessário que esta seja iluminada por uma ou mais fontes de luz, e que sejam definidas as propriedades de interação entre os objetos e a energia luminosa que incide em cada um deles.

Os modelos de iluminação, ou modelos de reflexão, são uma forma de representar como as imagens são percebidas pelos nossos olhos, e consistem em simular como os objetos refletem a luz. Sem um modelo de iluminação adequado, os objetos tendem a não apresentar características realistas.

Basicamente, os modelos de iluminação possuem dois tipos de representação da reflexão da luz:

- Reflexão Difusa ou Lambertiana: gera uma reflexão da luz igualmente em todas as direções, e depende do parâmetro de reflectância difuso do objeto e das posições das fontes de luz;
- Reflexão Especular: gera uma reflexão mais acentuada em determinados pontos da superfície – de acordo com a cor da luz e não apenas do objeto – e depende do parâmetro de reflectância especular do objeto, da posição da fonte de luz e da posição do observador.

Os modelos de iluminação estão divididos em:

Local a cor de um ponto na cena depende apenas das propriedades materiais do objeto e da iluminação primária, seja ambiente ou de uma fonte de luz específica. Toda informação necessária ao modelo é obtida localmente. Exemplo: modelo de *Phong* [Phong, 1975].

Global a cor de um ponto na cena depende das propriedades materiais do objeto e das iluminações primária e secundária, considerando os efeitos de inter-reflexão e transmissão da luz em outros objetos da cena, também chamada de iluminação indireta e que equivale ao segundo termo da Equação (1.1). Estes modelos são fisicamente mais abrangentes e produzem imagens com um grau de realismo muito próximo da cena real. Exemplo: Radiosidade [Hearn e Baker, 1994].

Como os modelos de iluminação local não são capazes de representar a interação entre diferentes superfícies, fenômenos que podem ocorrer em uma cena, tais como sombras, iluminação indireta refletida de outras superfícies, interação da luz com o meio em que é transmitida (água, ar, etc), simplesmente não são suportados por este modelo. A Figura 1.2 apresentou um exemplo comparativo da mesma cena em um modelo de iluminação local e global.

Entre os principais métodos de iluminação global encontram-se o método do *Traçador de Raios*, mais conhecido como *Ray Tracing*, e o método da *Radiosidade*.

O *Ray Tracing* é um método de renderização que calcula a imagem da cena simulando a forma como os raios de luz percorrem o seu caminho no mundo real. Os raios de luz são emitidos a partir de uma fonte de luz, de acordo com as características desta fonte de luz, e iluminam os objetos da cena. A luz é então refletida por estes objetos e pode ainda passar através de objetos transparentes e semi-transparentes. Esta luz refletida então chega a lente de uma câmera.

No *Ray Tracing*, um raio é lançado a partir de cada pixel no plano de imagem da câmera em direção à cena, seguindo o caminho inverso da luz. A cada intersecção deste raio com uma superfície da cena o mesmo é refletido e ou transmitido em direção a outras superfícies e assim por diante, “ricocheteando” na cena até ir em direção a uma fonte de luz, se perder no espaço ou atingir um número máximo de interações pré-estabelecido. Esses raios permitem construir uma árvore binária com a qual, conhecidas as intensidades e posições das fontes de luz, é possível calcular e armazenar em cada nó a contribuição daquela superfície interceptada para o valor final de intensidade do pixel. Por se tratar do modelo a ser utilizado nesta tese, maiores detalhes são dados no Capítulo 4, quando será apresentado o método proposto para a reflectância inversa.

A Radiosidade é um método de renderização baseado em uma análise detalhada das reflexões da luz em superfícies difusas e é baseado em um modelo de transferência de energia. A superfície é dividida em elementos (pequenas partes), todos com reflexão considerada difusa, e em cada superfície do modelo a quantidade de energia é a soma da energia que o elemento de uma superfície emite internamente e da quantidade de energia refletida pelos elementos das superfícies que interagem com a primeira, conforme a Equação (2.4) [Hearn e Baker, 1994]. Observe que essa equação é, basicamente, uma solução discreta e aproximada da *Equação Geral da Renderização* (Equação (1.1)), o que explica a semelhança entre os termos de ambas. Considerando todos os elementos de cada superfície, o método estabelece um sistema de equações lineares cuja solução é a distribuição da luz na superfície.

$$B_i = Be_i + \rho_i \sum_j F_{ij} B_j, \quad (2.4)$$

onde:

B_i é a Radiosidade no elemento i , considerando a luz direta e indireta (global);

Be_i é a energia emitida internamente pelo elemento i . Este termo existe quando o elemento emite luz própria, como uma lâmpada;

ρ_i é a BRDF difusa do elemento i . Observe que neste caso a superfície do elemento é considerada lambertiana;

F_{ij} é o *Fator de Forma*, que é uma função que depende da geometria das superfícies e que determina a quantidade da Radiosidade do elemento j que chega ao elemento i .

Ambos os métodos de iluminação global apresentados possuem vantagens e desvantagens. O *Ray Tracing* não representa bem a interação luminosa difusa, criando imagens altamente reflexivas, além disto pode se tornar lento para cenas complexas e de altas resoluções. Contudo, tem a vantagem de permitir o cálculo de cada pixel isoladamente, e de ser aplicado a superfícies especulares, com forte inter-reflexão e materiais translúcidos, esta última característica se configura uma das limitações da Radiosidade, que em contrapartida tem um custo computacional reduzido, se comparado ao *Ray Tracing*.

2.2 Renderização e Modelagem Baseada em Imagens

A renderização e modelagem baseada em imagens têm sido largamente utilizadas para recuperar a forma tridimensional (3D) dos objetos, seus parâmetros de reflectância (BRDF) e informações sobre as fontes de iluminação. Neste contexto, diversos autores têm contribuído para a resolução destes problemas, utilizando vários métodos e experimentações com diferenças em termos de técnicas aplicadas e resultados obtidos, que vão desde algoritmos e métodos de solução de sistemas a complicados equipamentos e dispositivos. Os trabalhos de maior interesse nesta tese são aqueles que consistem em obter um modelo de reflectância que explique uma ou mais superfícies de um ou mais objetos em uma mesma cena.

A Tabela 2.1 apresenta um visão geral dos principais trabalhos. É dado o resultado obtido em cada trabalho, a partir do modelo recuperado (forma e ou reflectância). É possível, ainda,

verificar se o método utilizado é local, global ou se faz uso de algum dispositivo especial, como um aparato mecânico, um scanner ou etc. Além disso, a tabela aponta a necessidade de imagens de entrada da cena, que podem ser múltiplas ou simplesmente uma única imagem.

<i>Método Utilizado</i>	<i>Iluminação</i>		<i>Imagens da Cena</i>		<i>Disp. Esp.</i>	<i>Recupera</i>		
	<i>Local</i>	<i>Global</i>	<i>Única</i>	<i>Múltip.</i>		<i>3D</i>	ρ_d	<i>BRDF</i>
[Tagare e deFigueiredo, 1991] [Woodham et al., 1991]	•			•		•	•	
[Sato et al., 2000] [Cipolla e Boyer, 1998] [Hertzmann e Seitz, 2003]	•			•		•		•
[Wong et al., 1997] [Nishino et al., 2001]	•			•				•
[Marschner e Greenberg, 1998] [Debevec, 1998] [Yu et al., 1999]		•		•				•
[Ikeuchi e Sato, 1991] [Tominaga e Tanaka, 2000] [Dror et al., 2001]	•		•				•	
[Fournier et al., 1993]		•	•				•	
[Drettakis et al., 1997]		•	•			•	•	
[Boivin e Gagalowicz, 2001] [Boivin e Gagalowicz, 2002]		•	•					•
[Debevec et al., 2000]		•		•	•			•
[Ward, 1992]	•			•	•			•
[Baribeau et al., 1992]	•			•	•		•	
[Miyazaki et al., 2003b]	•		•		•	•		•
[Matusik et al., 2003]	•			•	•	•		•
[Gardner et al., 2003]	•		•		•			•

Tabela 2.1: Comparação entre os diversos métodos de IBMR. Os trabalhos estão divididos quanto ao método: iluminação local ou global, com ou sem o uso de dispositivos especiais; quanto à entrada: única ou múltiplas imagens; e quanto ao resultado obtido: geometria (3D), reflectância difusa (ρ_d) e reflectância difusa e especular (BRDF).

A forma de classificação dos trabalhos desenvolvidos na área está dividida em métodos com iluminação local, métodos com iluminação global, e métodos com uso de dispositivos especiais, esses últimos aplicados à uma cena com iluminação local.

2.2.1 Métodos com Iluminação Local

Os métodos que se baseiam em um modelo de iluminação local utilizam uma simplificação da Equação de Renderização (Equação (1.1)), onde é considerada apenas a iluminação primária

incidente no ponto, resultando na reflectância dada unicamente pela Equação (2.3). Essa equação é invertida de forma que, conhecendo a radiância a partir de informações da imagem obtida da cena, e calculando a irradiância localmente a partir das fontes de luz e da geometria da cena, obtém-se a BRDF. Esse é o caso mais simples da renderização inversa, e é a base dos métodos com iluminação local.

Em seguida são apresentados alguns trabalhos, utilizando múltiplas ou uma única imagem. Em todos eles, a grande desvantagem está no fato de que não são suficientes para adquirir a BRDF em uma cena mais complexa, como no caso da Figura 1.2(b), o que torna os métodos bastante limitados no que diz respeito à complexidade das cenas, diferente do método proposto nesta tese, que utiliza um modelo de iluminação global.

2.2.1.1 Utilizando Múltiplas Imagens

O uso de múltiplas imagens é necessário para se obter as variações da radiância de um mesmo ponto da cena, em função da posição do observador ou das posições das fontes de luz. Com isso os métodos podem “perceber” a componente especular da reflectância.

Diversos trabalhos são baseados no uso do Método de Estéreo Fotométrico (*Photometric Stereo Method - PSM*), que utiliza múltiplas imagens obtidas com a mesma geometria mas sob condições de iluminação diferentes. Um sistema de equações é formado considerando as restrições extraídas da intensidade de cor das imagens e da diferença nas direções de incidência e reflexão da luz. A resolução deste sistema permite recuperar a geometria e a BRDF, em geral lambertiana [Woodham et al., 1991][Schlüns, 1997][Cipolla e Boyer, 1998]. Em [Tagare e deFigueiredo, 1991] é demonstrado que três fontes de iluminação distintas são suficientes para a inversão do sistema de equações.

Os métodos que utilizam o PSM possuem um custo computacional baixo, se comparado com métodos que utilizam uma única imagem, e embora não necessitem de várias posições geométricas da mesma cena, conseguem obter o modelo 3D da cena de forma bastante precisa, mas têm como desvantagem o fato de que as propriedades de reflectância da cena devem ser explicadas por modelos de iluminação predominantemente lambertianos ou modelos relativamente simples.

Outros métodos consideram que a geometria da cena é conhecida ou obtida por meio de algum método ou equipamento específico. Em [Sato et al., 2000] é proposto um sistema para estimar o modelo de iluminação utilizando um modelo 3D obtido por uma série de imagens de profundidade através de um *scanner*, uma imagem obtida de uma câmera, e conhecida a posição da fonte de luz. A partir dos ângulos de direção da energia incidente e refletida, obtidas com a imagens de profundidade, estabelecem-se quais pontos não possuem

especularidade e separam-se as componentes difusa e especular da reflectância, calculando-as em sistemas de equações separadamente, de forma a se obterem os parâmetros do modelo de iluminação. A principal limitação desse método é o fato de que é aplicado somente a um objeto, o que inviabiliza aplicações onde a cena possua mais de um objeto.

Em [Nishino et al., 2001] assume-se que a geometria da cena e os parâmetros da câmera são conhecidos. O método utiliza um conjunto de imagens e considera que o menor valor de intensidade das projeções visíveis de cada ponto, em todas as imagens, se refere exclusivamente à fração difusa da reflectância e, por subtração, obtém as imagens residuais que, para cada posição da câmera, contém a fração especular da reflectância. Com tais informações é obtido um campo de iluminação estimado, posteriormente utilizado em um sistema de equações para extrair os parâmetros de reflectância difuso e especular do modelo de iluminação simplificado de Torrance-Sparrow [Torrance e Sparrow, 1967]. Sua principal limitação é o fato de que cada ponto na cena deve aparecer, no mínimo, em duas imagens, sendo ao menos uma destas imagens capaz de exprimir a especularidade.

É possível, ainda, fazer uma boa estimativa da BRDF sem o conhecimento da geometria, como em [Wong et al., 1997], que descreve um método para recuperar a reflectância, considerando um conjunto de micro-faces, cada qual com sua própria BRDF, estimada a partir de um conjunto de imagens adquiridas em diferentes pontos de vista e condições de iluminação. A BRDF é dada como sendo a razão entre a intensidade do pixel pela intensidade da fonte de luz. O método é aplicado para re-iluminar a cena com novas fontes de luz e tem como vantagem não necessitar do conhecimento prévio do modelo 3D. Contudo, necessita de um sistema que permita modificar de maneira controlada a iluminação da cena, sem o qual não seria possível obter seus resultados.

Mais recentemente, [Hertzmann e Seitz, 2003] apresenta uma técnica que utiliza um objeto uniforme como referência para estimar a BRDF de outro objeto que possua o mesmo tipo de material, ou material semelhante. O método consiste em colocar o objeto de referência sob as mesmas condições de iluminação do objeto da cena. Com base no princípio de que dois pontos com a mesma normal e mesmo tipo de material, sob as mesmas condições de iluminação, possuem características semelhantes de reflectância, o método visa extrair um mapa de normais a cada uma das várias imagens adquiridas em posições diferentes. Esses mapas são utilizados para recuperar a forma e a reflectância do objeto. É um método aproximado e muito limitado, pois sua aplicação necessita sempre de um outro objeto de referência com as mesmas características materiais do objeto da cena que se deseja obter a reflectância.

2.2.1.2 Utilizando Uma Imagem

O uso de uma única imagem evita o custo operacional envolvido no uso de múltiplas imagens, seja pelos equipamentos ou pelas minúcias de posicionamento, controle e calibração da câmera em cada posição pré-definida. Isso facilita o processo experimental, entretanto a limitação de restrições dificulta a separação das componentes difusa e especular da reflectância, e em geral permitem a aquisição de apenas uma aproximação da BRDF.

Um dos primeiros trabalhos foi apresentado em [Ikeuchi e Sato, 1991], através de um método para obter a reflectância a partir de uma imagem de profundidade³ e uma imagem da cena. O método utiliza uma imagem de profundidade para compor os elementos geométricos necessários a um sistema de equações cujas restrições são dadas pela imagem da câmera e as incógnitas são os parâmetros do modelo de iluminação local de Torrance-Sparrow [Torrance e Sparrow, 1967]. Este método faz a separação das componentes difusa e especular com base em limiares determinados por uma metodologia de aproximação sucessiva, utilizando também parâmetros para determinar se um dado ponto é resultado de uma inter-reflexão ou sombreamento. Um sistema não linear resultante é resolvido por mínimos quadrados. A principal limitação do método é assumir uma reflectância difusa e uniforme em toda a superfície do objeto.

A utilização de uma única imagem torna a tarefa de determinar o coeficiente especular da reflectância complicada para estes métodos, entretanto em [Tominaga e Tanaka, 2000] é proposto um método para estimar os parâmetros de reflectância difusa e especular utilizando o modelo de iluminação local de Phong em um objeto de superfície homogênea e geometria simples (um cilindro). Por meio de uma análise do histograma de imagens, o método considera que a máxima intensidade de cor no objeto está relacionada à componente especular da reflectância, pois apresenta o maior brilho na imagem, e subtraindo essa intensidade máxima das intensidades de áreas na imagem que não apresentam esse brilho máximo, consegue fazer a separação das componentes difusa e especular da reflectância. O método tem grandes restrições, pois utiliza um único objeto de geometria simples (um cilindro) para o qual outros objetos devem ser aproximados.

Um trabalho que difere bastante dos outros é o dado em [Dror et al., 2001], que propõe o uso de uma base de imagens com objetos de reflectância conhecida e, por comparação, tenta recuperar e classificar a reflectância da cena a partir de uma única imagem sob condições de iluminação desconhecidas. O método extrai relações estatísticas entre a reflectância e as características da cena (distribuição dos coeficientes de *wavelets*) e as utiliza como discrimi-

³Neste tipo de imagem, os valores contidos em cada pixel referem-se à distância entre o ponto na cena e o sensor.

nantes em em um classificador. O trabalho é limitado no sentido em que permite cenas com apenas um objeto de superfície homogênea.

2.2.2 Métodos com iluminação global

Os métodos de iluminação global possuem a tarefa principal de extrair a BRDF de cenas mais complexas, que não podem ser explicadas por um modelo de iluminação local. Os métodos podem ser divididos em: (i) baseados na Radiosidade Inversa, (ii) baseados em *Ray Tracing*, e (iii) de formulação geral.

2.2.2.1 Radiosidade Inversa

Considere a Equação (2.4). A Radiosidade B_i pode ser extraída por diversas imagens adquiridas em pontos de vista diferentes; a emitância Be_i é obtida a partir das intensidades e posições das fontes de luz; e os fatores de forma F_{ij} podem ser calculados a partir da geometria, conhecida *a priori*. Neste caso, adquirir a reflectância se reduz a inverter a equação da Radiosidade e obter a reflectância difusa, dada por:

$$\rho_i = \frac{B_i - Be_i}{\sum_j F_{ij} B_j}. \quad (2.5)$$

O método de Radiosidade inversa consiste, então, em obter os dados necessários e resolver o sistema de equações formado em todos os elementos das superfícies da cena, a partir da Equação (2.5). Os métodos que utilizam a Radiosidade inversa necessitam de múltiplas imagens, obtidas de forma controlada, e que permitam visualizar diversas vistas da cena.

A complexidade do sistema formado pela Radiosidade inversa resulta da complexidade da BRDF das superfícies da cena. Os principais problemas dos métodos baseados em Radiosidade inversa são:

- A impossibilidade de utilizar os métodos para BRDF's complexas, como em cenas com superfícies translúcidas (Figura 1.6), pois neste caso a equação da Radiosidade não é suficiente para explicar a energia que se atravessa uma superfície e atinge outra.
- A preparação das imagens é cara e dispendiosa, ou seja, adquirir diversas imagens de uma cena, cada uma em uma posição específica, devidamente calibrada, é uma tarefa nem sempre viável. Considere como exemplo o caso da aplicação *Projeto Estrada Real Digital* apresentado na Seção 1.5.2. Nesse caso seria impraticável adquirir imagens internas de igrejas que não podem dispor de mais que poucas horas para a preparação,

posicionamento, calibração e aquisição das imagens, devido sua condição de visitação pública.

Um dos primeiros métodos de Radiosidade inversa, proposto por [Fournier et al., 1993], assume que a reflectância é constante em cada elemento da superfície e utiliza a seguinte heurística: a reflectância de cada elemento é dada pela reflectância média da superfície multiplicada pela razão entre a Radiosidade do elemento e a Radiosidade média da superfície. Os valores médios de reflectância são obtidos diretamente das imagens. O método falha quando a cena possui superfícies especulares, além de obter somente uma estimativa média da reflectância difusa.

Uma extensão do trabalho de Fournier é apresentada em [Drettakis et al., 1997]. Nesse caso, a reflectância de cada elemento é obtida a partir do mapeamento dos pontos de cada imagem na cena, utilizando para isto a projeção perspectiva inversa de cada câmera. Além disso, um algoritmo de Radiosidade hierárquica é utilizado de forma a aumentar o realismo da imagem sintética gerada. O método possui as mesmas limitações fundamentais do método de Fournier.

Quando se trata de uma cena mais complexa, que contenha superfícies com características especulares, um dos principais trabalhos é apresentado em [Yu et al., 1999]. O método consiste em se obterem diversas imagens de cada objeto, e a partir de cada imagem obter a energia de cada elemento da superfície. Utilizando uma coleção de imagens de um mesmo ponto na cena, o método considera que a imagem com menor intensidade corresponde à parcela difusa da reflectância, e a que possui a maior intensidade corresponde à soma da parcela difusa com a parcela especular. Uma subtração simples entre essas duas quantidades permite separar as duas componentes e tratá-las separadamente. O método usa mínimos quadrados para resolver um sistema de equações e estimar a parcela de especularidade iterativamente, tomando como partida a especularidade inicial adquirida do conjunto de imagens. A reflectância difusa é obtida resolvendo um sistema linear. O trabalho não apresenta prova teórica de que o algoritmo converge. Além disso, o método necessita de uma coleção de imagens representando todas as faces e fontes de luz primárias da cena, o que envolve um delicado processo de posicionamento e calibração de uma ou mais câmeras.

Outra abordagem é dada em [Loscos et al., 1999]. Em um primeiro momento, a geometria da cena é construída a partir de diversas imagens em vários pontos de vista. Em seguida, imagens adicionais são adquiridas do mesmo ponto de vista da cena para diversas posições da fonte de luz. Dessa forma, pode-se verificar quais imagens possuem a maior intensidade e a menor intensidade para um mesmo ponto da cena, o que permite estimar valores iniciais para as reflectâncias difusa e especular. Esta estimativa é utilizada para inicializar um

sistema obtido pela Radiosidade inversa. Este sistema é então otimizado iterativamente e comparado com os valores da cena até o método convergir para um erro mínimo.

2.2.2.2 Métodos Baseados em *Ray Tracing*

Nestes métodos, o *Ray Tracing* é utilizado como modelo de iluminação global para determinar a geometria relacionada com as iluminações primária e secundária. Diferente da Radiosidade, o *Ray Tracing* não tem uma relação matemática semelhante a da Equação (1.1), embora siga os mesmos princípios físicos. Em função disso, obter uma equação inversa que considere todos os caminhos percorridos pelos raios lançados pelo método não é uma tarefa trivial. A iluminação global inversa é obtida, então, a partir da árvore do *Ray Tracing*, somente após o raio lançado concluir seu percurso.

Em [Marschner e Greenberg, 1998], além de uma extensa e completa abordagem sobre renderização inversa, é utilizado como modelo de iluminação global o *Montecarlo Ray Tracing*. Seu método considera conhecidas a iluminação e geometria da cena, e utiliza várias imagens para montar e resolver um sistema de equações inversas da renderização. Além disso, o mesmo trabalho apresenta um método baseado em um aparato mecânico e filtros polarizadores que permitem identificar as regiões de especularidade da cena, separando a componente difusa e obtendo a imagem formada somente pela sua cor difusa, chamada de *mapa de textura*. A maior limitação do método é que o mesmo se aplica somente a um único material, que precisa ser formado a partir de uma superfície curva e homogênea.

2.2.2.3 Métodos de Formulação Geral

Estes métodos buscam obter a BRDF sem necessariamente estabelecerem um modelo de iluminação global inversa em particular. Geralmente, o modelo de iluminação é utilizado de forma direta para renderizar uma imagem da cena a partir de uma dada BRDF inicial. A imagem renderizada pode então ser comparada à imagem original da cena e, por meio de um método específico, modificar iterativamente os parâmetros da BRDF até alcançar um valor ótimo que minimize o erro entre duas imagens.

Um trabalho interessante que usa esta abordagem foi proposto em [Debevec, 1998]. Nele o autor apresenta um método que mistura a medida manual da radiância da cena e o uso da Radiosidade como modelo de iluminação global, como forma de adicionar novos objetos sintéticos ao modelo da cena real. Para inserir estes objetos, faz-se necessário considerar as inter-reflexões da cena e obter a BRDF das superfícies na cena que influenciam na cor do novo objeto. A iluminação do novo objeto sintético é obtida através de um espelho esférico

que que capta as fontes de luz, e que fica na posição do objeto que se deseja inserir. O método utiliza um modelo 3D parcial da cena, chamado de cena local, e manualmente calcula os parâmetros de reflectância de todos os objetos da cena. Neste caso o usuário modifica cada parâmetro interativamente até obter a melhor imagem renderizada possível. A partir destes parâmetros escolhidos pelo usuário como entrada, um procedimento automático é utilizado para refinar estes valores, comparando iterativamente a imagem renderizada com a imagem da cena.

Em [Boivin e Gagalowicz, 2001] [Boivin e Gagalowicz, 2002] é apresentado um método que requer como entrada o modelo 3D da cena, as informações de posição e intensidades das fontes de luz e uma única imagem da cena. Similarmente ao método anterior, utiliza a Radiosidade para gerar uma imagem sintética a partir de uma BRDF inicial e utilizá-la para fins de comparação. Devido ao fato de usar o modelo de reflectância de Ward [Ward, 1992], é aplicado a superfícies tanto isotrópicas quanto anisotrópicas. O algoritmo apresentado para a renderização inversa é iterativo e hierárquico. É hierárquico pois faz considerações sobre os tipos de reflectância das superfícies da cena: perfeitamente difusa, perfeitamente especular, difusa e especular, e assim por diante, resolvendo um caso de cada vez. É iterativo na medida em que utiliza um método de otimização para obter os parâmetros de superfícies não determinadas na etapa hierárquica, tomando como base a redução do erro entre a imagem da cena e a imagem renderizada pelo método da Radiosidade.

Este método tem uma estreita relação com o método de reflectância inversa proposto nesta tese, pois as entradas consideradas são as mesmas: modelo 3D da cena, informações de posição e intensidades das fontes de luz e uma única imagem da cena. Apesar dos bons resultados, em particular para superfícies anisotrópicas devido ao uso de um modelo de iluminação local mais robusto, o método dado em [Boivin e Gagalowicz, 2001] [Boivin e Gagalowicz, 2002] apresenta algumas deficiências que são resolvidas nesta tese, tais como:

- Não permite superfícies translúcidas na cena, sendo mais indicado para os casos de inter-reflexões difusas.
- Necessita fazer uma renderização da imagem a cada passo iterativo do algoritmo, o que é uma tarefa computacionalmente cara.
- Os parâmetros de reflectância são estimados separadamente para cada objeto, o que pode gerar resultados ruins para cenas com alta inter-reflexão entre os objetos.

2.2.3 Utilizando Dispositivos Especiais

Apesar de obterem os dados do modelo com base em imagens, alguns métodos utilizam dispositivos que auxiliam, direta ou indiretamente, na extração destes dados. Exemplos destes dispositivos são equipamentos de medição de forma, como *scanners* 3D, sensores que auxiliem na extração da BRDF, ou aparatos mecânicos para a geração de imagens rigorosamente controladas em termos de posição de câmeras e de fontes de luz.

Os métodos apresentados nesta seção têm dois problemas em comum: o primeiro consiste no fato de que os dispositivos utilizados possuem um custo considerável, seja em valores financeiros seja quanto à dificuldade de sua implementação; e o segundo se refere ao fato de que estes métodos são limitados a condições e considerações muito específicas da cena, sendo difícil generalizá-los para cenas mais gerais.

O método dado em [Ward, 1992] propõe medir diretamente a reflectância de uma superfície, utilizando um dispositivo de medição (gonioreflectômetro) desenvolvido especificamente para estimar os cinco parâmetros de um dado modelo de reflectância conhecido como *Modelo de Ward*. O equipamento utilizado por Ward é limitado ao uso para o modelo que ele mesmo desenvolveu.

Outro trabalho que utiliza um dispositivo específico é dado em [Baribeau et al., 1992], onde são medidos os parâmetros de reflectância de diversos objetos em uma mesma cena. Um laser policromático é utilizado para obter imagens de profundidade e de intensidade da cena. O método permite extrair a reflectância difusa do material, bem como os parâmetros de rugosidade de cada superfície, entretanto é limitado a objetos com superfícies lambertianas.

Em [Debevec et al., 2000] é apresentado um método para adquirir um campo de reflectância de uma face humana, a partir de um aparato mecânico que gira uma fonte de luz horizontal e verticalmente em torno da pessoa. As imagens são capturadas por uma câmera fixa. A BRDF é dada então pela parcela de luz refletida em cada direção (θ, ϕ) , gerando um campo ou mapa de reflectância. Esse método permite separar as componentes difusa e especular, facilitando a criação de um mapa de textura, e faz a interpolação de novos pontos de vista a partir dos pontos já conhecidos.

No trabalho de [Matusik et al., 2002] é desenvolvido um sistema que consiste de um scanner 3D baseado em imagens, composto por diversas cameras, luzes, monitores e uma mesa giratória, que é capaz de adquirir simultaneamente a forma aproximada e os parâmetros de reflectância e de refração de um objeto. Em uma extensão deste trabalho, o autor utiliza a mesma abordagem para medir a reflectância aproximada em vários tipos de materiais e superfícies [Matusik et al., 2003].

Em [Miyazaki et al., 2003a] os autores propuseram um método onde utilizam um filtro

polarizador que permite determinar as normais de superfícies, inclusive transparentes, e separar as componentes difusa e especular da reflectância. Com isso é possível obter o mapa de textura da cena e a distribuição da iluminação, tudo isso em uma abordagem integrada. O método não responde bem a inter-reflexões e é limitado à utilização em um único objeto.

Um mecanismo ainda mais especificado é apresentado em [Gardner et al., 2003], que utiliza uma fonte de luz que se movimenta linearmente, similar a um scanner, e cujo reflexo é usado para estimar a variação espacial das propriedades de reflectância de uma superfície. O método é limitado a superfícies planas, devido às características físicas do dispositivo.

2.3 Conceitos de Visão Estéreo

A visão estéreo, ou estéreo computacional, é um método utilizado para extrair informações de profundidade e da estrutura 3D de uma cena, permitindo reconstruir as coordenadas tridimensionais de pontos desta a partir de suas respectivas imagens em pelo menos duas câmeras [Trucco e Murino, 2000]. É dito estéreo denso quando o objetivo é encontrar para *todos* os pontos visíveis em uma imagem seus respectivos correspondentes em outra imagem. Os primeiros trabalhos nessa linha datam do final dos anos 70 e início dos anos 80 e em [Barnard e Fischler, 1982] é apresentada uma revisão com os fundamentos da reconstrução estéreo e as principais abordagens da época. O trabalho de [Koschan, 1993] discorre sobre a evolução das técnicas até 1993, incluindo pesquisas voltadas para o problema de oclusões, estéreo dinâmico, e implementações de estéreo em tempo real. Mais recentemente, [Brown et al., 2003] resumiram os novos avanços na área, apresentando e comparando os principais métodos utilizados.

A base fundamental do estéreo é o fato de que um ponto na cena, que possui uma localização tridimensional no mundo (P), é projetado em um par de imagens obtido por duas câmeras que possuem entre seus centros de projeção (O_L, O_R) uma distância conhecida (T), chamada de *baseline* (Figura 2.2). Como resultado, conhecendo a posição da projeção do ponto em cada imagem (p, p'), é possível determinar a localização 3D do ponto na cena.

Os problemas a serem resolvidos na visão estéreo são: calibração, correspondência e reconstrução. O primeiro consiste em determinar as posições e orientações relativas das câmeras (parâmetros extrínsecos) e informações sobre as distâncias focais, centros ópticos e distorção das lentes (parâmetros intrínsecos). Este é um problema bem conhecido e tratado e, neste trabalho, considerado resolvido. Trabalhos recentes em calibração de câmera podem ser obtidos em [Faugeras e Luong, 2001] e [Hartley e Zisserman, 2000], e portanto será dado maior atenção aos problemas da reconstrução e da correspondência.

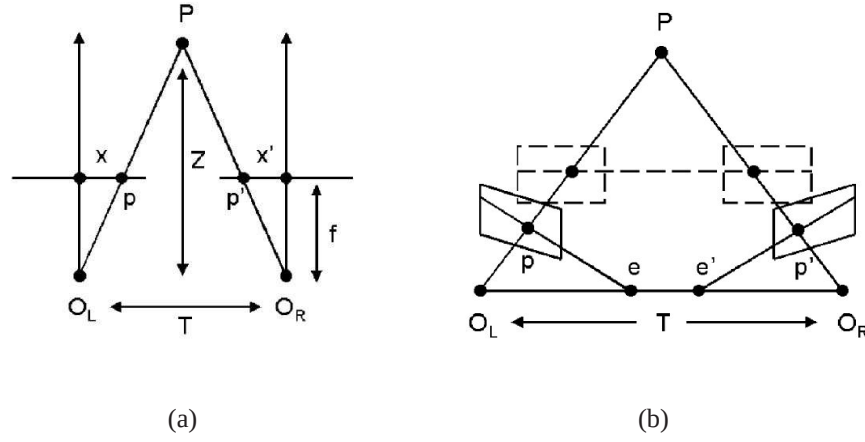


Figura 2.2: (a) Geometria do estéreo com imagens paralelas ao *baseline* (T). (b) Imagens retificadas utilizando a geometria epipolar. Figura obtida de [Brown et al., 2003].

2.3.1 O Problema da Reconstrução 3D

Considere a Figura 2.2(a). Seja o *baseline* T do par estéreo um segmento de linha que une os centros ópticos O_L e O_R . Ambas as câmeras estão alinhadas e possuem o eixo x paralelo a T . Isto significa que as projeções de qualquer ponto P nas duas imagens se encontram sempre na mesma linha. A diferença entre a coordenada x do pixel na imagem da esquerda e a coordenada x' do pixel na imagem da direita, que são as projeções p e p' de P , é chamada de *disparidade binocular*. O conjunto de todas as disparidades de todos os pixels das imagens é chamado de *mapa de disparidade*.

O problema da reconstrução consiste em determinar a estrutura 3D a partir de um mapa de disparidade, baseado na geometria da cena. A profundidade de um ponto P imageado por duas câmeras com centros ópticos O_L e O_R é definida pela intersecção de raios a partir desses centros ópticos que atravessam os planos de imagem em p e p' , respectivamente. Dado que a distância entre O_L e O_R (T) e a distância focal (f) são conhecidas, a profundidade Z do ponto é computada por semelhança de triângulos [Brown et al., 2003]:

$$Z = f \frac{T}{d}, \quad (2.6)$$

onde $d = x - x'$ é a disparidade, convertida em unidades métricas. É importante frisar que a consideração de paralelismo entre o eixo x das imagens e o *baseline* é inválida na maioria dos casos, mas pode ser facilmente resolvida com a retificação das imagens pelo uso da *Geometria Epipolar*: um ponto correspondente, se existir, tem que estar na linha epipolar,

definida como a intersecção do plano de imagem com o plano epipolar, que por sua vez é o plano que contém o ponto 3D procurado e o centro das duas lentes (Figura 2.2(b)). Através do conhecimento prévio dessa geometria, é possível aplicar uma transformação em cada imagem de tal forma que as linhas epipolares se tornam as linhas horizontais de cada imagem, o que permite reduzir o espaço de busca dos pontos correspondentes a uma linha, tal qual no caso da Figura 2.2(a). Maiores detalhes poder ser vistos em [Trucco e Murino, 2000].

2.3.2 O Problema da Correspondência

Como visto anteriormente, para obter a reconstrução 3D faz-se necessário encontrar a disparidade no par estéreo, o que implica estabelecer um casamento entre pixels das duas imagens que correspondam às projeções dos mesmos pontos da cena. Uma vez que existem muitas possibilidades para a escolha de pontos correspondentes, este problema é *ambíguo* [Faugeras et al., 1996]. Uma série de restrições podem ser impostas ao sistema estéreo para auxiliar na solução deste problema, em geral divididas em restrições geométricas e radiométricas.

As restrições geométricas podem ser impostas ao sistema de imagens, como a *restrição epipolar*, decorrente da existência de um único centro de projeção em cada câmera, que torna possível reduzir o espaço de busca por pontos correspondentes, ou podem serem assumidas a priori sobre os objetos da cena, como supor, por exemplo, que a variação na profundidade dos objetos com relação ao sistema de captura é lenta em praticamente toda a cena (*restrição de suavidade*).

As restrições radiométricas são impostas sobre os modelos que explicam a forma como a luz interage com os objetos da cena, o que significa utilizar modelos paramétricos de fontes de iluminação e da reflectância da superfície dos objetos. O modelo mais simples e mais amplamente utilizado é o modelo *Lambertiano*, ou seja, supõe-se que toda superfície é perfeitamente difusa.

É justamente nesta última consideração que os trabalhos de estéreo tradicional falham quando se utiliza uma cena que não é explicada por um modelo de iluminação local, pois neste caso os pixels correspondentes do par estéreo não necessariamente possuem a mesma intensidade, e podem variar de acordo com a posição de cada câmera, devido ao meio ou devido às propriedades materiais dos objetos da cena, o que interfere na determinação dos pontos correspondentes no par estéreo.

A busca da correspondência entre os pontos se divide em busca através de métodos locais ou globais. As Tabelas 2.2 e 2.3, baseadas em [Brown et al., 2003], apresentam os métodos atuais mais utilizados, com as principais referências que permitem um aprofundamento em

cada um deles.

<i>Método</i>	<i>Características Principais</i>	<i>Referências</i>
Casamento Simples	Faz a busca da correlação máxima, ou erro mínimo, numa pequena região da imagem. Tipicamente utiliza variações da correlação cruzada, diferença de intensidade, ou métricas de pontuação mais robustas.	[Zabih e Woodfill, 1994] [Faugeras et al., 1996] [Bhat e Nayar, 1998]
Otimização baseada no Gradiente	Minimiza uma função, em geral a soma das diferenças quadráticas, em torno de uma vizinhança da correspondência provável.	[Lucas e Kanade, 1981] [Kluth et al., 1992]
Casamento de Características	Neste método, o casamento depende de características específicas nas imagens, como arestas, curvas, etc., em lugar de se basear somente nas intensidades dos pixels.	[Dhond e Aggarwal, 1989] [Venkateswar e Chellappa, 1998] [Schmid e Zisserman, 1998] [Birchfield e Tomasi, 1999]

Tabela 2.2: Métodos de estéreo baseados em busca local da correspondência.

