



MAPAS DE PROFUNDIDADE DE SUPERFÍCIES TEXTURIZADAS
BASEADOS EM *DEPTH HALLUCINATION*

José Almeida Júnior

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Mariane Rembold Petraglia
José Gabriel Rodriguez
Carneiro Gomes

Rio de Janeiro
Março de 2012

MAPAS DE PROFUNDIDADE DE SUPERFÍCIES TEXTURIZADAS
BASEADOS EM *DEPTH HALLUCINATION*

José Almeida Júnior

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Mariane Rembold Petraglia, Ph.D.

Prof. José Gabriel Rodriguez Carneiro Gomes, Ph.D.

Prof. Antonio Alberto Fernandes de Oliveira, D.Sc.

Prof. Lisandro Lovisolo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2012

Almeida Júnior, José

Mapas de profundidade de superfícies texturizadas baseados em *Depth Hallucination*/José Almeida Júnior. – Rio de Janeiro: UFRJ/COPPE, 2012.

XIII, 60 p.: il.; 29,7cm.

Orientadores: Mariane Rembold Petraglia

José Gabriel Rodriguez Carneiro Gomes

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2012.

Referências Bibliográficas: p. 59 – 60.

1. *Depth Hallucination*. 2. Mapa de Profundidades.
3. Albedo. 4. *Shading Map*. 5. CSharp. I. Petraglia, Mariane Rembold *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Aos meus pais.

Agradecimentos

Inicialmente, gostaria de agradecer aos meus pais, Ana Maria de Barros Dias Almeida e José Almeida Filho, pela educação que foi dada à mim e à minha irmã, pelo incentivo na busca dos nossos sonhos e ideais, pela liberdade na escolha e no traçado do nosso futuro, além de todo apoio nas decisões que tomamos diariamente.

Agradeço à minha irmã, Mariana Dias Almeida, por ser minha eterna companheira em todos os momentos da minha vida, sejam eles bons ou ruins, além de ser um exemplo de força, de dedicação e de que devemos sempre fazer aquilo que amamos, superando todas as barreiras e empecílios para isso.

Agradeço à Ana Paula Puga por ter me passado sempre muito apoio, principalmente na etapa final do projeto, me acalmando e equilibrando. Como você já sabe, obrigado por ter ficado olhando, olhando, olhando...

Aos professores Mariane R. Petraglia e José Gabriel R. C. Gomes por terem sido meus orientadores, por confiarem no meu trabalho e, sobretudo, por me oferecerem trabalhar com processamento de imagem, o que resultou em uma nova paixão na minha vida.

À minha grande amiga Sandra Gonzaga por me ajudar em todos os momentos, principalmente os mais difíceis, esclarecendo minhas dúvidas e iluminando o meu caminho.

À toda equipe do laboratório PADS por oferecer toda a estrutura e suporte para este projeto. Ao Rafael Szendrodi por dar todo o suporte na instalação dos *softwares* no laboratório para a realização deste projeto e por manter tudo funcionando adequadamente.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MAPAS DE PROFUNDIDADE DE SUPERFÍCIES TEXTURIZADAS
BASEADOS EM *DEPTH HALLUCINATION*

José Almeida Júnior

Março/2012

Orientadores: Mariane Rembold Petraglia
José Gabriel Rodriguez Carneiro Gomes

Programa: Engenharia Elétrica

Esta dissertação tem como objetivo pesquisar, avaliar e principalmente testar as principais etapas que o compõem o processo de *Depth Hallucination*: a captura e processamento das imagens, além da visualização dos mapas de profundidades. Além disso, utilizando-se a linguagem de programação C# (CSharp), foi desenvolvido um *software* contendo todo o algoritmo estudado e criado através de uma interface gráfica amigável e de fácil manuseio, permitindo que os bons resultados, provenientes da utilização do processo, sejam utilizados no dia-a-dia de profissionais das áreas de efeitos visuais, arquitetura, *games*, entre outras.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SURFACE DEPTH MAP BASED ON DEPTH HALLUCINATION

José Almeida Júnior

March/2012

Advisors: Mariane Rembold Petraglia

José Gabriel Rodriguez Carneiro Gomes

Department: Electrical Engineering

This work has as main goals in research, evaluation and test of the Depth Hallucination process: capture the images, image processing and surface depth visualization. Using the *C#* (CSharp) programming language, we developed a software that contains all algorithms of this technique and presents a friendly and simple interface. This allows that the good results from the technique could be used by professionals from different areas, such as visual effects, architecture and games.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
1 Introdução	1
1.1 Organização da Dissertação	2
2 <i>Depth Hallucination</i> e Teoria Fotográfica	3
2.1 Textura, Albedo e Mapa de Profundidades	3
2.2 <i>Depth Hallucination</i>	4
2.2.1 Captura das Imagens	5
2.2.2 Calculando o Albedo e o <i>Shading Image</i>	7
2.2.3 Estimação do Mapa de Profundidades	8
2.2.4 Reiluminação da Superfície	15
2.3 Teoria Fotográfica	15
2.3.1 Câmera	15
2.3.2 Arquivo <i>Raw</i> ("Cru")	17
2.3.3 Diafragma	18
2.3.4 Velocidade do Obturador	22
2.3.5 ISO	23
2.3.6 Fotômetro	25
2.3.7 Balanço de Branco ou <i>WB</i> (White Balance)	29
2.3.8 Lentes Fotográficas (Objetivas)	30
3 Processo de Geração de Mapas de Profundidades	33
3.1 Como Capturar as Imagens	33
3.1.1 Ajustes Iniciais da Câmera e Acessórios	33
3.1.2 Capturando as Imagens	35
3.2 <i>Software</i> Desenvolvido	39
3.2.1 Interface Gráfica	39
3.2.2 Conversão das Imagens	40
3.2.3 Funções	42

4	Visualização do Mapa de Profundidades	48
4.1	Sigma e Número de Níveis do <i>Gaussian Blur</i>	49
4.2	Resultados e Limitações	51
5	Conclusão	57
	Referências Bibliográficas	59

Lista de Figuras

2.1	Exemplificação de textura, neste caso, o padrão da parede de tijolos.	4
2.2	Etapas da técnica <i>Depth Hallucination</i>	5
2.3	Exemplo de <i>diffuse-lit</i>	6
2.4	Exemplo de <i>flash-lit</i>	6
2.5	Exemplo de albedo.	7
2.6	Exemplo de superfície Lambertiana branca.	8
2.7	Exemplo de <i>shading image</i>	9
2.8	Exemplificação da utilização do modelo para picos e depressões localizados respectivamente acima e abaixo do plano que divide a superfície (baseado em [1]).	9
2.9	Porção do céu visível através do ponto P e incidência de luz por um ângulo sólido.	10
2.10	Variação da intensidade da luz conforme o cosseno do ângulo de incidência.	10
2.11	Modelo cilindro para os pontos localizados abaixo do plano médio da superfície (baseado em [1]).	11
2.12	Modelo hemisfério para os pontos localizados acima do plano médio da superfície (baseado em [1]).	12
2.13	Exemplificação do blur gaussiano em 4 (quatro) níveis.	14
2.14	Representação da incidência da luz dentro de uma DSLR.	16
2.15	Representação da incidência da luz ao capturarmos uma imagem.	16
2.16	Quanto maior o número de <i>f-stops</i> , menor a abertura do diafragma.	18
2.17	Aumento e diminuição da quantidade de luz com a variação da abertura do diafragma.	19
2.18	Exemplificação da região próxima ao plano focal que permanecerá nítida na imagem capturada.	20
2.19	Abertura do diafragma em $f3.5$. Apenas o primeiro elemento focado.	21
2.20	Abertura do diafragma em $f22$. Toda a cena focada.	21
2.21	Aumento e diminuição da quantidade de luz com a variação da velocidade do obturador.	22

2.22	Exemplificação de fotografias capturadas com diferentes velocidades do obturador.	23
2.23	ISO 100 e ISO 1600 respectivamente.	24
2.24	Aumento e diminuição da quantidade de luz com a variação do ISO.	24
2.25	Fotômetro (<i>display</i>).	25
2.26	Exemplo de fotografias capturadas com diferentes exposições.	26
2.27	Objeto de interesse.	27
2.28	Imagens capturadas de objetos localizados à frente de um fundo branco.	27
2.29	Imagens capturadas de objetos localizados à frente de um fundo escuro.	28
2.30	Exemplificação dos histogramas de fotografias capturadas com diferentes exposições.	28
2.31	Temperatura de cor para diferentes tipos de iluminação.	29
2.32	Resultados da aplicação de diferentes tipos de WB.	30
2.33	Ângulos de visão para diferentes distâncias focais, além da respectiva classificação das objetivas baseadas nas mesmas.	31
2.34	Fotografia capturada com uma lente grande-angular.	31
2.35	Fotografia capturada com uma lente teleobjetiva.	32
3.1	Câmera na função Manual (M).	34
3.2	Selecionando AF ou MF.	36
3.3	Exemplificação de uma imagem capturada utilizando-se uma forte potência do <i>flash</i>	37
3.4	Exemplificação de uma correta captura das imagens difusa e com <i>flash</i>	38
3.5	Símbolo indicando a localização do sensor na câmera.	39
3.6	Interface gráfica do <i>software</i> desenvolvido.	40
3.7	Imagens <i>raw</i> sendo exibidas no <i>Adobe Camera Raw</i>	41
3.8	Ajustando o processamento do <i>Photoshop</i> para 16 bits/canal.	42
3.9	Representação das funções desenvolvidas para o processo de <i>Depth Hallucination</i>	42
4.1	<i>Software Height Map Viewer</i> utilizado para a visualização dos mapas de profundidades. Cortesia de [2].	48
4.2	Mapa de profundidades obtido utilizando-se $\sigma = 10$ e níveis de <i>blur</i> igual a 3.	49
4.3	Mapa de profundidades obtido utilizando-se $\sigma = 50$ e níveis de <i>blur</i> igual a 4.	50
4.4	Mapa de profundidades obtido utilizando-se $\sigma = 100$ e níveis de <i>blur</i> igual a 5.	50
4.5	Mapa de profundidades obtido utilizando-se $\sigma = 100$ e níveis de <i>blur</i> igual a 6.	51

4.6	Pequeno pedaço de plástico presente na superfície de interesse.	52
4.7	Visualização do mapa de profundidades do piso de um estacionamento.	52
4.8	Resultado da presença do pequeno pedaço de plástico transparente na superfície.	53
4.9	Imagens difusa e com <i>flash</i> para a criação do mapa de profundidade da calçada.	53
4.10	Visualização do mapa de profundidades da calçada.	53
4.11	Resultado da presença do pequeno pedaço de papel na superfície.	54
4.12	Imagens difusa e com <i>flash</i> com incidência direta de luz para a criação do mapa de profundidade da parede.	54
4.13	Visualização do mapa de profundidades da parede com incidência direta de luz.	55
4.14	Imagens difusa e com <i>flash</i> para a criação do mapa de profundidade do bueiro.	55
4.15	Visualização do mapa de profundidades do bueiro.	56

Lista de Tabelas

3.1	Pseudo-código para leitura da imagens selecionadas pelo usuário. . . .	43
3.2	Pseudo-código para cálculo dos fatores de correção.	43
3.3	Pseudo-código para cálculo do albedo.	44
3.4	Pseudo-código para cálculo do <i>Shading</i>	45
3.5	Pseudo-código para cálculo do <i>depth map</i>	46
3.6	Pseudo-código para cálculo da pirâmide Laplaciana.	47
3.7	Pseudo-código para cálculo a função <i>Combined Model</i>	47

Capítulo 1

Introdução

A captura detalhada de características da geometria de uma determinada superfície é uma tarefa desafiadora. Superfícies texturizadas, como tijolos, pedras, madeira, entre outros, possuem características próprias na sua meso-estrutura que as diferenciam uma das outras.

Diferentes técnicas são utilizadas para a captura dessas características. Por exemplo, em uma delas, utilizam-se equipamentos especializados, como *scanners* a laser. Estes, apesar da sua alta precisão, são de alto custo, utilizam um intenso e longo trabalho manual e a malha 3D final geralmente apresenta falhas em sua estrutura, necessitando correções nas etapas posteriores.

Uma nova técnica, chamada *Depth Hallucination* [1], utilizando uma câmera DSLR (*Digital Single-Lens Reflex*) com um único posicionamento e através do *shape-from-shading* multi-escalar, produz, através de duas imagens, um mapa do albedo e um mapa de profundidades da textura da superfície de interesse com alta fidelidade e precisão (foto realismo). A principal diferença entre as duas imagens capturadas é que uma delas é obtida utilizando-se o *flash*, o qual pode ser da própria câmera fotográfica ou externo à mesma. Além disso, o custo é muito baixo, uma vez que são necessários uma câmera e alguns acessórios fotográficos.

Com esta técnica, a superfície 3D, posteriormente processada, pode ser reiluminada e vista sob qualquer ângulo, podendo ser utilizada em aplicações de arquitetura, *games* e efeitos visuais.

A criação de modelos 3D a partir de fotografias não é um processo novo [3]. Além de apresentar um ótimo resultado final, a relação custo benefício é extremamente vantajosa. Entretanto, nesse tipo de processo, as características detalhadas da superfície (meso-estrutura) que compõem os modelos raramente são levadas em consideração.

Apesar de apresentarem bons resultados, alguns métodos de reconstrução de meso-estrutura e obtenção de mapas das normais e texturas utilizam múltiplas fontes de captação para a realização da operação [4]. Existem outros métodos de

reconstrução do albedo e da meso-estrutura [5]. Entretanto, estes também necessitam de um conjunto de imagens e/ou equipamentos especializados para a realização das operações.

O processo de *Depth Hallucination* é inspirado em trabalhos bem sucedidos que conseguiram, sob certas circunstâncias, demonstrar a nossa falta de habilidade para interpretar corretamente a profundidade e a luz [6], permitindo a criação sintética de imagens a partir do conceito de aproximação *dark-is-deep* [7].

Esta dissertação tem como objetivo pesquisar, avaliar e principalmente testar todo o processo de *Depth Hallucination* nas principais etapas que compõem o processo: a captura e o processamento das imagens, além da visualização dos mapas de profundidades. Visamos não só a compreensão e aplicação da técnica, como também o preenchimento de possíveis lacunas existentes tanto na explicação teórica como na parte prática (captura das imagens). Além disso, foi desenvolvido um *software* contendo o algoritmo estudado e criado, permitindo a utilização deste método através de uma interface gráfica amigável e de fácil manuseio.

1.1 Organização da Dissertação

No Capítulo 2 será abordada a parte teórica dessa dissertação. Serão abordadas todas as formulações utilizadas na definição e desenvolvimento do *Depth Hallucination* [1]. Além disso, será apresentada uma parte básica, porém essencial, relativa à teoria fotográfica, uma vez que a obtenção de bons mapas de profundidade e textura dependem diretamente de uma captura precisa das imagens da superfície de interesse.

No Capítulo 3 abordaremos toda a metodologia para a captura correta das imagens da superfície (parte prática). Apresentaremos todos os acessórios fotográficos necessários, bem como algumas boas práticas durante o processo de aquisição das imagens. Em seguida, abordaremos algumas etapas de criação do *software* desenvolvido com todo o algoritmo do processo de *Depth Hallucination*. Explicaremos as principais características e funções desenvolvidas para obtermos uma interface amigável para o usuário, além do melhor desempenho possível tanto no processamento quanto na obtenção de uma representação foto real da superfície de interesse.

Os resultados serão apresentados no Capítulo 4 onde uma análise crítica será feita sobre os resultados obtidos a partir da avaliação da reconstrução da superfície de interesse utilizando-se os mapas de profundidade e textura obtidos.

Por fim, a conclusão e os trabalhos futuros serão apresentados no Capítulo 5.

Capítulo 2

Depth Hallucination e Teoria Fotográfica

Neste capítulo, apresentaremos a teoria envolvida na técnica de *Depth Hallucination* [1] (Seção 2.2), além da teoria fotográfica (Seção 2.3) necessária para uma correta captura das imagens, as quais serão posteriormente processadas para a obtenção dos mapas de profundidade e textura. Porém, primeiramente definiremos na Seção 2.1 textura, albedo e mapa de profundidades, características extremamente importantes para uma correta compreensão de todo o processo de *Depth Hallucination*.

2.1 Textura, Albedo e Mapa de Profundidades

Textura é a variação de informações em escalas menores que a escala de interesse [8]. Por exemplo, na Figura 2.1, se estivermos interessados em identificar o ambiente onde a fotografia foi tirada, o padrão da parede de tijolos pode ser considerado como textura. Entretanto, se estivermos interessados em identificar cada um dos tijolos, cada um destes pode ser considerado um objeto não-texturizado, pois com a imagem nesta escala, não podemos verificar os pequenos detalhes que diferenciam um tijolo do outro. Assim, textura é o aspecto de uma superfície ou seja, a "pele" de uma forma que permita identificá-la e distingui-la de outras formas.

O albedo de uma superfície pode ser visto como uma função que caracteriza o material que constitui uma determinada superfície. Ele fornece a fração de luz incidente que o material reflete para cada comprimento de onda [8].

O mapa de profundidades pode ser definido como uma matriz contendo os valores de depressão ou altura dos pontos que compõem uma determinada superfície. No caso do processo de *Depth Hallucination*, trata-se de uma matriz bidimensional, com a mesma resolução das imagens capturadas, contendo os valores das alturas

(depressões ou picos) de cada *pixel* da imagem da superfície de interesse.

