

Mapas de profundidade de superfícies texturizadas baseados em *Depth Hallucination*

José Almeida Jr, Mariane R. Petraglia e José Gabriel R. C. Gomes

Resumo—Este artigo 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.

Palavras-Chave—*Depth Hallucination*, Mapa de Profundidades, Albedo, *Shading Map*, CSharp.

Abstract—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.

Keywords—*Depth Hallucination*, Depth Map, Albedo, Shading Map, CSharp.

I. 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 processos são utilizados para a captura dessas características. Uma nova técnica, chamada *Depth Hallucination* [2], 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*.

II. *Depth Hallucination* E TEORIA FOTOGRÁFICA

A. Textura, Albedo e Mapa de Profundidades

Textura é a variação de informações em escalas menores que a escala de interesse [5]. É o aspecto de uma superfície ou seja, a "pele" de uma forma que permita identificá-la e distingui-la de outras formas.

José Almeida Jr, Mariane R. Petraglia e José Gabriel R. C. Gomes, Departamento de Engenharia Elétrica, COPPE-UFRJ, Rio de Janeiro-RJ, Brasil, E-mails: zealmeidajr@pads.ufrj.br, mariane@pads.ufrj.br, gabriel@pads.ufrj.br

O albedo fornece a fração de luz incidente que o material reflete para cada comprimento de onda [5].

O mapa de profundidades, 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 de cada *pixel* da imagem da superfície de interesse.

B. Depth Hallucination

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

- Captura das imagens;
- Estimação do albedo e *shading*;
- Estimação do mapa de profundidades;
- Reiluminação da superfície criada;

Na etapa de captura das imagens, faz-se necessária a utilização de uma câmera DSLR, pois, 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.

As superfícies capturadas devem ser planas, como por exemplo uma parede ou um piso. Todas as imagens devem ser capturadas no formato *RAW*, o qual não passou por nenhum processo de compressão com perda de dados.

A primeira imagem capturada deve estar sobre uma iluminação indireta (sombra ou dia nublado). Esta imagem é chamada de *diffuse-lit* (Figura 1). A segunda foto, chamada *flash-lit* (Figura 2), 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 capturada da imagem difusa.

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.

Por fim, captura-se a imagem de calibração, a qual é obtida de uma superfície Lambertiana branca (Figura 3) 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 [7] presentes nas imagens difusa e com *flash*.


 Fig. 1. Exemplo de *diffuse-lit*.

 Fig. 2. Exemplo de *flash-lit*.


Fig. 3. Exemplo de superfície Lambertiana branca.

C. Calculando o Albedo e o Shading Image

Após a aquisição das imagens, obtemos um fator de correção da exposição de cada imagem baseado na Equação (1) :

$$C_e = \frac{A^2}{T_s I}. \quad (1)$$

A imagem albedo I_a (Figura 4) é obtida a partir da Equação (2), sendo esta operação realizada em cada *pixel* j da imagem.

$$I_a(j) = \frac{I_f(j) - I_d(j)}{I_c(j)}. \quad (2)$$



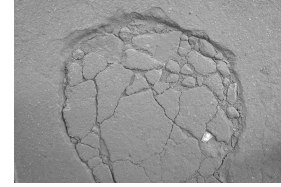
Fig. 4. Exemplo de albedo.

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* $I_c(j)$.

O *shading image* (Figura 5) é obtido através da razão entre a imagem difusa e do albedo anteriormente calculado. 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.

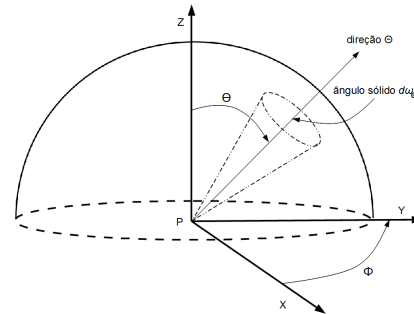
Entretanto, os resultados obtidos após as operações descritas resultam em um *shading image* extremamente claro, de tal forma que não conseguimos discernir os detalhes característicos da imagem. Dessa forma, baseado em testes empíricos, concluímos que antes da realização da normalização, o *shading image*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


 Fig. 5. Exemplo de *shading image*.