<i>Método</i>	<i>Características Principais</i>	<i>Referências</i>
Programação Dinâmica	Determina a disparidade de uma superfície através da busca de um caminho mínimo para linhas de varredura independentes, obtendo um mínimo global.	[Ohta e Kanade, 1985] [Intille e Bobick, 1994] [Hingorani et al., 1996] [Belhumeur, 1996] [Birchfield e Tomasi, 1998]
Curvas Intrínsecas	Faz um mapeamento das linhas epipolares para o espaço de curvas intrínsecas, convertendo o problema da busca do mínimo em um problema dos vizinhos mais próximos.	[Tomasi e Manduchi, 1996] [Tomasi e Manduchi, 1998]
Fluxo Máximo e Corte de Grafo	Encontra a correspondência a partir da minimização de um funcional de energia, o que é feito pelo corte mínimo ou fluxo máximo de um grafo cujos vértices e as arestas refletem o funcional de energia.	[Roy e Cox, 1998] [Thomos et al., 1998] [Boykov et al., 1999] [Zhao, 2000] [Kolmogorov e Zabih, 2001]

Tabela 2.3: Métodos de estéreo baseados em busca global da correspondência.

Além desses, outros métodos de otimização global têm sido apresentados, como o método de Difusão não-Linear [Shah, 1993][Mansouri et al., 1998][Scharstein e Szeliski, 1998], o método de Propagação Estimada [Sun et al., 2002], recomendado para casos de oclusões, e os métodos de reconstrução volumétrica [Fua e Leclerc, 1995] [Faugeras e Keriven, 1998] [Kutulakos e Seitz, 2000] [Carceroni e Kutulakos, 2002].

Cada um desses métodos tem vantagens e aplicações específicas que diferem entre si, entretanto o que possui os melhores resultados para uma situação geral, conforme estudo

comparativo em [Scharstein et al., 2001], é o de *Minimização de Energia via Corte de Grafos*. Observe na Figura 2.3 que a qualidade do mapa obtido por esse método, se comparado com o mapa de disparidade real, possui qualidade superior em relação aos outros métodos. Por esse motivo, esse método será utilizado como comparativo nos resultados obtidos nesta tese. Uma explicação de seu funcionamento básico é apresentada na Seção 2.3.3.

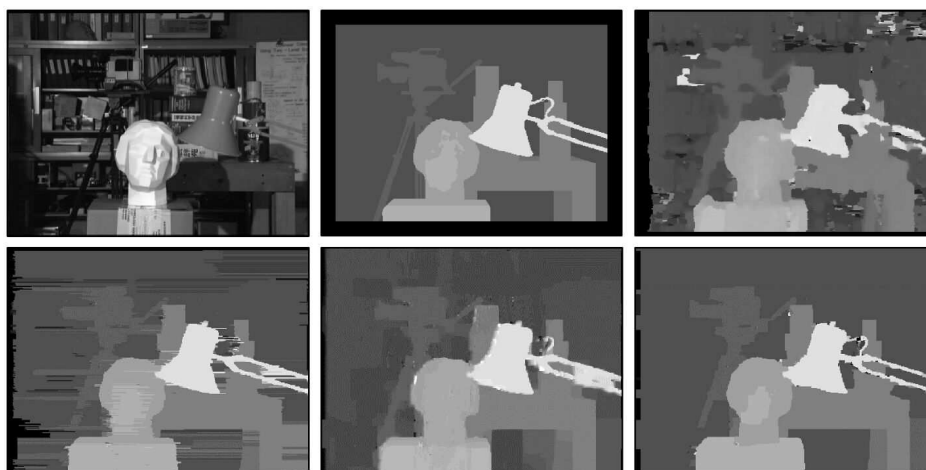


Figura 2.3: Resultados comparativos da imagem da Universidade de Tsukuba, obtida em [Scharstein et al., 2001]. Da esquerda para a direita: imagem esquerda do par estéreo; mapa de disparidade real; algoritmo de correlação local [Muhlmann et al., 2001]; programação dinâmica [Intille e Bobick, 1994]; fluxo máximo [Roy e Cox, 1998]; e corte de grafos [Kolmogorov e Zabih, 2001].

Todos os métodos citados anteriormente utilizam restrições radiométricas, mais especificamente, uma BRDF simplificada (lambertiana) onde as intensidades dos pixels correspondentes são iguais. Recentemente, trabalhos utilizando o que se convencionou chamar de *Estéreo Recíproco* fazem o uso de restrições que permitem aplicar o estéreo a superfícies com BRDF's arbitrárias [Zickler et al., 2002][Janko, 2003][Tu e Mendonca, 2003]. Para isso é utilizado o princípio da reciprocidade proposto por Helmholtz, explorando a simetria da reflectância das superfícies, onde as posições de câmeras e fontes de luz são escolhidas de tal forma que a relação entre os valores dos pixels correspondentes dependa somente da forma da superfície, e não da reflectância do material [Barros, 2004]. Essa restrição adicional, entretanto, só é válida para modelos de iluminação local, em condições cuidadosamente controladas.

Em [Kim et al., 2003] são desconsideradas as restrições radiométricas, utilizando uma variação do método de minimização de energia via corte de grafos, onde o custo de dados de um funcional de energia é dado por uma correlação feita por meio da máxima informação

mútua entre os pixels candidatos. Basicamente, esse método apresenta bons resultados quando as imagens estão sujeitas a ruídos que podem ser explicados por uma função linear, diferentemente dos ruídos gerados em um ambiente subaquático, que têm variação exponencial.

2.3.3 Método de Minimização via Corte de Grafos

O algoritmo de estéreo com minimização de energia via corte de grafos, aqui denominado EMGC (*Energy Minimization via Graph Cuts*) [Kolmogorov e Zabih, 2002], será utilizado nesta tese tanto para obtenção de resultados comparativos com um método tradicional, quanto para ser adaptado à proposta do estéreo subaquático. Esta escolha se deve ao fato de que o mesmo apresenta os melhores resultados na obtenção do mapa de disparidade de um estéreo denso, conforme o estudo comparativo apresentado em [Scharstein et al., 2001], e responde bem ao uso de um par estéreo com áreas oclusas, que são caracterizadas por pixels visíveis em somente uma das imagens.

O método EMGC manipula oclusões apropriadamente utilizando um funcional de energia que trata as imagens de entrada de forma simétrica e impõe suavidade espacial enquanto também preserva descontinuidades. A formulação do funcional de energia define uma configuração de disparidades $\mathcal{D}$ que pode ser vista como um conjunto discretos de valores de disparidade para uma dada cena. O funcional de energia em questão é então definido a partir de três termos [Kolmogorov e Zabih, 2001][Kolmogorov et al., 2003]:

$$\mathcal{E}(\mathcal{D}) = \mathcal{E}_{\text{dados}}(\mathcal{D}) + \mathcal{E}_{\text{suavidade}}(\mathcal{D}) + \mathcal{E}_{\text{visibilidade}}(\mathcal{D}). \quad (2.7)$$

Os três termos descritos incluem:

- um termo $\mathcal{E}_{\text{dados}}$, que resulta da diferença de intensidade entre pixels das imagens, em geral calculada com custo calculado pela diferença absoluta ou diferença quadrática, dada pela Equação (2.8):

$$\mathcal{E}_{\text{dados}}(\mathcal{D}) = \sum_{a \in \mathcal{A}} D(a), \quad (2.8)$$

sendo $a = \langle p_l, p_r \rangle$ um par de pontos onde se verifica a correspondência, sendo p_l relativo à imagem da esquerda e p_r relativo à imagem da direita, e $D(a)$ a diferença absoluta ou quadrática da intensidade dos pixels destes dois pontos.

- um termo para suavidade $\mathcal{E}_{suavidade}$, que faz com que os pixels da mesma vizinhança na imagem tendam a terem valores de disparidades similares, caso contrário sofrem uma penalidade que aumenta a energia deste termo, dado pela Equação (2.9):

$$\mathcal{E}_{suavidade}(\mathcal{D}) = \sum_{\{a_1, a_2\} \in \mathcal{N}} V_{a_1, a_2} \cdot T(f(a_1) \neq f(a_2)), \quad (2.9)$$

sendo V_{a_1, a_2} uma constante de penalidade calculada pela média das intensidades dos pixels em uma vizinhança $\mathcal{N}$ de cada imagem, $f(a)$ um assinalamento para um par de pontos, que pode ser ativo (1) se os pontos são correspondentes ou inativo (0) caso contrário, e $T(f(a_1) \neq f(a_2)) = 1$ para assinalamentos diferentes (disparidades distintas) e $T(f(a_1) \neq f(a_2)) = 0$ caso contrário.

- um termo para visibilidade $\mathcal{E}_{visibilidade}$, que verifica se um pixel é ocluso, e em caso afirmativo impõe uma penalidade que, na prática, equivale a levar a energia do funcional a um valor infinito.

$$\mathcal{E}_{visibilidade}(\mathcal{D}) = \sum_{p \in \{\mathcal{L} \cup \mathcal{R}\}} C_p \cdot T(|N_p(\mathcal{D})| = 0), \quad (2.10)$$

onde C_p uma constante de penalidade de valor tendendo a infinito, e $T(|N_p(\mathcal{D})| = 0) = 1$ para pixels oclusos, e $T(|N_p(\mathcal{D})| = 0) = 0$ caso contrário.

Este funcional de energia é, então, representado por um grafo com dois nós terminais (um fonte e um sorvedouro), cujas arestas representam os termos do funcional de energia e os nós internos representam as possibilidades de correspondência. O funcional é minimizado eficientemente por meio de um algoritmo de minimização via corte de grafos, desenvolvido especificamente para os tipos de aplicações de minimização de energia em visão computacional [Boykov e Kolmogorov, 2001]. O algoritmo de corte de grafos é apresentado no Apêndice A (Seção A.3). A complexidade do mesmo é polinomial no pior caso, contudo na prática é aproximadamente linear para grafos com diversos caminhos curtos entre os terminais, como no caso que se está utilizando. Maiores detalhes do método de [Kolmogorov e Zabih, 2001][Kolmogorov et al., 2003] são dados no Apêndice B.

2.4 A Visão Computacional em Imagens Subaquáticas

Sensoriamento, mapeamento e manipulação em ambientes subaquáticos são alguns tipos de aplicações de técnicas da visão computacional que possuem grande potencial de impacto

econômico e social. O uso da visão como uma modalidade de sensoriamento neste tipo de ambiente é muito útil em diversas tarefas como a manutenção de estruturas, monitoração da qualidade de água, e identificação e avaliação da fauna aquática [Luria e Kinney, 1970] [Negahdaripour e Yu, 1995] [Caccia, 1999] [Trucco e Murino, 2000]. Infelizmente, as imagens subaquáticas são sujeitas à forte influência das partículas em suspensão no meio líquido onde a luz se propaga e, em consequência disto, são tipicamente menos definidas que a maioria das imagens adquiridas na atmosfera [Yitzhaky et al., 1998]. Em geral, imagens subaquáticas não respondem bem às técnicas padrão de visão computacional. A principal dificuldade advém das características de absorção e dispersão da luz na água, que é similar ao efeito causado no ar por névoa ou outras partículas em suspensão.

Para entender esses fenômenos, considere que a radiância L é medida em um volume representado por um tubo estreito, que a luz atravessa como se fosse um feixe colimado. A radiância sofre os efeitos de atenuação devido à absorção e dispersão do meio e perde parte de sua energia ΔL ao longo de uma distância Δl , que podemos quantificar como $\Delta L = -cL\Delta l$, onde c é uma constante de proporcionalidade chamada *coeficiente de atenuação*, com unidade m^{-1} [Zaneveld et al., 2005]. No limite, quando $\Delta l \rightarrow 0$, é possível obter $dL/dl = -cL$. O coeficiente c depende do comprimento de onda λ da luz que se propaga no meio, e da localização no meio, sendo por isso representado como $c(\vec{x}, \lambda)$, onde $\vec{x}$ é o vetor na posição (x, y, z) .

Entretanto, devido ao efeito dispersivo da luz, a radiância que atravessa este tubo de uma localização $\vec{x}_1$ a outra localização $\vec{x}_2$ não é apenas atenuada, mas sofre a influência da luz dispersa no meio vinda de outras direções, inclusive na direção especificada pelas posições $\vec{x}_1$ e $\vec{x}_2$. Considerando uma semi-esfera, como no caso da Figura 2.1, cada direção é representada pelos ângulos (θ, ϕ) .

O aumento da radiância ao longo de l é proporcional à contribuição da luz dispersa que entra no tubo em ângulo sólido dw . Portanto, é possível dizer que a $dL/dl \propto Ldw$. A função de proporcionalidade que relaciona dL/dl a Ldw é chamada *função de dispersão volumétrica*, designada por $\beta(\vec{x}, \theta, \phi, \theta', \phi')$, que descreve a dispersão elástica da radiância de um comprimento de onda λ partindo da direção (θ, ϕ) à direção (θ', ϕ') :

$$\frac{dL(\vec{x}, \theta, \phi, \lambda)}{dl} = \beta(\vec{x}, \theta, \phi, \theta', \phi', \lambda) L(\vec{x}, \theta', \phi', \lambda) dw. \quad (2.11)$$

Esta equação dá a razão do incremento da radiância na direção $\hat{w}$ da localização $\vec{x}$, com um comprimento de onda λ , devido à dispersão da luz vinda na direção $\hat{w}'$ com a mesma localização e comprimento de onda. Como a luz pode vir de qualquer direção, é necessário integrar essa função para todas as direções (θ, ϕ) possíveis em uma semi-esfera. Sendo

assim, temos [Zaneveld et al., 2005]:

$$\frac{dL(\vec{x}, \theta, \phi, \lambda)}{dl} = \int_{\Omega} \beta(\vec{x}, \theta, \phi, \theta', \phi', \lambda) L(\vec{x}, \theta', \phi', \lambda) d\Omega. \quad (2.12)$$

Combinando a radiância perdida devido à atenuação, e a radiância adquirida devido à dispersão, obtemos a *Equação de Transferência Radiativa* para a propagação luz na água:

$$\frac{dL(\vec{x}, \theta, \phi, \lambda)}{dl} = -c(\vec{x}, \lambda) L(\vec{x}, \theta, \phi, \lambda) + \int_{\Omega} \beta(\vec{x}, \theta, \phi, \theta', \phi', \lambda) L(\vec{x}, \theta', \phi', \lambda) d\Omega, \quad (2.13)$$

que é válida sob as seguintes condições:

- A iluminação no meio é difusa, de tal forma que possui a mesma intensidade em qualquer volume infinitesimal do meio;
- A radiância no meio depende exclusivamente da profundidade (distância em espessura d'água entre o sensor o ponto observado), e não devido a reflexão da luz em algum objeto ou cena contida no meio;
- O meio é considerado elástico, ou seja, a luz se dispersa no meio mantendo o mesmo comprimento de onda;

Muitos autores tem apresentado soluções para a Equação (2.13), sendo a mais clássica apresentada em [Chandrasekhar, 1960]. Mais recentemente, em [Gordon, 2002] são discutidos diversos métodos desenvolvidos para estimar as propriedades ópticas inerentes às características da água, que o autor chama de *Problema Inverso da Óptica Hidráulica*: dadas as medidas de iluminação em um determinado corpo de água, determinar as características ópticas fundamentais do meio. Seus métodos consideram somente a iluminação natural na água, seja da luz solar direta ou indireta.

Uma outra forma de tentar obter uma aproximação dos parâmetros do meio é por meio da medição direta ou indireta, uma vez que efeitos de atenuação e dispersão estão relacionados ao nível de turbidez da água. Em [Davies-Colley e Smith, 2001] é apresentada uma comparação entre a forma de medição de turbidez da água utilizando um equipamento de medição próprio para tal, chamado *turbidímetro nefelométrico*, e uma forma mais subjetiva de medição utilizando um alvo conhecido submerso na água e obtendo uma medida de “clareza da água”. O método propõe a confecção de um alvo preto que é submerso na água até que o mesmo não possa mais ser visualizado, obtendo uma distância d chamada *hydrological range*. Esta distância se relaciona com a atenuação da água a partir da seguinte expressão:

$$\beta d = 4.8/c, \quad (2.14)$$

onde c é o coeficiente de atenuação da água, também dado por $c = a + b$, sendo a o coeficiente de absorção da luz na água e b o coeficiente de dispersão da luz na água.

Uma variação deste método bastante conhecida é do uso de um disco com regiões preta e branca, chamado de *sechi disk*, muito utilizado por biólogos marinhos e oceanógrafos para determinar um valor aproximado da turbidez da água, de forma simples e rápida.

Outros trabalho que apresentam métodos para obter as propriedades ópticas da água, ao todo ou em parte, são [Kirk, 2003] [Ackleson, 2003] [Carder et al., 2003], e mais recentemente [McCormick, 2004].

Contudo, estes valores são aproximados e específicos para uma dada solução simplificada da Equação (2.13). Em geral os trabalhos de visão desconsideram os efeitos do meio, ou utilizam alguma forma de minimizar os seus efeitos. Portanto, vamos classificar estes métodos em abordagens distintas: sem considerar os efeitos da água; considerando estes efeitos sem aplicar um modelo de propagação da luz na água; e considerando um modelo de propagação da luz na água.

2.4.1 Extraíndo Dados sem a Melhoria das Imagens

Os trabalhos que visam extrair informações de cenas subaquáticas utilizando técnicas de visão computacional sem considerar a influência do meio são aqueles onde a informação de cor ou mesmo o contraste das imagens obtidas da cena não é relevante, e portanto se considera a radiância como resultado de um modelo de iluminação local, ainda que as imagens violem essa consideração.

A maioria destes trabalhos [Caccia, 1999] estão voltados para controle de navegação de veículos autônomos, que consistem em executar travessias a velocidades e profundidades/altitudes determinadas, pairando sobre áreas de interesse, e adquirindo imagens, dados e amostras, minimizando a interação do veículo com o meio. Estes veículos utilizam imagens obtidas que são processadas de forma a obter referenciais para permitir sua navegação autônoma. A profundidade de alguns pontos da imagem é obtida combinando imagens de vídeo com outros dispositivos, como sonar e ou laser. Como as imagens são obtidas no fundo do oceano, faz-se necessário o uso de sistemas de iluminação que aumentam ainda mais o ruído devido, principalmente, à dispersão do meio.

Utilizando especificamente o estéreo, podemos citar dois trabalhos interessantes: (1) [Saito et al., 1995], que desenvolveu um método para a obtenção de um modelo 3D aproxi-

mado de um objeto submerso, utilizando *Motion Stereo*. Este método visa corrigir o efeito da refração da luz na água e obter o modelo 3D através do uso de um algoritmo de votação que possibilita determinar a melhor correspondência para o casamento de pontos no estéreo; e (2) [Tillet et al., 2000], que apresenta uma abordagem para estimar as dimensões (como peso e forma) de peixes contidos em um tanque de 3.5 m de diâmetro e 1m de profundidade, com imagens adquiridas por um par estéreo de câmeras. As câmeras são utilizadas para determinar pontos nos contornos visíveis do peixe e acompanhar a evolução destes pontos, de modo a obter a forma, tipo, orientação e posição do peixe.

Em ambos os casos, os autores apresentam alguns resultados interessantes, mas para uma situação estritamente controlada, uma vez que em situações reais, os efeitos causados nas imagens podem tornar inviável a aplicação do método [Ishimaru, 1999] a não ser que exista uma forma de restaurar a imagem obtida, permitindo assim estabelecer uma correspondência confiável entre os pontos e obter do par estéreo resultados consistentes da disparidade.

2.4.2 Melhoria das Imagens sem Considerar um Modelo de Propagação da Luz na Água

Alguns trabalhos consideram os efeitos de perda de contraste devido ao meio e procuram utilizar filtros ou técnicas de processamento digital de imagens [Gonzalez e Woods, 2000] que permitam corrigir ou reduzir este efeito, sem levar em consideração a sua natureza não linear e de origem não local.

Dependendo do nível de turbidez da água, esta pode ser uma abordagem razoável, entretanto não é geral. Usualmente se obtém uma “melhoria média” na imagem corrigida, ou seja, a melhoria do contraste é percebida no que seria uma distância média da cena, fazendo com que o aumento de contraste ocorra de forma mais expressiva em objetos na cena a menores distâncias, e de forma pior em objetos na mesma cena a distâncias maiores. O principal problema dessa abordagem consiste em não estabelecer um modelo que explique a propagação da luz na água que possa ser utilizado para corrigir os efeitos nas imagens.

Um caso típico do exposto anteriormente é o trabalho em [Garcia et al., 2002], que apresenta quatro técnicas para melhoria de contraste em imagens subaquáticas. A primeira consiste em uma expansão de contraste que pode ser linear ou não linear, dependendo da degradação da imagem; a segunda se dá através de equalização por histograma; a terceira consiste numa filtragem homomórfica, que utiliza a transformada de Fourier aplicada à imagem e filtra as altas frequências; e a última utiliza-se da subtração de um campo de iluminação que é obtido com base em curvas polinomiais.

O trabalho apresentado em [Harsdorf et al., 1999] propõe minimizar os efeitos do ambiente sub-aquático utilizando um dispositivo laser e um filtro polarizador. O laser serve como fonte de luz para um objeto e reduz os efeitos causados pela absorção, enquanto o filtro polarizador minimiza a contribuição do *backscattering light*. Isto proporciona uma substancial melhoria no contraste da imagem obtida. Ainda que não utilize um modelo de propagação de luz explicitamente, sua técnica considera os dois principais efeitos fotométricos do meio. A melhoria do contraste não necessariamente implica em uma restauração da imagem, e sim em minimizar os efeitos do meio no momento de adquirir esta imagem. O método tem algumas limitações, principalmente se as condições de turbidez da água forem expressivas e se a distância entre objetos de uma mesma cena for considerável, pois neste caso a não linearidade dos efeitos causados pela água irá influir negativamente no resultado obtido.

Alguns métodos conseguem obter melhorias nas imagens subaquáticas utilizando aparatos óptico/mecânicos, como em [Levoy et al., 2004], que utiliza um projetor, um painel de espelhos e uma câmera, onde as imagens refletidas pelos espelhos são capturadas da tal forma que a luz emitida pelo projetor esteja focalizada na cena, assim como a câmera, e com isto pontos imageados fora deste foco possuem menos contribuição para a composição da imagem, ou seja, a luz refletida no meio, e não na cena, é “enfraquecida” pelo método, chamado *Synthetic Aperture Confocal Imaging*. O método obtém resultados interessantes se considerado apenas o efeito *backscattering light*, entretanto possui um setup experimental bastante delicado, e não funciona bem para cenas com grandes distâncias em espessura d’água, pois não resolve o problema da atenuação da luz na água.

2.4.3 Melhoria das Imagens Utilizando um Modelo de Propagação da Luz na Água

Neste caso, os trabalhos utilizam um modelo baseado em uma simplificação da Equação (2.13), que é suficiente para obter resultados expressivos em termos de restauração das imagens e obtenção de mapas de disparidade. Quando se utiliza um modelo de propagação da luz, algumas grandezas precisam ser determinadas: a atenuação e dispersão do meio, e a distância de cada ponto da cena em relação ao sensor. Entretanto, para obter os parâmetros do meio é necessário o conhecimento da distância, e vice-versa. Em função disto, é comum que um mesmo método obtenha ao mesmo tempo um mapa de profundidades e a restauração da imagem da cena. As três maneiras conhecidas de fazer isso são: obter duas imagens da mesma cena em condições ambientais diferentes; obter duas imagens da mesma cena utilizando um filtro polarizador; obter um par estéreo da cena.

2.4.3.1 Imagens da Mesma Cena em Condições Ambientais Diferentes

Este é um trabalho desenvolvido em [Nayar e Narasimhan, 1999][Narasimhan et al., 2002] para imagens obtidas em cenas atmosféricas com condições adversas de clima, como neblina, névoa, entre outras. Esse método baseado em princípios físicos permite restaurar o contraste entre duas ou mais imagens obtidas de uma mesma cena sob condições climáticas substancialmente diferentes. Um modelo monocromático de dispersão atmosférica descreve como a cena é afetada quando condições adversas de clima estão presentes:

$$L(d) = L(\infty)\rho e^{-\beta d} + L(\infty)(1 - e^{-\beta d}), \quad (2.15)$$

onde:

$L(d)$ é a radiância percebida no sensor, em função da distância d entre esse e o ponto da cena;

$L(\infty)$ é radiância máxima devido ao efeito *backscattered light* no meio, considerando o termo medido em uma distância infinita;

ρ é um termo chamado *radiância normalizada*, e é função da reflectância difusa, da resposta espectral da câmera, e da geometria da iluminação no ponto, mas que não depende da mudança nas condições ambientais do meio. Em termos mais concretos, ρ é tal que a relação $L(\infty)\rho$ expressa a radiância devida unicamente ao ponto da cena;

β é o coeficiente de dispersão da luz, considerado constante em todo o meio.

Detalhes quanto ao modelo e suas considerações podem ser encontrados, entre outras publicações dos autores, em [Nayar e Narasimhan, 1999] e [Narasimhan e Nayar, 2000].

O método se baseia no princípio de que duas imagens estáticas obtidas da cena, à mesma distância mas sob condições diferentes do meio, apresentam diferentes radiâncias $L_1(d)$ e $L_2(d)$, $L_1(\infty)$ e $L_2(\infty)$, e coeficientes de dispersão β_1 e β_2 . Entretanto, como a superfície da cena não muda, ρ é o mesmo nas duas imagens. Utilizando essas duas imagens e a Equação (2.15) é possível obter uma razão entre $L_1(d)$ e $L_2(d)$ e, com algumas manipulações algébricas, determinar:

$$(\beta_2 - \beta_1)d = -\ln \frac{L_2(\infty) - L_2(d)}{L_1(\infty) - L_1(d)} - \ln \frac{L_1(\infty)}{L_2(\infty)}, \quad (2.16)$$

onde $(\beta_2 - \beta_1)d$ é chamada de profundidade óptica do meio, que basicamente é a distância módulo um fator de escala. Dessa forma, pode-se obter um mapa de profundidade da cena

em escala. No método básico, os valores de $L(d)$ são obtidos a partir das imagens, e os valores de $L(\infty)$ são obtidos a partir da seleção de um ponto de horizonte na imagem, como sendo aquele de máxima radiância (nesse caso, $L(d) = L(\infty)$). Observe que o método não obtém os valores de β .

Com o mapa de profundidade, é possível restaurar a imagem da cena a partir de um valor conhecido de radiância, como por exemplo um objeto totalmente escuro, cuja reflectância em ρ seja nula. Usando-se esse ponto, a componente de dispersão (segundo termo da Equação (2.16)) permite obter um valor referência para β .

Em uma extensão desse trabalho [Narasimhan et al., 2002], os autores apresentam sua metodologia para restaurar o contraste de imagens com cores. O método utilizado pode ser estudado e modificado para aplicações em imagens subaquáticas, uma vez que a maior diferença entre cenas do ar e cenas subaquáticas é que nestas últimas há menor visibilidade [Negahdaripour e Khamene, 2000]. Obter uma modelagem física para o comportamento da luz em meio subaquático é parte da solução para restauração das imagens obtidas neste meio.

Exceto em cenas muito particulares, esses algoritmos requerem a disponibilidade de duas imagens da mesma cena em que o meio de propagação da luz possui propriedades significativamente distintas (*e.g.*, sob condições climáticas diferentes), que podem levar dias para serem adquiridas. Assim, tais algoritmos não são úteis nas aplicações onde o conteúdo da cena muda mais rapidamente do que as propriedades ópticas do meio de propagação da luz, tais como a identificação e contagem de peixes nos mecanismos de transposição, e são mais aplicáveis somente a cenas estáticas.

2.4.3.2 Imagens da Mesma Cena com o Uso de um Filtro Polarizador

Esse é um trabalho recente apresentado em [Schechner e Karpel, 2004], cujo método se utiliza do mesmo modelo físico proposto por [Nayar e Narasimhan, 1999], porém aplicado a imagens subaquáticas, mantendo entretanto as mesmas simplificações e considerações, conforme a Equação (2.15).

Um filtro polarizador é um dispositivo óptico que pode ser acoplado a uma câmera e tem como característica principal impedir a passagem de uma forma de onda que não esteja na mesma polaridade ⁴ do filtro. O método se baseia no princípio de que a luz dispersa pelo meio tem uma polaridade diferente da luz que é refletida pela cena. Com isto, obtém-se duas imagens da cena: na primeira o filtro é ajustado para obter a máxima intensidade na imagem; e na segunda o filtro é ajustado para obter a mínima intensidade na imagem.

⁴a polaridade de uma forma de onda ocorre ao longo de um plano paralelo à direção de propagação da luz (*e.g.*, horizontal ou vertical).

Com isto, uma subtração entre as duas imagens equivaleria a obter o termo da luz dispersa no meio dado por $L(\infty)(1 - e^{-\beta d}) = (L_{max}(d) - L_{min}(d))$, o que leva a seguinte expressão:

$$\beta d = -\ln \left[1 - \frac{L_{max}(d) - L_{min}(d)}{L(\infty)} \right], \quad (2.17)$$

onde βd é distância óptica do ponto da cena ao sensor. A partir daí, similarmente como em [Nayar e Narasimhan, 1999], é possível utilizar essa distância óptica para gerar um mapa de profundidades e a restauração da imagem da cena.

Para conduzir as experiências, foi projetado e construído um sistema óptico com câmera, lente e um filtro polarizador. Testes feitos com imagens obtidas no oceano apresentam resultados bastante satisfatórios, em termos visuais, da restauração das imagens subaquáticas.

Apesar disto, o método possui duas limitações consideráveis, se comparado ao método proposto neste trabalho: (1) é aplicável somente para imagens com grandes distâncias em espessura d'água, sem o qual a diferença entre as imagens obtidas com e sem o filtro polarizador não é suficiente para obter resultados satisfatórios; (2) tem o problema de ser utilizado apenas em imagens estáticas, que devem ser obtidas uma em seguida a outra, com as mudanças no ajuste do filtro de polarização, o que inviabiliza sua utilização para um sistema aplicado à cenas dinâmicas, como no caso de um sistema automático de contagem de peixes.

2.4.3.3 Utilizando um Par Estéreo da Cena

Esse método, denominado *Estéreo Subaquático*, é uma das contribuições desta tese, e maiores detalhes serão dados no Capítulo 3. Basicamente, utiliza um par estéreo da cena a fim de se obter as correspondências entre os pontos e extrair as informações 3D. Como o meio influencia negativamente o contraste das imagens, um modelo de propagação da luz na água, baseado no modelo proposto em [Nayar e Narasimhan, 1999], será utilizado de modo a permitir a obtenção do mapa de profundidades da cena simultaneamente à restauração das imagens [Queiroz-Neto et al., 2004].

A principal diferença entre o estéreo subaquático e os demais trabalhos citados na Seção 2.4.3 é que o mesmo permite utilizar as imagens sem a limitação de que a cena seja estática, pois as duas imagens do par estéreo podem ser adquiridas ao mesmo tempo, além de propor um método experimental para estimar os parâmetros da água, o que permite uma restauração mais precisa de toda a imagem.

Capítulo 3

Obtendo Informações 3D em Cenas Subaquáticas Utilizando Estéreo

Uma forte motivação para a utilização deste cenário é o recente surgimento do interesse em técnicas de estéreo denso [[Scharstein e Szeliski, 2002](#)] dentro da comunidade de visão computacional, com aplicações em ambientes não convencionais. O estéreo denso é um grande problema de otimização que busca encontrar o mapeamento entre pixels de uma imagem e pixels correspondentes de uma outra imagem que minimize algum erro global de casamento pré-definido. A pesquisa nesta área é focalizada, em sua maior parte, em encontrar restrições no conjunto de possíveis mapeamentos, suficientemente fortes para tornar o problema tratável sob o aspecto de seu custo computacional, como nos trabalhos apresentados em [[Roy e Cox, 1998](#)], [[Chen e Medioni, 1999](#)] e [[Silva e Santos-Victor, 2000](#)].

O problema maior a ser observado, neste cenário, é o problema da correspondência, pois os métodos de estéreo denso não apresentam um bom resultado quando se utiliza uma cena que não é explicada por um modelo de iluminação local, uma vez que neste caso os pixels correspondentes do par estéreo não necessariamente possuem a intensidade mapeada por uma função bijetora que não dependa de outras variáveis que não as próprias intensidades, e podem variar de acordo com a posição de cada câmera. Isto ocorre devido à absorção e dispersão da luz na água, que reduzem significativamente o contraste na imagem, conforme apresentado na Seção [2.4](#).

Como exemplo, a Figura [3.1](#) apresenta os efeitos da água em uma cena obtida em um aquário, com três objetos geometricamente simples com textura multicolorida. Os objetos estão distantes entre si em torno de 30cm, sendo o objeto mais a esquerda o mais distante. As imagens foram adquiridas em três situações: (a) água limpa, (b) água com turbidez intermediária, e (c) água com turbidez alta. A perda de contraste é evidente à medida em que a

turbidez da água aumenta.

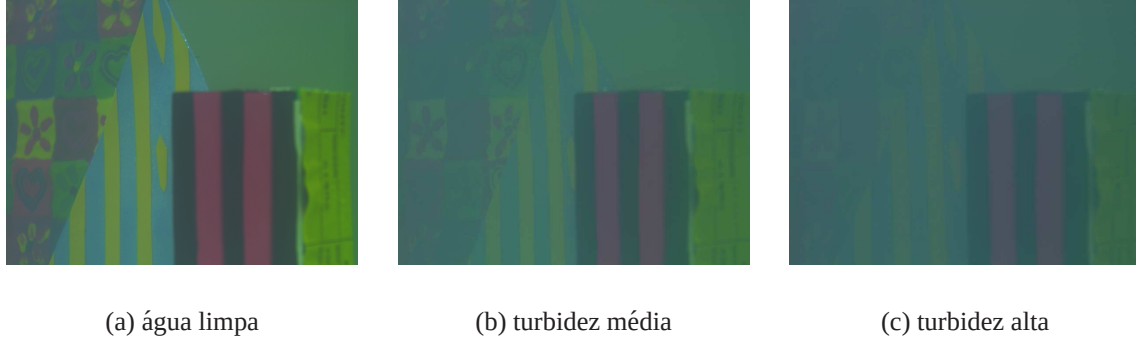


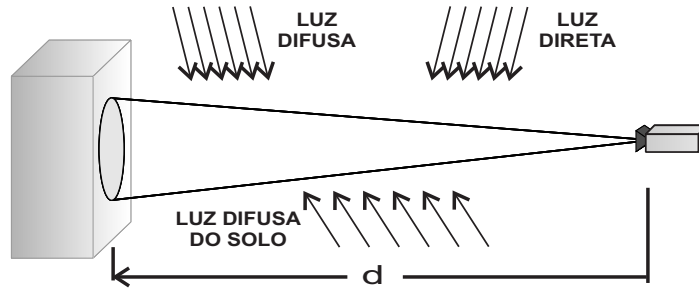
Figura 3.1: Exemplo dos efeitos da água nas imagens dos mesmos objetos submersos em um tanque com três situações diferentes de turbidez da água.

Se a absorção e as propriedades de dispersão do meio forem uniformes e conhecidas *a priori* e a distância entre a câmera e um objeto for também conhecida, é possível obter uma imagem real do objeto que foi afetado por estes fenômenos e remover artificialmente seus efeitos, a fim criar uma imagem artificial que pareça como uma imagem real adquirida em um meio livre dos efeitos de absorção e dispersão. Neste fase do trabalho, esta idéia é explorada da seguinte forma: (1) executando *a priori* a calibração da câmera (como no estéreo denso tradicional) e a estimação dos parâmetros de um modelo de propagação de propagação da luz na água, e (2) usando essa informação dentro de um algoritmo de estéreo denso modificado a fim de corrigir as cores do pixel ao comparar pares de pixels em linhas epipolares correspondentes. Observe que tais correções *devem* ser feitas durante a execução do estéreo denso, porque as profundidades reais dos vários objetos na cena com respeito a cada câmera *não* são conhecidas.