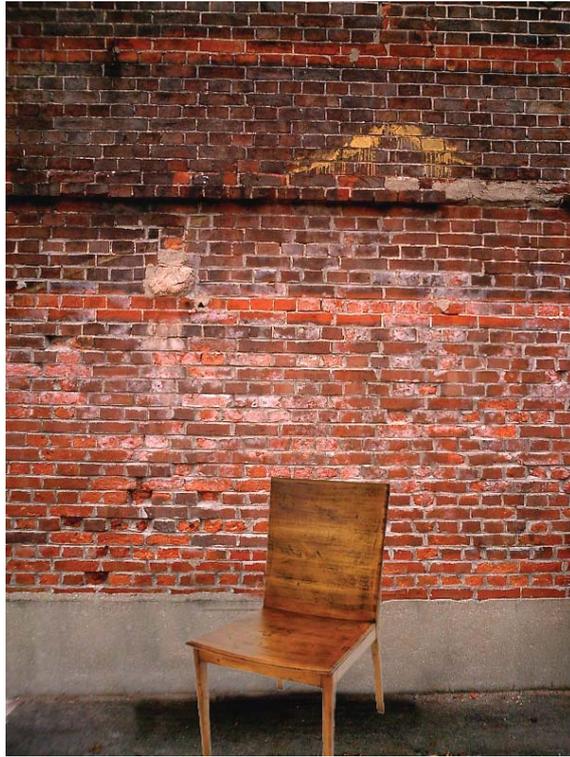


Figura 2.1: Exemplificação de textura, neste caso, o padrão da parede de tijolos.

2.2 *Depth Hallucination*

O processo de *Depth Hallucination* é constituído das seguintes etapas individuais:

- Captura das imagens (Seção 2.2.1);
- Estimação do albedo e *shading* (Seção 2.2.2);
- Estimação do mapa de profundidades (Seção 2.2.3);
- Reiluminação da superfície criada (Seção 2.2.4);

Na Figura 2.2 são apresentadas as etapas do processo *Depth Hallucination* bem como a relação entre as mesmas.

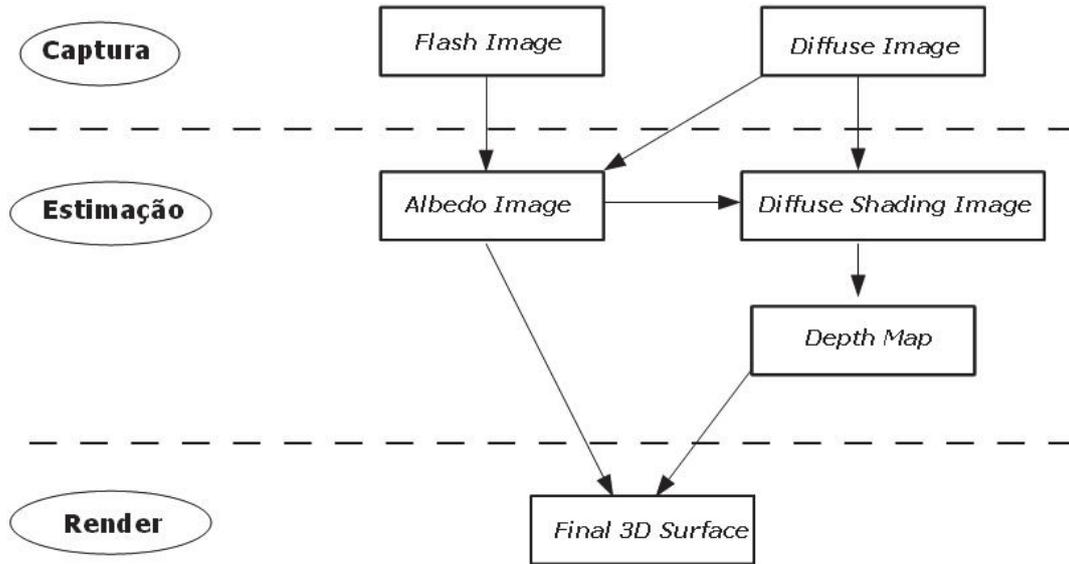


Figura 2.2: Etapas da técnica *Depth Hallucination*.

2.2.1 Captura das Imagens

Na etapa de captura das imagens, faz-se necessária a utilização de uma câmera DSLR (*Digital Single Lens Reflex*), pois a mesma, além de apresentar um sensor de melhor qualidade e maior resolução, permite ao usuário a configuração manual dos seus principais atributos, como abertura do diafragma, velocidade do obturador, ISO (sensibilidade do sensor), entre outros. Além da câmera, é necessária a utilização de um tripé onde a mesma deverá ser fixada. Na Seção 2.3 apresentaremos todas as teorias e definições básicas sobre fotografia.

Para o processo de *Depth Hallucination*, as superfícies capturadas devem ser planas, como por exemplo uma parede ou um piso. Se for verificada a presença de especularidades na superfície de interesse, um polarizador pode ser utilizado para minimizar tais efeitos [9]. Todas as imagens devem ser capturadas no formato *RAW*, o qual não passou por nenhum processo de compressão com perda de dados, ou seja, mantém todas as informações da imagem tal como captada pelo sensor da câmera.

A primeira imagem capturada deve estar sobre uma iluminação indireta (sombra ou dia nublado). Esta imagem é chamada de *diffuse-lit* (Figura 2.3). A segunda foto, chamada *flash-lit* (Figura 2.4), deve ser capturada do mesmo ponto da primeira imagem e utilizando o *flash* com uma potência suficiente para revelar os locais em sombra na superfície. Além disso, deve possuir a mesma configuração de abertura de diafragma e foco utilizados na captura da imagem difusa. As únicas características que podem ser alteradas são a velocidade do obturador, para compensar a superexposição causada pelo uso do *flash* na captura da imagem, e o ISO.

O *flash* utilizado para a captura do *flash-lit* pode ser tanto embutido na própria câmera, quanto um *flash* externo. O ideal é que o mesmo fique o mais próximo possível do corpo da câmera a fim de minimizar as possíveis sombras criadas.



Figura 2.3: Exemplo de *diffuse-lit*.



Figura 2.4: Exemplo de *flash-lit*.

2.2.2 Calculando o Albedo e o *Shading Image*

Após a aquisição das imagens, torna-se necessária a calibração das mesmas utilizando-se os valores de abertura do diafragma (A), ISO (I) e velocidade do obturador (T_s) selecionados para a captura das imagens difusa (*diffuse-lit*) e com *flash* (*flash-lit*). Dessa forma, obtemos um fator de correção da exposição de cada imagem baseado na Equação (2.1) :

$$C_e = \frac{A^2}{T_s I}. \quad (2.1)$$

A imagem albedo I_a (Figura 2.5) é obtida a partir da Equação (2.2), sendo esta operação realizada em cada *pixel* j da imagem.

$$I_a(j) = \frac{I_f(j) - I_d(j)}{I_c(j)}. \quad (2.2)$$



Figura 2.5: Exemplo de albedo.

O valor de cada *pixel* $I_d(j)$ da imagem difusa é subtraído dos respectivos *pixels* $I_f(j)$ da imagem com *flash* e o resultado é dividido pelos valores dos *pixels* da imagem de calibração do *flash* $I_c(j)$. Esta é obtida de uma superfície Lambertiana branca (Figura 2.6) utilizando-se as mesmas configurações de abertura do diafragma, ISO e velocidade do obturador utilizadas na captura da imagem com *flash*, além de estar posicionada a uma mesma distância da superfície. Uma superfície Lambertiana é aquela que dispersa a luz igualmente em todas as direções. Desta forma, é possível corrigir os efeitos de *vignetting*, *fall-off* e o fator cosseno global [10] presentes nas imagens difusa e com *flash*.

O *vignetting* é a redução da luminância nas extremidades inferiores e superiores da imagem causada pela lente fotográfica utilizada [11]. Já o efeito *fall-off* é definido como a queda da intensidade da luz com o inverso do quadrado da distância para a superfície de interesse. Por fim, o decaimento da intensidade de luz com o cosseno do ângulo de incidência define o fator cosseno global (Seção 2.2.3) [10].



Figura 2.6: Exemplo de superfície Lambertiana branca.

O *shading image* (Figura 2.7) é obtido através da razão entre a imagem difusa e do albedo anteriormente calculado (Equação (2.2)). A operação deve ser realizada apenas no canal de luminância, uma vez que o processo de *Depth Hallucination* utiliza apenas esse canal para a estimação do mapa de profundidades. Além disso, para essa estimação, o processo assume um valor de altura 0 para *pixels* de intensidade 0.5. Desta forma, o *shading image* deve ser normalizado nesse valor médio.

Entretanto, os resultados obtidos após as operações descritas anteriormente resultam em um *shading image* extremamente claro, de tal forma que não conseguimos discernir os detalhes característicos da imagem. Dessa forma, baseado em testes empíricos, concluímos que antes da realização da normalização em 0.5, o *shading image* deve ser dividido por um fator fixo. Os melhores resultados foram encontrados através da utilização de um valor igual a 20.

2.2.3 Estimação do Mapa de Profundidades

A técnica de *Depth Hallucination* baseia-se no desenvolvimento de dois modelos para a reconstrução da meso-estrutura de uma determinada superfície: os modelos de picos e os de depressões. Para o último, o fator cosseno, referente à orientação da superfície em relação ao céu, prevalece. Já para o primeiro, predomina o efeito *visible*

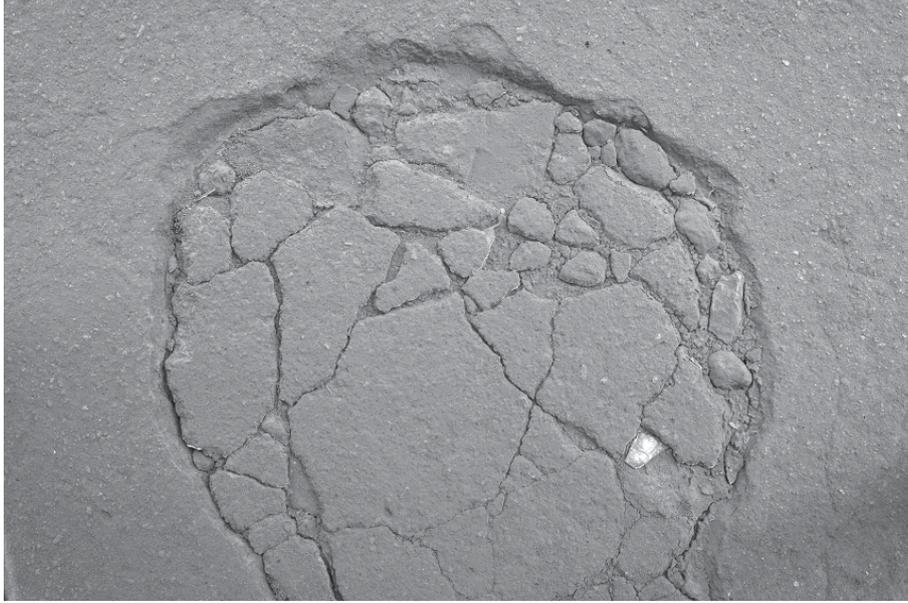


Figura 2.7: Exemplo de *shading image*.

aperture onde pelo menos parte da vizinhança do ponto de interesse encontra-se em sombra. A Figura 2.8 exemplifica onde cada modelo deve ser utilizado.

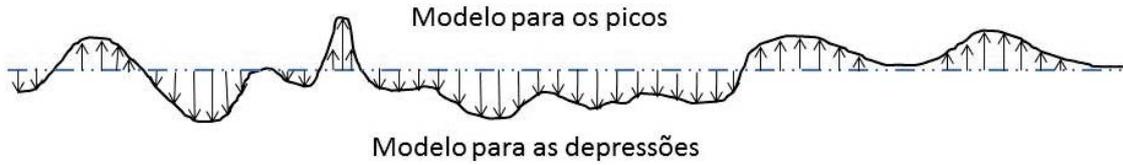


Figura 2.8: Exemplificação da utilização do modelo para picos e depressões localizados respectivamente acima e abaixo do plano que divide a superfície (baseado em [1]).

Considere a porção do céu visível através de um determinado ponto P e a incidência de um raio de luz através de um ângulo sólido dw_{Θ} , conforme exibido pela Figura 2.9:

Realizando a parametrização do ângulo sólido para (θ, ϕ) , obtemos:

$$dw_{\Theta} = \sin \theta d\theta d\phi. \quad (2.3)$$

Dessa forma, a integral sobre o hemisfério será:

$$\int_{\Omega} f(\Theta) dw_{\Theta} = \int_0^{2\pi} \int_0^{\theta} f(\theta, \phi) \sin \theta d\theta d\phi. \quad (2.4)$$

Para estimarmos a profundidade de um ponto localizado na cavidade da superfície de interesse, inicialmente calculamos o fator de iluminação E_c a partir da Equação (2.4), onde $f(\theta, \phi) = \cos \theta$, uma vez que a iluminação sobre um ponto na

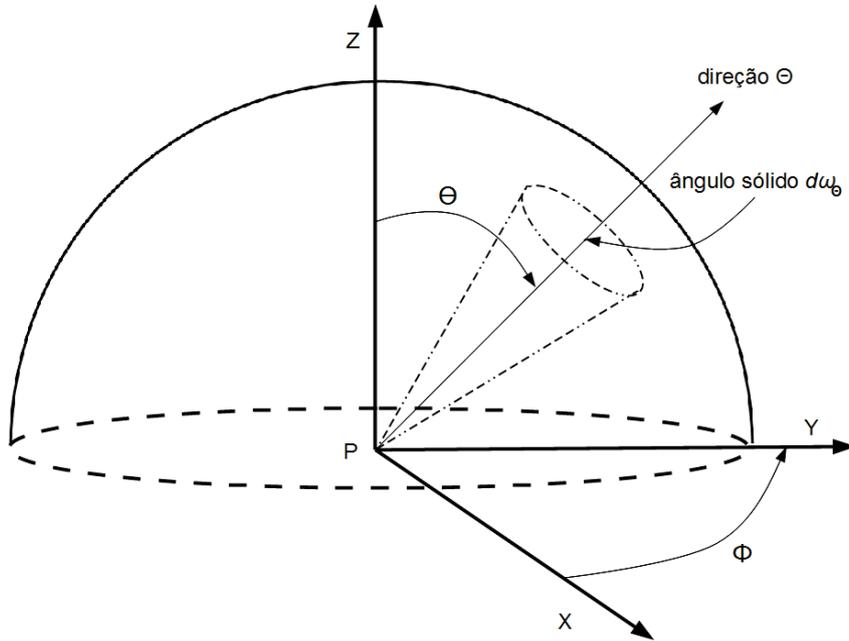


Figura 2.9: Porção do céu visível através do ponto P e incidência de luz por um ângulo sólido.

superfície decai com o cosseno do ângulo de incidência (Figura 2.10).

$$E_c = \int_0^{2\pi} \int_0^\theta \cos \theta \sin \theta d\theta d\phi. \quad (2.5)$$

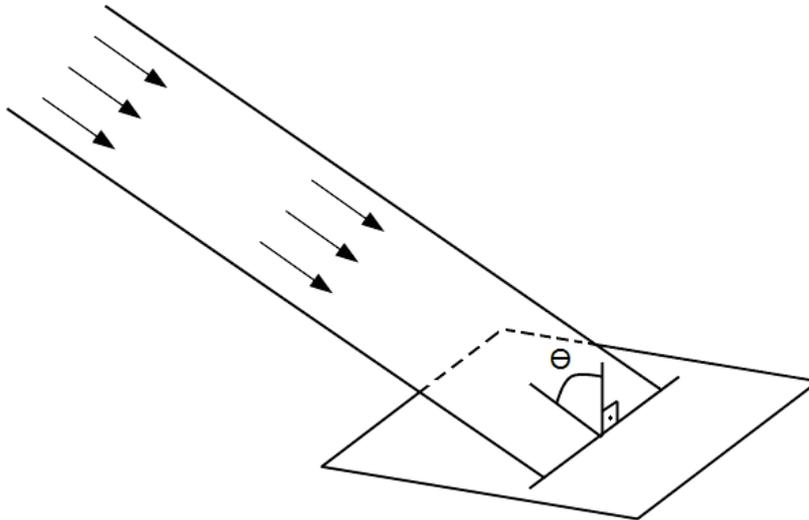


Figura 2.10: Variação da intensidade da luz conforme o cosseno do ângulo de incidência.

O modelo para depressões aproxima as cavidades da superfície como cilindros de

raio a e de altura d , como mostra a Figura 2.11. Dessa forma:

$$E_c = \int_0^{2\pi} \int_0^\theta \cos \theta \sin \theta d\theta d\phi = \pi \sin^2 \theta = \pi \frac{a^2}{a^2 + d^2}. \quad (2.6)$$

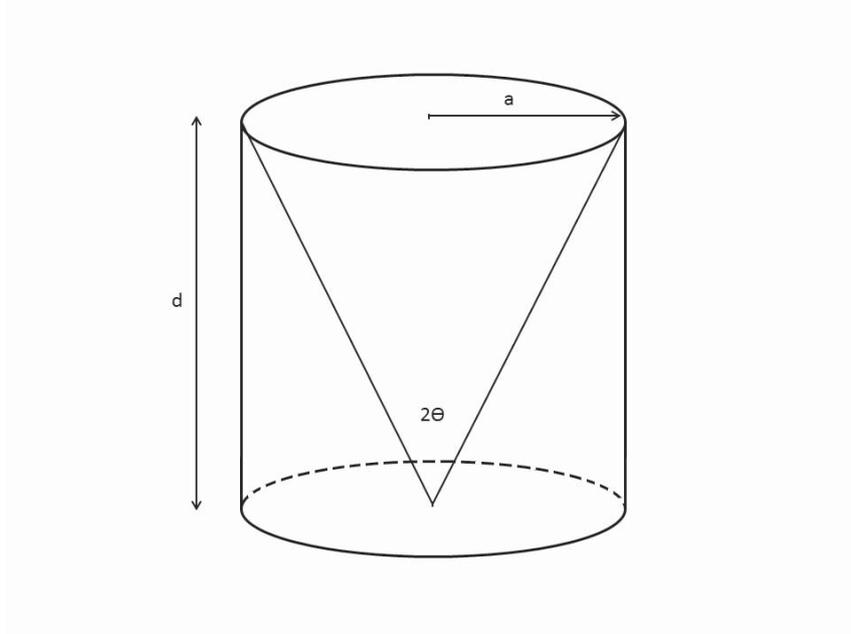


Figura 2.11: Modelo cilindro para os pontos localizados abaixo do plano médio da superfície (baseado em [1]).