D. Estimação do Mapa de Profundidades

A técnica de *Depth Hallucination* baseia-se no desenvolvimento dos modelos de picos e os de depressões para a reconstrução da meso-estrutura.

Para o modelo de depressões, considere a porção visível do céu (E_c) 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 6:


 Fig. 6. Porção do céu visível através do ponto P e incidência de luz por um ângulo sólido.

Realizando a parametrização do ângulo sólido para (θ, ϕ) e calculando a integral sobre o hemisfério, obtemos:

$$E_c = \int_{\Omega} f(\Theta) dw_\Theta = \int_0^{2\pi} \int_0^\theta f(\theta, \phi) \sin \theta d\theta d\phi. \quad (3)$$

Uma vez que a iluminação sobre um ponto na superfície decai com o cosseno do ângulo de incidência, temos que $f(\theta, \phi) = \cos \theta$. Assim:

$$E_c = \int_0^{2\pi} \int_0^\theta \cos \theta \sin \theta d\theta d\phi. \quad (4)$$

O modelo para depressões aproxima as cavidades da superfície como cilindros de raio a e de altura d , como mostra a Figura 7. Dessa forma:

$$E_c = \int_0^{2\pi} \int_0^\theta \cos \theta \sin \theta d\theta d\phi = \pi \sin^2 \theta = \pi \frac{a^2}{a^2 + d^2}. \quad (5)$$

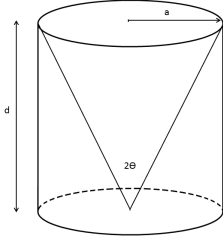


Fig. 7. Modelo cilindro para os pontos localizados abaixo do plano médio da superfície (baseado em [2]).

Para obtermos o fator *shading* S para o modelo de depressões, é necessário dividir E_c pelo fator de iluminação relativo ao céu, E_h , o qual é igual a π conforme a Equação (6):

$$E_h = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \cos \theta \sin \theta d\theta d\phi = \pi. \quad (6)$$

Dessa forma:

$$S = \frac{E_c}{E_h} = \frac{\pi \sin^2 \theta}{\pi} = \frac{a^2}{a^2 + d^2}. \quad (7)$$

Assim, resolvendo a Equação (7) para d , obtemos a profundidade do ponto localizado na cavidade da superfície:

$$d = a \sqrt{\frac{1}{S} - 1}. \quad (8)$$

Já para modelo de picos, o fator *shading* S é uma função da porção visível do hemisfério h_v sob o ângulo Φ (Equação (9)), somada à porção refletida h_r do mesmo hemisfério localizada fora do ângulo Φ (Equação (10)).

$$h_v = \frac{\pi}{2}(1 + \cos \Phi), \quad (9)$$

$$h_r = \nu \frac{\pi}{2}(1 - \cos \Phi). \quad (10)$$

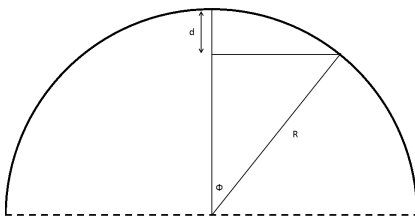


Fig. 8. Modelo hemisfério para os pontos localizados acima do plano médio da superfície (baseado em [2]).

Da mesma forma que na Equação (7), dividimos a soma entre h_v e h_r pelo fator de iluminação relativo ao céu, π . Assim:

$$S = \frac{\frac{\pi}{2}(1 + \cos \Phi) + \nu \frac{\pi}{2}(1 - \cos \Phi)}{\pi}, \quad (11)$$

onde ν é o fator de reflexão efetivo da superfície em torno do ponto de interesse.