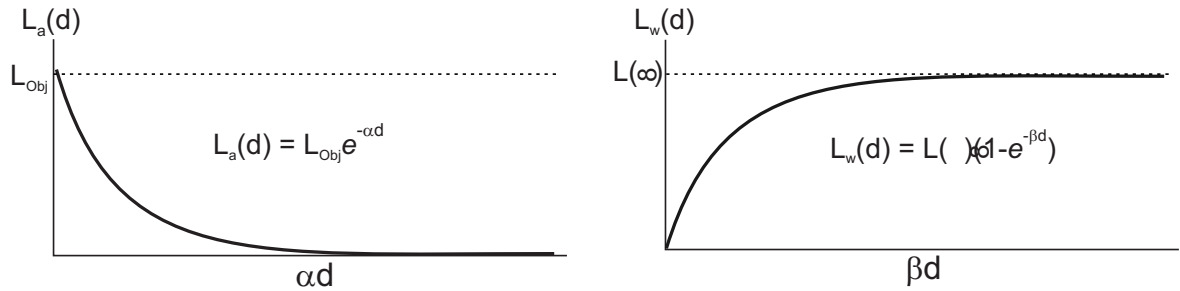
3.1 Modelo de Propagação da Luz na Água

Como descrito na Seção 1.2.1, os processos físicos de interação da luz em meio aquático causam uma forte perda de contraste nas imagens, principalmente devido aos fenômenos de *atenuação* e *dispersão* da luz na água. Uma forma de descrever estes fenômenos e os efeitos que causam nas imagens é utilizar um modelo simplificado de propagação da luz na água, que é uma variação da Equação de Transferência Radiativa da luz na água (Equação (2.13)) apresentada na Seção 2.4.

Mais especificamente, nesta tese é utilizado o modelo de propagação de luz dado em [Narasimhan e Nayar, 2002] para ambientes atmosféricos (Seção 2.4.3) e posteriormente uti-



(a) Aquisição de uma imagem em ambiente subaquático.



(b) Gráficos das componentes de atenuação e dispersão da luz na água.

Figura 3.2: Modelo simplificado de propagação da luz na água. Em (a) um objeto submerso do qual se adquire uma imagem. Em (b) os gráficos da radiancia incidente na câmera que possui duas componentes: a radiancia atenuada $L_a(d)$, dada pela radiancia L_{obj} do objeto reduzida exponencialmente em função do coeficiente de atenuação α do meio e da distância d ; e a radiancia dispersa no meio $L_w(d)$, que aumenta exponencialmente em função do coeficiente de dispersão β e a distância d , até um valor máximo de radiancia do meio $L(\infty)$.

lizado em [Schechner e Karpel, 2004] em ambientes subaquáticos. Porém, nesta tese o modelo foi adaptado para descrever a propagação da luz na água de forma mais precisa, sem a simplificação do modelo original no qual os parâmetros de atenuação e dispersão podem ser explicados por um único parâmetro.

Como ilustrado na Figura 3.2, este modelo assume que a intensidade de cada pixel na imagem da cena possui duas componentes: atenuação e dispersão. A primeira faz com que a radiancia na imagem diminua com a distância entre o sensor e a cena, principalmente devido à absorção e à dispersão da luz na água. A segunda proporciona um aumento da radiancia na imagem com o aumento da distância entre a cena e o sensor, devido à interação da luz com partículas em suspensão na água que reflete a luz em diversas direções, inclusive em direção ao sensor. Este último fenômeno será denominado *waterlight*. A radiancia total $L(d)$ recebida pelo sensor a partir da cena é, então, uma relação aditiva da radiancia gerada pelo objeto da cena e atenuada pelo meio e da radiancia gerada devido ao efeito *waterlight*,

que é dada pela Equação (3.1):

$$L(d) = L_a(d) + L_w(d), \quad (3.1)$$

onde d é a profundidade da cena em relação ao sensor. Os termos L_a e L_w são, respectivamente, a radiância direta atenuada e a radiância devido ao efeito *waterlight*, que possuem relação com os coeficientes de atenuação e de dispersão da luz na água, aqui denominados α e β . A radiância atenuada é dada pela Equação (3.2)

$$L_a(d) = L_{obj}e^{-\alpha d}, \quad (3.2)$$

onde L_{obj} é a radiância devido a um ponto da cena, dada em função da reflectância, da intensidade do espectro da luz, e da resposta espectral da câmera, mas não das condições adversas do meio onde se encontra a cena. α é o coeficiente de atenuação da luz no meio e d é a distância entre o sensor e o ponto na cena. Um modelo de radiância devido ao efeito *waterlight* é dada pela Equação (3.3), similarmente ao que ocorre no ambiente atmosférico [Narasimhan e Nayar, 2000].

$$L_w(d) = K(1 - e^{-\beta d}), \quad (3.3)$$

onde K é uma constante de proporcionalidade que depende da natureza da iluminação no meio. Se o objeto está a uma distância infinita (no horizonte da cena), a radiância devido a este efeito é máxima e é encontrada fazendo-se $d = \infty$, de onde obtém-se $K = L(\infty)$. Dessa forma, a radiância da imagem devido ao efeito *waterlight* a uma dada distância d do sensor é:

$$L_w(d) = L(\infty)(1 - e^{-\beta d}). \quad (3.4)$$

Utilizando as Equações (3.2) e (3.4), podemos re-escrever a Equação (3.1) e, por meio desta, obter a radiância em cada pixel da imagem obtida de uma cena em meio subaquático, que é dada por:

$$L(d) = L_{obj}e^{-\alpha d} + L(\infty)(1 - e^{-\beta d}). \quad (3.5)$$

3.2 Estimativa dos Parâmetros do Modelo de Propagação da Luz na Água

Os parâmetros do modelo são obtidos experimentalmente a partir de alvos submersos posicionados em distâncias conhecidas, o que permite gerar dados e gráficos utilizados para estimar os valores de α , β e $L(\infty)$. Para uma melhor compreensão, é apresentada a estimativa dos parâmetros do modelo utilizados na Seção 5.1.5.1. O experimento utilizou um alvo que possui uma região com intensidade máxima de cor (branco) e outra com intensidade mínima de cor (preto), uma câmera Sony DFW X-700, um aquário com 200l de água de dimensões 30x120x60 cm e pode ser explicado pelo Algoritmo 1:

Algoritmo 1 Algoritmo da estimativa dos parâmetros da água.

- 1: Obter o parâmetro $L(\infty)$ como sendo a radiância da maior distância em espessura d'água da cena, para a qual o aumento da distância praticamente não modifica a intensidade percebida pelo sensor;
 - 2: **para** cada distância conhecida d em espessura d'água **faça**
 - 3: Posicionar o alvo na distância d ;
 - 4: Adquirir a imagem do alvo submerso;
 - 5: Extrair as radiâncias das regiões de mínimo (preto) e máximo (branco) do alvo;
 - 6: Calcular os valores das distâncias escaladas αd e βd utilizando as Equações (3.7) e (3.8);
 - 7: **fim para**
 - 8: Plotar os gráficos de αd e βd em função de d e obter os valores de α e β extraíndo as inclinações das retas e ou utilizando regressão linear.
 - 9: Refinar os parâmetros obtidos através de uma otimização entre os valores da imagem submersa corrigida e sua equivalente em ambiente aéreo.
-

O valor de $L(\infty)$ é obtido a partir de uma imagem do alvo na maior distância em espessura d'água da cena, para a qual o aumento de d praticamente não modifica a intensidade percebida pelo sensor. Isto significa que para um valor de d grande, a radiância percebida pela câmera equivale a $L(\infty)$. Essa é uma aproximação que pode também ser obtida a partir da própria imagem da cena que se deseja modelar, como em [Schechner e Karpel, 2004], desde que a cena contenha uma região do horizonte ou de uma distância que reflita esse efeito.

3.2.1 Cálculo do Parâmetro de Dispersão

Após adquirir o conjunto de imagens do alvo, o parâmetro de dispersão é obtido considerando apenas a região de intensidade de cor mínima dos alvos. No caso ideal, a contribuição

dessa região do alvo na formação da imagem é nula, ou seja, $L_{obj} = 0$. Assim, a Equação (3.5) se reduz a:

$$L(d) = L(\infty)(1 - e^{-\beta d}), \quad (3.6)$$

de onde é possível obter a seguinte relação:

$$\beta d = -\ln \left(\frac{L(\infty) - L(d)}{L(\infty)} \right). \quad (3.7)$$

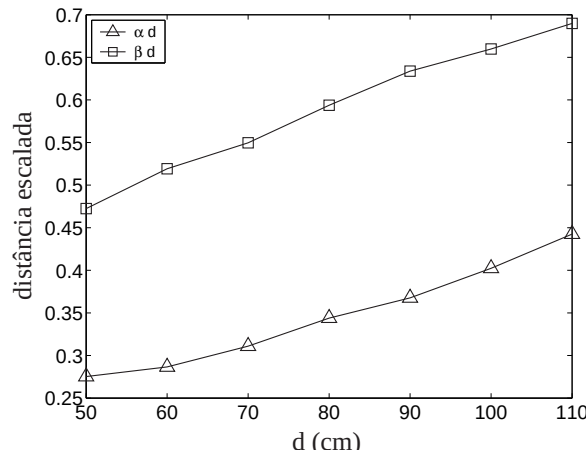


Figura 3.3: Evolução das distâncias escaladas αd e βd da banda verde.

Para cada distância d é calculado o valor médio da distância escalada βd utilizando a Equação (3.7), o valor de $L(d)$ extraído da região de intensidade mínima das imagens dos alvos, e o valor de $L(\infty)$ obtido anteriormente. Uma vez que a distância d é conhecida, obtemos o valor de β diretamente do gráfico da Figura 3.3 ou utilizando alguma técnica de regressão linear. O resultado para o experimento da Seção 5.1.5.1 pode ser visto na Tabela 3.1.

3.2.2 Cálculo do Parâmetro de Atenuação

Para o cálculo de α utilizamos a mesma consideração de que a região de intensidade mínima do alvo possui valor próximo de zero, e que portanto a radiância percebida pela câmera é devido ao efeito *waterlight*, dado pela Equação (3.4). Com isso, os valores de $L_w(d) = L(\infty)(1 - e^{-\beta d})$ podem ser obtidos a partir das imagens dos alvos (os mesmos valores de $L(d)$ da Seção 3.2.1). Manipulando-se a Equação (3.5), obtemos a distância escalada αd dada por:

$d \text{ (cm)}$	<i>distancia escalada (βd)</i>		
	R	G	B
50	0.6105	0.4725	0.8484
60	0.6569	0.5192	0.8972
70	0.6168	0.5496	0.9699
80	0.6452	0.5937	1.1384
90	0.6548	0.6338	1.2504
100	0.6660	0.6598	1.4232
110	0.6970	0.6898	1.4665
$\beta(\text{cm}^{-1})$	0.0086	0.0076	0.0144

Tabela 3.1: Distancias escaladas (cm) e cálculo de β médio (cm^{-1}) em cada canal RGB.

$$\alpha d = -\ln \left(\frac{L(d) - L(\infty)(1 - e^{-\beta d})}{L_{obj}} \right). \quad (3.8)$$

Similarmente ao caso anterior, para cada distância d é calculado o valor médio da distância escalada αd utilizando-se a Equação (3.8), o valor de $L(d)$ extraído da região de intensidade máxima das imagens dos alvos, e o valor de $L(\infty)$ obtido anteriormente. Uma vez que a distância d é conhecida, obtemos o valor de α diretamente do gráfico da Figura 3.3 ou utilizando alguma técnica de regressão linear. O resultado para o experimento da Seção 5.1.5.1 pode ser visto na Tabela 3.2.

$d \text{ (cm)}$	<i>distancia escalada (αd)</i>		
	R	G	B
50	0.2791	0.2753	0.3202
60	0.3145	0.2866	0.3142
70	0.3299	0.3109	0.3795
80	0.3506	0.3439	0.3950
90	0.3665	0.3676	0.4875
100	0.4397	0.4024	0.5715
110	0.4813	0.4425	0.6614
$\alpha(\text{cm}^{-1})$	0.0047	0.0044	0.0056

Tabela 3.2: Distancias escaladas (cm), cálculo de α médio (cm^{-1}) em cada canal RGB.

3.2.3 Refinamento dos Parâmetros do Modelo

Algumas considerações feitas na estimativa dos valores dos parâmetros de atenuação e dispersão são aproximadas, como no caso da obtenção de $L(\infty)$ e o fato de que a região de intensidade mínima de cor do alvo realmente possui uma contribuição nula. Portanto, faz-se

necessário refinar esses valores, o que em princípio consiste em aplicar a função inversa do modelo dado na Equação (3.5), recuperando-se L_{obj} na imagem subaquática (Equação (3.9)), e em seguida calcular o erro médio quadrático (RMSE) entre a imagem recuperada e uma imagem do mesmo alvo obtida em ambiente sem os efeitos de atenuação e dispersão (fora da água). Este procedimento é feito iterativamente por um método de otimização não linear que altera os parâmetros buscando minimizar a função de erro. Neste trabalho utilizamos a função de minimização *fminunc* do MatLab 6.1, que é apropriada para a busca do mínimo de uma função com mais de uma variável. Os resultados finais dos parâmetros são dados na Tabela 3.3. A Figura 3.4 apresenta o alvo utilizado nas posições $d_1 = 50cm$ e $d_2 = 100cm$ fora do aquário, submerso e corrigido com os parâmetros obtidos.

$$L_{obj} = [L(d) - L(\infty)(1 - e^{-\beta d})]e^{\alpha d}. \quad (3.9)$$

Canal	parametros refinados		
	α	β	$L(\infty)$
R	0.0063	0.0044	45.84
G	0.0042	0.0055	86,58
B	0.0087	0.0092	51,05

Tabela 3.3: otimização dos parâmetros α , β em cm^{-1} e $L(\infty)$ em intensidade de pixels.

3.3 Estéreo com Estimativa Experimental dos Parâmetros do Modelo de Propagação da Luz

A teoria geométrica da visão estéreo, que possibilita a obtenção de coordenadas tridimensionais de objetos a partir de pares de imagens, depende da capacidade de se resolver o problema da correspondência, que consiste em determinar os elementos do par de imagens que são projeções do mesmo elemento no mundo 3D. A existência de candidatos ambíguos para combinação, ou mesmo da inexistência de correspondentes devido à oclusão tornam o problema ainda mais complicado. O método estéreo utiliza algumas técnicas para reduzir o número de ambigüidades, entre essas o uso da geometria epipolar. Entretanto, frequentemente considera-se que pixels correspondentes possuem intensidades semelhantes, o que nem sempre é verdadeiro, mas pequenas variações devido à diferenças nos sensores podem ser corrigidas por meio de uma calibração radiométrica do par estéreo.

No ambiente subaquático o problema da correspondência é agravado pelo meio, pois cada sensor utilizado na aquisição do par estéreo encontra-se a distâncias diferentes em relação

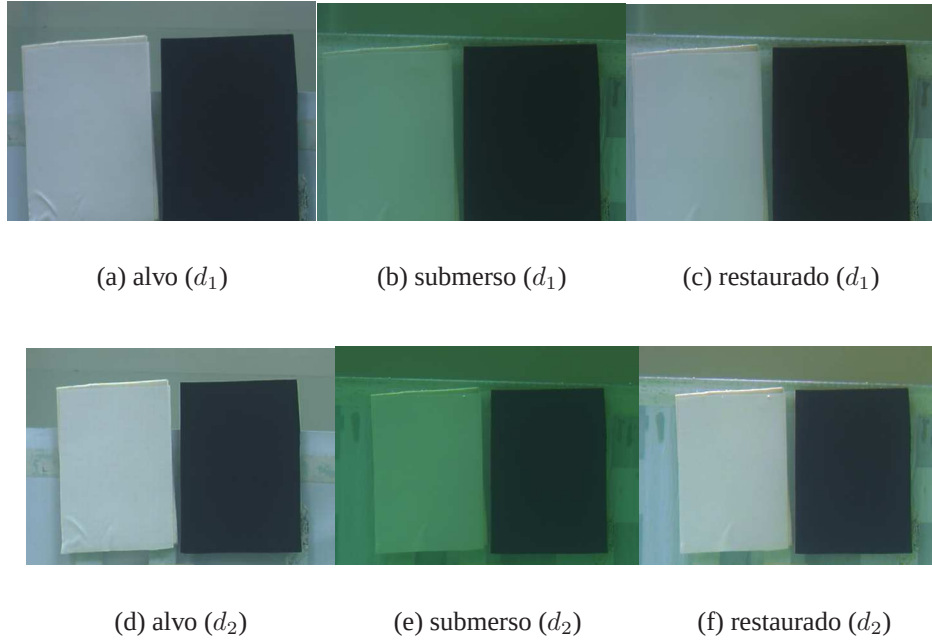


Figura 3.4: (a) alvo fora da água na distancia $d_1 = 50cm$, (b) e (c) alvo d_1 submerso na água e a mesma imagem restaurada pela Equação (3.9). Em (d), (e) e (f) o mesmo para $d_2 = 100cm$.

a um ponto da cena, e pela Equação (3.5) é possível observar que essa diferença influencia significativamente os valores de intensidade obtidos em cada imagem. Os métodos de calibração radiométrica não funcionam para este caso e se tornam piores quanto maior for a diferença entre as distâncias da cena a cada sensor. Isso ocorre por se tratar de uma alteração com característica não linear e dependente da distância não conhecida entre o sensor e a cena.

Não é possível aplicar uma restauração *a priori* das imagens, ainda que se possuam os parâmetros do modelo de propagação da luz na água, pois não são conhecidas as distâncias de cada ponto da cena aos sensores. Por outro lado, seria necessário que as imagens estivessem restauradas para que o estéreo obtivesse as distâncias desejadas, o que nos leva de volta ao problema da correspondência. O método desenvolvido nesta tese utiliza as vantagens da geometria epipolar para reduzir o espaço de busca e obter a melhor similaridade utilizando a Equação (3.9) aplicada a cada par de pontos verificados nas n distâncias possíveis (onde n é o tamanho da linha), conforme o Algoritmo 2.

Dessa forma, o método consiste em, basicamente, considerar uma disparidade candidata entre um par de pixels das duas imagens e calcular as distâncias entre o ponto da cena e os sensores, utilizando esses dados para restaurar as intensidades dos pixels das imagens e

Algoritmo 2 Algoritmo do estéreo subaquático.

-
- 1: Calibrar e retificar as imagens geométrica e radiometricamente;
 - 2: Calibrar α e β e medir $L(\infty)$, para cada banda;
 - 3: **para** cada ponto $p_1 = (i, j)$ na imagem 1 (imagem de referência) **faça**
 - 4: **para** cada ponto p_2 na linha i da imagem 2 **faça**
 - 5: Computar as distâncias em espessura d'água d_1 e d_2 , relativas às duas câmeras;
 - 6: Calcular L_{obj1} e L_{obj2} usando a Equação (3.9);
 - 7: Determinar a correspondência entre os pontos através de um algoritmo de estéreo convencional.
 - 8: **fim para**
 - 9: **fim para**
 - 10: A disparidade $\mathbf{d} = [d_1, d_2]$ de cada p_1 pode ser obtida minimizando um funcional de energia apropriado construído a partir do termo $\varepsilon_{(d_1, d_2)} = \text{abs}(L_{obj1} - L_{obj2})$ que mede o quão bom é o casamento de cada par de pixels candidatos p_1 e p_2 , que potencializa termos que retornam propriedades desejáveis da reconstrução tais como a suavização ou a preservação espacial das discontinuidades, e penaliza resultados indesejáveis tais como a diferença de intensidade dos pixels.
-

procurar a correspondência entre os pontos por meio de um algoritmo de estéreo denso, utilizando desde uma medida de correspondência simplificada com minimização local até um método mais elaborado com minimização global. A disparidade candidata que apresentar a melhor correspondência é utilizada para gerar o mapa de disparidades da cena. O Apêndice A (texto de [Barros, 2004]) apresenta alguns métodos para solução do problema da correspondência.

3.4 Estéreo com Estimativa Automática dos Parâmetros do Modelo de Propagação da Luz

A idéia central para evitar o processo experimental de estimativa dos parâmetros do modelo de propagação da luz na água, conforme Seção 3.2, é estender o Algoritmo 2 e inserir os parâmetros do modelo como variáveis a serem determinadas, de forma a se efetuar uma otimização global onde a combinação desses parâmetros e das disparidades minimize um dado critério de otimização.

Seja um o par estéreo de uma cena subaquática. A aplicação do modelo dado na Equação (3.5) para um pixel em cada imagem do par estéreo que corresponde ao mesmo ponto na cena leva ao seguinte sistema:

$$\begin{cases} L(d_1) = L_{obj}e^{-\alpha d_1} + L(\infty)(1 - e^{-\beta d_1}) \\ L(d_2) = L_{obj}e^{-\alpha d_2} + L(\infty)(1 - e^{-\beta d_2}), \end{cases}$$

onde $L(d_1)$ é relativo à imagem 1 que possui uma distância d_1 em espessura d'água e $L(d_2)$ é relativo à imagem 2 que possui uma distância d_2 em espessura d'água. Fazendo-se uma razão entre $L(d_1)$ e $L(d_2)$, e com algumas manipulações algébricas, obtém-se:

$$[L(d_1) - L(\infty)(1 - e^{-\beta d_1})]e^{\alpha d_1} - [L(d_2) - L(\infty)(1 - e^{-\beta d_2})]e^{\alpha d_2} = 0. \quad (3.10)$$

O valor $L(\infty)$ é obtido diretamente de uma das imagens, a partir da maior distância em espessura d'água, como em [Schechner e Karpel, 2004]. As distâncias d_1 e d_2 são calculadas a partir de um mapa de disparidade obtido pelo Algoritmo 2. Se forem considerados todos os pixels das imagens, e os valores de $L(d_1)$ e $L(d_2)$ extraídos das imagens do par estereo, é possível determinar os valores de α e β através de um método de minimização cuja função objetivo é dada por:

$$\arg \min_{\alpha, \beta} \epsilon \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \left([L_1^i - L(\infty)(1 - e^{-\beta d_1^i})]e^{\alpha d_1^i} - [L_2^i - L(\infty)(1 - e^{-\beta d_2^i})]e^{\alpha d_2^i} \right)^2. \quad (3.11)$$

O problema de encontrar o mínimo de uma função não linear bi-dimensional é uma tarefa que demanda um alto custo computacional. Uma forma de resolver este problema é utilizar um ciclo iterativo do Método do Gradiente Conjugado padrão [Press et al., 1992], o que no entanto não impede que um outro método possa ser utilizado. Esse método utiliza-se das derivadas parciais do erro para incrementar a direção de busca dos parâmetros, tornando a convergência mais rápida.

Com isso, a idéia básica contida no Algoritmo 3 considera que o espaço de parâmetros na busca das disparidades e dos parâmetros do modelo pode ser dividido em dois sub-espacos, de tal forma que exista uma intersecção entre as soluções obtidas em cada um deles, e que nesta intersecção se encontra a solução do problema. No primeiro sub-espaco são minimizadas as disparidades dados os parâmetros do modelo, e no segundo são minimizados os parâmetros do modelo, dadas as disparidades.

É importante ressaltar que neste método deve ser utilizado um método de estereo denso com minimização global. Isso é importante para garantir que o espaço de soluções das disparidades tenha uma solução mínima global, dado que a busca no espaço de soluções dos parâmetros do modelo considera toda a imagem do par estereo.

Algoritmo 3 Algoritmo de estéreo subaquático com estimativa automática dos parâmetros do modelo de propagação da luz na água.

- 1: Calibrar e retificar as imagens geométrica e radiometricamente;
 - 2: Estimar um valor inicial de α e β ;
 - 3: **enquanto** um critério de convergência ou número máximo de iterações não ocorre **faça**
 - 4: Executar o Algoritmo 2 para obter o mapa de disparidades globalmente mínimo utilizando os valores de α e β correntes;
 - 5: Computar as distâncias em espessura d'água d_1^i e d_2^i , relativas às duas câmeras;
 - 6: $\epsilon \leftarrow \frac{1}{n} \sum_{i=1}^n \left([L_1^i - L(\infty)(1 - e^{-\beta d_1^i})]e^{\alpha d_1^i} - [L_2^i - L(\infty)(1 - e^{-\beta d_2^i})]e^{\alpha d_2^i} \right)^2$;
 - 7: Encontrar $\arg \min_{\alpha, \beta}(\epsilon)$, utilizando ϵ , $\frac{\partial \epsilon}{\partial \alpha}$ e $\frac{\partial \epsilon}{\partial \beta}$;
 - 8: Atualizar α e β ;
 - 9: Calcular o critério de correspondência;
 - 10: **fim enquanto**
-

Devido às oclusões que ocorrem no par estéreo, as distâncias para alguns pixels das imagens simplesmente não podem ser calculadas, ou resultam em valores espúrios (*outliers*) que devem ser desconsiderados na otimização de α e β . Uma maneira de fazer isto é utilizar o método de casamento robusto dado em [Rousseeuw e Leroy, 1987] e [Black e Jepson, 1996].

Os valores de α e β obtidos não necessariamente são os parâmetros fisicamente consistentes da água, como obtidos na Seção 3.2, pois a correspondência mínima no Algoritmo 3 pode ser alcançada para um conjunto de valores de α e β que satisfaçam a Equação 3.10.

3.5 Análise de Complexidade dos Métodos

3.5.1 Complexidade de Custo

A implementação dos métodos propostos possui uma complexidade de custo computacional que depende do método de estéreo convencional utilizado. Em termos de custo, o método de estéreo subaquático dado pelo Algoritmo 2 será utilizado para obter as disparidades, seja através da aquisição experimental (Seção 3.2) ou automática (Algoritmo 3) dos parâmetros do modelo de propagação da luz na água, sendo que essa primeira tarefa precisa ser executada uma única vez para os diversos pares estéreo de um mesmo ambiente.

No método de estéreo subaquático do Algoritmo 2, o custo principal no ciclo entre os passos 3 e 9 é $\mathcal{O}(st)$, sendo st o custo do estéreo tradicional utilizado no processo de correspondência, portanto o custo total é $\mathcal{O}(n.l.st)$, onde n é o número de pixels da imagem e l é o número de pixels de uma linha da imagem;

No método de obtenção automática dos parâmetros dado no Algoritmo 3, o custo princi-

pal no ciclo entre os passos 2 e 8 é dado pela chamada do Algoritmo 2 no passo 3, portanto o custo total é $\mathcal{O}(iter.n.l.st)$, onde $iter$ é o número de iterações necessárias para a convergência do método, que não é uma variável estabelecida pelo usuário, mas indica que esse custo será sempre $iter$ vezes maior que o custo do método de estéreo subaquático.

3.5.2 Complexidade de Espaço

Em relação à complexidade de espaço, os métodos não possuem estruturas de dados que possam tornar crítico o uso de memória primária, e a complexidade de espaço é dada pela necessidade de uso de memória do método de estéreo convencional em uso.

3.6 Limitações dos Métodos

Não obstante os avanços que os métodos proporcionam ao que já existe para o problema tratado, os mesmos possuem algumas limitações.

O método de estimativa dos parâmetros do modelo é uma tarefa que exige a aquisição de imagens controladas (alvos) dentro do ambiente, o que pode não ser uma tarefa trivial para o caso de algumas situações reais, como no caso de rios de forte correnteza ou em mar aberto. Entretanto, isso pode ser resolvido com a confecção de um aparato mecânico que auxilie na aquisição nas imagens.

O modelo adotado não explica o efeito de borramento que o espalhamento da luz na água causa na imagem, mas isso pode ser feito por alguma técnica de processamento de imagens específica para esse fim, como indicado em [Schechner e Karpel, 2004].

O método de estimativa automática dos parâmetros necessita que as entradas iniciais de α e β sejam, ao menos, da mesma ordem de grandeza dos valores reais. Isso ocorre devido ao uso dos métodos de otimização que em geral são bastante dependentes da entrada. Uma forma de minimizar este problema é avaliar métodos de otimização que gerem resultados mais consistentes, mesmo para entradas distantes do valor mínimo da função objetivo.

Os métodos apresentados não são aplicáveis se a atenuação e a dispersão do meio for tal que não seja possível obter informações da cor do objeto da cena na imagem adquirida, ou seja, os métodos funcionam para situações em que $L(d) - L(\infty)(1 - e^{-\beta d}) > 0$. Essa é uma limitação universal para qualquer método que utilize visão computacional em ambiente subaquático a partir de imagens de intensidade.

Capítulo 4

Obtendo os Parâmetros de Reflectância de uma Cena com Iluminação Global

O problema da reflectância inversa, em particular para cenas com iluminação global, tem sido abordado por uma série de técnicas com bons resultados, como apresentado na Seção 2.2. Entre as principais dificuldades dos métodos atuais temos: (1) o uso de múltiplas imagens exige uma preparação experimental elaborada, necessitando que cada câmera seja calibrada e posicionada de forma controlada; (2) esta é uma tarefa computacionalmente cara, principalmente devido às inter-reflexões entre os objetos da cena; (3) a presença de superfícies translúcidas inviabiliza a maioria dos métodos atuais que se baseiam no uso da radiosidade. Nesta seção é apresentado um método para obter os parâmetros do modelo de reflectância que explicam as cores das superfícies de cada objeto da cena, de maneira computacionalmente eficiente, considerando objetos cuja superfície seja homogênea, reflexiva ou translúcida, a partir de uma única imagem da cena.

A idéia básica do método proposto nesta tese pode ser vista no diagrama da Figura 4.1. As entradas para o método são: (1) a geometria 3D da cena; (2) as informações sobre as fontes de luz, como localização e intensidade; (3) uma imagem da cena. Um modelo de iluminação global, combinado com um modelo de iluminação local apropriado, utilizam a geometria 3D e as informações sobre as fontes de luz para obter as restrições fotométricas necessárias ao cálculo do valor de intensidade de um pixel. Mais especificamente, as restrições são relacionadas com as contribuições vindas de outras superfícies, determinadas por meio do modelo de iluminação global e das propriedades materiais dos objetos da cena.

As informações da geometria e propriedades materiais da cena formam a equação de um pixel da imagem na qual as incógnitas são os parâmetros de reflectância primária e secundária, considerando constantes as restrições fotométricas relacionadas à geometria da

cena. Como as fontes de luz e a forma tridimensional da cena não mudam (a cena é estática), os coeficientes da equação de cada pixel da imagem são extraídos em uma única renderização e guardados em uma estrutura de dados apropriada. Após a renderização, os dados desta estrutura, juntamente com os valores reais de cada pixel da imagem da cena, são utilizados para gerar um sistema de equações não linear fatorado que permite obter o resultado de forma mais eficiente que o sistema não linear original, e cuja solução é o conjunto de parâmetros de reflectância dos objetos da cena.

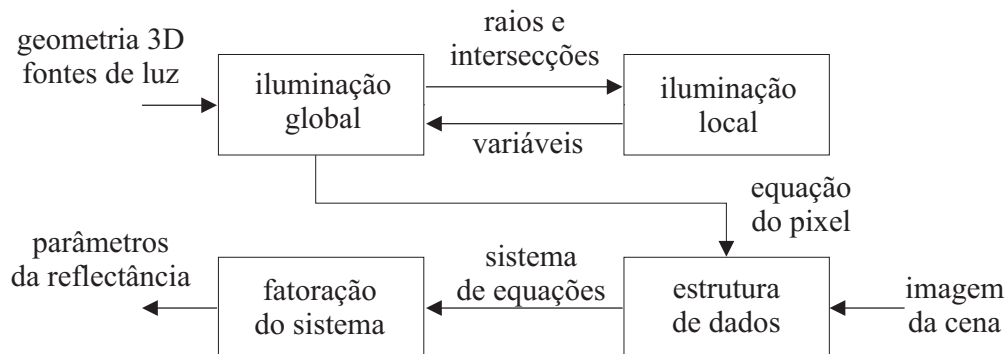


Figura 4.1: Visão geral do método de reflectância inversa proposto. Uma única renderização utilizando um método de iluminação global (*Ray Tracing*) em conjunto com um método de iluminação local, conhecidas a geometria e as fontes de luz da cena, adquire a equação de cada pixel de uma imagem da cena que são armazenados em uma estrutura de dados apropriada que, a partir de uma imagem da cena real, forma um sistema de equações que é fatorado e resolvido, obtendo assim os parâmetros de reflectância.

4.1 Modelo de Formação da Imagem

A fim de poder extrair as propriedades fotométricas de superfícies da cena diretamente de uma única imagem, é necessário definir como as intensidades individuais do pixel são determinadas em função da iluminação e da geometria 3D conhecidas *a priori*, e dos parâmetros de reflectância e transmissividade não conhecidos. Para esta finalidade é utilizado um *modelo de iluminação global* conhecido e bem estabelecido chamado *Ray Tracing* (*Traçador de Raios*), que permite explicar bem as inter-reflexões, efeitos da transparência, refrações e sombras devido às múltiplas fontes de iluminação.

O *Ray Tracing* é baseado nos princípios da geometria óptica clássica e assume que todos os objetos da cena são suficientemente maiores que o comprimento de onda da luz, e que essa se propaga por meio de linhas retas. Raios de luz refletidos pelas superfícies da cena emanam em todas as direções, e alguns deles atravessam o plano de imagem da câmera em

uma posição de um pixel. A idéia básica deste modelo é utilizar apenas um raio por pixel, que é o equivalente a ver a cena através de uma câmera *pinhole*, de tal forma que este raio parte de uma das I_k fontes de luz primárias em direção a um pixel particular, com uma série de D reflexões ou de refrações deste raio nas superfícies da cena [Glassner, 1995].

A Figura 4.2(a) apresenta um exemplo de *Ray Tracing*. Um raio primário é lançado a partir de cada pixel do plano de imagem e verificada a intersecção deste com todos os objetos da cena. Se uma superfície é interceptada, calculam-se as distâncias entre o pixel no plano de imagem e os pontos das superfícies interceptadas, cuja menor distância indica qual dessas é visível para o pixel. Se a superfície é reflexiva, então um raio de reflexão, R_i , é gerado. Se a superfície é translúcida, então um raio de transmissão, T_i , é gerado. Os raios de reflexão/transmissão são continuamente gerados até que o raio não intercepte mais nenhum objeto da cena ou um limiar de nível máximo de profundidade seja atingido.

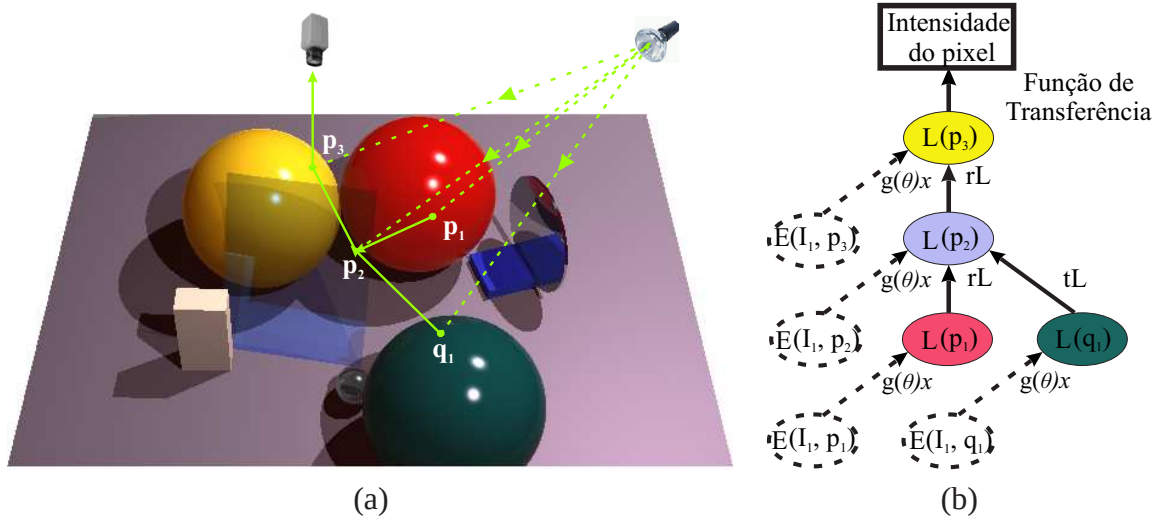


Figura 4.2: *Árvore de Computação da Intensidade do Pixel* com profundidade $D = 3$. (a) os trajetos que interceptam a cena e geram as reflexões / refrações da árvore a partir da fonte I_1 para um dado pixel. (b) A árvore resultante, com $E(I_1, p)$ dado pela Equação (4.3), e $L(p)$ dado pela Equação (4.6).

A fim de modelar superfícies ásperas com propriedades fotométricas não-ideais, é permitido que a primeira reflexão ou refração (primárias) em cada trajeto do raio seja não-ideal, *i.e.*, supõe-se que após a luz ser emitida por uma fonte de luz primária, sua primeira interação com uma superfície da cena pode mudar sua direção de propagação arbitrariamente. A fração de energia incidente que é refletida ao longo de cada sentido possível, após esta interação inicial com a cena, é determinada por um modelo predefinido da *Função de Distribuição da Reflectância e Transmissividade* (BRTDF), que é chamado de *modelo de iluminação local*.

Por outro lado, a fim preservar a praticabilidade computacional do modelo de iluminação global utilizado, supõe-se que as interações subseqüentes (secundárias) da luz com a cena ao longo de cada trajeto de propagação são geometricamente ideais, isto é, as interações subseqüentes da luz são refletidas ou refratadas somente ao longo das direções ideais $\hat{\mathbf{r}}$ e $\hat{\mathbf{t}}$ conforme a teoria de Fresnel, dadas pelas Equações (4.1) e (4.2) [Hearn e Baker, 1994].

$$\hat{\mathbf{r}} = \hat{\mathbf{u}} - (2\hat{\mathbf{u}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}, \quad (4.1)$$

$$\hat{\mathbf{t}} = \left(\frac{\eta_i}{\eta_r} \cos \phi_i - \cos \phi_r \right) \hat{\mathbf{n}} - \frac{\eta_i}{\eta_r} \hat{\mathbf{u}}, \quad (4.2)$$

onde os parâmetros η_i e η_r são os índices da refração incidente e refratada na superfície do material, ϕ_i e ϕ_r são, respectivamente, os ângulos de incidência e refração, $\hat{\mathbf{n}}$ é a normal da superfície e $\hat{\mathbf{u}}$ é o vetor unitário da direção incidente do raio.