Para obtermos o fator *shading* S é necessário dividir E_c pelo fator de iluminação relativo ao céu, E_h , o qual é igual a π conforme a Equação (2.7):

$$E_h = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \cos \theta \sin \theta d\theta d\phi = \pi. \quad (2.7)$$

Dessa forma:

$$S = \frac{E_c}{E_h} = \frac{\pi \sin^2 \theta}{\pi} = \frac{a^2}{a^2 + d^2}. \quad (2.8)$$

Assim, resolvendo a Equação (2.8) para d , obtemos a profundidade do ponto localizado na cavidade da superfície:

$$d = a \sqrt{\frac{1}{S} - 1}. \quad (2.9)$$

Já para os pontos localizados acima do plano médio da superfície (modelo para os picos), o fator *shading* S é uma função da porção visível do hemisfério h_v sob o ângulo Φ (Equação (2.10)), somada a porção refletida h_r do mesmo hemisfério localizada fora do ângulo Φ (Equação (2.11)).

$$h_v = \frac{\pi}{2}(1 + \cos \Phi), \quad (2.10)$$

$$h_r = \nu \frac{\pi}{2}(1 - \cos \Phi). \quad (2.11)$$

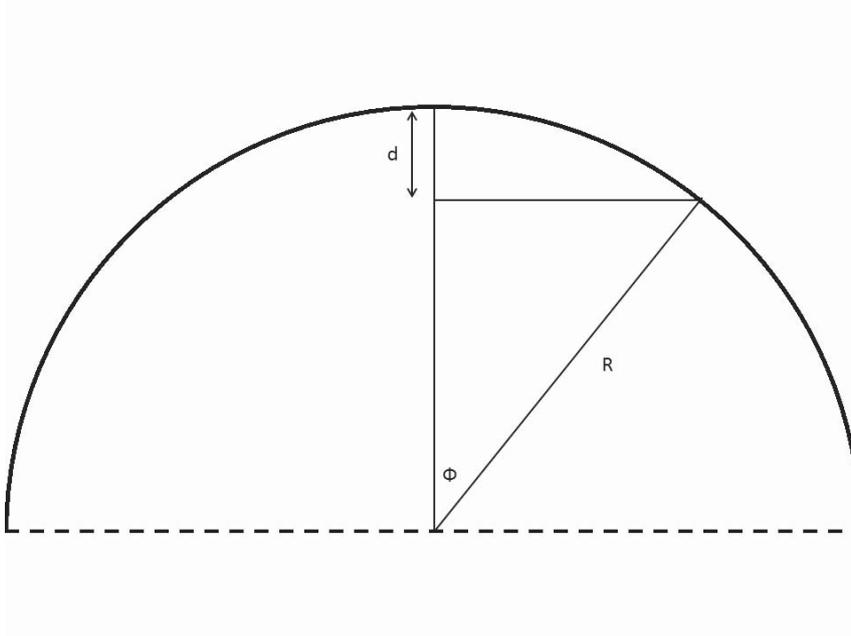


Figura 2.12: Modelo hemisfério para os pontos localizados acima do plano médio da superfície (baseado em [1]).

Da mesma forma que na Equação (2.8), dividimos a soma entre h_v e h_r pelo fator de iluminação relativo ao céu, π . Assim:

$$S = \frac{\frac{\pi}{2}(1 + \cos \Phi) + \nu \frac{\pi}{2}(1 - \cos \Phi)}{\pi}, \quad (2.12)$$

onde ν é o fator de reflexão efetivo da superfície em torno do ponto de interesse. Resolvendo a Equação (2.12) para $\cos \Phi$, obtemos:

$$\cos \Phi = \frac{2S - (1 + \nu)}{1 - \nu}. \quad (2.13)$$

Observando a Figura 2.12, temos que:

$$d = R - R \cos \Phi. \quad (2.14)$$

Substituindo a Equação (2.13) na Equação (2.14) acima, concluímos que:

$$d = 2R \frac{1 - S}{1 - \nu}, \quad (2.15)$$

onde R é o raio do pico do hemisfério.

Conforme descrito na Seção 2.2.2, o *Shading Image* obtido foi normalizado em 0.5. Dessa forma, substituindo $S = \frac{1}{2}$ e combinando as Equações (2.9) e (2.15), obtemos:

$$a = \frac{R}{(1 - \nu)}. \quad (2.16)$$

Tanto para o modelo acima do plano médio como para o abaixo do plano médio, a técnica de *Depth Hallucination* assume o fator de reflexão efetivo como 0 (zero), ou seja, $\nu = 0$. Assim, para cada ponto (*pixel*) existente no *Shading Image* com um valor S , chegamos ao seguinte modelo:

$$D(S) = \frac{d}{a} = \begin{cases} \sqrt{\frac{1}{S} - 1} & \text{se } S \leq \frac{1}{2}, \\ 2(1 - S) & \text{se } S > \frac{1}{2}. \end{cases} \quad (2.17)$$

Dada uma grande região da superfície de interesse, uma forte alteração de *shading* na mesma geralmente corresponde a uma maior diferença de profundidade se compararmos a mesma alteração em uma pequena região da superfície. Dessa forma, uma vez que o modelo de estimação da profundidade dos pontos é baseado no fator a , deve-se considerar diferentes escalas de a no *shading image*. Assim, são geradas algumas imagens com *blur* gaussiano a partir do *shading image* anteriormente normalizado (Figura 2.13). O número de imagens (níveis) de *blur* que serão gerados é escolhido pelo usuário do sistema. Quanto maior esse número, mais preciso e suave será o mapa de profundidades gerado. Porém, maior será o tempo de processamento. No Capítulo 4 apresentaremos os diferentes resultados obtidos para o mapa de profundidades a partir da escolha do número de níveis de *blur* pelo usuário.

Outra importante informação nesta etapa do processo diz respeito à escolha do valor de *sigma* para a filtragem gaussiana. Apesar de não ser informado maiores detalhes na documentação do *Depth Hallucination* [1], realizamos testes empíricos,

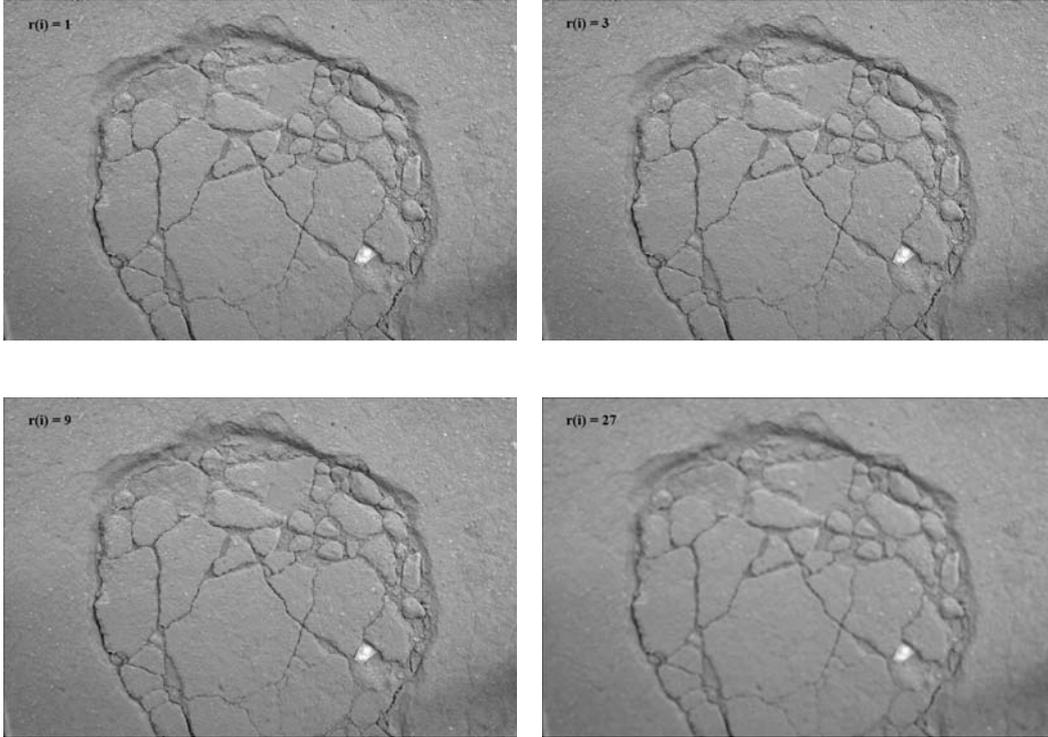


Figura 2.13: Exemplificação do blur gaussiano em 4 (quatro) níveis.

os quais também são exibidos na Seção 4.1 e constatamos que para valores acima de 100 não notamos variações perceptíveis no mapa de profundidades. Desta forma, assumiremos esse valor ao longo de todo o processo.

O raio da filtragem Gaussiana utilizado é um kernel radii, o qual, de um nível para o outro, aumenta em potências de três. Para cada nível, a imagem gerada é dividida pela imagem do próximo nível superior e, em seguida, normalizada em 0.5. Dessa forma, obtemos uma pirâmide Laplaciana de imagens de mesma resolução [12].

Utilizando a Equação (2.17), onde a é o raio do blur gaussiano utilizado em cada nível $l(i)$ gerado, o mapa de profundidades da superfície de interesse é obtido a partir do acúmulo das profundidades para cada *pixel* j em cada nível $l(i)$, conforme a Equação (2.18):

$$d_j = \sum_{i=1}^N r(i)[D(l_j(i)) - 1]. \quad (2.18)$$

Uma vez que utilizamos a normalização por 0.5 na Equação (2.17), se desejarmos que os valores médios do mapa de profundidade fiquem em torno de zero, a subtração por um faz-se necessária.

Uma vez que o processo de *Depth Hallucination* utiliza algumas simplificações, como por exemplo, ignorar o cálculo do albedo nas regiões acentuadamente inclina-

das da superfície de interesse, bem como a aproximação das reentrâncias da superfícies como depressões (covas) e não como fendas (mais apropriado), é permitido ao usuário aplicar um fator de escala uniforme no mapa de profundidades compensando as aproximações anteriormente descritas, além de permitir uma representação visual mais próxima do real. Essa escala fica entre 0.75 e 1.5.

2.2.4 Reiluminação da Superfície

Uma vez que a imagem albedo foi criada e o mapa de profundidades obtido, podemos reiluminar a superfície virtual e utilizá-la em aplicações voltadas tanto para arquitetura quanto para *games* e efeitos visuais (foto realismo).

2.3 Teoria Fotográfica

No processo de *Depth Hallucination*, a criação de um mapa de profundidades correto e preciso está totalmente atrelada a uma correta aquisição das imagens da superfície de interesse. Dessa forma, torna-se necessária uma explanação sobre os principais conceitos fotográficos básicos para uma correta captura das imagens.

2.3.1 Câmera

Para a captura das imagens podemos utilizar câmeras digitais compactas ou as DSLR (*Digital Single Lens Reflex*). Se for utilizada uma câmera compacta a mesma deverá permitir ao usuário a utilização de configurações manuais da velocidade do obturador, abertura do diafragma e ISO. Além disso, deve permitir uma captura da imagem em formato *raw* (Seção 2.3.2).

Idealmente, deve-se utilizar uma câmera DSLR. Este tipo de câmera tem como uma de suas principais características permitir que o usuário enxergue exatamente a mesma imagem que está entrando pela lente da câmera. Isso ocorre graças a um espelho e ao pentaprisma existente na mesma (Figura 2.14).

Inicialmente a luz entra pela objetiva (lente ótica), a qual na verdade é formada por um conjunto complexo de lentes de variados formatos e dimensões. Durante esse percurso, a luz passa pelo diafragma, o qual determina a quantidade de luz que atingirá o sensor da câmera a partir da regulagem da abertura do mesmo (Seção 2.3.3). Após passar pelas diversas lentes, a luz atinge um espelho localizado à frente do sensor e posicionado exatamente a 45° , o qual a reflete para cima. Em seguida, a luz passa por uma tela de foco e por outra lente antes de atingir o pentaprisma. Este prisma, além de inverter a imagem, a reflete em exatos 90° , permitindo ao usuário visualizá-la através de um visor analógico, conforme a objetiva está "enxergando".

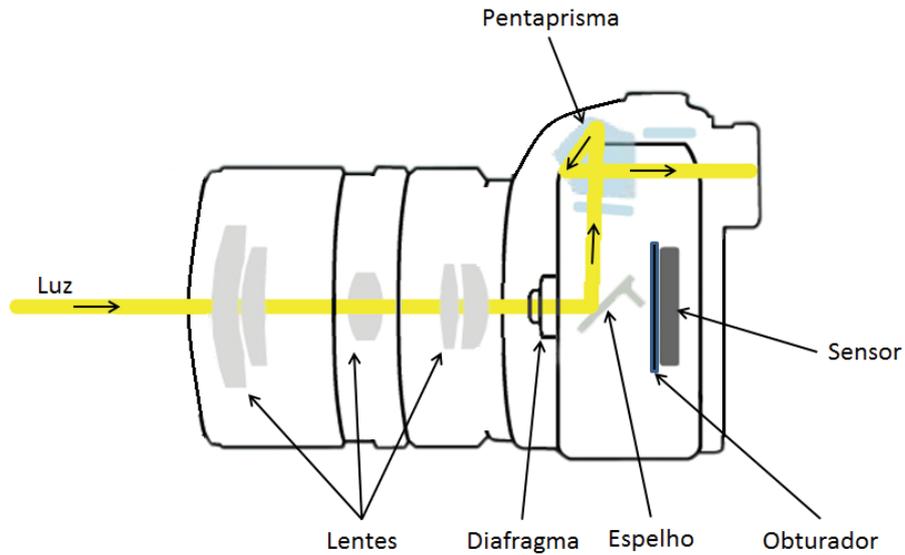


Figura 2.14: Representação da incidência da luz dentro de uma DSLR.

Quando pressionamos o botão para a captura da imagem (Figura 2.15), o espelho que reflete a luz é erguido, fazendo com que a mesma, que já passou pela abertura do diafragma, atinja o obturador da câmera. Este é responsável por controlar o tempo que o sensor ficará exposto à luz (Seção 2.3.4). Após esse tempo, o obturador se fecha novamente. Em seguida, o sensor envia as informações captadas para o processador, o qual inverte a imagem e a encaminha para visor eletrônico, permitindo que o usuário a visualize.

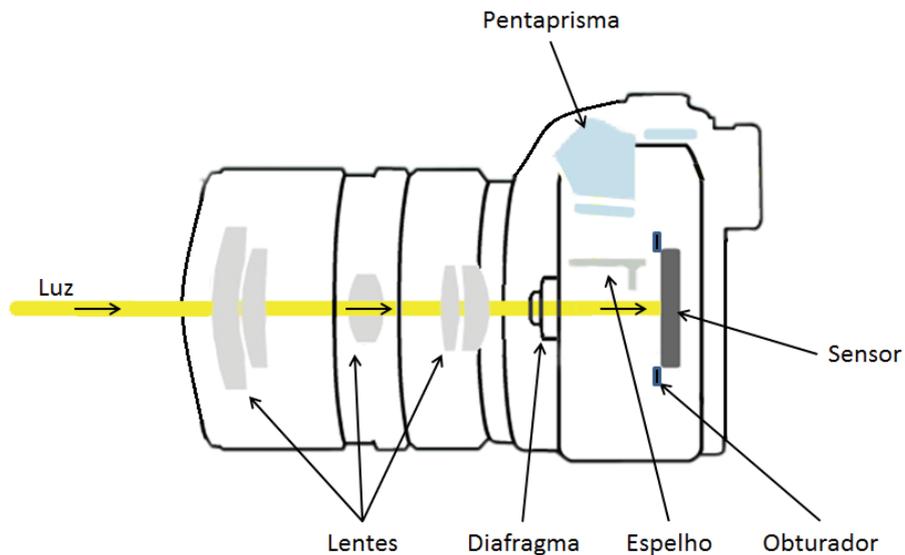


Figura 2.15: Representação da incidência da luz ao capturarmos uma imagem.

2.3.2 Arquivo *Raw* ("Cru")

Dependendo da câmera digital utilizada, a mesma permite a captura das imagens de interesse em diferentes formatos, sendo o *jpeg* o mais comum entre estes. Além deste, as câmeras semi-profissionais e profissionais permitem a captura e o registro de imagens no formato *raw*, o qual caracteriza-se por armazenar a imagem capturada pelo sensor da câmera sem nenhum tipo de compressão ou pós-processamento. O arquivo *raw* possui a mesma função do filme negativo da fotografia analógica, o qual não pode ser utilizado como a imagem final, mas contém todas as informações para que possamos revelá-la.

Ao fotografarmos utilizando o formato *jpeg*, por exemplo, a própria câmera digital processa a imagem capturada utilizando automaticamente, ou através de uma configuração prévia do usuário, valores pré-definidos de contraste, saturação e balanço de branco (Seção 2.3.7). Dessa forma, toda a informação capturada pelo sensor da câmera sofre um processamento após a captura.

Na captura em formato *raw*, a câmera também utiliza uma configuração automática ou manual para o contraste, a saturação e o *WB*. Esses valores são registrados no arquivo, porém não são aplicados em definitivo na imagem resultante. Através de *software* específico, esses valores pré-definidos poderão ser alterados, uma vez que o arquivo possui todas as informações da imagem conforme captura do sensor. Além disso, nos arquivos *raw* existe muito mais informação por *pixel*, dado que, dependendo da câmera utilizada, poderemos ter de 12 a 14 bits/pixel. Assim, possuímos muito mais informação de espaço de cor na imagem capturada, diferentemente da compressão estabelecida pelo formato *jpeg*.