Observando a Figura 8, temos que:

$$d = R - R \cos \Phi. \quad (12)$$

Resolvendo a Equação (11) para $\cos \Phi$ e substituindo na Equação (12) obtemos:

$$d = 2R \frac{1 - S}{1 - \nu}, \quad (13)$$

onde R é o raio do pico do hemisfério.

Dado que o *Shading Image* foi normalizado em 0.5, fazemos $S = \frac{1}{2}$. Além disso, a técnica de *Depth Hallucination* assume o fator de reflexão efetivo como zero ($\nu = 0$). Assim, combinando as Equações (8) e (13), obtemos:

$$a = R. \quad (14)$$

Dessa forma, para cada ponto (*pixel*) existente no *Shading Image* com um valor S , chegamos ao seguinte modelo:

$$D(S) = \frac{d}{a} = \sqrt{\frac{1}{S} - 1} \quad \text{se } S \leq \frac{1}{2}, \quad (15)$$

$$D(S) = \frac{d}{a} = 2(1 - S) \quad \text{se } S > \frac{1}{2}. \quad (16)$$

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. 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.

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 [6].

Utilizando as Equações (15) e (16), 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 (17):

$$d_j = \sum_{i=1}^N r(i) [D(l_j(i)) - 1]. \quad (17)$$

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

É permitido ao usuário aplicar um fator de escala uniforme, entre 0.75 e 1.5, no mapa de profundidades compensando algumas simplificações que o processo de *Depth Hallucination* utiliza, como por exemplo, ignorar o cálculo do albedo nas regiões acentuadamente inclinadas da superfície de interesse, bem como a aproximação das reentrâncias da superfícies como depressões (covas) e não como fendas (mais apropriado).

III. PROCESSO DE GERAÇÃO DE MAPAS DE PROFUNDIDADES

A. 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. Após o posicionamento da câmera, deve-se estabelecer o foco sobre a superfície.

O balanço do branco deve ser ajustado para *AWB*, diferentemente das recomendações para a utilização em *daylight* [2]. 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.

O ajuste de ISO deve ser escolhido 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, dependendo da distância a que a câmera estiver da superfície, o valor da abertura do diafragma não pode ser muito pequeno (maior diâmetro), 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, estabelecemos a velocidade do obturador de tal forma que tenhamos uma imagem balanceada. Recomenda-se a utilização da função histograma, presente na grande maioria das câmeras digitais, para a constatação de que os ajustes anteriores nos fornecerão uma imagem balanceada. Por fim, 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.

Apesar do processo de *Depth Hallucination* informar que a potência do *flash* deve ser a maior possível, após a realização de diversos testes, 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 specularidades geradas pelo *flash*, as quais resultarão em distorções no mapa de profundidades.

Para a captura da imagem *flash-lit*, os únicos parâmetros que podemos alterar são a velocidade do obturador e o ISO. Os demais 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.

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 necessária para o cálculo do albedo.

B. Software Desenvolvido

Visando facilitar a utilização do processo de *Depth Hallucination* pelo usuário final, foi desenvolvido um *software* com uma interface simples e amigável utilizando-se a linguagem C#.

A Figura 9 exhibe todas as funções criadas e a sequência em que cada uma delas será executada. Trata-se de uma sequência de funções em cascata, as quais são relativas às etapas do *Depth Hallucination*.

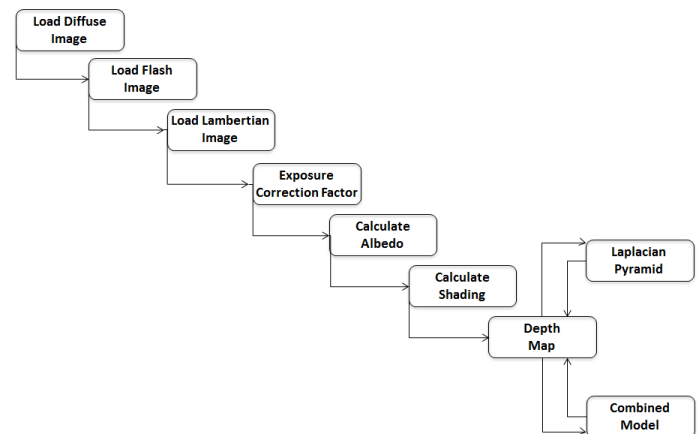


Fig. 9. Representação das funções desenvolvidas para o processo de *Depth Hallucination*.