Como mostrado na Figura 4.2(b), para um dado pixel, o conjunto de todos os trajetos consistentes com as suposições anteriores geram a *Árvore de Computação da Intensidade do Pixel*. As folhas dessa árvore armazenam a irradiância devido a cada trajeto do raio de propagação vindo de cada fonte de luz; cada nó interno calcula a radiância acumulada que sai de um ponto de reflexão ou refração ao longo de um dado trajeto do raio; a raiz da árvore computa a intensidade do pixel resultante.

Considerando as suposições anteriores, o valor em cada folha da *Árvore de Computação da Intensidade do Pixel* é conhecido *a priori*. Em particular, a irradiância que chega a um ponto $\mathbf{p}$ da cena devido uma fonte de luz $\mathbf{l}_k$ é dada por

$$E = \frac{I}{\text{dist}^2(\mathbf{l}_k, \mathbf{p})}, \quad (4.3)$$

onde I é a intensidade da fonte de luz $\mathbf{l}_k$ e dist é a função que mede a distância entre $\mathbf{l}_k$ e $\mathbf{p}$.

As arestas que conectam as folhas aos nós internos da árvore convertem os valores de irradiância para valores de radiância, cuja forma de conversão é definida pela escolha de um modelo BRTDF. É possível observar que alguns dos modelos mais populares de BRTDF na literatura são parcialmente lineares, de acordo com a seguinte definição:

Definição 4.1 [*BRTDF Parcialmente Linear*] Um modelo BRTDF é parcialmente linear se, e somente se, pode ser escrito como

$$\mathbf{g}(\theta) \cdot \mathbf{x}, \quad (4.4)$$

onde $\mathbf{x}$ é o vetor de parâmetros fotométricos lineares, θ é o vetor de parâmetros fotométricos não-lineares, e $\mathbf{g}(\theta)$ é uma função multi-variada que não depende de $\mathbf{x}$.

A metodologia apresentada nesta tese pode ser utilizada com diferentes modelos de BRTDF que estejam em conformidade com a definição 4.1. Em particular, um destes modelos é o *Modelo de Reflectância de Phong* [Phong, 1975], que expressa a reflectância como uma combinação linear de dois termos: (1) uma parcela difusa, que reflete a luz proveniente de uma fonte de iluminação igualmente em todas as direções; (2) e uma parcela especular que reflete a concentração da luz em uma região da superfície, introduzindo um termo cossenoidal para modelar os efeitos da rugosidade da superfície na especularidade.

Desse modo, para esse modelo temos $\mathbf{x} = \langle k_d, k_s \rangle$, as reflectâncias difusa e especular, respectivamente, $\theta = \langle n_s \rangle$, o expoente do cosseno chamado de *shininess*, e

$$g(\theta) = \left\langle \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \left(\hat{\mathbf{n}} \cdot \hat{\mathbf{h}} \right)^{n_s} \right\rangle,$$

onde $\hat{\mathbf{n}}$ é a normal da superfície, $\hat{\mathbf{l}}$ é a direção da fonte de luz e $\hat{\mathbf{h}}$ é o vetor unitário que dá a direção intermediária entre $\hat{\mathbf{l}}$ e a direção de visualização $\hat{\mathbf{v}}$ (Figura 4.3). Outros modelos BRTDF bastante conhecidos e que são parcialmente lineares e diferenciáveis são aqueles propostos em [Oren e Nayar, 1995], [Wolff et al., 1998] e [Ward, 1992].

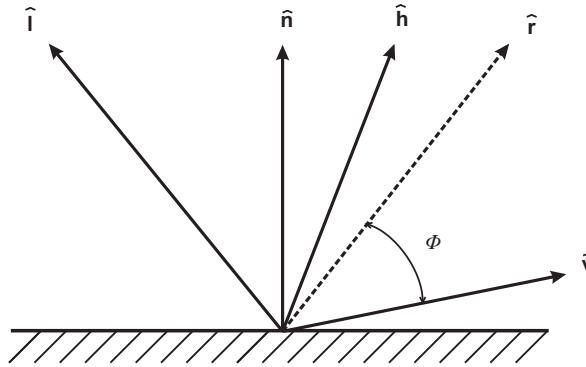


Figura 4.3: Geometria do modelo de Phong. Onde: $\hat{\mathbf{l}}$ é o vetor que dá a direção à fonte luz; $\hat{\mathbf{n}}$ é o vetor normal ao ponto da cena; $\hat{\mathbf{v}}$ é o vetor que dá a direção à câmera; $\hat{\mathbf{r}}$ é o vetor que dá a direção de reflexão ideal; e $\hat{\mathbf{h}}$ é o vetor que dá a direção intermediária entre $\hat{\mathbf{l}}$ e $\hat{\mathbf{v}}$.

As arestas entre os nós internos da Árvore de Computação de Intensidade do Pixel mapeiam radiâncias para radiâncias, a fim de modelar as reflexões e os refrações ideais em que a luz é refletida somente ao longo de uma única direção e refratada somente ao longo de uma outra direção. Em tais casos ideais, se os meios da propagação em ambos os lados da superfície forem transparentes, a fração da energia refletida é dada pelo fator de atenuação de Fresnel

$$F(\theta_i, \theta_r) \stackrel{\text{def}}{=} \frac{1}{2} \frac{\sin^2(\theta_i - \theta_r)}{\sin^2(\theta_i + \theta_r)} \left[1 + \frac{\cos^2(\theta_i + \theta_r)}{\cos^2(\theta_i - \theta_r)} \right], \quad (4.5)$$

onde θ_i e θ_r são, respectivamente, os ângulos de incidência e refração do raio com respeito à normal da superfície. Neste caso, a fração de energia transmitida é $1 - F(\theta_i, \theta_r)$. Entretanto, a fim de modelar o fato de que as superfícies da cena podem realmente absorver uma parte da energia incidente, esses fatores de atenuação são multiplicados, respectivamente, por um parâmetro secundário de reflectividade desconhecido, r , e um parâmetro secundário de transmissividade desconhecido, t , onde ambos devem ser recuperados juntamente com a BRTDF (primária) da superfície.

Desta forma, se um nó interno na árvore receber como entradas uma irradiância incidente E , uma radiância refletida L_r , e uma radiância transmitida L_t , este terá como saída a radiância

$$L_{\text{node}} = E \mathbf{g}(\theta) \cdot \mathbf{x} + r L_r F + t L_t (1 - F), \quad (4.6)$$

onde F é definido na Equação (4.5).

Finalmente, a aresta entre o nó interno superior e a raiz da árvore corresponde à função de transferência do pixel, que mapeia os valores de radiância às intensidades do pixel. Estas funções podem ser conhecidas *a priori* se a câmera for calibrada radiometricamente. Na prática, como mostrado no Capítulo 5, é possível utilizar imagens adquiridas por uma câmera sem calibração radiométrica para extrair as propriedades fotométricas da cena (a menos um fator de escala dado pelo ganho linear em cada canal RGB), uma vez que os ajustes da câmera são feitos para eliminar componentes fortemente não-lineares (tipicamente, a correção gamma) de sua função de transferência.

Como a imagem possui diversos pixels, a estrutura de dados completa utilizada no método é, na verdade, uma floresta com uma árvore para cada pixel da imagem.

4.2 Fatoração Fotométrica

Como discutido no Capítulo 1, o problema de recuperar a reflectância a partir de uma única imagem de uma cena arbitrária é ambíguo. Entretanto, sob certas condições de contorno este problema tem uma solução bem definida. Neste trabalho, em particular, é considerado o caso onde uma cena é composta de N superfícies, cada uma com propriedades fotométricas homogêneas. Além disso, supõe-se que todos os índices de refração dessas superfícies são conhecidos *a priori*, de modo que todos os trajetos possíveis da propagação da luz relevantes ao processo de formação da imagem podem ser computados mesmo antes que as proprieda-

des fotométricas das superfícies sejam conhecidas. Um ponto chave nesta metodologia é o fato que, sob estas circunstâncias, o sistema de equações obtido a partir das Árvore de Computação da Intensidade do Pixel de todos os pixels na imagem de entrada pode ser fatorado em: (1) uma matriz que depende somente da iluminação conhecida, da geometria conhecida e dos parâmetros fotométricos não-lineares desconhecidos; e (2) um vetor dos fatores polinomiais que envolvem somente os parâmetros fotométricos lineares.

Este fato pode ser verificado a partir da forma geral da Árvore de Computação de Intensidade do Pixel, dada na Figura 4.2 e na Equação (4.6). Mais especificamente, seja $\mathbf{x}^{(i)}$ e $\theta^{(i)}$ os parâmetros fotométricos da BRTDF da i -ésima superfície, $i = 1, \dots, N$, e seja $r^{(i)}$ e $t^{(i)}$ os parâmetros secundários de reflectividade e transmissividade, respectivamente. Seja, ainda,

$$\tilde{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{x}^{(1)} \Downarrow \dots \Downarrow \mathbf{x}^{(N)}, \quad \tilde{\theta} \stackrel{\text{def}}{=} \theta^{(1)} \Downarrow \dots \Downarrow \theta^{(N)},$$

onde o operador $\Downarrow$ denota uma concatenação vetorial, e

$$\tilde{\mathbf{r}} \stackrel{\text{def}}{=} \langle r^{(1)}, \dots, r^{(N)} \rangle, \quad \tilde{\mathbf{t}} \stackrel{\text{def}}{=} \langle t^{(1)}, \dots, t^{(N)} \rangle.$$

$$y_i = x_1 r_1 r_2 t_2$$

Considere, agora, um trajeto ascendente que conecta uma folha arbitrária à raiz por meio dos nós internos $D_{\text{path}} \leq D$ em uma Árvore de Computação de Intensidade do Pixel. Se esses nós internos corresponderem aos pontos nas superfícies com os índices $i_1, \dots, i_{D_{\text{path}}}$, a contribuição desse trajeto à radiância total na raiz da árvore é

$$L_{\text{path}} = \mathbf{c}(\tilde{\theta}) \cdot \mathbf{x}^{(i_1)} \prod_{d=2}^{D_{\text{path}}} s^{(i_d)}, \quad \text{onde :} \quad (4.7)$$

$$s^{(i_d)} \stackrel{\text{def}}{=} r^{(i_d)} \quad \text{ou} \quad t^{(i_d)},$$

respectivamente se a aresta entre o $(d-1)$ -ésimo e d -ésimo nós internos do trajeto se referem a uma reflexão ou refração secundária; $\mathbf{c}(\tilde{\theta})$ é o produto entre o termo $E g(\theta)$ da Equação (4.6) e todos os fatores conhecidos de atenuação de Fresnel no trajeto.

Adicionando todas as contribuições de todos os trajetos na árvore, a radiância total na raiz pode ser escrita como

$$L_{\text{pixel}} = \mathbf{a}(\tilde{\theta}) \cdot \mathbf{y}, \quad \text{onde :} \quad (4.8)$$

$$\mathbf{y} \stackrel{\text{def}}{=} \tilde{\mathbf{x}} \otimes \text{Kron}^{D-1} (\langle 1 \rangle \Downarrow \tilde{\mathbf{r}} \Downarrow \tilde{\mathbf{t}}), \quad (4.9)$$

o operador $\otimes$ denota o Produto de Kronecker (ou Produto Tensorial) de dois vetores coluna, e $\text{Kron}^n(\cdot)$ é a Exponencial de Kronecker, recursivamente definida como $\text{Kron}^0(\cdot) = \langle 1 \rangle$, $\text{Kron}^n(\mathbf{v}) = \text{Kron}^{n-1}(\mathbf{v}) \otimes \mathbf{v}$.

O vetor $\mathbf{y}$ na Equação (4.8) é composto dos fatores polinomiais que envolvem somente os parâmetros fotométricos lineares. Cada elemento em $\mathbf{y}$ é obtido multiplicando-se um elemento do vetor $\mathbf{x}$ por uma instância possível do elemento ($D_{\text{path}} - 1$) que aparece na Equação (4.7). Similarmente, o vetor $\mathbf{a}(\tilde{\theta})$ é obtido coletando os elementos de todas as instâncias do vetor $\mathbf{c}(\tilde{\theta})$ gerado pelos trajetos ascendentes que conectam as folhas à raiz da Árvore de Computação da Intensidade do Pixel. Empilhando todas as instâncias da Equação (4.8), uma para cada intensidade na imagem da entrada, obtêm-se o sistema não linear

$$\mathbf{A}(\tilde{\theta}) \mathbf{y} = \mathbf{b}, \quad (4.10)$$

onde $\mathbf{A}(\tilde{\theta})$ é uma matriz com tantas linhas quanto o número de pixels da imagem de entrada e com $\mathcal{O}(N^D)$ colunas, obtidas a partir dos vetores $\mathbf{a}(\tilde{\theta})$; e o vetor $\mathbf{b}$ consiste nas radiâncias obtidas aplicando-se a função de transferência inversa do pixel às intensidades da imagem.

A Equação (4.10), o resultado final da fatoração fotométrica, é a base de um método eficiente para recuperar todos os parâmetros fotométricos de uma cena globalmente iluminada a partir de uma única imagem, que é descrito na Seção 4.3.

4.3 Recuperando as Propriedades Fotométricas

Neste trabalho, o problema de recuperar as propriedades fotométricas da cena a partir de uma imagem de entrada é tratado como um problema de otimização. Mais especificamente, o objetivo do método proposto é minimizar a norma do vetor residual que é equivalente à diferença entre os dois lados da Equação (4.10). Escrevendo essa norma como sendo a norma Euclidiana (que é uma boa escolha se as radiâncias da imagem são contaminadas por um ruído que obedeça a uma distribuição gaussiana simples [Weng et al., 1993]), o objetivo do método torna-se, então

$$\text{encontre} \quad \arg \min_{\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}, \tilde{\theta}} \epsilon \stackrel{\text{def}}{=} \left\| \mathbf{A}(\tilde{\theta}) \mathbf{y}(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}) - \mathbf{b} \right\|^2. \quad (4.11)$$

Esse problema de encontrar o mínimo de uma função não linear bi-dimensional é uma tarefa que demanda um alto custo computacional. De forma a resolver este problema de maneira eficiente, é possível observar que dentro do espaço de parâmetros da medida do erro dado na Equação (4.11), existe um subespaço onde as projeções dessa medida se-

Algoritmo 4 Computando os parâmetros lineares ótimos $\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}$ e o erro fotométrico residual ϵ , para um dado vetor $\tilde{\boldsymbol{\theta}}$ de parâmetros não lineares.

requer: Pré-computação de $\hat{\mathbf{A}}, \hat{\mathbf{b}}, \|\mathbf{b}\|^2$ (Algoritmo 5).

- 1: Decomponha $\hat{\mathbf{A}}$ como $\mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^T$ onde $\boldsymbol{\Lambda}$ é uma matrix diagonal e $\mathbf{V}$ é uma matriz ortogonal;
 - 2: Compute $\boldsymbol{\Lambda}^*$ atribuindo zero a todos os elementos não-significativos de $\boldsymbol{\Lambda}$ e invertendo todos os elementos restantes diferentes de zero;
 - 3: $\mathbf{y} \leftarrow \mathbf{V} \boldsymbol{\Lambda}^* \mathbf{V}^T \hat{\mathbf{b}}$;
 - 4: Recupere os valores iniciais de $\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}$ a partir de $\mathbf{y}$ (Algoritmo 6);
 - 5: **para** um número máximo de iterações pré-definido **faça**
 - 6: Compute $\mathbf{y}$ a partir de $\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}$ como dado na Equação (4.9);
 - 7: $\epsilon \leftarrow (\hat{\mathbf{A}} \mathbf{y} - 2 \hat{\mathbf{b}}) \cdot \mathbf{y} + \|\mathbf{b}\|^2$;
 - 8: **se** ϵ não obteve um decréscimo significativo **então** Pare;
 - 9: $\frac{\partial \epsilon}{\partial \tilde{\mathbf{x}} \downarrow \tilde{\mathbf{r}} \downarrow \tilde{\mathbf{t}}} \leftarrow 2 (\hat{\mathbf{A}} \mathbf{y} - \hat{\mathbf{b}}) \cdot \frac{\partial \mathbf{y}}{\partial \tilde{\mathbf{x}} \downarrow \tilde{\mathbf{r}} \downarrow \tilde{\mathbf{t}}}$;
 - 10: Refine $\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}$ utilizando ϵ e $\frac{\partial \epsilon}{\partial \tilde{\mathbf{x}} \downarrow \tilde{\mathbf{r}} \downarrow \tilde{\mathbf{t}}}$;
 - 11: **fim para**
-

jam polinomiais, geralmente de um grau baixo. Encontrar os mínimos de tais projeções é mais fácil que encontrar os mínimos de funções arbitrárias não-lineares da mesma dimensão [Waki et al., 2005]. Esta observação é explorada decompondo a Equação (4.11) nos seguintes sub-problemas: (1) obter os parâmetros lineares $\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}$, dado um vetor fixo $\tilde{\boldsymbol{\theta}}$; e (2) otimizar os parâmetros não-lineares $\tilde{\boldsymbol{\theta}}$, supondo que há um oráculo para computar os mínimos no sub-espço polinomial da medida do erro. As soluções para estes dois sub-problemas são apresentadas nos Algoritmos 4 e 7.

4.3.1 Otimização Polinomial

A idéia principal do Algoritmo 4 é: uma vez que $\tilde{\boldsymbol{\theta}}$ é fixado, a Equação (4.10) pode ser tratada como um sistema linear, por meio do qual $\mathbf{y}$ pode ser obtido diretamente. Infelizmente, os vetores $\mathbf{y}$ encontrados desta maneira não serão necessariamente consistentes com sua forma polinomial definida na Equação (4.9), devido ao mal condicionamento inerente ao sistema. Assim, é necessário extrair um conjunto de parâmetros consistentes $\tilde{\mathbf{x}}, \tilde{\mathbf{r}}, \tilde{\mathbf{t}}$ desta solução linear inicial (como indicado no passo 4 do algoritmo) e então refiná-lo iterativamente a fim encontrar um mínimo no sub-espço polinomial da medida do erro.

Uma outra idéia essencial para tornar o Algoritmo 4 eficiente é expressar a medida do erro dado na Equação (4.11) em suas derivadas com respeito às incógnitas $\mathbf{y}$ e $\tilde{\boldsymbol{\theta}}$ em função

da matriz normalizada $\hat{\mathbf{A}}$ e do vetor normalizado $\hat{\mathbf{b}}$, definidos como

$$\hat{\mathbf{A}} \stackrel{\text{def}}{=} \mathbf{A}(\tilde{\theta})^T \mathbf{A}(\tilde{\theta}), \quad \hat{\mathbf{b}} \stackrel{\text{def}}{=} \mathbf{A}(\tilde{\theta})^T \mathbf{b}. \quad (4.12)$$

Isto é especialmente importante nas imagens onde o número das superfícies com propriedades fotométricas distintas (N) é muito menor do que o número de pixels da imagem, e a profundidade (D) do *Ray Tracing* é pequena. Neste cenário, enquanto a matriz $\mathbf{A}(\tilde{\theta})$ pode não ser alocada inteiramente na memória principal, a matriz $\hat{\mathbf{A}}$ é quadrada no número de colunas da matriz anterior, com um total de $\mathcal{O}(N^{2D})$ elementos que permitem tratar o sistema em memória primária e de forma mais eficiente.

A normalização das matrizes de cada banda é obtida por meio dos valores extraídos diretamente da Árvore de Computação de Intensidade do Pixel. Para minimizar o uso de memória, em vez de gerar a matriz $\mathbf{A}(\tilde{\theta})$ explicitamente e então pré-multiplicá-la pela sua transposta, é computado o produto externo de cada linha de $\mathbf{A}(\tilde{\theta})$ consigo mesma e então adicionado o resultado de cada operação a um acumulador com $\mathcal{O}(N^{2D})$ elementos, assim apenas uma linha da matriz original é extraída por vez e em seguida eliminada, como apresentado no Algoritmo 5.

Dois pontos do Algoritmo 4 serão particularmente discutidos nesta seção, os passos 10 e 4 na ordem de importância.

Algoritmo 5 Algoritmo de normalização do sistema de equações.

- 1: **para** m de 1 até o número de linhas de $\mathbf{A}(\tilde{\theta})$ **faça**
 - 2: Computa linha $\mathbf{A}(\tilde{\theta})[m]$ relativa ao pixel m da imagem, a partir da Árvore de Computação de Intensidade do Pixel;
 - 3: **para** i de 1 até o número de colunas de $\mathbf{A}(\tilde{\theta})$ **faça**
 - 4: **para** j de 1 até o número de colunas de $\mathbf{A}(\tilde{\theta})$ **faça**
 - 5: $\hat{\mathbf{A}}[i][j] = \hat{\mathbf{A}}[i][j] + \mathbf{A}(\tilde{\theta})[m][i]\mathbf{A}(\tilde{\theta})[m][j]$;
 - 6: **fim para**
 - 7: $\hat{\mathbf{b}}[i] = \mathbf{A}(\tilde{\theta})[m][i] * \mathbf{b}[i]$;
 - 8: **fim para**
 - 9: Descarta $\mathbf{A}(\tilde{\theta})[m]$ corrente;
 - 10: **fim para**
-

O passo 10 do Algoritmo 4 deve idealmente encontrar uma solução mais próxima de um mínimo global da medida do erro que a solução atual. Mesmo que seja difícil garantir isto para medidas arbitrárias do erro, um conjunto crescente de evidências sugere que isto pode ser feito eficientemente no caso onde a medida do erro é uma soma dos quadrados de polinômios de baixa ordem [Waki et al., 2005], como na Equação (4.11). Na presente

Algoritmo 6 Valores iniciais de $\tilde{x}$, $\tilde{r}$, $\tilde{t}$ a partir de $\mathbf{y}$.

assegure: Mudanças em $\mathbf{y}$ têm efeito apenas neste escopo.

- 1: **enquanto** $\exists x \in \tilde{x}, \tilde{r}, \tilde{t}$ com um valor desconhecido **faça**
 - 2: **se** $\exists x \in \tilde{x}, \tilde{r}, \tilde{t}, \exists y \in \mathbf{y} \mid y = x^n, n \geq 1$ **então**
 - 3: Escolha um certo x e compute seu valor para cada $y \in \mathbf{y} \mid y = x^n, n \geq 1$;
 - 4: Faça x igual à média dos valores computados no passo anterior;
 - 5: **senão**
 - 6: Escolha um $x \in \tilde{r}, \tilde{t}$ arbitrário;
 - 7: Faça x igual a um valor arbitrário fisicamente consistente;
 - 8: **fim se**
 - 9: Atualize os valores de x em $\mathbf{y}$ de acordo com a Equação (4.9);
 - 10: **fim enquanto**
-

implementação do Algoritmo 4, a etapa 10 é definida simplesmente como uma iteração do Método do Gradiente Conjugado padrão [Press et al., 1992]. Entretanto, num trabalho futuro pretende-se utilizar neste passo métodos numéricos projetados especificamente para a otimização polinomial.

O passo 4, por outro lado, deve idealmente encontrar o conjunto de parâmetros $\tilde{x}, \tilde{r}, \tilde{t}$ que venham a satisfazer a relação descrita na Equação (4.9), com respeito a um vetor arbitrário $\mathbf{y}$. Uma vez que o resultado deste procedimento é usado somente para inicializar os parâmetros $\tilde{x}, \tilde{r}, \tilde{t}$ (que são refinados mais tarde), o mesmo é implementado como uma heurística simples descrita no Algoritmo 6.

É importante frisar que os passos 6 e 7 do Algoritmo 6 são executados somente se a imagem da entrada for ambígua. Isto acontece, por exemplo, se a imagem de entrada contiver a projeção de um espelho ideal ($r = 1$) que reflete somente a imagem de uma única superfície, situada fora do campo de visão da câmera. Tal imagem é idêntica a uma outra ajustando-se o r do espelho a 0,5 e dobrando-se todos os parâmetros lineares de BRTDF da superfície refletida, por exemplo. Nesse tipo da situação, a imagem de entrada simplesmente não contém informações suficientes para definir de forma única todos os parâmetros fotométricos. Apesar disso, a metodologia proposta recupera uma das várias interpretações fisicamente consistentes de uma imagem de entrada ambígua.

4.3.2 Otimização Não Polinomial

A fim de manter o caráter geral desta formulação, é assumido que a parte não-linear das BRTDFs das superfícies – isto é, o fator $g(\theta)$ da Equação (4.4) – pode ser toda a função multi-variada diferenciável. Como consequência disto, o fator $A(\tilde{\theta})$ da Equação (4.11) não

Algoritmo 7 Reflectância inversa eficiente.

-
- 1: Escolha um valor inicial para $\tilde{\theta}$;
 - 2: **para** um número de iterações pré-definido **faça**
 - 3: Compute $\hat{\mathbf{A}}$, $\hat{\mathbf{b}}$ e $\|\mathbf{b}\|^2$ conforme o Algoritmo 5;
 - 4: Execute o Algoritmo 4 para obter $\tilde{\mathbf{x}}$, $\tilde{\mathbf{r}}$, $\tilde{\mathbf{t}}$, ϵ .
 - 5: **se** ϵ não obteve um decréscimo significativo **então** Pare;
 - 6: Compute $\mathbf{y}$ a partir de $\tilde{\mathbf{x}}$, $\tilde{\mathbf{r}}$, $\tilde{\mathbf{t}}$ conforme a Equação (4.9);
 - 7: $(\forall \phi \in \tilde{\theta}) \frac{\partial \epsilon}{\partial \phi} \leftarrow (\mathbf{A}(\tilde{\theta}) \mathbf{y} - \mathbf{b}) \cdot \left(\frac{\partial \mathbf{A}}{\partial \phi} \mathbf{y} \right)$;
 - 8: Encontre $\tilde{\theta}$ utilizando ϵ e $\frac{\partial \epsilon}{\partial \theta}$;
 - 9: **fim para**
-

possui nenhuma forma especial que possa ser explorada, o que faz com que o Algoritmo 7 seja mais simples que o Algoritmo 4.

A idéia principal do Algoritmo 7 é a de computar somente uma vez todos os coeficientes conhecidos das Árvores de Computação de Intensidade do Pixel, e então armazená-los em uma estrutura de dados que seja percorrida seqüencialmente todas as vezes que os passos 3 ou 7 são executados. Isto evita a necessidade de múltiplas execuções de um procedimento computacionalmente caro para renderizar uma imagem sintética da cena sob uma iluminação global, como feito em [Boivin e Gagalowicz, 2001] e [Boivin e Gagalowicz, 2002].

Mais uma vez, a metodologia é ajustada para ser especialmente eficiente quando o número de superfícies com propriedades fotométricas distintas é muito menor que o número de pixels da imagem da cena, e a profundidade do *Ray Tracing* é pequena. Por exemplo, no passo 3, utilizamos o procedimento do Algoritmo 5, o que permite maior eficiência computacional. Um procedimento similar é usado para garantir que em uma única visita a cada Árvore de Computação da Intensidade do Pixel, o Algoritmo 7 computa todas as derivadas parciais do passo 7, sempre sem gerar a matriz $\mathbf{A}(\tilde{\theta})$ explicitamente, que é potencialmente grande se consideramos o uso de memória primária.

Similar ao que é feito na implementação do Algoritmo 4, o passo 8 do Algoritmo 7 é executado simplesmente como uma iteração do Método do Gradiente Conjugado. Entretanto, desde que nesse último pouco se conhece *a priori* sobre a natureza da medida do erro, estabelecer um limite de iterações na otimização para ganhar alguma eficiência é uma atitude necessária e bastante razoável.

4.4 Análise de Complexidade do Método

4.4.1 Complexidade de Tempo

O método proposto possui uma complexidade de tempo que depende do número de iterações de um método de otimização, que por sua vez depende da convergência de sua função objetivo. Entretanto, para fins de cálculo da complexidade, é assumido que este número de iterações é estimável dentro de um limite razoável. A análise é feita a partir do Algoritmo 7, e é afetada, basicamente, pelos seguintes aspectos:

- O custo no passo 3 é dado pela normalização das matrizes feitas no Algoritmo 5, que é $\mathcal{O}(num.N^{2D})$ operações de ponto flutuante, sendo N é o número de superfícies com materiais distintos que formam a cena, D a quantidade de reflexões ou de refrações do raio nas superfícies da cena (profundidade do *Ray Tracing*) e num o número de pixels da imagem de entrada;
- O custo no passo 4 é dado pela otimização polinomial dos parâmetros lineares, conforme o Algoritmo 4, que é $\mathcal{O}(iter.N^{2D})$ operações de ponto flutuante, sendo $iter$ o número de iterações do método de otimização utilizado;
- O custo dos passos 6, 7 são inferiores ao do passo 8, que faz a busca dos valores de $\tilde{\theta}$ utilizando ϵ e $\frac{\partial \epsilon}{\partial \theta}$, e é $\mathcal{O}(iter.num.p)$ operações de ponto flutuante, sendo p o custo de computação de um pixel por meio da Árvore de Computação de Intensidade do Pixel.

O ciclo que envolve os passos 2 a 9 é dado por um número de ciclos nc de iterações determinado pelo usuário como uma relação de compromisso entre eficiência e qualidade do resultado, não dependendo portanto da convergência do erro. Com isto, o custo total do método é dado por: $\mathcal{O}(nc.[iter.num.p + \mathcal{O}(iter.N^{2D}) + \mathcal{O}(N^{2D})])$. Para o pior caso onde N e D são muito grandes, o custo computacional do método é exponencial dado por $\mathcal{O}(nc.iter.N^{2D})$. Entretanto, na prática N é limitado ao tamanho da cena e pequenos valores de D são suficientes para representá-la, o que leva o método a um custo polinomial dado por $\mathcal{O}(nc.iter.num.p)$ na maioria dos casos.

4.4.2 Complexidade de Espaço

Em relação à complexidade de espaço, o método possui duas estruturas de dados principais que podem tornar crítico o uso de memória primária:

- As Árvores de Computação de Intensidade do Pixel possuem uma complexidade de espaço que depende da profundidade permitida aos raios do *Ray Tracing*, da quantidade de inter-reflexões e transmissões entre os objetos, e do tamanho da imagem da cena, dada por $\mathcal{O}(num.D2^{D-1})$ no pior caso, se for considerado que todas as N superfícies possuem os parâmetros secundários $\langle r, t \rangle$ diferentes de zero.
- A matriz de restrições fotométricas lineares $A(\tilde{\theta})$, que originalmente teria uma complexidade proibitiva de $\mathcal{O}(num.N^D)$, mas devido ao Algoritmo 5 de normalização tem complexidade $\mathcal{O}(N^{2D})$, o que vem a ser um ganho para os casos onde $num \gg N^D$ (bastante comum na prática).

Portanto, com o uso da normalização, a complexidade de espaço do método é dada pelas Árvores de Computação de Intensidade do Pixel, sendo $\mathcal{O}(num.D2^{D-1})$, o que vem a ser um dado importante, pois dependendo do tamanho num da imagem e da profundidade D do *Ray Tracing*, a memória primária disponível pode não ser suficiente, o que implicaria em um aumento considerável no tempo de processamento do método devido ao acesso à memória secundária.

4.5 Limitações do Método

Não obstante o avanço que o método proporciona ao que já existe para o problema tratado, o mesmo possui algumas limitações que são discutidas a seguir.

O uso de modelos BRTDF é restrito aos modelos que estejam em conformidade com a definição 4.1, ao qual são denominados modelos parcialmente lineares. Embora seja uma restrição, a maioria dos modelos BRTDF estão de acordo com essa definição, o que torna o método bastante abrangente.

O método depende do conhecimento, *a priori*, dos índices de refração dos materiais contidos na cena. Os índices que são essenciais para o funcionamento do método são aqueles de objetos translúcidos e podem ser obtidos de uma tabela padrão de índices de refração, caso os materiais sejam conhecidos *a priori*. Caso contrário, os índices podem ser estimados iterativamente, antes da análise fotométrica, minimizando as distâncias entre as projeções na imagem real das características da cena que são visíveis através dos objetos transparentes e das projeções dessas características na imagem gerada pelo *Ray Tracing*.

O método considera que as superfícies são homogêneas, ou seja, não possuem textura. Esta limitação pode ser resolvida com uma extensão do método para os casos em que as superfícies possuam textura, o que entretanto será objeto de trabalhos futuros.

Capítulo 5

Implementação e Resultados

Neste capítulo são apresentados os detalhes de implementação e os resultados obtidos até o momento, como forma de validar as metodologias propostas nos Capítulos 3 e 4. O mesmo está dividido em duas seções principais, que envolvem os dois cenários contendo problemas distintos relacionados a modelagem em ambientes com iluminação não local.

5.1 Estéreo Subaquático

O método proposto no Capítulo 3 apresenta o estéreo subaquático como uma abordagem a ser utilizada enquanto solução do problema de modelagem geométrica em cenas com iluminação não local. Inicialmente, são apresentados detalhes importantes para a efetivação dos experimentos, como o método de calibração das câmeras e retificação das imagens, e em seguida apresentados os resultados comparativamente a um método de estéreo tradicional.

5.1.1 Ambiente Experimental

Os resultados experimentais foram obtidos em um ambiente contendo objetos geometricamente simples inseridos em um aquário com 200l de água de dimensões $30 \times 120 \times 60cm$, e cujas imagens foram adquiridas em duas câmeras Sony DFW-X900 dispostas ortogonalmente entre si para permitir a obtenção de diferentes distâncias em espessura d'água para cada imagem do par estéreo, dificultando ainda mais o problema da correspondência, conforme a Figura 5.1.

Essa configuração da disposição das câmeras permite verificar a robustez do método, uma vez que as distâncias entre a cena e as duas câmeras são bastante diferentes, gerando uma situação crítica na implementação de métodos de estéreo convencional. Nas imagens

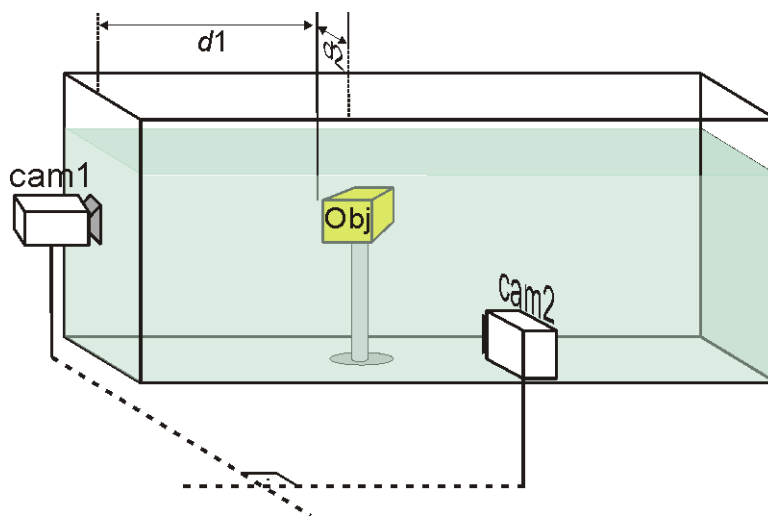


Figura 5.1: Ambiente experimental do estéreo subaquático. Um par de imagens do objeto submerso é adquirido pelas câmeras 1 e 2, dispostas ortogonalmente entre si de tal forma que $d_1 > d_2$.

obtidas, as cores dos objetos da cena variam com a distância em espessura d'água que nos experimentos chega a ser de 112cm para uma câmera e 30cm para outra.

A imagem de referência é adquirida pela câmera 1 que se encontra encostada no vidro do aquário, enquanto que a câmera 2 está numa posição ortogonal à primeira e se encontra a uma dada distância do aquário, que deve ser calculada e desconsiderada de forma a obter somente a distância relativa à espessura d'água. Isto é feito estabelecendo-se um plano que corresponda à parede do aquário que é fronto-paralela à câmera 2, e obtendo a intersecção entre esse plano e uma reta oriunda da câmera ortogonalmente em direção ao plano. A distância da câmera ao objeto menos a distância da câmera ao plano resultam na distância em espessura d'água d_2 .

5.1.2 Calibração das Câmeras

A calibração geométrica das câmeras, e por consequência do sistema estéreo, consiste em obter os parâmetros intrínsecos (distância focal, coeficiente de distorção radial, centro da imagem, e razão de aspecto do pixel) e extrínsecos (matrizes de rotação e translação) de cada uma das câmeras usadas [Trucco e Murino, 2000], e em seguida determinar os parâmetros extrínsecos do sistema estéreo.