No arquivo "cru", além das informações anteriormente citadas, diversas outras informações sobre as características da máquina fotográfica e objetiva, assim como as configurações da câmera no momento da captura da imagem, são registradas no mesmo (metadados). Dessa forma, utilizando *software* que interprete esses dados, conseguiremos acessar todas essas características e, a partir das mesmas, realizar pós-processamento. Por exemplo, uma vez que possuímos as informações da objetiva utilizada, é possível atenuar as distorções de lente produzidas pela mesma.

Por não sofrer qualquer tipo de compressão ou pós-processamento, os arquivos *raw* são maiores que o *jpeg*. Além disso, não há um padrão estabelecido para o formato cru. Em geral, cada fabricante de câmera possui seu formato proprietário. Por exemplo, *.cr2* para máquinas fotográficas do fabricante Canon e *.nef* para Nikon. Por fim, para leitura desses arquivos, faz-se necessária a utilização de *software* específico, o qual permite também a manipulação das características das imagens capturadas, além da conversão para outros formatos. O mais popular e completo entre esses programas é o *Adobe Photoshop*, o qual possui o *plugin* gratuito *Adobe*

Camera Raw que foi utilizado neste trabalho (Seção (3.2)).

2.3.3 Diafragma

O diafragma é responsável por controlar a quantidade de luz que atinge o sensor da câmera a partir do diâmetro da abertura do mesmo. É formado por várias lâminas sobrepostas e fica geralmente no centro ótico da lente. Nas câmeras fotográficas, a abertura do diafragma ou f -stop é representada por uma série de números chamados números f , os quais são obtidos a partir da Equação (2.19):

$$f = \frac{f'}{D}, \quad (2.19)$$

onde f' é a distância focal da lente e D é o diâmetro específico do diafragma. Dado que o diâmetro do diafragma é inversamente proporcional ao número f , temos que quanto maior for o valor escolhido para f , menor será a abertura do diafragma, como exemplifica a Figura 2.16.

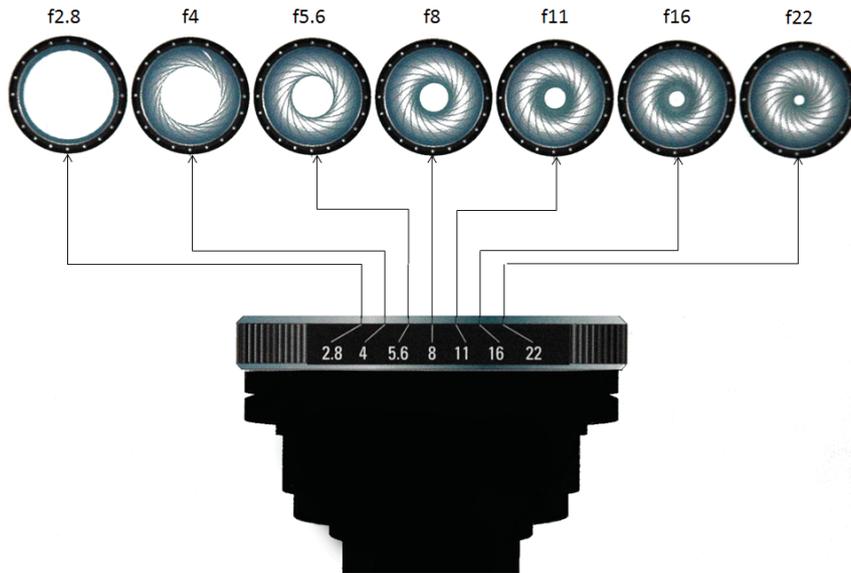


Figura 2.16: Quanto maior o número de f -stops, menor a abertura do diafragma.

Considere que desejamos calcular o número f necessário para dobrar a quantidade de luz recebida pelo sensor para uma mesma distância focal. Considerando a abertura do diafragma como um círculo perfeito, temos que:

$$A_1 = 2A_2 \quad (2.20)$$

onde A é a área da circunferência. Assim, temos que a relação entre os diâmetros será:

$$D_1 = \sqrt{2}D_2. \quad (2.21)$$

Conseqüentemente,

$$f_2 = \sqrt{2}f_1 \approx 1.4f_1. \quad (2.22)$$

É devido ao fator de multiplicação por 1.4 que as câmeras fotográficas geralmente apresentam a escala para o número f da seguinte forma: 1.4, 2, 2.8, 4, 5.6, 8, 11, 16 e assim por diante.

Portanto, ao mudarmos o diafragma de $f5.6$ para $f4.0$, por exemplo, estamos dobrando a quantidade de luz que o sensor captará. Dizemos que houve um aumento de 1.0 $f-stop$. Na situação contrária, mudando de $f4.0$ para $f5.6$, reduz-se pela metade a quantidade de luz. Assim, houve uma redução de 1.0 $f-stop$ (Figura 2.17). Ou seja, a alteração de 1.0 $f-stop$ para mais significa o dobro de luz atingindo o sensor (maior exposição), enquanto que 1.0 $f-stop$ para menos significa que metade da luz está entrando pela lente (menor exposição).

Nas câmeras digitais é muito comum encontramos valores para a abertura do diafragma divididos em $\frac{1}{2}$ e $\frac{1}{3}$ de $f-stop$. Assim, entre $f4$ e 5.6 podemos ter $f4.8$ para $\frac{1}{2}$ $f-stop$ e mais dois valores, $f4.5$ e $f5.0$, para $\frac{1}{3}$ de $f-stop$, por exemplo.

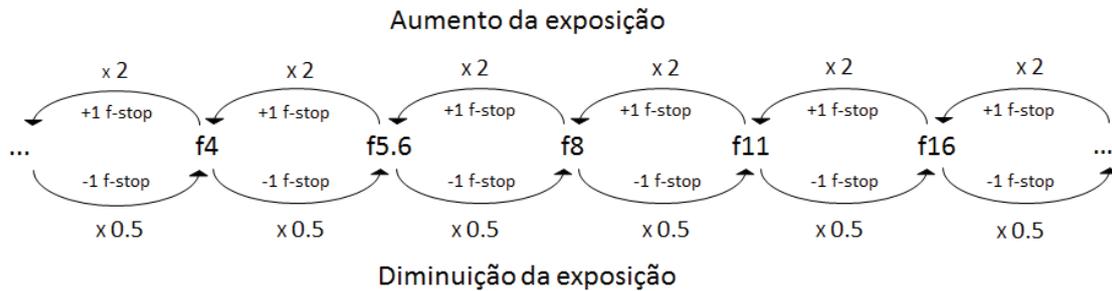


Figura 2.17: Aumento e diminuição da quantidade de luz com a variação da abertura do diafragma.

Ao focarmos em um determinado objeto presente em uma cena, estamos colocando sobre o mesmo o plano focal. Assim, ao capturarmos a imagem, o elemento estará completamente nítido e os demais elementos localizados a uma certa distância estarão totalmente desfocados. Entretanto, algumas áreas próximas ao plano focal ficam tão nítidas quanto o objeto de interesse. De um modo geral, $\frac{1}{3}$ dessa área fica localizada à frente do plano focal e o restante, $\frac{2}{3}$, ficam atrás do plano. A essa região que apresenta os elementos focados (nítidos) e que inclui o plano de foco dá-se o nome de profundidade de campo (Figura 2.18).

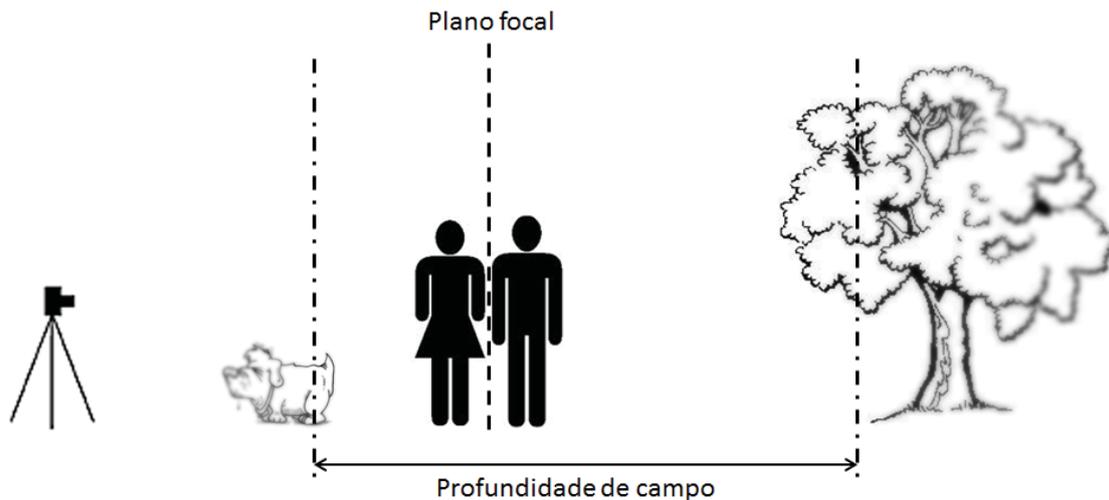


Figura 2.18: Exemplificação da região próxima ao plano focal que permanecerá nítida na imagem capturada.

A abertura do diafragma está diretamente relacionada à profundidade de campo. Quanto menor for o número f (maior abertura do diafragma), menor será a profundidade de campo e apenas uma pequena região próxima ao plano focal permanecerá nítida. Por outro lado, quanto maior for o número f (menor abertura do diafragma) maior será a profundidade de campo e, conseqüentemente, maior será o número de elementos nítidos na imagem.

Nas figuras a seguir, o plano focal selecionado encontra-se sobre o objeto localizado à frente da câmera. Na Figura 2.19, selecionamos uma abertura do diafragma pequena, neste caso, $f3.5$. Dessa forma, teremos uma profundidade de campo pequena e apenas o primeiro objeto permanecerá nítido. Entretanto, alterando-se a abertura para um valor elevado, neste caso $f22$, podemos perceber na Figura 2.20 que, apesar de o plano focal continuar localizado no elemento mais à frente da câmera, os demais elementos também estarão nítidos, pois a profundidade de campo é maior.



Figura 2.19: Abertura do diafragma em $f3.5$. Apenas o primeiro elemento focado.



Figura 2.20: Abertura do diafragma em $f22$. Toda a cena focada.

2.3.4 Velocidade do Obturador

O ajuste da velocidade do obturador permite controlar a duração ou intervalo de tempo em que o sensor fica exposto à luz. A sua localização pode variar de uma câmera para outra, porém a finalidade é sempre a mesma. Nas câmeras mais antigas, o obturador possui um formato circular, muito parecido com o do diafragma. Nas câmeras mais modernas, placas sobrepostas verticalmente constituem o obturador que, durante a captura da imagem, é retraído como uma cortina.

Normalmente a velocidade do obturador é fornecida como fração de 1.0 segundo ($\frac{1}{x}$). Assim, quando maior for x , maior será a velocidade do obturador e, consequentemente, menor será o tempo de exposição do sensor à luz. Por outro lado, quanto menor for x , menor será a velocidade do obturador, ocasionando um maior tempo de exposição.

Da mesma forma que a abertura do diafragma, a diferença entre dois valores de velocidade é também de dobro-metade (Figura 2.21). Por exemplo, a velocidade de $\frac{1}{500}$ deixa o sensor duas vezes mais exposto que $\frac{1}{1000}$ (aumento de 1.0 *f-stop*). Além disso, é comum encontrarmos nas câmeras digitais velocidades divididos em $\frac{1}{2}$ e $\frac{1}{3}$ de *f-stop*. Então, entre as velocidades de $\frac{1}{500}$ e $\frac{1}{1000}$, podemos ter $\frac{1}{750}$ para $\frac{1}{2}$ *f-stop* e mais duas velocidades, $\frac{1}{650}$ e $\frac{1}{800}$, para $\frac{1}{3}$ *f-stop*, por exemplo.

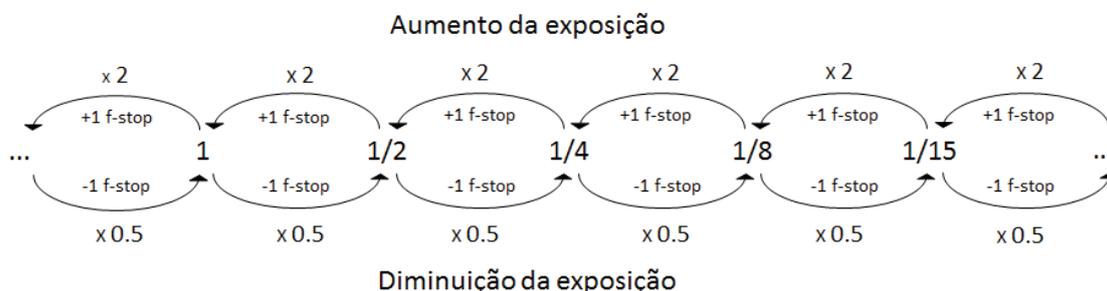


Figura 2.21: Aumento e diminuição da quantidade de luz com a variação da velocidade do obturador.

O ajuste da velocidade do obturador permite determinar se na imagem capturada desejamos "congelar" a ação de um ponto de interesse que está em movimento na cena. Dessa forma, quanto menor for a velocidade do obturador, mais borrado estará o objeto de interesse na imagem capturada (Figura 2.22a). Caso contrário, quanto maior for a velocidade do obturador, mais os objetos em movimento na cena aparecerão estáticos na imagem capturada (Figuras 2.22b e 2.22c).



(a) Velocidade do obturador: 1/60.



(b) Velocidade do obturador: 1/500.



(c) Velocidade do obturador: 1/2000.

Figura 2.22: Exemplificação de fotografias capturadas com diferentes velocidades do obturador.

2.3.5 ISO

Na fotografia analógica, o ISO (ou ASA) refere-se à sensibilidade à luz do filme que está sendo utilizado. Na fotografia digital, refere-se à sensibilidade do sensor à luz. Quanto maior for o número ISO escolhido, maior será a sensibilidade do sensor à luz e vice-versa.

Para baixos valores de ISO, precisaremos de uma maior abertura do diafragma para entrada de luz e/ou mais tempo de exposição para iluminarmos o sensor. Valores de 50 a 200 são utilizados em situações com muita luz ambiente, como um dia de sol, por exemplo. Para os menores valores de ISO, a imagem capturada apresentará os melhores resultados tanto na qualidade da imagem quanto na nitidez da mesma

(baixo ruído). Altos valores do ISO, acima de 800, são escolhidos para situações com pouquíssima luz presente no ambiente. Porém, dependendo da câmera que está sendo utilizada, um ISO muito elevado proporciona uma captura da imagem bastante granulada, pois as amostras capturadas pelo sensor são muito espaçadas umas das outras, havendo a captura de muito ruído. Assim, quanto maior o ISO, maior será o ruído na imagem capturada, conforme demonstra o exemplo da Figura 2.23.

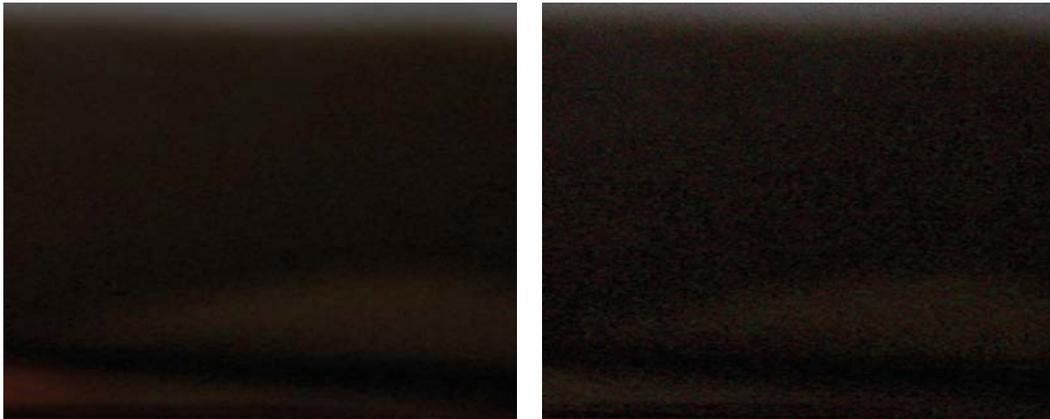


Figura 2.23: ISO 100 e ISO 1600 respectivamente.

Igualmente à abertura do diafragma e à velocidade do obturador existe a relação de dobro-metade entre os valores de ISO (Figura 2.24). Por exemplo, o ISO 400 é duas vezes mais exposto que o 200.



Figura 2.24: Aumento e diminuição da quantidade de luz com a variação do ISO.

Principalmente nas câmeras DSLR mais modernas, é possível encontrarmos valores de $\frac{1}{3}$ de *f-stops*. Portanto, entre os valores 200 e 400 para o ISO, podemos encontrar 250 e 320.

2.3.6 Fotômetro

Fotômetros são aparelhos que medem a quantidade de luz recebida ou refletida pela cena observada. Nas câmeras digitais, a luz refletida pelos objetos que estão presentes na cena atinge uma célula fotoelétrica, a qual registra a exposição da luz na cena e a exibe para o usuário através de uma escala numérica. Esta, dependendo da câmera que está sendo utilizada, apresenta valores negativos e positivos, geralmente de -2 até +2, como mostra o exemplo da Figura 2.25. A indicação da exposição da cena depende das configurações selecionadas para a abertura do diafragma, velocidade do obturador e ISO.



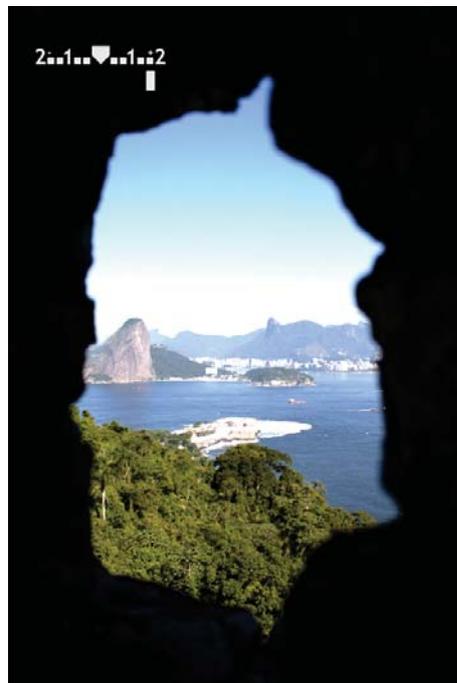
Figura 2.25: Fotômetro (*display*).