IV. VISUALIZAÇÃO DO MAPA DE PROFUNDIDADES

Após a correta captura das imagens e o processamento das mesmas, obtemos como resultado final uma matriz bidimensional com o mesmo número de linhas e colunas das imagens utilizadas. Essa matriz contém o mapa de profundidades de cada *pixel* existente na imagem.

A. Sigma e Número de Níveis do Gaussian Blur

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).

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.

Para um número de níveis de *blur* mais elevado, os resultados obtidos são tão satisfatórios quanto os obtidos para 6 níveis. Porém, quanto maior esse número, 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.

B. Resultados e Limitações

Como um primeiro exemplo, o piso escolhido é constituído por diversas rachaduras, além de outros detritos também presentes, dentre estes, um pequeno pedaço de plástico transparente.

A Figura 10 apresenta o resultado da visualização do mapa de desta superfície. Podemos observar que todos os elementos anteriormente descritos, e que constituem a superfície de interesse, apresentam uma altura/profundidade condizentes com o esperado. Entretanto, o pequeno pedaço de plástico transparente é representado como sendo parte integrante de uma das pequenas pedras que compõem o piso fotografado (Figura 11). Este resultado se deve à violação da condição de termos um elemento/superfície supostamente opaca. Apesar disto, o resultado foi bastante satisfatório.



Fig. 10. Visualização do mapa de profundidades do piso de um estacionamento.



Fig. 11. Resultado da presença do pequeno pedaço de plástico transparente na superfície.

De forma semelhante, podemos observar o bom resultado exibido na Figura 12.

Neste exemplo, podemos observar a presença de um pequeno pedaço de papel colorido localizado sobre a superfície. Por se tratar de um elemento não-translúcido, podemos notar que o pequeno detrito está nitidamente localizado sobre a mesma (Figura 13), não apresentando um resultado igual ao exemplo anterior com o pequeno pedaço de plástico.



Fig. 12. Visualização do mapa de profundidades da calçada.



Fig. 13. Resultado da presença do pequeno pedaço de papel na superfície.

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

V. CONCLUSÕES

Algumas melhorias e esclarecimentos foram acrescentados à técnica: a divisão do *shading image* por um fator 20 (vinte), antes do processo de normalização; a utilização de um balanço de branco em *AWB*, diferentemente da recomendada (*day-light*); 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.

REFERÊNCIAS

- [1] J. Almeida Junior, Mapas de profundidade de superfícies texturizadas baseados em *Depth Hallucination*, COPPE-UFRJ, 2012.
- [2] M. Glenross, G. J. Ward, C. Jay, J. Liu, F. Melendez e R. Hubbard, *A Perceptually Validated Model for Surface Depth Hallucination*. In ACM Transactions on Graphics (ACM SIGGRAPH), v.27(3), pp 59:1 - 59:8, 2008.
- [3] P. E. Debevec, C. J. Taylor e J. Malik, *Modeling and rendering architecture from photographs: a hybrid geometry and image-based approach*, In SIGGRAPH, ACM, pp. 11–20, 1996.
- [4] Y. Yu, P. Debevec, J. Malik e T. Hawkins, *Inverse global illumination: recovering reflectance models of real scenes from photographs*, In SIGGRAPH, ACM, pp. 215–224, 1999.
- [5] M. PETROU, P. G. SEVILLA, *Image Processing: Dealing with Texture*, Wiley, 2006.
- [6] P. Burt e Adelson, *The Laplacian pyramid as a compact image code*, *IEEE Transactions on Communications*, v. 31, pp. 532–540, 1983.
- [7] P. Dutre, *Global Illumination Compendium*, Virtual Book, 2003. Disponível em: <http://people.cs.kuleuven.be/~philip.dutre/GII/>, 2003.