A calibração foi efetuada a partir de um *ToolBox* de calibração para *Matlab*, desenvolvido por Jean-Yves Bouguet e disponibilizado em [Bouguet, 1999]. Para efetuar a calibração é necessário ter à disposição um alvo de calibração planar branco com quadrados negros de

<i>Câmera</i>	<i>WBU</i>	<i>WBV</i>	<i>Hue</i>	<i>Sat</i>	<i>Gain</i>	<i>Shutter</i>
<i>Sony 60996</i>	112	176	132	132	2102	2048
<i>Sony 60997</i>	116	169	132	132	2102	2048

Tabela 5.1: Parâmetros das câmeras utilizadas na aquisição do par estéreo.

dimensões e espaçamento entre estes conhecidos, neste caso de $25mm$. No mínimo duas imagens do alvo de calibração devem ser adquiridas pelas duas câmeras, simultaneamente, com o alvo em diferentes posições no espaço coberto pelo ângulo de visada do sistema estéreo (as duas câmeras devem “enxergar” o alvo).

Para exemplificar, a Figura 5.2 apresenta as posições de cada alvo de calibração e o resultado da calibração dos parâmetros extrínsecos de cada câmera e do par estéreo adquirido na configuração dada pela Figura 5.1.

O erro da calibração é obtido através da re-projeção dos pontos de cada quina do alvo de calibração nas quinas correspondentes no plano de imagem. Em todos os experimentos o erro geométrico foi inferior a 0.5 pixels.

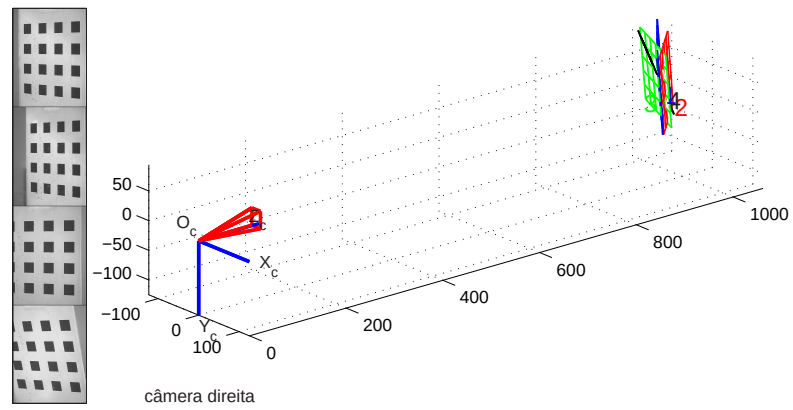
A calibração radiométrica das câmeras consiste em determinar uma transformação que explique a diferença entre as intensidades das cores de imagens obtidas da mesma cena, mas por diferentes câmeras. Em geral essa transformação é linear do tipo $ax + b$, e deve ser aplicada às imagens adquiridas como forma de compensação da diferença de resposta de uma câmera em relação à outra. Entretanto, como o meio já impõe uma diferença radiométrica entre as imagens das duas câmeras, de característica não linear, o uso dessa função de compensação torna-se inadequado, pois iria alterar significativamente as componentes de atenuação e dispersão do meio, descaracterizando o modelo de propagação de luz exposto na Seção 3.1.

Como opção, a calibração é feita através de um ajuste manual dos parâmetros de cada câmera, ambas obtendo um alvo padrão na mesma posição e sob as mesmas condições de iluminação, de forma a tornar as respostas das câmeras equivalentes, com os resultados conforme a Tabela 5.1 :

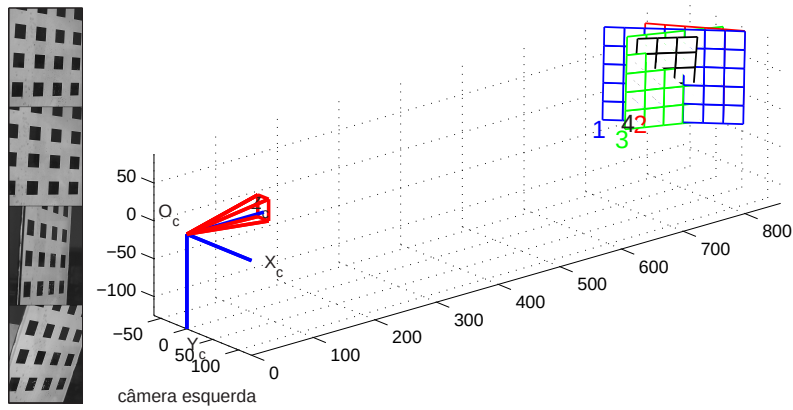
Após o ajuste foi medido o erro médio com base no alvo da Figura 5.3. O erro médio de cada banda é superior a 3 e inferior a 6 níveis de cinza.

5.1.3 Retificação das Imagens

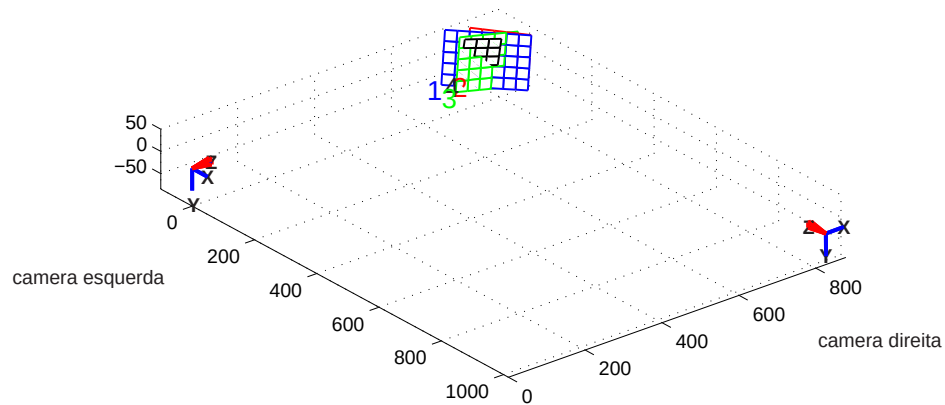
Esta etapa é importante pois permite fazer uso da restrição da geometria epipolar para reduzir o espaço de busca do algoritmo de estéreo. O processo de retificação das imagens consiste em alinhar as imagens, através de translações e rotações com relação às suas posições originais,



(a) calibração da câmera direita



(b) calibração da câmera esquerda



(c) calibração do sistema estéreo

Figura 5.2: Em (a) e (b) as posições dos alvos de calibração e o resultado gráfico das calibrações dos parâmetros extrínsecos das câmeras direita e esquerda, respectivamente. Em (c) o resultado gráfico da calibração dos parâmetros extrínsecos do sistema estéreo.

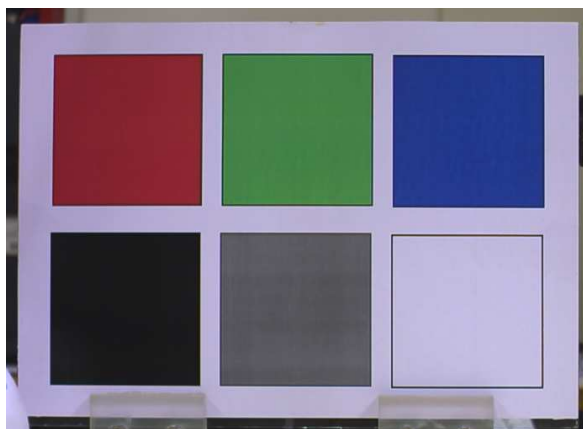


Figura 5.3: Alvo de calibração radiométrica.

de tal forma que pares de linhas epipolares correspondentes tornam-se colineares e paralelas a um dos eixos da imagem, em geral o eixo horizontal.

A retificação das imagens do par estéreo é efetuada com base no método desenvolvido em Fusiello et al. [Fusiello et al., 2000]. Os autores disponibilizaram um código implementado em *Matlab* para uso em pesquisas. O programa tem como entradas o par de imagens e as matrizes de projeções perspectivas, que podem ser definidas a partir das matrizes de parâmetros extrínsecos do sistema estéreo e dos parâmetros intrínsecos de cada câmera. Como resultado, obtém-se duas matrizes de transformação, uma para cada imagem do par estéreo, que são usadas na retificação das imagens.

As imagens retificadas podem ser projetadas em posições externas ao plano de imagem original, alterando dessa forma o tamanho das mesmas, o que pode implicar em perda de informações importantes dos objetos contidos na cena. Dessa forma, as matrizes geradas devem ser corrigidas com base nas dimensões das imagens de entrada, para que a operação de retificação retorne imagens com a mesma dimensão que as imagens originais. Os resultados da retificação são apresentados nas Figuras 5.5 e 5.8, nas seções específicas de cada experimento realizado.

5.1.4 Adaptação do Algoritmo de Estéreo Convencional no Estéreo Subaquático

O algoritmo de estéreo EMGC [Kolmogorov e Zabih, 2002] (Seção 2.3.3) é utilizado como parte integrante do método apresentado no Algoritmo 2, e irá atuar na determinação de um valor ótimo das disparidades entre o par estéreo e, conseqüentemente, influenciará no cálculo das distâncias d_1 e d_2 relativas à câmera da esquerda e da direita, respectivamente, em relação

à cena. A adaptação consiste na modificação da maneira de obtenção do termo $\mathcal{E}_{dados}$ descrito no funcional de energia dado pela Equação (2.7), que passa então a retornar o valor da diferença entre as intensidades corrigidas de dois pixels que estão separados por uma disparidade $\mathcal{D}$.

O problema da correspondência se mantém inalterado, porém novos cálculos devem ser efetuados para a determinação dos valores corrigidos de intensidade dos pixels. A obtenção do termo E_{dados} para dois pixels p_1 e p_2 em planos de imagens diferentes é descrito na Equação (5.1).

$$\mathcal{E}_{dados}(p_1, p_2) = |L_{Obj}(d_1) - L_{Obj}(d_2)|, \quad (5.1)$$

onde, $p_2 = p_1 + disp$, e d_1 e d_2 são obtidos através do método de triangulação, que consiste em determinar a menor distância entre duas retas que partem do centro de projeção de cada uma das câmeras (Seção 2.3.1). $L_{Obj}(d)$ é a radiância do pixel corrigido através da Equação (3.9), dado o valor de d e conhecidos os parâmetros do modelo de propagação da luz na água.

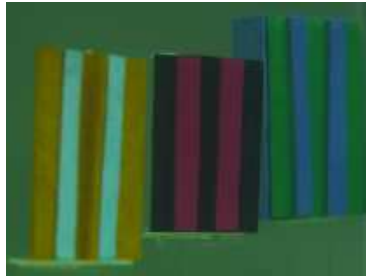
5.1.5 Resultados do Estéreo Subaquático com Estimativa Experimental dos Parâmetros da Água

Os resultados a seguir foram obtidos utilizando o método de estimativa dos parâmetros do modelo, dado pelo Algoritmo 1 na Seção 3.2, e o método de Estéreo Subaquático dado pelo Algoritmo 2 na Seção 3.3, aqui denominado de UWS (*Underwater Stereo*). São apresentados 3 experimentos com níveis de turbidez baixo, médio e alto, de pares de imagens obtidos em ambiente controlado e as cores dos objetos são definidas de forma a facilitar a correspondência para um algoritmo de estéreo convencional.

5.1.5.1 Cena com Água Limpa (Turbidez Baixa)

Neste primeiro experimento utilizamos uma cena com três objetos de geometria simples imersos em água tratada. A principal dificuldade deste experimento consiste em obter a correlação dos pixels para o objeto mais distante, que sofre maiores efeitos da degradação devido ao meio em relação aos objetos mais próximos. A cor de fundo (*background*) apresenta uma cor esverdeada em função do vidro do aquário e é acentuada pelos efeitos da água, gerando ambigüidades na correlação dos pixels da região de cor verde do último objeto. A Figura 5.4 apresenta o par de imagens obtido e a Figura 5.5 o par de imagens retificado. Os

dados de estimativa dos parâmetros de atenuação e dispersão deste experimento se encontram na Tabela 3.3 apresentada na Seção 3.2.

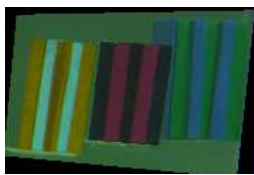


(a) imagem esquerda



(b) imagem direita

Figura 5.4: Imagens adquiridas (a) e (b) do experimento com turbidez baixa.



(a) imagem esquerda



(b) imagem direita

Figura 5.5: Imagens retificadas (a) e (b) do experimento com turbidez baixa

Como teste comparativo, utilizamos o EMGC original e o UWS utilizando o EMGC como parte integrante do método. A Figura 5.6 apresenta os mapas de disparidade obtidos. A imagem de referência utilizada é a imagem da esquerda, portanto cores mais escuras no mapa representam distâncias maiores, enquanto que cores mais claras representam distâncias menores. A ausência de cor, completamente branco neste caso, representa as oclusões.



(a) mapa de EMGC



(b) mapa de UWS

Figura 5.6: (a) e (b) os mapas de disparidade da cena do experimento com turbidez baixa.

Os resultados dados pelos mapas de disparidade da Figura 5.6 demonstram que o método

Canal	parâmetros		
	α	β	$E(\infty)$
R	0.0076	0.0031	81.57
G	0.0070	0.0051	110.47
B	0.0110	0.0044	119.15

Tabela 5.2: parâmetros α , β em cm^{-1} e $E(\infty)$ em pixels, do experimento com turbidez média.

EMGC de estéreo denso teve diversas falhas ao longo da cena, em particular no que se refere ao último objeto da cena (mais profundo). Isso ocorreu principalmente porque o último objeto é o que possui maior degradação do meio devido à maior distância em espessura d'água, dificultando a correspondência entre os pontos e obtenção do mapa de disparidades. Por outro lado, o método UWS apresentado nesta tese conseguiu obter resultados mais consistentes, apresentando no mapa as disparidades dos três objetos, ainda que com uma falha na correspondência das listas do primeiro objeto (mais próximo). Essa falha se deve ao fato de que as diferenças do nível de intensidade de cor nessa região entre as duas imagens é acentuada devido ao ruído da imagem e ao erro residual da estimativa dos parâmetros do modelo de propagação da luz na água, dificultando sua correspondência.

5.1.5.2 Cena com Água + Leite (Turbidez Média)

No segundo experimento utilizou-se uma cena com três objetos simples imersos em um meio modificado, adicionando-se 15ml de leite à água e assim aumentando sua turbidez, com conseqüente aumento da atenuação e do efeito *waterlight*. Neste caso os objetos possuem uma textura que visa facilitar o algoritmo de estéreo tradicional, sem repetições de cores que possam gerar ambigüidades. Isso ocorre para certificar que as falhas sejam exclusivamente devido ao meio e não devido às dificuldades da cena.

A Figura 5.7 apresenta o par de imagens obtidos, e a Figura 5.8 este mesmo par após o processo de retificação. Os dados de estimativa dos novos parâmetros para esse meio seguem a metodologia da Seção 3.2, e se encontram na Tabela 5.2.

A imagem de referência utilizada é a imagem da esquerda, portanto as cores no mapa que representam as distâncias e oclusões seguem o mesmo padrão da seção anterior. Os métodos EMGC e UWS foram aplicados ao par estéreo de imagens e os mapas de disparidade obtidos são apresentados na Figura 5.9. Como os objetos possuem uma textura que visa facilitar a correspondência, o método EMGC obteve um mapa com as disparidades dos três objetos, entretanto com falhas na correspondência em algumas regiões da cena, causando falsas oclusões em partes do mapa. O método UWS apresentou um mapa de disparidades

sem falsas oclusões e com resultados mais uniformes, apresentando uma considerável melhoria em relação ao método EMGC. É importante citar que ambos os métodos apresentaram uma falha no que deveria ser uma oclusão no primeiro objeto (a faixa marron não é visível na imagem da direita) e que foi casado com o terceiro objeto. Além disso, os mapas não apresentam uma uniformidade nas disparidades do mesmo objeto, o que entretanto é uma limitação própria dos algoritmos de estéreo denso quando a cena possui uma textura similar a do experimento em questão.

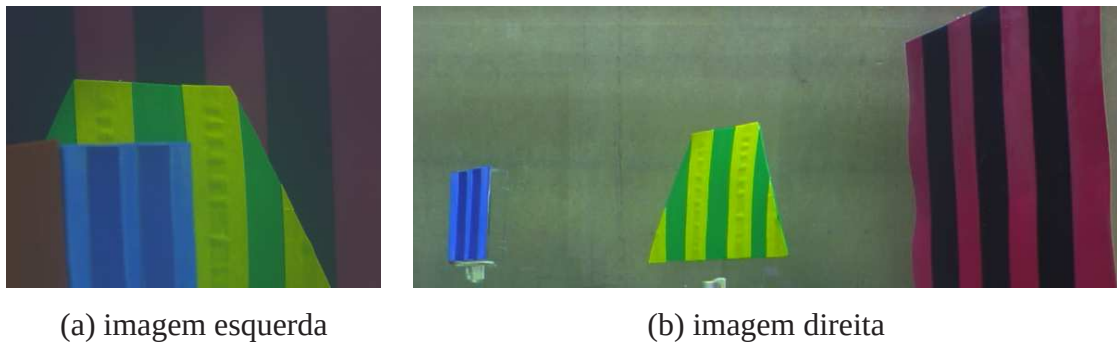


Figura 5.7: Imagens adquiridas (a) e (b) do experimento com turbidez média.

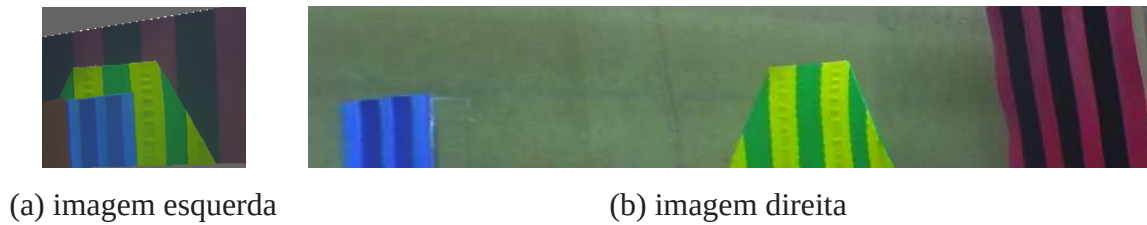


Figura 5.8: Imagens retificadas (a) e (b) do experimento com turbidez média.

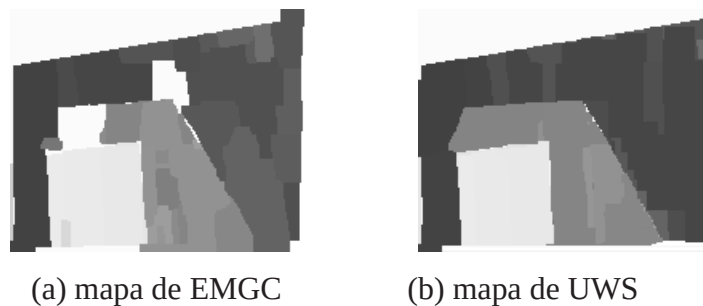


Figura 5.9: (a) e (b) os mapas de disparidade da cena do experimento com turbidez média.

5.1.5.3 Cena com Água + Argila (Turbidez Alta)

De modo a testar a robustez do método, realizamos um experimento onde aproximadamente 150g de argila foram adicionadas em 200 litros de água limpa, de forma a aumentar significativamente a turbidez do meio. Os objetos e suas posições dentro do aquário são praticamente as mesmas do experimento com turbidez média. Na Figura 5.10, podemos ver a imagem da cena real sem água e a mesma cena adquirida no aquário com água+argila, o que permite perceber o quão degradada é a imagem da cena.

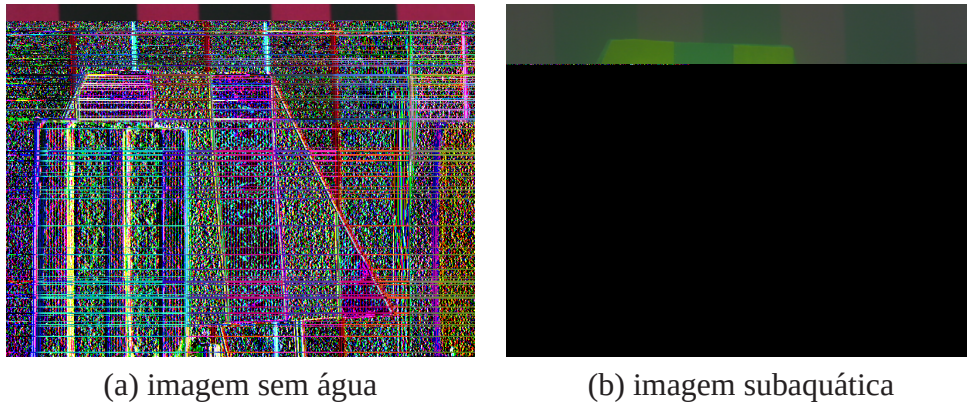


Figura 5.10: Imagens da cena utilizada no experimento com turbidez alta.

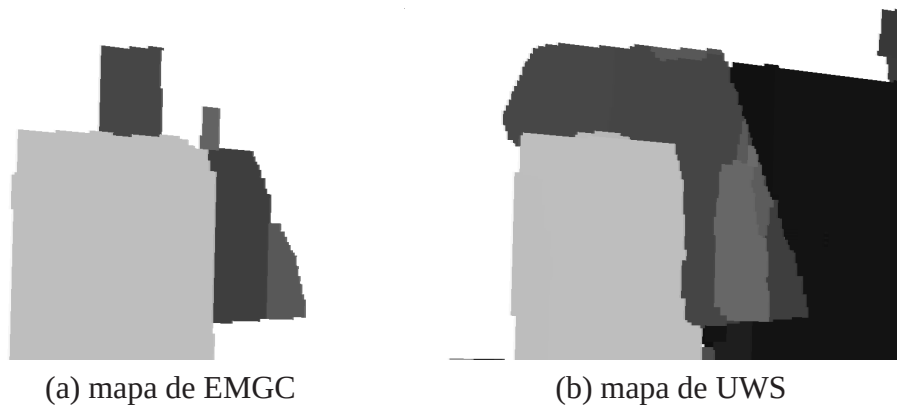


Figura 5.11: (a) e (b) os mapas de disparidade da cena do experimento com turbidez alta.

Os métodos EMGC e UWS foram aplicados a um par estéreo adquirido neste ambiente e os mapas de disparidade obtidos são apresentados na Figure 5.11. Devido à turbidez expressiva do meio, o último objeto da cena teve suas intensidades originais bastante deterioradas, o que fez com que a correspondência entre os pixels do par de imagens fosse bastante difícil, e levou o método EMGC a não realizar, inclusive, correspondência alguma para este objeto, além de apresentar também falhas significativas nas disparidades do segundo objeto. Por

<i>Canal</i>	<i>parâmetros</i>		
	α	β	$E(\infty)$
<i>R</i>	0.0097	0.0023	81.57
<i>G</i>	0.0093	0.0034	110.47
<i>B</i>	0.0099	0.0021	119.15

Tabela 5.3: Parâmetros α , β em cm^{-1} obtidos automaticamente e $E(\infty)$ em pixels obtido do experimento com turbidez média.

outro lado, o método UWS conseguiu recuperar uma parcela significativa da disparidade do terceiro objeto, além de adquirir corretamente a disparidade para os outros dois objetos, o que demonstra sua robustez a aplicabilidade mesmo quando o meio possui características de atenuação e dispersão expressivas.

Estes resultados, assim como os resultados dos experimentos com turbidez baixa e média, demonstram que o método proposto nesta tese apresenta avanços significativos na modelagem 3D de cenas em ambientes subaquáticos, principalmente quando comparado a um método de estéreo denso.

5.1.6 Resultados do Estéreo Subaquático com Estimativa Automática dos Parâmetros da Água

O resultado do experimentos desta seção foi obtido com o uso do Algoritmo 3 dado na Seção 3.4, denominado AUWS (*Automatic Underwater Stereo*). Para fins de comparação, foi utilizado o mesmo par estéreo de imagens adquiridos no experimento de turbidez média da Seção 5.1.5.2. Isto vai permitir verificar a qualidade do mapa de disparidades obtido com e sem a estimativa experimental dos parâmetros do modelo.

Os valores iniciais dos parâmetros α e β de atenuação e dispersão foram testados em diversas ordens de grandeza (0.0001, 0.001, 0.01, e 0.1), e observou-se que o algoritmo não converge para uma solução se a entrada não for, pelo menos, da mesma ordem de grandeza dos valores estimados dos parâmetros. A Tabela 5.3 apresenta os resultados finais dos parâmetros. A diferença entre estes valores e os obtidos pela estimativa experimental dada na Tabela 5.2 confirmam que, embora sejam uma solução para o problema da correspondência e permitam obter o mapa de disparidades, os parâmetros estimados automaticamente não necessariamente são os parâmetros reais do meio, ou seja, não é possível utilizá-los para restaurar fielmente as imagens. isso ocorre devido ao fato de que a otimização faz a busca para a menor energia que permita a maior correspondência, e isso pode acontecer para um conjunto de possibilidades de valores dos parâmetros.

Como nos casos anteriores, a imagem de referência utilizada é a imagem da esquerda, portanto as cores nos mapas que representam as distâncias e oclusões seguem o mesmo padrão da Seção 5.1.5.1. O método AUWS foi aplicado ao par estéreo adquirido com turbidez média e o mapa de disparidades obtido é apresentado na Figura 5.12, juntamente com o mapa de disparidade adquirido pelo método UWS para este mesmo par de imagens.

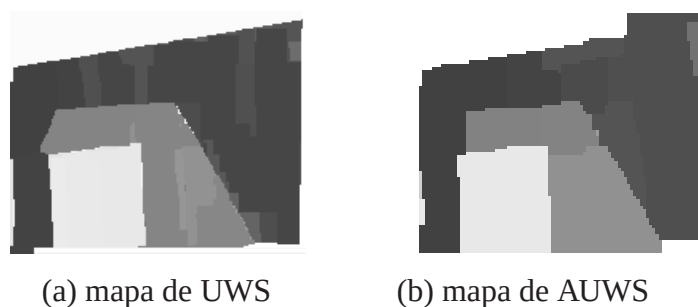


Figura 5.12: (a) e (b) os mapas de disparidade da cena do experimento com turbidez média.

Os resultados dos mapas de disparidade em ambos os métodos são similares, com pequenas diferenças que não chegam a serem significativas, o que demonstra que é o uso do método AUWS obtém resultados consistentes. Contudo, é importante lembrar que o método AUWS apresenta um custo computacional acima do método UWS, e que não possui uma comprovação de que o método converge para um mínimo que apresente um resultado consistente. Portanto, o resultado desse experimento permite concluir que o método UWS é mais robusto e preciso em termos de obter um bom resultado, enquanto que o método AUWS tem a facilidade de não precisar do processo anterior de estimativa experimental dos parâmetros do modelo de propagação da luz na água.

5.1.7 Restauração de Imagens Subaquáticas

Embora não seja um dos objetivos principais desta tese, a restauração de imagens subaquáticas é um resultado natural da aplicação do UWS, pois ao final do método tem-se o mapa de disparidades, com o qual facilmente se calculam as distâncias de cada ponto da cena ao sensor, e dispõe-se também dos parâmetros do modelo de propagação da água, obtidos experimentalmente conforme a Seção 3.2 de estimativa dos parâmetros α e β . Portanto, restaurar a imagem significa aplicar a Equação (3.9) (que restaura a cor do objeto) para cada pixel da imagem.

A principal diferença, em termos de restauração da imagem, entre este método e o proposto em [Schechner e Karpel, 2004] está no fato de que é possível restaurar imagens de cenas dinâmicas. Considerando isso e o fato de que o UWS terá aplicações futuras no projeto

Canal	parâmetros		
	α	β	$E(\infty)$
R	0.0025	0.00045	37.02
G	0.0020	0.00032	130.51
B	0.0024	0.00021	77.67

Tabela 5.4: Parâmetros α , β em cm^{-1} e $E(\infty)$ em pixels estimados experimentalmente.

Utilização Racional de Mecanismos de Transposição de Peixes (Seção 1.5.1), o experimento consiste em:

- Utilizar um aquário com 200l de água + 100g de argila para tornar a água turva;
- Estimar os parâmetros α e β experimentalmente;
- Posicionar as duas câmeras fronto-paralelamente em relação ao aquário, movendo a câmera 2 para uma posição aproximadamente paralela à posição da câmera 1 dada na Figura 5.1;
- Calibrar geométrica e radiometricamente o par estéreo;
- Colocar no aquário 5 peixes bagre de aproximadamente 10cm de comprimento cada;
- Adquirir duas seqüências de filmes, uma em cada câmera, idealmente ao mesmo instante ou garantindo que exista uma intersecção entre os instantes dos filmes adquiridos em cada câmera.

O Algoritmo UWS é aplicado a cada par estéreo por vez, portanto é necessário fazer um alinhamento temporal na duas seqüências de filme para estabelecer a correspondência entre os quadros (*frames*) e definir os pares estéreo. Neste experimento o alinhamento temporal foi executado a partir do método desenvolvido em [Carceroni et al., 2004].

A Tabela 5.4 apresenta os resultados da estimativa experimental dos parâmetros do modelo de propagação da luz na água, e A Figura 5.13 apresenta um par estéreo temporalmente alinhado e retificado extraído das duas seqüências de filme.

Os resultados dos métodos EMGC e UWS aplicados a este par estéreo são dados na Figura 5.14. Observe que, devido a alta turbidez e considerável espessura d'água (os peixes não se aproximavam das câmeras) o resultado do método EMGC não conseguiu fazer as correspondências e não obteve um mapa de disparidades consistente, enquanto que o método UWS obteve o mapa, ainda que com alguma imperfeições, mas é possível perceber nitidamente a forma dos dois peixes que se encontram visíveis na imagem.

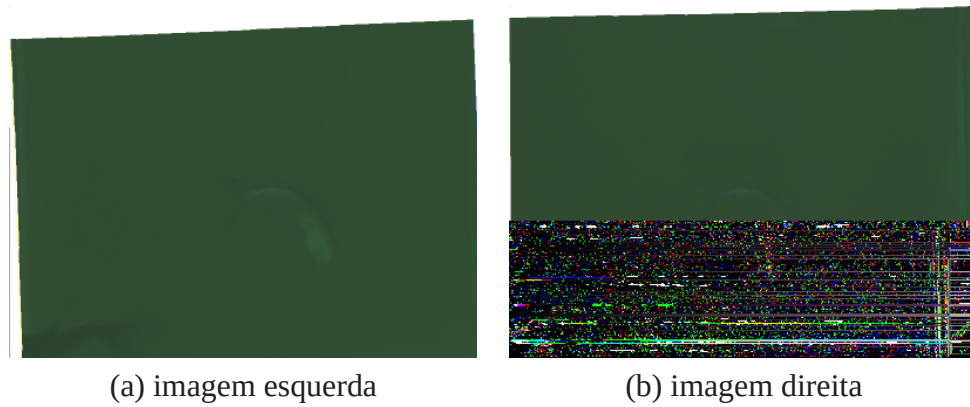


Figura 5.13: Par estéreo temporalmente alinhado. Em (a) o quadro 149 do filme da câmera esquerda e em (b) o quadro 072 do filme da câmera direita.

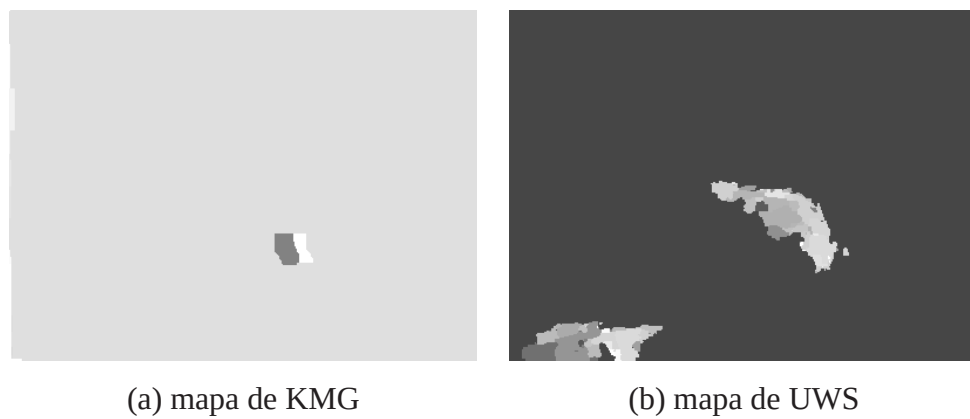


Figura 5.14: (a) e (b) os mapas de disparidade do par estéreo dado na Figura 5.13.

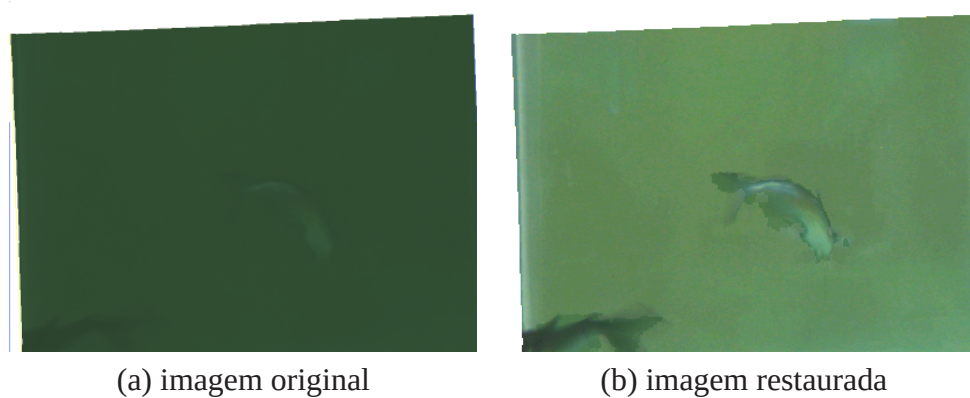


Figura 5.15: Em (a) a imagem da cena original, e em (b) a imagem restaurada.

Utilizando o mapa de disparidades do método UWS (Figura 5.14) para calcular as distâncias em espessura d'água de cada ponto da cena à câmera e conhecidos os parâmetros do modelo

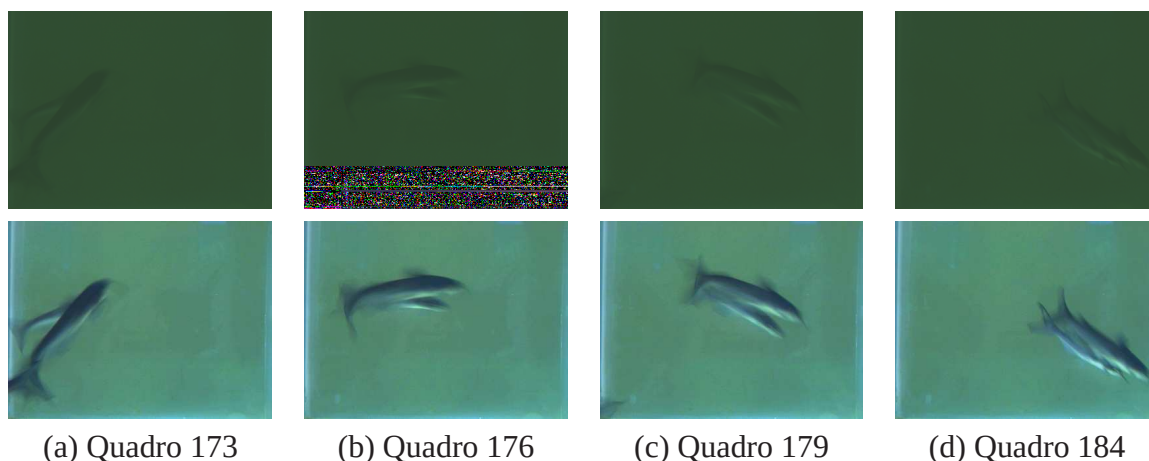


Figura 5.16: Exemplos de restauração de imagens subaquáticas para o filme da câmera 1, sendo acima o quadro original e abaixo o quadro restaurado.

de propagação da luz na água, dados na Tabela 5.4, a Figura 5.15 apresenta o resultado da restauração da imagem esquerda do par estéreo, repetida para fins de facilitar a comparação visual.

Como o mapa de disparidades possui imperfeições nos contornos dos peixes, principalmente devido ao efeito de borramento causado pelo movimento dos peixes (*motion blur*) [Gonzalez e Woods, 2000], uma abordagem alternativa é calcular a disparidade média do peixe e utilizá-la para obter uma distância média e aplicá-la na restauração da imagem, ainda que isso possa comprometer a restauração do fundo da cena. A Figura 5.16 apresenta exemplos desse caso.

5.2 Reflectância Inversa Eficiente

Nesta Seção é implementado o método descrito no Capítulo 4. Os resultados experimentais apresentados objetivam demonstrar a aplicabilidade e eficácia do método dado no Capítulo 4. Para tanto, o método foi aplicado a cenas previamente estabelecidas, com um número de objetos e tipos de superfícies que permitem verificar sua robustez através de dois experimentos: uma cena sintética, isenta de ruídos inerentes ao processo de aquisição; e uma cena real com todas as questões envolvidas com ruído, precisão e particularidades dos objetos da cena.