Se a imagem capturada apresenta uma exposição baixa, onde os pontos de interesse não estão bem iluminados, temos uma cena subexposta. O fotômetro indica que uma cena está subexposta ao posicionar-se nos valores abaixo de zero na sua escala (Figura 2.26a).

Quando a imagem capturada apresenta muita luz ambiente, deixando os objetos em cena excessivamente iluminados, temos uma cena superexposta, a qual é indicada por valores positivos na escala fotométrica (Figura 2.26b). Já quando o indicador do fotômetro está na posição zero, a imagem capturada apresentará a cena de interesse com uma exposição equilibrada, onde poderemos observar detalhes tanto nas regiões mais claras quanto nas mais escuras (Figura 2.26c). Entretanto, muitas vezes as informações passadas pelo fotômetro precisam ser reinterpretadas.



(a) Exemplo de foto subexposta.



(b) Exemplo de foto superexposta.



(c) Exemplo de foto balanceada.

Figura 2.26: Exemplo de fotografias capturadas com diferentes exposições.

Imaginemos que desejamos fotografar o objeto da Figura 2.27.



Figura 2.27: Objeto de interesse.

Posicionamos o mesmo à frente de um fundo neutro claro qualquer. Após a escolha de uma configuração de abertura do diafragma e velocidade do obturador, obtemos uma exposição balanceada. Entretanto, ao analisarmos a imagem capturada (Figura 2.28a), constatamos que a mesma apresenta-se subexposta.

Como sabemos, o fotômetro informa a quantidade de luz refletida por toda a cena. Uma vez que o fundo escolhido reflete uma quantidade de luz muito maior que a proveniente do objeto, o fotômetro acaba indicando que a cena como um todo está superexposta. Ao ajustarmos as configurações da câmera para compensarmos essa superexposição, diminuiremos o excesso de luz em toda a cena, melhorando a exposição do plano de fundo, porém tornando o objeto de interesse subexposto.

Se reajustarmos as configurações de forma a obtermos uma pequena superexposição na indicação do fotômetro, apesar de estarmos permitindo que uma quantidade maior de luz seja capturada pelo sensor, teremos obtido uma imagem com uma exposição mais balanceada, conforme podemos observar na Figura 2.28b.



(a) Apesar do fotômetro indicar uma exposição balanceada, obtemos uma imagem subexposta.



(b) Apesar do fotômetro indicar uma superexposição, obtemos uma imagem balanceada.

Figura 2.28: Imagens capturadas de objetos localizados à frente de um fundo branco.

Se posicionarmos o objeto à frente de um fundo escuro qualquer, o resultado obtido será exatamente o oposto. Ao configurarmos uma exposição balanceada para a cena, a imagem capturada apresentada será superexposta (Figura 2.29a), pois compensamos a pouca quantidade de luz que atinge o fotômetro com um maior tempo de exposição e/ou uma maior abertura do diafragma. Neste caso, uma imagem balanceada é obtida através de um ajuste subexposto da câmera, conforme exibido na Figura 2.29b.

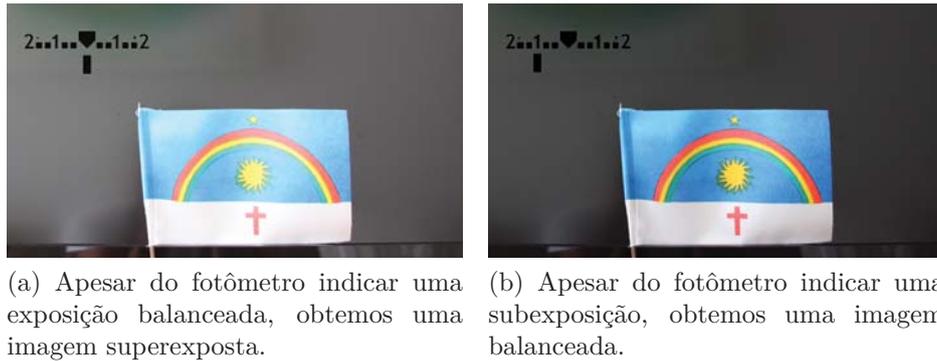


Figura 2.29: Imagens capturadas de objetos localizados à frente de um fundo escuro.

Uma outra forma de analisarmos a exposição das imagens capturadas é através da avaliação dos histogramas das mesmas. A grande maioria das câmeras digitais permite ao usuário a visualização do histograma das imagens capturadas. Algumas câmeras mais modernas e profissionais já permitem a visualização do mesmo em tempo real, não sendo necessária a realização de capturas para avaliação da exposição.

As Figuras 2.30a a 2.30c exibem os respectivos histogramas das Figuras 2.26a, 2.26c e 2.26b, respectivamente.

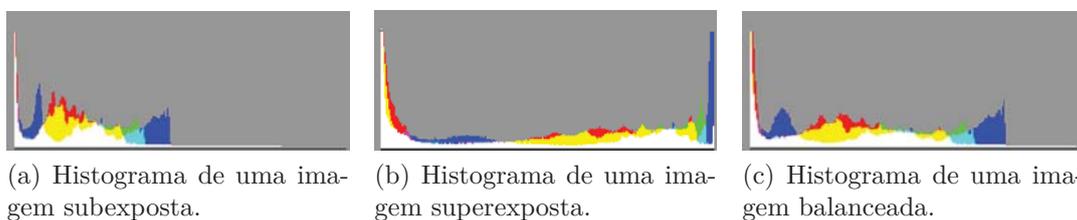


Figura 2.30: Exemplificação dos histogramas de fotografias capturadas com diferentes exposições.

Podemos observar que a Figura 2.30a apresenta uma quantidade maior de informações nos valores mais baixos de luminância, caracterizando uma subexposição. O oposto pode ser verificado na Figura 2.30b, onde temos mais informações nas regiões mais claras. Um exposição balanceada apresenta um histograma com informações

bem distribuídas ao longo de todo o gráfico, tanto nas regiões escuras quanto nas claras, conforme mostra a Figura 2.30c.

2.3.7 Balanço de Branco ou *WB* (White Balance)

O *WB* é o ajuste das cores presentes em uma cena tomando-se como referência uma cor neutra, neste caso, o branco. As fontes de luz possuem uma variação nas tonalidades de cores que influenciam diretamente na cor refletida pelo objetos. A luz do sol, por exemplo, apresenta uma cor levemente amarelada, enquanto que a luz do *flash* é mais branca. O olho humano interpreta automaticamente a luz branca do ambiente utilizando-a como referência para o ajuste das demais cores. As câmeras fotográficas realizam esse mesmo processo, porém de forma não tão eficaz.

Dependendo da fonte de luz que ilumina o ambiente, os objetos exibidos nas imagens capturadas podem não apresentar a fidelidade das cores que os mesmos apresentariam sob uma iluminação difusa. É sob esta iluminação que conseguimos capturar as cores dos objetos presentes em cena com fidelidade mais próxima do real, ou seja, próxima ao que estamos enxergando a olho nu. Na Figura 2.31, podemos observar no espectro apresentado a variação da temperatura de cor para diferentes tipos de iluminação. Temperatura de cor pode ser definida como a temperatura necessária para que um corpo negro emita radiação eletromagnética (luz) com comprimento de onda igual ao da cor que queremos usar como branco.

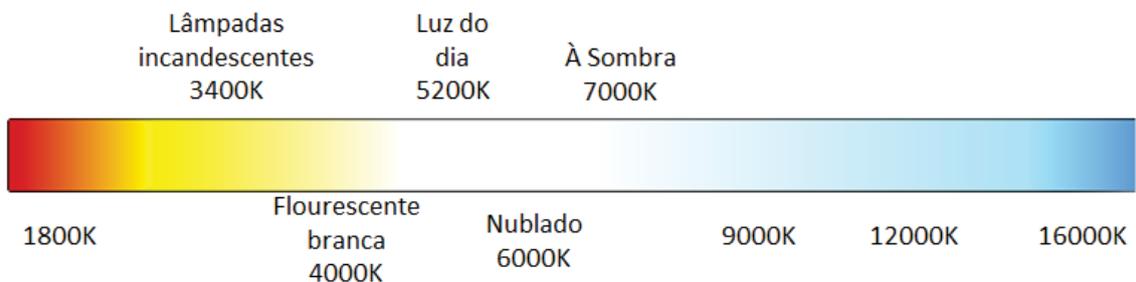


Figura 2.31: Temperatura de cor para diferentes tipos de iluminação.

O balanço de branco busca equilibrar as cores componentes do espectro da luz branca inserindo tons avermelhados (caso a cena apresente tons azulados) ou tons azulados (caso a cena apresente tons avermelhados) na imagem capturada.

Diferentemente das câmeras analógicas, as quais utilizavam filtros ou lentes específicas para correção e/ou compensação de cor, as câmeras digitais possuem um dispositivo que realiza essa atividade. Essa correção pode ser feita automaticamente (*AWB - Automatic White Balance*) ou manualmente, corrigindo a receptividade da câmera às diferentes cores. A Figura 2.32 exemplifica alguns resultados da aplicação de diferentes tipos de *WB*.

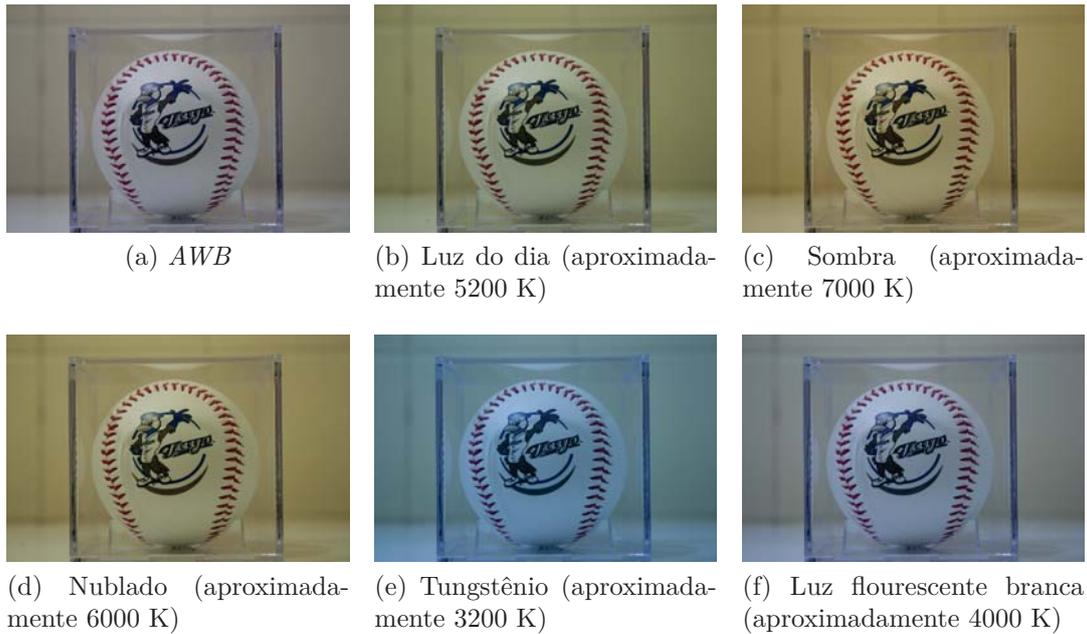


Figura 2.32: Resultados da aplicação de diferentes tipos de WB.

2.3.8 Lentes Fotográficas (Objetivas)

As objetivas são compostas por um conjunto de 6 a 12 lentes em seu interior. Estas são necessárias para a correção de diversos problemas na forma, cor e foco que ocorrem durante a captura da imagem. Além da qualidade dos cristais e produtos químicos que são aplicados nas lentes, a principal característica que as diferencia é a distância focal. Quanto maior for a distância focal (em mm), menor será a área visualizada da cena de interesse. Além disso, teremos a sensação de uma maior proximidade dos planos que compõem a imagem. Por outro lado, quanto menor for a distância focal, maior será a área capturada na imagem, porém os planos da imagem aparentarão estar situados a uma distância maior.

As objetivas são classificadas como normais, grande-angulares e teleobjetivas (Figura 2.33). As normais recebem esse nome pois são as que mais se aproximam do ângulo de visão equivalente à visão humana (aproximadamente 50°). Além disso, a distância entre os planos que compõem a imagem são bastante similares às estabelecidas pelo olho humano. Distâncias focais entre 35 e 70 mm caracterizam as lentes normais.

As grande-angulares são aquelas que possuem uma distância focal de até 35 mm. Caracterizam-se por permitir a visualização de um grande ângulo de visão, porém a distância aparente entre os planos que compõem a imagem são bem maiores. Além disso, essas objetivas caracterizam-se por apresentar distorções nos objetos fotografados, principalmente os localizados nos cantos do quadro fotográfico ou quando estão muito próximos à lente. Esta distorção é menos acentuada para distâncias

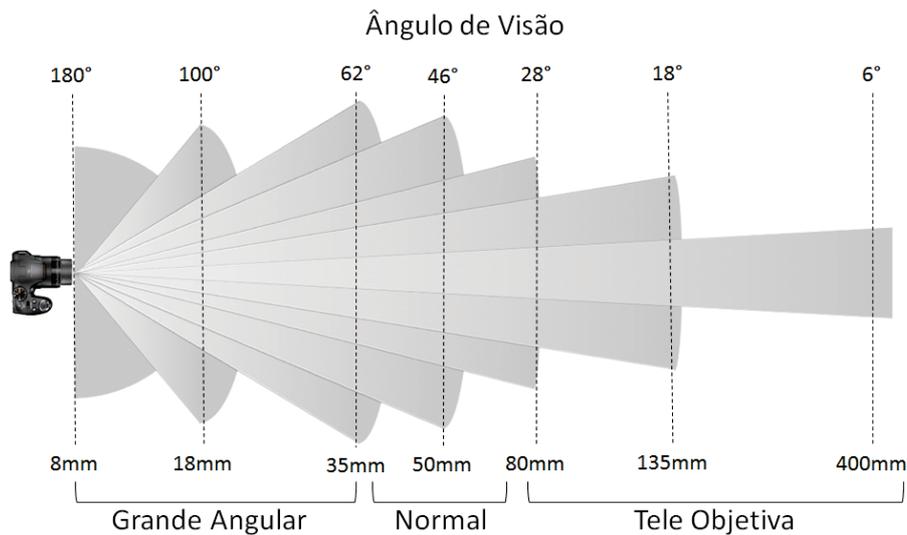


Figura 2.33: Ângulos de visão para diferentes distâncias focais, além da respectiva classificação das objetivas baseadas nas mesmas.

focais entre 24 e 35 mm. A Figura 2.34 exibe uma fotografia tirada com a lente grande-angular.



Figura 2.34: Fotografia capturada com uma lente grande-angular.

Outra importante característica das grande-angulares é a profundidade de campo. As cenas registradas com este tipo de lente, mesmo se utilizarmos baixos valores de abertura de diafragma, apresentam uma profundidade de campo maior (Seção 2.3.3), fazendo com que os elementos localizados em uma determinada região próxima ao plano focal estejam nítidos (focados).

As teleobjetivas possuem distância focal acima de 70 mm. Com estas lentes é possível registrar os objetos de interesse localizados a grandes distâncias (*zoom* ótico). Diferentemente das grande-angulares, a tele praticamente não distorce a imagem,

porém apresenta um achatamento dos planos que à compõem. Elas aparentam estar mais próximos entre si do que realmente estão. Além disso, a profundidade de campo para as cenas capturadas com esse tipo de objetiva é bastante pequena. Dessa forma, é possível obter apenas um elemento nítido, o qual está localizado sob o plano focal. As demais regiões permanecerão completamente desfocadas. A Figura 2.35 exibe a mesma cena anterior (Figura 2.34), porém tirada com a teleobjetiva.



Figura 2.35: Fotografia capturada com uma lente teleobjetiva.

Capítulo 3

Processo de Geração de Mapas de Profundidades

Para a correta utilização do processo de *Depth Hallucination* é necessária a captura precisa de todas as imagens, as quais serão posteriormente processadas. Dessa forma, utilizando os conceitos básicos de fotografia apresentados na Seção 2.3, explicaremos detalhadamente todo o processo de captura das imagens (Seção 3.1) da superfície de interesse, além de boas práticas e cuidados durante a realização do mesmo.

Por fim, com o objetivo de facilitar a utilização da técnica pelo usuário final, apresentaremos as principais características do *software* desenvolvido na linguagem C# (sharp), o qual possui uma simples e amigável interface de fácil utilização (Seção 3.2).

3.1 Como Capturar as Imagens

3.1.1 Ajustes Iniciais da Câmera e Acessórios

As câmeras DSLR possuem algumas opções de configuração para auxiliar o usuário na captura das imagens. Dependendo do fabricante, os nomes atribuídos podem ser diferentes, mas a função é a mesma.

Existem algumas funções de captura da imagem completamente automática, uma delas é representada geralmente pela letra P nas câmeras das marcas Canon e Nikon. Nesta opção, a câmera ajusta automaticamente todas as configurações da mesma (abertura do diafragma, velocidade do obturador, entre outras). Não podemos utilizar essa opção, pois algumas características da captura da imagem difusa precisam ser mantidas durante a captura da imagem com *flash*. Utilizando a câmera nessa configuração, todos os parâmetros sofrerão alterações, inviabilizando o processo.

As câmeras também possuem modos de prioridade em relação à abertura do diafragma (Av para Canon e A para Nikon) e à velocidade do obturador (Tv para Canon e T para Nikon). No primeiro caso, o usuário determina qual a abertura do diafragma ele deseja e, automaticamente, a câmera seleciona uma velocidade do obturador que atenda a escolha do usuário de tal forma a obter uma exposição balanceada (Seção 2.3.6). No segundo caso, ocorre o inverso: o usuário seleciona a velocidade que deseja para o obturador, enquanto a câmera ajusta a melhor abertura do diafragma.