A fim demonstrar a importância o uso das Árvores de Computação de Intensidade do Pixel e da fatoração fotométrica para recuperar os parâmetros de reflectância em uma cena globalmente iluminada, são executadas experiências comparativas entre o método descrito na Seção 4.3 – que será denominado *Factorized Tree-Based Optimization* (F-TBO) – e duas alternativas mais simples que foram aplicadas à mesma imagem de entrada. Uma destas alternativas – que será denominada *Non-Factorized Tree-Based Optimization* (NF-TBO) – utiliza as Árvores de Computação de Intensidade do Pixel para evitar a necessidade de executar múltiplas renderizações da cena, mas busca otimizar os parâmetros fotométricos lineares e não-lineares conjuntamente, usando o método do Gradiente Conjugado. Uma alternativa ainda mais simples – que será denominada *Rendering-Based Optimization* (RBO) – não gera as Árvores de Computação da Intensidade do Pixel, e simplesmente computa os erros renderizando a cena e comparando a imagem sintética resultante à imagem de entrada.¹ Neste sentido, RBO é conceitualmente similar à parte iterativa do método apresentado em [Boivin e Gagalowicz, 2002]. Todos os três métodos são instanciados com a utilização do modelo de reflectância de Phong [Phong, 1975]. A renderização direta é executada com uma versão modificada do *Ray Tracer* de Sam Buss [Buss, 2003], disponível para uso em pesquisas.

5.2.1 Aplicando o Método a uma Cena Sintética

Para verificar a funcionalidade do método é gerada a imagem de uma cena sintética onde os objetos apresentam todas as possibilidades de parâmetros do modelo de iluminação, inclusive reflectividade e transmissividade (Figura 1.2). O experimento faz uma comparação de eficiência (*speedup* e tempo) e eficácia (erro médio de intensidade de pixel) entre os métodos

¹Neste caso, é também utilizado o método do Gradiente Conjugado em RBO, entretanto não é possível computar suas derivadas parciais apenas a partir da renderização, e portanto as derivadas parciais são computadas analiticamente em cada iteração a partir de uma árvore de dados, sendo entretanto este tempo de cálculo das derivadas desconsiderado na apresentação dos resultados. Isto permite manter a mesma base para fins de comparação do tempo de execução de cada método.

F-TBO, NF-TBO e RBO, sendo que nesses dois últimos os resultados em termos de erro e de imagem final são equivalentes, sendo diferenciados basicamente pelo tempo de execução de cada um, pois utilizam o mesmo método de otimização não-linear com o cálculo da função de erro sem e com múltiplas renderizações.

As superfícies dos objetos da cena são homogêneas e divididas em quatro tipos diferentes de materiais. A esfera à esquerda tem características de reflectância essencialmente lambertianas, não apresentando portanto os parâmetros secundários de reflectividade e transmissividade; a esfera ao centro apresenta características de reflectância secundária fortemente reflexivas; a esfera a direita apresenta características de reflectância secundária fortemente transmissivas, e todas as esferas geram sombras e inter-reflexões na cena em um piso formado por dois tipos de materiais com características lambertianas.

Os experimentos foram executados em uma máquina Pentium IV 2.26GHz / 512Mb RAM na plataforma windows XP, obtendo os resultados para os três métodos. Em ambos os casos o método de otimização utilizado é o do Gradiente Conjugado [Press et al., 1992].

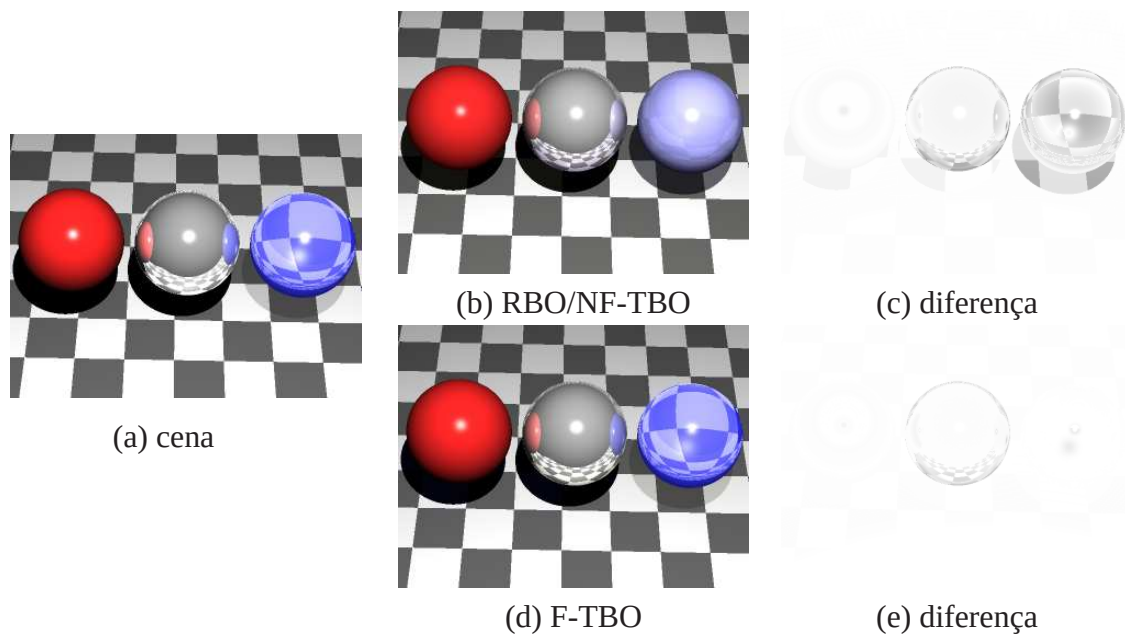


Figura 5.17: Em (a) a imagem da cena real. (b) apresenta o resultado do método NF-TBO, similar ao método RBO. (d) apresenta o resultado do método F-TBO. (c) e (e) apresentam a diferença entre cada método e a imagem da cena real.

Um primeiro resultado apresenta uma comparação entre a cena real e as imagens renderizadas com os parâmetros adquiridos no método F-TBO e RBO/NF-TBO (são equivalentes) cujas imagens e diferença entre estas e a imagem da cena real são apresentadas na

Figura 5.17. Os parâmetros dos objetos da cena e os resultados obtidos com a aplicação dos métodos, para cada banda RGB, são apresentados na Tabela 5.6. Os resultados consistem em valores aproximados dos parâmetros de BRTDF obtidos com os métodos RBO, NF-TBO, e F-TBO.

A imagem de diferença permite perceber que os métodos RBO/NF-TBO obtiveram as maiores falhas nos parâmetros secundários (reflectividade e transmissividade) das esferas ao centro e à esquerda, o que também pode ser observado na Tabela 5.6. Por outro lado, o método F-TBO obteve resultados mais consistentes, e os erros percebidos na esfera ao centro (vide imagem diferença) ocorreram em função do coeficiente de atenuação de Fresnel que influenciam nas intensidades mais próximas à borda do objeto, enquanto que o método busca obter um valor médio de reflectividade para todo o objeto.

É importante observar que os valores obtidos na Tabela 5.6 possuem casos exatos e casos com diferenças que podem chegar a 33%, o que no entanto acabam tendo pouca influência na composição da intensidade para os pixels daquele objeto em questão, e portanto não podem ser analisados isoladamente.

O Gráfico da 5.18 apresenta um comparativo entre os erros obtidos para diferentes entradas dos parâmetros de reflectância/transmissividade. O resultado foi obtido aplicando-se os métodos à diferentes entradas dos parâmetros lineares da BRTDF. Como o F-TBO independe de entrada para os parâmetros lineares, obteve valores constantes de erro para diferentes entradas, em contrapartida à variação do erro obtida pelos métodos RBO e NF-TBO. Para o caso de variação dos parâmetros não lineares, todos os métodos apresentam variação similar.

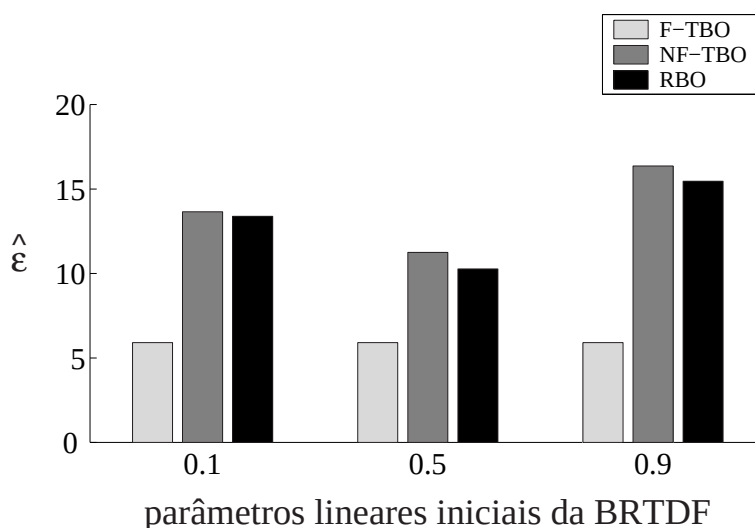


Figura 5.18: Erro fotométrico RMS em níveis de cinza para três diferentes entradas dos parâmetros de reflectância / transmitância.

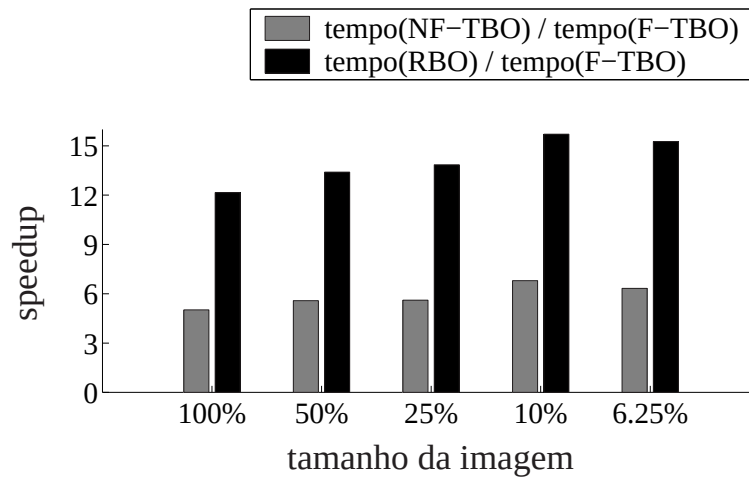


Figura 5.19: *Speedup* entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.

O ganho de eficiência é medido pelos *speedups* entre os métodos tendo o F-TBO como referência, e são obtidos da seguinte maneira: os métodos NF-TBO e RBO são executados até a convergência de um erro mínimo ou atingir um número máximo de iterações. O método F-TBO é executado um número de ciclos determinado, suficiente para obter um erro equivalente ao erro obtido nos métodos NF-TBO e RBO. Isto significa que o *speedup* entre os métodos é medido para a execução do algoritmo que atinja um mesmo patamar de erro. Os métodos são executados para cinco diferentes resoluções da imagem (100%, 50%, 25%, 10% e 6.25%) de maneira a avaliar a eficiência em relação ao tamanho da entrada de dados.

Métodos	<i>Speedup</i>
NF-TBO / F-TBO	5.02 a 6.80
RBO / F-TBO	12.15 a 15.71

Tabela 5.5: *Speedup* entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.

Os resultados obtidos são apresentados no gráfico da Figura 5.19 e na Tabela 5.5, bem como o tempo gasto em cada caso é apresentado na Figura 5.20. Em todos os casos de resolução o método F-TBO foi mais eficiente que os demais, confirmando o ganho que se supunha obter com a fatoração e resolução de maneira mais eficiente do sistema de equações gerado pelo método. Observe que o RBO leva cerca de 60 min contra aproximadamente 5 min do F-TBO na imagem de tamanho original, o que proporciona uma substancial economia de tempo de execução.

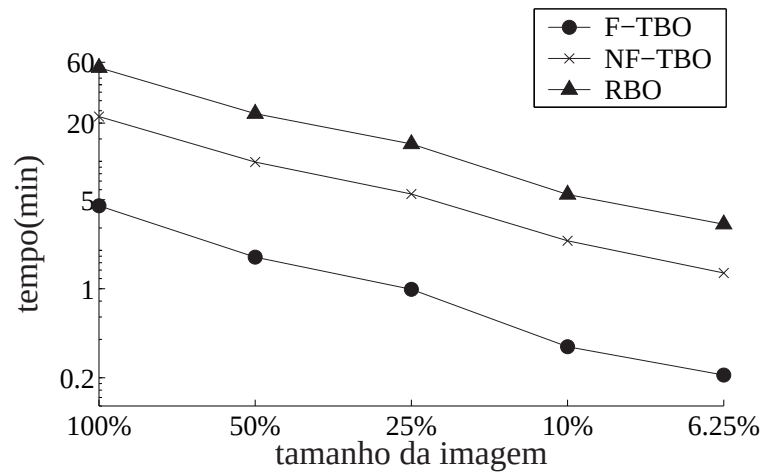


Figura 5.20: Tempo de processamento de CPU em minutos para os métodos RBO, NF-TBO e F-TBO, sob as mesmas condições de entrada em diferentes tamanhos de imagem.

Os resultados apresentados nessa seção demonstram que o método é funcional, tanto em termos de qualidade do resultado, quanto em eficiência. Para demonstrar a robustez do método e suas limitações quando aplicado a uma imagem com ruído, imperfeições na geometria e intensidades de fonte de luz imprecisas, a seção seguinte apresenta o mesmo experimento aplicado a uma cena real.

5.2.2 Aplicando o Método a uma Cena Real

Nesta seção são descritos, inicialmente, alguns detalhes experimentais necessários à aplicação do método a uma cena real, como forma de obter os dados necessários à sua execução: o modelo 3D da cena, e as informações sobre as fontes de luz e a imagem da cena. Em seguida, o método foi aplicado a uma cena previamente estabelecida, com um número de objetos e tipos de superfícies que permitem verificar sua robustez através de uma cena real com todas as questões envolvidas como ruído, imprecisão dos dados de entrada e particularidades dos objetos da cena. Assim como na seção anterior, os experimentos desta seção utilizam os métodos F-TBO, RBO e NF-TBO para fins de comparação.

5.2.2.1 Ambiente Experimental

A montagem experimental da cena real (Figura 1.6), foi criada em laboratório especificamente para este fim, utilizou duas fontes de luz aproximadamente pontuais, e nove diferentes objetos: três esferas plásticas (vermelha, amarela e verde), com um aparentemente

Parâmetros dos Objetos		Cena			F-TBO			RBO/NF-TBO		
		R	G	B	R	G	B	R	G	B
esfera vermelha	k_{d1}	0.70	0.10	0.10	0.59	0.10	0.11	0.52	0.10	0.11
	k_{s1}	0.50	0.50	0.50	0.41	0.43	0.41	0.36	0.41	0.41
	r_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	t_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	n_{s1}	60.0			57.0			64.0		
esfera cinza	k_{d2}	0.40	0.40	0.40	0.41	0.40	0.41	0.42	0.43	0.42
	k_{s2}	0.50	0.50	0.50	0.44	0.43	0.46	0.39	0.36	0.40
	r_2	1.00	1.00	1.00	0.69	0.70	0.67	0.55	0.52	0.59
	t_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	n_{s2}	60.0			57.2			64.0		
esfera azul	k_{d3}	0.10	0.10	0.50	0.09	0.10	0.51	0.42	0.42	0.62
	k_{s3}	0.50	0.50	0.50	0.35	0.35	0.33	0.37	0.37	0.34
	r_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	t_3	0.80	0.80	0.80	0.81	0.81	0.78	0.35	0.35	0.39
	n_{s3}	60.0			57.5			60.0		
plano	k_{d4}	0.80	0.80	0.80	0.77	0.77	0.73	0.73	0.73	0.74
	k_{s4}	0.20	0.20	0.20	0.18	0.17	0.18	0.51	0.51	0.42
	r_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	t_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	n_{s4}	12.0			10.4			64.0		
	k_{d5}	0.30	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.30
	k_{s5}	0.20	0.20	0.20	0.19	0.19	0.16	0.10	0.10	0.10
	r_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	t_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	n_{s5}	12.0			16.7			64.0		

Tabela 5.6: Valores dos parâmetros de reflectância do modelo para a cena, os métodos F-TBO e RBO/NF-TBO (similares), sendo $\mathbf{x} = \langle k_d, k_s \rangle$, os parâmetros primários de reflectância difusa e especular, respectivamente, $\langle r, t \rangle$ os parâmetros secundários de reflectividade e transmissividade, e $\theta = \langle n_s \rangle$ o parâmetro não linear da função $g(\theta)$.

alto coeficiente de *shininess*; um plástico azul translúcido; um espelho cilíndrico; um paralelepípedo azul de plástico que é refletido e sombreado pelo espelho; uma esfera de vidro próxima à esfera verde de plástico para mostrar o efeito de uma sombra sobre outra sombra; um paralelepípedo branco de borracha cujo canto inferior esquerdo representa o ponto de referência da cena no mundo; e um papel cinza em formato A4 como plano base da cena.

As duas fontes de luz se encontram posicionadas do mesmo lado da cena, evitando que o reflexo da luz incida diretamente no espelho. Essa é uma simplificação experimental em função da limitação do renderizador utilizado para modelar a cena, e não devido ao método.

Observe que existe uma concentração da luz refratada pela esfera de vidro, e que no entanto foi artificialmente retirada para estes experimentos, uma vez que o renderizador uti-

lizado (Ray Tracer dado em [Buss, 2003]) não permite modelar este efeito. Para recompor esta parte da cena, foi utilizada a média dos pixels vizinhos que não continham o efeito de concentração da luz. Os índices de refração das esferas foram considerados como os valores padrão para o plástico (1.46) e vidro (1.52).

5.2.2.2 Calibração das Câmeras e das Fontes de Luz

Para renderizar a cena sintética a partir da cena real faz-se necessário conhecer a geometria de cada objeto da cena, suas posições, orientações e a relação entre a referência da câmera e da cena, bem como as posições e intensidades das fontes de luz. Para obter estes dados, dois métodos de calibração foram efetuados: um para a calibração de câmera; e outro para a calibração das fontes de luz.

Para estimar os parâmetros extrínsecos e intrínsecos da câmera foi utilizado um alvo de calibração planar contendo círculos cujos centros estão distantes entre si de $15mm$ (Figura 5.21). O método de calibração utilizado foi o método de Tsai [Tsai, 1986] [Tsai, 1987] [Willson, 2003], que é um método conhecido e bem resolvido e que permitiu obter erros geométricos da re-projeção dos pontos inferiores a 0.6 pixel.

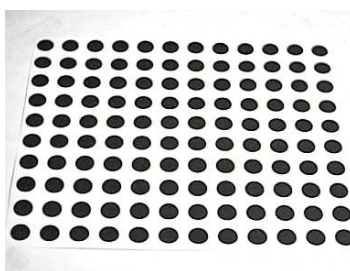


Figura 5.21: Alvo de calibração geométrica.

A iluminação da cena foi fornecida por dois bulbos de halogênio 100-W, cujas posições 3D foram estimadas a partir de um processo de calibração das fontes de luz, que visa estimar a posição de cada fonte, enquanto suas intensidades são estimadas pelos dados do fabricante [Bouquet e Perona, 1998]. Basicamente, o método consiste em adquirir diversas imagens da cena com um dispositivo mecânico (um pino) colocado de forma a ficar alinhado com o vetor normal do plano. Para cada imagem adquirida, o dispositivo é colocado em diferentes e aleatórias posições. Em cada imagem existem dois pontos importantes: o ponto da base do dispositivo e o ponto da sombra projetada no plano que representa o topo do dispositivo. Ambos os pontos devem estar visíveis à imagem. Feito isso, são recuperados vários vetores (descritos na referência do mundo) partindo do ponto da sombra, passando pelo topo do dispositivo, e em direção à fonte de luz, utilizando os dados obtidos no processo de calibração

da câmera. Um sistema sobredeterminado é, então, formulado e resolvido por um método linear, obtendo a posição da fonte de luz.

Ambos os processos de calibração (de câmera e de fonte de luz) foram executados com o uso do aplicativo desenvolvido por Fábio Marinho e Rodrigo Carceroni, de propriedade do Laboratório de Visão Computacional e Robótica do DCC/UFMG.

5.2.2.3 Modelagem 3D da Cena

A modelagem 3D da cena, que basicamente são as posições de pontos dos objetos sob coordenadas da cena que permitam a sua renderização, pode ser obtida através de diversos processos, inclusive manual. A forma mais precisa seria o uso de um scanner 3D, entretanto, considerando a indisponibilidade de acesso a essa tecnologia e a exemplo do trabalho de [Boivin e Gagalowicz, 2001], utilizamos um pacote de modelagem 3D baseado em imagens que utiliza o método *FACADE* desenvolvido em [Debevec et al., 1996].

O aplicativo tem como entrada dez fotografias sob diferentes pontos de vista da cena e, após indicações de correspondência de pontos feitas manualmente em cada imagem, obtêm-se os pontos que modelam a geometria de cada objeto da cena. Observe que estas imagens não são utilizadas no método de reflectância inversa, e tão somente para permitir a obtenção do modelo 3D.

5.2.2.4 Resultados para Cena Real

Os métodos F-TBO, NF-TBO e RBO são aplicados ao mesmo conjunto de dados de entrada, visando determinar a robustez e o ganho de eficiência do método F-TBO quando aplicado a uma cena real, a partir da comparação com os resultados obtidos nos outros dois métodos.

A imagem de entrada da cena foi obtida por uma câmera Hitachi DP-50 com resolução 640x480 pixels. Os experimentos foram executados em uma máquina Pentium IV 2.26GHz / 512Mb RAM na plataforma windows XP, obtendo os resultados para os três métodos. O método de otimização do Gradiente Conjugado [Press et al., 1992] é utilizado em todos os métodos, que são executados com 4 diferentes resoluções da imagem de entrada. O critério de parada dos algoritmos é dado pela diferença percentual do erro anterior em relação ao erro atual, com base em um limiar pré-definido.

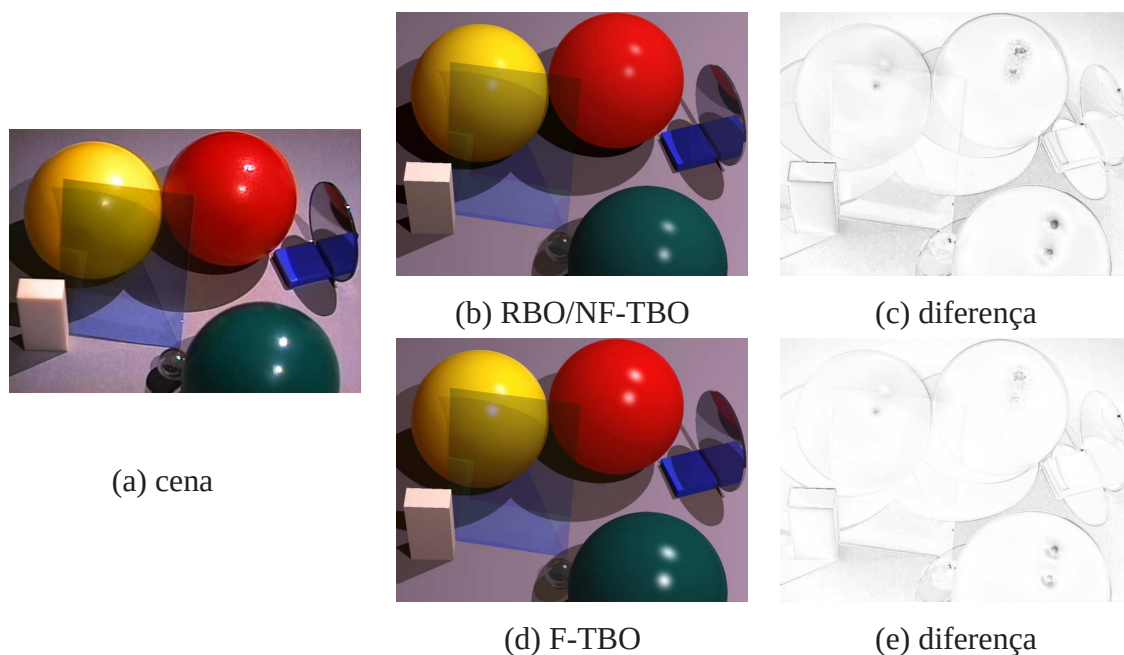


Figura 5.22: Em (a) a imagem da cena real. (b) apresenta o resultado do método NF-TBO, similar ao método RBO. (d) apresenta o resultado do método F-TBO. (c) e (e) apresentam a diferença entre cada método e a imagem da cena real.

O primeiro resultado apresenta uma comparação entre a cena real e as imagens renderizadas com os parâmetros adquiridos no método F-TBO, RBO e NF-TBO cujas imagens e diferença entre estas e a imagem da cena real são apresentadas na Figura 5.22, sendo que neste caso os resultados de RBO e F-TBO são equivalentes. As imagens diferença demonstram que os ruídos e imperfeições geométricas da cena geram um erro bem maior que no caso da imagem sintética, entretanto é possível observar que o método F-TBO obteve um resultado significativamente melhor em relação ao RBO/NF-TBO, em particular nas reflexões concentradas da luz devido ao parâmetro de *shininess* e dos objetos com transparência e alta reflexão (espelho).

Alguns fatores que aumentam o erro nas cenas reais devem ser considerados. Primeiramente, no processo do *Ray Tracing*, bordas difusas não são bem definidas porque o renderizador disponibilizado em [Buss, 2003] não possui técnicas anti-aliasing. Isso pode ser observado na imagem da Figura 5.22 que mostra a diferença absoluta entre a imagem da cena e a imagem renderizada. Existem erros de precisão na medida dos objetos reais que não são geometricamente ideais e, por causa disto, causam erros acentuados nas bordas dos objetos, em particular nas esferas. Além disso, as fontes de luz reais são apenas aproxima-

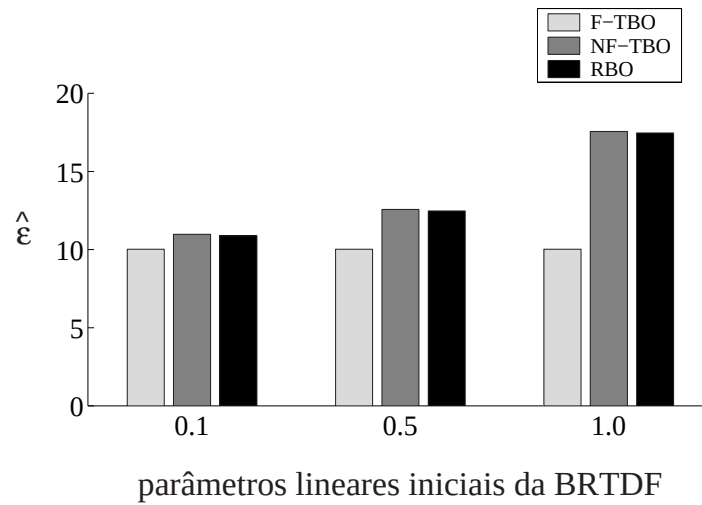


Figura 5.23: Erro fotométrico RMS em níveis de cinza para três diferentes entradas dos parâmetros de reflectância / transmitância.

damente pontuais, o que também provoca pequenas diferenças entre a cena real e a imagem renderizada. Uma simulação do cálculo do erro sem considerar os erros de precisão das medidas geométricas leva o erro a um valor em torno de 6 pixels, o que é bastante razoável no contexto desta aplicação.

O Gráfico da Figura 5.23 apresenta um comparativo entre os erros obtidos para diferentes entradas dos parâmetros lineares da BRTDF para os três métodos. Como o F-TBO independe destes valores de entrada, obteve valores constantes de erro para diferentes entradas, em contrapartida à variação do erro obtida pelos métodos RBO e NF-TBO. Para o caso de variação dos parâmetros não lineares, todos os métodos apresentam variação similar.

O ganho de eficiência é medido pelos *speedups* entre os métodos tendo o F-TBO como referência, e seguem o mesmo procedimento adotado no caso do experimento com imagem sintética, ou seja, estabelece-se uma quantidade de iterações do método F-TBO (6 iterações) que obtém um erro equivalente ao erro obtido nos métodos RBO e NF-TBO para um limiar de variação percentual do erro de 10^{-5} . Os métodos são executados para cinco diferentes resoluções da imagem (100%, 50%, 25%, 10% e 6.25%) de maneira a avaliar a eficiência em relação ao tamanho da entrada de dados.

Métodos	<i>Speedup</i>
NF-TBO / F-TBO	3.11 a 4.27
RBO / F-TBO	7.35 a 8.82

Tabela 5.7: *Speedup* entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.

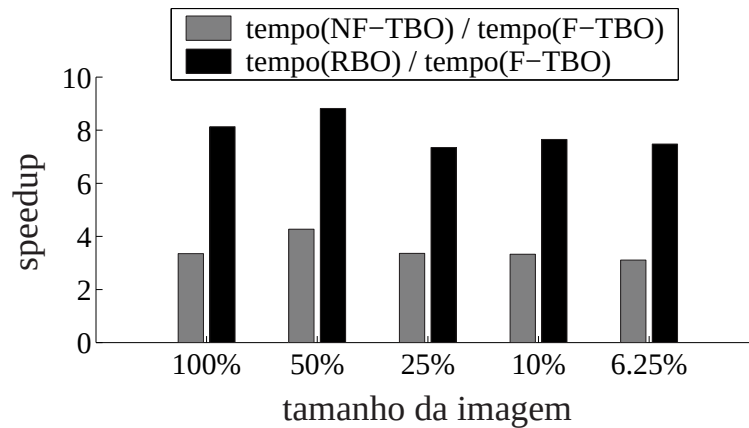


Figura 5.24: *Speedup* entre os métodos RBO e F-TBO e entre os métodos NF-TBO e F-TBO, considerando diferentes tamanhos da imagem original.

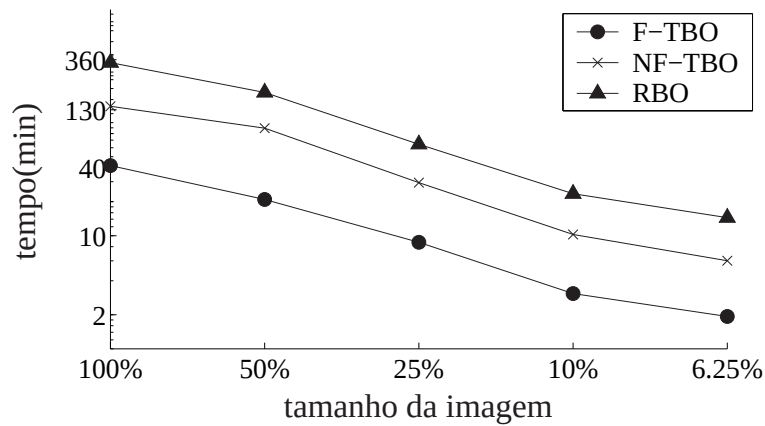
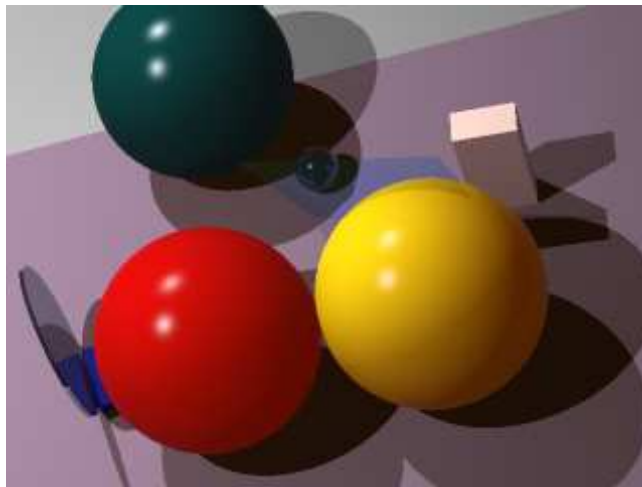


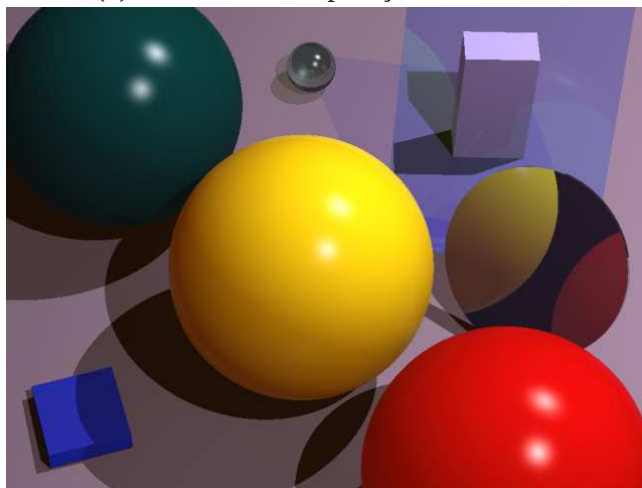
Figura 5.25: Tempo de processamento de CPU em minutos para os métodos RBO, NF-TBO e F-TBO, sob as mesmas condições de entrada em diferentes tamanhos de imagem.

Os resultados obtidos são apresentados no gráfico da Figura 5.24 e na Tabela 5.7, bem como o tempo gasto em cada caso é apresentado na Figura 5.25. Em todos os casos de resolução o método F-TBO foi mais eficiente que os demais, mantendo um ganho de eficiência significativo mesmo para o caso em que se utiliza uma imagem de uma cena real, com todos os problemas de ruído que a mesma contém. Observe que o método RBO leva cerca de 5.7h contra aproximadamente 42 min do método F-TBO na imagem de tamanho original, o que proporciona uma substancial economia de tempo de execução.

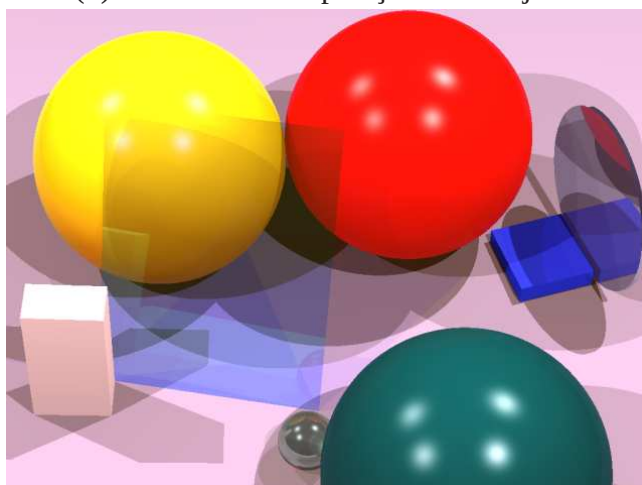
Uma vez conhecidos os parâmetros da BRTDF de cada objeto, é possível manipular a cena sintética de maneira a gerar diferentes condições de iluminação e pontos de vista, ou mesmo mudança de posições dos objetos (Figura 5.26), demonstrando o foto-realismo obtido em novas imagens sintéticas geradas com o resultado do método proposto nesta tese.



(a) modificando a posição da câmera



(b) modificando as posições dos objetos



(c) modificando a iluminação da cena

Figura 5.26: Em (a), (b) apresentam novas vistas da cena sintética obtida pelo método F-TBO. Em (c) a mesma cena sintética sob novas condições de iluminação.

Capítulo 6

Conclusão e Trabalhos Futuros

Esta tese apresentou a importância e dificuldades dos métodos de modelagem baseada em imagens aplicados à cenas que não podem ser explicadas pelo uso de um modelo de iluminação local. As dificuldades relacionadas ao problema geral foram tratadas em dois casos particulares apresentando soluções para dois cenários específicos dentro deste escopo. No primeiro caso é obtido o modelo geométrico de uma cena submersa onde a luz interage globalmente com o meio, e no segundo caso é obtido o modelo de reflectância de uma cena onde a luz interage globalmente com os objetos da cena.

O primeiro caso apresentou um método de estéreo subaquático que utiliza um modelo físico de propagação da luz na água para corrigir as intensidades dos valores dos pixels nas imagens de uma cena submersa, e um método de estéreo denso para obter as estimativas de profundidade das posições dos pontos de uma imagem de referência que estão correlacionados na outra imagem do par estéreo, ambos integrados para determinar um mapa de disparidades da cena (Capítulo 3).

O método foi utilizado para obter o mapa de disparidades de três cenas subaquáticas reais, e comparado com um algoritmo de estéreo denso de reconhecida eficiência. Em todos os casos o método proposto obteve resultados melhores e mais confiáveis em relação a um método de estéreo convencional de reconhecida qualidade, como pôde ser visto a partir de comparações dos mapas de disparidades (Seção 5.1.5).

Os resultados apontaram que uma dificuldade do estéreo subaquático é a maneira experimental de se obter os parâmetros do modelo de propagação da luz no meio, o que pode ser inviável para alguns ambientes mais agressivos. Para tanto, foi desenvolvida uma extensão do método de modo a permitir sua execução com a estimativa destes parâmetros a partir de uma otimização iterativa. O resultado apresentado demonstra que essa extensão do método possui um resultado consistente (Seção 5.1.6), entretanto a mesma é dependente dos valores

iniciais de entrada dos parâmetros, e portanto não pode ser ainda considerada uma solução definitiva.

Outra questão interessante é a respeito da distribuição da luz na água, considerada uniforme em toda a cena. Essa consideração pode ser feita no caso de águas profundas ou ambiente iluminado artificialmente, entretanto quando a cena se encontra próxima à superfície da água a luz torna-se mais acentuada na região da superfície e é gradualmente atenuada à medida que penetra na água. O método proposto pode ser estendido para esses casos, simplesmente calculando a distância em espessura d'água entre a superfície e o ponto da cena em análise e utilizando o parâmetro de atenuação da luz na água para corrigir a não uniformidade da iluminação, contudo faz-se necessário que o espelho d'água seja visível no par estéreo de imagens.

Embora não fosse um dos objetivos principais desta tese, uma consequência do método de estéreo subaquático é sua utilização para restaurar cenas subaquáticas com resultados bastante significativos (Seção 5.1.7), inclusive em cenas dinâmicas, principal limitação dos métodos atuais de restauração de imagens subaquáticas (Seção 2.4.3). Por outro lado, o custo computacional do método não permite que a restauração de uma cena dinâmica, ou mesmo a obtenção da geometria, seja feita em tempo real. Uma forma de melhorar a eficiência do método é utilizar algoritmos de estéreo denso mais eficientes, ainda que com resultados menos precisos que o algoritmo EMGC utilizado no método do estéreo subaquático.

O segundo caso apresentou um método para recuperar os parâmetros de reflectância de uma cena real com superfícies homogêneas, inter-reflexões entre objetos, sombras, objetos transparentes e reflexivos. O método utiliza uma única renderização do (*Ray Tracing*) que extrai as informações relativas à geometria e iluminação da cena, que são constantes uma vez que a cena é estática. Essas informações são guardados em uma estrutura de dados apropriada e posteriormente utilizadas, juntamente com os valores reais de cada pixel da imagem da cena, para gerar um sistema de equações não linear que é fatorado para permitir obter a solução de forma mais eficiente que se fosse utilizado o sistema não linear original (Capítulo 4).

Os resultados experimentais da aplicação do método de reflectância inversa eficiente apresentam comparações favoráveis em relação ao erro em intensidade de pixels, ganho em velocidade e imagem final obtida, quando considerados outros métodos com otimização sem fatoração do sistema e com otimização a partir de renderizações iterativas (Seção 5.2). Em alguns casos, o método proposto nesta tese chega a ser mais de 15 vezes mais rápido para uma cena sintética, e quase 9 vezes mais rápido para uma cena real, para uma comparação com o método de renderização iterativa.

O método utiliza como entradas a geometria da cena, as informações de posição e intensidade das fontes de luz e uma única imagem da cena. Além disso, para o caso de superfícies translúcidas cujo material é conhecido, os valores dos índices de refração são obtidos de tabelas com valores padrão de índices de refração. Contudo, é possível estimar os índices de refração para os materiais não conhecidos *a priori* através de um método iterativo de otimização, aplicado antes da análise fotométrica, que minimize as distâncias entre as projeções na imagem real das características da cena que são visíveis através dos objetos transparentes e das projeções dessas características na imagem gerada pelo *Ray Tracing*.

É importante frisar que o conhecimento dos índices de refração é essencial para aquelas superfícies que possuem transmissividade tal que influencia na direção refratada dos raios do *Ray Tracing*. Para quantificar o efeito dos índices de refração na aparência da cena em materiais que não possuem essa característica, foi feito um experimento adicional com uma cena sintética onde o método foi executado duas vezes, na primeira com os índices de refração conhecidos e na segunda com os índices de refração ajustados para 1. Os erros fotométricos RMS obtidos foram, respectivamente, 11.49 e 11.66 os níveis de cinza, o que sugere que o não conhecimento dos índices de refração de superfícies que não influenciam na direção refratada dos raios do *Ray Tracing* resultam apenas em uma perda de exatidão praticamente imperceptível.

A principal limitação do método de reflectância inversa está no fato do mesmo ser aplicado, até o momento, somente a cenas com objetos de superfícies homogêneas. Entretanto, em [Boivin e Gagalowicz, 2002] é mostrado que uma vez estimados todos os parâmetros fotométricos não-lambertianos, a textura lambertiana pode facilmente ser recuperada a partir de uma única imagem. Para verificar esta possibilidade, foram executados experimentos adicionais em uma versão sintética da cena mostrada na Figura 1.6 em que texturas foram adicionadas à esfera vermelha e ao plano de fundo. Ao ser aplicada a metodologia de reflectância inversa eficiente a esta nova cena, obteve-se como resultado os parâmetros fotométricos que explicam as cores médias destas superfícies texturizadas com uma precisão de aproximadamente 3 níveis de cinza. Isto sugere que é possível estender a metodologia apresentada nesta tese no caminho proposto por Boivin e Gagalowicz, o que provavelmente permitirá obter os parâmetros de reflectância também de cenas que possuam objetos com textura.

Ambos os casos tratados nesta tese, seja o de adquirir a geometria ou a reflectância da cena, apresentam resultados que trazem contribuições significativas para a solução do problema do geral, validando os métodos que foram propostos. Entretanto, é importante que se visualize a integração destes casos particulares em um método que possa, ao mesmo tempo, recuperar a geometria e a reflectância de uma cena globalmente iluminada. Embora

a princípio aparente ser um caminho natural, esta não é uma tarefa trivial, pois os algoritmos de estéreo denso não são aplicáveis a cenas com inter-reflexão entre os objetos, como no caso utilizado para adquirir a reflectância inversa, principalmente devido às ambigüidades contidas na cena que não permitem obter a correspondência entre os pontos de maneira correta. O caminho para obter um método que recupere ao mesmo tempo a geometria e a reflectância passa, necessariamente, pelo uso de múltiplas imagens da cena, como no caso de alguns trabalhos apresentados no Capítulo 2.

Assim sendo, o objetivo de avançar no estado da arte para modelagem de cenas com iluminação não local foi alcançado, mas não exaurido, pois melhorias significativas podem ainda ser adicionados aos métodos. Mais especificamente, alguns trabalhos futuros que se pretende desenvolver para estender as contribuições desta tese são:

- Pesquisar e utilizar métodos de otimização especialistas para os casos do estéreo subaquático, cuja função objetivo é exponencial, e da reflectância inversa eficiente, no caso da otimização polinomial;
- Utilizar outros métodos de estéreo convencional integrados ao estéreo subaquático, como maneira de estabelecer uma relação de compromisso entre a qualidade do resultado obtido e a eficiência computacional do método;
- Desenvolver uma extensão do método de reflectância inversa eficiente, tornando-o apto a ser utilizado em cenas com textura suave, estabelecendo suas limitações;
- Pesquisar a viabilidade de estender o método de reflectância inversa eficiente para objetos translúcidos que produzem a concentração da luz em determinado ponto (*caustics surfaces*).

Apêndice A

Métodos para Solução do Problema da Correspondência

O problema da correspondência apresenta muitas ambigüidades, e como consequência uma série de restrições são adotadas para tornar esse problema tratável, das quais podemos citar: a restrição da *invariância cromática*, as restrições de *unicidade*, *ordem* e *suavidade* e a *restrição epipolar*.

Fica evidente a forma como podemos fazer uso das restrições da invariância cromática e da geometria epipolar, casando cada ponto de uma imagem com todo pixel na outra imagem que apresentar a mesma cor na linha epipolar correspondente. Entretanto, restrições adicionais devem ser levadas em consideração como a unicidade, a suavidade e a restrição de ordem. Porém, como podemos aplicá-las a algoritmos de correspondência?

Sob a hipótese de que não existem objetos transparentes nas cenas observadas, a restrição da unicidade nos diz que cada ponto de uma imagem pode ter somente um ponto correspondente na outra imagem. Assim, a forma mais simples de se aplicar esta restrição, juntamente com a restrição da invariância cromática e a restrição epipolar, seria casar cada ponto de uma imagem com um único ponto da outra imagem da cor *mais similar* e pertencente à mesma linha epipolar. Essa ingênua implementação poderia ser bem aplicada a um caso ideal onde, além das superfícies da cena observada serem Lambertianas, cada ponto observado deveria ter uma cor única. Entretanto, mesmo nesse mundo ideal, ainda estaríamos sujeitos a “erros” ocasionados pela discretização dos valores de intensidades capturados pelos sensores das câmeras utilizadas.

Como exemplo da deficiência dessa metodologia, vamos considerar um *estereograma* preto e branco de pontos aleatórios, como na Figura A.1. Fica claro que, aplicando a correspondência como proposto anteriormente, teremos uma alta taxa de erro na obtenção de dados

corretos da disparidade. Isto porque, a observação de um único ponto isoladamente, não nos fornece informação suficiente para determinarmos de forma única o seu correspondente.



Figura A.1: Estereograma formado por pontos aleatórios.

A seguir iremos dar uma breve descrição de três métodos usados para resolver o problema da correspondência. Deve ficar claro que os métodos que serão descritos são somente uma pequena porção da grande variedade de métodos existentes para resolver o problema. Entretanto, selecionamos estes métodos devido a propriedades particulares de cada um, como a simplicidade, a eficiência computacional e a precisão.

A.1 Correlação

Métodos baseados na correlação se caracterizam por fazer uso explícito da restrição de suavidade partindo do princípio de que, na vizinhança de um ponto da imagem, todos os pixels podem ser associados a valores de disparidade similar. Essa idéia pode ser comparada a uma subdivisão da imagem em pequenas áreas contínuas que, se reprojetaadas na cena, resultariam em um conjunto de superfícies planas, ou seja, com uma única normal associada a cada superfície.

O princípio das técnicas de correlação é mostrado na Figura A.2. Neste caso, para encontrarmos um pixel p_2 da imagem I_2 que corresponde a um pixel p_1 da imagem I_1 consideramos uma *janela* retangular de tamanho $m \times n$ e, o critério de similaridade é a medida da correlação entre as janelas nas duas imagens. O ponto p_2 correspondente será dado pela janela que maximiza (ou minimiza, dependendo da medida adotada) o critério de similaridade entre as janelas nas duas imagens ao longo das linhas epipolares.

Muitas medidas para correlação têm sido propostas para aplicações diversas, sendo que a mais freqüentemente utilizada é a correlação dos valores das intensidades dos pixels.

Apesar da grande simplicidade e eficiência computacional muitos problemas ocorrem nesta abordagem [Faugeras, 1993]:

- assume-se que as linhas epipolares correspondem às linhas horizontais das imagens.

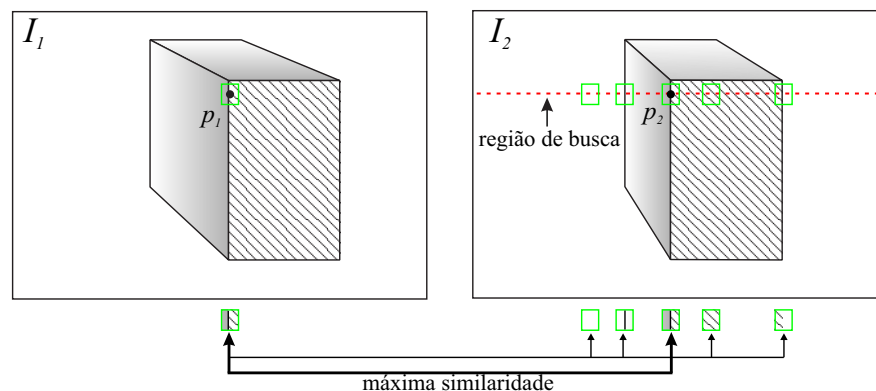


Figura A.2: Correlação por meio de uma janela retangular. Um pixel p_1 na imagem I_1 , corresponderá a um pixel p_2 , na região de busca da imagem I_2 , que estiver contido na janela de maior similaridade com relação à janela de p_1 .

Isto implica que as imagens utilizadas devem passar por um processo denominado *retificação*, que tem o objetivo de aplicar transformações nos planos da imagem de modo a reprojeta-los em um único plano comum;

- se a função de correlação não estiver bem definida, a disparidade não será bem determinada, pois a correspondência depende de uma função de correlação apropriada;
- assume-se que a disparidade é constante em toda a janela de correlação, o que implica em assumir que as disparidades são “localmente” constantes, ou equivalentemente, supor que toda a cena observada é composta de pequenos planos fronto-paralelos à cena observada.

Na prática, técnicas baseadas em correlação funcionam relativamente bem em regiões suaves e texturizadas, porém tendem a suavizar as bordas dos objetos observados, gerando “borrões” nas discontinuidades das superfícies.

A.2 Programação Dinâmica

A programação dinâmica pode ser aplicada a diversos problemas de otimização. Problemas dessa natureza geralmente apresentam muitas soluções possíveis, onde cada solução apresenta um valor de custo diferente. Dessa forma, estamos interessados em encontrar um valor ótimo para a solução referente a um valor de custo mínimo ou máximo (dependendo de como a função de custo for definida), dentre todas as soluções possíveis.

Em geral, o desenvolvimento de um algoritmo de programação dinâmica pode ser dividido em uma sequência de quatro etapas [Cormen et al., 1999]:

1. Caracterizar a estrutura de uma solução ótima;
2. Definir recursivamente o valor de uma solução ótima;
3. Calcular o valor de uma solução ótima em um processo *bottom-up*;
4. Construir uma solução ótima a partir de informações calculadas.

As etapas de 1 a 3 acima formam a base para a solução de um problema utilizando programação dinâmica. A etapa 4 pode ser omitida se apenas o valor de uma solução ótima é exigido, não a solução em si.

Podemos transformar o problema da correspondência em visão estéreo em um problema de minimização de uma função de custo. A Programação Dinâmica nos oferece uma forma de minimizar funções compostas por um grande número de variáveis discretas. Ela funciona como um método de divisão e conquista, ou seja, resolve problemas combinando as soluções dos diversos subproblemas envolvidos. Entretanto, ao contrário dos métodos de divisão e conquista, a Programação Dinâmica é aplicável quando os subproblemas são interdependentes, ou seja, compartilham soluções. Dessa forma, ao invés de tentarmos casar pares de pixels independentemente nas imagens, um conjunto de pixels em uma imagem pode ser associado, com base em um critério adotado, a um conjunto de pixels correspondentes na outra imagem.

Assumindo que as imagens usadas estão retificadas, o problema de obtermos a correspondência entre dois perfis de características pode ser formulado como um problema de encontrar um caminho ótimo em um grid bidimensional. Na Figura A.3 o eixo vertical representa o perfil de características de uma linha na imagem da esquerda enquanto que, o eixo horizontal exibe o perfil de uma linha na imagem da direita. Um ponto $P(i, j)$ no caminho exibido representa o custo da correspondência entre os pontos i e j , nas imagens da esquerda e da direita, respectivamente.

Técnicas de Programação Dinâmica são usadas para, eficientemente, encontrar o caminho de custo ótimo dentre os vários caminhos possíveis. Da observação da Figura A.3, percebe-se que, se a superfície da cena observada é contínua e toda parte da superfície pode ser vista em ambas as imagens, o caminho gerado é contínuo e monotonicamente crescente.

Vamos assumir que $D(i, j)$ representa a diferença de intensidade entre dois pixels i e j , influenciado por sua vizinhança (como nas técnicas de correlação). O custo de um caminho

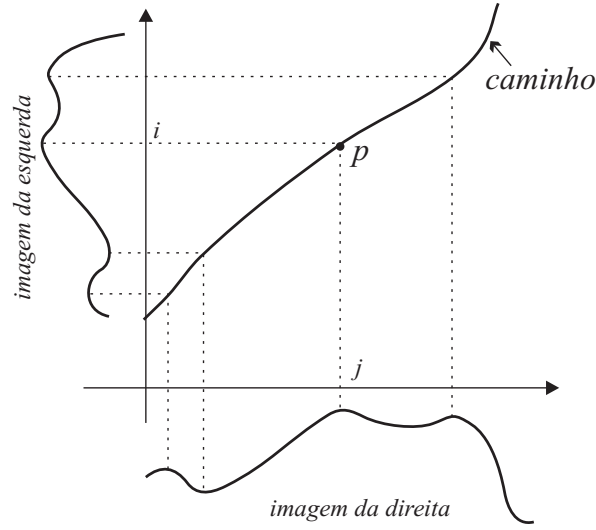


Figura A.3: Correspondência via Programação Dinâmica. Representação do caminho de custo ótimo na computação da correspondência entre linhas epipolares correspondentes nas imagens de entrada.

pode então ser definido como a integral dessas diferenças ao longo do caminho. Seja $F(i, j)$ o custo mínimo associado a um caminho ótimo do ponto inicial até o ponto (i, j) . Podemos definir um algoritmo de Programação Dinâmica para calcular o caminho mínimo com base no Algoritmo 8, onde n corresponde ao número de pixels em uma linha da imagem.

Algoritmo 8 Algoritmo para resolver o problema da correspondência a partir de técnicas de Programação Dinâmica.

{ PASSO (1) }

$F(i, 1) \leftarrow D(i, 1)$

$F(1, j) \leftarrow D(1, j)$

{ PASSO (2) }

para $2 < i, j < n$ **faça**

$F(i, j) \leftarrow D(i, j) + \min\{D(i-1, j-1), F(i-2, j-1) + D(i-1, j),$
 $F(i-1, j-2) + D(i, j-1)\}$

fim para

É importante notarmos que no Algoritmo 8, somente 3 arcos são possíveis de serem gerados em cada passo da iteração, veja Figura A.4. Essa é uma restrição intencional, pois poderíamos permitir a geração de arcos em todas as direções possíveis, entretanto, em qualquer caso, por restringirmos a direção, a área de busca também fica severamente restringida. Essas restrições são conseqüências das restrições de continuidade e ordem que o algoritmo impõe, que é equivalente a dizer que somente caminhos estritamente monotônicos

são admissíveis [Faugeras, 1993]. Como consequência, a correspondência estéreo usando Programação Dinâmica é mais adequada quando, por exemplo, as imagens são relativas a fotos aéreas de terrenos planos [Huguet et al., 2004].

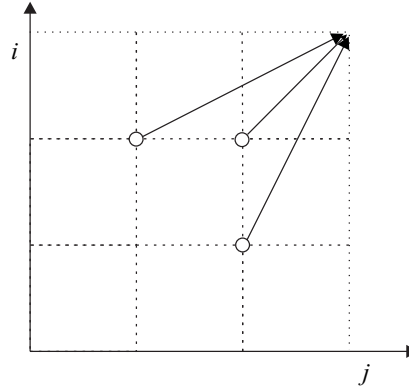


Figura A.4: Arcos possíveis que levam à correspondência (i, j) na correspondência via Programação Dinâmica.

Vale ressaltar que, se as cenas incluem descontinuidades nas superfícies, o passo (2) do algoritmo de programação dinâmica acima deve ser modificado para se adequar ao problema. Da mesma forma, se uma parte da superfície observada está visível somente na imagem da esquerda ou na da direita, o Algoritmo 8 também deve ser modificado de forma que no primeiro caso (visível na imagem da esquerda), o arco correspondente a esta parte da imagem seja vertical, e em caso contrário ele será horizontal. A dificuldade dessa abordagem é a definição de uma função de custo adequada, uma vez que em regiões parcialmente oclusas a diferença entre as regiões correspondentes não pode ser calculada. Assim, alguns trabalhos têm sido propostos procurando estabelecer a correspondência neste tipo de situação como em [Belhumeur e Mumford, 1992].

Outra observação que deve ser considerada é que até o momento, cada par de linhas correspondentes entre as duas imagens foi tratado independentemente das demais linhas. Entretanto, se uma aresta se prolonga através de várias linhas nas imagens, as correspondências em uma linha de varredura sofrem uma forte influência das correspondências em sua vizinhança. Forçar a consistência entre as linhas epipolares é equivalente a aplicar uma restrição de continuidade entre elas [Cox et al., 1996].

A.3 Fluxo Máximo e Cortes Mínimos em Grafos

Podemos interpretar um grafo orientado como uma *rede de fluxo* e, conseqüentemente, podemos usá-lo para resolver problemas relacionados a fluxos de materiais. Como exemplo, imagine um material percorrendo um sistema desde uma origem, onde o material é produzido, até um depósito, onde ele é consumido. A origem produz o material em alguma taxa fixa, e o depósito consome o material na mesma taxa. O *fluxo* do material em qualquer ponto do sistema é intuitivamente a taxa na qual o material se move. Uma rede de fluxo pode ser usada para modelar líquidos fluindo por tubos, peças por linhas de montagem, corrente por redes elétricas, informações por redes de comunicação [Cormem et al., 1999] e, também, o problema da correspondência em visão estéreo [Boykov et al., 1999], [Kolmogorov et al., 2003], [Kolmogorov e Zabih, 2002], [Roy e Cox, 1998].

Cada aresta orientada em uma rede de fluxo pode ser imaginada como um canal para o material. Cada canal tem uma capacidade estabelecida, dada como uma taxa máxima na qual o material pode fluir pelo canal. Vértices são junções de canais e, além da origem e do depósito, o material flui pelos vértices, sem acumulação. Em outras palavras, a taxa na qual o material entra no vértice deve ser igual à taxa em que ele deixa o vértice. Esta propriedade é conhecida por *conservação de fluxo*. No problema do fluxo máximo, desejamos calcular a maior taxa na qual o material pode ser enviado desde uma origem até um depósito sem violar quaisquer restrições de capacidade. A seguir daremos uma definição mais formal, baseada na teoria de grafos, para uma rede de fluxo [Cormem et al., 1999].

A.3.1 Redes de Fluxo e Fluxos

Uma *rede de fluxo* é um grafo orientado $G = (V, E)$ em que cada aresta $(u, v) \in E$ tem uma capacidade $c(u, v) \geq 0$. Se $(u, v) \notin E$, supomos que $c(u, v) = 0$. Distinguímos dois vértices em uma rede de fluxo: uma origem s e um depósito t . Por conveniência, supomos que cada vértice reside em algum caminho desde a origem até o depósito. Isto é, para todo vértice $v \in V$, existe um caminho $s \rightsquigarrow v \rightsquigarrow t$. Então, o grafo é fracamente conectado, e $|E| \geq |V| - 1$.

Seja $G = (V, E)$ uma rede de fluxo com uma função de capacidade c . Seja s a origem da rede e seja t o depósito. Um *fluxo* em G é uma função de valor real $f : V \times V \rightarrow \mathbf{R}$ que satisfaz às três propriedades seguintes:

Restrição de capacidade: Para todo $u, v \in V$, exigimos $f(u, v) \leq c(u, v)$.

Simetria oblíqua: Para todo $u, v \in V$, exigimos $f(u, v) = -f(v, u)$.

Conservação de fluxo: Para todo $u \in V - \{s, t\}$, exigimos $\sum_{v \in V} f(u, v) = 0$.

A quantidade $f(u, v)$, que pode ser positiva, 0 ou negativa, é chamada **fluxo do vértice** u até o vértice v . O valor de um fluxo f é definido como

$$|f| = \sum_{v \in V} f(s, v), \quad (\text{A.1})$$

ou seja, o fluxo total que sai da origem¹. No *problema de fluxo máximo*, temos uma rede de fluxo G com origem s e depósito t , e desejamos encontrar um fluxo de valor máximo.

A.3.2 O Método de Ford-Fulkerson

Em [Ford e Fulkerson, 1962] foi apresentado um método para resolver o problema do fluxo máximo baseado em um importante teorema conhecido por *teorema do fluxo máximo e corte mínimo*. Esse teorema caracteriza o valor de um fluxo máximo em termos de cortes na rede de fluxo. Para isto, o método de Ford-Fulkerson é dependente de três idéias essenciais para o teorema do fluxo máximo e corte mínimo e relevantes para muitos algoritmos de fluxo: redes residuais, caminhos em ampliação e cortes.

O método de Ford-Fulkerson é iterativo. Começamos com $f(u, v) = 0$ para todo $u, v \in V$, dando um fluxo inicial de valor 0. A cada iteração, aumentamos o valor do fluxo, encontrando um “caminho de ampliação”, que podemos imaginar como sendo um caminho desde a origem s até o depósito t onde ainda podemos incrementar o valor do fluxo. Esse processo é repetido até não ser possível encontrar nenhum caminho de ampliação. O teorema do fluxo máximo e corte mínimo assegura que, no final, esse processo produz um fluxo máximo.

Algoritmo 9 Método de Ford-Fulkerson para o problema do fluxo máximo.

```

MÉTODO FORD-FULKERSON( $G, s, t$ )
  inicializar fluxo  $f$  como 0
  enquanto existir um caminho de ampliação  $p$  faça
    ampliar fluxo  $f$  ao longo de  $p$ 
  fim enquanto
  retorne  $f$ 

```

A seguir, iremos dar uma descrição de três conceitos de extrema importância para o método descrito e, finalmente, uma descrição mais detalhada do Algoritmo 9 baseado em [Cormen et al., 1999].

¹A notação $|\cdot|$ é usada para indicar valor de fluxo e não valor absoluto ou cardinalidade

Redes residuais

Intuitivamente, dados uma rede de fluxo e um fluxo, a rede residual consiste em arestas que podem admitir mais fluxo. De modo mais formal, suponha que temos uma rede de fluxo $G = (V, E)$ com origem s e depósito t . Seja f um fluxo em G , e considere um par de vértices $u, v \in V$. A quantidade de fluxo *adicional* que podemos “empurrar” desde u até v antes de exceder a capacidade $c(u, v)$ é a *capacidade residual* de (u, v) , dada por:

$$c_f(u, v) = c(u, v) - f(u, v) \quad (\text{A.2})$$

Dados uma rede de fluxo $G = (V, E)$ e um fluxo f , a **rede residual** de G induzida por f é $G_f = (V, E_f)$, onde:

$$E_f = \{(u, v) \in V \times V : c_f(u, v) > 0\}$$

Caminhos de ampliação

Dados uma rede de fluxo $G = (V, E)$ e um fluxo f , um **caminho de ampliação** p é um caminho simples desde s até t na rede residual G_f . Pela definição de rede residual, cada aresta (u, v) em um caminho de ampliação admite algum fluxo positivo adicional de u até v sem violar a restrição de capacidade sobre a aresta. Chamamos a quantidade máxima pela qual podemos aumentar o fluxo em cada aresta de um caminho de ampliação p de **capacidade residual** de p , dada por:

$$c_f(p) = \min\{c_f(u, v) : (u, v) \in p\}.$$

Cortes em redes de fluxo

O método de Ford-Fulkerson aumenta repetidamente o fluxo ao longo de diversos caminhos de ampliação até ser encontrado um fluxo máximo. Para isto, o método faz uso do teorema do fluxo máximo e corte mínimo, que será exibido mais adiante.

Um *corte* (S, T) de uma rede de fluxo $G = (E, V)$ é uma partição de V em S e $T = V - S$ tal que $s \in S$ e $t \in T$. Se f é um fluxo, então o *fluxo líquido* pelo corte (S, T) é definido como $f(S, T)$. A *capacidade* do corte (S, T) é $c(S, T)$. Um *corte mínimo* de uma rede é um corte cuja capacidade é mínima dentre todos os cortes da rede.

Teorema do fluxo máximo e corte mínimo

Se f é um fluxo em uma rede de fluxo $G = (V, E)$ com origem s e destino t , então as seguintes condições são equivalentes:

1. f é um fluxo máximo em G .
2. A rede residual G_f não contém caminhos de ampliação.
3. $|f| = c(S, T)$ para algum corte (S, T) de G .

Em outras palavras, o teorema do fluxo máximo e corte mínimo que nos diz que um fluxo é máximo se, e somente se, sua rede residual não contém caminhos de ampliação.

O algoritmo básico de Ford-Fulkerson

Em cada iteração do método de Ford-Fulkerson, encontramos algum caminho de ampliação p e aumentamos o fluxo f em cada aresta de p pela capacidade residual $c_f(p)$. A implementação do método a seguir calcula o fluxo máximo em um grafo $G = (V, E)$ atualizando o fluxo $f(u, v)$ entre cada par u, v de vértices que estão conectados por uma aresta. Se u e v não estão conectados por uma aresta em um ou outro sentido, supomos implicitamente que $f(u, v) = 0$. As capacidades $c(u, v)$ são consideradas dadas juntamente com o grafo, e $c(u, v) = 0$ se $(u, v) \notin E$. A capacidade residual $c_f(u, v)$ é calculada de acordo com a Equação (A.2). A expressão $c_f(p)$ no código é na realidade apenas uma variável temporária que armazena a capacidade residual do caminho p .

Algoritmo 10 Algoritmo básico de Ford-Fulkerson.

```

1: FORD-FULKERSON( $G, s, t$ )
2: para cada aresta  $(u, v) \leftarrow E(G)$  faça
3:    $f(u, v) \leftarrow 0$ 
4:    $f(v, u) \leftarrow 0$ 
5: fim para
6: enquanto existir um caminho  $p$  de  $s$  até  $t$  na rede residual  $G_f$  faça
7:    $c_f(p) \leftarrow \min\{c_f(u, v) : (u, v) \text{ está em } p\}$ 
8:   para cada aresta  $(u, v)$  em  $p$  faça
9:      $f(u, v) \leftarrow f(u, v) + c_f(p)$ 
10:     $f(v, u) \leftarrow -f(u, v)$ 
11:   fim para
12: fim enquanto
13: retorne  $f$ 

```

O algoritmo de FORD-FULKERSON é simplesmente uma expansão do pseudocódigo MÉTODO FORD-FULKERSON apresentado no Algoritmo 9. As linhas 2 a 4 inicializam o

fluxo f como 0. O loop das linhas 5 a 9 encontra repetidamente um caminho de ampliação p em $G - f$ e amplia o fluxo f ao longo de p pela capacidade residual $c_f(p)$. Quando não existe nenhum caminho de ampliação, o fluxo f é um fluxo máximo.

A.3.3 Fluxo Máximo, Cortes Mínimos e Problema da Correspondência

Tendo como base os fundamentos teóricos descritos até o momento, vamos agora verificar como podemos aplicá-los à solução do problema da correspondência em visão estéreo. Para isso, vamos modelar nosso problema como um problema de rotulação de grafos. O principal objetivo é encontrar uma rotulação consistente (representando as disparidades) de um grafo (representando os pixels da imagem). Tal problema pode ser formulado com base na teoria apresentada de fluxo máximo e cortes mínimos em grafos e, geralmente, é separado em problemas de fluxo máximo e de cortes mínimos pois, embora tenhamos visto que o problema de fluxo máximo pode ser resolvido ao encontrarmos um corte mínimo, cada uma dessas abordagens define a estrutura do grafo de uma forma diferente.

A primeira formulação do problema da correspondência estéreo baseado em grafos foi proposta em [Roy e Cox, 1998]. O grafo é construído com base nas formulações propostas para fluxo máximo: o grafo é construído formando uma malha 3D de pontos (x', y', d) , onde (x', y') são coordenadas dos pixels na imagem e d é um valor de disparidade pertencente ao conjunto de valores possíveis, mais uma origem s e um depósito t . Mais formalmente, é definido um grafo $G = (V, E)$ onde V é definido como:

$$V = V^* \cup \{s, t\},$$

onde V^* é a malha 3D:

$$V^* = \{(x', y', d) : x' \in [0 \dots x'_{max}], y' \in [0 \dots y'_{max}], d \in [0 \dots d_{max}]\},$$

onde $(x'_{max} + 1, y'_{max} + 1)$ é dado pelo tamanho da imagem de entrada e d'_{max} é dado pela disparidade máxima. Internamente, a malha é 6-conectada e a origem s é conectada ao plano frontal, enquanto o plano de fundo é conectado ao depósito t . Veja Figura A.5.

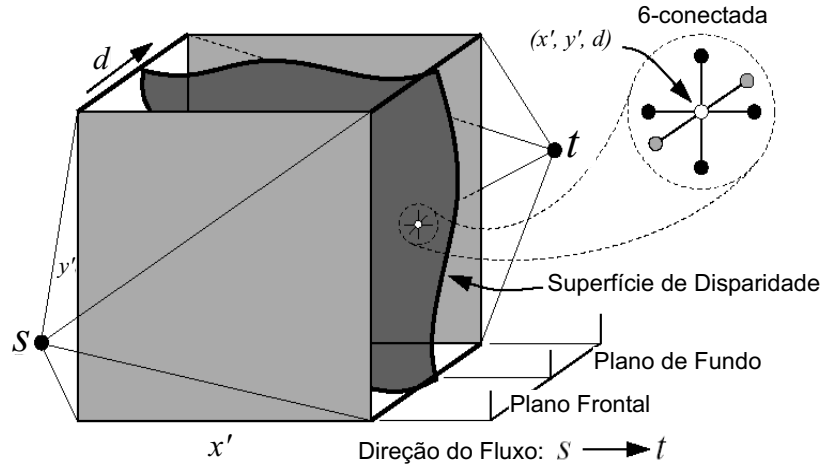


Figura A.5: Correspondência estereó como um problema de fluxo máximo. Imagem obtida em [Roy e Cox, 1998].

Os custos associados às arestas são representados por funções de custo. No grafo definido, o corte mínimo, separando a origem do depósito é encontrado. As disparidades são associadas à cada pixel a partir do corte mínimo da seguinte forma: para cada ponto (x', y') , a maior disparidade associada à aresta ao longo do corte mínimo é selecionada. Note que, dessa forma o problema pode ser interpretado como determinar uma superfície de custo ótimo que separe a origem s do depósito t no volume de disparidade criado.

Posteriormente, apresentando resultados mais promissores, temos uma formulação do problema do corte mínimo proposta em [Boykov et al., 1999] e posteriormente adaptado em [Kolmogorov e Zabih, 2002]. Essas abordagens definem o problema da correspondência estereó da seguinte forma: seja $\mathcal{L}$ o conjunto de pixels na imagem da esquerda, seja $\mathcal{R}$ o conjunto de pixels na imagem da direita, e seja $\mathcal{P}$ o conjunto de todos os pixels: $\mathcal{P} = \mathcal{L} \cup \mathcal{R}$. Um pixel p terá coordenadas (p_x, p_y) . Uma energia funcional apropriada é então construída em função de um conjunto de valores de disparidade. Neste caso, o objetivo é encontrar uma configuração de disparidade f que minimize uma função de energia global construída. Tal tarefa envolve o conceito de *movimentos*.

Considere uma disparidade particular (ou rótulo) α . Uma configuração f' é dita estar em um único *movimento de expansão- α* de f se, para todos os pixels $p \in \mathcal{L}$ $f'_p = f_p$ ou $f'_p = \alpha$ (f_p denota o valor de disparidade de um pixel p). Agora considere um par de disparidade α e β onde, $\alpha \neq \beta$. Uma configuração f' é dita estar em um único *movimento de troca- $\alpha\beta$* de f se, para todos os pixels $p \in \mathcal{L}$ que tenham rótulos α ou β , $f' = \alpha$ ou $f' = \beta$, e para todos os outros pixels $f'_p = f_p$.

O fato crucial acerca destes *movimentos* é que, para uma dada configuração f , é possível

encontrar eficientemente um forte mínimo local de energia; mais especificamente, a menor configuração de energia dentro de um único movimento de expansão- α ou de troca- $\alpha\beta$ de f , respectivamente. Esta *operação de aprimoramento local* pode ser obtida via cortes de grafos. O algoritmo de expansão consiste de uma seqüência de melhorias locais através de expansão- α , para diferentes valores de disparidade α , até que nenhuma expansão- α possa reduzir a energia. Da mesma forma, o algoritmo de troca consiste de uma seqüência de operações de trocas- $\alpha\beta$ para diferentes pares de disparidades α, β , até que nenhuma troca- $\alpha\beta$ possa reduzir a energia. [Kolmogorov e Zabih, 2002] propuseram uma melhoria na formulação da função de energia usada por Boykov, onde as oclusões foram explicitamente representadas.

Apêndice B

Estéreo com Minimização de Energia via Corte de Grafos

O algoritmo de estéreo com minimização de energia via corte de grafos, aqui denominado EMGC (Energy Minimization via Graph Cuts) [Kolmogorov e Zabih, 2002], é o que atualmente vem apresentando os melhores resultados na obtenção de mapas de disparidade de um estéreo denso, conforme estudo comparativo apresentado em [Scharstein et al., 2001]. Além disso, o EMGC responde bem ao uso de um par estéreo com áreas oclusas, que são caracterizadas por pontos da cena visíveis em somente uma das imagens.

B.1 Representação do Problema

Seja $\mathcal{A}$ o conjunto não ordenado de pares de pixels que potencialmente podem ser correspondentes. Para o estéreo com câmeras alinhadas, ou imagens retificadas, é possível fazer o uso da geometria epipolar de tal forma que:

$$\mathcal{A} = \{\langle p, q \rangle | p_y = q_y \text{ e } 0 \leq q_x - p_x < k\}.$$

Assumindo que as disparidades se encontram em um intervalo finito, então cada pixel na imagem da esquerda $\mathcal{L}$ pode potencialmente corresponder a um dos k possíveis pixels na imagem da direita $\mathcal{R}$. O objetivo é encontrar um subconjunto de $\mathcal{A}$ contendo apenas pares de pixels que efetivamente sejam correspondentes. Equivalentemente, é dado a cada assinalamento $a \in \mathcal{A}$ um valor f_a cujo valor é 1 se os pixels p e q se correspondem e 0 em caso contrário. Os assinalamentos que possuem valor 1 são ditos *ativos*.

Seja, então, $A(f)$ o conjunto de assinalamentos ativos de acordo com a configuração

f . Seja $N_p(f)$ o conjunto de assinalamentos ativos em f que envolvem o pixel p , (p.e. $N_p(f) = \{\langle p, q \rangle \in A(f)\}$). Uma configuração f é dita *única* se cada pixel está envolvido em, no máximo, um assinalamento ativo, ou seja, não exista ambigüidade:

$$\forall p \in \{\mathcal{L} \cup \mathcal{R}\} \quad |N_p(f)| \leq 1$$

É importante observar que $|N_p(f)| = 0$ significa que não há um assinalamento ativo e, por consequência, o pixel em questão é uma oclusão.