Da mesma forma que a opção automática, não podemos utilizar essas funções, uma vez que as características da imagem podem ser modificadas com o ajuste automático da câmera. Por exemplo, para a prioridade na abertura do diafragma, deveríamos mantê-la constante nas fotos capturadas (difusa, *flash* e Lambertiana). Entretanto, a variação automática dada pela câmera para a velocidade do obturador pode não ser suficiente para capturarmos uma imagem com *flash* balanceada, que compense o excesso de luz gerado pelo mesmo (Seção 2.3.6).

Já na situação com prioridade para a velocidade do obturador, teríamos uma variação na abertura do diafragma, o qual deve permanecer fixo em todas as imagens. Além disso, pode ocorrer uma variação da profundidade de campo (Seção 2.3.3) de uma foto para a outra, podendo tornar não nítidas algumas regiões da superfície de interesse.

Dessa forma, é necessário que a câmera esteja na opção manual (M) (Figura 3.1), a qual permite o ajuste de todas as características da câmera, da forma que o usuário ache mais conveniente para cada situação.



Figura 3.1: Câmera na função Manual (M).

Conforme descrito na Seção 2.3.2, a câmera deve estar configurada para a captura das imagens no formato *raw*, que permite o máximo registro possível de informação por pixel, além de possuir os metadados com as características e configurações da câmera e objetiva utilizadas.

Apesar das recomendações [1] para a utilização de um *WB* (Seção 2.3.7) para *daylight*, conseguimos obter bons resultados utilizando o ajuste do branco automático (*AWB*). Em muitos casos, foi apresentado um melhor balanço de cor no albedo, o qual ficou mais próximo das cores reais da superfície, em comparação com o recomendado.

Idealmente, a câmera utilizada deve possuir um *flash* embutido para a captura do *flash-lit*. Caso não possua, pode-se utilizar um *flash* externo, porém o mesmo deve ficar o mais próximo possível do corpo da câmera a fim de minimizar as sombras geradas por sua utilização.

É extremamente importante a utilização de um tripé para auxiliar na captura das fotografias. Uma vez que as fotos devem ser tiradas de uma mesma posição, o mesmo permitirá uma boa fixação da câmera, além de ajudar a evitar possíveis translações e/ou rotações entre as imagens capturadas. Esse tipo de deslocamento, em nível de *pixel*, é bastante comum quando abrimos o *flash* embutido na câmera, por exemplo. Se os *pixels* nas imagens capturadas não corresponderem precisamente entre si, isto comprometerá a criação do mapa de profundidades. Assim, é muito importante uma boa fixação da câmera no tripé, além do mesmo estar bem equilibrado em relação ao solo.

Em relação à lente fotográfica que deve ser utilizada, devemos dar preferência às grande-angulares. Apesar de gerarem distorções mais acentuadas nas extremidades, essas lentes permitem a captura de uma área maior, mesmo a curtas distâncias, além de uma maior profundidade de campo (Seção 2.3.8). Entretanto, as distorções não serão problemáticas, uma vez que podemos atenuá-las através da utilização do *software Adobe Camera Raw*, o qual obtém informações dos metadados sobre a objetiva utilizada na foto em questão.

É necessária a utilização de uma trena pois, para capturarmos a imagem da superfície Lambertiana, precisaremos posicionar o tripé e a câmera na mesma distância, à superfície plana, em que estavam, do objeto, no momento da captura do *diffuse-lit* e *flash-lit*.

3.1.2 Capturando as Imagens

Após a escolha da superfície de interesse, posiciona-se a câmera já fixada no tripé de tal forma que a mesma fique o mais perpendicular possível à superfície. A primeira foto a ser capturada é a *diffuse-lit*.

Após o posicionamento da câmera, deve-se estabelecer o foco sobre a superfície. O mesmo pode ser feito manualmente (MF - *Manual Focus*) ou automaticamente (AF - *Auto Focus*) (Figura 3.2). Na opção automática, a câmera estabelece o foco baseado no contraste de determinados pontos na cena observada. Dependendo do fabricante da mesma, o número de pontos para avaliação do foco pode variar. Algumas câmeras profissionais usam até 49 pontos para estimação do foco.

Caso o ajuste do foco seja feito automaticamente (AF), após estabelecer o mesmo, recomenda-se colocar em MF. Dessa forma, o foco estabelecido permanecerá fixo, não permitindo que haja um novo ajuste do mesmo antes da captura de cada imagem.



Figura 3.2: Selecionando AF ou MF.

Apesar da superfície estar sob uma iluminação indireta, conforme explicação na Seção 2.2.1, primeiramente devemos escolher um ajuste de ISO de tal forma que não haja, nas imagens capturadas, muito ruído (imagem granulada). Assim, um baixo valor de ISO é o ideal, entre 50 e 200.

Em seguida, devemos escolher um ajuste de abertura do diafragma e velocidade do obturador. Dependendo da distância a que a câmera estiver da superfície, o valor da abertura do diafragma não pode ser muito pequeno (maior diâmetro, conforme Seção 2.3.3), pois dessa forma teremos uma pequena profundidade de campo e, dependendo das características da superfície, podemos ter um desfoque em algumas regiões da mesma (picos ou depressões).

Por fim, uma vez que a abertura do diafragma já foi escolhida, estabelecemos o valor da velocidade do obturador de tal forma que tenhamos uma imagem balanceada. Mesmo que a velocidade selecionada seja baixa, não devemos nos preocupar com possíveis borrões na imagem final pois a câmera encontra-se sob o tripé, dando-lhe a fixação necessária mesmo para imagens com longo tempo de exposição.

Antes da captura em definitivo da imagem *diffuse-lit*, deve-se capturar uma primeira imagem para avaliar se a mesma está corretamente exposta, dado que, conforme Seção 2.3.6, apesar de termos uma leitura do fotômetro balanceada, dependendo das características da superfície de interesse, como especularidades por exemplo, poderemos ter uma imagem subexposta ou superexposta. Para uma melhor avaliação, pode-se utilizar a função histograma presente na grande maioria das câmeras digitais. Após a constatação de que os ajustes nos fornecerão uma imagem balanceada, capturamos a imagem difusa.

A próxima captura é da imagem *flash-lit*. Caso a câmera possua um *flash* embutido, devemos abri-lo com o máximo de cuidado, pois poderá haver um pequeno deslocamento (em nível de *pixel*) da câmera, impossibilitando um pós-processamento das imagens uma vez que os *pixels* das imagens difusa e com *flash* não corresponderão uns aos outros.

Apesar do processo de *Depth Hallucination* informar que a potência do *flash* deve ser a maior possível, após a realização de diversas capturas, concluímos que essa não é a configuração ideal. A potência deve ser ajustada de tal forma que as regiões sombreadas na imagem difusa sejam reveladas. Dependendo da distância de

captura e das características dos materiais que compõem a superfície de interesse, ao utilizarmos a potência máxima, acabamos obtendo uma imagem superexposta, além de possíveis pontos com especularidades geradas pelo *flash*, as quais resultarão em distorções no mapa de profundidades. A Figura 3.3 exemplifica a captura de uma imagem utilizando uma alta potência de *flash*. Como resultado, além de obtermos uma *flash-lit* superexposta, podemos observar vários pontos com especularidades na superfície de interesse, dado que a mesma possui características reflexivas.



Figura 3.3: Exemplificação de uma imagem capturada utilizando-se uma forte potência do *flash*.

Para a captura da imagem *flash-lit*, os únicos parâmetros que podemos alterar são a velocidade do obturador e o ISO. Tanto a abertura do diafragma como o foco, escolhidos para a captura da imagem difusa, devem ser mantidos. A velocidade deve ser maior (menor tempo de exposição) que a escolhida para a imagem difusa, uma vez que devemos compensar o excesso de luz gerado pelo *flash*. Dessa forma, a escolha da nova velocidade está atrelada à potência do *flash* selecionado. Algumas capturas prévias devem ser feitas para avaliar se as regiões sombreadas foram reveladas e se a combinação velocidade do obturador e potência do *flash* resultam numa imagem bem equilibrada na exposição.

A Figura 3.4 apresenta os resultados de uma correta captura das imagens difusa (Figura 3.4a) e com *flash* (Figura 3.4b) baseadas na correta configuração dos ajustes da câmera, além dos cuidados durante a captura (conforme explicação anterior). Podemos observar que a principal diferença entre as duas imagens é a revelação dos

detalhes das depressões da superfície de interesse, dado que na imagem difusa os mesmos estavam sob uma região de penumbra ou sombra, enquanto que na imagem com *flash* eles foram revelados, sem necessariamente obtermos uma imagem superexposta .



(a) *Diffuse-lit*



(b) *Flash-lit*

Figura 3.4: Exemplificação de uma correta captura das imagens difusa e com *flash*.

Após a captura da *flash-lit*, deve-se medir a distância entre a superfície e o sensor da câmera, pois precisamos capturar mais uma imagem: a de uma superfície Lambertiana branca (Seção 2.2.2), necessária para o cálculo do albedo. A mesma é responsável por corrigir os efeitos de *vignetting*, *fall-off* e o fator cosseno global presentes nas imagens difusa e com *flash*. O símbolo, mostrado na Figura 3.5, é encontrado na câmeras digitais e indica onde está posicionado o sensor das mesmas.



Figura 3.5: Símbolo indicando a localização do sensor na câmera.

Deve-se posicionar a câmera fixada no tripé na distância anteriormente medida. Os ajustes da mesma (velocidade, diafragma, ISO e potência do *flash*) devem ser os mesmos utilizados na captura da *flash-lit*. Vale observar que, uma vez que a superfície é homogênea monocromática, a realização do foco automático fica impossibilitada, pois, conforme explicação anterior, o mesmo baseia-se no contraste presente em determinados pontos na cena. Assim, a regulagem do foco deve ser manual (MF).

3.2 *Software* Desenvolvido

Visando facilitar a utilização do processo de *Depth Hallucination* pelo usuário final, foi desenvolvido um *software* com uma interface simples e amigável. A linguagem C# (sharp) foi escolhida pois a mesma é de fácil aprendizagem e uso, robusta e com boa performance. Além disso, em conjunto com os demais recursos da arquitetura .Net, o C# também é a linguagem ideal para a criação de uma nova categoria de programas que aproveitam as oportunidades trazidas da internet. Para o desenvolvimento do *software*, foi utilizado o Visual C# 2010 Express Edition, o qual é fornecido e registrado gratuitamente na internet.

3.2.1 Interface Gráfica

A interface gráfica criada para o usuário é de fácil compreensão e utilização (Figura 3.6). Ela possui alguns campos específicos, os quais devem ser preenchidos para que todas as etapas seguintes, relativas ao processamento das imagens, sejam realizadas.

Os primeiros campos são relativos às características das configurações da câmera utilizadas para a aquisição das imagens difusa e com *flash*, ou seja, velocidade do obturador, abertura do diafragma e ISO. Não é necessário o preenchimento das características relativas à imagem Lambertiana, uma vez que a mesma possui os mesmos ajustes da imagem com *flash*.

Outra importante informação que deve ser fornecida ao *software* é a resolução das imagens que serão utilizadas. O usuário deve informar primeiramente a largura da imagem, seguido da altura da mesma, ambas em *pixels*. Essas informações são

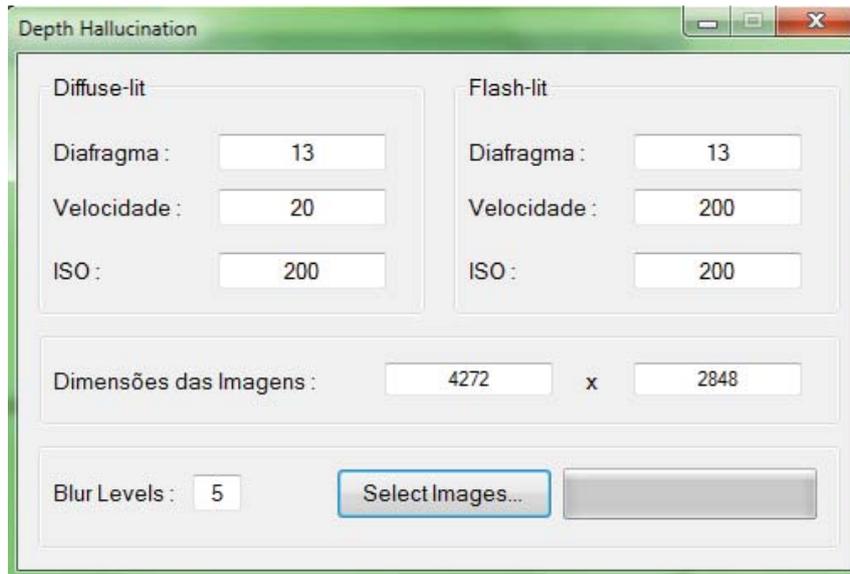


Figura 3.6: Interface gráfica do *software* desenvolvido.

necessárias para podermos determinar o tamanho do *array* que conterà os dados relativos aos *pixels* de cada imagem utilizada no processo.

Por fim, a última informação a ser fornecida para o *software* é o número de níveis de *blur* (*blur levels*) que devem ser utilizados durante a etapa de filtragem Gaussiana (2.2.3). Quanto maior esse número, mais preciso e suave será o mapa de profundidades gerado. Porém, maior será o tempo de processamento.

Após preencher todos os campos anteriores, o usuário deve pressionar o botão *Select Images*, o qual exibirá um *Open File Dialog* para que seja informado o arquivo relativo à primeira imagem, neste caso a *diffuse-lit*. Em seguida, novamente é exibida a mesma janela do *Open File Dialog* para que seja informado qual é o arquivo relativo à imagem *flash-Lit*. Por fim, o mesmo deve ser feito para a imagem Lambertiana.

Visando informar ao usuário sobre o andamento das etapas do processamento do *Depth Hallucination*, uma barra de progresso foi inserida ao lado do botão *Select Images*. A mesma retratará, apenas visualmente, o percentual já finalizado para cada nível de *blur* processado. Essa etapa foi a escolhida para representar a porcentagem já finalizada do processo pois trata-se da atividade mais longa durante o processamento.

3.2.2 Conversão das Imagens

Conforme explicação na Seção 2.3.2, as imagens capturadas devem estar no formato *raw*, o qual possui extensões diferentes de arquivo dependendo do fabricante da câmera utilizada. Dessa forma, visando padronizar o formato do arquivo de entrada para o *software* desenvolvido, devemos converter os arquivos *raw* referentes a cada uma das imagens adquiridas. Para isso, utilizamos o *Adobe Camera Raw*, o

qual trata-se de um *plugin* que funciona em conjunto com o *Adobe Photoshop*.

Uma vez que, a partir do arquivo *raw*, e utilizando o *plugin* acima, podemos converter os arquivos para qualquer extensão, a mesma deve ser escolhida de tal forma que os dados originais, com todas as informações e características das imagens, não sejam alterados. Além disso, a extensão a ser escolhida deve ser de fácil utilização, uma vez que precisaremos acessar todas as informações referentes aos valores de cada *pixel* que constituem a imagem. Dessa forma, o formato escolhido para a conversão dos arquivos foi o *Photoshop RAW (.raw)*.

O processo de conversão das imagens é muito simples. Inicialmente selecionamos os arquivos *.cr2* (para câmeras Canon) ou *.nef* (para câmeras Nikon) ou outras extensões dependendo da câmera utilizada. Em seguida, abrimos as mesmas utilizando o *Photoshop*. Por se tratarem de imagens "cruas", automaticamente as mesmas são exibidas no *Adobe Camera Raw*, conforma Figura 3.7.

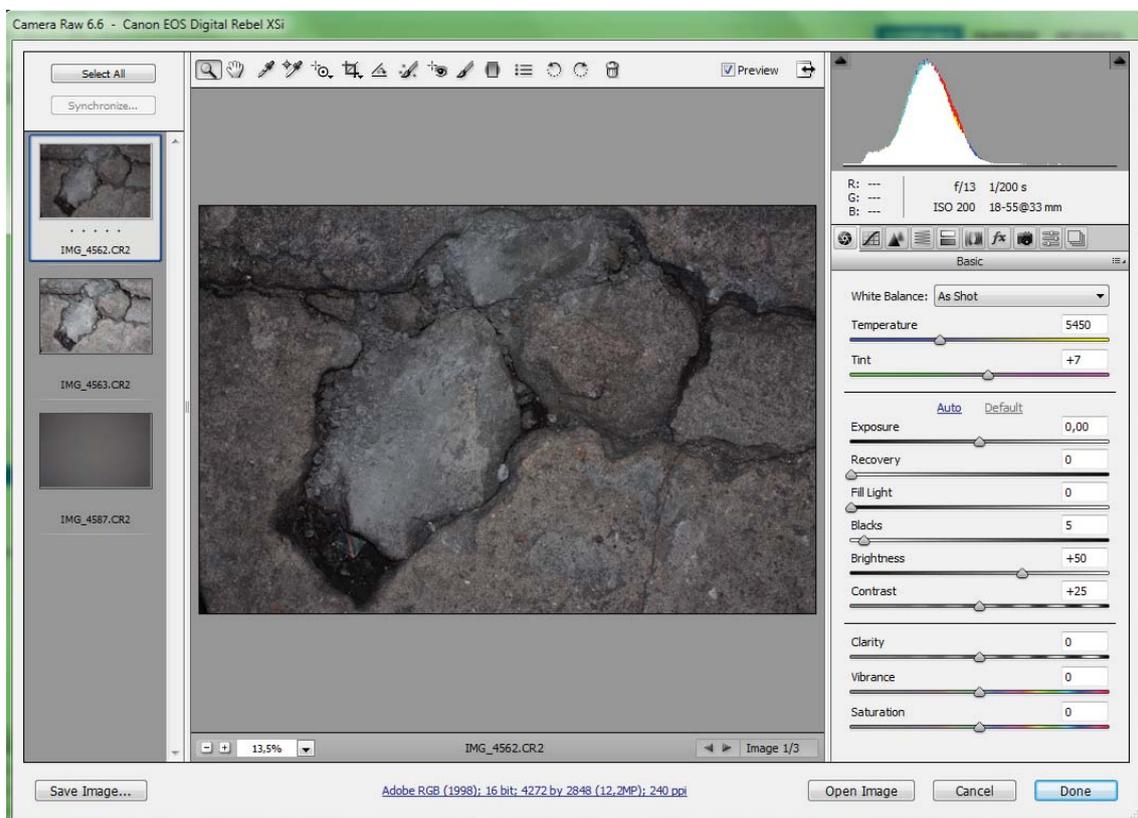


Figura 3.7: Imagens *raw* sendo exibidas no *Adobe Camera Raw*.