O problema consiste, então, em obter o conjunto de assinalamentos ativos f de $A(f)$ que representem o conjunto de disparidades do par estéreo.

B.2 O algoritmo de Estéreo Utilizando Movimentação de Expansão da Disparidade

A representação da Seção B.1 é utilizada para permitir a discretização da busca das disparidades em um valor finito k de possibilidades. Isto é feito a partir de um algoritmo de movimento de expansão de disparidade denominado *expansão γ* ¹.

Para um assinalamento $a = \langle p, q \rangle$, seja $d_a = (q_x - p_x, q_y - p_y)$ sua disparidade, e A^γ o conjunto de todos os assinalamentos em A com disparidade γ . A configuração f' é uma expansão γ de f se $A(f')$ é um subconjunto de $A(f) \cup A^\gamma$. Em outras palavras, f' é uma expansão γ de f se alguns assinalamentos ativos em f são desativados e alguns assinalamentos com disparidade γ se tornam ativos.

O método é explicado pelo Algoritmo 11 [Kolmogorov e Zabih, 2002], e consiste em selecionar (ordenada ou aleatoriamente) uma disparidade γ e encontrar uma configuração f' que seja uma expansão γ de f (passo local). Se esta configuração reduz um funcional de energia (passo global) então o algoritmo vai para a próxima iteração, caso contrário o algoritmo termina e retorna o valor de f obtido.

B.3 O Funcional de Energia

O Algoritmo 11 manipula oclusões apropriadamente utilizando uma função de energia que trata as imagens de entrada de forma simétrica e impõe suavidade espacial preservando

¹o nome adotado na literatura é expansão α , entretanto este trabalho utiliza γ para diferenciar do parâmetro α de atenuação da água.

Algoritmo 11 Algoritmo de estéreo convencional (EMGC).

```

1: Inicia com uma configuração única  $f$  arbitrária;
2:  $sucesso := 0$ ;
3: para cada disparidade  $\gamma$  de 1 até  $k$  faça
4:   Encontre  $\hat{f} = \arg \min \mathcal{E}(f')$ , sendo  $f'$  uma expansão  $\gamma$  de  $f$ ;
5:   se  $\mathcal{E}(\hat{f}) < \mathcal{E}(f)$  então
6:      $f := \hat{f}$ ;
7:      $sucesso := 1$ ;
8:   fim se
9: fim para
10: se  $sucesso = 1$  então
11:   vá para 2;
12: fim se
13: Retorna  $f$ ;

```

descontinuidades. A formulação do funcional de energia define uma configuração de disparidade f que pode ser vista como um conjunto discreto de valores de disparidade para uma dada cena. O funcional de energia em questão é então definido a partir de três termos [Kolmogorov e Zabih, 2001][Kolmogorov et al., 2003]:

$$\mathcal{E}(f) = \mathcal{E}_{\text{dados}}(f) + \mathcal{E}_{\text{suavidade}}(f) + \mathcal{E}_{\text{visibilidade}}(f), \quad (\text{B.1})$$

onde:

- $\mathcal{E}_{\text{dados}}$ é o termo que resulta da diferença de intensidade entre pixels das imagens, em geral calculada com custo L_1 (diferença absoluta) ou L_2 (diferença quadrática), dada pela Equação (B.2):

$$\mathcal{E}_{\text{dados}}(f) = \sum_{a \in \mathcal{A}} D(a), \quad (\text{B.2})$$

sendo $a = \langle p, q \rangle$, $D(a) = (I_p - I_q)^2$ e I o valor em níveis de cinza do pixel.

- $\mathcal{E}_{\text{suavidade}}$ é um termo para suavidade que faz com que os pixels da mesma vizinhança na imagem tendam a terem valores de disparidades similares, caso contrário sofrem uma penalidade que aumenta o custo deste termo, dado pela Equação (B.3):

$$\mathcal{E}_{\text{suavidade}}(f) = \sum_{\{a_1, a_2\} \in \mathcal{N}} V_{a_1, a_2} \cdot T(f(a_1) \neq f(a_2)), \quad (\text{B.3})$$

sendo V_{a_1, a_2} uma constante de penalidade calculada pela média das intensidades dos pixels da imagem, e $T(f(a_1) \neq f(a_2)) = 1$ para assinalamentos diferentes (disparidades distintas) em uma vizinhança $\mathcal{N}$, e $T(f(a_1) \neq f(a_2)) = 0$ caso contrário.

- $\mathcal{E}_{visibilidade}$ é um termo para visibilidade que verifica se um pixel é ocluso, e em caso afirmativo impõe uma penalidade que, na prática, equivale a levar a energia do funcional a um valor infinito.

$$\mathcal{E}_{visibilidade}(f) = \sum_{p \in \{\mathcal{L} \cup \mathcal{R}\}} C_p \cdot T(|N_p(f)| = 0), \quad (\text{B.4})$$

onde C_p uma constante de penalidade de valor tendendo a infinito, e $T(|N_p(f)| = 0) = 1$ para pixels oclusos, e $T(|N_p(f)| = 0) = 0$ caso contrário.

B.4 A Minimização via Corte de Grafos

Muitos problemas em visão computacional podem ser formulados em termos de minimização de energia, como estéreo e restauração de imagens. O objetivo principal é encontrar um assinalamento f que minimize um funcional de energia [Veksler, 1999]. Alguns destes problemas podem ser computados eficientemente utilizando *programação dinâmica* (vide Seção A.2), contudo esse método é restrito ao uso de funções de energia de uma única dimensão e em geral não é uma boa solução para funções de energia de duas dimensões. Outra opção é o uso do método *simulated annealing*, entretanto seu custo computacional para uma função arbitrária é exponencial e na prática o algoritmo se torna lento.

O passo crítico do Algoritmo 11 é obter eficientemente $(\hat{f})$ (linha 4). Para tanto, é utilizado o algoritmo de minimização via corte de grafos [Boykov e Kolmogorov, 2001]. Esse método é aplicável apenas a um conjunto limitado de funcionais de energia, como no caso do funcional da Equação (B.1).

Seja $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ um grafo com dois vértices terminais s, t denominados fonte (*source*) e sorvedouro (*sink*). Um corte $\mathcal{C} = \mathcal{V}^s, \mathcal{V}^t$ é uma partição dos vértices em dois conjuntos tais que $s \in \mathcal{V}^s$ e $t \in \mathcal{V}^t$, ou seja, um corte que desconecta s e t formando dois subgrafos. O custo de $\mathcal{C}$ é dado pela soma dos custos de todas as arestas que vão de $\mathcal{V}^s$ a $\mathcal{V}^t$:

$$c(\mathcal{V}^s, \mathcal{V}^t) = \sum_{u \in \mathcal{V}^s, v \in \mathcal{V}^t, (u, v) \in \mathcal{E}} c(u, v). \quad (\text{B.5})$$

O problema do corte mínimo consiste em determinar um corte $\mathcal{C}$ com o menor custo. Este problema é equivalente a computar o fluxo máximo entre os dois terminais, de acordo com o teorema de Ford e Fulkerson [Ford e Fulkerson, 1962].

O algoritmo de fluxo máximo e corte de grafos é apresentado no Apêndice A (Seção A.3). A complexidade do mesmo é polinomial no pior caso, contudo na prática é aproximadamente linear para grafos com diversos caminhos curtos entre os terminais, como no caso que se está utilizando.

A construção do grafo pode ser dada da seguinte forma [Kolmogorov e Zabih, 2001]: Suponha que o algoritmo inicie com uma configuração única f^0 . Os assinalamentos ativos para uma nova configuração em uma expansão γ será um subconjunto de $\tilde{\mathcal{A}} = \mathcal{A}^0 \cup \mathcal{A}^\gamma$, onde $\mathcal{A}^0 = \{a \in \mathcal{A}(f^0) | d(a) \neq \gamma\}$ e $\mathcal{A}^\gamma = \{a \in \mathcal{A} | d(a) = \gamma\}$ (em uma expansão γ ²). Define-se então a configuração $\tilde{f}$ tal que $\mathcal{A}(\tilde{f}) = \tilde{\mathcal{A}}$.

O grafo direcionado $\mathcal{G}$ é construído com vértices que correspondem aos assinalamentos e custos das arestas derivados da Equação (B.1). Para cada vértice $a \in \tilde{\mathcal{A}}$ existirão arestas (s, a) e (a, t) , ou seja, todos estarão conectados aos terminais. Se $\{a_1, a_2\} \in \mathcal{N}$ (são vizinhos), existirão arestas (a_1, a_2) e (a_2, a_1) onde ambas estarão em $\mathcal{A}^0$ ou em $\mathcal{A}^\gamma$. Finalmente, considerando um par de vértices a_1, a_2 que sejam comuns ao mesmo pixel p (p.e., $a_1 = \langle p, q \rangle$ e $a_2 = \langle p, r \rangle$), existirão arestas entre cada par de assinalamentos [Kolmogorov e Zabih, 2001].

Os custos das arestas fazem com que o método minimize o funcional de energia, e podem ser obtidos com maiores detalhes em [Kolmogorov et al., 2003]. A configuração $f^{\mathcal{C}}$ que corresponde ao corte $\mathcal{C}$ é definida por:

$$\begin{aligned} \forall a \in \mathcal{A}^0 \quad f_a^{\mathcal{C}} &= \begin{cases} 1, & \text{se } a \in \mathcal{V}^s \\ 0, & \text{se } a \in \mathcal{V}^t, \end{cases} \\ \forall a \in \mathcal{A}^\gamma \quad f_a^{\mathcal{C}} &= \begin{cases} 1, & \text{se } a \in \mathcal{V}^t \\ 0, & \text{se } a \in \mathcal{V}^s. \end{cases} \end{aligned}$$

Assim, a configuração resultante do corte mínimo $f^{\mathcal{C}}$ será a configuração $\hat{f}$ que minimiza o funcional de energia da Equação (B.1).

²lembrando que assinalamentos ativos podem tornar-se inativos, e assinalamentos inativos cuja disparidade é γ podem tornar-se ativos

Referências Bibliográficas

- [Ackleson, 2003] Ackleson, S. G. (2003). Light in shallow waters: A brief research review. *Journal of Limnology and Oceanography*, 48:323–328.
- [Baribeau et al., 1992] Baribeau, R., Rioux, M., e Godin, G. (1992). Color reflectance modeling using a polychromatic laser range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:263–269.
- [Barnard e Fischler, 1982] Barnard, S. T. e Fischler, M. A. (1982). Computacional stereo. *ACM Computing Surveys*, 14:553–572.
- [Barros, 2004] Barros, W. F. (2004). Reconstrução 3d de cenas com reflectância arbitrária usando o princípio da reciprocidade de helmholtz. Dissertação de mestrado, Universidade Federal de Minas Gerais.
- [Beckmann e Spizzichino, 1963] Beckmann, P. e Spizzichino, A. (1963). *The scattering of electromagnetic waves from rough surfaces*. Pergamon Press.
- [Belhumeur, 1996] Belhumeur, P. N. (1996). A bayesian approach to binocular stereopsis. *International Journal in Computer Vision*, 19:237–262.
- [Belhumeur e Mumford, 1992] Belhumeur, P. N. e Mumford, D. (1992). A bayesian treatment of the stereo correspondence problem using half-occluded regions. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 506–512.
- [Bhat e Nayar, 1998] Bhat, D. N. e Nayar, S. K. (1998). Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):415–423.
- [Birchfield e Tomasi, 1998] Birchfield, S. e Tomasi, C. (1998). Depth discontinuities by pixel-to-pixel stereo. In *IEEE International Conference on Computer Vision*, p. 1073–1080.
- [Birchfield e Tomasi, 1999] Birchfield, S. e Tomasi, C. (1999). Multiway cut for stereo and motion with slanted surfaces. In *IEEE International Conference on Computer Vision*, p. 489–495.

- [Black e Jepson, 1996] Black, M. J. e Jepson, A. D. (1996). Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation. In *European Conference on Computer Vision*, p. 329–342.
- [Blinn, 1977] Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. *ACM SIGGRAPH Conference Proceedings*, 11(4):192–198.
- [Boivin e Gagalowicz, 2001] Boivin, S. e Gagalowicz, A. (2001). Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *ACM SIGGRAPH Conference Proceedings*, p. 107–116.
- [Boivin e Gagalowicz, 2002] Boivin, S. e Gagalowicz, A. (2002). Inverse rendering from a single image. In *European Conference on Color in Graphics, Imaging and Vision*.
- [Born, 1999] Born, M. (1999). *Principles of optics*. Cambridge U. Press.
- [Bouguet, 1999] Bouguet, J.-Y. (1999). Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.
- [Bouguet e Perona, 1998] Bouguet, J.-Y. e Perona, P. (1998). 3D photography on your Desk. In *IEEE International Conference on Computer Vision*, p. 43–50.
- [Boykov e Kolmogorov, 2001] Boykov, Y. e Kolmogorov, V. (2001). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, p. 359–374.
- [Boykov et al., 1999] Boykov, Y., Veksler, O., e Zabih, R. (1999). Fast approximate energy minimization via graph cuts. In *IEEE International Conference on Computer Vision*, p. 377–384.
- [Brown et al., 2003] Brown, M. Z., Burschka, D., e Hager, G. D. (2003). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:993–1008.
- [Buss, 2003] Buss, S. R. (2003). *3D Computer Graphics: A Mathematical Introduction with OpenGL*. Cambridge University Press.
- [Caccia, 1999] Caccia, M. (1999). Vision for estimating the slow motion of unmanned underwater vehicles. Preliminary studies carried out in the framework of a CNR-NATO Grant.
- [Carceroni, 2001] Carceroni, R. L. (2001). *Recovering Non-Rigid 3D Motion, Shape and Reflectance from Multi-View Image Sequences: A Differential-Geometric Approach*. Phd thesis, University of Rochester.
- [Carceroni e Kutulakos, 2002] Carceroni, R. L. e Kutulakos, K. N. (2002). Multi-view scene capture by surfel sampling: From video streams to non-rigid 3d motion, shape and reflectance. *International Journal of Computer Vision*, 49(2-3):175–214.

- [Carceroni et al., 2004] Carceroni, R. L., Pádua, F. L. C., Santos, G. A. M. R., e Kutulakos, K. N. (2004). Linear sequence-to- sequence alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, p. 746–753.
- [Carder et al., 2003] Carder, K. L., Liu, C.-C., Lee, Z., English, D. C., Patten, J., Chen, F. R., e Ivey, J. E. (2003). Illumination and turbidity effects on observing faceted bottom elements with uniform lambertian albedos. *Journal of Limnology and Oceanography*, 48:355–363.
- [Chandrasekhar, 1960] Chandrasekhar, S. (1960). *Radiative Transfer*. Dover Publications.
- [Chen e Medioni, 1999] Chen, Q. e Medioni, G. (1999). A volumetric stereo matching method: Application to image-based modeling. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 29–34.
- [Cipolla e Boyer, 1998] Cipolla, R. e Boyer, E. (1998). 3d model acquisition from uncalibrated images. In *IAPR W. on Machine Vision Applications*, p. 559–568.
- [Cook e Torrance, 1981] Cook, R. e Torrance, K. (1981). A reflectance model for computer graphics. *ACM SIGGRAPH Conference Proceedings*, 15(4):307–316.
- [Cormem et al., 1999] Cormem, T. H., Leiserson, C. E., e Rivest, R. L. (1999). *Introduction to Algorithms*. The MIT Press, Cambridge, Mass., 2nd edição.
- [Corrêa et al., 2002] Corrêa, W. T., Oliveira, M. M., Silva, C. T., e Wang, J. (2002). Modeling and rendering of real environments. *Revista de Informática Teórica e Aplicada*, 9(2):127–156.
- [Cox et al., 1996] Cox, I. J., Hingorani, S. L., Rao, S. B., e Maggs, B. (1996). A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567.
- [Criminisi et al., 1999] Criminisi, A., Reid, I., e Zisserman, A. (1999). Single view metrology. In *IEEE International Conference on Computer Vision*, p. 434–442.
- [Davies-Colley e Smith, 2001] Davies-Colley, R. J. e Smith, D. G. (2001). Turbidity, suspended sediment, and water clarity: A review. *Journal of the American Water Resources Association*, 37:1085–1101.
- [Debevec, 1998] Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *ACM SIGGRAPH Conference Proceedings*, p. 189–198.
- [Debevec et al., 2000] Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., e Sagar, M. (2000). Acquiring the reflectance field of a human face. In *ACM SIGGRAPH Conference Proceedings*, p. 145–156.
- [Debevec et al., 1996] Debevec, P. E., C.J.Taylor, e Malik, J. (1996). Modeling and rendering architecture from photographs:a hybrid geometry-and image-based approach. In *ACM SIGGRAPH Conference Proceedings*, p. 11–20.

- [Dhond e Aggarwal, 1989] Dhond, U. R. e Aggarwal, J. (1989). Structure from stereo - a review. In *IEEE Transactions on Systems, Man, and Cybernetics*, 19:1489–1510.
- [Drettakis et al., 1997] Drettakis, G., Robert, L., e Bougnoux, S. (1997). Interactive common illumination for computer augmented reality. In *Proc. 8th Eurographics W. on Rendering*.
- [Dror et al., 2001] Dror, R., Adelson, E., e Willsky, A. (2001). Surface reflectance estimation and natural illumination statistics.
- [Faugeras, 1993] Faugeras, O. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, MA.
- [Faugeras et al., 1996] Faugeras, O., Hotz, B., Mathieu, H., Viéville, T., Zhang, Z., Fua, P., Théron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P., e Proy, C. (1996). Real time correlation based stereo: algorithm implementations and applications. *The International Journal of Computer Vision*.
- [Faugeras e Keriven, 1998] Faugeras, O. e Keriven, R. (1998). Variational principles, surface evolution, pde's, level set methods, and stereo problem. *IEEE Transactions on Image Processing*, 7:336–344.
- [Faugeras e Luong, 2001] Faugeras, O. e Luong, Q. T. (2001). *The Geometry ou Multipli Images*. The MIT Press.
- [Ford e Fulkerson, 1962] Ford, L. e Fulkerson, D. (1962). *Flows in Networks*. Princeton University Press.
- [Fournier et al., 1993] Fournier, A., Gunawan, A. S., e Romanzin, C. (1993). Common illumination between real and computer generated scenes. In *Proc. Graphics Interface 1993*.
- [Fua e Leclerc, 1995] Fua, P. e Leclerc, Y. G. (1995). Object-centered surface reconstruction: Combining multi-image stereo and shading. *Int'l J. Computer Vision*, 16:35–56.
- [Fusiello et al., 2000] Fusiello, A., Trucco, E., e Verri, A. (2000). A compact algorithm for rectification of stereo pairs. *Journal of Machine Vision and Applications*, 12(1):16–22.
- [Garcia et al., 2002] Garcia, R., Nicosevici, T., e Cufí, X. (2002). On the way to solve lighting problems in underwater imaging. In *Proc. IEEE OCEANS Conference*, p. 1018–1024.
- [Gardner et al., 2003] Gardner, A., Tchou, C., Hawkins, T., e Debevec, P. (2003). Linear light source reflectometry. *ACM Trans. on Graphics*, 22(3):749–758.
- [Glassner, 1995] Glassner, A. S. (1995). *Principles of Digital Images Synthesis*, volume 2. Morgan Kaufmann.

- [Gonzalez e Woods, 2000] Gonzalez, R. C. e Woods, R. E. (2000). *Processamento de Imagens Digitais*. Edgard Blücher.
- [Gordon, 2002] Gordon, H. W. (2002). Inverse methods in hydrologic optics. *Oceanologia*, 44(1):9–58.
- [Harsdorf et al., 1999] Harsdorf, S., Reuter, R., e Töneböen, S. (1999). Contrast enhanced optical imaging of hazardous substances on the seafloor. In *In: Environmental Sensing and Applications. Proc. EUROPTO Series, SPIE*, p. 378–383.
- [Hartley e Zisserman, 2000] Hartley, R. e Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press.
- [Hearn e Baker, 1994] Hearn, D. e Baker, M. P. (1994). *Computer Graphics - C Version*. PrenticeHall, 2nd edição.
- [Hertzmann e Seitz, 2003] Hertzmann, A. e Seitz, S. M. (2003). Shape and materials by example: A photometric stereo approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 533–540.
- [Hingorani et al., 1996] Hingorani, S. L., Cox, I. J., e Rao, S. B. (1996). A maximum likelihood stereo algorithm. In *Computer Vision and Image Understanding*, 63:542–567.
- [Horn, 1986] Horn, B. K. P. (1986). *Robot Vision*. McGraw-Hill.
- [Huguet et al., 2004] Huguet, A. B., Andrade, M. C., Carceroni, R. L., e Araújo, A. A. (2004). Color-based watershed segmentation of low-altitude aerial images. In *SIBGRAPI Conference Proceedings, IEEE Computer Society Press*, p. 138–145.
- [Ikeuchi e Sato, 1991] Ikeuchi, K. e Sato, K. (1991). Determining reflectance properties of an object using range and brightness images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:1139–1153.
- [Intille e Bobick, 1994] Intille, S. S. e Bobick, A. F. (1994). Incorporating intensity edges in the recovery of occlusion regions. In *ICPR94*, p. A:674–677.
- [Ishimaru, 1999] Ishimaru, A. (1999). *Wave propagation and scattering in random media*. Wiley-IEEE Press.
- [Janko, 2003] Janko, Z. (2003). Helmholtz reciprocity in surface reconstruction: an overview. In *Procendings of Grafika 2003*, p. 103–108.
- [Jenkins e White, 1997] Jenkins, F. A. e White, H. E. (1997). *Fundamentals of optics*. McGraw-Hill.
- [Kajiya, 1985] Kajiya, J. T. (1985). Anisotropic reflection models. *ACM SIGGRAPH Conference Proceedings*, 19(4):15–22.

- [Kim et al., 2003] Kim, J., Kolmogorov, V., e Zabih, R. (2003). Visual correspondence using energy minimization and mutual information. In *IEEE International Conference on Computer Vision*, volume 2, p. 13–16.
- [Kirk, 2003] Kirk, J. T. O. (2003). The vertical attenuation of irradiance as a function of the optical properties of the water. *Journal of Limnology and Oceanography*, 48:9–17.
- [Kluth et al., 1992] Kluth, V., Kunkel, G., e Rauhala, U. (1992). Global least squares matching. In *Proceedings Int'l Geoscience and Remote Sensing Symp.*, volume 2, p. 1615–1618.
- [Kolmogorov e Zabih, 2001] Kolmogorov, V. e Zabih, R. (2001). Computing visual correspondence with occlusions via graph cuts. In *IEEE International Conference on Computer Vision*, p. 508–515.
- [Kolmogorov e Zabih, 2002] Kolmogorov, V. e Zabih, R. (2002). Multi-camera scene reconstruction via graph cuts. In *European Conference on Computer Vision*, p. 82–96.
- [Kolmogorov et al., 2003] Kolmogorov, V., Zabih, R., e Gortler, S. (2003). Generalized multi-camera scene reconstruction using graph cuts. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2683, p. 506–512.
- [Koschan, 1993] Koschan, A. (1993). What is new in computational stereo since 1989: A survey of current stereo papers. Technical Report 93-22, Technical Univ. of Berlin.
- [Kutulakos e Seitz, 2000] Kutulakos, K. N. e Seitz, S. M. (2000). A theory of shape by space carving. *International Journal in Computer Vision*, 38(3):199–218.
- [Lafortune e Willems, 1994] Lafortune, E. P. e Willems, Y. D. (1994). Using the Modified Phong BRDF for Physically Based Rendering. Technical Report CW197, Department of Computing Science.
- [Lambert, 1760] Lambert, J. H. (1760). *Photometria sive de mensura de gratibus luminis, colorum umbrae*. Eberhard Klett.
- [Levoy et al., 2004] Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., e Bolas, M. (2004). Synthetic aperture confocal imaging. *ACM Trans. Graph.*, 23(3):825–834.
- [Lewis, 1993] Lewis, R. R. (1993). Making shaders more physically plausible. In *Fourth Eurographics W. on Rendering*, p. 47–62.
- [Loscos et al., 1999] Loscos, C., Frasson, M.-C., Drettakis, G., Walter, B., Granier, X., e Poulin, P. (1999). Interactive virtual relighting and remodeling of real scenes. In *Proc. 10th Eurographics W. on Rendering*, volume 10, p. 235–246.
- [Lucas e Kanade, 1981] Lucas, B. e Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI81*, p. 674–679.

- [Luria e Kinney, 1970] Luria, S. M. e Kinney, J. A. (1970). Underwater vision. *Science*, 167:1454–1461.
- [Mansouri et al., 1998] Mansouri, A. R., Mitiche, A., e Konrad, J. (1998). Selective image diffusion: Application to disparity estimation. In *IEEE International Conference on Computer Vision*, volume 3, p. 284–288.
- [Marschner e Greenberg, 1998] Marschner, S. R. e Greenberg, D. P. (1998). *Inverse Rendering for Computer Graphics*. Phd thesis, Cornell University.
- [Matusik et al., 2003] Matusik, W., Pfister, H., Brand, M., e McMillan, L. (2003). A data-driven reflectance model. *ACM Trans. on Graphics*, 22(3):759–769.
- [Matusik et al., 2002] Matusik, W., Pfister, H., Ziegler, R., Ngan, A., e McMillan, L. (2002). Acquisition and rendering of transparent and refractive objects. In *In Eurographics '02 Proc.*, p. 19–24.
- [McCluney, 1994] McCluney, W. R. (1994). *Introduction to Radiometry and Photometry*. Artech House.
- [McCormick, 2004] McCormick, N. J. (2004). Analytic inverse radiative transfer equations for atmospheric and hydrologic optics. *Journal of Optical Society of America*, 21:1009–1017.
- [Miyazaki et al., 2003a] Miyazaki, D., Kagesawa, M., e Ikeuchi, K. (2003a). Polarization-based transparent surface modeling from two views. In *IEEE International Conference on Computer Vision*, p. 1381–1386.
- [Miyazaki et al., 2003b] Miyazaki, D., Tan, R. T., Hara, K., e Ikeuchi, K. (2003b). Polarization-based inverse rendering from a single view. In *IEEE International Conference on Computer Vision*, p. 982–987.
- [Muhlmann et al., 2001] Muhlmann, K., Maier, D., Hesser, J., e Manner, R. (2001). Calculating dense disparity maps from color stereo images, an efficient implementation. In *Proc. IEEE Workshop Stereo and Multi-Baseline Vision*, p. 30–36.
- [Narasimhan e Nayar, 2000] Narasimhan, S. G. e Nayar, S. K. (2000). Chromatic framework for vision in bad weather. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, p. 1598–1603.
- [Narasimhan e Nayar, 2002] Narasimhan, S. G. e Nayar, S. K. (2002). Vision and the atmosphere. *International Journal in Computer Vision*, 48(3):233–254.
- [Narasimhan et al., 2002] Narasimhan, S. G., Wang, C., e Nayar, S. K. (2002). All the images of an outdoor scene. In *European Conference on Computer Vision*.
- [Nayar et al., 1991] Nayar, S. K., Ikeuchi, K., e Kanade, T. (1991). Surface reflections: Physical and geometrical perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):611–634.

- [Nayar e Narasimhan, 1999] Nayar, S. K. e Narasimhan, S. G. (1999). Vision in bad weather. In *IEEE International Conference on Computer Vision*, volume 2, p. 820–827.
- [Negahdaripour e Khamene, 2000] Negahdaripour, S. e Khamene, A. (2000). Motion-based compression of underwater video imagery for the operations of unmanned submersible vehicles. *Special Issue on Underwater Computer Vision and Pattern Recognition*, 79(1).
- [Negahdaripour e Yu, 1995] Negahdaripour, S. e Yu, C. (1995). *Underwater Robotic vehicles: Design and Control*. TSI Press.
- [Nery, 2004] Nery, M. S. (2004). Identificação de espécies de peixes utilizando visão computacional: Estudo de caso da escada de peixes da usina hidrelétrica de Igarapava. Dissertação de mestrado, Universidade Federal de Minas Gerais.
- [Nishino et al., 2001] Nishino, K., Zhang, Z., e Ikeuchi, K. (2001). Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *IEEE International Conference on Computer Vision*, p. 599–606.
- [Ohta e Kanade, 1985] Ohta, Y. e Kanade, T. (1985). Stereo by intra- and intra-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:139–154.
- [Oliveira, 2002] Oliveira, M. M. (2002). Image-based modeling and rendering techniques: A survey. *Revista de Informática Teórica e Aplicada*, 9(2):37–66.
- [Oren e Nayar, 1995] Oren, M. e Nayar, S. K. (1995). Generalization of the lambertian model and implications for machine vision. *International Journal in Computer Vision*, 14(3):227–251.
- [Phong, 1975] Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 18(8):311–317.
- [Press et al., 1992] Press, W. H., Flannery, B. P., Teukolsky, S. A., e Vetterling, W. T. (1992). *Numerical Recipes in C*. Cambridge University Press, 2nd edição.
- [Queiroz-Neto et al., 2004] Queiroz-Neto, J. P., Carceroni, R. L., Barros, W. F., e Campos, M. F. M. (2004). Underwater stereo. In *SIBGRAPI Conference Proceedings, IEEE Computer Society Press*, p. 170–177.
- [Queiroz-Neto et al., 2005] Queiroz-Neto, J. P., Carceroni, R. L., e Coelho, L. C. R. (2005). Recovering Photometric Properties of Multiple Strongly-Reflective, Partially-Transparent Surfaces from a Single Image. Aceito na *X IEEE International Conference on Computer Vision*, Beijing, China.
- [Rousseeuw e Leroy, 1987] Rousseeuw, P. J. e Leroy, A. M. (1987). *Robust Regression & Outlier Detection*. Academic Press.

- [Roy e Cox, 1998] Roy, S. e Cox, I. J. (1998). A maximum-flow formulation of the N-camera stereo correspondence problem. In *IEEE International Conference on Computer Vision*, p. 492–499.
- [Saito et al., 1995] Saito, H., Kawamura, H., e Nakajima, M. (1995). 3d shape measurement of underwater objects using motion stereo. In *21th International Conference on Industrial Electronics, Control, and Instrumentation*, p. 1231 – 1235.
- [Santos, 2001] Santos, M. (2001). *Estradas Reais: introdução ao estudo dos caminhos do ouro e do diamante no Brasil*. Editora Estrada Real.
- [Sato et al., 2000] Sato, Y., Wheeler, M. D., e Ikeuchi, K. (2000). Object shape and reflectance modeling from observation. In *ACM SIGGRAPH Conference Proceedings*, p. 379–387.
- [Scharstein e Szeliski, 1998] Scharstein, D. e Szeliski, R. (1998). Stereo matching with non-linear diffusion. *International Journal in Computer Vision*, 28(2):155–174.
- [Scharstein e Szeliski, 2002] Scharstein, D. e Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal in Computer Vision*, 47(1):7–42.
- [Scharstein et al., 2001] Scharstein, D., Szeliski, R., e Zabih, R. (2001). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision, Kauai, HI*.
- [Schechner e Karpel, 2004] Schechner, Y. Y. e Karpel, N. (2004). Clear underwater vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, p. 536–543.
- [Schlick, 1993] Schlick, C. (1993). A Customizable Reflectance Model for Everyday Rendering. In *Fourth Eurographics W. on Rendering*, p. 73–84.
- [Schlüns, 1997] Schlüns, K. (1997). The irradiance error and its effect in photometric stereo. In *Proc. First Joint Australia and New Zealand Biennial Conf. on Digital Image and Vision Computing: Techniques and Applications*, p. 539–544.
- [Schmid e Zisserman, 1998] Schmid, C. e Zisserman, A. (1998). The geometry and matching of curves in multiple views. In *European Conference on Computer Vision*, p. 394–409.
- [Shah, 1993] Shah, J. (1993). A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 34–40.
- [Silva e Santos-Victor, 2000] Silva, C. e Santos-Victor, J. (2000). Intrinsic images for dense stereo matching with occlusions. In *European Conference on Computer Vision*, volume 1, p. 100–114.

- [Sun et al., 2002] Sun, J., Shum, H. Y., e Zheng, N. N. (2002). Stereo matching using belief propagation. In *European Conference on Computer Vision*, p. 510–524.
- [Tagare e deFigueiredo, 1991] Tagare, H. e deFigueiredo, R. (1991). A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:133–152.
- [Thomos et al., 1998] Thomos, I., Malasiotis, S., e Srinatzis, M. (1998). Optimized block based disparity estimation in stereo systems using a maximumflow approach. In *SIB-GRAPI Conference Proceedings, IEEE Computer Society Press*.
- [Tillet et al., 2000] Tillet, R., McFarlane, N., e Lines, J. (2000). Estimating dimensions of free-swimming fish using 3d point distribution models. *Special Issue on Underwater Computer Vision and Pattern Recognition*, 79(1):123–141.
- [Tomasi e Manduchi, 1996] Tomasi, C. e Manduchi, R. (1996). Stereo without search. In *European Conference on Computer Vision*, p. 452–465.
- [Tomasi e Manduchi, 1998] Tomasi, C. e Manduchi, R. (1998). Stereo matching as a nearest-neighbor problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:333–340.
- [Tominaga e Tanaka, 2000] Tominaga, S. e Tanaka, N. (2000). Estimating reflection parameters from a single color image. *IEEE Comp. Graph. and Applications*, 20:58–66.
- [Torrance e Sparrow, 1967] Torrance, K. E. e Sparrow, E. M. (1967). Theory for off-specular reflection from roughened surfaces. *J. of the Optical Society of America*, 57(9):1105–1114.
- [Trucco e Murino, 2000] Trucco, A. e Murino, V. (2000). Guest editors' introduction. *Special Issue on Underwater Computer Vision and Pattern Recognition*, 79(1).
- [Tsai, 1986] Tsai, R. Y. (1986). An efficient and accurate camera calibration technique for 3d machine vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. 364–374.
- [Tsai, 1987] Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, RA-3(4):323–344.
- [Tu e Mendonca, 2003] Tu, P. e Mendonca, P. (2003). Surface reconstruction via helmholtz reciprocity with a single image pair. In *IEEE Conference on Computer Vision and Pattern Recognition*, p. I: 541–547.
- [Veksler, 1999] Veksler, O. (1999). *Efficient Graph-based Energy Minimization Methods in Computer Vision*. Phd thesis, Cornell University.

- [Venkateswar e Chellappa, 1998] Venkateswar, V. e Chellappa, R. (1998). Hierarchical stereo and motion correspondence using feature groupings. *International Journal in Computer Vision*, 15:245–269.
- [Waki et al., 2005] Waki, H., Kim, S., e Kojima, M. (2005). Sums of squares and semi-definite programming relaxations for polynomial optimization problems with structured sparsity. Technical Report B-411, Dep. of Math. and Comp. Sciences - Tokyo Inst. of Technology.
- [Ward, 1992] Ward, G. J. (1992). Measuring and modeling anisotropic anisotropic reflection. In *ACM SIGGRAPH Conference Proceedings*, p. 265–272.
- [Weng et al., 1993] Weng, J., Ahuja, N., e Huang, T. S. (1993). Optimal motion and structure estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):864–884.
- [Willson, 2003] Willson, R. (2003). Tsai camera calibration software - release 3.0b3. Available in: <<http://www.ph.tn.tudelft.nl/PRInfo/software/msg00054.html>>. Acessado em 2003.
- [Wolff et al., 1998] Wolff, L. B., Nayar, S. K., e Oren, M. (1998). Improved diffuse reflection models for computer vision. *International Journal in Computer Vision*, 30(1):55–71.
- [Wong et al., 1997] Wong, T., Heng, P., Or, S., e Ng, W. (1997). Image-based rendering with controllable illumination. In *Proc. of the 8-th Eurographics W. on Rendering*, p. 13–22.
- [Woodham et al., 1991] Woodham, R. J., Iwahori, Y., e Barman, R. A. (1991). Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength. In *UBC Computer Science TR-91-18*.
- [Yitzhaky et al., 1998] Yitzhaky, Y., Dror, I., e Kopeika, N. S. (1998). Restoration of atmospherically blurred images according to weather-predicted atmospheric modulation transfer functions. *Optical Engineering*, 36:3064–3072.
- [Yu et al., 1999] Yu, Y., Debevec, P., Malik, J., e Hawkins, T. (1999). Inverse global illumination: Recovering reflectance models of real scene from photographs. In *ACM SIGGRAPH Conference Proceedings*, volume 19, p. 215–224.
- [Zabih e Woodfill, 1994] Zabih, R. e Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision*, p. 151–158.
- [Zaneveld et al., 2005] Zaneveld, J., Twardowski, M., Barnard, A., e Lewis, M. (2005). Introduction to radiative transfer. In *In Remoting Sensing of Coastal Aquatic Environments - chapter 1*, p. 1–21.
- [Zhao, 2000] Zhao, H. (2000). Global optimal surface from stereo. In *International Conference on Pattern Recognition*, p. 101–104.

- [Zickler et al., 2002] Zickler, T., Belhumeur, P. N., e Kriegman, D. J. (2002). Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *European Conference on Computer Vision*, p. 869–884.