Antes de abrirmos as imagens em definitivo no *Photoshop* e realizarmos a conversão, devemos ajustar o mesmo para que o seu processamento seja em 16 bits/canal, conforme é exibido na Figura 3.8. Dessa forma, não haverá nenhuma perda de informação ao abrirmos e convertermos as imagens dentro do programa.

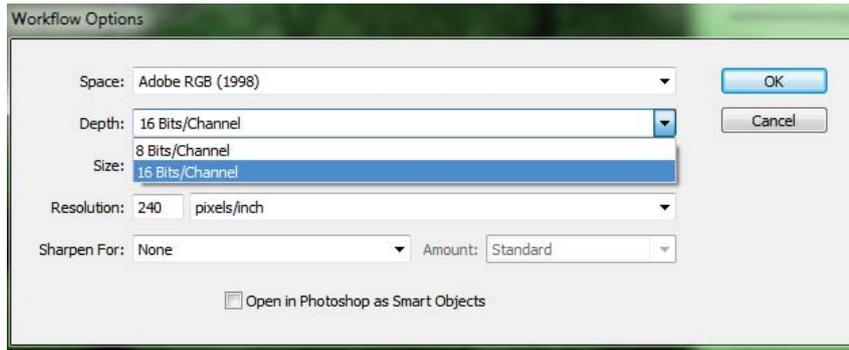


Figura 3.8: Ajustando o processamento do *Photoshop* para 16 bits/canal.

3.2.3 Funções

A Figura 3.9 exibe todas as funções e a sequência em que cada uma delas será executada após o usuário pressionar o botão *Select Images*. Trata-se de uma sequência de funções em cascata as quais são relativas às etapas necessárias para o processamento do *Depth Hallucination*.

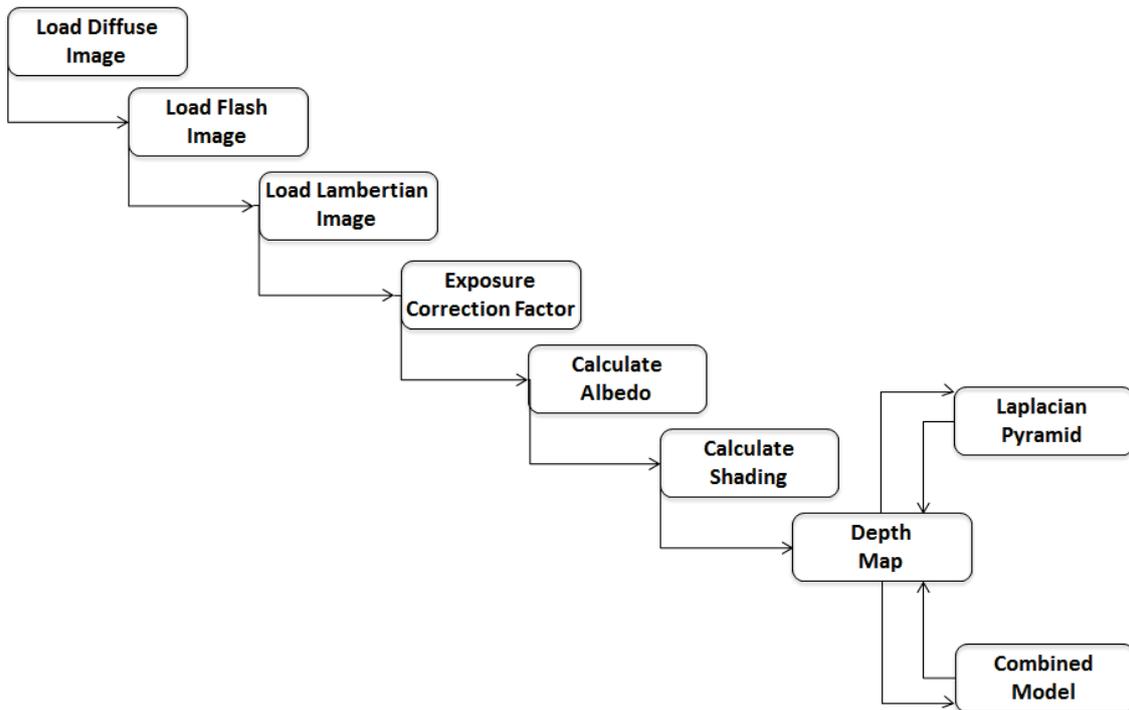


Figura 3.9: Representação das funções desenvolvidas para o processo de *Depth Hallucination*.

A seguir, serão descritas a partir de pseudo-códigos todas as principais funções desenvolvidas para o *software* em questão.

Load Images

Baseado na interface gráfica desenvolvida (Seção 3.2.1), o pseudo-código na Tabela 3.1 realiza a chamada da mesma. Em seguida, captura as informações fornecidas pelo usuário através da interface gráfica e, por fim, armazena as imagens selecionadas pelo mesmo em *arrays*.

Tabela 3.1: Pseudo-código para leitura da imagens selecionadas pelo usuário.

```
Função LoadImages()
{
  Exibir_Interface();
  Se: Botao.Click(), Então:
  {
    Armazenar_Infos_Velov_Diafragma_ISO;
    Para i = 0 Até i < 3:
    {
      Exibir_OpenFileDialog();
      Se: ImagemSelecionada.Existir(), Então:
        array = LerImagem();
      Senão:
      {
        ExibirMensagemErro();
        Retornar();
      }
    }
  }
}
```

Exposure Correction Factor

Conforme descrito na Seção 2.2.2, essa função, exibida na Tabela 3.2, simplesmente calcula o fator de correção da exposição de cada uma das imagens capturadas baseado nos valores estabelecidos para a velocidade do obturador, abertura do diafragma e ISO.

Tabela 3.2: Pseudo-código para cálculo dos fatores de correção.

```
Função ExposureCorrectionFactor(veloc, diafragma, iso)
{
  Retornar ((diafragma)2/(veloc*iso))
}
```

Calculate Albedo

Utilizando as imagens anteriormente armazenadas em *arrays* e os fatores de correção da exposição, calculamos o albedo (Tabela 3.3) conforme teoria na Seção 2.2.2.

Tabela 3.3: Pseudo-código para cálculo do albedo.

```
Função CalculateAlbedo()
{
  Para i = 0 Até i < último_pixel:
  {
    albedo_array[i] = ((flash_array[i] * fator_correção_flash) -
                      (diffuse_array[i] * fator_correção_diffuse)) /
                      (lambertian_array[i] * fator_correção_Lambertian);
  }

  max = máximo_valor_existente_em_albedo_array;

  Para i = 0 Até i < último_pixel:
  {
    albedo_array[i] = (albedo_array[i] / max );
  }
}
```

Calculate Shading

Após o cálculo do albedo, segue-se a etapa de criação do *shading image* (Tabela 3.4), conforme explicação teórica na Seção 2.2.2.

Depth Map

Na etapa de cálculo do *Depth Map* (Seção 2.2.3), cujo pseudo-código é apresentado na Tabela 3.5, são necessárias várias operações matemáticas, dentre estas, convoluções. Visando facilitar a visualização dos resultados numéricos obtidos, o *gray_array*, utilizado anteriormente na Tabela 3.4, é convertido em uma matriz bidimensional com o número de linhas e colunas iguais as dimensões das imagens utilizadas no processo.

A variável *matrix3D* consiste numa matriz de três dimensões onde são armazenados os resultados de cada convolução para cada um dos níveis de *blur*. A função *Add* é a responsável por adicionar a matriz resultante do processo de convolução na *matrix3D*.

Tabela 3.4: Pseudo-código para cálculo do *Shading*.

```

Função CalculateShading()
{
  Para i = 0 Até i < último_pixel:
  {
    shading_array[i] = ((diffuse_Array[i] * fator_correção_diffuse)/
                        albedo_array[i]); /
    shading_array[i] = shading_array[i] / 20;
  }

  Para i = 0 Até i < último_pixel:
  {
    Se: shading_array[i] < 0, Então: shading_array[i] = 0;
    Se: shading_array[i] > 65535, Então: shading_array[i] = 65535;
    Senão: Truncar(shading_array[i]);
  }

  Para i = 0 Até i < último_pixel Incremento = i + 3:
  {
    gray_array[k] = (0.2989 * shading_array[i]) +
                    (0.5870 * shading_array[i + 1]) +
                    (0.1140 * shading_array[i + 2]);

    k++;
  }

  max = máximo_valor_existente_em_gray_array;
  min = mínimo_valor_existente_em_gray_array;

  Para i = 0 Até i < último_pixel:
  {
    gray_array[i] = (gray_array[i] - min) / (max - min);
  }
}

```

Tabela 3.5: Pseudo-código para cálculo do *depth map*.

```
Função DepthMap()
{
  Para i = 0 Até i < número_linhas:
  {
    Para j = 0 Até j < número_coluna:
    {
      shading_matrix[i][j] = shading_array[k];
      k++;
    }
  }

  matrix3D = Add(shading_matrix)

  Para i = 1 Até i < níveis_blur:
  {
    H = Kernel(raio_kernel, sigma);
    resultado = Convolução(shading_matrix, H);
    matrix3D = Add(resultado);
    raio_kernel =  $3^i$  + raio_kernel;
  }

  Para i = 1 Até 1 < (níveis_blur - 1):
  {
    laplacian_matrix = LaplacianPyramid(matrix3D, i);
  }

  Para k = 0 Até k < níveis_blur:
  {
    Para i = 0 Até i < número_linhas:
    {
      Para j = 0 Até j < número_colunas:
      {
        profundidade = CombinedModel(laplacian_matrix[i][j][k]);
        depth_matrix[i][j] = depth_matrix[i][j] +
          ( $3^k$  * (profundidade - 1));
      }
    }
  }
}
```

Laplacian Pyramid

Baseado na teoria da Seção 2.2.3, a partir de cada nível de *blur* gerado, a imagem criada é dividida pela imagem do próximo nível superior e, em seguida, normalizada em 0.5. Obtemos assim uma pirâmide Laplaciana, conforme pseudo-código na Tabela 3.6.

Tabela 3.6: Pseudo-código para cálculo da pirâmide Laplaciana.

```
Função LaplacianPyramid(matrix, index)
{
  Para i = 0 Até i < número_linhas:
  {
    Para j = 0 Até j < número_coluna:
    {
      laplacian_matrix[i][j][index] = 0.5 * (matrix[i][j][index]/
      matrix[i][j][index + 1]);
      k++;
    }
  }
}
```

Combined Model

A função relativa a esta etapa foi desenvolvida baseada na Equação (2.17). Dessa forma, temos o pseudo-código na Tabela 3.7.

Tabela 3.7: Pseudo-código para cálculo a função *Combined Model*.

```
Função CombinedModel(value)
{
  Se: value <= 0.5, Então:
    Retornar Sqrt( ( 1 / value ) - 1 );
  Senão: Retornar ( 2 * ( 1 - value ) );
}
```

Capítulo 4

Visualização do Mapa de Profundidades

Após a correta captura das imagens, baseada nas boas práticas explicadas na Seção 3.1.2, e das operações realizadas pelas funções exibidas na Seção 3.2.3, obtemos como resultado final uma matriz bidimensional com o mesmo número de linhas e colunas das imagens com e sem *flash* e da imagem Lambertiana. Essa matriz contém o mapa de profundidades de cada *pixel* existente na imagem. A última etapa do processo consiste em exibir esse mapa de profundidades para o usuário.

Para a visualização do mesmo, foi utilizado o *software Height Map Viewer*, o qual utiliza a biblioteca *TerrainLib* [2] desenvolvida para a renderização de terrenos e florestas utilizando *OpenGL*. A Figura 4.1 exibe a interface do programa.



Figura 4.1: *Software Height Map Viewer* utilizado para a visualização dos mapas de profundidades. Cortesia de [2].

A Seção 4.1 exibe os diferentes resultados obtidos nos testes empíricos a partir de diversas combinações entre o valor do *sigma* utilizado na filtragem gaussiana, assim como o número de níveis de *blur* utilizados. Por fim, a Seção 4.2 apresentará alguns exemplos dos bons resultados obtidos utilizando-se a técnica do *Depth Hallucination*, bem como algumas limitações da mesma.

4.1 Sigma e Número de Níveis do *Gaussian Blur*

Conforme a explicação apresentada na Seção 2.2.3, foram realizados testes empíricos com os valores da variável *sigma* e do número de níveis de *blur* referentes à filtragem gaussiana, etapa esta necessária para a criação de um correto mapa de profundidades. O teste consistia em inicialmente fixarmos o valor do número de níveis e variarmos o valor do *sigma*. Em seguida, repetimos o processo incrementando-se o número de níveis de *blur* até chegarmos a um resultado próximo ao real (fotoreal).

Para o número de níveis de *blur* igual a 3 (três), independentemente do valor do *sigma* escolhido, podemos notar a presença de muito ruído nos mapas de profundidades resultantes. Além disso, não há a presença de diferenças de alturas no resultado observado, ou seja, a superfície reconstruída assemelha-se à uma superfície plana, sem desníveis. A Figura 4.2 exemplifica o resultado obtido.

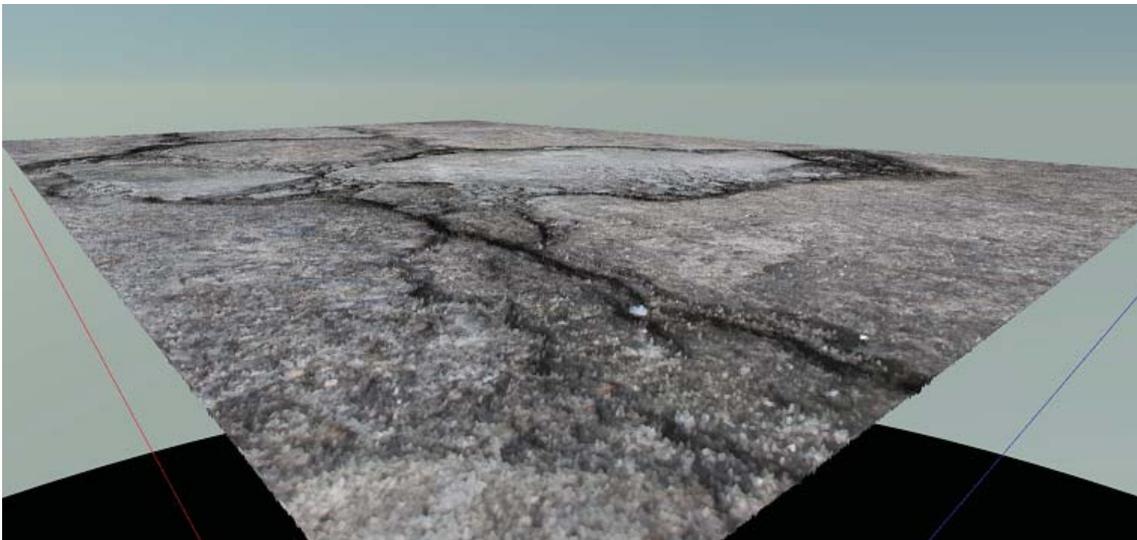


Figura 4.2: Mapa de profundidades obtido utilizando-se $\sigma = 10$ e níveis de *blur* igual a 3.

Aumentando-se o número de níveis para 4 (quatro), podemos observar uma variação mínima de altura na superfície reconstruída. Entretanto, a presença de um forte ruído, ainda existente e em destaque (independente do valor de *sigma*), faz com que o resultado obtido não seja fiel às características reais observadas na superfície fotografada. Na Figura 4.3 podemos observar os problemas anteriormente descritos.

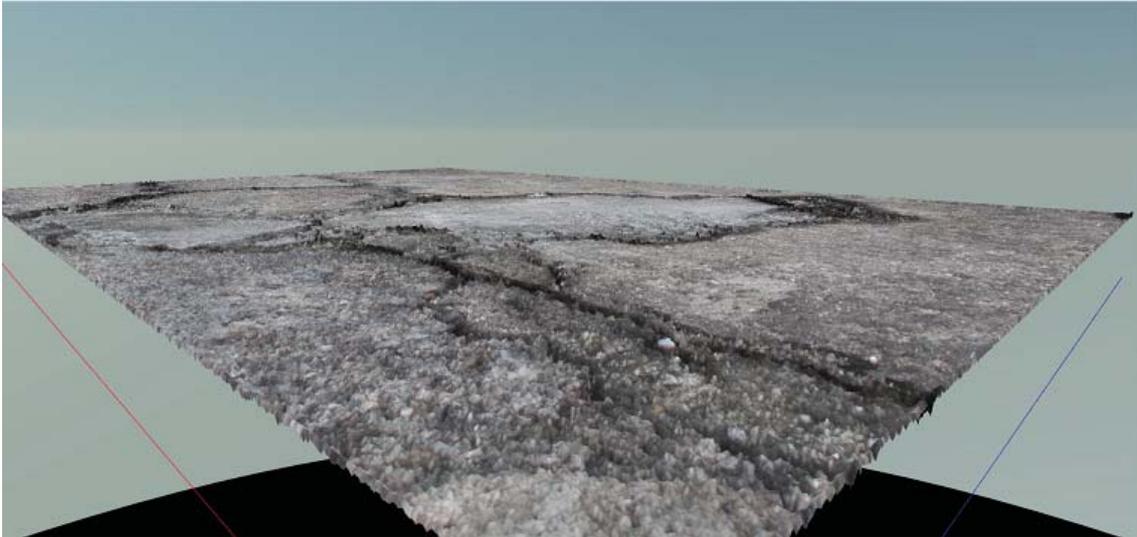


Figura 4.3: Mapa de profundidades obtido utilizando-se $\sigma = 50$ e níveis de *blur* igual a 4.

Para 5 (cinco) níveis de *blur*, podemos notar variações nos resultados obtidos para diferentes valores de *sigma* utilizados. Com este variando de 1 até 10, podemos notar os problemas anteriormente relatados: pequena variação na altura e a presença de muito ruído. Já para os valores de 20 até 50, observamos uma diminuição no ruído anteriormente característico, além de notarmos os buracos mais profundos e os picos mais elevados. Porém o melhor resultado foi obtido para $\sigma = 100$, onde podemos notar uma maior variação na altura dos elementos característicos da superfície e um ruído menos evidente que nas situações anteriores (Figura 4.4). Para valores de *sigma* acima de 100, mínimas alterações foram observadas.

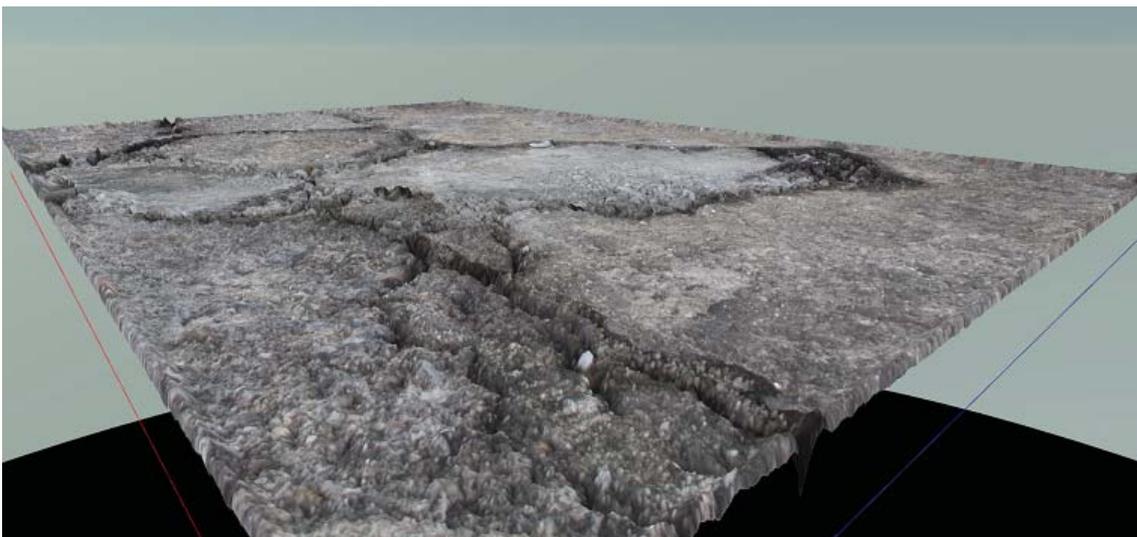


Figura 4.4: Mapa de profundidades obtido utilizando-se $\sigma = 100$ e níveis de *blur* igual a 5.

Com 6 (seis) níveis de *blur* e *sigma* igual a 100, conseguimos obter um resultado

bastante satisfatório e muito próximo do observado no real. Todas as características da superfície fotografada estão bem representadas e apresentam as alturas condizentes com o observado. Em relação ao ruído, podemos notar uma acentuada diminuição do mesmo, não afetando o resultado final, diferentemente dos testes anteriores. A Figura 4.5 exibe o resultado obtido.

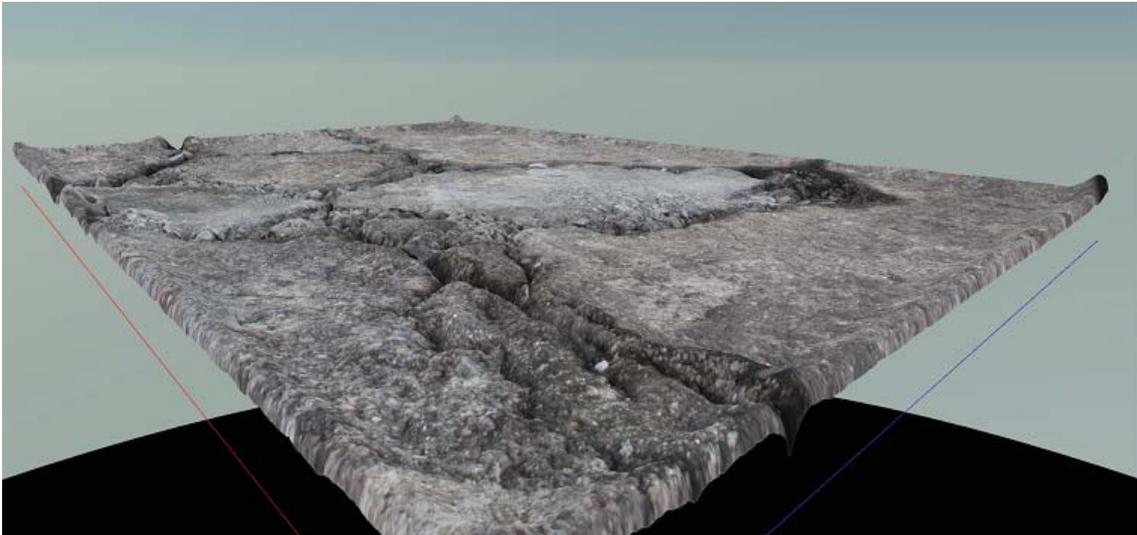


Figura 4.5: Mapa de profundidades obtido utilizando-se $\sigma = 100$ e níveis de *blur* igual a 6.

Para um número de níveis de *blur* mais elevado, os resultados obtidos são tão satisfatórios quanto o observado na Figura 4.5. Porém, quanto maior o número de níveis, mais longa será a etapa de processamento das imagens. Desta forma, baseado no objetivo final do usuário na utilização do mapa de profundidades, cabe ao mesmo decidir qual a melhor configuração que atende as suas necessidades.

4.2 Resultados e Limitações

Como um primeiro exemplo de resultado, a superfície escolhida é uma falha localizada no piso de um estacionamento. As Figuras 3.4a e 3.4b exibem, respectivamente, as imagens difusa e com *flash* capturadas. A superfície é constituída por diversas rachaduras, além de outros detritos também presentes, dentre estes, um pequeno pedaço de plástico transparente (Figura 4.6). Por se tratar de uma superfície opaca e sem características reflexivas, não obtivemos, durante a captura da imagem com *flash*, regiões com especularidades, as quais causariam distorções no mapa final criado.

A Figura 4.7 apresenta o resultado da visualização do mapa desta superfície. Podemos observar que todos os elementos anteriormente descritos, e que constituem a superfície de interesse, apresentam uma altura/profundidade condizentes com o



Figura 4.6: Pequeno pedaço de plástico presente na superfície de interesse.

esperado, como por exemplo: as rachaduras, as depressões que separam o piso plano do estacionamento e a região do piso danificado, além das pedras soltas e dos demais detritos que o constituem. Entretanto, o pequeno pedaço de plástico transparente, anteriormente citado, por possuir tal característica, é representado como sendo parte integrante de uma das pequenas pedras que compõem o piso fotografado (Figura 4.8). Este resultado deve-se à violação da condição de termos um elemento/superfície supostamente opaca. Apesar disto, o resultado foi bastante satisfatório.



Figura 4.7: Visualização do mapa de profundidades do piso de um estacionamento.

De forma semelhante, a partir da Figura 4.9a e da Figura 4.9b , podemos observar o bom resultado exibido na Figura 4.10. A superfície faz parte de uma calçada e apresenta diversas características com variações de profundidade.

Neste exemplo, podemos observar a presença de um pequeno pedaço de papel colorido localizado sobre a superfície. Por se tratar de ua elemento não translúcido, podemos notar que o pequeno detrito está nitidamente localizado sobre a mesma



Figura 4.8: Resultado da presença do pequeno pedaço de plástico transparente na superfície.



(a) *Diffuse-lit*



(b) *Flash-lit*

Figura 4.9: Imagens difusa e com *flash* para a criação do mapa de profundidade da calçada.



Figura 4.10: Visualização do mapa de profundidades da calçada.

(Figura 4.11), não apresentando um resultado igual ao exemplo anterior com o pequeno pedaço de plástico.

Os resultados a seguir foram obtidos não respeitando algumas premissas básicas



Figura 4.11: Resultado da presença do pequeno pedaço de papel na superfície.

para a utilização da técnica de *Depth Hallucination*. Como um primeiro exemplo, escolhemos uma parede para obtenção do mapa de profundidades. Entretanto, ao invés da mesma estar sob uma iluminação indireta, conforme explicação na Seção 2.2.1, havia a presença da luz do sol incidindo lateralmente à superfície. As Figuras 4.12a e 4.12b exibem as imagens difusa e com *flash* obtidas.

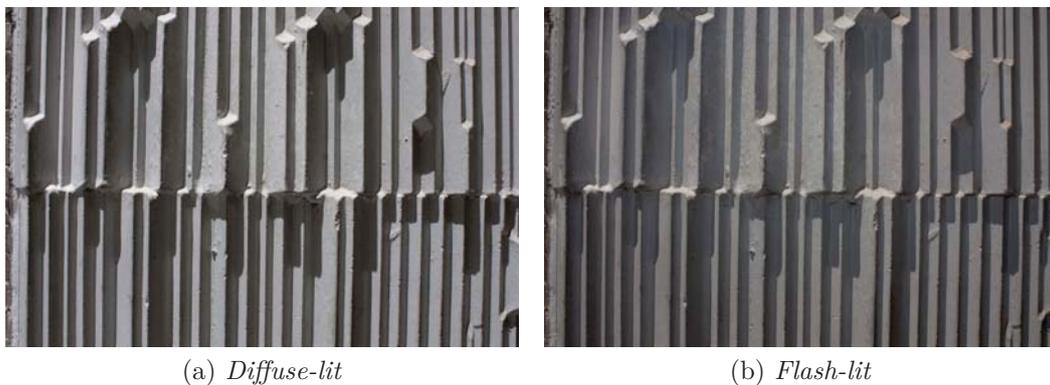


Figura 4.12: Imagens difusa e com *flash* com incidência direta de luz para a criação do mapa de profundidade da parede.

Observando atentamente a Figura 4.12b podemos perceber que, apesar de estarmos utilizando o *flash*, o mesmo não foi suficiente para compensar as sombras resultantes da incidência da luz do sol. Conforme explicação anterior (Seção 3.1.2), a potência do *flash* escolhida deve ser suficiente para revelar as regiões em penumbra, exibindo os detalhes existentes nas mesmas. Se a potência fosse aumentada com o intuito de compensarmos as sombras criadas pelo sol, obteríamos uma imagem bastante superexposta, o que inviabilizaria a criação do mapa de profundidades. Desta forma, a incidência da luz solar ocasiona distorções no *shading image* e, consequentemente, gera alterações no mapa de profundidades. Observando a Figura 4.13,

podemos perceber a presença de uma variação da altura exatamente nas regiões afetadas pelas sombras.



Figura 4.13: Visualização do mapa de profundidades da parede com incidência direta de luz.

As Figuras 4.14a e 4.14b exibem, respectivamente, as imagens difusa e com *flash* capturadas de um bueiro. Apesar de terem sido capturadas sob uma iluminação indireta, por se tratar de uma superfície constituída por um material altamente reflexivo, mais uma vez temos a violação da premissa de termos uma superfície não refletora. Conforme podemos observar na Figura 4.15, o resultado obtido não é satisfatório, dado que, devido às várias especularidades observadas, o mapa de profundidades apresenta diversas distorções exatamente nessas regiões específicas, não representando, de forma verossímil, a superfície de interesse.



Figura 4.14: Imagens difusa e com *flash* para a criação do mapa de profundidade do bueiro.

A partir dos resultados anteriormente apresentados, podemos concluir que, respeitadas as premissas básicas das características das superfícies de interesse, além



Figura 4.15: Visualização do mapa de profundidades do buraco.

da realização de uma correta captura das imagens difusa e com *flash*, podemos criar um mapa de profundidades condizente com as características existentes no terreno fotografado.

Capítulo 5

Conclusão

Esta dissertação teve como objetivo pesquisar, avaliar e, principalmente, testar todo o processo principal da técnica de *Depth Hallucination*: a captura das imagens difusa, com *flash* e a Lambertiana, o processamento das mesmas, a obtenção do respectivo mapa de profundidades da superfície de interesse e a visualização do mesmo.

Algumas melhorias e esclarecimentos foram acrescentados à técnica: a divisão do *shading image* por um fator 20 (vinte), antes do processo de normalização, evitando problemas na obtenção do mesmo; a utilização de um balanço de branco em *AWB*, diferentemente da recomendada (*daylight*), apresentando uma imagem albedo com as cores mais fiéis às da superfície de interesse; o ajuste da potência do *flash* com o objetivo principal de apenas iluminar as regiões que estavam mais escuras na imagem difusa (capturada anteriormente), diferentemente da utilização da máxima potência do *flash*, conforme recomendava o processo original do *Depth Hallucination*.

Além disso, podemos constatar que, apesar de toda a importância da teoria matemática, a captação correta das imagens, a partir de um conhecimento mínimo das técnicas de fotografia, destaca-se como parte de extrema importância para a realização de todo o processo. Por fim, respeitando-se as premissas básicas estipuladas pela técnica, conseguimos obter resultados extremamente satisfatórios e bem fiéis às características existentes nas superfícies de interesse fotografadas.

Visando aproximar a técnica do usuário final, foi desenvolvido um *software* contendo todo o algoritmo necessário para a implementação da mesma. O mesmo caracteriza-se por possuir uma simples interface gráfica de fácil entendimento e utilização, permitindo assim que os bons resultados provenientes da utilização do processo sejam utilizados no dia-a-dia de profissionais das áreas de efeitos visuais, arquitetura, *games*, entre outras.

Como propostas futuras para esse trabalho podemos destacar a análise, a avaliação e os testes na construção do mapa de profundidades da superfície de interesse a partir da análise do histograma de uma superfície com características semelhantes e

anteriormente processada (etapa esta já existente no processo de *Depth Hallucination*). Realizar uma análise mais profunda de toda a técnica desenvolvida, visando resultados ainda melhores que os já obtidos. A criação de novas funções no *software* desenvolvido que acelerem o processamento dos dados, uma vez que as imagens de entrada do sistema possuem altíssima resolução. Por fim, a adaptação do mesmo como um *plugin* à *softwares* profissionais, como 3ds Max e Maya, permitindo que o profissional gráfico execute todas as operações internamente nos *softwares* que já estão sendo utilizados.

Referências Bibliográficas

- [1] MASHHUDA GLENROSS, GREGORY J. WARD, C. J. J. L. F. M., HUBBOLD, R. “A Perceptually Validated Model for Surface Depth Hallucination”, *In ACM Transactions on Graphics (ACM SIGGRAPH)*, v. 27(3), pp. 59:1 – 59:8, 2008.
- [2] MARQUES ALMEIDA DA SILVA, R., F. B. *TerrainLib*. Virtual Book, 2009. Disponível em: <<http://www.marquess.com.br/terrainlib/>>. Acesso em: 14 mar 2012.
- [3] DEBEVEC, P. E., T. C. J., MALIK, J. “Modeling and rendering architecture from photographs: a hybrid geometry and image-based approach”, *In SIGGRAPH, ACM*, pp. 11–20, 1996.
- [4] RUSHMEIER, H., BERNARDINI, F. “Computing consistent normals and colors from photometric data”, *In Second Conference on 3-D Imaging and Modeling 3DIM, IEEE*, pp. 99–108, 1999.
- [5] YU, Y., D. P. M. J., HAWKINS, T. “Inverse global illumination: recovering reflectance models of real scenes from photographs”, *In SIGGRAPH, ACM*, pp. 215–224, 1999.
- [6] KHAN, E. A., R. E. F. R. W. E. A. “Image-based material editing”, *In SIGGRAPH, ACM*, pp. 654–663, 2006.
- [7] LANGER, M. S., BULTHOFF, H. H. “Depth discrimination from shading under diffuse lighting”, *Perception*, v. 29, 6, pp. 649–660, 2000.
- [8] PETROU, M., S. P. G. “Image Processing: Dealing with Texture”, *Wiley*, 2006.
- [9] HERSHBERGER, W. *Taming those high-lights: cross-polarization flash macro photography*. Virtual Book, 2008. Disponível em: <<http://www.naturescape.net/042004/wh0404.htm>>. Acesso em: 18 set 2011.

- [10] DUTRE, P. *Global Illumination Compendium*. Virtual Book, 2003. Disponível em: <<http://people.cs.kuleuven.be/~philip.dutre/GI/>>. Acesso em: 06 mar 2012.
- [11] WALREE, P. V. *Vignetting*. Virtual Book, 2011. Disponível em: <<http://toothwalker.org/optics/vignetting.html>>. Acesso em: 06 mar 2012.
- [12] BURT, P., ADELSON. “The Laplacian pyramid as a compact image code”, *IEEE Transactions on Communications*, v. 31, pp. 532 – 540, 1983.
- [13] LANGER, M. S., Z. S. W. “Shape-from-shading on a cloud day”, *Journal of the Optical Society of America*, v. 11, 2, pp. 467–478, 1994.
- [14] PETSCHNIGG, G., S. R. A. M. E. A. “Digital photography with flash and no-flash image pairs”, *In SIGGRAPH, ACM*, pp. 664–672, 2004.
- [15] EISEMANN, E., D. F. “Flash photography enhancement via intrinsic relighting”, *In SIGGRAPH, ACM*, pp. 673–678, 2004.